

# Inferencia Estadística

Maestría en Análisis Estadístico y Computación

## Capítulo 1: Variables aleatorias y distribuciones de probabilidad

---

Dra. L. Leticia Ramírez Ramírez

Enero–Mayo, 2020



# Contenido

Sobre el curso

Distribuciones de probabilidad de variables aleatorias discretas

Distribución uniforme discreta

Distribución Bernoulli

Distribución Binomial

Distribución Hipergeométrica

Distribución Geométrica y Binomial Negativa

Proceso de Poisson

Función Generatriz de momentos

Distribuciones de probabilidad de variables aleatorias continuas

Distribución uniforme continua

El modelo Normal o Gaussiano

El modelo exponencial y Gamma

El modelo Beta

Transformación de variables

Suma

Censura y Truncamiento

Funciones de variables aleatorias

Métodos de transformación de variables

Método de la función de distribución acumulada

Método de la transformación o cambio de variable: Ejemplo

## Sobre el curso

---

# Agradecimientos

En forma de agradecimiento, se enlistan personas que han contribuido de una u otra forma en la construcción de todo el material utilizado, a través de los años:

- Graciela González Farías
- Ulises Márquez
- 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 en la bibliografía.**

## **Distribuciones de probabilidad de variables aleatorias discretas**

---

# Modelos probabilísticos

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**.

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).

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).



---

## Distribución uniforme discreta

## 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

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

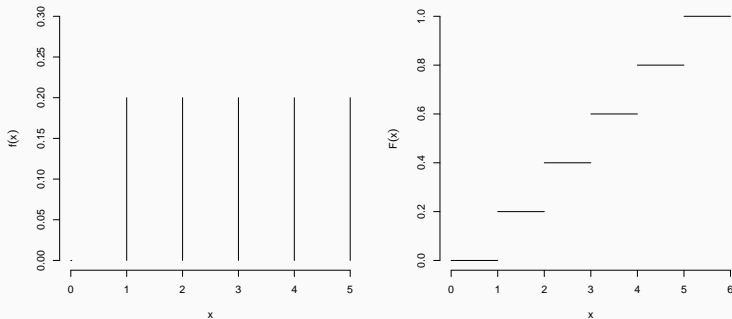
El caso más común de esta distribución es cuando la variable toma los valores enteros

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

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

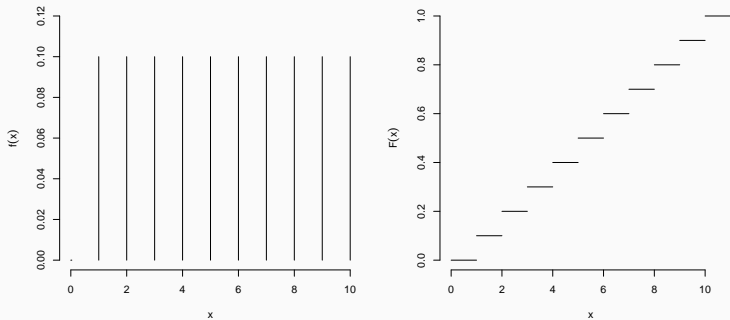
## Distribución uniforme discreta

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**.



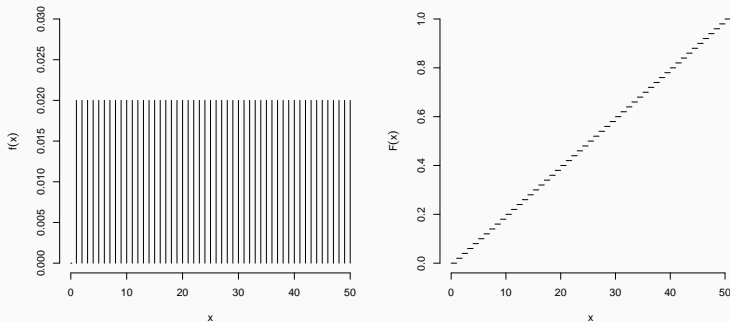
**Figure 1:** Distribución uniforme para  $n = 5$ .

# Distribución uniforme discreta



**Figure 2:** Distribución uniforme  $n = 10$ .

# Distribución uniforme discreta



**Figure 3:** 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 Bernoulli



## 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

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

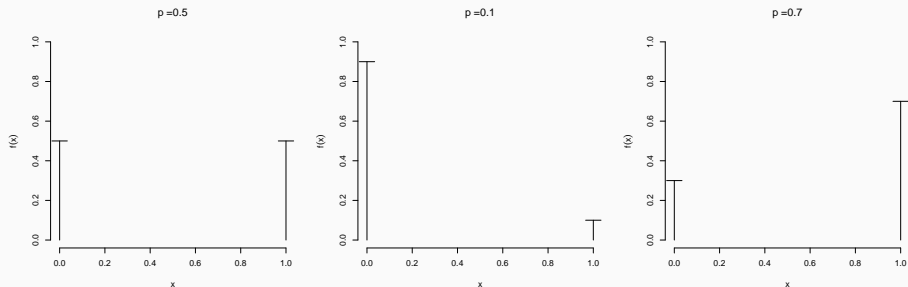
o bien,

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

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

# Distribución Bernoulli

Por ejemplo:



**Figure 4:** 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**.

## Media y Varianza:

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

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

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

---

## Distribución Binomial

## 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:

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

# 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

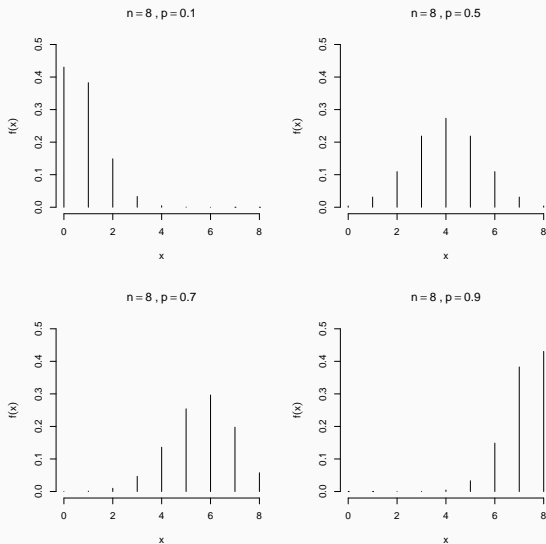


Figure 5: Distribución binomial

## 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

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

## 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. **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

# 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  $X = \#$  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.

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  $N$ .

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

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

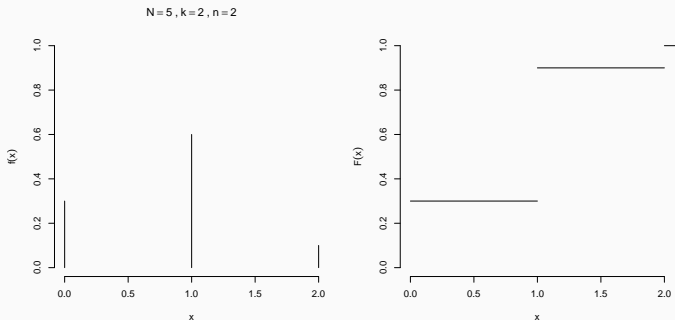
Por lo tanto, la probabilidad de observar  $x$  “éxitos” en  $n$  pruebas está dada por la función de probabilidad (**Distribución Hipergeométrica**):

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

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

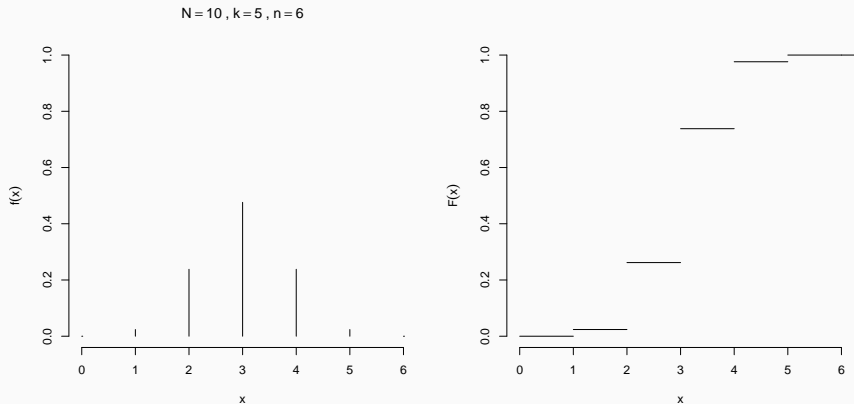
# Distribución Hipergeométrica

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



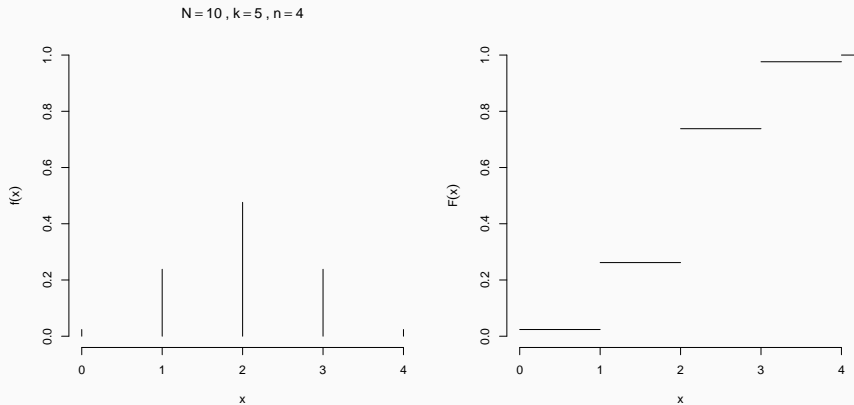
**Figure 6:** Funciones de masa y de distribución acumulada para  $X \sim \text{Hiper}(5, 2, 2)$ .

# Distribución Hipergeométrica



**Figure 7:** Funciones de masa y de distribución acumulada para  $X \sim \text{Hiper}(10, 5, 6)$ .

# Distribución Hipergeométrica



**Figure 8:** Funciones de masa y de distribución acumulada para  $X \sim \text{Hiper}(10, 5, 4)$ .

## 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

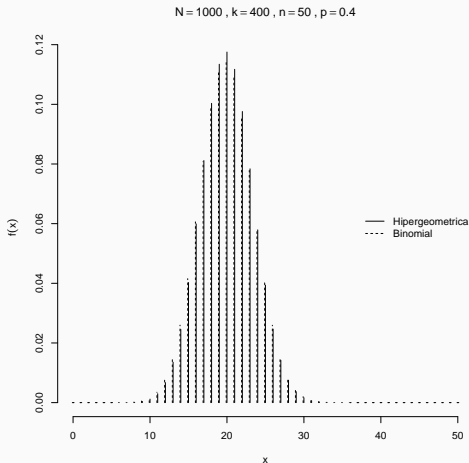
$$\lim_{N \rightarrow \infty} \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x}.$$

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 Geométrica y Binomial Negativa

# 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  $r$ .

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.

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  $(1 - p = q)$ :

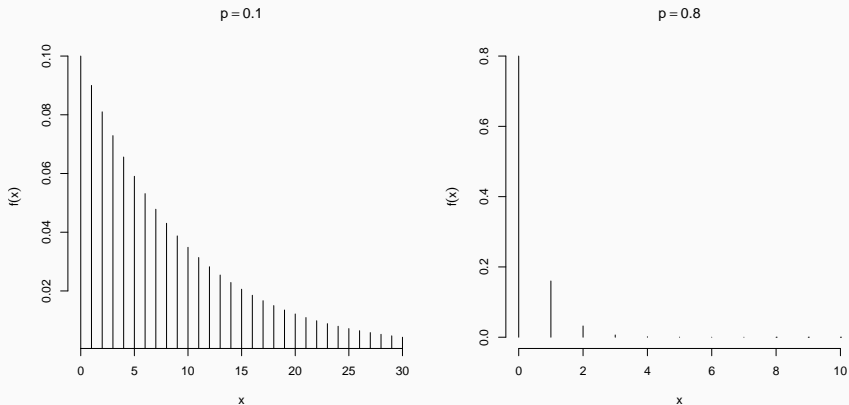
$$f(x) = q^{x-1}p, \quad x = 1, 2, 3, \dots \quad (\text{Distribución Geométrica})$$

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  $p$ .

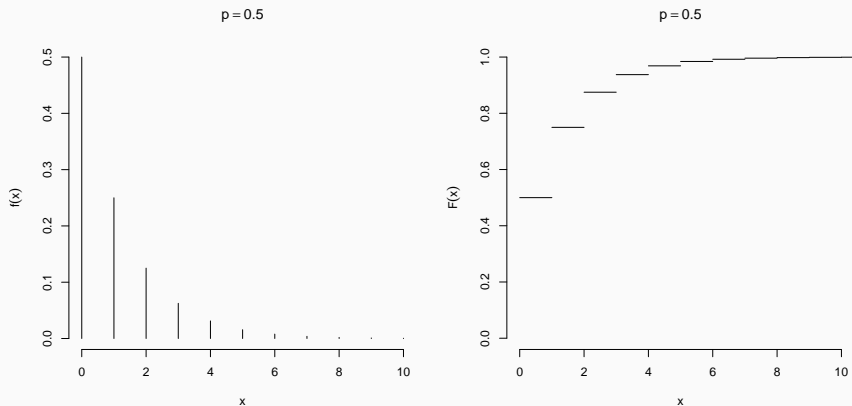


# Distribución Geométrica



**Figure 9:** Funciones de masa para  $\text{Geom}(0.1)$  y  $\text{Geom}(0.8)$ .

# Distribución Geométrica



**Figure 10:** 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*.

**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,

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

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.



## Media y Varianza

Ya que

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

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

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

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

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

**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  $X = \# \text{ de viajes necesarios hasta obtener 5 “éxitos”}$ , entonces

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

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 75), 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

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  $r$ .

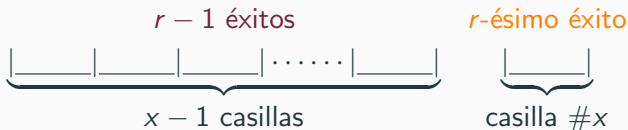
*Puesto que la condición es encontrar  $r$  defectuosos, se tienen que realizar **mínimo**  $r$  pruebas.*

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

---

Recuerda que

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

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

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

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

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

$X \sim \text{B}(n = 10, p = 0.05)$  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

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.

# Proceso de Poisson

---

# 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  $1\text{ m}^2$ ,
- Número de bacterias en  $1\text{ cm}^3$  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**”,  $\lambda$  (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  $\lambda_1$  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.*

*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  $\lambda$  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  $\lambda$  en el intervalo considerado.

Notemos que  $\lambda = np$  implica que  $p = \lambda/n$  y como  $\lambda$  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

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

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

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  $\lambda^x$  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

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

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

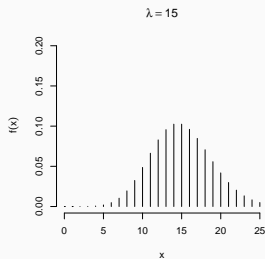
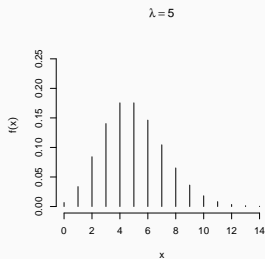
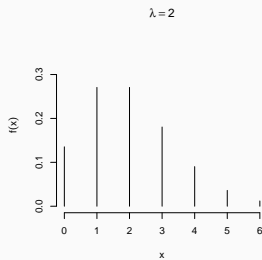
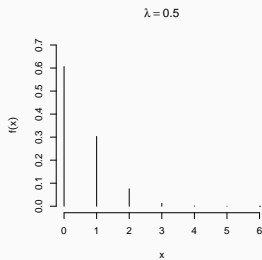
- $\left(1 - \frac{\lambda}{n}\right)^{-\frac{n}{\lambda}} \rightarrow e$ , cuando  $n \rightarrow \infty$ .

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  $\lambda$ .

# 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.**

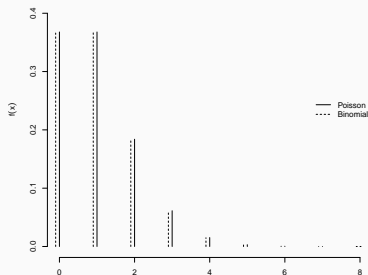


# Distribución Poisson

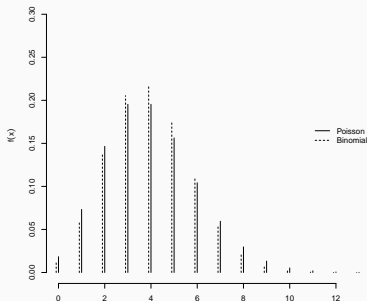
La variable aleatoria Poisson se puede usar para calcular probabilidades binomiales en forma aproximada cuando  $n$  es muy grande y  $p$  es pequeña.

**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

**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  $\pi r^2$ . La v.a.  $X$  se distribuye  $\text{Poisson}(\lambda = 2 \frac{\text{chapulines}}{m^2} \cdot \pi r^2)$ .  
(¿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.  $\text{Poisson}(\lambda = 2 \frac{\text{averías}}{\text{semana}})$ . 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  $\text{Poisson}(\lambda = 8)$ , así

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 Poisson

Consideremos una fuente de material radiactivo que emite partículas  $\alpha$ .

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  $p_n(t) = P(X_t = n)$ ,  $n = 0, 1, \dots$

## Distribución Poisson: Hipótesis

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  $\Delta t$  es suficientemente pequeña, donde  $\lambda$  es una constante positiva. Es decir que  $p_1(\Delta t) \sim \lambda \Delta t$ . También supondremos que  $\Delta t > 0$ .

*Esto es, si el intervalo es suficientemente pequeño, la probabilidad de obtener exactamente una emisión durante ese intervalo es directamente proporcional a la longitud del intervalo.*

## Distribución Poisson: Hipótesis (cont)

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  $p_k(\Delta t) \rightarrow 0$ ,  $k \leq 2$ .  
*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.

Saquemos algunas conclusiones de las hipótesis anteriores:

1. Las hipótesis 1 y 2 juntas implican que la variable aleatoria  $X_t$  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

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



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  $P(X_t = 0)$ .

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

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

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

$$q'_1(t) = \lambda, \text{ y por tanto } q_1(t) = \lambda t;$$

$$q'_2(t) = \lambda q_1(t) = \lambda^2 t, \text{ y por tanto } q_2(t) = \frac{(\lambda t)^2}{2}.$$

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  $\lambda t$ .**

## Distribución Poisson: Observaciones

- 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.*

- b) La constante  $\lambda$  **apareció originalmente como una constante de proporcionalidad en la hipótesis 3.**

*Vale la pena mencionar las siguientes interpretaciones de  $\lambda$ :*

- *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.*

- 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.*

## Distribución 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.



## Distribución 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.

## Distribución 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  $\alpha t$ . 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,

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

## Distribución 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  $(1-p)\alpha t$ .

### Teorema

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.

**Una forma más compleja, pero más realista es pensar que  $\lambda$  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.

## Distribución Poisson: Ejemplo

**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  $\Delta t$  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.**

## Distribución Poisson

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  $[0, 0.9907]$ ?

# Distribución Poisson

Solución.

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

$$\text{En } [0,4]: \lambda = \int_{t_1}^{t_2} \alpha(t) dt = \int_0^4 e^{2+0.6t} dt = 123.44$$

$$\text{En } [2,6]: \lambda = \int_{t_1}^{t_2} \alpha(t) dt = \int_2^6 e^{2+0.6t} dt = 409.82$$

2. Primero obtengamos el valor de  $\lambda$  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: 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.

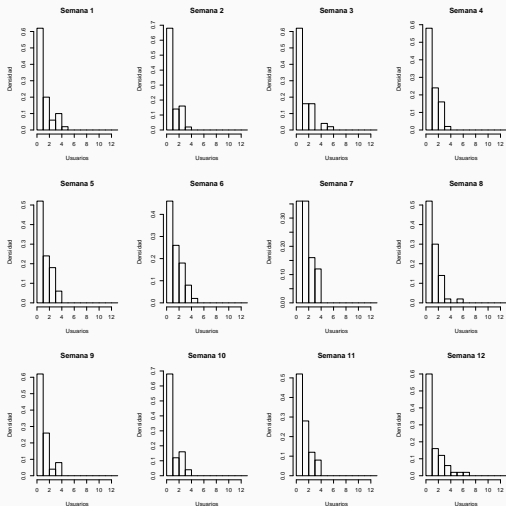
Con esta finalidad, la empresa recopiló información sobre el número de usuarios que se conectan por hora (24 horas) durante 12 semanas.

# Distribución Poisson

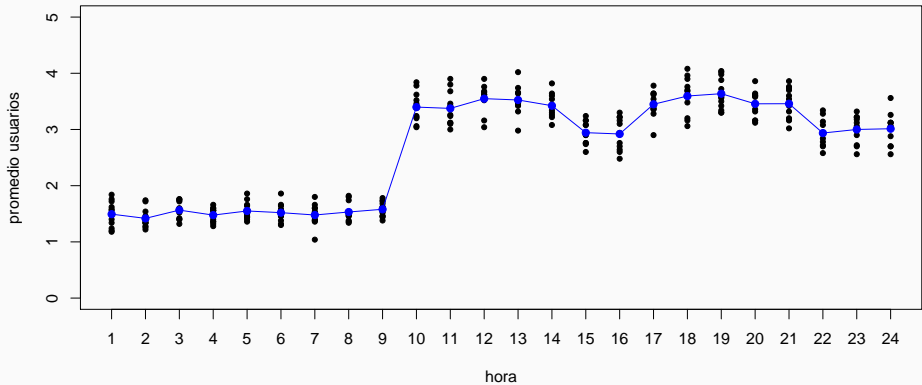
Por ejemplo, para la semana 1:

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	...

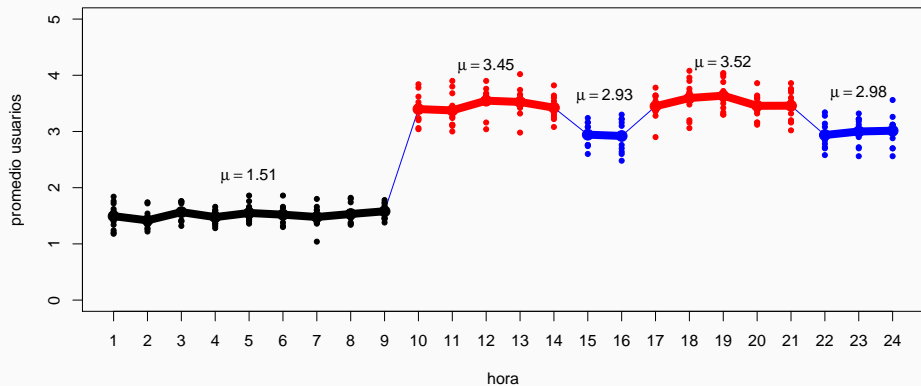
# Ejemplo. Análisis exploratorio



## 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  $\lambda k$ , que indica la intensidad (es decir, el número de usuarios esperado) del proceso en el intervalo deseado.

## 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  $\lambda k$ , 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:

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

## 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.

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

## 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

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

## 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

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

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

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

donde

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

Entonces

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

## Ejemplo. Aplicación del modelo

Entonces

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



## 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.

## 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, supongamos 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:

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

## Ejemplo. Aplicación del modelo

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

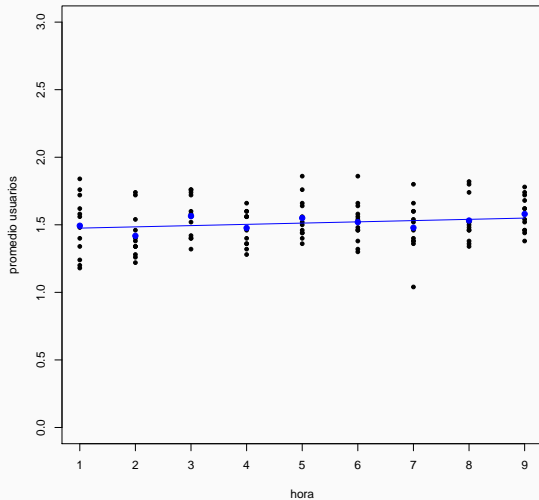
Probemos con un modelo lineal. Para los datos de promedios en ese intervalo de tiempo, obtenemos:

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

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:

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

donde

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

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.

- Hemos analizado la información para averiguar cómo se comportan las conexiones de los usuarios durante el día completo.



## 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.

## 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.

- 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?)

- 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

- 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.

## 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.

# Funcion Generatriz de momentos

---

## Funcion Generatriz de momentos

La función Generatriz es una de las funciones generadoras, que nos permiten caracterizar de forma única las distribuciones de v.a.'s. Además nos permiten calcular en forma rápida sus momentos y obtener fácilmente las funciones generatrices para operaciones como la suma.

### Definición

La función generatriz de la v.a.  $X$  (discreta o continua) se define como

$$M_X(t) = E(e^{Xt}), \quad t \in \mathbb{R}$$

siempre que  $E(|e^{Xt}|) < \infty$ .



## Funcion Generatriz de momentos

Si la función Generatriz existe en una vecindad de 0, entonces se puede utilizar para generar los momentos de  $X$ .

### Teorema

Si  $E(|e^{Xt}|) < \infty$  para toda  $t \in (-a, a)$  para  $a > 0$ , entonces

$$\mathbb{E}(X^n) = M_X^{(n)}(0) = \frac{d^n M_X}{dt^n}(0).$$

Con la FGM podemos también fácilmente obtener la FGM de la sum de v.a.'s independientes y que cuenten con sus FGM. Esto es,

### Resultado

Sea  $X_1, \dots, X_n$  v.a.'s independientes, donde  $M_{X_i}(t)$  existe para  $X_i \in \{X_1, \dots, X_n\}$ . Si  $Y = \sum_1^n X_i$ , entonces

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$$

## Funcion Generatriz de momentos: Uniforme

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

### Función Generatriz de Momentos:

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

## Funcion Generatriz de momentos: 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  $a = q$ ,  $b = e^t p$ :

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

# Funcion Generatriz de momentos: 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,

$$E(X) = np.$$

## Funcion Generatriz de momentos: 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

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

## 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

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



## 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  $|qe^t| < 1$ .

## Funcion Generatriz de momentos: 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:

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

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

## Funcion Generatriz de momentos: Binomial Negativa

La generatriz de momentos es entonces

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

## Funcion Generatriz de momentos: 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,

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

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

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

## Funcion Generatriz de momentos: Binomial Negativa

Por lo tanto,

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

Por otra parte,

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

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

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

# **Distribuciones de probabilidad de variables aleatorias continuas**

---

---

## Distribución uniforme continua

## Distribución uniforme

- Es uno de los modelos más 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.



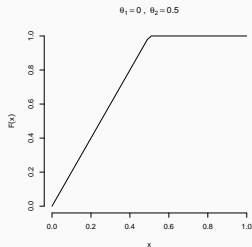
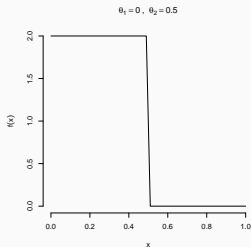
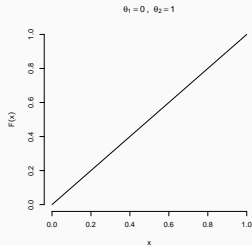
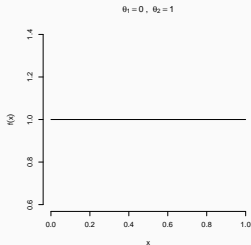
## Definición

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

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

# Distribución uniforme continua

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



## 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  $\frac{a+b}{2}$ .

## 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:

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

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...

## 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).

## Usando la Distribución uniforme(0,1) para simular v.a.'s

### 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.

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

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

## Usando la Distribución uniforme(0,1) para simular v.a.'s

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

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

## Usando la Distribución uniforme(0,1) para simular v.a.'s

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.

## Usando la Distribución uniforme(0,1) para simular v.a.'s

### 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<sup>1</sup>

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

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

---

<sup>1</sup> $y = 1 - e^{-\frac{1}{2}x} \Rightarrow 1 - y = e^{-\frac{1}{2}x} \Rightarrow -\frac{1}{2}x = \log(1 - y) \Rightarrow x = -2 \log(1 - y)$

## Usando la Distribución uniforme(0,1) para simular v.a.'s

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,  $F_Y(y) = y$ .



## Usando la Distribución uniforme(0,1) para simular v.a.'s

### Demostración

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

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

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

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

---

## El modelo Normal o Gaussiano

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  $d$ ,

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

# 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.

## 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  $\pm 1$ .
- Su valor máximo ocurre en  $z = 0$  y es  $\frac{1}{\sqrt{2\pi}}$ .
- $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 1$ .



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

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

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  $\infty$  el resultado es 1.

## Variable aleatoria normal con cualquier media y varianza.

Cada problema tiene su distribución propia. En particular, tiene su asociada una media  $\mu$  y una varianza  $\sigma^2$  propias.

Tenemos que la medida de localización  $\mu$  y la medida de dispersión  $\sigma$  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

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

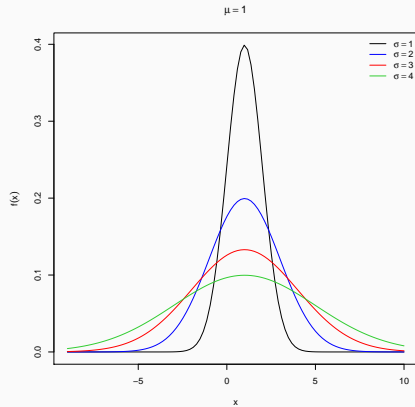
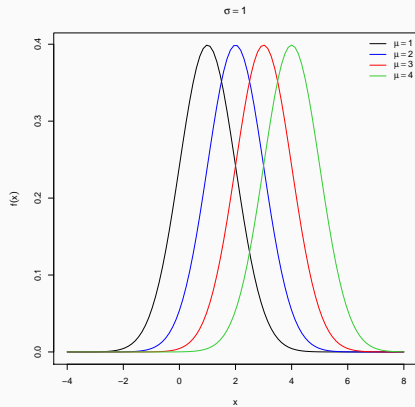
y su varianza

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

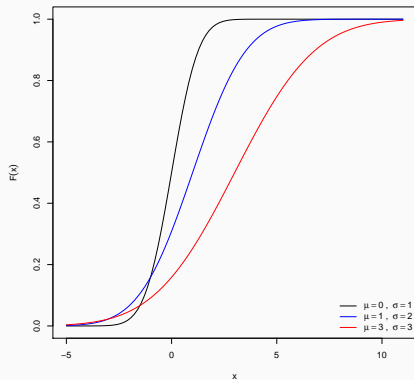
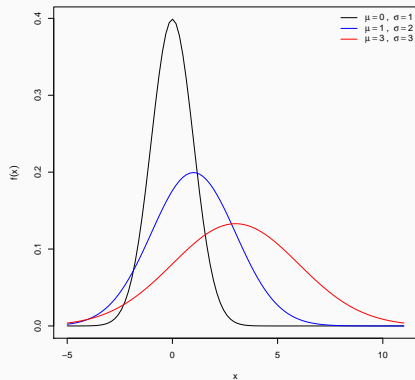
Entonces, la media y la varianza de  $X = Z\sigma + \mu$  son, respectivamente,  $\mu$  y  $\sigma^2$ . Dado que solo se está recorriendo y escalando la función original, la gráfica de esta nueva variable tendrá la misma forma (campana), solo que no estará centrada en cero, y tal vez esté mas “chata” 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  $\mu$  y dividiendo por  $\sigma$ , 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  $\mu$  y varianza  $\sigma^2$ , y se dice que se distribuye normal con parámetros  $\mu$  y  $\sigma$ .

**Notación:**  $X \sim N(\mu, \sigma^2)$ .

# El modelo Normal o Gaussiano

La generatriz de momentos es

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

donde usamos la propiedad demostrada en conceptos básicos.

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

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

**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  $\mu$  y  $\sigma^2$ , 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  $N(\mu, \sigma)$ .

### 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  $3.0 \pm 0.01$  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

## 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  $\chi^2$  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 exponencial y Gamma

## El modelo exponencial

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

### Definición

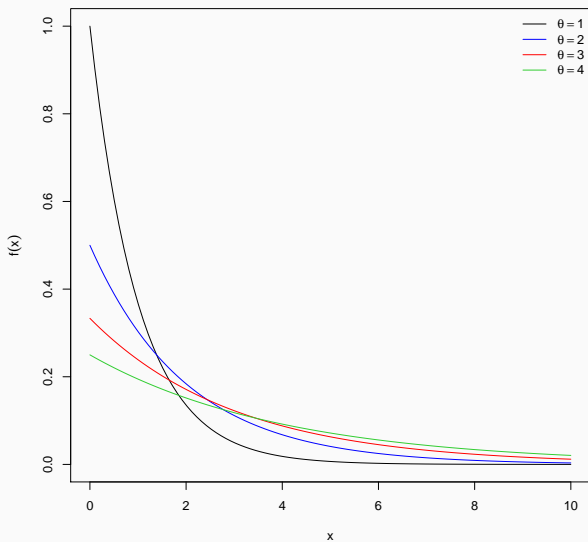
Sea  $X$  una v.a. continua. Decimos que  $X$  se distribuye de acuerdo a una ley exponencial con parámetro  $\theta$  si su función de densidad es

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

Notación:  $X \sim \exp(\theta)$ .

Observa que antes habíamos usado  $\beta$  en lugar de  $\theta$ . ¡En gustos se rompen géneros!

# El modelo exponencial



## El modelo exponencial

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

$$E(X) = \int_0^{\infty} x \frac{1}{\theta} e^{-x/\theta} dx;$$

$$u = x \quad dv = \frac{1}{\theta} e^{-x/\theta} dx$$

$$du = dx \quad v = -e^{-x/\theta}$$

$$= -xe^{-x/\theta} \Big|_0^{\infty} - \int_0^{\infty} (-e^{-x/\theta}) dx$$

$$= \int_0^{\infty} e^{-x/\theta} dx$$

$$= -\theta e^{-x/\theta} \Big|_0^{\infty}$$

$$= \theta$$

$$\therefore E(X) = \theta$$



## El modelo exponencial

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

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

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

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

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

## Notas:

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

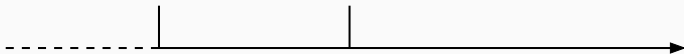
# El modelo exponencial

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



Se pueden presentar exitos previos pero al momento de ocurrir uno pone el reloj en ceros y comienza el conteo de nuevo

exito



# El modelo exponencial

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

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

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

De aquí concluimos que

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

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

## El modelo exponencial

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

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

## El modelo exponencial

En efecto, denotemos con

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

Entonces,

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



- Bajo este esquema, a  $1 - F_X(t_2) = P(X > t_2) = R(t)$  se le conoce como la **confiabilidad** de  $X$ .

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

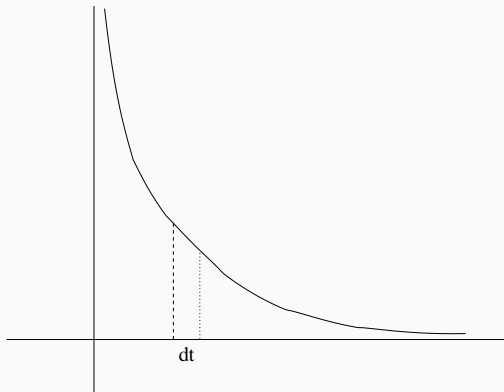
La **función de riesgo** (hazard function) o **tasa de fallo** se define como

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

y cuando se considera sobre un intervalo de tiempo muy pequeño, se le puede interpretar como la **probabilidad instantánea de fallo**.

## El modelo exponencial

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



## El modelo exponencial

Sea

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

Entonces,

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

Ahora, como

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

se sigue que

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

que es conocida como la tasa instantánea de falla.

## El modelo exponencial

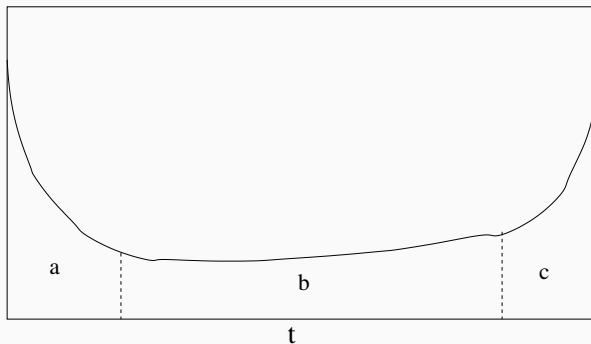
$h(t)\Delta(t) \approx$  la probabilidad de que se presente una “falla” al tiempo  $t + dt$  dado que al tiempo  $t$  no ha habido fallas.

En el caso particular del modelo exponencial,

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

que es una constante.

Ver gráficas sobre el comportamiento esperado de funciones de riesgo.



**Figure 13:** Gráfica de “tina de baño”. (a) etapa de ajuste inicial o “mortalidad infantil”. (b) etapa constante. (c) etapa de desgaste o envejecimiento.

## El modelo exponencial

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

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

# El modelo exponencial

## Ejemplo

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

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

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

a)

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

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



## El modelo exponencial

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

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

Entonces

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

y además, para el evento

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

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

# El modelo exponencial

## Ejercicio

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



Tan pronto como un componente falla, el sistema completo falla.

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

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

### Ejercicio (continua...)

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

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

# El modelo exponencial

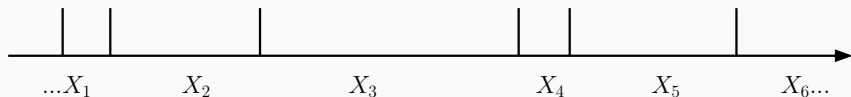
## Ejemplo

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

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

## El modelo exponencial

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



**Figure 14:** Proceso Poisson. Los éxitos ocurren en forma aleatoria sobre el intervalo de definición.

## El modelo exponencial

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

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

Sea  $X$  el tiempo hasta observar  $r$  éxitos. Entonces

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

## El modelo exponencial

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

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

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

# Función de distribución

## Ejemplo

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

1. Determinar la función de distribución inherente al problema. En este caso

$$F(x) = 1 - e^{-x/2}.$$

2. Calcular la función inversa, es decir, resolver  $F^{-1}(x) = y$

$$x = 1 - e^{-y/2} \Leftrightarrow 1 - x = e^{-y/2} \Leftrightarrow \log(1 - x) = -y/2$$

3. Finalmente, graficar

$$X_{(i)} \text{ vs } -2 \log \left( 1 - \frac{i}{n+1} \right)$$



## El modelo Gamma

Sea  $X$  una v.a. continua. Diremos que  $X$  se distribuye de acuerdo a una función de probabilidad Gamma con parámetros  $\alpha$  y  $\beta$  si su función de densidad es

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

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

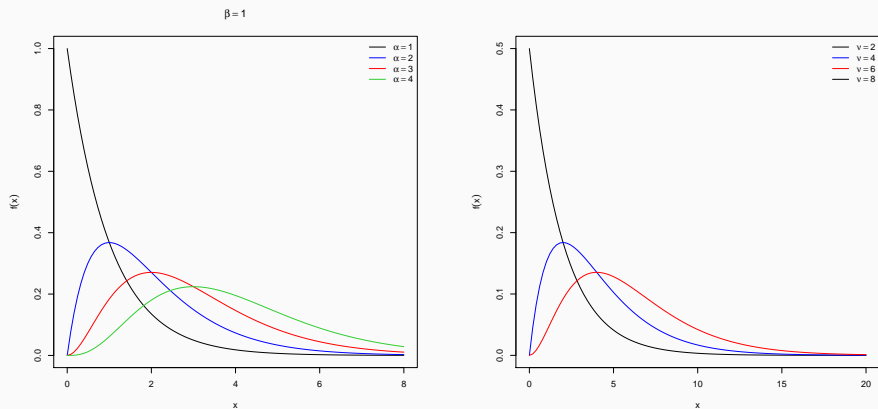
## El modelo Gamma

**Notación:**  $X \sim \text{Gamma}(\alpha, \beta)$ .

Algunas propiedades de la función Gamma:

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

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



**Figure 15:** Distribución Gamma

---

## El modelo Beta

## El modelo Beta

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

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

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

**Notación:**  $X \sim \text{Beta}(\alpha, \beta)$ .

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

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

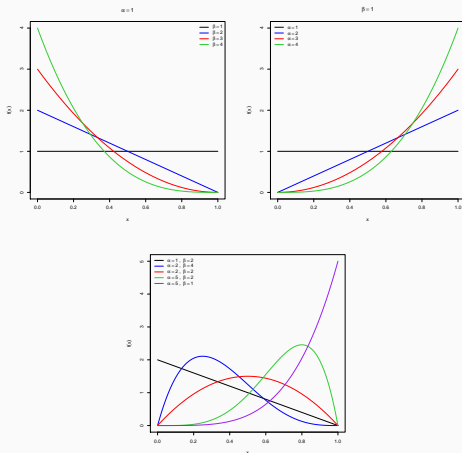
esto es,  $X \sim U(0, 1)$ .

Para demostrar que la función  $f$  integra 1, se sugiere revisar la página de wikipedia sobre la función beta

[https://en.wikipedia.org/wiki/Beta\\_function](https://en.wikipedia.org/wiki/Beta_function).

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

# El modelo Beta



**Figure 16:** Distribución Beta

## El modelo Beta

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

Sea  $r$  un número real positivo,

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



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

Con  $r = 1$ :

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

Con  $r = 2$ :

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

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

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

La función generatriz de momentos expandida en series de potencias es

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

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

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

### Resultado

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

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

esto es,

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

### Demostración

Se demostrará que

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

para la variable aleatoria  $X \sim \text{Binomial}$ .

## Demostración (continúa...)

$$\begin{aligned}P(X \geq k) &= \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} \\ \frac{d}{dp} P(X \geq k) &= \sum_{j=k}^n \binom{n}{j} \left\{ p^j (n-j) (1-p)^{n-j-1} (-1) + j p^{j-1} (1-p)^{n-j} \right\} \\ &= \sum_{j=k}^n \left\{ \frac{n!}{j!(n-j)!} j p^{j-1} (1-p)^{n-j} - \frac{n!}{j!(n-j)!} (n-j) p^j (1-p)^{n-j-1} \right\} \\ &= \sum_{j=k}^n \left\{ \frac{n(n-1)!}{(j-1)!(n-j)!} p^{j-1} (1-p)^{n-j} - \frac{n(n-1)!}{j!(n-j-1)!} p^j (1-p)^{n-j-1} \right\} \\ &= \sum_{j=k}^n \left\{ n \binom{n-1}{j-1} p^{j-1} (1-p)^{n-j} - n \binom{n-1}{j} p^j (1-p)^{n-j-1} \right\} \\ &\quad \underbrace{\hspace{15em}}_{\text{es una suma telescópica}}\end{aligned}$$

## Demostración (continúa...)

entonces,

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

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

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

### Ejemplo

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

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



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

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

a) El valor medio está dado por

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

es decir, el 20%.

b) Además,

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

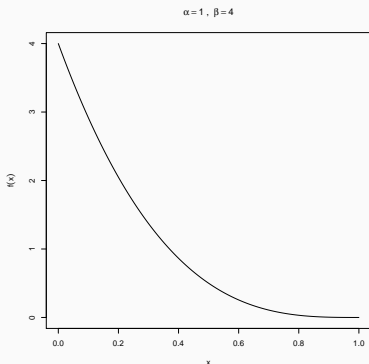
Sea  $u = 1 - x$ ,  $du = -dx$ , así  $u = .75$  y  $u = 0$ .

# El modelo Beta

Entonces

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

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



## Ejercicio

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

### Ejercicio (Continúa)

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

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

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

# Transformación de variables

---

---

**Suma**

---

## Censura y Truncamiento

## Censura y Truncamiento

La censura es una modificación a una distribución. Existen varios tipos de censura. Por ahora sólo nos referimos a la censura por la derecha tipo I. En la censura por la derecha, a las observaciones que rebasan un valor  $a$  sólo se sabe que rebasan este valor. Si la variable sin censura  $X$  tiene fd  $f(x)$  entonces, podemos definir a  $Y$  como  $X$ , si el  $X < a$  y  $a$  si  $X \geq a$ . En este caso la fd de  $Y$  es

$$f_Y(y) = f_X(y)I_{(-\infty, a)}(y) + P(X \geq a)I_a(y)$$

El truncamiento más común es por la izquierda. Al igual que la censura, las observaciones son imperfectas, pero el truncamiento por la izquierda ignora totalmente el número de eventos con un valor menor a un valor  $a$ . Si  $Z$  es el truncamiento de  $X$  en  $a$ , entonces su fd es

$$f_Z(z) = \frac{f_X(z)}{P(X > a)}I_{(a, \infty)}(z)$$



---

## Funciones de variables aleatorias

## Funciones de variables aleatorias

Anteriormente hemos trabajado con funciones de variables aleatorias, que de una u otra forma nos han dado información valiosa en la modelación de problemas del tipo aleatorio. Por ejemplo, dada una v. a.  $X$ , definimos:

- $g(X) = (X - \mu)^2$ , y obtuvimos la varianza de  $X$ :  $V(X) = E[g(X)]$ .
- $g(X) = \frac{X-\mu}{\sqrt{V(X)}}$ ,  $g(X) = (\frac{X-\mu}{\sigma})^3$ ,  $g(X) = (\frac{X-\mu}{\sigma})^4$ .
- $g(X) = e^{tX}$ ,  $g(X) = 1 - e^{-\beta X}$ .
- $g(X) = \sum_{i=1}^n X_i$ , etc.

Incluso definimos,  $g(Z) = Z\sigma + \mu$ .

## Funciones de variables aleatorias

Cada una de estas cantidades han servido para aumentar nuestro conocimiento de la v. a. de interés. Parece pues natural indagar el comportamiento típico de esas cantidades, esto es, encontrar y explotar sus **funciones de densidad de probabilidad respectivas** (de las más útiles).

En principio, las variables aleatorias que surjan serán susceptibles de ser analizadas para encontrar sus **funciones de distribución, valores esperados, varianzas, etc.** Sin embargo nos concentraremos en solo algunas transformaciones particulares que generan resultados de gran interés para la puesta en marcha de procedimientos estadísticos prácticos.

Daremos a continuación una lista de resultados de transformaciones que nos llevan a las distribuciones más empleadas en la construcción de inferencias estadísticas.

Lo que debe quedar en mente son **las condiciones que permiten la derivación de dichas distribuciones**.

Los métodos de transformaciones y otros resultados de importancia, podrás consultarlos en el apéndice del capítulo.

# Funciones de variables aleatorias

## Resultado 1: Distribución Ji-Cuadrada ( $\chi^2$ ).

Sea  $X$  una variable aleatoria **normal estándar**, entonces la densidad de  $Y = X^2$  es la función de densidad de una variable aleatoria con distribución Gamma con  $\beta = 2$ ,  $\alpha = \frac{1}{2}$ ; esto es, la de una  $\chi_1^2$ , **Ji-cuadrada con 1 grado de libertad**. En símbolos,

$$X \sim N(0,1) \quad \Rightarrow \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < +\infty.$$

y para  $Y = X^2$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}, \quad y > 0.$$

esto es,

$$Y \sim \text{Gamma}(\alpha = \frac{1}{2}, \beta = 2), \quad \text{i.e. } Y \sim \chi_1^2.$$

Es decir,

$$X \sim N(0, 1) \Rightarrow X^2 \sim \chi_1^2 \equiv \text{Gamma}(\alpha = \frac{1}{2}, \beta = 2).$$

Esta distribución surge en el trabajo inferencial cuando se trata de establecer la distribución de **varianzas muestrales**, como veremos más adelante.

## Resultado 2: Distribución t-Student.

Sean  $X$  y  $Y$  variables aleatorias independientes tales que:

$$X \sim N(0, 1), \quad Y \sim \chi_n^2.$$

Definamos  $T = \frac{X}{\sqrt{Y/n}}$  entonces

$$f_T(t) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})(1+\frac{t^2}{n})^{(n+1)/2}} & -\infty < t < \infty, n > 0, (n \in \mathbb{N}) \\ 0 & \text{en otro caso.} \end{cases}$$

Esta distribución juega un papel predominante en estimación por intervalo y pruebas de hipótesis sobre la media  $\mu$ , particularmente cuando la muestra proviene de una **población normal** y su varianza no se conoce, sin importar el tamaño de la muestra considerada.

Al igual que la distribución normal, la función de distribución de la t-Student no tiene forma cerrada, por lo que se será necesario usar software para calcular probabilidades.



# Funciones de variables aleatorias

Observaciones:

- La distribución es simétrica con respecto a cero.
- $E(T) = 0$ ,  $V(T) = \frac{n}{n-2}$ ,  $n > 2$ .
- Si  $n = 1$ , entonces

$$f_T(t) = \frac{1}{\pi(1+t^2)} \quad -\infty < t < \infty,$$

lo cual corresponde a la **distribución de *Cauchy*(0, 1)**.

- Notemos que

$$f_T(t) = \underbrace{\frac{1}{\sqrt{2\pi}} \frac{(\frac{n-1}{2})\Gamma(\frac{n-1}{2})}{\sqrt{\frac{n}{2}}\Gamma(\frac{n}{2})}}_{\star} \underbrace{\left(1 + \frac{t^2/2}{n/2}\right)^{-n/2}}_{\star\star} \underbrace{\left(1 + \frac{t^2}{n}\right)^{-1/2}}_{\star\star\star}$$

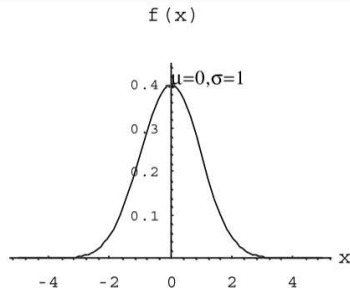
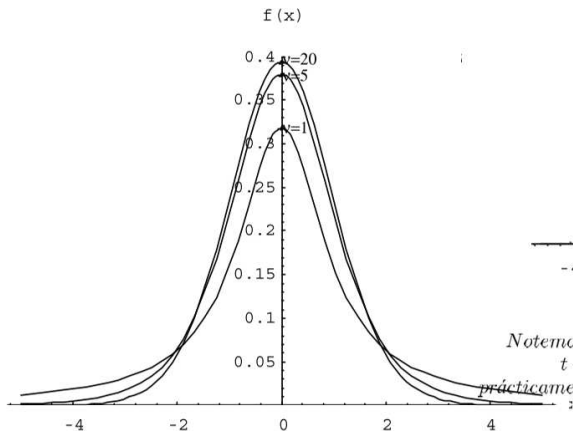
Obsérvese que, cuando  $n \rightarrow \infty$ ,

$$\star \rightarrow \frac{1}{\sqrt{2\pi}}, \quad \star\star \rightarrow e^{-t^2/2}, \quad \star\star\star \rightarrow 1.$$

Luego  $f_T(t) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ , es decir que cuando  $n \rightarrow \infty$   $f_T(t)$  converge a una normal estándar.

Notación: Escribimos  $t_n$  para referirnos a la distribución  $t$  con  $n$  grados de libertad.

# Funciones de variables aleatorias



Notemos cómo la última de las curvas  $t$  - Student, con  $n = 20$ , prácticamente se ve igual a la gráfica de una  $N(0,1)$

**Resultado 3:** Distribución  $F$  de Fisher o de Snedecor.

Sean  $X$  y  $Y$  dos variables aleatorias **independientes** tales que

$$X \sim \chi_n^2, \quad Y \sim \chi_m^2.$$

Definamos  $F = \frac{\frac{X}{n}}{\frac{Y}{m}} = \frac{mX}{nY}$ , entonces

$$f_F(v) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} v^{\frac{n}{2}-1} \left(v \frac{n}{m} + 1\right)^{-\frac{m+n}{2}} \quad 0 \leq v < \infty.$$

Esta es llamada **distribución  $F$**  con  $n$  grados de libertad del numerador y  $m$  grados de libertad del denominador.

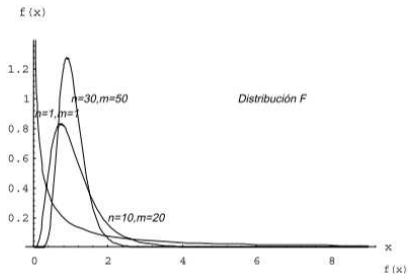
Notación: Escribimos  $F \sim F_{n,m}$  para denotar a la  $F$  con  $n$  grados de libertad del numerador y  $m$  grados de libertad del denominador.

Será necesario utilizar software para calcular probabilidades asociadas a esta distribución.

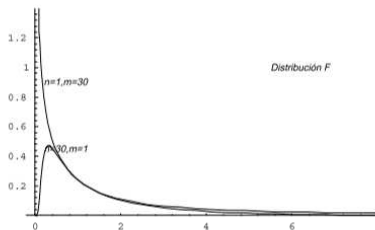
Como veremos en el capítulo de inferencia, este resultado tendrá su impacto en **la descripción de los valores de varianzas muestrales**,

- Comparaciones entre 2 o más poblaciones;
- Cuando se quiere ver si hay diferencias en variabilidad al aplicar dos métodos de medición;
- En regresión para comparar dos modelos;
- En diseños, para comparar efectos de “tratamientos”, (diferentes niveles de temperatura, insecticidas, concentraciones etc.);
- Entre muchos otros usos...

# Funciones de variables aleatorias



Notemos que la distribución cambia si se intercambian los grados de libertad.



## Ejercicios:

1. Demostrar que si  $X \sim t_v$ , entonces la distribución de la v.a.  $Y = X^2$  es una  $F_{1,v}$
2. En el último ejemplo, considera la división  $\frac{1}{F} = R = \frac{\frac{Y}{m}}{\frac{X}{n}}$  y demuestra que  $R \sim F_{m,n}$ , una  $F$  con los grados de libertad intercambiados.



## Resultado 4: Muestra Aleatoria y Estadísticos de Orden.

Un concepto fundamental es el siguiente: al tomar una muestra aleatoria

$$X_1, X_2, \dots, X_n;$$

de alguna población, lo que se está considerando es **un conjunto de variables aleatorias i.i.d.**

Las características principales de una m.a. son:

- Cada v. a. está definida sobre la misma población, por lo tanto la distribución de cada una será exactamente la misma.
- Los posibles resultados de cada una no afectan los posibles resultados de las otras.

# Funciones de variables aleatorias

## Ejemplo (Estadísticos de Orden)

Existen muchos casos en los que la cantidad importante es el valor menor de una m.a., que por el mismo carácter aleatorio de la m.a. también será una v. a. Equivalentemente, la cantidad importante puede ser el valor mayor de la m. a., y también es una v.a. Las dos variables aleatorias así definidas se llaman respectivamente

- primer estadístico de orden  $X_{(1)}$ ,
- n-ésimo estadístico de orden  $X_{(n)}$ .

Estos estadísticos de orden resultan ser una transformación de variables aleatorias:

- $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$ ,
- $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ .

## Aplicaciones:

- Cuando un sistema deja de funcionar al fallar alguna de sus  $n$  componentes, la duración total puede ser vista como la duración más corta de cualquiera de sus componentes. Cada componente puede tener asignada una v.a. de duración  $X_i$ , entonces, podríamos definir como la v.a. de interés a  $X_{(1)}$  = *el menor valor de una m.a. de tamaño  $n$* ; en símbolos,  $X_{(1)} = \min\{X_1, X_2, \dots, X_n\}$ . (**SISTEMA EN SERIE**)
- Un sistema que deja de funcionar hasta que el último componente falla; aquí interesaría  $X_{(n)}$  el valor mayor de una m.a. de tamaño  $n$ . En símbolos,  $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ . (**SISTEMA EN PARALELO**)

## Aplicaciones (continúa):

- En una prueba de resistencia de material de construcción, la variable crítica puede ser la máxima resistencia posible si se toma una m.a. de tamaño  $n$  ( $n$  mediciones). Nos interesaría,  $X_{(n)}$  = *el mayor valor de la m. a. de tamaño  $n$* ; en símbolos,  $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ .
- También puede interesar el rango de la m.a., esto es  $R = X_{(n)} - X_{(1)}$ . Esta es una medición alternativa de la dispersión muy usada en la industria de la manufactura, la cual representa otra transformación de variables aleatorias.

## Funciones de variables aleatorias

Las densidades de  $X_{(1)}$  y  $X_{(n)}$  se deducen a continuación usando el método de la acumulada.

**Densidad de  $X_{(n)}$ :** Llamaremos  $Y_n = X_{(n)}$ ,  $y$  a sus valores y  $F_{Y_n}(y)$  a su función de distribución acumulada. Entonces,

$$F_{Y_n}(y) = P\{Y_n \leq y\} = P\{X_1 \leq y, X_2 \leq y, \dots, X_n \leq y\};$$

si  $Y_n$  es menor que  $y$  entonces cada uno de los valores de las  $X_i$  tienen que ser menor que  $y$ . Puesto que las  $X_i$  son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_n}(y) &= \prod_{i=1}^n P(X_i \leq y) \quad (\text{independencia}) \\ &= [P\{X \leq y\}]^n \quad (\text{distribuciones idénticas}) \\ &= [F_X(y)]^n. \end{aligned}$$

Entonces,

$$\begin{aligned}f_{Y_n}(y) &= \frac{d}{dy} [F_X(y)]^n \\&= n[F_X(y)]^{n-1} \frac{d}{dy} [F_X(y)] \quad (\text{Regla de la Cadena}) \\&= n[F_X(y)]^{n-1} f_X(y). \quad (\text{Teorema Fundamental del Cálculo})\end{aligned}$$

## Funciones de variables aleatorias

**Densidad de  $X_{(1)}$ :** Llamaremos  $Y_1 = X_{(1)}$ , y a sus valores y  $F_{Y_1}(y)$  a su función de distribución acumulada. Entonces,

$$F_{Y_1}(y) = P\{Y_1 \leq y\} = 1 - P\{Y_1 > y\} = 1 - P\{X_1 > y, X_2 > y, \dots, X_n > y\};$$

el valor más pequeño  $Y_1$  es mayor que  $y$  cuando cada uno de los valores de las  $X_i$  lo sean. Puesto que las variables aleatorias son independientes, la anterior probabilidad se calcula como el producto de las probabilidades de cada evento:

$$\begin{aligned} F_{Y_1}(y) &= 1 - \prod_{i=1}^n P(X_i > y) && \text{(independencia)} \\ &= 1 - [P\{X > y\}]^n && \text{(distribuciones idénticas)} \\ &= 1 - [1 - F_X(y)]^n. && \text{(complemento)} \end{aligned}$$

Después,

$$\begin{aligned}f_{Y_1}(y) &= \frac{d}{dy}[1 - [1 - F_X(y)]^n] \\&= -n[1 - F_X(y)]^{n-1} \frac{d}{dy}[1 - F_X(y)] \quad (\text{Regla de la Cadena}) \\&= -n[1 - F_X(y)]^{n-1}[-f_X(y)] \quad (\text{Teorema Fundamental del Cálculo}) \\&= n[1 - F_X(y)]^{n-1}f_X(y).\end{aligned}$$



## Funciones de variables aleatorias

### Ejemplo

Considera una m.a.  $X_1, X_2, \dots, X_n$  de tamaño  $n$ , tomada de una población  $Unif(0, 1)$ . Calcula las densidades de los estadísticos de orden primero y  $n$ -ésimo. Recordando que

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{c.o.p.} \end{cases} \quad F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & x > 1 \end{cases}$$

Lo que implica,

$$f_{Y_n}(y) = n[F_X(y)]^{n-1}f_X(y) = ny^{n-1} \cdot 1 \quad \text{para } 0 < y < 1,$$

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1}f_X(y) = n(1 - y)^{n-1} \cdot 1 \quad \text{para } 0 < y < 1.$$

# Funciones de variables aleatorias

## Ejemplo (Sistema en Serie)

El siguiente sistema en serie funcionará siempre que cada una de las componentes (idénticas) mostradas fallen, cada componente funciona independientemente de las otras. Supóngase que  $X$  = tiempo de vida de cada componente del sistema y que  $X \sim \exp(\beta)$ . Determinar el comportamiento del tiempo de vida del sistema, es decir, de  $Y_1 = \min(X_1, X_2, X_3)$ .



Recordemos que

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\beta} e^{-\frac{x}{\beta}} & x > 0 \end{cases} \quad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\frac{x}{\beta}} & x > 0 \end{cases}$$

Entonces,

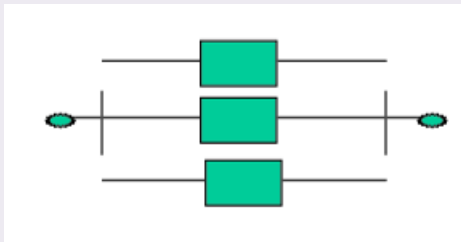
$$\begin{aligned} f_{Y_1}(y) &= n[1 - F_X(y)]^{n-1} f_X(y) = 3[1 - (1 - e^{-\frac{y}{\beta}})]^{3-1} \cdot \frac{1}{\beta} e^{-\frac{y}{\beta}} \\ &= \frac{3}{\beta} (e^{-\frac{y}{\beta}})^2 e^{-\frac{y}{\beta}} = \frac{3}{\beta} e^{-\frac{3y}{\beta}} \quad \text{para } y > 0. \end{aligned}$$

Es decir que  $Y_1 \sim \exp(\frac{\beta}{3})$ .

## Funciones de variables aleatorias

### Ejemplo (Sistema en Paralelo)

El siguiente sistema en paralelo funcionará hasta que fallen todas las componentes (idénticas) mostradas, cada componente funciona independientemente de las otras. Supóngase que  $X$  = tiempo de vida de cada componente del sistema y que  $X \sim \exp(\beta)$ . Determinar el comportamiento del tiempo de vida del sistema, es decir, de  $Y_3 = \max\{X_1, X_2, X_3\}$ .



Luego,

$$\begin{aligned}f_{Y_3}(y) &= n[F_X(y)]^{n-1}f_X(y) = 3[1 - e^{-\frac{y}{\beta}}]^{3-1} \cdot \frac{1}{\beta}e^{-\frac{y}{\beta}} \\&= \frac{3}{\beta}[1 - e^{-\frac{y}{\beta}}]^2e^{-\frac{y}{\beta}} \quad \text{para } y > 0.\end{aligned}$$

Para un tratamiento más amplio sobre estadísticos de orden puedes revisar: *David, H.A.; Nagaraja, H.N. (2003). Order Statistics. Wiley Series in Probability and Statistics. doi:10.1002/0471722162.*

## Resultado 5: Distribuciones de Sumas de Variables Aleatorias Independientes.

**El Método de la Función Generatriz de Momentos** se aplica tanto a variables continuas como discretas. El método consiste en

- **Encontrar** la función generatriz de momentos de la variable transformada;
- **Identificarla** como la generatriz de alguna variable conocida;

dándonos entonces la distribución de la transformación.

Está basado en la unicidad de la generatriz de momentos (si existe), y en el teorema que dice: “...igualdad de generatrices para dos variables aleatorias implica igualdad de distribuciones”.

- Si **no** podemos identificar la generatriz de momentos de nuestra variable transformada con alguna de un catálogo, el método no nos proporciona ninguna información y será necesario utilizar el método de transformaciones de variables para poder determinar la distribución de la nueva variable.
- Ya hemos usado este procedimiento anteriormente en la derivación de la distribución binomial negativa vista como una suma de variables aleatorias geométricas con probabilidad de éxito constante  $p$ , y en la construcción de una Gamma a partir de sumas de Exponenciales.

### Resultado 6.

Sean  $X_1, X_2, \dots, X_n$ ,  $n$  variables aleatorias independientes y  $a_1, a_2, \dots, a_n$  constantes reales. Se define  $Y = a_1X_1 + a_2X_2 + \dots + a_nX_n$  (combinación lineal) entonces la generatriz de momentos de  $Y$  es

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t),$$

donde  $M_{X_i}(t)$  es la generatriz de momentos de  $X_i$ .

**Demostración:** Se hace para el caso continuo, el caso discreto es análogo (sólo es necesario cambiar la integral por suma).



# Funciones de variables aleatorias

$$\begin{aligned}M_Y(t) &= E(e^{Yt}) = E(e^{(a_1 X_1 + a_2 X_2 + \dots + a_n X_n)t}) = E(e^{a_1 X_1 t} \cdot e^{a_2 X_2 t} \cdot e^{a_3 X_3 t} \dots e^{a_n X_n t}) \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (e^{a_1 X_1 t} \cdot e^{a_2 X_2 t} \cdot e^{a_3 X_3 t} \dots e^{a_n X_n t}) \underbrace{(f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n))}_{\text{Densidad Conjunta}} dx_1 dx_2 dx_3 \dots dx_n\end{aligned}$$

y dado que  $X_1, X_2, \dots, X_n$  son independientes,

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (e^{a_1 x_1 t} \cdot e^{a_2 x_2 t} \dots e^{a_n x_n t}) [f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n)] dx_1 dx_2 \dots dx_n$$

como las variables solo intervienen en una funcion, se puede cambiar la integral multiple por un producto de integrales simples.

$$\begin{aligned} &= \underbrace{\int_{-\infty}^{\infty} e^{a_1 x_1 t} f_{X_1}(x_1) dx_1}_{M_{X_1}(a_1 t)} \int_{-\infty}^{\infty} e^{a_2 x_2 t} f_{X_2}(x_2) dx_2 \cdots \int_{-\infty}^{\infty} e^{a_n x_n t} f_{X_n}(x_n) dx_n \\ &= M_{X_1}(a_1 t) \cdot M_{X_2}(a_2 t) \cdots M_{X_n}(a_n t) \quad \therefore \quad M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t) \end{aligned}$$

El método es ampliamente utilizado en transformaciones definidas como combinaciones lineales de muestras aleatorias, recuerda que una m. a. está compuesta de un conjunto de variables aleatorias i.i.d.

## Funciones de variables aleatorias

**Ejemplos:** 1) Sean  $X_1, X_2, \dots, X_n$  variables Bernoulli con parámetro  $p$  e independientes. Considerando  $Y = X_1 + X_2 + \dots + X_n$ , encuentra la distribución de  $Y$ .

La generatriz de momentos de  $X_i$  es

$$M_{X_i}(t) = (1 - p) + pe^t,$$

con  $a_i = 1$ . Así que,

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t) = \prod_{i=1}^n [(1 - p) + pe^t] = ((1 - p) + pe^t)^n,$$

la cual representa la generatriz de momentos de una variable aleatoria binomial con parametros  $n$  y  $p$

$$Y \sim \text{Bin}(n, p).$$

La suma de v.a. Bernoulli con parámetro  $p$  = una v.a. *Binomial*( $n, p$ ). 280

2) Si  $X \sim \text{Gamma}(\alpha, \beta)$ , encontrar la distribución de  $Y = aX$ , con  $a > 0$  y distinta de 1.

La generatriz de una  $X \sim \text{Gamma}(\alpha, \beta)$  es  $M_X(t) = (\frac{1}{1-\beta t})^\alpha$ , entonces

$$M_Y(t) = M_{aX}(t) = M_X(at) = \left(\frac{1}{1-\beta(at)}\right)^\alpha = \left(\frac{1}{1-(\beta a)t}\right)^\alpha,$$

que podemos identificar como la generatriz de una v. a.  $\text{Gamma}(\alpha, \beta^*)$ , si  $a > 0$  con  $\beta^* = a\beta$

$$Y \sim \text{Gamma}(\alpha, a\beta).$$

## Funciones de variables aleatorias

3) Considera una v.a.  $\chi_n^2$  Ji-cuadrada con  $n$  grados de libertad, es decir,

$$X \sim \chi_n^2 \equiv \text{Gamma}(\alpha = \frac{n}{2}, \beta = 2).$$

Encontrar la distribución de  $Y = aX$ , con  $a > 0$  y diferente de 1.

Del ejemplo anterior,

$$Y \sim \text{Gamma}(\alpha = \frac{n}{2}, \beta^* = 2a) = \text{Gamma}(\frac{n}{2}, 2a).$$

Esta no corresponde a la generatriz de una  $\chi^2$ . Una v.a.  $\chi_n^2$  multiplicada por una constante (distinta de la unidad), no resulta en otra v. a.  $\chi^2$ . Este resultado es útil cuando se estudia el comportamiento de la varianza muestral  $S^2$

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2, \quad \text{pero } S^2 = \frac{\sigma^2 Y}{n-1} \sim \text{Gamma}(\frac{n-1}{2}, 2\frac{\sigma^2}{n-1}).$$

## Funciones de variables aleatorias

4) Sean  $X_1, X_2, \dots, X_n$  variables aleatorias **independientes normales**,  $X_i \sim N(\mu_i, \sigma_i^2)$ . Encontrar la distribución de  $Y = a_1X_1 + \dots + a_nX_n$ , donde  $a_i$  es constante.

La generatriz de momentos de  $X_i$

$$M_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}.$$

De acuerdo al teorema, la generatriz de momentos de  $Y = a_1X_1 + a_2X_2 + \dots + a_nX_n$  es

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^n M_{X_i}(a_i t) \\ \Rightarrow M_{X_i}(a_i t) &= e^{\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}}. \end{aligned}$$

## Funciones de variables aleatorias

Por lo tanto,

$$M_Y(t) = \prod_{i=1}^n [e^{\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}}] = e^{\sum_{i=1}^n [\mu_i a_i t + \frac{\sigma_i^2 a_i^2 t^2}{2}]} = e^{t(\sum_{i=1}^n a_i \mu_i) + (\sum_{i=1}^n a_i^2 \sigma_i^2) \frac{t^2}{2}}.$$

Esta es la generatriz de momentos de una variable aleatoria normal con media  $\sum_{i=1}^n a_i \mu_i$  y varianza  $\sum_{i=1}^n a_i^2 \sigma_i^2$ , por lo tanto

$$Y \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right);$$

es decir, cualquier combinacion lineal de normales independientes es a su vez una normal.

En particular, nos gustaría analizar el comportamiento que tendrían los resultados de medias obtenidas de muestras aleatorias seleccionadas de una población normal de tamaño  $n$ ,  $X_1, X_2, \dots, X_n$ , esto es, de la variable aleatoria,  $\bar{X} = \sum_{i=1}^n X_i/n$ , cuando  $X \sim N(\mu, \sigma^2)$ . Es una variable aleatoria porque todavía no sabemos que valores veremos para cada  $X_i$ .



**Resultado 7:** La distribución de  $\bar{X}$  (Población Normal).

Observa que si en el ejemplo anterior la m. a. es extraída de una población normal con media  $\mu$  y varianza  $\sigma^2$  entonces cada  $X_i \sim N(\mu, \sigma^2)$ , y la v.a. definida como

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Para este resultado, no importa el tamaño de la muestra, solo que la población de estudio *de donde se obtiene la muestra*, sea **normal**.

**Ejemplo:** Sea  $X_1, X_2, \dots, X_n$  i.i.d.  $Poisson(\lambda)$ . Calcular  $P(\bar{X} < 2)$ .

Para calcular esta probabilidad buscaremos la f.d.p. de  $\bar{X}$ .

Con el método de la generatriz, si hacemos  $Y = \bar{X}$  y recordamos que la generatriz de una  $Poisson(\lambda)$  es  $M_X(t) = e^{\lambda(e^t - 1)}$ , tenemos

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t) = \prod_{i=1}^n M_{X_i}\left(\frac{t}{n}\right) = \prod_{i=1}^n e^{\lambda(e^{\frac{t}{n}} - 1)} = [e^{\lambda(e^{\frac{t}{n}} - 1)}]^n = e^{\lambda n(e^{\frac{t}{n}} - 1)}.$$

## Funciones de variables aleatorias

Puesto que **no** es una generatriz conocida, para este caso no podemos obtener la función de densidad de  $Y = \bar{X}$ . Esto ejemplifica que **la f.d.p. de  $\bar{X}$  no siempre es una normal, depende de la distribución asociada con la población.** Sin embargo, podríamos intentar encontrar la f.d.p. de la v. a.

$$Y = \sum_{i=1}^n X_i,$$

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(a_i t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n e^{\lambda(e^t - 1)} = [e^{\lambda(e^t - 1)}]^n = e^{\lambda n(e^t - 1)},$$

que es la generatriz de una distribución Poisson con parámetro  $\lambda n$ .

Entonces  $Y \sim \text{Poisson}(\lambda^*)$ , donde  $\lambda^* = n\lambda$ . Podemos calcular la probabilidad pedida como:

$$P(\bar{X} < 2) = P\left(\frac{1}{n} \sum_{i=1}^n X_i < 2\right) = P\left(\sum_{i=1}^n X_i < 2n\right) = P(Y < 2n).$$

## Funciones de variables aleatorias

Con  $\lambda = 1.5$ ,  $n = 4$ , tenemos que  $Y \sim \text{Poisson}(\lambda^* = 6)$ . Entonces:

$$P(\bar{X} < 2) = P(Y < 8) = 0.7439.$$

```
ppois(q=7,lambda=6)
```

### Resultado 8: Distribución de una suma de variables aleatorias Ji-cuadradas independientes.

Sean  $X_1, \dots, X_n$  variables aleatorias independientes cada una de ellas con distribución  $\chi^2$  con  $\nu_i$  grados de libertad. Sea  $Y = X_1 + X_2 + \dots + X_n$ , encontrar su distribución.

La generatriz de momentos para  $X_i$  es:

$$M_{X_i}(t) = (1 - 2t)^{-\frac{\nu_i}{2}}.$$

La generatriz de momentos de  $Y$  es entonces

$$\begin{aligned}M_Y(t) &= \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n (1 - 2t)^{-\frac{\nu_i}{2}} \\&= (1 - 2t)^{\sum_{i=1}^n (-\frac{\nu_i}{2})} \\&= (1 - 2t)^{-\frac{(\sum_{i=1}^n \nu_i)}{2}},\end{aligned}$$

que corresponde a la generatriz de momentos de una variable aleatoria  $\chi^2$  con  $\sum_{i=1}^n \nu_i$  grados de libertad, es decir, **la suma de  $\chi^2$ 's independientes con  $\nu_i$  grados de libertad cada una, es a su vez una  $\chi^2$  con  $\nu = \sum_{i=1}^n \nu_i$  grados de libertad.**

## Ejercicio:

1. Generar 500 muestras aleatorias de tamaños 5, 10, 100, 500, 1000 de una población  $N(0,1)$ .
  - 1.1 En cada caso, calcular  $\bar{X}$  para cada muestra.
  - 1.2 Graficar su histograma, así como su Q-Q plot. En todos los casos deberá verse como una normal pues esa es la distribución exacta de esos valores. Las imperfecciones del ajuste que puedes notar se deben a que estás haciendo una simulación, sólo estás viendo 500 de un  $\#$  infinito de posibles resultados.
2. Repetir el ejercicio anterior para  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ , la desviación estándar, y concluye.

# Métodos de transformación de variables

---

# Métodos de transformación de variables

## Técnicas para encontrar las densidades de probabilidad de funciones de variables aleatorias.

Nos interesara saber como se distribuye una variable aleatoria que ha sido definida como una transformación (función) de otra variable aleatoria.

Para ser más precisos, supongamos que  $X$  es una variable aleatoria cuya distribución ya conocemos, queremos encontrar la distribución de la variable  $Y$  definida como  $Y = g(X)$ , donde  $g$  es una función de  $X$ .

Para ello existen 3 metodos clásicos :

- **Método de la función de distribución acumulada,**
- **Método de la transformación,**
- **Método de la generatriz de momentos.**



---

## Método de la función de distribución acumulada

## Método de la función de distribución acumulada

### Ejemplo (Distribución $\chi_1^2$ con un grado de libertad)

Si suponemos que  $X \sim N(0, 1)$ , hemos visto que  $Y = X^2$  se distribuye como una  $\chi_1^2$ . A saber,

$$\begin{aligned}F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\&= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \quad y \geq 0.\end{aligned}$$

Derivando podemos encontrar la densidad de  $Y$ . Aplicando la regla de la cadena,

$$\begin{aligned}f_Y(y) &= F'_Y(y) = \Phi'(\sqrt{y})\left(\frac{1}{2}y^{-1/2}\right) - \Phi'(-\sqrt{y})\left(-\frac{1}{2}y^{-1/2}\right) \\&= \frac{1}{2\sqrt{y}}[\Phi'(\sqrt{y}) + \Phi'(-\sqrt{y})] \quad y > 0.\end{aligned}$$

## Método de la función de distribución acumulada

Como  $\Phi'(x)$ -la derivada de la acumulada de la normal- es la función de densidad de  $X$ , entonces

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left[ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \right] = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}, \quad y > 0.$$

Como hicimos notar antes, la función de densidad anterior corresponde a la densidad de una v.a. con distribución  $Gamma(\alpha = \frac{1}{2}, \beta = 2)$ .

Es decir, si  $X \sim N(0, 1) \Rightarrow X^2 \sim \chi_1^2 \equiv Gamma(\alpha = \frac{1}{2}, \beta = 2)$ . Como ya hemos mencionado, este resultado es muy utilizado en la práctica.

## Método de la función de distribución acumulada

### Ejemplo

Si  $X$  tiene una densidad uniforme con  $a = -1$  y  $b = 3$ , obtener la densidad de  $Y = |X|$ .

Tenemos que

$$f_X(x) = \begin{cases} \frac{1}{4} & -1 < x < 3 \\ 0 & \text{otra parte} \end{cases}.$$

Así que

$$F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y),$$

para  $y > 0$ . Derivando con respecto a  $y$  obtenemos que

$$f_Y(y) = f_X(y) + f_X(-y) \quad y > 0.$$

En este caso particular tenemos

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y) = \frac{1}{4} + \frac{1}{4} & 0 < y < 1 \\ f_X(y) + f_X(-y) = \frac{1}{4} + 0 & 1 < y < 3 \end{cases}.$$

es decir,

$$f_Y(y) = \begin{cases} \frac{1}{2} & 0 < y < 1 \\ \frac{1}{4} & 1 < y < 3 \\ 0 & \text{otra parte} \end{cases}.$$

---

## Método de la transformación o cambio de variable: Ejemplo

## Método de la transformación o cambio de variable

Este método puede aplicarse a ambos casos (discreto y continuo).

En el caso discreto queda determinado en forma directa.

Ejemplo. (variable aleatoria discreta): Considera a  $X$  la suma del resultado de lanzar dos dados. Encontrar:

a) La función de probabilidad de  $Y = 3X - 2$  dado 1.

dado1 dado 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

## Método de la transformación o cambio de variable

La función de probabilidad para  $X$  es

$x$	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Con  $Y = 3X - 2$ , tenemos

$y$	4	7	10	13	16	19	22	25	28	31	34
$f(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



b) Así, la función de probabilidad de la v.a.

$$Y = \begin{cases} 0 & \text{si } X \text{ par} \\ 1 & \text{si } X \text{ non} \end{cases}$$

está dada por

y	1	0
f(y)	$\frac{18}{36}$	$\frac{18}{36}$

# Método de la transformación o cambio de variable

## Teorema del Cambio de Variable.

Aunque recibe un tratamiento especial, esta técnica realmente es una forma resumida de presentar lo que se hace en el método de la acumulada, como veremos a continuación.

### Teorema

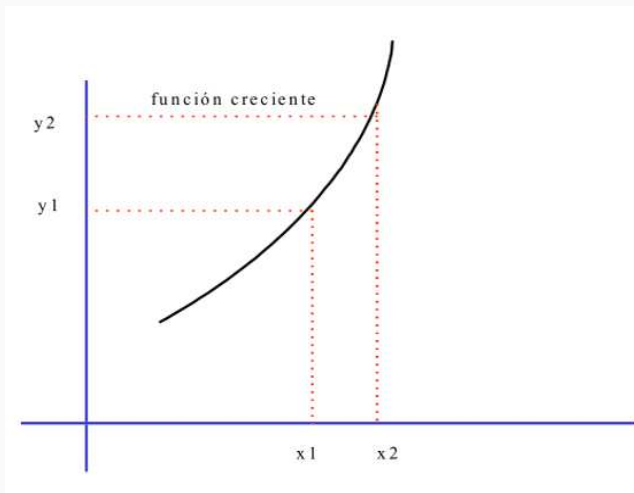
Supongamos que  $X$  es una variable aleatoria continua y que  $Y \equiv u(X)$ . Si  $u$  es una función monótona en todos los valores de la variable  $X$ , de tal forma que se puede despejar  $X$  en la ecuación  $Y = u(X)$ , digamos  $X = w(Y)$ ; entonces la densidad de  $Y$  esta dada por:

$$f_Y(y) = f_X(w(y)) \cdot |w'(y)|,$$

donde  $w'(y) = \frac{d}{dy} w(y)$  y las barras indican que se debe sacar el valor absoluto de  $w'(y)$ .

# Método de la transformación o cambio de variable

Demostración: Monótona (siempre con una misma tendencia) creciente.



## Método de la transformación o cambio de variable

Supongamos que  $Y = u(X)$  es creciente, entonces

$$x_1 \leq x_2 \Rightarrow y_1 \leq y_2,$$

ya que al ser  $u$  creciente se preserva la desigualdad. Así, la acumulada de  $Y$  es

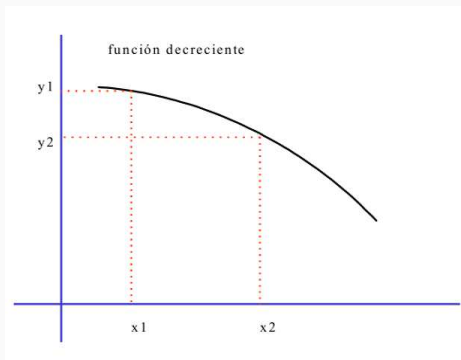
$$F_Y(y) = P(Y \leq y) = P(u(X) \leq y) = P(X \leq w(y)) = F_X(w(y)).$$

Luego la densidad de  $Y$  es

$$\begin{aligned} f_Y(y) &= F'_Y(y) = F'_X(w(y)) \cdot w'(y) = f_X(w(y)) \cdot w'(y) \\ &\quad \text{(Regla de la cadena)} \\ &= f_X(w(y)) \cdot w'(y), \end{aligned}$$

que también es monótona decreciente.

# Método de la transformación o cambio de variable



Supongamos ahora que  $Y = u(X)$  es decreciente. Luego,

$$x_1 \leq x_2 \Rightarrow y_1 \geq y_2,$$

es decir que se invierte la desigualdad.

## Método de la transformación o cambio de variable

La acumulada de  $Y$  es entonces

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

mientras que su densidad es

$$\begin{aligned}f_Y(y) &= F_Y(y) = -F_X(w(y)) \cdot w'(y) \\&= -f_X(w(y)) \cdot w'(y) \\&= f_X(w(y)) \cdot [-w'(y)].\end{aligned}$$

# Método de la transformación o cambio de variable

Combinando  $u$  creciente o decreciente, obtenemos que:

$$f_Y(y) = f_X(w(y)) \cdot |w'(y)|,$$

lo que prueba el Teorema.

Los pasos que seguimos fueron

- Despejar  $X$  en la relación  $Y = u(X)$  obteniendo  $X = w(Y)$  y derivar  $\frac{dx}{dy} = w'(y)$ .
- Evaluar la f.d.p. de  $X$  en  $w(y)$  y multiplicarla por  $|w'(y)|$ .

# Método de la transformación o cambio de variable

## Ejemplo

Dada la variable aleatoria  $X$  con función de densidad

$$f_X(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{en otra parte} \end{cases},$$

encontrar la distribución de probabilidad de  $Y = 8X^3$ .

Tenemos que

$$Y = u(X) = 8X^3; \quad y = 8x^3 \quad \Rightarrow \quad x = \left(\frac{y}{8}\right)^{1/3} = \frac{1}{2}(y)^{1/3} \quad (u(X) \text{ es monótona})$$

Entonces,

$$\therefore \quad w(y) = \frac{1}{2}(y)^{1/3} \quad w'(y) = \frac{1}{6}y^{-2/3} \quad (\text{despeje y derivación})$$



Luego, la función de densidad de  $Y$  está dada por

$$\begin{aligned} f_Y(y) &= f_X(w(y)) \cdot |w'(y)| \\ &= \begin{cases} 2\left(\frac{y^{\frac{1}{3}}}{2}\right)\left(\frac{y^{-\frac{2}{3}}}{6}\right) & 0 < w(y) < 1 \\ 0 & \text{en otra parte} \end{cases} \\ &= \begin{cases} \frac{1}{6}(y^{-\frac{1}{3}}) & 0 < y < 8 \\ 0 & \text{otra parte} \end{cases} . \end{aligned}$$

## Método de la transformación o cambio de variable

### Ejercicio

Cuando  $X \sim N(0, 1)$  y  $Y = X^2$ , sabemos que  $Y$  tiene una distribución  $\chi_1^2$ . Aplica la técnica anterior notando que  $Y$  es decreciente para valores negativos de  $X$  y creciente para los valores positivos de  $X$  -es decir, que será necesario partir la región y sumar las contribuciones en ambas regiones.

Es decir que, si  $x_i = w_i(y)$ ,

$$f_Y(y) = \sum_{i=1}^k f(x_i) \left| \frac{dx_i}{dy} \right|.$$

(la derivada de la  $i$ -ésima transformación  
inversa que tengamos que calcular de las  $k$  regiones)

## Caso bivariado:

Sean  $X_1, X_2$  variables aleatorias continuas con función de densidad conjunta  $f_{X_1, X_2}(x_1, x_2)$  y sea  $\mathcal{X} = \{(x_1, x_2) : f_{X_1, X_2}(x_1, x_2) > 0\}$ . Supongamos que:

$$y_1 = g_1(x_1, x_2)$$

$$y_2 = g_2(x_1, x_2)$$

Asumiremos que  $g$  satisface las siguientes condiciones.

## Método de la transformación o cambio de variable

- $g$  define una **transformación biyectiva** de  $\mathcal{X}$  a  $\mathcal{Y} = \{y_1, y_2\}$ , donde  $y_1$  y  $y_2$  están definidas por la transformación.
- El Jacobiano de  $g^{-1}$ , denotado por  $J$  y dado por el determinante de orden 2

$$J = \left| \frac{\partial [g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)]}{\partial \underset{\sim}{y}} \right| \quad \underset{\sim}{y} = (y_1, y_2)$$
$$= \begin{vmatrix} \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_1^{-1}(y_1, y_2)}{\partial y_2} \\ \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_1} & \frac{\partial g_2^{-1}(y_1, y_2)}{\partial y_2} \end{vmatrix},$$

tiene derivadas parciales de primer orden continuas sobre  $\mathcal{Y}$ .

Si el Jacobiano es diferente de cero para  $(y_1, y_2) \in \mathcal{Y}$ , la función de densidad conjunta de  $(Y_1, Y_2)$  está dada por:

$$f_{Y_1 Y_2}(y_1, y_2) = \begin{cases} f_{X_1, X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) |J| & (y_1, y_2) \in \mathcal{Y} \\ 0 & \text{en otro caso} \end{cases}$$

# Método de la transformación o cambio de variable

## Ejemplo (distribución t-Student)

Usada por primera vez en 1908 por W. S. Gosset quién estableció el comportamiento de algunas medidas muestrales, en forma experimental en principio. Fisher derivó la prueba matemática en 1924 y la generalización a varias variables fue dada por Hotelling y se le conoce en el análisis multivariado por “ $T^2$  de Hotelling”.

Esta distribución se deriva de la siguiente manera. Sean  $X$  y  $Y$  variables aleatorias **independientes** tales que:

$$X \sim N(0, 1), \quad Y \sim \chi_n^2,$$

y definamos

$$T = \frac{X}{\sqrt{Y/n}}.$$

## Método de la transformación o cambio de variable

Observa que se trata del cociente de una normal estándar y la raíz cuadrada de una  $\chi^2_{(n)}$  dividida por sus grados de libertad, siendo estas variables independientes. Por lo tanto, su distribución conjunta es el producto de sus densidades:

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, \quad -\infty < x < \infty, y > 0.$$

Si la transformación es

$$t = \frac{x}{\sqrt{y/n}} \quad \text{y si} \quad u = y,$$

entonces la transformación inversa es

$$x = t\sqrt{u/n}, \quad y = u.$$

## Método de la transformación o cambio de variable

Claramente esto nos da una complicación extra: primero elegir esta segunda coordenada de tal forma que el problema de inversión de funciones no se complique y segundo obtener la función conjunta de donde obtendremos la marginal de la variable principal, en este caso  $t$ .

$$\mathcal{X} = \{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}^+\} \longrightarrow \mathcal{Y} = \{(t, u) | t \in \mathbb{R}, u \in \mathbb{R}^+\}.$$

Entonces,

$$J = \begin{vmatrix} \sqrt{\frac{u}{n}} & \frac{t}{\sqrt{n}} \frac{1}{2} u^{-\frac{1}{2}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}} \implies |J| = \sqrt{\frac{u}{n}}.$$



Así,

$$f_{T,U}(t, u) = \sqrt{\frac{u}{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2(\frac{u}{n})} \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} e^{-\frac{1}{2}u}$$

para

$$0 < u < \infty, \quad -\infty < t < \infty, \quad n \in \mathbb{N}.$$

## Método de la transformación o cambio de variable

Pero queremos  $f_T(t)$ , así que necesitamos la marginal

$$\begin{aligned} f_T(t) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2(\frac{u}{n})} \frac{1}{n^{\frac{1}{2}}(\frac{1}{2})^{-(n/2)}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1+\frac{1}{2}} e^{-\frac{1}{2}u} du \\ &= \frac{(\frac{1}{2})^{n/2}}{\sqrt{2\pi}\Gamma(\frac{n}{2})\sqrt{n}} \int_0^\infty u^{\frac{n+1}{2}-1} e^{-\frac{1}{2}(1+\frac{t^2}{n})u} du. \end{aligned}$$

Es decir que

$$f_T(t) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})(1+\frac{t^2}{n})^{(n+1)/2}} & -\infty < t < \infty, n > 0, (n \in \mathbb{N}) \\ 0 & \text{en otro caso} \end{cases}.$$

## Ejemplo: Distribución F de Fisher o de Snedecor.

Sean dos variables aleatorias independientes  $X \sim \chi_n^2$ ,  $Y \sim \chi_m^2$ . Obtener la f.d.p. de la v.a.

$$F = \frac{\frac{X}{n}}{\frac{Y}{m}} = \frac{mX}{nY}.$$

Como  $X$  y  $Y$  son independientes su distribución de probabilidad es el producto de sus densidades, es decir,

$$f_{X,Y}(x,y) = \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \cdot \frac{1}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}} y^{\frac{m}{2}-1} e^{-\frac{y}{2}} \quad x, y > 0.$$

## Método de la transformación o cambio de variable

Si hacemos

$$V = \frac{mX}{nY}, \quad U = Y,$$

la transformación inversa es

$$x = \frac{nvu}{m}, \quad y = u, \quad u, v > 0.$$

Esto último define una transformación uno a uno

$$\mathcal{X} = \{(x, y) | x \in \mathbb{R}^+, y \in \mathbb{R}^+\} \longrightarrow \mathcal{Y} = \{(v, u) | v \in \mathbb{R}^+, u \in \mathbb{R}^+\}$$

con Jacobiano

$$J = \begin{vmatrix} \frac{n}{m}u & \frac{n}{m}v \\ 0 & 1 \end{vmatrix} = \frac{n}{m}u \neq 0$$

Entonces

$$f_{U,V}(u, v) = f_{X,Y}\left(\frac{nuv}{m}, u\right)|J| = f_{X,Y}\left(\frac{nuv}{m}, u\right)\left|\frac{n}{m}u\right| = \frac{n}{m}uf_{X,Y}\left(\frac{nuv}{m}, u\right)$$

así

$$\begin{aligned} f_{U,V}(v, u) &= \frac{n}{m}u \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} \left(\frac{nuv}{m}\right)^{\frac{n}{2}-1} e^{-\frac{nuv}{2m}} \cdot \frac{1}{\Gamma(\frac{m}{2})2^{\frac{m}{2}}} u^{\frac{m}{2}-1} e^{-\frac{u}{2}} \\ &= \frac{n}{m}u \frac{1}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \left(\frac{nuv}{m}\right)^{\frac{n}{2}-1} e^{-\frac{nuv}{2m}} \cdot u^{\frac{m}{2}-1} e^{-\frac{u}{2}} \quad u, v > 0. \end{aligned}$$

## Método de la transformación o cambio de variable

Nos interesa la marginal de  $V$ , ya que  $V = F = \frac{X}{Y} = \frac{\frac{X}{Y}}{m}$ . Luego integraremos para todo valor de  $U$

$$\begin{aligned} f_F(v) &= \int_0^\infty f_{U,V}(v, u) du \\ &= \frac{\frac{n}{m} \left(\frac{n}{m}\right)^{\frac{n}{2}-1} v^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{m}{2}\right) 2^{\frac{n}{2}+\frac{m}{2}}} \int_0^\infty u^{\frac{n+m}{2}-1} e^{-u\left(\frac{nv}{2m}+\frac{1}{2}\right)} du. \end{aligned}$$

Recordando que para una v.a.  $G \sim \text{Gamma}(\alpha, \beta)$  se tiene

$$\int_0^\alpha \text{Gamma}(\alpha, \beta) dg = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} g^{\alpha-1} e^{-\frac{g}{\beta}} dg = 1,$$

entonces

$$\int_0^\infty g^{\alpha-1} e^{-\frac{g}{\beta}} dg = \Gamma(\alpha)\beta^\alpha.$$

## Método de la transformación o cambio de variable

Así, en la expresión para  $f_F(v)$ , podemos identificar la integral con esta última expresión si

$$\alpha = \frac{n+m}{2}, \quad \beta = \left( \frac{nv}{2m} + \frac{1}{2} \right)^{-1},$$

de donde

$$\begin{aligned} f_F(v) &= \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} v^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \Gamma\left(\frac{m+n}{2}\right) \left(\frac{vn}{2m} + \frac{1}{2}\right)^{-\frac{m+n}{2}} \\ &= \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})2^{\frac{n}{2}+\frac{m}{2}}} \left(\frac{n}{m}\right)^{\frac{n}{2}} v^{\frac{n}{2}-1} \left(v\frac{n}{m} + 1\right)^{-\frac{m+n}{2}}. \end{aligned}$$

Probabilidades bajo la distribución  $F_{n,m}$  sólo pueden calcularse a través de métodos numéricos.

## Método de la transformación o cambio de variable

**Ejemplo:** Consideremos  $X_1$  y  $X_2$  variables independientes con distribución  $N(0, 1)$ . Se define

$$Y_1 = \frac{X_1 + X_2}{2}, \quad Y_2 = \frac{(X_1 - X_2)^2}{2}.$$

Encuentra la distribución conjunta de  $(Y_1, Y_2)$ , sus marginales y muestra que son independientes.

**Nota:** Si  $(X_1, X_2)$  es una muestra aleatoria de tamaño 2 de una población normal,  $Y_1$  representa a  $\bar{X}$  y  $Y_2$  a  $S^2$ . Así, este resultado muestra que para  $n = 2$ ,  $(\bar{X}, s^2)$  son variables aleatorias independientes. Este resultado puede generalizarse para cualquier  $n$ .



## Método de la transformación o cambio de variable

La transformación definida por  $y_1 = \frac{x_1+x_2}{2}$ ,  $y_2 = \frac{(x_1-x_2)^2}{2}$  mapea

$$\mathcal{X} = \{(x_1, x_2) | x_1, x_2 \in \mathbb{R}\} \longrightarrow \mathcal{Y} = \{(y_1, y_2) | y_1 \in \mathbb{R}, y_2 \in [0, \infty)\}$$

pero no es una transformación uno a uno

$$T : y_1 = \frac{x_1 + x_2}{2} \qquad T_1^{-1} : x_1 = y_1 + \sqrt{\frac{y_2}{2}} \qquad T_2^{-1} : x_1 = y_1 - \sqrt{\frac{y_2}{2}}$$

$$y_2 = \frac{(x_1 - x_2)^2}{2} \qquad x_2 = y_1 - \sqrt{\frac{y_2}{2}} \qquad x_2 = y_1 + \sqrt{\frac{y_2}{2}}$$

Primero, podemos excluir  $y = 0$ , pero aún así la transformación definida no nos permite separar a  $\mathcal{X}$  en dos conjuntos disjuntos pues todos los puntos en  $x_1 = x_2$  coinciden.

## Método de la transformación o cambio de variable

En cada uno de esos puntos se tendrá  $y_2 = 0$ . Para evitar esta dificultad se puede definir  $f_{X_1, X_2}(x_1, x_2) = 0$  para todos los puntos donde  $x_1 = x_2$ , lo cual es posible sin alterar la función de probabilidad pues la probabilidad de una recta es cero (el volumen de una recta es cero). Así,

$$\mathcal{X} = \{(x_1, x_2) | x_i \in (-\infty, \infty) \ i = 1, 2, \ x_1 \neq x_2\},$$

y lo podemos partirlo como

$$\mathcal{X}_1 = \{(x_1, x_2) | x_1 > x_2\}, \quad \mathcal{X}_2 = \{(x_1, x_2) | x_1 < x_2\}.$$

Así, las transformaciones  $T_1^{-1}$  y  $T_2^{-1}$  están bien definidas. Entonces:

$$T_1 : \quad |J_1| = \begin{vmatrix} 1 & \sqrt{\frac{1}{2}} \frac{y_2^{-1/2}}{2} \\ 1 & -\sqrt{\frac{1}{2}} \frac{y_2^{-1/2}}{2} \end{vmatrix} = \left| -\frac{y_2^{-1/2}}{\sqrt{2}} \right| = \frac{1}{\sqrt{2y_2}}.$$

## Método de la transformación o cambio de variable

Notemos que

$$T_2 : \quad J_2 = \begin{vmatrix} 1 & -\sqrt{\frac{1}{2}} \frac{y_2^{-1/2}}{2} \\ 1 & \sqrt{\frac{1}{2}} \frac{y_2^{-1/2}}{2} \end{vmatrix} = \frac{y_2^{-1/2}}{\sqrt{2}} = \frac{1}{\sqrt{2}y_2} = |J_1|.$$

Lo cual implica

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{\sqrt{2}y_2} \frac{1}{2\pi} e^{-\frac{1}{2} \left\{ (y_1 + \sqrt{\frac{y_2}{2}})^2 + (y_1 - \sqrt{\frac{y_2}{2}})^2 \right\}} \\ &\quad + \frac{1}{\sqrt{2}y_2} \frac{1}{2\pi} e^{-\frac{1}{2} \left\{ (y_1 - \sqrt{\frac{y_2}{2}})^2 + (y_1 + \sqrt{\frac{y_2}{2}})^2 \right\}}, \end{aligned}$$

de donde

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{2}{2\pi\sqrt{2}y_2} e^{-\frac{1}{2}(2y_1^2 + y_2)}.$$

## Método de la transformación o cambio de variable

Notemos que la distribución conjunta se puede reescribir, multiplicando y dividiendo por las constantes adecuadas, como

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} \frac{\sqrt{2}}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{y_1^2}{2}} \frac{1}{2^{1/2} \Gamma(\frac{1}{2})} y_2^{\frac{1}{2}-1} e^{-\frac{1}{2} y_2} & y_1 \in \mathbb{R}, y_2 > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$\therefore Y_1 \sim N(0, 1/2)$  y  $Y_2 \sim \chi_1^2$  mostrando además que son independientes.

**Nota:** Dado que  $X_1 \sim N(0, 1)$  y  $X_2 \sim N(0, 1)$  son independientes, directamente es posible establecer la distribución de

$$Y_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2$$

lo cual no es más que una combinación lineal de normales con  $a = b = \frac{1}{2}$ .

## Método de la transformación o cambio de variable

$$\Rightarrow Y_1 \sim \text{Normal} \left( \frac{1}{2}(0) + \frac{1}{2}(0), \frac{1}{4}(1) + \frac{1}{4}(1) \right) \text{ i.e } Y_1 \sim N \left( 0, \frac{1}{2} \right)$$

Ademas, si  $Y_2 = \frac{(X_1 - X_2)^2}{2}$ , notemos que tambien  $\frac{X_1 - X_2}{\sqrt{2}}$  cae en el caso anterior con

$$a = \frac{1}{\sqrt{2}}, b = -\frac{1}{\sqrt{2}},$$

de donde

$$\frac{X_1}{\sqrt{2}} - \frac{X_2}{\sqrt{2}} \sim N \left( 0, \frac{1}{2}(1) + \frac{1}{2}(1) \right) = N(0, 1).$$

Aquí,  $Y_2$  representa una variable aleatoria  $[N(0, 1)]^2 \therefore Y_2 \sim \chi_{(1)}^2$ .

Ademas la conjunta se factoriza como  $f(y_1) \cdot f(y_2)$  lo cual nos da la independencia.

## Otros Ejemplos:

**Ejemplo 1.** Si  $X_1$  y  $X_2$  son variables aleatorias independientes que tienen densidad uniforme con  $\alpha = 0$  y  $\beta = 1$ , encontrar la densidad de  $Y = X_1 + X_2$ .

Como son independientes, la densidad conjunta es el producto de las densidades marginales, o sea,

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1, \quad 0 < x_2 < 1 \\ 0 & \text{otra parte} \end{cases}.$$

## Método de la transformación o cambio de variable

Así, la acumulada de  $Y$  está dada por

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \int_0^y \int_0^{y-x_2} 1 dx_1 dx_2 & 0 < y < 1 \\ 1 - \int_{y-1}^1 \int_{y-x_2}^1 1 dx_1 dx_2 & 1 < y < 2 \\ 1 & y > 2 \end{cases}.$$

Obsérvese que

$$\begin{aligned} \int_0^y \int_0^{y-x_2} 1 dx_1 dx_2 &= \int_0^y [x_1]_0^{y-x_2} dx_2 = \int_0^y (y - x_2) dx_2 = [yx_2 - \frac{x_2^2}{2}]_0^y \\ &= y^2 - \frac{y^2}{2} = \frac{y^2}{2}. \end{aligned}$$

## Método de la transformación o cambio de variable

$$\begin{aligned}\int_{y-1}^1 \int_{y-x_2}^1 1 dx_1 dx_2 &= \int_{y-1}^1 [x_1]_{y-x_2}^1 dx_2 = \int_{y-1}^1 [1 - (y - x_2)] dx_2 \\&= [x_2 - yx_2 + \frac{x_2^2}{2}]_{y-1}^1 \\&= [(1 - y + \frac{1}{2}) - ((y - 1) - y(y - 1) + \frac{(y - 1)^2}{2})] \\&= \frac{3}{2} - y - y + 1 + y(y - 1) - \frac{(y - 1)^2}{2} \\&= \frac{5}{2} - 2y + y^2 - y - \frac{(y^2 - 2y + 1)}{2} \\&= \frac{5}{2} + \frac{-4y + 2y^2 - 2y - y^2 + 2y - 1}{2} \\&= \frac{4 - 4y + y^2}{2} = \frac{(y - 2)^2}{2}.\end{aligned}$$



## Método de la transformación o cambio de variable

Por lo tanto

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 < y < 1 \\ 1 - \frac{(y-2)^2}{2} & 1 < y < 2 \\ 1 & y > 2 \end{cases}.$$

Así, la densidad está dada

$$f_Y(y) = \begin{cases} y & 0 < y < 1 \\ 2 - y & 1 < y < 2 \\ 0 & \text{c.o.p.} \end{cases}.$$

**La densidad de la suma de dos variables aleatorias independientes uniformes es triangular.**

## Método de la transformación o cambio de variable

**Ejemplo 2:** Si  $P$ , el precio de un producto (en dólares), y  $S$ , las ventas totales (en unidades de 10,000), son variables aleatorias cuya distribución conjunta puede obtenerse aproximadamente como la densidad de probabilidad conjunta

$$f(p, s) = \begin{cases} 5pe^{-ps} & 0.20 < p < 0.40, s > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}.$$

Obtener la densidad de probabilidad de la variable aleatoria  $V = SP$ , que es la cantidad total de dinero que se gasta en este producto en unidades de \$10,000.

$$\begin{aligned}F_V(v) &= P(V \leq v) = P(SP \leq v) = P(S \leq \frac{v}{p}) \\&= \int_{0.2}^{0.4} \int_0^{\frac{v}{p}} 5pe^{-ps} ds dp = -5 \int_{0.2}^{0.4} e^{-sp} \Big|_0^{\frac{v}{p}} dp = 5 \int_{0.2}^{0.4} (e^{-v} - 1) dp \\&= 1 - e^{-v}, \quad v > 0\end{aligned}$$

Derivando,  $f_V(v) = e^{-v}$ ,  $v > 0$ . Es decir,  $V \sim \text{Exp}(\beta = 1)$