Inferencia Estadística

Dr. Edgar Jimenez
Dr. Ulises Márquez

Maestría en Cómputo Estadístico

CIMAT Monterrey

Agradecimientos

La mayor parte de estas notas fueron preparadas por la Dra. Graciela González Farías.

En forma de agradecimiento, se enlistan personas que han contribuido de una u otra forma en la construcción de estas notas a través de los años:

- Víctor Muñiz
- Rogelio Ramos Quiroga
- Juan Antonio López
- Sigfrido Iglesias González
- Rodrigo Macías Paéz
- Benito Hernández Chaudary
- Todos los estudiantes que han colaborado con sugerencias y comentarios sobre estas notas desde 1998.

Estas notas son de uso exclusivo para enseñanza y no pretende la sustitución de los textos y artículos involucrados.

Temario

Objetivos del curso: Proporcionar las bases de la estadística inferencial, orientadas al manejo y análisis de datos.

1. Variables aleatorias y distribuciones de probabilidad.
    1. Distribuciones de probabilidad de variables aleatorias discretas.
    2. Procesos de Poisson.
    3. Distribuciones de probabilidad de variables aleatorias continuas.
    4. Métodos gráficos para la identificación de distribuciones.
    5. Estimación de densidades vía kernels.
    6. Desigualdades para variables aleatorias.
    7. Distribuciones de probabilidad de vectores aleatorios.
    8. Esperanzas condicionales y regresión.
    9. Modelos jerárquicos, compuestos y mezclas de variables aleatorias.
    10. Transformaciones de variables aleatorias.
    11. Simulación de variables aleatorias.
    12. Convergencia de variables aleatorias y el Teorema del Límite Central.

Temario

2. Distribuciones muestrales y métodos de estimación.
   1. Estimación puntual.
   2. Distribuciones muestrales.
   3. Propiedades de los estimadores.
   4. Estimadores de momentos.
   5. Estimadores de máxima verosimilitud y sus propiedades.
   6. Estimación por intervalos.
   7. Método Bootstrap.
3. Pruebas de Hipótesis e intervalos de confianza.
   1. Definición de conceptos.
   2. Potencia de la prueba.
   3. Pruebas para dos poblaciones normales independientes.
   4. Pruebas para medias en muestras pareadas.
   5. Pruebas básicas de varianzas.
   6. Pruebas para proporciones.

Temario

4. Lecturas complementarias:
   1. Cambio de variables para transformaciones de más de una variable
   2. Maximización de la función de verosimilitud.
   3. Otras que los profesores del curso consideren apropiadas y que se le proporcionarán a lo largo del curso.
5. Temas optativos de modelos para presentaciones finales, por ejemplo:
   1. Pruebas no-paramétricas clásicas.
   2. Pruebas de permutaciones.
   3. Estimación no paramétrica (suavizadores y splines).
   4. Pruebas de bondad de ajuste.

Entre muchas otras opciones que se discutirán con el grupo un mes antes de la fechas de exámenes finales.

Evaluación y acreditación

- Un examen parcial de medio término, 29 de septiembre: 30 %.
- Evaluación de las tareas (de 2 tipos) y actividades en clase y asistencia: 40 %.
- Un examen final, consistente en una exposición donde se entrega un reporte y se hace una presentación de 1/2 hora. La presentación debe incluir antecedentes, metodología, un ejemplo práctico y compartir el código. Deberán entregar a los instructores y a sus compañeros el resumen. Adicionalmente, deberán dejar un ejercicio sobre el tema a sus compañeros que calificarán en forma honesta. La fecha del examen final es 4, 5 y 6 de diciembre: 30 %.

Las tareas tienen una frecuencia quincenal e incluyen TODOS los ejercicios dejados en las notas y requerirán en general el uso de recursos computacionales.

Textos

- **Larry Wasserman (2004) . All of Statistics, A concise course in Statistical Inference. Springer.**
- F.M. Dekking, C. Kraaikamp, H.P. Lopuhaa L.E. Meester (2005). A Modern Introduction to Probability and Statistics, Understanding Why and How. Springer text in Statistics.
- John A. Rice (1995). Mathematical Statistics and Data Analysis, Second Edition. Duxbury Press.
- Casella & Berger. (2002). Statistical Inference, Second Edition . Duxbury Press.
- Richard J. Larsen and Morris L. Marx (2011). An Introduction to Mathematical Statistics and its Applications. Fifth Edition. Prentice Hall.

Modelos probabilísticos

Modelos probabilísticos

Recordemos que una variable aleatoria discreta es aquella que sólo toma un número contable de valores (finito o infinito). Por ejemplo:

- El número de plantas con daños visibles producidos por una plaga.
- El número de individuos a favor de un partido político.
- El número de televisores con defectos en su selector de canales de un lote de 100 televisores.
- El número de personas en la fila en un centro de servicio al público, entre las 9 y 10 de la mañana.

Notamos que en cada una de esas situaciones uno lleva a cabo algún tipo de **conteo**.

Modelos probabilísticos

En un principio uno debería:

- Examinar cada caso;
- Ver cuáles son las condiciones específicas en que se realiza el muestreo;
- Establecer los supuestos de simplicidad que sean factibles; y,
- Determinar el modelo probabilístico que mejor describa el comportamiento de la característica bajo estudio (verificando su validez).

Modelos probabilísticos

Este procedimiento general ha dado lugar a un cierto número de modelos que aparecen frecuentemente en las aplicaciones. Así, lo que haremos aquí, es construir estos modelos particulares formando un catálogo básico que nos permita referenciar nuestras situaciones particulares a alguno de estos. En la construcción del catálogo, contemplamos varios puntos:

1. Supuestos necesarios para identificar el uso del modelo.
2. Construcción del modelo: función de probabilidad y de probabilidad acumulada.
3. Momentos: Media, Varianza, Función generatriz de momentos (cuando aplique).

Modelos de variables discretas

Distribución uniforme discreta

Cuando se tiene un número **finito** de resultados de un experimento, $x_1, x_2, \ldots, x_n$, cada uno de ellos **igualmente probable**, se dice que se tiene una variable aleatoria que puede ser modelada por una distribución uniforme discreta. Notación:

$$X \sim U(\, x_1, x_2, \ldots, x_n \,).$$

Se lee: X se distribuye como una variable aleatoria uniforme.

Este tipo de modelo aparece típicamente en la selección de **muestras al azar**.

El único parámetro de la distribución es n, el número posible de resultados.

Distribución uniforme discreta

Como la suma de todas las probabilidades debe ser uno y todos los valores son equiprobables, entonces la función de masa está dada por

$$f(x) = \begin{cases} 1/n & \text{para } x = x_1, x_2, \ldots, x_n \\ 0 & \text{en otro caso} \end{cases}$$

Bajo esta definición, claramente $f(x) \geq 0$ para toda x, y

$$\sum_{i=1}^{n} f(x_i) = n\left(\frac{1}{n}\right) = 1.$$

El caso más común de esta distribución es cuando la variable toma los valores enteros

$$\{1, 2, \ldots, n\}.$$

El resto de los resultados los estableceremos para cuando la variable aleatoria uniforme toma estos valores.

Distribución uniforme discreta

El parámetro (valor que identifica unívocamente al modelo) de la distribución es n, el número total de objetos. Se dice en este caso que se trata de un espacio de probabilidad equiprobable.

Figura: Distribución uniforme para $n = 5$.

Distribución uniforme discreta

Figura: Distribución uniforme $n = 10$.

Distribución uniforme discreta

Figura: Distribución uniforme $n = 50$.

Distribución uniforme discreta

Media y Varianza:

$$\mathrm{E}(X) = \sum_{x=1}^{n} x \frac{1}{n} = \frac{1+2+\cdots+n}{n} = \frac{n+1}{2},$$

y

$$\begin{aligned}
\mathrm{V}(X) = & \ \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \sum_{x=1}^{n} x^2 \frac{1}{n} - \left(\frac{n+1}{2}\right)^2 \\
= & \ \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{2n^2+3n+1}{6} - \frac{n^2+2n+1}{4} \\
= & \ \frac{2(2n^2+3n+1) - 3(n^2+2n+1)}{12} = \frac{4n^2+6n+2-3n^2-6n-3}{12} \\
= & \ \frac{n^2-1}{12}.
\end{aligned}$$

Distribución uniforme discreta

Función Generatriz de Momentos:

$$
\begin{aligned}
M_X(t) = & \ \mathrm{E}\left(e^{tX}\right) = \sum_{x=1}^{n} e^{tx} \cdot f(x) = \sum_{x=1}^{n} e^{tx} \cdot \frac{1}{n} \\
= & \ \frac{1}{n} \sum_{x=1}^{n} \left(e^{t}\right)^{x} = \frac{1}{n} \left[\frac{e^{t} - (e^{t})^{n+1}}{1 - e^{t}}\right] = \frac{1}{n} \cdot \frac{e^{t}(1 - e^{nt})}{1 - e^{t}}, \quad \forall t.
\end{aligned}
$$

Nota: $M_X(0) = 0$, aplicando la regla de L'Hopital.

Se puede consultar más información sobre la función característica en:
`en.wikipedia.org/wiki/Characteristic_function_(probability_theory)`

Distribución Bernoulli

Definamos un experimento en el cual hay únicamente dos posibles resultados: a uno de ellos le llamamos "éxito", al otro "fracaso". Este tipo de variables aparece frecuentemente en nuestros conteos, por ejemplo, si pensamos en clasificar nuestros productos como: defectuoso, no defectuoso; grande o pequeño; azul o blanco; sí o no, etc.

Generalmente le asignamos un valor de 1 al "éxito" y un valor de 0 al "fracaso". Notemos que la asignación de los valores numéricos es arbitraria. Este procedimiento sirve de base para la construcción de otras distribuciones de gran utilidad.

Distribución Bernoulli

Definamos $X = \#$ de "éxitos", entonces

$$x = \begin{cases} 1 & \text{si } X = \text{"éxito"}, \text{ esto con probabilidad } p \\ 0 & \text{si } X = \text{"fracaso"}, \text{ esto con probabilidad } (1-p) = q \end{cases}$$

o bien,

$$f(x) = p^x(1-p)^{1-x}, \quad \text{para} \quad x = 0, 1; \qquad (\text{Distribución Bernoulli}).$$

Notación: $X \sim Bernoulli(p)$. El único parámetro de la distribución de ésta variable aleatoria es p, la probabilidad de éxito.

Distribución Bernoulli

Por ejemplo:

Figura: Función de densidad para $X \sim Bernoulli(p)$.

El experimento que guió al modelo, dos posibles resultados con probabilidades p y q $(p + q = 1)$, se denomina experimento Bernoulli.

Distribución Bernoulli

Media y Varianza:

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{1} x \cdot f(x) = 0 \cdot (1-p) + 1 \cdot p = p \\
\mathrm{E}(X^2) &= \sum_{x=0}^{1} x^2 \cdot f(x) = 0^2 \cdot (1-p) + 1^2 \cdot p = p \\
\mathrm{V}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = p - p^2 = p(1-p) = pq
\end{aligned}$$

Función Generatriz de Momentos:

$$\begin{aligned}
M_X(t) &= \mathrm{E}\left(e^{tX}\right) = \sum_{x=0}^{1} e^{tx} \cdot f(x) = e^{t \cdot 0}(1-p) + e^{t \cdot 1} p \\
&= (1-p) + pe^t = q + pe^t, \quad \forall t.
\end{aligned}$$

Distribución Binomial e Hipergeométrica

Para las siguentes distribuciones definiremos nuestra variable aleatoria de la siguiente manera:

$$X = \# \text{ de } \text{“éxitos”} \text{ en } n \text{ observaciones.}$$

X es el número de unos o “éxitos” que se presenten en la muestra de tamaño n. Lo que asigna un patrón distinto de comportamiento, es la **forma como se realizan (condiciones) las n observaciones**.

Distribución Binomial

Supongamos que podemos hacer:

1. n observaciones independientes.
2. la probabilidad de éxito en cada observación permanece constante, esto es, siempre es p.

En otras palabras, estamos asumiendo que nuestra población es suficientemente grande como para tomarla como "infinita" y que podemos garantizar la independencia entre observaciones, esto es, no obtenemos información adicional para predecir el siguiente resultado sólo porque ya observamos al (los) anterior (es). Esto refleja un comportamiento de procesos que lo podríamos denominar estable. Estas condiciones deben estar presentes al menos durante el período en que se realiza el estudio.

Distribución Binomial

Por ejemplo, pensemos en un proceso industrial, producción de mangueras para gas. Las clasificaremos como defectuosas (éxito) o no defectuosas, de acuerdo a si cumplen o no con el tamaño requerido. Se toma una muestra de tamaño tres y se hacen las mediciones de cada una de las mangueras.

- Asumimos que nuestra población son todas las mangueras que pasan por ese proceso (# muy grande),
- Se asume que no hay ninguna causa que motive desperfectos sistemáticos,
- También notemos que, bajo estas condiciones, la probabilidad de que una manguera no sea de las medidas requeridas deberá permanecer constante para las 3 observaciones y, más aún, esta probabilidad está dada por otro mecanismo independiente al conteo que nos atañe en este momento.

Distribución Binomial

Por ejemplo, p se podría determinar como una proporción observada a través del tiempo, o bien porque conozcamos la ley de fallas en cortes, de la maquinaria que se emplea. Para nuestros propósitos, p es un valor dado.

Notemos primero que el número posible de valores que puede tomar X es $\{1, 2, \ldots, n\}$. Ahora obtengamos la ley de comportamiento asociada con esta variable:

$$\begin{aligned} \mathrm{P}(X = 0) &= \mathrm{P}(\text{ninguna manguera defectuosa en la muestra de } n = 3) \\ &= \mathrm{P}(X_1 = 0, X_2 = 0, X_3 = 0) \end{aligned}$$

donde X_i denota la i-ésima observación. Cabe notar que cada X_i toma el valor de 0 ó 1 con probabilidades constantes q y p, respectivamente. Esto es, cada X_i **representa una variable Bernoulli**(p) **y son independientes entre sí**.

Distribución Binomial

Entonces:

$$\begin{aligned} \mathrm{P}(X_1 = 0, X_2 = 0, X_3 = 0) &= \mathrm{P}(X_1 = 0)\mathrm{P}(X_2 = 0)\mathrm{P}(X_3 = 0) \\ &= qqq = q^3 = \binom{3}{0} q^3 \end{aligned}$$

A manera de ejercicio, puedes verificar que:

$$\begin{aligned} \mathrm{P}(X = 1) &= \mathrm{P}(\{X_1 = 1, X_2 = 0, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 0, X_2 = 0, X_3 = 1\}) \\ &= 3pqq = 3pq^2 = \binom{3}{1} pq^2, \qquad \mathbf{1} \ \textbf{“éxito”}; \\ \mathrm{P}(X = 2) &= \mathrm{P}(\{X_1 = 1, X_2 = 1, X_3 = 0\} \text{ ó } \{X_1 = 1, X_2 = 0, X_3 = 1\} \text{ ó } \{X_1 = 0, X_2 = 1, X_3 = 1\}) \\ &= 3ppq = 3p^2q = \binom{3}{2} p^2q, \qquad \mathbf{2} \ \textbf{“éxitos”}; \\ \mathrm{P}(X = 3) &= \mathrm{P}(\{X_1 = 1, X_2 = 1, X_3 = 1\}) \\ &= ppp = 3p^3 = \binom{3}{3} p^3, \qquad \mathbf{3} \ \textbf{“éxitos”}. \end{aligned}$$

Distribución Binomial

Esto es, tenemos la Distribución Binomial

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, 1, \ldots, n \\ \\ 0 & \text{en otro caso} \end{cases}$$

En este caso se dice que X sigue una distribución binomial con parámetros (n, p). Notación: $X \sim B(n, p)$.

Distribución Binomial

Distribución Binomial

Se puede verificar que la suma, sobre todos los valores de X, de $f(x)$ es uno haciendo uso del **Teorema del Binomio de Newton**:

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x,$$

en dónde si se reemplaza a por q y b por p se obtiene

$$(q+p)^n = \sum_{x=0}^{n} \binom{n}{x} q^{n-x} p^x = \sum_{x=0}^{n} f(x),$$

y como $q + p = (1-p) + p = 1$, se concluye que

$$\sum_{x=0}^{n} f(x) = (1)^n = 1.$$

Distribución Binomial

Su **Función Generatriz de Momentos** es:

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \sum_{x=0}^{n} e^{tx} \cdot f(x) = \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (e^t)^x p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x}.$$

De acuerdo al binomio de Newton con $a = q,\ b = e^t p$:

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = (q + pe^t)^n.$$

Distribución Binomial

Media y Varianza

Utilizaremos la generatriz de momentos. Sabemos que $\mathrm{E}(X) = M'_X(0)$ y $\mathrm{E}(X^2) = M''_X(0)$, entonces

$$\begin{aligned} M'_X(t) = & \ n(q + pe^t)^{n-1}(pe^t), \\ M'_X(0) = & \ n(q + pe^0)^{n-1}(pe^0). \end{aligned}$$

Puesto que $e^0 = 1$ y $p + q = 1$, tenemos:

$$M'_X(0) = \ n(1)^{n-1}p = np,$$

y, por lo tanto,

$$\mathrm{E}(X) = np.$$

Distribución Binomial

Ahora,

$$\begin{aligned} M_X''(t) &= n(q+pe^t)^{n-1}pe^t + pe^t\left[n(n-1)(q+pe^t)^{n-2}pe^t\right], \\ \mathrm{E}(X^2) &= M_X''(0) = np + n(n-1)p^2. \end{aligned}$$

entonces

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = np + n(n-1)p^2 - (np)^2 \\ &= np + n^2p^2 - np^2 - n^2p^2 = np - np^2 = np(1-p) \\ &= npq. \end{aligned}$$

Distribución Binomial

La variable aleatoria Binomial como una suma de variables aleatorias Bernoulli

Hemos mencionado que cada X_i es una v.a. Bernoulli y que estas son independientes. Considera la variable aleatoria Y definida como la suma de n variables aleatorias Bernoulli(p), esto es, como la suma de ceros y unos producidos por los posibles resultados de las variables Bernoulli

$$Y = \sum_{i=1}^{n} X_i, \qquad \text{con } X_i \sim \text{Bernoulli}(p) \text{ e independientes entre sí.}$$

Podemos imaginarnos lo anterior como realizando n repeticiones de un experimento Bernoulli en el que cada resultado será independiente de los otros, un **muestreo aleatorio**.

Distribución Binomial

Recordando las propiedades del valor esperado y el valor que toma la esperanza de una v.a. *Bernoulli*(p) , tenemos que el valor esperado de Y es

$$\mathrm{E}(Y) = \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np$$

y su varianza

$$\mathrm{V}(Y) = \mathrm{V}\left(\sum_{i=1}^{n} X_i\right).$$

Veremos posteriormente que la varianza de una suma de variables aleatorias **independientes** es la suma de las varianzas de cada v.a. individual (multiplicada por el cuadrado de su coeficiente respectivo), entonces

$$\mathrm{V}(Y) = \sum_{i=1}^{n} \mathrm{V}(X_i) = \sum_{i=1}^{n} pq = npq.$$

Distribución Binomial

Vemos que estos resultados coinciden con los de la v.a. Binomial(n, p) pero **no** es suficiente para afirmar que $Y \sim$ Binomial(n, p). Si su generatriz es la misma que la de una v.a. binomial podremos afirmar que Y es binomial:

$$\begin{aligned} M_Y(t) &= \mathrm{E}(e^{tY}) = \mathrm{E}\left(e^{t\sum_{i=1}^{n} X_i}\right) = \mathrm{E}\left(e^{t(X_1+X_2+\cdots+X_n)}\right) \\ &= \mathrm{E}\left(e^{tX_1}e^{tX_2}\ldots e^{tX_n}\right), \end{aligned}$$

de nuevo, como las variables aleatorias X_i son independientes, el último término se puede escribir como

$$\begin{aligned} M_Y(t) &= \mathrm{E}\left(e^{tX_1}\right)\mathrm{E}\left(e^{tX_2}\right)\cdots\mathrm{E}\left(e^{tX_n}\right) \\ &= (q+pe^t)(q+pe^t)\cdots(q+pe^t) = (q+pe^t)^n, \end{aligned}$$

que es idéntica a la generatriz de momentos de una binomial.

Distribución Binomial

Por lo tanto nuestra v.a. Y, producto de un muestreo determinado, es una v.a. Binomial(n, p).

¿Qué se obtiene si en la sumatoria que define Y usamos $n = 1$?

Es muy importante recalcar el hecho de que cada posible resultado del experimento es una v.a. con las mismas propiedades: misma distribución, mismo parámetro, y que el resultado de cada una de ellas será independiente del resultado de las otras.

Estas condiciones se resumen diciendo que se realizó un **Muestreo Aleatorio**.

Distribución Hipergeométrica

Si cambiamos la forma en que se realiza el muestreo y mantenemos como nuestra v.a. a

$$X = \# \text{ de } \text{“éxitos”} \text{ en } n \text{ observaciones},$$

nuestro modelo tiene que ser adaptado a las nuevas condiciones.

Supongamos que podemos hacer:

- n observaciones de un conjunto total de N posibles,
- la probabilidad de éxito en cada observación cambia “paso a paso”.

Ahora estamos asumiendo que nuestra población es finita (N) y no hay independencia entre observaciones, esto es, el resultado de cada observación será afectado por los resultados anteriores.

Distribución Hipergeométrica

Como **ejemplo**, considera un lote de 25 "botes" de leche comercial de 1 litro, cada uno susceptible de ser clasificado como en mal estado (éxito = agria) o en buen estado. Se toma una muestra de tamaño n, se considera la posibilidad de que se tengan k botes con leche agria y se hace la revisión de cada uno de los botes. Aquí $X = \#$ de botes con leche en mal estado.

- Nuestra población inicial son los 25 "botes",
- La probabilidad de que un bote seleccionado contenga leche en mal estado cambiará conforme vamos haciendo el muestreo debido a que en el lote habrá inicialmente una cierta proporción de botes en mal estado (botes de leche en mal estado / N total de botes en el lote) y al ir sacando los botes esta proporción se verá afectada por los resultados obtenidos con anterioridad. Esto es, se realiza un muestreo sin reemplazo.

Distribución Hipergeométrica

Este tipo de experimento es muy usado cuando en la revisión se tiene que destruir o alterar el objeto, en nuestro caso hay que abrir el bote de leche para revisarlo.

Considera lo siguiente: si $k = 4$, la proporción inicial es de $\frac{4}{25}$, pero si sacamos un bote la proporción cambia a $\frac{3}{24}$ ó $\frac{4}{24}$, dependiendo de si el que se sacó contenía leche en mal estado o no, respectivamente.

Distribución Hipergeométrica

Para poder establecer el conjunto de posibles valores de X, necesitamos conocer, en principio, la proporción exacta de botes en mal estado, esto es, cuántos botes hay en mal estado en el lote de 25, y cuántas observaciones se hacen (n):

- Si hay $k = 3$ botes con leche agria y se toman $n = 5$ botes para revisarlos, X podrá tomar los valores $\{0, 1, 2, 3\}$, i.e. $0 \leq x \leq k$; pero si se revisan $n = 2$, X podrá tomar los valores $\{0, 1, 2\}$, i.e. $0 \leq x \leq n$.
- Algo similar ocurriría si el lote fuera de $N = 10$ botes, si hubiesen $k = 4$ botes con leche agria y si se revisaran $n = 7$: puesto que se están revisando 7 y sólo hay 6 en buen estado, X podría ser $\{1, 2, 3, 4\}$, i.e. $n - (N - k) \leq x \leq k$. ¿Qué valores toma X si $k = 5$ y $n = 7$?

Observa que x **llega hasta el menor de los números** n **y** k**, y que el menor valor de** x **es 0 ó** $n - (N - k)$. Estos casos muestran que los valores de la v.a. dependerán de los valores de k, n y N.

Distribución Hipergeométrica

Las condiciones de muestreo mencionadas suelen aplicarse en la práctica en problema de muestreo de **aceptación en control de calidad**, así como en la **estimación del tamaño de una población finita** N en **procesos de captura-recaptura**. (Ver [4] ejemplo 3.2.10 pág. 86, ejemplo 3.2.9 págs. 84-85. También te recomendamos (deberías leer) el capítulo 4, pág. 42 de [5], para muestreo de aceptación.)

No es de sorprender este tipo de aplicaciones ya que, como dijimos, al tomar n observaciones en las cuales habrá x "éxitos", conoceremos la proporción $\frac{x}{n}$, la cual nos dará información de la verdadera proporción $\frac{k}{n}$.

Distribución Hipergeométrica

Esquemáticamente podemos imaginarnos los parámetros de la distribución hipergeométrica como en la siguiente figura

Población con N objetos		Muestra de n objetos tomados al azar
k elementos con cierta característica de interés	$\leftarrow$ Inferencia	x elementos con la característica de interés
$N-k$ sin la característica	$\leftarrow$ Inferencia	$n-x$ sin la característica de interés

Distribución Hipergeométrica

La ley de comportamiento asociada con nuestra v.a. se obtiene utilizando el enfoque de frecuencia relativa en el que contamos el número de casos favorables y dividimos entre el número de casos totales:

- Nos interesan x "éxitos" de un total de k posibles, y esto lo podemos hacer de $\binom{k}{x}$ formas; pero por cada uno de estos resultados, como se seleccionaron n, obtendríamos $n-x$ "fracasos" de un total de $N-k$ "fracasos" posibles, y esto lo podemos hacer de $\binom{N-k}{n-x}$ formas. Entonces, usando la regla multiplicativa, el número total de **casos favorables** es $\binom{k}{x}\binom{N-k}{n-x}$.
- Al seleccionar n objetos de un total de N, podemos hacerlo de $\binom{N}{n}$ formas.

Distribución Hipergeométrica

Por lo tanto, la probabilidad de observar x "éxitos" en n pruebas está dada por la función de probabilidad (Distribución Hipergeométrica):

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \text{máx}\left\{0, n-(N-k)\right\} \leq x \leq \text{mín}\left\{n, k\right\}.$$

Sus parámetros son n, k y N. Notación $X \sim \text{Hiper}(N, k, n)$.

Distribución Hipergeométrica

En las siguientes gráficas se muestra la funciones de masa y de distribución acumulada de la distribución hipergeométrica para varios valores de los parámetros.

Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(5, 2, 2)$.

Distribución Hipergeométrica

Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(10, 5, 6)$.

Distribución Hipergeométrica

Figura: Funciones de masa y de distribución acumulada para $X \sim$Hiper$(10, 5, 4)$.

Distribución Hipergeométrica

Media y Varianza

La media y la varianza de una v.a. hipergeométrica están dados por:

$$\mathrm{E}(X) = \frac{nk}{N}, \qquad \mathrm{V}(X) = \frac{nk}{N}\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Debido a la especial complejidad y a su poco uso práctico, no se muestra la función generatriz de momentos de la hipergeométrica. (Ver referencias adicionales en Clase 4).

Distribución Hipergeométrica

Aproximación de la distribución binomial a la distribución hipergeométrica

Existe una relación entre la distribución binomial y la hipergeométrica: bajo ciertas condiciones, una puede aproximar los valores de probabilidad de la otra.

Observa que, a partir del enfoque de probabilidad como frecuencia relativa, la razón $\frac{k}{N}$ representa la proporción de éxitos p, mientras que $1 - \frac{k}{N}$ la proporción de fracasos q, de lo cual las anteriores fórmulas pueden ser escritas como

$$\mathrm{E}(X) = np, \qquad \mathrm{V}(X) = npq\left(\frac{N-n}{N-1}\right).$$

Distribución Hipergeométrica

Se puede apreciar una cierta similitud con la media y la varianza de la distribución binomial. Si N fuese muy grande, y el tamaño n de la muestra pequeño, tendríamos que

- el factor $\frac{N-n}{N-1}$ tendería a 1;
- p equivaldría a la probabilidad de "éxito" como en la distribución binomial, pudiéndose considerar como **"cuasi-fija"**, ya que no cambiaría mucho al ir realizando las observaciones;
- q equivaldría a la probabilidad de "fracaso", con las mismas consideraciones que para p; e,
- igualmente, podríamos considerar una **"cuasi-independencia"**, suponiendo que no tendríamos "rachas sistemáticas de resultados", esto es, el resultado de una observación no nos daría información del resultado de la siguiente observación.

Distribución Hipergeométrica

Con las condiciones anteriores nuestro modelo sería aproximadamente binomial, y podemos usar esta distribución como una aproximación al modelo hipergeométrico.

Es posible demostrar formalmente que si $p = k/N$ se mantiene constante entonces

$$\lim_{N\to\infty} \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}.$$

La pregunta es: ¿qué tan grande deberá ser N comparada con n? Se ha encontrado de la experiencia que cuando la proporción $\frac{n}{N}$ es del orden del 10 %, se tiene una buena aproximación, mejorando cuando $\frac{n}{N}$ disminuye.

Distribución Hipergeométrica

Por ejemplo, tenemos que la aproximación binomial para Hiper$(1000, 400, 50)$ es excelente.

Distribución Hipergeométrica

En este segundo ejemplo la relación entre N y n es del 50 %, por lo que no se espera una buena aproximación entre las probabilidades asociadas bajo cada distribución.

Distribución Hipergeométrica

El factor $\frac{N-n}{N-1}$, cantidad que nos recuerda la finitud de N, es llamado el **factor de corrección para población finita**. ¿Cuál es $\lim_{N\to\infty} \frac{N-n}{N-1}$, con n fija?

Nota: Entre los materiales del curso se encuentra un texto sobre métodos de captura y recaptura asociados a la distribución hipergeométrica y sobre como calcular los momentos de tal distribución. Es importante leerlos.

Distribución Geométrica y Binomial Negativa

En la mayoría de los procesos industriales existen procedimientos de "monitoreo" a través de gráficos de control. Para ello pueden emplearse diversos criterios, por ejemplo

- detener la producción (o ajustar el proceso) hasta el momento en que se detecta el primer artículo defectuoso (al obtener una "falsa alarma"),
- o bien, hasta que se detectan un cierto número r con $(r > 1)$, de artículos defectuosos.

Analizaremos el modelo correspondiente a un "éxito", y posteriormente lo generalizaremos a r "éxitos", para cualquier valor de r.

Distribución Geométrica

Aunque usamos el ejemplo de la línea de producción para introducir las v.a.'s de esta sección, también las definiremos en términos de la cantidad de experimentos Bernoulli realizados hasta obtener un determinado número de "éxitos".

Uno de los criterios arriba mencionados para mantener bajo control las líneas de producción es detener el proceso (o ajustarlo) hasta obtener un artículo defectuoso. Este criterio puede parecer un tanto estricto, pero si se sabe que el obtener un artículo defectuoso es un evento muy raro, esto es, hay una probabilidad pequeña de que ocurra; o bien, que la obtención de un artículo defectuoso es una situación en extremo crítica o costosa; es factible poner en marcha dicha política.

Distribución Geométrica

La variable aleatoria de interés en este caso es

$$X = \# \text{ de observaciones hasta obtener } \underline{\text{un}} \text{ éxito}.$$

Observa que en ningún momento se está fijando el número de observaciones, como en la binomial y la hipergeométrica, sino que más bien el número de observaciones es la variable aleatoria de interés y lo que es fijo en este caso es el número de éxitos.

La v.a. geométrica se define como el número de observaciones de variables aleatorias Bernoulli(p) independientes <u>hasta obtener un éxito</u>.

Distribución Geométrica

Nuestros supuestos son:

1. Cada vez que se realice una observación la probabilidad de que ocurra un "éxito" es fija e igual a p.
2. Las observaciones son independientes.
3. El número de observaciones puede continuar indefinidamente antes de encontrar el primer defectuoso.

Distribución Geométrica

El supuesto 1 equivale a decir que estamos pensando en hacer observaciones sobre el mismo proceso y, por lo tanto, todos sus elementos deben tener en principio las mismas características.

Por ejemplo, en la línea de producción, el proceso "no sabe" o "no se acuerda" de lo que ha pasado antes de la observación que se esté realizando, ya que los productos defectuosos ocurren por mero azar (una variación inherente e inevitable), no por alguna falla sistemática. Cuando uno detecta un artículo defectuoso, sabiendo que el modelo asigna una cierta probabilidad a ese evento, se toma una decisión respecto al proceso. Ver [15].

El supuesto 2 equivale a decir que el que una observación resulte o no en "éxito", **no** da información sobre el resultado de la siguiente observación.

Distribución Geométrica

Bajo las condiciones y supuestos anteriores podemos deducir inmediatamente el mecanismo para asignar probabilidades, donde llamamos p a la probabilidad de "éxito" y por ende $(1 - p = q)$:

$$f(x) = q^{x-1}p, \quad x = 1, 2, 3, \ldots. \qquad \text{(Distribución Geométrica)}$$

Notación: $X \sim \text{Geom}(p)$. El único parámetro es p, la probabilidad de éxito.

A continuación se presentan algunas gráficas de funciones de distribución acumulada y de densidad para $\text{Geom}(p)$ y distintos valores de p.

Distribución Geométrica

Figura: Funciones de masa para Geom(0.1) y Geom(0.8).

Distribución Geométrica

Figura: Funciones de masa y de distribución acumulada para Geom(0.5).

Distribución Geométrica

Esta distribución regularmente se utiliza en la modelación de:

1. Fenómenos con eventos raros, i.e. aquéllos en los que p es **pequeña**.
2. tiempos de espera discretos.

Ejemplo del uso 1: cuando se quiere detectar una enfermedad rara, en lugar de muestrear mediante una campaña masiva toda la población, se puede ir muestreando aleatoriamente de uno en uno, y si conforme se realiza el muestreo no se detecta a nadie enfermo, se puede pensar que no hay problema; pero si se encuentra a alguien con la enfermedad, es motivo para hacer un muestreo más exhaustivo. Algo equivalente se considera en los errores de *transmisión en comunicaciones*, o cualquier sistema considerado como muy *confiable*.

Distribución Geométrica

Ejemplo del uso 2: cuando se tiene que inspeccionar un sistema para ver si está en buenas condiciones para seguir operando, las inspecciones se realizan programadamente cada cierto intervalo de tiempo fijo, y aún cuando una falla haya ocurrido en el lapso entre inspecciones se considera, que esta ocurre al momento de realizar la inspección ("tiempo discreto"). Entonces, la v.a. sería *el número de intervalos de tiempo que se tuvo que esperar hasta encontrar un "éxito"*. Nuevamente se espera que el defecto ocurra por la aleatoriedad inherente y no por desgaste.

Distribución Geométrica

Se ha dicho que el proceso "no sabe" o "no se acuerda" de lo que ha pasado antes de la observación que se esté realizando. Cuando un proceso tiene esta característica, se dice que "NO TIENE MEMORIA". Lo mismo se aplica para la v.a. que describe al proceso.

Matemáticamente: si $X \sim \text{Geom}(p)$, calculemos la probabilidad de que un "éxito" ocurra para una X mayor que un cierto número de pruebas, denotado como la suma de dos números $a + b$, dado que se sabe que el éxito ocurre después de las a pruebas (a y b son enteros positivos). Si el resultado es independiente de que el éxito ocurra después de las primeras a pruebas significará que al proceso "no le importa" lo que haya ocurrido antes del valor que se esté observando.

Distribución Geométrica

En símbolos,

$$P(X > a+b | X > a) = \frac{P(X > a+b, X > a)}{P(X > a)} \quad \text{pues } P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P(X > a+b)}{P(X > a)} \quad \text{ya que } X > a+b \text{ es la intersección entre los dos intervalos del numerador}$$

$$= \frac{1 - \mathrm{P}(X \leq a+b)}{1 - \mathrm{P}(X \leq a)}$$

$$= \frac{1 - F(a+b)}{1 - F(a)}.$$

Por otro lado, la función de distribución de una v.a. Geom(p) es

$$F(x) = \sum_{n=1}^{x} q^{n-1} p = p \sum_{m=0}^{x-1} q^m = p \left(\frac{1 - q^x}{1 - q} \right) = 1 - q^x.$$

Distribución Geométrica

De lo anterior,

$$P(X > a+b|X > a) = \frac{1 - F(a+b)}{1 - F(a)} = \frac{q^{a+b}}{q^a} = q^b = 1-F(b) = P(X > b);$$

esto es, X "**no tiene memoria**".

Como ejemplos numéricos

$$\begin{aligned} P(X > 1+2|X > 1) &= P(X > 2), \\ P(X > 1+2|X > 2) &= P(X > 1), \\ P(X > 1+1|X > 1) &= P(X > 1). \end{aligned}$$

Distribución Geométrica

Bosquejo: Otra forma de visualizar esto es pensando que se "recorre el valor más bajo posible de X".

En el primer ejemplo numérico es como si desechará el 1 como origen y pusiera mi origen en 2 (que ahora sería 1). *Esto equivale a decir que si están haciendo inspecciones cada hora para verificar la vida útil de algún componente, el hecho de saber que no ha fallado en la observación 100, por ejemplo, (esto es, después de 100 horas de uso) no nos dice nada para responder a la pregunta: ¿cuál será la probabilidad de que falle en las siguientes 20 horas? La probabilidad de que falle es la misma que cuando el componente es nuevo y te haces la misma pregunta: ¿cuál será la probabilidad de que falle en las siguientes 20 horas?*

Esto sólo tiene sentido si se está trabajando con un proceso que se encuentra en estado de estabilidad.

Distribución Geométrica

Media y Varianza

Ya que

$$\sum_{n=1}^{+\infty} nz^{n-1} = \frac{1}{(1-z)^2} \qquad |z| < 1,$$

el valor esperado de $X \sim \text{Geom}(p)$ es

$$\begin{aligned} \text{E}(X) &= \sum_{x=1}^{+\infty} x \cdot f(x) = \sum_{x=1}^{+\infty} x(\underbrace{1-p}_{q})^{x-1} p = p \sum_{x=1}^{+\infty} x q^{x-1} \\ &= p\left(\frac{1}{(1-q)^2}\right) = p\left(\frac{1}{p^2}\right) = \frac{1}{p}. \end{aligned}$$

Distribución Geométrica

Para el cálculo de la varianza, nos queda

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2, \\ \mathrm{E}(X^2) &= \mathrm{E}(X^2 - X + X) = \mathrm{E}(X^2 - X) + \mathrm{E}(X) \\ &= \mathrm{E}(X(X-1)) + \mathrm{E}(X), \end{aligned}$$

pero

$$\begin{aligned} \mathrm{E}(X(X-1)) &= \sum_{x=1}^{+\infty} x(x-1) \cdot f(x) = \sum_{x=1}^{+\infty} x(x-1) q^{x-1} p \\ &= p \sum_{x=1}^{+\infty} x(x-1) q^{x-1} = pq \sum_{x=1}^{+\infty} x(x-1) q^{x-2} \\ &= pq \left(\frac{2}{(1-q)^3} \right) = pq \left(\frac{2}{p^3} \right) = \frac{2(1-2p)}{p^2}. \end{aligned}$$

Distribución Geométrica

La cantidad $\mathrm{E}(X(X-1))$ recibe el nombre de **Segundo Momento Factorial**, de aquí tenemos que

$$\mathrm{E}(X^2) = \mathrm{E}(X(X-1)) + \mathrm{E}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-2p+p}{p^2} = \frac{2-p}{p^2},$$

y entonces

$$\mathrm{V}(X) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.$$

Distribución Geométrica

Función Generatriz de Momentos:

$$M_X(t) = \mathrm{E}(e^{tX}) = \sum_{x=1}^{+\infty} e^{tx} \cdot f(x) = \sum_{x=1}^{+\infty} e^{tx} \cdot q^{x-1} p = p \sum_{x=1}^{+\infty} (e^t)^x \cdot q^{x-1}$$

$$= pe^t \sum_{x=1}^{+\infty} (e^t)^{x-1} \cdot q^{x-1} = pe^t \sum_{x=1}^{+\infty} (e^t q)^{x-1} \underset{(j=x-1)}{=} pe^t \sum_{j=0}^{+\infty} (e^t q)^j,$$

usando la identidad $\sum_{n=0}^{+\infty} z^n = \frac{1}{1-z}$, para $|z| < 1$, nos queda

$$M_X(t) = pe^t \left(\frac{1}{1 - qe^t} \right) = \frac{pe^t}{1 - qe^t} \qquad \text{si } |qe^t| < 1.$$

Distribución Geométrica

Ejemplo: Una tienda ofrece un premio a las personas que logren juntar las cinco letras de la palabra "valor"; para ello, se les regala una letra al azar de entre las cinco en cada compra que realicen. Lo que interesa aquí es la cantidad de viajes necesarios (compras) para juntar las cinco letras distintas.

Se podría pensar que la v.a. de interés, $X = \#$ de viajes necesarios hasta juntar las cinco letras de la palabra "valor", es geométrica ya que:

- cada resultado es independiente del anterior,
- el número posible de viajes, realizados entre cada consecución de letras distintas, se puede considerar como "infinito".

Distribución Geométrica

Sin embargo, la probabilidad de un "éxito" cada vez que compre irá cambiando: la primera vez la probabilidad de juntar una letra es 1; la segunda vez que compre ya tendrá en su poder una letra, con lo que la probabilidad de "éxito" cambiaría de 1 a la probabilidad de que reciba una letra distinta de la que se le dio inicialmente y esta sería 4/5 (4 favorables de 5 posibles); y así seguiría hasta obtener otro "éxito" (letra distinta), con lo que p cambiaría a 3/5, etc.

Entonces, se puede pensar que en cada viaje la probabilidad es mantenida fija hasta que haya "éxito", y juntando esto con los dos supuestos anteriores, se tendrían v.a.'s geométricas (exceptuando la primera), con p's distintas, cada vez que se lograra un "éxito". Se puede visualizar lo anterior usando una recta numérica:

Distribución Geométrica

Así, nuestra v.a. de interés podría considerarse como la suma de todos los viajes a la tienda, esto es, la suma de cada una de las v.a.'s geométricas definidas:

$$X = X_1 + X_2 + X_3 + X_4 + X_5,$$

donde $X_i \sim \text{Geom}(p_i)$ $i = 2, 3, 4, 5$, con $p_2 = 4/5$, $p_3 = 3/5$, $p_4 = 2/5$, $p_5 = 1/5$. Entonces, podemos encontrar el **número esperado de viajes** calculando

$$\begin{aligned} \text{E}(X) = & \ \text{E}\left(\sum_{i=1}^{5} X_i\right) = \text{E}(X_1) + \text{E}(X_2) + \text{E}(X_3) + \text{E}(X_4) + \text{E}(X_5) \\ = & \ 1 + \frac{1}{p_2} + \frac{1}{p_3} + \frac{1}{p_4} + \frac{1}{p_5} = 1 + \frac{1}{4/5} + \frac{1}{3/5} + \frac{1}{2/5} + \frac{1}{1/5} \\ = & \ 11.42 \text{ viajes}, \end{aligned}$$

esto es, entre 11 ó 12 viajes.

Distribución Geométrica

Ejemplo: Para el Sorteo Tec, si cada vez se expenden 150,000 boletos y 4 automóviles como premio, ¿en cuántos sorteos en promedio hay que participar hasta sacarse un automóvil? ¿Cuánto hay que invertir en promedio si cada boleto cuesta 1,050 pesos?

Como p es fija, hay independencia y "podemos" comprar boletos por siempre, entonces con $X = \#$ de participaciones hasta sacarse un automóvil, $\mathrm{E}(X) = \frac{150{,}000}{4} = 37,500$. Además, habría que invertir $1,050 \times 37,500 = 39,375,000$ pesos.

Con dos rifas por año, uno esperaría ganarse un auto en unos ¡18,750 años! Cuántos años tendrías que participar en promedio si quieres ganarte la casa del primer premio?

Distribución Binomial Negativa

Considera nuevamente el ejemplo del premio al juntar las letras de la palabra "valor", pero suponte que las probabilidades de éxito p_i se mantienen constantes (por ejemplo pensando que el premio se da a la persona que junte 5 letras "V", con lo que p_i siempre sería $1/5$). Como antes, la variable aleatoria $X =$ **# de viajes necesarios hasta obtener 5 "éxitos"**, entonces

$$X = \sum_{i=1}^{5} X_i, \quad \text{donde } X_i \sim \text{Geom}(p) \text{ son independientes y } p \text{ constante.}$$

Esto es equivalente al segundo criterio mencionado para la detención o ajuste de un proceso (ver diapositiva 56): hasta obtener r (por ejemplo 3) artículos defectuosos (debido tal vez a lo costoso de parar la línea al obtener el primer defectuoso), en donde

Distribución Binomial Negativa

- cada éxito tiene la misma probabilidad p, ya que se trata del mismo proceso y se asume que éste es estable (no hay fallas sistemáticas),
- cada X_i es independiente;

en este caso la variable aleatoria de interés es

$$X = \ \# \text{ de observaciones hasta encontrar } r \text{ “éxitos”},$$

la cual se puede redefinir como

$$X = \sum_{i=1}^{r} X_i, \text{ donde cada } X_i \sim \text{Geom}(p) \text{ es independiente y p constante.}$$

La v.a. X definida como la suma anterior se dice que sigue una distribución Binomial Negativa: $X \sim \text{BN}(r, p)$. Aquí, inclusive $X_1 \sim \text{Geom}(p)$.

Distribución Binomial Negativa

A continuación se muestra un dibujo similar al hecho para la distribución geométrica (ver diapositiva 76), aquí obs = observaciones y Geo=Geom:

obs	éxito 1	obs	éxito 2	$\cdots$	obs	éxito r-1	obs	éxito r
$\cdots$	$\bullet$	$\cdots$	$\bullet$	$\cdots$	$\cdots$	$\bullet$	$\cdots$	$\bullet$
$X_1 \sim$Geo(p)		$X_2 \sim$Geo(p)		$\cdots$	$X_{r-1} \sim$ Geo(p)		$X_r \sim$Geo(p)	

Media y Varianza

El valor esperado de la v.a. $X \sim \mathrm{BN}(r, p)$ es

$$\mathrm{E}(X) = \mathrm{E}\left(\sum_{i=1}^{r} X_i\right) = \sum_{i=1}^{r} \mathrm{E}(X_i) = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p};$$

para la varianza usamos el hecho de que las X_i's son independientes

$$\mathrm{V}(X) = \mathrm{V}\left(\sum_{i=1}^{r} X_i\right) = \sum_{i=1}^{r} \mathrm{V}(X_i) = \sum_{i=1}^{r} \frac{1-p}{p^2} = r\left(\frac{1-p}{p^2}\right).$$

Distribución Binomial Negativa

Función Generatriz de Momentos

Usando la estructura de suma se tiene que la función generatriz de momentos de una distribución BN(r,p) está dada por

$$\begin{aligned} M_X(t) = & \ \mathrm{E}(e^{tX}) = \mathrm{E}\left(e^{t\sum_{i=1}^{r} X_i}\right) = \mathrm{E}\left(e^{tX_1+tX_2+\cdots+tX_r}\right) \\ = & \mathrm{E}(e^{tX_1})\,\mathrm{E}(e^{tX_2})\,\cdots\,\mathrm{E}(e^{tX_r}) \\ = & \left(\frac{pe^t}{1-qe^t}\right)\left(\frac{pe^t}{1-qe^t}\right)\cdots\left(\frac{pe^t}{1-qe^t}\right) = \left(\frac{pe^t}{1-qe^t}\right)^r, \end{aligned}$$

para $|qe^t| < 1$.

Distribución Binomial Negativa

Recalquemos los supuestos:

- Cada vez que se realice una observación la probabilidad de que ocurra un éxito es fija e igual al valor p:
  Nuevamente, el proceso "no sabe" o "no se acuerda" de lo que ha pasado antes de la observación que se esté realizando. Cuando uno detecte artículos defectuosos, sabiendo la probabilidad que el modelo asigna a ese caso, se toma una decisión respecto al proceso. Ver[15].
- Las observaciones son independientes.
  Como antes, el que una observación haya sido o no "éxito", no da información sobre el resultado de la siguiente observación.
- El primer posible valor de la variable aleatoria es r.
  Puesto que la condición es encontrar r defectuosos, se tienen que realizar mínimo r pruebas.
- El número de observaciones puede continuar indefinidamente antes de encontrar el r-ésimo defectuoso.

Distribución Binomial Negativa

Una forma de obtener explícitamente a la función de probabilidad de la v.a. Binomial Negativa se da a continuación.

El evento de que el r-ésimo "éxito" ocurra en $X = x$ equivale a pensar que en las $x - 1$ observaciones anteriores hay $r - 1$ "éxitos", los cuales pudieron ocurrir en cualquiera de las anteriores $x - 1$ observaciones. Si agrupamos en casillas las observaciones, lo anterior puede visualizarse de la siguiente manera

$$\overbrace{\underbrace{|_____|_____|_____|\cdots\cdots|_____|}_{x-1 \text{ casillas}}}^{r-1 \text{ éxitos}} \quad \overbrace{\underbrace{|_____|}_{\text{casilla } \#x}}^{r\text{-ésimo éxito}}$$

Distribución Binomial Negativa

Debido a la independencia sólo se tienen que multiplicar las probabilidades de cada casilla para obtener la probabilidad de r "éxitos" en x pruebas:

- En la última prueba hay "éxito" con probabilidad p.
- Los $r-1$ "éxitos" anteriores pueden ocurrir en cualquier orden. El número de maneras en que $r-1$ "éxitos" se pueden acomodar en $x-1$ casillas está dado por $\binom{x-1}{r-1}$, y cada uno de ellos tiene asignada una probabilidad p.
- Hay $(x-1)-(r-1)=x-r$ fracasos cada uno con probabilidad $(1-p)$.

Distribución Binomial Negativa

Entonces, la probabilidad de obtener r éxitos en x pruebas es

$$\begin{aligned} \mathrm{P}(X = x) &= p\left[\binom{x-1}{r-1} p^{r-1}\right](1-p)^{x-r} \\ &= \binom{x-1}{r-1}(1-p)^{x-r}p^{r}, \end{aligned}$$

con $x = r, r+1, r+2, \ldots$

Esta es la Distribución Binomial Negativa cuyos parámetros son r y p. Notemos que n sigue siendo aleatorio.

Distribución Binomial Negativa

Relación entre una v.a. $X \sim \mathbf{B}(n, p)$ con una v.a. $Y \sim \mathbf{BN}(r, p)$

Recuerda que

$X =$ # de éxitos en n pruebas Bernoulli independientes,

$Y =$ # de observaciones hasta tener r éxitos en pruebas Bernoulli independientes.

Consideremos la probabilidad de que en los n intentos se observen a lo más $r - 1$ "éxitos", esto es,

$$\mathrm{P}(X \leq r - 1).$$

Observemos que en los n intentos hay a lo más $r - 1$ éxitos si y sólo si en los n intentos se observan menos de r éxitos, es decir que la probabilidad anterior es igual a la probabilidad

$$\mathrm{P}(Y > n).$$

Distribución Binomial Negativa

Entonces

$$P(X \leq r-1) = P(Y > n) = 1 - P(Y \leq n).$$

Cabe recalcar que esto **no** es una aproximación, es una identidad.

Ejemplo: Si un proceso nos da dos equipos con defectos "severos" se detiene la producción. La probabilidad de ese tipo de defectos es 0.05. ¿Cuál es la probabilidad de hacer más de 10 equipos sin detener la producción?

La variable aleatoria es $Y = \#$ de equipos revisados hasta obtener 2 con defectos "severos".

Distribución Binomial Negativa

Supuestos:

- probabilidad fija e igual a p (no hay fallas sistemáticas)
- resultados independientes

Entonces $Y \sim \mathrm{BN}(r = 2, p = 0.05)$, de donde

$$\mathrm{P}(Y > 10) = 1 - \mathrm{P}(Y \leq 10) = 1 - \sum_{y=2}^{10} \binom{y-1}{2-1} (0.05)^2 (0.95)^{y-2} = 0.914.$$

Utilizando la relación con la distribución binomial, para $X \sim \mathrm{B}(n = 10, p = 0.05)$ tenemos que

$$\mathrm{P}(Y > 10) = \mathrm{P}(X \leq 2 - 1) = \sum_{x=0}^{1} \binom{10}{x} (0.05)^x (0.95)^{10-x} = 0.914.$$

Distribución Binomial Negativa

Los cálculos de la generatriz de momentos, media y varianza se pueden obtener a partir de la función de probabilidad de la binomial negativa, aunque de hecho ya lo hicimos mediante la representación en sumas de la distribución binomial negativa. Por completez se incluye el siguiente este material.

Deduciremos la generatriz de momentos para obtener la media y la varianza de $X \sim \mathrm{BN}(r, p)$. Para calcular la generatriz de momentos se utilizarán los siguientes resultados:

$$\binom{n}{z} = \binom{n}{n-z}, \tag{1}$$

$$(1-x)^{-r} = \sum_{j=0}^{+\infty} \binom{j+r-1}{j} x^j, \quad \text{para } |x| < 1. \tag{2}$$

Distribución Binomial Negativa

La generatriz de momentos es entonces

$$
\begin{aligned}
M_X(t) = \mathrm{E}(e^{tX}) &= \sum_{x=r}^{+\infty} e^{tx} \cdot f(x) = \sum_{x=r}^{+\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} p^r \\
&= p^r \sum_{x=r}^{+\infty} e^{tx} \binom{x-1}{r-1} q^{x-r} = p^r (e^t)^r \sum_{x=r}^{+\infty} \binom{x-1}{r-1} (qe^t)^{x-r} \\
&= p^r (e^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{r-1} (qe^t)^{x-r} \qquad j = x - r \\
&= (pe^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{(j+r-1)-(r-1)} (qe^t)^{x-r} \qquad \text{de (1)} \\
&= (pe^t)^r \sum_{j=0}^{+\infty} \binom{j+r-1}{j} (qe^t)^{x-r} \\
&\overset{\text{de (2)}}{=} (pe^t)^r (1-qe^t)^{-r} = \frac{(pe^t)^r}{(1-qe^t)^r} = \left(\frac{pe^t}{1-qe^t}\right)^r, \qquad |qe^t| < 1.
\end{aligned}
$$

Distribución Binomial Negativa

Por lo tanto

$$M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r, \quad \text{para } |qe^t| < 1.$$

Cuando $r = 1$ la binomial negativa es equivalente a la Geométrica; es decir, la binomial negativa se puede considerar como una generalización de la geométrica. Así,

$$\mu = \mu_1' = \mathrm{E}(X) = M_X'(t)|_{t=0},$$

$$\begin{aligned} M_X'(t) &= r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{(1-qe^t)pe^t - pe^t(-qe^t)}{(1-qe^t)^2} \\ &= r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{pe^t - pqe^{2t} + pqe^{2t}}{(1-qe^t)^2} = r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{pe^t}{(1-qe^t)^2}, \\ &= \frac{r(pe^t)^r}{(1-qe^t)^{r+1}}. \end{aligned}$$

$$M_X'(0) = \frac{r(pe^0)^r}{(1-qe^0)^{r+1}} = \frac{rp^r}{(1-q)^{r+1}} = \frac{rp^r}{p^{r+1}} = \frac{r}{p}.$$

Distribución Binomial Negativa

Por lo tanto,

$$\mathrm{E}(X) = r/p.$$

Por otra parte,

$$\mu_2' = \mathrm{E}(X^2) = M_X''(t)|_{t=0},$$

$$M_X''(t) = \frac{(1-qe^t)^{r+1} r \cdot r(pe^t)^{r-1} pe^t - r(pe^t)^r (r+1)(1-qe^t)^r(-qe^t)}{[(1-qe^t)^{r+1}]^2},$$

$$M_X''(0) = \frac{(1-q)^{r+1} r^2 p^{r-1} p - rp^r (r+1)(1-q)^r(-q)}{(1-q)^{2(r+1)}}$$

$$= \frac{p^{r+1} r^2 p^r + r(r+1) p^r p^r q}{p^{2r+2}}$$

$$= \frac{p^{2r+1} r^2 + r(r+1) p^{2r} q}{p^{2r+2}} = \frac{rp^{2r}(rp + (r+1)q)}{p^{2r+2}}$$

$$= \frac{r\left(\overbrace{rp + rq}^{r} + q\right)}{p^2} = \frac{r(r+q)}{p^2}.$$

Distribución Binomial Negativa

Así que

$$\mathrm{E}(X^2) = \frac{r(r+q)}{p^2}.$$

Entonces,

$$\begin{aligned}\mathrm{V}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 &= \frac{r(r+q)}{p^2} - \left(\frac{r}{p}\right)^2 \\ &= \frac{r(r+q) - r^2}{p^2} = \frac{r^2 + rq - r^2}{p^2} = \frac{rq}{p^2} = \frac{r(1-p)}{p^2},\end{aligned}$$

lo que implica

$$\mathrm{V}(X) = \frac{r(1-p)}{p^2}.$$

Distribución Binomial Negativa

A modo de resumen, notemos lo siguiente:

- Notemos que en un contexto teórico el valor de r puede ser distinto de un número entero.
- Además, la variable aleatoria BINOMIAL, cuando $n = 1$, nos da una variable aleatoria que tiene distribución Bernoulli; es decir, la Bernoulli es un caso particular de la binomial.
- Similarmente, una variable aleatoria BINOMIAL NEGATIVA, con $r = 1$, nos da una variable aleatoria con distribución GEOMÉTRICA; es decir, la geométrica es un caso particular de la binomial negativa.

Distribución Poisson

Supóngase ahora que estamos interesado en el número de "éxitos", pero ahora preguntándonos por la posibilidad de que ocurran en un cierto intervalo de tiempo o espacio, por ejemplo

- Número de defectos en una plancha de acero de 1 m^2,
- Número de bacterias en 1 cm^3 de agua potable,
- Número de entregas de materia prima entre las 8:00 A.M. y la 1:00 P.M.,
- Número de llamadas que llegan a un conmutador en un minuto.

En estos casos interesa la variable aleatoria

$X =$ # de "éxitos" obtenidos en un cierto intervalo (temporal o espacial).

Distribución Poisson

Nuestros supuestos son:

- El número promedio de "éxitos", λ (la intensidad), sobre el intervalo dado se mantiene constante.
  Se considera que el comportamiento es estable. Incluso, si se estuviera interesado en un múltiplo t (no necesariamente entero) del intervalo original, el valor λ_1 correspondiente será $\lambda_1 = \lambda t$.
- Los éxitos aparecen en forma aleatoria en intervalos de la misma magnitud, los intervalos no se traslapan y son independientes entre ellos.
  Se considera que si en un sub-intervalo del intervalo dado ocurre un determinado número de "éxitos", no implica que en otro sub-intervalo de la misma amplitud tendrá ese mismo número, mayor que él, o menor. Sólo se espera que en promedio se tenga el mismo número.

Distribución Poisson

Observa que no se hace un número de pruebas hasta obtener "éxitos", sino que se cuenta el número de éxitos obtenidos en un continuo espacial o temporal y por ende, cuando el intervalo tiende a cero la probabilidad de éxito también tiende a cero.

- El número posible de "éxitos" es infinito.

Distribución Poisson

Mecanismo para asignación de probabilidades

Supongamos que tenemos un proceso Poisson y lo observamos en un cierto intervalo, digamos de una unidad, y asumimos que λ es el número promedio de éxitos en esa unidad observada.

Por otra parte, consideremos esa misma unidad y que:

- Se realizan 10 observaciones y se cuenta el número de "éxitos", cada uno con probabilidad p, obtenidos en 10 subdivisiones. Es natural pensar que hay una probabilidad relativamente pequeña de observar más de un éxito, en cada subdivisión.
- Se realizan 10 observaciones y se cuenta el número de "éxitos", cada uno con probabilidad p (más pequeña que en el anterior caso), obtenidos en mil subdivisiones del intervalo de tal manera que p será tan pequeña que hay una probabilidad de cero de observar más de un éxito, en cada subdivisión.

Distribución Poisson

El número de "éxitos" Y en n pruebas independientes con probabilidad p se puede modelar ahora como una v.a. Binomial, cuya media es $\mathrm{E}(Y) = np$, que debe coincidir con el número promedio de éxitos λ en el intervalo considerado.

Notemos que $\lambda = np$ implica que $p = \lambda/n$ y como λ es fija, p será más y más pequeña a medida que n crece. Esto nos lleva a establecer el siguiente resultado:

La distribución de Y es Poisson con $\lambda = np$ constante, cuando el número de observaciones tiende a infinito en el intervalo, con p siendo más pequeña conforme n aumenta.

Distribución Poisson

Demostración:
Si $X \sim \mathrm{B}(n,p)$, entonces

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad np = \lambda \to p = \frac{\lambda}{n}$$

$$\begin{aligned}
f(x) &= \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)(n-2)\cdots(n-(x-1))(n-x)!}{x!(n-x)!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}\lambda^x\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x} \\
&= \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)}{x!}\lambda^x\left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-x}.
\end{aligned}$$

Distribución Poisson

Por otro lado,

- $1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\to 1 \quad \text{cuando } n\to+\infty$
- $\left(1-\frac{\lambda}{n}\right)^{-x}\to 1 \quad \text{cuando } n\to+\infty$
- $x!$ y λ^x quedan igual

El único término faltante es $\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$. Observe que $\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$ tiene la forma $1^{-\infty}$ que es una forma indeterminada.

Distribución Poisson

Sea $y = \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$, entonces

$$\log(y) = -\frac{n}{\lambda}\log\left(1 - \frac{\lambda}{n}\right) = \frac{-\log\left(1 - \frac{\lambda}{n}\right)}{\frac{\lambda}{n}}.$$

De aquí tenemos,

$$\lim_{n\to\infty} \log(y) = \lim_{n\to\infty} \frac{-\log\left(1 - \frac{\lambda}{n}\right)}{\frac{\lambda}{n}}, \quad \text{forma indeterminada del tipo } \frac{0}{0};$$

aplicando la Regla de L'Hopital

$$\lim_{n\to\infty} -\frac{\frac{1}{(1-\lambda/n)} \cdot \left(\frac{\lambda}{n^2}\right)}{\left(-\frac{\lambda}{n^2}\right)} = \lim_{n\to\infty} \frac{1}{\left(1 - \frac{\lambda}{n}\right)} = 1 \qquad \Rightarrow \qquad \lim_{n\to\infty} \log(y) = 1$$

Distribución Poisson

Ahora, usando el hecho de que si f es una función continua se cumple

$$\lim_{x \to a} f(y) = f\left(\lim_{x \to a} y\right);$$

y, por ser la función logaritmo natural una función continua, tenemos

$$\log\left(\lim_{n \to \infty} y\right) = 1 \qquad \Rightarrow \qquad \lim_{n \to \infty} y = e.$$

De aquí que

- $\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} \to e$, cuando $n \to \infty$.

Distribución Poisson

Usando todos los resultados anteriores tenemos que la función de probabilidad de la **Distribución Poisson** está dada por

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{para } x = 0, 1, 2, \ldots$$

Notación: $X \sim \text{Poisson}(\lambda)$. El único parámetro del modelo es λ.

Distribución Poisson

Distribución Poisson

¿Qué podemos aprender del procedimiento anterior?

1. La fórmula para asignar probabilidades, como siempre, es válida bajo una serie de supuestos que debemos verificar se cumplan razonablemente en nuestros estudios particulares. Por otro lado, es el resultado de un proceso matemático no necesariamente intuitivo, como lo había sido en los casos anteriores.
2. Aunque no siempre se tiene que pensar que un modelo Poisson proviene de una variable binomial tomada en circunstancias límite como se mostró arriba, sí nos deja una herramienta de aproximación muy útil. Más aún, la aproximación es utilizada a favor nuestro en el sentido de que, los cálculos bajo una Poisson son, en general, más simples que con una binomial.

Distribución Poisson

La variable aleatoria Poisson se puede usar para calcular probabilidades binomiales en forma aproximada cuando n es muy grande y p es pequeña.

Ejemplos: La aproximación es buena cuando $n \geq 20$ y $p \leq 0.05$. Cuando $n \geq 100$ y $np < 10$ la aproximación es excelente.

Distribución Poisson

Función generatriz de momentos:

$$M_X(t) = \mathrm{E}(e^{tX}) = \sum_{x=0}^{+\infty} e^{tx} f(x) = \sum_{x=0}^{+\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\lambda^x e^{tx}}{x!}.$$

Recordando el desarrollo en series de potencias de la función exponencial

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \cdots,$$

obtenemos

$$M_X(t) = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

Distribución Poisson

Ahora, utilizando la relación entre $M_X(t)$ y $\mathrm{E}(X^r)$, obtenemos **la media y la varianza**

$$\mathrm{E}(X) = \mu = \mu_1' = M_X'(t)|_{t=0}$$
$$\Rightarrow M_X'(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t \Rightarrow M_X'(0) = e^{\lambda(e^0-1)} \cdot \lambda e^0 = \lambda.$$

Por lo tanto, $\mathrm{E}(X) = \lambda$.

Ahora,

$$\mathrm{E}(X^2) = \mu_2' = M_X''(t)|_{t=0}$$
$$\Rightarrow M_X''(t) = e^{\lambda(e^t-1)} \cdot \lambda e^t + \lambda e^t e^{\lambda(e^t-1)} \cdot \lambda e^t \Rightarrow M_X''(0) = \lambda + \lambda^2.$$

Por lo tanto, $\mathrm{E}(X^2) = \lambda + \lambda^2$.

Distribución Poisson

Entonces

$$V(X) = E(X^2) - [E(X)]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Nota: Demuestre que $C_X(t) = \log M_X(t)$, $C'_X(0) = \mu$, $C''_X(0) = \sigma^2$.

Usando lo anterior

$$C_X(t) = \log\left(e^{\lambda(e^t-1)}\right) = \lambda(e^t - 1),$$

de donde

$$\begin{aligned} C'_X(t) = \lambda e^t \quad &\Rightarrow \quad C'_X(0) = \lambda e^0 = \lambda = \mu \\ C''_X(t) = \lambda e^t \quad &\Rightarrow \quad C''_X(0) = \lambda e^0 = \lambda = \sigma^2 \end{aligned}$$

Distribución Poisson

Ejemplo. Sabiendo que, según las condiciones de un bosque, se esperan encontrar 2 chapulines por m^2: ¿qué tan grande debe ser el radio de una región circular de muestreo dentro del bosque para que la probabilidad de encontrar al menos un chapulín sea de 0.99?

Solución: En este caso, $X = \#$ de chapulines en una área circular determinada de πr^2. La v.a. X se distribuye Poisson($\lambda = 2\frac{\text{chapulines}}{m^2} \cdot \pi r^2$). (¿Porqué?) Así,

$$\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X \leq 0) = 1 - \mathrm{P}(X = 0) = 1 - \frac{(2\pi r^2)^0 e^{-2\pi r^2}}{0!} = 0.99,$$

entonces

$$e^{-2\pi r^2} = 0.01 \qquad \Rightarrow \qquad r = 0.8561\, m.$$

Distribución Poisson

Ejemplo. Una compañía renta máquinas fotocopiadoras. Se sabe que el número de reportes por averías por semana que recibe es una v.a. Poisson$\left(\lambda = 2\frac{\text{averías}}{\text{semana}}\right)$. a) ¿Cuál es la probabilidad de no recibir queja en una semana? b) ¿Cuál es la probabilidad de recibir más de tres reportes dado que ya se recibió una queja? c) ¿Cuál es la probabilidad de recibir menos de seis quejas en un mes?

Solución:

$$Y = \# \text{ de quejas por semana.}$$

1. Se pide $\mathrm{P}(Y = 0) = 2^0 e^{-2}/0! = 0.1354$.
2. Aquí lo que interesa es $\mathrm{P}(X > 3|X \geq 1) = \frac{\mathrm{P}(X>3, X\geq 1)}{\mathrm{P}(X\geq 1)} = \frac{\mathrm{P}(X>3)}{\mathrm{P}(X\geq 1)} = \frac{1-F(3)}{1-F(0)} = \frac{1-0.857}{1-0.1354} = 0.1654$.
3. En este caso $\lambda = 4 \cdot 2 = 8$, ya que en un mes hay 4 semanas. Entonces, $Y = \#$ de quejas en un mes, es Poisson$(\lambda = 8)$, así

$$\mathrm{P}(Y < 6) = \mathrm{P}(Y \leq 5) = F(5) = 0.191.$$

Distribución Poisson

Estos ejemplos muestran la importancia de las unidades que se están manejando para el intervalo de interés.

La distribución de probabilidad Poisson es derivada en el libro de Meyer, en una forma muy interesante, además de remarcar su importancia como "proceso" y no sólo como modelo probabilístico. Se discute más adelante.

Distribución Binomial Negativa: Sobredispersión

La distribución binomial negativa es útil para modelar datos de conteos con **sobredispersión**. En este contexto, decimos que un conjunto de datos de conteo exhiben sobredispersión si muestran mayor variabilidad de la que se esperaría con una distribución Poisson.

En la práctica, la sobredispersión aparece en datos de conteo que surgen en epidemias o en dinámicas poblacionales y es causada por la aleatoriedad en el movimiento de la población o en las tasa de contacto, o por deficiencias en el modelo para capturar adecuadamente en la dinámica poblacional.

Es posible incluir a la sobredispersión como un paramétro en la función de masa de una binomial negativa, para esto es necesario realizar la siguiente reparametrización.

Distribución Binomial Negativa: Sobredispersión

Sea $Y \sim \mathrm{BN}(r, p)$, entonces

$$P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r} \qquad y = r, r+1, \ldots,$$

si tomamos $H = Y - r$, $y = h + r$, la ecuación anterior se puede reescribir como

$$\begin{aligned} P(H = h) &= \binom{h+r-1}{r-1} p^r (1-p)^h \\ &= \frac{\Gamma(h+r)}{\Gamma(r)h!} p^r (1-p)^h \qquad h = 0, 1, \ldots \end{aligned}$$

Poniendo $X = H$, $x = h$, $k = r$, la igualdad anterior es equivalente a

$$P(X = x) = \frac{\Gamma(k+x)}{\Gamma(r)x!} p^k (1-p)^x \qquad x = 0, 1, \ldots$$

Distribución Binomial Negativa: Sobredispersión

La esperanza m de X está dada por $m = (1-p)k/p$, lo que implica que $p = k/(m+k)$. Así,

$$\begin{aligned} P(X=x) &= \frac{\Gamma(k+x)}{\Gamma(r)x!} p^k (1-p)^x \\ &= \frac{\Gamma(k+x)}{\Gamma(r)x!} \left(\frac{k}{m+k}\right)^k \left(\frac{m}{m+k}\right)^x \\ &= \frac{\Gamma(k+x)}{\Gamma(r)x!} \left(1+\frac{m}{k}\right)^{-k} \left(\frac{m}{m+k}\right)^x \qquad x = 0, 1, \ldots \end{aligned}$$

Se puede ver que

$$E(X) = \mu = m$$

$$\mathrm{Var}(X) = \sigma^2 = m + \frac{m^2}{k}.$$

Tendremos sobredispersión cuando $\sigma^2 > \mu$ (sobredispersión relativa a la Poisson). El parámetro k se conoce como el **parámetro de dispersión**.

Procesos de Poisson

Consideremos una fuente de material radiactivo que emite partículas α.

Definamos a X_t como el número de partículas emitidas durante un periodo de tiempo específico $[0, t]$.

Vamos a hacer algunas hipótesis acerca de la variable aleatoria (discreta) X_t que nos permitirán determinar la distribución de probabilidades de X_t . La posibilidad de estas hipótesis se justifica por el hecho de que la evidencia empírica sostiene una cantidad considerable de resultados teóricos que vamos a derivar.

Observa que la variable aleatoria X_t puede tomar los valores $\{0, 1, 2, \ldots\}$. Pongamos $p_n(t) = P(X_t = n)$, $n = 0, 1, \ldots$

Procesos de Poisson

Las hipótesis sobre X_t que vamos a enunciar son cinco:

1. El número de partículas emitidas durante intervalos de tiempo no sobrepuestos son variables aleatorias independientes.
2. Si Y_t se define es igual al número de partículas emitidas durante $[t_1, t_1 + t]$, para cualquier $t_1 > 0$, las variables aleatorias X_t y Y_t tienen la misma distribución de probabilidades.
   En otras palabras, la distribución del número de partículas emitidas durante cualquier intervalo depende sólo de la longitud del intervalo y no de los puntos extremos.
3. $p_1(\Delta t)$ es igual aproximadamente a $\lambda \Delta t$, si Δt es suficientemente pequeña, donde λ es una constante positiva. Es decir que $p_1(\Delta t) \sim \lambda \Delta t$. También supondremos que $\Delta t > 0$.
   Esta hipótesis expresa que si el intervalo es suficientemente pequeño, la probabilidad de obtener exactamente una emisión durante ese intervalo es directamente proporcional a la longitud del intervalo.

Procesos de Poisson

Recordatorio de notación: $a(\Delta t) \sim b(\Delta t)$ significa que $a(\Delta t) = b(\Delta t) + o(\Delta t)$ para una función indeterminada $o(\Delta t)$ que representa el resto y satisface $o(\Delta t)/\Delta t \to 0$ cuando $\Delta t \downarrow 0$.

Continuación de las hipótesis:

4. $\sum_{k=2}^{\infty} p_k(\Delta t) \sim 0$. Nótese que esto implica que $p_k(\Delta t) \to 0$, $k \geq 2$.
   Esto significa que la probabilidad de obtener dos o más emisiones en un intervalo suficientemente pequeño es despreciable.
5. $X_0 = 0$, o de manera equivalente $p_0(0) = 1$.
   Esto equivale a una condición inicial para el modelo que estamos describiendo.

Las cinco hipótesis anteriores harán posible que deduzcamos una expresión para $p_n(t) = P(X_t = n)$, como veremos más adelante.

Procesos de Poisson

Saquemos algunas conclusiones de las hipótesis anteriores:

1. Las hipótesis 1 y 2 juntas implican que la variable aleatoria $X_{\Delta t}$ y $X_{t+\Delta t} - X_t$ son variables aleatorias independientes con la misma distribución de probabilidades.
2. De las hipótesis 3 y 4 podemos concluir que

$$p_0(\Delta t) = 1 - p_1(\Delta t) - \sum_{k=2}^{\infty} p_k(\Delta t) \sim 1 - \lambda \Delta t.$$

3. Podemos escribir

$$\begin{aligned} p_0(t+\Delta t) =& P(X_{t+\Delta t} = 0) \\ =& P(X_t = 0 \text{ y } X_{t+\Delta t} - X_t = 0) \\ =& p_0(t) p_0(\Delta t) \\ \sim& p_0(t)(1 - \lambda \Delta t) \end{aligned}$$

Procesos de Poisson

Continuación de las conclusiones de las hipótesis anteriores:

4. Luego tenemos

$$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t} \sim -\lambda p_0(t).$$

Haciendo $\Delta t \downarrow 0$, lo anterior implica

$$p_0'(t) = -\lambda p_0(t) \Longleftrightarrow \frac{p_0'(t)}{p_0(t)} = -\lambda.$$

Resolviendo la ecuación diferencial anterior llegamos a

$$p_0(t) = \exp\left(-\lambda t\right).$$

Así, nuestras hipótesis nos han conducido a una expresión para $P(X_t = 0)$.

Procesos de Poisson

Continuación de las conclusiones de las hipótesis anteriores:

5. Considerando $p_n(t+\Delta t) = P(X_{t+\Delta t} = n)$, tenemos que

$$\begin{aligned}
p_n(t+\Delta t) =& P(X_{t+\Delta t} = n) = \sum_{x=0}^{n} P(X_t = x, X_{t+\Delta t} = n) \\
=& \sum_{x=0}^{n} P(X_t = x, X_{t+\Delta t} - X_t = n - x) \\
=& \sum_{x=0}^{n} P(X_t = x) \cdot P(X_{t+\Delta t} - X_t = n - x) = \sum_{x=0}^{n} p_x(t) p_{n-x}(\Delta t) \\
=& \sum_{x=0}^{n-2} p_x(t) p_{n-x}(\Delta t) + p_{n-1}(t) p_1(\Delta t) + p_n(t) p_0(\Delta t)
\end{aligned}$$

Usando las hipótesis 3 y 4 y la conclusión 2, obtenemos

$$p_n(t+\Delta t) \sim p_{n-1}(t)\lambda\Delta t + p_n(t)(1 - \lambda\Delta t).$$

Procesos de Poisson

5. Continuación: Luego tenemos

$$\frac{p_n(t+\Delta t)-p_n(t)}{\Delta t} \sim \lambda p_{n-1}(t) - \lambda p_n(t).$$

Haciendo $\Delta t \downarrow 0$, como en la consecuencia anterior, obtenemos

$$p_n'(t) = \lambda p_{n-1}(t) - \lambda p_n(t), \;\; n = 1, 2, \ldots$$

Definamos $q_n(t) = e^{\lambda t} p_n(t)$. Así el sistema de ecuaciones diferenciales anterior se transforma en $q_n'(t) = \lambda q_{n-1}(t)$, $n = 1, 2, \ldots$ Puesto que $p_0(t) = e^{-\lambda t}$, encontramos que $q_0(t) = 1$. Nótese que también $q_n(0) = 0$, para $n > 0$. Así, recursivamente obtenemos

$$q_1'(t) = \lambda, \text{ y por tanto } q_1(t) = \lambda t;$$

$$q_2'(t) = \lambda q_1(t) = \lambda^2 t, \text{ y por tanto } q_2(t) = \frac{(\lambda t)^2}{2}.$$

Procesos de Poisson

5. Continuación: En general, $q_n'(t) = \lambda q_{n-1}(t)$ y por tanto $q_n(t) = (\lambda t)^n/n!$. Recordando la definición de q_n, finalmente obtenemos
$$p_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad \text{n=0,1,2,...}$$

Hemos demostrado así que el número de partículas emitidas durante el intervalo de tiempo $[0, t)$ de una fuente radioactiva, con las suposiciones hechas anteriormente, es una variable aleatolria con una distribución de Poisson con parámetro λt.

Procesos de Poisson

Observaciones sobre lo que hicimos arriba:

a) Es importante darse cuenta de que la distribución de Poisson apareció como una consecuencia de ciertas suposiciones que hicimos. Esto significa que cada vez que dichas suposiciones sean válidas (o al menos lo sean aproximadamente) la distribución de Poisson debe usarse como un modelo apropiado.
Resulta que hay una gran cantidad de fenómenos para los cuales es adecuado el modelo de Poisson:
   - *Representemos por X_t el número de llamadas telefónicas que llegan a una central telefónica durante un periodo de tiempo de longitud t. Las suposiciones anteriores se satisfacen aproximadamente, en especial durante el "periodo congestionado" del día. Luego, X_t tiene una distribución de Poisson.*
   - *Representemos por X_t el número de electrones que salen del cátodo de un tubo al vacío. Nuevamente las suposiciones son apropiadas y, por tanto, X_t tiene una distribución de Poisson.*

Procesos de Poisson

Observaciones sobre lo que hicimos arriba:

b) La constante λ apareció originalmente como una constante de proporcionalidad en la hipótesis 3.

Vale la pena mencionar las siguientes interpretaciones de λ:

- *Si X_t representa el número de ocurrencias de un evento durante un intervalo de tiempo de longitud t, entonces, $E(X_t) = \lambda t$ y, por tanto, $\lambda = E(X_t)/t$ representa la razón esperada con la cual se emiten las partículas.*
- *Si X_v representa el número de ocurrencias de algún evento dentro de un volumen especificado V, entonces $E(X_v) = \lambda V$, por lo tanto, $\lambda = E(X_v)/V$ representa la densidad esperada con la cual aparecen las estrellas.*

Procesos de Poisson

Observaciones sobre lo que hicimos arriba:

c) Es importante señalar que nuestra exposición no se refirió sólo a una variable aleatoria X que posee una distribución de Poisson, sino que para cada $t > 0$, encontramos que X_t tenía una distribución de Poisson con un parámetro dependiente de t. **Tal colección (infinita) de variables aleatorias también se conoce como proceso de Poisson**.

   De igual forma, se genera un proceso de Poisson cada vez que ocurre un evento en algún intervalo de tiempo de modo que se satisfagan las hipótesis 1 hasta 5.

Procesos de Poisson

Ejemplo: Una complicada maquinaria, cuando funciona perfectamente, puede producir una utilidad de C dólares por hora ($C > 2$) a una compañía. Sin embargo, esta máquina tiene una tendencia a fallar en momentos inesperados e impredecibles. Supóngase que el número de fallas durante cualquier periodo de longitud t horas es una variable aleatoria con una distribución de Poisson con parámetro t.

Si la máquina falla x veces durante t horas, la pérdida ocasionada (la improductividad de la máquina más la reparación) es igual a $x^2 + x$ dólares. Luego, la utilidad total P durante cualquier periodo de t horas es igual a $P = Ct - (X^2 + X)$, donde X es la variable aleatoria que representa el número de fallas de la máquina.

Procesos de Poisson

Ejemplo (cont): Por tanto, P es una variable aleatoria, y podría ser interesante elegir t (lo que está a nuestra voluntad) de manera tal que la utilidad esperada sea maximizada. Tenemos

$$E(P) = Ct - E(X^2 + X).$$

Por lo estudiado sobre la distribución Poisson sabemos que $E(X) = t$ y $E(X^2) = t + t^2$. Luego se deduce que $E(P) = Ct - 2t - t^2$. Para encontrar el valor de t, para el cual se maximiza $E(P)$, diferenciamos $E(P)$ e igualamos a cero la expresión resultante. Obtenemos $C - 2 - 2t = O$, obteniendo $t = \frac{1}{2}(C - 2)$ horas.

Procesos de Poisson

Ejemplo: Sea X_t igual al número de partículas emitidas por una fuente radiactiva durante un intervalo de tiempo de longitud t. Supóngase que X_t tiene una distribución de Poisson con parámetro αt. Se instala un instrumento para anotar el número de partículas emitidas. Supóngase que hay una probabilidad constante p de que cualquier partícula emitida no se cuente. Si R_t es igual al número de partículas contadas durante el intervalo específico, cuál es la distribución de probabilidades de R_t?

Para $X_t = x$ dada, la variable aleatoria R_t tiene una distribución binomial con parametros $(x, 1-p)$. Esto es,

$$P(R_t = k | X_t = x) = \binom{x}{k}(1-p)^k p^{x-k}.$$

Procesos de Poisson

Ejemplo (cont.): Usando la fórmula de probabilidad total,

$$
\begin{aligned}
P(R_t = k) &= \sum_{x=k}^{\infty} P(R_t = k | X_t = x) P(X_t = x) \\
&= \sum_{x=k}^{\infty} \binom{x}{k} (1-p)^k p^{x-k} e^{-\alpha t} \frac{(\alpha t)^x}{x!} \\
&= \left(\frac{1-p}{p}\right)^k \frac{e^{-\alpha t}}{k!} \sum_{x=k}^{\infty} (p\alpha t)^x \frac{1}{(x-k)!} \\
&= \left(\frac{1-p}{p}\right)^k \frac{e^{-\alpha t}}{k!} \sum_{i=0}^{\infty} (p\alpha t)^{i+k} \frac{1}{i!} = \left(\frac{1-p}{p}\right)^k \frac{e^{-\alpha t}}{k!} (p\alpha t)^k e^{p\alpha t} \\
&= e^{-\alpha t(1-p)} \frac{(\alpha t(1-p))^k}{k!}.
\end{aligned}
$$

Es decir que R_t tiene una distribución de Poisson con parámetro $(1-p)\alpha t$.

Procesos de Poisson

Theorem

Si $\{N(s), s \geq 0\}$ *es un proceso de Poisson, entonces*

1. $N(0) = 0$.
2. $N(t+s) - N(s) \sim Poisson(\lambda t)$.
3. $N(t)$ *tiene incrementos independientes.*

Recíprocamente, si 1, 2 y 3 valen, entonces $\{N(s), s \geq 0\}$ *es un proceso de Poisson.*

Procesos de Poisson

Una forma más compleja, pero más realista es pensar que λ cambia con el tiempo.

Eso es fácil de imaginar, piensa por ejemplo en: # de llegadas de clientes a un banco en un período de una hora. Claramente, si tomamos períodos de observación entre las 9-10 a.m. comparados con 12-1 p.m., no tendríamos porque esperar que el número promedio de clientes fuera el mismo.

Lo mismo si el día de la semana es lunes, o viernes o peor aún, día de pago.

Procesos de Poisson

Ejemplo. En el contexto de la fuente de material radioactivo (diapositiva 115), ahora cambiemos la hipótesis 3 a que la probabilidad de que exactamente un evento ocurra en el intervalo $[t, t + \Delta t]$ es $\alpha(t) \cdot \Delta t + O(\Delta t)$, para Δt pequeño.

Repitiendo el procedimiento que se hizo para las hipótesis originales, se puede mostrar que el número de eventos que ocurren durante el intervalo $[t_1, t_2]$ sigue una distribución Poisson con parámetro

$$\lambda = \int_{t_1}^{t_2} \alpha(t) dt.$$

La ocurrencia de eventos en el tiempo en una situación como esta es llamada: **Proceso Poisson No-homogéneo**.

Procesos de Poisson

Ejemplo: Un artículo ("Inference Based on Retrospective Ascertainment", J. Amer. Stat. Assoc., 1989, pp. 360-372) consideró la siguiente función de intensidad:

$$\alpha(t) = e^{a+bt}$$

como apropiada para eventos que involucran la transmisión del virus del SIDA a través de transfuciones sanguíneas.

Supongamos que $a = 2$ y $b = 0.6$ (tiempo en años)

1. ¿Cuál es el número esperado de eventos en el intervalo [0,4]? ¿En el [2,6]?
2. ¿Cuál es la probabilidad de que a lo más 15 eventos ocurran en el intervalo [0, 0.9907]?

Procesos de Poisson

Solución.

1. Como se trata de un proceso Poisson, λ sigue representando al valor medio (hecho que puede demostrarse), por lo que sólo nos restaría evaluar λ para cada intervalo:

   En [0,4]: $\lambda = \int_{t1}^{t_2} \alpha(t)dt = \int_0^4 e^{2+0.6t}dt = 123.44$

   En [2,6]: $\lambda = \int_{t1}^{t_2} \alpha(t)dt = \int_2^6 e^{2+0.6t}dt = 409.82$

2. Primero obtengamos el valor de λ en este intervalo:

$$\lambda = \int_0^{0.9907} e^{2+0.6t}dt = 9.99 \sim 10$$

   $\mathrm{P}(X \leq 15) = F(15, 10) = 0.978$, esto es, aproximadamente en el primer año, la probabilidad de que a lo más 15 eventos se presenten, es del 97.8 %.

Procesos de Poisson

Ejemplo

Una empresa proveedora de servicios de Internet cuenta actualmente con una infraestructura de 7 líneas de un solo teléfono, las cuales van siendo ocupadas a medida que los usuarios se van conectando.

Una vez que las 7 líneas están siendo utilizadas simultáneamente, no se puede realizar otra conexión, por lo tanto, la infraestructura resulta ineficiente para dar un servicio adecuado.

La empresa desea aumentar el número de líneas de conexión a red, por lo que necesita saber en qué momento es más probable que se sature el servicio y la intensidad de la saturación.

Con esta finalidad, la empresa recopiló información sobre el número de usuarios que se conectan por hora (24 horas) durante 12 semanas.

Procesos de Poisson

Por ejemplo, para la semana 1:

1 hrs	2.00	1.00	2.00	0.00	1.00	0.00	3.00	0.00	0.00	2.00	$\cdots$
2 hrs	3.00	5.00	1.00	2.00	0.00	0.00	0.00	1.00	2.00	4.00	$\cdots$
3 hrs	0.00	0.00	2.00	3.00	2.00	0.00	0.00	2.00	1.00	0.00	$\cdots$
4 hrs	2.00	1.00	4.00	1.00	0.00	1.00	0.00	1.00	1.00	3.00	$\cdots$
5 hrs	2.00	0.00	1.00	3.00	4.00	1.00	0.00	1.00	2.00	1.00	$\cdots$
6 hrs	1.00	0.00	1.00	1.00	1.00	1.00	2.00	3.00	1.00	0.00	$\cdots$
7 hrs	0.00	1.00	2.00	3.00	0.00	1.00	3.00	0.00	3.00	2.00	$\cdots$
8 hrs	1.00	1.00	1.00	1.00	2.00	2.00	1.00	2.00	1.00	2.00	$\cdots$
9 hrs	0.00	1.00	3.00	2.00	5.00	2.00	2.00	1.00	1.00	1.00	$\cdots$
10 hrs	2.00	4.00	1.00	8.00	5.00	2.00	2.00	3.00	3.00	5.00	$\cdots$
11 hrs	3.00	2.00	2.00	3.00	1.00	3.00	1.00	2.00	2.00	1.00	$\cdots$
12 hrs	3.00	4.00	1.00	3.00	5.00	0.00	4.00	0.00	4.00	1.00	$\cdots$
13 hrs	0.00	8.00	6.00	3.00	3.00	5.00	0.00	3.00	3.00	1.00	$\cdots$
14 hrs	2.00	2.00	3.00	3.00	2.00	6.00	1.00	5.00	4.00	1.00	$\cdots$
15 hrs	0.00	4.00	4.00	1.00	3.00	1.00	1.00	2.00	6.00	2.00	$\cdots$
16 hrs	5.00	1.00	4.00	3.00	3.00	4.00	1.00	3.00	2.00	5.00	$\cdots$
17 hrs	4.00	4.00	4.00	3.00	5.00	4.00	5.00	2.00	2.00	6.00	$\cdots$
18 hrs	1.00	1.00	2.00	1.00	4.00	0.00	4.00	1.00	2.00	1.00	$\cdots$
19 hrs	2.00	2.00	3.00	1.00	4.00	5.00	1.00	2.00	2.00	4.00	$\cdots$
20 hrs	1.00	1.00	2.00	2.00	7.00	3.00	3.00	3.00	1.00	3.00	$\cdots$
21 hrs	2.00	3.00	10.00	5.00	3.00	6.00	4.00	1.00	1.00	3.00	$\cdots$
22 hrs	3.00	1.00	3.00	4.00	3.00	5.00	3.00	3.00	2.00	4.00	$\cdots$
23 hrs	2.00	4.00	2.00	2.00	2.00	2.00	5.00	4.00	3.00	5.00	$\cdots$
24 hrs	8.00	1.00	2.00	4.00	5.00	2.00	2.00	4.00	3.00	5.00	$\cdots$

Ejemplo. Análisis exploratorio (Lo que no se debe de hacer con un proceso estocástico)

Ejemplo. Análisis exploratorio

Ejemplo. Análisis exploratorio

Ejemplo. Selección del modelo

El fenómeno de interés es el número de usuarios por hora, o de forma más general, el **número de usuarios en el intervalo** $(s, s+k)$. Por lo tanto, dada la naturaleza del experimento, es lógico asumir una distribución Poisson con parámetro λk, que indica la intensidad (es decir, el número de usuarios esperado) del proceso en el intervalo deseado.

Del análisis exploratorio, podemos ver que **la intensidad no es constante durante el periodo de estudio de 24 hrs.** Por lo tanto, tenemos un **Proceso Poisson no homogéneo**:

$$X \sim \text{Poisson}(\lambda(t)),$$

para $t = 0, 1, 2, ...$

Ejemplo. Selección del modelo

En nuestro caso:

$$\lambda(t) = \begin{cases} 1.51 & \text{0 a 9 hrs} \\ 3.48 & \text{9 a 14 hrs y 16 a 21 hrs} \\ 2.96 & \text{14 a 16 hrs y 21 a media noche} \end{cases}$$

Ejemplo. Aplicación del modelo

Supongamos que queremos averiguar la probabilidad de que haya 2 conexiones en un intervalo de tiempo de 2 horas, entre las 5 y las 7 de la mañana. En este caso:

$$P(X_{(7)} - X_{(5)} = 2) = P(X_{(2)} = 2) = \frac{(2 \times 1.51)^2 e^{-(2\times 1.51)}}{2!} = 0.2225.$$

En R:

```
dpois(2,2*1.51)
```

Ejemplo. Aplicación del modelo

```
> ?dpois
>
Poisson                 package:stats                  R Documentation

The Poisson Distribution

Description:

     Density, distribution function, quantile function and random
     generation for the Poisson distribution with parameter lambda.
Usage:

     dpois(x, lambda, log = FALSE)
     ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
     qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
     rpois(n, lambda)

Arguments:
       x: vector of (non-negative integer) quantiles.
       q: vector of quantiles.
       p: vector of probabilities.
       n: number of random values to return.
  lambda: vector of (non-negative) means.
log, log.p: logical; if TRUE, probabilities p are given as log(p).
lower.tail: logical; if TRUE (default), probabilities are P[X <= x],
          otherwise, P[X > x].
```

Ejemplo. Aplicación del modelo

Algo más interesante para la empresa, es averiguar las probabilidades de usuarios conectados en las horas pico. Por ejemplo, la probabilidad de que haya al menos 5 usuarios conectados entre las 4 y 6 de la tarde.

$$\begin{aligned} P(X_{(18)} - X_{(16)} \geq 5) &= P(X_{(2)} \geq 5) \\ &= 1 - \sum_{x=0}^{4} \frac{(2\times 3.48)^x e^{-(2\times 3.48)}}{x!} \\ &= 0.823 \end{aligned}$$

En R:

```
1-ppois(4,3.48*2)
```

o

```
ppois(4,3.48*2,lower.tail=FALSE)
```

Ejemplo. Aplicación del modelo

En general, si tenemos un proceso Poisson no homogéneo, el número de eventos en el intervalo (s, t) se distribuye Poisson con parámetro

$$\alpha = \int_s^t \lambda(u)du.$$

Por ejemplo, si queremos saber cuál es la probabilidad de que haya 3 usuarios conectados entre las 8 y 10 de la noche:

$$P(X_{(22)} - X_{(20)} = 3) = \frac{\alpha^3 e^{-(\alpha)}}{3!},$$

donde

$$\alpha = \int_{20}^{21} 3.48du + \int_{21}^{22} 2.96du = 6.44.$$

Entonces

$$P(X_{(2)} = 3) = 0.072.$$

Ejemplo. Aplicación del modelo

Entonces

$$P(X_{(2)} = 3) = 0.072.$$

Observa que no se suman las probabilidades, sino las contribuciones de las intensidades en cada intervalo.

Del mismo modo, la probabilidad de que haya al menos 5 usuarios conectados entre las 8 y las 10:

$$P(X_{(22)} - X_{(20)} \geq 5) = 0.77.$$

Ejemplo. Aplicación del modelo

Ahora, calculemos la probabilidad de que haya al menos 5 usuarios conectados entre las 8 y 10 de la noche y entre las 5 y 7 de la mañana

$$P(X_{(22)} - X_{(20)} \geq 5 \text{ y } X_{(7)} - X_{(5)} \geq 5).$$

Como son intervalos disjuntos, las intensidades para ambos eventos son:

$$\alpha_{22-20} = 6.44 \text{ y } \alpha_{7-5} = 1.51 \times 2,$$

por lo tanto:

$$P(X_{(22)} - X_{(20)} \geq 5 \text{ y } X_{(7)} - X_{(5)} \geq 5) = 0.77 \times 0.188 = 0.145.$$

Ejemplo. Aplicación del modelo

Ahora, supon que queremos ajustar un modelo a la intensidad del primer grupo de conexiones observadas (0 a 9 hrs).

Probemos con un modelo lineal. Para los datos de promedios en ese intervalo de tiempo, obtenemos:

$$\lambda(t) = 1.466 + 0.01t$$

En R:

```
## las conexiones promedio de 0 a 9 hrs
y <- mmed[1:9]
## ajusta un modelo polinomial
mod <- lm(y ~ poly(1:9, degree=1, raw=TRUE))
```

Ejemplo. Aplicación del modelo

Ejemplo. Aplicación del modelo

Con este modelo para la intensidad, calculemos la probabilidad de que haya 2 usuarios conectados entre las 5 y las 7 de la mañana:

$$P(X_{(7)} - X_{(5)} = 2) = P(X_{(2)} = 2) = \frac{(\alpha)^2 e^{-\alpha}}{2!}$$

donde

$$\alpha = \int_5^7 (1.466 + 0.01t)dt = 3.052.$$

Entonces

$$P(X_{(7)} - X_{(5)} = 2) = P(X_{(2)} = 2) = \frac{(3.052)^2 e^{-3.052}}{2!} = 0.22,$$

que es prácticamente igual a la probabilidad obtenida anteriormente cuando asumiamos intensidad constante, como era de esperarse.

Ejemplo. Conclusiones

- Hemos analizado la información para averiguar cómo se comportan las conexiones de los usuarios durante el día completo.
- Usamos un modelo adecuado para el fenómeno que observamos, considerando el comportamiento de los usuarios durante distintos periodos de tiempo. Con esto, pudimos calcular probabilidades de usuarios conectados en distintos horarios del día, algo que puede ser de utilidad para la compañía.
- Con el modelo usado, podemos calcular los periodos donde hay mayor probabilidad de saturación, es decir, donde hay muchos usuarios conectados, y en base a esto, habilitar mayor número de líneas.

Ejemplo. Conclusiones

- Sin embargo, no sabemos aún cuántas conexiones están dejando de hacerse (intentos fallidos o usuarios rechazados) por la saturación del sistema (¿Cómo se les ocurre que podríamos averiguarlo?)
- Tampoco sabemos aún si la intensidad varía en ciertas temporadas del año
- Nos falta saber también cuál sería el número (óptimo) de líneas de conexión necesarias para aumentar la capacidad de servicio.

Referencias extras...

En el sitio del curso se incluyen dos referencias sobre la modelación del servicio en call centers.

La primera es una referencia ya clásica (**Telephone Call Centers Tutorial, Review, and Research Prospects**) y contiene un tutorial sobre modelos para la administración de la operación de un call center.

La segunda referencia (**Data-Driven Appointment-Scheduling Under Uncertainty The Case of an Infusion Unit in a Cancer Center**) incluye un enfoque moderno de un problema similar en un centro de cáncer.

Distribución Poisson

Nota Histórica

S. D. Poisson, en 1837 publicó un trabajo en el cual incluía el comportamiento límite de la distribución binomial:

$$\lim_{n\to\infty} \text{Bin}(x; n, p) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{con } p = \frac{\lambda}{n},\ \lambda \text{ fija}$$

Ver [10] págs 193-201. Durante 50 años no atrajo el interés de nadie hasta que L. von Bortkiewicz lo incluyó en una publicación suya, y en la que usaba la distribución de Poisson para modelar un problema del mundo real. Bortkiewicz fue pionero en el uso de distribuciones de probabilidad como modelos de datos, que es precisamente lo que tratamos de hacer en este curso. Establecer modelos que sirven para describir situaciones que ocurren en el mundo real. La distribución Poisson apareció en libros de Leyes, como un medio para estudiar el comportamiento de las demandas legales.

Modelos de variables continuas

(Texto, Sección 2.4)

Variables continuas

Recuerda que en el caso de variables aleatorias continuas el patrón de comportamiento no puede deducirse de la forma como se realiza (condiciones) el muestreo, como en el caso discreto, sino que se basa en gran medida en

- Los supuestos que se hagan sobre la variable (características físicas, etc.).
- El comportamiento manifestado a través del experimento e interpretado con la ayuda de histogramas, diagramas de caja, polígonos de frecuencia, pruebas de bondad de ajuste, Q-Q plots, etc.

Distribución Uniforme continua

Distribución uniforme

- Es uno de los modelos mas simples. Modela problemas prácticos en los que no hay ninguna "preferencia" en cuanto a las probabilidades que puede tomar la v.a. de interés, en el sentido de que les asigna la misma probabilidad a cualquier intervalo del mismo tamaño que sea considerado.
- Tiene un gran uso en simulación de números aleatorios de otras distribuciones, en estadística Bayesiana y en problemas donde no se requiere plantear un modelo complejo desde el inicio, para poder dar respuesta a preguntas complejas, siempre y cuando la solución sea aceptable.
- La definición de la misma es sencilla. Se define el intervalo de interés (debe ser finito) y se asigna la misma probabilidad, esto es, si el intervalo es (a, b), entonces el área debe ser uno, lo que nos forza a asignar una probabilidad de $\frac{1}{b-a}$ a todo el intervalo.

Distribución uniforme continua

Definición (Distribución uniforme continua)

Una v.a. X tiene una distribución uniforme continua, denotada con $X \sim Unif(a, b)$ si y solo si su función de densidad de probabilidad está dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & \textit{otra parte.} \end{cases}$$

Distribución uniforme continua

En las gráficas siguientes, $\theta_1 = a$ y $\theta_2 = b$.

Distribución uniforme continua

Media y Varianza:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx$$
$$= \left[\frac{1}{b-a}\frac{x^2}{2}\right]_a^b = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{1}{2}\frac{(b+a)(b-a)}{b-a}$$
$$= \frac{b+a}{2} \quad \text{(y por intuición).}$$

La media se puede obtener directamente de la gráfica: es el "punto de equilibrio" y es $\frac{a+b}{2}$.

Distribución uniforme continua

Media y Varianza:

$$
\begin{aligned}
V(X) =& E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\frac{b+a}{2}\right)^2 \\
=& \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{b+a}{2}\right)^2 \\
=& \frac{1}{b-a} \int_a^b x^2 dx - \left(\frac{b+a}{2}\right)^2 \\
=& \frac{1}{3} \frac{b^3 - a^3}{b-a} - \left(\frac{b+a}{2}\right)^2 \\
=& \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 \\
=& \frac{(a-b)^2}{12}.
\end{aligned}
$$

Distribución uniforme continua

Generatriz de momentos: Tenemos que,

$$\begin{aligned} M_X(t) =& E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_a^b e^{tx} \frac{1}{b-a} dx \\ =& \frac{1}{b-a} \int_a^b e^{tx} dx = \frac{1}{b-a} \frac{e^{tx}}{t} \Big|_a^b \\ =& \frac{e^{tb} - e^{ta}}{t(b-a)}, \quad t \neq 0, \end{aligned}$$

en el caso cuando $t = 0$, utilizamos la regla de L'Hopital:

$$\lim_{t \to 0} M_X(t) = \lim_{t \to 0} \frac{e^{tb} - e^{ta}}{t(b-a)} = \lim_{t \to 0} \frac{be^{tb} - ae^{ta}}{(b-a)} = \frac{b-a}{b-a} = 1.$$

Distribución uniforme continua

Por lo tanto,

$$M_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & t \neq 0 \\ \\ 1 & t = 0 \end{cases}$$

En este caso, la f.g.m. es mas compleja de lo que nos gusta! y por lo general no recurrimos a ella salvo en algunos problemas de caracter más teóricos...

Distribución uniforme continua

Función de distribución acumulada.
Se calculó en el propedéutico:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$
$$= \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$

Esta representación es muy empleada, como veremos mas adelante.

Distribución uniforme continua

Ejemplo

Una persona llega los días hábiles a una estación del metro entre las 9 y las 9:15 de la mañana, y logra llegar a tiempo a su trabajo (cuando el metro no se retrasa). Proponer una distribución de probabilidad factible para las v.a. $X =$ tiempo de llegada entre las 9 y las 9:15, y calcular la probabilidad de que la persona llegue entre las 9:01 y las 9:05, y entre las 9:07 y las 9:11 a la estación del metro.

Supuesto: *El momento de llegada puede darse en cualquier minuto dentro del intervalo inicialmente considerado.*

Distribución uniforme continua

La persona no se preocupa por llegar en los primeros minutos después de las 9:00, o los últimos, etc. Se puede hacer un muestreo de los tiempos de llegada de la persona y analizar los datos mediante, por ejemplo, un histograma, para corroborar nuestro supuesto.

Entonces, $X \sim Unif(0, 15)$

$$f(x) = \begin{cases} \frac{1}{15} & 0 < x < 15 \\ \\ 0 & \text{otra parte} \end{cases},$$

lo que implica que

$$P(1 < X < 5) = \int_1^5 \frac{1}{15} dx = \frac{4}{15}, \qquad P(7 < X < 11) = \int_7^{11} \frac{1}{15} dx = \frac{4}{15}.$$

Distribución uniforme continua

- Observa que las probabilidades dependen solamente del tamaño del intervalo, no de su "localización".
- Se ha dicho que la densidad uniforme sirve para generar números aleatorios de otras distribuciones. La generación de números aleatorios uniformes se puede hacer en casi todas las calculadoras de bolsillo y paquetes computacionales estadísticos. Los libros de simulación nos proveen con más de un algoritmo para su generación.
- Incluso uno puede generarlos mediante algún experimento como el de sacar números de una urna en forma aleatoria, que representen los dígitos de un número con n cifras con cierta cantidad de decimales, siempre y cuando la selección se haga aleatoriamente y que cada número tenga la misma probabilidad de salir (el muestreo es con reemplazo).

Distribución uniforme continua

Ejemplo

Se sabe que cierta v.a. X tiene una distribución $f_X(x) = \frac{1}{2}e^{-\frac{1}{2}x}$, $x > 0$. Calcula:

a) *Su distribución acumulada $F_X(x)$.*
b) *Defina la v.a. $Y = F_X(X)$ y calcula su distribución acumulada, es decir, $F_Y(y)$.*
c) *Obtén la función de densidad $f_Y(y)$ utilizando el hecho de que $f_Y(y) = \frac{dF_Y(y)}{dy}$.*

Los subíndices se utilizan para distinguir a las funciones respectivas y para recalcar la distribución que se está manejando.

Distribución uniforme continua

a) La función de distribución acumulada ya fue calculada en el capítulo de conceptos básicos (con $\beta = 2$),

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\frac{1}{2}x} & x > 0 \end{cases}.$$

Distribución uniforme continua

b) Definimos $Y = F_X(X) = 1 - e^{-\frac{1}{2}X}$. Observa que $0 < y < 1$, ya que $0 < F_X(x) < 1$, ya que es una probabilidad. Entonces,

$$\begin{aligned} F_Y(y) =& P(Y \leq y) = P(1 - e^{-\frac{1}{2}X} \leq y) = P(X \leq -2\log(1-y)) \\ =& F_X(x = -2\log(1-y)) = 1 - e^{-\frac{1}{2}(-2\log(1-y))} \\ =& 1 - e^{\log(1-y)} = 1 - (1-y) \\ =& y \end{aligned}$$

c) Derivando obtenemos que

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}(y) = 1, \qquad \text{con } 0 < y < 1.$$

Distribución uniforme continua

Observa en el ejemplo anterior que Y tiene una f.d.p. $Unif(0,1)$, esto es,

$$f(y) = \begin{cases} 1 & 0 < y < 1 \\ \\ 0 & \text{otra parte} \end{cases}.$$

Entonces, si se pudieran generar números aleatorios $Y \sim Unif(0,1)$, se podría utilizar $Y = 1 - e^{-\frac{1}{2}X}$ para generar números aleatorios X simplemente despejándola.

Distribución uniforme continua

Ejercicio

i) *Crea una columna con 100 valores de una Unif*$(0,1)$ *en el software de tu preferencia.*

ii) *Construye otra columna con la fórmula*[a]

$$x = -2\log(1-y)$$

iii) *Construye el histograma de esta nueva columna y concluye.*

[a] $y = 1 - e^{-\frac{1}{2}X} \Rightarrow 1 - y = e^{-\frac{1}{2}X} \Rightarrow -\frac{1}{2}X = \log(1-y) \Rightarrow x = -2\log(1-y)$

Distribución uniforme continua

De hecho, este ejemplo es un caso particular de un teorema importante que se menciona a continuación.

Teorema (Teorema de la transformación integral de probabilidad)

Sea X una v.a. continua con función de distribución acumulada $F_X(x)$, y definamos la v.a. Y como $Y = F_X(X)$. Entonces Y tiene una distribución $Unif(0,1)$, esto es, $F_Y(y) = y$.

Distribución uniforme continua

Demostración

Tenemos que, para $0 < y < 1$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y, \end{aligned}$$

de donde vemos que la distribución acumulada de Y *corresponde a la distribución acumulada de una v.a.* $Unif(0,1)$, *como se quería demostrar.*

Este resultado es útil cuando $F(x)$ es una función de la cual podemos despejar a x, algo que no siempre es posible, como veremos para muchos de los modelos del catálogo. Sin embargo, modificaciones de esta idea base siguen ayudando en simulación de variables aleatorias.

El modelo Normal o Gaussiano

El modelo Normal o Gaussiano

Pues bien, las cosas comenzaron allá por los 1700. La historia alrededor de la distribución normal podría titularse: "Las aventuras de la distribución normal. Surgimiento, fama y abuso." Sin exagerar, se podría decir que esta distribución es *la piedra angular de la estadística*, y tendremos oportunidad de comprobarlo. A manera de introducción veremos un poco de esa fascinante historia.

Aparte de la aproximación límite de Poisson a la binomial, se había derivado otra en una publicación de 1718 a manos de DeMöivre, y que posteriormente retomó y generalizó Laplace en 1812.

El modelo Normal o Gaussiano

Teorema (DeMöivre-Laplace)

Sea X una v.a. binomial definida en n pruebas independientes, cada una con probabilidad de éxito p. Para cualquiera par de números c y d,

$$\lim_{n\to\infty} P\left(c < \frac{X - np}{\sqrt{npq}} < d\right) = \frac{1}{\sqrt{2\pi}} \int_c^d e^{-\frac{1}{2}z^2} dz.$$

Como puedes ver, se encontró que una probabilidad binomial, en un caso límite, puede ser calculada como el área bajo la curva $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$.

Entonces, en sí misma, ésta curva debería representar una f.d.p. En efecto, la función así definida cumple con todas las propiedades de una f.d.p. para $-\infty < z < \infty$.

El modelo Normal o Gaussiano

La función se hizo tan popular en la descripción de una gran variedad de fenómenos que se acostumbró llamarla "la distribución normal". Es tan usada que incluso en ocasiones *pecamos* de querer ajustar la función a fenómenos que no tienen ese comportamiento (**warning**).

Una de las primeras aplicaciones es debida a K.F. Gauss publicada en 1809, quien la utilizó para modelar los errores observados en la astronomía. Posteriormente fue L.A. Quetelet el que "arrebató" la "exclusividad" de su uso a los físicos, ya que mostró su aplicación en el campo de la sociología y fenómenos antropológicos.

El modelo Normal o Gaussiano

Definición

Una v.a. Z tiene una distribución **Normal estándar** *si su f.d.p. está dada por*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \qquad -\infty < z < \infty.$$

El modelo Normal o Gaussiano

De entre sus muchas propiedades podemos señalar:

- Es simétrica alrededor de un eje vertical que pasa por $z = 0$.
- Su mediana coincide con su media, es decir, $m = \mu = 0$.
- Su varianza es 1 y tiene puntos de inflexión en ± 1.
- Su valor máximo ocurre en $z = 0$ y es $\frac{1}{\sqrt{2\pi}}$.
- $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1.$

12 videos sobre como calcular la integral:

https://www.youtube.com/playlist?list=PLJb1qAQIrmmCgLyHWMXGZnioRHLqOk2bW

El modelo Normal o Gaussiano

De entre sus muchas propiedades podemos señalar:

- La integral $\int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$, que representa $P(a \leq Z \leq b)$, con a y b números reales, no puede evaluarse analíticamente; esto es, no tiene una antiderivada, solo puede hacerse mediante métodos numéricos. Ya en épocas de DeMöivre se calculaba de esa manera. Nosotros usaremos software que tiene implementado tales métodos.
- La distribución acumulada es $F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$ y se denota con $\Phi(z)$.

El modelo Normal o Gaussiano

Media y Varianza.

El valor esperado de Z es

$$\mu = E(Z) = \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 0,$$

ya que el integrando es una función impar (si lo deseas puedes integrar mediante un cambio de variable).

Su varianza es

$$\sigma^2 = V(Z) = E\left((z-\mu)^2\right) = E(z^2) = \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1,$$

la cual se obtiene integrando por partes dos veces.

El modelo Normal o Gaussiano

Entonces, la media y la varianza son cero y uno respectívamente, esto se escribe abreviadamente como

$$Z \sim N(0,1);$$

en palabras, Z tiene una distribución normal estándar con media cero y varianza uno. El uso de la letra Z para una v.a. normal estándar se ha generalizado en la literatura.

Función generatriz de momentos:

Tenemos que

$$M_Z(t) = E(e^{tz}) = \int_{-\infty}^{\infty} e^{tz} f(z)dz = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

El modelo Normal o Gaussiano

Este es un buen momento para hacerte una recomendación respecto al cálculo de integrales que se ven complicadas: trata de **identificar** la integral que estés resolviendo con alguna que hayas manejado anteriormente, haciendo las analogías necesarias puedes obtener su valor.

Por ejemplo, teníamos que

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2+tz} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2tz)} dz,$$

El modelo Normal o Gaussiano

completando el cuadrado y separando,

$$
\begin{aligned}
M_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(z^2+2tz+(\frac{2t}{2})^2-(\frac{2t}{2})^2\right)} dz \\
&= e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z+t)^2} dz \\
&= e^{\frac{1}{2}t^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \\
&= e^{\frac{1}{2}t^2}(1) = e^{\frac{1}{2}t^2},
\end{aligned}
$$

donde se usa el hecho de que la última integral es uno, ya que es como si U fuera una v.a. con distribución normal estándar, y al integrarla desde $-\infty$ a ∞ el resultado es 1.

El modelo Normal o Gaussiano

Variable aleatoria normal con cualquier media y varianza.

Cada problema tiene su distribución propia. En particular, tiene su asociada una media μ y una varianza σ^2 propias.

Tenemos que la medida de localización μ y la medida de dispersión σ sirven para trasladar o escalar las funciones mediante alguna transformación (¿recuerdas la expresión $\frac{X-\mu}{\sqrt{V(X)}}$?).

En la siguiente transformación haremos uso de esas ideas con el fin de generalizar la distribución normal estándar y no estar sujetos a problemas en que la media sea cero y la varianza uno.

El modelo Normal o Gaussiano

Queremos trasladar la distribución normal estándar de forma tal que su media no sea cero, y escalarla de tal forma que su varianza no sea uno. Consideremos la siguiente transformación:

$$X = Z\sigma + \mu.$$

El valor esperado de la v.a. X es

$$E(X) = E(Z\sigma + \mu) = \sigma E(Z) + \mu = 0 + \mu = \mu,$$

y su varianza

$$V(X) = V(Z\sigma + \mu) = \sigma^2 V(Z) = \sigma^2(1) = \sigma^2.$$

El modelo Normal o Gaussiano

Entonces, la media y la varianza de $X = Z\sigma + \mu$ son, respectívamente, μ y σ^2. Dado que solo se está recorriendo y escalando la función original, la gráfica de esta nueva varible tendrá la misma forma (campana), solo que no estará centrada en cero, y tal vez esté mas "chata" o mas "espigada".

El modelo Normal o Gaussiano

El modelo Normal o Gaussiano

El modelo Normal o Gaussiano

Entonces, cabe la pregunta: si para calcular probabilidades con la nomal estándar teníamos que evaluar la integral respectiva numéricamente, ¿cómo haremos con la variable $X = Z\sigma + \mu$?

Observa que se puede despejar Z de esta relación: $\frac{X-\mu}{\sigma} = Z$. Entonces

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\ &= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right). \end{aligned}$$

En desigualdades como la anterior, si realizamos la misma operación en cada término a, X y b, la desigualdad resultante es totalmente equivalente a la original (respetando el sentido de la desigualdad si se multiplica por un número positivo y cambiando el sentido si se multiplica por un número negativo).

El modelo Normal o Gaussiano

Entonces, calcuar una probabilidad para X se reduce a calcular una probabilidad para Z:

$$
\begin{aligned}
P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= P\left(Z \leq \frac{b-\mu}{\sigma}\right) - P\left(Z \leq \frac{a-\mu}{\sigma}\right) \\
&= F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).
\end{aligned}
$$

El modelo Normal o Gaussiano

Podemos expresar lo anterior de la siguiente manera: si X se distribuye como una normal estándar escalada y recorrida, digamos "acampanada", y deseamos calcular probabilidades para X, utilizaremos la distribución normal estándar, simplemente restando μ y dividiendo por σ, esto es, "relocalizando" y escalando la v.a. X adecuadamente, esto es, **estandarizando**.

De hecho, la distribución de $X = Z\sigma + \mu$ es

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

con media μ y varianza σ^2, y se dice que se distribuye normal con parámetros μ y σ.

Notación: $X \sim N(\mu, \sigma^2)$.

El modelo Normal o Gaussiano

La generatriz de momentos es

$$\begin{aligned} M_X(t) = M_{Z\sigma+\mu}(t) &= e^{\mu t} M_Z(t\sigma) \\ &= e^{\mu t} e^{\frac{1}{2}(t\sigma)^2} \\ &= e^{\mu t + \frac{1}{2} t^2 \sigma^2}, \end{aligned}$$

donde usamos la propiedad demostrada en conceptos básicos.

$$M_{\frac{X-a}{b}}(t) = M_{\frac{X}{b} - \frac{a}{b}}(t) = e^{-\frac{a}{b}} M_X\left(\frac{t}{b}\right),$$

con $\frac{1}{b} = \sigma$, $-\frac{a}{b} = \mu$.

El modelo Normal o Gaussiano

Observación: se pudieron haber usado otras letras en $X = Z\sigma + \mu$, por ejemplo, $X = Z\alpha + \delta$, el resultado para la media y la varianza hubiera sido $E(X) = \delta$, $V(X) = \alpha^2$, pero como la notación usual es μ y σ^2, se pensó en usar sus símbolos desde el inicio (porque ya se sabía la respuesta). Cabe comentar que la notación $N(\mu, \sigma^2)$ puede variar, algunos autores usan $N(\mu, \sigma)$.

El modelo Normal o Gaussiano

Ejemplo

Se tiene un proceso industrial donde se producen baleros y se toma la medición de su diámetro. La especificación del producto toma como bueno un balero con diámetro en el rango 3.0 ± 0.01 *cm. Si la variable diámetro se puede considerar normal con media* $\mu = 3$ *y* $\sigma = 0.005$.

a) *¿Qué proporción de baleros caen dentro de especificación?*
b) *Se realiza una investigación en la que es necesario revisar el 5 % de los producidos a la izquierda de 3.01. ¿A partir de qué valor de la v.a.* X *se efectuaría esta revisión?*

El modelo Normal o Gaussiano

Sea la v.a. X el diámetro de un balero producido. Se sabe que $X \sim N(3, 0.005^2)$, entonces nos están preguntando el porcentaje de baleros que caen dentro de los límites de especificación; esto es,

$$\begin{aligned}
P(3 - 0.01 < X < 3 + 0.01) &= P(2.99 < X < 3.01) \\
&= P\left(\frac{2.99 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{3.01 - \mu}{\sigma}\right) \\
&= P\left(\frac{2.99 - 3}{0.005} < Z < \frac{3.01 - 3}{0.005}\right) \\
&= P\left(-2 < Z < 2\right) \\
&= 0.9544.
\end{aligned}$$

Por lo tanto la proporción de baleros que caen dentro de las especificaciones es 95.44 %.

El modelo Normal o Gaussiano

Observa que coincide con la regla empírica con $k = 2$.

En la segunda pregunta se pide un valor x_p tal que

$$P(X > x_p) = 0.05 + P(X > 3.01) = 0.05 + 0.0228,$$

entonces el 5 % de los datos a la izquierda del límite superior corresponde a un valor de X tal que

$$P(X < x_p) = 1 - (0.05 + 0.0228) = 0.9272.$$

Si estandarizamos,

$$P(X < x_p) = P(Z < z_p) = 0.9272,$$

de tablas obtenemos que $z_p = 1.455$, de donde $x_p = z_p\sigma + \mu$ $= 1.455(0.005) + 3 = 3.0072$. La cantidad $z_p(x_p)$ es llamado percentil y ésta forma de encontrarlos se usa exhaustivamente (dada la probabilidad, hallar los valores de la variable).

El modelo Normal o Gaussiano

Ejercicio

Se han realizado ciertas pruebas de resistencia en ladrillos obteniéndose las mediciones que a continuación se muestran, agrupadas en una tabla de frecuencias.

Interv.	*De*	*Hasta*	*Frecuencia*	*Frec. relativa (%)*
1	28.70	32.65	5	5.56
2	32.65	36.60	6	6.67
3	36.60	40.55	11	12.22
4	40.55	44.50	17	18.89
5	44.50	48.45	19	21.11
6	48.45	52.40	19	21.11
7	52.40	56.35	7	7.78
8	56.35	60.30	2	2.22
9	60.30	64.25	3	3.33
10	64.25	68.20	1	1.11

El modelo Normal o Gaussiano

Ejercicio (continúa...)

a) *Traza el histograma.*

b) *Calcula la probabilidad de cada intervalo de clase de la tabla de frecuencias, asumiendo que las resistencias siguen una distribución normal con media 45.47 y varianza 58.19.*

c) *Compara las frecuencias relativas con las probabilidades bajo normalidad ¿Qué se puede concluir? Esta es la idea base de una prueba de Bondad de Ajuste conocida como χ^2 de Pearson. Si la media y varianza de la normal no se conocen de antemano, podemos usar los valores muestrales correspondientes como una aproximación de los mismos. A esto lo llamamos estimación de parámetros y ya hablaremos más adelante de las cualidades de los estimadores.*

El modelo Normal o Gaussiano

La normal truncada.

Ejemplo

Supón que se fabrica cierto tipo de perno, su longitud Y es una v.a. con distribución $N(2.2, 0.01)$. De un gran lote de estos pernos se saca un nuevo lote, descartando todos aquellos para los cuales $Y > 2$. Obtén la distribución de probabilidad de X, la v.a. que representa el largo de los pernos en el nuevo lote.

El modelo Normal o Gaussiano

Si encontramos $F(x)$, su distribución acumulada, podremos obtener $f(x)$ derivando. Se podría pensar a X como igual a Y excepto porque la probabilidad de que X sea mayor a 2 es cero:

$$
\begin{aligned}
F(x) &= P(X \leq x) \\
&= \begin{cases} P(Y \leq x | Y \leq 2) = 1 & \text{si } x > 2 \\ \\ P(Y \leq x)/P(Y \leq 2) & \text{si } x \leq 2 \end{cases}.
\end{aligned}
$$

El modelo Normal o Gaussiano

- El primer renglón se considera cuando el valor de x está a la derecha de 2, en tal caso se sabe que Y no puede ser mayor que 2, esto es, $Y \leq 2$ en el nuevo lote. Por lo tanto, $P(X \leq 2)$ debe calcularse como la probabilidad de que Y sea menor que x dado que $Y \leq 2$: $P(Y \leq x | Y \leq 2)$, pero esta probabilidad es 1, ya que en el nuevo lote, todos miden menos de 2.
- El segundo renglón se puede visualizar como una "frecuencia relativa", casos posibles entre casos totales, para el caso contínuo.

El modelo Normal o Gaussiano

Derivando:

$$
\begin{aligned}
f(x) &= \frac{d}{dx}P(X \leq x) \\
&= \begin{cases} 0 & \text{si } x > 2 \\ \frac{d}{dx}P(Y \leq x)/P(Y \leq 2) & \text{si } x \leq 2 \end{cases} \\
&= \begin{cases} 0 & \text{si } x > 2 \\ \frac{d}{dx}\left(\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}(.1)} e^{-\frac{1}{2}\left(\frac{y-2.2}{.1}\right)^2} dy\right)/P(Y \leq 2) & \text{si } x \leq 2 \end{cases} \\
&= \begin{cases} 0 & \text{si } x > 2 \\ \left(\frac{1}{\sqrt{2\pi}(.1)} e^{-\frac{1}{2}\left(\frac{x-2.2}{.1}\right)^2}\right)/\Phi(-2) & \text{si } x \leq 2 \end{cases},
\end{aligned}
$$

El modelo Normal o Gaussiano

donde usamos el Teorema Fundamental del cálculo: $\frac{d}{dx}\int_a^x f(u)du = f(x)$, y la función de probabilidad acumulada de la normal estándar $\Phi(z) = P(Y \leq 2) = F_Y(2) = \Phi\left(\frac{2-\mu}{\sigma}\right) = \Phi\left(\frac{2-2.2}{0.1}\right) = \Phi(-2)$.

La media y la varianza de esta nueva v.a. que se genera al descartar ciertos valores de la v.a. original, depende de si el truncamiento es a la derecha, a la izquierda o en ambas partes. En nuestro ejemplo, el truncamiento se hizo a la derecha.

El modelo Normal o Gaussiano

Queremos recalcar el hecho del cambio de estas cantidades. Supongamos que en un proceso, un producto tiene una distribución normal con varianza σ^2. Al rechazar sistemáticamente productos a partir de cierto valor, la varianza de los productos seleccionados no coincidirá con la correspondiente al proceso, sino que será menor. Intuitivamente es así, ya que se está disminuyendo el rango de variación: la nueva variable en nuestro ejemplo va de $-\infty$ a 2, y antes iba de $-\infty$ a ∞.

Si un comprador realiza una investigación acerca de la variabilidad del producto recibido, obtendrá una idea falsa de la misma en caso de no estar consciente del truncamiento, esto es, debido al truncamiento la variabilidad medida sobre el producto recibido, no corresponde a la verdadera varianza del proceso. Por esa razón, muchas veces preferimos saber cuál es la capacidad del proceso (si es que puede ser obtenida). Para más información de la normal truncada, ver las lecturas de apoyo.

El modelo Normal o Gaussiano

En el documento **Mean and Variance of Truncated Normal Distributions**, subido en el site del curso, se presenta el cálculo de la media y la varianza para la normal truncada.

Implementación de Q-Q Plots

Q-Q plots

Las gráficas Quantile-Quantile (QQ) son útiles para comparar funciones de distribución.

Definición

Sea X una v.a. con función de distribución (fd) F. La fd inversa o función de cuantiles es

$$F^{-1}(x) = \text{ínf}\left\{x : F(x) > q\right\},$$

para que $q \in [0,1]$. Si F es estrictamente creciente, entonces el $p-$ésimo cuantil es el único número real x tal que

$$F(x) = p,$$

o equivalentemente

$$x_p = F^{-1}(p).$$

Q-Q plots

Sea X una v.a. continua con fd F_X y definamos $Y = aX + b$, con $a, b \in \mathbb{R}$. Denotemos con F_X^{-1} la función de cuantiles de X. Observemos lo siguiente

$$\begin{aligned} F_Y(y) = p \quad &\Leftrightarrow \quad P(aX + b \leq y) = P(X \leq (y-b)/a) = p \\ &\Leftrightarrow \quad F_X((y-b)/a) = p, \end{aligned}$$

de donde obtenemos

$$F_Y^{-1}(p) = aF_X^{-1}(p) + b.$$

Función de distribución empírica

Un primer problema de inferencia consiste en la estimación no paramétrica de la función de distribución empírica (fde) de F. Sea $X_1, \ldots, X_n \sim F$ una muestra independiente e idénticamente distribuida (iid), donde F es una fd sobre la recta real. Estimaremos F con la fde que se define a continuación.

Definición

La fde $\hat{F}_n$ es la función de distribución acumulada que asigna masa $1/n$ en cada punto X_i. Formalmente,

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} 1_{\{X_i \leq x\}}}{n},$$

donde

$$1_{\{X_i \leq x\}} = \begin{cases} 1 & si\ X_i \leq x \\ 0 & si\ X_i > x \end{cases}$$

Función de distribución empírica

Teorema

En cualquier valor fijo x,

$$\begin{aligned} E(\hat{F}_n(x)) &= F(x) \\ Var(\hat{F}_n(x)) &= \frac{F(x)(1-F(x))}{n} \to 0 \\ \hat{F}_n(x) &\xrightarrow{P} F(x). \end{aligned}$$

Teorema (Glivenko-Cantelli)

Sean $X_1, \ldots, X_n \sim F$. Entonces

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{P} 0$$

Función de distribución empírica

Las gráficas Q-Q son una herramienta visual extremadamente útil para establecer cualitativamente el ajuste de un conjunto de datos a una distribución teórica.

Supongamos que la hipótesis de que X sigue cierta distribución F es de interés. Dada una muestra $X_1, \ldots X_n$, y usando el teorema de Glivenko-Cantelli, tenemos un buen estimador de la distribución de estos datos:

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} 1_{\{X_i \leq x\}}}{n} \xrightarrow{P} F(x).$$

Función de distribución empírica

Entonces, graficando

$$F(x) \text{ vs } \hat{F}_n(x)$$

para x apropiados, podremos determinar visualmente la validez de la hipótesis.

Sea $x_{(i)}$ el $i-$ésimo estadístico de orden, entonces

$$\hat{F}_n(x_{(i)}) = \frac{\sum_{i=1}^{n} 1_{\{x_i \leq x_{(i)}\}}}{n} = \frac{i}{n},$$

Función de distribución empírica

podemos intentar graficar el PP plot

$$F(x_{(i)}) \text{ vs } \frac{i}{n},$$

o el QQ plot

$$x_{(i)} \text{ vs } F^{-1}\left(\frac{i}{n}\right) = q_i, \quad i = 1, \ldots, n.$$

Sin embargo, esto no es del todo correcto. ¿Por qué?
Operativamente, debemos graficar

$$x_{(i)} \text{ vs } F^{-1}\left(\frac{i}{n+1}\right), \quad i = 1, \ldots, n.$$

Función de distribución empírica

x_i	rango$= i$	$p_i = \frac{i}{n+1}$	$z_{p_i} = q_i$ tal que $P(Z \leq z_{p_i}) = p_i$
31	1	0.0625	-1.53412
32	2	0.125	-1.15035
33.2	3	0.1875	-0.88715
34.5	4	0.25	-0.67449
35	5	0.3125	-0.48878
35.5	6	0.375	-0.31864
36.5	7	0.4375	-0.15731
37.2	8	0.5	0
37.5	9	0.5625	0.157311
37.8	10	0.625	0.318639
38	11	0.6875	0.488776
38.3	12	0.75	0.67449
38.5	13	0.8125	0.887147
39	14	0.875	1.150349
39.5	15	0.9375	1.534121

Función de distribución empírica

Función de distribución empírica

En el caso del ajuste de una Normal, no se grafica directamente el percentil sino su valor estandarizado, por lo que solo interesa ver una recta aunque no sea de 45º -así evitamos el cálculo de parámetros.

$$(\overline{x}, s^2) \to (\mu, \sigma^2)$$

$$z_{p_i} = \frac{y_i - \overline{x}}{s} \quad \text{vs.} \quad x_i$$

El modelo exponencial y Gamma

El modelo exponencial

El modelo exponencial. El modelo exponencial es muy común en la descripción del comportamiento de tiempos de espera, crecimiento de poblaciones, tiempos de vida, etcétera.

Definición

Sea X una v.a. continua. Decimos que X se distribuye de acuerdo a una ley exponencial con parámetro θ si su función de densidad es

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

Notación: $X \sim exp(\theta)$.

Observa que antes habíamos usado β en lugar de θ. ¡En gustos se rompen géneros!

El modelo exponencial

El modelo exponencial

Ahora calculemos media, varianza y función generatriz de momentos solo por completez de la presentación.

$$
\begin{aligned}
E(X) &= \int_0^\infty x\frac{1}{\theta}e^{-x/\theta}dx; \\
&= -xe^{-x/\theta}|_0^\infty - \int_0^\infty \left(-e^{-x/\theta}\right)dx \\
&= \int_0^\infty e^{-x/\theta}dx \\
&= -\theta e^{-x/\theta}\Big|_0^\infty \\
&= \theta
\end{aligned}
$$

$$
\begin{aligned}
u &= x & dv &= \tfrac{1}{\theta}e^{-x/\theta}dx \\
du &= dx & v &= -e^{-x/\theta}
\end{aligned}
$$

$\therefore E(X) = \theta$

El modelo exponencial

$$
\begin{aligned}
E(X^2) &= \int_0^\infty x^2 \frac{1}{\theta} e^{-x/\theta} dx; \qquad u = x^2 \quad dv = \tfrac{1}{\theta} e^{-x/\theta} dx \\
& \qquad\qquad\qquad\qquad\qquad\qquad du = 2xdx \quad v = -e^{-x/\theta} \\
&= -x^2 e^{-x/\theta}|_0^\infty - \int_0^\infty -e^{-x/\theta}(2x)dx \\
&= 2\int_0^\infty x e^{-x/\theta} dx; \qquad u = 2x \quad dv = e^{-x/\theta} dx \\
& \qquad\qquad\qquad\qquad\qquad\qquad du = 2xdx \quad v = -\theta e^{-x/\theta} \\
&= -2x\theta e^{-x/\theta}|_0^\infty - \int_0^\infty -\theta e^{-x/\theta} 2dx \\
&= \int_0^\infty 2\theta e^{-x/\theta} dx \\
&= 2\theta \left(-\theta e^{-x/\theta}\right)\Big|_0^\infty = 2\theta^2
\end{aligned}
$$

$$V(X) = E(X^2) - (E(X))^2 = 2\theta^2 - \theta^2 = \theta^2$$

$\therefore V(X) = \theta^2.$

El modelo exponencial

$$
\begin{aligned}
M_x(t) &= E(e^{tx}) = \int_0^\infty e^{tx}\frac{1}{\theta}e^{-x/\theta}dx \\
&= \frac{1}{\theta}\int_0^\infty e^{tx-\frac{x}{\theta}}dx = \frac{1}{\theta}\int_0^\infty e^{x\left(t-\frac{1}{\theta}\right)}dx \\
&= \frac{1}{\theta}\frac{e^{\left(t-\frac{1}{\theta}\right)x}}{t-\frac{1}{\theta}}\Bigg|_0^\infty = -\frac{1}{\theta}\frac{1}{t-\frac{1}{\theta}} \\
&= -\frac{1}{\theta\left(t-\frac{1}{\theta}\right)} \\
& \quad t-\frac{1}{\theta} < 0 \\
&= \frac{1}{1-\theta t} = (1-\theta t)^{-1}
\end{aligned}
$$

$\therefore M_x(t) = (1-\theta t)^{-1}$ para $t < \frac{1}{\theta}$.

El modelo exponencial

Notas:

- La variable aleatoria exponencial está ligada fuertemente con la v.a. Poisson. Cuando la v.a. Poisson cuenta el número de ocurrencias de un cierto evento en un periodo de tiempo fijo, el tiempo que pasa entre una ocurrencia y la siguiente es una v.a. exponencial.
  Así, si consideramos un proceso Poisson (λ) en cierto intervalo dado, y cronometramos el tiempo entre éxitos consecutivos, podemos observar que:

El modelo exponencial

En el proceso Poisson los éxitos se presentan aleatoriamente en el intervalo bajo estudio

El modelo exponencial

Sea T el tiempo hasta observar el siguiente éxito. Notemos que

$$\begin{aligned}
P(T > t) =& P(\text{no se ha observado ningún éxito en el intervalo } (0, t)) \\
=& P(X = 0 | X \sim P(\lambda t)) \text{ la media del proceso Poisson se ajusta al intervalo } (0, t) \\
=& \frac{e^{-\lambda t}(\lambda t)^0}{0!} \\
=& e^{-\lambda t}.
\end{aligned}$$

De aquí que $P(T \leq t) = 1 - e^{-\lambda t}$. Esto es, $F_T(t) = 1 - e^{-\lambda t}$, y por lo tanto, $f_T(t) = \frac{d}{dt} F_T(t) = \lambda e^{-\lambda t}$.

El modelo exponencial

De aquí concluimos que

$$T \sim Exp(\theta = 1/\lambda).$$

Al revés también es válida, si el tiempo entre ocurrencias es exponencial entonces el número de ocurrencias en un tiempo fijo es Poisson.

El modelo exponencial

Al igual que con la distribución geométrica, el modelo exponencial es empleado cuando el fenómeno bajo estudio es estable, esto es, cuando los eventos se presentan por azar y no por causas sistemáticas.

Es decir, si modelamos los tiempos de vida de una bateria LTalgo, esperaríamos que las fallas no se presenten siguiendo un patrón fijo; esto es, que el proceso de donde vienen esté trabajando bajo condiciones homogéneas en la producción de todas ellas. Entonces, deberá verificarse que esta variable exponencial tampoco "tiene memoria".

El modelo exponencial

En efecto, denotemos con

$A =$ el evento de que *la batería dure mas de* t_1+t_2 *meses, dado que sabemos que ya ha funcionado por* t_1 *meses.*

Entonces,

$$
\begin{aligned}
P(A) =& P(X > t_1 + t_2 | X > t_1) = \frac{P(X > t_1 + t_2 \ y \ X > t_1)}{P(X > t_1)} \\
=& \frac{P(X > t_1 + t_2)}{P(X > t_1)} = \frac{1 - F_X(t_1 + t_2)}{1 - F_X(t_1)} = \frac{1 - \left(1 - e^{-\left(\frac{t_1+t_2}{\theta}\right)}\right)}{1 - \left(1 - e^{-\left(\frac{t_1}{\theta}\right)}\right)} \\
=& e^{-\left(\frac{t_2}{\theta}\right)} = 1 - \left(1 - e^{-\left(\frac{t_2}{\theta}\right)}\right) = 1 - F_X(t_2) \\
=& P(X > t_2) \\
=& P(\textit{la batería dure más de } t_2 \textit{ meses}).
\end{aligned}
$$

El modelo exponencial

- Bajo este esquema, a $1 - F_X(t) = P(X > t) = R(t)$ se le conoce como la **confiabilidad** de X.

  Notemos que $R(0) = 1$ y $R(\infty) = 0$. Mide "lo que queda de vida al tiempo t".

  La **función de riesgo** (hazard function) o **tasa de fallo** se define como
  $$h(t) = \frac{f(t)}{R(t)},$$
  y cuando se considera sobre un intervalo de tiempo muy pequeño, se le puede interpretar como la **probabilidad instantánea de fallo**.

El modelo exponencial

Nota que $h(t)$ puede tomar valores mayores a 1, es decir, $h(t)$ no es una probabilidad. En efecto,

El modelo exponencial

Sea

$$\Delta(t) = (t, t + dt).$$

Entonces,

$$f(t)|\Delta(t)| \approx P(T \in (t, t + dt)).$$

Ahora, como

$$h(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)},$$

se sigue que

$$h(t)|\Delta(t)| \approx \frac{P(T \in (t, t + dt))}{P(T > t)} = P(T \in (t, t + dt)|T > t),$$

que es conocida como la tasa instantánea de falla.

El modelo exponencial

$h(t)\Delta(t) \approx$ la probabilidad de que se presente una "falla" al tiempo $t + dt$ dado que al tiempo t no ha habido fallas.

En el caso particular del modelo exponencial,

$$h(t) = \frac{1}{\theta} \frac{e^{-\frac{t}{\theta}}}{e^{-\frac{t}{\theta}}} = \frac{1}{\theta}$$

que es una constante.

Ver gráficas sobre el comportamiento esperado de funciones de riesgo.

El modelo exponencial

Figura: Gráfica de "tina de baño". (a) etapa de ajuste inicial o "mortalidad infantil". (b) etapa constante. (c) etapa de desgaste o envejecimiento.

El modelo exponencial

Si uno piensa en términos de la vida de seres humanos, es claro el comportamento de la tasa de riesgo.

La distribución exponencial es válida para modelar la parte en donde se espera que si una falla se presenta es por azar o "accidente", la parte plana de la gráfica anterior.

El modelo exponencial

Ejemplo

Una larga experiencia con abanicos de cierta marca usados con motores diesel ha sugerido que el modelo exponencial es suficientemente bueno para describir el tiempo a la falla de dichos abanicos. También se ha visto que el tiempo medio a la falla es de 25000 horas.

a) *Para un abanico seleccionado al azar, ¿cuál es la probabilidad de que dure entre 20000 y 30000 horas?*
b) *Dado que se sabe que ha durado 10000 horas, ¿cuál es la probabilidad de que dure mas de 30000?*
c) *¿Cuál es la probabilidad de que el tiempo de vida del abanico exceda el valor medio en más de una desviación estándar? ¿En más de 2 desviaciones estándar?*

El modelo exponencial

Solución: Sea X el tiempo a la falla del abanico, $x > 0$. $E(X) = \theta = 25$ (en miles por hora).

a)

$$P(20 < X < 30) = \int_{20}^{30} \frac{1}{25} e^{-x/25} dx = F(39) - F(20)$$
$$= 1 - e^{-30/25} - (1 - e^{-20/25}) = 0.148$$

b) $P(X > 30 | X > 10) = P(X > 20) = e^{-20/25} = 0.449$

El modelo exponencial

c) Sabemos que $E(X) = 25$ y $\sqrt{V(X)} = \sqrt{\theta^2} = 25$. Sea el evento

$$A = \{\text{el tiempo de vida del abanico excede el valor medio en más de una desviación estándar}\}.$$

Entonces

$$P(A) = P(X > 50) = e^{-50/25} = 0.135,$$

y además, para el evento

$$B = \{\text{el tiempo de vida del abanico excede el valor medio en más de dos desviaciones estándar}\}.$$

$$P(B) = P(X > 75) = e^{-75/25} = 0.05.$$

El modelo exponencial

Ejercicio

Consideremos un sistema formado por 5 componentes idénticos conectados en serie tal como se muestra a continución:

Tan pronto como un componente falla, el sistema completo falla. Supongamos que cada componente sigue un modelo de tiempo a la falla exponencial con $\theta = 100$*, y que los componentes fallan en forma independiente una de la otra. Definamos los eventos*

$$A_i = \{i - \text{ésimo componente dura al menos } t \text{ horas}\}, i = 1, 2, 3, 4, 5.$$

El modelo exponencial

Ejercicio (continua...)

Las A_i's son independientes e identicamente distribuidas. Sea X el tiempo de la falla del sistema.

a) *El evento $\{X \geq t\}$, ¿a cuál evento, en términos de las A_i's, es equivalente?*
b) *Usando la independencia de las A_i's, calcula $P(X \geq t)$. Obtén $F(t) = P(X \leq t)$. ¿Cuál es la distribución de X?*
c) *Si en lugar de 5 componentes tenemos n, ¿cuál es la distribución de X?*

El modelo exponencial

Ejemplo

Supongamos ahora que tenemos un inventario de 6 componentes claves (un paquete) de un cierto sistema. El **tiempo a la falla de cada componente sigue un modelo exponencial con $\theta = 1$ (en cientos de horas)**, *y podemos asumir que se comportan de forma* **independiente**. *Cuando el componente en el sistema falla es reemplazado en forma inmediata por uno del almacén.*

Si los tiempos empleados en el cambio son suficientemente pequeños para ignorarlos, ¿cuál es la distribución del tiempo a la falla del sistema? Con esto podríamos calcular el tiempo de vida medio y así saber cuándo debe volver a surtir para que el sistema no se detenga por causa de este componente.

El modelo exponencial

Solución: Sea $X = \sum_{i=1}^{6} X_i$ la suma de los tiempos de vida de cada componente, donde $X_i \sim exp(\theta = 1)$, que además son independientes.

Figura: Proceso Poisson. Los éxitos ocurren en forma aleatoria sobre el intervalo de definición.

El modelo exponencial

Entonces, claramente $E(X) = \sum_{i=1}^{6} E(X_i)$, pero ¿cuál es la distribución de X?

Podríamos tomar a X como el tiempo hasta obtener $r = 6$ éxitos. Entonces si consideramos un proceso Poisson(λ), y contamos el tiempo hasta que ocurran 6 éxitos, podríamos construir nuestra función de la siguiente manera:

Sea X el tiempo hasta observar r éxitos. Entonces

$$\begin{aligned} F_X(x) &= P(X \leq x) = 1 - P(X > x) \\ &= 1 - P(\# \text{ éxitos en el intervalo } [0, x] \text{ es a lo más } r-1) \\ &= 1 - P(W \leq r-1), \quad \text{donde } W \sim Poisson(\lambda x) \\ &= 1 - \sum_{w=0}^{r-1} \frac{(\lambda x)^w e^{-\lambda x}}{w!}. \end{aligned}$$

El modelo exponencial

Para determinar la densidad de X lo único que nos resta es derivar la función acumulada:

$$
\begin{aligned}
f(x) = \frac{d}{dx}F(x) &= -\frac{d}{dx}\left(e^{-\lambda x} + \sum_{w=1}^{r-1}\frac{(\lambda x)^w e^{-\lambda x}}{w!}\right) \\
&= \lambda e^{-\lambda x} + \lambda\sum_{w=1}^{r-1}\frac{(\lambda x)^w e^{-\lambda x}}{w!} - \lambda\sum_{w=1}^{r-1}\frac{w(\lambda x)^{w-1}e^{-\lambda x}}{w(w-1)!} \\
&= \lambda\sum_{w=0}^{r-1}\frac{(\lambda x)^w e^{-\lambda x}}{w!} - \lambda\sum_{y=0}^{r-2}\frac{(\lambda x)^y e^{-\lambda x}}{y!}, \quad \text{tomando } y = w-1 \\
&= \frac{\lambda(\lambda x)^{r-1}e^{-\lambda x}}{(r-1)!}. \qquad (3)
\end{aligned}
$$

¿La conocemos? Aún no, pero ésta es una de las formas como se puede presentar un modelo de probabilidad conocido como *la distribución de probabilidad Gamma*.

Función de distribución

Ejemplo

Supongamos que la hipótesis de que $X_1, \ldots, X_n \sim exp(1/2)$ *es de interés y nos piden una gráfica Q-Q para validar éste supuesto. Seguimos los siguientes pasos:*

1. *Determinar la función de distribución inherente al problema. En este caso*
$$F(x) = 1 - e^{-x/2}.$$
2. *Calcular la función inversa, es decir, resolver* $F^{-1}(x) = y$
$$x = 1 - e^{-y/2} \Leftrightarrow 1 - x = e^{-y/2} \Leftrightarrow \log(1-x) = -y/2$$
3. *Finalmente, graficar*
$$X_{(i)} \ vs \ -2\log\left(1 - \frac{i}{n+1}\right)$$

Función de distribución empírica

Ejemplo

Si la hipótesis fuese cierta, deberiamos observar algo así

El modelo Gamma

El modelo Gamma

Sea X una v.a. continua. Diremos que X se distribuye de acuerdo a una función de probabilidad Gamma con parámetros α y β si su función de densidad es

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{para } x \geq 0 \\ \\ 0 & \text{para } x < 0 \end{cases},$$

con α, β números reales positivos y $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du$ es la función Gamma o función factorial generalizada.

El modelo Gamma

Notación: $X \sim Gamma(\alpha, \beta)$.

Algunas propiedades de la función Gamma:

1. $\Gamma(1) = 1$
2. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. Esto se demuestra integrando por partes la función $\Gamma(\alpha + 1)$.
3. Si n es un entero positivo, $\Gamma(n) = (n - 1)!$
4. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Los parámetros de la distribución α y β se conocen como parámetros de forma y escala, y no en todos los casos tienen una interpretación directa. En el ejemplo de los tiempos de vida, α representa el número de componentes y β el tiempo de vida medio de cada componente, pero, repetimos, no siempre existen asociaciones como éstas.

El modelo Gamma

Figura: Distribución Gamma

El modelo Gamma

Si $\alpha = 1$

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{para } x \geq 0 \\ \\ 0 & \text{para } x < 0 \end{cases},$$

de aquí que una v.a. exponencial es un caso particular de la Gamma cuando $\alpha = 1$ y $\beta = \theta$.

El modelo Gamma

El modelo Gamma

Notemos además que (3) en el ejemplo anterior, el del inventario con los 6 componentes, puede reescribirse como

$$f(x) = \frac{\lambda(\lambda x)^{r-1}e^{-\lambda x}}{(r-1)!} = \frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} = \frac{1}{\Gamma(r)\left(\lambda^{-1}\right)^r}x^{r-1}e^{-(x/\lambda^{-1})},$$

de donde se sigue que

$$X = \text{tiempo hasta observar } r \text{ éxitos} \sim Gamma\left(\alpha = r, \beta = 1/\lambda\right).$$

Esto es,

$$\begin{aligned} X &= \text{Suma de variables aleatorias exponenciales } 1/\lambda, \\ &\quad \text{independientes} \sim Gamma\left(\alpha = r, \beta = 1/\lambda\right), \end{aligned}$$

donde α es un número entero en este caso.

Este resultado es de suma utilidad para calcular probabilidades de forma sencilla (sin integrar o recurrir a software). La única condición es que α sea entero.

El modelo Gamma

Entonces

$$\begin{aligned} P(X > x) &= P(W \leq \alpha - 1) \\ &= \sum_{w=0}^{\alpha-1} \frac{(x/\beta)^w e^{-(x/\beta)}}{w!}, \end{aligned}$$

con $w \sim Poisson(\lambda x)$, $\lambda = 1/\beta$.

Aunque no exista ningún proceso Poisson detrás de este argumento, es solo una herramienta intermedia para hacer el cálculo, y **no** es una aproximación sino una igualdad.

El modelo Gamma

Veamos ahora algunos resultados asociados con este modelo,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx \\ &= \int_{0}^{\infty} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)} dx \\ &= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_{0}^{\infty} x^{\alpha-1} e^{-(x/\beta)} dx, \end{aligned}$$

y la integral $\int_{0}^{\infty} x^{\alpha-1} e^{-(x/\beta)} dx = \Gamma(\alpha)\beta^{\alpha}$, de donde, en efecto, $f(x)$ representa una función de probabilidad.

El modelo Gamma

Primero obtendremos la fgm y con ella los momentos de la distribución, dado que esto facilita los cálculos de los mismos.

$$
\begin{aligned}
M_X(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \\
&= \int_0^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{tx-(x/\beta)} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-\overbrace{\left(\frac{1}{\beta}-t\right)}^{\frac{1}{\beta^{(*)}}} x} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Gamma(\alpha)\beta^{(*)^{\alpha}} \quad \text{donde} \frac{1}{\beta^{(*)}} = \left(\frac{1}{\beta}-t\right) = \frac{1-\beta t}{\beta},\ \beta^{(*)} = \frac{\beta}{1-\beta t} \\
&= = \frac{1}{\beta^{\alpha}} \left(\frac{\beta}{1-\beta t}\right)^{\alpha} = \frac{1}{(1-\beta t)^{\alpha}} \\
&= (1-\beta t)^{-\alpha}.
\end{aligned}
$$

$\therefore M_X(t) = (1-\beta t)^{-\alpha}$.

El modelo Gamma

Ahora,

$$\begin{aligned} E(X) &= M'_X(t)|_{t=0} \\ &= -\alpha\left(1-\beta t\right)^{-\alpha-1}\left(-\beta\right)|_{t=0} \\ &= \alpha\beta\left(1-\beta t\right)^{-\alpha-1}|_{t=0} = \alpha\beta \end{aligned}$$

$\therefore E(X) = \alpha\beta.$

$M''_X(0) = \alpha\beta^2\left(\alpha+1\right)$, por lo tanto, $E(X^2) = \alpha\beta^2\left(\alpha+1\right)$.
Entonces

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \alpha\beta^2\left(\alpha+1\right) - (\alpha\beta)^2 \\ &= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2, \end{aligned}$$

$\therefore V(X) = \alpha\beta^2.$

El modelo Gamma

Otro caso particular de la distribución Gamma se presenta cuando los parámetros toman los valores

$$\alpha = \frac{\nu}{2} \qquad \text{y} \qquad \beta = 2, \quad \nu \text{ un número entero,}$$

y recibe el nombre de variable χ^2 (ji-cuadrada o chi-cuadrada) con ν grados de libertad.

Notación: $X \sim \chi^2(\nu)$. Su función de densidad está dada por

$$f(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}} x^{\left(\frac{\nu}{2}-1\right)} e^{-\frac{x}{2}} & x \geq 0 \\ \\ 0 & x < 0 \end{cases}.$$

El modelo Gamma

De los resultados anteriores, por simple sustitución obtenemos:

$$
\begin{aligned}
M_X(t) &= (1-2t)^{-\frac{\nu}{2}} \\
E(X) &= \nu && \text{(los grados de libertad)} \\
V(X) &= 2\nu && \text{(dos veces los grados de libertad)}
\end{aligned}
$$

La variable χ^2 es importante en el muestreo (y hablaremos de ella más adelante). Nos permite establecer comportamientos de características de interés como la varianza muestral y su relación con inferencia estadística.

Nota que en este caso sólo hay un parámetro: ν.

El modelo Gamma

El modelo Gamma

Algunas chi-cuadradas son muy asimétricas, esto depende básicamete de los grados de libertad ν: entre más grande es este valor, la forma de la gráfica tiende a verse más simétrica.

Ejercicio

¿Cuál es la densidad de una v.a. χ^2 con $\nu = 2$?

El modelo Gamma

Ejemplo

En cierta ciudad, el consumo diario de energía (en millones de kilowatts/hora) sigue un modelo Gamma$(\alpha = 3, \beta = 2)$*. Esto es un promedio de consumo diario de 6 millones de kilowatts/hora (K/h).*

Si la planta de esta ciudad tiene una capacidad de producción máxima de 12 millones K/h ¿Cuál es la probabilidad de que el suministro sea insuficiente en un día dado?

El modelo Gamma

Sea X el consumo de energía en millones de K/h. Entonces,

$$P(X > 12) = \int_{12}^{\infty} \frac{1}{\Gamma(3)2^3} x^{3-1} e^{-\frac{x}{2}} dx.$$

Para resolver este problema tenemos 3 opciones:

1. Resolver directamente la integral. En este caso es sencillo, ya que solo hay que integrar por partes 2 veces, ya que α es un valor entero y eso facilita las cosas. Recordemos que $\Gamma(3) = (3-1)! = 2$. (HACERLO).
2. Reconocer que, ya que $\beta = 2$ y α un entero, siempre puede identificarse como un elemento de la distribución χ^2, en particular, $\alpha = 6/2$, y estamos hablando de una χ^2 con $\nu = 6$ grados de libertad. Entonces, si se cuenta.

El modelo Gamma

3. α es un valor entero, entonces podemos usar el intercambio de probabilidades Poisson.

   Recordemos que $[0, x] = [0, 12]$; $\beta = 1/\lambda$, entonces $\lambda = 1/2$.

$$\begin{aligned} P(X > 12) &= P(W \leq 3 - 1), \text{ donde } W \sim Poisson(\lambda x = 12/2) = Poisson(6) \\ &= \sum_{w=0}^{2} \frac{6^w e^{-6}}{w!} = 0.062. \end{aligned}$$

   Nota: *cuando α no es entero hay que utilizar tablas de la Gamma incompleta o software estadístico*.

El modelo Gamma

Existe una identidad con funciones Gamma dada en la siguiente ecuación:

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.$$

Esto nos dice que

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = 1,$$

de donde esta función puede servir como función de densidad.

El modelo Beta

El modelo Beta

Es un modelo que se aplica a v.a. (esto es, características bajo estudio) que tienen un rango de definición finito y pueden ser entonces codificadas entre $(0, 1)$, o bien a proporciones tales como porcentajes en compuestos químicos, porcentajes de tiempos utilizados en el desempeño de una labor dada, etc. Todas estas situaciones pueden ser modeladas mediante esta familia de distribuciones. En estadística Bayesiana es donde también ha encontrado un amplio uso.

Se dice que X sigue una distribución Beta con parámetros $\alpha > 0$ y $\beta > 0$ si

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{otro caso.} \end{cases}$$

Notación: $X \sim Beta(\alpha, \beta)$.

El modelo Beta

Notemos que si $\alpha = 1$ y $\beta = 1$,

$$f(x) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} x^{1-1}(1-x)^{1-1} = 1 \qquad 0 < x < 1,$$

esto es, $X \sim U(0,1)$.

Para demostrar que la función f integra 1, se sugiere revisar la página de wikipedia sobre la función beta `https://en.wikipedia.org/wiki/Beta_function`.

La distribución Beta *es una familia rica en formas*, como lo muestran las siguientes gráficas para distintos valores de los parámetros:

El modelo Beta

Figura: Distribución Beta

El modelo Beta

Para calcular la media y varianza de este modelo es mejor hacer el cálculo de $E(X^r)$ y de allí obtener todo lo que necesitamos.

Sea r un número real positivo,

$$
\begin{aligned}
E(X^r) &= \int_{-\infty}^{\infty} x^r f(x)dx = \int_0^1 x^r \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+r-1}(1-x)^{\beta-1}dx; \quad \alpha^* = \alpha + r \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha^*)\Gamma(\beta)}{\Gamma(\alpha^*+\beta)} \\
&= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)}
\end{aligned}
$$

$$
\therefore E(X^r) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)}.
$$

El modelo Beta

Observa que siempre tratamos de evitar las integrales reconociendo las funciones que están en ellas y completándolas para obtener su valor en forma inmediata.

Con $r = 1$:

$$\begin{aligned} E(X) &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} \\ &= \frac{\Gamma(\alpha+\beta)\alpha\Gamma(\alpha)}{\Gamma(\alpha)(\alpha+\beta)\Gamma(\alpha+\beta)} \\ &= \frac{\alpha}{\alpha+\beta}. \end{aligned}$$

El modelo Beta

Con $r = 2$:

$$E(X^2) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(\alpha+\beta+2)} = \frac{\Gamma(\alpha+\beta)(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)}$$
$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2}$$
$$= \frac{(\alpha+\beta)(\alpha^2+\alpha) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\therefore V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

El modelo Beta

La función generatriz de momentos expandida en series de potencias es

$$M_X(t) = 1 + tE(X) + \frac{t^2}{2}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots \quad t \in \mathbb{R},$$

entonces, usando la fórmula que se obtuvo para $E(X^r)$ se puede dejar indicada $M_X(t)$ como una serie de potencias.

Existe una relación entre las v.a. binomiales y las beta, sin embargo, a diferencia de las relaciones que hemos mostrado antes entre otras variables, ésta no tiene un caracter intuitivo, como verás a continuación. Sólo usa el resultado, la demostración se presenta por completez. Nota que esta relación es válida si los valores de los parámetros de la Beta son números enteros.

El modelo Beta

Resultado

Si $Y \sim Beta(\alpha = k, \beta = n - (k-1))$ *y* $X \sim Binomial(n, p)$, *entonces*

$$P(Y > p) = P(X \leq k - 1),$$

esto es,

$$\int_p^1 \frac{n!}{(k-1)!(n-k)!} y^{k-1}(1-y)^{n-k} dy = \sum_{x=0}^{k-1} \binom{n}{x} p^x (1-p)^{n-x}.$$

El modelo Beta

Demostración

Se demostrará que

$$
\begin{aligned}
P(X \geq k) &= \sum_{j=k}^{n} \binom{n}{j} p^j (q)^{n-j} \\
&= \int_0^p \frac{1}{B(k, n-k+1)} u^{k-1}(1-u)^{n-k} du \\
&= \frac{1}{B(k, n-k+1)} \int_0^p u^{k-1}(1-u)^{n-k} du
\end{aligned}
$$

para la variable aleatoria $X \sim$ *Binomial*.

El modelo Beta

Demostración (continúa...)

$$
\begin{aligned}
P(X \geq k) &= \sum_{j=k}^{n} \binom{n}{j} p^j (q)^{n-j} \\
\frac{d}{dp} P(X \geq k) &= \sum_{j=k}^{n} \binom{n}{j} \left\{ p^j (n-j)(1-p)^{n-j-1}(-1) + jp^{j-1}(1-p)^{n-j} \right\} \\
&= \sum_{j=k}^{n} \left\{ \frac{n!}{j!(n-j)!} jp^{j-1}(1-p)^{n-j} - \frac{n!}{j!(n-j)!}(n-j)p^j(1-p)^{n-j-1} \right\} \\
&= \sum_{j=k}^{n} \left\{ \frac{n(n-1)!}{(j-1)!(n-j)!} p^{j-1}(1-p)^{n-j} - \frac{n(n-1)!}{j!(n-j-1)!} p^j(1-p)^{n-j-1} \right\} \\
&= \underbrace{\sum_{j=k}^{n} \left\{ n\binom{n-1}{j-1} p^{j-1}(1-p)^{n-j} - n\binom{n-1}{j} p^j(1-p)^{n-j-1} \right\}}_{\text{es una suma telescópica}}
\end{aligned}
$$

El modelo Beta

Demostración (continúa...)

entonces,

$$\frac{d}{dp}P(X \geq k) = n\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}.$$

Si $f(p) = P(X \geq k)$, *se tiene que* $f(0) = \sum_{j=k}^{n}\binom{n}{j}0^j(1)^{n-j} = 0$. *Entonces, si se integran ambos lados de la última ecuación obtenemos*

$$\begin{aligned}
P(X \geq k) &= \int_0^p \frac{d}{dp}P(X \geq k)dp = n\binom{n-1}{k-1}\int_0^p p^{k-1}(1-p)^{n-k}dp \\
&= \frac{n!}{(k-1)!(n-k)!}\int_0^p u^{k-1}(1-u)^{n-k}du \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)}\int_0^p u^{k-1}(1-u)^{n-k}du \\
&= \frac{1}{B(k,n-k+1)}\int_0^p u^{k-1}(1-u)^{n-k}du.
\end{aligned}$$

El modelo Beta

Ejemplo

Si la proporción anual de restaurantes nuevos que fracasan en la ciudad de Monterrey es una v.a. que puede ser modelada mediante una Beta con $\alpha = 1$ *y* $\beta = 4$*, determinar:*

a) *La proporción media de restaurantes que fracasan en un año dado.*
b) *La probabilidad de que al menos el 25 % de los restaurantes fracasen.*

El modelo Beta

Solución. Sea X el porcentaje de restaurantes que quiebran en un año dado.

$$f(x) = \frac{\Gamma(5)}{\Gamma(1)\Gamma(4)} x^{1-1}(1-x)^{4-1} = \frac{4!}{3!}(1-x)^3 = 4(1-x)^3.$$

a) El valor medio está dado por

$$E(X) = \frac{\alpha}{\alpha+\beta} = 0.2,$$

es decir, el 20 %.

b) Además,

$$P(X > 0.25) = \int_{.25}^{1} 4(1-x)^3 dx.$$

Sea $u = 1 - x$, $du = -dx$, así $u = .75$ y $u = 0$.

El modelo Beta

Entonces

$$P(X > 0.25) = -\int_0^{.75} -4u^3 du = u^4\Big|_0^{.75} = .316,$$

como era de esperarse de la gráfica y del valor medio.

El modelo Beta

Ejercicio

- *En el ejemplo anterior usa la relación con la Binomial para hacer el cálculo de la probabilidad en b).*
- *En muchos proyectos se emplea un método llamado PERT (Program Evaluation and Review Technique) para coordinar varias actividades en proyectos grandes y/o complejos (por ejemplo, fue empleado en la construcción de los Apolo). Un supuesto estándar de esta técnica es que el tiempo necesario para completar cualquier actividad una vez que ha comenzado, se puede modelar mediante una distribución Beta con*
  - $A =$ *tiempo optimista (si todo va bien)*
  - $B =$ *tiempo pesimista (si todo va mal)*

El modelo Beta

Ejercicio (Continúa)

Por ejemplo, supongamos que estamos construyendo casas habitación y que el tiempo necesario, en días, para poner los cimientos de una casa sigue una distribución Beta con $A = 2$ *y* $B = 5$*. Además, se puede calcular que* $\alpha = 2$ *y* $\beta = 3$*.*

Calcula la probabilidad de que el tiempo para terminar los cimientos sea menor a 3 días.

Nota: A y B no son los valores de los parámetros, recuerda que primero debe codificarse la variable para que tome valores entre $(0, 1)$*.*

El modelo Weibull

El modelo Weibull

Volviendo sobre la definición de la función de riesgo o tasa de falla:

$$h(t) = \frac{f(t)}{R(t)}.$$

Recordemos que definimos a la CONFIABILIDAD de un componente, producto, sistema, etcétera, como la probabilidad de que funcionará correctamente al menos durante un tiempo especificado t.

$$\begin{aligned} R(t) &= P(T \geq t) = P(T > t) \qquad \text{para variable continuas} \\ &= \int_t^{\infty} f(x)dx \qquad \text{donde } f \text{ es la función de densidad de } T \\ &= 1 - F(t) \qquad \text{donde } F(t) \text{ es la acumulada de } T. \end{aligned}$$

El modelo Weibull

También vimos que la probabilidad condicional de que una componente falle en el intervalo $(t, t+\Delta t)$ dado que no falló hasta el tiempo t es

$$c = P(\text{falla en } (t, t+\Delta t) | \text{ no falló hasta } t)$$

y que cumple

$$\begin{aligned} c &= \frac{P(\text{intersección})}{P(\text{no falle antes de } t)} = \frac{P(\text{falla en } (t, t+\Delta t))}{R(t)} \\ &= \frac{1}{R(t)} \int_t^{t+\Delta t} f(x)dx \\ &= \frac{F(t+\Delta t) - F(t)}{R(t)}. \end{aligned}$$

El modelo Weibull

Dividiendo entre Δt y tomando el límite cuando $\Delta t \to 0$ se obtiene una **tasa o razón de falla instantánea** y representa una medida del riesgo de que el producto falle en el siguiente instante dado que ha funcionado bien hasta el tiempo t.

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t \cdot R(t)} \\ &= \frac{1}{R(t)} \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} = \frac{F'(t)}{R(t)} \\ &= \frac{f(t)}{R(t)}. \end{aligned}$$

El modelo Weibull

La tasa de falla $h(t)$ se puede encontrar a partir de la densidad $f(t)$.

Sin embargo, algo que resulta más importante todavía es que, si se conoce la tasa de falla, podemos encontrar la función de densidad. Primero observamos que, dado que $R(t) = 1 - F(t)$, entonces

$$R'(t) = -F'(t) = -f(t),$$

y por lo tanto

$$h(t) = -\frac{R'(t)}{R(t)},$$

la cual puede ser vista como una ecuación diferencial que puede ser resuelta para $h(t)$.

El modelo Weibull

Integrando en ambos lados de la ecuación obtenemos:

$$\int h(t)dt = -\log R(t) + C_1$$

$$\Rightarrow -\int h(t)dt = \log R(t) + C_2$$

$$\Rightarrow \exp\left\{-\int h(t)dt\right\} = K \cdot R(t)$$

$$\therefore R(t) = Ce^{-\int h(t)dt}.$$

El modelo Weibull

Entonces, en resumen, conociendo $R(t)$ se puede encontrar $h(t)$ y conociendo $h(t)$ se puede encontrar $R(t)$ mediante

$$R(t) = Ce^{-\int h(t)dt},$$

y la constante C se determina por la condición inicial $R(0) = 1$.

Ahora, consideremos la tasa de falla dada por

$$h(t) = \alpha\beta t^{\beta-1}, \quad t > 0.$$

Confiabilidad

Figura: Gráfica de "tina de baño". (a) etapa de ajuste inicial o "mortalidad infantil". (b) etapa constante. (c) etapa de desgaste o envejecimiento.

El modelo Weibull

Según esta definición $h(t)$ será

$$h(t) = \begin{cases} \alpha : \text{constante} & \text{si } \beta = 1 \\ \text{decreciente} & \text{si } \beta < 1 \\ \text{creciente} & \text{si } \beta > 1 \end{cases}.$$

Entonces, dependiendo de los valores de los parámetros obtendremos una familia muy rica para describir toda clase de comportamientos asociados con la función de riesgo, como los descritos en la "tina de baño".

El modelo Weibull

Entonces la pregunta es, ¿cuál es la densidad que tiene esta $h(t)$ cómo su función de riesgo?

Sólo tenemos que evaluar $R(t)$ y de allí, derivando, podemos encontrar la función deseada, la cual es llamada **Función Weibull** de dos parámetros

$$\begin{aligned} R(t) &= Ce^{-\int h(t)dt} \\ &= Ce^{-\int \alpha\beta t^{\beta-1}dt} \\ &= Ce^{-\alpha t^{\beta}}; \end{aligned}$$

y con la condición inicial $R(0) = 1$, tenemos que $C = 1$. Por lo tanto $R(t) = e^{-\alpha t^{\beta}}$, y esto implica que

$$f(t) = h(t)R(t) = \alpha\beta t^{\beta-1}\left(e^{-\alpha t^{\beta}}\right), \quad t > 0,$$

lo cual corresponde a la función de densidad de una variable aleatoria Weibull(α, β).

El modelo Weibull

Definición

La variable aleatoria continua X tiene una distribución Weibull con parámetros α y β si su función de densidad es

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}} & x \geq 0 \\ \\ 0 & x < 0 \end{cases}$$

donde α y β son constantes positivas.

El modelo Weibull

Figura: Distribución Weibull

El modelo Weibull

Cuando $\beta = 1$, se obtiene la densidad $\alpha e^{-\alpha x}$, la cual es exponencial con $\theta = \frac{1}{\alpha}$, i.e. $\frac{1}{\theta} e^{-x/\theta}$, que corresponde a $h(t) = \alpha =$ constante $= 1/\theta$.

Calculamos el valor esperado de X^n:

$$\begin{aligned} E(X^n) &= \int_{-\infty}^{\infty} x^n f(x) dx \\ &= \int_{0}^{\infty} x^n \alpha \beta x^{\beta-1} e^{-\alpha x^\beta} dx \\ &= \int_{0}^{\infty} \alpha \beta x^{n+\beta-1} e^{-\alpha x^\beta} dx. \end{aligned}$$

El modelo Weibull

Haciendo el cambio de variable $u = \alpha x^{\beta}$, $x = \left(\frac{u}{\alpha}\right)^{1/\beta}$,
$dx = \frac{1}{\beta}\left(\frac{u}{\alpha}\right)^{1/\beta-1}\left(\frac{1}{\alpha}\right)du$, obtenemos:

$$
\begin{aligned}
E(X^n) &= \int_0^{\infty} \alpha\beta\left(\left(\frac{u}{\alpha}\right)^{1/\beta}\right)^{n+\beta-1} e^{-u}\frac{1}{\beta}\left(\frac{u}{\alpha}\right)^{1/\beta-1}\left(\frac{1}{\alpha}\right)du \\
&= \int_0^{\infty}\left(\frac{u}{\alpha}\right)^{\frac{n+\beta-1}{\beta}+\frac{1}{\beta}-1} e^{-u}du = \int_0^{\infty}\left(\frac{u}{\alpha}\right)^{\frac{n}{\beta}+1-\frac{1}{\beta}+\frac{1}{\beta}-1} e^{-u}du \\
&= \int_0^{\infty}\left(\frac{u}{\alpha}\right)^{\frac{n}{\beta}} e^{-u}du = \alpha^{-\frac{n}{\beta}}\int_0^{\infty} u^{\frac{n}{\beta}} e^{-u}du \\
&= \alpha^{-\frac{n}{\beta}}\int_0^{\infty} u^{\left(\frac{n}{\beta}+1\right)-1} e^{-u}du \\
&= \alpha^{-\frac{n}{\beta}}\Gamma\left(\frac{n}{\beta}+1\right).
\end{aligned}
$$

El modelo Weibull

La media de la distribución Weibull es entonces:

$$E(X) = \alpha^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta} + 1\right).$$

Para encontrar la varianza, encontramos primero $E(X^2)$:

$$E(X^2) = \alpha^{-\frac{2}{\beta}} \Gamma\left(\frac{2}{\beta} + 1\right).$$

El modelo Weibull

Ahora:

$$
\begin{aligned}
V(X) &= E(X^2) - (E(X))^2 \\
&= \alpha^{-\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right) - \left(\alpha^{-\frac{1}{\beta}}\Gamma\left(\frac{1}{\beta}+1\right)\right)^2 \\
&= \alpha^{-\frac{2}{\beta}}\Gamma\left(\frac{2}{\beta}+1\right) - \alpha^{-\frac{2}{\beta}}\Gamma^2\left(\frac{1}{\beta}+1\right) \\
&= \alpha^{-\frac{2}{\beta}}\left(\Gamma\left(\frac{2}{\beta}+1\right) - \Gamma^2\left(\frac{1}{\beta}+1\right)\right).
\end{aligned}
$$

El modelo Weibull

La función generatriz de momentos no es muy usada y su forma es complicada.

La distribución Weibull se utiliza para modelar tiempos de vida o el tiempo hasta que ocurre una falla de un elemento o componente. Es decir, si T es el tiempo de falla, entonces T suele considerarse como una variable aleatoria con distribución Weibull

$$T \sim \mathit{Weibull}(\alpha, \beta).$$

Notemos que en este caso, las expresiones para la media y la varianza en términos de los parámetros son bastante complejas, al menos no es inmediato saber cuánto valen α y β, aún si conocemos μ y σ, o valores aproximados de ellos a través de estimaciones de la muestra.

El modelo Weibull

Existen otras formas de escribir la función Weibull dependiendo del autor. Otras dos formas comunes de encontrar en la literatura son

$$f(x) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, \qquad \text{ó}$$

$$f(x) = \frac{\theta}{\alpha} x^{\theta-1} e^{-\frac{x^{\theta}}{\alpha}}.$$

Entonces, *siempre es conveniente verificar cuál densidad está siendo usada para poder discernir sobre el valor de los parámetros, así como para hacer las modificaciones correspondientes a las fórmulas de media y varianza*.

El modelo Weibull

También existe la Weibull con tres parámetros, el tercero siendo el mínimo valor que puede tomar las variables, esto es, $x > \tau > 0$, en lugar de $x > 0$.

Por ejemplo, si estamos midiendo los tiempos que un trabajador tarda en construir una escoba en una fábrica determinada, debe haber un tiempo mínimo τ que debe consumirse, aún para el empleado más diestro. Esto quiere decir que debemos modificar la densidad para que la integral, ahora de τ a ∞, nos de uno.

El modelo Weibull

Ejercicio

1. *Graficar la función Weibull para los siguientes valores de los parámetros:*

$$\alpha = 1, \beta = 2$$
$$\alpha = 2, \beta = 2$$
$$\alpha = 3, \beta = 4$$

   Algebraicamente, o bien generando variables con ésta distribución y construyendo un histrograma.

2. *Grafica las funciones de confiabilidad y riesgo para cada uno de los casos anteriores.*
3. *Si $h(t) = a + bt$, ¿cuál es la densidad asociada?*
4. *Estudiar las funciones de riesgo para la distribución Normal, Gamma y Lognormal. ¿Cuál es la principal dificultad en estos casos?*

Ejemplo

El artículo *"The Load-Life Relationship for M50 bearings with Silicon Nitride Ceramic Balls"* (Lubrication Engr., 1984, pp. 153-159), reporta datos de tiempo de vida de rodamientos (millones de revoluciones) probados a 6.45 kN de carga.

47.1	68.1	68.1	90.8
103.6	106.0	115.0	126.0
146.6	229.0	240.0	240.0
278.0	278.0	289.0	289.0
367.0	385.9	392.0	505.0

a) Construye un gráfico de probabilidad Weibull. ¿Es plausible la familia de distribución Weibull?

b) Construye un gráfico de probabilidad normal. ¿Es plausible la normalidad?

Ejemplo

La distribución Weibull está dada por

$$f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}},$$

entonces la acumulada es

$$\begin{aligned} p = F(x) &= 1 - e^{-\alpha x^{\beta}} \\ \Rightarrow \log(1-p) &= -\alpha x^{\beta} \end{aligned}$$

de donde

$$\underbrace{\log(-\log(1-p))}_{y} = \underbrace{\log(\alpha)}_{\beta_0} + \underbrace{\beta \log(x)}_{\beta x^*}. \tag{4}$$

Ejemplo

Transformación de la empírica
y
x*
Q−Q plot normal
Cuantiles muestrales
Cuantiles teóricos
20
19
Histogram of x
Frecuencias
Observaciones
BoxPlot

Ejemplo

El ajuste lineal queda como

$$y = -8.61 + 1.56x^*,$$

es decir

$$\alpha = \exp\{\beta_0\} = \exp\{-8.61\} = 0.00018, \qquad \beta = 1.56.$$

Por otra parte, como

$$\mu = \alpha^{-1/\beta}\Gamma(1 + 1/\beta), \qquad \sigma = \left(\alpha^{-2/\beta}\left(\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)\right)\right)^{1/2},$$

entonces

$$\hat{\mu} = 250.56 \cdot \Gamma(1.64) = 223, \qquad \hat{\sigma} = 147.4.$$

Ejemplo

Retomando el Q-Q plot que hicimos hace algunas diapositivas, podemos retomar los datos tal cual y ajustarles una regresión a los datos contra los z_i

Ejemplo

El ajuste lineal queda como

$$y = -1.45 + 0.0066x.$$

Por otra parte, como

$$z_i = \frac{q_i - \mu}{\sigma} = \frac{1}{\sigma} q_i - \frac{\mu}{\sigma} = \beta q_i + \beta_0$$

entonces

$$\hat{\sigma} = \frac{1}{0.0066} = 150.37, \qquad \frac{\hat{\mu}}{150.37} = 1.4514 \Rightarrow \hat{\mu} = 218.25.$$

Ejemplo

Resumiendo, tenemos que el modelo normal y el modelo Weibull producen los siguientes momentos:

Distribución	$\hat{\mu}$	$\hat{\sigma}$
Normal	218.25	150.37
Weibull	223	147.4

¿Qué distribución debería un estadístico recomendar según la evidencia? ¿Por qué?

Referencia sobre confiabilidad:

1. **Statistical Methods for Reliability Data. William Q. Meeker and Luis A. Escobar. Wiley.**

Ejemplo

Transformación de la empírica
y
x*
Q−Q plot normal
Cuantiles muestrales
Cuantiles teóricos
20
19
Histogram of x
Frecuencias
Observaciones
BoxPlot

El modelo Lognormal

El modelo Lognormal

Supongamos que estamos interesados en estudiar los niveles de contaminación en el agua del Arroyo Seco (que si existe...).

Sabemos, en principio, que nuestra variable es intrínsecamente positiva y esperamos muchos valores *pequeños*, pero también, alguna proporción no despreciable de valores altos. Esto es, esperamos que el comportamiento de la variable sea relativamente sesgado a la derecha, pero que la cola de la distribución no caiga muy rápido, es decir, una distribución de cola pesada, mejor conocida como *de valores extremos*.

En situaciones como ésta, es común que una transformación logarítmica de nuestros valores pueda ser bien modelada por alguna normal $N(\mu, \sigma^2)$. Este también es el comportamiento típico de muchas variables económicas, tales como ingresos.

El modelo Lognormal

Entonces, nos enfrentamos a una situación donde, si llamamos Y a nuestra variable original (por ejemplo las concentraciones de contaminantes) y X su transformación,

$$X = \log Y \sim N(\mu, \sigma^2).$$

se dice que

$$Y = e^X \sim Lognormal(\mu, \sigma^2).$$

Observemos que necesariamente $Y > 0$, de otra manera X no está bien definida.

El modelo Lognormal

El modelo Lognormal

Para calcular probabilidades Lognormales **no** es necesario conocer explícitamente su densidad, se pueden usar las tablas de la normal. Sea $Y \sim Lognormal(\mu, \sigma^2)$ y $X \sim N(\mu, \sigma^2)$,

$$\begin{aligned} P(Y \leq y) &= P(e^X \leq y) \\ &= P(X \leq \log y) \qquad \text{pues log es una función creciente} \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{\log y - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{\log y - \mu}{\sigma}\right). \end{aligned}$$

El modelo Lognormal

De igual manera se puede probar que

$$P(y_1 \leq Y \leq y_2) = P\left(\frac{\log y_1 - \mu}{\sigma} \leq Z \leq \frac{\log y_2 - \mu}{\sigma}\right)$$

$$P(Y \geq y) = 1 - P\left(Z \leq \frac{\log y - \mu}{\sigma}\right).$$

El modelo Lognormal

Ahora bien, si estamos interesados en los valores medios y varianzas de la variable original Y, puede resultar de utilidad calcular su densidad, pero no es necesario (*¿por qué?*). Mostraremos al mismo tiempo una técnica que ya la hemos usado en más de una ocasión, conocida como la *técnica de transformación de variables a través de la función acumulada*. Conociendo la distribución acumulada de una varible continua se puede obtener la función de densidad derivando la acumulada:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt,$$

y usando el teorema fundamental del cálculo:

$$F'_X(x) = f(t)dt.$$

El modelo Lognormal

μ y σ^2 pueden considerarse la media y varianza de la normal X, pero no necesariamente son la media y varianza de la Lognormal Y.
Para la variable transformada Lognormal $Y = e^X$, tenemos que

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(e^X \leq y) = P(X \leq \log y) \\ &= F_X(\log y), \end{aligned}$$

y por tanto

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\log y) \\ &= f_X(\log y) \cdot \frac{1}{y} \quad \text{(aplicando la regla de la cadena)} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2} \cdot \frac{1}{y}. \end{aligned}$$

El modelo Lognormal

Entonces,

- La función de densidad de Y es

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}, \qquad y > 0.$$

- Si $T \sim Lognormal(\mu, \sigma^2)$ la función generatriz de momentos no existe.
- Ya que $X \sim N(\mu, \sigma^2)$, y como sabemos que $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$,

$$E(Y) = E(e^X) = M_X(1) = e^{\mu + \frac{\sigma^2}{2}}.$$

El modelo Lognormal

Para calcular $V(X)$ calculamos primero $E(Y^2)$:

$$\begin{aligned} E(Y^2) &= E\left((e^X)^2\right) = E(e^{2X}) = M_X(2) \\ &= e^{2\mu+2\sigma^2}. \end{aligned}$$

Ahora,

$$\begin{aligned} V(Y) &= E(Y^2) - (E(Y))^2 \\ &= e^{2\mu+2\sigma^2} - \left(e^{\mu+\frac{\sigma^2}{2}}\right)^2 \\ &= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right). \end{aligned}$$

Nota: Teoría de valores extremos

Tomado de `https://en.wikipedia.org/wiki/Extreme_value_theory`.

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed random variables with cumulative distribution function F and let $M_n = \text{máx}(X_1, \ldots, X_n)$ denote the maximum.

In practice, we might not have the distribution function F but the *Fisher–Tippett–Gnedenko theorem* provides an asymptotic result. If there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\Pr\{(M_n - b_n)/a_n \leq z\} \to G(z)$$

as $n \to \infty$ then

$$G(z) \propto \exp\left[-(1 + \zeta z)^{-1/\zeta}\right]$$

where ζ depends on the tail shape of the distribution.

Nota: Teoría de valores extremos

When normalized, G belongs to one of the following non-degenerate distribution families:

- Weibull distribution/Weibull law

$$G(z) = \begin{cases} \exp\left\{-\left(-\left(\frac{z-b}{a}\right)\right)^{\alpha}\right\} & z < b \\ 1 & z \geq b \end{cases}$$

  when the distribution of M_n has a light tail with finite upper bound. Also known as Type 3.

- Gumbel distribution/Gumbel law

$$G(z) = \exp\left\{-\exp\left(-\left(\frac{z-b}{a}\right)\right)\right\} \text{ for } z \in \mathbb{R}.$$

  when the distribution of M_n has an exponential tail. Also known as Type 1.

El modelo Lognormal

Continuación:

- Frechet distribution/Fréchet Law

$$G(z) = \begin{cases} 0 & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & z > b \end{cases}$$

when the distribution of M_n has a Heavy-tailed distribution (including polynomial decay). Also known as Type 2.

Referencia: Coles S. (2001) "An Introduction to Statistical Modeling of Extreme Values". Springer, London.

El modelo Lognormal

Ejercicio

El diseño y planeación de obras hidráulicas está relacionado con eventos hidrológicos futuros; por ejemplo, la avenida de diseño para el vertedor de una presa.
Se planea construir un bordo para protección contra inundaciones y se requiere saber lo siguiente:

- *¿Cuál es la probabilidad de que en un año cualquiera el gasto exceda* $7500m^3/s$ *?*
- *¿Cuál es el gasto de diseño para una avenida con período de retorno de 60 años para una inundación?*

De una estación hidrométrica se obtuvieron los gastos máximos anuales registrados. La literatura al respecto señala que este tipo de datos sigue distribuciones como la Normal, la Lognormal, la Pearson III (Gamma de tres parámetros), entre otras.

El modelo Lognormal

Se dispone de información de los gastos máximos anuales desde el año de 1950 hasta 2017. El primer paso consiste en identificar con qué tipo de población nos enfentamos. Para ello, exploraremos los datos con la ayuda de los Q-Q plots y así identificar la población de referencia.

Del Q-Q plot y los valores muestrales relevantes, podemos observar que los datos **no** parecen ser normales.

$$\text{media} = 3669.705, \qquad \text{SD} = 1262.401,$$

Prob.	10 %	25 %	50 %	75 %	90 %
Quantil	2067.687	2626.999	3599.823	4501.240	5515.856

El modelo Lognormal

El modelo Lognormal

Sabemos que si

$$X \sim Lognormal(\mu, \sigma),$$

entonces:

$$\log X \sim Normal(\mu, \sigma).$$

Tomando esta transformación, tenemos que

$$\text{media} = 8.14, \qquad \text{SD} = 0.35,$$

Prob.	10 %	25 %	50 %	75 %	90 %
Quantil	7.633744	7.873587	8.188640	8.412025	8.615374

El modelo Lognormal

El modelo Lognormal

Gráficamente podemos concluir que la transformación logarítmica sí es normal, y de esto último, que los datos son Lognormales.

Ahora, para calcular la probabilidad que se desea,

$$P(X > 7500),$$

utilizamos la transformación $\log X$ y calculamos la probabilidad deseada directamente de $P(\log X > \log 7500)$ en base a una $Normal(\mu, \sigma^2)$.

El modelo Lognormal

Así, tenemos que

$$\begin{aligned} P(\log X > \log 7500) &= P\left(Z > \frac{\log 7500 - \hat{\mu}}{\hat{\sigma}}\right) \\ &= 1 - P\left(Z \leq \frac{\log 7500 - \hat{\mu}}{\hat{\sigma}}\right) \\ &= 0.0150. \end{aligned}$$

En el cálculo anterior utilizamos los estimadores de μ y σ.

Conclusión: La población de la cual provienen los gastos máximos anuales es Lognormal y, en base a la estimación de los parámetros que provee la muestra, la probabilidad de el gasto máximo anual sea mayor a $7500 m^3/s$ es relativamente baja.

El modelo Lognormal

Para la segunda pregunta se requiere entender el concepto de período de retorno usado en ingeniería.

Un período de retorno, también conocido como intervalo de recurrencia, es un estimado de la plausibilidad de un evento, como un temblor o una inundación, de ocurrir.

Su medición estadística está basada en datos históricos y usualmente se usa para el análsis de riesgo. Esto asumiendo que la probabilidad de que que ocurra el evento no varía con el tiempo y es independiente de eventos pasados.

Formalmente, el período de retorno T se define como el inverso multiplicativo de la probabilidad de que el evento ocurra en un año particular.

El modelo Lognormal

Para entender el concepto de período de retorno observemos que la v.a. que indica si en un año dado ocurrió la inundación se distribuye $Bernoulli(1/T)$. Luego, por año uno espera $1/T$ inundaciones. Así la frecuencia esperada es T.

En el ejemplo del bordo para protección contra inundaciones, podríamos pensar que si el gasto supera los $7500m^3/s$ en un año, entonces el bordo es insuficiente y se tiene una inundación. Luego, considerando lo desarrollado anteriormente, la probabilidad de haya una inundación en un año en particular es $P(X > 7500) = 0.015$.

Así el período de retorno de una inundación en el área donde se contruyó el bordo es $T = 1/(0.015)$ años $= 66.66$ años.

El modelo Lognormal

El gasto de diseño para una avenida con período de retorno de $T = 60$ años se puede entender como el gasto máximo g que el bordo está diseñado para soportar sin causar una inundación. Luego, en este caso la probabilidad de inundación será $p = P(X > g)$ y, por tanto, $T = 1/p$.

Luego el gasto de diseño calcular resolviendo la ecuación para g

$$\begin{aligned} 0.016 = \frac{1}{T} &= P(X > g) = P(\log X > \log g) \\ &= P\left(\frac{\log X - \mu}{\sigma} > \frac{\log g - \mu}{\sigma}\right) = P\left(Z > \frac{\log g - \mu}{\sigma}\right). \end{aligned}$$

Usando la función cuantil obtenemos que

$$\frac{\log g - \mu}{\sigma} = 2.12 \Rightarrow g \approx 7221.48 m^3/s.$$

Relación con análsis del riesgo

El período de retorno es útil en análisis de riesgos (como natural, inherente o el riesgo de fallo hidrológico). Cuando se consideran las expectativas del diseño de una estructura, el período de retorno es útil para calcular el riesgo de la estructura.

El riesgo asociado a una estructura se define como

$$R = 1 - (1 - \frac{1}{T})^n$$

donde n es la vida esperada de la estructura y $\frac{1}{T}$ es la probabilidad de ocurrencia del evento en cuestión.

En el contexto del bordo, notemos que si asumimos que cada año tiene la misma probabilidad de sufrir una inundación y estos son independientes, el riesgo es la probabilidad de que suceda al menos una inundación en el período de vida de la estructura.

Relación con análsis del riesgo

Así, si se proyecta que el bordo tenga una vida útil de 60 años y $T = 1/(0.015)$ años $= 66.66$ años, el riesgo de que haya una inundación en la zona del bordo es

$$R = 1 - (1 - (\frac{1}{T}))^{60} = 1 - (1 - 0.015)^{60} = 0.59.$$

Si deseamos que el riesgo del bordo sea de $R = 0.1$ para una vida útil de $n = 60$ años, podemos despegar T de la definición de riesgo

$$T = \frac{1}{1 - (1 - R)^{1/n}}.$$

De lo anterior podemos calcular el gasto de diseño como lo hicimos antes.

Desigualdades en probabilidad

Desigualdades

Las desigualdades en teoría de probabilidad son útiles para acotar cantidades que en cualquier otro caso resultan difíciles de calcular.

Teorema (Desigualdad de Markov)

Sea X una v.a. no negativa y supongamos que $E(X)$ existe. Para cualquier $t > 0$,

$$P(X > t) \leq \frac{E(X)}{t}.$$

Demostración

Como $X > 0$,

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx = \int_0^t x f(x) dx + \int_t^\infty x f(x) dx \\ &\geq \int_t^\infty x f(x) dx \geq t \int_t^\infty f(x) dx = tP(X > t). \end{aligned}$$

Desigualdades

Teorema (Desigualdad de Chebyshev)

Sea $\mu = E(X)$ *y* $\sigma^2 = V(X)$. *Entonces, si* $t > 0$,

$$P\left(|X-\mu| \geq t\right) \leq \frac{\sigma^2}{t^2} \text{ y } P(|Z| \geq k) \leq \frac{1}{k^2},$$

donde $Z = \frac{X-\mu}{\sigma}$. *En particular,* $P(|Z| > 2) \leq 1/4$ *y* $P(|Z| > 3) \leq 1/9$.

Demostración

$$P\left(|X-\mu| \geq t\right) = P\left(|X-\mu|^2 \geq t^2\right) \leq \frac{E(X-\mu)^2}{t^2} = \frac{\sigma^2}{t^2}.$$

La segunda parte se demuestra tomando $k\sigma = t$.

Desigualdades

Ejemplo

Se prueba un método de predicción (por ejemplo, redes neuronales) con un conjunto de prueba (n casos), donde

$$\begin{aligned} X_i &= 1 \quad \textit{si la predicción es errónea} \\ X_i &= 0 \quad \textit{si la predicción es correcta.} \end{aligned}$$

Entonces, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *representa la tasa de error observado.*

Consideraremos $X_i \sim$*Bernoulli con media desconocida* p*. Quisieramos saber que tan bien* $\bar{X}_n$ *estima a* p*, en otras palabras: ¿Qué tan cerca* (ϵ) *está* $\bar{X}$ *de* p*?*

Desigualdades

Ejemplo (Cont...)

Sabemos que $V(\bar{X}_n) = \frac{V(X_i)}{n} = \frac{p(1-p)}{n}$ *y*

$$P(|\bar{X}_n - p| > \epsilon) \underbrace{\leq}_{Chebyshev} \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2},$$

ya que $p(1-p) \leq 1/4 \ \forall p$. *Entonces para* $\epsilon = 0.2$ *y* $n = 100$, *la tasa de error es 0.0625.*

Notemos que

$$\max_p p(1-p) \Rightarrow \frac{d(p-p^2)}{dp} = 1 - 2p = 0 \Rightarrow p = 1/2, \quad \therefore p(1-p) \leq 1/4.$$

Desigualdades

Podemos usar la desigualdad de Chebyshev para determinar que tan concentrada, en términos de su desviación estándar, está una variable aleatoria alrededor de su media.

Sea X una variable con media $\mu < \infty$ y varianza $\sigma^2 < \infty$ y pongamos $Z = (X - \mu)/\sigma$. Notemos que la desigualdad de Chebyshev implica que

$$P(|X - \mu| \geq k \cdot \sigma) = P(|Z| \geq k) \leq \frac{1}{k^2},$$

de donde se sigue que

$$P(|X - \mu| < k \cdot \sigma) = 1 - P(|X - \mu| \geq k \cdot \sigma) \geq 1 - \frac{1}{k^2}.$$

Desigualdades

Así, para $k = 1, 2, 3, 4, 5$, se obtienen las siguientes cotas

$$P(|X - \mu| < \sigma) \geq 1 - \frac{1}{1} = 0,$$
$$P(|X - \mu| < 2 \cdot \sigma) \geq 1 - \frac{1}{4} = 0.75,$$
$$P(|X - \mu| < 3 \cdot \sigma) \geq 1 - \frac{1}{9} = 0.8888,$$
$$P(|X - \mu| < 4 \cdot \sigma) \geq 1 - \frac{1}{16} = 0.9375,$$
$$P(|X - \mu| < 5 \cdot \sigma) \geq 1 - \frac{1}{25} = 0.96.$$

Para $k = 1$, la desigualdad de Chebyshev no da información valiosa.

Desigualdades

Para los otros casos nos da una cota para saber que tan concentrada está la v.a. Por ejemplo, para $k = 3$, en palabras, dice que la probabilidad de que los valores de X estén a una distancia de cuando mucho 3σ de su media es al menos del 89 %.

Observa que la información relevante está en el intervalo para la v.a. X, el cual depende de los valores de μ y σ, y que en general k puede ser cualquier número real positivo.

Desigualdades

Reglas empíricas.

Con ks denotando la distancia a partir de $\bar{x}$, se da el porcentaje, aproximado, de los datos contenidos en el intervalo de $\bar{x} - ks$ a $\bar{x} - ks$.

k	Intervalo	Probabilidad (porcentaje)	Teo. de Chebyshev
1	$(\bar{x} - s,\ \bar{x} + s)$	68 %	≥ 0 %
2	$(\bar{x} - 2s,\ \bar{x} + 2s)$	95 %	≥ 75 %
3	$(\bar{x} - 3s,\ \bar{x} + 3s)$	99.7 %	≥ 88.88 %

Las reglas empíricas son el "equivalente muestral" del Teorema de Chebyshev, en el cual se reemplaza $(\mu, \sigma) = (\bar{x}, s)$, y el lado derecho por las probabilidades de la tabla anterior para la k correspondiente.

Desigualdades

Las reglas fueron establecidas para distribuciones en forma de campana pero su utilidad rebasa a ésta como puedes comprobarlo en las gráficas que se presentan a continuación. Es importante mencionar que el resultado de Chebyshev se establece para poblaciones donde μ y σ son conocidas y requiere modificaciones cuando se aplica a valores muestrales $\bar{x}$ y s.

A continuación, se presentan algunas gráficas para 6 tipos de funciones de densidad (v.a. continuas) y el área que contienen en los intervalos $(\bar{x} - ks, \bar{x} - ks)$ para $k = 1, 2, 3$.

En dichas gráficas, podemos comprobar que en general la regla empírica da buenas aproximaciones.

Desigualdades

Desigualdades

Una desigualdad que se usa con cierta frecuencia en reconocimiento de patrones es la llamada *desigualdad de Hoeffding*, que es más exacta que la de Markov.

Teorema (Desigualdad de Hoeffding)

Sea $Y_1, Y_2, \ldots, Y_n$ v.a. independientes tales que $E(Y_i) = 0$, $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para cualquier $t > 0$

$$P\left(\sum_{i=1}^{n} Y_i \geq \epsilon\right) \leq e^{-t\epsilon} \prod_{i=1}^{n} e^{t^2(b_i - a_i)^2/8}.$$

En particular, si $X_1, X_2, \ldots X_n \sim$Bernoulli$(p)$ independientes, entonces, para cualquier $\epsilon > 0$

$$P(|\bar{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2},$$

donde $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Desigualdades

La demostración de la desigualdad anterior está en el libro de texto.

Ejemplo

$X_1, X_2, \ldots, X_n \sim$*Bernoulli*$(p)$ *independientes. Tomemos* $n = 100$ *y* $\epsilon = 0.2$.

- *Chebyshev:*

$$P(|\bar{X}_n - p| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{400(0.2)^2} = 0.0625.$$

- *Hoeffding:*

$$P(|\bar{X}_n - p| > 0.2) \leq 2e^{-2(100)(0.2)^2} = 0.00067,$$

que es mucho más cercana a la real.

Desigualdades

En particular, para un $\alpha > 0$, tomando $\epsilon_n = \sqrt{\frac{1}{2n}\log(2/\alpha)}$, la desigualdad de Hoeffding nos da

$$P(|\bar{X}_n - p| > \epsilon_n) \leq 2e^{-2n\epsilon_n^2} = \alpha.$$

Si $C = (\bar{X}_n - \epsilon_n, \bar{X}_n + \epsilon_n)$,

$$P(p \notin C) = P(|\bar{X}_n - p| > \epsilon_n) \leq \alpha,$$

por lo tanto, $P(p \in C) \geq 1 - \alpha$, es decir, que el intervalo C "atrapa" al verdadero parámetro p con probabilidad $1 - \alpha$.

Decimos que C es un $1 - \alpha$ intervalo de confianza.

Desigualdades

Teorema (Desigualdad de Cauchy-Schwartz)

Si X y Y tienen varianzas finitas,

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}.$$

Una forma más común de escribirla es como

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)},$$

donde la igualdad se cumple si y solo si $X - \mu_X = \lambda(Y - \mu_Y)$, con $\lambda = Cov(X,Y)/Var(Y)$.

"Si la varianza de X y Y es pequeña, la covarianza también es pequeña."

Desigualdades

Nota: Una función g es convexa si para cada x, y y cada $\alpha \in [0, 1]$,

$$g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y).$$

En particular, si g es dos veces diferenciable y $g''(x) \geq 0 \ \forall x$, entonces g es convexa.

Una función es cóncava si $-g$ es convexa.

Ejemplos de convexas: $g(x) = x^2$, $g(x) = e^x$.

Ejemplos de cóncavas: $g(x) = -x^2$, $g(x) = \log(x)$.

Desigualdades

Teorema (Desigualdad de Jensen)

Si g es convexa, entonces

$$E(g(X)) \geq g(E(X)).$$

Si g es cóncava, entonces

$$E(g(X)) \leq g(E(X)).$$

Demostración

Sea $L(x) = a + bx$ la línea tangente a $g(x)$ en el punto $E(X)$, si g es convexa

$$E(g(X)) \geq E(L(X)) = E(a + bX) = a + bE(X) = L(E(X)) = g(E(X)).$$

Desigualdades

De aquí que

- $E(X^2) \geq (E(X))^2 \Rightarrow Var(X) = E(X^2) - (E(X))^2 \geq 0$.
- Si X es positiva, $E(\frac{1}{X}) \geq \frac{1}{E(X)}$.
- $E(\log(X)) \leq \log(E(X))$, ya que $g = \log$ es cóncava.

Función de distribución empírica

Teorema (Desigualdad de Dvoretzky-Kiefer-Wolfowitz)

Sean $X_1, \ldots, X_n \sim F$ independientes. Entonces, para todo $\epsilon > 0$,

$$P\left(\sup_x |\hat{F}_n(x) - F(x)| > \epsilon\right) \leq 2e^{-2n\epsilon^2}$$

A partir de la desigualdad anterior, podemos construir un conjunto de confianza como sigue: Definamos

$$L(x) = \text{máx}\left\{\hat{F}_n(x) - \epsilon_n, 0\right\},$$
$$U(x) = \text{mín}\left\{\hat{F}_n(x) + \epsilon_n, 1\right\},$$

dónde

$$\epsilon_n = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}.$$

Función de distribución empírica

Puede verse que para cualquier F

$$P(L(x) \leq F(x) \leq U(x) \text{ para todo } x) \geq 1 - \alpha.$$

Estimación de densidad no paramétrica (EDNP). Notas de Víctor Muñiz.

Introducción

Dada una secuencia de variables aleatorias independientes e identicamente distribuidas $x_1, x_2, \ldots x_n$ con función de densidad común $f(x)$,

¿Cómo podemos estimar $f(x)$?

(Parzen, E. *On estimation of a Probability Density Function and Mode*, 1962)

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Estimación de la velocidad de la luz experimentalmente (Michaelson & Morley, 1986)

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Estimación de la velocidad de la luz experimentalmente (Michaelson & Morley, 1986)

- Estimamos su densidad
- ¿Qué modelo es el adecuado?
- Esto dependerá de varios factores, por ejemplo, la naturaleza de los datos, el conocimiento apriori del fenómeno estudiado o la experiencia del analista.

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Estimación de la velocidad de la luz experimentalmente (Michaelson & Morley, 1986)

- Estimamos su densidad
- ¿Qué modelo es el adecuado?
- Esto dependerá de varios factores, por ejemplo, la naturaleza de los datos, el conocimiento apriori del fenómeno estudiado o la experiencia del analista.

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Estimación de la velocidad de la luz experimentalmente (Michaelson & Morley, 1986)

- Estimamos su densidad
- ¿Qué modelo es el adecuado?
- Esto dependerá de varios factores, por ejemplo, la naturaleza de los datos, el conocimiento apriori del fenómeno estudiado o la experiencia del analista.

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Estimación de la velocidad de la luz experimentalmente (Michaelson & Morley, 1986)

- Ajustemos un modelo normal, los parámetros $\theta = (\mu, \sigma)$ los estimamos de la muestra.

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Longitud de ríos de Norteamérica según la US Geological Survery (1975).

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Longitud de ríos de Norteamérica según la US Geological Survery (1975).

- Estimamos su densidad
- ¿Qué modelo parece apropiado?

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Longitud de ríos de Norteamérica según la US Geological Survery (1975).

- Estimamos su densidad
- ¿Qué modelo parece apropiado?

Introducción

Un esquema tradicional de análisis de datos.

Ejemplo: Longitud de ríos de Norteamérica según la US Geological Survery (1975).

- Ajustamos un modelo Gamma(α, λ), con α el parámetro de forma y λ el parámetro de escala.

Introducción

- Aquí, todo el conocimiento que obtengamos del fenómeno bajo estudio proviene de **EL modelo** que escojamos, que a su vez nos proveé **LA función de densidad** de los datos.
- La modelación estadística tradicional escoge algún modelo paramétrico conocido para $f(x)$, por ejemplo, Normal, Gamma, Exponencial, etc...
- Esto tiene mucho sentido para fenómenos muy estudiados y analizados (estudios de confiabilidad, por ejemplo)
- Sin embargo, para fenómenos o datos más complejos, no siempre es conveniente o válido suponer una distribución de antemano. Verás muchos ejemplos más adelante.

Introducción

- Aquí, todo el conocimiento que obtengamos del fenómeno bajo estudio proviene de **EL modelo** que escojamos, que a su vez nos proveé **LA función de densidad** de los datos.
- La modelación estadística tradicional escoge algún modelo paramétrico conocido para $f(x)$, por ejemplo, Normal, Gamma, Exponencial, etc...
- Esto tiene mucho sentido para fenómenos muy estudiados y analizados (estudios de confiabilidad, por ejemplo)
- Sin embargo, para fenómenos o datos más complejos, no siempre es conveniente o válido suponer una distribución de antemano. Verás muchos ejemplos más adelante.

Introducción

- Aquí, todo el conocimiento que obtengamos del fenómeno bajo estudio proviene de **EL modelo** que escojamos, que a su vez nos proveé **LA función de densidad** de los datos.
- La modelación estadística tradicional escoge algún modelo paramétrico conocido para $f(x)$, por ejemplo, Normal, Gamma, Exponencial, etc...
- Esto tiene mucho sentido para fenómenos muy estudiados y analizados (estudios de confiabilidad, por ejemplo)
- Sin embargo, para fenómenos o datos más complejos, no siempre es conveniente o válido suponer una distribución de antemano. Verás muchos ejemplos más adelante.

Introducción

- Aquí, todo el conocimiento que obtengamos del fenómeno bajo estudio proviene de **EL modelo** que escojamos, que a su vez nos proveé **LA función de densidad** de los datos.
- La modelación estadística tradicional escoge algún modelo paramétrico conocido para $f(x)$, por ejemplo, Normal, Gamma, Exponencial, etc...
- Esto tiene mucho sentido para fenómenos muy estudiados y analizados (estudios de confiabilidad, por ejemplo)
- Sin embargo, para fenómenos o datos más complejos, no siempre es conveniente o válido suponer una distribución de antemano. Verás muchos ejemplos más adelante.

EDNP

En EDNP, no hacemos supuestos distribucionales sobre los datos.

Dado un conjunto de datos $\mathbf{X} \in \mathbb{R}^d$ (por simplicidad, empezaremos consideraremos el caso univariado):

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \sim f(\mathbf{x}),$$

queremos estimar **UNA** distribución de densidad $\hat{f}(\mathbf{x})$ que aproxime a $f(\mathbf{x})$ tal que

$$\hat{f}(\mathbf{x}) \geq 0$$

$$\int_{\mathbb{R}} \hat{f}(\mathbf{x}) d\mathbf{x} = 1,$$

(*bona fide* density).

EDNP

¿Qué nos gustaría de $\hat{f}(\mathbf{x})$?

- Que se parezcan: $E\hat{f}(\mathbf{x}) = f(\mathbf{x})$.
  Si $\hat{f}(\mathbf{x})$ es una estimación basada en una muestra de tamaño n, ésta característica nos asegura que

$$E\hat{f}(\mathbf{x}) \to f(\mathbf{x}) \text{ cuando } n \to \infty$$

  (insesgado)
- Que converja a la vedadera distribución:

$$\hat{f}(\mathbf{x}) \xrightarrow{P} f(\mathbf{x})$$

  (consistente)

Para esto, debemos definir una forma de medir la diferencia entre ambas distribuciones...

EDNP

¿Qué nos gustaría de $\hat{f}(\mathbf{x})$?

- Que se parezcan: $E\hat{f}(\mathbf{x}) = f(\mathbf{x})$.
  Si $\hat{f}(\mathbf{x})$ es una estimación basada en una muestra de tamaño n, ésta característica nos asegura que

$$E\hat{f}(\mathbf{x}) \to f(\mathbf{x}) \text{ cuando } n \to \infty$$

  (insesgado)

- Que converja a la vedadera distribución:

$$\hat{f}(\mathbf{x}) \xrightarrow{P} f(\mathbf{x})$$

  (consistente)

Para esto, debemos definir una forma de medir la diferencia entre ambas distribuciones...

EDNP

¿Qué nos gustaría de $\hat{f}(\mathbf{x})$?

- Que se parezcan: $E\hat{f}(\mathbf{x}) = f(\mathbf{x})$.
  Si $\hat{f}(\mathbf{x})$ es una estimación basada en una muestra de tamaño n, ésta característica nos asegura que

$$E\hat{f}(\mathbf{x}) \to f(\mathbf{x}) \text{ cuando } n \to \infty$$

  (insesgado)

- Que converja a la vedadera distribución:

$$\hat{f}(\mathbf{x}) \xrightarrow{P} f(\mathbf{x})$$

  (consistente)

Para esto, debemos definir una forma de medir la diferencia entre ambas distribuciones...

EDNP

¿Qué nos gustaría de $\hat{f}(\mathbf{x})$?

- Que se parezcan: $E\hat{f}(\mathbf{x}) = f(\mathbf{x})$.
  Si $\hat{f}(\mathbf{x})$ es una estimación basada en una muestra de tamaño n, ésta característica nos asegura que

  $$E\hat{f}(\mathbf{x}) \to f(\mathbf{x}) \text{ cuando } n \to \infty$$

  (insesgado)
- Que converja a la vedadera distribución:

  $$\hat{f}(\mathbf{x}) \xrightarrow{P} f(\mathbf{x})$$

  (consistente)

Para esto, debemos definir una forma de medir la diferencia entre ambas distribuciones...

EDNP: El histograma

El histograma.

Es quizá el método no paramétrico más usado para estimar y visualizar una densidad. El método es muy simple.

Supongamos que $\mathbf{x} \in [a, b]$

- Crea una partición fija de M celdas disjuntas $T_0, T_1, \ldots, T_{M-1}$ que comprendan el intervalo $[a, b]$, cada una con un ancho h.
- La densidad se estima mediante:

$$\hat{f}(x) = \frac{1}{nh} \sum_{m=0}^{M-1} N_m I_{T_m}(x),$$

donde

- I_{T_m} es la función indicadora del intervalo m,
- $N_m = \sum_{i=1}^{n} I_{T_m}(x_i)$, es decir, el número de valores que caen en la celda T_m

EDNP: El histograma

Las desventajas del histograma como un estimador de la densidad han sido mencionadas por varios autores. Por ejemplo:

- Celdas fijas
- Discontinuidades en las fronteras de las celdas
- Selección del origen del histograma

Ejemplo...

EDNP: El histograma

Las desventajas del histograma como un estimador de la densidad han sido mencionadas por varios autores. Por ejemplo:

- Celdas fijas
- Discontinuidades en las fronteras de las celdas
- Selección del origen del histograma

Ejemplo...

EDNP: KDE

Estimación usando un Kernel (KDE).

Es el método mas popular. Para el caso univariado, el KDE está dado por

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \qquad x \in \mathbb{R}, h > 0.$$

K es la función Kernel, y h es el **ancho de banda**, que determina la suavidad de la estimación.

Bajo condiciones no muy restrictivas (h debe decrecer cuando n aumenta), puede mostrarse que KDE converge en probabilidad a la verdadera densidad.

EDNP: KDE

Estimación usando un Kernel (KDE).

Es el método mas popular. Para el caso univariado, el KDE está dado por

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \qquad x \in \mathbb{R}, h > 0.$$

K es la función Kernel, y h es el **ancho de banda**, que determina la suavidad de la estimación.
Bajo condiciones no muy restrictivas (h debe decrecer cuando n aumenta), puede mostrarse que KDE converge en probabilidad a la verdadera densidad.

EDNP: KDE

EDNP: KDE

Dado un kernel K y un ancho de banda h, la KDE es *única* para un conjunto de datos específico, entonces, **no depende** de la selección del origen, como pasa con los histogramas.

El Kernel K

- Puede ser una función de densidad también, generalmente se selecciona una función unimodal y simétrica.
- El centro del kernel se coloca sobre cada dato x_i
- La influencia de cada dato se propaga en su vecindad
- La contribución de cada punto se suma para la estimación total

EDNP: KDE

Dado un kernel K y un ancho de banda h, la KDE es *única* para un conjunto de datos específico, entonces, **no depende** de la selección del origen, como pasa con los histogramas.

El Kernel K

- Puede ser una función de densidad también, generalmente se selecciona una función unimodal y simétrica.
- El centro del kernel se coloca sobre cada dato x_i
- La influencia de cada dato se propaga en su vecindad
- La contribución de cada punto se suma para la estimación total

EDNP: KDE

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Más importante que el Kernel, es la elección del ancho de banda.

El ancho de banda h

- Es un factor de escala
- Controla la suavidad o rugosidad de la estimación
- Introducimos un concepto importante: **sobreestimación**
- Esto a su vez, lleva a otro concepto aún mas importante: **Bias-Variance tradeoff**
- Veamos un ejemplo

EDNP: KDE

Cómo elegimos h

- A "ojo" (¿qué quieres ver?)
- Dos criterios:
  - Normal scale rule (también conocida como "rule of thumb")
  - Validación cruzada (CV)

EDNP: KDE

Cómo elegimos h

- A "ojo" (¿qué quieres ver?)
- Dos criterios:
  - Normal scale rule (también conocida como "rule of thumb")
  - Validación cruzada (CV)

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
  $$h^{ROT} = 1.06sn^{-1/5},$$
  con s es la estimación de σ.
  - La opción por default en R
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
$$h^{ROT} = 1.06sn^{-1/5},$$
con s es la estimación de σ.
  - La opción por default en R
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
  $$h^{ROT} = 1.06sn^{-1/5},$$
  con s es la estimación de σ.
  - La opción por default en R
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
  $$h^{ROT} = 1.06sn^{-1/5},$$
  con s es la estimación de σ.
  - La opción por default en R
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
  $$h^{ROT} = 1.06sn^{-1/5},$$
  con s es la estimación de σ.
  - La opción por default en `R`
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Cómo elegimos h

- Introducimos un criterio: MISE (mean integrated squared error). Ver notas.
- Normal scale rule (también conocida como "rule of thumb")
  - Asume que f es Normal, y calcula h óptima que minimice MISE según este supuesto.
  - Puede mostrarse que, si usamos un kernel Gaussiano, el h óptimo bajo este esquema es
  $$h^{ROT} = 1.06sn^{-1/5},$$
  con s es la estimación de σ.
  - La opción por default en `R`
  - Bien para un primer vistazo, pero tiende a suavizar de más cuando f es multimodal o claramente no-Gaussiana.
- Validación cruzada: utiliza el criterio *leave-one-out* CV.

EDNP: KDE

Ejemplo

Sea $Y \in \{0, 1\}$ una variable aleatoria binaria que indica si una persona tiene coronary heart disease (CHD) (1) o no la tiene (0), y X una variable aleatoria continua que representa su medición de systolic blood pressure (SBP).

1. *Estima $P(X = x \mid Y = 0)$ y $P(X = x \mid Y = 1)$ usando estimaciones basadas en un kernel Gaussiano.*
2. *Realiza una estimación de la probabilidad posterior de que un paciente tenga la enfermedad CHD en base a su medición de SBP, es decir:*

$$P(Y = 1 \mid X = x) = \frac{P(X = x \mid Y = 1)\, P(Y = 1)}{P(X = x \mid Y = 1)\, P(Y = 1) + P(X = x \mid Y = 0)\, P(Y = 0)}.$$

Utiliza las estimaciones usadas en el inciso anterior y distintos valores de ancho de banda del kernel. ¿Qué valor tomarías para la apriori $P(Y = 0)$ y $P(Y = 1)$? ¿Qué efecto tiene el parámetro del ancho de banda en los resultados?

EDNP: KDE

KDE multivariado.

La extensión al caso multivariado es sencilla:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|} \sum_{i=1}^{n} K(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)), \quad \mathbf{x} \in \mathbb{R}^d,$$

donde $\mathbf{H}$ es una matriz de $d \times d$ no singular que generaliza el ancho de banda h, y K es una función con media $\mathbf{0}$ e integra 1.

Ejemplo

Ejemplo: sampling KDE.

EDNP: KDE

KDE multivariado.

La extensión al caso multivariado es sencilla:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|} \sum_{i=1}^{n} K(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)), \quad \mathbf{x} \in \mathbb{R}^d,$$

donde $\mathbf{H}$ es una matriz de $d \times d$ no singular que generaliza el ancho de banda h, y K es una función con media $\mathbf{0}$ e integra 1.

Ejemplo

Ejemplo: sampling KDE.

EDNP: KDE

KDE multivariado.

La extensión al caso multivariado es sencilla:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|} \sum_{i=1}^{n} K(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{x}_i)), \quad \mathbf{x} \in \mathbb{R}^d,$$

donde $\mathbf{H}$ es una matriz de $d \times d$ no singular que generaliza el ancho de banda h, y K es una función con media $\mathbf{0}$ e integra 1.

Ejemplo

Ejemplo: sampling KDE.

Modelos para más de una Variable Aleatoria.

Texto, Cap. 3

Introducción

Hasta aquí hemos construido modelos para una sola variable aleatoria como si existieran siempre per se, sin nada que interactúe en su medio.

Sin embargo, en muchos casos, estaremos interesados en estudiar en forma conjunta más de una característica.

Por ejemplo, si medimos el porcentaje de zinc puro en una cierta reacción, no podemos modelar el comportamiento de este porcentaje sin tomar en cuenta las condiciones bajo las cuales fue producido, algunas de ellas serán "constantes", pero otras pueden variar, esto es, uno esperaría cambios en los valores de los porcentajes de zinc si las revoluciones por minuto a las que se mueve el reactor cambian, o la velocidad del extractor varía, o si la temperatura no se puede mantener fija, etc.

Introducción

En casos como éstos, algunas de las variables determinan el comportamiento de otra (es un modelo del tipo Input-Output). Formas simplificadas de estos modelos son los que usamos frecuentemente en análisis de regresión, diseño de experimentos, etc.

Considera

$$y = \mu_y + e$$

Aquí μ_y representa a la media de la variable aleatoria Y. Sin embargo, esta media cambia por la acción o presencia de otras variables, esto es:

$$\mu_y = f(x_1, x_2, x_3)$$

la media es una función (tal vez lineal, tal vez no) de las otras variables que de alguna manera quedan involucradas en el estudio.

Introducción

Por mencionar un ejemplo clásico en la literatura, supongamos que de la altura de los padres queremos saber cuál será la altura de los hijos, (predecir). Galton, en un estudio realizado sobre 942 padres e hijos, encontró algo como lo que se muestra en la gráfica siguiente:

Introducción

¿Qué quiere decir esto?

La línea horizontal representa el valor promedio de las alturas de los hijos, ignorando las alturas de los padres, sin embargo la línea con pendiente creciente nos da una mejor idea de lo que está pasando con el comportamiento de las alturas de los hijos.

Se esperaba que, entre más altos sean los padres, más lo serán los hijos, relación que sólo se cumple en cierta medida.

Vean el alto nivel de dispersión que tenemos en este caso, lo cual, típicamente oscurece las conclusiones.

Introducción

Otro caso podría ser el crecimiento de los anillos en los troncos de los árboles (Dendrocronología) como un medio para determinar su edad.

Se muestra una gráfica de 72 mediciones en donde se tomaron árboles de la misma variedad con diferente número de años:

Introducción

Aquí, μ_y, el radio promedio del anillo de esa variedad de árboles, es una función no-lineal del tiempo, lo cual es completamente lógico, se espera tienda a estabilizarse después de un cierto número (grande) de años.

Las curvas muestran diferentes opciones de cómo modelar el comportamiento medio de los radios. La línea horizontal de nuevo, representa la media de los radios, sin tomar en cuenta la edad. Claramente *you are missing something!*

¿Es la edad una variable aleatoria? La respuesta en general es sí. El proceso es el siguiente, seleccionamos al azar árboles de esta variedad en una zona donde creemos que podemos encontrar tanto árboles jóvenes como viejos. Mediante algún procedimiento alternativo (por ejemplo pruebas con carbono 14) establecemos la edad del árbol y finalmente hacemos las mediciones de los diámetros de los anillos.

Introducción

En este caso es inoperante pensar edad como un factor fijo y controlable por el investigador. Ambas mediciones son entonces variables aleatorias. De otra forma implicaría que para completar el estudio tendríamos que esperar 80 años!!!

En la industria por ejemplo, existen muchas situaciones en donde sí se puede tener el control de los factores. Recordemos el caso de producción de zinc puro, aquí, el valor de la temperatura, presión, revoluciones por minuto del revolvedor, del abanico de succión, etc. al menos en *teoría* permanecen fijos o pueden ser manipulados por el operador. Esto resulta en una simplifición muy útil cuando se plantea el modelo y como dijimos antes, son modelos que se trabajan en diseños de experimentos.

Introducción

En otros casos, no existe una relación de *dependencia*, sino que todas afectan a todas en algún nivel y nuestro interés se centra en entender tales *interdependencias*.

Consideremos otro ejemplo clásico en la literatura (**Fisher Iris data**), la construcción de un criterio de clasificación de variedades de plantas *setosa*, *versicolor* y *virginica* en base a *las mediciones del sépalo y el pétalo* (longitud y ancho).

Una vista de los datos en términos de las primeras 3 de estas características se presenta en la siguiente diapositiva.

Introducción

Se ve claramente la separación entre variedades y más aún la diferencia de versicolor y virginica con setosa.

En cada planta se miden dos características, por lo que no se espera que esos valores entre sí sean independientes.

Nos interesa saber en qué medida estas 3 mediciones nos permiten decidir si la variedad es setosa, versicolor o virginica, y lo que explotamos es la estructura de dependencias entre las mediciones de sépalos y pétalos. Esto se conoce como análisis discriminante (se estudia en Análisis Multivariado).

Introducción

Para poder establecer soluciones a problemas como los mencionados antes, y muchos otros, necesitamos una serie de definiciones y conceptos básicos que son los que trataremos en esta sección y que, cabe mencionar, es lo único que estableceremos en el curso sobre estructuras multivariadas.

Muchos de estos conceptos también nos ayudarán a hacer las conexiones entre los modelos y las inferencias que podamos hacer sobre ellos a partir de información muestral.

Para simplificar la notación trabajaremos casi exclusivamente el caso bivariado dado que la mayoría de los resultados son directamente extendibles al caso de variables aleatorias.

Introducción

Nuestros objetos de estudio son ahora parejas de variables (X, Y).

Existen diferentes combinaciones, por ejemplo, ambas variables son discretas (modelos discretos), ambas continuas (modelos continuos) o una variable siendo discreta y la otra continua (modelos de mezclas).

Daremos una explicación extensiva de los primeros dos y sólo trataremos algunos casos especiales de modelos de mezclas.

Conceptos Básicos

Modelos de Probabilidad Conjunta Discretos

Modelos de Probabilidad Conjunta Discretos

Sólo para fines de ilustrar los conceptos, comenzaremos con un ejemplo muy simple. Supongamos que estamos muestreando 3 baterías de un lote con las siguientes características:

- 3 baterías son nuevas
- 4 son reconstruidas
- 5 tienen algún tipo de problema

¡Es un lote que no desaríamos tener en las manos! Por desgracia en la práctica no sabemos lo que hay, sólo lo que vemos en la muestra y de allí tratamos de inferir sobre las cualidades del lote completo; más aún, sobre el proceso de donde provienen!!!!!.

Modelos de Probabilidad Conjunta Discretos

Si definimos

$$\begin{aligned} X &= \# \text{ de baterías nuevas en la muestra,} \\ Y &= \# \text{ de baterías reconstruidas en la muestra,} \end{aligned}$$

no es necesario definir una tercer variable Z para el resto de las baterías dado que con dos es suficiente para describir el comportamiento total de los elementos de la muestra, como veremos a continuación.

Podemos en principio, calcular las probabilidades

$$p(x, y) = P(X = x, Y = y),$$

para $x = 0, 1, 2, 3$ y $y = 0, 1, 2, 3$.

Modelos de Probabilidad Conjunta Discretos

Calculemos algunas de ellas:

$$p(0,0) = \frac{\#\text{ de casos favorables}}{\#\text{ de casos totales}} = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}.$$

Aunque, realmente,

$$\#\text{ de casos favorables} = \binom{5}{3}\binom{3}{0}\binom{4}{0}.$$

Esto define lo que podríamos llamar una distribución Hipergeométrica Multivariada.

Modelos de Probabilidad Conjunta Discretos

Continuando con el proceso anterior

$$p(0,1) = \frac{\binom{5}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{40}{220}, \qquad p(0,2) = \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220},$$

$$p(0,3) = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}, \qquad p(1,0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220},$$

$$p(1,1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}, \qquad p(1,2) = \frac{\binom{3}{1}\binom{4}{2}\binom{5}{0}}{\binom{12}{3}} = \frac{18}{220},$$

$$p(2,0) = \frac{\binom{3}{2}\binom{4}{0}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220}, \qquad p(2,1) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220},$$

$$p(3,0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220}.$$

Modelos de Probabilidad Conjunta Discretos

Arreglando los resultados anteriores en una tabla, se tiene que

x/y	0	1	2	3
0	10/220	40/220	30/220	4/220
1	30/220	60/220	18/220	0
2	15/220	12/220	0	0
3	1/220	0	0	0

Lo que obtenemos es lo que se conoce como *distribución conjunta* de X y Y. Cada entrada en la tabla no es más que la probabilidad de la intersección de dos eventos $\{X = x, Y = y\}$.

¿Cuándo sabemos que hemos construído un modelo de probabilidad conjunta válido? La respuesta, al igual que antes, radica en el cumplimiento de las reglas básicas de probabilidad.

Modelos de Probabilidad Conjunta Discretos

Definición

Sean X y Y dos variables aleatorias discretas. Se define la función función de probabilidad conjunta de X y Y como

$$f(x, y) = P(X = x, Y = y),$$

para cada par de valores (x, y) en el rango de X, Y. La cantidad representa la probabilidad de la intersección de los dos eventos: $X = x$ y $Y = y$.

La función de probabilidad conjunta satisface que:

1. $0 \leq f(x, y) \leq 1 \; \forall (x, y)$,
2. $\sum_x \sum_y f(x, y) = 1$.

Modelos de Probabilidad Conjunta Discretos

Notemos que nuestra tabla cumple con ambas propiedades:

x/y	0	1	2	3	$P\{X = x\}$
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
$P\{Y = y\}$	56/220	112/220	48/220	4/220	1

De aquí, se pueden definir otras distribuciones de interés. Por ejemplo, es claro que el último renglón en la tabla representa la distribución de $Y =$ *"# de baterías reconstruidas en la muestra"* y que la última columna es la distribución de $X =$ *"# de baterías nuevas en la muestra"*. A éstas distribuciones se les conoce como las distribuciones *marginales (por estar en los márgenes)*.

Modelos de Probabilidad Conjunta Discretos

Notemos que para obtener las distribuciones marginales lo que se hizo fue ignorar la presencia de la otra variable y, por ende, sumar sobre cada renglón o cada columna, según sea el caso.

Definición

Las distribuciones marginales de X y Y, denotadas por $p(x) = f_X(x)$ y $p(y) = f_Y(y)$, se definen como

$$f_X(x) = \sum_y p(x, y) \quad \textit{para cada } x,$$

y

$$f_Y(y) = \sum_x p(x, y) \quad \textit{para cada } y.$$

Modelos de Probabilidad Conjunta Discretos

Bajo estos casos (más de una variable) también podemos contestarnos otras preguntas que involucran el conocimiento previo de la información, por ejemplo:

Si se sabe que X tomó el valor de 2, ¿cuál es el comportamiento de Y?

Claramente esto restringe sus posibles valores y por tanto sus probabilidades cambian.

Modelos de Probabilidad Conjunta Discretos

A saber, tendríamos que calcular

$$P\{Y=0 \mid X=2\} = \frac{P\{X=2, Y=0\}}{P\{X=2\}} = \frac{\frac{15}{220}}{\frac{27}{220}} = \frac{15}{27},$$

y así para cada valor de Y.

x/y	0	1	2	3	$P\{X=x\}$
0					
1					
2	15/220	12/220	0	0	27/220
3					
$P\{Y=y \mid X=2\}$	15/27	12/27	0	0	1

Modelos de Probabilidad Conjunta Discretos

Es como olvidarnos del resto de la información: ahora, en lugar de las 220 posibles muestras iniciales, solo nos quedan 27, aquellas que tienen el valor de 2 asignado a X.

Este proceso de asignar probabilidades es de particular importancia, pues responde a un tipo de problemas, los más comunes.

Regresa sobre los ejemplos con los que iniciamos este capítulo y verás que nos preguntamos, por ejemplo, sobre el % de zinc cuando la temperatura es xxxx, etc. En otra palabras, estamos frecuentemente interesados en saber cómo se comporta nuestra característica bajo estudio sujeta a la presencia de otras variables, las cuales podemos fijar a diferentes valores y estudiar su efecto en la variable respuesta.

Estas distribuciones reciben el nombre de **distribuciones Condicionales.**

Modelos de Probabilidad Conjunta Discretos

Definición

Sean X y Y dos variables aleatorias con distribución conjunta $f(x,y)$ y marginales definidas por $f_X(x)$ y $f_Y(y)$. Entonces

i) *La distribución condicional de Y dado $X = x$ es*

$$P\{Y = y \mid X = x\} = \frac{f(x,y)}{f_X(x)},$$

para cada valor de y, donde $f_X(x) = \sum_y f(x,y)$ para cada x.

ii) *La distribución condicional de X dado Y es:*

$$P\{X = x \mid Y = y\} = \frac{f(x,y)}{f_Y(y)},$$

para cada valor de x, donde $f_Y(y) = \sum_x f(x,y)$ para cada y.

Modelos de Probabilidad Conjunta Discretos

Así las cosas, se pueden obtener tantas distribuciones condicionales para Y como valores se tengan de X, y viceversa. En nuestro ejemplo, podríamos calcular 3 distribuciones condicionales para Y y otras 3 para X.

Al igual que en el caso de una variable, son las condiciones del muestreo las que nos dirán como construir el modelo conjunto.

En muchas situaciones es a través de la recopilación histórica, de archivo o planeando un experimento para estos fines, que se pueden construir tablas de distribuciones conjuntas, utilizando frecuencias relativas, que se consideran estables (léase, correctas).

Modelos de Probabilidad Conjunta Discretos

Ejemplo

Se quiere **verificar la programación** *del tiempo que debe durar la luz verde en un semáforo que se encuentra en una cierta intersección, con vuelta a la izquierda. Como no se tiene información al respecto, se envía a un estudiante graduado a hacer observaciones sobre el número de carros X y el número de camiones Y que llegan en un ciclo (entre verde y verde); y con esto se construye la tabla de distribuciones conjunta. Se plantean una serie de preguntas y con ello confirmas qué tan eficiente fue el ciclo planeado.*

Modelos de Probabilidad Conjunta Discretos

Los resultados fueron los siguientes

		x		
	$p(x,y)$	0	1	2
y	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

Modelos de Probabilidad Conjunta Discretos

Ejercicio:

1. Verifica que es una tabla válida de probabilidades conjuntas.
2. Calcula las distribuciones marginales para carros y camiones.
3. ¿Cuál es la probabilidad de que se tengan exactamente un carro y un camión en un ciclo dado?
4. Supongamos que la vuelta a la izquierda tiene una capacidad máxima de 5 carros y un camión es equivalente a 3 carros. ¿ Cuál es la probabilidad de que se sature la línea en un ciclo dado?
5. ¿Cuál es la probabilidad de que se tenga exactamente un carro, dado que ya se tiene un camión en un ciclo predeterminado y se sabe que no habrá ningún otro camión más? Contesta esta misma pregunta para $Y = 0, 1, 2, 3, 4, 5$.

Modelos de Probabilidad Conjunta Discretos

Con el concepto de distribuciones condicionales, podemos refinar nuestra definición, hasta ahora operativa, de independencia. Este es posiblemente uno de los supuestos más usados en la construcción de modelos y por lo mismo uno de los que más se abusa.

¿Qué significaría que $P(Y = y \mid X = x) = P(Y = y)$?

Básicamente nos dice que X, al menos cuando se fija en su nivel x, no agrega información adicional sobre el comportamiento de Y. Entonces, si esta condición se cumple sin importar el valor que propongamos de X, se dirá que X y Y son independientes.

Modelos de Probabilidad Conjunta Discretos

Más formalmente tenemos la siguiente definición.

Definición

Dos variables aleatorias X y Y se dice que son independientes si, para todo par de valores x y y

$$f(x,y) = f_X(x)f_Y(y)$$

Nota:

$$P(Y=y|X=x) = \frac{f(x,y)}{f_X(x)} \quad \Rightarrow \begin{aligned} f(x,y) =& P(Y=y \mid X=x)f_X(x) \\ =& P(Y=y)f_X(x) = f_Y(y)f_X(x) \end{aligned}$$

"La conjunta es el producto de las marginales."

¿Es el número de carros en un ciclo independiente del número de camiones?

Modelos de Probabilidad Conjunta Continuos

Modelos de Probabilidad Conjunta Continuos

Los conceptos son los mismos, sólo hay que ajustar las definiciones a las características de las variables continuas.

Comencemos con un caso concreto:

Ejemplo

Un grupo de estudiantes del Programa Emprendedor establecieron un restaurante de comida rápida (fast food) e hicieron un estudio de tiempos para evaluar el servicio al cliente. Definamos Y_1 como el tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio y Y_2 el tiempo que espera el cliente en la fila hasta llegar a la ventilla de servicio. Claramente $Y_1 \geq Y_2$.

Modelos de Probabilidad Conjunta Continuos

Ejemplo (Continuación)

Con un poco más de esfuerzo y aprovechando lo que aprendieron en sus cursos de matemáticas, lograron establecer el modelo del comportamiento conjunto dado por:

$$f(y_1, y_2) = \begin{cases} e^{-y_1} & 0 \leq y_2 \leq y_1 \leq \infty \\ 0 & \textit{en otra parte} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

Ejemplo (Continuación)

Esperaban un comportamiento de caída exponencial, así que tomaron datos y ajustaron el mejor modelo en esta familia, con un software especializado.

Modelos de Probabilidad Conjunta Continuos

Notas:

- Por supuesto que es irreal pensar que tome valores infintos. No puede sobrepasar lo que dura abierto el restaurante y ningún cliente en un negocio de este tipo, soportaría una espera larga. Sin embargo, y como ya lo hicimos notar desde el incio del curso, ésta es una hipótesis de simplicidad. Notemos que a los tiempos grandes, si el modelo es adecuado, les corresponde una probabilidad prácticamente de cero por lo que resulta irrelevante si se quitan de antemano. El eliminar esos valores conllevaría por otro lado, a establcer modelos que son bastante más complejos de manejar desde su planteamiento y, a fin de cuentas, generarían respuestas casi idénticas en la región de interés.

Modelos de Probabilidad Conjunta Continuos

Notas:

- Resulta evidente de la forma de la función y de la gráfica que ahora tenemos definida una superficie y en lugar de valores por intervalo, tenemos regiones en el plano que definen a los valores de las variables. Por ejemplo, aquí, la región en donde la función toma valores distintos de cero es:

$$0 \leq y_2 \leq y_1 \leq \infty$$

esto se representa como:

Modelos de Probabilidad Conjunta Continuos

Ahora bien, necesitamos primero que nada verificar si éste es un modelo de probabilidad. Para ello y viendo las definiciones de la sección previa, extendamos la definición al caso continuo.

Definición

Una función de 2 variables $f(x, y)$ definida sobre el plano xy se llama función de densidad de probabilidad conjunta (densidad conjunta) de X y Y si y sólo si :

$$P((x, y) \in A) = \int_A f(x, y)dxdy \qquad \forall \text{ región } A \text{ del plano } xy.$$

Esto nos dice que las probabilidades se obtienen como volumenes bajo la superficie definida. La superficie debe verse siempre sobre el plano xy.

Modelos de Probabilidad Conjunta Continuos

La función de densidad conjunta satisface que:

1. $f(x,y) \geq 0$, $\forall (x,y)$.
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$

Esto nos dice que el volumen total debe ser la unidad.

a) Mostrar que la función determinada por nuestros emprendedores, es realmente una función de densidad conjunta.

Esto significa que la

$$\begin{aligned}\int_0^{\infty}\int_0^{y_1} e^{-y_1}dy_2dy_1 &= \int_0^{\infty} e^{-y_1}y_2|_0^{y_1}dy_1 = \int_0^{\infty} y_1e^{-y_1}dy_1\\ &= -y_1e^{-y_1}|_0^{\infty} + \int_0^{\infty} e^{-y_1}dy_1\\ &= 0 - e^{-y_1}|_0^{\infty} = 0+1 = 1,\end{aligned}$$

por lo que la distribución conjunta está bien definida.

Modelos de Probabilidad Conjunta Continuos

A continuación estudiaremos el concepto de función acumulada, el cual es muy útil pues nos permite calcular probabilidades. Daremos su definición pero no nos preocuparemos, en general, de calcular dicha función explícitamente; sólo nos enfocaremos en la idea que de ella se desprende que nos facilite los cálculos posteriormente.

Definición

La función de distribución conjunta o función de distribución acumulada conjunta de X y Y se define como:

$$F(x,y) = P(X \leq x, Y \leq y) = \begin{cases} \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)dsdt & X \text{ y } Y \text{ continuas} \\ \\ \sum_{t \leq y} \sum_{s \leq x} f(s,t) & X \text{ y } Y \text{ discretas} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

b) Calcular las densidades marginales de Y_1 y Y_2. Esto es, determinar los comportamientos por separado del tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio, y el tiempo que dura en la fila hasta llegar a la ventanilla de servicio.

Para responder esta pregunta, primero tenemos que definir que entendemos ahora por marginales: sólo recordemos que significa ignorar la presencia de la(s) otra(s) variable(s) y esto se logra "sumando" sobre todos los valores de las variables que no son de nuestro interés en ese momento. Geométricamente se vería como "aplastar" la gráfica sobre el eje que nos interesa.

Modelos de Probabilidad Conjunta Continuos

Definición

Las distribuciones marginales de v.a. continuas X y Y, denotadas por $f_X(x)$ y $f_Y(x)$, se definen como

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy,$$

y

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx.$$

Nota: Para el caso discreto intercambiamos las $\sum$ por $\int$.

Modelos de Probabilidad Conjunta Continuos

Ahora sí, volviendo sobre el ejemplo, tenemos

$$f_{Y_1}(y_1) = \int_0^{y_1} e^{-y_1} dy_2 = y_2 e^{-y_1}\Big|_0^{y_1}$$
$$= y_1 e^{-y_1} \quad 0 \leq y_1 \leq \infty \quad \text{una } Gamma(\alpha = 2, \beta = 1)$$

y

$$f_{Y_2}(y_2) = \int_{y_2}^{\infty} e^{-y_1} dy_1 = -e^{-y_1}\Big|_{y_2}^{\infty}$$
$$= e^{-y_2} \quad 0 \leq y_2 \leq \infty \quad \text{una } exp(\beta = 1)$$

Modelos de Probabilidad Conjunta Continuos

No siempre las distribuciones marginales resultan en distribuciones conocidas...

Definición

Dos variables aleatorias continuas se dice que son **independientes** *si su distribución conjunta se puede expresar como el producto de sus marginales*

$$f(x, y) = f_X(x) f_Y(y).$$

Más adelante volveremos sobre esta definición.

Ahora plantearemos una serie de preguntas, para algunas de las cuales tendremos que definir algunos conceptos en el contexto varias variables.

Modelos de Probabilidad Conjunta Continuos

Recordemos que: Y_1: tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio

Y_2: tiempo que espera el cliente en la fila hasta llegar a la ventanilla de servicio.

Definimos el evento $A =$ *el tiempo total es menor a 2 minutos y el tiempo en la fila es mayor a uno*. Entonces

$$P(A) = P(Y_1 < 2, Y_2 > 1).$$

Modelos de Probabilidad Conjunta Continuos

Así tenemos que

$$
\begin{aligned}
P(Y_1 < 2, Y_2 > 1) &= \int_1^2 \int_{y_2}^2 e^{-y_1} dy_1 dy_2 \\
&= \int_1^2 [-e^{-y_1}]_{y_2}^2 dy_2 \\
&= \int_1^2 [-e^{-2} + e^{-y_2}] dy_2 \\
&= \left[-e^{-2} y_2 - e^{-y_2}\right]_1^2 \\
&= -2e^{-2} - e^{-2} - [-e^{-2} - e^{-1}] \\
&= e^{-1} - 2e^{-2} \\
&= 0.0972088.
\end{aligned}
$$

Modelos de Probabilidad Conjunta Continuos

c) Ahora definimos el evento $A =$ el tiempo total de espera sea mayor al doble del tiempo en la fila. Determinar

$$P(A) = P(Y_1 \geq 2Y_2).$$

$$P(Y_1 \geq 2Y_2) = \int_0^\infty \int_{2y_2}^\infty e^{-y_1} dy_1 dy_2 = \int_0^\infty [-e^{-y_1}]_{2y_2}^\infty dy_2$$
$$= \int_0^\infty e^{-y_2} dy_2 = -\frac{e^{-2y_2}}{2} \Big|_0^\infty = \frac{1}{2}.$$

Modelos de Probabilidad Conjunta Continuos

d) La variable aleatoria $Y_1 - Y_2$ representa el tiempo que se tarda en la ventanilla de servicio. Sea el evento $A =$ el tiempo en la ventanilla sea mayor o igual que un minuto. Calcular:

$$P(A) = P(Y_1 - Y_2 \geq 1)$$

$$P(Y_1 - Y_2 \geq 1) = \int_0^\infty \int_{y_2+1}^\infty e^{-y_1} dy_1 dy_2 = \int_0^\infty [-e^{-y_1}]_{y_2+1}^\infty dy_2$$

$$= \int_0^\infty e^{-(y_2+1)} dy_2 = e^{-(y_2+1)} \Big|_0^\infty = e^{-1} = 0.36788.$$

Modelos de Probabilidad Conjunta Continuos

e) Si se sabe que el tiempo total de espera y de atención para un cliente será de más de dos minutos, calcular la probabilidad de que el tiempo que dicho cliente espera para que le atiendan sea menor de un minuto.

Se sabe que $Y_1 > 2$, queremos la probabilidad de que $Y_2 < 1$. Esto es, queremos calcular una probabilidad condicional del tipo

$$P(Y_2 < 1 | Y_1 > 2) = \frac{P(Y_2 < 1, Y_1 > 2)}{P(Y_1 > 2)}.$$

En este caso sólo nos resta calcular la probabilidad indicada en el numerador y usar el resultado de las marginales, para obtener la probabilidad en el denominador.

Modelos de Probabilidad Conjunta Continuos

La región de integración para el numerador es

$$P(Y_2 < 1, Y_1 > 2) = \int_0^1 \int_2^\infty e^{-y_1} dy_1 dy_2 = \int_0^1 [-e^{-y_1}]_2^\infty dy_2$$
$$= \int_0^1 e^{-2} dy_2$$
$$= y_2 e^{-2}|_0^1 = e^{-2}.$$

Modelos de Probabilidad Conjunta Continuos

Para el denominador:

$$P(Y_1 > 2) = \int_2^\infty f_{Y_1}(y_1)dy_1 = \underbrace{\int_2^\infty y_1 e^{-y_1} dy_1}_{\substack{u = y_1 \qquad dv = e^{-y_1}dy_1 \\ du = dy_1 \qquad v = -e^{-y_1}}} = -y_1 e^{-y_1}|_2^\infty + \int_2^\infty e^{-y_1} dy_1$$

$$= 2e^{-2} - e^{-y_1}|_2^\infty = 2e^{-2} + e^{-2} = 3e^{-2}$$

$$\therefore \qquad P(Y_2 < 1 | Y_1 > 2) = \frac{e^{-2}}{3e^{-2}} = \frac{1}{3}$$

Modelos de Probabilidad Conjunta Continuos

Nota: Podrás observar que en todo lo anterior (y en lo que sigue), no necesitamos ver el comportamiento de la superficie definida por la densidad conjunta, pero lo que sí es crucial es establecer la región de interés en el plano para definir el volumen que nos de la probabilidad requerida.

f) Calcular el tiempo medio que un cliente estará en la ventanilla.

¿Será muy probable que un cliente pase más de dos minutos en la ventanilla de servicio?

Nos preguntamos por $E(Y_1 - Y_2)$. Nosotros hemos aprendido cómo calcular valores de sumas de esperanzas en principio para cuando sólo teníamos una variable aleatoria. Primero daremos algunas definiciones para el caso de dos (o más variables aletorias) y algunas de sus propiedades y volveremos más adelante para responder a las preguntas en este apartado.

Modelos de Probabilidad Conjunta Continuos

Valores Esperados

Demos un pequeño recorrido por lo que sabemos de valores esperados:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx \quad \text{y} \quad E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy.$$

Si tenemos dos variables aleatorias, estos mismos valores se pueden obtener directamente (es como calcular la marginal dentro del procedimiento de obtención del esperado).

Modelos de Probabilidad Conjunta Continuos

Recordemos que ahora existe una densidad conjunta (o probabilidad conjunta):

$$E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy = \int_{-\infty}^{+\infty} x\left\{\int_{-\infty}^{+\infty} f(x,y)dy\right\}dx = \int_{-\infty}^{+\infty} xf(x)dx.$$

La integral en rojo, no es otra cosa que la marginal $f(x)$, de donde volvemos a la definición inicial.

Análogamente para Y:

$$\begin{aligned} E(X+Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} yf(y)dy \\ &\text{(y usando los resultados vistos arriba)} \\ &= E(X)+E(Y). \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

Con esto podemos ya responder a la primera parte de nuestra pregunta, sobre el tiempo medio que el cliente estará en la ventanilla, *pues el resultado anterior es válido, ya sea si sumamos o restamos*:

$$E(Y_1 - Y_2) = E(Y_1) - E(Y_2)$$

y estos valores se pueden obtener directamente de sus marginales.

Como vimos que la distribución de Y_1 es una Gamma(2,1) y la de Y_2 una exponencial(1)

$$E(Y_1 - Y_2) = E(Y_1) - E(Y_2) = 2 \cdot 1 - 1 = 1. \quad \text{(1 minuto en promedio)}$$

Modelos de Probabilidad Conjunta Continuos

Por otra parte, como $Y_1 - Y_2$ es el tiempo que pasa en la ventanilla de servicio

$$\Longrightarrow \text{queremos} \qquad P(Y_1 - Y_2 > 2) = \int_0^\infty \int_{y_2+2}^\infty e^{-y_1} dy_1 dy_2$$
$$= \int_0^\infty [-e^{-y_1}]_{y_2+2}^\infty dy_2$$
$$= \int_0^\infty e^{-(y_2+2)} dy_2$$
$$= -e^{-y_2+2}|_0^\infty = e^{-2} = 0.135335,$$

o sea que sólo el 13.53 % de las veces un cliente tendrá que esperar en la ventanilla más de dos minutos.

Modelos de Probabilidad Conjunta Continuos

Volvamos sobre las definiciones

A pesar de que estamos trabajando con parejas (o vectores), no definimos valores esperados sobre el vector más allá de aquellas operaciones sobre los elementos del mismo que nos den una valor real, esto es, trabajaremos sólo con los casos donde:

$$g(x, y) \to \mathbb{R} \qquad \text{(un valor en los reales).}$$

Ejemplos de tales funciones son sumas, restas, productos, logaritmos, etc. Cualquier operación que no nos regrese un vector como resultado.

Así, podemos definir en forma un poco más general, el cálculo del valor esperado con distribuciones conjuntas.

Modelos de Probabilidad Conjunta Continuos

Definición

Sean X y Y variables aleatorias con función de distribución conjunta $f(x,y)$ (o densidad conjunta) y $g(X,Y)$, una función real, entonces

$$E(g(X,Y)) = \begin{cases} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy & \text{si } x,y \text{ son continuas} \\ \\ \sum_{\forall x}\sum_{\forall y} g(x,y)p(x,y) & \text{si } x,y \text{ son discretas} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

A manera de ejercicio y para establecer un resultado que es muy útil que ya hemos usado repetidas veces, calculemos $E(XY)$

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy$$

sin más infomación, sólo nos restaría integrar para cada caso particular.

Sin embargo si sabemos que X y Y son independientes, podemos reducir mucho el trabajo de la siguiente manera:

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(y)f(y)dxdy$$

(por la independencia)

$$= \int_{-\infty}^{+\infty} xf(x)dx \int_{-\infty}^{+\infty} yf(y)dy \quad = E(X)E(Y).$$

Modelos de Probabilidad Conjunta Continuos

Lo que el resultado anterior nos dice es que, bajo independencia, el valor esperado del producto es el producto de los valores esperados (esto lo usamos para establecer propiedades de la generatriz de momentos).

El resultado también se sostiene para el caso de variables discretas.

Esta definición general de valor esperado nos da también la oportunidad de extender el concepto a el cálculo de varianzas. Recordemos que conceptualmente estamos haciendo lo mismo, esto es, construyendo una medida de la variabilidad o dispersión de los valores de las variables alrededor de su valor medio; sólo que aquí podemos hablar de la dispersión de X, de la de Y, o de cualquier nueva variable definida por la transformación bajo estudio, por ejemplo, $X + Y$, $\ln(XY)$, etc.

Modelos de Probabilidad Conjunta Continuos

Tenemos que

$$V(X) = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx \qquad \text{con } \mu_x = E(X).$$

Tomando $g(x, y) = (x - \mu_x)^2$, obtenemos que

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} (x - \mu_x)^2 \{ \int_{-\infty}^{+\infty} f(x, y) dy \} dx \\ &= \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx. \end{aligned}$$

Lo mismo para Y.

Modelos de Probabilidad Conjunta Continuos

Definición

Sean X y Y variables aleatorias con función de distribución conjunta (o densidad conjunta) $f(x,y)$ y $g(x,y)$ una función en los reales,

$$V(g(X,Y)) = \begin{cases} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[g(x,y)-\mu_{g(x,y)}]^2 f(x,y)dxdy & \text{si } x,y \text{ son continuas} \\ \\ \sum_{\forall x}\sum_{\forall y}[g(x,y)-\mu_{g(x,y)}]^2 p(x,y) & \text{si } x,y \text{ son discretas} \end{cases}$$

donde $\mu_{g(x,y)} = E(g(X,Y))$.

En forma compacta:

$$V(g(X,Y)) = E[g(x,y)-\mu_{g(x,y)}]^2 = E(g(X,Y))^2 - \{\mu_{g(x,y)}\}^2.$$

Modelos de Probabilidad Conjunta Continuos

Calculemos, por ejemplo, $V(X+Y)$

$$\begin{aligned} V(X+Y) &= E[X+Y-\{E(X+Y)\}]^2 = E[X-E(X)+Y-E(Y)]^2 \\ &= E(X-E(X))^2 + E(Y-E(Y))^2 + 2E[(X-E(X))(Y-E(Y))] \\ &= V(X)+V(Y)+2Cov(X,Y). \end{aligned}$$

Definiendo

$$\begin{aligned} Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \quad \text{la covarianza entre X y Y} \\ &= E(XY)-E(X)E(Y). \end{aligned}$$

Modelos de Probabilidad Conjunta Continuos

¿Qué mide este término generado como el producto cruzado de las diferencias de las desviaciones de los valores con respecto a sus respectivas medias?

La mayoría de nosotros hemos visto este término antes. Veamos que pasa con éste cuando asumimos que X y Y son **independientes**

$$E[(X - E(X))(Y - E(Y))] = E(X - E(X))E(Y - E(Y)) = 0,$$

esto es, si las variables son independientes entonces la *covarianza* es cero. Esto nos dice que este término nos da una forma de medir dependencias entre X y Y.

Por desgracia, sólo mide un tipo de dependencias: *las dependencias lineales*. **Esto nos dice que si $Cov(X, Y) = 0$, no necesariamente es cierto que las variables sean independientes, simplemente dice que no hay dependencia lineal, pero puede ser de otro tipo.**

Modelos de Probabilidad Conjunta Continuos

Por ejemplo si consideramos las variables X y $Y = X^2$, *claramente existe una relación cuadrática entre X y Y.*

Resultado: Si X y Y son variables aleatorias independientes

$$Cov(X, Y) = 0 \qquad y \qquad Var(X \pm Y) = Var(X) + Var(Y),$$

de otra forma

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y).$$

Ya antes habíamos invocado este resultado sin demostración. Estos resultados se extienden directamente a n variables aleatorias discretas o continuas. Ver Texto [0] pág. 276-277 Teorema 5.2.

Modelos de Probabilidad Conjunta Continuos

Como con medias y varianzas, en donde tenemos sus formas correspondientes en términos de valores muestrales, podemos definir lo que se conoce como la *covarianza muestral*.

Sea $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ *una muestra aleatoria*, entonces *la covarianza muestral* se define como:

$$d = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Notemos que la muestra ahora, está formada de parejas de valores.

Modelos de Probabilidad Conjunta Continuos

En general es difícil darse una idea adecuada del nivel de dependencia lineal a partir de los valores de covarianzas dado que estos dependen de la escala en que manejamos la información. Por esa razón se ideó un índice ρ conocido como coeficiente de correlación:

$$\rho = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = E\left(\frac{X - E(X)}{\sigma_x}\right)\left(\frac{Y - E(Y)}{\sigma_y}\right).$$

La última igualdad implica que el coeficiente de correlación se puede pensar como la covarianza de las variables estandarizadas.

Modelos de Probabilidad Conjunta Continuos

Esta definición tiene varias ventajas:

1. ρ es invariantes bajo cambios de escala;
2. $-1 \leq \rho \leq 1$;
3. Es 1 ó -1 sólo cuando la relación lineal es exacta;
4. $\rho = 0$ (esto es Cov(X,Y)=0) indica que no hay dependencia lineal.

El correspondiente coeficiente de correlación muestral se define como

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}.$$

Modelos de Probabilidad Conjunta Continuos

Algunas ilustraciones del comportamiento de este coeficiente:

Modelos de Probabilidad Conjunta Continuos

$Y_1 - Y_2$ el tiempo que pasa el cliente en la ventanilla. ¿Qué tanta variabilidad hay en estos tiempos?

$$V(Y_1 - Y_2) = V(Y_1) + V(Y_2) - 2Cov(Y_1, Y_2)$$

Ya conocemos las varianzas, porque corresponden a las de las distribuciones marginales

$$V(Y_1) = 2 \cdot 1^2 = 2 \qquad y \qquad V(Y_2) = 1$$

Mostraremos todos los cálculos sólo a manera de ejercicio de las definiciones que hemos dado hasta aquí.

Para calcular estas varianzas y covarianzas necesitamos obtener

$$E(Y_1^2), \qquad E(Y_2^2) \qquad y \qquad E(Y_1 Y_2).$$

Modelos de Probabilidad Conjunta Continuos

$$E(Y_1^2) = \int_{-\infty}^{+\infty} y_1^2 f_{Y_1}(y_1)dy_1 = \underbrace{\int_0^{\infty} y_1^3 e^{-y_1} dy_1}_{\substack{u = y_1^3 \quad dv = e^{-y_1}dy_1 \\ du = 3y_1^2 dy_1 \quad v = -e^{-y_1}}}$$

$$= -y_1^3 e^{-y_1}|_0^{\infty} + \int_0^{\infty} 3y_1^2 e^{-y_1} dy_1$$

$$= \underbrace{\int_0^{\infty} 3y_1^2 e^{-y_1} dy_1}_{\substack{u = 3y_1^2 \quad dv = e^{-y_1}dy_1 \\ du = 6y_1 dy_1 \quad v = -e^{-y_1}}} = -3y_1^2 e^{-y_1}|_0^{\infty} + \underbrace{\int_0^{\infty} 6y_1 e^{-y_1} dy_1}_{\substack{u = 6y_1 \quad dv = e^{-y_1}dy_1 \\ du = 6dy_1 \quad v = -e^{-y_1}}}$$

$$= -6y_1 e^{-y_1}|_0^{\infty} + \int_0^{\infty} 6e^{-y_1} dy_1$$

$$= -6e^{-y_1}|_0^{\infty} = 6$$

Modelos de Probabilidad Conjunta Continuos

$$E(Y_2^2) = \underbrace{\int_0^\infty y_2^2 e^{-y_2} dy_2}_{\substack{u = y_2^2 \qquad dv = e^{-y_2} dy_2 \\ du = 2y_2 dy_2 \qquad v = -e^{-y_2}}}$$

$$= -y_2^2 e^{-y_2}|_0^\infty + \underbrace{\int_0^\infty 2y_2 e^{-y_2} dy_2}_{\substack{u = 2y_2 \qquad dv = e^{-y_2} dy_2 \\ du = 2dy_2 \qquad v = -e^{-y_2}}}$$

$$= -2y_2 e^{-y_2}|_0^\infty + \int_0^\infty 2e^{-y_2} dy_2$$

$$= -2e^{-y_2}|_0^\infty = 2$$

Modelos de Probabilidad Conjunta Continuos

$$E(Y_1Y_2) = \int_0^\infty \underbrace{\int_{y_2}^\infty y_1 y_2 e^{-y_1} dy_1}_{\substack{u = y_1y_2 \quad dv = e^{-y_1}dy_1 \\ du = y_2 dy_1 \quad v = -e^{-y_1}}} dy_2$$

$$= \int_0^\infty \left[-y_1y_2e^{-y_1}\Big|_{y_2}^\infty + \int_{y_2}^\infty y_2 e^{-y_1} dy_1 \right] dy_2$$

$$= \int_0^\infty \{y_2^2 e^{-y_2} + [-y_2e^{-y_1}]_{y_2}^\infty\} dy_2$$

$$= \int_0^\infty (y_2^2 e^{-y_2} + y_2 e^{-y_2}) dy_2$$

$$= 2 + 1 = 3$$

$\therefore \quad V(Y_1) = 6 - 2^2 = 2 \qquad V(Y_2) = 2 - 1^2 = 1$

$\therefore \quad Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = 3 - 2 * 1 = 1$

$\therefore \quad V(Y_1 - Y_2) = 2 + 1 - 2(1) = 1$

Modelos de Probabilidad Conjunta Continuos

Esto es, se tiene una desviación estándar de minuto. De acuerdo a Chebyshev, al menos el 75 % de los tiempos estarán en el intervalo $(2 - 2(1), 2 + 2(1)) = (0, 4)$: *pero por las reglas empíricas*, sabemos que la probabilidad contenida en dos desviaciones estándar alrededor de la media es del orden del 90 y tantos %. La pregunta sería, ¿esa variación se puede reducir sin aumentar los costos, ni para ellos, ni para los clientes? ¿O, podemos vivir con ella, sin mayores consecuencias?

Nota: El coeficiente de correlación $\rho = \frac{1}{\sqrt{2}\sqrt{1}} = 0.71$

¿De qué otra forma podemos medir dependencias? Bueno, una forma no siempre reconocida como tal, pero ampliamente usada en lo que a sus consecuencias se refiere, es mediante los valores medios asociados a las distribuciones condicionales:

Modelos de Probabilidad Conjunta Continuos

Definición. Las distribuciones condicionales de $X|Y = y$ y $Y|X = x$ denotadas por $f_{X|y}(x)$ y $f_{Y|x}(y)$ se definen de la siguiente forma.
Para cada valor de y

$$f_{X|y}(x) = \frac{f(x,y)}{f(y)}$$

Es una función de densidad en X. Sus valores pueden depender (en el sentido de función matemática , de y, pero como este valor es dado, no representa un factor aleatorio). De hecho:

$$\int_{-\infty}^{+\infty} f_{X|y}(x)dx = \int_{-\infty}^{+\infty} \frac{f(x,y)}{f(y)}dx = \frac{1}{f(y)}\int_{-\infty}^{+\infty} f(x,y)dx = \frac{1}{f(y)}f(y) = 1$$

Análogamente, *para cada valor de x*

$$f_{Y|x}(y) = \frac{f(x,y)}{f(x)}$$

Modelos de Probabilidad Conjunta Continuos

Nota: de aquí podemos ver cómo se deriva el concepto de independencia: Y no modifica su comportamiento por saber como se comporta X:

$$f_{Y|x}(y) = \frac{f(x,y)}{f(x)} \text{ si X y Y son independientes}$$

$$\Rightarrow f(x,y) = f_{Y|x}(y)f(x) = f(x)f(y)$$

Del ejemplo del "fast food", si sequiere conocer cuál sería el comportamiento de los tiempos que el cliente toma en la fila, dado que el tiempo total debería ser de 2 minutos, tendríamos:

$$f_{Y_2|y_1}(y_2) = \frac{f(y_1,y_2)}{f(y_1)} = \frac{e^{-y_1}}{y_1 e^{-y_1}} = \frac{1}{y_1} \qquad 0 \leq y_2 \leq y_1 \leq \infty$$

Modelos de Probabilidad Conjunta Continuos

Notas:

1) si existen condiciones en el rango de la variable, éstas deben aparecer con las condicionales. No así cuando calculamos marginales, pues en ese caso, se *barre* sobre la otra variable, mientras que aquí, *condicionas* con la otra variable.

2) como recomendación general, calcula primera la condicional en forma genérica, y después evalúa en los valores particulares.

Modelos de Probabilidad Conjunta Continuos

Como $y_1 = 2$, sólo evaluamos y obtenemos la distribución condicional para estos tiempos :

$$f_{Y_2|y_1=2}(y_2) = \frac{1}{2} \qquad 0 \leq y_2 \leq 2$$

Esto es, la distribución resultante es una $U(0,2)$. *(No siempre resultan distribuciones conocidas)*

Bajo estas circunstancias es muy sencillo obtener su valor medio, $\mu_{Y_2|y_1=2} = 1$ *(1 minuto en promedio, dado que tarda exactamente 2 minutos, en tiempo total)*

Modelos de Probabilidad Conjunta Continuos

La notación para los valores esperados calculados a partir de las condicionales es:

- $E(Y|X=x) = \int_{-\infty}^{+\infty} y f_{Y|x}(y)dy$, o bien
- $E(X|Y=y) = \int_{-\infty}^{+\infty} x f_{X|y}(x)dx$

A $E(Y|X=x)$ *SE LE CONOCE COMO LA REGRESIÓN DE Y SOBRE X.*

Cuando $E(Y|X=x)$ es una función lineal, entonces se le llama la regresión lineal (teórica) de Y sobre X.

Cuando planteamos nuestros modelos, tales como $y = \mu_y + e$, deberíamos entender

$$\mu_{Y|x} = E(Y|X=x)$$

Modelos de Probabilidad Conjunta Continuos

En general es difícil obtener las distribuciones condicionales y de allí, sus valores esperados, los cuales además, casi nunca son funciones lineales de las x's. Entonces lo que planteamos a través de modelos como estos es simplemente una forma de aproximar al comportamiento medio de Y dada la infomación en X (que puede ser más de una variable).

Ejercicio: Calcula la regresión de Y_1 sobre $Y_2 = y_2 = 1$ en el ejemplo. Tendría ésta un significado valorable? explica tanto como puedas.

Lectura: Estudiar las definiciones y ejemplos sobre aplicaciones de las esperanzas condicionales en el libro de Texto. Capítulo 4, secciones 4.4-4.5

Modelos de Probabilidad Conjunta Continuos

Un ejemplo más: Sean X y Y dos variables aleatorias con función de densidad conjunta:

$$f(x,y) = \begin{cases} kx(y-x) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

a) Encontrar el valor de la constante k.

$$\Rightarrow 1 = \int_0^1 \int_{-x}^{x} kx(y-x)dydx = \int_0^1 kx(\frac{y^2}{2} - xy)|_{-x}^{x} dx$$

$$= \int_0^1 kx[(\frac{x^2}{2} - x^2) - (\frac{x^2}{2} + x)]dx$$

$$= \int_0^1 kx(\frac{-x^2}{2} - \frac{3x^2}{2})dx = \int_0^1 -2kx(x^2)dx = \int_0^1 -2kx^3 dx$$

$$= -2k[\frac{x^4}{4}]_0^1 = -\frac{2k}{4} = -\frac{k}{2} = 1. \qquad \text{por lo tanto } k = -2$$

Modelos de Probabilidad Conjunta Continuos

Y por lo tanto la función de densidad es:

$$f(x,y) = \begin{cases} -2x(y-x) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

o equivalentemente,

$$f(x,y) = \begin{cases} 2x(x-y) & 0 < x < 1 \quad -x < y < x \\ 0 & \text{en otra parte} \end{cases}$$

Modelos de Probabilidad Conjunta Continuos

Nota sobre marginales:

Para el caso de más de dos variables se define la marginal en forma análoga. Por ejemplo, si tenemos tres variables aleatorias X, Y y Z con función de probabilidad o de densidad $f(x,y,z) \Rightarrow \sum_y \sum_z f(x,y,z)$ ó $\int_y \int_z f(x,y,z)dxdy$ según sea el caso. Aquí es posible definir también la marginal conjunta de dos de las tres variables, por ejemplo X y Y . En ese caso la marginal conjunta se define como:

$$f_{X,Y}(x,y) = \sum_z f(x,y,z) \qquad o \qquad \int_z f(x,y,z)dz$$

Estamos listos para ver dos modelos particulares de los cuales encontramos multitud de aplicaciones: El modelo Multinomial y la Normal Bivariada.

El modelo Normal Bivariado

El modelo normal bivariado

Si $X \sim N(\mu_x, \sigma_x^2)$ y $Y \sim N(\mu_y, \sigma_y^2)$ son independientes, entonces

$$\begin{aligned} f_{(X,Y)}(x,y) &= f_X(x) f_Y(y) \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \\ &= \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}, \end{aligned}$$

este es un caso particular de la distribución conjunta Normal Bivariada (caso particular).

El modelo normal bivariado

Recordemos que si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes (una muestra aleatoria, por ejemplo) y definimos $Y = \sum_{i=1}^{n} X_i$, entonces

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{t(X_1+X_2+\ldots+X_n)}) \\ &= E(e^{tX_1+tX_2+\ldots+tX_n}) \\ &= E(e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_n}) \\ &= E(e^{tX_1})E(e^{tX_2}) \ldots E(e^{tX_n}) \text{ por independencia} \\ &= M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t) \end{aligned}$$

Más aún, si $X_1, X_2, \ldots, X_n$ tienen la misma distribución

$$M_Y(t) = [M_X(t)]^n$$

El modelo normal bivariado

Nota: Además es fácil verificar que si $W = aX + bY$ entonces, (una combinación lineal de dos variables normales independientes)

$$\begin{aligned} M_W(t) &= E[e^{(aX+bY)t}] = E(e^{atX})E(e^{btY}) = M_X(at) \cdot M_Y(bt) \\ &= [e^{a\mu_x t + \frac{1}{2}(a^2\sigma_x^2)t^2}][e^{b\mu_y t + \frac{1}{2}(b^2\sigma_y^2)t^2}] \\ &= e^{(a\mu_x + b\mu_y)t + \frac{1}{2}(a^2\sigma_x^2 + b^2\sigma_y^2)t^2}, \end{aligned}$$

i.e. $W \sim Normal(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$.

Esto quiere decir que cualquier combinación lineal de variables aleatorias normales independientes también sigue una distribución normal.

No sólo somos capaces de saber cuál es la media y la varianza de la suma, sino que además nos permite conocer su distribución!!. Este es un resultado particularmente útil para las distribuciones muestrales.

El modelo normal bivariado

Ejemplo

Las resistencias de ciertos resistores tienen una distribución normal con media de 100 ohms y desviación estándar de 10 ohms. Dos de tales resistencias se conectan en serie resultando que la resistencia total del circuito es la suma de las resistencias individuales. Determinar la probabilidad de que la resistencia total sea mayor que 220 ohms.

Solución: Tenemos que

$$R_i \sim N(100, 10) \quad i = 1, 2 \text{ son independientes}$$

y, por tanto,

$$M_{R_i(t)} = e^{100t + \frac{1}{2}100t^2}.$$

El modelo normal bivariado

Entonces $R_T = R_1 + R_2$ y

$$M_{R_T}(t) = (e^{100t+\frac{1}{2}100t^2})^2 = e^{200t+\frac{1}{2}200t^2},$$

lo que nos dice que $R_T \sim N(200, \sigma^2_{R_T} = 200)$. Esto implica

$$P(R_T > 220) = P(Z > \frac{220-200}{\sqrt{200}}) = P(Z > 1.41) = 0.0793.$$

Lectura: Ver Montgomery D.C.(1985) Introduction to Statistical Quality Control, pág. 396-403.

El modelo normal bivariado

Ejercicio

En un área determinada, cierto material es seleccionado al azar y pesado en dos tiempos distintos. Sea $W =$ peso del material (verdadero) y X_1, X_2 las dos mediciones hechas. Uno puede pensar estas mediciones como

$$\begin{aligned} X_1 &= W + e_1 \\ X_2 &= W + e_2, \end{aligned}$$

donde e_1 y e_2 son los errores de medición. Supongamos que e_1 y e_2 son independientes entre sí e independientes de W y que $V(e_1) = V(e_2) = \sigma_e^2$. Expresar a ρ, el coeficiente de correlación entre X_1 y X_2, en términos de σ_W^2 (la varianza del peso verdadero) y σ_e^2; interpreta el resultado.

El modelo normal bivariado

Ahora consideraremos el caso siguiente: X y Y no son variables independientes pero (X, Y) tiene una distribución Normal Bivariada.

En este caso la función de densidad conjunta de las variables es

$$f_{X,Y}(x,y) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}.$$

El modelo normal bivariado

Figura: Gráficas de Normales Bivariadas, con dependencia cero, positiva y negativa.

El modelo normal bivariado

- Si cada una sigue una distribución normal, y no son independientes, no podemos asegurar que la distribución conjunta sea normal bivariada.
- De la expresión anterior es inmediato verificar que si $\rho = 0$ entonces

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

  y por lo tanto X y Y son independientes. Esto es, si X y Y tienen una distribución normal bivariada y la covarianza (ó correlación) es cero entonces X y Y deben ser independientes.
- Bajo normalidad bivariada, la única forma de dependencia permitida es la lineal.

Estos resultados parecen intrascendentes, pero son la base de la mayoría de los modelos que trabajamos en regresión o diseño de experimentos, y la razón de la popularidad de la distribución normal.

El modelo normal bivariado

La distribución marginal de X es

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$
$$= \int_{-\infty}^{+\infty} \frac{e^{-\frac{1}{2(1-\rho^2)}\left\{(\frac{x-\mu_X}{\sigma_X})^2+(\frac{y-\mu_Y}{\sigma_Y})^2-2\rho(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y})\right\}}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}dy$$

Considerando las transformaciones $u = \frac{x-\mu_X}{\sigma_X}$ y $v = \frac{y-\mu_Y}{\sigma_Y} \Rightarrow dv = \frac{1}{\sigma_Y}dy$, la notación se simplifica,

El modelo normal bivariado

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_x(1-\rho^2)} e^{-\frac{1}{2(1-\rho^2)}u^2} e^{-\frac{1}{2(1-\rho^2)}\{v^2-2\rho uv\}} dv \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{u^2}{2(1-\rho^2)}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\{(v-\rho u)^2-\rho^2u^2\}} dv \\
&\quad \text{(se completa el cuadrado en el exponente)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{u^2}{(1-\rho^2)}-\frac{\rho^2u^2}{(1-\rho^2)}\right)} \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{(v-\rho u)^2}{1-\rho^2}} dv}_{N(\rho u,(1-\rho^2))} \\
&\quad \text{(se reconoce la distribución para evitar la integración)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}u^2} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}
\end{aligned}
$$

$\therefore\ X \sim N(\mu_x, \sigma_x^2)$

El modelo normal bivariado

Similarmente, $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} \therefore\ Y \sim N(\mu_y, \sigma_y^2)$

En resumen

Normal Bivariada $\Longrightarrow$ Marginales Normales,
pero Marginales normales **no implican** Normal Bivariada (a menos que sean variables aleatorias normales independientes)

El modelo normal bivariado

Ahora, consideremos la distribución condicional de Y dado X

$$
\begin{aligned}
f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} &= \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{u^2}{2(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\{v^2-2\rho uv\}}}{\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}u^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{v^2-2\rho uv+\rho^2u^2\right\}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[\frac{v-\rho u}{\sqrt{1-\rho^2}}\right]^2}.
\end{aligned}
$$

El modelo normal bivariado

Es decir,

$$f_{Y|X}(y|x) = \frac{\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y - \left\{\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right\}\right]^2\right\}}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}},$$

(truco algebraico para reconocer la distribución)

lo cual implica que

$$(Y|X=x) \sim N\left(\mu_{Y|X} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x), \sigma^2_{Y|X} = \sigma_y^2(1-\rho^2)\right),$$

y, por tanto,

$$E(Y|X=x) = \rho\frac{\sigma_y}{\sigma_x}x + (\mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x).$$

Regresión

Regresión

- La distribución normal multivariada tuvo sus comienzos con estudios de la distribución normal bivariada. Tales estudios, se cree, iniciaron a mediados del siglo XIX, y a partir de allí, hubo un desarrollo dramático, fundamentalmente cuando Galton (1888) publicó su trabajo sobre los usos del análisis de la correlación en genética.
- Como Pearson notó, en 1885 Galton había terminado la teoría de la correlación normal bivariada pero, dada su personalidad, *(era muy modesto y a través de su vida subestimo sus propias capacidades matemáticas)*, no se percató en forma inmediata la importancia de la función de densidad normal bivariada y por ende, todas las consecuencias que esto tendría en el desarrollo de los modelos estadísticos.

Regresión

- Consecuentemente, fue Pearson (1896) quién dio la fórmula matemática definitiva de la distribución normal bivariada.
- El desarrollo de la teoría de la distribución normal multivariada se originó principalmente de los estudios del análisis de regresión y del análisis de correlación múltiple y parcial.
- Veamos con software moderno, cuál fue el problema que estudio Galton.

Regresión

- Sea (X, Y) una variable aleatoria continua, bidimensional que toma todos los valores en el plano euclidiano esto es, $-\infty \leq x \leq +\infty, -\infty \leq y \leq +\infty$. Decimos que (X, Y) tiene una distribución normal bivariada si su fdp conjunta está dada por la siguiente expresión:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

- Notemos que se tienen las siguientes restricciones para los parámetros: $-\infty \leq \mu_x \leq +\infty$, $-\infty \leq \mu_y \leq +\infty$; $\sigma_x > 0$; $\sigma_y > 0$; $-1 \leq \rho \leq 1$

Regresión

Recordemos que

- El parámetro ρ es el coeficiente de correlación entre X y Y;
- Se puede observar que si $\rho = 0$, la fdp conjunta de (X, Y) puede factorizarse y, por tanto, X y Y son independientes.
- Entonces en el caso de la distribución normal bivariada, se puede ver que:

  correlación cero **sí y sólo sí** son independientes.

Regresión

- Si la variable aleatoria (X, Y) sigue una distribución normal bivariada, los momentos de la función de distribución condicional de Y, dado $X = x$, son normal con media

$$\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

y varianza

$$\sigma_y^2(1 - \rho^2)$$

- Notemos que la media condicional puede escribirse como $\beta_0 + \beta_1 x$

Regresión

- Esto demuestra que las medias son funciones lineales en la variable que condiciona
- También demuestra que la varianza de la distribución condicional se reduce en proporción a $(1 - \rho^2)$.
- Esto es, si ρ está próxima a cero, la varianza condicional es esencialmente la misma que la varianza incondicional, mientras que si ρ está próximo a ± 1, la varianza condicional está próxima a cero.

Regresión

- LA ÚNICA FORMA DE DEPENDENCIA POSIBLE ES LINEAL

$$f_{y|x}(y \mid x) = \frac{1}{\sqrt{2\pi\sigma_y^2(1-\rho^2)}} \times \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right]^2\right\}$$

- Se inicia con la distribución conjunta y se derivan estos resultados. Esto se puede generalizar para la normal multivariada en forma directa.

Regresión

Para el caso general

- Si $\Sigma_{22} > 0$, entonces la distribución condicional de $\mathbf{Y}_1$ dado $\mathbf{Y}_2$ es

$$Y_1 | Y_2 \sim N_n(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Los objetos de estudio:

- $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2)$
- $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Regresión

- La respuesta a este problema fue :
  El mejor predictor es **la esperanza condicional** de Y_1 dado Y_2
- Este predictor puede ser lineal o no.
- Solo si sabemos que las variables tienen una distribución normal multivariada, podemos asegurar que es lineal.
- El problema entonces es conocer éstas distribuciones conjuntas, cosa que en la práctica resulta bastante difícil

Regresión

- El problema se replantó de la siguiente forma:

  Encontrar una función lineal de las Y_2 que fuese el mejor predictor de Y_1, sin imponer la restricción de normalidad multivariada, pero (Y_1, Y_2) deben tener una distribución conjunta.
- $\min_{f(y_2) lineal} \mathrm{E}\left(y_1 - f(y_2) \right)^2$
- La respuesta bajo esta restricción de buscar sólo entre las funciones lineales, sorprendentemente nos da la misma solución que en el caso normal multivariado, esto es:
  - $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2)$
    La media es lineal
  - $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$
    La misma estructura del error de predicción

El modelo normal bivariado

Para $t = (t_1, t_2)$, la función generatriz de momentos conjunta del vector $W = (X, Y)$ está dada por

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E(e^{t'W}) = E(e^{t_1 x + t_2 y}) \\ &= \exp\left\{ t_1\mu_x + t_2\mu_y + \frac{1}{2}(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2) \right\} \end{aligned}$$

Ejercicio: Demostrar la igualdad anterior y entregar su desarrollo junto con la Tarea 4.

Ejercicio: Sean X y Y v.a. Mostrar que $E(Y|X)$ es el mejor predictor de Y entre todas las funciones de X.

El modelo normal bivariado

Ejercicio

1. *Determina qué variables aleatorias corresponden a las generatrices* $M_{X,Y}(t,0)$ *y* $M_{X,Y}(0,t)$
2. *Si* X *y* Y *tienen una distribución normal bivariada, y* $U = X + Y, W = X - Y$.
   Determina el coeficiente de correlación de U *y* W.

El modelo normal bivariado

Ejemplo

La distribución de dos elementos de un tipo de material (en gramos) sigue una distribución normal bivariada, donde

$$\mu_x = 4, \mu_y = 6 \; \sigma_x^2 = \sigma_y^2 = 1 \; y \; Cov(X, Y) = 0.8$$

¿Cuál es el valor esperado del elemento Y si se observan 6 grs. del elemento X?

Solución:

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y} = \frac{0.8}{1 \cdot 1} = 0.8$$

$\Longrightarrow E(Y|X) = E(Y|X = 6) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x) = 6 + 0.8(\frac{1}{1})(6 - 4) = 6 + 0.8(2) = 7.6$ grs.

El modelo Multinomial,
Cap 3. Secciones 3.9 - 3.10

El modelo Multinomial

La distribución multinomial es una generalización de la distribución binomial. Este tipo de modelos es adecuado para encuestas de opinión, investigación de mercados, inspección industrial, etc. Además tiene otro muchos usos para generar pruebas estadísticas como las de bondad de ajuste, entre otras.

El modelo Multinomial

Un experimento multinomial consiste en:

1. n pruebas idénticas;
2. El resultado de cada observación puede ser clasificado en una de k-clases o casillas;
3. La probabilidad de ser clasificado en la casilla i es p_i $i = 1, 2, \ldots, k$, y permanece constante de observación a observación. Más aún $p_1 + p_2 + ... + p_k = 1$;
4. Las pruebas son independientes;
5. Las variables aleatorias son $X_1, X_2, \ldots, X_k$ donde $X_i =\#$ de observaciones en la casilla i, $i = 1, 2, \ldots, k$. Nótese que $X_1 + X_2 + ...X_k = n$.

El modelo Multinomial

La función de probabilidad conjunta de $X_1, X_2, ..., X_k$ está dada por

$$p(x_1, x_2, ..., x_K) = \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}$$

donde $\sum_{i=1}^{k} p_i = 1$ y $\sum_{i=1}^{k} x_i = n$.

Nota: Una generalización para multinomios del desarrollo del binomio de Newton está dada por

$$\sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k} = (p_1 + p_2 + ... + p_k)^n = 1^n = 1.$$

Esto asegura que la suma de la función de probabilidad sobre todos los valores posibles de las variables es 1.

El modelo Multinomial

Resultado

La distribución marginal de X_i, para $i = 1, 2, \ldots, k$, es Binomial(n, p_i).

Si sólo queremos la distribución de X_i, i.e. el # de éxitos en la casilla i, podemos pensar que sólo hay dos tipos de casillas: i o no i, con p_i como la probabilidad asociada a la casilla i y $1 - p_i$ como la probabilidad asociada las casillas complementarias.

Para ilustrar conceptos, la demostración se hará sólo para el caso $k = 3$. El resultado es intuitivamente fácil de entender y manejar.

El modelo Multinomial

Sea (X, Y, Z) un vector aleatorio con probabilidad conjunta:

$$f(x, y, z) = \frac{n!}{x!y!z!} p_X^x p_Y^y p_Z^z,$$

con $x + y + z = n$. Entonces, tenemos que

$$\begin{aligned} x &= 0, 1, \ldots, n; \\ \Rightarrow y &= 0, 1, \ldots, n - x; \\ \Rightarrow z &= 0, 1, \ldots, n - x - y. \end{aligned}$$

El modelo Multinomial

Ahora

$$
\begin{aligned}
f_X(X) &= \sum_{y=0}^{n-x}\sum_{z=0}^{n-x-y}\frac{n!}{x!y!z!}p_X^x p_Y^y p_Z^z \\
&= \sum_{y=0}^{n-x}\frac{n!}{x!y!(n-x-y)!}p_X^x p_Y^y (1-p_X-p_Y)^{n-x-y} \\
&= \frac{n!}{x!(n-x)!}p_X^x(1-p_X)^{n-x}\sum_{y=0}^{n-x}\frac{(n-x)!}{y!(n-x-y)!}p_Y^y\frac{(1-p_X-p_Y)^{n-x-y}}{(1-p_X)^{n-x}} \\
&= \frac{n!}{x!(n-x)!}p_X^x(1-p_X)^{n-x}\cdot \\
&\quad \underbrace{\sum_{y=0}^{n-x}\frac{(n-x)!}{y!(n-x-y)!}\left(\frac{p_Y}{1-p_X}\right)^y\left(1-\frac{p_Y}{1-p_X}\right)^{n-x-y}}_{1 \text{ por ser la suma de una v.a } \textit{Binomial}\left(n-x,\ \frac{p_Y}{1-p_X}\right)} \\
&= \frac{n!}{x!(n-x)!}p_X^x(1-p_X)^{n-x} \\
&= \binom{n}{x}p_X^x(1-p_X)^{n-x}.
\end{aligned}
$$

El modelo Multinomial

El mismo argumento se utiliza para X y Z.

De este resultado se sigue entonces que

$$E(X_i) = np_i, \qquad V(X_i) = np_i q_i.$$

Más aún, notemos que el número de resultados por casilla no son independientes y

$$Cov(X_i, X_j) = -np_i p_j$$

Para explicar esto, usaremos los siguientes argumentos.

El modelo Multinomial

1. Supongamos que $X = (X_1, \ldots, X_K)$ una variable aleatoria multinomial con distribución

$$\mathrm{P}(X = x) \;=\; \mathrm{P}(X_1 = x_1, \ldots, X_K = x_K) \;=\; \frac{n!}{x_1! \cdots, x_K!}\, p_1^{x_1} \cdots p_K^{x_K}$$

donde $x_1 + \cdots + x_K = n$.

2. Consideremos el caso específico de una variable multinomial como la de arriba, pero con una sola tirada ($n = 1$):

$$\mathrm{P}(W = w) \;=\; \mathrm{P}(W_1 = w_1, \ldots, W_K = w_K) \;=\; p_1^{w_1} \cdots p_K^{w_K}$$

donde $w_1 + \cdots + w_K = 1$. Como es una sola tirada, es claro que si, por ejemplo, ocurre la primera casilla ($w_1 = 1$) entonces todas las otras casillas son cero. También, cada casilla (i.e. W_i) es una Bernoulli, con probabilidad de éxito igual a p_i, o sea que $\mathrm{E}(W_i) = p_i$ y $\mathrm{Var}(W_i) = p_i(1 - p_i)$.

El modelo Multinomial

Además, la variable W además cumple:

- Media de W (usando notación de vectores como vectores columna)

$$\mathrm{E}(W) = \mathrm{E}\begin{bmatrix} W_1 \\ \vdots \\ W_K \end{bmatrix} = \begin{bmatrix} \mathrm{E}(W_1) \\ \vdots \\ \mathrm{E}(W_K) \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix}$$

- Varianza de W (i.e. la matriz de varianzas y covarianzas) (o, simplemente, la matriz de covarianza de W)

$$\mathrm{Var}(W) = \begin{bmatrix} \mathrm{Var}(W_1) & \mathrm{Cov}(W_1, W_2) & \cdots & \mathrm{Cov}(W_1, W_K) \\ \mathrm{Cov}(W_2, W_1) & \mathrm{Var}(W_2) & \cdots & \mathrm{Cov}(W_2, W_K) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(W_K, W_1) & \mathrm{Cov}(W_K, W_2) & \cdots & \mathrm{Var}(W_K) \end{bmatrix}.$$

El modelo Multinomial

Los elementos diagonales de la matriz anterior sabemos que son $\mathrm{Var}(W_i) = p_i(1-p_i)$. Ahora, los elementos fuera de la diagonal son

$$\mathrm{Cov}(W_i, W_j) = \mathrm{E}(W_i W_j) - \mathrm{E}(W_i)\mathrm{E}(W_j) = 0 - p_i p_j.$$

La razón por la cual $\mathrm{E}(W_i W_j) = 0$ es la siguiente: el producto $W_i W_j$ sólo puede tomar el valor 0, pues (W_i, W_j) puede valer (0,0), (1,0), (0,1) y en cada caso $W_i W_j = 0$, pero nunca puede ser (1,1), pues se trata de una sola tirada.

Entonces

$$\mathrm{Var}(W) = \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_K \\ -p_2 p_1 & p_2(1-p_2) & \cdots & -p_2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_K p_1 & -p_K p_2 & \cdots & p_K(1-p_K) \end{bmatrix}.$$

El modelo Multinomial

3. Finalmente, si X es una variable aleatoria multinomial con n tiradas (independientes), entonces X se puede escribir como la suma de n multinomiales de una sola tirada

$$X = Z_1 + Z_2 + \cdots + Z_n$$

donde cada Z_j es como la W del punto anterior.

Entonces,

- Media de X

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{E}(Z_1 + \cdots + Z_n) = \mathrm{E}(Z_1) + \cdots + \mathrm{E}(Z_n) \\ &= \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} + \cdots + \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} = \begin{bmatrix} np_1 \\ \vdots \\ np_K \end{bmatrix}. \end{aligned}$$

El modelo Multinomial

- Varianza de X

$$\mathrm{Var}(X) = \mathrm{Var}(Z_1 + \cdots + Z_n) = \mathrm{Var}(Z_1) + \cdots + \mathrm{Var}(Z_n)$$

(varianza de una suma es la suma de las varianzas bajo independencia), entonces

$$\mathrm{Var}(X) = \begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_K \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_K \\ \vdots & \vdots & \ddots & \vdots \\ -np_Kp_1 & -np_Kp_2 & \cdots & np_K(1-p_K) \end{bmatrix}.$$

El modelo Multinomial

- Por definición, la función generatriz de momentos del vector $X = (X_1, X_2, \ldots, X_K)$ está dada por

$$\begin{aligned} M_{X_1,X_2,\ldots,X_K}(t) &= E(e^{t'X}) \\ &= E(e^{t_1X_1+t_2X_2+\ldots+t_kX_k}), \end{aligned}$$

  para $t = (t_1, \ldots, t_K)$.
- Notemos que t tiene la misma dimensión que $X = (X_1, \ldots, X_K)$.
- También es factible tomar derivadas cruzadas, las cuales al evaluarse en cero nos dan momentos cruzados que sirven para formar las covarianzas, esto es, nos dan $E(X^iX^j)$.

El modelo Multinomial

La generatriz de momentos de la Multinomial, usando el teorema de Newton para multinomio

$$\begin{aligned} M_{X_1,X_2,\ldots,X_K}(t_1,t_2,\ldots,t_K) &= E(e^{t_1X_1+t_2X_2+\ldots+t_kX_k}) \\ &= (p_1e^{t_1}+p_2e^{t_2}+\ldots+p_ke^{t_k})^n \end{aligned}$$

Así, por ejemplo, si tomamos

$$\begin{aligned} M_{X_1,X_2,\ldots,X_K}(t_1,0,\ldots,0) &= (p_1e^{t_1}+p_2e^{0}+\ldots+p_ke^{0})^n \\ &= (p_1e^{t_1}+p_2+\ldots+p_k)^n \\ &= (p_1e^{t_1}+(1-p_1))^n \\ &= (p_1e^{t_1}+q_1)^n, \end{aligned}$$

que es la función generatriz de momentos de una $Binomial(n,p_1)$. Esto certifica el hecho de que las marginales son binomiales.

El modelo Multinomial

También se pueden obtener otros resultados , por ejemplo

$$\begin{aligned} M_{X_1,X_2,\ldots,X_K}(t,t,0,\ldots,0) &= (p_1e^t + p_2e^t + p_3 \ldots + p_k)^n \\ &= ((p_1+p_2)e^t + (1-p_1-p_2))^n \end{aligned}$$

la cual corresponde a la función generatriz de momentos de una v.a. $Binomial(n, p_1 + p_2)$.

Evaluar en la misma t, equivale a juntar las clasificaciones, a sumar las variables. En la Teoría de la Probabilidad esto implica la convolución de las variables aleatorias.

El modelo Multinomial

Ejemplo

Se fabrica una barra de un largo específico. Supongamos que el largo verdadero X (en pulgadas) sigue una distribución uniforme $[10, 12]$. *Si se fabrican 10 de tales barras, calcular la probabilidad de obtener 5 de longitud menor a 10.5 pg. y 2 de longitud mayor que 11.8 pg.*

Definamos:

$$A_1 = \{X < 10.5\}, A_2 = \{10.5 \leq X \leq 11.8\}, A_3 = \{X > 11.8\}$$

y las variables

$$X_1 = \# \text{ de barras de la clase } A_1$$
$$X_2 = \# \text{ de barras de la clase } A_2$$
$$X_3 = \# \text{ de barras de la clase } A_3$$

El modelo Multinomial

Ejemplo

Se fabrica una barra de un largo específico. Supongamos que el largo verdadero X (en pulgadas) sigue una distribución uniforme $[10, 12]$. *Si se fabrican 10 de tales barras, calcular la probabilidad de obtener 5 de longitud menor a 10.5 pg. y 2 de longitud mayor que 11.8 pg.*

Definamos:

$$A_1 = \{X < 10.5\}, A_2 = \{10.5 \leq X \leq 11.8\}, A_3 = \{X > 11.8\}$$

y las variables

$$X_1 = \# \text{ de barras de la clase } A_1$$
$$X_2 = \# \text{ de barras de la clase } A_2$$
$$X_3 = \# \text{ de barras de la clase } A_3$$

El modelo Multinomial

De aquí $p_1 = 0.25$, $p_2 = 0.65$ y $p_3 = 0.1$. Estas probabilidades se calculan a partir de la distribución Uniforme que modela la longuitud de las barras. Por ejemplo, para p_1 tenemos que

$$p_1 = \int_{10}^{10.5} \frac{1}{2} dx = \left[\frac{1}{2}x\right]_{10}^{10.5} = \frac{1}{2}(10.5 - 10) = 0.5\frac{1}{2} = 0.25.$$

Ahora obtenemos

$$P(X_1 = 5, X_2 = 3, X_3 = 2) = \frac{10!}{5!3!2!}(0.25)^5(0.65)^3(0.1)^2 = 0.0067583.$$

El modelo Multinomial

Una propiedad importante de la distribución Multinomial es que, *cualquier variable aleatoria, continua o discreta* puede ser reducida a este modelo, particionando su rango en k intervalos (sin traslape). Por ejemplo, si X es una v. a. continua:

$$(-\infty, c_1); [c_1, c_2); [c_2, c_3); ...[c_{k-1}, \infty)$$

con $p_i = \int_{c_{i-1}}^{c_i} f_X(x)dx \;\; i = 1, 2, ..., k$ donde $c_0 = -\infty$ y $c_k = \infty$.

Esto es útil para la pruebas de Bondad de Ajuste.

El modelo Multinomial

Ejercicio

La tasa de muertes por accidente de tráfico de un año reciente (fatalidades por 100 millones de millas recorridas) se listan para cada uno de los 50 estados de los Estados Unidos. Construyeron un histograma de estas observaciones usando como clases 2.0-2.9, 3.0-3.9, etc.

Suponiendo que la distribución de estas tasas puede ser aproximada por una distribución normal, determinaron las frecuencias esperadas para cada una de las clases. (Con $\mu = 5.3$ y $\sigma = 1.3$).

El modelo Multinomial

Tasa de Muertes	Frecuencia Observada	Frecuencia Esperada
< 2.9	1	1.8
3.0-3.9	4	5.7
4.0-4.9	16	12.2
5.0-5.9	15	14.9(*)
6.0-6.9	7	10.3
7-0-7.9	4	4.1
> 8.0	3	1.0
		$\sum = 50$

El modelo Multinomial

- Cada intervalo en la tabla es una de las clasificaciones posibles, las cuales son mutuamente excluyentes.
- Las 50 observaciones se asumen independientes en el sentido de que una no debería dar información sobre donde se clasificarán las otras.
- Las probabilidades se obtienen de la distribución que se asume que siguen los datos. Por ejemplo, para calcular la frecuencia esperada (*) se tiene que:

$$\begin{aligned} P(5 < X < 5.9) &\approx P\left(\frac{4.95 - 5.3}{1.3} \leq Z \leq \frac{5.95 - 5.3}{1.3}\right) \\ &= P(-0.27 \leq Z \leq 0.50) \\ &\approx P(Z \leq 0.50) - P(Z \leq -0.27) \\ &= 0.6915 - 0.3936 = 0.2979, \end{aligned}$$

$\Rightarrow (0.2979) \cdot 50 = 14.895 = 14.9.$

El modelo Multinomial

Como tenemos las frecuencias observadas para cada clasificación, lo que se hace ahora es compararlas, por ejemplo se usa el criterio siguiente:

$$\sum_{i=1}^{k} \frac{(\text{frec. observada - frec. esperada})^2}{\text{frec. esperada}},$$

esto nos da una medida de la discrepancia que nos permite decir si es suficientemente grande como para creer que los datos no puede provenir de esa distribución.

Estas ideas del cómo tomar decisiones las discutiremos más a fondo en el último capítulo, cuando hablemos de Pruebas de Hipótesis y alguno de ustedes expondrá el tema de Bondad de Ajuste.

Esperanzas Condicionales

Esperanzas condicionales

Supongamos que X y Y son v.a. con función de probabilidad conjunta $f_{X,Y}(x,y)$ y funciones de probabilidad marginales $f_X(x)$ y $f_Y(y)$. Sea g una función tal que $E[g(X)] < \infty$. Definimos la esperanza condicional de $g(X)$ dado $Y = y$ por la fórmula

$$E[g(X)|Y = y] = \sum_x g(x) f_{X|Y}(x|y) \qquad \text{si } f_Y(y) > 0,$$

y la esperanza condicional no está definida para valores y tales que $f_Y(y) = 0$. La ley de la probabilidad total para esperanzas condicionales es

$$E[g(X)] = \sum_y E[g(X)|Y = y] f_Y(y).$$

Esperanzas condicionales

La esperanza condicional $E[g(X)|Y = y]$ es una función de la variable real y, que denotaremos $\varphi(y)$. Si evaluamos esta función φ en la variable aleatoria Y obtenemos una nueva variable aleatoria $\varphi(Y)$, que denotamos $E[g(X)|Y]$:

$$E[g(X)|Y](\omega) = E[g(X)|Y = Y(\omega)].$$

Podemos ahora escribir la ley de la probabilidad total como

$$E[g(X)] = \sum_y E[g(X)|Y = y]f_Y(y) = E[E[g(X)|Y]].$$

Como la esperanza condicional de $g(X)$ dado $Y = y$ es la esperanza respecto a la densidad de probabilidad condicional $f_{X|Y}(x|y)$, las esperanzas condicionales se comportan en muchos aspectos como esperanzas ordinarias.

Esperanzas condicionales

Supongamos que $c \in \mathbb{R}$, g es una función tal que $E[|g(X)|] < \infty$, h es una función acotada y ν es una función en $\mathbb{R}^2$ tal que $E[|\nu(X,Y)|] < \infty$. La esperanza condicional cumple las siguientes propiedades:

1. $E[c_1 g(X_1) + c_2 g(X_2)|Y = y] = c_1 E[g(X_1)|Y = y] + c_2 E[g(X_2)|Y = y]$.
2. Si $g \geq 0$ entonces $E[g(X)|Y = y] \geq 0$.
3. $E[\nu(X,Y)|Y = y] = E[\nu(X,y)|Y = y]$.
4. $E[g(X)|Y = y] = E[g(X)]$ si X, Y son v.a. independientes.
5. $E[g(X)h(Y)|Y = y] = h(y)E[g(X)|Y = y]$.
6. $E[g(X)h(Y)] = \sum_y h(y)E[g(X)|Y = y]f_Y(y) = E[h(Y)E[g(X)|Y]]$.

Como consecuencia de 1, 5 y 6, obtenemos

- $E[c|Y = y] = c$.
- $E[h(Y)|Y = y] = h(y)$.
- $E[g(X)] = E[E[g(X)|Y]]$.

Esperanzas condicionales

En el caso en que X y Y son continuas, la esperanza condicional se define de forma análoga, solo cambian sumas por integrales. Si g una función tal que $E[g(X)] < \infty$, la esperanza condicional de $g(X)$ dado $Y = y$ por la fórmula

$$E[g(X)|Y = y] = \int g(x) f_{X|Y}(x|y) dx \qquad \text{si } f_Y(y) > 0.$$

Estas esperanzas condicionales satisfacen también las propiedades 1-5 que listamos anteriormente. La propiedad 6 es en este caso,

$$E[g(X)h(Y)] = E[h(Y)E[g(X)|Y]] = \int h(y) E[g(X)|Y = y] f_Y(y) dy$$

válida para cualquier h acotada y suponiendo $E[g(X)] < \infty$. El bullet 3 de las propiedades de la esperanza condicional en el caso discreto también se recupera en este caso, cambiando la suma por una integral.

Esperanzas condicionales

Caso continuo-discreto: Sean X y N v.a. con distribución conjunta donde N toma valores $0, 1, 2, \ldots$ La función de distribución condicional $F_{X|N}(x|n)$ de X dado que $N = n$ es

$$F_{X|N}(x|n) = \frac{P(X \leq x, N = n)}{P(N = n)} \qquad \text{si } P(N = n) > 0,$$

y la función de distribución condicional no está definida para otros valores de n. Es sencillo verificar que $F_{X|N}(x|n)$ es una f.d. en x para cada valor fijo de n para el cual esté definida.

Supongamos que X es continua y $F_{X|N}(x|n)$ es diferenciable en x para todo n con $P(N = n) > 0$. Definimos la densidad condicional de X dado $N = n$ por

$$f_{X|N}(x|n) = \frac{d}{dx} F_{X|N}(x|n).$$

Esperanzas condicionales

De nuevo, $f_{X|N}(x|n)$ es una densidad en x para los n para los cuales está definida.

Usando la ley de la probabilidad total obtenemos la densidad marginal de X,

$$f_X(x) = \sum_n f_{X|N}(x|n) f_N(n)$$

Supongamos que g es una función para la cual $E[|g(X)|] < \infty$. La esperanza condicional de $g(X)$ dado que $N = n$ se define por

$$E[g(X)|N = n] = \int g(x) f_{X|N}(x|n) dx.$$

Esta esperanza condicional satisface las propiedades anteriores y en este caso la ley de la probabilidad total es

$$E[g(X)] = \sum_n E[g(X)|N = n] f_N(n) = E[E[g(X)|N]].$$

Esperanzas condicionales

Caso discreto-continuo: Consideremos ahora un vector (N, X). Supongamos que X tiene una distribución continua de densidad $f_X(x)$ y, dado el valor x de X, N es discreta con función de probabilidad $f_{N|X}(n|x)$ para $n \geq 0$. Podemos pensar que X es un parámetro (aleatorio) de la distribución de N, y una vez conocido el valor de este parámetro la distribución de N está completamente determinada.

La función de probabilidad condicional de N dado X es

$$f_{N|X}(n|x) = P(N = n|X = x) = \frac{f_{N,X}(n, x)}{f_X(x)}$$

siempre que $f_X(x) > 0$, donde $f_{N,X}(n, x)$ es la densidad de probabilidad conjunta del vector (N, X). La función de distribución condicional correspondiente es

$$F_{N|X}(n|x) = \frac{1}{f_X(x)} \sum_{k=0}^{n} P(N = k|X = x) = \frac{1}{f_X(x)} = \sum_{k=0}^{n} f_{N|X}(k|x).$$

Modelos Mixtos.

Modelos Mixtos

Otras formas en que aparecen distribuciones con dependencias es a través de distribuciones compuestas y jerárquicas.

Describiremos algunas de estas situaciones y sus distribuciones asociadas.

Mas adelante en una de las presentaciones se hablará explícitamente de valores medios y variabilidades en modelos compuestos.

Modelos jerárquicos

Modelos jerárquicos

Modelos jerárquicos

Distribuciones cuyos parámetros son aleatorios.

Pensemos en los siguientes casos:

- **Caso 1:** En un plan de control de calidad se seleccionan diariamente 3 productos (la producción se puede considerar como suficientemente grande) y se observa el número de defectuosos. Sin embargo, la proporción de defectuosos p varía día con día, y después de estudiar su comportamiento se decide un modelo $Uniforme(0, 1)$.

  Para un día seleccionado al azar, ¿Cuál sería la probabilidad de encontrar dos productos defectuosos en la muestra?

Modelos jerárquicos

Modelos jerárquicos

Distribuciones cuyos parámetros son aleatorios.

Pensemos en los siguientes casos:

- **Caso 1:** En un plan de control de calidad se seleccionan diariamente 3 productos (la producción se puede considerar como suficientemente grande) y se observa el número de defectuosos. Sin embargo, la proporción de defectuosos p varía día con día, y después de estudiar su comportamiento se decide un modelo $Uniforme(0, 1)$.

  Para un día seleccionado al azar, ¿Cuál sería la probabilidad de encontrar dos productos defectuosos en la muestra?

Modelos jerárquicos

- **Caso 1 (continuación):** Notemos que, X es el número de defectuosos en la muestra de 3 dado p:

$$X|p \sim Binomial(n, p)$$

y

$$p \sim Uniforme(0, 1).$$

Entonces, si queremos saber cuál es la probabilidad de x defectuosos, debe calcularse para cualquier posible valor de p, esto es, nos estamos preguntando por **la probabilidad no-condicionada, la probabilidad conjunta.**

Modelos jerárquicos

- **Caso 1 (continuación):**

$$\begin{aligned} P(X=x) &= P(X=x, 0<p<1) \\ &= \int_0^1 f(x,p)dp \\ &= \int_0^1 \binom{3}{x} p^x (1-p)^{3-x} f(p)dp. \end{aligned}$$

Recordemos que p es uniforme en $(0,1)$ por lo que $f(p)$ toma el valor de 1 en ese intervalo.

Evaluando en $X=2$, verifica que la probabilidad deseada es de $1/4$.

Modelos jerárquicos

- **Caso 1 (continuación):**

$$
\begin{aligned}
P(X = x) &= \int_0^1 3p^2(1-p)dp \\
&= 3\left[\frac{p^3}{3} - \frac{p^4}{4}\right]_0^1 \\
&= \tfrac{3}{12}.
\end{aligned}
$$

Modelos jerárquicos

- **Caso 2:** Se sabe que el número de defectos por yarda, digamos Y, de cierta tela, sigue una distribución Poisson con parámetro λ, sin embargo, λ presenta un comportamiento aleatorio que a su vez puede ser modelado por la siguiente función:

$$f(\lambda) = \begin{cases} e^{-\lambda} & \lambda \geq 0 \\ 0 & \text{otro caso} \end{cases} .$$

¿Cuál es la distribución del número de defectos por yarda?

Modelos jerárquicos

- **Caso 2 (continuación):**

$$
\begin{aligned}
P(Y=y) &= P(Y=y, \lambda \geq 0) \\
&= \int_0^\infty f(y,\lambda)d\lambda \\
&= \int_0^\infty \underbrace{\frac{\lambda^y e^{-\lambda}}{y!}}_{P(y|\lambda)} \underbrace{e^{-\lambda}}_{f(\lambda)} d\lambda \\
&\quad \text{notemos que } y \text{ está fija en esta integral} \\
&= \int_0^\infty \frac{\lambda^y e^{-2\lambda}}{y!} d\lambda.
\end{aligned}
$$

Modelos jerárquicos

- **Caso 2 (continuación):** hacemos el cambio de variable $\beta = 2\lambda$,

$$
\begin{aligned}
P(Y = y) &= \frac{1}{2}\int_0^\infty \frac{(\beta/2)^y e^{-\beta}}{y!} d\beta \\
&= \frac{1}{2}\left(\frac{1}{2}\right)^y \int_0^\infty \frac{\beta^y e^{-\beta}}{y!} d\beta \\
&= \left(\frac{1}{2}\right)^{y+1}, \qquad y = 0, 1, 2, \ldots
\end{aligned}
$$

Modelos compuestos

Modelos compuestos

Modelos compuestos

Distribuciones de sumas de variables aleatorias.

Existen diversas situaciones donde las distribuciones condicionales toman un papel preponderante, como es el caso de la *suma de variables aleatorias*. Por ejemplo

- **Teoría de colas**. Si N es el número de clientes que llegan a solicitar un servicio en un periodo de tiempo y X_i es el tiempo de servicio requerido por el $i-$ésimo cliente, entonces la demanda total de tiempo de servicio es $X = X_1 + \cdots + X_N$.

  Observa que N es discreta y X se modela de manera continua.

Modelos compuestos

Modelos compuestos

Distribuciones de sumas de variables aleatorias.

Existen diversas situaciones donde las distribuciones condicionales toman un papel preponderante, como es el caso de la *suma de variables aleatorias*.

Por ejemplo

- **Teoría de colas**. Si N es el número de clientes que llegan a solicitar un servicio en un periodo de tiempo y X_i es el tiempo de servicio requerido por el $i-$ésimo cliente, entonces la demanda total de tiempo de servicio es $X = X_1 + \cdots + X_N$.

  Observa que N es discreta y X se modela de manera continua.

Modelos compuestos

Modelos compuestos

Distribuciones de sumas de variables aleatorias.

Existen diversas situaciones donde las distribuciones condicionales toman un papel preponderante, como es el caso de la *suma de variables aleatorias*. Por ejemplo

- **Teoría de colas**. Si N es el número de clientes que llegan a solicitar un servicio en un periodo de tiempo y X_i es el tiempo de servicio requerido por el $i-$ésimo cliente, entonces la demanda total de tiempo de servicio es $X = X_1 + \cdots + X_N$.

  Observa que N es discreta y X se modela de manera continua.

Modelos compuestos

- **Teoría de riesgos**. Considera un número total de N reclamos en una compañía de seguros durante una semana. Si X_i es el monto del $i-$ésimo reclamo, entonces $X = X_1 + \cdots + X_N$ es la responsabilidad total de la compañía.
  En este caso, N es discreta y X es continua.
- **Modelos de población**. Considera N como el número de plantas de cierta especie en una área específica. Sea X_i el número de semillas producidas por la $i-$ésima planta. Entonces $X = X_1 + \cdots + X_N$ es el número total de semillas producidas en el área.
  A diferencia de los dos ejemplos anteriores, tanto N como X son discretas.

Modelos compuestos

- **Biometría**. Se tiene un esquema de muestreo de un número N aleatorio de cierta especie animal. Sea X_i el peso del $i-$ésimo especimen capturado. Entonces $X = X_1 + \cdots + X_N$ es el peso total capturado.

  De nuevo, N es discreta y X es continua.

Discutamos el caso donde tenemos una variable aleatoria discreta N y una variable aleatoria continua X.

Suponemos que las X_i's son independientes de N.

Modelos compuestos

- **Biometría**. Se tiene un esquema de muestreo de un número N aleatorio de cierta especie animal. Sea X_i el peso del $i-$ésimo especimen capturado. Entonces $X = X_1 + \cdots + X_N$ es el peso total capturado.

  De nuevo, N es discreta y X es continua.

Discutamos el caso donde tenemos una variable aleatoria discreta N y una variable aleatoria continua X.

Suponemos que las X_i's son independientes de N.

Modelos compuestos

Se puede mostrar que la distribución condicional para $X = X_1 + \cdots + X_N$, dado $N = n$, está dada por

$$f^{(n)}(x),$$

donde

$$\begin{aligned} f^{(1)}(z) &= f(z) \\ f^{(n)}(z) &= \int f^{(n-1)}(z-u)f(u)du. \end{aligned}$$

Supón que $P(N = 0) = 0$. La distribución marginal de $X = X_1 + \cdots + X_N$ estará dada por

$$f_X(x) = \sum_{n=1}^{\infty} f^{(n)}(x)p_N(n).$$

Modelos compuestos

Para calcular probabilidades, usamos

$$P(a < X < b) = \int_a^b \left(\sum_{n=1}^{\infty} f^{(n)}(x) p_N(n) \right) dx.$$

Supón que tenemos una suma geométrica de variables aleatorias exponenciales. Conocemos las formas de las funciones para cada una de las variables que hemos mencionado. Ahora,

$$f^{(n)}(x) = \begin{cases} \dfrac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases},$$

que corresponde a la densidad de una distribución Gamma.

Modelos compuestos

Así,

$$
\begin{aligned}
f_X(x) &= \sum_{n=1}^{\infty} f^{(n)}(x) p_N(n) \\
&= \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} p(1-p)^{n-1} \\
&= \lambda p e^{-\lambda x} \sum_{n=1}^{\infty} \frac{(\lambda(1-p)x)^{n-1}}{(n-1)!} \\
&= \lambda p e^{-\lambda x} e^{\lambda(1-p)x} \\
&= \lambda p e^{-\lambda p x}, \quad x \geq 0,
\end{aligned}
$$

que es una densidad exponencial.

Mezlcas de distribuciones.

Mezclas de distribuciones

Ejemplo

Denotemos con Y el tiempo (en cientos de horas) de cierto tipo de componentes electrónicos.

Estos componentes frecuentemente fallan cuando son colocados en el sistema. Se ha observado que las probabilidades de falla inmediata son de $1/4$. *Cuando el componente no ha fallado, sus tiempos de vida se pueden modelar mediante una distribución exponencial*$(\beta = 1)$.

¿Para este sistema, cuál es la distribución de los tiempos de vida de estos componentes?

Mezclas de distribuciones

Y es una mezcla de dos variables aleatorias. **La función de probabilidad se puede escribir como:**

$$F(y) = c_1 F_1(y) + c_2 F_2(y),$$

donde $c_1 + c_2 = 1$.

Para nuestro ejemplo, X_1 tiene probabilidad 1 en el punto 0 y X_2 tiene una distribución exponencial:

$$F_1(y) = \begin{cases} 0 & y < 0 \\ 1 & y \geq 0 \end{cases},$$

$$F(y) = \frac{1}{4} F_1(y) + \frac{3}{4} F_2(y),$$

$$F_2(y) = \int_0^y e^{-x} dx = 1 - e^{-y}, \quad y > 0.$$

Mezclas de distribuciones

Y es una mezcla de dos variables aleatorias. **La función de probabilidad se puede escribir como:**

$$F(y) = c_1 F_1(y) + c_2 F_2(y),$$

donde $c_1 + c_2 = 1$.

Para nuestro ejemplo, X_1 tiene probabilidad 1 en el punto 0 y X_2 tiene una distribución exponencial:

$$F_1(y) = \begin{cases} 0 & y < 0 \\ 1 & y \geq 0 \end{cases},$$

$$F(y) = \frac{1}{4} F_1(y) + \frac{3}{4} F_2(y),$$

$$F_2(y) = \int_0^y e^{-x} dx = 1 - e^{-y}, \quad y > 0.$$

Funciones de variables aleatorias.

Funciones de variables aleatorias

Anteriormente hemos trabajado con funciones de variables aleatorias,que de una u otra forma nos han dado información valiosa en la modelación de problemas del tipo aleatorio. Por ejemplo, dada una v. a. X, definimos:

- $g(X) = (X - \mu)^2$, y obtuvimos la varianza de X: $V(X) = E[g(X)]$.
- $g(X) = \frac{X-\mu}{\sqrt{V(X)}}$, $g(X) = (\frac{X-\mu}{\sigma})^3$, $g(X) = (\frac{X-\mu}{\sigma})^4$.
- $g(X) = e^{tX}$, $g(X) = 1 - e^{-\beta X}$.
- $g(X) = \sum_{i=1}^{n} X_i$, etc.

Incluso definimos, $g(Z) = Z\sigma + \mu$.

Funciones de variables aleatorias

Cada una de estas cantidades han servido para aumentar nuestro conocimiento de la v. a. de interés. Parece pues natural indagar el comportamiento típico de esas cantidades, esto es, encontrar y explotar sus funciones de densidad de probabilidad respectivas (de las más útiles).

En principio, las variables aleatorias que surjan serán susceptibles de ser analizadas para encontrar sus funciones de distribución, valores esperados, varianzas, etc. Sin embargo nos concentraremos en solo algunas transformaciones particulares que generan resultados de gran interés para la puesta en marcha de procedimientos estadísticos prácticos.

Funciones de variables aleatorias

Daremos a continuación una lista de resultados de transformaciones que nos llevan a las distribuciones más empleadas en la construcción de inferencias estadísticas.

Lo que debe quedar en mente son **las condiciones que permiten la derivación de dichas distribuciones**.

Los métodos de transformaciones y otros resultados de importancia, podrás consultarlos en el apéndice del capítulo.

Funciones de variables aleatorias

Resultado 1: Distribución Ji-Cuadrada (χ^2).

Sea X una variable aleatoria normal estándar, entonces la densidad de $Y = X^2$ es la función de densidad de una variable aleatoria con distribución Gamma con $\beta = 2$, $\alpha = \frac{1}{2}$; esto es, la de una χ^2_1, Ji-cuadrada con 1 grado de libertad. En símbolos,

$$X \sim N(0,1) \quad \Rightarrow \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < +\infty.$$

y para $Y = X^2$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}, \qquad y > 0.$$

esto es,

$$Y \sim Gamma(\alpha = \frac{1}{2}, \beta = 2), \qquad \text{i.e.} \quad Y \sim \chi^2_1.$$

Funciones de variables aleatorias

Es decir,

$$X \sim N(0,1) \quad \Rightarrow \quad X^2 \sim \chi_1^2 \equiv Gamma(\alpha = \frac{1}{2}, \beta = 2).$$

Esta distribución surge en el trabajo inferencial cuando se trata de establecer la distribución de varianzas muestrales, como veremos más adelante.

Funciones de variables aleatorias

Resultado 2: Distribución t-Student.

Sean X y Y variables aleatorias independientes tales que:

$$X \sim N(0,1), \qquad Y \sim \chi^2_n.$$

Definamos $T = \frac{X}{\sqrt{Y/n}}$ entonces

$$f_T(t) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})(1+\frac{t^2}{n})^{(n+1)/2}} & -\infty < t < \infty,\ n > 0,\ (n \in \mathbb{N}) \\ 0 & \text{en otro caso.} \end{cases}$$

Funciones de variables aleatorias

Esta distribución juega un papel predominante en estimación por intervalo y pruebas de hipótesis sobre la media μ, particularmente cuando la muestra proviene de una población normal y su varianza no se conoce, sin importar el tamaño de la muestra considerada.

Al igual que la distribución normal, la función de distribución de la t -Student no tiene forma cerrada, por lo que se será necesario usar software para calcular probabilidades.

Funciones de variables aleatorias

Observaciones:

- La distribución es simétrica con respecto a cero.
- $E(T) = 0,\ V(T) = \frac{n}{n-2},\ n > 2.$
- Si $n = 1$, entonces

$$f_T(t) = \frac{1}{\pi(1+t^2)} \qquad -\infty < t < \infty,$$

  lo cual corresponde a la distribución de $Cauchy(0,1)$.
- Notemos que

$$f_T(t) = \underbrace{\frac{1}{\sqrt{2\pi}} \frac{(\frac{n-1}{2})\Gamma(\frac{n-1}{2})}{\sqrt{\frac{n}{2}}\Gamma(\frac{n}{2})}}_{\star} \underbrace{(1+\frac{t^2/2}{n/2})^{-n/2}}_{\star\star} \underbrace{(1+\frac{t^2}{n})^{-1/2}}_{\star\star\star}$$

Funciones de variables aleatorias

Obsérvese que, cuando $n \to \infty$,

$$\bigstar \to \frac{1}{\sqrt{2\pi}}, \quad \bigstar\bigstar \to e^{-t^2/2}, \quad \bigstar\bigstar\bigstar \to 1.$$

Luego $f_T(t) \to \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, es decir que cuando $n \to \infty$ $f_T(t)$ converge a una normal estándar.

Notación: Escribimos t_n para referirnos a la distribución t con n grados de libertad.

Funciones de variables aleatorias

Notemos cómo la última de las curvas $t-Student$, con $n=20$, prácticamente se ve igual a la gráfica de una $N(0,1)$

Funciones de variables aleatorias

Resultado 3: Distribución F de Fisher o de Snedecor.

Sean X y Y dos variables aleatorias independientes tales que

$$X \sim \chi^2_n, \qquad Y \sim \chi^2_m.$$

Definamos $F = \frac{\frac{X}{n}}{\frac{Y}{m}} = \frac{mX}{nY}$, entonces

$$f_F(v) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}(\frac{n}{m})^{\frac{n}{2}} v^{\frac{n}{2}-1}(v\frac{n}{m}+1)^{-\frac{m+n}{2}} \qquad 0 \leq v < \infty.$$

Esta es llamada distribución F con n grados de libertad del numerador y m grados de libertad del denominador.

Funciones de variables aleatorias

Notación: Escribimos $F \sim F_{n,m}$ para denotar a la F con n grados de libertad del numerador y m grados de libertad del denominador.

Será necesario utilizar software para calcular probabilidades asociadas a esta distribución.

Como veremos en el capítulo de inferencia, este resultado tendrá su impacto en la descripción de los valores de varianzas muestrales,

- Comparaciones entre 2 o más poblaciónes;
- Cuando se quiere ver si hay diferencias en variabilidad al aplicar dos métodos de medición;
- En regresión para comparar dos modelos;
- En diseños, para comparar efectos de "tratamientos", (diferentes niveles de temperatura, insecticidas, concentraciones etc.);
- Entre muchos otros usos...

Funciones de variables aleatorias

Notemos que la distribucibución cambia si se intercambian los grado de libertad.

Funciones de variables aleatorias

Ejercicios:

1. Demostrar que si $X \sim t_v$, entonces la distribución de la v.a. $Y = X^2$ es una $F_{1,v}$
2. En el último ejemplo, considera la división $\frac{1}{F} = R = \frac{\frac{Y}{m}}{\frac{X}{n}}$ y demuestra que $R \sim F_{m,n}$, una F con los grados de libertad intercambiados.

Funciones de variables aleatorias

Resultado 4: Muestra Aleatoria y Estadísticos de Orden.

Un concepto fundamental es el siguiente: al tomar una muestra aleatoria

$$X_1, X_2, \ldots, X_n;$$

de alguna población, lo que se está considerando es **un conjunto de variables aleatorias i.i.d**.

Las características principales de una m.a. son:

- Cada v. a. está definida sobre la misma población, por lo tanto la distribución de cada una será exactamente la misma.
- Los posibles resultados de cada una no afectan los posibles resultados de las otras.

Funciones de variables aleatorias

Ejemplo (**Estadísticos de Orden**)

Existen muchos casos en los que la cantidad importante es el valor menor de una m.a., que por el mismo carácter aleatorio de la m.a. también será una v. a. Equivalentemente, la cantidad importante puede ser el valor mayor de la m. a., y también es una v.a. Las dos variables aleatorias así definidas se llaman respectivamente

- *primer estadístico de orden* $X_{(1)}$,
- *n-ésimo estadístico de orden* $X_{(n)}$.

Estos estadísticos de orden resultan ser una transformación de variables aleatorias:

- $X_{(1)} = \text{mín}\{X_1, X_2, \ldots, X_n\}$,
- $X_{(n)} = \text{máx}\{X_1, X_2, \ldots, X_n\}$.

Funciones de variables aleatorias

Aplicaciones:

- Cuando un sistema deja de funcionar al fallar alguna de sus n componentes, la duración total puede ser vista como la duración más corta de cualquiera de sus componentes. Cada componente puede tener asignada una v.a. de duración X_i, entonces, podríamos definir como la v.a. de interés a $X_{(1)}$=*el menor valor de una m.a. de tamaño* n; en símbolos, $X_{(1)} = \text{mín}\{X_1, X_2, \ldots, X_n\}$. (SISTEMA EN SERIE)
- Un sistema que deja de funcionar hasta que el último componente falla; aquí interesaría $X_{(n)}$ el valor mayor de una m.a. de tamaño n. En símbolos, $X_{(n)} = \text{máx}\{X_1, X_2, \ldots, X_n\}$. (SISTEMA EN PARALELO)

Funciones de variables aleatorias

Aplicaciones (continúa):

- En una prueba de resistencia de material de construcción, la variable crítica puede ser la máxima resistencia posible si se toma una m.a. de tamaño n (n mediciones). Nos interesaría, $X_{(n)}$=*el mayor valor de la m. a. de tamaño n*; en símbolos, $X_{(n)} = max\{X_1, X_2, \ldots, X_n\}$.
- También puede interesar el rango de la m.a., esto es $R = X_{(n)} - X_{(1)}$. Esta es una medición alternativa de la dispersión muy usada en la industria de la manufactura, la cual representa otra transformación de variables aleatorias.

Funciones de variables aleatorias

Las densidades de $X_{(1)}$ y $X_{(n)}$ se deducen a continuación usando el método de la acumulada.

Densidad de $X_{(n)}$: Llamaremos $Y_n = X_{(n)}$, y a sus valores y $F_{Y_n}(y)$ a su función de distribución acumulada. Entonces,

$$F_{Y_n}(y) = P\{Y_n \leq y\} = P\{X_1 \leq y\,, X_2 \leq y\,, \ldots, X_n \leq y\};$$

si Y_n es menor que y entonces cada uno de los valores de las X_i tienen que ser menor que y. Puesto que las X_i son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_n}(y) &= \prod_{i=1}^{n} P(X_i \leq y) \quad \text{(independencia)} \\ &= [P\{X \leq y\}]^n \quad \text{(distribuciones idénticas)} \\ &= [F_X(y)]^n. \end{aligned}$$

Funciones de variables aleatorias

Entonces,

$$
\begin{aligned}
f_{Y_n}(y) =& \frac{d}{dy}[F_X(y)]^n \\
=& n[F_X(y)]^{n-1}\frac{d}{dy}[F_X(y)] \quad \text{(Regla de la Cadena)} \\
=& n[F_X(y)]^{n-1}f_X(y). \quad \text{(Teorema Fundamental del Cálculo)}
\end{aligned}
$$

Funciones de variables aleatorias

Densidad de $X_{(1)}$: Llamaremos $Y_1 = X_{(1)}$, y a sus valores y $F_{Y_1}(y)$ a su función de distribución acumulada. Entonces,

$$F_{Y_1}(y) = P\{Y_1 \leq y\} = 1 - P\{Y_1 > y\} = 1 - P\{X_1 > y, X_2 > y, \ldots, X_n > y\};$$

el valor más pequeño Y_1 es mayor que y cuando cada uno de los valores de las X_i lo sean. Puesto que las variables aleatorias son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_1}(y) &= 1 - \prod_{i=1}^{n} P(X_i > y) && \text{(independencia)} \\ &= 1 - [P\{X > y\}]^n && \text{(distribuciones idénticas)} \\ &= 1 - [1 - F_X(y)]^n. && \text{(complemento)} \end{aligned}$$

Funciones de variables aleatorias

Después,

$$
\begin{aligned}
f_{Y_1}(y) &= \frac{d}{dy}[1-[1-F_X(y)]^n] \\
&= -n[1-F_X(y)]^{n-1}\frac{d}{dy}[1-F_X(y)] \quad \text{(Regla de la Cadena)} \\
&= -n[1-F_X(y)]^{n-1}[-f_X(y)] \quad \text{(Teorema Fundamental del Cálculo)} \\
&= n[1-F_X(y)]^{n-1}f_X(y).
\end{aligned}
$$

Funciones de variables aleatorias

Ejemplo

Considera una m.a. $X_1, X_2, \ldots, X_n$ *de tamaño* n*, tomada de una población* $Unif(0,1)$*. Calcula las densidades de los estadísticos de orden primero y* n*-ésimo. Recordando que*

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & c.o.p. \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$$

Lo que implica,

$$f_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y) = ny^{n-1} \cdot 1 \qquad para\ 0 < y < 1,$$

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y) = n(1-y)^{n-1} \cdot 1 \qquad para\ 0 < y < 1.$$

Funciones de variables aleatorias

Ejemplo (Sistema en Serie)

El siguiente sistema en serie funcionará siempre que cada una de las componentes (idénticas) mostradas fallen, cada componente funciona independientemente de las otras. Supóngase que X =tiempo de vida de cada componente del sistema y que $X \sim exp(\beta)$. Determinar el comportamiento del tiempo de vida del sistema, es decir, de $Y_1 = min(X_1, X_2, X_3)$.

Funciones de variables aleatorias

Recordemos que

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\beta} e^{-\frac{x}{\beta}} & x > 0 \end{cases} \qquad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\frac{x}{\beta}} & x > 0 \end{cases}$$

Entonces,

$$\begin{aligned} f_{Y_1}(y) &= n[1 - F_X(y)]^{n-1} f_X(y) = 3[1 - (1 - e^{-\frac{y}{\beta}})]^{3-1} \cdot \frac{1}{\beta} e^{-\frac{y}{\beta}} \\ &= \frac{3}{\beta}(e^{-\frac{y}{\beta}})^2 e^{-\frac{y}{\beta}} = \frac{3}{\beta} e^{-\frac{3y}{\beta}} \quad \text{para } y > 0. \end{aligned}$$

Es decir que $Y_1 \sim exp(\frac{\beta}{3})$.

Funciones de variables aleatorias

Ejemplo (Sistema en Paralelo)

El siguiente sistema en paralelo funcionará hasta que fallen todas las componentes (idénticas) mostradas, cada componente funciona independientemente de las otras. Supóngase que X =tiempo de vida de cada componente del sistema y que $X \sim exp(\beta)$. Determinar el comportamiento del tiempo de vida del sistema, es decir, de $Y_3 = max\{X_1, X_2, X_3\}$.

Funciones de variables aleatorias

Luego,

$$f_{Y_3}(y) = n[F_X(y)]^{n-1} f_X(y) = 3[1 - e^{-\frac{y}{\beta}}]^{3-1} \cdot \frac{1}{\beta} e^{-\frac{y}{\beta}}$$
$$= \frac{3}{\beta}[1 - e^{-\frac{y}{\beta}}]^2 e^{-\frac{y}{\beta}} \quad \text{para } y > 0.$$

Para un tratamiento más amplio sobre estadísticos de orden puedes revisar: *David, H.A.; Nagaraja, H.N. (2003). Order Statistics. Wiley Series in Probability and Statistics. doi:10.1002/0471722162.*

Funciones de variables aleatorias

Resultado 5: Distribuciones de Sumas de Variables Aleatorias Independientes.

El Método de la Función Generatriz de Momentos se aplica tanto a variables continuas como discretas. El método consiste en

- **Encontrar** la función generatriz de momentos de la variable transformada;
- **Identificarla** como la generatriz de alguna variable conocida;

dándonos entonces la distribución de la transformación.

Está basado en la unicidad de la generatriz de momentos (si existe), y en el teorema que dice: "...igualdad de generatrices para dos variables aleatorias implica igualdad de distribuciones".

Funciones de variables aleatorias

- Si **no** podemos identificar la generatriz de momentos de nuestra variable transformada con alguna de un catálogo, el método no nos proporciona ninguna información y será necesario utilizar el método de transformaciones de variables para poder determinar la distribución de la nueva variable.
- Ya hemos usado este procedimiento anteriormente en la derivación de la distribución binomial negativa vista como una suma de variables aleatorias geométricas con probabilidad de éxito constante p, y en la construcción de una Gamma a partir de sumas de Exponenciales.

Funciones de variables aleatorias

Resultado 6.

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias independientes y $a_1, a_2, \ldots, a_n$ constantes reales. Se define $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ (combinación lineal) entonces la generatriz de momentos de Y es

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t),$$

donde $M_{X_i}(t)$ es la generatriz de momentos de X_i.

Demostración: Se hace para el caso continuo, el caso discreto es análogo (sólo es necesario cambiar la integral por suma).

Funciones de variables aleatorias

$$M_Y(t) = E(e^{Yt}) = E(e^{(a_1X_1+a_2X_2+\ldots+a_nX_n)t}) = E(e^{a_1X_1t} \cdot e^{a_2X_2t} \cdot e^{a_3X_3t} \ldots e^{a_nX_nt})$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(e^{a_1X_1t} \cdot e^{a_2X_2t} \cdot e^{a_3X_3t} \ldots e^{a_nX_nt})\underbrace{(f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n))}_{\text{Densidad Conjunta}}dx_1dx_2dx_3\ldots dx_n$$

y dado que $X_1, X_2, \ldots, X_n$ son independientes,

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(e^{a_1x_1t} \cdot e^{a_2x_2t} \ldots e^{a_nx_nt})[f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)]dx_1dx_2\ldots dx_n$$

como las variables solo intervienen en una funcion, se puede cambiar la integral multiple por un producto de integrales simples.

Funciones de variables aleatorias

$$= \underbrace{\int_{-\infty}^{\infty} e^{a_1 x_1 t} f_{X_1}(x_1) dx_1}_{M_{X_1}(a_1 t)} \int_{-\infty}^{\infty} e^{a_2 x_2 t} f_{X_2}(x_2) dx_2 \cdots \int_{-\infty}^{\infty} e^{a_n x_n t} f_{X_n}(x_n) dx_n$$

$$= M_{X_1}(a_1 t) \cdot M_{X_2}(a_2 t) \ldots M_{X_n}(a_n t) \quad \therefore \quad M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

El método es ampliamente utilizado en transformaciones definidas como combinaciones lineales de muestras aleatorias, recuerda que una m. a. está compuesta de un conjunto de variables aleatorias i.i.d.

Funciones de variables aleatorias

Ejemplos: 1) Sean $X_1, X_2, \ldots, X_n$ variables Bernoulli con parámetro p e independientes. Considerando $Y = X_1 + X_2 + \ldots + X_n$, encuentra la distribución de Y.

La generatriz de momentos de X_i es

$$M_{X_i}(t) = (1-p) + pe^t,$$

con $a_i = 1$. Así que,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} [(1-p) + pe^t] = ((1-p) + pe^t)^n,$$

la cual representa la generatriz de momentos de una variable aleatoria binomial con parametros n y p

$$Y \sim Bin(n, p).$$

La suma de v.a. Bernoulli con parámetro p =una v.a. $Binomial(n, p)$.

Funciones de variables aleatorias

2) Si $X \sim Gamma(\alpha, \beta)$, encontrar la distribución de $Y = aX$, con $a > 0$ y distinta de 1.

La generatriz de una $X \sim Gamma(\alpha, \beta)$ es $M_X(t) = (\frac{1}{1-\beta t})^\alpha$, entonces

$$M_Y(t) = M_{aX}(t) = M_X(at) = \left(\frac{1}{1-\beta(at)}\right)^\alpha = \left(\frac{1}{1-(\beta a)t}\right)^\alpha,$$

que podemos identificar como la generatriz de una v. a. $Gamma(\alpha, \beta^*)$, si $a > 0$ con $\beta^* = a\beta$

$$Y \sim Gamma(\alpha, a\beta).$$

Funciones de variables aleatorias

3) Considera una v.a. χ^2_n Ji-cuadrada con n grados de libertad, es decir,

$$X \sim \chi^2_n \equiv Gamma(\alpha = \frac{n}{2}, \beta = 2).$$

Encontrar la distribución de $Y = aX$, con $a > 0$ y diferente de 1.

Del ejemplo anterior,

$$Y \sim Gamma(\alpha = \frac{n}{2}, \beta^* = 2a) = Gamma(\frac{n}{2}, 2a).$$

Esta no corresponde a la generatriz de una χ^2. Una v.a. χ^2_n multiplicada por una constante (distinta de la unidad), no resulta en otra v. a. χ^2. Este resultado es útil cuando se estudia el comportamiento de la varianza muestral S^2

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \text{pero} \quad S^2 = \frac{\sigma^2 Y}{n-1} \sim Gamma(\frac{n-1}{2}, 2\frac{\sigma^2}{n-1}).$$

Funciones de variables aleatorias

4) Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes normales, $X_i \sim N(\mu_i, \sigma_i^2)$. Encontrar la distribución de $Y = a_1X_1 + \ldots + a_nX_n$, donde a_i es constante.

La generatriz de momentos de X_i

$$M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}.$$

De acuerdo al teorema, la generatriz de momentos de $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ es

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

$$\Rightarrow M_{X_i}(a_i t) = e^{\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}}.$$

Funciones de variables aleatorias

Por lo tanto,

$$M_Y(t) = \prod_{i=1}^{n}[e^{\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}}] = e^{\sum_{i=1}^{n}[\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}]} = e^{t(\sum_{i=1}^{n} a_i \mu_i) + (\sum_{i=1}^{n} a_i^2 \sigma_i^2)\frac{t^2}{2}}.$$

Esta es la generatriz de momentos de una variable aleatoria normal con media $\sum_{i=1}^{n} a_i \mu_i$ y varianza $\sum_{i=1}^{n} a_i^2 \sigma_i^2$, por lo tanto

$$Y \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2);$$

es decir, cualquier combinacion lineal de normales independientes es a su vez una normal.

Funciones de variables aleatorias

En particular, nos gustaría analizar el comportamiento que tendrían los resultados de medias obtenidas de muestras aleatorias seleccionadas de una población normal de tamaño n, $X_1, X_2, \ldots, X_n$, esto es, de la variable aleatoria, $\overline{X} = \sum_{i=1}^{n} X_i/n$, cuando $X \sim N(\mu, \sigma^2)$. Es una variable aleatoria porque todavía no sabemos que valores veremos para cada X_i.

Funciones de variables aleatorias

Resultado 7: La distribución de $\overline{X}$ (Población Normal).

Observa que si en el ejemplo anterior la m. a. es extraída de una población normal con media μ y varianza σ^2 entonces cada $X_i \sim N(\mu, \sigma^2)$, y la v.a. definida como

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n \sim N(\mu, \frac{\sigma^2}{n})$$

Para este resultado, no importa el tamaño de la muestra, solo que la población de estudio *de donde se obtiene la muestra*, sea normal.

Funciones de variables aleatorias

Ejemplo: Sea $X_1, X_2, \ldots, X_n$ i.i.d. *Poisson*(λ). Calcular $P(\overline{X} < 2)$.

Para calcular esta probabilidad buscaremos la f.d.p. de $\overline{X}$.

Con el método de la generatriz, si hacemos $Y = \bar{X}$ y recordamos que la generatriz de una Poisson(λ) es $M_X(t) = e^{\lambda(e^t - 1)}$, tenemos

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} M_{X_i}(\frac{t}{n}) = \prod_{i=1}^{n} e^{\lambda(e^{\frac{t}{n}} - 1)} = [e^{\lambda(e^{\frac{t}{n}} - 1)}]^n = e^{\lambda n(e^{\frac{t}{n}} - 1)}.$$

Funciones de variables aleatorias

Puesto que **no** es una generatriz conocida, para este caso no podemos obtener la función de densidad de $Y = \overline{X}$. Esto ejemplifica que la f.d.p. de $\overline{X}$ no siempre es una normal, depende de la distribución asociada con la población. Sin embargo, podríamos intentar encontrar la f.d.p. de la v. a. $Y = \sum_{i=1}^{n} X_i$,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda(e^t-1)} = [e^{\lambda(e^t-1)}]^n = e^{\lambda n(e^t-1)},$$

que es la generatriz de una distribución Poisson con parámetro λn.

Entonces $Y \sim Poisson(\lambda^*)$, donde $\lambda^* = n\lambda$. Podemos calcular la probabilidad pedida como:

$$P(\overline{X} < 2) = P(\frac{1}{n}\sum_{i=1}^{n} X_i < 2) = P(\sum_{i=1}^{n} X_i < 2n) = P(Y < 2n).$$

Funciones de variables aleatorias

Con $\lambda = 1.5, n = 4$, tenemos que $Y \sim Poisson(\lambda^* = 6)$. Entonces:

$$P(\overline{X} < 2) = P(Y < 8) = 0.7439.$$

```
ppois(q=7,lambda=6)
```

Resultado 8: Distribución de una suma de variables aleatorias Ji-cuadradas independientes.

Sean $X_1, \ldots, X_n$ variables aleatorias independientes cada una de ellas con distribución χ^2 con ν_i grados de libertad. Sea $Y = X_1 + X_2 + \ldots + X_n$, encontrar su distribución.

La generatriz de momentos para X_i es:

$$M_{X_i}(t) = (1 - 2t)^{-\frac{\nu_i}{2}}.$$

Funciones de variables aleatorias

La generatriz de momentos de Y es entonces

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} (1-2t)^{-\frac{\nu_i}{2}} \\ &= (1-2t)^{\sum_{i=1}^{n}(-\frac{\nu_i}{2})} \\ &= (1-2t)^{-\frac{(\sum_{i=1}^{n}\nu_i)}{2}}, \end{aligned}$$

que corresponde a la generatriz de momentos de una variable aleatoria χ^2 con $\sum_{i=1}^{n} \nu_i$ grados de libertad, es decir, la suma de χ^2's independientes con ν_i grados de libertad cada una, es a su vez una χ^2 con $\nu = \sum_{i=1}^{n} \nu_i$ grados de libertad.

Funciones de variables aleatorias

Ejercicio:

1. Generar 500 muestras aleatorias de tamaños 5, 10, 100, 500, 1000 de una población N(0,1).
   1. En cada caso, calcular $\overline{X}$ para cada muestra.
   2. Graficar su histograma, así como su Q-Q plot. En todos los casos deberá verse como una normal pues esa es la distribución exacta de esos valores. Las imperfecciones del ajuste que puedes notar se deben a que estás haciendo una simulación, sólo estás viendo 500 de un # infinito de posibles resultados.
2. Repetir el ejercicio anterior para $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$, la desviación estándar, y concluye.

Métodos de transformación de variables

Métodos de transformación de variables

Técnicas para encontrar las densidades de probabilidad de funciones de variables aleatorias.

Nos interesara saber como se distribuye una variable aleatoria que ha sido definida como una transformación (función) de otra variable aleatoria.

Para ser más precisos, supongamos que X es una variable aleatoria cuya distribución ya conocemos, queremos encontrar la distribución de la variable Y definida como $Y = g(X)$, donde g es una función de X. Para ello existen 3 metodos clásicos :

- Método de la función de distribución acumulada,
- Método de la transformación,
- Método de la generatriz de momentos.

Método de la función de distribución acumulada.

Método de la función de distribución acumulada

Ejemplo (Distribución χ^2_1 con un grado de libertad)

Si suponemos que $X \sim N(0,1)$, hemos visto que $Y = X^2$ se distribuye como una χ^2_1. A saber,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \qquad y \geq 0. \end{aligned}$$

Derivando podemos encontrar la densidad de Y. Aplicando la regla de la cadena,

$$\begin{aligned} f_Y(y) = F_Y(y) &= \Phi'(\sqrt{y})(\frac{1}{2}y^{-1/2}) - \Phi'(-\sqrt{y})(-\frac{1}{2}y^{-1/2}) \\ &= \frac{1}{2\sqrt{y}}[\Phi'(\sqrt{y}) + \Phi'(-\sqrt{y})] \qquad y > 0. \end{aligned}$$

Método de la función de distribución acumulada

Como $\Phi'(x)$-la derivada de la acumulada de la normal- es la función de densidad de X, entonces

$$f_Y(y) = \frac{1}{2\sqrt{y}}[\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y} + \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y}] = \frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{1}{2}y}, \quad y > 0.$$

Como hicimos notar antes, la función de densidad anterior corresponde a la densidad de una v.a. con distribución $Gamma(\alpha = \frac{1}{2}, \beta = 2)$.

Es decir, si $X \sim N(0,1) \Rightarrow X^2 \sim \chi_1^2 \equiv Gamma(\alpha = \frac{1}{2}, \beta = 2)$. Como ya hemos mencionado, este resultado es muy utilizado en la práctica.

Método de la función de distribución acumulada

Ejemplo

Si X tiene una densidad uniforme con $a = -1$ y $b = 3$, obtener la densidad de $Y = |X|$.

Tenemos que

$$f_X(x) = \begin{cases} \frac{1}{4} & -1 < x < 3 \\ 0 & \textit{otra parte} \end{cases}.$$

Así que

$$F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y),$$

para $y > 0$. Derivando con respecto a y obtenemos que

$$f_Y(y) = f_X(y) + f_X(-y) \qquad y > 0.$$

Método de la función de distribución acumulada

En este caso particular tenemos

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y) = \frac{1}{4} + \frac{1}{4} & 0 < y < 1 \\ f_X(y) + f_X(-y) = \frac{1}{4} + 0 & 1 < y < 3 \end{cases}.$$

es decir,

$$f_Y(y) = \begin{cases} \frac{1}{2} & 0 < y < 1 \\ \frac{1}{4} & 1 < y < 3 \\ 0 & \text{otra parte} \end{cases}.$$

Método de la transformación o cambio de variable

Método de la transformación o cambio de variable

Este metodo puede aplicarse a ambos casos (discreto y continuo).

En el caso discreto queda determinado en forma directa.

Ejemplo. (variable aleatoria discreta): Considera a X la suma del resultado de lanzar dos dados. Encontrar:

a) La función de probabilidad de $Y = 3X - 2$ dado 1.

dado1 / dado 2	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Método de la transformación o cambio de variable

La función de probabilidad para X es

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Con $Y = 3X - 2$, tenemos

y	4	7	10	13	16	19	22	25	28	31	34
$f(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Método de la transformación o cambio de variable

b) Así, la funcion de probabilidad de la v.a.

$$Y = \begin{cases} 0 & si\ X\ par \\ 1 & si\ X\ non \end{cases}$$

está dada por

y	1	0
f(y)	$\frac{18}{36}$	$\frac{18}{36}$

Método de la transformación o cambio de variable

Teorema del Cambio de Variable.

Aunque recibe un tratamiento especial, esta técnica realmente es una forma resumida de presentar lo que se hace en el método de la acumulada, como veremos a continuación.

Teorema

Supongamos que X es una variable aleatoria continua y que $Y \equiv u(X)$. Si u es una función monótona en todos los valores de la variable X, de tal forma que se puede despejar X en la ecuación $Y = u(X)$, digamos $X = w(Y)$; entonces la densidad de Y esta dada por:

$$f_Y(y) = f_X(w(y)) \cdot |w'(y)|,$$

donde $w'(y) = \frac{d}{dy} w(y)$ y las barras indican que se debe sacar el valor absoluto de $w'(y)$.

Método de la transformación o cambio de variable

Demostración: Monótona (siempre con una misma tendencia) creciente.

Método de la transformación o cambio de variable

Supongamos que $Y = u(X)$ es creciente, entonces

$$x_1 \leq x_2 \Rightarrow y_1 \leq y_2,$$

ya que al ser u creciente se preserva la desigualdad. Así, la acumulada de Y es

$$F_Y(y) = P(Y \leq y) = P(u(X) \leq y) = P(X \leq w(y)) = F_X(w(y)).$$

Luego la densidad de Y es

$$\begin{aligned} f_Y(y) =& F_Y'(y) = F_X'(w(y)) \cdot w'(y) = f_X(w(y)) \cdot w'(y) \\ & \text{(Regla de la cadena)} \\ =& f_X(w(y)) \cdot w'(y), \end{aligned}$$

que también es monótona decreciente.

Método de la transformación o cambio de variable

Supongamos ahora que $Y = u(X)$ es decreciente. Luego,

$$x_1 \leq x_2 \Rightarrow y_1 \geq y_2,$$

es decir que se invierte la desigualdad.

Método de la transformación o cambio de variable

La acumulada de Y es entonces

$$\begin{aligned} F_Y(y) =& P(Y \leq y) = P(u(X) \leq y) = P(X \geq w(y)) \\ =& 1 - P(X \leq w(y)) \\ =& 1 - F_X(w(y)), \end{aligned}$$

mientras que su densidad es

$$\begin{aligned} f_Y(y) &= F_Y'(y) = -F_X'(w(y)) \cdot w'(y) \\ &= -f_X(w(y)) \cdot w'(y) \\ &= f_X(w(y)) \cdot [-w'(y)]. \end{aligned}$$

Método de la transformación o cambio de variable

Combinando u creciente o decreciente, obtenemos que:

$$f_Y(y) = f_X(w(y)) \cdot |w'(y)|,$$

lo que prueba el Teorema.

Los pasos que seguimos fueron

- Despejar X en la relación $Y = u(X)$ obteniendo $X = w(Y)$ y derivar $\frac{dx}{dy} = w'(y)$.
- Evaluar la f.d.p. de X en $w(y)$ y multiplicarla por $|w'(y)|$.

Método de la transformación o cambio de variable

Ejemplo

Dada la variable aleatoria X con funcion de densidad

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \textit{en otra parte} \end{cases},$$

encontrar la distribución de probabilidad de $Y = 8X^3$.

Tenemos que

$$Y = u(X) = 8X^3;\ y = 8x^3 \quad \Rightarrow \quad x = (\frac{y}{8})^{1/3} = \frac{1}{2}(y)^{1/3}\ (u(X)\ \textit{es monótona})$$

Entonces,

$$\therefore \qquad w(y) = \frac{1}{2}(y)^{1/3} \qquad w'(y) = \frac{1}{6}y^{-\frac{2}{3}} \qquad (\textit{despeje y derivación})$$

Método de la transformación o cambio de variable

Luego, la función de densidad de Y está dada por

$$
\begin{aligned}
f_Y(y) &= f_X(w(y)) \cdot |w'(y)| \\
&= \begin{cases} 2\left(\frac{y^{\frac{1}{3}}}{2}\right)\left(\frac{y^{-\frac{2}{3}}}{6}\right) & 0 < w(y) < 1 \\ 0 & \text{en otra parte} \end{cases} \\
&= \begin{cases} \frac{1}{6}\left(y^{-\frac{1}{3}}\right) & 0 < y < 8 \\ 0 & \text{otra parte} \end{cases}.
\end{aligned}
$$

Método de la transformación o cambio de variable

Ejercicio

Cuando $X \sim N(0,1)$ y $Y = X^2$, sabemos que Y tiene una distribución χ^2_1. Aplica la técnica anterior notando que Y es decreciente para valores negativos de X y creciente para los valores positivos de X -es decir, que será necesario partir la región y sumar las contribuciones en ambas regiones.

Es decir que, si $x_i = w_i(y)$,

$$f_Y(y) = \sum_{i=1}^{k} f(x_i) \left| \frac{dx_i}{dy} \right|.$$

(la derivada de la i-ésima transformación inversa que tengamos que calcular de las k regiones)

Método de la transformación o cambio de variable

Caso bivariado:

Sean X_1, X_2 variables aleatorias continuas con función de densidad conjunta $f_{X_1,X_2}(x_1, x_2)$ y sea $\mathcal{X} = \{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$. Supongamos que:

$$\begin{aligned} y_1 =& g_1(x_1, x_2) \\ y_2 =& g_2(x_1, x_2) \end{aligned}$$

Asumiremos que g satisface las siguientes condiciones.

Método de la transformación o cambio de variable

- g define una **transformación biyectiva** de $\mathcal{X}$ a $\mathcal{Y} = \{y_1, y_2\}$, donde y_1 y y_2 estan definidas por la transformación.
- El Jacobiano de g^{-1}, denotado por J y dado por el determinante de orden 2

$$J = \left| \frac{\partial [g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)]}{\partial \underset{\sim}{y}} \right| \qquad \underset{\sim}{y} = (y_1, y_2)$$

$$= \begin{vmatrix} \frac{\partial g_1^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1,y_2)}{\partial y_2} \\ \\ \frac{\partial g_2^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1,y_2)}{\partial y_2} \end{vmatrix},$$

tiene derivadas parciales de primer orden continuas sobre $\mathcal{Y}$.

Método de la transformación o cambio de variable

Si el Jacobiano es diferente de cero para $(y_1, y_2) \in \mathcal{Y}$, la función de densidad conjunta de (Y_1, Y_2) está dada por:

$$f_{Y_1 Y_2}(y_1, y_2) = \begin{cases} f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2))|J| & (y_1, y_2) \in \mathcal{Y} \\ \\ 0 & \text{en otro caso} \end{cases}$$

Método de la transformación o cambio de variable

Ejemplo (distribución t-Student)

Usada por primera vez en 1908 por W. S. Gosset quién estableció el comportamiento de algunas medidas muestrales, en forma experimental en principio. Fisher derivó la prueba matemática en 1924 y la generalización a varias variables fue dada por Hotelling y se le conoce en el análisis multivariado por "T^2 de Hotelling".

Esta distribución se deriva de la siguiente manera. Sean X y Y variables aleatorias **independientes** *tales que:*

$$X \sim N(0,1), \qquad Y \sim \chi^2_n,$$

y definamos

$$T = \frac{X}{\sqrt{Y/n}}.$$

Método de la transformación o cambio de variable

Observa que se trata del cociente de una normal estándar y la raíz cuadrada de una $\chi^2_{(n)}$ dividida por sus grados de libertad, siendo estas variables independientes. Por lo tanto, su distribución conjunta es el producto de sus densidades:

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, \qquad -\infty < x < \infty,\ y > 0.$$

Si la transformación es

$$t = \frac{x}{\sqrt{y/n}} \qquad \text{y si} \qquad u = y,$$

entonces la transformación inversa es

$$x = t\sqrt{u/n}, \qquad y = u.$$

Método de la transformación o cambio de variable

Claramente esto nos da una complicación extra: primero elegir esta segunda coordenada de tal forma que el problema de inversión de funciones no se complique y segundo obtener la función conjunta de donde obtendremos la marginal de la variable principal, en este caso t.

$$\mathcal{X} = \{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}^+\} \longrightarrow \mathcal{Y} = \{(t, u) | t \in \mathbb{R}, u \in \mathbb{R}^+\}.$$

Entonces,

$$J = \begin{vmatrix} \sqrt{\frac{u}{n}} & \frac{t}{\sqrt{n}} \frac{1}{2} u^{-\frac{1}{2}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}} \quad \Longrightarrow \quad |J| = \sqrt{\frac{u}{n}}.$$

Método de la transformación o cambio de variable

Así,

$$f_{T,U}(t,u) = \sqrt{\frac{u}{n}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2\left(\frac{u}{n}\right)}\frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma\left(\frac{n}{2}\right)}u^{\frac{n}{2}-1}e^{-\frac{1}{2}u}$$

para

$$0 < u < \infty, \qquad -\infty < t < \infty, \qquad n \in \mathbb{N}.$$

Método de la transformación o cambio de variable

Pero queremos $f_T(t)$, así que necesitamos la marginal

$$f_T(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2(\frac{u}{n})} \frac{1}{n^{\frac{1}{2}}(\frac{1}{2})^{-(n/2)}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1+\frac{1}{2}} e^{-\frac{1}{2}u} du$$

$$= \frac{(\frac{1}{2})^{n/2}}{\sqrt{2\pi}\Gamma(\frac{n}{2})\sqrt{n}} \int_0^\infty u^{\frac{n+1}{2}-1} e^{-\frac{1}{2}(1+\frac{t^2}{n})u} du.$$

Es decir que

$$f_T(t) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})(1+\frac{t^2}{2})^{(n+1)/2}} & -\infty < t < \infty, n > 0, (n \in \mathbb{N}) \\ 0 & \text{en otro caso} \end{cases}.$$

Método de la transformación o cambio de variable

Ejemplo: Distribución F de Fisher o de Snedecor.

Sean dos variables aleatorias independientes $X \sim \chi^2_n$, $Y \sim \chi^2_m$. Obtener la f.d.p. de la v.a.

$$F = \frac{\frac{X}{n}}{\frac{Y}{m}} = \frac{mX}{nY}.$$

Como X y Y son independientes su distribución de probabilidad es el producto de sus densidades, es decir,

$$f_{X,Y}(x,y) = \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \cdot \frac{1}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}} y^{\frac{m}{2}-1} e^{-\frac{y}{2}} \qquad x, y > 0.$$

Método de la transformación o cambio de variable

Si hacemos

$$V = \frac{mX}{nY}, \qquad U = Y,$$

la transformación inversa es

$$x = \frac{nvu}{m}, \qquad y = u, \qquad u, v > 0.$$

Esto último define una transformación uno a uno

$$\mathcal{X} = \{(x, y) | x \in \mathbb{R}^+, y \in \mathbb{R}^+\} \longrightarrow \mathcal{Y} = \{(v, u) | v \in \mathbb{R}^+, u \in \mathbb{R}^+\}$$

con Jacobiano

$$J = \begin{vmatrix} \frac{n}{m}u & \frac{n}{m}v \\ 0 & 1 \end{vmatrix} = \frac{n}{m}u \neq 0$$

Método de la transformación o cambio de variable

Entonces

$$f_{U,V}(u,v) = f_{X,Y}(\frac{nuv}{m}, u)|J| = f_{X,Y}(\frac{nuv}{m}, u)\left|\frac{n}{m}u\right| = \frac{n}{m}uf_{X,Y}(\frac{nuv}{m}, u)$$

así

$$\begin{aligned} f_{U,V}(v,u) =& \frac{n}{m}u\frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}(\frac{nuv}{m})^{\frac{n}{2}-1}e^{-\frac{nuv}{2m}} \cdot \frac{1}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}}u^{\frac{m}{2}-1}e^{-\frac{u}{2}} \\ =& \frac{n}{m}u\frac{1}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}}(\frac{nuv}{m})^{\frac{n}{2}-1}e^{-\frac{nuv}{2m}} \cdot u^{\frac{m}{2}-1}e^{-\frac{u}{2}} \qquad u, v > 0. \end{aligned}$$

Método de la transformación o cambio de variable

Nos interesa la marginal de V, ya que $V = F = \frac{\frac{X}{n}}{\frac{Y}{m}}$. Luego integraremos para todo valor de U

$$
\begin{aligned}
f_F(v) &= \int_0^\infty f_{U,V}(v,u)du \\
&= \frac{\frac{n}{m}\left(\frac{n}{m}\right)^{\frac{n}{2}-1} v^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \int_0^\infty u^{\frac{n+m}{2}-1} e^{-u(\frac{nv}{2m}+\frac{1}{2})} du.
\end{aligned}
$$

Recordando que para una v.a. $G \sim Gamma(\alpha, \beta)$ se tiene

$$
\int_0^\alpha Gamma(\alpha,\beta)dg = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} g^{\alpha-1} e^{-\frac{g}{\beta}} dg = 1,
$$

entonces

$$
\int_0^\infty g^{\alpha-1} e^{-\frac{g}{\beta}} dg = \Gamma(\alpha)\beta^\alpha.
$$

Método de la transformación o cambio de variable

Así, en la expresión para $f_F(v)$, podemos identificar la integral con esta última expresión si

$$\alpha = \frac{n+m}{2}, \qquad \beta = (\frac{nv}{2m} + \frac{1}{2})^{-1},$$

de donde

$$\begin{aligned} f_F(v) &= \frac{(\frac{n}{m})^{\frac{n}{2}} v^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \Gamma\left(\frac{m+n}{2}\right)\left(\frac{vn}{2m}+\frac{1}{2}\right)^{-\frac{m+n}{2}} \\ &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \left(\frac{n}{m}\right)^{\frac{n}{2}} v^{\frac{n}{2}-1}\left(v\frac{n}{m}+1\right)^{-\frac{m+n}{2}}. \end{aligned}$$

Probabilidades bajo la distribución $F_{n,m}$ sólo pueden calcularse a través de métodos numéricos.

Método de la transformación o cambio de variable

Ejemplo: Consideremos X_1 y X_2 variables independientes con distribución $N(0,1)$. Se define

$$Y_1 = \frac{X_1 + X_2}{2}, \qquad Y_2 = \frac{(X_1 - X_2)^2}{2}.$$

Encuentra la distribución conjunta de (Y_1, Y_2), sus marginales y muestra que son independientes.

Nota: Si (X_1, X_2) es una muestra aleatoria de tamaño 2 de una población normal, Y_1 representa a $\bar{X}$ y Y_2 a S^2. Así, este resultado muestra que para $n = 2$, $(\bar{X}, s^2)$ son variables aleatorias independientes. Este resultado puede generalizarse para cualquier n.

Método de la transformación o cambio de variable

La transformación definida por $y_1 = \frac{x_1+x_2}{2}$, $y_2 = \frac{(x_1-x_2)^2}{2}$ mapea

$$\mathcal{X} = \{(x_1, x_2)|x_1, x_2 \in \mathbb{R}\} \longrightarrow \mathcal{Y} = \{(y_1, y_2)|y_1 \in \mathbb{R}, y_2 \in [0, \infty)\}$$

pero no es una transformación uno a uno

$$T: \quad y_1 = \frac{x_1 + x_2}{2} \qquad T_1^{-1}: \quad x_1 = y_1 + \sqrt{\frac{y_2}{2}} \qquad T_2^{-1}: \quad x_1 = y_1 - \sqrt{\frac{y_2}{2}}$$

$$y_2 = \frac{(x_1 - x_2)^2}{2} \qquad x_2 = y_1 - \sqrt{\frac{y_2}{2}} \qquad x_2 = y_1 + \sqrt{\frac{y_2}{2}}$$

Primero, podemos excluir $\underset{\sim}{y} = \underset{\sim}{0}$, pero aún así la transformación definida no nos permite separar a $\tilde{\mathcal{X}}$ en dos conjuntos disjuntos pues todos los puntos en $x_1 = x_2$ coinciden.

Método de la transformación o cambio de variable

En cada uno de esos puntos se tendrá $y_2 = 0$. Para evitar esta dificultad se puede definir $f_{X_1,X_2}(x_1, x_2) = 0$ para todos los puntos donde $x_1 = x_2$, lo cual es posible sin alterar la función de probabilidad pues la probabilidad de una recta es cero (el volumen de una recta es cero). Así ,

$$\mathcal{X} = \{(x_1, x_2)|x_i \in (-\infty, \infty)\ i = 1, 2,\ x_1 \neq x_2\},$$

y lo podemos partirlo como

$$\mathcal{X}_1 = \{(x_1, x_2)|x_1 > x_2\}, \qquad \mathcal{X}_2 = \{(x_1, x_2)|x_1 < x_2\}.$$

Así, las transformaciones T_1^{-1} y T_2^{-1} estan bien definidas. Entonces:

$$T_1 : \qquad |J_1| = \left\| \begin{array}{cc} 1 & \sqrt{\frac{1}{2}}\frac{y_2^{-1/2}}{2} \\ \\ 1 & -\sqrt{\frac{1}{2}}\frac{y_2^{-1/2}}{2} \end{array} \right\| = \left| -\frac{y_2^{-1/2}}{\sqrt{2}} \right| = \frac{1}{\sqrt{2y_2}}.$$

Método de la transformación o cambio de variable

Notemos que

$$T_2: \qquad J_2 = \begin{vmatrix} 1 & -\sqrt{\frac{1}{2}}\frac{y_2^{-1/2}}{2} \\ 1 & \sqrt{\frac{1}{2}}\frac{y_2^{-1/2}}{2} \end{vmatrix} = \frac{y_2^{-1/2}}{\sqrt{2}} = \frac{1}{\sqrt{2y_2}} = |J_1|.$$

Lo cual implica

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) = & \frac{1}{\sqrt{2y_2}}\frac{1}{2\pi}e^{-\frac{1}{2}\left\{(y_1+\sqrt{\frac{y_2}{2}})^2+(y_1-\sqrt{\frac{y_2}{2}})^2\right\}} \\ & + \frac{1}{\sqrt{2y_2}}\frac{1}{2\pi}e^{-\frac{1}{2}\left\{(y_1-\sqrt{\frac{y_2}{2}})^2+(y_1+\sqrt{\frac{y_2}{2}})^2\right\}}, \end{aligned}$$

de donde

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{2}{2\pi\sqrt{2y_2}}e^{-\frac{1}{2}(2y_1^2+y_2)}.$$

Método de la transformación o cambio de variable

Notemos que la distribución conjunta se puede reescribir, multiplicando y dividiendo por las constantes adecuadas, como

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{\sqrt{2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{y_1^2}{\frac{1}{2}}} \frac{1}{2^{1/2}\Gamma(\frac{1}{2})} y_2^{\frac{1}{2}-1} e^{-\frac{1}{2}y_2} & y_1 \in \mathbb{R},\ y_2 > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$\therefore\ Y_1 \sim N(0, 1/2)$ y $Y_2 \sim \chi^2_1$ mostrando además que son independientes.

Nota: Dado que $X_1 \sim N(0,1)$ y $X_2 \sim N(0,1)$ son independientes, directamente es posible establecer la distribución de

$$Y_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

lo cual no es más que una combinación lineal de normales con $a = b = \frac{1}{2}$.

Método de la transformación o cambio de variable

$$\implies Y_1 \sim Normal(\frac{1}{2}(0) + \frac{1}{2}(0), \frac{1}{4}(1) + \frac{1}{4}(1)) \; i.e \; Y_1 \sim N(0, \frac{1}{2})$$

Ademas, si $Y_2 = \frac{(X_1 - X_2)^2}{2}$, notemos que tambien $\frac{X_1 - X_2}{\sqrt{2}}$ cae en el caso anterior con

$$a = \frac{1}{\sqrt{2}}, b = -\frac{1}{\sqrt{2}},$$

de donde

$$\frac{X_1}{\sqrt{2}} - \frac{X_2}{\sqrt{2}} \sim N(0, \frac{1}{2}(1) + \frac{1}{2}(1)) = N(0, 1).$$

Aquí, Y_2 representa una variable aleatoria $[N(0,1)]^2 \quad \therefore Y_2 \sim \chi^2_{(1)}$. Ademas la conjunta se factoriza como $f(y_1) \cdot f(y_2)$ lo cual nos da la independencia.

Método de la transformación o cambio de variable

Otros Ejemplos:

Ejemplo 1. Si X_1 y X_2 son variables aleatorias independientes que tienen densidad uniforme con $\alpha = 0$ y $\beta = 1$, encontrar la densidad de $Y = X_1 + X_2$.

Como son independientes, la densidad conjunta es el producto de las densidades marginales, o sea,

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1, \quad 0 < x_2 < 1 \\ 0 & \text{otra parte} \end{cases}.$$

Método de la transformación o cambio de variable

Así, la acumulada de Y está dada por

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \int_0^y \int_0^{y-x_2} 1 dx_1 dx_2 & 0 < y < 1 \\ 1 - \int_{y-1}^1 \int_{y-x_2}^1 1 dx_1 dx_2 & 1 < y < 2 \\ 1 & y > 2 \end{cases}.$$

Obsérvese que

$$\int_0^y \int_0^{y-x_2} 1 dx_1 dx_2 = \int_0^y [x_1]_0^{y-x_2} dx_2 = \int_0^y (y - x_2) dx_2 = [yx_2 - \frac{x_2^2}{2}]_0^y$$
$$= y^2 - \frac{y^2}{2} = \frac{y^2}{2}.$$

Método de la transformación o cambio de variable

$$\int_{y-1}^{1}\int_{y-x_2}^{1} 1 dx_1 dx_2 = \int_{y-1}^{1} [x_1]_{y-x_2}^{1} dx_2 = \int_{y-1}^{1} [1-(y-x_2)] dx_2$$

$$= [x_2 - yx_2 + \frac{x_2^2}{2}]_{y-1}^{1}$$

$$= [(1-y+\frac{1}{2}) - ((y-1) - y(y-1) + \frac{(y-1)^2}{2})]$$

$$= \frac{3}{2} - y - y + 1 + y(y-1) - \frac{(y-1)^2}{2}$$

$$= \frac{5}{2} - 2y + y^2 - y - \frac{(y^2-2y+1)}{2}$$

$$= \frac{5}{2} + \frac{-4y+2y^2-2y-y^2+2y-1}{2}$$

$$= \frac{4-4y+y^2}{2} = \frac{(y-2)^2}{2}.$$

Método de la transformación o cambio de variable

Por lo tanto

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 < y < 1 \\ 1 - \frac{(y-2)^2}{2} & 1 < y < 2 \\ 1 & y > 2 \end{cases}.$$

Así, la densidad está dada

$$f_Y(y) = \begin{cases} y & 0 < y < 1 \\ 2 - y & 1 < y < 2 \\ 0 & \text{c.o.p.} \end{cases}.$$

La densidad de la suma de dos variables aleatorias independientes uniformes es triangular.

Método de la transformación o cambio de variable

Ejemplo 2: Si P, el precio de un producto (en dólares), y S, las ventas totales (en unidades de 10,000), son variables aleatorias cuya distribución conjunta puede obtenerse aproximadamente como la densidad de probabilidad conjunta

$$f(p,s) = \begin{cases} 5pe^{-ps} & 0.20 < p < 0.40,\ s > 0 \\ \\ 0 & \text{en cualquier otra parte} \end{cases}.$$

Obtener la densidad de probabilidad de la variable aleatoria $V = SP$, que es la cantidad total de dinero que se gasta en este producto en unidades de \$10,000.

Método de la transformación o cambio de variable

$$F_V(v) = P(V \leq v) = P(SP \leq v) = P(S \leq \frac{v}{P})$$

$$= \int_{0.2}^{0.4} \int_0^{\frac{v}{p}} 5pe^{-ps} ds dp = -5 \int_{0.4}^{0.2} e^{-sp}\Big|_0^{-\frac{v}{p}} dp = 5 \int_{0.4}^{0.2} (e^{-v} - 1) dp$$

$$= 1 - e^{-v}, \quad v > 0$$

Derivando, $f_V(v) = e^{-v}$, $v > 0$. Es decir, $V \sim Exp(\beta = 1)$

El modelo normal bivariado

Ejercicio: Sean $Z_1, Z_2 \sim N(0,1)$ v.a. independientes. Además, sean $\mu_1, \mu_2 \in \mathbb{R}$, $\rho \in (-1,1)$ y $\sigma_1, \sigma_2 \in (0,\infty)$. Consideremos la transformación dada por

$$X = \mu_1 + \sigma_1 Z_1, \qquad Y = \mu_2 + \sigma_2[\rho Z_1 + \sqrt{1-\rho^2} Z_2].$$

Demuestra que $X \sim N(\mu_1, \sigma_1)$ y $Y \sim N(\mu_2, \sigma_2)$. Además, usa el teorema de la tranformación de variables para demostrar que la densidad de (X, Y) está dada por

$$f_{X,Y}(x,y) = \frac{\exp\left(-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right)}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}},$$

para $x, y \in \mathbb{R}$.

Apéndice 1: Método de la distribución acumulada para más de una variable

Método de la dist. acumulada para más de una variable

Ya mencionamos arriba como trabajar el caso bivariado, ahora extenderemos el resultado al caso general:

Método de Transformaciones para más de dos variables. Sea $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio de dimensión $n > 2$, con función de densidad conjunta $f_{\underset{\sim}{X}}(\underset{\sim}{x})$, tal que $\mathcal{X} = \{\underset{\sim}{x} | f_{\underset{\sim}{X}}(\underset{\sim}{x}) > 0\}$. Supongase que:

1. $\mathcal{X}$ puede ser particionado en los conjuntos $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_m$ tal que la transformación $\underset{\sim}{y} = (g_1(\underset{\sim}{x}), g_2(\underset{\sim}{x}), \ldots, g_n(\underset{\sim}{x}))$ es uno a uno de $\mathcal{X}_i$ sobre $\mathcal{Y}, i = 1, 2 \ldots m$, con transformación inversa denotada por $\underset{\sim}{x} = \{g_{1i}^{-1}(\underset{\sim}{y}), \ldots, g_{ni}^{-1}(\underset{\sim}{y})\}$, para $i = 1, 2, \ldots, m$.
2. Todas las derivadas parciales de primer orden son continuas y cada Jacobiano

Método de la dist. acumulada para más de una variable

$$J_i = \begin{vmatrix} \frac{\partial g_{1i}^{-1}}{\partial y_1} & \frac{\partial g_{1i}^{-1}}{\partial y_2} & \cdots & \frac{\partial g_{1i}^{-1}}{\partial y_n} \\ \vdots & & & \\ \frac{\partial g_{ni}^{-1}}{\partial y_1} & \frac{\partial g_{ni}^{-1}}{\partial y_2} & \cdots & \frac{\partial g_{ni}^{-1}}{\partial y_n} \end{vmatrix}$$

con $i = 1, 2, \ldots m$ es diferente de cero en el rango de la transformación, entonces la funcion de densidad conjunta de $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n) = g(\underset{\sim}{X}) = (g_1(\underset{\sim}{X}), \ldots, g_n(\underset{\sim}{X}))$ está dada por:

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \begin{cases} \sum_{i=1}^{m} |J_i| f_{\underset{\sim}{X}}(g_{1i}^{-1}(\underset{\sim}{y}), \ldots, g_{ni}^{-1}(\underset{\sim}{y})) & y \in \mathcal{Y} \\ 0 & \text{en otro caso} \end{cases}$$

Método de la dist. acumulada para más de una variable

Ejemplos

Ejemplos 1. Sean X y Y dos variables aleatorias normales estandar e independientes, encontrar la densidad conjunta de $U = X + Y$, $V = X - Y$.

Entonces,

$$u + v = 2x \qquad u - v = 2y$$
$$\Rightarrow x = \frac{u+v}{2} \qquad y = \frac{u-v}{2},$$

por lo tanto $w_1(u,v) = \frac{u+v}{2}$, $w_2(u,v) = \frac{u-v}{2}$ y

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}.$$

Método de la dist. acumulada para más de una variable

Como son independientes, la densidad conjunta de X y Y es

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y).$$

Así, como

$$\Rightarrow \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

se tiene

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \\ &= \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}} \qquad -\infty < x < \infty, \ -\infty < y < \infty. \end{aligned}$$

Método de la dist. acumulada para más de una variable

Así que la funcion de densidad conjunta de U y V es

$$
\begin{aligned}
f_{U,V}(u,v) &= f_{X,Y}(w_1(u,v), w_2(u,v)) \cdot |J| \\
&= \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2\right]\right)\left|-\frac{1}{2}\right| \\
&= \frac{1}{4\pi} e^{-\frac{1}{2}[\frac{u^2+2uv+v^2}{4} + \frac{u^2-2uv+v^2}{4}]} \\
&= \frac{1}{4\pi} e^{-\frac{1}{2(4)}[2u^2+2v^2]} \\
&= \frac{1}{4\pi} e^{-\frac{1}{4}[u^2+v^2]}, \quad -\infty < u < \infty, -\infty < v < \infty \\
&= \left[\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2(\sqrt{2})^2}u^2}\right]\left[\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2(\sqrt{2})^2}v^2}\right].
\end{aligned}
$$

Método de la dist. acumulada para más de una variable

Entonces, cada variable (U, V) corresponde a una normal con media (μ) igual a cero y varianza (σ^2) igual a 2. Además, U y V son independientes.

Si estuvieramos interesados en calcular la densidad de U, se puede hacer encontrando la marginal de U a partir de la conjunta de U y V

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u, v)dv.$$

Para este ejemplo:

$$f_U(u) = \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2(\sqrt{2})^2}u^2}\right] \left[\frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2(\sqrt{2})^2}v^2}\right] dv$$

por lo tanto $f_U(u) = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2(\sqrt{2})^2}u^2}$, $-\infty < u < \infty$.

Método de la dist. acumulada para más de una variable

Ejemplo 2. Sean X_1 y X_2 variables aleatorias continuas con densidad conjunta

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 4x_1x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{en otra parte} \end{cases}.$$

a) Encontrar la densidad conjunta de $Y_1 = X_1^2$ y $Y_2 = X_1X_2$.
b) Encontrar la densidad de $Y_2 = X_1X_2$.

Método de la dist. acumulada para más de una variable

SOLUCIÓN:

a)

$$y_1 = u_1(x_1, x_2) = x_1^2, \qquad y_2 = u_2(x_1, x_2) = x_1 x_2$$

$$x_1 = \sqrt{y_1} \qquad x_2 = \frac{y_2}{x_1} = \frac{y_2}{\sqrt{y_1}}$$

$$\Rightarrow \quad w_1(y_1, y_2) = y_1^{1/2} \qquad w_2(y_1, y_2) = \frac{y_2}{y_1^{1/2}}$$

Así que,

$$\therefore \qquad J = \begin{vmatrix} \frac{1}{2} y_1^{-1/2} & 0 \\ -\frac{1}{2} y_1^{-3/2} y_2 & y_1^{-1/2} \end{vmatrix} = (\frac{1}{2} y_1^{-1/2})(y_1^{-1/2}) - 0 = \frac{1}{2y_1}.$$

Método de la dist. acumulada para más de una variable

La densidad conjunta de Y_1 y Y_1 esta dada por

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(w_1(y_1,y_2), w_2(y_1,y_2))|J|.$$

Así que

$$\therefore \qquad f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 4y_1^{1/2}\frac{y_2}{y_1^{1/2}}\frac{1}{2y_1} & 0 < y_1 < 1,\ 0 < y_2 < \sqrt{y_1} \\ 0 & \text{en cualquier otra parte} \end{cases}$$

$$= \begin{cases} 2\frac{y_2}{y_1} & 0 < y_1 < 1,\ 0 < y_2 < \sqrt{y_1} \\ 0 & \text{en cualquier otra parte} \end{cases}.$$

Método de la dist. acumulada para más de una variable

b) La densidad de $Y_2 = X_1 X_2$ se puede calcular sacando la marginal de Y_2. Así,

$$\begin{aligned} f_{Y_2}(y_2) &= \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_1 \\ &= \int_{y_2^2}^{1} \frac{2y_2}{y_1} dy_1 = [2y_2 \ln(y_1)]_{y_2^2}^{1} = 0 - 2y_2 \ln(y_2^2) \\ &= -4y_2 \ln(y_2) \quad 0 < y_1 < 1. \end{aligned}$$

Método de la dist. acumulada para más de una variable

Ejemplo 3. Cuando la corriente I pasa por una resistencia R, la potencia generada esta dada por $W = I^2 R$. Supongase que I tiene una distribución uniforme en el intervalo $[0, 1]$ y que R tiene la funcion de densidad dada por

$$f_R(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & \text{en otro caso} \end{cases}.$$

Obtener la funcion de densidad de W, asumiendo que R e I son independientes. Notese que la potencia generada siempre es menor que la resistencia.

Método de la dist. acumulada para más de una variable

SOLUCION: Tenemos que

$$I \sim Unif(0,1).$$

Es decir que,

$$f_I(i) = 1 \qquad 0 \leq i \leq 1$$
$$f_R(r) = 2r \qquad 0 \leq r \leq 1.$$

Como R e I son independientes, entonces la densidad conjunta esta dada por

$$f_{R,I}(r,i) = 2r \quad 0 \leq i \leq 1,\ 0 \leq r \leq 1.$$

Como $W = I^2 R$, los valores posibles para W son entre 0 y 1 solamente.

Método de la dist. acumulada para más de una variable

Así,

$$
\begin{aligned}
F_W(w) = P(W \leq w) = P(I^2 R \leq w) = P(R \leq \frac{w}{I^2}) \\
&= \int_0^w \int_0^1 2rdidr + \int_w^1 \int_0^{\sqrt{w/r}} 2rdidr \\
&= \int_0^w [2ir]_0^1 dr + \int_w^1 [2ir]_0^{\sqrt{w/r}} dr = \int_0^w 2rdr + \int_w^1 2\sqrt{\frac{w}{r}} rdr \\
&= r^2|_0^w + 2\sqrt{w} \int_w^1 r^{1/2} dr \\
&= w^2 + \frac{4}{3}\sqrt{w} - \frac{4}{3}w^2 = \frac{4}{3}w^{1/2} - \frac{1}{3}w^2 \quad 0 < w \leq 1.
\end{aligned}
$$

$$
\therefore \quad f_W(w) = \frac{2}{3}w^{-1/2} - \frac{2}{3}w = \frac{2}{3}(w^{-1/2} - w) \quad 0 < w \leq 1.
$$

Método de la dist. acumulada para más de una variable

<u>OTRA FORMA (por el Teorema de la Transformación):</u> Consideremos la transformación $U = I$, $W = I^2 R$; es decir,

$$\begin{aligned} u &= i, & w &= i^2 r, \\ \Rightarrow i &= u, & r &= w/u^2. \end{aligned}$$

El Jacobiano de la transformación es:

$$J = \begin{vmatrix} 1 & 0 \\ -\frac{2w}{u^3} & \frac{1}{u^2} \end{vmatrix} = \frac{1}{u^2}$$

Así que la densidad conjunta de u y w es

$$f_{U,W}(u,w) = 2\left(\frac{w}{u^2}\right)\frac{1}{u^2} = \frac{2w}{u^4} \qquad \begin{array}{c} 0 \leq u \leq 1 \\ 0 \leq \frac{w}{u^2} \leq 1 \Rightarrow 0 \leq w \leq u^2. \end{array}$$

Método de la dist. acumulada para más de una variable

Entonces, la densidad de W es la marginal

$$f_W(w) = \int_{\sqrt{w}}^{1} \frac{2w}{u^4} du = 2w \left[-\frac{u^{-3}}{3} \right]_{\sqrt{w}}^{1} = -\frac{2}{3}w + \frac{2}{3}w(w)^{-3/2}.$$

$$\therefore \qquad f_W(w) = \frac{2}{3}w^{-1/2} - \frac{2}{3}w = \frac{2}{3}(w^{-1/2} - w) \quad 0 < w \leq 1$$

Apéndice 2:Generación de números aleatorios

Generación de números aleatorios

El método de Box-Müller para generación de valores de variables aleatorias normales es de gran utilidad.

En la generación de números aleatorios, de una v.a. exponencial a partir de una v.a. uniforme, hicimos uso del teorema de la transformación integral de probabilidad. El método en sí es simple, aunque tiene sus bemoles. Uno de ellos es la necesidad de conocer la distribución acumulada $F_X(x)$ de la v.a. de la cual queremos generar los números aleatorios, e incluso si se conociera, a la hora de intentar despejar X de $Y = F_X(x)$ puede ser que no se logre hacerlo en forma "cerrada".

Generación de números aleatorios

En el caso de la distribución normal, por ejemplo, no se puede encontrar una fórmula "cerrada" para $F_X(x)$ y por lo tanto el método no es fácilmente ejecutable. Una técnica común que se utiliza con frecuencia para generar observaciones de variables aleatorias normales es la conocida con el nombre de técnica de Box-Müller. No es el método más eficiente pero es ampliamente utilizado por su simplicidad. Este método se basa en el siguiente resultado:

Teorema

Sean X y Y variables aleatorias independientes con distribución normal estándar cada una de ellas. Sean (R, θ) las coordenadas polares del punto en el plano con coordenadas cartesianas (X, Y). Entonces R^2 y θ son variables aleatorias independientes con $R^2 \sim Exp(\beta = 2)$ y $\theta \sim Unif(0, 2\pi)$

Generación de números aleatorios

Demostración:

La función de densidad conjunta de las variables X y Y está dada por

$$f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, \ -\infty < x < \infty, -\infty < y < \infty$$

ya que X y Y son independientes $N(0,1)$. Usando las relaciones entre coordenadas polares y coordenadas cartesianas, tenemos que

$$R^2 = X^2 + Y^2$$

$$\theta = \tan^{-1}(\frac{Y}{X})$$

Generación de números aleatorios

Encontraremos la función de densidad conjunta de R^2 y θ. Por simplicidad en la notación, definimos

$$U = R^2 = X^2 + Y^2$$
$$V = \theta = \tan^{-1}(\frac{Y}{X})$$

la transformación inversa es

$$X = \sqrt{U}Cos(V)$$
$$Y = \sqrt{U}Sen(V),$$

con Jacobiano

$$J = \begin{vmatrix} \frac{1}{2}u^{-\frac{1}{2}}Cos(v) & -\sqrt{u}Sen(v) \\ \frac{1}{2}u^{-\frac{1}{2}}Sen(v) & \sqrt{u}Cos(v) \end{vmatrix} = \frac{1}{2}Cos^2(V) + \frac{1}{2}Sen^2(V) = \frac{1}{2}.$$

Generación de números aleatorios

Por lo tanto, la densidad conjunta de U y V es

$$f_{U,V}(u,v) = \frac{1}{2\pi} e^{-\frac{1}{2}u} \cdot \frac{1}{2}, \quad 0 < u < \infty, 0 < v < 2\pi$$

Las densidades marginales son

$$f_U(u) = \int_0^{2\pi} \frac{1}{4\pi} e^{-\frac{1}{2}u} dv = \frac{1}{4\pi} e^{-\frac{1}{2}u} v|_0^{2\pi} = \frac{1}{2} e^{-\frac{1}{2}u}, \quad 0 < u < \infty,$$
$$f_V(v) = \int_0^{\infty} \frac{1}{4\pi} e^{-\frac{1}{2}u} du = \frac{1}{2\pi} [e^{-\frac{1}{2}u}]_0^{\infty} = \frac{1}{2\pi}, \quad 0 < v < 2\pi,$$

por lo tanto las variables U y V son independientes y además $U \sim Exp(\beta = 2)$ y $V \sim Unif(0, 2\pi)$.

Generación de números aleatorios

Esto significa que, si se pueden generar una observación u_1 de una v.a. exponencial con parámetro $\beta = 2$ y, en forma independiente, una observación v_1 de una v.a. uniforme en el intervalo $(0, 2\pi)$, al tomar la transformación

$$\begin{aligned} x_1 &= \sqrt{u_1}\,Cos(v_1), \\ y_1 &= \sqrt{u_1}\,Sen(v_1) \end{aligned},$$

obtendremos <u>observaciones independientes de dos variables aleatorias normales estándar.</u>

Generación de números aleatorios

Aprovechando el resultado anterior, si recordamos que se pueden obtener valores de una v.a. exponencial a partir de una v.a. uniforme y notamos que una observación de una v.a. $Unif(0, b)$ se puede obtener a partir de una $Unif(0, 1)$, podemos usar únicamente la generación de variables aleatorias $Unif(0, 1)$ para generar una v.a. normal. Esto se considera a continuación:

Si

$$U_1 \sim Unif(0, 1) \text{ entonces } U = -\beta \ln(U_1) \sim Exp(\beta)$$

y si

$$U_2 \sim Unif(0, 1) \text{ entonces } V = 2\pi U_2 \sim Unif(0, 2\pi).$$

por lo tanto, las variables

$$X = \sqrt{-2ln(U_1)}Cos(2\pi U_2)$$
$$Y = \sqrt{-2ln(U_1)}Sen(2\pi U_2)$$

son variables aleatorias normales estándar e independientes.

Generación de números aleatorios

Podemos resumir el método de Box-Müller en los siguientes pasos:

1. Generar dos observaciones independientes u_1 y u_2 de variables aleatorias $Unif(0,1)$.
2. Evaluar $x_1 = \sqrt{-2ln(u_1)}Cos(2\pi u_2)$ y $x_2 = \sqrt{-2ln(u_1)}Sen(2\pi u_2)$.

Este método es bastante simple para ser programado y no requiere de cálculos de funciones inversas o memoria computacional, sin embargo es lento porque es necesario tomar un logaritmo, una raíz cuadrada, un coseno y un seno para cada dos observaciones.

Una vez generadas las observaciones de las variables aleatorias normales estándar, se pueden transformar en valores de variables aleatorias normales con una media y una varianza específicas, es decir, si queremos observaciones de una v.a. $N(\mu, \sigma^2)$ y x_1 es una observación de una normal estándar, entonces $w_1 = \sigma x_1 + \mu$ será una observación de esta distribución.

Generación de números aleatorios

Ejercicios (entregar con la tarea 5):

1. Sean $U_1 \sim Unif(0,1)$ y $U_2 \sim Unif(0,1)$ independientes. Encontrar la función de densidad conjunta de $X = \sqrt{-2ln(U_1)}Cos(2\pi U_2)$ y $Y = \sqrt{-2ln(U_1)}Sen(2\pi U_2)$ mediante el teorema de la transformación para dos variables y verificar que X y Y son variables aleatorias $N(0,1)$ independientes.
2. Usar este resultado para generar 100 valores de una v.a. $N(3,16)$ y hacer un histograma con los valores obtenidos. ¿Podrías hacer un Q-Q plot para estos datos?

Interrelaciones entre distribuciones más Comunes

Interrelaciones entre distribuciones más comunes

Figura: Tomado de "Statistical Inference", George Casella and Roger Berger Wadsworth & Brooks/Cole 1990.

Introducción: Modos de convergencia

- Ya que la estadística tiene que ver con recolectar información, naturalmente estamos interesados en saber que pasa cuando juntamos más y más datos.
- ¿Qué podemos decir del compartamiento límite de una sucesión de variables aleatorias $X_1, X_2, \ldots$?
- El estudio del comportamiento de sucesiones de variables aleatorias es uno de los aspectos más importantes de la teoría de la probabilidad. Esta parte de la probabilidad recibe los nombres de Teoría límite, Teoría asintótica o Teoría de grandes muestras.

Introducción: Modos de convergencia

Recordemos de Cálculo que una sucesión $\{x_n\}$ converge a un límite x si

$$\forall \epsilon > 0,\ \exists N > 0 \text{ tal que } |x_n - x| < \epsilon\ \forall n > N.$$

Veamos con un ejemplo porque la definición anterior es insuficiente para trabajar con variables aleatorias.

Si $x_n = x \in \mathbb{R}$, obtenemos que $\lim_{n\to\infty} x_n = x$. Consideremos una versión probabilistica de esto: Supongamos que $\{X_n\}$ es una sucesión de v.a. aleatorias independientes con distribución $N(0,1)$. ¿$X_n \to X \sim N(0,1)$?

Lo anterior no puede ser cierto pues $P(X_n = X) = 0$ para todo $n \in \mathbb{N}$. Lo anterior deja claro que en probabilidad necesitamos algo más sútil para establecer convergencias.

Introducción: Modos de convergencia

Tomemos otro ejemplo. Consideremos $X_1, X_2, \ldots$ donde $X_i \sim N(0, 1/i)$.

X_n está muy concentrada alrededor de 0 para n grande, por lo que nos gustaría decir $X_n \to 0$. Sin embargo, $P(X_n = 0) = 0$ para todo n.

Estudiaremos cinco conceptos de convergencia de variables aleatorias: convergencia en probabilidad, convergencia en distribución, convergencia en L_2, convergencia casi segura y convergencia en L_1.

La convergencia en L_1 y L_2 son casos particulares de la convergencia L_p. Debido a que queda fuera de los objetivos del curso, no estudiaremos está convergencia. Sin embargo las ideas son similares a los casos L_1 y L_2.

Tipos de convergencia

Definición

Sea X_n una sucesión de v.a. y sea X una v.a. Denotemos por F_n a la función de distribución de X_n y por F a la función de distribución de X.

1. *X_n converge a X en probabilidad, denotado por $X_n \xrightarrow{P} X$, si para cada $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

2. *X_n converge a X en distribución, denotado por $X_n \xrightarrow{d} X$, si*

$$\lim_{n\to\infty} F_n(t) = F(t)$$

para todo t donde F es continua.

3. *X_n converge a X en media cuadrática (o en L_2), denotado por $X_n \xrightarrow{L_2} X$, si*

$$\lim_{n\to\infty} E\left[(X_n - X)^2\right] = 0.$$

Tipos de convergencia

Ejemplo (Convergencia en distribución)

Sea $X_n \sim N(0, 1/n)$. *Tenemos que*

$$F_n(t) \equiv P(X_n < t) = P(\sqrt{n}X_n < \sqrt{n}t) = P(Z < \sqrt{n}t).$$

Luego,

$$\lim_{n\to\infty} F_n(t) = \begin{cases} 0 & si\ t < 0, \\ 1 & si\ t > 0. \end{cases}$$

¿Conocemos alguna función de distribución parecida? Consideremos la v.a. constante $Y = 0$. *Se tiene que la función de distribución de* Y *está dada por*

$$F_Y(t) = \begin{cases} 0 & si\ t < 0, \\ 1 & si\ t \geq 0. \end{cases}$$

Debido a que F_Y *no es continua en* 0, *podemos escribir que* $X_n \xrightarrow{d} Y \equiv 0$.

Tipos de convergencia

Notemos que en el ejemplo anterior $F_n(0) = 1/2$ para todo n y que $F_n(0) = 1/2 \neq 1 = F_Y(0)$. Es decir que la convergencia falla en $t = 0$; sin embargo esto no importa porque F_Y no es continua en $t = 0$.

Ejemplo (Convergencia en probabilidad)

Como antes, sea $X_n \sim N(0, 1/n)$. La desigualdad de Markov implica que, para $\epsilon > 0$,

$$P(|X_n| > \epsilon) = P(|X_n|^2 > \epsilon^2) \leq \frac{E(X_n^2)}{\epsilon^2} = \frac{(1/n)}{\epsilon^2} \to 0$$

cuando $n \to \infty$. Así que $X_n \xrightarrow{P} Y \equiv 0$.

Tipos de convergencia

El siguiente resultado establece relaciones entre los distintos tipos de convergencia.

Teorema

Se tienen las siguientes relaciones:

a) $X_n \xrightarrow{L_2} X$ *implica que* $X_n \xrightarrow{P} X$.

b) $X_n \xrightarrow{P} X$ *implica que* $X_n \xrightarrow{d} X$.

c) *Si* $X_n \xrightarrow{d} X$ *y si* $P(X = c) = 1$ *para algún número real* c, *entonces* $X_n \xrightarrow{P} X$.

En general ninguna de las implicaciones inversas se cumplen, a excepción del caso especial en c).

Tipos de convergencia

Demostración

a) Supongamos que $X_n \xrightarrow{L_2} X$. *Sea* $\epsilon > 0$ *fijo. De la desigualdad de Markov se sigue que*

$$P(|X_n - X| > \epsilon) = P(|X_n - X|^2 > \epsilon^2) \leq \frac{E\left[|X_n - X|^2\right]}{\epsilon^2} \to 0.$$

b) Sea $\epsilon > 0$ *y* x *un punto de continuidad de* F. *Entonces*

$$\begin{aligned} F_n(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, X \leq x + \epsilon) + P(X_n \leq x, X > x + \epsilon) \\ &\leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon). \end{aligned}$$

Tipos de convergencia

Demostración (Continuación)

También

$$\begin{aligned} F(x-\epsilon) &= P(X \leq x-\epsilon) \\ &= P(X \leq x-\epsilon, X_n \leq x) + P(X \leq x-\epsilon, X_n > x) \\ &\leq F_n(x) + P(|X_n - X| > \epsilon). \end{aligned}$$

Por lo tanto,

$$F(x-\epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) - P(|X_n - X| > \epsilon).$$

Tomando el límite es $n \to \infty$, concluimos que

$$F(x-\epsilon) \leq \lim_{n\to\infty} F_n(x) \leq F(x+\epsilon).$$

Tipos de convergencia

Demostración (Continuación)

Debido a que F es continua en x, al hacer $\epsilon \downarrow 0$, llegamos a que

$$\lim_{n\to\infty} F_n(x) = F(x).$$

c) Sea $\epsilon > 0$ fijo. Entonces,

$$\begin{aligned} P(|X_n - c| > \epsilon) &= P(X_n < c - \epsilon) + P(X_n > c + \epsilon) \\ &\leq P(X_n \leq c - \epsilon) + P(X_n > c + \epsilon) \\ &= F_n(c - \epsilon) + 1 - F_n(c + \epsilon) \\ &\to F(c - \epsilon) + 1 - F(c + \epsilon) \\ &= 0 + 1 - 1 = 0. \end{aligned}$$

Ahora demostraremos que las implicaciones inversas no se cumplen.

Tipos de convergencia

Demostración (Continuación)

Convergencia en probabilidad no implica convergencia en L_2. *Sea* $U \sim Uniforme(0,1)$ *y* $X_n = \sqrt{n}1_{(0,1/n)}(U)$. *Entonces, si* $\epsilon > 0$,

$$P(|X_n| > \epsilon) = P(\sqrt{n}1_{(0,1/n)}(U) > \epsilon) = P(0 \leq U < 1/n) = 1/n \to 0.$$

Por lo que, $X_n \xrightarrow{P} 0$. *Sin embargo* $E(X_n^2) = n\int_0^{1/n} = 1$ *para todo* n, *lo que implica que* X_n *no converge en* L_2.

Tipos de convergencia

Demostración (Continuación)

Convergencia en distribución no implica convergencia en probabilidad. *Sea* $X \sim N(0,1)$ *y* $X_n = -X$ *para* $n \in \mathbb{N}$. *Así,* $X_n \sim N(0,1)$. *Como* X_n *tiene la misma distribución que* X *para todo* n, *luego* $\lim F_n(x) = F(x)$ *para todo* x. *Por lo tanto* $X_n \xrightarrow{d} X$. *Pero, si* $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) = P(|2X| > \epsilon) = P(|X| > \epsilon/2) \neq 0.$$

Así que X_n *no converge en probabilidad.*

Tipos de convergencia

Uno podría pensar que si $X_n \xrightarrow{P} b$, para b constante, entonces $E(X_n) \to b$. Sin embargo esto no se cumple.

Sea X_n una variable aleatoria definida por $P(X_n = n^2) = 1/n$ y $P(X_n = 0) = 1 - 1/n$. Notemos que

$$P(|X_n| < \epsilon) = P(X_n = 0) = 1 - (1/n) \to 1,$$

para $\epsilon > 0$ pequeño. Así que $X_n \xrightarrow{P} 0$. Sin embargo,

$$E(X_n) = n^2(1/n) + 0 \cdot (1 - 1/n) = n.$$

Así, $E(X_n) \to \infty$.

Tipos de convergencia

Algunas propiedades de convergencia se conservan bajo transformaciones.

Teorema

Sean X_n, X, Y_n, Y v.a. Sea g una función continua y $c \in \mathbb{R}$.

a) *Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$ entonces $X_n + Y_n \xrightarrow{P} X + Y$.*
b) *Si $X_n \xrightarrow{L_2} X$ y $Y_n \xrightarrow{L_2} Y$ entonces $X_n + Y_n \xrightarrow{L_2} X + Y$.*
c) *Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} c$ entonces $X_n + Y_n \xrightarrow{d} X + c$.*
d) *Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$ entonces $X_n Y_n \xrightarrow{P} XY$.*
e) *Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} c$ entonces $X_n Y_n \xrightarrow{d} cX$.*
f) *Si $X_n \xrightarrow{P} X$ entonces $g(X_n) \xrightarrow{P} g(X)$.*
g) *Si $X_n \xrightarrow{d} X$ entonces $g(X_n) \xrightarrow{d} g(X)$.*

Tipos de convergencia

- Los incisos c) y e) son conocidos como el Teorema de Slutzky.
- $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$ en general no implican que $X_n + Y_n \xrightarrow{d} X + Y$.

Existen otros tipos de convergencia que estudiaremos: la convergencia casi segura (c.s.) y la convergencia en L_1.

Definición

Diremos que X_n converge casi seguramente a X, denotado por $X_n \xrightarrow{c.s.} X$, si

$$P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1.$$

Además diremos que X_n converge en L_1 a X, denotado por $X_n \xrightarrow{L_1} X$ si

$$E\left[|X_n - X|\right] \to 0$$

cuando $n \to \infty$.

Tipos de convergencia

A continuación establecemos relaciones entre los nuevos tipos de convergencia introducidos y los introducidos al principio de las diapositivas.

Teorema

Sean $\{X_n\}$ una sucesión de variables aleatorias y X una variable aleatoria. Entonces,

a) $X_n \xrightarrow{c.s.} X$ *implica que* $X_n \xrightarrow{P} X$.

b) $X_n \xrightarrow{L_2} X$ *implica que* $X_n \xrightarrow{L_1} X$.

c) $X_n \xrightarrow{L_1} X$ *implica que* $X_n \xrightarrow{P} X$.

Tipos de convergencia

- Convergencia en probabilidad no implica convergencia casi segura (Borel-Cantelli).
- Convergencia en probabilidad no implica convergencia L_1.
- Convergencia en probabilidad de una sucesión uniformemente integrable implica convergencia en L_1.
- Convergencia casi segura no implica convergencia en L_1. Tomemos X_n de tal manera que $P(X_n = 0) = 1 - 1/n^2$ y $P(X_n = 2^n) = 1/n^2$. Claramente $X_n \xrightarrow{c.s.} 0$ pero $E(X_n)$ no converge.

Ley de los Grandes Números

A continuación estudiaremos un resultado muy importante de Teoría de la Probabilidad: **La Ley de Grandes Números**.

Este Teorema nos dice que la media de una muestra grande es cercana a la media de la distribución.

La primera versión y prueba formal se debe a Jacob Bernoulli. Bernoulli dio una prueba rigurosa para el caso en el que las variables aleatorias son independientes e identicamente distribuidas con distribución *Bernoulli*(1/2). Le tomó 20 años poderlo demostrar.

La designación de Ley de los Grandes Números se debe a Poisson.

Ley de los Grandes Números

Teorema (Ley Débil General de los Grandes Números)

Supongamos que $\{X_n, n \geq 1\}$ *son variables aleatorias independientes y definamos* $S_n = \sum_{j=1}^{n} X_j$. *Si*

1. $\sum_{j=1}^{n} P(|X_n| > n) \to 0$,
2. $\frac{1}{n^2} \sum_{j=1}^{n} E[|X_j| 1_{\{|X_j| \leq n\}}] \to 0$;

entonces si definimos

$$a_n = \sum_{j=1}^{n} E[X_j 1_{\{|X_j| \leq n\}}],$$

obtenemos que

$$\frac{S_n - a_n}{n} \xrightarrow{P} 0.$$

Ley de los Grandes Números

Un caso particular importante del resultado anterior es la llamada Ley Débil de los Grandes Números de Khintchin.

Teorema (Ley Débil de los Grandes Números de Khintchin)

Supongamos que $\{X_n, n \geq 1\}$ *son v.a.i.i.d. con* $E(|X_1|) < \infty$ *y* $E(X_1) = \mu$. *Entonces,*

$$\frac{S_n}{n} \xrightarrow{P} \mu.$$

Observemos que no estamos requiriendo nada de la varianza en el resultado anterior. Si suponemos que $\mathrm{Var}(X_1) < \infty$, la desigualdad de Chebyshev implica que

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2},$$

que tiende a 0 cuando $n \to \infty$. Esto constituye una demostración de un caso particular de LDGNK.

Ley de los Grandes Números

Ejemplo: Consideremos el lanzamiento de una moneda con probabilidad p de obtener cara. Denotemos por X_i el resultado de un lazamiento (0 ó 1). Sabemos que $E(X_1) = p$. La fracción de caras después de n lanzamientos es $\overline{X}_n$.

De acuerdo a la LDGNK, $\overline{X}_n$ converge a p en probabilidad. Sin embargo, esto no quiere decir que converja numéricamente (ver LFGNK) a p. Significa que cuando n es grande, la distribución de $\overline{X}_n$ está concentrada alrededor de p.

Supongamos que $p = 1/2$. ¿Qué tan grande debe de ser n para que $P(0.4 \leq \overline{X}_n \leq 0.6) \geq 0.7$?

Ley de los Grandes Números

Notemos que $E(\overline{X}_n) = p = 1/2$ y que $\text{Var}(\overline{X}_n) = p(1-p)/n = 1/(4n)$. De la desigualdad de Chebyshev

$$\begin{aligned} P(0.4 \leq \overline{X}_n \leq 0.6) &= P(|\overline{X}_n - \mu| \leq 0.1) \\ &= 1 - P(|\overline{X}_n - \mu| > 0.1) \\ &\geq 1 - \frac{1}{4n(0.1)^2} \\ &= 1 - \frac{25}{n} \end{aligned}$$

La expresión anterior será mayor que 0.7 si $n = 84$.

Ley de los Grandes Números

Teorema (Ley Fuerte de Grandes Números de Kolmogorov)

Sea $\{X_n, n \geq 1\}$ *una sucesión de v.a.i.i.d. y pongamos* $S_n = \sum_{i=1}^{n} X_i$. *Existe* $c \in \mathbb{R}$ *tal que*

$$\overline{X}_n \xrightarrow{c.s} c$$

ssi $E(|X_1|) < \infty$, *en cuyo caso* $c = E(X_1)$.

Teorema (Corolario de la LFGNK)

Si $\{X_n\}$ *es una sucesión de v.a.i.i.d., entonces*

$$E(|X_1|) < \infty \quad \textit{implica que } \overline{X}_n \xrightarrow{c.s} \mu = E(X_1),$$

y

$$E(X_1^2) < \infty \quad \textit{implica que } \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2 \xrightarrow{c.s.} \sigma^2 = Var(X_1).$$

Ley de los Grandes Números

Veamos una aplicación de la LFGNK a Procesos de Poisson.

Supongamos que $\{X_n, n \geq 1\}$ una sucesión de v.a.i.i.d. tal que $X_i \sim Exponencial(\lambda)$. Poniendo $S_0 = 0$ y

$$N(t) := \sum_{j=0}^{\infty} 1_{\{S_j \leq t\}}.$$

Lo anterior define un proceso de Poisson (¿Por qué?). Observemos que

$$\{N(t) \leq n\} = \{S_n \geq t\}$$

y

$$S_{N(t)} \leq t \leq S_{N(t)+1}.$$

Ley de los Grandes Números

Ya que para todo $m \in \mathbb{N}$,

$$\lim_{t\to\infty} P(N(t) \leq m) = \lim_{t\to\infty} P(S_m > t) = 0,$$

se tiene que $N(t) \xrightarrow{P} \infty$. Definamos

$$\Lambda_1 = \{\omega \in \Omega : \frac{S_n(\omega)}{n} \to \lambda\},$$

$$\Lambda_2 = \{\omega \in \Omega : N(t,\omega) \to \infty\}.$$

Entonces $\Lambda := \Lambda_1 \cap \Lambda_2$ cumple que $P(\Lambda) = 1$ (por la LFGNK) y, por tanto, para cada $\omega \in \Lambda$

$$\frac{S_{N(t,\omega)}(\omega)}{N(t,\omega)} \to \lambda \qquad \text{cuando } t \to \infty;$$

es decir que

$$\frac{S_{N(t)}}{N(t)} \xrightarrow{c.s.} \lambda \qquad \text{cuando } t \to \infty.$$

Ley de los Grandes Números

Las desigualdades para $S_{N(t)}$ y t implican que

$$\frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)} = \frac{S_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)} \xrightarrow{c.s.} \lambda \qquad \text{cuando } t \to \infty$$

y

$$\frac{t}{N(t)} \geq \frac{S_{N(t)}}{N(t)} \xrightarrow{c.s.} \lambda \qquad \text{cuando } t \to \infty.$$

De donde

$$\frac{N(t)}{t} \xrightarrow{c.s.} \frac{1}{\lambda},$$

lo que implica que en el largo plazo la taza 'éxitos' en un proceso de Poisson es λ^{-1}.

Ley de los Grandes Números

Una consecuencia muy importante de la LFGNK es el llamado Teorema de Glivenko-Cantelli. Este Teorema nos dice que la función distribución empirica es una aproximación uniforme de la verdadera función de distribución.

Teorema (Glivenko-Cantelli)

Sea $X_1, \ldots, X_n$ una muestra de variables aleatorias con función de distribución F. Además, sea F_n la función de distribución empírica de $X_1, \ldots, X_n$. Definamos

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Entonces

$$D_n \xrightarrow{c.s.} 0,$$

cuando $n \to \infty$.

Distribución del Estimador de la Media

Teorema del Límite Central (TLC) Clásico: Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria (i.e. independientes e identicamente distribuidas) de una población arbitraria que tiene media μ y varianza σ^2, entonces cuando $n \to \infty$

$$Z_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

Por distribución límite debemos entender que sólo es un comportamiento aproximado, esto es, calcular probabilidades con la distribución muestral exacta o con la normal estándar, bajo muestras grandes, nos dará cantidades muy similares entre sí.

Teorema del Límite Central

A continuación veremos algunas versiones alternativas o más generales del TLC. Omitiremos las demostraciones pues están quedan fuera de los objetivos del curso.

Estos teoremas nos darán condiciones sobre cuando podemos aproximar la distribución de la media muestral por una normal estándar.

Teorema del Límite Central

Sean $\{X_n, n \geq 1\}$ independientes (pero no necesariamente idénticamente distribuidas) y supongamos que X_k tiene distribución F_k y que $E(X_k) = 0$ y $\mathrm{Var}(X_k) = \sigma_k^2$. Definamos

$$s_n^2 = \sum_{j=1}^{n} \sigma_j^2 = \mathrm{Var}(\sum_{j=1}^{n} X_j).$$

Diremos que $\{X_k\}$ satisface la condición de Lindeberg si, para todo $t > 0$, cuando $n \to \infty$

$$\frac{1}{s_n^2} \sum_{j=1}^{n} E(X_j^2 1_{\{|X_j/s_n| > t\}}) \to 0.$$

Teorema del Límite Central

- La condición de Lindeberg nos dice que para cada k, la mayoría de la masa de X_k está centrada en un intervalo alrededor de la media $(= 0)$ y que este intervalo es pequeño relativo a s_n.
- La condición de Lindeberg implica que

$$\sup_{k \leq n} \frac{\sigma_k^2}{s_n^2} \to 0$$

cuando $n \to \infty$.

- De la desigualdad de Chebyshev, el límite anterior implica que

$$\max_{k \leq n} P(|X_k|/s_n > \epsilon) \to 0.$$

Esta condición es típica en teoremas del límite central y asegura de que no existe un sumando que domina y que cada sumando contribuye un pequeña porción del total.

Teorema del Límite Central

Teorema (TLC de Lindeberg-Feller)

Bajo la notación usada arriba, la condición de Lindeberg implica

$$\frac{S_n}{s_n} \equiv \frac{\sum_{i=1}^{n} X_i}{s_n} \xrightarrow{d} N(0,1).$$

El Teorema anterior puede extenderse de la siguiente manera.

Teorema (Condición de Liapunov)

Sea $\{X_k, k \geq 1\}$ *una sucesión de v.a.i. satisfaciendo* $E(X_k) = 0$, $Var(X_k) = \sigma_k^2 < \infty$ *y* $s_n^2 = \sum_{j=1}^{n} \sigma_k^2$. *Si para algún* $\delta > 0$

$$\frac{\sum_{k=1}^{n} E\left[|X_k|^{\delta+2}\right]}{s_n^{\delta+2}} \to 0,$$

entonces la condición de Lindeberg se cumple y, por lo tanto, también el TLC.

Teorema del Límite Central

Ejemplo (Récords)

Sea $\{X_n, n \geq 1\}$ *una sucesión de v.a. i.i.d. continuas. Decimos que* X_n *es un récord de la sucesión si*

$$X_n > \text{máx}\{X_1, \ldots, X_{n-1}\}$$

y definamos el evento

$$A_k = \{X_k \text{ es un récord}\}.$$

Un resultado debido a Renyi dice que los eventos $\{A_k, k \geq 1\}$ *son independientes y que*

$$P(A_j) = \frac{1}{j} \qquad j \geq 2.$$

Teorema del Límite Central

Ejemplo (Récords)

Definamos

$$1_k = 1_{A_k}, \qquad \mu_n = \sum_{j=1}^{n} 1_j.$$

Notemos que

$$E(1_k) = \frac{1}{k}, \qquad Var(1_k) = \frac{1}{k} - \frac{1}{k^2},$$

de donde

$$s_n^2 = Var(\mu_n) = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k^2} \right) = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k^2} \sim \log n.$$

Así que

$$s_n^3 \sim (\log n)^{3/2}.$$

Teorema del Límite Central

Ejemplo (Récords)

Por otra parte,

$$E|1_k - E(1_k)|^3 = E\left|1_k - \frac{1}{k}\right|^3 = \left|1 - \frac{1}{k}\right|^3 \frac{1}{k} + \frac{1}{k^3}\left(1 - \frac{1}{k}\right) \leq \frac{1}{k} + \frac{1}{k^3},$$

y por lo tanto

$$\frac{\sum_{k=1}^{n} E|1_k - E(1_k)|^3}{s_n^3} \leq \frac{\sum_{k=1}^{n}\left(\frac{1}{k} + \frac{1}{k^3}\right)}{(\log n)^{3/2}} \sim \frac{\log n}{(\log n)^{3/2}} \to 0.$$

Esto implica que $\{X_k\}$ satisface la condición de Liapunov.

Teorema del Límite Central

Ejemplo (Récords)

Entonces tenemos que

$$\frac{\mu_n - E(\mu_n)}{\sqrt{Var(\mu_n)}} \xrightarrow{d} N(0,1).$$

Notemos que

$$\sqrt{Var(\mu_n)} = s_n \sim \sqrt{\log n}$$

y que

$$\frac{E(\mu_n) - \log n}{\sqrt{\log n}} = \frac{\sum_{k=1}^{n} \frac{1}{k} - \log n}{\sqrt{\log n}} \sim \frac{\gamma}{\sqrt{\log n}} \to 0,$$

donde γ es la constante de Euler.

Teorema del Límite Central

Ejemplo (Récords)

Lo que finalmente da como resultado (Convergence to Types Theorem)

$$\frac{\mu_n - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1).$$

Si quisieramos darle formalidad a está ultima parte, una forma de hacerlo es aplicando el "Convergence to Types Theorem".

Teorema del Límite Central

Es natural preguntarse que tan precisa es la aproximación normal del TLC. EL siguiente Teorema responde esta inquietud.

Teorema (Desigualdad de Berry-Essèen)

Sean $\{X_n, n \geq 1\}$ *i.i.d. y supongamos que* $E(X_k) = 0$, $E(X_1^2) = \sigma^2 < \infty$ *y* $E(|X_1|^3) < \infty$. *Entonces,*

$$\sup_x \left| P\left(\frac{\sqrt{n} \cdot \overline{X}_n}{\sigma} \leq z \right) - \Phi(z) \right| \leq \frac{33}{4} \frac{E\left[|X_1|^3\right]}{\sqrt{n}\sigma^3}.$$

Teorema del Límite Central

Decimos que una sucesión $\{X_n, n \geq 1\}$ es estrictamente estacionaria si, para cada k, la distribución conjunta de

$$(X_{n+1}, \ldots, X_{n+k})$$

es independiente de n para $n = 0, 1, \ldots$. Por otra parte, llamamos a la sucesión m-dependiente si para cada t,

$$(X_1, \ldots, X_{t-1}), \qquad (X_{t+m+1}, X_{t+m+2}, \ldots)$$

son independientes. Así, si las variables son lo suficientemente lejanas son independientes.

Un ejemplo de un sucesión estacionaria y m-dependiente es el modelo de series de tiempo llamado medias móviles de orden m: Sean $\{Z_n\}$ v.a.i.i.d. y, para constantes dadas $c_1, c_2, \ldots, c_m$, definamos el proceso

$$X_t := \sum_{i=1}^{m} c_i Z_{t-j}, \qquad t = 0, 1, \ldots.$$

Teorema del Límite Central

Teorema (Hoeffding and Robbins)

Supongamos que $\{X_n, n \geq 1\}$ *es una sucesión estrictamente estacionaria y m-dependiente con* $E(X_1) = 0$ *y*

$$Cov(X_t, X_{t+h}) = E(X_t X_{t+h}) =: \gamma(h).$$

Supongamos que

$$v_n := \gamma(0) + 2\sum_{k=1}^{n} \gamma(k) \neq 0.$$

Entonces

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t \xrightarrow{d} N(0, v_m)$$

y

$$nVar(\overline{X}_n) \to v_m.$$

Teorema del Límite Central

Finalmente, estudiaremos un TLC multivariado.

Teorema (Teorema del Límite Central Multivariado)

Supongamos que $X_1, X_2, \ldots, X_n$ son vectores aleatorios i.i.d. donde

$$X_i = (X_{1i}, X_{2i}, \ldots, X_{ki})'$$

con media

$$\mu = (\mu_1, \mu_2, \ldots, \mu_k)' = (E(X_{1i}), E(X_{2i}), \ldots, E(X_{ki}))'$$

y matriz de varianza Σ. Sea

$$\overline{X} = \left(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k\right)'$$

donde $\overline{X}_j = n^{-1} \sum_{i=1}^{n} X_{ji}$. Entonces,

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \Sigma).$$

Distribuciones Asociadas al Muestreo y Estimación Puntual

Distribución Muestral

Obtener información muestral y reconstruir variables de interés es el objetivo de la inferencia. Para ello definimos lo que es un estadístico de prueba

$$Y = h(X_1, X_2, \ldots, X_n),$$

y notamos que depende sólo de la muestra y no de los parámetros.

Distribución Muestral

Notemos además que: Y el estadístico, es una función de variables aleatorias, por lo que en sí mismo es una variable aleatoria y con los métodos de la sección anterior, sería posible establecer su distribución en forma exacta. A ésta se le llama Distribución Muestral

Ahora, más que obtener cualquier función de la muestra, estamos interesados en aquéllas que nos digan algo "inteligente" sobre los parámetros u otras características de la población.

Definición: Dada una población que tiene parámetro θ (notación general) y $X_1, X_2, \ldots, X_n$ una muestra aleatoria de esta población, entonces un estimador para θ es un estadístico que de alguna manera nos da valores próximos al valor real de θ.

Distribución Muestral

Ejemplos:

- En cualquier población $\overline{X}$ es un estimador de la media de la población.
- En cualquier población $\tilde{X}$ (mediana) es también un estimador de la media de la población.
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ es un estimador de σ^2.
- $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ es otro estimador para σ^2.
- En una población distribuida Uniforme en el intervalo (a, b), $\hat{a}_0 = X_{(1)} = \text{mín}\{X_1, X_2, \ldots, X_n\}$ es un estimador para a y $\hat{b}_0 = X_{(n)} = \text{máx}\{X_1, X_2, \ldots, X_n\}$, es un estimador para b.

Distribución Muestral

Ejemplo. (Sólo para fijar ideas): Supongamos que el modelo poblaciónal está dado por:

x	0	1	2	3
f(x)	.2	.4	.3	.1

a) Considera una muestra de tamaño 2: X_1, X_2(todas las posibles parejas que se pueden formar a partir de los 4 valores que toma la variable X). Construir la distribución de la media muestral:

$$\overline{X} = \frac{X_1 + X_2}{2}$$

Para ello, comenzaremos formando la distribución conjunta de X_1 y X_2.

Distribución Muestral

Notando que son independientes, construir la conjunta es sólo tomar el producto de los valores marginales.

x_2, x_1	0	1	2	3	suma
0	.04	.08	.06	.02	.2
1	.08	.16	.12	.04	.4
2	.06	.12	.09	.03	.3
3	.02	.04	.03	.01	.1
suma	.2	.4	.3	.1	1

Distribución Muestral

Lo que falta es calcular todos los posibles valores que se pueden obtener de medias muestrales, basados en las 16 parejas factibles, esto quiere decir que cuando uno realmente va y toma la muestra y calcula el valor medio de las dos observaciones entonces pueden darse los siguientes valores:

valor de $\overline{x}$	0	.5	1	1.5	2	2.5	3	Total
probabilidad	0.04	0.16	0.28	0.28	0.17	0.06	0.01	1

Distribución Muestral

Distribución muestral de $\overline{X}$

No es difícil ver de la función de probabilidad $f(x)$ que la media y varianza poblaciónales son $\mu = 1.3, \sigma^2 = 0.81$, y de la tabla anterior podemos calcular cuáles serían estas mismas características para la distribución de $\overline{X}$:

$$E(\overline{X}) = 0(0.04) + .5(0.16) + 1(0.28) + 1.5(.28) + 2(0.17) + 2.5(0.06) + 3(0.01) = 1.3 = \mu$$

y

$$\begin{aligned} V(\overline{X}) =& E(\overline{X})^2 - [E(\overline{X})]^2 = 0^2(0.04) + .5^2(0.16) + 1^2(0.28) + 1.5^2(0.28) + \\ &+ 2^2(0.17) + (2.5)^2(0.06) + 3^2(0.01) - (1.3)^2 = .405 = \frac{0.81}{2} = \frac{\sigma^2}{n} \end{aligned}$$

Distribución Muestral

En general para conocer los valores medios o de varianza de $\overline{X}$ o de S^2, o cualquier otro estimador que se exprese en términos de sumas o sumas de cuadrados , no es necesario conocer su distribución, sencillamente utilizamos propiedades de valores esperados

$$E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu$$
$$= \frac{n\mu}{n} = \mu \quad .$$

(Por el concepto de muestra aleatoria, misma distribución).

Distribución Muestral

Además,

$$V(\overline{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

(por el concepto de muestra aleatoria: independencia).

Estos resultados son válidos sin importar si la variable es discreta o continua (i.e. para cualquier distribución muestreada).

Pero la distribución muestral dependerá del comportamiento de donde se obtiene la muestra.

Distribución del Estimador de la Media

Distribución del Estimador de la Media

Nota que en este sentido $\overline{X}$ es un estimador natural de μ, dado que en promedio $\overline{X}$ toma ese valor (μ), y la variabilidad asociada con los valores de $\overline{X}$ se puede controlar mediante el tamaño de muestra . Esto es, entre más grande sea n, los valores de $\overline{X}$ se agrupan más alrededor de su media, que es μ.

Ya vimos, como un ejemplo en el capítulo anterior, que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria tomada de una población Normal(μ ,σ) entonces:

$$\overline{X} \sim Normal(\mu, \frac{\sigma^2}{n})$$

Aquí se hereda la distribución, aunque los parámetros hay que ajustarlos.

Distribución del Estimador de la Media

- Cuando, por ejemplo, se muestrea de una Poisson o de una Gamma vimos que no era sencillo establecer la distribución de $\overline{X}$ (lo hicimos sólo para la suma y concluimos que $\overline{X}$ es un múltiplo de la distribución de la suma).

- Un resultado que es de capital importancia, debido a la simplicidad que se deriva de éste, en cuanto al establecimiento de distribuciones muestrales es "El Teorema de Límite Central". Existen muchas versiones, la que verás aquí, es la más simple de todas ellas. Para otros casos ver [2], y algunos serán discutidos más adelante.

Distribución del Estimador de la Media

Teorema del Límite Central (TLC): Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población arbitraria que tiene media μ y varianza σ^2 (y su función generatriz de momentos es $M_X(t)$), entonces la distribución límite cuando $n \to \infty$ de

$$Z_n = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma}$$

es la distribución normal estándar.

Por distribución límite debemos entender que sólo es un comportamiento aproximado, esto es, calcular probabilidades con la distribución muestral exacta o con la normal estándar, bajo muestras grandes, nos dará cantidades muy similares entre sí.

Distribución del Estimador de la Media

Nota: La población no cambia de distribución, es el comportamiento de sus valores medios el que puede ser modelado en forma aproximada por una Normal Estándar.

¿Qué tan buena es la aproximación? Esto depende de dos factores entrelazados:

i) La forma de la distribución de donde se obtiene la muestra (básicamente, qué tan asimétrica es) y,

ii) el tamaño de la muestra.

Entrelazados porque entre más asimétrica sea la distribución original, los valores medios obtenidos de muestras de esas poblaciones, más tardarán en comportarse como una Normal. Aquí, "más" significa: más grande debe ser la muestra de donde calculemos el valor medio.

Distribución del Estimador de la Media

Las siguientes gráficas ilustran el razonamiento anterior:

- Se simularon 10,000 valores de un modelo dado.
- Se tomaron 1000 muestras de tamaños $n = 2, 5, 10, 30$ y 60 cada una.
- Se calcularon los valores medios en cada una de las 1000 muestras para cada uno de los diferentes casos (n's).
- Se graficaron los histogramas para visualizar la distribución muestral (estandarizando los valores de cada $\bar{x}$, recordemos que nosotros sabemos cuál es la media y varianza poblaciónal para estas simulaciones).

Distribución del Estimador de la Media

Figura: Una población con distribución triangular. Esta distribución es simétrica y "medio parecida" a una Normal.

Distribución del Estimador de la Media

Figura: Distribución Uniforme. Esta distribución es simétrica, pero no se parece al comportamiento de una Normal pues las colas de la distribución no "decaen".

Distribución del Estimador de la Media

Figura: Distribución Triangular inversa. Esta distribución es simétrica pero su comportamiento en el centro y colas de la distribución son los opuestos al de una Normal.

Distribución del Estimador de la Media

Figura: Distribución Exponencial. Esta distribución no es simétrica, ni se parece a la normal.

Distribución del Estimador de la Media

¿Qué conclusiones puedes obtener de estas gráficas?

¿La afirmación: "generalmente se considera como valor $n \geq 30$ que da una "buena aproximación" tiene más sentido? Pues sí, pero no siempre es necesario muestrear tanto para obtener buenos resultados.

Ahora daremos una demostración analítica del teorema, para satisfacer a los fans de la formalidad matemática (opcional)

Demostración: Se utiliza el método de la generatriz de momentos. Probaremos que $M_Z(t) \to e^{t^2/2}$ ya que $e^{t^2/2}$ es la generatriz de momentos de una variable aleatoria con distribución normal estándar.

Distribución del Estimador de la Media

$$
\begin{aligned}
M_Z(t) = E(e^{Zt}) = E(e^{t(\frac{\sqrt{n}}{\sigma}(\overline{X}-\mu))}) &= E(e^{t\frac{\sqrt{n}\bar{X}}{\sigma}} \cdot e^{-t\frac{\sqrt{n}\mu}{\sigma}}) \\
&= e^{-t\frac{\sqrt{n}\mu}{\sigma}} E(e^{t\frac{\sqrt{n}\bar{X}}{\sigma}}) = e^{-t\frac{\sqrt{n}\mu}{\sigma}} E(e^{\frac{t}{\sqrt{n}\sigma}\sum_{i=1}^{n} X_i}) \\
&= e^{-t\frac{\sqrt{n}\mu}{\sigma}} E(e^{\frac{t}{\sqrt{n}\sigma}X_1} \cdot e^{\frac{t}{\sqrt{n}\sigma}X_2} \cdots e^{\frac{t}{\sqrt{n}\sigma}X_n}),
\end{aligned}
$$

dado que las X_i son independientes

$$
M_Z(t) = e^{-t\frac{\sqrt{n}\mu}{\sigma}} \left[M_{X_1}(\frac{t}{\sigma\sqrt{n}}) \cdot M_{X_2}(\frac{t}{\sigma\sqrt{n}}) \cdots M_{X_n}(\frac{t}{\sigma\sqrt{n}}) \right]
$$

dado que son distribuciones idénticas

Distribución del Estimador de la Media

$$M_Z(t) = e^{-t\frac{\sqrt{n}\mu}{\sigma}}\left[M_X(\frac{t}{\sigma\sqrt{n}})\cdot M_X(\frac{t}{\sigma\sqrt{n}})\cdots M_X(\frac{t}{\sigma\sqrt{n}})\right]$$
$$= e^{-t\frac{\sqrt{n}\mu}{\sigma}}\left[M_X(\frac{t}{\sigma\sqrt{n}})\right]^n$$

Ahora expandemos $M_X(\frac{t}{\sigma\sqrt{n}})$ en series de potencias alrededor de $t = 0$ (Taylor):

$$M_X\left(\frac{t}{\sigma\sqrt{n}}\right) = M_X(0)+M_X'(0)\left(\frac{t}{\sigma\sqrt{n}}\right)+\frac{M_X''(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2+\frac{M_X'''(0)}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3+\cdots$$

$$M_X(t) = E(e^{tx})$$
$$M_X(0) = E(e^0) = 1$$

Distribución del Estimador de la Media

$$\Longrightarrow M_X\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n^{3/2}} + \cdots$$

Entonces la generatriz de Z esta dada por:

$$M_Z(t) = e^{-t\frac{\sqrt{n}\mu}{\sigma}}\left[1 + \mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n^{3/2}} + \cdots\right]^n$$

y tomando logaritmos:

$$\ln M_Z(t) = -\frac{t\sqrt{n}}{\sigma}\mu + n\ln\left[1 + \mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \mu_3'\frac{t^3}{6\sigma^3 n^{3/2}} + \cdots\right]$$

Distribución del Estimador de la Media

Sabiendo que $\ln(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} + \cdots$

$$
\begin{aligned}
\ln M_Z(t) =& \frac{t\sqrt{n}}{\sigma}\mu + n\left\{\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \cdots\right] - \frac{1}{2}\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \cdots\right]^2 + \right. \\
& \left. + \frac{1}{3}\left[\mu_1'\frac{t}{\sigma\sqrt{n}} + \mu_2'\frac{t^2}{2\sigma^2 n} + \cdots\right]^3 - \cdots\right\} \\
=& t\left[\frac{-\mu\sqrt{n}}{\sigma} + \frac{n\mu_1'}{\sigma\sqrt{n}}\right] + t^2\left[\frac{n\mu_2'}{2\sigma^2 n} - \frac{n(\mu_1')^2}{2\sigma^2 n}\right] \\
& + t^3\left[\frac{n\mu_3'}{6\sigma^3 n^{3/2}} - \frac{2n\mu_1'\mu_2'}{4\sigma^3 n^{3/2}} + \frac{n(\mu_1')^3}{3\sigma^3 n^{3/2}}\right] + \cdots
\end{aligned}
$$

Distribución del Estimador de la Media

como $\mu_1' = E(X) = \mu$ y $\mu_2' - \mu_1^2 = E(X^2) - [E(X)]^2 = V(X) = \sigma^2$

$$\Rightarrow \ln M_Z(t) = \frac{t^2}{2} + \frac{t^3}{\sigma^3\sqrt{n}}\left[\frac{\mu_3'}{6} - \frac{\mu_1'\mu_2'}{2} + \frac{(\mu_1')^3}{3}\right] + \cdots$$

Cuando $n \to \infty$, en la expresión anterior obtenemos que:

$$\lim_{n\to+\infty} \ln M_Z(t) = \frac{t^2}{2}$$

$$\ln[\lim_{n\to+\infty} M_Z(t)] = \frac{t^2}{2}$$

$$\therefore \lim_{n\to+\infty} M_Z(t) = e^{t^2/2}$$

Distribución del Estimador de la Media

Algunos casos particulares del TLC que son de importancia práctica:

1. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población $Bernoulli(p)$, i.e. cada X_i puede tomar solo valores $\{0, 1\}$. Recordemos que para cada X_i, $E(X_i) = p$ y $V(X_i) = pq$. La cantidad

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\#\text{ número de éxitos en la muestra}}{\#\text{total de observaciones}}$$

es la proporción de éxitos en muestra (%) $\overline{X} = \hat{p}$. Además

$$Y = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$$

ya que las X_i son una muestra aleatoria i.i.d.

Distribución del Estimador de la Media

El TLC implica

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0,1).$$

Haciendo álgebra llegamos a que

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{n(\overline{X} - \mu)}{n\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

$$\Rightarrow \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \sim N(0,1), \quad \text{cuando } n \to \infty$$

Distribución del Estimador de la Media

De aquí, haciendo los ajustes para la población Bernoulli, se llega a que

$$Z = \frac{\hat{p} - p}{\sqrt{pq}/\sqrt{n}} = \frac{Y - np}{\sqrt{npq}} \sim N(0,1) \quad \text{cuando } n \to \infty.$$

Este es el denominado Teorema de De-Moivre-Laplace.

Se pueden calcular probabilidades binomiales a través de aproximarlas con valores de probabilidad obtenidos de la distribución Normal. Que tan buena es la aproximación de nuevo depende de lo asimétrica que sea la Binomial considerada, esto es, depende del valor de p: si p es pequeña o muy grande, deberíamos tener muestras relativamente grandes para contrarrestar el efecto del sesgo.

Estudiar esta aproximación, así como la idea de en el libro de Factor de Corrección por continuidad, Texto[0].

Distribución del Estimador de la Media

2. Una aproximación semejante se puede encontrar para una muestra de tamaño n obtenida de una población Poisson. La razón del porque funciona se encuentra en el hecho de que, dado un proceso Poisson (Y) éste se puede subdividir en partes iguales y en cada subdivisión se auto define de nuevo un Proceso Poisson con λ restringido al subintervalo. Así, Y se puede expresar como una suma de variables Poisson i.i.d., y sobre éstas se aplica el TLC.

Distribución del Estimador de la Media

3. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con media μ_1 y varianza σ_1^2. Además, sea $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de otra población con media μ_2 y varianza σ_2^2, tal que las muestras sean independientes entre sí.

   El TLC implica que:

$$\frac{\overline{X} - \mu_1}{\frac{\sigma_1}{\sqrt{n}}} \to N(0,1) \quad \text{cuando } n \to \infty$$

   y

$$\frac{\overline{Y} - \mu_2}{\frac{\sigma_2}{\sqrt{n}}} \to N(0,1) \quad \text{cuando } n \to \infty.$$

Distribución del Estimador de la Media

En otras palabras $\overline{X}$ y $\overline{Y}$ se pueden aproximar como

$$\overline{X}_{n_1} \dot{\sim} N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \qquad \overline{Y}_{n_2} \dot{\sim} N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right).$$

cuando n_1 y n_2 es grande. Esto implica que, bajo las mismas premisas,

$$\Rightarrow \overline{X} - \overline{Y} \dot{\sim} N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

y

$$\overline{X} + \overline{Y} \dot{\sim} N\left(\mu_1 + \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Distribución del Estimador de la Media

Luego, cuando n_1 y n_2 son grandes, tenemos la siguiente aproximación

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \dot{\sim} N(0,1).$$

Se puede hacer algo similar para $\overline{X} + \overline{Y}$.

Distribución del Estimador de la Media

- Si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población Bernoulli(θ_1) y $Y_1, Y_2, \ldots, Y_n$ es una muestra de otra población Bernoulli(θ_2), tenemos los siguientes estimadores

$$\overline{X} = \hat{\theta}_1, \qquad \bar{Y} = \hat{\theta}_2.$$

Aplicando lo hecho antes llegamos a que

$$\frac{\hat{\theta}_1 - \hat{\theta}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}} \dot{\sim} N(0,1), \qquad \text{cuando } n_1 \to \infty,\ n_2 \to \infty.$$

Distribución del Estimador de la Media

El TLC nos dice que la distribución de

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

se aproxima a la de una normal para n grande.

Existe un resultado similar cuando lo que interesa es una función de $\overline{X}$, digamos $g(\overline{X})$, método que estudiaremos después.

A continuación aplicaremos el TLC en algunos ejemplos.

Distribución del Estimador de la Media

Ejemplo 1: Treinta componentes electronicos estan conectados de la siguiente forma: Tan pronto como D_1 falle, D_2 empieza a funcionar, y así sucesivamente. Supón que cada componente tiene un tiempo de vida exponencial con parametro $\theta = 10$ horas y que sus tiempos de vida son independientes. Calcular la probabilidad de que el sistema funcione cuando menos 350 horas.

Sea

$$\begin{aligned} T_1 &= \text{Tiempo de vida del componente } D_1 \\ T_2 &= \text{Tiempo de vida del componente } D_2, \\ &\vdots \\ T_{30} &= \text{Tiempo de vida del componente } D_{30}. \end{aligned}$$

Distribución del Estimador de la Media

Sabemos que $T_i \sim Exp(\theta = 10)$ $i = 1, 2, \ldots, 30$ y que además son independientes entre sí. Definamos T como el tiempo de vida de todo el sistema. Entonces $T = T_1 + T_2 + \ldots + T_{30}$ y por tanto

$$P(T \geq 350) = P\left(\sum_{i=1}^{30} T_i \geq 350\right)$$
$$= P\left(\frac{\sum_{i=1}^{30} T_i}{30} \geq \frac{350}{30}\right) = P(\overline{T} \geq \frac{350}{30}).$$

En este caso tenemos que los primeros momentos están dados por $\mu = \theta$ y$\sigma^2 = \theta^2$. Luego, aplicando el TLC llegamos a que

$$P\left(\frac{\overline{T} - \theta}{\frac{\theta}{\sqrt{n}}} \geq \frac{\frac{350}{30} - 10}{\frac{10}{\sqrt{30}}}\right) \cong P(Z \geq 0.9128)$$
$$= 1 - P(Z < 0.9128) = 1 - 0.8186 = 0.1814.$$

Distribución del Estimador de la Media

Ejemplo 2: Suponegamos que $X_1, \ldots, X_{50}$ son i.i.d. $Poisson(\lambda = 0.03)$ y sea $S = \sum_{i=1}^{50} X_i$.

a) Calcular $P(S > 3)$ usando TLC.

$$P(S > 3) = P(\sum_{i=1}^{50} X_i > 3) = P\left(\frac{\sum_{i=1}^{50} X_i}{50} > \frac{3}{50}\right)$$

$$= P(\overline{X} > \frac{3}{50}), \quad \mu = \lambda, \quad \sigma^2 = \lambda$$

$$= P\left(\frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{n}}} > \frac{\frac{3}{50} - 0.03}{\sqrt{\frac{0.03}{50}}}\right) \cong P(Z > 1.22) = 1 - P(Z < 1.122)$$

$$= 1 - 0.8888 = 0.1112$$

Con correccion por continuidad, con $S > 3.5$ queda 0.0516.

Distribución del Estimador de la Media

b) Calcular $P(S > 3)$ usando la distribucion de la suma de Poisson's.

Tenemos que

$$S \sim Poisson\left(\sum_{i=1}^{50} \lambda_i\right) \Longrightarrow S \sim Poisson(50(0.03)) = Poisson(1.5)$$

Así que

$$\begin{aligned} P(S > 3) &= 1 - P(S \leq 3) \\ &= 1 - [P(S = 0) + P(S = 1) + P(S = 2) + P(S = 3)] \\ &= 0.065. \end{aligned}$$

La aproximación se recomienda más en la medida que λ sea grande (> 5). Esto porque entonces la distribución Poisson es menos asimétrica.

Distribución del Estimador de la Media

Ejercicio 4: La proporcion real de familias en cierta ciudad que viven en casa propia es 0.7. Si se escogen al azar 84 familias de esa ciudad y se les pregunta si viven o no en casa propia, ¿Con qué probabilidad podemos asegurar que el valor que se obtendra de la proporcion muestral caerá entre 0.64 y 0.76?

Denotemos por X_i a la variable de que la i-ésima familia viva en en casa propia (1 ó 0). En este caso tenemos que $X_i \sim Bernoulli(0.7)$. La proporción muestral está dada por

$$\hat{\theta}_{84} = (84)^{-1} \sum_{i=1}^{84} X_i.$$

Distribución del Estimador de la Media

$$
\begin{aligned}
P(0.64 < \hat{\theta}_{84} < 0.76) &= P\left(\frac{0.64 - 0.7}{\sqrt{\frac{0.7(0.3)}{84}}} < \frac{\hat{\theta}_{84} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} < \frac{0.76 - 0.7}{\sqrt{\frac{0.7(0.3)}{84}}}\right) \\
&\simeq P(-1.2 < Z < 1.2) = 2P(0 < Z < 1.2) \\
&= 2(0.3849) = 0.7698
\end{aligned}
$$

De nuevo, esta aproximación puede mejorarse usando la corrección por continuidad y se deja como ejercicio.

Distribución del Estimador de la Media

En resumen, la distribución muestral de $\overline{X}$ puede trabajarse como:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \begin{cases} N(0,1) & \text{si la muestra viene de una población normal} \\ N(0,1) & \text{si la muestra viene de una población arbitraria} \\ & \text{y } n \text{ es grande} \end{cases}.$$

Distribución del Estimador de la Varianza

Distribución del Estimador de la Varianza

Otra distribución muestral que es de nuestro interés es la de la varianza muestral S^2 , puesto que S^2 es un estimador natural de σ^2.

Recordemos nuestra definición de la varianza muestral:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left\{(X_i - \mu) - (\overline{X} - \mu)\right\}^2.$$

Entonces,

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n}\left\{(X_i - \mu) - (\overline{X} - \mu)\right\}^2 \\ &= \sum_{i=1}^{n}(X_i - \mu)^2 - 2\sum_{i=1}^{n}(X_i - \mu)(\overline{X} - \mu) + n(\overline{X} - \mu)^2 \\ &= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2.\end{aligned}$$

Distribución del Estimador de la Varianza

Además,

$$
\begin{aligned}
E((n-1)S^2) &= E\left\{\sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2\right\} \\
&= \sum_{i=1}^{n} E(X_i-\mu)^2 - nE(\overline{X}-\mu)^2 \\
&= \sum_{i=1}^{n} \sigma^2 - n\frac{\sigma^2}{n} = n\sigma^2 - \sigma^2 = \sigma^2(n-1),
\end{aligned}
$$

pues $E(X_i-\mu)^2 = \sigma^2$ y $E(\overline{X}-\mu)^2 = \frac{\sigma^2}{n}$.

Por lo tanto,

$$E(S^2) = \sigma^2.$$

Distribución del Estimador de la Varianza

Establecer la varianza de S^2 es bastante más complejo, por lo que en principio lo haremos sólo para el caso cuando la muestra aleatoria $X_1, X_2, \ldots, X_n$ haya sido tomada de una población $Normal(\mu, \sigma)$; esto se hará derivando primero la distribución muestral y, posteriormente, sus momentos, media y varianza.

La media ya sabemos que es la varianza de la población, sin importar de dónde hayamos tomado la muestra $E(S^2) = \sigma^2$ siempre.

Sabemos que

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

también sabemos que, cuando la muestra es tomada de una población normal,

$$\frac{(X_i - \mu)}{\sigma} \sim N(0,1), \qquad \frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1).$$

Distribución del Estimador de la Varianza

Con la información anterior en mente, consideremos lo siguiente.

Si en la expresión para la varianza multiplicamos ambos lados por $n-1$, restamos y sumamos μ dentro del paréntesis, tenemos

$$(n-1)S^2 = \sum_{i=1}^{n} \left\{(X_i - \mu + \mu - \overline{X})\right\}^2.$$

Agrupando y elevando al cuadrado llegamos a

$$\begin{aligned}(n-1)S^2 &= \sum_{i=1}^{n} \left\{(X_i - \mu) - (\overline{X} - \mu)\right\}^2 = \sum_{i=1}^{n}[(X_i - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu) + (\overline{X} - \mu)^2] \\ &= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{n}(\overline{X} - \mu)^2.\end{aligned}$$

Distribución del Estimador de la Varianza

En el segundo término se dejó sólo el término que tiene índice (sobre el que corre la sumatoria), ahora, puesto que

Observemos que

$$\sum_{i=1}^{n} X_i = n\overline{X} \qquad \text{(por definición de media muestra)}$$

$$\sum_{i=1}^{n} \mu = n\mu \qquad \text{(se suma n veces una constante)}$$

$$\sum_{i=1}^{n} (\overline{X} - \mu)^2 = n(\overline{X} - \mu)^2 \qquad \text{(se suma n veces una constante)}$$

Distribución del Estimador de la Varianza

Usando la observación anterior, podemos escribir que

$$\begin{aligned}(n-1)S^2 &= [\sum_{i=1}^{n}(X_i-\mu)^2] - 2(\overline{X}-\mu)(n\overline{X}-n\mu) + n(\overline{X}-\mu)^2 \\ &= [\sum_{i=1}^{n}(X_i-\mu)^2] - 2n(\overline{X}-\mu)^2 + n(\overline{X}-\mu)^2 \\ &= [\sum_{i=1}^{n}(X_i-\mu)^2] - n(\overline{X}-\mu)^2.\end{aligned}$$

Entonces

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

Distribución del Estimador de la Varianza

Dividiendo por σ^2 nos queda

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 - \left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 .$$

El primer término del lado derecho es la suma de n v.a.i. $N(0,1)$ elevadas al cuadrado, y el segundo término es una v.a. $N(0,1)$ también elevada al cuadrado. ¿Qué distribución tiene una suma de variables aleatorias ji-cuadradas independientes con un grado de libertad? Una Ji-cuadrada con n grados de libertad.

Nota: De hecho una suma de l variables aleatorias Ji-cuadradas independientes con n_i grados de libertad $(i = 1, \ldots, l)$ es una variable aleatoria Ji-cuadrada con $m = n_1 + n_2 + \ldots + n_l$ grados de libertad. Verificar.

Distribución del Estimador de la Varianza

Como tenemos que $\frac{(n-1)S^2}{\sigma^2}$ es igual a la resta de dos variables aleatorias con distribuciones χ^2_n y χ^2_1, no conocemos su distribución.

Aunque en la nota anterior se menciona cual es la distribución de una suma de variables aleatorias Ji-cuadrada independientes, no es tan evidente cuál es la distribución de las variables restadas. En primer lugar no hay garantía de independencia, y en segundo lugar si las variables inmiscuidas fueran independientes, la generatriz de la combinación no es una reconocible.

Distribución del Estimador de la Varianza

Sin embargo, si vemos la ecuación de la siguiente manera

$$\frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2,$$

existe un teorema que garantiza la independencia de los dos términos de la izquierda. Considerando esto podemos calcular la generatriz de la combinación de las dos variables e igualarla a la generatriz de la variable de la derecha.

Si llamamos $U = \frac{(n-1)S^2}{\sigma^2}$, $V = (\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}})^2$ y $W = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2$, tenemos que $U + V = W$ y entonces

$$M_U(t) \cdot M_V(t) = M_{U+V}(t) = M_W(t).$$

Distribución del Estimador de la Varianza

Como $V \sim \chi^2_1 \equiv Gamma(\alpha = \frac{1}{2}, \beta = 2)$ y $W \sim \chi^2_n \equiv Gamma(\frac{n}{2}, 2)$, podemos sustituir las generatrices correspondientes

$$M_U(t) \cdot (1-2t)^{-\frac{1}{2}} = (1-2t)^{-\frac{n}{2}}.$$

Despejando $M_U(t)$, se tiene que

$$M_U(t) = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{n-1}{2}},$$

la cual corresponde a la generatriz de una distribución Gamma con $n-1$ grados de libertad. Es decir que

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \equiv Gamma(\alpha = \frac{n-1}{2}, \beta = 2).$$

(Nota que cuando formamos el estimador de la varianza como S^2, la suma esta dividida por $n-1$ y los grados de libertad asociados aquí también son $n-1$. Esto no se obtiene por casualidad).

Distribución del Estimador de la Varianza

Ahora bien, recordemos que anteriormente se vio que si se tiene una v.a. $Y \sim Gamma(\alpha = \frac{\nu}{2}, \beta = 2)$, entonces un múltiplo de Y cumple $aY \sim Gamma(\alpha = \frac{\nu}{2}, \beta^* = a\beta = 2a)$, que es distinta de una χ^2.

Si identificamos a $\frac{(n-1)S^2}{\sigma^2}$ con Y y tomamos $a = \frac{\sigma^2}{n-1}$, entonces $aY = S^2$. Así,

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim Gamma(\alpha = \frac{n-1}{2}, \beta = 2),$$

lo que implica que

$$aY = \frac{\sigma^2}{n-1} \cdot \frac{n-1}{\sigma^2} S^2 = S^2 \sim Gamma(\alpha = \frac{n-1}{2}, \beta = \frac{2\sigma^2}{n-1}).$$

En general preferimos trabajar con una Ji-cuadrada que con una Gamma.

Distribución del Estimador de la Varianza

Como conocemos que la varianza de una v.a. $Gamma(\alpha, \beta)$ es $\alpha\beta^2$, tenemos que

$$V(S^2) = \underbrace{\frac{n-1}{2}}_{\alpha} \underbrace{\left(\frac{2\sigma^2}{n-1}\right)^2}_{\beta^2} = \frac{2\sigma^4}{n-1}.$$

La siguiente figura muestra en forma esquemática las distribuciones muestrales para los estimadores de la media y de la varianza, obtenidos de muestras de poblaciones normales.

Distribución del Estimador de la Varianza

Distribución del Estimador de la Varianza

Observación:

Hemos demostrado que la distribución de una v.a. S^2 es una distribución Gamma, la cual usualmente es sesgada a la derecha. Por ello el valor promedio es "jalado" por la cola "pesada" y existe una mayor posibilidad de obtener un valor alejado del promedio (σ^2); esto es, S^2 generalmente subestima el valor de la varianza.

Sin embargo, si el tamaño de muestra n es grande, la variación de S^2 es menor y la curva tiende a ser más simétrica, más "acampanada", algo como una normal, y la subestimación tiende a desaparecer. Esto concuerda con el T.L.C. ya que S^2, en última instancia, se expresa como un promedio de variables aleatorias independientes, $Y_i = (X_i - \overline{X})^2$.

Distribución del Estimador de la Varianza

El hablar de algo llamado precisión en la estimación implica que deberá existir su contraparte: error en la estimación. Esto depende al menos de dos factores, tamaño de muestra y distribución de la población.

Según nuestro grado de precisión establecido, la forma de la población y los objetivos planteados, podremos fijar un tamaño de muestra. Más adelante estudiaremos algunos casos particulares.

Nota: En el caso de no tener una población normal se tiene la siguiente relación para la varianza de S^2

$$V(S^2) = \sigma^4 \left(\frac{2}{n-1} + \frac{E(X-\mu)^4 - 3\sigma^4}{n\sigma^4} \right) = \sigma^4 \left(\frac{2}{n-1} + \frac{\delta}{n} \right),$$

donde $\delta =$ coeficiente de curtosis.

Distribución del Estimador de la Varianza

Habíamos dicho que si δ era aproximadamente cero, nuestra curva se asemejaba más a la de una normal. En esta fórmula, si $\delta = 0$, la varianza de S^2 se reduce a la varianza correspondiente para una población normal $(\frac{2\sigma^4}{n-1})$.

Denotamos con $\chi^2_{\alpha,\nu}$ al numero real que satisface que

$$P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$$

Por ejemplo, si $\alpha = 0.95$ y los grados de libertad son $\nu = 15$, tenemos que $\chi^2_{0.95,15} = 7.261 \Rightarrow P(\chi^2_{15} > 7.26) = 0.95$.