

Astrometría I: Probabilidad y Estadística

Parte I

28 de abril de 2011

Índice

1. Probabilidad: Nociones Básicas	5
1.1. Fenómenos y modelos	5
1.1.1. Modelos determinísticos	5
1.1.2. Modelos aleatorios	6
1.2. Conceptos útiles	6
1.2.1. Experimento aleatorio	6
1.2.2. Espacio muestral	6
1.2.3. Evento o Suceso	7
1.3. La Probabilidad	8
1.3.1. Axiomas	10
1.3.2. Reglas para el cálculo	11
1.3.3. Cálculo	12
1.4. Probabilidad Condicional	13
1.5. Eventos Independientes	14
1.6. Teorema de Bayes	15
2. Variables Aleatorias	19
2.1. Definición	19
2.2. Discretas y Continuas	21
2.3. Función de Probabilidad	21
2.3.1. Función de probabilidad de una variable aleatoria discreta	22
2.3.2. Función de probabilidad de una variable aleatoria continua	25
2.4. Función de Distribución Acumulada	26
2.4.1. Función acumulada para variables discretas	26
2.4.2. Función acumulada para variables continuas	28
3. Distribuciones de Probabilidad	30
3.1. Modelos probabilísticos discretos	30
3.1.1. Modelo de Bernoulli	30
3.1.2. Modelo Binomial	31
3.1.3. Modelo de Poisson	34
3.1.4. Otros modelos discretos	37
3.2. Modelos probabilísticos continuos	39
3.2.1. Modelo Normal	39
3.2.2. Modelo Exponencial	41
3.2.3. Otros modelos continuos	43
3.3. Generadores de números (pseudo) aleatorios	44
3.3.1. Números aleatorios uniformes	44
3.3.2. Variables aleatorias discretas	46
3.3.3. Método de Inversión	46
3.3.4. Método de Rechazo	47

3.3.5.	Método de Box-Müller	47
3.4.	Caracterización completa de las distribuciones de probabilidades . . .	48
3.4.1.	Momentos de una distribución	48
3.4.2.	Función generatriz de momentos	49
3.4.3.	Cumulantes de una distribución	50
4.	Inferencia Estadística	51
4.1.	Conceptos importantes	51
4.1.1.	Universos, población y muestra	51
4.1.2.	Parámetros y estadísticos	53
4.2.	Muestra y Muestreo	54
4.2.1.	Muestra representativa	54
4.2.2.	Muestreo aleatorio	54
4.3.	Distribuciones Muestrales	55
4.3.1.	Distribución de la media muestral	56
4.3.2.	Distribución de la diferencia de medias muestrales	59
4.4.	Métodos Inferenciales	61
5.	Inf. Est.: Estimación (I)	62
5.1.	Estimación puntual	62
5.1.1.	Estimador insesgado	62
5.1.2.	Estimador consistente	63
5.1.3.	Estimador eficiente	64
5.2.	Intervalos de confianza (IC)	64
5.2.1.	IC para una media poblacional	64
5.2.2.	IC para la diferencia de dos medias poblacionales	70
6.	Inf. Est.: Estimación (II)	75
6.1.	Histogramas	75
6.1.1.	Definición	75
6.1.2.	Intervalo de confianza	75
6.1.3.	Histogramas para variables continuas	77
6.1.4.	Funciones "kernel" para histogramas de variables continuas	77
6.2.	Técnicas de Remuestreo	79
6.2.1.	Método Bootstrap	79
6.2.2.	Método Jackknife	87
7.	Inf. Est.: Prueba de Hipótesis (I)	88
7.1.	PH: un procedimiento de decisión	88
7.2.	Procedimiento general para la PH	90
7.2.1.	Hipótesis	90
7.2.2.	Nivel de significación	92
7.2.3.	Estadístico de prueba	94
7.2.4.	Zona de aceptación	95

7.2.5.	Cómputos	96
7.2.6.	Decisión	96
7.2.7.	Conclusión	96
7.3.	PH para una media poblacional	97
7.3.1.	PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente y con varianza conocida	97
7.3.2.	PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente con varianza desconocida y tamaño de muestra grande ($n \geq 30$)	98
7.3.3.	PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente con varianza desconocida y tamaño de muestra pequeño ($n < 30$)	99
7.3.4.	PH para una media pobl. cuando la muestra proviene de una población con distribución no normal y tamaño de muestra grande ($n \geq 30$)	100
7.4.	PH para dos medias poblacionales	101
7.4.1.	PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente y con varianza conocidas	101
7.4.2.	PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente, con varianza desconocidas y tamaño de muestras grandes ($n_1, n_2 \geq 30$)	102
7.4.3.	PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente, con varianza desconocidas y tamaño de muestras pequeñas ($n_1, n_2 < 30$)	103
7.4.4.	PH para dos medias pobl. cuando las muestras provienen de poblaciones con distribución no normal y tamaño de muestras grandes ($n_1, n_2 \geq 30$)	105
7.5.	PH para dos varianzas poblacionales	106
8.	Inf. Est.: Prueba de Hipótesis (II)	109
8.1.	Método Chi-cuadrado	109
8.2.	Método de Kolmogorov-Smirnov	112
8.3.	Independencia estadística	115
8.3.1.	El método χ^2 ... el regreso	116
8.3.2.	Coefficiente de correlación lineal de Pearson	119
8.3.3.	Función de correlación	120
9.	Estimadores Generales	121
9.1.	Máxima Probabilidad	121
9.2.	Ajuste de datos	123
9.2.1.	Cuadrados mínimos como estimador de máxima probabilidad	123
9.2.2.	Ajuste por chi-cuadrado	125
9.2.3.	Ajustando datos con una recta usando chi-cuadrado	127

1. Probabilidad: Nociones Básicas

Para emprender el estudio de la estadística y su alcance a la hora de analizar un conjunto de datos es necesario tener a mano las nociones básicas de probabilidad. La probabilidad y la estadística son dos disciplinas íntimamente conectadas. Inicialmente el único punto de unión que se puede establecer es que ambas disciplinas tienen en común el estudio de los fenómenos aleatorios. La teoría de probabilidades tiene como problema general describir mediante un modelo matemático cada tipo de fenómeno aleatorio, mientras que la inferencia estadística tiene planteado el problema inverso, es decir, a partir del conocimiento de una parte del fenómeno pretende establecer sus propiedades, para lo cual forzosamente debe utilizar algún modelo probabilístico que describa el fenómeno. Es esta dependencia de la estadística con la teoría de probabilidad lo que justifica profundizar el estudio de esta última.

1.1. Fenómenos y modelos

Un fenómeno natural es toda manifestación natural que puede ser percibida mediante los sentidos o instrumentos. Los fenómenos naturales se pueden clasificar en *determinísticos* y *aleatorios*. Los determinísticos se pueden definir como toda manifestación natural que observada repetidamente bajo las mismas condiciones, produce siempre resultados idénticos. Por ejemplo, el tiempo que tarda un objeto en llegar al suelo invariablemente será el mismo, si las condiciones son iguales en cada repetición de la experiencia. Los aleatorios, en cambio, son todo proceso que al observarlo repetidamente bajo el mismo conjunto de condiciones, producen resultados diferentes. Tirar un dado es un ejemplo de este fenómeno, ya que aunque se conozcan todos los resultados posibles, no se puede predecir con completa certeza uno en particular.

Una manera de estudiar estos fenómenos es mediante la construcción de modelos matemáticos, los cuales intentan (simplificando u omitiendo algunos detalles) representar, mediante expresiones cuantitativas, las características, propiedades y/o funcionamiento de los procesos naturales. De acuerdo con los fenómenos ya mencionados los modelos existentes pueden ser determinísticos o aleatorios.

1.1.1. Modelos determinísticos

Estos modelos establecen que las condiciones en las cuales se realiza un experimento determinan la ocurrencia de un resultado particular. Por ej., si observamos el desplazamiento de un móvil cierta distancia (d), podemos utilizar como modelo matemático para describir la velocidad media desarrollada (v) la ecuación $v = d/t$ (con t el tiempo transcurrido). Éste es un modelo determinístico, porque cada vez que se repita la experiencia y se obtengan los mismos valores d y t , se producirá el mismo valor de v . Obviamente, este es un modelo simplificado en el que muchos factores no han sido tenidos en cuenta (temperatura del aire, presión atmosférica, etc.), sin embargo, las pequeñas desviaciones que se podrían llegar a obtener no invalidan el modelo.

1.1.2. Modelos aleatorios

En estos modelos las condiciones de un experimento no determinan un resultado particular, sino su probabilidad de ocurrencia dentro de un conjunto de resultados posibles. Es decir, que estos modelos son fórmulas que permiten obtener la distribución de probabilidad de los resultados posibles del experimento. Por ej., cuántas veces saldrá el número 6 al lanzar un dado 5 veces? En éste caso se debe utilizar un modelo probabilístico, que permite conocer cual es la probabilidad de obtener cualquiera de los resultados posibles. El modelo que buscamos es el siguiente: $p(x) = C_x^n p^x q^{n-x}$ con x el num. de veces o ensayos donde ocurre el resultado esperado, n el num. total de ensayos, p la probabilidad de éxito, $q = (1 - p)$ la probabilidad de fracaso y $C_x^n = n!/x!(n-x)!$. Por ej., la probabilidad de obtener 3 veces el num. 6 en 5 lanzamientos de un dado es $p(3) = C_3^5 (1/6)^3 (5/6)^2 = 0,0312$.

1.2. Conceptos útiles

A continuación detallaremos ciertos conceptos y nomenclaturas que nos serán útiles cada vez que enfrentemos un problema probabilístico.

1.2.1. Experimento aleatorio

Un experimento, desde el punto de vista estadístico, está constituido por uno o más ensayos, término que identifica cualquier acto repetible que produce un resultado único cada vez que se ejecuta. Cualquier experimento que puede tener más de un resultado se califica como aleatorio y es posible encontrar un modelo que permita determinar la probabilidad de ocurrencia de cada resultado. Las características comunes de experimento aleatorio son:

- Pueden repetirse indefinidamente manteniendo las condiciones en las que se realiza
- Previo a cualquier ensayo no es posible predecir un resultado particular
- Previo al experimento es posible predecir el conjunto de posibles resultados
- La frecuencia de aparición de los diferentes resultados tiende a regularizarse al aumentar el número de repeticiones.

Ejemplos de experimentos aleatorios: lanzar una o más monedas, tirar un dado, determinar el número de individuos en varias unidades de muestreo, etc.

1.2.2. Espacio muestral

Asociado a cualquier experimento aleatorio (E) existe un espacio muestral (S) que se define como el conjunto de todos los posibles resultados de E.

Ejemplo: Si el experimento fuese determinar el número de hijas mujeres en familias con 4 hijos, se puede identificar el resultado de cada ensayo con las letras V=varón y M=mujer. El espacio muestral estaría integrado por todas las posibles formas de ocurrencia del

experimento:

$$S = \left\{ \begin{array}{l} VVVV \\ VVVM, \mathbf{VVMV}, VMVV, MVVV \\ \mathbf{VVMM}; VMVM, \mathbf{VMMV}, MVMV, MMVV, MVVM \\ VMMM, \mathbf{MVMM}, MMVM, \mathbf{MMM}V \\ MMMM \end{array} \right\}$$

Si las posibles ocurrencias son numerosas se pueden representar los resultados con un número. En nuestro ejemplo, si cada resultado es el número de mujeres entonces tendremos que $VVVV$ le corresponde el 0, a la segunda línea le corresponderá el 1 y así sucesivamente de modo que el espacio muestral se puede representar como

$$S = \{0, 1, 2, 3, 4\}$$

Cuando se describe el espacio muestral de esta manera se dice que se lo ha hecho por extensión o descripción. Otra manera de hacerlo es por comprensión como

$$S = \{x \in \mathbb{N} / 0 \leq x \leq 4\}$$

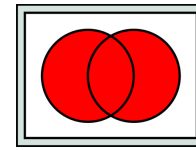
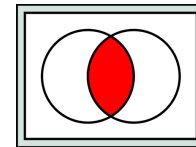
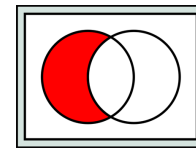
De acuerdo a la naturaleza de la variable que se esté utilizando los espacios muestrales pueden ser *discretos* o *continuos*. Es discreto si está formado por elementos numerables, es decir que son consecuencia de contar los resultados individuales de un experimento. A su vez el número de elementos contables puede ser finito o infinito. Ejemplo de espacio discreto y finito es el que usamos con anterioridad, es decir, el número de mujeres en familias de 4 hijos, mientras que si el experimento es el número de veces que hay que lanzar una moneda hasta obtener cara por primera vez, entonces se genera un espacio discreto e infinito. Por otro lado, el espacio muestral es continuo si esta formado por elementos no numerables. Entonces, por naturaleza, todo espacio continuo es infinito. Ejemplos de este tipo de espacio resultan con las variables de un proceso de medición (tiempo, altura, peso, densidad, temperatura, etc.)

1.2.3. Evento o Suceso

Cualquier conjunto de resultados dentro de un espacio muestral se denomina *evento* o *suceso*. En la terminología de conjuntos se puede decir que un evento (A) es un subconjunto del espacio muestral (S). El evento integrado por todos los resultados es igual al espacio muestral. A continuación especificamos terminología:

- *Evento elemental*: es cada resultado que conforma un espacio muestral
- *Evento complemento*: Dado un evento A en el espacio muestral S , el evento complemento de A (\bar{A}), está constituido por todos los elementos que pertenecen a S y que no están en A .
- *Evento vacío*: es el evento que no tiene elementos y que por lo tanto no puede ocurrir (\emptyset).

Con los eventos de un mismo espacio muestral se pueden realizar operaciones que resultan en la formación de nuevos eventos, los cuales siguen siendo subconjuntos del espacio muestral. Existen dos operaciones básicas: la unión y la intersección de eventos, que en cierto modo son paralelas a las operaciones de suma y multiplicación respectivamente. La *unión* de dos eventos A y B , se representa $A \cup B$, y da como resultado otro evento, el cual está formado por todos los elementos que pertenecen al evento A , al evento B o a ambos a la vez (fig. (a)). Cuando la unión de dos eventos equivale a todo el espacio muestral, se dice que los dos eventos son mutuamente exhaustivos. La *intersección* de dos eventos A y B se representa $A \cap B$, y da como resultado otro evento, el cual está formado por los elementos que pertenecen a ambos eventos a la vez (fig. (b)). Cuando la intersección de dos eventos es vacía, se dice que los dos eventos son mutuamente excluyentes. Por último, los elementos de un evento A que no se encuentran en el evento B , forman otro evento llamado *diferencia* de A y B , representado por $A - B$ (fig. (c)).

(a) $A \cup B$ (b) $A \cap B$ (c) $A - B$

1.3. La Probabilidad

La teoría del azar consiste en reducir todos los acontecimientos del mismo tipo a un cierto número de casos igualmente posibles, es decir, tales que estemos igual de indecisos respecto a su existencia, y en determinar el número de casos favorables al acontecimiento cuya probabilidad se busca. La proporción entre este número y el de todos los casos posibles es la medida de esta probabilidad, que no es, pues, más que una fracción cuyo numerador es el número de casos favorables y cuyo denominador el de todos los posibles.

Pierré Simon Laplace (1749-1827)

Ha llegado el momento de establecer que entendemos como probabilidad. La noción de probabilidad es algo con lo que convivimos diariamente haciendo conjeturas acerca de que esperamos que pase y consecuentemente, tomando decisiones. Por lo que nuestra primera definición de probabilidad sería cualquier probabilidad establecida es una afirmación que indica cuán posible se cree que es que un evento ocurra". Pero, más allá de establecer una definición intuitiva necesitamos convertir la intuición al lenguaje matemático. Por lo tanto empezaremos reescribiendo la definición y diremos que "la probabilidad es un valor numérico que cuantifica la posibilidad o factibilidad de ocurrencia de un resultado determinado dentro de un conjunto de resultados posibles". A un resultado imposible de ocurrir se le asigna una probabilidad de 0, si por el contrario es segura su ocurrencia, se le asigna una probabilidad de 1. A las probabilidades intermedias se les asocian valores entre 0 y 1.

Hay dos enfoques diferentes sobre cómo asignar la probabilidad a un evento: la asignación *objetiva* o *frecuentista* y la asignación *subjetiva* o *bayesiana*.

- *Asignación Objetiva*: Se basa en el conocimiento fáctico del espacio muestral y de la frecuencia relativa de ocurrencia de sus eventos elementales. El conocimiento de estas dos características puede ser de dos maneras:

- *Probabilidad a priori*: Este enfoque supone que la probabilidad de ocurrencia de un resultado particular se conoce antes de producirse el mismo. Para esto es necesario asumir que todos los resultados elementales son igualmente probables y excluyentes.

Si el espacio muestral S tiene n elementos e_i equiprobables, es decir con probabilidad $1/n$, y además se define un suceso A formado por r eventos elementos, la probabilidad de ocurrencia de A será :

$$P(A) = \sum_{i=1}^n P(e_i) = \sum_{i=1}^n 1/n = r/n$$

es decir, en esta concepción (usualmente llamada clásica), la probabilidad de un evento es igual al número de resultados en que el evento ocurre dividido por el número de resultados posibles. A modo de ayuda, también puede ser útil pensar la probabilidad de un conjunto como el tamaño relativo del mismo con respecto al evento seguro.

Ejemplo: Si de un mazo de cartas de poker se extrae aleatoriamente una carta, se quiere saber la probabilidad con la cual pueden ocurrir los siguientes eventos: a) sale un As, b) sale una espada negra o c) sale una J o una Q. El espacio muestral está formado por 52 eventos elementales equiprobables:

$$S = \left\{ \begin{array}{l} 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (\diamond \text{ rojo}) \\ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (\heartsuit \text{ rojo}) \\ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (\clubsuit \text{ negro}) \\ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (\spadesuit \text{ negro}) \end{array} \right\}$$

Entonces: a) $A = \{ 1 \diamond, 1 \heartsuit, 1 \clubsuit, 1 \spadesuit \} \longrightarrow P(A) = 4/52 = 0.077$

b) $B = \{ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K (\spadesuit \text{ negra}) \} \longrightarrow P(B) = 13/52 = 0.25$

c) $C = \{ J \diamond, J \heartsuit, J \clubsuit, J \spadesuit, Q \diamond, Q \heartsuit, Q \clubsuit, Q \spadesuit \} \longrightarrow P(C) = 8/52 = 0.154$

- *Probabilidad a posteriori*: Cuando no se tiene un experimento con un número finito de resultados equiprobables el concepto anterior no sirve, por lo que se requiere una definición más general de probabilidad. La concepción de probabilidad a posteriori surgió de la comprobación empírica. Es una observación común que en los experimentos aleatorios repetidos muchas veces la frecuencia relativa con la cual se produce un resultado se estabiliza alrededor de un cierto valor. Por lo tanto, si un experimento aleatorio se repite indefinidamente, la frecuencia relativa (fr) con la cuales aparecen los resultados se pueden

hacer equivalentes a su probabilidad de ocurrencia, ya que

$$\lim_{n \rightarrow \infty} fr(A) = P(A)$$

Esta forma de proceder permite acercarnos al verdadero valor de la probabilidad de un evento, pero obviamente, en términos prácticos, este valor es imposible de obtener. Aún así, se puede asumir que es una buena aproximación, que mejorará mientras más repeticiones del experimento existan.

Ejemplo: Se quiere conocer la probabilidad de obtener cara al lanzar una moneda cargada. El espacio muestral está dado por las dos posibilidades *cara* o *seca*, $S = \{c, s\}$, pero no es equiprobable. Para poder conocer la probabilidad de ocurrencia de los eventos, es necesario lanzar la moneda una gran cantidad de veces, anotar el resultado y calcular la frecuencia relativa. Si se lanzó 200 veces la moneda, de las cuales el evento cara ocurrió 75 veces, entonces $fr(c) = 75/200 = 0,375$ y $fr(s) = 125/200 = 0,625$. Por lo tanto, estas frecuencias se asignan como probabilidades de ocurrencia de los eventos considerados.

- *Asignación Subjetiva:* Muchos fenómenos puede que nunca hayan ocurrido o que se hayan producido muy pocas veces. Por ejemplo, una carrera de caballos es un hecho único, que nunca puede repetirse bajo las mismas condiciones o el descubrimiento de una droga nueva para curar una enfermedad. En estos casos, la asignación de la probabilidad no puede estar basada ni en el conocimiento previo del espacio muestral, ni en la frecuencia de ocurrencia de los hechos, de modo que el enfoque objetivo es obsoleto. Por lo tanto, aquí es cuando entra en acción el método de asignación subjetiva. De acuerdo a esta visión, el valor de probabilidad es asignado por una persona de acuerdo al grado de confianza que ella tenga en la ocurrencia del hecho. Bajo este punto de vista, diferentes individuos disponiendo de la misma información pueden tener distintos grados de confianza acerca de la ocurrencia de un evento (un ejemplo de esto son las apuestas deportivas). Aún cuando parezca que este método de asignación esta fuera del ámbito científico, no hay otra cosa más alejada de la realidad, ya que actualmente el enfoque subjetivo tiene gran utilidad en la Teoría Bayesiana de la desición, área de la estadística en pleno desarrollo.

1.3.1. Axiomas

Los axiomas o postulados que debe cumplir la probabilidad son los siguientes:

- *De positividad:* la probabilidad de un evento nunca es un número negativo, es cero (evento imposible de ocurrir) o un real positivo. Este axioma puede denotarse como: $P(A) \geq 0$.
- *De certidumbre:* la probabilidad de todo el espacio muestral es uno, $P(S) = 1$, es decir, la probabilidad de todo evento con un certidumbre total de ocurrencia es uno. Estos dos axiomas en conjunto establecen que $0 \leq P(A) \leq 1$.

- *De la adición:* la probabilidad de un evento A es igual a la suma de las probabilidades de los eventos elementales que lo conforman.

Ejemplo: En familias de 4 hijos, cuál es la probabilidad de encontrar una que tenga menos de 3 hijos varones ? Del espacio muestral que ya habíamos especificado en la página 7, sabemos que posee 16 elementos equiprobables y que el evento que buscamos posee 11 elementos:

$$A = \left\{ \begin{array}{l} \mathbf{VVMM}; VMVM, \mathbf{VMMV}, MVMV, \mathbf{MMVV}, MVVM \\ VMMM, \mathbf{MVMM}, MMVM, \mathbf{MMM}V \\ MMMM \end{array} \right\}$$

por lo que la probabilidad del evento A será igual a la suma de las probabilidades de los 11 elementos, $P(A) = 11/16 = 0,6875$

1.3.2. Reglas para el cálculo

A partir de los axiomas anteriores se pueden deducir algunas reglas básicas para calcular las probabilidades de diferentes tipos de eventos:

- *Del conjunto vacío:* Si \emptyset es el conjunto vacío, entonces $P(\emptyset) = 0$, es decir representa un evento que no puede ocurrir.
- *De adición para eventos mutuamente excluyentes:* Si A y B son dos eventos mutuamente excluyentes, la probabilidad de ocurrencia de A o de B es la suma de sus probabilidades separadas, es decir, $P(A \cup B) = P(A) + P(B)$.
- *De adición para eventos solapados:* Si A y B son dos eventos cualesquiera que pueden ocurrir juntos, significa que algunos de los eventos elementales que los conforman pertenecen tanto a A como a B , es decir forman parte de la intersección de los dos eventos. Por el 3er axioma sabemos que la probabilidad de ocurrencia de la unión de dos eventos es la suma de las probabilidades de los eventos elementales que los forman. Ahora, si solo se suman las probabilidades de los eventos A y B para el cálculo de la probabilidad de la unión, estaremos contando dos veces las probabilidades de los eventos elementales que pertenecen a la intersección, por lo tanto es necesario sustraer sus probabilidades una vez, es decir,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- *De la complementación:* Sean A y \bar{A} dos eventos complementarios en un espacio muestral S . Ya que los eventos complementarios son mutuamente excluyentes, se deduce de los axiomas 2do y 3ro que la probabilidad de la unión de A con \bar{A} es

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = P(S) = 1$$

por lo tanto, $P(A) = 1 - P(\bar{A})$.

1.3.3. Cálculo

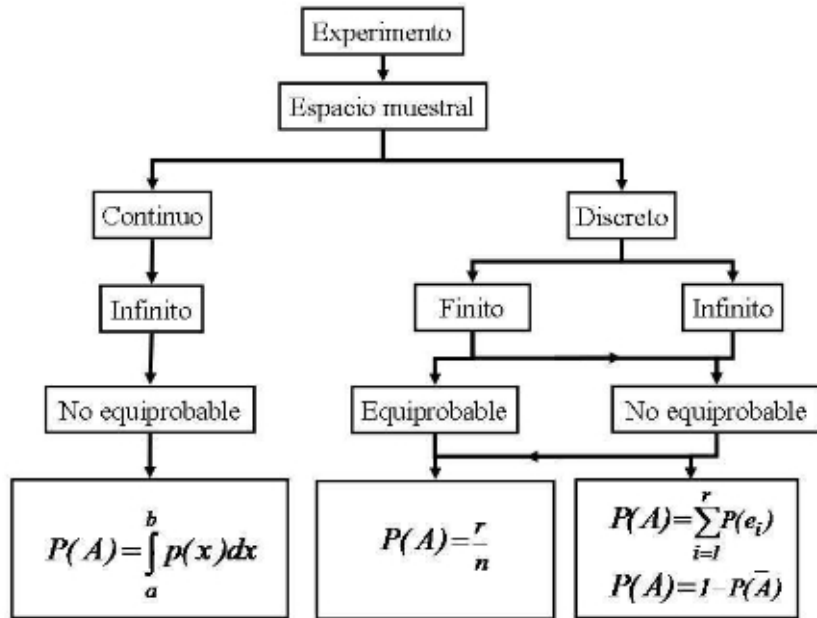
A continuación se detalla un procedimiento general que puede facilitar el cálculo de la probabilidad.

Paso 1: En primer término se debe definir correctamente el espacio muestral. En la figura de la derecha se muestra un esquema de los distintos tipos de espacios muestrales que puede generar un experimento aleatorio.

Paso 2: Se asigna un valor de probabilidad a cada evento elemental de modo que cumpla que $\sum_S p(e_i) = 1,0$.

Paso 3: Se define el o los eventos de interés en función de los eventos elementales que los componen.

Paso 4: Se calcula la probabilidad del evento o los eventos de nuestro interés de acuerdo a las formulaciones dadas en la figura.



Ejemplo 1: Cuál es la probabilidad de obtener dos números pares cuando se lanzan dos dados?

Paso 1: Se tiene un espacio muestral discreto, finito y con 36 resultados equiprobables:

$$S = \left\{ \begin{array}{l} (1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6) \\ (2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6) \\ (3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6) \\ (4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6) \\ (5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6) \\ (6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6) \end{array} \right\}$$

Paso 2: Cada evento elemental tiene la misma probabilidad de ocurrencia, $P(e_i) = 1/36$ de modo que se cumpla que $\sum_S p(e_i) = 1,0$.

Paso 3: El evento definido es: $A = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}$

Paso 4: La probabilidad de A es el número de veces que ocurre A sobre el total de posibles resultados: $P(A) = 9/36 = 1/4$.

Ejemplo 2: En el transcurso de una investigación efectuada para evaluar el efecto de una droga sobre cierta enfermedad parasitaria, se seleccionaron 200 grupos de cinco ratas, que después de dos días de haber sido inoculadas con el parásito se les aplicó una dosis de la droga y al cabo de dos semanas se registró el número de animales muertos. Se quiere conocer cuál es la probabilidad de que muera alguna rata si se repite la experiencia.

Paso 1: El espacio muestral es discreto, finito y con 6 resultados no equiprobables.

Paso 2: En éste caso es necesario recurrir al concepto de frecuencia relativa (num. de grupos con x ratas muertas), para asignar un valor de probabilidad a cada evento elemental. En la 3era columna de la tabla pueden verse dichas probabilidades que cumplen que $\sum_S p(e_i) = 1,0$

Num. de ratas muertas x	Frecuencia fr	Probabilidad $fr/200$
0	120	0.60
1	40	0.20
2	20	0.10
3	10	0.05
4	6	0.03
5	4	0.02

Paso 3: El evento definido es $A = \{\text{una o más ratas muertas}\} = \{1, 2, 3, 4, 5\}$

Paso 4: Para calcular la probabilidad del evento A se puede recurrir a la regla de la complementación, sabiendo que $\bar{A} = \{\text{ninguna rata muerta}\} = \{0\}$. Entonces tendremos que $P(A) = 1 - P(\bar{A}) = 1 - P(0) = 1 - 0,60 = 0,40$. Observar que la regla de la adición arroja el mismo resultado.

1.4. Probabilidad Condicional

En muchas ocasiones la probabilidad de ocurrencia de un evento depende de la ocurrencia o no de otro suceso. Supongamos que de grupo de 100 ratones 80 hembras y 20 machos; se eligen aleatoriamente dos individuos y se verifica su sexo. Cuál es la probabilidad de que el segundo ratón sea hembra? Definamos los eventos $A = \{\text{1er ratón hembra}\}$ y $B = \{\text{2do ratón hembra}\}$. Si elegimos aleatoriamente un ejemplar y después de verificar su sexo se regresa al lote, la probabilidad de obtener una hembra siempre será $P(A) = P(B) = 80/100 = 0,8$. Pero supongamos que se decide que si en la primera extracción el ratón es macho debe regresar al lote, entonces la probabilidad del 2do resultado dependerá del 1ero. Así, seguirá siendo $P(A) = 0,8$ pero $P(B)$ vendrá dada por las siguientes opciones: a) $P(B) = 80/100$ si A no ocurrió, es decir si el 1er individuo fue macho; b) $P(B) = 79/99$ si A ocurrió, es decir si el 1er individuo fue hembra.

En otras palabras, para poder calcular la $P(B)$ debemos saber si A ocurrió o no. Este tipo de probabilidad se llama *condicional*, se indica $P(B/A)$ y se lee la probabilidad de B dado que ocurrió A . Lo más importante de notar es que se está calculando la probabilidad de B sobre un nuevo espacio muestral, el cual es mas reducido.

Veamos otro ejemplo: En familias de 4 hijos, cuál es la probabilidad de que 2 y solo 2 sean mujeres si se sabe que la familia tiene 2 o más mujeres?

Recordemos que el espacio muestral para este ejemplo ya fué detallado por extensión en la página 7. El evento del cual se quiere conocer la probabilidad es:

$$A = \{2 \text{ y solo 2 mujeres}\} = \{\mathbf{VVMM}; VMVM, VMMV, MVMV, MMVV, MVVM\}$$

La probabilidad de A sin ninguna condición es $P(A) = 6/16 = 0,375$. Sin embargo como ya se conoce que la familia seleccionada tiene 2 o más hijas, la información es mayor. El evento que ya ocurrió lo designamos como B y sus 11 resultados que lo integran son:

$$B = \left\{ \begin{array}{l} \mathbf{VVMM}; VMVM, VMMV, MVMV, MMVV, MVVM \\ VMMM, MVMM, MMVM, MMMV \\ MMMM \end{array} \right\}$$

De modo que la probabilidad de obtener 2 y sólo 2 mujeres dado que se sabe que hay 2 o más mujeres, se obtiene dividiendo el número de elementos de A entre el nuevo número de resultados posibles, es decir, $P(A/B) = 6/11 = 0,545$.

Si observamos detenidamente los dos eventos involucrados, nos podremos dar cuenta que los elementos de A están incluidos en B , y esto no es otra cosa que el conjunto $A \cap B$. De modo que la probabilidad condicionada se puede expresar en forma general como:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

1.5. Eventos Independientes

Se dice que una serie de eventos que ocurren unidos o en secuencia son *independientes* si el resultado de uno no afecta al otro. Hay casos en los cuales se puede precisar fácilmente que dos eventos son independientes. Por ejemplo, si le preguntamos a unas cuantas personas en la calle si sufren de miopía y si comen ravioles a la bolognesa, podríamos asegurar que los resultados a las preguntas son eventos independientes ya que dichas acciones no están relacionadas. Pero si les preguntamos a dichas personas si les gusta el fútbol y si han visto TyC Sports alguna vez, no es posible responder con certeza si los dos eventos son independientes, porque es muy posible que la frecuencia de personas que miran partidos de fútbol por dicho canal sea alta. Una manera objetiva de decidir si dos eventos son independientes es comparar las probabilidades de ocurrencia de uno de los eventos antes y después que el otro evento ocurra. En términos formales, dos eventos A y B , se dice que son independientes si se cumple:

$$P(A/B) = P(A)$$

es decir, la probabilidad del evento A no cambia cuando haya ocurrido el evento B .

Observar que dicha relación también puede expresarse como $P(A \cap B)/P(B) = P(A)$ por lo tanto se deduce que la ocurrencia conjunta de dos eventos independientes es igual a $P(A \cap B) = P(A)P(B)$, lo que constituye otra manera de definir la independencia de dos eventos (siempre que las probabilidades sean mayores que cero).

Ejemplo: En un estudio sobre la calidad del agua de los ríos que conforman cierta cuenca hidrográfica, se encontró que el 28 % de los ríos tienen una altitud superior a los 2500 m; un 20 % tienen temperatura del agua menor a 12°C y un 10 % tienen ambas características.

Son independientes los eventos *altitud* ≥ 2500 m (A) y *temperatura* $\leq 12^\circ\text{C}$ (B)?

Los valores de probabilidad se asignan a partir de las frecuencias relativas:

$$P(A) = 0,28 \quad P(B) = 0,20 \quad P(A \cap B) = 0,10$$

La comprobación de la independencia o dependencia de los eventos A y B , se puede hacer a partir de la igualdad que establece que la probabilidad de ocurrencia conjunta de dos eventos independientes es igual al producto de sus probabilidades individuales. Tenemos entonces que

$$P(A \cap B) = 0,10 \quad P(A)P(B) = 0,20 \times 0,28 = 0,06$$

Al ser $P(A \cap B) \neq P(A)P(B)$ se concluye que los eventos A y B no son independientes. Es decir, el hecho de que un río tenga una altitud superior a 2500 m aumenta la probabilidad de que tenga una temperatura menor a 12°C .

1.6. Teorema de Bayes

En el año 1763, dos años después de la muerte de Thomas Bayes (1702-1761), se publicó una memoria en la que aparece, por vez primera, la determinación de la probabilidad de las causas a partir de los efectos que han podido ser observados. El cálculo de dichas probabilidades recibe el nombre de teorema de Bayes. Este teorema proporciona la probabilidad condicional de un evento A dado otro evento B (probabilidad posteriori), en función de la probabilidad condicional del evento B dado A y de la probabilidad marginal del evento A (probabilidad apriori).

Recordemos que la probabilidad condicional de 2 eventos A y B está definida como $P(A/B) = P(A \cap B)/P(B)$, por lo que $P(A/B)P(B) = P(A \cap B)$. Análogamente por simetría también podemos escribir $P(B/A)P(A) = P(A \cap B)$. Combinando ambas ecuaciones obtenemos lo que se conoce como el teorema de Bayes:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

Notar que el denominador $P(B)$ puede ser reescrito de la siguiente manera:

$$P(B) = P(B \cap (A \cup \bar{A})) = P((B \cap A) \cup (B \cap \bar{A})) = P(B \cap A) + P(B \cap \bar{A})$$

usando las formulas para la probabilidad condicional e introduciendo el resultado para $P(B)$ en la ecuación del teorema nos queda

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/\bar{A})P(\bar{A})}$$

Observar que el denominador es una sumatoria sobre los eventos A y \bar{A} que conforman todo el espacio muestral. Una manera general de escribir el Teorema de Bayes es la siguiente: Sea A_1, A_2, \dots, A_n un sistema completo de sucesos (es decir que abarca todo el espacio muestral S), tales que la probabilidad de cada uno de ellos es distinta de cero, y sea B un suceso cualquiera del que se conocen las probabilidades condicionales $P(B/A_i)$. entonces la probabilidad $P(A_i/B)$ viene dada por la expresión:

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^n P(B/A_i)P(A_i)}$$

En resumen, este teorema nos permite, si conocemos la probabilidad de que ocurra un suceso, modificar su valor cuando disponemos de nueva información.

Ejemplo: Un ejemplo clásico del uso del teorema de Bayes es el problema de oro y plata. Hay tres bolsas que tienen, cada una 2 monedas. Las de la primera son de oro, las de la segunda son de plata y las de la tercera son una de plata y otra de oro. Se escoge una bolsa al azar y de ella una moneda también al azar. Si la moneda es de oro, cuál es la probabilidad de que la otra moneda en la bolsa sea de oro también?

Primero, notemos que la segunda bolsa no pudo haber sido elegida (porque no tiene monedas de oro), sólo pudo haber sido seleccionada la primera o la tercera. Si la bolsa elegida hubiese sido la tercera, el evento cuya probabilidad nos interesa no se realiza. De modo que el evento que nos interesa es equivalente a que se haya elegido la primera bolsa. Una vez establecido lo anterior, apliquemos el teorema de Bayes para calcular:

$$P(1^\circ|Au) = \frac{P(1^\circ)P(Au|1^\circ)}{P(1^\circ)P(Au|1^\circ) + P(2^\circ)P(Au|2^\circ) + P(3^\circ)P(Au|3^\circ)}$$

Las probabilidades que entran al lado derecho de la igualdad las sacamos, inmediatamente, de las condiciones del problema y después de hacer cuentas tenemos que

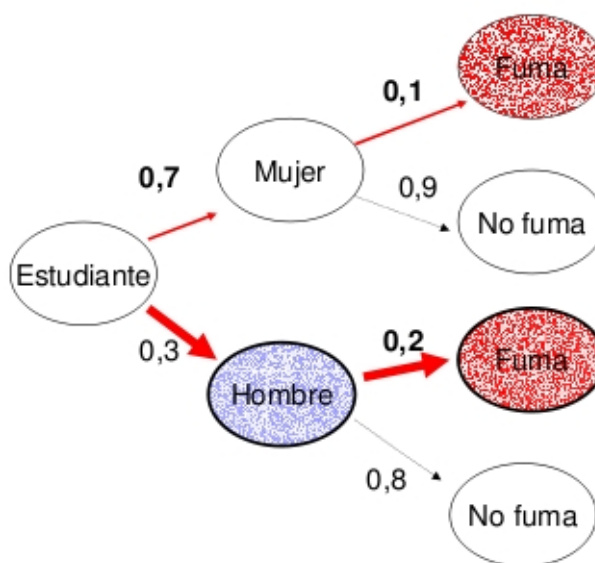
$$P(1^\circ|Au) = 2/3$$

Este problema es clásico porque existe una solución a la que muchas personas llegan y es falsa. El argumento es el siguiente. Como todas las bolsas son igualmente posibles, y el hecho de que la primer moneda extraída sea de oro, nos indica que no se trata de la segunda bolsa. Concluimos que las dos bolsas restantes tienen igual probabilidad y, por tanto, la probabilidad de que la otra moneda sea de oro es $1/2$. Si Ud. piensa de acuerdo a este razonamiento (erróneo!), es muy difícil que encuentre en qué se equivoca. Lo que está mal es que lo que averiguamos, al saber que la moneda extraída es de oro, es algo más que el rechazo de la segunda bolsa. Si sólo nos dijeran que la bolsa escogida al azar

no fué la segunda, sin informarnos del metal de la moneda sacada, todavía tendríamos incertidumbre respecto a la primera moneda; todavía podríamos apostar a si ésta es de oro o de plata. Al decirnos que la moneda fué de oro, estamos aprendiendo algo más, y eso echa por tierra el argumento de igual probabilidad para las dos bolsas restantes. La información con la que contamos nos indica que nos hallamos frente a un caso en el que la bolsa era la primera y sacamos, la primera de las monedas que contenía, o la segunda, (ya llevamos 2 posibilidades), o bien la bolsa era la tercera y en ese caso tan solo podría ser que sacáramos en primer lugar la moneda de oro, luego la que queda dentro es de plata (una única posibilidad). Tenemos 3 posibles sucesos en los que en 2 de ellos sacaríamos a continuación una moneda de oro ($2/3$ de probabilidad), y tan sólo una de las veces la nueva moneda sería de plata ($1/3$ de probabilidad). Lo interesante del problema es que, si nos hubieran dicho que la moneda sacada fué de plata, aplicando la fórmula de Bayes, llegamos a la conclusión de que la probabilidad de que la otra moneda sea también de plata es $2/3!$. Es decir, si vamos a apostar al metal de la otra moneda, nos conviene apostar por el metal de la primera. Este ejemplo nos lleva a reflexionar sobre el uso adecuado de la información contenida en "lo dado." en el cálculo de la probabilidad condicional.

Otro ejemplo:

En este ejemplo veremos una herramienta útil a la hora de estimar las probabilidades usando el Teorema de Bayes. Esta herramienta es la construcción del árbol de probabilidades. Veamos: En un aula el 70% de los alumnos son mujeres. De ellas, el 10% son fumadoras. De los varones, son fumadores el 20%. En la figura de la derecha puede verse la construcción de dicho árbol con la información brindada por el problema. Por lo tanto, formulemos el evento que nos interesa resolver: Si se elige a un individuo al azar y es fumador, que probabilidad hay de que sea un hombre?



Según el Teorema de Bayes la probabilidad de que siendo fumador F sea hombre H es $P(H/F) = \frac{P(H)P(F/H)}{P(F)}$. El numerador de esta fracción se puede calcular siguiendo la línea de flechas

gruesas rojas y multiplicando sus probabilidades, ya que $P(H) = 0,3$ y $P(F/H) = 0,2$. Por último, la probabilidad de ser fumador es $P(F) = P(M)P(F/M) + P(H)P(F/H) = 0,7 \times 0,1 + 0,3 \times 0,2 = 0,13$, en consecuencia la respuesta a nuestro problema es $P(H/F) = (0,3 \times 0,2)/0,13 = 0,46$.

Curiosidad Bayesiana

Aunque probablemente todos razonamos de una forma más parecida a la metodología bayesiana que a la frecuentista, resulta difícil traducirlo en términos matemáticos y difícil de evaluar y de transmitir, por lo que para finalizar voy a citar un artículo escrito por el matemático John Allen Paulos sobre la utilización de las estadísticas que efectuó el abogado defensor en el famoso juicio del jugador y actor norteamericano O.J. Simpson, acusado del asesinato de su mujer, donde vemos que la comprensión del concepto de probabilidad condicional, y al menos una idea intuitiva del teorema de Bayes, es de utilidad y aplicación en la vida diaria:

El abogado defensor Alan Dershowitz afirmaba que, puesto que menos del uno por mil de las mujeres maltratadas por sus compañeros mueren a manos de éstos (cálculo frecuentista), los malos tratos producidos en el matrimonio Simpson no tenían que ver con el caso. Aunque las cifras son correctas, las palabras del señor Dershowitz son de una incongruencia apabullante; no tienen en cuenta un hecho ineludible: Nicole Simpson murió de muerte violenta. Dadas ciertas suposiciones fácticas razonables de homicidio y malos tratos conyugales, se puede ver fácilmente, empleando el teorema de Bayes, que si un hombre maltrata a su mujer o novia, y ésta muere asesinada después, el vapuleador es el homicida más del 80% de las veces. Así pues estaba matemáticamente justificado, a falta de otros indicios, que la policía sospechara inmediatamente del señor Simpson. No estoy defendiendo en modo alguno la derogación de los derechos de nuestra cuarta enmienda; me limito a puntualizar que señalar con el dedo al señor Simpson no era, tal como estaban las cosas, ilógico, ni fue como sostenía el defensor una muestra de racismo. Me pregunto, serían frecuentistas o bayesianos los miembros del jurado?

2. Variables Aleatorias

La identificación de cada resultado, en algunos experimentos aleatorios, obedece a un reconocimiento de las propiedades que lo caracterizan. Por ejemplo, la condición de ser hembra en un recién nacido es un resultado cuya calificación depende de una serie de características cualitativas específicas, al igual que con la raza o la salud. En otros tipos de experimentos aleatorios no basta con calificar los resultados, sino que es necesario caracterizarlos cuantitativamente. En algunos casos esta cuantificación resulta de un proceso de conteo, así se habla del número de hijos, de dientes, de cromosomas, de electrones, de emisiones radiactivas, etc. En otros casos, al determinar características como el peso, la talla, la temperatura o la concentración de alguna sustancia en ciertos objetos o elementos, se asigna a cada resultado un valor dentro de una escala de medición específica. Cada una de esas características cuantificables por conteo o por medición recibe el nombre genérico de variables aleatorias; son variables porque su valor cambia de un elemento a otro; y son aleatorias porque su comportamiento es impredecible. Las variables aleatorias son importantes porque ellas caracterizan los fenómenos o procesos naturales, por lo que resulta muy valioso comprender en la forma más completa posible sus propiedades y comportamiento. Una primera aproximación a este conocimiento se logra estableciendo el conjunto de posibles valores que puede asumir la variable y su respectiva probabilidad de ocurrencia.

2.1. Definición

Hasta el momento a los resultados de un experimento aleatorio los hemos calificado como caras de una moneda, lados del dado, colores de ojos, etc. En matemáticas, es frecuentemente más fácil manejar números que objetos arbitrarios. Por eso, la idea es representar los resultados de un experimento random por números que pueden ser asignados mediante funciones especiales. Veamos como funciona.

Supongamos el espacio muestral de lanzar 3 monedas. Los 8 resultados posibles son:

$$S = \{ccc, ccs, csc, scc, css, scs, ssc, sss\}$$

Este mismo espacio muestral se puede expresar en números. Para esto, es necesario definir una regla o norma que al aplicarla le adjudique a cada resultado un valor. Por ejemplo, se puede establecer la siguiente regla: *contar el número de secas que aparecen en cada resultado del espacio muestral*. La asociación de números a cada resultado puede verse en el caso (a) de la siguiente tabla. Si seguimos viendo la tabla, cuáles serán las reglas definidas para los casos (b) y (c)? Los 3 espacios numéricos mostrados

(a)	(b)	(c)
ccc → 0	ccc → 1	ccc → 0
ccs → 1	ccs → 2	ccs → 1
csc → 1	csc → 2	csc → 1
scc → 1	scc → 2	scc → 1
css → 2	css → 3	css → 4
scs → 2	scs → 3	scs → 4
ssc → 2	ssc → 3	ssc → 4
sss → 3	sss → 4	sss → 9

en la tabla pueden ser expresados como

$$S_1 = \{0, 1, 2, 3\} \quad S_2 = \{1, 2, 3, 4\} \quad S_3 = \{0, 1, 4, 9\}$$

Si adoptamos como x la letra que significa *cantidad de número de sellos* entonces las funciones matemáticas que generan dichos espacios son:

$$f_1(x) = x \quad f_2(x) = x + 1 \quad f_3(x) = x^2$$

Si cada una de estas reglas se define en forma genérica como una variable aleatoria, y a su vez sabemos que cada regla es una función matemática, la definición de variable aleatoria se puede enunciar como:

Sea E un experimento aleatorio y S su espacio muestral, toda función que asigne a cada uno de los elementos de S un número real $X(s)$, se llama variable aleatoria.

Las variables aleatorias se identifican con letras mayúsculas, por lo que nuestros ejemplos podrían ser identificados de la siguiente manera:

$$X = n^\circ \text{ de sellos} \quad Y = n^\circ \text{ de sellos} + 1 \quad Z = \text{cuadrado del } n^\circ \text{ de sellos}$$

El resultado de definir una variable aleatoria es que genera un nuevo espacio muestral numérico que se denomina recorrido o rango espacial y se identifica con la letra R . En nuestro ejemplo tendríamos:

$$R_x = \{0, 1, 2, 3\} \quad R_y = \{1, 2, 3, 4\} \quad R_z = \{0, 1, 4, 9\}$$

Es importante puntualizar algunas cosas con relación al concepto de variable aleatoria:

1. Para un mismo experimento es posible definir diferentes variables aleatorias. En nuestro ejemplo se pudieron especificar otras variables aleatorias como el número de lanzamientos o la distancia entre las monedas.
2. En muchos casos el resultado de un experimento es directamente un número. Por ej., si se mide la altura de un individuo se obtiene directamente un valor.
3. En términos prácticos, en el estudio de una variable aleatoria es más importante conocer los valores que ella asume que saber cuáles son los elementos que conforman su espacio muestral.
4. Los valores que asumen las variables aleatorias se identifican con letras minúsculas por ej. x , y , z . Si se define como variable aleatoria $X = \text{tamaño de una persona}$, y se quiere indicar la probabilidad de que esa persona supere determinada altura, este evento se puede expresar como $P(X > x)$, donde x asume el valor que se especifique.

2.2. Discretas y Continuas

De acuerdo con las características del rango espacial, las variables aleatorias se clasifican en *discretas* y *continuas*.

- *Discretas*: Se denomina discreta si el rango espacial esta constituido por un número finito o infinito contable de valores:

$$R_x = \{x_1, x_2, x_3, \dots, x_r, \dots, x_n, \dots\}$$

Éstas se generan a partir del recuento de elementos: número de hijos, de partículas, de átomos, etc.

Ejemplo: Se registra el número de varones nacidos en los primeros 4 partos ocurridos el primer día del año. El espacio muestral S estará formado por 16 resultados equiprobables. La variable aleatoria *número de varones* origina un espacio R_x formado por 5 resultados numerables.

$$S = \left\{ \begin{array}{l} \text{MMMM} \\ \text{VMMM, MVMM, MMVM, MMMV} \\ \text{VVMM; VMVM, VMMV, MVMV, MMVV, MVVM} \\ \text{VVVM, VVMV, VMVV, MVVV} \\ \text{VVVV} \end{array} \right\} \implies R_X = \left\{ \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \right\}$$

- *Continuas*: Se denomina continua si el rango espacial está constituido por un número infinito de valores en un intervalo dado:

$$R_x = \{X_{(S)} = x / x_1 \leq X \leq x_2\}$$

Estas se generan por la medición de magnitudes como la longitud, el peso, el volumen, la densidad, la temperatura, etc.

Ejemplo: Se atrapó una trucha en un río y se le determinó el tamaño. En éste experimento el espacio muestral R_X se origina inmediatamente como resultado de determinar la longitud del cuerpo del pez, que es una característica propia de cada individuo. El rango espacial R_X está formado por infinitos resultados dentro de un determinado intervalo.

$$S = \{ \text{tamaño de las truchas} \} \implies R_X = \{x_i = \text{tamaño} / 10 \text{ cm} \leq x_i \leq 15 \text{ cm}\}$$

2.3. Función de Probabilidad

Recordemos que dijimos que para tener un buen conocimiento de una variable aleatoria no basta con saber cuáles son los valores que puede asumir sino también es necesario describir su comportamiento en término de probabilidades. Para ello se requiere una nueva función, conocida como función de probabilidad con la cual es posible asignar un valor de probabilidad a cada resultado de un rango espacial.

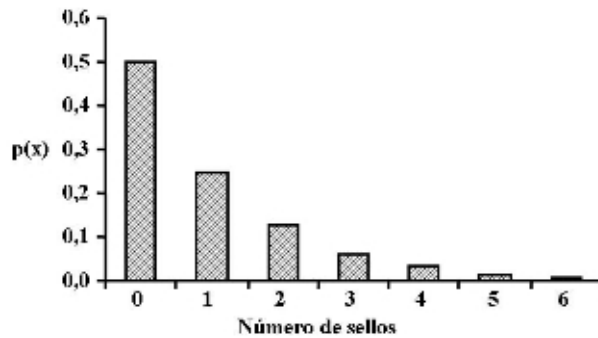
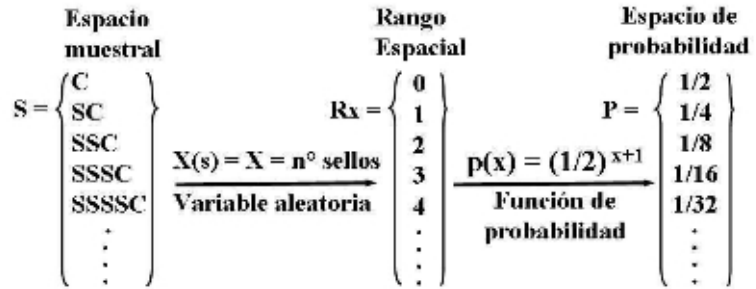
2.3.1. Función de probabilidad de una variable aleatoria discreta

En el caso de variables discretas, la función de probabilidad se denota como $p(x)$ y se interpreta como la probabilidad de que la variable aleatoria tome el valor x_i , es decir, $p(x) = P(X = x_i)$. Obviamente, como la función de probabilidad genera un valor de probabilidad, estos números deben satisfacer las siguientes condiciones:

$$0 \leq p(x) \leq 1 \quad \sum_{R_x} p(x) = 1 \quad P(x_1 \leq X \leq x_2) = \sum_{x_1}^{x_2} p(x)$$

Las dos primeras condiciones son equivalentes a los axiomas probabilísticos de positividad y certidumbre. La tercera propiedad simplemente establece que si se conoce la función de probabilidad de una variable aleatoria discreta, entonces se puede calcular la probabilidad correspondiente a cualquier intervalo abierto o cerrado entre dos puntos x_1 y x_2 .

Ejemplo: Aquí podemos ver el experimento de lanzar una moneda hasta obtener cara por primera vez. En la figura de la derecha pueden observarse los distintos espacios generados por el experimento aleatorio: el espacio muestral S , el rango espacial R_x (generado por la variable aleatoria número de sellos) y el espacio de probabilidad P . El conjunto de pares ordenados $[x_i, p(x_i)]$ para una variable discreta se denomina distribución de probabilidad. En la parte inferior de la figura también puede observarse una representación gráfica de dicha distribución de probabilidad. En este momento es fácil responder a interrogantes relativas a la variable aleatoria. Por ej., cuál es la probabilidad de obtener menos de 3 sellos?. La respuesta se tiene sumando las probabilidades que hay en el espacio de probabilidad:



$$P(X < 3) = P(X \leq 2) = p(0) + p(1) + p(2) = 0,50 + 0,25 + 0,125 = 0,875$$

2.3.1.1. Parámetros de la distribución de una variable aleatoria discreta

La mayoría de las veces resulta poco práctico manejar toda la distribución de probabilidades para determinar el comportamiento de una variable, por lo que es conveniente conocer algunos parámetros que caracterizan a la variable aleatoria. Esta idea se aprecia claramente cuando se tiene una función determinística como es la ecuación de una recta, $f(x) = \alpha x + \beta$, caracterizada por la pendiente α y la ordenada al origen β , los cuales definen completamente el comportamiento funcional. Dos de los parámetros más importantes para caracterizar las variables aleatorias son el *valor promedio* y la *varianza*, los que proporcionan una rápida visión de la naturaleza de la variable.

Valor promedio: Veamos este concepto a través de un ejemplo. En un estudio de campo se determinó, para cierta región, el número de crías por madriguera para una determinada especie de roedor y la probabilidad con la cual esto ocurre. En la tabla de la derecha podemos ver las probabilidades en función del número de crías por madriguera. Si después de un tiempo se revisan N madrigueras en la misma región, es posible estimar en forma aproximada el número de individuos por madriguera que se espera encontrar. Si el número de madrigueras revisado es $N = 300$, el número de madrigueras con un cierto número x de crías (frecuencia esperada) puede observarse en la 3era columna de la tabla. Ahora, si se quiere conocer el número promedio de crías por madriguera se debe multiplicar la frecuencia esperada por el número de crías y su total se divide por el número total de madrigueras

N° crías x	Prob. $p(x)$	Frec. $p(x)N$	fx	fx/N
1	0.25	75	75	0.25
2	0.40	120	240	0.80
3	0.20	60	180	0.60
4	0.08	24	96	0.32
5	0.05	15	75	0.25
6	0.02	6	36	0.12
Total	1.00	300	702	2.34

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{702}{300} = 2,34$$

Si en la fórmula del cálculo de \bar{x} se sustituye $\sum_{i=1}^n f_i$ por N y se aplica el concepto de frecuencia relativa a la probabilidad que establece que $fr_{(x)} = p(x)$, se obtiene una nueva fórmula de cálculo para la media a partir de los valores de probabilidad

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n f_i x_i}{N} = \sum_{i=1}^n \frac{f_i x_i}{N} = \sum_{i=1}^n fr_{(x_i)} x_i = \sum_{i=1}^n p(x_i) x_i$$

La conclusión es que el valor promedio de la distribución de una variable discreta es igual a la suma del producto de cada valor de la variable por su probabilidad de ocurrencia.

Si a este concepto lo extrapolamos de la muestra a la población, el valor promedio de la distribución de valores de una variable discreta es

$$\mu = \sum_{i=1}^n p(x_i)x_i$$

A este valor promedio también se lo conoce como *Esperanza matemática* o *Valor esperado* y se suele denotar como $E(x)$.

Varianza: Si de una población se extrae un niño y se le determina el número de caries, cabrían las siguientes preguntas: El número de caries será igual al valor promedio de la población? El valor estará cercano o alejado al valor promedio?. Si sólo conocemos el valor promedio no podremos responder ninguna de estas preguntas. A lo sumo sabremos que tendrá un número de caries mayor o menor al promedio y que sus probabilidades de ocurrencia dependen de la forma de la distribución de la variable. De modo que no basta conocer el valor medio de una variable aleatoria para poder describir desde un punto de vista práctico alguna de sus características más interesantes. Las preguntas hechas anteriormente hacen pensar que se requiere otro tipo de medida que cuantifique la dispersión de valores alrededor del valor medio. Lo más simple sería determinar la desviación de cada valor respecto al valor medio, es decir sería necesario obtener para cada x_i la desviación $x_i - \mu$. Como se quiere tener un valor de desviación para toda la distribución, la suma de las mismas podría representar una medida general de desviación. Sin embargo, como el término $\sum(x_i - \mu) = 0$, la mejor manera de evadir este problema es elevando al cuadrado cada desviación: $(x_i - \mu)^2$, de modo que el valor promedio de todas las diferencias cuadráticas se podría usar como esa medida única de dispersión. Puesto que $(x_i - \mu)^2$ es también una variable aleatoria, su valor promedio será:

$$\sigma^2 = \sum_{i=0}^n p(x_i)(x_i - \mu)^2$$

Esta medida de dispersión se denomina *varianza*. Una fórmula más simple para el cálculo de σ^2 se obtiene desarrollando el binomio cuadrado presente en la fórmula anterior, obteniéndose

$$\sigma^2 = \left(\sum_{i=1}^n x_i^2 p(x_i) \right) - \mu^2$$

Volviendo a nuestro ejemplo, si nos dijeran que la distribución de probabilidades en función del número de caries por niño es

N° caries	0	1	2	3	4	5	6	7
$p(x)$	0.19	0.29	0.21	0.15	0.09	0.04	0.02	0.01

Se quiere conocer la probabilidad de que un niño tenga más de 2 y menos de 6 caries, el número promedio de caries por niño y la varianza de la distribución. Puede verse que $P(2 < X < 6) = P(3 \leq X \leq 5) = p(3) + p(4) + p(5) = 0,15 + 0,09 + 0,04 = 0,28$. Hacer el cálculo y probar que $\mu = 1,91$ y $\sigma^2 = 2,48$.

2.3.2. Función de probabilidad de una variable aleatoria continua

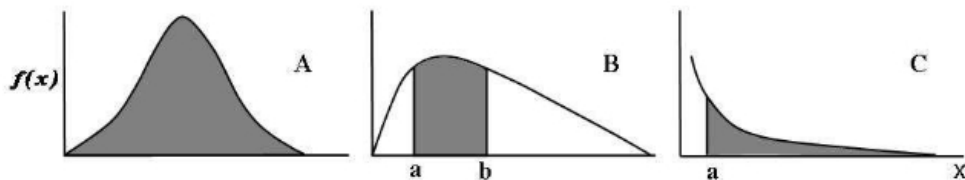
En el caso de variables aleatorias continuas, la función de probabilidad se identifica como $f(x)$. Para las variables continuas no tiene sentido encontrar la probabilidad exacta de un valor puntual puesto que su rango espacial está formado por infinitos valores, de modo que la expresión $P(X = x_i)$ carece de sentido.

Por ejemplo, supongamos que queremos medir la temperatura en la superficie de un lago. Un termómetro con una apreciación en grados puede determinar que la temperatura del agua es $28^\circ C$. Sin embargo debido a la apreciación tan gruesa, cualquier valor entre $27,5^\circ C$ y $28,5^\circ C$ el instrumento lo aprecia como $28^\circ C$. Si se cambia el termómetro por otro con una apreciación de $0,1^\circ C$, el valor de temperatura tendrá una décima más de apreciación, digamos que fue de $28,2^\circ C$. Pero la incertidumbre se mantiene porque cualquier valor entre $28,15$ y $28,25$ es medido como $28,2^\circ C$. Esta falta de seguridad sobre cuál es el verdadero valor de temperatura del agua siempre estará presente, en primer lugar porque teóricamente la apreciación del termómetro puede incrementarse indefinidamente y en segundo término porque el rango espacial de la temperatura, igual que el de todas las variables continuas, está formado por infinitos valores.

Al no poderse definir para una variable aleatoria continua una función $p(x)$ que asigne una probabilidad a cada valor x_i de su rango espacial, es necesario establecer una nueva función $f(x)$ que fije la probabilidad de todos los valores x_i . Esta función debe satisfacer las siguientes condiciones:

$$0 \leq f(x) \leq 1 \quad \int_{x \leq x_i} f(x) dx = 1 \quad P(a \leq X \leq b) = \int_a^b f(x) dx$$

La función $f(x)$ representa la distribución de probabilidad y el área bajo dicha función equivale a su probabilidad de ocurrencia.



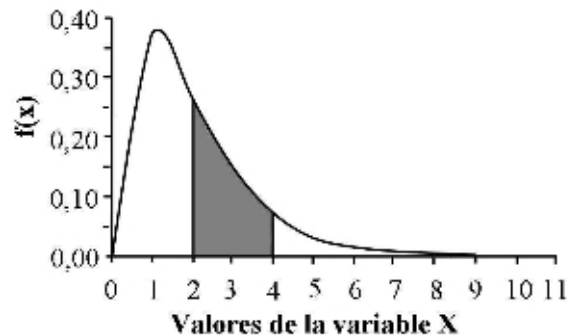
En la figura superior el caso A ejemplifica la condición de que el área sombreada bajo la curva debe ser igual a la unidad; el caso B muestra que el área sombreada representa la probabilidad de que la variable se encuentre entre a y b ; y el caso C indica que el área sombreada bajo la curva representa la probabilidad de que la variable sea igual o mayor al valor a . Por último, observar que una consecuencia de la naturaleza de las variables continuas, es que las probabilidades $P(a < X < b)$, $P(a < X \leq b)$, $P(a \leq X < b)$ y $P(a \leq X \leq b)$ son todas iguales.

Ejemplo:

Encuentre la probabilidad de que una variable aleatoria sea mayor a 2 y menor a 4 si se sabe que su función de probabilidad es

$$f(x) = \begin{cases} x e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Para encontrar la probabilidad solicitada es necesario hallar el área bajo la curva localizada entre los valores 2 y 4 (ver figura). Para ello se procede a integrar por partes la función



$$P(2 \leq X \leq 4) = \int_2^4 x e^{-x} dx = [-x e^{-x} - e^{-x}]_2^4 = [-e^{-x}(x+1)]_2^4 = -5e^{-4} + 3e^{-2} \simeq 0,3144$$

2.3.2.1. Parámetros de la distribución de una variable aleatoria continua

Los significados de la media y la varianza de la distribución de una variable aleatoria continua siguen siendo los mismo que tienen para las variables aleatorias discretas, sólo que en lugar de sumar un número definido de valores enteros, es necesario sumar infinitos valores, de modo que sus fórmulas de cálculo son las siguientes:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

Ejemplo: Siguiendo con la función del ejemplo anterior, su media y varianza son:

$$\mu = \int_0^{\infty} x f(x) dx = \int_0^{\infty} x^2 e^{-x} dx = 2$$

$$\sigma^2 = \int_0^{\infty} x^3 e^{-x} dx - \mu^2 = 6 - 4 = 2$$

2.4. Función de Distribución Acumulada

Probablemente la función de distribución de probabilidades acumuladas sea una de las funciones con más aplicación en la práctica estadística porque la mayoría de las tablas usadas en esta disciplina se generan a partir de funciones acumuladas.

2.4.1. Función acumulada para variables discretas

Al rango espacial de cualquier experimento se le puede asociar otra función que cuantifica la probabilidad de que la variable aleatoria X asuma un valor igual o menor a x_i .

Esta función se simboliza como $F(x)$ y se denomina función de distribución acumulativa. Para el caso de variables aleatorias discretas la función acumulativa queda definida como

$$F(x) = P(X \leq x_i) = \sum_{x \leq x_i} p(x)$$

Ejemplo: Sea la variable aleatoria $X =$ la suma de la cara de 2 dados, determine la distribución de probabilidades acumuladas y calcule las probabilidades siguientes:

- | | |
|-------------------------|-----------------------------------------------------|
| 1) $P(X \leq 6)$ | 5) $P(2 \leq X \leq 8 \text{ y } 5 \leq X \leq 10)$ |
| 2) $P(3 \leq X \leq 8)$ | 6) $P(X > 8 \text{ o } X < 4)$ |
| 3) $P(X > 3)$ | 7) $P(5 < X < 10 \text{ o } X > 7)$ |
| 4) $P(2 < X < 8)$ | 8) $P(4 \leq X \leq 7 / X \leq 6)$ |

a) El espacio muestral está formado por 36 posibles resultados equiprobables

$$S = \left\{ \begin{array}{l} (1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6) \\ (2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6) \\ (3, 1)(3, 2)(3, 3)(3, 4)(3, 5)(3, 6) \\ (4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6) \\ (5, 1)(5, 2)(5, 3)(5, 4)(5, 5)(5, 6) \\ (6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6) \end{array} \right\}$$

b) El rango espacial de la variable aleatoria es el siguiente:

$$R_x = \{ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \}$$

c) Las distribuciones de probabilidad y acumulativas son las que figuran en la tabla de la derecha.

Entonces, las probabilidades solicitadas son:

[[1]] $P(x \leq 6) = F_{(6)} = 0,41667$

[[2]] $P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2) = F_{(8)} - F_{(2)} = 0,72222 - 0,02778 = 0,69$

[[3]] $P(X > 3) = 1 - P(X \leq 3) = 1 - 0,08333 = 0,92$

[[4]] $P(2 < X < 8) = P(3 \leq X \leq 7) = P(X \leq 7) - P(X \leq 2) = F_{(7)} - F_{(2)} = 0,58333 - 0,02778 = 0,56$

[[5]] $P(2 \leq X \leq 8 \text{ y } 5 \leq X \leq 10) = P(5 \leq X \leq 8) = P(X \leq 8) - P(X \leq 4) = F_{(8)} - F_{(4)} = 0,72222 - 0,16667 = 0,56$

[[6]] $P(X > 8 \text{ o } X < 4) = 1 - P(X \leq 8) + P(X \leq 3) = 1 - F_{(8)} + F_{(3)} = 1 - 0,72222 + 0,08333 = 0,36$

[[7]] $P(5 < X < 10 \text{ o } X > 7) = P(6 \leq X \leq 9) + P(X \geq 8) - P(8 \leq X \leq 9) = P(X \leq 9) - P(X \leq 5) + 1 - P(X \leq 7) - P(X \leq 9) + P(X \leq 7) = 1 - P(X \leq 5) = 1 - F_{(5)} = 1 - 0,28 = 0,72$

[[8]] $P(4 \leq X \leq 7 / X \leq 6) = \frac{P(4 \leq X \leq 6)}{P(X \leq 6)} = \frac{P(X \leq 6) - P(X \leq 3)}{P(X \leq 6)} = \frac{0,41667 - 0,08333}{0,41667} = 0,80$

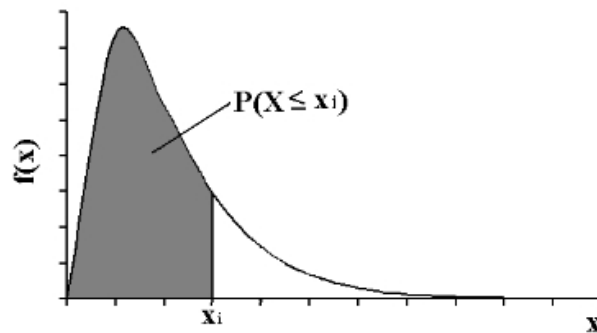
x_i	$p(x_i)$	$F(x_i)$
2	0.02778	0.02778
3	0.05556	0.08333
4	0.08333	0.16667
5	0.11111	0.27778
6	0.13889	0.41667
7	0.16667	0.58333
8	0.13889	0.72222
9	0.11111	0.83333
10	0.08333	0.91667
11	0.05556	0.97222
12	0.02778	1.00000

2.4.2. Función acumulada para variables continuas

Cuando se trata de variables continuas la función acumulativa se define como:

$$\Phi(x_i) = P(X \leq x_i) = \int_{x \leq x_i} f(x) dx$$

En el caso de variables discretas la $P(X \leq x_i)$ se obtiene sumando los valores de probabilidad de todos los resultados iguales o menores a x_i . Para las variables continuas esta probabilidad se obtiene calculando el área que se encuentra por debajo de $f(x)$ y a la izquierda del valor x_i (ver figura de la derecha). Dado que estos cálculos pueden llegar a ser bastante complejos dependiendo de la naturaleza de $f(x)$, se han desarrollado para las funciones de probabilidad más usadas, tablas con las probabilidades acumuladas. Estas facilitan el cálculo de probabilidades.



x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.011521	0.20	0.324718	0.40	0.913659	0.60	0.999268
0.01	0.014561	0.21	0.358065	0.41	0.927102	0.61	0.999467
0.02	0.018268	0.22	0.392531	0.42	0.938882	0.62	0.999615
0.03	0.022750	0.23	0.427863	0.43	0.949118	0.63	0.999724
0.04	0.028125	0.24	0.463782	0.44	0.957941	0.64	0.999804
0.05	0.034518	0.25	0.500000	0.45	0.965482	0.65	0.999862
0.06	0.042059	0.26	0.536218	0.46	0.971875	0.66	0.999903
0.07	0.050882	0.27	0.572137	0.47	0.977250	0.67	0.999933
0.08	0.061118	0.28	0.607469	0.48	0.981732	0.68	0.999954
0.09	0.072898	0.29	0.641935	0.49	0.985439	0.69	0.999968
0.10	0.086341	0.30	0.675282	0.50	0.988479	0.70	0.999979
0.11	0.101557	0.31	0.707280	0.51	0.990952	0.71	0.999986
0.12	0.118639	0.32	0.737730	0.52	0.992947	0.72	0.999990
0.13	0.137656	0.33	0.766471	0.53	0.994543	0.73	0.999994
0.14	0.158655	0.34	0.793373	0.54	0.995810	0.74	0.999996
0.15	0.181651	0.35	0.818349	0.55	0.996807	0.75	0.999997
0.16	0.206627	0.36	0.841345	0.56	0.997585	0.76	0.999998
0.17	0.233529	0.37	0.862344	0.57	0.998188	0.77	0.999999
0.18	0.262270	0.38	0.881361	0.58	0.998650	0.78	0.999999
0.19	0.292720	0.39	0.898443	0.59	0.999002	0.79	1.000000

Ejemplo: Supongamos que la variable $X =$ contenido de plomo en sangre de personas, tiene la función de probabilidades

$$f(x) = \begin{cases} \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\frac{(x-\mu)^2}{2\sigma^2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Usando la tabla de probabilidades acumuladas, calcule la probabilidad de que un individuo seleccionado aleatoriamente a) tenga una concentración superior a 0.40 ppm b) tenga una concentración menor a 0.30 ppm si se sabe que forma parte de un grupo de personas cuya concentración de plomo en la sangre se encuentra entre 0.25 y 0.45 ppm.

a) La probabilidad $P(X \geq 0,40)$ se obtiene calculando el área debajo de $f(x)$ por encima de 0.40. Es decir

$$P(X \geq 0,40) = 1 - P(X \leq 0,40) = 1 - \Phi_{(0,40)} = 1 - 0,913659 = 0,08634$$

En términos prácticos se puede decir que aproximadamente el 8,6 % de los individuos de esa población tienen más de 0.40 ppm de plomo en la sangre.

b) La segunda probabilidad solicitada es condicionada. Interesa obtener dos áreas, la que se encuentra entre 0.25 y 0.30, que representa la intersección de los dos eventos y el área entre 0.25 y 0.45 que es el nuevo espacio muestral reducido. Es decir

$$\begin{aligned} P(X \leq 0,30 / 0,25 \leq X \leq 0,45) &= \frac{P[(X \leq 0,30) \cap (0,25 \leq X \leq 0,45)]}{P(0,25 \leq X \leq 0,45)} = \frac{P(0,25 \leq X \leq 0,30)}{P(0,25 \leq X \leq 0,45)} = \\ &= \frac{P(X \leq 0,30) - P(X \leq 0,25)}{P(X \leq 0,45) - P(X \leq 0,25)} = \frac{\Phi(0,30) - \Phi(0,25)}{\Phi(0,45) - \Phi(0,25)} = \frac{0,685272 - 0,500}{0,965482 - 0,500} = 0,3980 \end{aligned}$$

3. Distribuciones de Probabilidad

La estadística inferencial tiene como problema general establecer las propiedades de un fenómeno aleatorio estudiando una parte del mismo. Para esto es necesario conocer la distribución de probabilidad de la variable aleatoria que se está estudiando. Esto puede ser complicado si no existe otra alternativa que deducir teóricamente la función de probabilidad. Afortunadamente, existen numerosos modelos de probabilidad, muchos de los cuales, aunque hayan sido generados con otros fines, pueden ser usados para describir el comportamiento de la mayoría de las variables aleatorias que son estudiadas en las ciencias naturales. Los modelos de distribuciones de probabilidad se clasifican de acuerdo con la naturaleza de la variable aleatoria en modelos probabilísticos discretos y continuos. En este punto es necesario enfatizar que el esfuerzo que se haga en entender las propiedades de estos modelos permitirá, por una parte, comprender mejor el funcionamiento de los métodos de inferencia estadística y por otro lado, contar con más y mejores criterios para elegir el modelo apropiado en la aplicación de algún método estadístico.

3.1. Modelos probabilísticos discretos

3.1.1. Modelo de Bernoulli

Una gran cantidad de situaciones que se presentan en distintos campos de acción tienen en común algunas cosas. Por ejemplo: lanzar una moneda y determinar si sale cara en cada lanzamiento; lanzar un dado y verificar cada vez si sale un número par; elegir aleatoriamente un individuo y determinar su sexo; determinar si un elemento es metálico; etc. Todos estos experimentos y otros similares reciben el nombre genérico de *Ensayos de Bernoulli*, y tienen las siguientes características:

1. Cada vez que se repite el experimento se producen 2 resultados mutuamente excluyentes. Estos resultados se identifican generalmente como *éxito* y *fracaso*.
2. Cada vez que se repite el experimento la probabilidad de ocurrencia del éxito p o del fracaso q no cambian.
3. Los resultados son independientes. El hecho de que ocurra un fracaso o un éxito, no afecta la probabilidad de ocurrencia de un nuevo resultado al repetir el experimento.

En consecuencia, el espacio muestral para los distintos ensayos de Bernoulli está formado por dos resultados, éxito (E) y fracaso (F), es decir $S = \{E, F\}$. Si definimos la variable aleatoria $X = \text{número de éxitos en un ensayo}$ entonces tendremos el siguiente rango espacial, $R_X = \{0, 1\}$. Si p es la probabilidad de éxito y $1 - p$ la de fracaso, entonces sabemos que la función de probabilidad debe cumplir que $P(X = 0) = p^0(1 - p)^1$ y $P(X = 1) = p^1(1 - p)^0$. Por lo tanto, se deduce que la función de probabilidad para la distribución de Bernoulli es

$$p_{(x)} = p^x(1 - p)^{1-x}$$

El valor esperado y la varianza de esta distribución son: $\mu = p$ y $\sigma^2 = pq$ respectivamente.

3.1.2. Modelo Binomial

Un experimento binomial consta de "varios" ensayos de Bernoulli, por ej, lanzar una moneda n veces y determinar si sale cara en cada lanzamiento. En cada repetición del experimento se mantienen las propiedades de los ensayos de Bernoulli y que la variable aleatoria que los caracteriza es el número de veces que ocurre el éxito (o el fracaso) en n repeticiones del experimento. La función de probabilidad para este tipo de variable la vamos a deducir a partir del siguiente ejemplo.

Ejemplo: En una investigación de cierta parasitosis, pequeñas dosis de una vacuna experimental se inyectaron en ratones de laboratorio. Los resultados encontrados demostraron que 4 de cada 20 ratones mueren a causa de la vacuna. Si la misma dosis de la vacuna se aplica a 4 ratones, cuál es la probabilidad de que mueran x ratones?.

- En primer lugar, se verifica que se trata de un ensayo de Bernoulli.
 - Tiene 2 resultados posibles: ratón muere (éxito); ratón sobrevive (fracaso).
 - La probabilidad de éxito es $p = 1/5$ y del fracaso es $q = 4/5$ (invariantes).
 - El experimento se repitió 5 veces ($n = 4$).
- El espacio muestral consta de 16 resultados. Si se representa con m el evento morir y con s el evento sobrevivir, entonces

$$S = \left\{ \begin{array}{l} \text{ssss} \\ \text{sssm, ssms, smss, msss} \\ \text{ssmm; smsm, mssm, smms, msms, mmss} \\ \text{smmm, msmm, mmsm, mmsm} \\ \text{mmmm} \end{array} \right\}$$

La variable aleatoria $X = \text{número de ratones muertos}$ genera el rango espacial $R_X = \{0, 1, 2, 3, 4\}$.

- Si $p = \text{probabilidad de morir}$ y $q = \text{probabilidad de sobrevivir}$; la probabilidad con la cual ocurrirán los resultados del espacio muestral S son:

$$\begin{aligned} P(X = 0) &= p(\text{ssss}) &&= 1 \text{ qqqq} = 1 \text{ } p^0 q^4 \\ P(X = 1) &= p(\text{sssm}) + p(\text{ssms}) + p(\text{smss}) + p(\text{msss}) &&= 4 \text{ } pqqq = 4 \text{ } p^1 q^3 \\ P(X = 2) &= p(\text{ssmm}) + p(\text{smsm}) + p(\text{mssm}) + p(\text{smms}) + p(\text{msms}) + p(\text{mmss}) &&= 6 \text{ } ppqq = 6 \text{ } p^2 q^2 \\ P(X = 3) &= p(\text{smmm}) + p(\text{msmm}) + p(\text{mmsm}) + p(\text{mmms}) &&= 4 \text{ } pppq = 4 \text{ } p^3 q^1 \\ P(X = 4) &= p(\text{mmmm}) &&= 1 \text{ } pppp = 1 \text{ } p^4 q^0 \end{aligned}$$

- Puede observarse que los valores de p están elevados a una potencia que coincide con el valor de x de la variable aleatoria, mientras que los de q están elevados a una potencia que es igual a $4 - x$. Observar que 4 también es el número de repeticiones del experimento, por lo que una expresión general sería

$$p^x q^{n-x}$$

5. También puede verse que cada término está multiplicado por un coeficiente que representa el número de secuencias diferentes de cómo pueden morir x ratones. Este número no es otra cosa que el número de permutaciones de n elementos diferentes, siendo x elementos de una clase (ratones que mueren) y $n-x$ de otra clase (ratones que sobreviven). Esto también se conoce como combinatoria ${}_n C_x$ y viene descripto por la siguiente fórmula

$${}_n C_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Por lo tanto, la función de probabilidad del modelo binomial puede ser expresada de la siguiente manera

$$p_{(x)} = {}_n C_x p^x q^{n-x} = \binom{n}{x} p^x q^{n-x}$$

Su aplicación permitirá calcular la probabilidad de que un resultado ocurra x veces en n repeticiones. Para finalizar con el ejemplo, podemos utilizar la fórmula encontrada para calcular las probabilidades:

$$\begin{aligned} P(X = 0) &= p_{(0)} = \binom{4}{0} (1/5)^0 (4/5)^{4-0} = (1)(1)(0,4096) = 0,4096 \\ P(X = 1) &= p_{(1)} = \binom{4}{1} (1/5)^1 (4/5)^{4-1} = (4)(0,2)(0,5120) = 0,4096 \\ P(X = 2) &= p_{(2)} = \binom{4}{2} (1/5)^2 (4/5)^{4-2} = (6)(0,04)(0,64) = 0,1536 \\ P(X = 3) &= p_{(3)} = \binom{4}{3} (1/5)^3 (4/5)^{4-3} = (4)(0,008)(0,8) = 0,0256 \\ P(X = 4) &= p_{(4)} = \binom{4}{4} (1/5)^4 (4/5)^{4-4} = (1)(0,0016)(1) = 0,0016 \end{aligned}$$

Distribución de probabilidades

El conjunto de pares ordenados $[x_i; p_{(x_i)}]$ genera una distribución binomial, nombre que se le da porque los sucesivos términos de la distribución de probabilidad son semejantes a los obtenidos con la expansión del binomio de Newton $(p + q)^n$. Cuando una variable aleatoria se distribuye en forma binomial con parámetros n y p se puede representar mediante la siguiente expresión: $X : b(n; p)$. La forma de la distribución binomial cambia para cada combinación de valores diferentes de n y/o p . En la

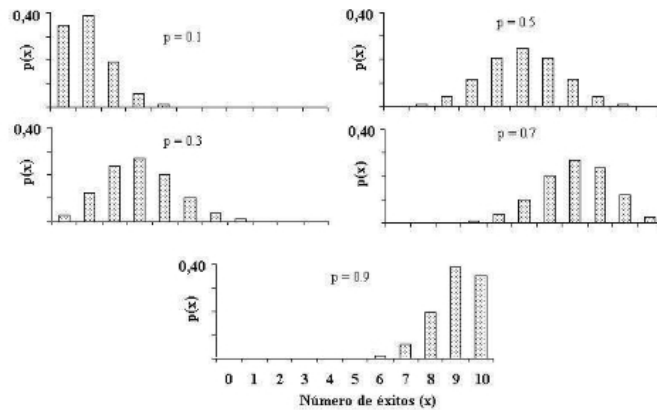


figura puede verse un ejemplo de dicha variación cuando se toma $n = 10$ y diferentes valores de p .

Función de probabilidad acumulada

La función de probabilidad acumulada para el modelo binomial puede escribirse como

$$F_{(x)} = P(X \leq x) = \sum_{R_x} {}_n C_x p^x q^{n-x}$$

Para facilitar la aplicación de la distribución binomial, existen tablas con las probabilidades acumuladas. A continuación damos un ejemplo de dichas tablas. La tabla tiene tres entradas, el valor del parámetro p (probabilidad de éxito), el valor de n (número de repeticiones) y el valor de x (número de éxitos).

		p										
n	X	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6
15	0	0,2059	0,0874	0,0352	0,0134	0,0047	0,0016	0,0005	0,0001	0	0	
15	1	0,549	0,3186	0,1671	0,0802	0,0353	0,0142	0,0052	0,0017	0,0005	0,0001	0
15	2	0,8159	0,6042	0,398	0,2361	0,1268	0,0617	0,0271	0,0107	0,0037	0,0011	0,0003
15	3	0,9444	0,8227	0,6482	0,4613	0,2969	0,1727	0,0905	0,0424	0,0176	0,0063	0,0019
15	4	0,9873	0,9383	0,8358	0,6465	0,5155	0,3519	0,2173	0,1204	0,0592	0,0255	0,0093
15	5	0,9978	0,9832	0,9389	0,8516	0,7216	0,5643	0,4032	0,2608	0,1509	0,0769	0,0338
15	6	0,9997	0,9964	0,9819	0,9434	0,8689	0,7548	0,6098	0,4522	0,3036	0,1818	0,095
15	7	10,000	0,9994	0,9958	0,9827	0,95	0,8868	0,7869	0,6535	0,5	0,3465	0,2131
15	8		0,9999	0,9992	0,9958	0,9848	0,9578	0,905	0,8182	0,6964	0,5478	0,3902
15	9		10,000	0,9999	0,9992	0,9963	0,9876	0,9662	0,9231	0,8491	0,7392	0,5968
15	10			10,000	0,9999	0,9993	0,9972	0,9907	0,9745	0,9408	0,8796	0,7827
15	11				10,000	0,9999	0,9995	0,9981	0,9937	0,9824	0,9576	0,9095
15	12					10,000	0,9999	0,9997	0,9989	0,9963	0,9893	0,9729
15	13						10,000	10,000	0,9999	0,9995	0,9983	0,9948
15	14							10,000	10,000	0,9999	0,9995	
15	15									10,000	10,000	

La tabla mostrada tiene como parámetro $n = 15$ y su uso es relativamente sencillo. Si tenemos un experimento con una variable aleatoria que se distribuye binomialmente con $n = 15$ y $p = 0,5$, y quisiéramos calcular la probabilidad, por ej., $P(X > 5)$, inspeccionando la tabla podríamos calcular que:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - 0,1509 = 0,8491$$

Valor esperado y Varianza

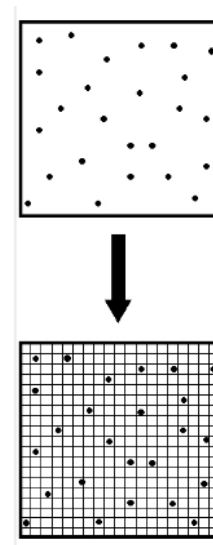
El valor esperado y la varianza de la distribución binomial son los siguientes:

$$\mu = np \qquad \sigma^2 = npq$$

3.1.3. Modelo de Poisson

Esta distribución fue introducida por el matemático francés S.D. Poisson en 1837. El modelo de Poisson, a semejanza del binomial, consta de varios ensayos de Bernoulli. La diferencia estriba en que el modelo binomial sirve para calcular la probabilidad de ocurrencia de un resultado particular en un número finito de repeticiones, mientras que con el modelo de Poisson se determina la probabilidad de ocurrencia de un determinado evento en el tiempo o el espacio y no en un número definido de repeticiones del experimento. En estos eventos que se producen aleatoriamente en el espacio o el tiempo, la frecuencia de ocurrencia de un evento es tan baja con relación a la frecuencia de no ocurrencia que se consideran como *sucesos raros*. Tratar de describir la distribución de una variable aleatoria de este tipo mediante el modelo binomial sería impráctico puesto que el número de ensayos tendría que ser extraordinariamente grande para que ocurriera el resultado esperado. Analicemos el siguiente caso.

Ejemplo: Un biólogo está colectando individuos de una especie de planta cuyos individuos están distribuidos aleatoriamente e independientemente en una sabana. Es de suma importancia conocer la distribución de probabilidades de la variable $X = \text{número de plantas}$. Para obtener esta distribución se podría usar el modelo binomial. Sólo se necesitaría considerar cada punto muestreado como una repetición del proceso, sin embargo, esto implicaría trabajar con un número de repeticiones extremadamente grande, puesto que la presencia de una planta en un punto del área de búsqueda es un hecho muy poco frecuente con relación al número de puntos donde no se encuentra. Bajo el supuesto de que se pudiera superar la dificultad del elevado número de repeticiones, se tendría otro problema, como el de que la función binomial está caracterizada por un valor de n muy grande y un valor de p muy pequeño, lo que hace sumamente tedioso el cálculo de probabilidades por tener que usar factoriales de números muy grandes. Afortunadamente, situaciones como la planteada donde $n \rightarrow \infty$ y $p \rightarrow 0$, se pueden resolver usando el modelo probabilístico de Poisson. Para deducir la función de probabilidad de Poisson se hará uso de dos supuestos: el primero es que en esta sabana se delimitó una parcela de terreno que tiene un número promedio de plantas igual a λ ; y el segundo es que el área de la parcela se corresponde con una unidad de superficie, de forma que λ representa el número promedio de plantas por unidad de superficie. El mayor interés es el de conocer la probabilidad con la cual la variable aleatoria asume los valores de su rango espacial, el cual es $R_x = \{0, 1, 2, 3, \dots, N\}$. Una manera útil de encontrar las probabilidades para cada resultado en R_x sería dividir la parcela en n unidades del mismo tamaño lo suficientemente pequeñas para que en cada uno de ellas se produzca uno de dos resultados: presencia o ausencia de plantas (ver figura de la derecha). Bajo estas nuevas condiciones



el experimento presenta las características de un experimento binomial. Por lo tanto es posible utilizar la función de probabilidad del modelo binomial para el cálculo de probabilidades. Pero para poder hacer esto, hace falta conocer el valor de p . Éste se puede deducir a partir de λ , que es el número promedio de plantas por parcela o por unidad de superficie. Puesto que la parcela se dividió en n subparcelas, la probabilidad de que ocurra una planta en cada una de las n subparcelas de tamaño $1/n$ será $p = \lambda/n$ y la probabilidad de que no ocurra será $q = 1 - (\lambda/n)$, de modo que la función distribución binomial queda:

$$p(x) = {}_n C_x (\lambda/n)^x (1 - \lambda/n)^{n-x}$$

Sin embargo, esta función sólo es una aproximación, pues toma en cuenta n subparcelas. Como la superficie es una variable continua, el área de la parcela se puede dividir en infinitas subparcelas, de modo que cuando n tiende a infinito, la función de probabilidad binomial se aproxima a

$$\lim_{n \rightarrow \infty} {}_n C_x (\lambda/n)^x (1 - \lambda/n)^{n-x} = \frac{e^{-\lambda} \lambda^x}{x!}$$

donde λ es el número de ocurrencia del evento de interés en una unidad de espacio (o tiempo). Para cualquier otro valor de espacio (o tiempo) la función de probabilidad será:

$$p(x) = \frac{e^{-\lambda a} (\lambda a)^x}{x!}$$

donde a es un factor de proporcionalidad que permite calcular el número de ocurrencias del éxito en un tiempo o espacio dado diferente a la unidad. Si se hace $\lambda a = \mu$ la función de probabilidades para el modelo de Poisson queda

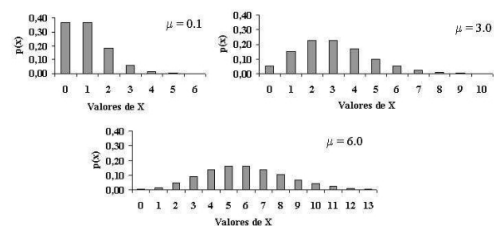
$$p(x) = \frac{e^{-\mu} \mu^x}{x!}$$

con μ el número promedio de ocurrencias en un espacio o tiempo dado y x el número de veces que ocurre el éxito en ese mismo espacio o tiempo.

Distribución de probabilidades

La distribución de probabilidades de Poisson está formada por los pares ordenados de valores $[x_i; p(x_i)]$ y la misma está caracterizada por un sólo parámetro: el promedio μ . En forma similar a la distribución binomial, la distribución Poisson es una familia de curvas, cuya forma depende de μ (ver figura).

Su aplicación puede verse a través del siguiente ejemplo. Supóngase que el número de partículas radiactivas emitidas por cierto material durante una hora tiene una distribución Poisson cuyo promedio es de 0.8 partículas por hora.



Cuál es la probabilidad de que en 5 horas se emitan más de 3 y menos de 7 partículas?
Para encontrar la probabilidad solicitada se deberá calcular

$$P(3 < X < 7) = P(4 \leq X \leq 6) = p_{(4)} + p_{(5)} + p_{(6)}$$

Ahora, si $\lambda = \text{emisiones/hora}$, el número promedio esperado para 5 horas será $\mu = \lambda t = 0,8 \times 5 = 4 \text{ emisiones}$. Entonces, las probabilidades requeridas son

$$p_{(4)} = e^{-4}4^4/4! = 0,1954$$

$$p_{(5)} = e^{-4}4^5/5! = 0,1562$$

$$p_{(6)} = e^{-4}4^6/6! = 0,1041$$

por lo que, la probabilidad total buscada es $P(3 < X < 7) = 0,4557$.

Función de probabilidad acumulada

Las probabilidades acumuladas de la función de probabilidad de Poisson también pueden venir tabuladas, donde las entradas usuales son el parámetro μ y el número de éxitos x .

X	μ									
	1,8	1,9	2	2,1	2,2	2,3	2,4	2,5	2,6	2,7
0	0,1653	0,1496	0,1353	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672
1	0,4628	0,4337	0,406	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487
2	0,7306	0,7037	0,6767	0,6496	0,6227	0,596	0,5697	0,5438	0,5184	0,4936
3	0,8913	0,8747	0,8571	0,8386	0,8194	0,7993	0,7787	0,7576	0,736	0,7141
4	0,9636	0,9559	0,9473	0,9379	0,9275	0,9162	0,9041	0,8912	0,8774	0,8629
5	0,9896	0,9868	0,9834	0,9796	0,9751	0,97	0,9643	0,958	0,951	0,9433
6	0,9974	0,9966	0,9955	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794
7	0,9994	0,9992	0,9989	0,9985	0,998	0,9974	0,9967	0,9958	0,9947	0,9934
8	0,9999	0,9998	0,9998	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981
9	10.000	10.000	10.000	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995
10				10.000	10.000	10.000	10.000	0,9999	0,9999	0,9999

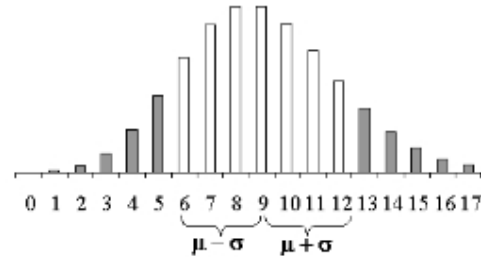
Su utilización es la misma que la realizada con las tablas binomiales, pero veamos un ejemplo para clarificar. Supóngase que el número de impulsos que recibe una central telefónica es una variable que se distribuye como Poisson. El promedio es de 120 impulsos recibidos por hora. La central tiene una capacidad máxima de 4 impulsos por minuto. Cuál es la probabilidad de que en un minuto determinado la central se congestione?. La central comenzará a fallar cuando el número de impulsos sea superior a 4 por minuto, de modo que la probabilidad solicitada es $P(X > 4)$. Si $\lambda = 120 \text{ impulsos/hora} = 120 \text{ impulsos}/ 60 \text{ minutos} = 2 \text{ impulsos/minuto}$, entonces se tiene que $\mu = \lambda t = (2 \text{ impulsos/minuto})(1 \text{ minuto}) = 2 \text{ impulsos}$. Entonces entramos en la tabla con los valores $x = 4$ y $\mu = 2$ y tenemos que la probabilidad buscada es

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F(4) = 1 - 0,9473 = 0,0527$$

Valor esperado y Varianza

El valor esperado y la varianza de la distribución Poisson son iguales: $\mu = \sigma^2$. Apliquemos esto en el siguiente ejemplo. Sea X una variable que se distribuye según el modelo de Poisson, sabiendo que $\mu = 9$ calcule la probabilidad que tiene la variable aleatoria de ser mayor o menor a la media en más de una desviación estándar (σ , es decir, la raíz de la varianza). La probabilidad solicitada es

$$P[X < (\mu - \sigma) \text{ o } X > (\mu + \sigma)]$$



Como se sabe que en el modelo Poisson $\mu = \sigma^2$, se deduce que $\sigma^2 = 9$. Por lo tanto, la desviación estándar es $\sigma = \sqrt{9} = 3$, de modo que la probabilidad que buscamos es:

$$P[X < (9 - 3) \text{ o } X > (9 + 3)] = P[X < 6 \text{ o } X > 12] = P(X < 6) + P(X > 12) = \\ = P(X \leq 5) + 1 - P(X \leq 12) = 0,1157 + 1 - 0,8758 = 0,2399$$

Relación entre los modelos Binomial y Poisson

Observar que la deducción de la función de probabilidad del modelo de Poisson se hizo a partir de la función de probabilidad del modelo binomial. Con este propósito, en un experimento binomial se aumentó infinitamente el número de repeticiones n , y la probabilidad de ocurrencia del éxito se disminuyó proporcionalmente a este aumento, $p = \lambda/n$. Si siguiésemos dentro del marco binomial, el cálculo mediante la función de probabilidad se dificulta porque hay que trabajar con factoriales muy grandes. De modo que en cualquier ensayo de Bernoulli donde n sea muy grande y p muy pequeño, se puede utilizar la función de Poisson para calcular las probabilidades de ocurrencia del éxito, sabiendo que $\mu = np$.

3.1.4. Otros modelos discretos

A continuación se mencionan las características principales de 3 modelos discretos usados comúnmente.

Modelo geométrico

Supongamos que ensayos independientes, cada uno teniendo probabilidades p , son realizados hasta que un evento dado ocurre por primera vez, sin límite en el número de ensayos realizados. Si un evento es observado por primera vez después de x ensayos, significa que falló $x - 1$ veces, es decir, esto pasa con probabilidad $(1 - p)^{x-1}$. Cuando el evento finalmente ocurre, lo hace con probabilidad p . Entonces puede verse que la

función probabilidad vendrá dada por

$$P(x) = (1 - p)^{x-1}p \quad x = 1, 2, \dots$$

Cualquier variable aleatoria cuya probabilidad venga dada por esta ecuación se dice que es una variable aleatoria geométrica. El valor esperado para esta distribución es $\mu = 1/p$, mientras que la varianza es $\sigma^2 = (1 - p)/p^2$.

Modelo binomial negativo

Supongamos que ensayos independientes, cada uno teniendo probabilidades p , son realizados hasta que un total de r eventos éxitos se han acumulado. Para que el r -ésimo evento ocurra en el ensayo x , debe haber habido $r - 1$ éxitos en los primeros $x - 1$ ensayos y el x -ésimo ensayo debe ser un éxito. Por lo que la función de probabilidad es

$$P(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad x = r, r+1, \dots$$

Cualquier variable aleatoria cuya probabilidad venga dada por esta ecuación se dice que es una variable aleatoria binomial negativa con parámetro (r, p) . Observar que una variable aleatoria geométrica es una binomial negativa con parámetro $(1, p)$. El valor esperado para esta distribución es $\mu = r/p$, mientras que la varianza es $\sigma^2 = r(1-p)/p^2$.

Modelo hipergeométrico

Supongamos que una muestra de tamaño n es elegida aleatoriamente (sin remplazar) de una urna que contiene N pelotas, de las cuales m son blancas y $N - m$ son negras. Si llamamos X el número de pelotas blancas seleccionadas, entonces

$$P(x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}} \quad x = 0, 1, \dots, n$$

Cualquier variable aleatoria cuya probabilidad venga dada por esta ecuación se dice que es una variable aleatoria hipergeométrica. El valor esperado para esta distribución es $\mu = nm/N$, mientras que la varianza es $\sigma^2 = \frac{N-n}{N-1} np(1-p)$ con $p = m/N$. Observar que si el número N de pelotas es considerablemente grande comparado con n , entonces el número de pelotas blancas elegidas tiene aproximadamente una función de probabilidad binomial (el cual es un experimento que se realiza con remplazos).

3.2. Modelos probabilísticos continuos

3.2.1. Modelo Normal

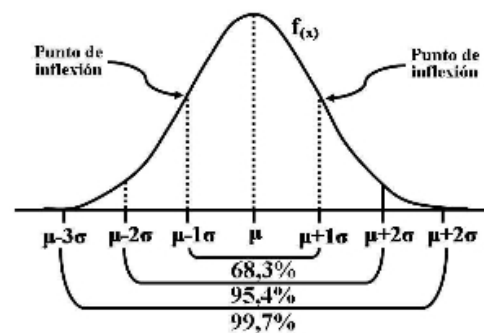
La distribución *normal* fué introducida por el matemático francés Abraham De Moivre en 1733. De Moivre, quien usó esta distribución para aproximar las probabilidades conectadas con lanzar una moneda, la llamó curva exponencial con forma de campana. Su utilidad, sin embargo, fué demostrada en 1809, cuando el famoso matemático alemán Karl Friedrich Gauss la usó como una parte integral de su aproximación para predecir la ubicación de objetos astronómicos. Como resultado, resultó común después de esto que la denominaran distribución Gaussiana. Durante la segunda mitad del siglo XIX, la mayoría de los estadistas comenzaron a creer que la mayoría de los conjuntos de datos tenían histogramas con la forma de campana de una distribución gaussiana, por lo que comenzó a ser aceptado que es *normal* para cualquier conjunto de datos con forma de campana estar descripto por esta curva. Como resultado de esto, y siguiendo el camino del estadista británico Karl Pearson, la gente comenzó a referirse a la distribución gaussiana como la curva normal.

La función de probabilidad de la distribución normal sirve de modelo para una gran cantidad de variables continuas naturales, tales como la temperatura, la humedad, la precipitación, la altura, el peso, la concentración, el coeficiente de inteligencia, los errores instrumentales, etc. Igualmente, la distribución de muchos estadísticos tiende hacia la distribución normal, por lo cual esta distribución adquiere una gran importancia en el análisis de datos mediante la inferencia estadística.

Una variable aleatoria X se encuentra distribuida normalmente si su función de probabilidad es la siguiente:

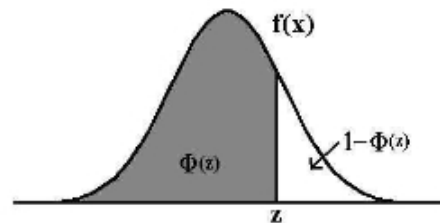
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Esta función está caracterizada por 2 parámetros: la media μ y la desviación estándar σ . El valor de μ define la posición de la distribución y el valor de σ define la forma de la distribución. La distribución normal es simétrica, con un valor máximo para $x = \mu$ y presenta dos puntos de inflexión para $x = \pm\sigma$. En la figura de la derecha pueden verse dichos puntos, como así también las áreas contenidas por los intervalos definidos por 1, 2 y 3 desviaciones estándar alrededor de μ . La función de probabilidad $f(x)$ tiende a cero a medida que x tiende a $\pm\infty$, por lo que las dos colas de la distribución se aproximan asintóticamente a cero. Cuando una variable aleatoria sigue la distribución normal se indica $X : N(\mu; \sigma)$. Por tratarse de un modelo para variables continuas, la probabilidad de que la variable se encuentre en un intervalo se obtiene integrando la función $f(x)$ entre los límites del intervalo. Igualmente,



se puede calcular la probabilidad utilizando la función acumulativa $\Phi(x)$ (ver Sección 2). En el caso de distribuciones discretas, los valores de la función acumulativa están tabulados para diferentes valores de los parámetros que caracterizan estas distribuciones. Esto no es posible en el caso de la distribución normal porque al ser aplicable a variables continuas existen infinitos valores de μ y σ .

Afortunadamente, esta situación se resolvió tabulando las probabilidades acumuladas para una única distribución, con valores de μ y σ específicos, y mediante el procedimiento de tipificación se puede transformar cualquier variable normal en esta variable estándar o patrón. La variable que se seleccionó como estándar es aquella cuya función tiene como parámetros $\mu = 0$ y $\sigma = 1$, por lo cual se le denominó variable normal estándar, unitaria o tipificada, identificándose con la letra Z para diferenciarla de las otras variables cuyas distribuciones de probabilidad tienen $\mu \neq 0$ y $\sigma \neq 1$. La función de probabilidad de la variable Z es la siguiente:



$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

La probabilidad de encontrar un valor de Z en un intervalo dado, se obtiene calculando el área que se encuentra entre la curva y el intervalo definido en el eje de coordenadas. Pero en lugar de integrar $f(z)$ entre los límites del intervalo, esta área se puede calcular utilizando la tabla de la función acumulada $\Phi(z)$, que proporciona los valores de integración entre $-\infty$ y un dado valor de Z .

Transformación de una variable X en la variable Z

Al transformar la función $f(x)$ en la función $f(z)$, lo que realmente se hizo fue sustituir el término $\frac{x-\mu}{\sigma}$ por la variable z

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \xrightarrow[\sigma = 1]{Z = \frac{x-\mu}{\sigma}} f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

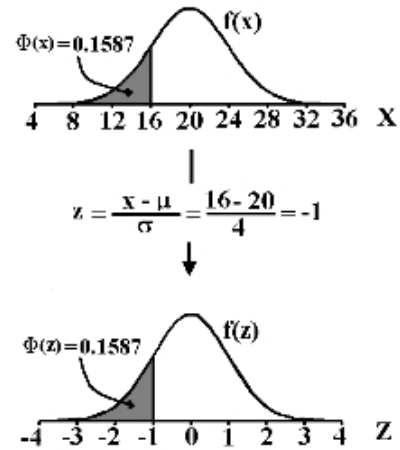
Entonces, cualquier variable X que se distribuye normalmente con $\mu \neq 0$ y $\sigma \neq 1$, se puede convertir en la variable Z , restando a todo valor de X su media μ y dividiendo esta diferencia con su desviación estándar σ . Observar que los valores de Z expresan la distancia de X respecto a su media μ en términos de desviación estándar. Por ejemplo si un valor de una variable X al transformarse produce un valor de $z = 1,5$, este último indica que el valor de X está a $1,5\sigma$ a la derecha de μ .

Ejemplo: Sea $X : N(20; 4)$ y se quiere conocer la probabilidad de que la variable tenga un valor menor a 16. La probabilidad que nos interesa es $P(X \leq 16)$. Para poder determinar el valor de esta área mediante la tabla de probabilidades acumuladas de la distribución normal estándar, se debe convertir el valor de x en su respectivo valor z , lo cual se hace mediante la siguiente operación:

$$z = \frac{x - \mu}{\sigma} = \frac{16 - 20}{4} = -1$$

Ahora se busca el área que se encuentra a la izquierda de $z = -1$ en la tabla de probabilidades acumuladas para la variable Z y se toma dicha área como la probabilidad con que la variable aleatoria X asume un valor igual o menor a 16. En consecuencia se tiene

$$P(X \leq 16) = P\left(Z \leq \frac{16 - 20}{4}\right) = P(Z \leq -1) = \Phi(-1) = 0,1587$$



3.2.2. Modelo Exponencial

Una distribución exponencial aparece frecuentemente, en la práctica, cuando se mide la cantidad de tiempo hasta que un evento específico ocurre. Por ejemplo, la cantidad de tiempo (comenzando ... ahora!) hasta que sucede un terremoto, o hasta que estalle una nueva guerra, o hasta que se reciba una llamada telefónica que resulte ser número equivocado. Todos estos casos son variables aleatorias que tienden, en la práctica, a tener distribuciones exponenciales.

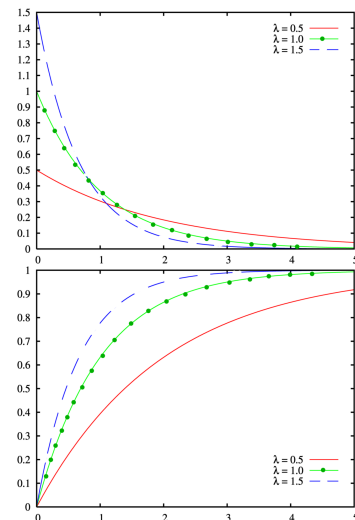
Una variable aleatoria continua cuya función de probabilidad viene dada, para algún $\lambda > 0$, por

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

se dice que es una variable aleatoria exponencial con parámetro λ . La función distribución acumulada exponencial viene dada por

$$F(x_i) = P(X \leq x_i) = \int_0^{x_i} \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^{x_i} = 1 - e^{-\lambda x_i}$$

En la figura de la derecha pueden verse ejemplos de las funciones de probabilidad exponencial (panel superior) y sus correspondientes funciones acumuladas (panel inferior).



El valor esperado para esta distribución es $\mu = 1/\lambda$, mientras que la varianza es $\sigma^2 = 1/\lambda^2$. Una característica importante que poseen las variables aleatorias continuas con una distribución exponencial es que *no tienen memoria*. Que significa esto? Se dice que una variable aleatoria no-negativa X no tiene memoria si

$$P(X > s + t \mid X > t) = P(X > s) \quad \forall s, t \geq 0$$

Si pensamos que X es el período de vida de algún instrumento, esta ecuación establece que la probabilidad de que el instrumento sobreviva por al menos $s+t$ horas, dado que ya sobrevivió t horas, es la misma que la probabilidad inicial de haber sobrevivido al menos s horas. En otras palabras, si el instrumento sobrevivió hasta la edad t , la distribución del tiempo restante de sobrevivida es la misma que la distribución del período original de vida, es decir, es como si el instrumento no recordara que ya ha sido utilizado por un tiempo t . Observar que la ecuación antes escrita, es equivalente a la siguiente

$$\frac{P[(X > s + t) \cap (X > t)]}{P(X > t)} = P(X > s) \longrightarrow P(X > s + t) = P(X > s)P(X > t)$$

con esta ecuación, es fácil corroborar que una distribución exponencial *no tiene memoria*, ya que $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$. Por último, resulta que no sólo la distribución exponencial no tiene memoria, sino que puede demostrarse que es la *única* distribución continua que tiene esta característica.

Ejemplo: Consideremos una oficina de correos que es atendida por 2 empleados. Supongamos que cuando el señor Pérez entra en el sistema, descubre que el señor González está siendo atendido por un empleado y el señor Díaz por otro. Supongamos también que el señor Pérez sabe que será atendido cuando alguno de los otros clientes se vaya. Si la cantidad de tiempo que un empleado emplea con un cliente está distribuido exponencialmente con parámetro λ , cuál es la probabilidad de que, de los 3 clientes, el señor Pérez sea el último en irse del correo?

La respuesta se obtiene haciendo el siguiente razonamiento: consideremos el tiempo en el cual el señor Pérez encuentra un empleado libre. En este punto, alguno de los otros 2 clientes se habrá ido y el otro todavía estará siendo atendido. Sin embargo, debido a la falta de memoria de la distribución exponencial, se concluye que la cantidad de tiempo adicional que esta otra persona (ya sea González o Díaz) tendrá que esperar todavía en el correo también está regulada por una distribución exponencial con parámetro λ . Esto significa, que es la misma cantidad tiempo que faltaría si es que el servicio de esta persona recién estuviese empezando. Por lo tanto, por simetría, la probabilidad de que la persona restante termine antes que el señor Pérez debe ser igual a $1/2$.

3.2.3. Otros modelos continuos

Modelo Gamma

Una variable aleatoria se dice que tiene una distribución gamma con parámetros (t, λ) (ambos mayores que 0) si su función de probabilidad viene dada por

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

donde $\Gamma(t)$ es la llamada función gamma, que esta definida como

$$\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy = (t-1)\Gamma(t-1)$$

Si t tiene un valor entero n , entonces $\Gamma(n) = (n-1)!$. Cuando esto pasa, la distribución gamma con parámetros (t, λ) surge, en la práctica, como la distribución de la cantidad de tiempo que uno tiene que esperar hasta que un total de n eventos haya ocurrido. Más específicamente, si los eventos ocurren aleatoriamente en el tiempo y de acuerdo con los axiomas de un modelo de Poisson, entonces resulta que la cantidad de tiempo que uno tiene que esperar hasta que un total de n eventos haya ocurrido será una variable aleatoria gamma con parámetros (n, λ) . Observar que el valor esperado y la varianza serán $\mu = t/\lambda$ y $\sigma^2 = t/\lambda^2$ respectivamente.

Modelo Beta

Una variable aleatoria se dice que tiene una distribución beta si su función de probabilidad viene dada por

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & (-\infty, 0] \cup [1, +\infty) \end{cases}$$

donde

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

La distribución beta puede ser usada para modelar fenómenos cuyo conjunto de posibles valores es un intervalo finito $[c, d]$, donde c denota el origen y $d-c$ es la unidad de medida que puede transformarse en un intervalo $[0, 1]$. Cuando $a = b$, la función beta es simétrica alrededor de $1/2$, dando más y más peso a las regiones alrededor de $1/2$ a medida que el valor de a crece. Cuando $b > a$ el pico de la función se corre hacia la izquierda, mientras que si $a > b$ el pico de la función se corre hacia la derecha. El valor esperado de la función beta es $\mu = a/(a+b)$, mientras que la varianza viene dada por $\sigma^2 = ab/[(a+b)^2(a+b+1)]$.

3.3. Generadores de números (pseudo) aleatorios

Una herramienta importante para el entendimiento de los fenómenos naturales es simular un proceso natural en forma computacional. Para ello es muy importante contar con generadores de números aleatorios. Estas aplicaciones se realizan en muy variados campos con el fin de emular distintos comportamientos: física (por ejemplo, para simular colisiones entre partículas), ingeniería (diseño de obras hidráulicas, puentes, etc.), inversiones de capital, redes, servicios a clientes, call centers, etc. La simulación a través de la computadora es una herramienta poderosa para comprender la naturaleza de sistemas complejos.

A continuación analizaremos un conjunto de métodos que permitirán generar dichos números basados en reglas o funciones predeterminadas. Ahora, si esto es así, hay una pregunta que cabe hacerse y es por qué motivo un número generado por una fórmula, la cuál es determinística, va a resultar aleatorio. La respuesta es que el número no es aleatorio, pero parece serlo, en el sentido en que, en una aplicación, la relación real entre un número y el siguiente no tiene ningún significado físico. Las secuencias de números generadas en forma determinística reciben el nombre de secuencias *pseudo-aleatorias* o *quasi-aleatorias*, si bien nosotros nos referiremos a ellas como secuencias aleatorias, sobreentendiendo que sólo "parecen" aleatorias. Números aleatorios generados en forma determinística en una computadora funcionan muy bien en muchísimas aplicaciones, siempre que el método de generación sea bueno.

Como dijimos en el párrafo anterior, la idea es generar números pseudo aleatorios a partir de reglas determinísticas. Estos números "lucen como aleatorios y deberían tener muchas de sus propiedades. Con esto, uno podría decir que son "buenos". Ahora, qué significan las palabras "lucen como bueno" en este contexto es algo que debería ser especificado. Uno quisiera tener números aleatorios de tal manera que cada número tenga verdaderamente la misma probabilidad de ocurrencia. Además, si dos números generados difieren muy poco, los números aleatorios que surgen a partir de ellos deberían diferir sustancialmente, es decir, números consecutivos deberían tener baja correlación. A continuación, se describen algunos de los generadores más conocidos que intentan cumplir con estas condiciones.

3.3.1. Números aleatorios uniformes

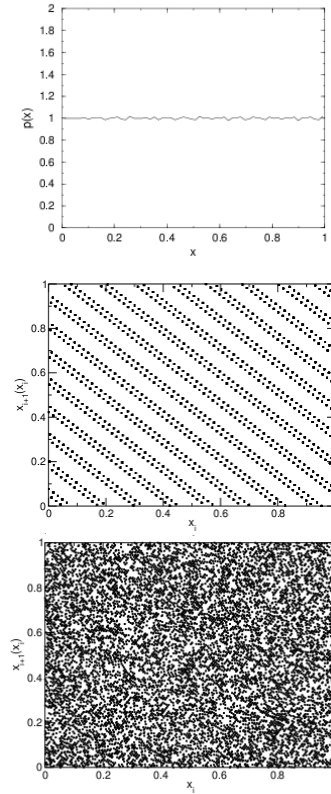
El método más simple para generar números aleatorios son los *generadores de congruencia lineal*. Ellos generan una secuencia x_1, x_2, \dots de números enteros entre 0 y $m-1$ usando una regla recursiva

$$x_{n+1} = (ax_n + c) \bmod(m) = \text{mod}(ax_n + c; m)$$

El valor inicial x_0 se le llama *semilla*. Para generar r números aleatorios distribuidos en el intervalo $[0, 1)$ se tiene que dividir el número aleatorio por el módulo de m .

Observar que se tienen que elegir los parámetros a, c, m de manera de obtener "buenos" números aleatorios, donde "bueno" significa con poca correlación".

Ejemplo: Para ver que significa "generador malo", consideremos un ejemplo con parámetros $a = 12351$, $c = 1$, $m = 2^{15}$ y una semilla con valor $x_0 = 1000$. Se generaron $r=10000$ números aleatorios dividiendo cada uno de ellos por m . Ellos están distribuidos en el intervalo $[0, 1)$. Su distribución puede verse en el panel superior de la figura de la derecha. La distribución luce casi plana, pero si se mira de cerca presenta ciertas irregularidades. Estas irregularidades pueden ser estudiadas analizando las k -tuplas de k números aleatorios $(x_i, x_{i+1}, \dots, x_{i+k-1})$. Un buen generador de números aleatorios no mostraría correlaciones y llenaría el espacio k -dimensional uniformemente. Desafortunadamente, los generadores de este tipo, yacen en planos de $(k-1)$ dimensiones. Puede demostrarse que hay a lo sumo $m^{1/k}$ de esos planos y un mal generador tiene muchos menos. Ese es el caso de nuestro ejemplo. La distribución de los números de nuestro ejemplo puede verse en la correlación de dos puntos $x_{i+1}(x_i)$ entre sucesivos números aleatorios x_i, x_{i+1} mostrada en el panel central de la figura. Es bastante evidente que existen pares sucesivos de números que están altamente correlacionados, es decir, que los números no son tan aleatorios como quisiéramos. Un ejemplo extremo sería usar $a = 1234500$, ya que solo 15 números diferentes pueden crearse (con 1000 de semilla) hasta que la iteración se detiene en un punto fijo. Ahora, si eligiésemos $a = 12349$, la correlación de dos puntos luce como la mostrada en el panel inferior de la figura. Obviamente, el comportamiento es mucho más irregular, aunque algunas correlaciones pobres podrían ser visibles para altas k -tuplas. Un generador que ha pasado varias pruebas empíricas es con $a = 7^5 = 16807$, $m = 2^{31} - 1$ y $c = 0$. Si se desea implementar dicho generador debe tenerse cuidado porque durante el cálculo se generan números que sobrepasan los 32 bits.



Hasta ahora, hemos visto como generar números aleatorios que estén distribuidos aleatoriamente entre $[0, 1)$. En general, estamos interesados en obtener números aleatorios que estén distribuidos de acuerdo con una dada distribución de probabilidades $p(x)$. A continuación varios métodos que nos permitirán lograr esto.

3.3.2. Variables aleatorias discretas

En el caso de distribuciones discretas con un número finito de resultados, se pueden crear tablas de posibles resultados junto con sus probabilidades $p(x_i)$, asumiendo que los x_i están elegidos en orden ascendente. Para elegir un número x , uno tiene que elegir un número aleatorio u el cual tiene que estar distribuido uniformemente entre $[0, 1)$ y tomar la entrada j de la tabla tal que si definimos la suma $s_j \equiv \sum_{k=1}^j p(x_k)$, entonces $s_{j-1} < u < s_j$. Notar que se puede buscar rápidamente en la tabla usando el método de bisección. El vector s_i puede dividirse en 2 mitades, elegir la mitad que contiene u , dividirla en 2 nuevas mitades, elegir la mitad que contiene u , y así sucesivamente, continuar hasta llegar a la condición deseada, es decir determinar j . De esta manera, generar un número aleatorio tiene una complejidad temporal que crece logarítmicamente con el número máximo de entradas de la tabla.

3.3.3. Método de Inversión

Dado un generador de números aleatorios, el cual se asume que genera números aleatorios U uniformemente distribuidos en $[0, 1)$, queremos generar números aleatorios Z con probabilidad $p_Z(z)$. La correspondiente función distribución es

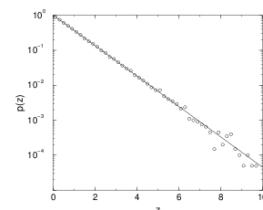
$$F_Z(z) \equiv P(Z \leq z) \equiv \int_{-\infty}^z p_Z(z') dz'$$

El objetivo es encontrar la función $g(u)$ tal que, después de la transformación $Z = g(U)$, los resultados Z estén distribuidos de acuerdo con la ecuación anterior. Se asume que g puede ser invertida y que es estrictamente creciente. Entonces se obtiene

$$F_Z(z) = P(Z \leq z) = P(g(U) \leq z) = P(U \leq g^{-1}(z))$$

Ya que la función distribución $F_U(u) = P(U \leq u)$, para una variable distribuida uniformemente se tiene que $F_U(u) = u$, entonces $F_Z(z) = g^{-1}(z)$. Entonces, uno sólo debe elegir $g(z) = F_Z^{-1}(z)$ como función transformación para obtener números aleatorios que estén distribuidos según $F_Z(z)$. Por lo tanto, este método sólo funciona si F_Z puede ser invertida. Observar que si este no es el caso, entonces debermos usar alguno de los métodos que describiremos en las subsecciones siguientes, o se pueden generar tablas de la función distribución y usar el método para variables discretas que vimos anteriormente.

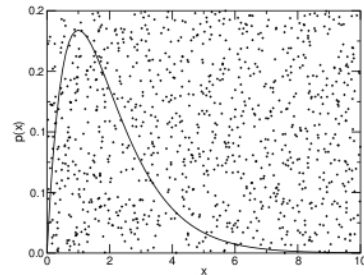
Ejemplo: Dada una distribución exponencial con parámetro μ , y función distribución acumulada $F_Z(z) = 1 - \exp(-z/\mu)$, se pueden obtener números aleatorios distribuidos exponencialmente Z generando números aleatorios uniformemente distribuidos u y eligiendo $z = -\mu \ln(1 - u)$. En la figura de la derecha puede verse este ejemplo usando 10^5 números aleatorios generados con $\mu = 1$. La gráfica está en escala logarítmica en el eje y . Solo para valores grandes se observan desviaciones de la distribución teórica. Esto se debe a fluctuaciones estadísticas ya que $p_Z(z)$ es muy chica en ese rango.



3.3.4. Método de Rechazo

Como ya se mencionó anteriormente, el método de inversión sólo funciona cuando la función distribución puede invertirse analíticamente. Cuando tenemos funciones distribución que no cumplen con esta condición, algunas veces se puede solucionar este problema creando números aleatorios y combinándolos de una manera inteligente.

El método de rechazo funciona para variables aleatorias donde la función de probabilidad $p(x)$ puede limitarse en un cuadrado $[x_0, x_1] \times [0, y_{max}]$, es decir, $p(x) = 0$ para $x \notin [x_0, x_1]$ y $p(x) \leq y_{max}$. La idea básica para generar números aleatorios distribuidos de acuerdo con $p(x)$ es generar pares (x, y) , distribuidos uniformemente en $[x_0, x_1] \times [0, y_{max}]$ y aceptar sólo aquellos números x tales que $y \leq p(x)$, es decir, los pares de puntos que se encuentran por debajo de $p(x)$ (ver figura). De esta manera, la probabilidad de los x elegidos es proporcional a $p(x)$, como se deseaba. El método de rechazo puede ser aplicado siempre que la función de probabilidad se pueda encuadrar, pero tiene la desventaja que muchos más números aleatorios han sido generados que los que son usados. Si el área cuadrada es $A = (x_1 - x_0)y_{max}$, uno tiene que generar, en promedio, $2A$ números aleatorios auxiliares para obtener un número aleatorio que caiga dentro de la distribución. Si esto provoca que el método sea poco eficiente, se puede optar por considerar varias regiones cuadradas para diferentes partes de la función probabilidad.

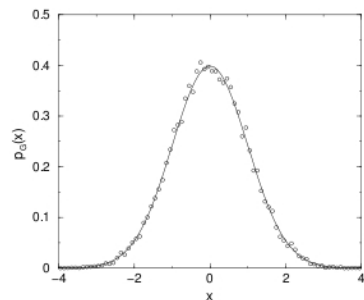


3.3.5. Método de Box-Müller

En el caso de que la función distribución no pueda invertirse ni que la probabilidad pueda encuadrarse, se tiene que aplicar métodos especiales. Tomemos como ejemplo el caso en el que necesitemos generar números aleatorios a partir de una distribución normal o gaussiana. En la figura de la derecha puede verse una distribución normal con media $\mu = 0$ y varianza $\sigma^2 = 1$. Esta función no puede invertirse, ni puede encuadrarse, ya que va desde $-\infty$ a $+\infty$. Por lo tanto, para lograr nuestro objetivo utilizaremos el método de Box-Müller. Se necesitan 2 variables aleatorias U_1 y U_2 uniformemente distribuidas en $[0, 1)$ para generar dos variables gaussianas independientes N_1 y N_2 . Esto puede lograrse generando u_1 y u_2 a partir de U_1 y U_2 y asignando

$$n_1 = \sqrt{-2 \log(1 - u_1)} \cos(2\pi u_2) \quad n_2 = \sqrt{-2 \log(1 - u_1)} \sin(2\pi u_2)$$

Los puntos que se muestran en la figura son 10^4 números aleatorios generados con este método.



3.4. Caracterización completa de las distribuciones de probabilidades

3.4.1. Momentos de una distribución

Hemos visto en las secciones anteriores, que una manera de caracterizar una distribución de probabilidades es establecer dos parámetros conocidos como el valor esperado μ y la varianza σ^2 definidos por:

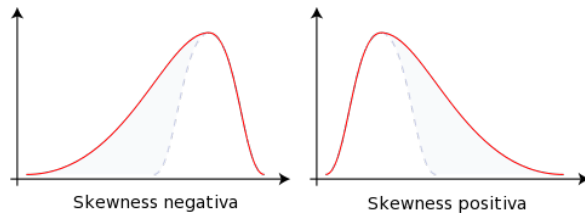
$$\mu = \sum_{i=1}^N x_i p(x_i) = E(X) \quad \sigma^2 = \left(\sum_{i=1}^N x_i^2 p(x_i) \right) - \mu^2 = E(X^2) - E(X)^2$$

los valores $E(X)$ y $E(X^2)$ se denominan los valores de expectación de la variable X de primer y segundo orden respectivamente, y son mejor conocidos con los *momentos de primer y segundo orden* de una distribución. Una caracterización completa de una dada distribución se logra cuando se estudian todos los *momentos* hasta orden n de una distribución, es decir, cuando se conocen los

$$E(X^n) = \sum_{i=1}^N x_i^n p(x_i) \quad \text{con } n = 1, 2, 3, 4, \dots$$

Hasta ahora sabemos que $E(X)$ está directamente relacionado con el valor promedio de una distribución (μ), mientras que con $E(X^2)$ y $E(X)$ podemos tener una idea de la dispersión de los valores de la variable alrededor de su media (σ). A modo de ejemplo, podemos ver que la información brindada por los momentos $E(X^3)$ y $E(X^4)$ está íntimamente relacionada con la forma de una distribución. Estos dos momentos generan dos parámetros conocidos como *Skewness* y *Kurtosis*. La skewness es una medida de la asimetría de una distribución respecto a su valor medio y está definida por

$$\gamma = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} = \frac{E[(x - \mu)^3]}{\sigma^3}$$



En la figura de la derecha pueden verse dos ejemplos de skewness negativa y positiva. Por otro lado, la kurtosis es una medida de cuán sobresaliente es el pico de una distribución y suele venir definido como

$$\kappa = \frac{E(X^4) - 4\mu E(X^3) + 3\mu^2 E(X^2) + \mu^4}{\sigma^4} = \frac{E[(x - \mu)^4]}{\sigma^4}$$

Si a esta definición le restamos el valor 3, estaríamos haciendo una comparación de cuán sobresaliente es el pico de la distribución comparado con una distribución normal. Bajo esta condición se suele denominar *mesokúrtica* a la distribución igual a la normal, *leptokúrtica* cuando $\kappa - 3$ es positivo y *platikúrtica* cuando $\kappa - 3$ es negativo.

Observar que hemos detallado las ecuaciones para el cálculo de los momentos en el caso de distribuciones discretas solamente. De manera análoga podemos definir los momentos para distribuciones continuas como

$$E(X^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

Por último, es necesario notar la importancia del conocimiento de los momentos de una distribución en la estadística. Los momentos caracterizan de tal forma a las distribuciones que si los momentos de dos distribuciones son iguales, diremos que las distribuciones son iguales. Además, podemos decir que dos distribuciones son más semejantes cuanto mayor sea el número de los momentos que coinciden.

3.4.2. Función generatriz de momentos

Una manera de calcular fácilmente los momentos de una distribución, es mediante el uso de la *función generatriz de momentos*. Para una variable aleatoria X , la función generatriz de momentos se define como

$$G_X(k) = E[e^{ikX}] = \int_{R_x} e^{ikx} f(x) dx$$

donde R_x es el rango espacial de la variable X . Observar que cuando x va de menos a más infinito, la función generatriz es la transformada de Fourier de $f(x)$ por lo que la función distribución de probabilidades sería la transformada inversa de Fourier de la función generatriz.

A partir de la función generatriz podemos calcular todos los momentos de la variable aleatoria X . Cuando estos momentos X^n existen, $G_X(k)$ puede ser desarrollada como una serie de Taylor alrededor de $k = 0$, y por lo tanto se deduce que

$$E[e^{ikX}] = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} M_n$$

con

$$M_n = E(X^n) = \left. \frac{1}{i^n} \frac{d^n G_X(k)}{dk^n} \right|_{k=0}$$

en esta última ecuación tenemos todos los momentos de la variable aleatoria X expresados en términos de su función generatriz y, por lo tanto, esta expresión resulta muy útil para el cálculo, en particular, del valor promedio y la varianza de X .

Por último, y a los fines puramente prácticos, podemos simplificar la expresión para G_X , olvidándonos de la parte compleja y escribir las ecuaciones de la siguiente manera, tanto para distribuciones continuas como discretas:

$$G_X(k) = E[e^{kX}] = \int_{R_x} e^{kx} f(x) dx \quad G_X(k) = E[e^{kX}] = \sum_{R_x} e^{kx} p(x)$$

$$M_n = E(X^n) = \left. \frac{d^n G_X(k)}{dk^n} \right|_{k=0}$$

3.4.3. Cumulantes de una distribución

Otra manera de analizar las distribuciones es mediante el uso de los *cumulantes*. Los cumulantes K_n de una variable aleatoria X están definidos por las siguientes relaciones

$$G_X(k) = E[e^{ikX}] = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} M_n = 1 + \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} M_n = \exp\left(\sum_{n=0}^{\infty} \frac{(ik)^n}{n!} K_n\right)$$

$$\ln(G_X(k)) = \ln\left(1 + \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} M_n\right) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} K_n$$

De estas relaciones se deduce que los primeros n cumulantes pueden ser expresados por los primeros n momentos y viceversa. Estas relaciones hasta $n = 4$ se escriben explícitamente como:

$$\begin{aligned} K_1 &= M_1 \\ K_2 &= M_2 - M_1^2 \\ K_3 &= M_3 - 3M_1M_2 + 2M_1^3 \\ K_4 &= M_4 - 3M_2^2 - 4M_1M_3 + 12M_1^2M_2 - 6M_1^4 \\ \\ M_1 &= K_1 \\ M_2 &= K_2 + K_1^2 \\ M_3 &= K_3 + 3K_1K_2 + K_1^3 \\ M_4 &= K_4 + 3K_2^2 + 4K_1K_3 + 6K_1^2K_2 + K_1^4 \end{aligned}$$

Observar que el primer cumulante es igual al primer momento y el segundo cumulante a la varianza. A veces resulta útil considerar densidades de probabilidad para las cuales todos los cumulantes, excepto los dos primeros, se anulan. Cuando ese es el caso, puede verse que la función generatriz queda

$$G_X(k) = \exp\left(ikK_1 - \frac{k^2}{2}K_2\right)$$

Haciendo la anti-transformada llegaríamos a que la función distribución descrita por esta función generatriz, es la distribución normal o gaussiana. Con lo cual concluimos que una variable aleatoria tiene una función de probabilidad normal si y solo si todos sus cumulantes K_n se anulan para $n > 2$.

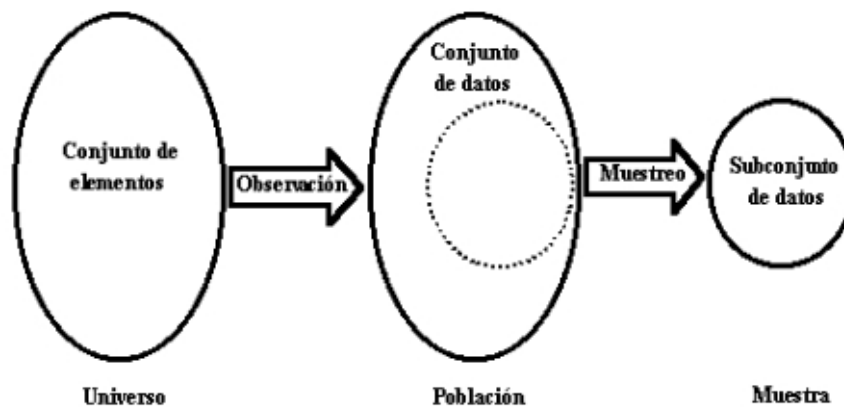
4. Inferencia Estadística

Se mencionó con anterioridad que la inferencia estadística tiene como problema general el establecimiento de las propiedades de un fenómeno aleatorio estudiando una parte del mismo. La teoría de probabilidad genera los modelos que describen la distribución de probabilidades de los resultados de un experimento aleatorio, mientras que los métodos de inferencia estadística evalúan las características de una parte del fenómeno y utilizando esos mismos modelos de probabilidad producen por inducción, conclusiones sobre la totalidad del fenómeno. En la estadística inferencial existe toda una terminología que identifica las diferentes partes y procesos involucrados. Con el propósito de manejar adecuadamente esta terminología será necesario definir algunos conceptos básicos, para luego estudiar algunas propiedades de la porción estudiada del fenómeno, así como la relación funcional que existe entre ella y el colectivo.

4.1. Conceptos importantes

4.1.1. Universos, población y muestra

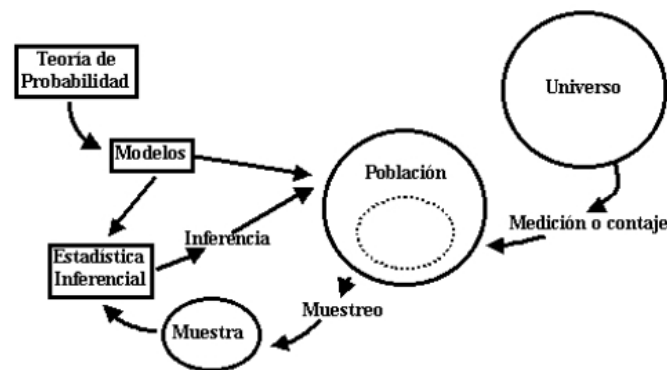
Un fenómeno aleatorio es toda manifestación material susceptible de observarse o medirse mediante los sentidos o instrumentos en individuos, cosas o elementos similares que forman parte de un colectivo denominado *Universo*. Este colectivo puede estar formado por un número finito o infinito de tales unidades. Una *Observación* es un dato o valor numérico que se obtiene al calificar o cuantificar una característica en las diferentes unidades. El conjunto de observaciones origina una *Población*, la cual puede estar formada por un número finito o infinito de datos o valores numéricos. Una *Muestra* es un conjunto formado por n observaciones extraídas de la población. El número n de observaciones define el tamaño de la muestra.



Ejemplos:

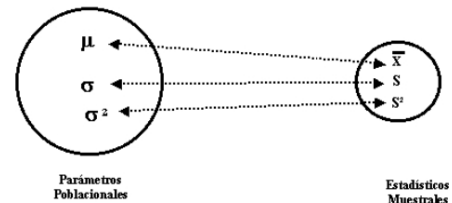
<p>(I) Un productor agrícola quiere conocer algunas características de las mazorcas de maíz producidas en una parcela. Para tal fin selecciona 50 mazorcas y cuenta el número de granos en cada mazorca .</p>	<ul style="list-style-type: none"> ▪ Universo: todas las mazorcas de maíz que produjo la parcela. Universo finito. ▪ Población: todos los valores de la característica número de granos de cada mazorca. Población finita. ▪ Muestra: 50 valores de la característica número de granos.
<p>(II) El mismo productor seleccionó 20 mazorcas y determinó el peso de cada una.</p>	<ul style="list-style-type: none"> ▪ Universo: el mismo del ejemplo anterior. ▪ Población: todos los valores de peso de cada mazorca. ▪ Muestra: 20 valores de la característica peso de cada mazorca.
<p>(III) Un biólogo quiere conocer algunas características de los rabipelados <i>Didelphus marsupialis</i>. Seleccionó 100 individuos y le determinó a cada uno el número de glándulas sebáceas en los miembros anteriores.</p>	<ul style="list-style-type: none"> ▪ Universo: conjunto de rabipelados de la especie <i>Didelphus marsupialis</i>. Universo infinito formado por todos los ejemplares que viven vivieron y los que van a existir en el futuro. ▪ Población: todos los valores de la característica número de glándulas sebáceas. Población infinita. ▪ Muestra: 100 valores de la característica número de glándulas sebáceas.
<p>(IV) El biólogo del ejemplo anterior midió el contenido de hemoglobina en la sangre de 500 rabipelados.</p>	<ul style="list-style-type: none"> ▪ Universo: igual al anterior. ▪ Población: todos los valores de la concentración de hemoglobina. Población infinita. ▪ Muestra: 500 valores de la característica concentración de hemoglobina.
<p>(V) Otro biólogo desea conocer el tamaño de los sapos del género <i>Atelopus</i> que viven en la selva de Monte Zerpa. Capturó 35 individuos y midió la longitud del cuerpo de cada ejemplar.</p>	<ul style="list-style-type: none"> ▪ Universo: todos los sapos del género <i>Atelopus</i> que viven hoy en Monte Zerpa. Universo finito. ▪ Población: todos los valores del tamaño. Población finita. ▪ Muestra: 35 valores de longitud o tamaño.

De los ejemplos anteriores se pueden obtener dos conclusiones importantes: la primera es que los conceptos de universo y población son relativos y es el investigador quien determina, según su interés, la extensión del universo y, consecuentemente, la de la población a estudiar. Así vemos como en los ejemplos (III) y (IV) el biólogo al decidir estudiar los rabipelados al nivel taxonómico de especie, estaba también decidiendo estudiar un universo infinito. Por el contrario, en el ejemplo (V) limitó su estudio a los sapos del género *Atelopus* que viven en un sitio determinado, es decir que decidió trabajar con un universo finito. La segunda conclusión que puede obtenerse es que de un universo se pueden generar varias poblaciones. Así vimos que del mismo universo de mazorcas se generó una población de números de granos (I) y otra de peso de los granos (II). En la siguiente figura puede verse un esquema relacionando la probabilidad y la estadística, ahora incluyendo los conceptos nuevos.



4.1.2. Parámetros y estadísticos

Cuando estudiamos un fenómeno aleatorio, realmente lo que estamos haciendo es analizar las propiedades de las diferentes poblaciones de las variables que lo caracterizan. Muchas de las propiedades poblacionales son descritas por valores que reciben el nombre genérico de *Parámetros*. Por lo general los parámetros se identifican mediante una letra griega y son valores únicos que no cambian entre tanto no cambie la composición de la población. Algunos de los parámetros poblacionales más importantes son: el promedio (μ), la varianza (σ^2) y la desviación (σ). Las muestras también tienen características propias y relacionadas funcionalmente con las propiedades de la población. Estas características muestrales reciben el nombre de *Estadísticos*, y a diferencia de los parámetros son variables y cambian de muestra a muestra. Los estadísticos se identifican con letras del alfabeto romano y entre los más importantes se pueden señalar la media aritmética (\bar{X}); la varianza (S^2) y la desviación estándar (S).



4.2. Muestra y Muestreo

4.2.1. Muestra representativa

Las muestras deben proporcionar la información necesaria (estadísticos), a partir de la cual se infieren las propiedades (parámetros) de la población. En una buena muestra debe estar representada toda o al menos una gran parte de la información presente en la población. Para que una muestra sea representativa debe incluir los valores de la variable en la misma proporción como ellos se encuentran repartidos en la población.

En la tabla de la derecha se representa la producción porcentual de cuatro diferentes variedades de soja obtenida en una determinada región y los valores de esta misma producción de acuerdo con lo estimado con dos muestras. De la tabla se deduce

Variedad	Produc. real	Muestra 1	Muestra 2
A	52 %	25 %	49 %
B	24 %	35 %	26 %
C	18 %	22 %	17 %
D	6 %	18 %	8 %

que la distribución de la producción de soja evidenciada por la muestra 2 y la distribución de la producción real son muy parecidos, por lo tanto, se puede decir que la muestra 2 es representativa de la producción de la población. Por el contrario, la muestra 1 proporciona una distribución de la producción que no se corresponde con la de la región y, obviamente, no es representativa. Lograr que una muestra sea representativa es una tarea difícil, especialmente si se trata de poblaciones infinitas. Una manera de hacerlo es tomando muestras grandes, ya que se incrementa la posibilidad de que todos los grupos de valores de la variable que caracteriza la población estén representados. Sin embargo, este procedimiento, además de desvirtuar el fundamento de la estadística inferencial, puede significar incrementos importantes en los costos, en el tiempo o en la dificultad para manejar una mayor cantidad de información.

4.2.2. Muestreo aleatorio

Otra manera de lograr que una muestra sea representativa es eligiendo aleatoriamente los valores que van a formar parte de la muestra. Mediante el muestreo aleatorio todos los valores de la población tienen la misma posibilidad de ser elegidos, de modo que si se toma una muestra de un tamaño adecuado y se eligen aleatoriamente los elementos que la conforman se está asegurando la representatividad de la muestra. El muestreo aleatorio puede ser simple o restringido.

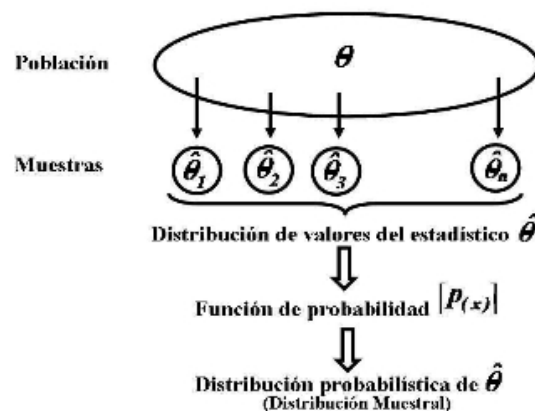
El siguiente ejemplo puede aclarar el funcionamiento del muestreo aleatorio simple. Supongamos que se quieren seleccionar 24 ratones de un grupo de 100 con el propósito de determinar la concentración promedio de una hormona en el grupo de animales. En primer lugar es necesario advertir que un universo de este tipo puede ser bastante heterogéneo, puesto que puede estar formado por individuos con diferentes progenitores, sexo, tamaño, peso, edad, etc. Consecuentemente la población de valores de la hormona también es heterogénea. Para que la muestra sea representativa es necesario que en ella

estén presentes valores provenientes de todas las categorías y en la misma proporción como están repartidas en la población. Si elegimos aleatoriamente los ratones, cada uno de ellos tiene la misma posibilidad de ser seleccionado y la probabilidad de que cada característica sea escogida es proporcional a su tamaño. Estas dos cualidades del proceso de elección deben hacer que la composición de la muestra se aproxime a la de la población.

En muchas ocasiones el tamaño de la muestra no es lo suficientemente grande para asegurar que las distintas categorías de valores de una población estén representadas proporcionalmente. Si no es posible aumentar el tamaño de la muestra, se puede recurrir a un muestreo aleatorio restringido, el cual aumenta la posibilidad de obtener muestras representativas. Entre los varios tipos de muestreo restringido que existen se pueden mencionar los siguientes: el muestreo estratificado, el muestreo por agrupamiento, el muestreo sistemático, el muestreo secuencial, etc. A modo de ejemplo contaremos el procedimiento para el muestreo estratificado. En este tipo de muestreo se divide la población en estratos o subpoblaciones dentro de las cuales se procede a realizar un muestreo aleatorio simple. El tamaño de las muestras pueden ser proporcional al tamaño de los estratos o todas pueden ser del mismo tamaño independientemente del tamaño de los estratos. Volvamos al ejemplo de los ratones. Las mismas características ya nombradas nos pueden servir para estratificar la población. Por ejemplo, podemos clasificar los ratones de acuerdo al estado de desarrollo del proceso reproductivo en tres categorías: inmaduros, maduros y post-reproductivos. La muestra de 24 valores de la hormona que se está estudiando se puede medir seleccionando aleatoriamente el mismo número de ratones dentro de cada una de estas categorías, o seleccionando dentro de cada categoría un número de ratones que sea equivalente a su proporción en la población.

4.3. Distribuciones Muestrales

Como ya sabemos un estadístico es una propiedad muestral cuyo valor cambia de muestra a muestra, por lo cual se comporta como una variable aleatoria. En consecuencia debe existir un modelo o función de probabilidad que describa su distribución de probabilidades, la cual se denomina distribución muestral. En la figura de la derecha puede verse un esquema que describe la idea. La importancia de conocer y comprender las distribuciones muestrales resulta del valor que ellas tienen para la inferencia estadística. En esta primera parte, solo nos interesa conocer las principales distribuciones y familiarizarnos con sus propiedades.



4.3.1. Distribución de la media muestral

Si de una población de valores de una variable aleatoria X que se distribuye normalmente con media μ_x y varianza σ_x^2 se extrae una muestra de tamaño n , entonces se puede calcular la media (\bar{x}) de la muestra. Esta media representa una de las muchas medias muestrales que se pueden extraer de la población de valores de la variable X . Por lo tanto, la media muestral, a su vez, es una nueva variable aleatoria \bar{X} que conforma una nueva población cuyos parámetros $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}^2$ se pueden deducir mediante la aplicación de la denominada *propiedad reproductiva de la distribución normal*.

4.3.1.1. Propiedad reproductiva de la distribución normal

Sean $X_1, X_2, X_3, \dots, X_n$, variables que se distribuyen normalmente, con la misma media: $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$ y la misma varianza: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_n^2$. La variable que resulta de la suma de las n variables individuales: $X = X_1 + X_2 + X_3 + \dots + X_n$, también se distribuye normalmente con una media: $\mu_x = \mu_1 + \mu_2 + \mu_3 + \dots + \mu_n = n\mu$ y una varianza: $\sigma_x^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots + \sigma_n^2 = n\sigma^2$.

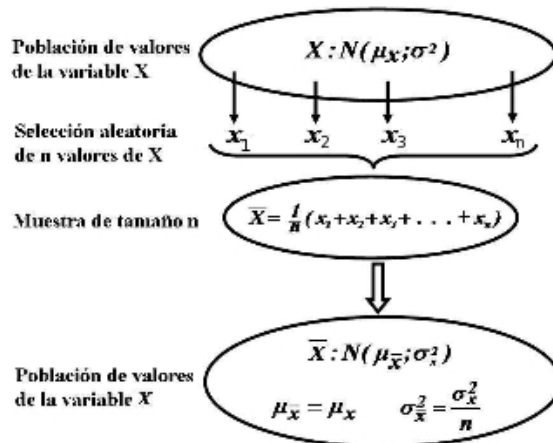
Puesto que es posible demostrar que cada uno de los valores ($x_1, x_2, x_3, \dots, x_n$) que forman parte de una muestra son una variable aleatoria que proviene de una misma población, se puede concluir que la media muestral es una variable que resulta de la suma de varias variables que tienen la misma μ y la misma varianza σ^2 .

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1}{n} + \frac{x_2}{n} + \frac{x_3}{n} + \dots + \frac{x_n}{n}$$

Por lo tanto, la media y la varianza de la distribución de medias muestrales serán:

$$\mu_{\bar{x}} = \frac{n\mu}{n} = \mu \quad \sigma_{\bar{x}}^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Por lo tanto, si se toman muestras aleatorias de la población de una variable X que se distribuye normalmente, la distribución de las medias muestrales también es normal con una media igual a la media de la población de la variable X , y una varianza igual a la de la población dividida con el tamaño de la muestra. La desviación de la distribución de medias muestrales se le denomina error estándar y se obtiene como el cociente entre la desviación de la población de la variable X y la raíz cuadrada del tamaño de la muestra $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.



Ejemplo: Sea una población de una variable que se encuentra distribuida normalmente con una media y una varianza igual a 800 y 1600 respectivamente, de la cual se seleccionan aleatoriamente 16 valores. Cuál es la probabilidad de que la muestra tenga un valor medio menor a 775?

Por la propiedad reproductiva sabemos que la media de una muestra obtenida de una población de valores distribuidos normalmente, también se distribuye normalmente con una media y una varianza igual a:

$$\mu_{\bar{x}} = \mu_x = 800 \quad y \quad \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{1600}{16} = 100$$

Por otro lado sabemos que para poder encontrar la probabilidad de ocurrencia de la variable aleatoria \bar{X} que sigue una distribución normal es necesario hallar un valor equivalente en términos de la variable Z , para lo cual recurrimos al estadístico

$$z = \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}}$$

Por lo tanto, la probabilidad deseada es:

$$P(\bar{X} \leq 775) = P\left(Z \leq \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}}\right) = P\left(Z \leq \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}}\right) = P\left(Z \leq \frac{775 - 800}{40 / \sqrt{16}}\right) = 0,0062$$

4.3.1.2. Teorema del Límite Central

La primera versión de este teorema fué propuesta por DeMoivre en 1733 para el caso especial donde las variables aleatorias eran de un ensayo de Bernoulli con $p = 1/2$. Esta fué subsecuentemente extendida por Laplace al caso de probabilidad arbitraria. Laplace también descubrió la manera más general de escribir el teorema. Su demostración, sin embargo, no era completamente rigurosa y, de hecho, no puede hacerse rigurosa fácilmente. Una demostración completamente rigurosa del teorema del límite central fué presentada por primera vez por el matemático ruso Liapounoff en el periodo de 1901-1902.

A continuación se se plantea el problema, y el correspondiente enunciado del teorema, acompañado por unos ejemplos gráficos.

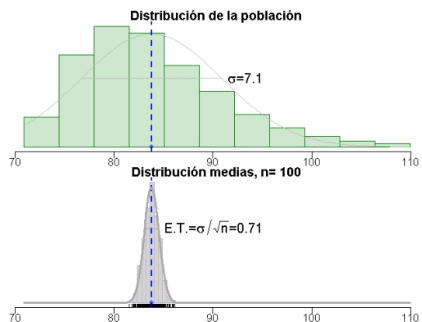
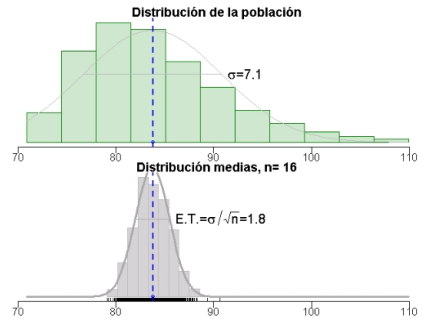
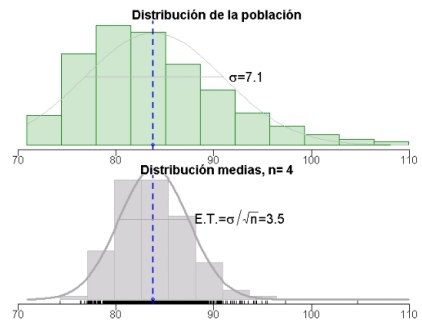
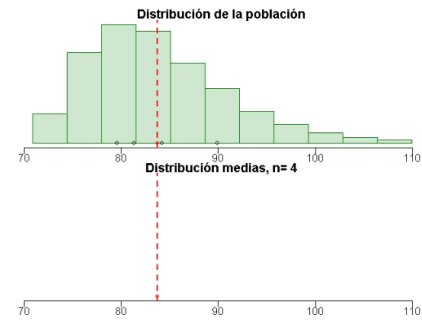
Supongamos que se tiene una variable de la cual se conoce la media μ_x y la varianza σ_x^2 pero no la forma de su distribución. Esto impide la aplicación de la propiedad reproductiva y consecuentemente la deducción de los parámetros que caracterizan la distribución de las medias muestrales. Sin embargo, se puede recurrir a otra propiedad de la distribución normal conocida como el *Teorema del Límite Central*, que establece lo siguiente:

Sean $X_1, X_2, X_3, \dots, X_n$ variables independientes con una misma función de probabilidad y por tanto con una misma distribución e igual $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$, e igual varianza $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_n^2$. La variable que resulta de la suma de las n variables independientes $X = X_1 + X_2 + \dots + X_n$ también se distribuye normalmente con una media: $\mu_x = \mu_1 + \mu_2 + \mu_3 + \dots + \mu_n = n\mu$ y una varianza: $\sigma_x^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots + \sigma_n^2 = n\sigma^2$ cuando n es grande.

En términos menos formales, el teorema anterior establece que las medias provenientes de muestras grandes tomadas de poblaciones con una distribución desconocida, se distribuyen normalmente con media y varianza igual a:

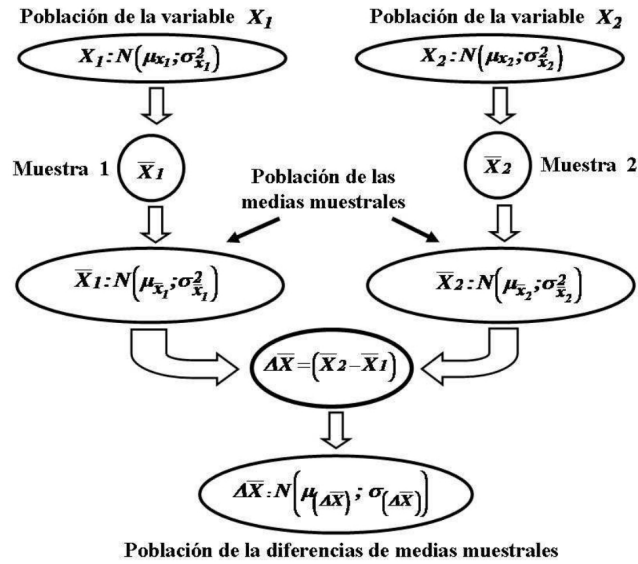
$$\mu_{\bar{x}} = \mu_x \quad y \quad \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

Por lo tanto, si se desconoce la distribución de una variable se puede suponer que aumentando el tamaño de la muestra, la distribución de la media muestral se aproximará progresivamente a una normal. En la práctica, se considera que una muestra de tamaño $n \geq 30$ es lo suficientemente grande para que se cumpla el teorema.



4.3.2. Distribución de la diferencia de medias muestrales

Muchas veces es necesario estudiar dos poblaciones de una misma variable. Supongamos que la variable se distribuye normalmente en ambas poblaciones y que de cada una se extrae independientemente una muestra aleatoria con tamaños n_1 y n_2 respectivamente, y que además se calcula la media de las dos muestras. A partir de éstas dos medias muestrales es posible generar nuevas variables que relacionen las dos poblaciones. Por ejemplo, se pueden sumar, restar, multiplicar o dividir los valores de las dos medias muestrales y originar otras variables cuyos valores estarían representados por el resultado de las operaciones realizadas. De estas nuevas variables, la más conveniente para la inferencia estadística es la diferencia de medias muestrales debido que se conocen las propiedades de su distribución de frecuencia. Cuando el muestreo de una variable se hace a partir de poblaciones que se distribuyen normalmente, la diferencia de medias muestrales es una nueva variable que de acuerdo con la propiedad reproductiva también se distribuye normalmente con media y varianza igual a:



$$\mu_{(\bar{x}_2 - \bar{x}_1)} = \mu_{\bar{x}_2} - \mu_{\bar{x}_1} = \mu_{x_2} - \mu_{x_1}$$

$$\sigma_{(\bar{x}_2 - \bar{x}_1)}^2 = \sigma_{\bar{x}_2}^2 + \sigma_{\bar{x}_1}^2 = \frac{\sigma_{x_2}^2}{n_2} + \frac{\sigma_{x_1}^2}{n_1}$$

Conocido el modelo de probabilidad que describe la distribución de la diferencia de medias muestrales, se puede calcular la probabilidad de ocurrencia de un determinado valor de la diferencia de medias muestrales, utilizando la transformación de Z

$$Z = \frac{(\bar{x}_2 - \bar{x}_1) - \mu_{(\bar{x}_2 - \bar{x}_1)}}{\sigma_{(\bar{x}_2 - \bar{x}_1)}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_{\bar{x}_2} - \mu_{\bar{x}_1})}{\sqrt{\sigma_{\bar{x}_2}^2 + \sigma_{\bar{x}_1}^2}} = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_{x_2} - \mu_{x_1})}{\sqrt{\frac{\sigma_{x_2}^2}{n_2} + \frac{\sigma_{x_1}^2}{n_1}}}$$

Ejemplo: Una muestra de tamaño 5 se obtiene aleatoriamente en una población de una variable normalmente distribuida con media igual a 50 y varianza igual a 9 y se registra la media muestral. Otra muestra aleatoria de tamaño 4 se selecciona en una segunda población de la misma variable cuya media es igual a 40 y su varianza igual a 4. Encuentre la probabilidad de que el valor de la diferencia de las medias muestrales sea menor a 8,2. Por la propiedad reproductiva de la distribución normal sabemos que $(\bar{X}_2 - \bar{X}_1)$ se distribuye normalmente con una media y una varianza igual a:

$$\begin{aligned}\mu_{(\bar{x}_2 - \bar{x}_1)} &= \mu_{x_2} - \mu_{x_1} = 50 - 40 = 10 \\ \sigma_{(\bar{x}_2 - \bar{x}_1)}^2 &= \frac{\sigma_{x_2}^2}{n_2} + \frac{\sigma_{x_1}^2}{n_1} = \frac{9}{5} + \frac{4}{4} = \frac{14}{5} = 2,8 \\ Z &= \frac{(\bar{x}_2 - \bar{x}_1) - \mu_{(\bar{x}_2 - \bar{x}_1)}}{\sigma_{(\bar{x}_2 - \bar{x}_1)}} = \frac{8,2 - 10}{\sqrt{2,8}} = \frac{-1,8}{1,6733} = -1,08\end{aligned}$$

Por lo tanto, la probabilidad buscada es

$$P(\bar{X}_2 - \bar{X}_1 \leq 8,2) = P(Z \leq -1,08) = 0,1401$$

4.3.2.1. La diferencia de medias muestrales y el Teorema del Límite Central

Cuando se desconoce la distribución de la variable, se pueden deducir las propiedades de la distribución de la diferencia de medias muestrales a partir del Teorema del Límite Central. Por lo tanto, si el muestreo se realiza a partir de poblaciones con distribución desconocida y el tamaño de las muestras es grande (n_1 y $n_2 \geq 30$), se aplica el teorema y la distribución de la diferencia de medias muestrales tendrá una media y una varianza igual a:

$$\begin{aligned}\mu_{(\bar{x}_2 - \bar{x}_1)} &= \mu_{\bar{x}_2} - \mu_{\bar{x}_1} = \mu_{x_2} - \mu_{x_1} \\ \sigma_{(\bar{x}_2 - \bar{x}_1)}^2 &= \sigma_{\bar{x}_2}^2 + \sigma_{\bar{x}_1}^2 = \frac{\sigma_{x_2}^2}{n_2} + \frac{\sigma_{x_1}^2}{n_1}\end{aligned}$$

4.4. Métodos Inferenciales

Hasta el momento hemos sentado las bases para el estudio de los fenómenos aleatorios en la naturaleza estableciendo: conceptos básicos para la probabilidad, sus principios y reglas de cálculo; definición del concepto de variable aleatoria y las funciones de probabilidad que de ellas se derivan; los modelos probabilísticos más conocidos y sus limitaciones; y las relaciones existentes entre el universo que describe un fenómeno y las muestras experimentales, de las cuales obtenemos estadísticos que pretenden describir los parámetros reales de una población. La pregunta que surge inmediatamente es: cómo debemos interpretar los valores que brindan los estadísticos obtenidos a partir de las distribuciones muestrales? Para responder a este interrogante contamos con los métodos de inferencia, los cuales sirven para determinar la probabilidad de que cualquier conclusión sobre una población que se haya derivado de la información aportada por un grupo de datos sea correcta. Los valores de los estadísticos muestrales, por muy bueno que haya sido el muestreo, siempre presentarán diferencias con respecto al respectivo valor poblacional o parámetro, debido fundamentalmente a que se está tratando con variables aleatorias que asumen valores distintos y que ocurren en la población con frecuencias diferentes. De modo que al ser imposible eliminar la aleatoriedad y si se quieren hacer generalizaciones a partir de la información obtenida de una muestra se debe establecer la confianza que se tiene en la muestra. Es decir se debe determinar que tan buena es la aproximación entre el valor del estadístico y el valor del parámetro respectivo. En este punto la estadística inferencial es de gran ayuda al ofrecer métodos que cuantifican el grado de confianza requerido para hacer las generalizaciones mencionadas anteriormente. Son dos los métodos de inferencia:

- **Estimación:** usa la información proporcionada por los estadísticos muestrales para estimar con cierta probabilidad el valor de un parámetro poblacional
- **Prueba de Hipótesis:** usa esa misma información para decidir, con una probabilidad conocida, si el parámetro poblacional es igual a algún valor preconcebido.

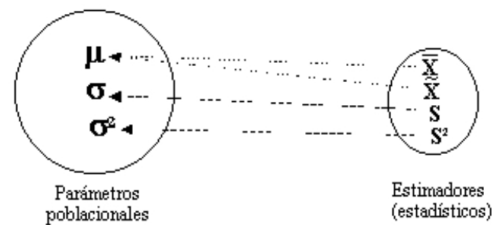
En las siguientes secciones analizaremos en detalle dichos métodos de inferencia estadística.

5. Inf. Est.: Estimación (I)

A continuación empezaremos a desarrollar los métodos inferenciales que nos permitirán cuantificar el grado de confianza que se puede tener de un estadístico, y de esa manera saber cuán acertadas serán nuestras conclusiones sobre los parámetros de la población. Para ello empezaremos con los métodos de estimación, los cuales pueden subdividirse en cuatro categorías: estimación puntual, intervalos de confianza o confianza, histogramas y técnicas de remuestreo. En esta sección nos concentraremos en los dos primeros de los métodos de estimación.

5.1. Estimación puntual

Una estimación puntual consiste en calcular el valor de un estadístico en una muestra, y considerar que el mismo es la mejor aproximación que se tiene a la magnitud del parámetro poblacional correspondiente. Por ejemplo, un valor cualquiera de una media muestral (\bar{x}) es una estimación puntual de la media poblacional (μ). Un mismo parámetro puede tener varios estimadores. Así tenemos que la media poblacional (μ) además de poder ser estimada por la media muestral (\bar{x}), también es estimada por la mediana (\tilde{x}) y por la moda (Mo) para una variable que se distribuye en forma simétrica. Elegir el mejor estimador de un parámetro se dificulta, porque además de existir varios estimadores para un mismo parámetro, ellos son variables aleatorias que pueden tener una amplia distribución de valores. El mejor estimador siempre será aquel que esté más cerca del valor del parámetro que se estima. Como esto no se puede conocer, la calidad de un estimador se debe evaluar en términos de algunas de sus propiedades como son: la insesgabilidad, la consistencia y la eficiencia.



5.1.1. Estimador insesgado

Se dice que un estimador $\hat{\theta}$ del parámetro θ es insesgado cuando el valor esperado o promedio de la distribución de $\hat{\theta}$ coincide con el valor del parámetro θ , es decir, $E(\hat{\theta}) = \theta$. Por ejemplo, la media muestral (\bar{x}) es un estimador insesgado de μ , debido a que la media de las medias muestrales $\mu_{\bar{x}}$ es igual a la media poblacional μ_x , es decir, $E(\bar{x}) = \mu_{\bar{x}} = \mu_x$. Igualmente, la mediana de una muestra (\tilde{x}) es un estimador insesgado de μ , porque la media de las medianas muestrales es igual a la media poblacional, cuando la distribución de la variable estudiada es simétrica, $E(\tilde{x}) = \mu_x$.

En cambio la varianza muestral puede ser un estimador sesgado si para su cálculo se usan n grados de libertad, es decir,

$$E(S^2) \neq \sigma^2 \quad \text{si} \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Esto puede demostrarse fácilmente. Haciendo el cálculo tenemos

$$E(S^2) = E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right] = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}^2)$$

como

$$\begin{aligned} \sigma^2 &= E(X_i^2) - E(X_i)^2 = E(X_i^2) - \mu^2 \longrightarrow E(X_i^2) = \sigma^2 + \mu^2 \\ \frac{\sigma^2}{n} &= E(\bar{X}^2) - E(\bar{X})^2 = E(\bar{X}^2) - \mu^2 \longrightarrow E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

nos queda que

$$E(S^2) = \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n} \sigma^2$$

Para hacer insesgada la varianza muestral, la misma debe calcularse multiplicándola por $n/(n-1)$, es decir

$$S^2 = \frac{n}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

esto es lo mismo que decir que estamos usando $n-1$ grados de libertad, de modo que ahora $E(S^2) = \sigma^2$.

5.1.2. Estimador consistente

Se dice que un estimador $\hat{\theta}$ del parámetro θ es consistente si el valor absoluto de la diferencia entre los valores del estimador y del parámetro es menor a medida que aumenta el tamaño de la muestra (n). Es decir,

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| \leq \epsilon) = 1 \quad \forall \epsilon > 0$$

Sabemos que la media y la mediana muestrales son estimadores insesgados de μ , pero, son igualmente consistentes?. La respuesta es afirmativa si la distribución de la variable estudiada es simétrica. Pero si la variable se distribuye asimétricamente la mediana muestral se aproximará más al valor de la mediana poblacional cuando n aumenta y la media muestral se acercará más a la media poblacional (μ). Recordemos que la media poblacional y la mediana poblacional son dos parámetros diferentes. De lo dicho anteriormente se puede concluir que la media muestral es más consistente que la mediana muestral como estimador de la media poblacional (μ).

5.1.3. Estimador eficiente

Se dice que un estimador $\hat{\theta}_1$ del parámetro θ es el más eficiente si no existe otro estimador $\hat{\theta}_2$ cuya varianza sea menor a la de θ , es decir

$$E[(\hat{\theta}_1 - \theta)^2] < E[(\hat{\theta}_2 - \theta)^2]$$

Si continuamos con la comparación entre media y mediana muestral como estimadores de μ , es necesario determinar para el caso de poblaciones con una distribución simétrica, cual de los dos estadísticos es mejor estimador de la media poblacional. Por lo tanto es necesario usar otras propiedades diferentes a la insesgabilidad y la consistencia. Cuando se examina la eficiencia de los dos estimadores, se encuentra que la varianza de la media muestral es menor que la varianza de la mediana muestral: $\sigma_{\bar{x}}^2 = 1,57\sigma_x^2$. Por lo tanto, en función de la insesgabilidad, consistencia y eficiencia, la media muestral (\bar{x}) es un mejor estimador de μ que la mediana muestral (\tilde{x}) para variables con distribuciones tanto simétricas como asimétricas.

5.2. Intervalos de confianza (IC)

Aunque un estimador como la media muestral sea insesgado, consistente y eficiente, lo más probable es que, aún en muestras grandes, el valor del estimador ($\hat{\theta}$) no coincida con el valor del parámetro (θ). Por lo tanto se utiliza otro procedimiento más seguro para inferir el valor del parámetro, como es la estimación por intervalo. Con este método se construye un intervalo a partir del valor de un estimador puntual ($\hat{\theta}$), mediante la definición de dos límites: uno superior (*LS*) y otro inferior (*LI*). Se supone que el intervalo contiene el parámetro poblacional (θ) con cierta probabilidad.

5.2.1. IC para una media poblacional

La deducción de un intervalo de confianza para la media poblacional depende de varios aspectos que combinados de cierta manera conforman una situación particular que determina la forma del intervalo. Los aspectos a considerar en la construcción de un intervalo de confianza son:

- el tipo de distribución de la variable estudiada,
- el conocimiento de la varianza poblacional, y
- el tamaño de la muestra.

A continuación estudiaremos las distintas situaciones o casos que se pueden presentar en el desarrollo de un intervalo de confianza.

5.2.1.1. Caso 1: Muestreo en una población distribuida normalmente y con varianza conocida

Supóngase que se desea estimar el valor de la media poblacional de una variable que se distribuye normalmente con varianza conocida (σ_x^2), para lo cual se extrae una muestra de tamaño n y se calcula la media de la muestra (\bar{x}). El valor de \bar{x} es uno del total que conforman la población de valores de la variable aleatoria \bar{X} que como se sabe se distribuye normalmente alrededor de una media $\mu_{\bar{x}}$ con varianza σ_x^2/n . En esta población se pueden encontrar dos valores \bar{x}_1 y \bar{x}_2 separados simétricamente de $\mu_{\bar{x}}$ que definen un intervalo dentro del cual queda incluida una proporción $(1-\alpha)$ del total de valores de \bar{X} . Los valores \bar{x}_1 y \bar{x}_2 se encuentran transformando la variable \bar{X} en la variable Z , es decir

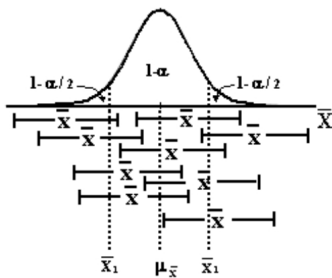
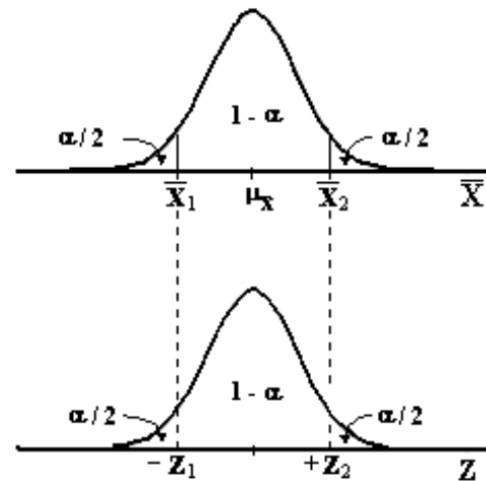
$$-z_1 = -z_{(1-\alpha/2)} = \frac{\bar{x}_1 - \mu_{\bar{x}}}{\sigma_x/\sqrt{n}}$$

$$+z_2 = +z_{(1-\alpha/2)} = \frac{\bar{x}_2 - \mu_{\bar{x}}}{\sigma_x/\sqrt{n}}$$

donde, por ejemplo, $+z_{(1-\alpha/2)}$ es el valor de Z a la izquierda del cual se encuentra una fracción del área igual a $1 - \alpha/2$. Estos valores de Z se encuentran en la tabla de la distribución acumulada de Z , por lo que despejando, los valores que necesitamos son

$$\bar{x}_1 = \mu_{\bar{x}} - z_{(1-\alpha/2)}\sigma_x/\sqrt{n} \quad y \quad \bar{x}_2 = \mu_{\bar{x}} + z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$$

Los valores \bar{x}_1 y \bar{x}_2 representan el límite inferior y superior del intervalo que contiene el $(1 - \alpha)100\%$ de los valores de \bar{X} .



límites del intervalo $[\mu_{\bar{x}} \pm z_{(1-\alpha/2)}\sigma_x/\sqrt{n}]$ (ver figura de la izquierda). Solamente aquellos intervalos generados a partir de aquellas pocas medias muestrales que se encuentran muy alejados de la media poblacional no incluyen a esta última.

De modo que un intervalo de la forma $[\bar{x} \pm z_{(1-\alpha/2)}\sigma_x/\sqrt{n}]$ recibe el nombre de intervalo de confianza del $(1 - \alpha)100\%$. Los valores extremos se denominan límites de confianza, existiendo un límite superior (LS= $\bar{x} + z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$) y un límite inferior (LI= $\bar{x} - z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$). El término $z_{(1-\alpha/2)}$ recibe el nombre de coeficiente de confiabilidad. La fracción $1 - \alpha$ se denomina nivel de confianza y representa la probabilidad de que el intervalo contenga el parámetro poblacional. Consecuentemente, α representa la probabilidad de que el intervalo no contenga el parámetro poblacional.

Observar que, a mayor amplitud del intervalo, aumenta la probabilidad de que el parámetro esté incluido dentro del intervalo dado, pero también es mayor la incertidumbre sobre el valor del parámetro. Lo ideal sería construir intervalos estrechos con un alto nivel de confianza. Cuando en una situación real se construye un intervalo de confianza, la media poblacional puede o no estar incluida dentro del intervalo. Sin embargo existe una probabilidad igual a $1 - \alpha$ de que el parámetro quede incluido. Otra forma de decirlo, si se construyen infinidad de intervalos similares, el $(1 - \alpha)100\%$ de los mismos contendrán a la media poblacional. Es importante advertir que es un error generalizado el señalar que la media poblacional se encuentra entre los valores de los límites del IC, porque la media poblacional como cualquier otro parámetro es un valor fijo, y la afirmación anterior sugiere que el parámetro puede asumir cualquier valor entre los dos límites con cierta probabilidad. Si se analiza con un poco más de detalle la relación entre los intervalos construidos a partir de las medias muestrales y la media poblacional, se observa que ambas cantidades se encuentran alejadas cierta distancia ϵ .

La distancia ϵ se denomina error de estimación. Para que un intervalo contenga a la media poblacional con una probabilidad igual a $1 - \alpha$, ese error debe ser menor a la distancia $z_{(1-\alpha/2)}\sigma_x/\sqrt{n}$, con lo cual el módulo de dicha distancia queda definido como el error máximo (ϵ_m). Una consecuencia directa de conocer ϵ_m es que permite determinar cuál debe ser el tamaño muestral adecuado para cometer ese error máximo un $(1 - \alpha)100\%$ de las veces, es decir

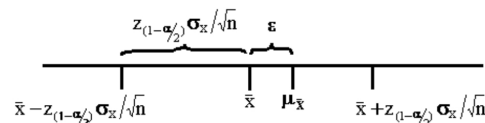
$$n = \left(\frac{z_{(1-\alpha/2)}\sigma_x}{\epsilon_m} \right)^2$$

Ejemplo: Al examinar 9 porciones de agua se encontró una concentración promedio de ión nitrato igual a $0,5 \mu\text{g}/\text{ml}$. Se desea estimar mediante un intervalo de confianza del 95% la concentración promedio del nitrato en el agua, si se sabe que la desviación del método para éste análisis es de $0,15 \mu\text{g}/\text{ml}$.

El intervalo que se quiere es de la forma $[\bar{x} \pm z_{(1-\alpha/2)}\sigma_x/\sqrt{n}]$ teniendo como límites los valores siguientes:

$$LI = \bar{x} - z_{(0,975)}\sigma_x/\sqrt{n} = 0,5 - 1,96(0,15/\sqrt{9}) = 0,4020 \mu\text{g}/\text{ml}$$

$$LS = \bar{x} + z_{(0,975)}\sigma_x/\sqrt{n} = 0,5 + 1,96(0,15/\sqrt{9}) = 0,5980 \mu\text{g}/\text{ml}$$



Entonces el intervalo buscado es $[0,4020; 0,5980]$. Se concluye que se tiene un 95 % de confianza de que la concentración promedio del ión nitrato en el agua se encuentre incluida dentro de este intervalo.

También se puede decir que el error máximo de estimación con un 95 % de confianza es:

$$\epsilon_m = \left| z_{(1-\alpha/2)} \sigma_x / \sqrt{n} \right| = \left| 1,96(0,15/\sqrt{9}) \right| = 0,098 \mu_g/ml$$

Ahora bien, si se desea aumentar el nivel de confianza, por ejemplo a un 99 %, sin aumentar el error de estimación, el tamaño de la muestra debe ser igual a:

$$n = \left(\frac{z_{(0,995)} \sigma_x}{\epsilon_m} \right)^2 = \left(\frac{2,58(0,15)}{0,098} \right)^2 = 16$$

Por otra parte, si se quiere reducir el error de estimación a unos $0,05 \mu_g/ml$, manteniendo el nivel de confianza del 95 %, entonces el tamaño muestral debe ser

$$n = \left(\frac{z_{(0,975)} \sigma_x}{\epsilon_m} \right)^2 = \left(\frac{1,96(0,15)}{0,05} \right)^2 = 35$$

5.2.1.2. Caso 2: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra grande ($n \geq 30$) .

La situación más común cuando se trata de estimar el valor de una media poblacional mediante un intervalo de confianza es que no slo se desconoce el valor de μ sino también el de la varianza poblacional σ_x^2 . Cuando se presenta una situación como la descrita, se puede utilizar la varianza de la muestra (S_x^2) como una estimación puntual de la varianza poblacional (σ_x^2). Si el tamaño de la muestra es grande ($n \geq 30$), el estadístico $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ se distribuye normalmente, quedando el intervalo de confianza de la forma $[\bar{x} \pm z_{(1-\alpha/2)} S_x/\sqrt{n}]$.

5.2.1.3. Caso 3: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra pequeño ($n < 30$) .

Una nueva situación se presenta si de una población que se distribuye normalmente con varianza desconocida se toma una muestra pequeña ($n < 30$). En éste caso, S_x ya no es un buen estimador de σ_x y el estadístico $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ no se distribuye normalmente. Afortunadamente, existe otro modelo que describe su distribución de probabilidades, conocido como distribución de T o de *Student*. En este caso, se dice que la variable $(\bar{x} - \mu_{\bar{x}})/(S_x/\sqrt{n})$ se distribuye como T con $n - 1$ grados de libertad. El intervalo de confianza vendrá dado por la expresión

$$[\bar{x} \pm t_{(1-\alpha/2;n-1)} S_x/\sqrt{n}]$$

donde $t_{(1-\alpha/2;n-1)}$ es el valor de T a la izquierda del cual se encuentra el $(1 - \alpha/2)100\%$ de los valores de T .

5.2.1.4. Distribución de T

La distribución de Student fue descrita en 1908 por William Sealy Gosset. Recordemos que si tenemos X_1, \dots, X_n variables aleatorias independientes distribuidas normalmente, con media μ y varianza σ^2 , entonces la distribución de las medias muestrales también se distribuye normalmente con media μ y varianza σ^2/n . Entonces

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

sigue una distribución normal de media 0 y varianza 1. Sin embargo, dado que la desviación estándar no siempre es conocida de antemano, Gosset estudió un cociente relacionado,

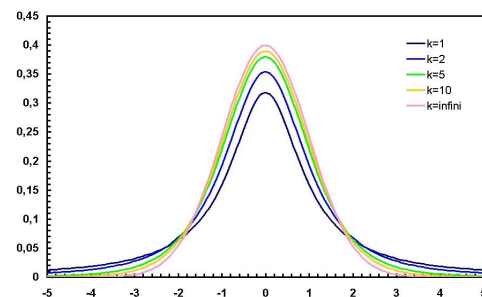
$$T = \frac{\bar{X} - \mu}{S_x/\sqrt{n}} \quad \text{donde} \quad S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

es la varianza muestral, y demostró que la función distribución de T es

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi} \Gamma(\nu/2)} (1+t^2/\nu)^{-(\nu+1)/2}$$

donde ν es igual a $n-1$. La distribución de T se llama ahora la distribución-t de Student. Gosset trabajaba en una fábrica de cerveza Guinness que prohibía a sus empleados la publicación de artículos científicos debido a una difusión previa de secretos industriales. De ahí que Gosset publicase sus resultados bajo el seudónimo de Student. El parámetro ν representa el número de grados de libertad. La distribución depende de ν , pero no de μ o σ , lo cual es muy importante en la práctica.

En la figura de la derecha pueden verse varias distribuciones T con distintos grados de libertad (k en el caso de la figura). La distribución T se caracteriza por: tomar valores entre $(-\infty, +\infty)$; los valores de T se distribuyen simétricamente alrededor de la media $\mu = 0$; y su forma es parecida a la distribución normal pero más prominente y con colas más elevadas. Es importante tener en mente que cuando el número de grados de libertad es grande, la distribución T tiende a una distribución normal (como era de esperarse). Para cada valor de ν existe una



distribución T . Las tablas de la distribución acumulativa de T tienen como entradas los grados de libertad y la probabilidad de tener un valor menor a t . Cualquier valor de t se identifica de la siguiente manera: $t_{(1-\alpha; n-1)}$. Por ejemplo $t_{(0,975;6)} = 2,447$ (ver tabla) es el valor de t a la izquierda del cual se encuentra una proporción del área igual a 0.975 con 6 grados de libertad, o dicho de otra manera: existe una probabilidad igual a 0.975 de encontrar un valor igual o menor a $t=2.447$ para 6 grados de libertad.

ν	$1-\alpha$	$t_{(0,90)}$	$t_{(0,95)}$	$t_{(0,975)}$	$t_{(0,99)}$	$t_{(0,995)}$	$t_{(0,9975)}$	$t_{(0,999)}$	$t_{(0,9995)}$
1		3,078	6,314	12,706	31,821	63,656	127,321	318,289	636,578
2		1,886	2,920	4,303	6,965	9,925	14,089	22,328	31,600
3		1,638	2,353	3,182	4,541	5,841	7,453	10,214	12,924
4		1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5		1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6		1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7		1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408

5.2.1.5. Caso 4: Muestreo a partir de una población con distribución desconocida, con varianza conocida y tamaño de muestra grande ($n \geq 30$) .

Cuando se desconoce la forma de la distribución de valores de una variable no se puede predecir como será la distribución de la media muestral, a menos que el tamaño de la muestra sea grande. Si este es el caso, es decir, $n \geq 30$, entonces es aplicable el Teorema del Límite Central y la variable \bar{X} tenderá a distribuirse normalmente con media $\mu_{\bar{x}} = \mu_x$ y varianza $\sigma_{\bar{x}}^2 = \sigma_x^2/n$, de modo que el intervalo de confianza será de la forma $[\bar{x} \pm z_{(1-\alpha/2)}\sigma_x/\sqrt{n}]$.

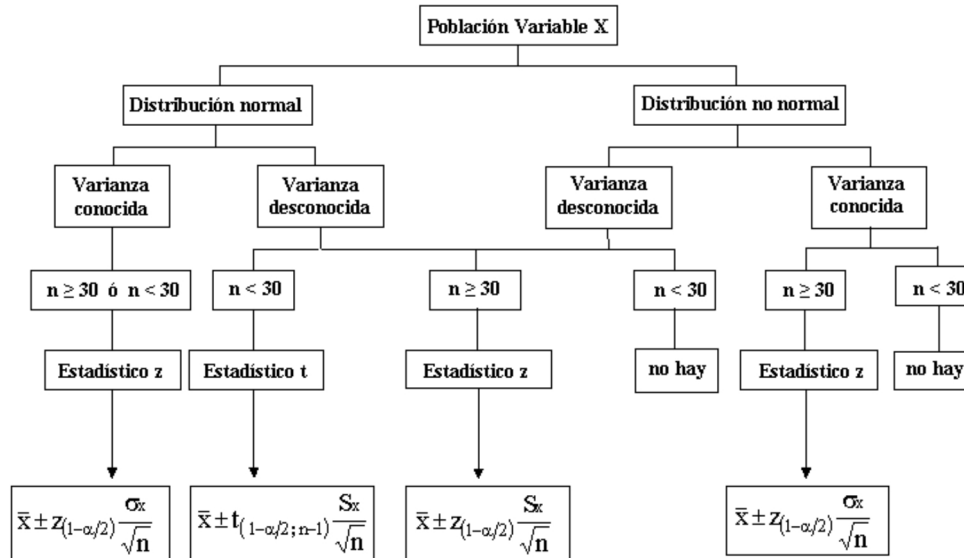
5.2.1.6. Caso 5: Muestreo a partir de una población con distribución y varianza desconocida y tamaño de muestra grande ($n \geq 30$) .

Como en el caso anterior al ser $n \geq 30$, es aplicable el Teorema del Límite Central por lo que la media muestral se distribuye normalmente. La varianza de la muestra S_x^2 se usa como estimador de σ_x^2 y el intervalo de confianza será de la forma $[\bar{x} \pm z_{(1-\alpha/2)}S_x/\sqrt{n}]$.

5.2.1.7. Caso 6: Muestreo a partir de una población con distribución desconocida y tamaño de muestra pequeño ($n < 30$) .

Cuando no se conoce la distribución de la variable y el tamaño de la muestra es pequeño ($n < 30$), no es posible predecir la distribución que asume la media muestral. Por lo tanto, no se puede construir un intervalo de confianza, a menos que los datos sean transformados y se logren aproximar a una distribución normal.

A continuación se presenta un esquema con la combinación de los diferentes aspectos que determinan la construcción de un IC.



5.2.2. IC para la diferencia de dos medias poblacionales

Al igual que en la estimación de una media poblacional a través de la construcción de un intervalo, para estimar la diferencia de medias poblacionales es necesario considerar el tipo de distribución de la variable, el conocimiento de las varianzas poblacionales y el tamaño de las muestras.

5.2.2.1. Caso 1: Muestreo a partir de poblaciones distribuidas normalmente y con varianzas conocidas

Recordemos que cuando se hace un muestreo de dos poblaciones distribuidas normalmente, se puede generar una nueva variable conocida como diferencia de medias muestrales, cuya distribución de valores se caracteriza por tener también una distribución normal, siendo su media y varianza las siguientes:

$$\mu_{(\bar{x}_2 - \bar{x}_1)} = \mu_{x_2} - \mu_{x_1} \quad \sigma_{(\bar{x}_2 - \bar{x}_1)}^2 = \frac{\sigma_{x_1}^2}{n_1} + \frac{\sigma_{x_2}^2}{n_2}$$

La deducción del intervalo de confianza para la diferencia de medias poblacionales se puede comenzar estableciendo que la probabilidad de que la variable $\bar{X}_2 - \bar{X}_1$ se encuentre entre dos valores cualquiera es igual a $1 - \alpha$.

$$P [(\bar{x}_2 - \bar{x}_1)_1 \leq \bar{X}_2 - \bar{X}_1 \leq (\bar{x}_2 - \bar{x}_1)_2] = 1 - \alpha$$

Ésta es la misma probabilidad de que la variable Z se encuentre entre dos valores

$$P [z_1 \leq Z \leq z_2] = P [-z_{(1-\alpha/2)} \leq Z \leq +z_{(1-\alpha/2)}] = 1 - \alpha$$

ahora si $Z = \frac{(\bar{x}_2 - \bar{x}_1) - \mu_{(\bar{x}_2 - \bar{x}_1)}}{\sigma_{(\bar{x}_2 - \bar{x}_1)}}$, haciendo un poco de álgebra nos queda

$$P [(\bar{x}_2 - \bar{x}_1) - z_{(1-\alpha/2)}\sigma_{(\bar{x}_2 - \bar{x}_1)} \leq \mu_{(\bar{x}_2 - \bar{x}_1)} \leq (\bar{x}_2 - \bar{x}_1) + z_{(1-\alpha/2)}\sigma_{(\bar{x}_2 - \bar{x}_1)}] = 1 - \alpha$$

sustituyendo $\sigma_{(\bar{x}_2 - \bar{x}_1)} = \sqrt{\frac{\sigma_{x_1}^2}{n_1} + \frac{\sigma_{x_2}^2}{n_2}}$ nos queda que el IC para estimar la diferencia entre dos medias poblacionales tiene la forma general

$$\left[(\bar{x}_2 - \bar{x}_1) \pm z_{(1-\alpha/2)} \sqrt{\frac{\sigma_{x_1}^2}{n_1} + \frac{\sigma_{x_2}^2}{n_2}} \right]$$

Ejemplo: En un trabajo de investigación se encontró que el contenido promedio de ácido úrico en 12 niños con el Síndrome de Down fue de 4,75 mg/100ml, mientras que en 18 niños normales el valor promedio encontrado fue de 3,95 mg/100 ml.. Mediante trabajos previos se había determinado que las varianzas de ambos grupos son 1,02 y 0,98 respectivamente. Suponiendo que la concentración de ácido úrico es una variable que se distribuye normalmente, construya un intervalo de confianza del 98 % para la diferencia de medias poblacionales.

Si las muestras provienen de poblaciones distribuidas normalmente y con varianza conocida, y el nivel de confianza $1 - \alpha = 0,98$, entonces sus límites son los siguientes:

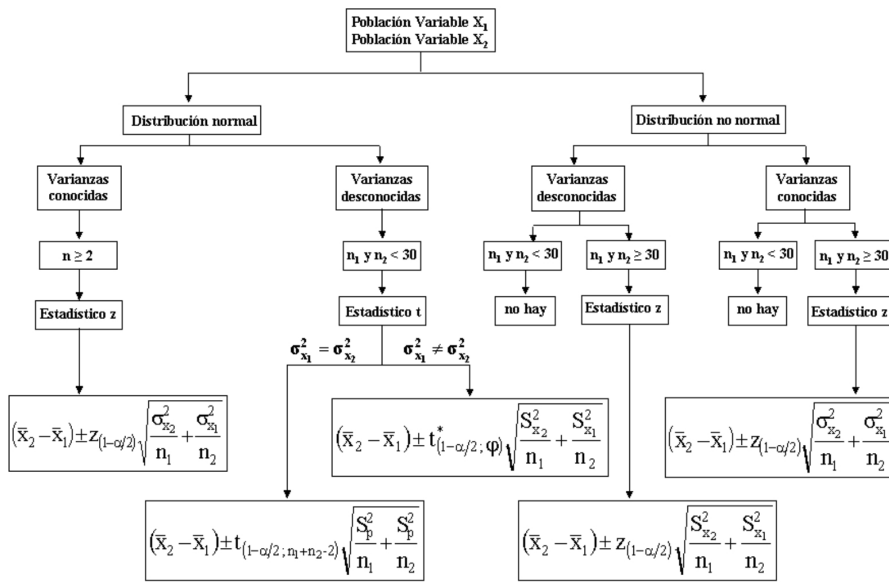
$$LI = \left[(\bar{x}_2 - \bar{x}_1) - z_{(0,99)} \sqrt{\frac{\sigma_{x_1}^2}{n_1} + \frac{\sigma_{x_2}^2}{n_2}} \right] = \left[0,80 - 2,33 \sqrt{\frac{1,02}{12} + \frac{0,98}{18}} \right] = 0,1099$$

$$LS = \left[(\bar{x}_2 - \bar{x}_1) + z_{(0,99)} \sqrt{\frac{\sigma_{x_1}^2}{n_1} + \frac{\sigma_{x_2}^2}{n_2}} \right] = \left[0,80 + 2,33 \sqrt{\frac{1,02}{12} + \frac{0,98}{18}} \right] = 1,8501$$

El intervalo buscado es [0.1099;1.8501]. Se concluye que se tiene un 98 % de confianza de que el valor de la diferencia de medias poblacionales sea un punto dentro de ese intervalo.

5.2.2.2. Otros casos

Los otros tipos de intervalos de confianza para la diferencia de medias poblacionales que resultan de la combinación de varias situaciones se muestran en el siguiente esquema.



5.2.2.3. IC para $\mu_{x_2} - \mu_{x_1}$ y el estadístico t .

Del esquema anterior puede verse que el uso del estadístico t está condicionado por la suposición de que dichas varianzas sean iguales o diferentes. De modo que la primera tarea antes de construir un intervalo, es determinar cuál de las dos situaciones se tiene: varianzas iguales y desconocidas o diferentes y desconocidas. Se puede establecer una regla práctica que permita decidir rápidamente esta cuestión. Lo primero que se debe hacer es calcular la razón de varianzas RV , como el cociente entre la varianza muestral mayor y la varianza muestral menor.

$$RV = \frac{s_1^2}{s_2^2} \quad \text{si} \quad s_1^2 > s_2^2$$

Luego se toma una decisión sobre la base de las siguientes reglas: Las varianzas son diferentes si

$$\alpha = 0,10 \quad y \quad RV > 2,0$$

$$\alpha = 0,05 \quad y \quad RV > 2,5$$

$$\alpha = 0,01 \quad y \quad RV > 3,5$$

5.2.2.3.1. Varianzas Iguales .

Cuando se acepta la suposición que las dos varianzas poblacionales aunque desconocidas son iguales, se pueden promediar las varianzas de las muestras para hacer una mejor estimación de la varianza poblacional. Para obtener el promedio, el valor de las varianzas muestrales debe ser ponderado por el tamaño de las muestras de acuerdo a la siguiente fórmula:

$$S_p^2 = \frac{(n_1 - 1)S_{x_1}^2 + (n_2 - 1)S_{x_2}^2}{n_1 + n_2 - 2}$$

De manera que la desviación de la distribución de diferencias de medias muestrales queda como

$$S_{(\bar{x}_2 - \bar{x}_1)} = \sqrt{\frac{S_p^2}{n_2} + \frac{S_p^2}{n_1}}$$

y el intervalo de confianza es

$$\left[(\bar{x}_2 - \bar{x}_1) \pm t_{(1-\alpha/2; n_1+n_2-2)} \sqrt{\frac{S_p^2}{n_2} + \frac{S_p^2}{n_1}} \right]$$

5.2.2.3.2. Varianzas diferentes

Si se asume que las varianzas de dos poblaciones, de una variable que se distribuye normalmente, son diferentes aunque desconocidas, no se puede usar el estadístico

$$\frac{(\bar{x}_2 - \bar{x}_1) - (\mu_{x_2} - \mu_{x_1})}{\sqrt{\frac{S_{x_2}^2}{n_2} + \frac{S_{x_1}^2}{n_1}}}$$

para calcular el coeficiente de confiabilidad $t_{(1-\alpha/2)}$, porque su distribución no sigue el modelo de distribución T . Sin embargo es posible calcular un nuevo coeficiente de confiabilidad $t_{(1-\alpha/2)}^*$, usando la siguiente fórmula

$$t_{(1-\alpha/2)}^* = \frac{w_1 t_{(1-\alpha/2; n_1-1)} + w_2 t_{(1-\alpha/2; n_2-1)}}{w_1 + w_2} \quad \text{con} \quad w_1 = \frac{s_1^2}{n_1} \quad \text{y} \quad w_2 = \frac{s_2^2}{n_2}$$

por lo que el intervalo de confianza se obtiene haciendo

$$\left[(\bar{x}_2 - \bar{x}_1) \pm t_{(1-\alpha/2)}^* \sqrt{\frac{S_{x_2}^2}{n_2} + \frac{S_{x_1}^2}{n_1}} \right]$$

Ejemplo: Al comparar dos métodos para determinar la concentración de boro en un material vegetal se efectuaron varias mediciones que figuran en la siguiente tabla.

Construya un intervalo de confianza del 99% para $\mu_{x_2} - \mu_{x_1}$. Suponga que la variable concentración se distribuye normalmente.

Las condiciones del problema indican que las muestras son pequeñas y provienen de dos poblaciones que se distribuyen nor-

	Concentración de Boro	
	Espectrofotometría	Fluorimetría
n	10	16
Media	26.00 $\mu\text{g}/\text{l}$	28.00 $\mu\text{g}/\text{l}$
Desviación	0.23 $\mu\text{g}/\text{l}$	1.30 $\mu\text{g}/\text{l}$

malmente y con varianzas desconocidas. Para escoger el intervalo adecuado, es necesario decidir si las desconocidas varianzas poblacionales son iguales o diferentes. Como $\alpha = 0,01$ y $RV = s_1^2/s_2^2 = (1,3)^2/(0,23)^2 = 31,9$ es mayor a 3.5 se acepta que las dos varianzas son diferentes. Por lo tanto el intervalo a construir debe ser el siguiente:

$$\left[(\bar{x}_2 - \bar{x}_1) \pm t_{(1-\alpha/2)}^* \sqrt{\frac{S_{x_2}^2}{n_2} + \frac{S_{x_1}^2}{n_1}} \right]$$

El primer paso es calcular el coeficiente de confiabilidad $t_{(1-\alpha/2)}^*$. Sabiendo que

$$t_{(1-\alpha/2; n_1-1)} = t_{(0,995; 9)} = 3,25 \quad \text{y} \quad t_{(1-\alpha/2; n_2-1)} = t_{(0,995; 15)} = 2,947$$

$$w_1 = \frac{s_1^2}{n_1} = \frac{(0,23)^2}{10} = 0,0053 \quad \text{y} \quad w_2 = \frac{s_2^2}{n_2} = \frac{(1,30)^2}{16} = 0,1056$$

el valor del coeficiente de confiabilidad será:

$$t_{(1-\alpha/2)}^* = \frac{w_1 t_{(1-\alpha/2; n_1-1)} + w_2 t_{(1-\alpha/2; n_2-1)}}{w_1 + w_2} = \frac{(0,0053) 3,25 + (0,1056) 2,947}{0,0053 + 0,1056} = 2,96$$

Con lo cual, los límites del intervalo de confianza son los siguientes:

$$LI = \left[(\bar{x}_2 - \bar{x}_1) - t_{(1-\alpha/2)}^* \sqrt{\frac{S_{x_2}^2}{n_2} + \frac{S_{x_1}^2}{n_1}} \right] = 2 - 2,96 \sqrt{\frac{(1,30)^2}{16} + \frac{(0,23)^2}{10}} = 2 - 0,9858 = 1,0142$$

$$LS = \left[(\bar{x}_2 - \bar{x}_1) + t_{(1-\alpha/2)}^* \sqrt{\frac{S_{x_2}^2}{n_2} + \frac{S_{x_1}^2}{n_1}} \right] = 2 + 2,96 \sqrt{\frac{(1,30)^2}{16} + \frac{(0,23)^2}{10}} = 2 + 0,9858 = 2,9858$$

El intervalo buscado es [1.0142;2.9858]. Se concluye que se tiene un 99% de confianza que el intervalo anterior incluya el valor de $\mu_{x_2} - \mu_{x_1}$.

6. Inf. Est.: Estimación (II)

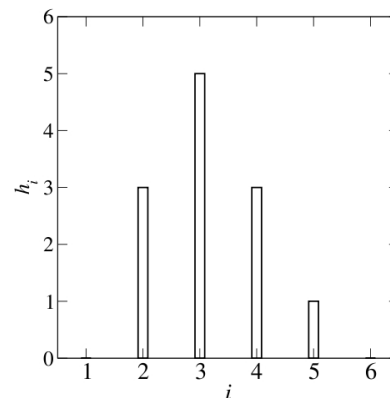
Continuando con los métodos de estimación, en esta sección nos enfocaremos en otros dos métodos que nos ayudaran a visualizar las distribuciones de probabilidad y cuantificar el grado de confianza que se puede tener de un estadístico. Estos métodos son los histogramas y las técnicas de remuestreo.

6.1. Histogramas

Algunas veces uno no solo quiere estimar los momentos de la distribución, si no que se puede querer tener una idea de la distribución completa. En este caso uno puede hacer uso de los histogramas.

6.1.1. Definición

Un histograma viene dado por el conjunto de intervalos disjuntos $B_k = [l_k, u_k)$ los cuales son los llamados *bines* y los contadores h_k de cada bin. Para una dada muestra de n puntos medidos, el valor h_k del bin contiene el número de puntos de la muestra que están contenidos en B_k . En principio, los bins pueden ser elegidos arbitrariamente. Lo que hay que tener en cuenta es que la unión de todos los intervalos cubra todos los puntos posibles de la muestra. Observar que el ancho $b_k = u_k - l_k$ de cada bin puede ser diferente. Sin embargo, frecuentemente se usan bins con anchos uniformes. Además, para varias aplicaciones, por ejemplo, cuando se consideran diferentes métodos de asignación de pesos a diferentes puntos, es útil considerar los contadores como valores variables reales.



6.1.2. Intervalo de confianza

Formalmente, para una dada variable X , el recuento h_k del bin k puede ser visto como un experimento aleatorio para una variable aleatoria binomial $H_k \sim B(n, p_k)$ con parámetros n y p_k , donde $p_k = P(X \in B_k)$ es la probabilidad de que el experimento aleatorio para X resulte un valor que está contenido en B_k . Esto significa que el intervalo de confianza para un bin puede ser obtenido, en principio, a partir de una distribución binomial. Sin embargo, para cada muestra el verdadero valor de p_k es desconocido y solo puede estimarse por $q_k = h_k/n$. Por lo tanto, la verdadera distribución binomial es desconocida. Por otro lado, una variable aleatoria binomial es la suma de n variables aleatorias de Bernoulli con parámetro

p_k . Entonces, el estimador q_k es la media muestral para una variable aleatoria de Bernoulli. Si el número de puntos de la muestra es grande, a partir del Teorema del Límite Central, la distribución de las medias muestrales (la cual de hecho es binomial) es aproximadamente normal o gaussiana. Por lo tanto, uno puede usar el intervalo de confianza estándar

$$P(q_k - zS/\sqrt{n} \leq p_k \leq q_k + zS/\sqrt{n}) \simeq 1 - \alpha$$

Recordar que una variable aleatoria de Bernoulli tiene una varianza muestral igual a $s^2 = q_k(1 - q_k) = (h_k/n)(1 - h_k/n)$.

Ahora, surge la pregunta: que es suficientemente "grande" como para que se pueda confiar en los IC estimados a partir de una gaussiana? Consideremos, por ejemplo, que no se encuentra ningún punto en cierto bin B_k . Esto puede pasar fácilmente en regiones donde p_k es más pequeño que $1/n$ pero distinto de cero, es decir, en las regiones del histograma que se usan para mostrar las colas de la función distribución de probabilidades. En ese caso, la fracción estimada es $q_k = 0$ con un intervalo de confianza de ancho cero, lo cual es ciertamente equivocado. Esto significa que el número de muestras n necesario para tener un IC creíble para el bin B_k depende del número de entradas en los bins. Una regla puesta a dedo por los estadistas es que se debe cumplir que $nq_k(1 - q_k) > 9$. Si esto no se cumple, el IC correcto $[q_{i,l}; q_{i,u}]$ para q_k , tiene que ser obtenido a partir de la distribución binomial y esto es bastante complicado, ya que hace uso de una nueva distribución denominada *distribución F*. Esta distribución de probabilidades viene descrita por la siguiente función

$$f(x) = d_1^{d_1/2} d_2^{d_2/2} \frac{\Gamma(d_1/2 + d_2/2)}{\Gamma(d_1/2)\Gamma(d_2/2)} \frac{x^{d_1/2-1}}{(d_1x + d_2)^{d_1/2+d_2/2}}$$

para $x > 0$ y $f(x) = 0$ para $x \leq 0$. Los parámetros d_1 y d_2 son los grados de libertad que describen a la variable X .

Volviendo a nuestro problema, si calculamos las distribuciones acumuladas correspondientes a la distribución F como

$$F_1 = F(1 - \alpha/2; 2n - 2h_k + 2, 2h_k) \quad \text{y} \quad F_2 = F(1 - \alpha/2; 2h_k + 2, 2n - 2h_k)$$

donde $F(\beta; r_1, r_2)$ establece el valor x tal que la función distribución para una función F con grados de libertad r_1 y r_2 , alcance el valor β , entonces, el IC buscado puede calcularse así:

$$q_{i,l} = \frac{h_k}{h_k + (n - h_k + 1)F_1} \quad \text{y} \quad q_{i,u} = \frac{(h_k + 1)F_2}{(h_k + 1)F_2 + n - h_k}$$

Si siempre se usaran estos IC, los cuales son antisimétricos respecto a q_k , uno no se equivocaría nunca. Sin embargo, para la mayoría de las aplicaciones las barras de error gaussianas funcionan bien.

6.1.3. Histogramas para variables continuas

Finalmente, en caso de que se quiera usar un histograma para representar una muestra extraída a partir de una variable aleatoria continua, se puede interpretar al histograma como una muestra para la función distribución de probabilidades, que puede representarse por el conjunto de pares $(x_k, p(x_k))$. Para simplificar, se asume que los puntos medio de cada intervalo son usados como coordenadas x . Para la normalización, se debe dividir por el número total de recuentos (como se hizo con $q_k = h_k/n$) y por el ancho del bin. Esto asegura que la integral del histograma, aproximada por la suma de los intervalos, dé como resultado la unidad. Por lo tanto tendremos que

$$x_k = (l_k + u_k)/2$$

$$p(x_k) = h_k/(nb_k)$$

El intervalo de confianza, cualquiera sea el tipo que se elija, debe ser normalizado de la misma manera. Para variables aleatorias discretas, el histograma puede ser usado para estimar la función distribución de probabilidades. En este caso, la elección de los bins, en particular su ancho, es fácil, ya que todos los posibles resultados de un experimento aleatorio son conocidos. En un histograma para variables continuas, la elección del ancho de los bins es importante. Básicamente, se debe ajustar el ancho manualmente, de tal manera que los datos esten representados lo mejor posible. Por lo tanto, el ancho de los bins no debe ser ni muy chico ni muy grande. Algunas veces elegir anchos no uniformes para los bins es lo mejor. Una manera apropiada de proceder sería tratar de que el ancho de los bins sea lo suficientemente grande en aquellos bins donde el número de puntos muestrados es pobre. Esto sugiere que cada bin debería contener aproximadamente el mismo número de puntos muestrados. Un ejemplo de regla para asignar ancho a los bins es $b = 3,49n^{1/3}$, la cuál proviene de minimizar la media integrada de las diferencias cuadradas entre una distribución gaussiana y un muestra extraída a partir de ella. En consecuencia, mientras más grande la varianza S de la muestra, más grande será el ancho del bin, por otro lado, incrementar el número de puntos en la muestra permite que el ancho del bin se reduzca.

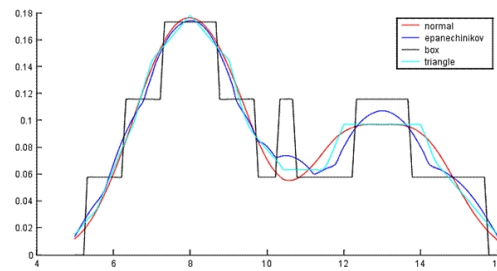
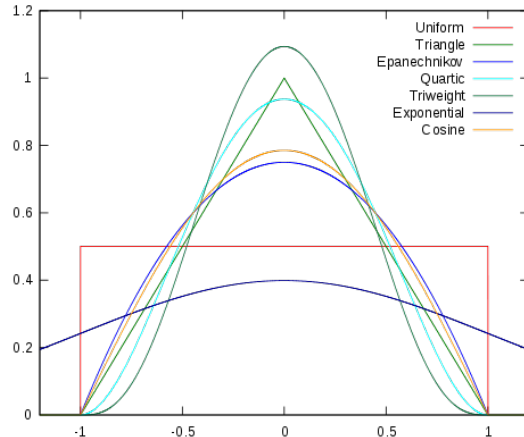
6.1.4. Funciones "kernel" para histogramas de variables continuas

Debe tenerse en cuenta que, los histogramas para describir distribuciones de probabilidades de variables continuas son solo una aproximación de la distribución real, debido al número finito de puntos y a la naturaleza discreta del proceso de binnedo. Este problema puede ser solucionado mediante el uso de las *funciones kernel*. Cada punto x_i puede representarse por una función kernel, la cual tiene las siguientes características: es puntiaguda; tiene el máximo en 0; cae a cero a una distancia h ; y su integral está normalizada para que su resultado sea la unidad. El estimador $\hat{p}(x)$, para la distribución de una variable continua, es una suma normalizada de todas las funciones kernel, una por cada punto de la muestra

$$\hat{p}(x) = \frac{1}{nh} \sum_i K\left(\frac{x - x_i}{h}\right)$$

La ventaja de estos estimadores kernel es que usualmente terminan siendo una función suave \hat{p} y para un dado valor $\hat{p}(x)$ también contribuyen los puntos que se encuentran más alejados de x , con peso decreciente a medida que aumenta la distancia. El parámetro más importante es el ancho h , ya que un valor pequeño haría que la función fuese una sucesión de picos distinguibles, mientras que un valor grande provocaría que se pierdan los detalles importantes de la distribución. A continuación se listan las funciones kernel más conocidas:

- Uniforme:
 $K(u) = \frac{1}{2}I(|u| \leq 1)$
- Triángulo:
 $K(u) = (1 - |u|)I(|u| \leq 1)$
- Epanechnikov:
 $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$
- Cuartica:
 $K(u) = \frac{15}{16}(1 - u^2)^2I(|u| \leq 1)$
- Triple peso:
 $K(u) = \frac{35}{32}(1 - u^2)^3I(|u| \leq 1)$
- Gaussiana:
 $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$
- Coseno:
 $K(u) = \frac{\pi}{4} \cos(\frac{\pi}{2}u) I(|u| \leq 1)$



La función $I(|u| \leq 1)$ es la función que asigna el valor 1 para todos los u que cumplen que $|u| \leq 1$, para el resto de los valores se define igual 0. En la figura superior puede verse la forma de dichas funciones kernel, mientras en la figura inferior puede verse un ejemplo en el que el histograma es transformado, por las distintas funciones kernel, en una distribución suave. Observar que se puede calcular el intervalo de confianza para el estimador $\hat{p}(x)$ haciendo

$$\left[\hat{p}(x) \pm z_{(1-\alpha/2)} \sqrt{\text{Var}[\hat{p}(x)]} \right]$$

donde

$$\text{Var}[\hat{p}(x)] = \frac{\hat{p}(x)}{nh} \int K^2(u) du$$

Los valores de la integral involucrada en el cálculo de la varianza están perfectamente determinados dependiendo de la función kernel utilizada. Para las funciones kernel enumeradas anteriormente son: uniforme: $1/2$; triangular: $2/3$; Epanechnikov: $3/5$; Cuartica: $5/7$; Triple peso: $350/429$; Gaussiana: $1/(2\sqrt{\pi})$; Coseno: $\pi^2/16$.

6.2. Técnicas de Remuestreo

Todos los métodos usados hasta aquí usan, de una u otra manera, una distribución normal para los datos. Sin embargo, nunca los datos están distribuidos exactamente de acuerdo con una normal. El procedimiento t es útil en la práctica porque es robusto, es decir, es bastante insensible a desviaciones respecto de la distribución normal por parte de los datos. Aún así, no se pueden usar los IC construidos con t si los datos distribuidos tienen un alto valor de skewness (distribuciones con colas), a menos que las muestras sean muy grandes. Los métodos que se describirán a continuación tienen la ventaja de que no necesitan de datos distribuidos normalmente o muestras muy grandes. Estos métodos prácticamente carecen de fórmulas y funcionan de la misma manera para muchas diferentes estadísticas. Estos métodos permiten, con la suficiente potencia computacional, obtener resultados que muchas veces son más exactos que aquellos obtenidos por métodos tradicionales. Es más, los intervalos que se obtienen con las técnicas de remuestreo, son conceptualmente más simples que los IC y las pruebas basadas en distribuciones normales, debido a que están directamente relacionados con la base del proceso inferencial: las distribuciones muestrales "muestran" qué debería pasar si se tomaran muchas muestras bajo las mismas condiciones.

6.2.1. Método Bootstrap

6.2.1.1. Definición

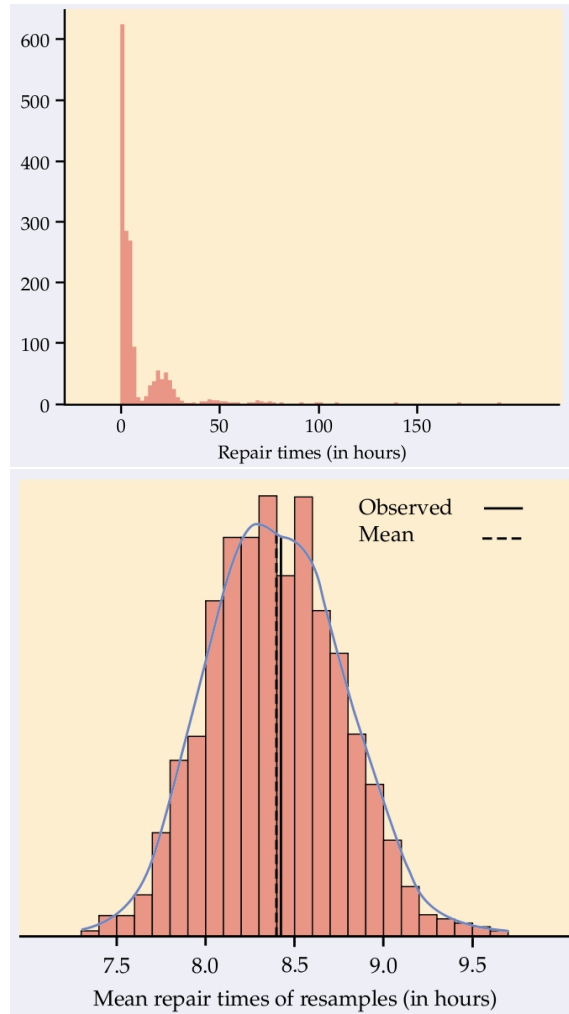
La inferencia estadística se basa en las distribuciones muestrales de una muestra de estadísticos. El método bootstrap es, en primer lugar, una manera de encontrar la distribución muestral, al menos aproximadamente, solo a partir de una muestra. Este es el procedimiento a seguir:

- *Remuestreo*: Una distribución muestral está basada en muchas muestras extraídas a partir de una población. Si tenemos una sola muestra aleatoria, se realizan muchos remuestreos, repitiendo el muestreo con repeticiones a partir de la única muestra aleatoria que disponemos. Cada remuestreo debe tener el mismo tamaño que la muestra aleatoria original.
- *Distribución bootstrap*: La distribución muestral de un estadístico colecciona los valores de dicho estadístico proveniente de muchas muestras. La distribución bootstrap de un estadístico colecciona sus valores a partir de muchos remuestreos. La distribución bootstrap nos da información acerca de la distribución muestral.

Por lo tanto la idea del bootstrap se puede describir de la siguiente manera:

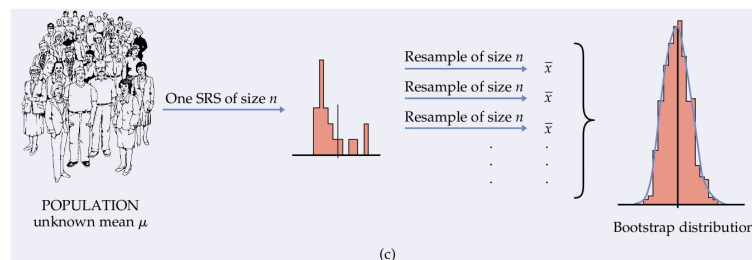
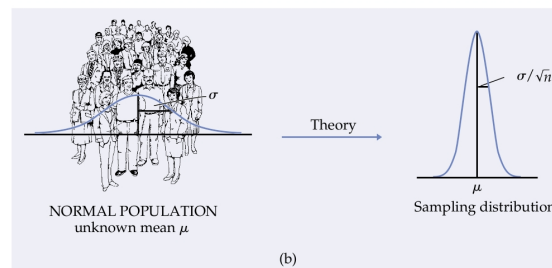
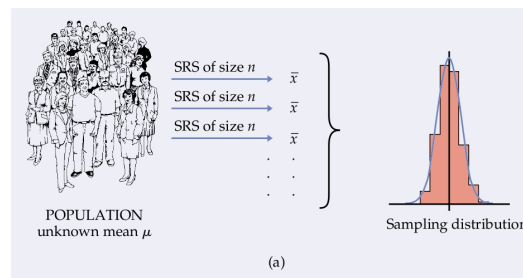
La muestra original representa la población a partir de la cuál fué extraída. Por lo que los remuestreos a partir de dicha muestra, representan que se obtendría si tomáramos muchas muestras extraídas de la población. La distribución bootstrap de un estadístico, basada en muchos remuestreos, representa la distribución muestral de dicho estadístico, basado en muchas muestras.

Ejemplo: En la mayoría de los países desarrollados, varias empresas de teléfonos ofrecen sus servicios en una dada ciudad. Para evitar que cada empresa tenga que instalar sus propias líneas, la empresa primaria de servicio de cada región debe compartir sus líneas con sus competidoras. A su vez, la empresa primaria debe encargarse de reparar las líneas de sus competidoras, por lo que surge la pregunta si dicha empresa repara con la misma velocidad sus líneas como las de sus competidoras. Para saber esto, se requiere implementar un test de significancia que permita comparar los tiempos de reparación para dos grupos de clientes. En la figura superior se observa la distribución de los tiempos de reparación registrados para 1664 clientes de empresas competidoras. Como puede verse, la distribución de los tiempos de reparación es bastante diferente a una distribución normal. La mediana es 3.59 horas y la media es 8.41 horas y el tiempo más largo de reparación es 191.6 horas. Para este análisis desistimos de usar el procedimiento t , especialmente porque el tamaño de la muestra clientes competidores es mucho menor que el correspondiente a la muestra de clientes de la empresa primaria. Si quisieramos estimar la media de la población μ , sabemos que el estadístico que corresponde es la media muestral \bar{x} , el cuál hemos dicho que tiene un valor igual a 8.41 horas. Ahora, usemos el método bootstrap sobre la muestra para calcular distintos valores de \bar{x} , así como si estuvieramos extrayendo diferentes muestra de la población. En la figura inferior puede verse el resultado de realizar 1000 remuestreos a partir de la muestra original. La línea solida vertical marca el valor original de 8.41, mientras que la línea a rayas marca la media de las medias bootstrap. Podemos comparar la distribución bootstrap con lo que sabemos de la distribución muestral:



- *Forma*: se ve que la distribución bootstrap es casi normal. El teorema del límite central dice que la distribución muestral de la media muestral es aproximadamente normal si n es grande. Por lo que la forma de la distribución bootstrap es cercana a la forma que esperamos que tenga la distribución muestral.
- *Centro*: la distribución bootstrap está centrada cerca a la media de la muestra original. Esto es, la media de la distribución bootstrap tiene un sesgo pequeño como estimador de la media de la población original (insesgabilidad).
- *Dispersión*: se puede obtener una medición numérica del ancho calculando la desviación estándar. Esta se denomina *error estándar bootstrap de \bar{x}* . El valor numérico para este ejemplo es 0.367. Por otro lado sabemos que la desviación estándar de la muestra original es $s/\sqrt{n} = 14,69/\sqrt{1664} = 0,360$. Por lo que el error estándar bootstrap está en acuerdo con la estimación teórica.

El pesado cálculo computacional necesario para producir la distribución bootstrap reemplaza la pesada teoría que nos habla acerca de la distribución muestral. La gran ventaja de la idea del remuestreo es que funciona frecuentemente cuando la teoría falla. Por supuesto, la teoría tiene sus ventajas: conocemos exactamente cuando funciona. Y por ahora, no sabemos cuando el remuestreo funciona.



6.2.1.2. La idea del bootstrap .

Pareciera que el remuestreo crea nuevos datos de la nada. Esto parece sospechoso. Pero las observaciones remuestreadas no son usadas como si fuesen nuevos datos. La distribución bootstrap de las medias remuestreadas se usa solamente para estimar de que manera la media muestral de la muestra original varía debido al muestreo aleatorio. Usar la muestra de datos con un propósito doble, estimar un parámetro y su variabilidad, es perfectamente legítimo. Hacemos exactamente lo mismo cuando calculamos \bar{x} para estimar μ y después calculamos s/\sqrt{n} a partir de los mismos datos para estimar la variabilidad de \bar{x} .

Entonces, ¿qué es lo novedoso de este método? Primero que nada, no se confía en la fórmula s/\sqrt{n} para estimar la desviación estándar de \bar{x} . Se adopta como estimador de la variabilidad, la desviación estándar ordinaria de los muchos valores \bar{x} obtenidos a partir de los remuestreos. Otra cosa que es nueva es que no se recurre al teorema del límite central o cualquier otra teoría para saber si la distribución muestral es aproximadamente normal. Lo que se hace es mirar la distribución bootstrap para saber si es aproximadamente normal o no. En la mayoría de los casos, la distribución bootstrap tiene aproximadamente la misma forma y dispersión que la distribución muestral, pero se encuentra centrada en el valor original del estadístico en vez de estar sobre el valor del parámetro poblacional. El método bootstrap nos permite calcular errores estándar para estadísticas para las cuales no tenemos fórmulas, y corroborar normalidad para estadísticas que la teoría no puede manejar con facilidad.

6.2.1.3. Primeros pasos para usar el bootstrap .

El método bootstrap es más útil para establecer condiciones cuando no conocemos como es la distribución muestral del estadístico. Los principios son:

- *Forma*: debido a que la forma de la distribución bootstrap se aproxima a la forma de la distribución muestral, podemos usar la distribución bootstrap para corroborar la normalidad de la distribución muestral.
- *Centro*: un estadístico es sesgado como una estima del parámetro, si su distribución muestral no está centrada en el verdadero valor del parámetro. Se puede corroborar sesgo viendo donde la distribución bootstrap de un estadístico está centrada con respecto al valor del estadístico para la muestra original. Más precisamente, el sesgo del estadístico es la diferencia entre la media de su distribución muestral y el verdadero valor del parámetro. La estima del sesgo para una distribución bootstrap es la diferencia entre la media de dicha distribución y el valor del estadístico de la muestra original.
- *Dispersión*: el error estándar bootstrap del estadístico es la desviación estándar de su distribución bootstrap. Entonces, el error estándar bootstrap estima la desviación estándar de la distribución muestral del estadístico.

6.2.1.3.1. IC bootstrap con el estadístico t

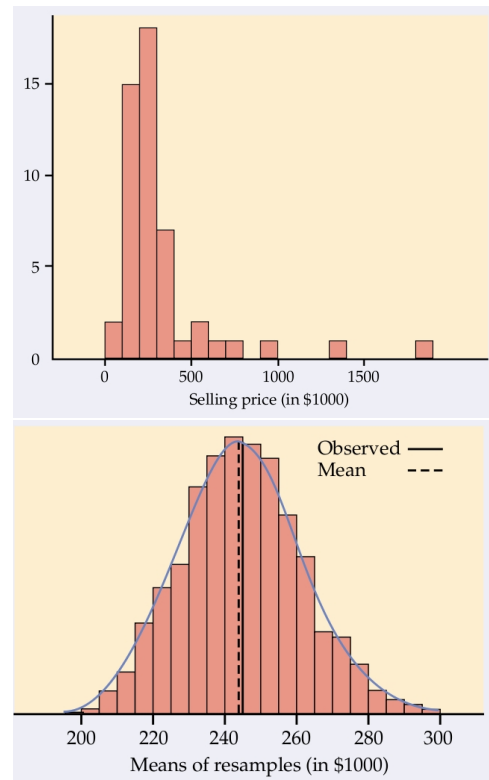
Si la distribución bootstrap de un estadístico muestra una forma normal y sesgo pequeño, se puede obtener un IC para el parámetro usando el error estándar bootstrap y la distribución t .

Ejemplo:

Se está interesado en los precios de venta de casas residenciales en una dada ciudad. Se tienen una muestra de 50 precios tomados, durante el año pasado, por un contador. Desafortunadamente los datos no distinguen entre casas residenciales o comercios. La mayoría de las ventas fueron residenciales, pero unas cuantas ventas de comercios a alto precio pueden incrementar considerablemente el resultado de la media de precios de venta. En la figura superior se observa la distribución de los 50 precios de venta con los que cuenta la muestra. La distribución, obviamente, dista de ser normal, con unos cuantos valores atípicos que podrían ser ventas comerciales. La muestra es relativamente chica, la distribución tiene una alta asimetría y está contaminada por un número desconocido de ventas comerciales. Como se podría estimar el centro de la distribución apesar de estos inconvenientes?

El primer paso es abandonar la media como medida del centro en favor de un estadístico que sea menos sensible a los valores atípicos. Podríamos escoger la mediana, pero en este caso se elije usar la *media recortada al 25%*.

Éste estadístico es la media de solo la parte central de las observaciones en un conjunto de datos. En particular, la $\bar{x}_{25\%}$ ignora los valores menores al 25% y los mayores al 75%, es decir, es la media del 50% del medio de las observaciones. En nuestro ejemplo, ya que el 25% de 50 es 12.5, desechamos los 12 valores más bajos y los 12 más altos de la lista de precios. Entonces, se obtiene que $\bar{x}_{25\%} = 244,0019$. No podemos decir mucho acerca de la distribución del estadístico $\bar{x}_{25\%}$ cuando solo se tienen 50 datos de una distribución muy asimétrica. Afortunadamente, no necesitamos saber nada para aplicar el método bootstrap. Realizamos 1000 remuestreos de 50 precios cada uno y calculamos la media y la forma de la distribución bootstrap (figura inferior). El calculo arroja los siguientes resultados: $\bar{x}_{25\%}^b = 244,7$, sesgo=0.7171 y error estándar bootstrap $S_b = 16,83$.



Ahora, que es lo que vemos? Con respecto a la *forma*, la distribución bootstrap es muy similar a una distribución normal. Esto sugiere que la distribución muestral de la media recortada es también muy similar a una normal. Si analizamos el *centro*, la estima del sesgo bootstrap es 0.7171, el cuál es pequeño en comparación al valor del 244 que toma el estadístico. Por lo tanto, el estadístico tiene un sesgo pequeño como estimador del parámetro poblacional. Y por último, viendo la *dispersión*, ésta es una estima de la desviación estándar de la distribución muestral de la media recortada. Recordemos que el IC, cuando una muestra proviene de una distribución normal con varianza desconocida, es

$$\bar{x} \pm t_{(1-\alpha/2; n-1)} S / \sqrt{n}$$

Cuando una distribución bootstrap es aproximadamente normal y tiene sesgo pequeño, podemos usar esencialmente la misma receta introduciendo como desviación estándar, el error estándar bootstrap para obtener el IC de cualquier parámetro, es decir

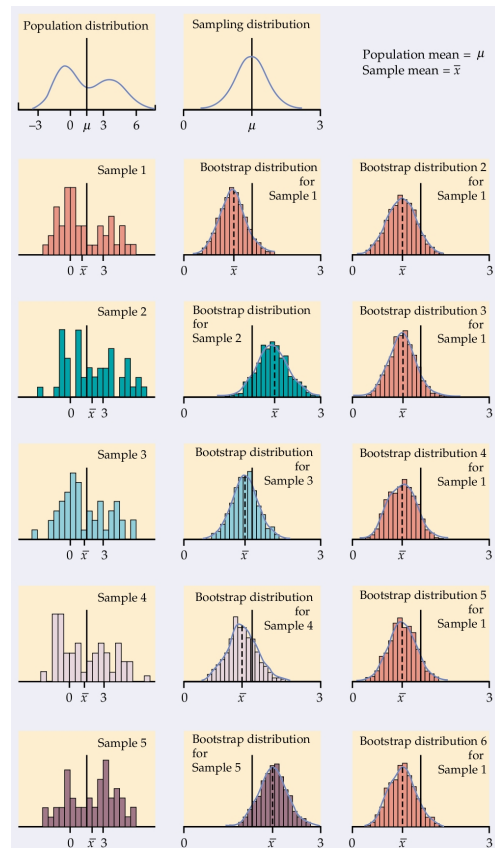
$$\bar{x} \pm t_{(1-\alpha/2; n-1)} S_b$$

Volviendo a nuestro ejemplo, si deseamos el IC con un 95% de probabilidad de que la media recortada este incluida en él, entonces

$$\begin{aligned} \bar{x}_{25\%} \pm t_{(0,975;49)} S_b &= 244 \pm 2,009(16,83) = \\ &= 244 \pm 33,81 \longrightarrow [210,19 ; 277,81] \end{aligned}$$

6.2.1.4. Qué tan exacta es una distribución bootstrap?

Las distribuciones muestrales de un estadístico muestran la variación del estadístico debido a la selección de distintas muestras aleatorias a partir de la población. Ahora hemos usado la distribución bootstrap como un sustituto de la distribución muestral. Esto introduce una segunda fuente de variabilidad aleatoria: el remuestreo es elegido aleatoriamente a partir de la muestra original. La inferencia bootstrap genera una distribución bootstrap y la usa para informarnos acerca de la distribución muestral. Podemos confiar en esa inferencia? En la figura de la derecha puede verse un ejemplo del proceso completo. La distribución de la población tiene dos picos y esta lejos de ser normal. La distribución muestral aparece a la derecha de la anterior, y es aproximadamente normal, como se espera por el teorema del límite central. Los histogramas en la columna de la izquierda son 5 muestras de 50 puntos extraídas de la población mientras que la columna central muestra el resamplero de cada una de esas muestras (1000 resampleros cada una).

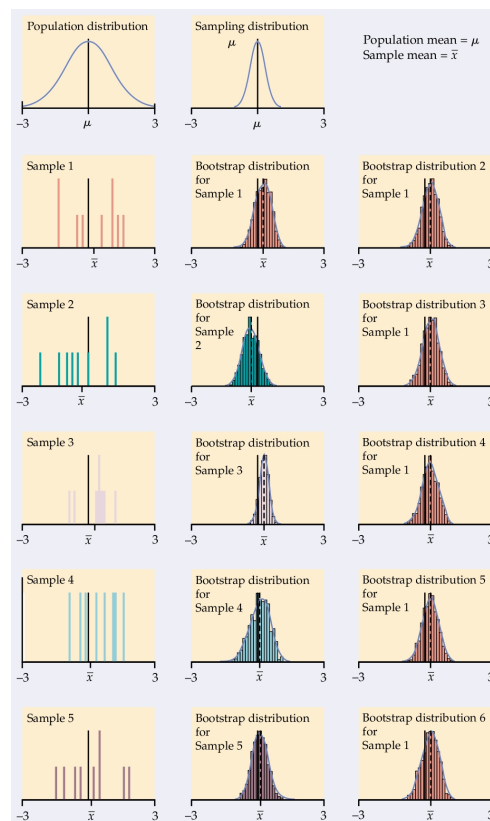


Finalmente, la columna de la derecha son distintos remuestreos de la muestra 1. Por lo tanto, si comparamos las 5 distribuciones bootstrap de la columna central, veremos el efecto de la elección aleatoria de las muestras originales, mientras que si comparamos las 6 distribuciones bootstrap realizadas a partir de la muestra 1, veremos el efecto de el remuestreo aleatorio. Las conclusiones son las siguientes:

- Cada distribución bootstrap está centrada cerca del valor del estadístico original. Esto significa que las estimas bootstrap del sesgo son pequeñas en todos los casos.
- Los 5 remuestreos (columna central) son similares a la distribución muestral en forma y dispersión. La variación muestra a muestra no es importante.
- Los 6 remuestreos de la muestra 1 son muy similares en forma, centro y dispersión. Esto significa que el remuestreo aleatorio introduce muy poca variación debido a la elección aleatoria de la muestra original a partir de la población.

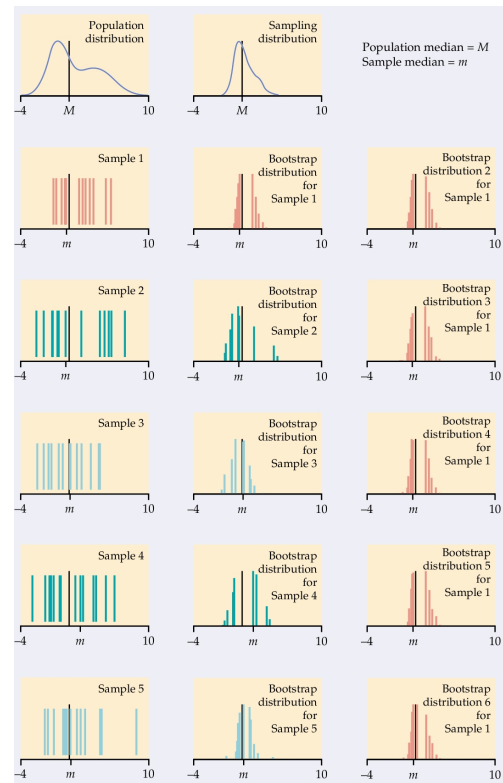
Por lo tanto, si una distribución bootstrap esta basada en una muestra moderadamente grande a partir de la población, su forma y dispersión no dependen fuertemente de la muestra original e imitan la forma y dispersión de la distribución muestral.

Ahora sabemos que casi todas las variaciones entre distribuciones bootstrap para un estadístico, tal como la media, proviene de la selección aleatoria de la muestra original a partir de la población. También sabemos que en general los estadistas prefieren muestras grandes porque las muestras pequeñas dan resultados más variables. Este hecho general también es cierto para los procedimientos bootstrap. Veamos un nuevo ejemplo, el cuál se encuentra graficado en la figura de la derecha. El esquema es el mismo que el del ejemplo anterior, salvo que ahora las muestras seleccionadas son de tamaño $n = 9$. La distribución de la población es normal, por lo tanto, por más que el tamaño de las muestras sea pequeño, la distribución muestral es normal. Las distribuciones bootstrap (columna central) muestran mucha más variación en forma y dispersión que las del ejemplo anterior. Por ejemplo, el remuestreo de la muestra 4 nos da una distribución bootstrap muy asimétrica. Por lo tanto, las distribuciones bootstrap no son similares a la distribución muestral. Es decir, no podemos confiar en una distribución bootstrap realizada a partir de una muestra pequeña para que reproduzca la forma y la dispersión de



la distribución muestral. Por otro lado, los 6 remuestreos de la muestra 1 son todos similares. Esto se debe a que cada distribución bootstrap está hecha con 1000 remuestreos. En conclusión, el método bootstrap no puede sobrellevar la debilidad de muestras pobres como una base para la inferencia. Algunos procedimientos bootstrap son usualmente más exactos que métodos estándar, pero incluso ellos no son lo suficientemente exactos para muestras muy pequeñas.

Por último, analizaremos el caso de aplicar el método bootstrap cuando se usa como estadístico a la mediana. Cuando hicimos el ejemplo de los precios de las propiedades elegimos como estadístico la media recortada en vez de la mediana. En parte, esto se hizo porque el procedimiento bootstrap no funciona bien con la mediana a menos que la muestra original sea bastante grande. Para entender mejor esto, veamos un ejemplo. El esquema del ejemplo es muy parecido a los ejemplos anteriores, con la diferencia que ahora el estadístico es la mediana (figura de la derecha). La letra M identifica al valor de la mediana en la población mientras que m denota la mediana muestral. Las 5 muestras son de tamaño $n = 15$. Como puede verse en la columna central, las 5 distribuciones bootstrap difieren marcadamente una de otra y de la distribución muestral. Éste es el porqué. La mediana de un remuestreo de 15 puntos es la 8 observación más grande. Esta siempre es una de las 15 observaciones en la muestra original y es usualmente una de las observaciones del medio. Entonces cada distribución bootstrap repite los mismos pocos valores, y estos valores dependen de la muestra original. La distribución muestral, por otro lado, contiene todas las medianas de todas las posibles muestras y por eso no está confinada a unos pocos valores. La dificultad disminuye cuando el tamaño de la muestra es par, ya que la mediana surge del promedio de las dos observaciones centrales. Es mucho menos notable, además, si las muestras son moderadamente grandes, digamos $n = 100$ o más. Los errores estándar bootstrap provenientes de esas muestras y los IC son razonablemente exactos, aunque las formas de las distribuciones bootstrap aún se vean raras. Esta misma dificultad se encontrará para otros estadísticos, como por ejemplo los cuartiles, los cuales son calculados por una o dos observaciones de una muestra.



6.2.2. Método Jackknife

Por último, describiremos brevemente otra técnica de remuestreo muy conocida. El método se denomina *Jackknife* y es principalmente útil cuando la dispersión de una distribución es grande o existe la presencia de valores atípicos (outliers) en la muestra.

Supongamos que tenemos una muestra de tamaño n extraída a partir de una población y estimamos el estadístico, por ejemplo, la media muestral \bar{x} . El procedimiento es similar al descrito para el caso del método bootstrap, en el sentido que el método Jackknife también remuestrea la muestra original de datos de manera de obtener varias muestras. La diferencia es que el remuestreo se hace eliminando un elemento x_i a la vez de la muestra y calculando el estadístico correspondiente para la nueva muestra de tamaño $n - 1$. Este procedimiento genera que el número de remuestreos que se pueden lograr este limitado por el tamaño de la muestra original. Con los nuevos n valores para el estadístico (\bar{x}_i^J) surgidos del procedimiento Jackknife, se puede calcular el error estándar de dicho estadístico, haciendo

$$S_J = \sqrt{\frac{n-1}{n} \sum_{i=1}^n (\bar{x}_i^J - \bar{x})^2}$$

Este método también es capaz de dar una estima del sesgo. Si tenemos una situación en la cuál una cantidad estimada tiende a salirse por arriba o por abajo del valor verdadero en una muestra muy pequeña. Entonces la estima de \bar{x} con los n puntos será más grande o más chica que el valor verdadero. Si esto pasa, uno podría esperar que, eliminar una medición, como se hace en el Jackknife, disminuya el sesgo. Este efecto se mide comparando la media de los valores del Jackknife con la media de toda la muestra. Si hay diferencias, se puede corregir por el sesgo usando

$$\bar{x}_C = \bar{x} - (n-1)(\bar{x}^J - \bar{x})$$

Observar que el método jackknife también suele aplicarse no eliminando un elemento a la vez, sino, eliminando un conjunto de elementos cada vez.

7. Inf. Est.: Prueba de Hipótesis (I)

Los métodos de estimación estudiados en anteriormente usan la información proporcionada por los estadísticos muestrales para estimar con cierta probabilidad el valor de un parámetro poblacional. Ahora se analizará el método de prueba de hipótesis que es un enfoque diferente. En éste caso, se supone a priori el valor del parámetro y sobre la base de la información obtenida en una muestra se somete a prueba la suposición, para luego tomar con cierta probabilidad, la decisión de rechazar o no rechazar la hipótesis. La prueba de hipótesis (desde ahora, PH) o contrastación de hipótesis es uno de los métodos estadísticos más usados en las ciencias naturales por ser un procedimiento que le proporciona al investigador un criterio objetivo para tomar decisiones con base a un número limitado de observaciones. Algunos de los problemas que podemos resolver con este método son los siguientes:

1. Determinación del nivel de confiabilidad de un estadístico.
2. Comparación de dos distribuciones para variables aleatorias discretas.
3. Comparación de dos distribuciones para variables aleatorias continuas.
4. Determinar la independencia estadística de dos poblaciones.

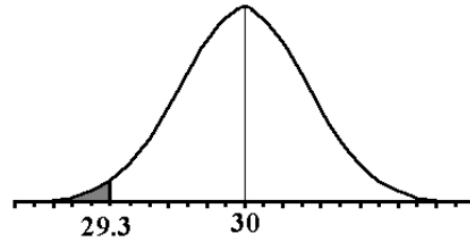
En esta sección solo desarrollaremos el primero de los puntos, mientras que dejaremos el tratamiento de los demás casos para la siguiente sección.

7.1. PH: un procedimiento de decisión

Antes de estudiar las distintas etapas y casos de las que consta el procedimiento para la PH, consideraremos un ejemplo que servirá para mostrar los fundamentos del proceso y la toma de decisiones.

Ejemplo: Con el propósito de determinar el efecto de una nueva dieta se forman varios lotes de 36 ratones con un peso aproximado a los 30 g. Para verificar si los grupos son homogéneos en cuanto al peso, vuelve a pesar cuidadosamente los 36 ratones de cada grupo y le calcula el valor promedio y la desviación estándar. Ahora el investigador se encuentra ante una disyuntiva: *a)* si el valor promedio de peso para cada grupo se considera como una simple desviación fortuita de los 30 g dada la variabilidad característica de las muestras aleatorias, no hay necesidad de reorganizar el grupo, y *b)* si el valor medido está verdaderamente desviado del valor esperado de 30 g es necesario reorganizar el grupo sustituyendo los ratones causantes de la desviación. A fin de tener un criterio objetivo que le ayude a tomar la mejor decisión, el investigador establece como premisa que el peso promedio μ de la población es de 30 g. Si esto es cierto es de esperar que el valor promedio \bar{x} del grupo o muestra sea muy cercano a dicho valor y su probabilidad de ocurrencia sea alta. Si esto sucede se acepta la hipótesis y se considera que la desviación del peso promedio de la muestra con respecto a la media esperada es producto de la naturaleza aleatoria de la variable peso, siendo innecesario reorganizar el grupo de ratones. Pero aún siendo cierto que $\mu = 30$, es posible,

aunque poco probable, que los 36 ratones tengan un peso promedio alejado del peso esperado de 30 g. En éste caso, el investigador puede aceptar que $\mu = 30$ y considerar que ocurrió un hecho poco probable o alternativamente decidir que en lugar de haber sucedido algo improbable considerar que el valor de la media poblacional es menor a 30. Entonces, supongamos que el investigador encontró que uno de los grupos dió como resultado un promedio de 29.3 g con una desviación de 2 g. De acuerdo a lo dicho anteriormente, para poder tomar la decisión de reorganizar o no el grupo de ratones, se debe proceder a determinar si 29.3 ocurre con una probabilidad alta o baja teniendo como hipótesis que $\mu = 30$. Como el peso promedio observado es menor a 30 se debe proceder a hallar la $P(\bar{X} \leq 30)$. Como el tamaño de la muestra es grande ($n = 36$) se puede afirmar, de acuerdo al Teorema del Límite Central, que dicha variable se distribuye normalmente con media igual a 30 y desviación igual a $S_{\bar{x}} = 2/\sqrt{36} = 0,33$. Por lo tanto la probabilidad buscada será:

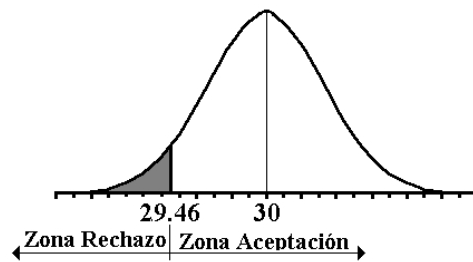


$$P(\bar{X} \leq 29,3) = P\left(Z \leq \frac{29,3 - 30}{0,33}\right) = 0,0179$$

Esta probabilidad tan baja (figura de la derecha), tiene dos explicaciones: *a*) es cierta la hipótesis y ocurrió un hecho casi imposible como el de obtener un peso promedio igual a 29.3 que está muy alejado del valor esperado de 30 g., y *b*) no es cierta la hipótesis anterior y el valor esperado es mucho menor a 30. La explicación *b* resulta obviamente más razonable. Si el valor de la media muestral hubiese sido de 29.9, la probabilidad de ocurrencia sería de 0.382. Esta probabilidad es alta siempre y cuando $\mu = 30$. Por lo tanto resulta razonable aceptar la presunción de que el peso promedio del grupo todavía es igual a 30 g. Pero si el valor de la media muestral fuese 29.5? La probabilidad de ocurrencia daría 0.1151. En este caso la probabilidad no es tan baja para rechazar de inmediato que $\mu = 30$ y tampoco es tan alta para aceptar sin mayores consideraciones. La mejor manera de resolver el problema es estableciendo previamente un valor límite para aceptar o rechazar la hipótesis y así poder tomar una decisión inmediata. Este valor límite debe excluir los valores que ocurren con menor probabilidad. Por lo general se excluyen aquellos valores cuya probabilidad de ocurrencia es igual o menor a 0.05. Una vez que se elige el valor de probabilidad que sirve de criterio para tomar una decisión, se pueden conocer cuáles valores de la variable cumplen con ésta decisión. Si decidimos que el valor de probabilidad crítico es 0.05, todos los valores que rechazan la hipótesis establecida son aquellos cuya $P(\bar{X} \leq \bar{x}) = 0,05$. Esta probabilidad es equivalente a $P(Z \leq z) = 0,05$. Buscando en la tabla se encuentra que el valor de Z que tiene a su izquierda una área de distribución de 0.05 es -1.64. Por lo tanto, tenemos que

$$\bar{x} = \mu_x + z_{0,05}S_x/\sqrt{n} = 30 + (-1,64)2/\sqrt{36} = 29,46$$

Este valor es ahora nuestro límite para tomar la decisión de aceptar o rechazar la presunción de que $\mu = 30$. Si la media del grupo de ratones es menor a 29.46 se rechaza la premisa y si es mayor se acepta (figura de la derecha). Ahora sabemos que 0.54 es la máxima desviación que se puede aceptar para concluir que la diferencia entre la media observada y la esperada es simplemente aleatoria. Volviendo al caso de los ratones, el investigador ahora conociendo el peso promedio de cada grupo puede tomar rápidamente una decisión para mantener o reorganizar el grupo, simplemente comparando la media obtenida con el valor crítico de 29.46 g.



7.2. Procedimiento general para la PH

En el procedimiento usado para resolver el ejemplo anterior se pueden identificar varias etapas fundamentales, las cuales se pueden reordenar e identificar en la siguiente forma:

1. Hipótesis
2. Nivel de significación
3. Estadístico de prueba
4. Zona de aceptación
5. Cálculos necesarios
6. Decisión
7. Conclusión

En lo que sigue supondremos que todas las variables usadas siguen una distribución normal y la mayoría de las veces usaremos la media poblacional μ como ejemplo del parámetro a estudiar.

7.2.1. Hipótesis

Por lo general toda investigación en el campo de las ciencias naturales se inicia a partir de una hipótesis la cual es una explicación tentativa que se da a un hecho observado. Ahora bien, en la formulación de cualquier hipótesis está implícita una hipótesis alternativa. Por ejemplo, se puede plantear como hipótesis de investigación que el ejercicio constante disminuye el nivel de colesterol en el plasma sanguíneo, pero asociada a esta hipótesis existe otra premisa alterna que se opone, en éste caso la alternativa sería que el ejercicio constante no disminuye el nivel de colesterol en el plasma sanguíneo. Estas hipótesis de investigación, para poderse someter a prueba, deben concretarse en términos cuantitativos, transformándose en hipótesis estadísticas. En forma general las hipótesis estadísticas son afirmaciones que involucran una propiedad de la distribución probabilística de la variable aleatoria que se está estudiando, propiedades como son la

media, la varianza, un valor de proporción o la forma de la distribución. De modo que el primer paso en un proceso de decisión es formular las hipótesis estadísticas, las cuales reciben el nombre de hipótesis nula (H_0) e hipótesis alternativa (H_1). La hipótesis nula se dice que es una hipótesis simple, porque es una afirmación de igualdad con un valor específico, mientras que la hipótesis alternativa se dice que es compuesta porque puede asumir diferentes valores. Si se representa un parámetro poblacional por letra griega θ y con θ_0 un valor cualquiera del parámetro, la forma genérica de la hipótesis nula sería una igualdad entre el parámetro y un valor específico del mismo:

$$H_0 : \theta = \theta_0$$

Por su parte la hipótesis alternativa se puede representar con una de las tres posibilidades siguientes:

$$H_1 : \begin{cases} \theta > \theta_0 \\ \theta < \theta_0 \\ \theta \neq \theta_0 \end{cases}$$

La utilidad de plantear las hipótesis de ésta manera se explica porque el rechazo de H_0 es un veredicto mucho más robusto que su no rechazo, puesto que es necesario acumular evidencia científica muy fuerte para poder rechazar una hipótesis nula. Por lo tanto la consecuencia de rechazar una hipótesis nula es un gran apoyo a la hipótesis alternativa. Ilustremos esta situación con la analogía siguiente: en los procesos judiciales donde hay alguien acusado de un delito, hay dos hipótesis: inocente (H_0) y culpable (H_1). El fiscal público tiene interés en probar que el acusado es culpable. Para poder llegar a una decisión de culpable es necesario presentar suficientes evidencias que garanticen que la decisión es correcta. De no tenerse evidencias fuertes la hipótesis nula de inocencia no puede ser rechazada, pero esto no significa que se comprobó la inocencia del acusado, sino que no se logró acumular suficientes elementos para rechazar H_0 . De hecho es posible que con nuevas investigaciones se determine la culpabilidad del acusado. Por el contrario habiéndose obtenido fuertes evidencias de culpabilidad, se acepta la hipótesis alternativa, decisión que es mucho más difícil revertir. En otras palabras la probabilidad de cometer un error es mucho menor al rechazar H_0 que al no rechazarla. En la práctica jurídica, si la evidencia es débil es preferible equivocarse declarando inocente a alguien culpable que condenando a un inocente. Un razonamiento similar a éste es el que usan los investigadores cuando plantean como hipótesis alternativa el evento que se quiere probar. Si los datos usados para probar las hipótesis proporcionan suficiente evidencia para rechazar la hipótesis nula, como consecuencia inmediata la hipótesis alternativa recibe un respaldo muy fuerte. Pero si el investigador hubiese planteado el mismo evento como hipótesis nula, su no rechazo no demuestra que el evento de interés sea verdad, sino que los datos no proporcionaron evidencia para rechazarla, dejando abierta la posibilidad de poder ser refutada con otro conjunto de datos o que otra hipótesis sea la verdadera. Por esta razón, es que la sustitución del término no rechazar H_0 por el término aceptar H_0 , no es muy conveniente y de hacerlo se debe estar consciente que la aceptación de H_0 es sólo temporal. El ejemplo que sigue puede aclarar la temporalidad de una aceptación de H_0 . Suponga que alguien afirma que todos los granos de porotos que hay en un saco son

de color negro. Para probarlo toma un puñado de granos y observa su color. Si todos los porotos del puñado son negros, no significa que probó su premisa, solamente le dio apoyo. Puede repetir el ensayo muchas veces con el mismo resultado, pero mientras existan granos de porotos en el saco su hipótesis no está probada, porque si en alguno de los ensayos encuentra un solo grano de otro color, la hipótesis nula queda definitivamente negada y por el contrario la hipótesis alternativa implícita de que no todos los granos de porotos del saco son negros queda plenamente confirmada.

La formulación de una hipótesis no siempre es una tarea fácil debido a que no todas las situaciones son obvias. Al no existir normas ni procedimientos que se puedan aplicar para plantear correctamente las hipótesis estadísticas, el investigador debe apelar a la experiencia y a su conocimiento del sistema bajo estudio.

7.2.2. Nivel de significación

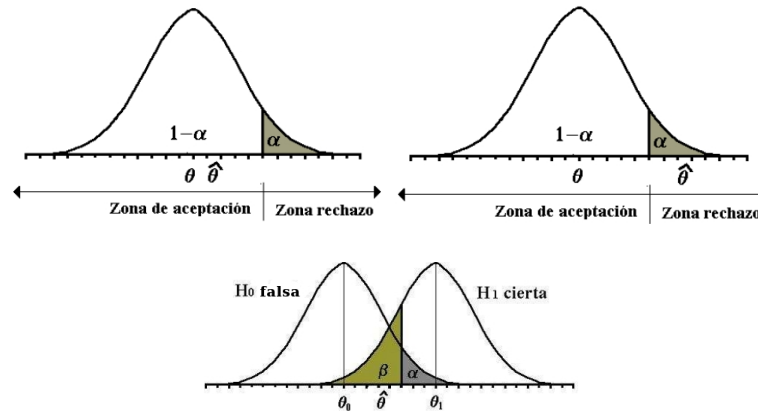
El proceso de PH se basa fundamentalmente en determinar si la diferencia que existe entre el valor del estadístico muestral y el valor del parámetro poblacional es lo suficientemente grande que no pueda atribuirse simplemente al azar, sino a la falsedad de la hipótesis nula. A fin de determinar el tamaño que debe tener esta diferencia para que sea significativa se establece un criterio o límite de significación. Cualquier valor del estadístico que supere este límite se dice que alcanzó una diferencia significativa con respecto al valor del parámetro. El límite se establece de forma que sólo alcanzan la significación aquellos valores que ocurren con una probabilidad igual o menor a 0.05 (podría ser 0.10 o 0.01). El establecimiento del límite de significación define de inmediato dos zonas en la distribución de valores del estadístico: *a*) una zona de aceptación de H_0 , dentro de la cual las diferencias entre el estadístico y el parámetro no son significativas, y *b*) una zona de rechazo de H_0 dentro de la cual las diferencias entre el estadístico y el parámetro son significativas.

7.2.2.1. Errores de tipos I y II

Cualquier decisión dentro del proceso de prueba de hipótesis lleva asociado cierto riesgo de fallar. Es decir que siempre existe la posibilidad de tomar una decisión equivocada, sólo que en este tipo de prueba se tiene la ventaja de conocer de antemano la probabilidad de equivocarse. Las posibles situaciones al tomar una decisión pueden verse en la siguiente tabla.

CONDICIÓN REAL	DECISIÓN	
	Rechazar H_0	No Rechazar H_0
H_0 cierta	<i>Error (Tipo I)</i>	<i>Acierto</i>
H_0 falsa	<i>Acierto</i>	<i>Error (Tipo II)</i>

El razonamiento básico del proceso de PH supone que si el planteamiento de la hipótesis nula es cierto, la mayoría de las muestras proporcionarán valores del estadístico muestral $\hat{\theta}$ muy próximos al parámetro θ , y por lo tanto caerán dentro de la zona de aceptación (figura superior izquierda).



Pero también una minoría de observaciones puede no caer en la zona de aceptación a pesar que H_0 sea cierta, provocando que se tome una decisión errada, aunque se tiene a favor que se conoce la magnitud del error (figura superior derecha). Por ejemplo cuando se define una zona de aceptación donde se espera caigan el 95 % de las observaciones si H_0 es cierta, también se está determinando que en un 5 % de los casos se puede cometer una equivocación al rechazar H_0 cuando de hecho es cierta. Es decir que la probabilidad de cometer una falla es igual a 0.05. Este tipo de error se llama Error Tipo I y su probabilidad se identifica con la letra α .

También se puede cometer un error si se acepta H_0 cuando de hecho es falsa. Esto sucede cuando una observación cae dentro de la zona de aceptación de H_0 , siendo la hipótesis H_1 la verdadera (figura inferior). Este tipo de error se conoce como Error Tipo II y su probabilidad se identifica con la letra β . En términos de probabilidad los dos tipos de errores se expresa de la forma siguiente:

$$P(ET I) = P(\hat{\theta} \text{ Zona rechazo} / H_0 \text{ cierta}) = \alpha$$

$$P(ET II) = P(\hat{\theta} \text{ Zona aceptación} / H_1 \text{ cierta}) = \beta$$

Como se puede notar, tanto α como β son probabilidades condicionadas. En cualquier PH lo más conveniente será que ambos tipos de errores sean lo más pequeños posible, pero esto no es fácil de lograr porque al intentar disminuir uno el otro aumenta proporcionalmente. Afortunadamente, al aumentar el tamaño de la muestra disminuye la probabilidad de cometer el Error Tipo II y se mantiene constante la probabilidad de cometer el Error Tipo I. De acuerdo a lo visto hasta ahora, sería lógico concluir que es necesario conocer la magnitud con la cual ambos errores operan en una PH. Lamentablemente, esto sólo es posible para el Error Tipo I. Debido a la naturaleza del procedimiento, al formular una hipótesis nula no sólo se supone el valor de un parámetro, sino que se presume la ubicación de la distribución de probabilidades del estadístico

de prueba. La consecuencia de esto es que puede fijarse un valor de α y establecerse la respectiva región de rechazo de H_0 . Esto no es posible para el caso del Error Tipo II. Aun cuando se rechace H_0 se desconoce el valor de la hipótesis alternativa y por lo tanto la ubicación de la distribución probabilística del estadístico de prueba, no pudiéndose fijar el valor de β . Por tales razones en toda PH una vez que se han formulado la hipótesis se fija el valor de α con el cual se cuantifica el riesgo que se está dispuesto a correr al rechazar una hipótesis nula cierta. El valor de α se conoce como nivel de significación, término con el cual se quiere destacar que cualquier estadístico cuya probabilidad de ocurrencia sea igual o menor al valor de α , mantiene una diferencia tan grande con el valor del parámetro supuesto que se puede concluir que no pertenece a la distribución con la cual se está trabajando y por lo tanto asegurar que H_0 es falsa y otra hipótesis es la verdadera.

7.2.3. Estadístico de prueba

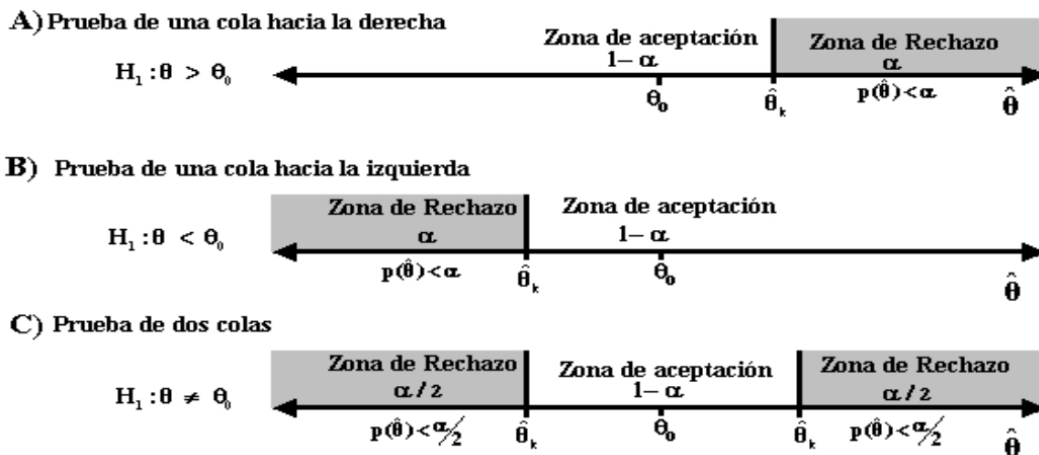
Para poder someter a prueba las hipótesis formuladas, es necesario usar alguna propiedad o estadístico de las muestras que esté relacionado con el parámetro objeto de la inferencia. Estas propiedades muestrales reciben el nombre genérico de estadísticos de prueba. Sin embargo, por razones prácticas, muchas veces los estadísticos de prueba no se usan en su forma original sino con otras formas equivalentes o derivadas (ver tabla).

Parámetro	Estadístico de prueba	Estadísticos de prueba derivados
Media (μ)	\bar{x}	$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$
		$z = (\bar{x} - \mu) / (s / \sqrt{n})$
		$t = (\bar{x} - \mu) / (s / \sqrt{n})$
Diferencia de medias ($\mu_2 - \mu_1$)	$\bar{x}_2 - \bar{x}_1$	$Z = (\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1) / \sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}$
		$Z = (\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1) / \sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}$
		$T = (\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1) / \sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}$
Varianza	S^2	$\chi^2 = (n-1)S^2 / \sigma_0^2$
Razón de varianzas	S_2^2 / S_1^2	$F = (s_2^2 \sigma_2^2) / (s_1^2 \sigma_1^2)$

La utilidad de estos y otros estadísticos de prueba se verá cuando se traten particularmente las PH para algunos parámetros.

7.2.4. Zona de aceptación

Una vez conocido el estadístico de prueba a utilizar, así como su distribución, es necesario definir en la distribución del estadístico muestral una zona de aceptación y una zona de rechazo de la hipótesis nula. La zona de aceptación de H_0 está formada por todos los valores del estadístico de prueba que ocurren con una probabilidad mayor a la establecida en el nivel de significación. Por el contrario, la zona de rechazo está formada por todos los valores del estadístico de prueba cuya probabilidad de ocurrencia es igual o menor al valor establecido en el nivel de significación. La zona de rechazo a diferencia de la zona de aceptación, y dependiendo de la hipótesis alternativa planteada, puede estar orientada en diferentes direcciones a lo largo del eje de valores de la variable aleatoria. Las definiciones serían: zona de rechazo a la derecha, a la izquierda y doble (ver figura).



Para concretar una decisión, es necesario encontrar un valor crítico ($\hat{\theta}_k$), el cuál es el valor del estadístico de prueba que separa la región de aceptación de la región de rechazo. Esto explica la importancia de conocer la distribución del estadístico de prueba. El valor requerido se obtiene usando las tablas de probabilidad acumulada de las distribuciones de probabilidad que estos estadísticos siguen. La cuantía del valor crítico depende, además de la distribución de probabilidad, del valor de α (ver tabla).

$\alpha = 0,100$	$\pm z_{(0,900)} = 1,29$	$\pm t_{(0,90; 10)} = 1,372$
$\alpha = 0,050$	$\pm z_{(0,950)} = 1,65$	$\pm t_{(0,95; 10)} = 1,812$
$\alpha = 0,025$	$\pm z_{(0,975)} = 1,96$	$\pm t_{(0,975; 10)} = 2,228$
$\alpha = 0,010$	$\pm z_{(0,990)} = 2,33$	$\pm t_{(0,99; 10)} = 2,764$

7.2.5. Cómputos

Con los datos proporcionados por una muestra de tamaño n se calcula el estadístico de prueba. La mayoría de las veces no se usa el estadístico de prueba directamente sino alguna de sus formas equivalentes, algunas de las cuales requieren para su uso que también se calcule la desviación estándar. La otra cantidad que hay que cuantificar es el valor crítico el cual depende del nivel de significación especificado y de la distribución probabilística que siga el estadístico de prueba.

7.2.6. Decisión

En la última etapa en el procedimiento de PH se debe tomar la decisión de rechazar o no la hipótesis nula. Si el estadístico de prueba cae dentro de la región de rechazo, se considera que la diferencia entre el parámetro y el estadístico de prueba es significativa y que la misma no puede atribuirse únicamente a las variaciones aleatorias de las muestras, por lo tanto se rechaza la hipótesis nula y se declara como falsa. Si por el contrario el estadístico de prueba se ubica en la zona de aceptación se considera que la diferencia entre el parámetro que y el estadístico de prueba es no significativa y que dicha diferencia es simplemente aleatoria, en consecuencia se puede aceptar la hipótesis nula planteada. Aquí es necesario recordar que la decisión de aceptar H_0 es una forma corta de decir que no existe suficiente evidencia para rechazarla y que en modo alguno se está concluyendo que la hipótesis nula es verdadera. Sólo se está aceptando temporalmente, hasta que se pruebe lo contrario.

7.2.7. Conclusión

En los inicios de ésta sección se dijo que la resolución de todo problema científico comenzaba con la formulación de las hipótesis de investigación, que luego eran transformadas en hipótesis estadísticas, que como hemos visto son las premisas sometidas al proceso de PH. De modo que para cerrar el ciclo del proceso, es necesario que las conclusiones estadísticas se transformen en conclusiones de investigación.

Finalmente es importante enfatizar que las decisiones de un investigador no tienen que ser siempre consecuentes con las decisiones estadísticas. Los métodos estadísticos sólo proporcionan elementos de juicios objetivos y poderosos, que deben ser tomados en cuenta por el investigador al momento de decidir, pero no son los únicos, hay otros elementos de juicio de naturaleza no estadística que el científico puede considerar para tomar una decisión. En otras palabras decidir entre dos o más alternativas siempre queda a juicio del investigador.

7.3. PH para una media poblacional

7.3.1. PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente y con varianza conocida

Ejemplo: Un médico traumatólogo afirma que el contenido de calcio en los huesos de mujeres que padecen osteoporosis después de aplicársele cierto tratamiento es mayor al valor promedio observado para la población femenina que padece esta enfermedad, el cual se sabe es igual a 270 mg/g con una desviación de 120 mg/g. Para probar su premisa el investigador determinó el contenido de calcio en los huesos de 36 individuos que fueron sometidos al tratamiento y pudo determinar que dicha muestra arroja un valor promedio de calcio igual a 310 mg/g. La concentración de calcio es una variable que se distribuye normalmente.

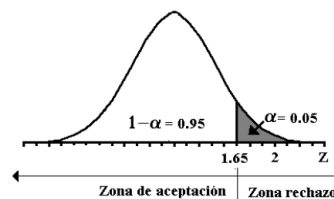
Las hipótesis de investigación son las siguientes:

H_0 : el tratamiento para la osteoporosis no tiene ningún efecto.

H_1 : el tratamiento para la osteoporosis aumenta los niveles de calcio en los huesos.

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: $H_0 : \mu = 270$ y $H_1 : \mu > 270$.
2. Especificación del valor crítico o nivel de significación: $\alpha = 0,05$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: ya que la variable se distribuye normalmente con varianza conocida lo más conveniente es usar $Z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$.
4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu > \mu_0$, se trata de una prueba de una cola hacia la derecha, siendo la zona de aceptación $ZA = \{Z/Z \leq z_{1-\alpha}\}$.



5. Cálculos necesarios:

$$Z = (\bar{x} - \mu)/(\sigma/\sqrt{n}) = (310 - 270)/(120/\sqrt{36}) = 40/20 = 2$$

$$ZA = \{Z/Z \leq z_{0,95}\} = \{Z/Z \leq 1,65\}$$

6. Decisión: Como $Z = 2 > z_{0,95} = 1,65$ el valor del estadístico de prueba se encuentra dentro de la zona de rechazo. Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Podemos afirmar que se tiene un 95 % de confianza que el tratamiento aplicado a los pacientes enfermos de osteoporosis aumenta el nivel de calcio en los tejidos óseos.

7.3.2. PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente con varianza desconocida y tamaño de muestra grande ($n \geq 30$)

Ejemplo: Un entomólogo sospecha que en cierta zona endémica para el dengue el valor de la tasa neta reproductiva (R_0) de una población del mosquito *Aedes aegypti* vector de dicha enfermedad, ha cambiado en relación con el valor determinado hace 5 años el cual era igual a 205 individuos. Con tal propósito determinó el valor de R_0 a 40 hembras criadas en el laboratorio y pertenecientes a una cepa desarrollada a partir de mosquitos capturados en la zona estudiada. Los resultados pueden verse en la tabla. El investigador sabe que la variable se distribuye normalmente y quiere someter a prueba su hipótesis no queriendo equivocarse en más del 5% de las veces.

Nº	R_c	Nº	R_c	Nº	R_c	Nº	R_c
1	228	11	201	21	141	31	144
2	173	12	212	22	169	32	226
3	182	13	162	23	163	33	228
4	197	14	282	24	159	34	192
5	205	15	216	25	192	35	205
6	260	16	181	26	231	36	237
7	233	17	249	27	257	37	223
8	289	18	174	28	174	38	226
9	158	19	196	29	206	39	182
10	199	20	220	30	149	40	195

Las hipótesis de investigación son las siguientes:

H_0 : la tasa neta de reproducción no ha cambiado.

H_1 : la tasa neta de reproducción se modificó después de 5 años.

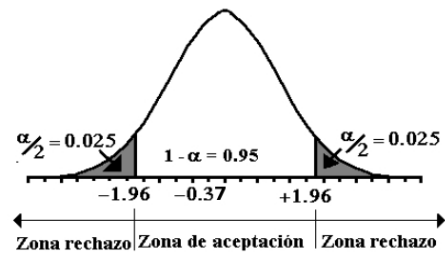
Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: $H_0 : \mu = 205$ y $H_1 : \mu \neq 205$.
2. Especificación del valor crítico o nivel de significación: $1 - \alpha = 0,95$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: ya que la variable se distribuye normalmente con varianza desconocida y tamaño grande, lo más conveniente es usar $Z = (\bar{x} - \mu)/(s/\sqrt{n})$.
4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu \neq \mu_0$, se trata de una prueba de dos colas, siendo la zona de aceptación $ZA = \{Z/ - z_{1-\alpha/2} < Z < +z_{1-\alpha/2}\}$.
5. Cálculos necesarios: $\bar{x} = 202,9$, $s = 36,17$,

$$Z = (\bar{x} - \mu)/(s/\sqrt{n}) = (202,9 - 205)/(36,17/\sqrt{40}) = -2,1/5,719 = -0,37$$

$$ZA = \{Z/ - z_{(0,975)} < Z < +z_{(0,975)}\} = \{Z/ - 1,96 < Z < +1,96\}$$

6. Decisión: Como $Z = -0,37$, el valor del estadístico de prueba se encuentra dentro de la zona de aceptación de H_0 . Por lo tanto se concluye que los datos no proporcionan suficiente evidencia para rechazar H_0 (ver figura).
7. Conclusión: La sospecha del investigador que la tasa de reproducción de la población de mosquitos se había modificado fue rechazada con un 95 % de confianza a la luz de la información proporcionada por la muestra.



7.3.3. PH para una media pobl. cuando la muestra proviene de una población distribuida normalmente con varianza desconocida y tamaño de muestra pequeño ($n < 30$)

Ejemplo: Un fisiólogo vegetal desea verificar si el contenido de nitrógeno en las hojas jóvenes de la especie *Rhizophora mangle*, es menor en las plantas que viven en una zona ambientalmente protegida con relación a las que viven en una zona que está siendo afectada por la contaminación con fertilizantes y cuyo valor promedio se cuantificó en 14,6 mg/g de nitrógeno. El análisis de 25 hojas jóvenes provenientes de la zona protegida produjo una media muestral de 10,48 con una desviación estándar de 2,41. Si la concentración de nitrógeno se distribuye normalmente, apoya la evidencia proporcionada por la muestra la presunción que las plantas de la zona protegida contienen menos nitrógeno?. El error tipo I no debe ser mayor al 1 %.

Las hipótesis de investigación son las siguientes:

H_0 : la concentración de N_2 en las hojas jóvenes de *R. mangle* en ambas regiones es la misma.

H_1 : la concentración de N_2 en las hojas jóvenes de *R. mangle* es menor en la región protegida.

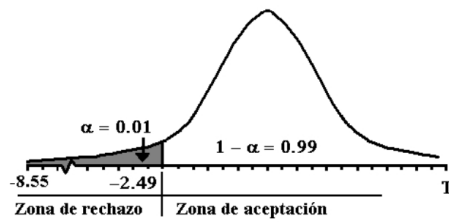
Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: $H_0 : \mu = 14,6$ y $H_1 : \mu < 14,6$.
2. Especificación del valor crítico o nivel de significación: $1 - \alpha = 0,99$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: ya que la variable se distribuye normalmente con varianza desconocida y tamaño pequeño, lo más conveniente es usar $T = (\bar{x} - \mu)/(s/\sqrt{n})$.
4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu < \mu_0$, se trata de una prueba de una cola hacia la izquierda, siendo la zona de aceptación $ZA = \{T/ - t_{(1-\alpha;n-1)} \leq T\}$.
5. Cómputos necesarios:

$$T = (\bar{x} - \mu)/(s/\sqrt{n}) = (10,48 - 14,6)/(2,41/\sqrt{25}) = -4,12/0,482 = -8,55$$

$$ZA = \{T/ - t_{(0,99;24)} \leq T\} = \{T/ - 2,492 \leq T\}$$

6. Decisión: Como $t = -8,55 \ll -t_{(0,99;24)} = -2,492$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 (ver figura).
7. Conclusión: Se puede afirmar con un 99% de confianza que la concentración de nitrógeno en las hojas de *Rhizophora mangle* en ambas regiones es diferente.



7.3.4. PH para una media pobl. cuando la muestra proviene de una población con distribución no normal y tamaño de muestra grande ($n \geq 30$)

Ejemplo: En cierto nervio del cuerpo humano, los impulsos eléctricos viajan a una velocidad promedio de 4,3 m/seg con una desviación igual a 1,2 m/seg. Un fisiólogo observó que la velocidad promedio de conducción del impulso eléctrico en 45 individuos con una distrofia fue de 3,7 m/seg. Basado en estos resultados el investigador presume que con relación a los individuos sanos en los individuos con distrofia el impulso eléctrico viaja a menor velocidad en el nervio estudiado. Soportan ésta hipótesis los resultados obtenidos?.

Las hipótesis de investigación son las siguientes:

H_0 : la velocidad del impulso nervioso es igual en los individuos con distrofia y en los individuos normales.

H_1 : la velocidad del impulso nervioso es menor en los individuos con distrofia que en los individuos normales..

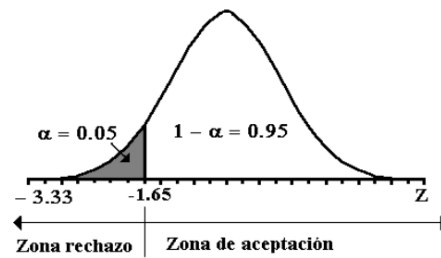
Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: $H_0 : \mu = 4,3$ y $H_1 : \mu < 4,3$.
2. Especificación del valor crítico o nivel de significación: $1 - \alpha = 0,95$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: aunque no se conoce la distribución de la variable, como el tamaño de la muestra es grande se aplica el Teorema del Límite Central y por lo tanto la media muestral se distribuye normalmente, por lo que lo más conveniente es usar $Z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$.
4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu < \mu_0$, se trata de una prueba de una cola hacia la izquierda, siendo la zona de aceptación $ZA = \{Z/ - z_{(1-\alpha)} \leq Z\}$.
5. Cómputos necesarios:

$$Z = (\bar{x} - \mu)/(\sigma/\sqrt{n}) = (3,7 - 4,3)/(1,2/\sqrt{45}) = -0,6/0,18 = -3,354$$

$$ZA = \{Z/ - z_{(0,95)} \leq Z\} = \{Z/ - 1,65 \leq Z\}$$

6. Decisión: Como $z = -3,354 < -z_{(0,95)} = -1,65$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 (ver figura).
7. Conclusión: Los datos soportan la suposición del investigador que en los individuos con distrofia la velocidad de transmisión del impulso nervioso es menor a la observada en individuos normales.



7.4. PH para dos medias poblacionales

Posiblemente la situación más frecuente de investigación en el campo de las ciencias naturales sea la de decidir entre dos alternativas. Por lo general cuando se requiere escoger entre dos métodos se recurre a una prueba de hipótesis para dos medias poblacionales. Esta prueba consiste básicamente en determinar si dos muestras estiman la misma media poblacional, ya sea porque se supone que las muestras provienen de una misma población o de poblaciones diferentes con la misma media.

7.4.1. PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente y con varianza conocidas

Ejemplo: De acuerdo a los estudios efectuados sobre el contenido de estroncio en los seres humanos se sabe que ésta variable se distribuye normalmente con varianza 144. Los mismos estudios indican que el contenido de este elemento en los huesos disminuye con la edad de las personas. En una investigación relacionada con éste problema, un químico determinó mediante la espectrofotometría de absorción atómica, el contenido de estroncio en muestras de huesos fracturados de pacientes femeninos pertenecientes a dos grupos etáreos diferentes. Los resultados pueden verse en la tabla. Estos resultados apoyan la hipótesis de la disminución de los niveles de estroncio en el tejido óseo al incrementar la edad de las personas? Use $\alpha = 0.03$.

Niveles de estroncio $\mu\text{g/g}$	
35-44 años	45-54 años
40,45	48,21
55,15	23,37
67,59	25,42
80,58	41,94
78,09	40,65
68,09	44,75
72,06	51,69

Las hipótesis de investigación son las siguientes:

H_0 : el contenido de estroncio en los huesos no se modifica con la edad de las personas.

H_1 : el contenido de estroncio en los huesos disminuye con la edad de las personas.

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: si se considera que la población de edades entre 35 y 44 años tiene una media μ_1 y que la población con edades entre 45 y 54 años tiene

una media μ_2 , las hipótesis estadísticas a probar son:

$$H_0 : \mu_1 = \mu_2 \text{ o } \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 > \mu_2 \text{ o } \mu_1 - \mu_2 > 0$$

2. Especificación del valor crítico o nivel de significación: $\alpha = 0,03$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: puesto que la variable se distribuye normalmente con varianza conocida y como se trata de una PH sobre diferencia de medias poblacionales lo más conveniente es usar

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu_1 > \mu_2$, se trata de una prueba de una cola hacia la derecha, siendo la zona de aceptación $ZA = \{Z/Z \leq z_{(1-\alpha)}\}$.
5. Cálculos necesarios: $\bar{x}_1 = 66,0$, $\bar{x}_2 = 39,43$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(66,0 - 39,43) - 0}{\sqrt{\frac{144}{7} + \frac{144}{7}}} = \frac{26,57}{6,41} = 4,14$$

$$ZA = \{Z/Z \leq z_{(0,970)}\} = \{Z/Z \leq 1,88\}$$

6. Decisión: Como $z = 4,14 \gg z_{(0,970)} = 1,88$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Se puede concluir con un 97% de confianza que la evidencia aportada por la muestra apoya la hipótesis de la disminución del nivel de estroncio en los huesos de las personas con la edad.

7.4.2. PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente, con varianza desconocidas y tamaño de muestras grandes ($n_1, n_2 \geq 30$)

Ejemplo: En el departamento de toxicología del ministerio de salud se necesita saber si el contenido de nicotina en dos marcas de cigarrillos importados es la misma. Con el propósito de resolver la situación se le determina el contenido de nicotina a un lote de cigarrillos de cada marca, encontrándose los resultados de la tabla. Si se sabe que la cantidad de nicotina se distribuye normalmente, determine con un nivel de confianza del 10% si las dos marcas tienen la misma cantidad de nicotina.

	Contenido de nicotina (mg)	
	Marca "Kill me softly"	Marca "Little life"
n	49	36
Media	24,0	25,2
Desviación estándar	2,30	2,90

Las hipótesis de investigación son las siguientes:

H_0 : la cantidad de nicotina en los cigarrillos de las dos marcas es la misma.

H_1 : la cantidad de nicotina en los cigarrillos de las dos marcas es diferente.

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: si se considera μ_1 y μ_2 como el valor promedio del contenido de nicotina en los cigarrillos "Kill me softly" "Little lifes" respectivamente, las hipótesis a probar son

$$H_0 : \mu_1 = \mu_2 \text{ o } \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \text{ o } \mu_1 - \mu_2 \neq 0$$

2. Especificación del valor crítico o nivel de significación: $\alpha = 0,10$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: puesto que la variable se distribuye normalmente con varianza desconocida y tamaño grande, lo más conveniente es usar

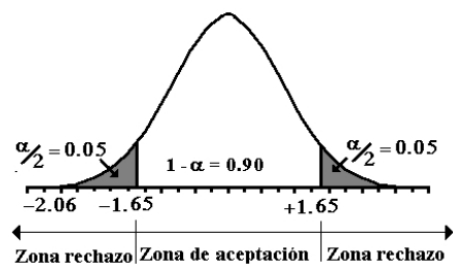
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu_1 \neq \mu_2$, se trata de una prueba de dos colas, siendo la zona de aceptación $ZA = \{Z/ - z_{(1-\alpha/2)} \leq Z \leq +z_{(1-\alpha/2)}\}$.
5. Cálculos necesarios: $s_1^2 = 5,29$, $s_2^2 = 8,41$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(24,0 - 25,2) - 0}{\sqrt{\frac{5,29}{49} + \frac{8,41}{36}}} = -2,06$$

$$ZA = \{Z/ - z_{(0,95)} \leq Z \leq +z_{(0,95)}\} = \{Z/ - 1,65 \leq Z \leq 1,65\}$$

6. Decisión: Como $z = -2,06 < z_{(0,95)} = -1,65$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Se puede concluir que la evidencia aportada por la muestra apoya como hipótesis que el contenido de nicotina en las dos marcas es diferente.



7.4.3. PH para dos medias pobl. cuando las muestras provienen de poblaciones distribuidas normalmente, con varianza desconocidas y tamaño de muestras pequeñas ($n_1, n_2 < 30$)

Ejemplo: En un estudio sobre la condición ecológica de los ríos altiandinos, se determinó la temperatura del agua en ríos de páramo y de selva nublada, obteniéndose los resultados siguientes:

	Temperatura del agua (°C)															
Ríos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Páramo	10,5	15,0	14,5	8,5	7,5	13,5	15,0	11,5	17,0	13,0	13,5	14,5	13,5	15,0	10,5	10,0
Selva	19,5	17,0	13,5	9,0	12,0	16,5	16,5	18,0	18,0	18,0	12,0	16,0	12,0	14,5	16,5	17,0

Conociendo que la temperatura del agua es una variable que se distribuye normalmente, se quiere poner a prueba la hipótesis que predice que la temperatura promedio de los ríos de selva nublada supera la temperatura de los ríos de páramo.

Las hipótesis de investigación son las siguientes:

H_0 : la temperatura del agua en los ríos es la misma en las dos unidades ecológicas.

H_1 : la temperatura del agua es mayor en los ríos de la zona selva.

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: si se considera a μ_1 y μ_2 como el valor promedio de la temperatura del agua en los ríos de páramo y de selva respectivamente, las hipótesis estadísticas a probar son:

$$H_0 : \mu_2 - \mu_1 = 0$$

$$H_1 : \mu_2 - \mu_1 > 0$$

2. Especificación del valor crítico o nivel de significación: $\alpha = 0,05$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: como la variable se distribuye normalmente con varianzas desconocidas y el tamaño es pequeño, para poder seleccionar el estadístico de prueba a usar, se debe determinar si las varianzas poblacionales se pueden considerar iguales o diferentes. Para esto se puede hacer uso de las reglas prácticas para la comparación de varianzas. Como $\alpha = 0,05$ y $RV = s_2^2/s_1^2 = (2,9)^2/(2,66)^2 = 1,19$ es menor a 2.5 se acepta que las dos varianzas son iguales. Por lo tanto se debe usar como estadístico de prueba

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu_2 - \mu_1 > 0$, se trata de una prueba de una cola hacia la derecha, siendo la zona de aceptación $ZA = \{T/T \leq t_{(1-\alpha; n_1+n_2-2)}\}$.
5. Cálculos necesarios: $\bar{x}_1 = 12,688$, $\bar{x}_2 = 15,375$, $s_1 = 2,66$, $s_2 = 2,9$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(16 - 1)(2,66)^2 + (16 - 1)(2,9)^2}{16 + 16 - 2} = 7,74$$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} = \frac{(15,38 - 12,69) - 0}{\sqrt{\frac{7,74}{16} + \frac{7,74}{16}}} = \frac{2,69}{0,9836} = 2,73$$

$$ZA = \{T/T \leq t_{(0,95;30)}\} = \{T/T \leq 1,697\}$$

6. Decisión: Como $z = 2,73 > t_{(0,95;30)} = 1,697$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Se puede concluir que se tiene un 95 % de confianza que la temperatura del agua es mayor en los ríos de selva nublada, que en los ríos de páramo.

7.4.4. PH para dos medias pobl. cuando las muestras provienen de poblaciones con distribución no normal y tamaño de muestras grandes ($n_1, n_2 \geq 30$)

Ejemplo: Se sabe que el contenido de calcio en los huesos de los animales de cierta especie se distribuye normalmente con una varianza 57.6 para las hembras y 51.2 para los machos. Con el propósito de determinar si existen diferencias en el contenido de calcio entre machos y hembras se le determinó a 31 hembras y 33 machos el contenido de calcio en el tejido óseo, encontrándose que para la muestra de hembras el valor promedio fue de 400.45 $\mu\text{g/g}$ y para la muestra de machos fue de 395.24 $\mu\text{g/g}$. Cuál debe ser la respuesta?. Use $\alpha = 0,05$.

Las hipótesis de investigación son las siguientes:

H_0 : el contenido de calcio en los huesos de los animales de los dos sexos es el mismo.

H_1 : el contenido de calcio en los huesos de los animales de ambos sexos es diferente.

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis: si se considera μ_1 y μ_2 como el valor promedio de la concentración de calcio en hembras y machos respectivamente, las hipótesis a probar son

$$H_0 : \mu_1 = \mu_2 \text{ o } \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \text{ o } \mu_1 - \mu_2 \neq 0$$

2. Especificación del valor crítico o nivel de significación: $\alpha = 0,05$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: aunque no se conoce la distribución de la variable, como el tamaño de la muestra es grande, se aplica el Teorema del Límite Central. Por lo tanto, las medias muestrales se distribuyen normalmente por lo que lo más conveniente es usar

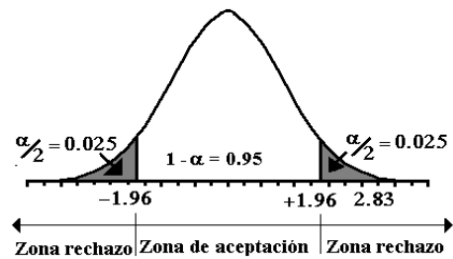
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4. Establecer una zona de aceptación para H_0 : Como $H_1 : \mu_1 \neq \mu_2$, se trata de una prueba de dos colas, siendo la zona de aceptación $ZA = \{Z/ -z_{(1-\alpha/2)} \leq Z \leq +z_{(1-\alpha/2)}\}$.
5. Cómputos necesarios:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(400,45 - 395,24) - 0}{\sqrt{\frac{57,6}{31} + \frac{51,2}{33}}} = \frac{5,21}{1,84} = 2,83$$

$$ZA = \{Z/ - z_{(0,975)} \leq Z \leq +z_{(0,975)}\} = \{Z/ - 1,96 \leq Z \leq 1,96\}$$

6. Decisión: Como $z = 2,83 > z_{(0,975)} = 1,96$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Se puede afirmar con un 95 % de confianza que el nivel de calcio en los huesos de los animales de los dos sexos es diferente.



7.5. PH para dos varianzas poblacionales

Para efectuar algunas comparaciones de medias poblacionales se debe averiguar si las muestras proceden de poblaciones con la misma varianza. Sin embargo este conocimiento también es importante para otro tipo de situación. Por ejemplo, al comparar la precisión de dos métodos, o al confrontar la variabilidad característica presente en dos individuos, dos poblaciones, dos procesos, etc. De modo que es muy valioso disponer de un método estadístico que, con mayor formalidad que las reglas prácticas dadas, precise si dos varianzas son o no homogéneas. Una forma de hacerlo es comparar mediante una prueba de hipótesis las varianzas poblacionales. Para esto es necesario, además de plantear las hipótesis, disponer de un estadístico de prueba y del modelo de distribución de probabilidad que este estadístico sigue. Afortunadamente, ambas cosas se conocen. Esta PH tiene como condición que las muestras sean independientes y las dos poblaciones estén distribuidas normalmente.

El planteamiento de las hipótesis sobre las varianzas es algo particular por el hecho de que las varianzas no son aditivas. Por lo que escribiremos las hipótesis de la siguiente manera:

$$H_0 : \sigma_2^2 = \sigma_1^2 \text{ o } \sigma_2^2/\sigma_1^2 = 1$$

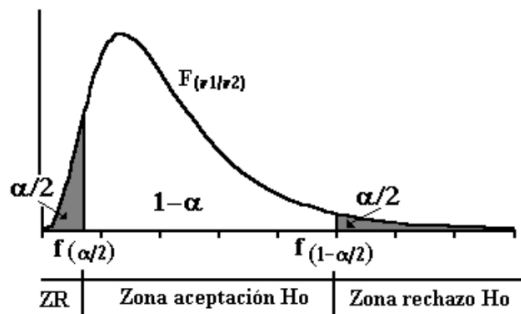
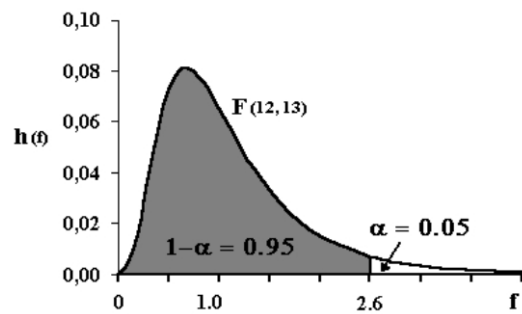
$$H_1 : \begin{cases} \sigma_2^2 \neq \sigma_1^2 \text{ o } \sigma_2^2/\sigma_1^2 \neq 1 \\ \sigma_2^2 > \sigma_1^2 \text{ o } \sigma_2^2/\sigma_1^2 > 1 \\ \sigma_2^2 < \sigma_1^2 \text{ o } \sigma_2^2/\sigma_1^2 < 1 \end{cases}$$

Como estadístico de pruebas se usa la razón de las varianzas muestrales, $F_0 = s_2^2/s_1^2$. Si las muestras provienen de dos poblaciones con la misma varianza o de una misma población, la distribución de probabilidades de la razón de varianzas sigue el modelo probabilístico conocido como distribución F , cuya función de probabilidad es

$$h(f) = d_1^{d_1/2} d_2^{d_2/2} \frac{\Gamma(d_1/2 + d_2/2)}{\Gamma(d_1/2)\Gamma(d_2/2)} \frac{f^{d_1/2-1}}{(d_1 f + d_2)^{d_1/2+d_2/2}}$$

para $f > 0$ y $h(f) = 0$ para $f \leq 0$. Los parámetros d_1 y d_2 son los grados de libertad que describen a la variable f y son estimados a partir de los tamaños de las muestras menos uno: $d_1 = n_1 - 1$, y $d_2 = n_2 - 1$.

Dada la utilidad de la distribución F para muchos métodos estadísticos, se han elaborado tablas de la su función acumulada para diferentes valores de d_1 y d_2 . Por ejemplo, si se tiene que $d_1 = 12$ y $d_2 = 13$, entonces un 0.95 del área bajo la curva de F se encuentra a la izquierda del percentil $f_{2,6}$ (ver figura de la derecha). Suponiendo que la razón de varianzas de dos muestras es menor al valor límite 2.6, eso significa que su probabilidad de ocurrencia es mayor a 0,05. En éste caso se considera que las diferencias entre las dos varianzas muestrales son aleatorias. Pero si la razón de varianzas es mayor a 2.6, es porque su probabilidad de ocurrencia es menor a 0,05, de lo que se deduce que las diferencias entre las dos varianzas muestrales no son simplemente fortuitas y por tanto las varianzas son diferentes.



En términos generales se puede decir que cuando se trata de una prueba con una cola a la derecha el valor $f_{(1-\alpha;d_1/d_2)}$ define el límite entre las zonas de aceptación y rechazo de la hipótesis nula. Cuando la prueba de hipótesis es de dos colas, debido a la asimetría de la distribución F, la zona de rechazo de H_0 es diferente para ambos lados de la distribución. El valor $f_{(1-\alpha/2;d_1/d_2)}$ sería el límite de la derecha y el valor $f_{(\alpha/2;d_1/d_2)}$ el límite de la izquierda (ver figura de la izquierda).

Aquí surge un pequeño inconveniente, porque las tablas de la función acumulada sólo presentan valores de f para la cola derecha. Esta situación se puede solventar de dos maneras. La forma más fácil es plantear las hipótesis de modo que la varianza muestral mayor siempre quede en el numerador. La otra solución es calcular el valor crítico de la cola izquierda mediante la expresión siguiente:

$$f_{(\alpha/2;d_1/d_2)} = \frac{1}{f_{(1-\alpha/2;d_2/d_1)}}$$

Por ejemplo, $f_{(0,975;8/12)} = 3,5$ es el límite crítico para la cola de la derecha, sin embargo en las tablas no se encuentra el valor de $f_{(0,025;8/12)}$ que sería el límite crítico de la cola de la izquierda, pero se puede calcular usando la relación anterior. En primer lugar se encuentra el valor $f_{(0,975;12/8)} = 4,2$, luego se obtiene el inverso de 4.2, siendo entonces $f_{(0,025;8/12)} = 0,238$. Este mismo procedimiento se debe usar para calcular el valor crítico de la zona de rechazo cuando la prueba de hipótesis es de una cola a la izquierda.

Ejemplo: En un estudio taxonómico sobre una especie de insecto se quiere usar una característica morfológica del cuerpo para estimar el tamaño de los adultos. Se escogerá como característica aquella que tenga la menor variabilidad. Con éste propósito se midieron en 10 individuos la longitud del ala anterior y la longitud total del cuerpo. Con base a los resultados que se presentan a continuación y sabiendo que las dos variables se distribuyen normalmente, escoja la que mejor estima el tamaño de los insectos.

Nº de Individuo	1	2	3	4	5	6	7	8	9	10
Alas anteriores (mm)	17,1	17	17,1	16,3	16,9	15,9	16,2	17,2	17,1	16,8
Tamaño del cuerpo (mm)	17,6	16,5	15,5	16,9	17,1	15,2	16,7	17,7	16,9	15,1

Ahora describamos el proceso de PH:

1. Formulación de la hipótesis:

$$H_0 : \sigma_2^2/\sigma_1^2 = 1$$

$$H_1 : \sigma_2^2/\sigma_1^2 \neq 1$$

2. Especificación del valor crítico o nivel de significación: $\alpha = 0,05$.
3. Elección de un estadístico de la muestra y de su distribución para someter a prueba la hipótesis: puesto que se trata de la comparación de dos varianzas, el estadístico de prueba es

$$F_0 = \frac{s_2^2}{s_1^2}$$

4. Establecer una zona de aceptación para H_0 : Como $H_1 : \sigma_2^2/\sigma_1^2 \neq 1$, se trata de una prueba de dos colas, siendo la zona de aceptación $ZA = \{F/f_{(\alpha/2;n_2-1/n_1-1)} \leq F \leq f_{(1-\alpha/2;n_2-1/n_1-1)}\}$.
5. Cómputos necesarios: $s_2^2 = 0,8907$, $s_1^2 = 0,2093$, $d_2 = n_2 - 1 = 9$, $d_1 = n_1 - 1 = 9$

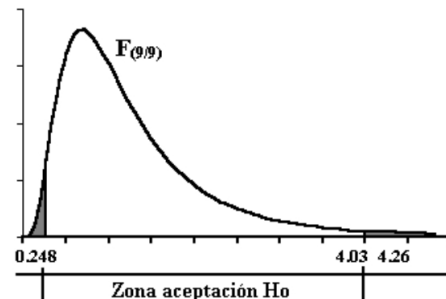
$$F_0 = \frac{s_2^2}{s_1^2} = \frac{0,8907}{0,2093} = 4,26$$

$$ZA = \{F/f_{(0,025;9/9)} \leq F \leq f_{(0,975;9/9)}\}$$

$$f_{(0,025;9/9)} = 1/f_{(0,975;9/9)} = 1/4,03 = 0,248$$

$$ZA = \{F/0,248 \leq F \leq 4,03\}$$

6. Decisión: Como $F_0 = 4,26 > f_{(0,975;9/9)} = 4,03$, el valor del estadístico de prueba se encuentra dentro de la zona de rechazo de H_0 . Por lo tanto se concluye que los datos proporcionan suficiente evidencia para rechazar H_0 .
7. Conclusión: Se puede afirmar con un 95% de confianza que las varianzas de las dos variables morfométricas son diferentes, siendo la longitud de las alas una variable más homogénea.



8. Inf. Est.: Prueba de Hipótesis (II)

Siguiendo con el método de PH, en esta sección nos concentraremos en dos problemas muy comunes en la estadística inferencial. El primero de estos problemas es poder distinguir cuando dos distribuciones surgen de la misma función distribución o provienen de funciones de distribución diferentes. Determinar que dos distribuciones son diferentes o mostrar que ellas son consistentes es una tarea que surge constantemente en muchas áreas de la investigación. Este problema además puede subdividirse en dos casos: cuando los datos provienen de variables discretas o de variables continuos. A continuación, describiremos dos métodos para afrontar cada uno de estos casos: el método Chi-cuadrado para tratar datos binned de variable discreta, y el método de Kolmogorov-Smirnov para analizar datos provenientes de variables aleatorias continuas como función de una sola variable. Por último, el segundo problema que trataremos de resolver consta en poder cuantificar la independencia estadística de un conjunto de datos.

8.1. Método Chi-cuadrado

El método Chi-cuadrado, usualmente denominado "de Pearson." o "de asociación", es considerado como una prueba no paramétrica que mide la discrepancia entre una distribución observada y otra teórica, indicando en qué medida las diferencias existentes entre ambas, de haberlas, se deben al azar en la PH. En particular, el método de Chi-cuadrado compara histogramas con distribuciones de probabilidad discretas. El método también funciona para funciones de distribución discretizadas, donde las probabilidades se obtienen integrando las funciones distribución sobre los diferentes bins. El método tiene dos variantes:

1. para comparar un histograma con una función de probabilidad acumulada discretizada. La hipótesis nula H_0 es *la muestra sigue la distribución de probabilidades dada*.
2. para comparar dos histogramas obtenidos de dos muestras diferentes. La hipótesis nula H_0 es *las dos muestras siguen la misma distribución*.

Caso 1: El estadístico del método viene dado por la siguiente fórmula:

$$\chi^2 = \sum_i \frac{(\text{observada}_i - \text{teórica}_i)^2}{\text{teórica}_i}$$

Cuanto mayor sea el valor de χ^2 , menos verosímil es que la hipótesis sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de chi-cuadrado, más ajustadas están ambas distribuciones. Los grados de libertad ν vienen dados por $\nu = N - 1$ donde N es el número de bins.

La función distribución relacionada con este estadístico es la función homónima chi-cuadrado, $q(\chi^2; \nu)$, y viene dada por la fórmula

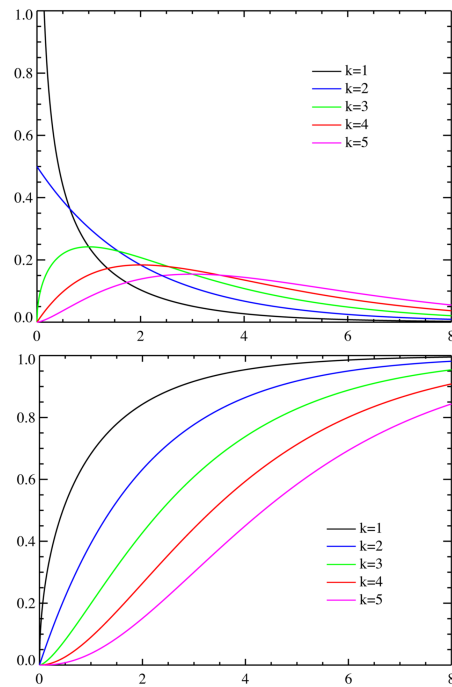
$$q(\chi^2; \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} (\chi^2)^{(\nu/2)-1} e^{-\chi^2/2}$$

para $\chi^2 \geq 0$ y $q(\chi^2; \nu) = 0$ para $\chi^2 < 0$, donde Γ es la función gamma (figura superior). Su función de distribución acumulada es

$$Q(\chi^2; \nu) = \frac{\gamma(\nu/2, \chi^2/2)}{\Gamma(\nu/2)}$$

donde γ es la función gamma incompleta (figura inferior). El valor esperado y la varianza de una variable aleatoria χ^2 con distribución chi-cuadrado son, respectivamente, ν y 2ν .

Estrictamente hablando, $Q(\chi^2; \nu)$ es la probabilidad de que la suma de los cuadrados de ν variables aleatorias normales, por unidad de varianza, sea mayor que χ^2 . Los términos en la sumatoria del estadístico χ^2 no son individualmente normales. Sin embargo, si el número de los bins es grande o el número de eventos en cada bin es grande, entonces la función de probabilidad chi-cuadrado es una buena aproximación a la distribución del estadístico en el caso de la hipótesis nula. Por lo tanto, las tablas de la distribución acumulada de esta función es la que se usa para hacer las estimas necesarias en el método chi-cuadrado. En consecuencia, se acepta H_0 cuando $\chi^2 < \chi^2_{(1-\alpha, \nu)}$. En caso contrario se rechaza. Observar que α representa el nivel de significación estadística elegido. A continuación se muestra una tabla donde se muestran los valores de χ^2 para 10 valores de grados de libertad. En la última fila se incluye la $P(\chi^2 \leq \chi^2_{(1-\alpha, \nu)})$.



ν	χ^2										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P	0.05	0.10	0.20	0.30	0.50	0.70	0.80	0.90	0.95	0.99	0.999

Ejemplo: Supongamos que en una escuela las estadísticas de años pasados muestran que, la comisión de admisión tiende a aceptar 4 alumnos por 1 que se rechaza. Este año una comisión constituida por un grupo diferentes de personas, aceptó 275 y rechazó 55. Se puede decir que esta nueva comisión difiere de manera significativa con la razón de rechazo de la comisión anterior?

La prueba estadística para determinar la significatividad de la diferencia en las frecuencias observadas es la prueba de chi-cuadrado. Lo que se hace al aplicar la fórmula de chi-cuadrado es restar al número de frecuencias observadas, el número de frecuencias esperadas; elevar esta diferencia al cuadrado, lo que hace que todos los valores asuman un valor positivo, y luego se divide el cuadrado obtenido entre las frecuencias esperadas. Esto se hace de manera independiente para cada una de las categorías. Una vez terminado este paso, se suman los resultados obtenidos en cada categoría y ese valor resultante de la suma es el valor χ^2 observado, el cual deberá ser comparado con el valor chi-cuadrado crítico, $\chi^2_{(1-\alpha, \nu)}$ según el nivel de significación escogido y los grados de libertad correspondientes.

En nuestro caso, tenemos 330 casos en total. Si la comisión anterior hubiera actuado se esperarían que aceptaran 264 alumnos y rechazaran 66. Así pues tomamos estos números (razón 4:1) como las frecuencias esperadas en cada caso. El cálculo correspondiente es

$$\chi^2 = \frac{(275 - 264)^2}{264} + \frac{(55 - 66)^2}{66} = 0,4589 + 1,83 = 2,29$$

El grado de libertad del problema viene de analizar que los datos están distribuidos en una tabla 2×2 , por lo que $\nu = (\text{filas} - 1)(\text{columnas} - 1) = 1 \times 1 = 1$.

Al comparar el valor χ^2 obtenido con el valor crítico de un grado de libertad y 0,05 de significación (ver tabla de la página anterior), es decir $\chi^2_{(0,95;1)} = 3,841$, vemos que el valor crítico es mayor que el observado, por lo que no se puede desacreditar la hipótesis nula y se concluye que la nueva comisión no muestra una política diferente a la de la comisión anterior.

Caso 2: Por último, en el caso de que quisieramos comparar dos histogramas, el estadístico viene dado por

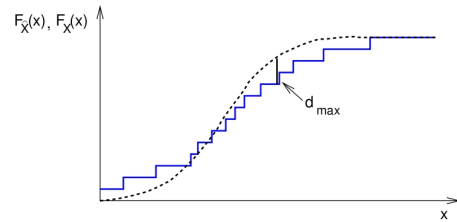
$$\chi^2 = \sum_i \frac{(\text{observada}_{1,i} - \text{observada}_{2,i})^2}{\text{observada}_{1,i} + \text{observada}_{2,i}}$$

donde la sumatoria corre sobre todos los bins que contribuyen. Observar que el denominador no es el promedio de las dos observaciones, es dos veces el promedio. La razón de esto es que cada término de una suma chi-cuadrado se supone que se aproxima al cuadrado de una cantidad normalmente distribuida con varianza unitaria. La varianza de la diferencia de dos cantidades normalmente distribuidas es la suma de sus varianzas individuales, no el promedio.

8.2. Método de Kolmogorov-Smirnov

Consideremos el caso donde las propiedades estadísticas de una muestra obtenidas a partir de experimentos repetidos usando variables aleatorias continuas, se quiere comparar con una función distribución de probabilidades F_X . Uno podría, en principio, comparar un histograma y su correspondiente distribución de probabilidades bineada usando el método de chi-cuadrado explicado anteriormente. Desafortunadamente, el bineado es artificial y tiene gran influencia en los resultados. Consecuentemente, el método presentado en esta sección es más útil ya que no requiere de ningún bineado.

El método se denomina de Kolmogorov-Smirnov (KS) y compara funciones distribución F_X con funciones de distribución empíricas $F_{\hat{X}}$. Uno podría elegir diferentes maneras para comparar las distribuciones, por ejemplo, calcular el área entre las curvas F_X y $F_{\hat{X}}$. El método KS eligió una simple medición: definir el valor máximo del modulo de la diferencia entre dos funciones de distribución acumuladas. Es decir, el estadístico es



$$d_{max} \equiv \max_{-\infty < x < \infty} |F_X(x) - F_{\hat{X}}(x)|$$

Así mismo, si se quisieran comparar dos distribuciones acumuladas observadas, el estadístico sería

$$d_{max} \equiv \max_{-\infty < x < \infty} |F_{\hat{X}_1}(x) - F_{\hat{X}_2}(x)|$$

Lo que hace útil al método KS es que su distribución, en el caso de la hipótesis nula (datos extraídos de la misma distribución), puede ser calculada, al menos una aproximación, dando la significación de cualquier valor distinto de cero para d_{max} . Una característica del método KS es que es invariante bajo reparametrizaciones de la variable x , es decir, se puede comprimir o alargar el eje x , y la distancia máxima permaneciera invariante. La función acumulada involucrada en el cálculo de la significancia puede escribirse como

$$Q_{KS}(x) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2jx^2}$$

Esta función es monótona con valores límites, $Q_{KS}(0) = 1$ y $Q_{KS}(\infty) = 0$. La probabilidad acumulada, como la conocemos, se escribe como

$$P(d_{max} \leq x) = 1 - Q_{KS}(x)$$

Al igual que en los métodos anteriores, la bondad del método KS se construye usando un valor crítico. Por lo tanto, la hipótesis nula es rechazada a nivel α si

$$d_{max}^{observ} > d_{max}^{\alpha}$$

donde d_{max}^α se encuentre a partir de

$$P(d_{max} \leq d_{max}^\alpha) = 1 - \alpha$$

Los valores de d_{max}^α se extraen a partir de tablas como la siguiente:

n	Nivel de significación α							
	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	0.90000	0.95000	0.97500	0.99000	0.99500	0.99750	0.99900	0.99950
2	0.68337	0.77639	0.84189	0.90000	0.92929	0.95000	0.96838	0.97764
3	0.56481	0.63604	0.70760	0.78456	0.82900	0.86428	0.90000	0.92065
4	0.49265	0.56522	0.62394	0.68887	0.73424	0.77639	0.82217	0.85047
5	0.44698	0.50945	0.56328	0.62718	0.66853	0.70543	0.75000	0.78137
6	0.41037	0.46799	0.51926	0.57741	0.61661	0.65287	0.69571	0.72479
7	0.38148	0.43607	0.48342	0.53844	0.57581	0.60975	0.65071	0.67930
8	0.35831	0.40962	0.45427	0.50654	0.54179	0.57429	0.61368	0.64098
9	0.33910	0.38746	0.43001	0.47960	0.51332	0.54443	0.58210	0.60846
10	0.32260	0.36866	0.40925	0.45562	0.48893	0.51872	0.55500	0.58042
$n > 50$	1.07	1.22	1.36	1.52	1.63	1.73	1.85	1.95
	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}

donde n representa los grados de libertad del problema. Otra manera común de entrar expresado el nivel de significación de un valor observado d_{max}^{observ} (para rechazar la hipótesis nula de que la distribuciones son iguales) viene dado aproximadamente por la siguiente fórmula

$$P(d_{max} > d_{max}^{observ}) = Q_{KS}([\sqrt{n_e} + 0,12 + 0,11/\sqrt{n_e}] d_{max}^{observ})$$

donde n_e es el número efectivo de datos. Para el caso de una distribución observada, $n_e = n$, mientras que cuando comparamos dos distribuciones observadas tendremos que $n_e = (n_1 n_2)/(n_1 + n_2)$. Por lo tanto, cuando este valor de P exceda el nivel de significancia α , la hipótesis nula será aceptada.

Ejemplo: Una investigación consiste en medir la altura de 100 niños de 5 años de edad. Se desea saber si las observaciones provienen de una población normal. El valor promedio de la muestra es 99.2 con desviación estándar 2.85.

Planteamiento de la hipótesis:

H_0 : No hay diferencias entre los valores observados y los teóricos de la distribución normal.

H_1 : Los valores observados de las frecuencias para cada clase son diferentes de las frecuencias teóricas de una distribución normal.

Nivel de significación: $\alpha = 0,05$

Zona de rechazo: Para todo valor de probabilidad mayor que 0.05, se acepta H_0 y se rechaza H_1 .

Aplicación de la prueba estadística: Primero se elaboran los cálculos de los valores teóricos esperados para la distribución normal. Inicialmente se determina el valor Z de

Serie de clases (talla en cm)	F	Fa
De 90 a 93	5	5
De 94 a 97	21	26
De 98 a 101	48	74
De 102 a 105	19	93
De 106 a 109	7	100
Total	100	

los límites de cada clase en la serie, por ejemplo: en la primera clase se determinan el límite inferior y el superior (90 y 93), y en las subsecuentes sólo los límites superiores (97, 101, 105 y 109). Para cada valor de Z , se localiza el área bajo la curva normal. Estos valores se utilizan para estimar, por medio de la diferencia del límite superior e inferior, el valor de la función teórica para ese bin. Estos resultados de diferencias se multiplican por el tamaño de la muestra (100 niños), luego se obtienen las frecuencias teóricas y después se arreglan en frecuencias acumuladas.

Límites de clases	Valor Z de los límites	Área bajo la curva tipificada	Diferencias entre clases	Diferencias $N(100) = F$	Fa
90	-3.23	-0.4994			
93	-2.18	-0.4854	0.014	1.4	1.4
97	-0.77	-0.2794	0.206	20.6	22.0
101	0.63	0.2357	0.5151	51.5	73.5
105	2.04	0.4793	0.2436	24.4	77.9
109	3.44	0.4997	0.0200	2.0	99.9
Total				99.9	

Las frecuencias acumuladas teóricas y las observadas se arreglan en los rangos correspondientes, como se muestra en la siguiente tabla, y posteriormente se aplica la fórmula de Kolmogorov-Smirnov.

Rangos	1	2	3	4	5
f_t	<u>1.4</u>	<u>22</u>	<u>73.5</u>	<u>97.9</u>	<u>99.9</u>
acumulada	100	100	100	100	100
f_{obs}	<u>5</u>	<u>26</u>	<u>74</u>	<u>93</u>	<u>100</u>
acumulada	100	100	100	100	100
$f_t - f_{obs}$	-0.036	-0.04	-0.005	0.049	-0.001

La diferencia máxima $d_{max} = 0,049$, valor que se compara con los valores críticos para el método Kolmogorov-Smirnov y se obtiene la probabilidad de la existencia de esa magnitud. El valor N es 100 por lo que, según la información que se puede extraer de las tablas (ver tabla de la página anterior), tenemos que

$$d_{max}^{\alpha} = d_{max}^{0,05} = \frac{1,36}{\sqrt{100}} = 0,136$$

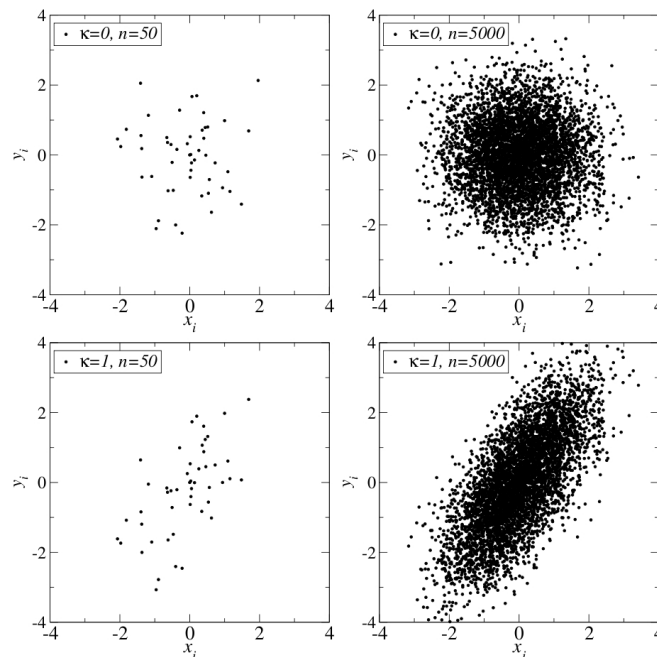
Decisión: En virtud de lo anterior, el estadístico de Kolmogorov-Smirnov obtenido es menor que el crítico y su probabilidad mayor que 0.05, por lo tanto, se acepta H_0 .

Conclusión: Las frecuencias observadas y las teóricas calculadas no difieren significativamente. Por lo tanto, las observaciones tienen una distribución normal.

8.3. Independencia estadística

Consideremos ahora muestras que consisten de pares de datos (x_i, y_i) con $i = 0, 1, \dots, n - 1$. La pregunta es, cuándo los valores y_i dependerán de los valores x_i (o viceversa). En el caso de que exista la dependencia, se dirá que estos valores están estadísticamente relacionados, y significa que conociendo uno de los valores podemos predecir el valor del otro con alta exactitud. Un ejemplo de dependencia estadística sucede en las simulaciones del clima. La cantidad de nieve caída está estadísticamente relacionada con la temperatura: si está muy caluroso o muy frío, no nevará. Esto también muestra que la dependencia de dos variables no necesariamente es monótona. Cuando uno está interesado en una dependencia monótona o lineal, usualmente se dice que las variables están *correlacionadas*.

Es importante darse cuenta que debemos distinguir entre la *significancia* estadística de una dependencia estadística y la *potencia* de la dependencia. Decir que una prueba nos dice que los valores x están estadísticamente relacionados con alta probabilidad, significa, usualmente, que tenemos una muestra grande. Por otro lado, la potencia de la dependencia estadística puede ser pequeña. Por ejemplo, que un dado valor de x tenga influencia en la distribución de probabilidades de y sólo levemente. En contrapartida, si la potencia es grande, significa, por ejemplo, que conociendo x casi se puede determinar y . Ahora, si solo tenemos unos pocos puntos en la muestra, no podemos estar seguros si los datos de la muestra están relacionados o no. Sin embargo, existe una conexión: mientras más grande sea la potencia, más fácil será probar que la dependencia es significativa. Para ilustrar lo dicho, consideremos una muestra donde los valores x_i son generados a partir de una distribución gaussiana (con $\mu = 0$ y $\sigma^2 = 1$), mientras que cada valor y_i se obtiene a partir de una distribución gaussiana con valor de expectativa κx_i (y $\sigma^2 = 1$). Entonces, si $\kappa = 0$, los datos son independientes. En la figura de la derecha pueden verse distintas distribuciones de pares (x_i, y_i) creados de esa manera. Se han creado 4 posibilidades, $\kappa = 0/1$ combinado con $n = 50/5000$. A continuación analizaremos que



pueden decirnos sobre estas muestras los métodos que describiremos en esta sección.

Primero presentaremos una variante del método chi-cuadrado, la cuál nos permitirá probar cuando un conjunto de datos es independiente. Luego, se darán a conocer los que se denominan *coeficientes de correlación lineal*, los cuales establecen la potencia de una correlación lineal. Finalmente, discutiremos como se puede cuantificar la dependencia dentro de una muestra, por ejemplo entre puntos muestrales $x_i, x_i + r$.

8.3.1. El método χ^2 ... el regreso

Para probar la dependencia estadística de una muestra $\{(x_0, y_0), (x_1, y_1), \dots, (x_{n-1}, y_{n-1})\}$, se considera usualmente la hipótesis H_0 : la muestra de puntos x y la muestra de puntos y son independientes. Para probar H_0 se ponen los pares de puntos muestrales en histogramas bidimensionales $\{h_{kl}\}$. Los recuentos $\{h_{kl}\}$ aumentan en una unidad, si para el dato (x_i, y_i) tenemos $x_i \in B_k^{(x)}$ y $y_i \in B_l^{(y)}$, para bins apropiadamente determinados $\{B_k^{(x)}\}$ y $\{B_l^{(y)}\}$. Sean k_x y k_y el número de bins en la dirección x e y respectivamente. Luego, se pueden calcular los histogramas unidimensionales $\{\hat{h}_k^{(x)}\}$ y $\{\hat{h}_l^{(y)}\}$ definidos por

$$\hat{h}_k^{(x)} = \sum_l h_{kl} \quad ; \quad \hat{h}_l^{(y)} = \sum_k h_{kl}$$

Estos histogramas unidimensionales describen como se distribuyen los recuentos para una variable, sin tener en cuenta el valor de la otra variable.

Las frecuencias relativas, que son estimas de probabilidades, son obtenidas normalizando con n , es decir, $\hat{h}_k^{(x)}/n$ y $\hat{h}_l^{(y)}/n$. Si dos variables son independientes, entonces la frecuencia relativa para obtener un par de valores (x, y) , en los bins $\{B_k^{(x)}\}$ y $\{B_l^{(y)}\}$, debe ser el producto de las frecuencias relativas simples de cada variable. Consecuentemente, multiplicando por n , se obtiene el correspondiente número esperado de recuentos n_{kl} , bajo la suposición de que H_0 se mantiene:

$$n_{kl} = n \frac{\hat{h}_k^{(x)}}{n} \frac{\hat{h}_l^{(y)}}{n} = \frac{\hat{h}_k^{(x)} \hat{h}_l^{(y)}}{n}$$

Estos recuentos esperados se pueden comparar con los valores medidos para los recuentos en el histograma bidimensional $\{h_{kl}\}$ por medio del estadístico χ^2 de la siguiente manera

$$\chi^2 = \sum_{kl} \frac{(h_{kl} - n_{kl})^2}{n_{kl}}$$

La interpretación estadística de χ^2 , nuevamente, viene dada por la distribución chi-cuadrado. El número de grados de libertad esta determinado por el número de bins $(k_x k_y)$ en el histograma bidimensional menos el número de restricciones y estimaciones. A la restricción que establece que $\sum_{kl} h_{kl} = n$ se le debe sumar las estimaciones que vienen dadas por las cantidades $\hat{h}_k^{(x)}$ y $\hat{h}_l^{(y)}$, es decir, por cada estimación, $(k_x - 1)$ en total por las filas, ya que la k_x -ésima queda determinada por las primeras $(k_x - 1)$,

análogamente, por cada estimación, $(k_y - 1)$ en total por las columnas. Por lo tanto, se obtiene el número de grados de libertad del estadístico haciendo:

$$\nu = k_x k_y - 1 - (k_x - 1) - (k_y - 1) = (k_x - 1)(k_y - 1)$$

Entonces, bajo la suposición de que los puntos muestrales x e y son independientes, $p = 1 - Q(\chi^2, \nu)$ da la probabilidad de tener un estadístico χ^2 o mayor. Comparando el valor de p con el nivel de significación, si $p < \alpha$, la hipótesis nula será rechazada.

Volviendo a los ejemplos de mostrados en la última figura, los valores p obtenidos para cada muestra son:

$$p(\kappa = 0, n = 50) = 0,077$$

$$p(\kappa = 0, n = 5000) = 0,457$$

$$p(\kappa = 1, n = 50) = 0,140$$

$$p(\kappa = 1, n = 5000) < 10^{-100}$$

Por lo tanto, la hipótesis nula de independencia no sería rechazada (con $\alpha = 0,05$) par el caso $\kappa = 1$, $n = 50$, el cuál esta correlacionado. Por otro lado, si las muestras son lo suficientemente grandes, no hay ninguna duda.

Veamos un ejemplo que describa el procedimiento completo.

Ejemplo: Se clasificaron los defectos de los muebles producidos en una planta de fabricación, primero, de acuerdo al tipo de defecto y segundo, de acuerdo al turno de producción. Lo que deseamos investigar es una posible dependencia entre las dos clasificaciones. Varían las proporciones de los diversos tipos de defectos de un turno a otro?. Por ejemplo, se observa un total de $n = 309$ muebles con defectos y se clasifican en cuatro tipos de defectos : A , B , C , D . Al mismo tiempo, cada mueble se identifica de acuerdo al turno de producción en el que es fabricado.

Turnos	Defecto A	Defecto B	Defecto C	Defecto D	Total
1	15(22.51)	21(20.99)	45(38.94)	13(11.56)	94
2	26(22.99)	31(21.44)	34(39.77)	5(11.81)	96
3	33(28.50)	17(26.57)	49(49.29)	20(14.63)	119
Total	74	69	128	38	309

Denotamos por p_A la probabilidad de que el defecto sea del tipo A , análogamente para p_B , p_C , p_D ; estas probabilidades las llamaremos probabilidades de las columnas de la tabla y se satisface:

$$p_A + p_B + p_C + p_D = 1$$

Análogamente p_i , $i = 1, 2, 3$ es la probabilidad de que ocurra un defecto en el turno i (probabilidad de la fila i) donde:

$$p_1 + p_2 + p_3 = 1$$

. Si las clasificaciones son independientes, entonces la probabilidad correspondiente a una celda debe ser el producto de las probabilidades de la fila y de la columna correspondiente a dicha celda. Por ejemplo, la probabilidad de que un defecto particular ocurra en el turno 1 y sea del tipo A debe ser $p_1 p_A$. La hipótesis nula se refiere a la independencia de las dos clasificaciones. No se especifican los valores numéricos de las probabilidades de las celdas. Por lo tanto, debemos estimar las probabilidades de las filas y de las columnas para poder estimar las frecuencias de celdas esperadas. Los estimadores de las probabilidades correspondientes a las columnas, son:

$$p_A = \frac{74}{309}, \quad p_B = \frac{69}{309}, \quad p_C = \frac{128}{309}, \quad p_D = \frac{38}{309}$$

Similarmente, las probabilidades para las filas son:

$$p_1 = \frac{94}{309}, \quad p_2 = \frac{96}{309}, \quad p_3 = \frac{119}{309}$$

Aplicando el estimador para el recuento esperado en caso de independencia, $n_{kl} = np_k p_l = 309 p_k p_l$ y los recuentos observados que figuran en la tabla, h_{kl} , podemos calcular el estadístico χ^2 .

$$\chi^2 = \sum_{k=1}^3 \sum_{l=A}^D \frac{(h_{kl} - n_{kl})^2}{n_{kl}} = 19,18$$

Como el grado de libertad para nuestro problema es $\nu = 6$ y $\alpha = 0,05$ tenemos que $\chi_{0,05,6}^2 = 12,60$. En consecuencia nuestro estadístico cae más allá del valor crítico, por lo tanto se rechaza la hipótesis nula, es decir, se concluye que no hay independencia entre el turno y el tipo de defecto.

8.3.2. Coeficiente de correlación lineal de Pearson

Una vez que se ha demostrado que una muestra contiene datos dependientes, uno puede tratar de medir la potencia de esa dependencia. Una manera estándar es usar el *coeficiente de correlación lineal de Pearson*, dados por

$$r \equiv \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

Este coeficiente asume, como lo indica su nombre, que existe una correlación lineal entre los datos. Para nuestro ejemplo de la figura, los coeficientes de correlación obtenidos son:

$$r(\kappa = 0, n = 50) = 0,009$$

$$r(\kappa = 0, n = 5000) = 0,009$$

$$r(\kappa = 1, n = 50) = 0,653$$

$$r(\kappa = 1, n = 5000) = 0,701$$

Aquí, también en los dos casos donde la estadística es baja, el valor de r refleja cuando los datos están correlacionados o no. Sin embargo, esto se da así porque estamos comparando datos correlacionados fuertemente con datos que no están correlacionados. Es decir estamos comparando extremos. Si comparásemos datos correlacionados débilmente, todavía tendríamos valores pequeños para r . Por lo tanto, para probar significancia, es mejor usar la prueba de hipótesis basado en el método χ^2 .

8.3.3. Función de correlación

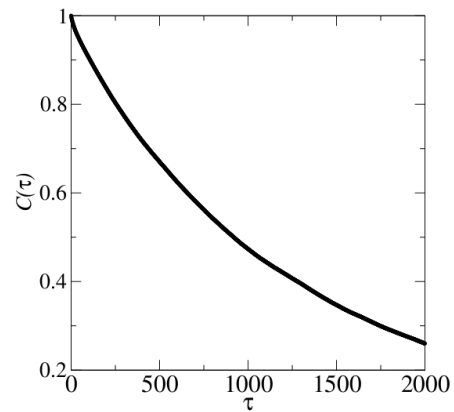
Finalmente, se puede notar que un tipo diferente de correlación puede surgir: hasta ahora hemos asumido siempre que los puntos muestrales x_i, x_j son estadísticamente independientes unos de otros. Sin embargo, podría ser el caso, por ejemplo, de que la muestra sea generada usando una simulación de una cadena de Markov de Monte Carlo, donde cada punto x_{i+1} es calculado usando un proceso aleatorio, pero también depende del valor del punto anterior x_i , entonces el índice i es un tipo de tiempo artificial muestral de la simulación. Esta dependencia disminuye cuando aumenta la distancia temporal entre puntos de la muestra. Una manera de ver que tan rápido esta dependencia disminuye es usar una variación del coeficiente de correlación, es decir, la función de correlación

$$\begin{aligned}\tilde{C}(\tau) &= \frac{1}{n-\tau} \sum_{i=0}^{n-1-\tau} x_i x_{i+\tau} \\ &\quad - \left(\frac{1}{n-\tau} \sum_{i=0}^{n-1-\tau} x_i \right) \times \left(\frac{1}{n-\tau} \sum_{i=0}^{n-1-\tau} x_{i+\tau} \right)\end{aligned}$$

El segundo término convergerá a \bar{x}^2 para $n \rightarrow \infty$ si se asume que la distribución de puntos muestrales es estacionaria, es decir, que no depende del tiempo muestral. Entonces, la función de correlación puede aproximarse por

$$\tilde{C}(\tau) = \frac{1}{n-\tau} \sum_{i=0}^{n-1-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x})$$

que resulta ser similar al numerador del coeficiente de Pearson visto anteriormente. Usualmente esta función se normaliza al valor que tiene en el origen ($C(\tau) = \tilde{C}(\tau)/\tilde{C}(0)$). Entonces la función $C(\tau)$ decrece a medida que aumenta la diferencia τ (ver figura de la derecha). Frecuentemente, la forma funcional es similar a una exponencial del tipo $\sim \exp(-\tau/\tau_c)$. En teoría, $C(\tau)$ converge a cero cuando $\tau \rightarrow \infty$, pero debido al tamaño finito de la muestra, aparecen fuertes fluctuaciones cuando τ se aproxima a n . Un tiempo típico τ_c que mide cuán rápido la dependencia de los puntos muestrales disminuye, viene dado por $C(\tau) = 1/e$. Al doble de la distancia, la correlación ya ha disminuido lo suficiente ($1/e^2$). Por lo tanto, si se quiere obtener barras de error para muestras obtenidas a partir de datos dependientes, se pueden incluir sólo puntos $x_0, x_{2\tau_c}, x_{4\tau_c}, x_{6\tau_c}, \dots$ en la muestra, o solo usar $n/(2\tau_c)$ en vez de n en cualquier cálculo de barras de error. Aunque estas barras de error son diferentes de las que se obtendrían a partir de una muestra realmente independiente, da una razonablemente buena impresión del error estadístico.



9. Estimadores Generales

Hasta aquí, se han presentado diferentes métodos para estimar parámetros los cuales pueden ser obtenidos directa y simplemente a partir de una dada muestra. En esta sección, se considera un método general que permite obtener estimadores para parámetros arbitrarios de las distribuciones de probabilidad. El método se basa en el principio de *máxima probabilidad* (maximum-likelihood). Este principio puede extenderse para modelar los datos muestrales donde usualmente se tienen tripletes del tipo $\{(x_0, y_0, \sigma_0), (x_1, y_1, \sigma_1), \dots, (x_{n-1}, y_{n-1}, \sigma_{n-1})\}$. En general, modelar los datos significa que se quiere determinar una relación del tipo $y = y(x)$. A este proceso se lo suele conocer como *ajuste de datos*.

9.1. Máxima Probabilidad

Consideremos la siguiente tarea: para una dada muestra $\{x_0, x_1, \dots, x_{n-1}\}$ y distribuciones de probabilidad representadas por $p_{\underline{\theta}}(x)$ y $f_{\underline{\theta}}(x)$, queremos determinar los parámetros $\underline{\theta} = (\theta_1, \dots, \theta_{n_p})$ tales que las distribuciones de probabilidad representen "mejor" los datos. Pero no hay una única manera de definir lo que significa "mejor", o algún procedimiento matemático para derivar un criterio apropiado. Ahora, si no se asume ningún conocimiento acerca de los parámetros, se puede usar el siguiente principio.:

El principio de máxima probabilidad establece que los parámetros $\underline{\theta}$ deben ser elegidos de manera que la probabilidad del conjunto de datos, especificados los parámetros, sea máxima

En el caso de una variable aleatoria discreta, si se asume que los diferentes datos puntuales son independientes, entonces la probabilidad de los datos viene dada por el producto de las probabilidades individuales de los datos puntuales. Esto define la función de máxima probabilidad como

$$L(\underline{\theta}) \equiv p_{\underline{\theta}}(x_1)p_{\underline{\theta}}(x_2)\dots p_{\underline{\theta}}(x_{n-1}) = \prod_{i=0}^{n-1} p_{\underline{\theta}}(x_i)$$

Para el caso continuo, la probabilidad de obtener, durante un experimento aleatorio, un dado valor es cero. Sin embargo, para un parámetro pequeño de incerteza ϵ , la probabilidad en el intervalo $[\tilde{x} - \epsilon, \tilde{x} + \epsilon]$ es $P(\tilde{x} - \epsilon \leq X \leq \tilde{x} + \epsilon) = \int_{\tilde{x}-\epsilon}^{\tilde{x}+\epsilon} f_{\underline{\theta}}(x)dx \approx f_{\underline{\theta}}(\tilde{x})2\epsilon$. Ya que 2ϵ introduce solo un factor, no es relevante para determinar el máximo. Por lo tanto, para el caso continuo, la función de máxima probabilidad es

$$L(\underline{\theta}) \equiv f_{\underline{\theta}}(x_1)f_{\underline{\theta}}(x_2)\dots f_{\underline{\theta}}(x_{n-1}) = \prod_{i=0}^{n-1} f_{\underline{\theta}}(x_i)$$

Para encontrar el máximo de la función de probabilidad $L(\theta)$ analíticamente, se deben calcular las derivadas primeras con respecto a todos los parámetros, respectivamente, e igualarlos a cero. Como la derivada de un producto involucra aplicar la regla del producto para la derivada, es más conveniente considerar el logaritmo de la función probabilidad

$$l(\theta) \equiv \log L(\theta)$$

Esto genera que la productoria se transforme en una sumatoria, para la cuál, las derivadas resultan mucho más simples de obtener. Además, como la función logaritmo es monótona, el máximo de la función probabilidad es igual al máximo del logaritmo de la función probabilidad. Por lo tanto, los parámetros "más apropiados" son determinados por el conjunto de ecuaciones

$$\frac{\partial l(\theta)}{\partial \theta_k} = 0 \quad (k = 1, \dots, n_p)$$

Observar que el hecho de que las derivadas primeras se hagan cero solo asegura que el punto obtenido es un extremo. Es más, estas ecuaciones generalmente tienen varias soluciones. Por lo tanto, se deben corroborar explícitamente cuáles soluciones son en verdad máximos, y quedarse con la mayor. Notar además, que los estimadores de máxima probabilidad, ya que son función de la muestra, también son variables aleatorias.

Como ejemplo simple, consideremos una distribución exponencial con parámetro μ . El logaritmo de la función probabilidad para una muestra $\{x_0, x_1, \dots, x_{n-1}\}$ es

$$l(\mu) = \log \prod_{i=0}^{n-1} f_{\mu}(x_i) = \sum_{i=0}^{n-1} \log \left[\frac{1}{\mu} \exp \left(-\frac{x_i}{\mu} \right) \right] = \sum_{i=0}^{n-1} \left[\log \left(\frac{1}{\mu} \right) - \frac{x_i}{\mu} \right] = n \log \left(\frac{1}{\mu} \right) - \frac{n\bar{x}}{\mu}$$

Si tomamos la derivada primera con respecto a μ se obtiene

$$0 = \frac{\partial l(\mu)}{\partial \mu} = \frac{-n\mu}{\mu^2} - \frac{-n\bar{x}}{\mu^2} = \frac{-n}{\mu^2}(\mu - \bar{x})$$

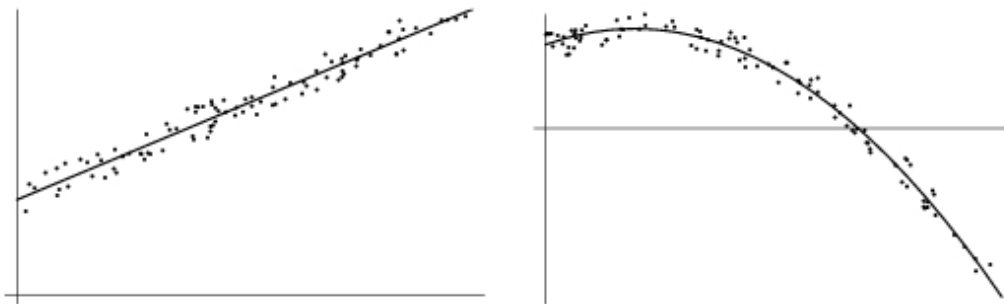
Lo cual implica que $\mu = \bar{x}$. Es fácil verificar que este valor corresponde a un máximo. Como el valor de expectación de una distribución exponencial es μ , esto es compatible con lo visto anteriormente donde se demostró que la media muestral es un estimador insesgado del valor de expectación.

Si se aplicase el principio de máxima probabilidad a una distribución gaussiana con parámetros μ y σ^2 , se obtiene como estimadores de máxima probabilidad la media muestral \bar{x} y la varianza muestral s^2 , respectivamente. Esto significa que el estimador de máxima probabilidad para σ^2 está sesgado, ya que, recordemos, $s^2 = (n-1/n)\sigma^2$. Afortunadamente, sabemos que el sesgo desaparece asintóticamente cuando $n \rightarrow \infty$. En general, puede demostrarse que, bajo ciertas condiciones de suavidad en las distribuciones subyacentes, todos los estimadores de máxima probabilidad para un parámetro θ_k son asintóticamente no sesgados.

En contraste con los casos de las funciones distribución exponencial y gaussiana, para muchas aplicaciones, los parámetros de máxima probabilidad no pueden ser relacionados con estimadores estándar de la muestra. Es más, usualmente no pueden ser determinados analíticamente. En ese caso, la solución es optimizar numéricamente el logaritmo de la función probabilidad para poder estimar sus correspondientes máximos.

9.2. Ajuste de datos

En la sección anterior los parámetros de la distribución de probabilidades son elegidos de manera que la distribución describa lo mejor posible los datos. Ahora consideremos un caso más general llamado modelado de los datos. Como ya mencionamos al principio, tenemos una muestra del tipo $\{(x_0, y_0, \sigma_0), (x_1, y_1, \sigma_1), \dots, (x_{n-1}, y_{n-1}, \sigma_{n-1})\}$. Tipicamente, los valores y_i son mediciones obtenidas a partir de una simulación con algún parámetro de control (por ej., la temperatura) fijado para diferentes valores de x_i ; σ_i es la correspondiente barra de error de y_i . Lo que se quiere es determinar los parámetros $\underline{\theta} = (\theta_1, \dots, \theta_{n_p})$ tal que la función parametrizada dada $y_{\underline{\theta}}(x)$ ajuste los datos lo "mejor" posible. Nuevamente, queda por definir que significa lo "mejor" posible.



9.2.1. Cuadrados mínimos como estimador de máxima probabilidad

Para lograr esto, recurriremos nuevamente al principio de máxima probabilidad y trataremos de estimar la correspondiente función de probabilidad. Supongamos que cada dato puntual y_i tiene una medición de error que es idenpedientemente aleatoria y distribuida de acuerdo a una distribución normal alrededor del valor del modelo $y_{\underline{\theta}}(x)$. Supongamos además, que la desviación estándar σ de estas distribuciones normales es la misma para todos los puntos. Entonces la función de probabilidad para el conjunto de datos será

$$L(\underline{\theta}) = \prod_{i=0}^{n-1} \exp \left[-\frac{1}{2} \left(\frac{y_i - y_{\underline{\theta}}(x_i)}{\sigma} \right)^2 \right] \Delta y$$

por lo que el logaritmo de esta función es

$$l(\underline{\theta}) = - \left[\sum_{i=0}^{n-1} \frac{(y_i - y_{\underline{\theta}}(x_i))^2}{2\sigma^2} \right] + (n-1) \log \Delta y$$

donde n , σ y Δy son todas constantes. Observar que, maximizar el logaritmo de la función de probabilidad es equivalente a minimizar el logaritmo negativo de dicha función, entonces lo que hay que minimizar son las diferencias cuadradas medias

$$\sum_{i=0}^{n-1} (y_i - y_{\underline{\theta}}(x_i))^2$$

Esto es lo que se denomina el método de ajuste por cuadrados mínimos. Lo que vimos entonces es que este método es una estimación de máxima probabilidad de los parámetros ajustados "si" los errores medidos son independientes y distribuidos normalmente con desviación estándar constante.

Por cientos de años, el hecho de que la distribución de probabilidades de una gran suma de pequeñas desviaciones aleatorias casi siempre converge a una distribución normal, ha fascinado a los estadistas. Sin embargo, esta característica tiende a hacer olvidar el hecho de que, para datos reales, la distribución normal es pobremente reproducida, o no reproducida en lo absoluto. En algunos casos, las desviaciones de la normalidad son fáciles de entender y cuantificar. Por ejemplo, si en dado un problema uno puede reconocer que los errores están distribuidos según Poisson, uno puede saber que si el número de recuentos es grande, la distribución Poisson converge hacia una gaussiana. Sin embargo, la convergencia no es uniforme y ocasiona predicciones equivocadas, provocando que el ajuste por cuadrados mínimos esté más distorsionado de lo que debería. Por otro lado, hay problemas en los que las desviaciones de los errores respecto de una distribución normal, no son fáciles de entender en detalle. Este es el caso de la existencia de valores atípicos (outliers), los cuales perjudican el ajuste de cuadrados mínimos. Para tratar estos problemas, es decir, tratar con casos en los que la distribución normal o gaussiana es una mala aproximación o en caso de que existan los valores de medición atípicos, existen las que se denominan estadísticas robustas. En el resto de esta sección, seguiremos asumiendo que los errores se encuentran distribuidos de acuerdo a una distribución normal, sin embargo es importante estar conciente de las limitaciones de estos modelos.

9.2.2. Ajuste por chi-cuadrado

Ahora generalizaremos un poco nuestro resultado de la sección anterior asumiendo que las desviaciones estándar σ_i de cada y_i son todas diferentes. Si ese es el caso, entonces en la ecuación del logaritmo de la función de probabilidad descrita en la sección anterior, las constantes solo son n y Δy . Por lo tanto, la ecuación a minimizar será

$$\chi^2 = \sum_{i=0}^{n-1} \left(\frac{y_i - y_{\underline{\theta}}(x_i)}{\sigma_i} \right)^2$$

Este método es conocido como ajuste por chi-cuadrado o ajuste por cuadrados mínimos pesados. Según la ecuación de χ^2 , los parámetros $\underline{\theta}$ son determinados tal que la función $y_{\underline{\theta}}(x)$ siga los datos puntuales $\{(x_0, y_0), \dots, (x_{n-1}, y_{n-1})\}$ lo más exactamente posible, donde las desviaciones son medidas en términos de las barras de error σ_i . Por lo tanto, los puntos con barras de error pequeñas entran con más peso. Una vez que se han ajustado los $\underline{\theta} = (\theta_1, \dots, \theta_{n_p})$ para minimizar el valor de χ^2 , los términos en la sumatoria no son todos independientes. Para modelos que son lineales en $\underline{\theta}'s$, sin embargo, la distribución de probabilidades para diferentes valores de χ^2 en su mínimo pueden ser derivados analíticamente, siendo una distribución chi-cuadrado con $n - n_p$ grados de libertad. En la sección 8 aprendimos cómo calcular la función de probabilidad acumulada $Q(\chi^2, \nu)$ en función de la función gamma incompleta, donde ν son los grados de libertad del problema, en este caso, $\nu = n - n_p$. Recordar que tanto Q como su complemento $P = 1 - Q$, se encuentran tabuladas. Es bastante común (y usualmente no está tan errado) asumir que la distribución chi-cuadrado es válida aún cuando los modelos no sean estrictamente lineales en los $\underline{\theta}'s$.

La probabilidad calculada da una medida cuantitativa de la bondad del ajuste del modelo. Si Q es un valor muy pequeño para algún conjunto particular de datos, entonces las aparentes discrepancias son poco probables que se deban a fluctuaciones aleatorias. Entonces, las posibilidades son: el modelo está mal, o la medición de los errores σ_i están mal, y en realidad son más grandes de lo establecido. Otra posibilidad es que la distribución de los errores no sea normal. Esto se debe a que en el cálculo de la probabilidad Q , se asume que los errores están distribuidos normalmente, con lo cual, si esto no pasa, la presencia de valores atípicos ocasiona valores bajos de Q . Esta última posibilidad es bastante común y también bastante benigna. Es por esta razón que hay algunos experimentos que son a menudo bastante tolerantes con las bajas probabilidades. No es raro que resulten aceptables, en términos de igualdad cualquier modelo con $Q > 0,001$. Los modelos verdaderamente malos serán rechazados cuando $Q \sim 10^{-18}$.

En el otro extremo, a veces pasa que la probabilidad Q es muy grande, cercana a 1, literalmente muy bueno para ser cierto. Los errores no normales no son causales para esto. Casi siempre, que el ajuste de chi-cuadrado sea tan bueno se debe a que el investigador, en un ataque de conservadurismo, haya sobreestimado sus mediciones de los errores. Muy raramente, un chi-cuadrado muy bueno es señal de fraude, es decir, que los datos fueron manipulados para que ajustasen el modelo.

Una regla a dedo, es asumir que un valor típico para χ^2 que refleje un ajuste moderado se obtiene cuando $\chi^2 \approx \nu$. Mucho más preciso es establecer que el estadístico χ^2 tenga

media ν y una desviación estándar $\sqrt{2\nu}$, y asintóticamente para grandes valores de ν , la distribución sea normal.

En algunos casos, las incertezas asociadas con el conjunto de mediciones no se conocen de antemano, y consideraciones relacionadas con el ajuste χ^2 son usadas para derivar un valor de σ . Si se asume que todas las mediciones tiene la misma desviación estándar y el modelo ajusta bien, entonces se puede proceder a asignar un valor arbitrario constante para σ a todos los puntos, luego se ajustan los parámetros del modelo minimizando χ^2 , y finalmente se recalcula

$$\sigma^2 = \sum_{i=0}^{n-1} \frac{(y_i - y_{\theta}(x_i))^2}{n - n_p}$$

Obviamente, esto nos impide tener una determinación independiente de la bondad del ajuste. Sin embargo, cuando los errores no se conocen, este procedimiento permite asignar algún tipo de barra de error a los puntos.

Por último, si derivamos la ecuación para χ^2 con respecto a los parámetros θ_k , obtenemos las siguientes ecuaciones que minimizan χ^2

$$0 = \sum_{i=0}^{n-1} \left(\frac{y_i - y_{\theta}(x_i)}{\sigma_i^2} \right) \left(\frac{\partial y_{\theta}(x_i)}{\partial \theta_k} \right) \quad k = 1, \dots, n_p$$

9.2.3. Ajustando datos con una recta usando chi-cuadrado

Para ejemplificar lo establecido en la sección anterior, veamos un ejemplo de su aplicación. Consideremos que se quiere ajustar a un conjunto de n puntos (x_i, y_i) , un recta o modelo lineal del tipo

$$y(x; a, b) = ax + b$$

A este problema se le suele llamar regresión lineal. Se asume que se conocen las incertezas σ_i asociadas con cada medición de y_i . Para medir que tan bien el modelo está de acuerdo con los datos, se usa el estadístico χ^2 , que en este caso es

$$\chi^2(a, b) = \sum_{i=0}^{n-1} \left(\frac{y_i - b - ax_i}{\sigma_i} \right)^2$$

Para minimizar esta expresión, se deberá resolver el siguiente sistema de ecuaciones

$$0 = \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=0}^{n-1} \frac{y_i - b - ax_i}{\sigma_i^2}, \quad 0 = \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=0}^{n-1} \frac{x_i(y_i - b - ax_i)}{\sigma_i^2}$$

Estas ecuaciones pueden reescribirse usando las siguientes definiciones

$$S \equiv \sum_{i=0}^{n-1} \frac{1}{\sigma_i^2}, \quad S_x \equiv \sum_{i=0}^{n-1} \frac{x_i}{\sigma_i^2}, \quad S_y \equiv \sum_{i=0}^{n-1} \frac{y_i}{\sigma_i^2}, \quad S_{xx} \equiv \sum_{i=0}^{n-1} \frac{x_i^2}{\sigma_i^2}, \quad S_{xy} \equiv \sum_{i=0}^{n-1} \frac{x_i y_i}{\sigma_i^2}$$

con lo cual, el sistema de ecuaciones se reduce a

$$bS + aS_x = S_y, \quad bS_x + aS_{xx} = S_{xy}$$

Por lo que, las soluciones de este sistema de ecuaciones son

$$b = \frac{S_{xx}S_y - S_xS_{xy}}{\Delta}, \quad a = \frac{SS_{xy} - S_xS_y}{\Delta}$$

donde $\Delta = SS_{xx} - (S_x)^2$. Pero aún no hemos terminado. Debemos estimar las incertezas en las estimas de los parámetros a y b , ya que los errores de medición en los datos deben introducir alguna incerteza en la determinación de dichos parámetros. Si los datos son independientes, entonces cada uno contribuye con su propia incerteza a la incerteza de los parámetros. Si consideramos la fórmula de propagación de errores para una función f cualquiera tenemos que

$$\sigma_f^2 = \sum_{i=0}^{n-1} \sigma_i^2 \left(\frac{\partial f}{\partial y_i} \right)^2$$

Para el caso de una recta, usando las soluciones obtenidas podemos escribir las derivadas parciales que van en la fórmula de propagación como

$$\frac{\partial b}{\partial y_i} = \frac{S_{xx} - S_x x_i}{\sigma_i^2 \Delta}, \quad \frac{\partial a}{\partial y_i} = \frac{S x_i - S_x}{\sigma_i^2 \Delta}$$

Por lo que, realizando la suma en la fórmula de propagación, se obtienen las incertezas para los parámetros

$$\sigma_b^2 = \frac{S_{xx}}{\Delta}, \quad \sigma_a^2 = \frac{S}{\Delta}$$

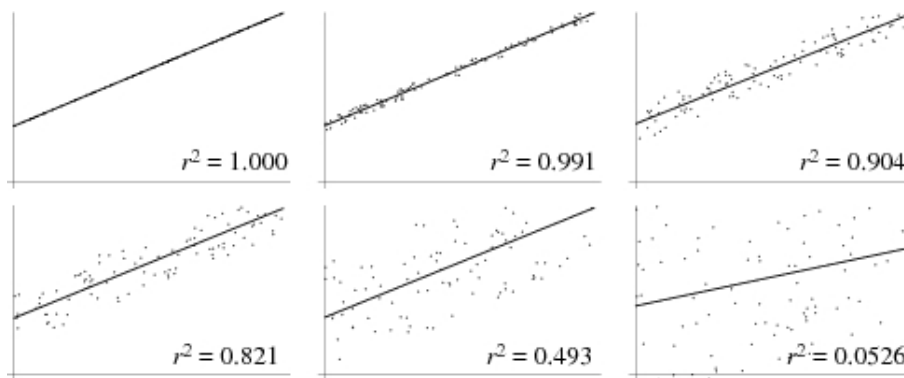
Puede verse que hace falta calcular un número adicional que caracterize apropiadamente la probable incerteza de la estimación de los parámetros. Ese número se denomina la *covarianza* de a y b y viene dada por

$$\text{Cov}(a, b) = \frac{-S_x}{\Delta}$$

El coeficiente de correlación entre la incerteza en a y la incerteza en b , el cuál es un número entre -1 y 1, proviene de la ecuación anterior, y es

$$r_{ab} = \frac{-S_x}{\sqrt{SS_{xx}}}$$

Un valor positivo de r_{ab} indica que los errores de a y b es probable que tengan el mismo signo, mientras que un valor negativo indica que los errores están anticorrelacionados, es decir, es probable que tengan distinto signo.



Pero todavía no hemos terminado. Debemos estimar la bondad del ajuste del modelo a los datos. Si no hacemos esto, no tenemos ninguna indicación de que los parámetros a y b obtenidos tengan algún significado. La probabilidad Q de que un valor de chi-cuadrado tan pobre como χ^2 , ocurra por azar es

$$Q = \frac{\gamma\left(\frac{n-2}{2}, \frac{\chi^2}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}$$

Si Q es más grande que 0.1, entonces la bondad del ajuste es creíble. Si es más grande que 0.001, el ajuste puede ser aceptable si los errores no son normales o han sido moderadamente subestimados. Si Q es menor a 0.001, entonces el modelo y/o el procedimiento de estimación son puestos en duda.

Por último, es útil notar que, computacionalmente hablando, las fórmulas anteriormente dadas son susceptibles de errores de redondeo. Para salvar este problema, usualmente se reescriben dichas ecuaciones usando que

$$t_i = \frac{1}{\sigma_i} \left(x_i - \frac{S_x}{S} \right) \quad i = 0, \dots, n-1$$

y

$$S_{tt} = \sum_{i=0}^{n-1} t_i^2$$

Sustituyendo, puede verse que

$$a = \frac{1}{S_{tt}} \sum_{i=0}^{n-1} \frac{t_i y_i}{\sigma_i}, \quad b = \frac{S_y - S_x a}{S}$$

$$\sigma_a^2 = \frac{1}{S_{tt}}, \quad \sigma_b^2 = \frac{1}{S} \left(1 + \frac{S_x^2}{S S_{tt}} \right), \quad Cov(a, b) = \frac{-S_x}{S S_{tt}}, \quad r_{ab} = \frac{Cov(a, b)}{\sigma_a \sigma_b}$$