

Inferencia Estadística

Dra. Graciela González Farías
Dr. Ulises Márquez



Maestría en Cómputo
Estadístico.

CIMAT Monterrey.



Agradecimientos

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
- Juan Antonio López
- Sigfrido Iglesias González
- Rodrigo Macías Paéz
- Edgar Jiménez
- Todos los estudiantes que han colaborado con sugerencias y comentarios sobre estas notas.

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

Temario

- 1 Variables aleatorias y distribuciones de probabilidad.
 - a) Distribuciones de probabilidad de variables aleatorias discretas.
 - b) Procesos de Poisson.
 - c) Distribuciones de probabilidad de variables aleatorias continuas.
 - d) Métodos gráficos para la identificación de distribuciones.
 - e) Estimación de densidades.
 - f) Distribuciones de probabilidad de vectores aleatorios.
 - g) Esperanzas condicionales y regresión.
 - h) Modelos jerárquicos, compuestos y mezclas de variables aleatorias.
 - i) Transformaciones de variables aleatorias.
 - j) Simulación de variables aleatorias.
 - k) Convergencia de variables aleatorias y el Teorema del Límite Central.

Temario

- 2 Distribuciones muestrales y métodos de estimación.
 - a) Propiedades de los estimadores.
 - b) Estimadores no insesgados
 - c) Distribuciones muestrales.
 - d) Principio de máxima verosimilitud.
 - e) Estimación puntual.
 - f) Bootstrap y jackknife.

Temario

- ③ Pruebas de Hipótesis e intervalos de confianza.
 - a) Definición de conceptos.
 - b) Potencia de la prueba.
 - c) Pruebas para dos poblaciones normales independientes.
 - d) Pruebas para medias en muestras pareadas.
 - e) Pruebas básicas de varianzas.
 - f) Pruebas para proporciones.
 - g) Conceptos de estimación bayesiana.
 - h) 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 Modelos gráficos probabilistas.
 - ⑤ Entre muchos otros.

Evaluación y acreditación

- Dos exámenes parciales, 18 de septiembre y 6 de noviembre: **15 %, cada uno.**
- 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: **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, \dots, 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:

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

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

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

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

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

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**.

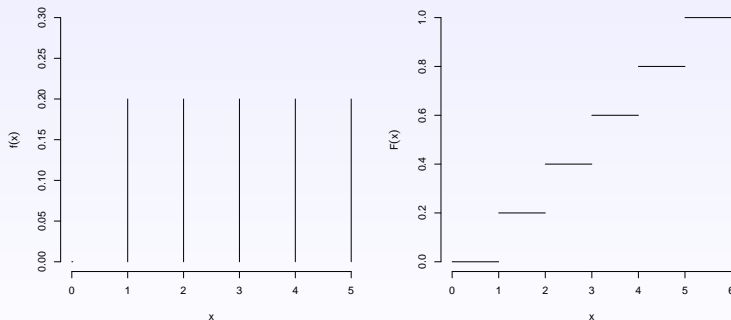


Figura: Distribución uniforme para $n = 5$.

Distribución uniforme discreta

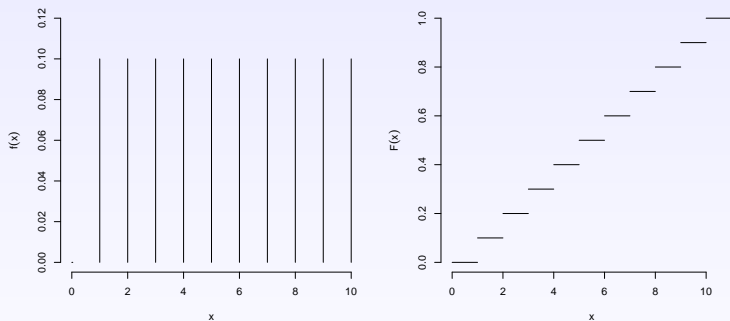


Figura: Distribución uniforme $n = 10$.

Distribución uniforme discreta

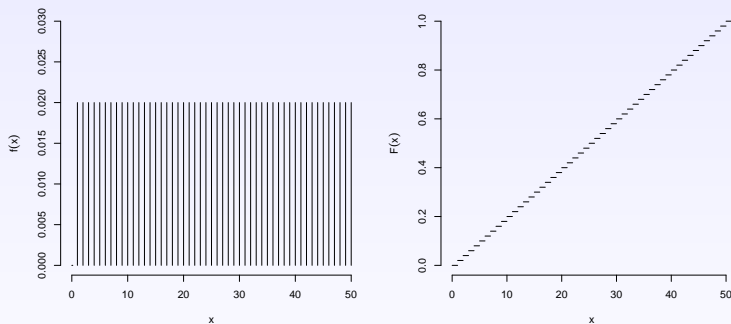


Figura: Distribución uniforme $n = 50$.

Distribución uniforme discreta

Media y Varianza:

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

y

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

Distribución uniforme discreta

Función Generatriz de Momentos:

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

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_\(probability_theory\)](https://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,

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

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

Distribución Bernoulli

Por ejemplo:

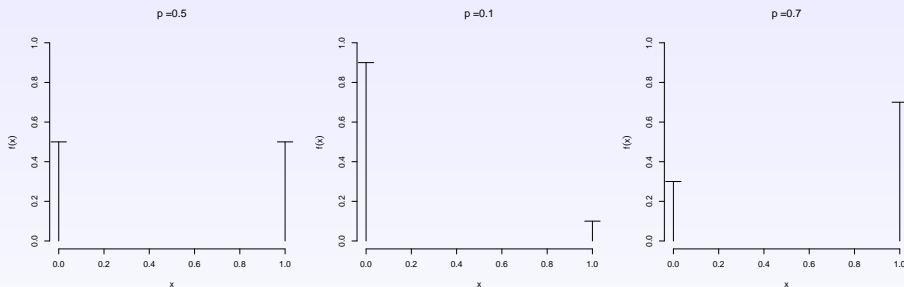


Figura: Función de densidad para $X \sim \text{Bernoulli}(p)$.

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:

$$E(X) = \sum_{x=0}^1 x \cdot f(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$E(X^2) = \sum_{x=0}^1 x^2 \cdot f(x) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

$$V(X) = E(X^2) + [E(X)]^2 = p - p^2 = p(1 - p) = pq$$

Función Generatriz de Momentos:

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

Distribución Binomial e Hipergeométrica

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

$X = \#$ de “éxitos” en n 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 $\{1, 2, \dots, n\}$. Ahora obtengamos la ley de comportamiento asociada con esta variable:

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

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:

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

A manera de ejercicio, puedes verificar que:

$$\begin{aligned} P(X = 1) &= P(\{X_1 = 1, X_2 = 0, X_3 = 0\} \cup \{X_1 = 0, X_2 = 1, X_3 = 0\} \cup \{X_1 = 0, X_2 = 0, X_3 = 1\}) \\ &= 3pq^2 = 3pq^2 = \binom{3}{1} p^1 q^2, \quad \mathbf{1 \text{ "éxito"}}; \end{aligned}$$

$$\begin{aligned} P(X = 2) &= P(\{X_1 = 1, X_2 = 1, X_3 = 0\} \cup \{X_1 = 1, X_2 = 0, X_3 = 1\} \cup \{X_1 = 0, X_2 = 1, X_3 = 1\}) \\ &= 3p^2q = 3p^2q = \binom{3}{2} p^2 q, \quad \mathbf{2 \text{ "éxitos"}}; \end{aligned}$$

$$\begin{aligned} P(X = 3) &= P(\{X_1 = 1, X_2 = 1, X_3 = 1\}) \\ &= ppp = 3p^3 = \binom{3}{3} p^3, \quad \mathbf{3 \text{ "éxitos"}}. \end{aligned}$$

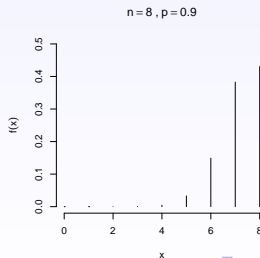
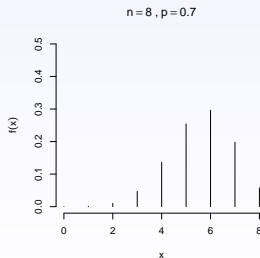
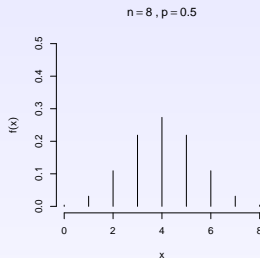
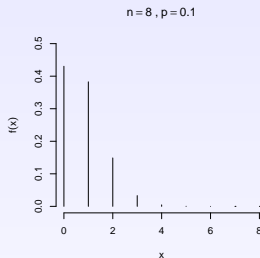
Distribución Binomial

Esto es, tenemos la **Distribución Binomial**

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

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:

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

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

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

Distribución Binomial

Media y Varianza

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

$$M'_X(t) = n(q + pe^t)^{n-1}(pe^t),$$

$$M'_X(0) = n(q + pe^0)^{n-1}(pe^0).$$

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

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

y, por lo tanto,

$$E(X) = np.$$

Distribución Binomial

Ahora,

$$M_X''(t) = n(q + pe^t)^{n-1}pe^t + pe^t [n(n-1)(q + pe^t)^{n-2}pe^t],$$
$$E(X^2) = M_X''(0) = np + n(n-1)p^2.$$

entonces

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

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

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

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

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

y su varianza

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

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

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

Distribución Binomial

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

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

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

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

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. $\text{Binomial}(n, p)$.

¿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

$X = \#$ de “éxitos” en n 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 $X = \# \text{ de botes con leche en mal estado}$.

- 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 .

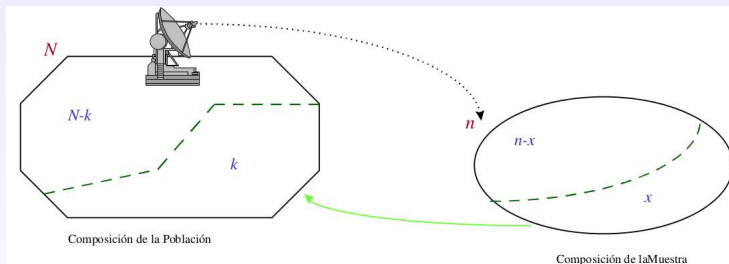
Distribución Hipergeométrica

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 | ← Inferencia | x elementos con la característica de interés |
| $N - k$ sin la característica | ← 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](#)):

$$f(x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}, \quad \max\{0, n - (N - k)\} \leq x \leq \min\{n, k\}.$$

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.

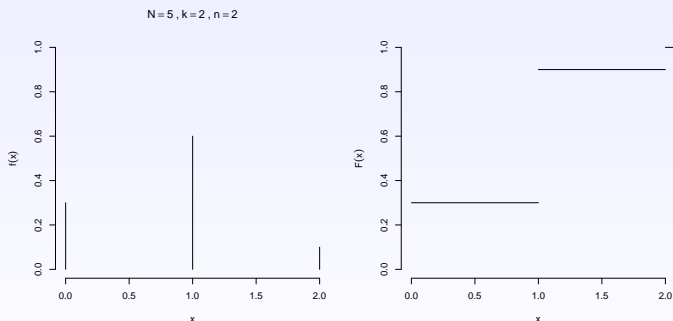


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(5, 2, 2)$.

Distribución Hipergeométrica

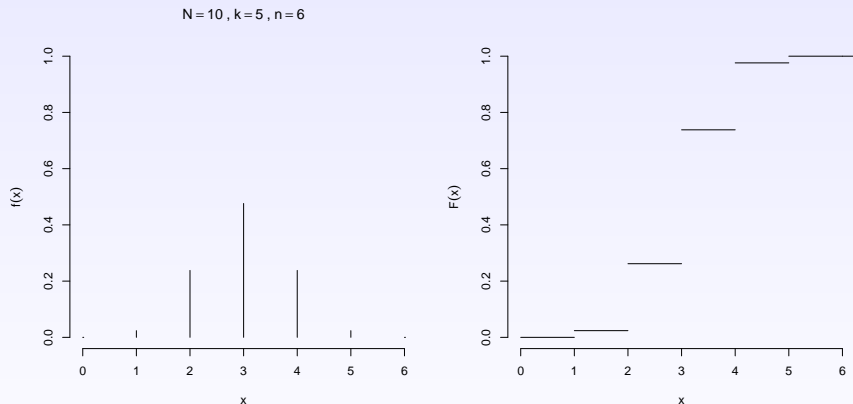


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(10, 5, 6)$.

Distribución Hipergeométrica

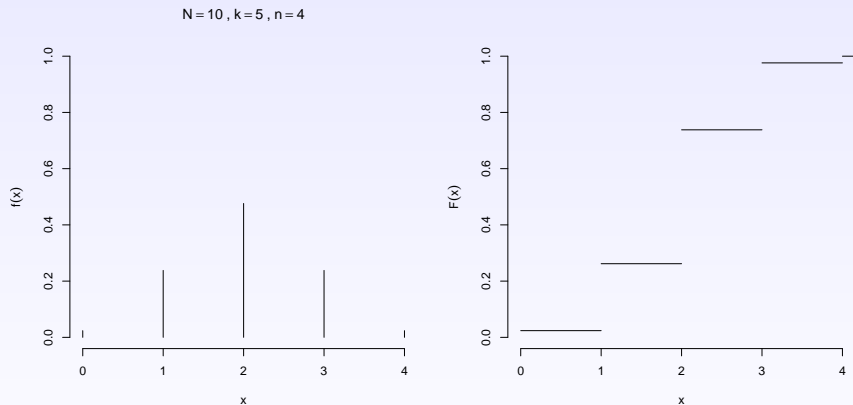


Figura: Funciones de masa y de distribución acumulada para $X \sim \text{Hiper}(10, 5, 4)$.

Distribución Hipergeométrica

Media y Varianza

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

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

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 2).

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

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

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.

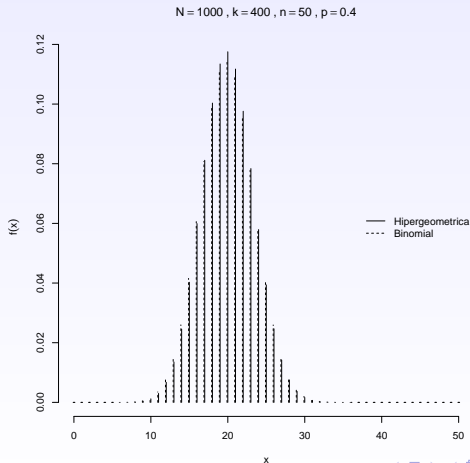
Es posible demostrar formalmente que si $p = k/N$ se mantiene constante entonces

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

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.

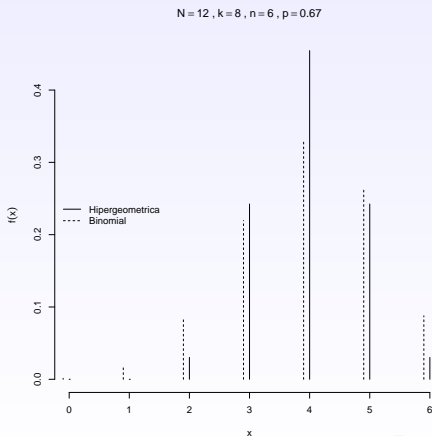
Distribución Hipergeométrica

Por ejemplo, tenemos que la aproximación binomial para $\text{Hiper}(1000, 400, 50)$ 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 $\lim_{N \rightarrow \infty} \frac{N-n}{N-1}$, 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

$X = \#$ de observaciones **hasta obtener un 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 hasta obtener un éxito.

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 $(1 - p = q)$:

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

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

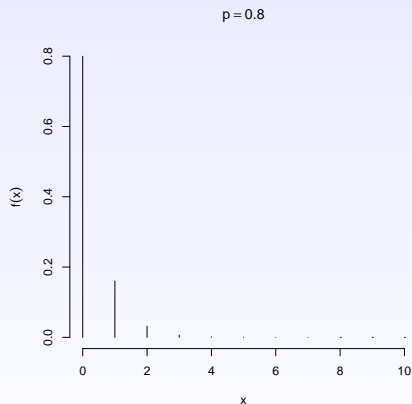
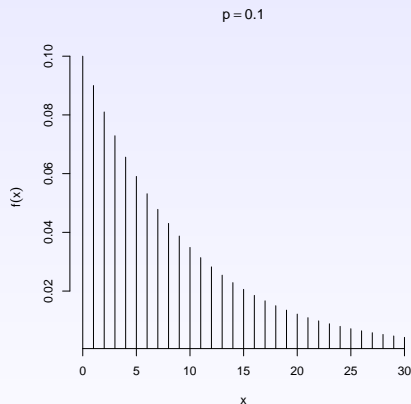


Figura: Funciones de masa para $\text{Geom}(0.1)$ y $\text{Geom}(0.8)$.

Distribución Geométrica

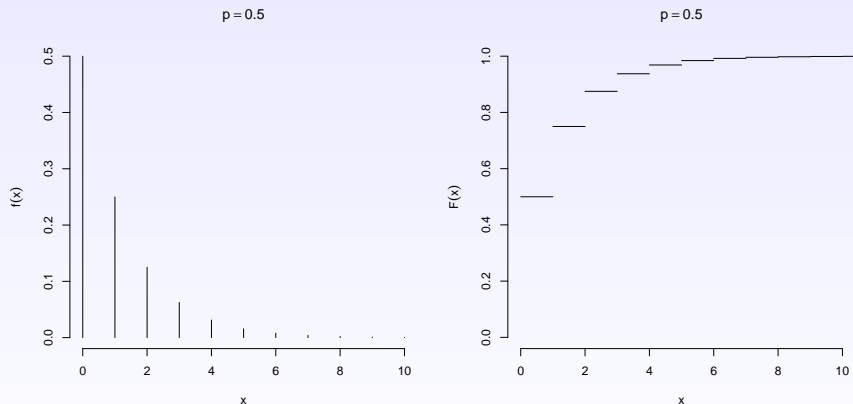


Figura: Funciones de masa y de distribución acumulada para $\text{Geom}(0.5)$.

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,

$$\begin{aligned}
 P(X > a + b | X > a) &= \frac{P(X > a + b, X > a)}{P(X > a)} && \text{pues } P(A|B) = \frac{P(A \cap B)}{P(B)} \\
 &= \frac{P(X > a + b)}{P(X > a)} && \text{ya que } X > a + b \text{ es la intersección entre los dos intervalos del numerador} \\
 &= \frac{1 - P(X \leq a + b)}{1 - P(X \leq a)} \\
 &= \frac{1 - F(a + b)}{1 - F(a)}.
 \end{aligned}$$

Por otro lado, la función de distribución de una v.a. $\text{Geom}(p)$ 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

$$P(X > 1 + 2 | X > 1) = P(X > 2),$$

$$P(X > 1 + 2 | X > 2) = P(X > 1),$$

$$P(X > 1 + 1 | X > 1) = P(X > 1).$$

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

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

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

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

Distribución Geométrica

Para el cálculo de la varianza, nos queda

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

pero

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

Distribución Geométrica

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

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

y entonces

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

Distribución Geométrica

Función Generatriz de Momentos:

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \sum_{x=1}^{+\infty} e^{tx} \cdot f(x) = \sum_{x=1}^{+\infty} e^{tx} \cdot q^{x-1} p = p \sum_{x=1}^{+\infty} (e^t)^x \cdot q^{x-1} \\
 &= pe^t \sum_{x=1}^{+\infty} (e^t)^{x-1} \cdot q^{x-1} = pe^t \sum_{x=1}^{+\infty} (e^t q)^{x-1} \underset{(j=x-1)}{=} pe^t \sum_{j=0}^{+\infty} (e^t q)^j,
 \end{aligned}$$

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

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

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

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

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, $E(X) = \frac{150,000}{4} = 37,500$. 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

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

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

$X = \#$ de observaciones **hasta** encontrar r “éxitos”,

la cual se puede redefinir como

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

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:



Media y Varianza

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

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

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

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

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

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

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 .

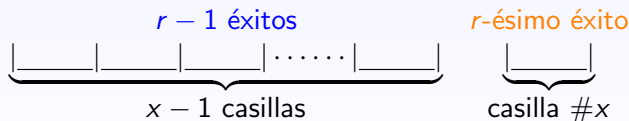
Puesto que la condición es encontrar r defectuosos, se tienen que realizar mínimo r pruebas.

- 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



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

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

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

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

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,

$$P(X \leq r - 1).$$

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

$$P(Y > n).$$

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

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

Utilizando la relación con la distribución binomial, para

$X \sim B(n = 10, p = 0.05)$ tenemos que

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

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 $X \sim \text{BN}(r, p)$. Para calcular la generatriz de momentos se utilizarán los siguientes resultados:

$$\binom{n}{z} = \binom{n}{n-z}, \quad (1)$$

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

Distribución Binomial Negativa

La generatriz de momentos es entonces

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

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,

$$\mu = \mu'_1 = E(X) = M'_X(t)|_{t=0},$$

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

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

Distribución Binomial Negativa

Por lo tanto,

$$E(X) = r/p.$$

Por otra parte,

$$\mu'_2 = E(X^2) = M''_X(t)|_{t=0},$$

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

$$\begin{aligned} M''_X(0) &= \frac{(1 - q)^{r+1} r^2 p^{r-1} p - rp^r (r+1)(1 - q)^r (-q)}{(1 - q)^{2(r+1)}} \\ &= \frac{p^{r+1} r^2 p^r + r(r+1)p^r p^r q}{p^{2r+2}} \\ &= \frac{p^{2r+1} r^2 + r(r+1)p^{2r} q}{p^{2r+2}} = \frac{rp^{2r}(rp + (r+1)q)}{p^{2r+2}} \\ &= \frac{r \left(\overbrace{rp + rq}^r + q \right)}{p^2} = \frac{r(r+q)}{p^2}. \end{aligned}$$

Distribución Binomial Negativa

Así que

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

Entonces,

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

lo que implica

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

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 $E(Y) = np$, 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 $X \sim B(n, p)$, entonces

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

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

Distribución Poisson

Por otro lado,

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

El único término faltante es $\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$. Observe que $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}$ 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,

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

aplicando la Regla de L'Hopital

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

Distribución Poisson

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

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

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

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

De aquí que

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

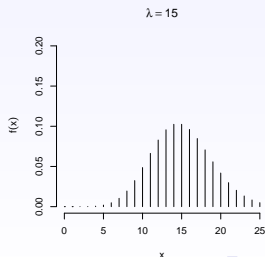
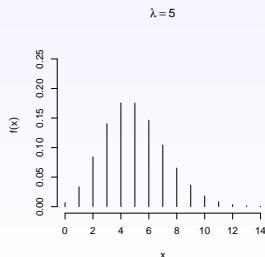
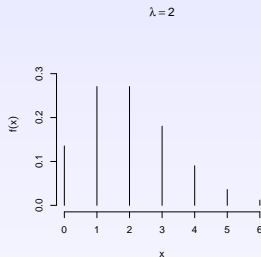
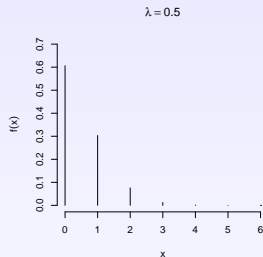
Distribución Poisson

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

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

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

Distribución Poisson



Distribución Poisson

¿Qué podemos aprender del procedimiento anterior?

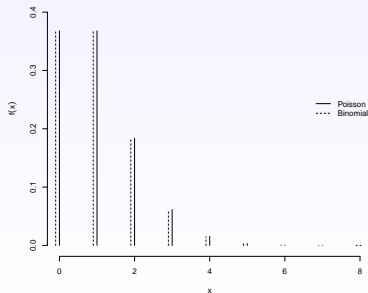
- 1 La fórmula para asignar probabilidades, como siempre, es **válida bajo una serie de supuestos que debemos verificar se cumplan razonablemente** en nuestros estudios particulares. Por otro lado, es el resultado de un proceso matemático no necesariamente intuitivo, como lo había sido en los casos anteriores.
- 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

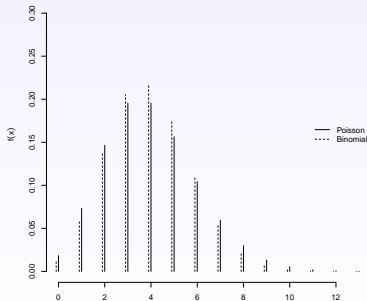
La variable aleatoria Poisson se puede usar para calcular probabilidades binomiales en forma aproximada cuando n es muy grande y p es pequeña.

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.

$\lambda = 1$ $n = 100$, $p = 0.01$



$\lambda = 4$ $n = 20$, $p = 0.2$



Distribución Poisson

Función generatriz de momentos:

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

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 $E(X^r)$, obtenemos la media y la varianza

$$\begin{aligned} E(X) &= \mu = \mu'_1 = M'_X(t)|_{t=0} \\ \Rightarrow M'_X(t) &= e^{\lambda(e^t-1)} \cdot \lambda e^t \Rightarrow M'_X(0) = e^{\lambda(e^0-1)} \cdot \lambda e^0 = \lambda. \end{aligned}$$

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

Ahora,

$$\begin{aligned} E(X^2) &= \mu'_2 = M''_X(t)|_{t=0} \\ \Rightarrow M''_X(t) &= e^{\lambda(e^t-1)} \cdot \lambda e^t + \lambda e^t e^{\lambda(e^t-1)} \cdot \lambda e^t \Rightarrow M''_X(0) = \lambda + \lambda^2. \end{aligned}$$

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

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

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

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,

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

entonces

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

Distribución Poisson

Ejemplo. Una compañía renta máquinas fotocopadoras. Se sabe que el número de reportes por averías por semana que recibe es una v.a. Poisson($\lambda = 2 \frac{\text{averías}}{\text{semana}}$). 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:

$Y = \#$ de quejas por semana.

- 1 Se pide $P(Y = 0) = 2^0 e^{-2} / 0! = 0.1354$.
- 2 Aquí lo que interesa es $P(X > 3 | X \geq 1) = \frac{P(X > 3, X \geq 1)}{P(X \geq 1)} = \frac{P(X > 3)}{P(X \geq 1)} = \frac{1 - F(3)}{1 - F(0)} = \frac{1 - 0.857}{1 - 0.1354} = 0.1654$.
- 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

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

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 parámetro en la función de masa de una binomial negativa, para esto es necesario realizar la siguiente reparametrización.

Distribución Binomial Negativa: Sobredispersión

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

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

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

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

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

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

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,

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

Se puede ver que

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

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

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 $\{0, 1, 2, \dots\}$. Pongamos $p_n(t) = P(X_t = n)$, $n = 0, 1, \dots$

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 $o(\Delta t)/\Delta t \rightarrow 0$ 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 $p_k(\Delta t) \rightarrow 0$, $k \geq 2$.
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

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

Resolviendo la ecuación diferencial anterior llegamos a

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

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

$$p'_n(t) = \lambda p_{n-1}(t) - \lambda p_n(t), \quad n = 1, 2, \dots$$

Definamos $q_n(t) = e^{\lambda t} p_n(t)$. Así el sistema de ecuaciones diferenciales anterior se transforma en $q'_n(t) = \lambda q_{n-1}(t)$, $n = 1, 2, \dots$. 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

$$p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n=0,1,2,\dots$$

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 aleatoria 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 $C - 2 - 2t = 0$, 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 $N(t + s) - N(s) \sim \text{Poisson}(\lambda t)$.
- 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 transfusiones 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:

$$\text{En } [0,4]: \lambda = \int_{t1}^{t2} \alpha(t)dt = \int_0^4 e^{2+0.6t} dt = 123.44$$

$$\text{En } [2,6]: \lambda = \int_{t1}^{t2} \alpha(t)dt = \int_2^6 e^{2+0.6t} dt = 409.82$$

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

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

$P(X \leq 15) = F(15, 10) = 0.978$, 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.

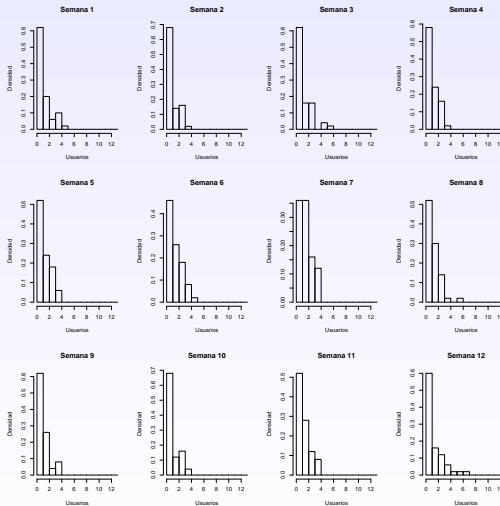
Con esta finalidad, la empresa recopiló información sobre el número de usuarios que se conectan por hora (24 horas) durante 12 semanas.

Procesos de Poisson

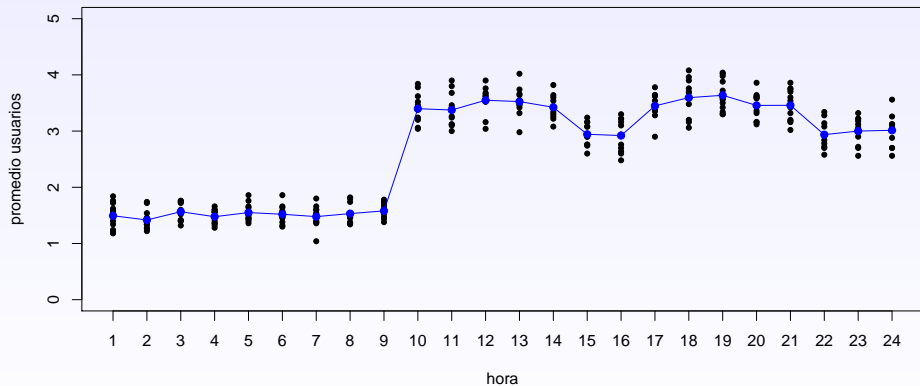
Por ejemplo, para la semana 1:

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

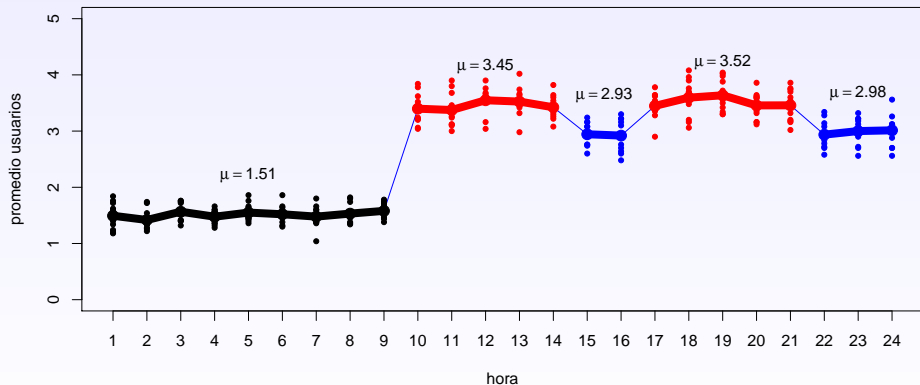
Ejemplo. Análisis exploratorio



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 $t = 0, 1, 2, \dots$

Ejemplo. Selección del modelo

En nuestro caso:

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

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 $P[X \leq x]$,
 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

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

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).

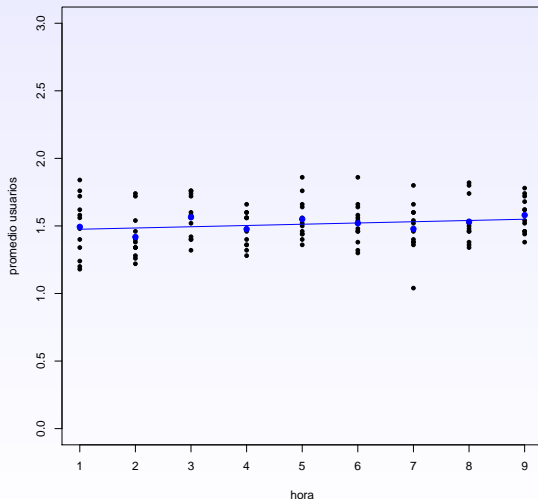
Probemos con un modelo lineal. Para los datos de promedios en ese intervalo de tiempo, obtenemos:

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

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 asumíamos 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.

Proceso de Poisson: Retomando...

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 transfusiones sanguíneas.

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

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

Proceso de Poisson: Retomando...

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:

$$\text{En } [0,4]: \lambda = \int_{t1}^{t2} \alpha(t)dt = \int_0^4 e^{2+0.6t} dt = 123.44$$

$$\text{En } [2,6]: \lambda = \int_{t1}^{t2} \alpha(t)dt = \int_2^6 e^{2+0.6t} dt = 409.82$$

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

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

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

Distribución Poisson

Nota Histórica

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

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

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 fuerza a asignar una probabilidad de $\frac{1}{b-a}$ a todo el intervalo.

Distribución uniforme continua

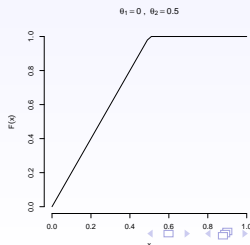
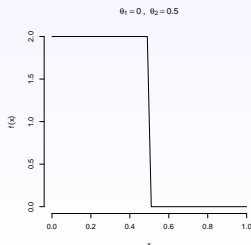
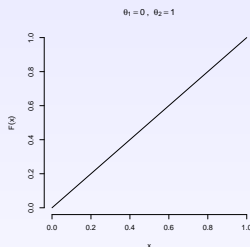
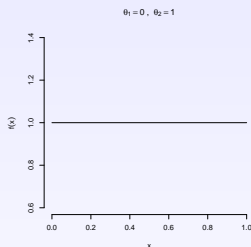
Definición

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

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

Distribución uniforme continua

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



Distribución uniforme continua

Media y Varianza:

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

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:

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

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:

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

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

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

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

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

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. $\text{Unif}(0, 1)$, 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 ,

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

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 “arrebato” 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

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

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 $\frac{1}{\sqrt{2\pi}}$.

- $$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1.$$

12 videos sobre como calcular la integral:

<https://www.youtube.com/playlist?list=PLJb1qAQIrrmmCgLyHWMXGZnioRHLqOk2bW>

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

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

la cual se obtiene integrando por partes dos veces.

El modelo Normal o Gaussiano

Entonces, la media y la varianza son cero y uno respectivamente, 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,

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

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

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

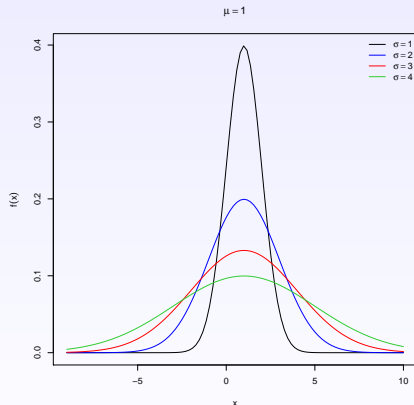
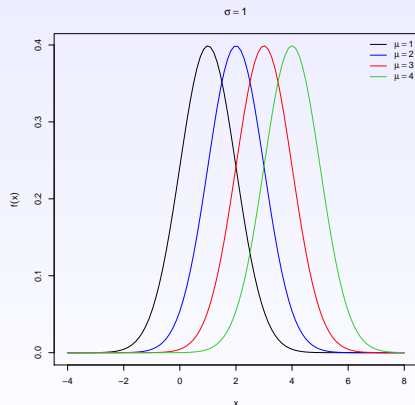
y su varianza

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

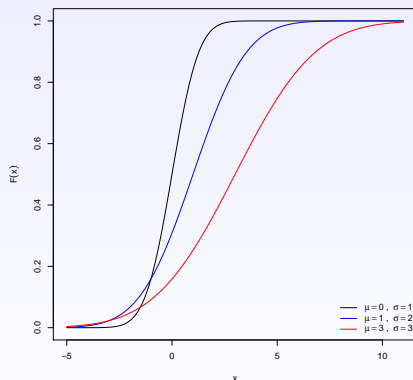
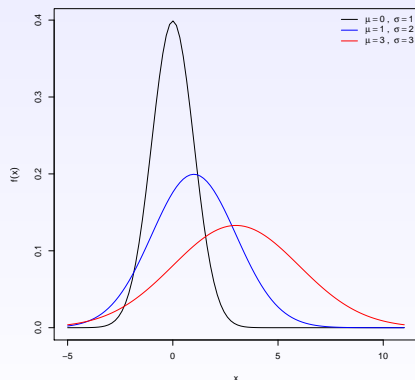
El modelo Normal o Gaussiano

Entonces, la media y la varianza de $X = Z\sigma + \mu$ son, respectivamente, μ y σ^2 . Dado que solo se está recorriendo y escalando la función original, la gráfica de esta nueva variable 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 normal 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, calcular 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.

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

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,

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

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 continuo.

El modelo Normal o Gaussiano

Derivando:

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

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

$$F^{-1}(x) = \inf \{x : F(x) > q\},$$

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

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

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 $X_1, \dots, X_n \sim F$ 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

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

Función de distribución empírica

Teorema

En cualquier valor fijo x ,

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

Teorema (Glivenko-Cantelli)

Sean $X_1, \dots, X_n \sim F$. 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, \dots, 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

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

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

Operativamente, debemos graficar

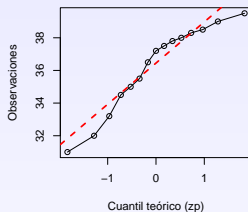
$$x_{(i)} \text{ vs } F^{-1}\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n.$$

Función de distribución empírica

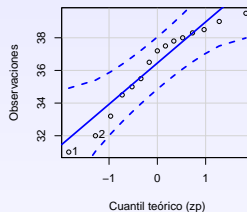
| x_i | rango= i | $p_i = \frac{i}{n+1}$ | $z_{p_i} = q_i$ tal que $P(Z \leq z_{p_i}) = p_i$ |
|-------|------------|-----------------------|---|
| 31 | 1 | 0.0625 | -1.53412 |
| 32 | 2 | 0.125 | -1.15035 |
| 33.2 | 3 | 0.1875 | -0.88715 |
| 34.5 | 4 | 0.25 | -0.67449 |
| 35 | 5 | 0.3125 | -0.48878 |
| 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

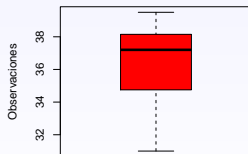
Normal Q-Q Plot



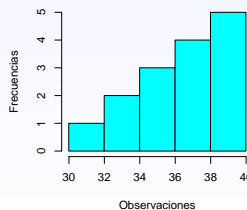
Bandas de confianza para Q-Q plot



Boxplot



Histograma



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 45° -así evitamos el cálculo de parámetros.

$$(\bar{X}, s^2) \rightarrow (\mu, \sigma^2)$$

$$Z_{p_i} = \frac{y_i - \bar{X}}{s} \quad \text{vs.} \quad X_i$$