

Introducción a la Ciencia de Datos

Repaso estadística y Probabilidades

INF-396

Prof: Juan G. Pavez S.
(Basada en clase profesor H.Allende)

Variables Aleatorias Discretas

- Sea $C \subseteq \Omega$ (soporte contable o denumerable)
- $f: C \rightarrow R$
$$C = \{c_i : i \in I \subseteq N\}$$
- $i) f(c_i) \geq 0$
- $ii) \sum_{i \in I} f(c_i) = 1$
- El conjunto $(x_i, f(x_i))$ se le denomina función de probabilidad o cuantía.

Función de Cuantía de v.a discreta

$$x(c_i) = x_i$$

$$P(A) = \sum_{j \in \{i : c_i \in C \cap A\}} f(c_j) = \sum_i P(X = x_i)$$

➤ Propiedades función de cuantía:

➤ $P(X = x_i) \geq 0$

➤ $\sum_i P(X = x_i) = 1$

➤ Función de Distribución

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

Función de Cuantía de v.a discreta

➤ Esperanza de una v.a.d. X

$$E[X] = \sum_i x_i P(X = x_i)$$

➤ Varianza de una v.a.d X

$$V[X] = \sum_i (x_i - E[X])^2 P(X = x_i)$$

- **Covarianza** entre X y Y (mide la asociación o dependencia lineal entre dos variables aleatorias)
- $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E[XY] - E[X]E[Y]$

Propiedades de la esperanza y Varianza

- **Propiedades de la esperanza**
- $E[ag(X) + bh(Y)] = aE[g(X)] + bE[h(Y)]$
- Sea X e Y cualesquier variables aleatoria
- $E[X + Y] = E[X] + E[Y]$
- Sea X e Y variables aleatorias **independientes** y g,h funciones
- $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Propiedades de la esperanza y Varianza

- Propiedades de la varianza
- $V[ag(X) + b] = a^2V[g(x)] + b$
- Sea X e Y variables aleatorias **independientes**
- $V[X + Y] = V[X] + V[Y]$
- Sea X e Y variables aleatorias **no independientes**
- $V[X + Y] = V[X] + V[Y] + 2COV[X, Y]$

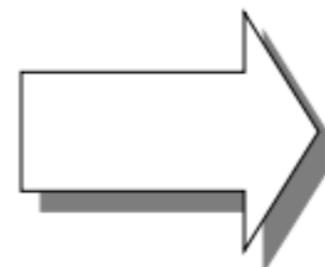
Distribución de Bernoulli

Consideremos un solo experimento ε

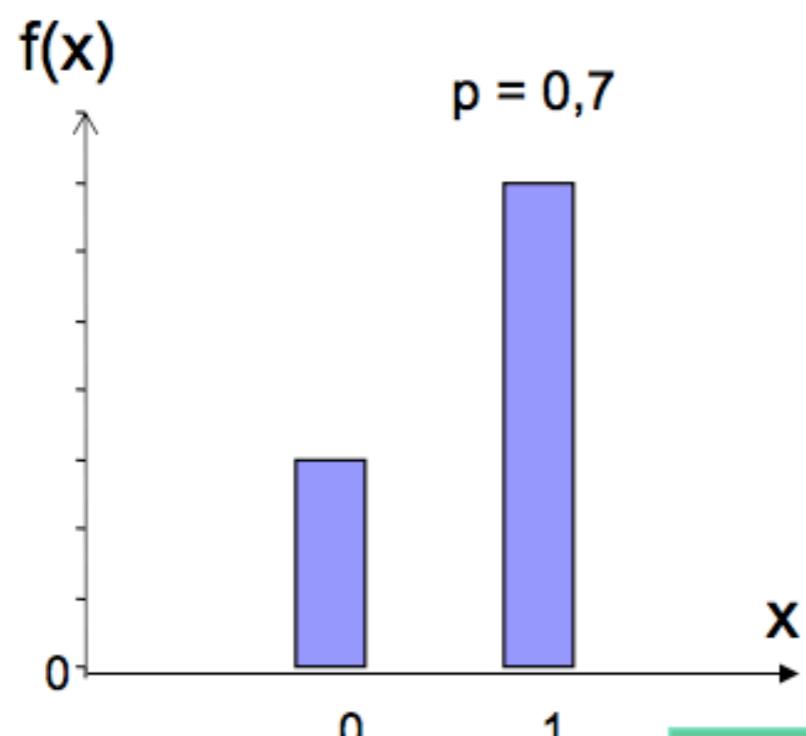
sea A un evento asociado con tal experimento.

supongamos que $P(A) = p$; luego $P(A^c) = 1 - p$

Sea la v.a. $X(A) = 1$
 $X(A^c) = 0$



$P(X = 1) = p$
 $P(X = 0) = 1 - p$



Entonces su función de cuantía es

$$f(x) = P(X=x) = p^x (1-p)^{1-x}$$

$$\begin{aligned} X &= 0, 1 \\ 0 &< p < 1 \end{aligned}$$

Distribución de Bernoulli

- Esperanza de la distribución de Bernoulli
- Varianza de la distribución de Bernoulli

Distribución Binomial

- Supongamos que de una línea de producción se extraen n piezas con reemplazo, las cuales pueden ser defectuosas o no con una probabilidad “ p ”.
- X : N° de piezas defectuosas en las n extracciones

Entonces

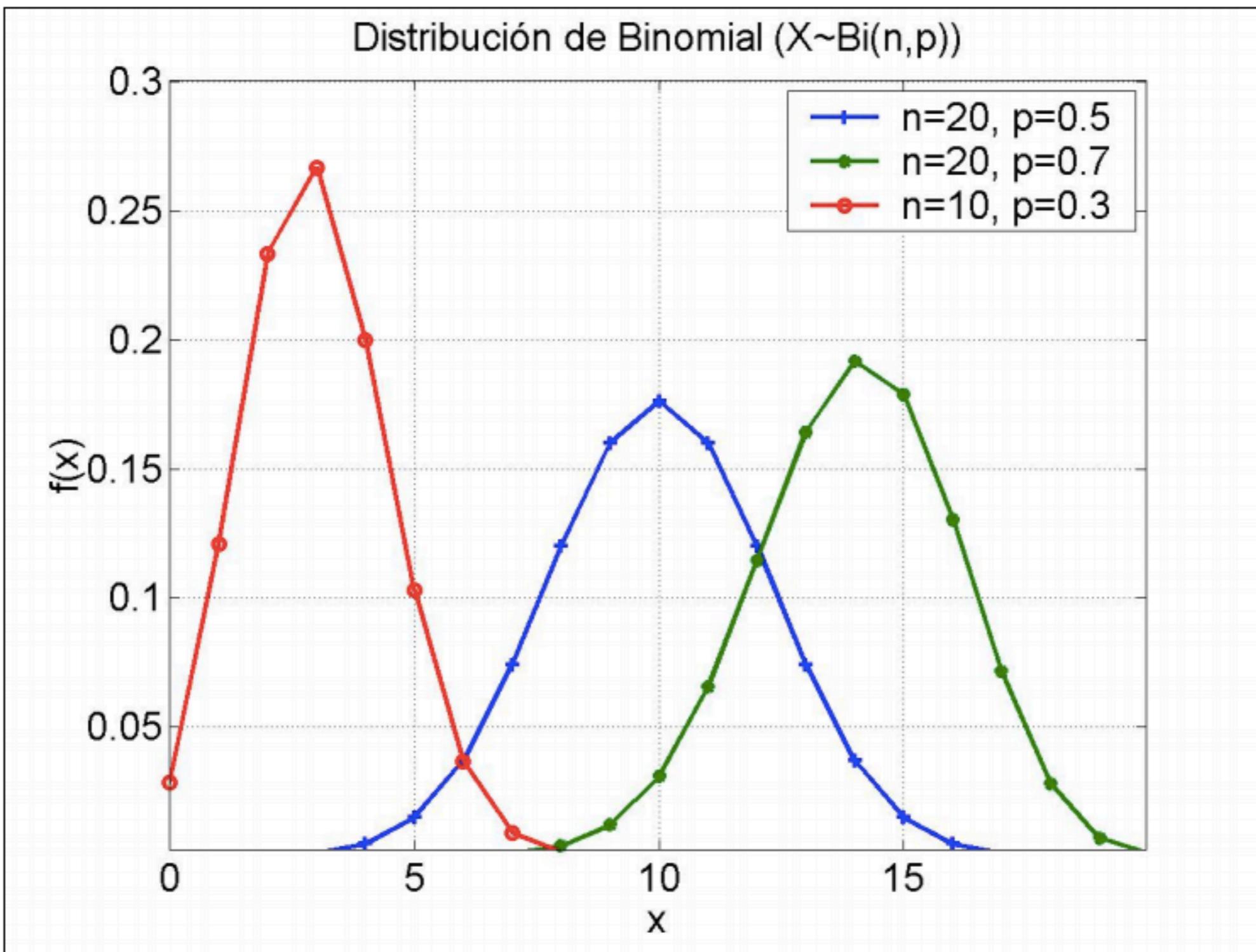
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n$$

Distribución Binomial

- Esperanza de la distribución de Binomial

-

Distribución Binomial



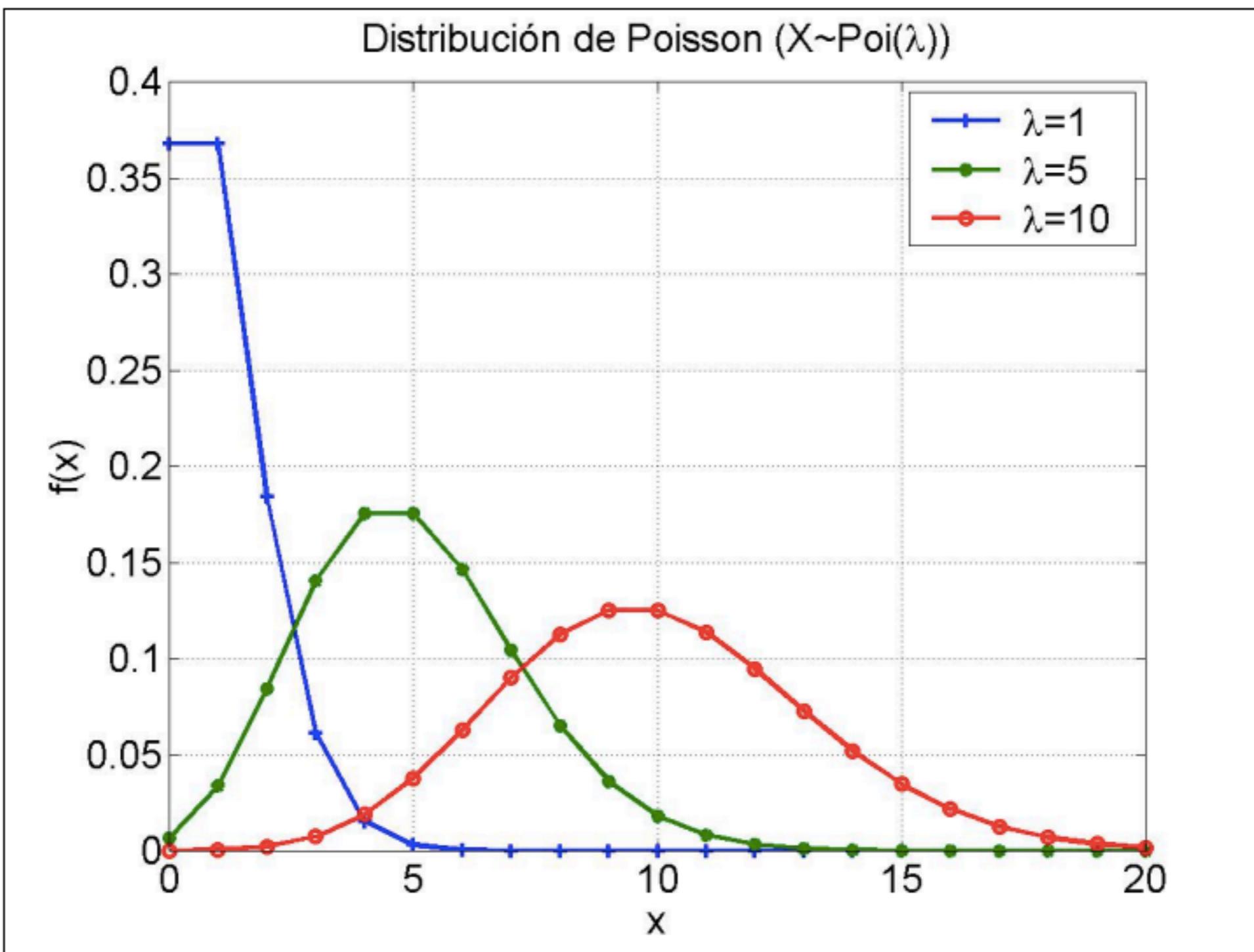
Distribución Poisson

- Supongamos que tenemos una muestra de tamaño grande, para lo cual la probabilidad de encontrar un artículo defectuoso es pequeño " p ", y por lo tanto " np " el número total de artículos defectuosos en la muestra.
 np =frecuencia de ocurrencia media
 k = número de eventos que ocurrirán
- Sea $\lambda = np$.

Entonces

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0,1,2,\dots$$

Distribución Poisson



Distribución Poisson

- Esperanza: $E[X] = \lambda$
- Varianza: $V[X] = \lambda$

- Caso límite: $X \sim B(n, p)$

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} I_{\{0,1,2,\dots,n\}}^{(k)}$$

con $n \rightarrow \infty$ y $p \approx 0$

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} I_{N_0}(k)$$

Variable Aleatoria Continua

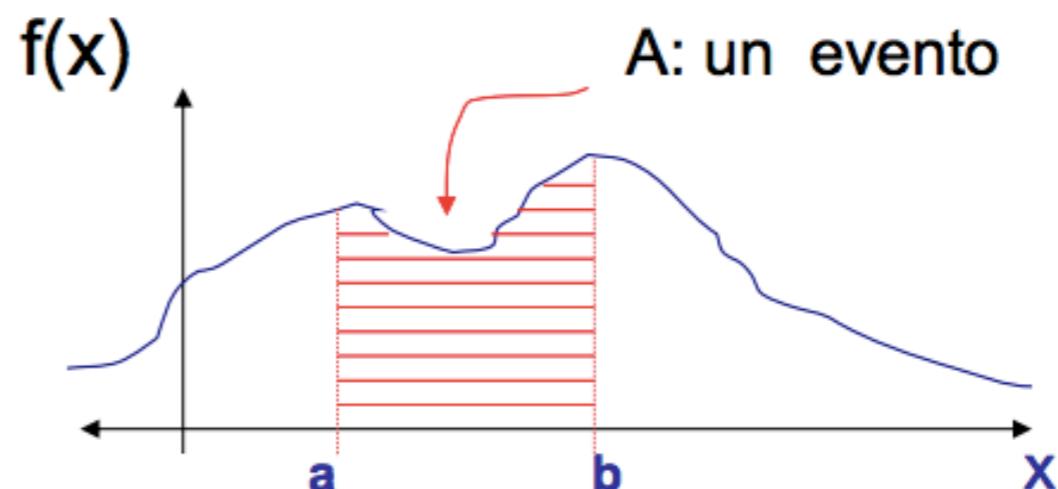
- Cuando el experimento se realiza sobre un espacio muestral Ω relacionado a escalas intervalares (mediciones de distancia, volúmenes, pesos, tiempos, ...)
- En este caso R_x es un intervalo o un conjunto de intervalos; existe una función continua especial:
- $f: R \rightarrow R$
- $$f(x) = \lim_{h \rightarrow 0} \frac{P(x < X < x + h)}{h} > 0$$

Variable Aleatoria Continua

Sea X una variable aleatoria continua. La función densidad de probabilidad (pdf) es una función que satisface:

$$f(x) > 0; \quad \forall x \in R_x \in -\infty, +\infty$$

$$\int_{R_x} f(x) dx = 1$$



$$A: \{x | a < x \leq b\}$$

$$P(A) = P(a < x < b) = \int_a^b f(x) dx$$

H. Allende

Variables Aleatorias Continuas

- Sea $C \subseteq \Omega$ (soporte contable o denumerable)
- $f: C \rightarrow R$ $C = \{c_i : i \in I \subseteq N\}$
- i) $f(x) \geq 0$
- ii) $\int_{-\infty}^{\infty} f(x)dx = 1$
- El conjunto $(x_i, f(x_i))$ se le denomina función de probabilidad o cuantía.

Variables Aleatorias Discretas

$$1. P(a \leq x \leq b) = \int_a^b f(x)dx$$

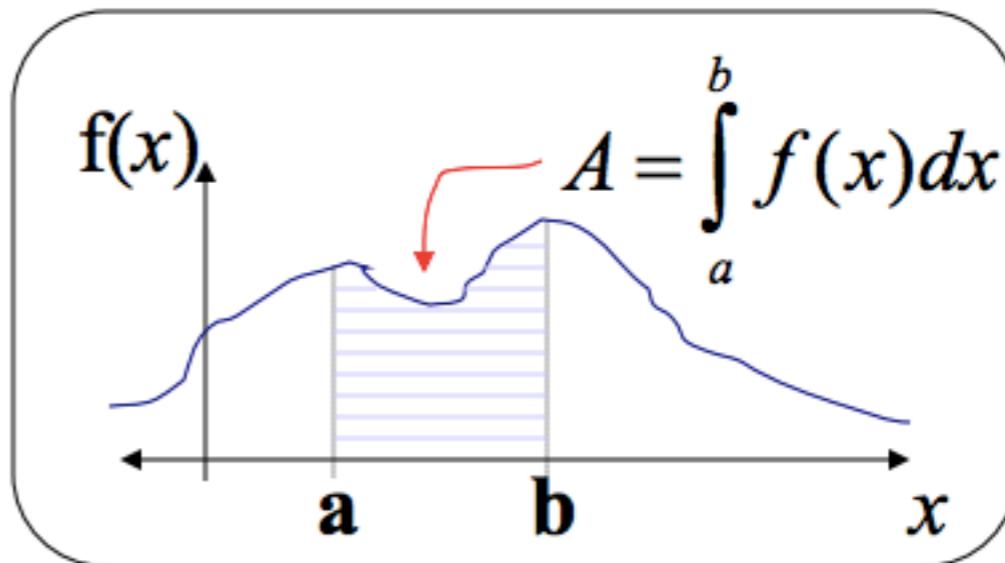
$$2. F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$$

$$3. F(-\infty) = 0 ; F(\infty) = 1$$

4. F_x es no decreciente

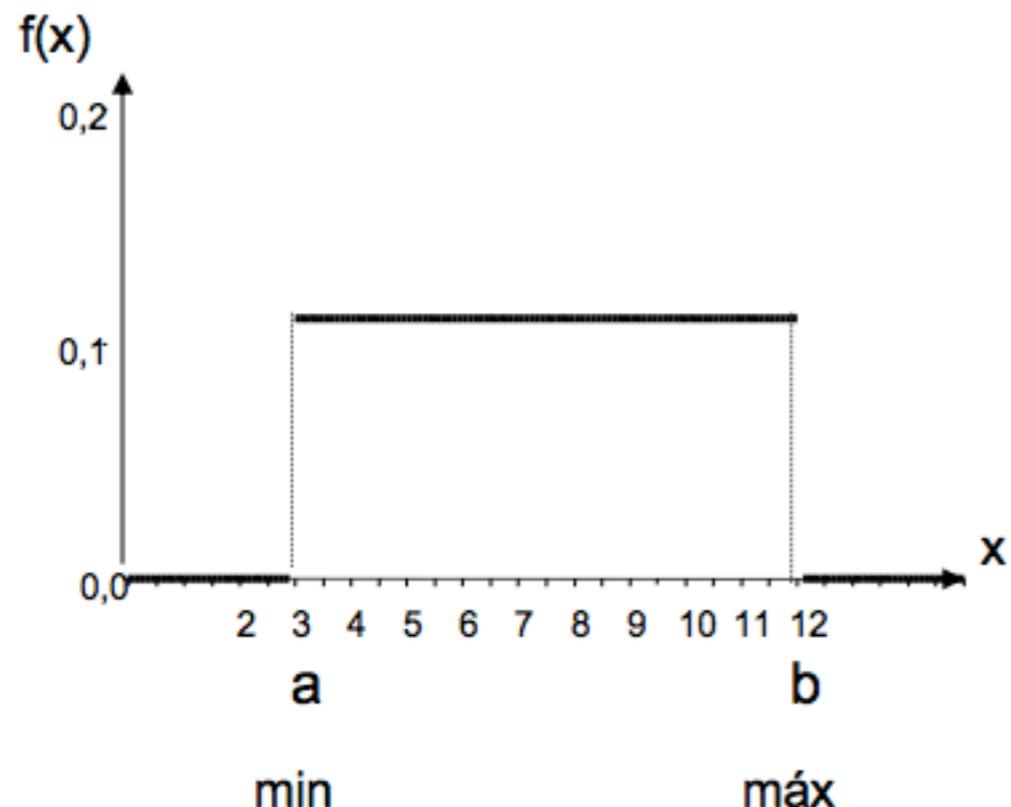
$$5. E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$6. V[X] = \int_R (x - E[X])^2 f(x)dx$$



Distribución Uniforme

Sea X una variable aleatoria continua que puede tomar cualquier valor entre $a \leq x \leq b$; cuya pdf es:



$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

Sea $a = 3; b = 12$

A: el evento $\{4 < x < 7\}$

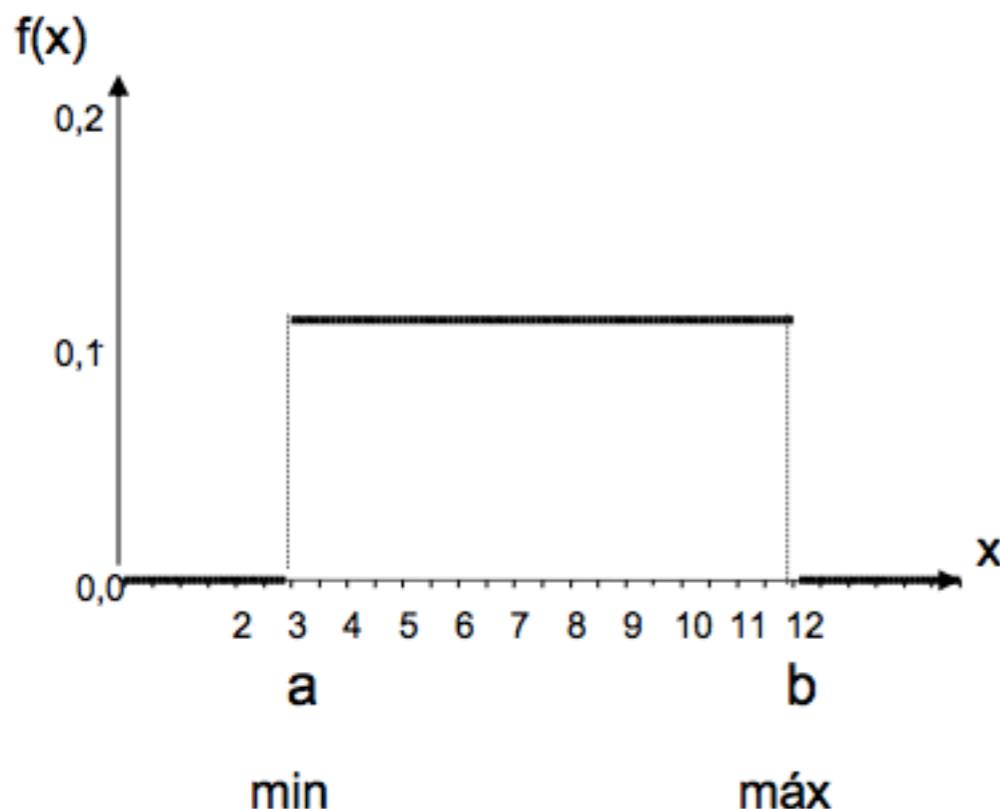
Entonces:

$$P(A) = P(4 < x < 7) = \int_{4}^{7} \frac{1}{9} dx$$

$$P(A) = \frac{1}{3}$$

Distribución Uniforme

Sea X una variable aleatoria continua que puede tomar cualquier valor entre $a \leq x \leq b$; cuya pdf es:



$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

Sea $a = 3; b = 12$

A: el evento $\{4 < x < 7\}$

Entonces:

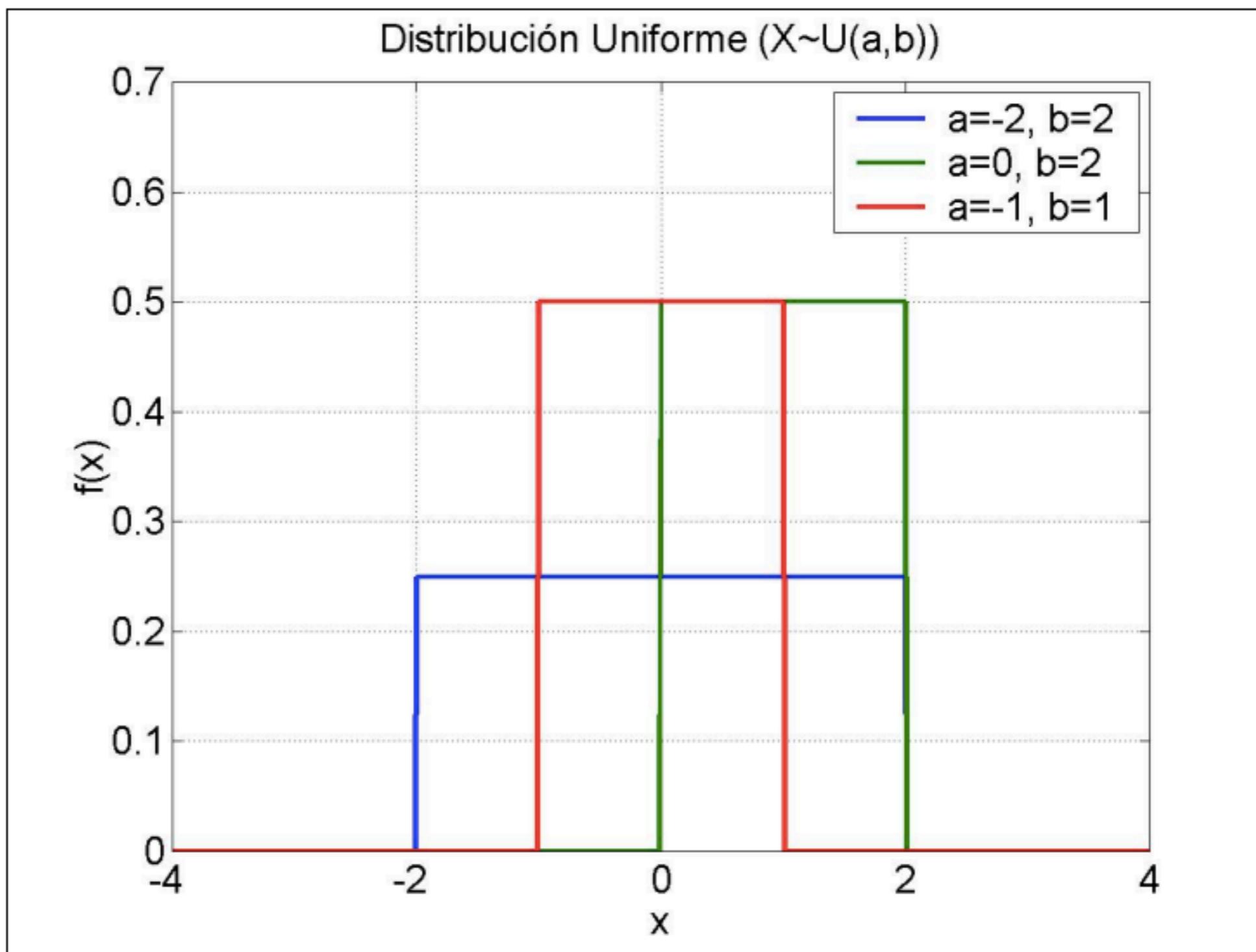
$$P(A) = P(4 < x < 7) = \int_{4}^{7} \frac{1}{9} dx$$

$$P(A) = \frac{1}{3}$$

Distribución Uniforme

- Esperanza de la distribución de Uniforme
- Varianza de la distribución Uniforme
-

Distribución Uniforme



Distribución Normal

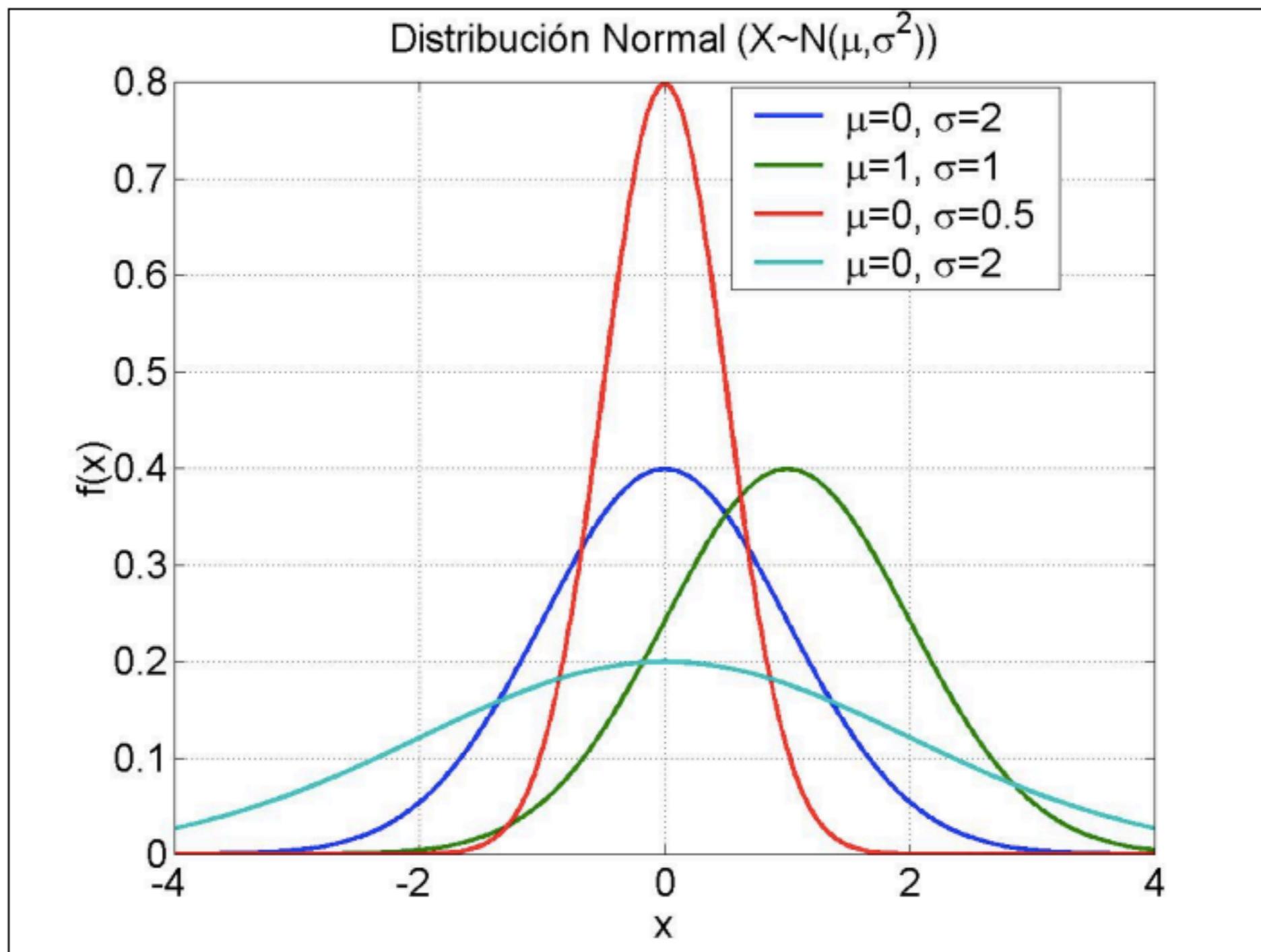
➤ Función de densidad

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

- La función de Distribución no tiene expresión analítica. (Usar tablas o calculadoras)

- Esperanza $E[X] = \mu$
- Varianza $V[X] = \sigma^2$
- Notación: $X \sim N(\mu, \sigma^2)$

Distribución Normal



Distribución Normal

Estandarización

Haciendo

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

se tiene que:

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, z \in R$$

y $F_z(z)$ se obtiene de tablas !

Distribución Normal

Estandarización

Haciendo

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

se tiene que:

$$f_z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, z \in R$$

y $F_z(z)$ se obtiene de tablas !

Distribución Normal

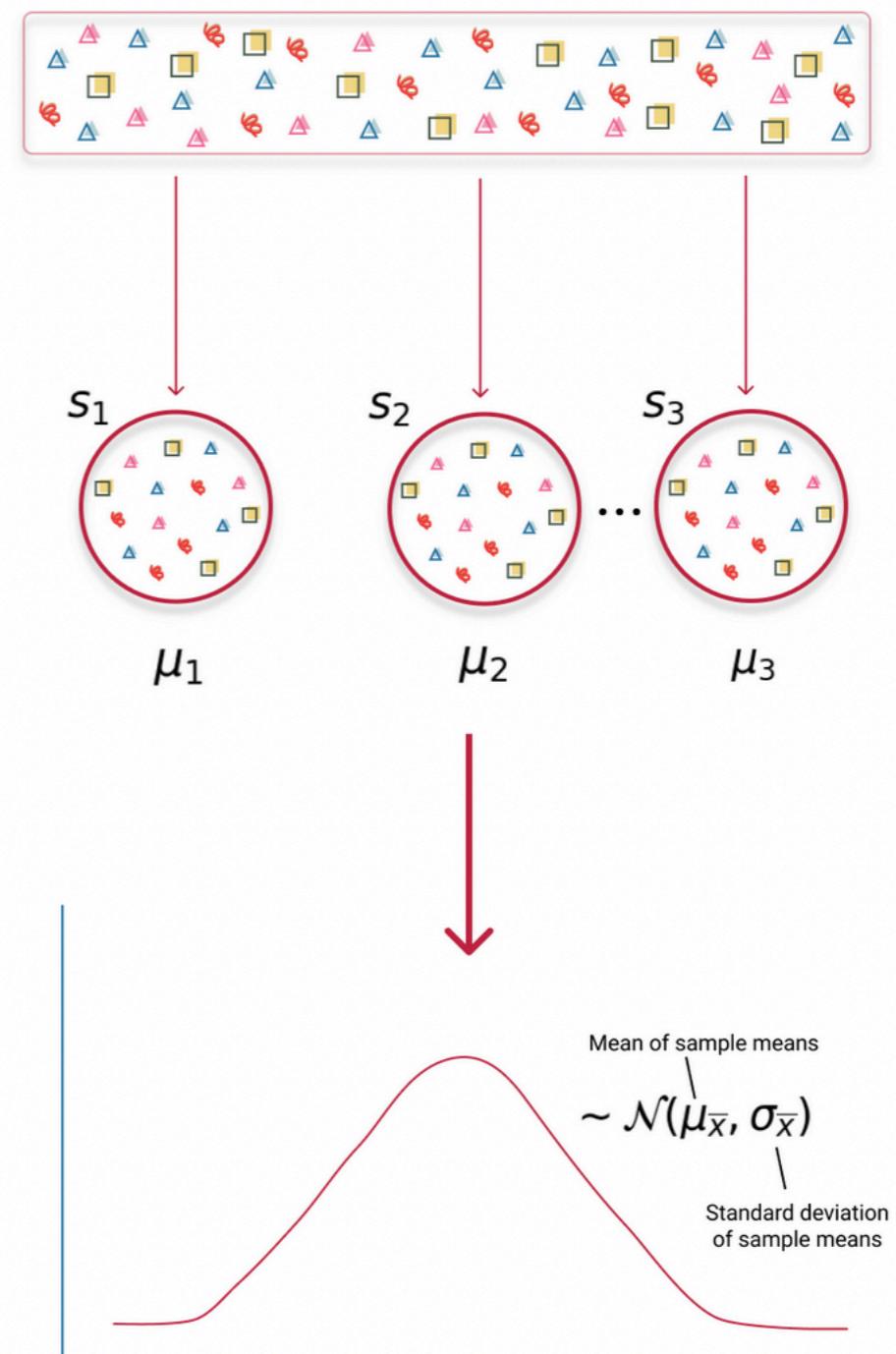
- **Teorema del límite central**
- Sea X_1, X_2, \dots, X_n un conjunto de variables aleatorias iid con valor esperado $E[X] = \mu < \infty, 0 < V[x] = \sigma^2 < \infty$ entonces la variable aleatoria que representa la media de la distribución converge a

$$\bar{X} \rightarrow N(\mu, \sigma^2/n)$$

<https://seeing-theory.brown.edu/probability-distributions/index.html#section3>

Distribución Normal

- Teorema del límite central

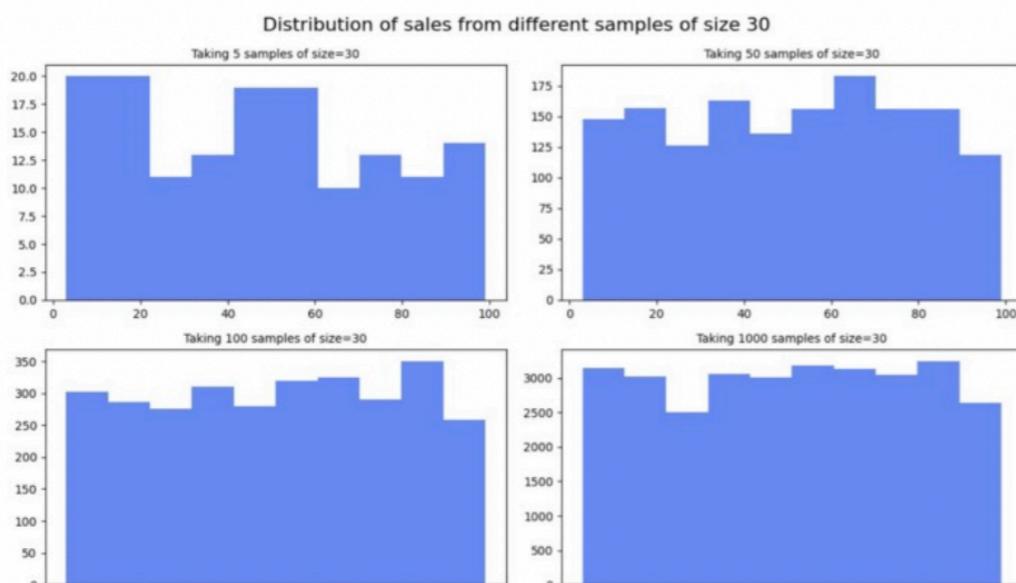


Distribución Normal

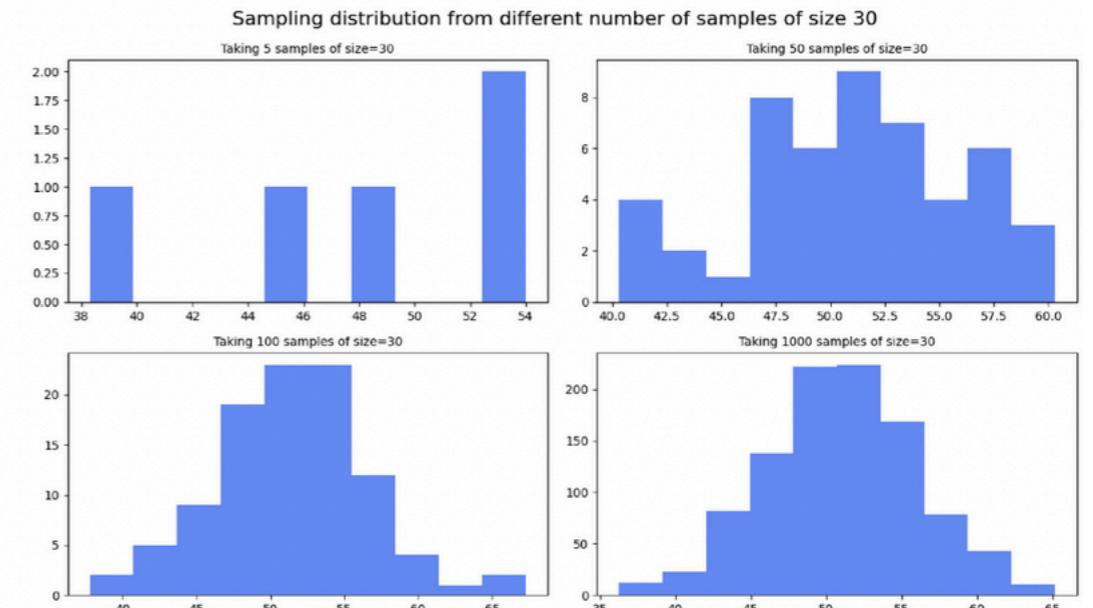
- **Teorema del límite central**
- Eres un manager regional de una cadena de super mercados, y debes reponer todas las semanas el agua embotellada. Te gustaría calcular cuantas botellas debes reponer semanalmente a cada tienda de manera de minimizar la pérdida.

Distribución Normal

- **Teorema del límite central**
- Eres un manager regional de una cadena de super mercados, y debes reponer todas las semanas el agua embotellada. Te gustaría calcular cuantas botellas debes reponer semanalmente a cada tienda de manera de minimizar la pérdida.



Distribution of different number of samples taken from the population, each sample with 30 data points.



Distribución Exponencial

➤ Función de densidad

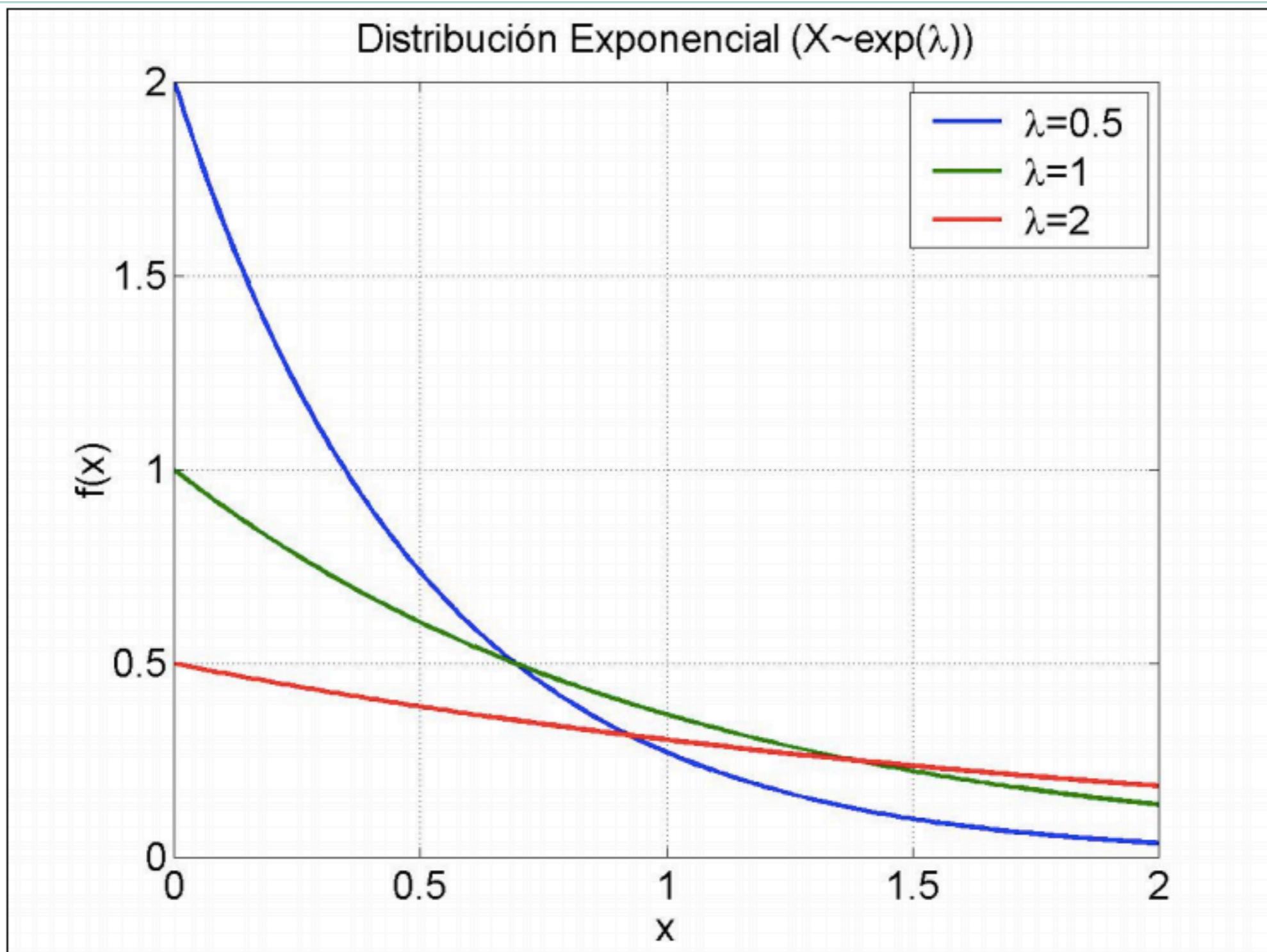
$$f_X(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \quad \text{si } x \geq 0, \lambda > 0$$

➤ Función de Distribución es

$$F_X(x) = 1 - e^{-\frac{x}{\lambda}} \quad x \geq 0$$

- Esperanza $E[X] = \lambda$
- Varianza $V[X] = \lambda^2$
- Notación: $X \sim \exp(\lambda)$

Distribución Exponencial



Distribución Gamma

➤ Función de densidad

$$f_X(x, \alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha) \beta^\alpha} I_{R^+}(x)$$

➤ Función de Distribución es

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t, \alpha, \beta) dt$$

➤ Esperanza

$$E[X] = \alpha\beta$$

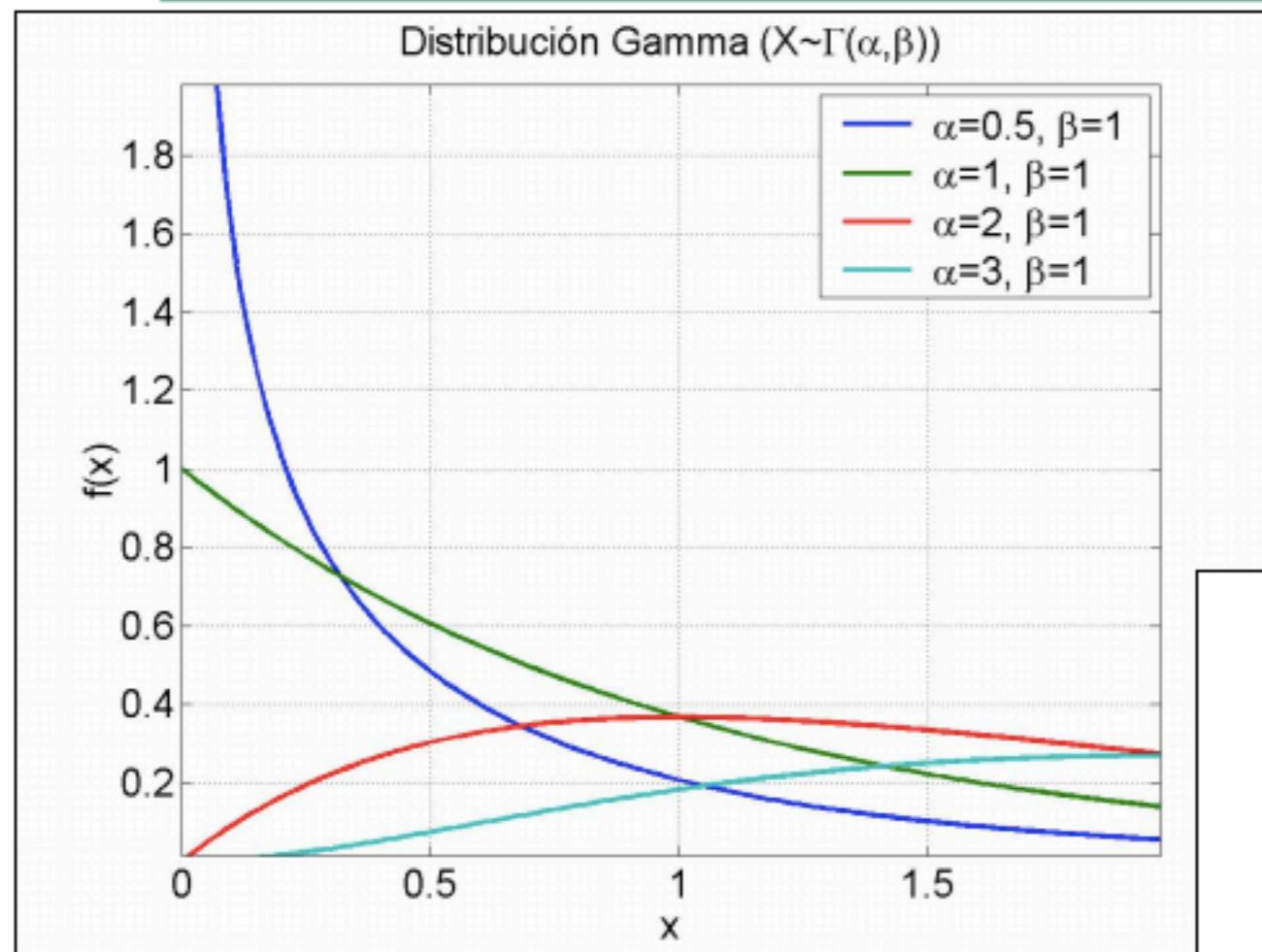
➤ Varianza

$$V[X] = \alpha\beta^2$$

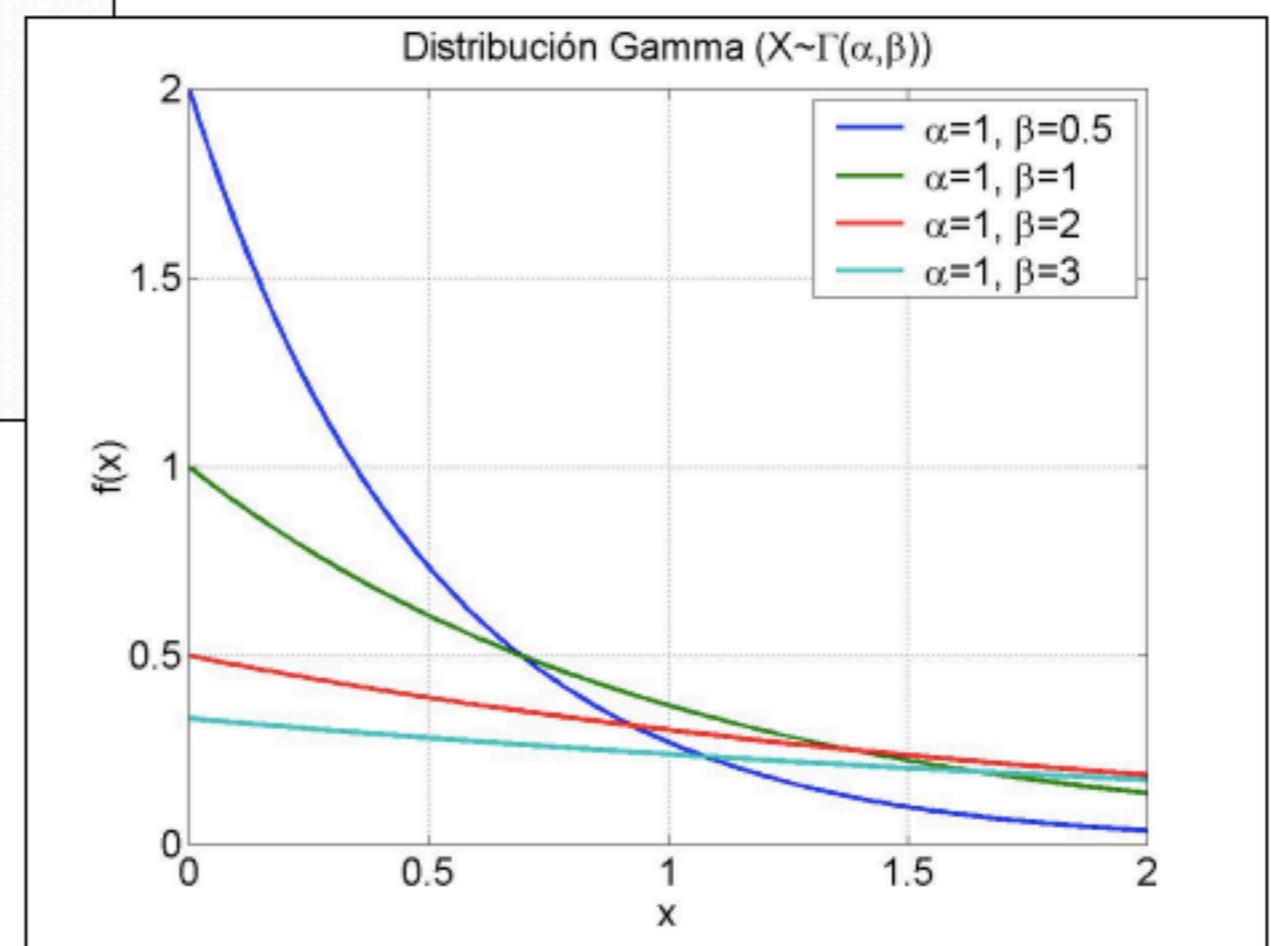
$$\Gamma(n) = \int_0^{\infty} y^{n-1} e^{-y} dy \quad n > 0$$

➤ Notación: $X \sim \text{Gamma}(\alpha, \beta) = \Gamma(\alpha, \beta)$

Distribución Gamma



$$X \sim \Gamma(\alpha, \beta)$$



Distribución Beta

➤ Función de densidad

$$f_X(x, r, s) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1} (1-x)^{s-1} I_{[0,1]}(x)$$

➤ Función de Distribución es

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(u, r, s) du$$

$$\beta(r, s) = \int_0^1 x^{r-1} (1-x)^{s-1} dx$$

➤ Esperanza

$$E[X] = \frac{r}{r+s}$$

$$E[X^\mu] = \frac{\Gamma(r+s)\Gamma(r+\mu)}{\Gamma(r)\Gamma(r+s+\mu)}$$

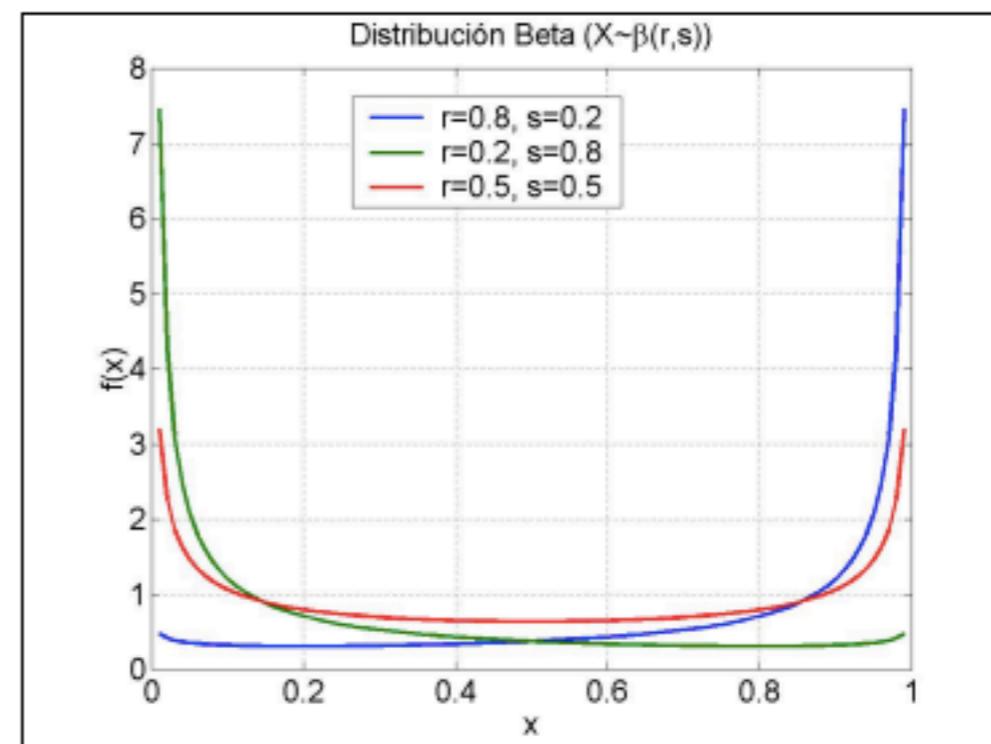
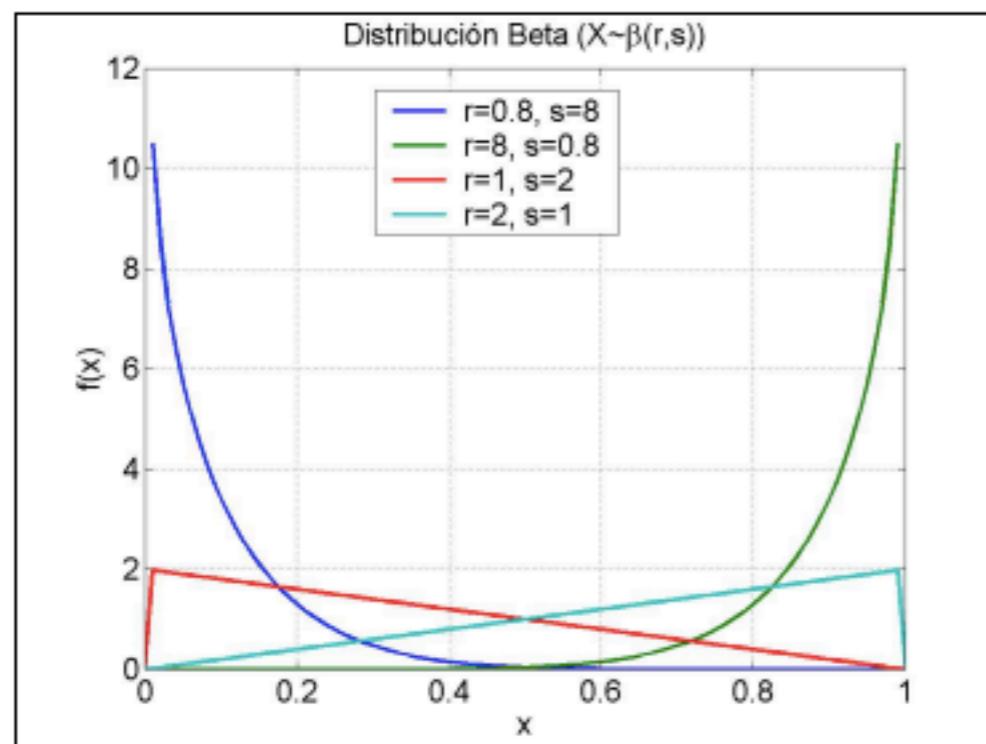
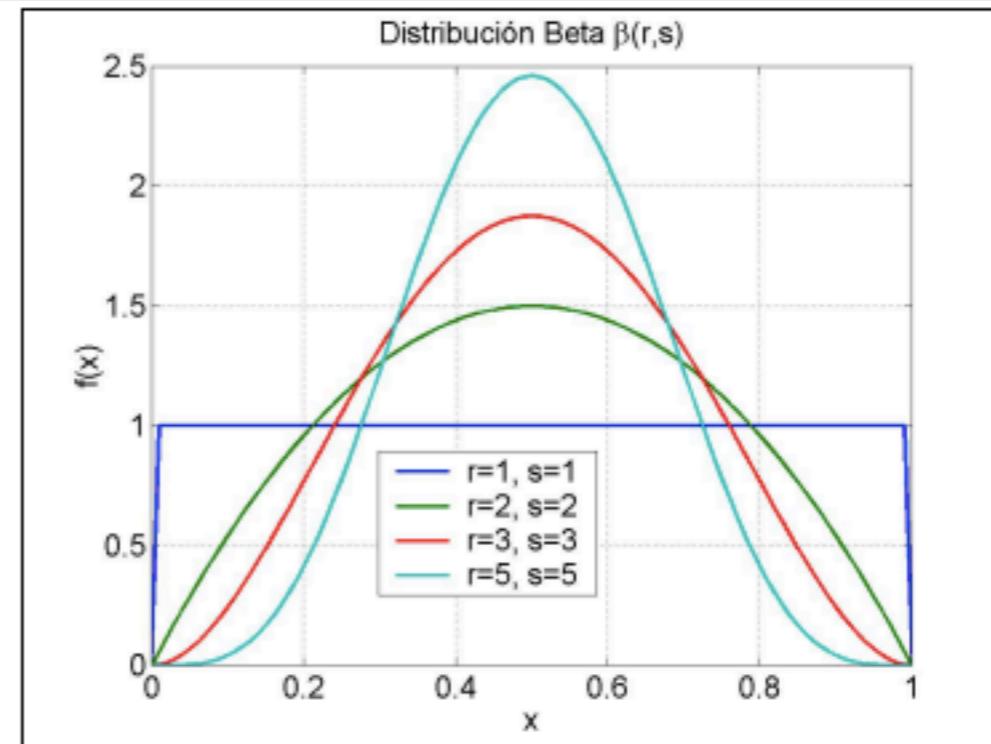
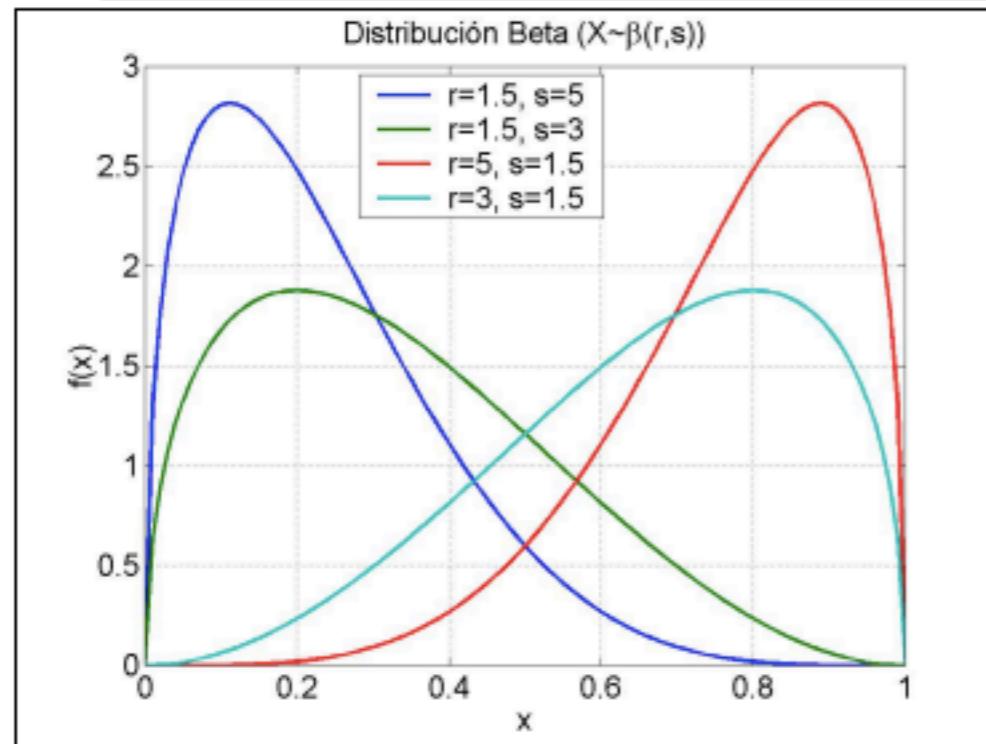
➤ Varianza

$$V[X] = \frac{rs}{(r+s)^2(r+s+1)}$$

➤ Notación:

$$X \sim Beta(r, s) = \beta(r, s)$$

Distribución Beta



Estimación de Parámetros

- Sea un modelo estadístico $f(x, \theta)$ con parámetros desconocidos θ .
- Un estimador es una regla que nos indica cómo obtener un parámetro de un modelo, basándose en la información contenida en una muestra aleatoria
- $T : X \rightarrow \tau \in \Xi$
$$T : x \rightarrow T(x) = T(x_1, x_2, \dots, x_n)$$
- $T(x)$ es el estimador de $\theta \in \Xi$

Método de Momentos

Momentos Observados

$$m_1 = 1/n \sum_{i=1}^n X_i^1 ,$$

$$m_2 = 1/n \sum_{i=1}^n X_i^2 ,$$

⋮

$$m_k = 1/n \sum_{i=1}^n X_i^k ,$$

Momentos Observados
(centrados en cero)

$$\mu_1 = E[X^1]$$

$$\mu_2 = E[X^2]$$

⋮

$$\mu_k = E[X^k]$$

y resolvemos el sistema de ecuaciones:

$$m_r = \mu_r, \quad r = 1, \dots, k$$

Método de Máxima Verosimilitud

- El método de máxima verosimilitud (MV) es el más popular para derivar estimadores.
- Sea $x_1, x_2, x_3, \dots, x_n$ una muestra i.i.d desde una población con función de densidad
 - $f(x | \theta_1, \theta_2, \dots, \theta_k)$
- Se define la **función de verosimilitud** cómo:
$$L(X | \theta) = L(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_k)$$
$$= \prod_{i=1}^n f(x_i | \theta_1, \theta_2, \dots, \theta_k)$$

Método de Máxima Verosimilitud

- Se define la **función de verosimilitud** cómo:

$$\begin{aligned} L(X \mid \theta) &= L(x_1, x_2, \dots, x_n \mid \theta_1, \theta_2, \dots, \theta_k) \\ &= \prod_{i=1}^n f(x_i \mid \theta_1, \theta_2, \dots, \theta_k) \end{aligned}$$

- El estimador de máxima verosimilitud $\hat{\theta}$ es el valor de los parámetros θ en los cuales **la verosimilitud alcanza su valor máximo**.

<https://seeing-theory.brown.edu/bayesian-inference/index.html#section2>

Método de Máxima Verosimilitud

- El estimador de máxima verosimilitud $\hat{\theta}$ es el valor de los parámetros θ en los cuales **la verosimilitud alcanza su valor máximo.**
- Es decir: Es el estimador que resuelve:
 - $\frac{\partial}{\partial \theta_i} L(X | \theta) = 0 \quad i = 1,..,k$
 - $\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(X | \theta) \right] |_{\theta=\hat{\theta}} < 0$

Método de Máxima Verosimilitud

- Muchas veces facilita el trabajo trabajar con la función de **log verosimilitud**, que mantiene el máximo de la optimización
- $$l(x | \theta) = \ln L(x | \theta) = \sum_{i=1}^n \ln f(x_i | \theta_1, \theta_2, \dots, \theta_k)$$

Método de Máxima Verosimilitud

- Considerar la distribución Binomial

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

Método de Máxima Verosimilitud

- Considerar la distribución normal

$$L(\mu, \sigma^2; x) = \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

Método de Máxima Verosimilitud

- Igualando a 0:

$$\frac{\partial l(x | \theta)}{\partial \theta} |_{\hat{\theta}} = 0$$

- $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

Método de Máxima Verosimilitud

- Su verosimilitud es:

$$\bullet \frac{\partial l(x | \mu, \sigma^2)}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu)$$

$$\bullet \frac{\partial l(x | \mu, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} - \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2$$

- Ahora la matriz Hesiana evaluada en los estimadores debe ser :

$$\bullet \frac{\partial^2 l(x | \theta)}{\partial \theta \partial \theta^T} = \begin{pmatrix} \frac{\partial^2 l(x | \theta)}{\partial \mu^2} & \frac{\partial^2 l(x | \theta)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 l(x | \theta)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 l(x | \theta)}{\partial \sigma^4} \end{pmatrix} \Big|_{\hat{\theta}} \text{ definida negativa}$$

Método de Máxima Verosimilitud

- Su verosimilitud es:

$$\bullet \frac{\partial l(x | \mu, \sigma^2)}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \mu)$$

$$\bullet \frac{\partial l(x | \mu, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} - \frac{1}{2\sigma^4} \sum_{i=1}^N (x_i - \mu)^2$$

$$\bullet \frac{\partial^2 l(x | \theta)}{\partial \theta \partial \theta^T} = \begin{pmatrix} \frac{\partial^2 l(x | \theta)}{\partial \mu^2} & \frac{\partial^2 l(x | \theta)}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 l(x | \theta)}{\partial \sigma^2 \partial \mu} & \frac{\partial^2 l(x | \theta)}{\partial \sigma^4} \end{pmatrix} = \begin{pmatrix} -\frac{N}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{N}{2\hat{\sigma}^2} \end{pmatrix}$$

que es definida negativa

Multivariate Distributions

Bivariate Distribution

- $p(x, y) = P(X = x \text{ and } Y = y)$

Considerar las siguientes variables aleatorias para 2 monedas lanzadas 3 veces

X: # caras para moneda 1, Y: # de sellos para moneda 2,
Z: #sellos para moneda 1

Entonces $p(x, y) = p(x)p(y)$ y

		y						z				
		0	1	2	3			0	1	2	3	
		0	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$			0	0	0	$\frac{1}{8}$
		1	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$			1	0	0	$\frac{3}{8}$
		2	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$			2	0	$\frac{3}{8}$	0
		3	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$			3	$\frac{1}{8}$	0	0

$p(x, y)$ $p(x, z)$

Multivariate Distributions

Marginal Distribution

- $$p(x) = \sum_y p(x, y)$$

$$p(y) = \sum_x p(x, y)$$

		y				$p_X(x)$			$p_X(x)$				
		0	1	2	3		x	0	1	2	3		
x	0	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$		
	1	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{3}{8}$	1	0	0	$\frac{3}{8}$	$\frac{3}{8}$		
	2	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{3}{8}$	2	0	$\frac{3}{8}$	0	$\frac{3}{8}$		
	3	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$	$\frac{1}{8}$	3	$\frac{1}{8}$	0	0	$\frac{1}{8}$		
		$P_y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$			$p_z(z)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Multivariate Distributions

Marginal Distribution

- $$p(x) = \sum_y p(x, y)$$

$$p(y) = \sum_x p(x, y)$$

		y				$p_X(x)$			$p_X(x)$				
		0	1	2	3		x	0	1	2	3		
x	0	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$		
	1	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{3}{8}$	1	0	0	$\frac{3}{8}$	$\frac{3}{8}$		
	2	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{3}{8}$	2	0	$\frac{3}{8}$	0	$\frac{3}{8}$		
	3	$\frac{1}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$	$\frac{1}{8}$	3	$\frac{1}{8}$	0	0	$\frac{1}{8}$		
		$P_y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$			$p_z(z)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Multivariate Distributions

Conditional Distribution

- $p(y|x) = \frac{p(x,y)}{p(x)}$ $f(y|x) = \frac{f(x,y)}{f(x)}$
Intuitivamente simboliza la probabilidad de que $Y=y$ (para algún valor de Y) dado que $X=x$
- Considerar la distribución
$$f(x,y) = y\left(\frac{1}{2} - x\right) + x \quad \text{si } 0 < x < 1, 0 < y < 2 \text{ o } 0 \text{ de otra forma}$$
- Chequear que esta función es una distribución bivariada y calcular la distribución marginal de x

Multivariate Distributions

Covarianza

$$\begin{aligned} COV(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= \int_{x \in \Omega_x} \int_{y \in \Omega_y} (x - E(x))(y - E(y))f(x, y)dxdy \end{aligned}$$

- Esta nos dice:
 - Si es positiva: valores grandes de X se asocian a valores grandes de Y y viceversa.
 - Si es negativa: Valores grandes de X se asocian a valores pequeños de Y y viceversa.
 - 0 covarianza no es lo mismo que independencia

Multivariate Distributions

Covarianza

- En más dimensiones es común utilizar la matriz de covarianza para resumir la información

-

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \text{Cov}(X_2, X_3) & \cdots & \text{Cov}(X_2, X_n) \\ \text{Cov}(X_1, X_3) & \cdots & & & \vdots \\ \vdots & & & \ddots & \vdots \\ \text{Cov}(X_1, X_n) & \cdots & \cdots & \cdots & \text{Var}(X_n) \end{pmatrix}$$

Multivariate Distributions

Normal Distribution

- $$N(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
Suponer que tenemos $x \sim N(x | \mu_1, \sigma^2)$, $y \sim N(y | \mu_2, \sigma^2)$ independientes.

$$p(x, y) = p(x)p(y)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma^2}}$$

Entonces

$$= (2\pi\sigma^2)^{-1} e^{\frac{-1}{2} \left[(x-\mu_1)(\sigma^2)^{-1}(x-\mu_1) + (y-\mu_2)(\sigma^2)^{-1}(y-\mu_2) \right]}$$

- $$= (2\pi\sigma^2)^{-1} e^{\frac{-1}{2} \left[(x-\mu_1)(y-\mu_2) \right]} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}^{-1} \begin{bmatrix} (x-\mu_1) \\ (y-\mu_2) \end{bmatrix}$$

Multivariate Distributions

The multivariate normal distribution.

$$p(x | y = c, \mu_c, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^2} \exp\left(-\frac{1}{2}(x - u)^T \Sigma^{-1} (x - u)\right)$$

The sample covariance matrix:

$$S_{p \times p} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{12} & s_{11} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}$$

where

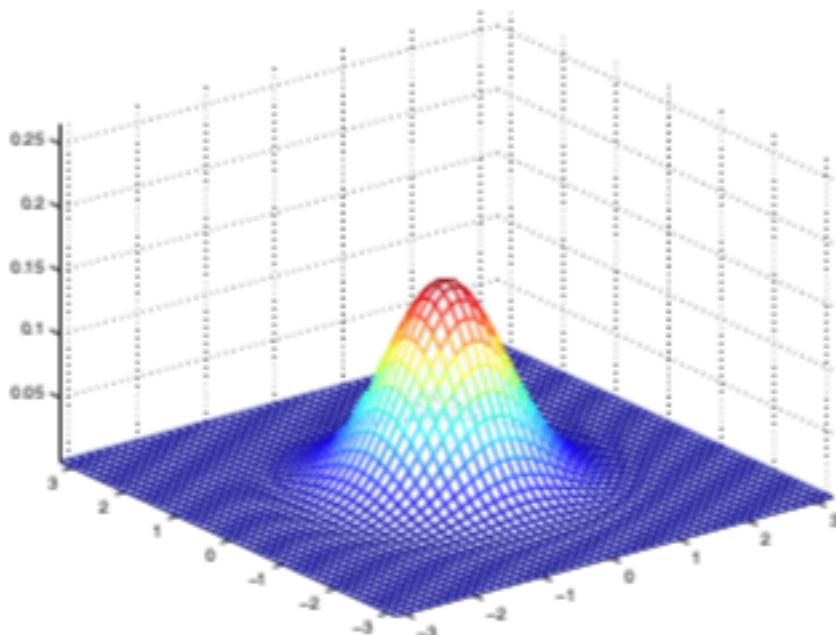
$$s_{ik} = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)$$

$$\hat{\mu}_{mle} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\hat{\Sigma}_{mle} = \frac{1}{N} \left(\sum_{i=1}^N x_i x_i^T \right) - \bar{x} \bar{x}^T$$

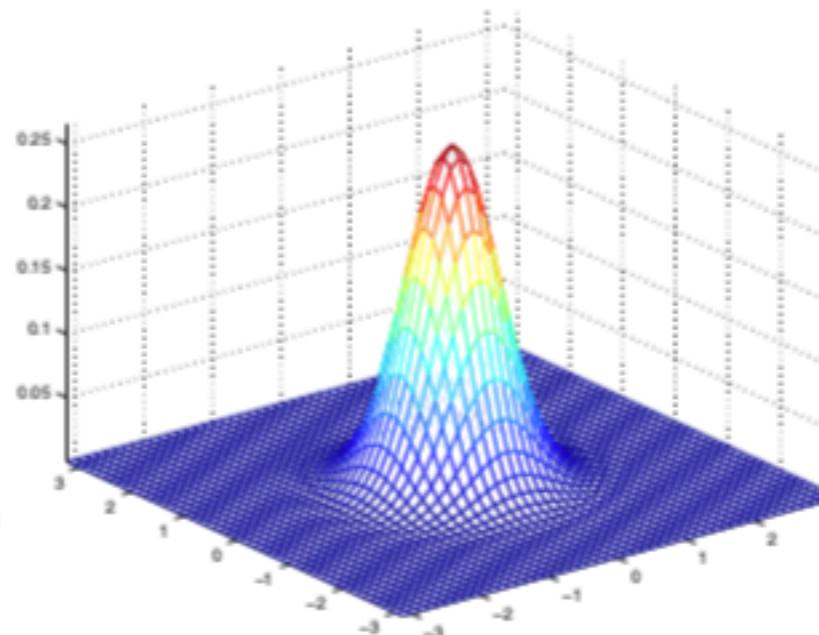
Multivariate Distributions

The multivariate normal distribution.

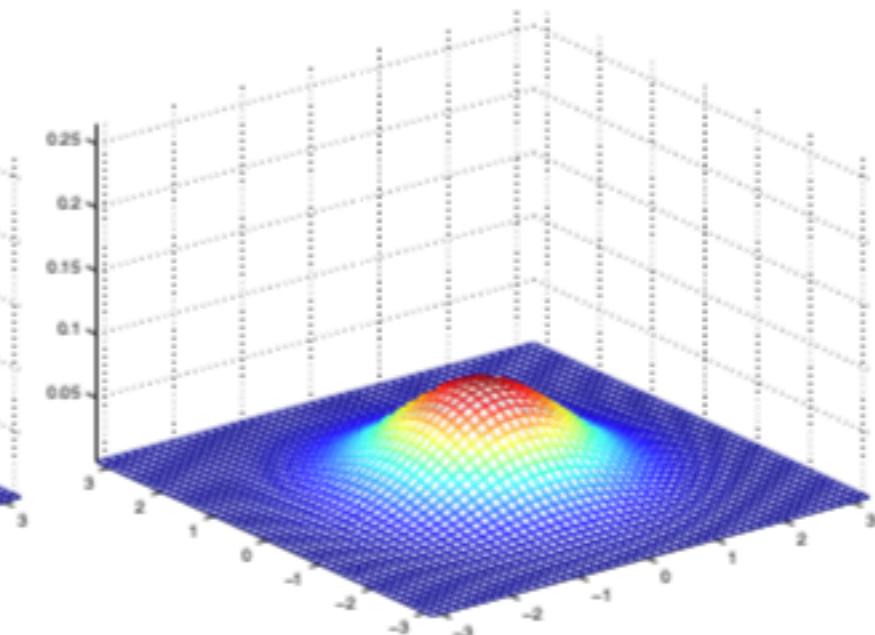


$$\Sigma = I$$

$$\mu^T = [0., 0.]$$



$$\Sigma = 0.6I$$

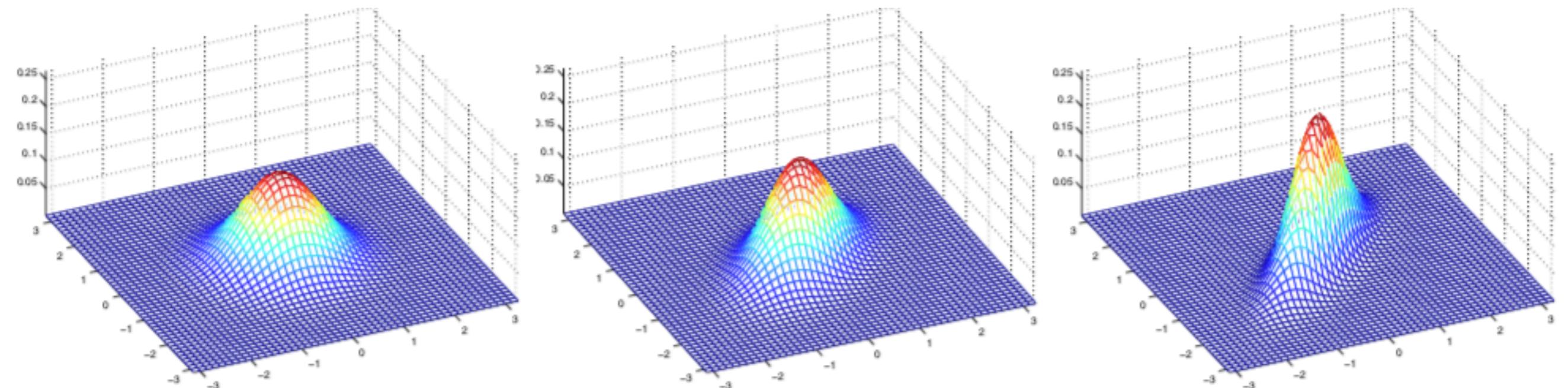


$$\Sigma = 2I$$

Credit: Machine Learning. Ng, A.

Multivariate Distributions

The multivariate normal distribution.



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

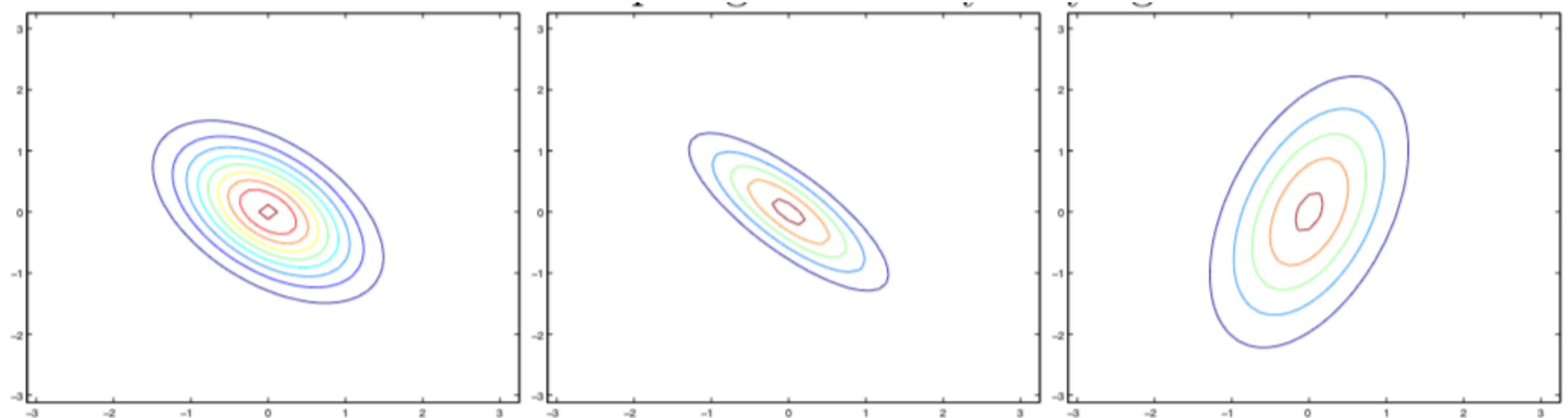
$$\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

$$\mu^T = [0., 0.]$$

Credit: Machine Learning. Ng, A.

Multivariate Distributions

The multivariate normal distribution.



$$\Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

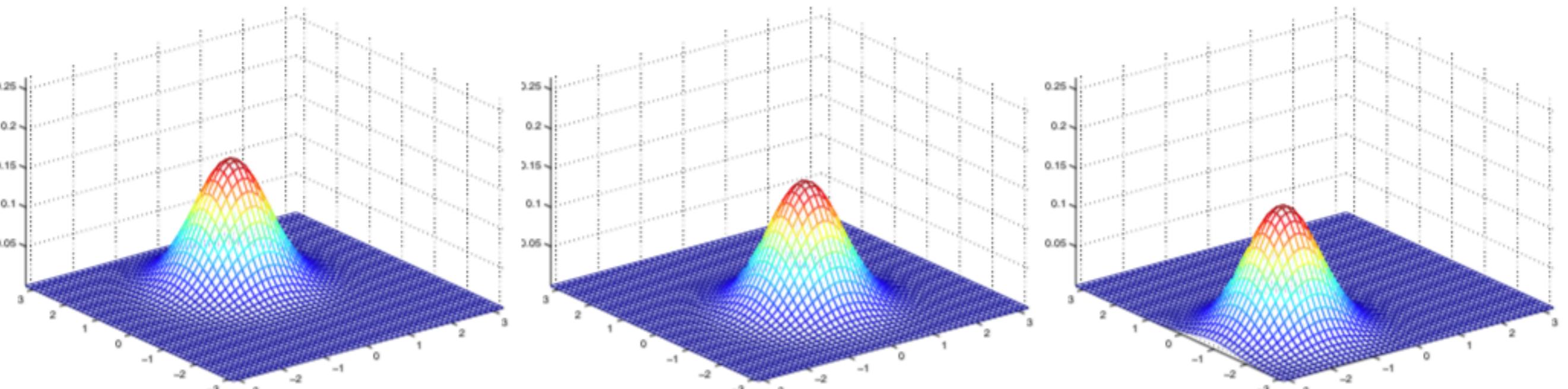
$$\Sigma = \begin{bmatrix} 3 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

$$\mu^T = [0., 0.]$$

Credit: Machine Learning. Ng, A.

Multivariate Distributions

The multivariate normal distribution.



$$\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} -1 \\ -1.5 \end{bmatrix}$$

$$\Sigma = I$$

Credit: Machine Learning. Ng, A.