# Estadística Aplicada y Cálculo Numérico (Grado en Química)

Valeri Makarov

Dept. de Matemática Aplicada, U.C.M.

10/02/2015 – 29/05/2015

F.CC. Matemáticas, Desp. 420
http://www.mat.ucm.es/∼vmakarov
e-mail: vmakarov@mat.ucm.es

## Capítulo 4

# Variables aleatorias y distribuciones de probabilidad

## Variables aleatorias

Una variable aleatoria viene de un experimento aleatorio.

Denotaremos con **mayúsculas las variables** (por ejemplo $X$) y con **minúsculas sus valores** concretos

$X$ puede tomar diferentes valores:

$$x \in (a, b)$$

**Sucesos**

- A: $X = x_1$
- B: $X \leq x_1$
- C: $x_1 < X \leq x_2$

Querremos calcular la **probabilidad**

# Función de distribución

Definamos una función:

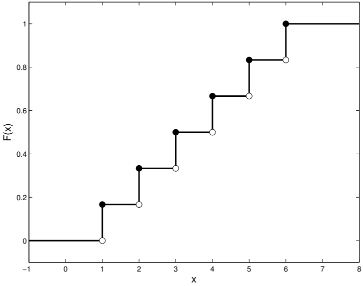
$$F(x) = P(X \leq x)$$

$F(x)$ se llama la función de distribución de $X$

$F(x)$ es la probabilidad de que la variable $X$ tome un valor menor o igual a $x$.

**Ejemplo:** Experimento con un dado
$F(x)$ puede tener saltos. Los puntos **siempre** "entran por la izquierda"

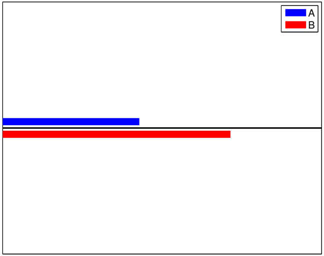
Función de distribución:



# Propiedades de $F(x)$

1. $0 \leq F(x) \leq 1$
2. $F(x)$ es creciente (es decir $F(x_1) \leq F(x_2)$ si $x_1 < x_2$)
3. $\lim_{x\to-\infty} F(x) = 0 \quad \lim_{x\to\infty} F(x) = 1$

¿Cómo calcular $P(x_1 < X \leq x_2)$?
Representamos sucesos:




$$A = (X \leq x_1) \quad B = (X \leq x_2)$$

$$C = (x_1 < X \leq x_2) = B - A$$

$$B - A = B \cap \overline{A}$$

$$P(B \cap \overline{A}) = P(B) + P(\overline{A}) - P(B \cup \overline{A}) = P(B) + 1 - P(A) - P(S)$$

$$P(x_1 < X \leq x_2) = F(x_2) + 1 - F(x_1) - 1 = F(x_2) - F(x_1)$$

# Variables aleatorias discretas

Una variable aleatoria discreta $X$ sólo puede tomar valores discretos (numerables):

$$X \in \{x_1, x_2, x_3, \dots\} \qquad x_1 < x_2 < x_3 \cdots$$

Función de probabilidad:

$$p_i = P(X = x_i) \qquad \sum_i p_i = 1$$

Podemos hacer un diagrama de barras análogo al diagrama de frecuencias relativas (recordemos: $p_i = \lim f_i$.)

$F(x)$ es una función escalonada

# Esperanza matemática

Sea $X \in \{x_1, \ldots, x_n\}$ una variable aleatoria discreta con las probabilidades $p_1, \ldots, p_n$

La **media aritmética o esperanza** matemática de $X$ es:

$$\mu_X = E[X] = \sum_{i=1}^{n} x_i p_i$$

Si $n = \infty$, entonces suponemos que la serie es convergente.

Comparar con la media muestral (de tamaño $m$):

$$\overline{x}(m) = \sum_{i=1}^{n} x_i f_i(m) \quad \Rightarrow \quad \mu_X = \lim_{m\to\infty} \overline{x}(m)$$

## Varianza de variable aleatoria

Si medimos $Y = aX + b$ tenemos:

$$\mu_Y = E[Y] = E[aX + b] = \sum_{i=1}^{n}(ax_i + b)p_i = aE[X] + b = a\mu_X + b$$

**Varianza** de $X$

$$\sigma^2 = V[X] = \sum_{i=1}^{n}(x_i - \mu_X)^2 p_i$$

Es decir

$$\sigma^2 = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$$

$$\sigma_Y^2 = a^2\sigma_X^2$$

# Variables aleatorias continuas

La función de distribución $F(x)$ de una variable continua aleatoria también es continua.

La probabilidad de encontrar $X$ entre $x_1$ y $x_2$:

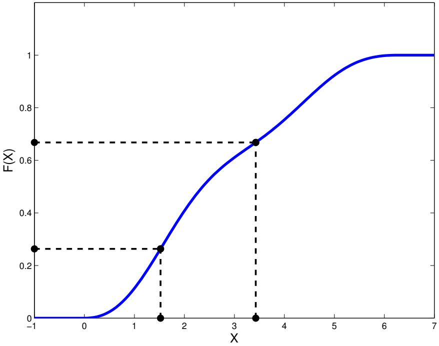
$$P(x_1 < X < x_2) = F(x_2) - F(x_1)$$

**Observación:** Si

$$x_2 \to x_1$$

Entonces

$$P = F(x_2) - F(x_1) \to 0$$



# Densidad de probabilidad

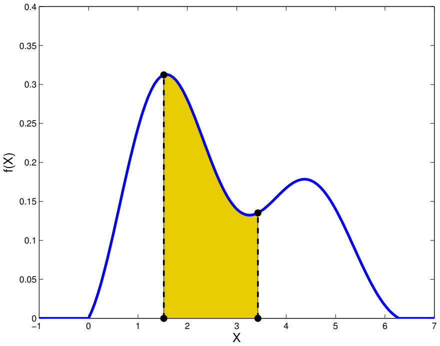
$$f(x) = \lim_{\epsilon \to 0} \frac{F(x+\epsilon) - F(x)}{\epsilon} = F'(x)$$

De las propiedades de $F(x)$:

1. $$f(x) \geq 0$$

2. $$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

3. $$\int_{-\infty}^{\infty} f(x)dx = 1$$



# Momentos

**Esperanza**

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

**Varianza**

$$\sigma^2 = V[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = E[X^2] - \mu^2$$

**Momentos centrales** de orden $k$

$$\gamma_k = \int_{-\infty}^{\infty} (x-\mu)^k f(x) dx$$

Sesgo: $\gamma_3/\sigma^3$; Curtosis: $\gamma_4/\sigma^4$

# Distribución binomial

Experimento aleatorio: observamos si ocurre o no un determinado suceso $A$.

*Ejemplo:* Lanzamiento de una moneda. $A$ - cara.

**Variable binomial:** $X$ (variable aleatoria discreta) - **número de éxistos** en $n$ pruebas independientes.

$$P(A) = p$$

¿Cómo calcular $P(X = k)$?

## Casos particulares

- $n = 1$
  Posibles valore de $X$: $\{0, 1\}$

$$P(X = 1) = P(A) = p \qquad P(X = 0) = P(\overline{A}) = 1 - p$$

- $n = 2$ Posibles valores de $X$: $\{0, 1, 2\}$

$$P(X = 0) = P(\overline{A} \cap \overline{A}) = P(\overline{A})P(\overline{A}) = (1 - p)^2$$

Aquí hemos usado la independencia de resultados de experimentos: $P(C|B) = P(C)$

$$P(X = 1) = P((A \cap \overline{A}) \cup (\overline{A} \cap A)) = 2P(A)P(\overline{A}) = 2p(1 - p)$$

$$P(X = 2) = P(A \cap A) = P(A)P(A) = p^2$$

# Distribución binomial

**Caso general:**

1. Se hace $n$ experimentos aleatorios.
2. $X =$ "número de éxitos". (entonces $n - X$ fracasos)
3. ¿Cuál es la probabilidad de que $X = k$?

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

donde $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ es el coeficiente binomial:

$$\begin{array}{c} 1 \\ 1\ 2\ 1 \\ 1\ 3\ 3\ 1 \\ 1\ 4\ 6\ 4\ 1 \end{array}$$

## Media y varianza de la variable binomial

1. Observación (para $\forall n$):

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = 1$$

2. Por definición:

$$\mu = E[X] = \sum_{k=0}^{n} kP(X=k) = \sum_{k=1}^{n} \frac{k\, n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$\mu = pn \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

Cambio: $k-1=i$, $n-1=m \quad \Rightarrow \quad n-k=m-i$

**Media:**

$$\mu = pn\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}p^i(1-p)^{m-i} = pn$$

**Varianza** (de forma análoga):

$$E[X^2] = n^2p^2 + np(1-p)$$

$$\sigma^2 = E[X^2] - \mu^2 = np(1-p)$$

Variables binomiales denotaremos:

$$X \sim Bin(n,p)$$

**Problema 3:** Calcula la función de probabilidad para $X =$ "número de caras al lanzar moneda 3 veces". Dos casos: $p = 0{,}5$ y $p = 0{,}6$.

**Solución:** Es una variable binomial con $n = 3$:

$$X \sim Bin(n, p) = Bin(3, p)$$

Posibles valores $X \in \{0, 1, 2, 3\}$. La función de probabilidad:

$$P(X = 0) = \frac{3!}{0!(3-0)!} p^0 (1-p)^{3-0} = (1-p)^3; \;\; P(X = 1) = 3p(1-p)^2$$

$$P(X = 2) = 3p^2(1-p); \quad P(X = 3) = p^3$$

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Moneda equilibrada: | 0,125 | 0,375 | 0,375 | 0,125 |
| Moneda cargada: | 0,064 | 0,288 | 0,432 | 0,216 |

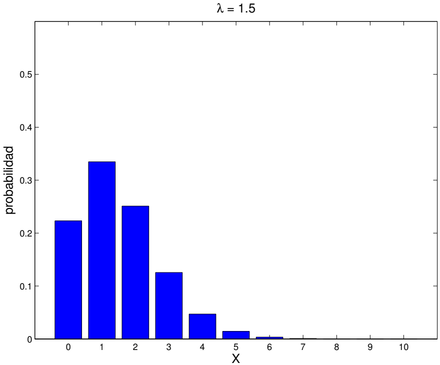
# Distribución de Poisson

Una variable aleatoria discreta:

$$X \in \{0, 1, 2, 3, \ldots\}$$

con la función de probabilidad

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \lambda > 0$$




Expresa la probabilidad de $k$ eventos en un tiempo fijo si estos eventos ocurren con una frecuencia media conocida y son independientes del tiempo discurrido desde el último evento.

Distribución de Poisson es el **caso límite de la distribución binomial**. Si $n$ es grande, entonces

$$X \sim Poi(\lambda) \approx B(n, \lambda/n)$$

**Comprobemos** que es una distribución

$$\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

**Esperanza**

$$E[X] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda$$

**Varianza**

$$E[X^2] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{k^2 \lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{k \lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{(m+1)\lambda^m}{m!}$$

$$= \lambda \left[ E[X] + e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \right] = \lambda^2 + \lambda$$

Finalmente:

$$\sigma^2 = V[X] = E[X^2] - E[X]^2 = \lambda$$

**Problema 10:** Una sustancia radiactiva emite, de media, una partícula $\alpha$ cada 1,94 s. Sea $X =$ "el número de partículas emitidas por segundo".

a) ¿Qué tipo de distribución de probabilidad sigue $X$?

$$P(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$$

b) La probabilidad de que se emita al menos una partícula en 1s

$$P(X \geq 1) = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - P(X = 0) = 1 - e^{-\lambda}$$

Ahora

$$E[X] = \lambda = 1/1{,}94 \approx 0{,}515 \quad \Rightarrow \quad P(X \geq 1) \approx 0{,}40$$

# Distribución normal

Se llama la distribución normal o distribución de Gauss o Gaussiana

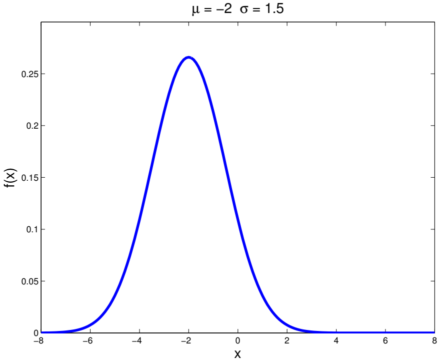
Una variable aleatoria continua:

$$X \in (-\infty, +\infty)$$

con la densidad de probabilidad

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \sigma > 0$$

La denotaremos: $X \sim N(\mu, \sigma)$



# Variable normal estándar

Supongamos que tenemos $X \sim N(\mu, \sigma)$. Definimos otra variable aleatoria:

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

A partir de las características de $Z$ podemos hallar las de $X$

$$P(X < x_1) = P\left(Z < \frac{x_1 - \mu}{\sigma}\right)$$

**Problema 13**: El diámetro de los tubos $d$ sigue una distribución normal con $\mu = 950$ mm y $\sigma = 10$ mm. ¿Cuál es la probabilidad de que un tubo escogido al azar tenga $d \geq 960$ mm?

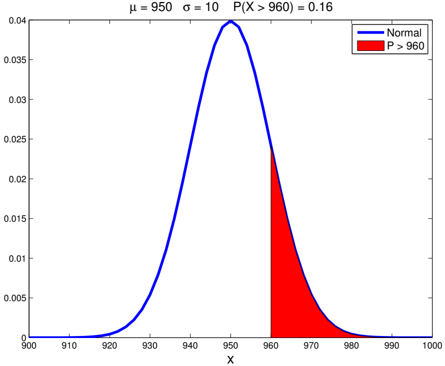
$$d \sim N(950, 10) \quad P(d \geq 960) = ?$$

**Resolución:**

$$P = \frac{1}{\sqrt{2\pi}\sigma} \int_{960}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

o usando la normalizada:
$Z = \frac{X-950}{10}$

$$P = \frac{1}{\sqrt{2\pi}} \int_{1}^{\infty} e^{-\frac{z^2}{2}} dz \approx 0{,}16$$

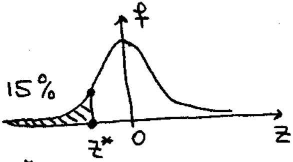


<u>Problema</u>: ¿Cuál es el DIÁMETRO MÁXIMO DE LOS 15% DE LOS TUBOS MÁS pequeños?

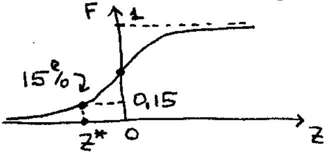
Variable NORMALIZADA: $Z = \frac{X-\mu_x}{\sigma_x} = \frac{X-950}{10}$

$z^*$ – DIÁMETRO MÁXIMO

Fun. DENSIDAD




FUN. DISTRIBUCIÓN



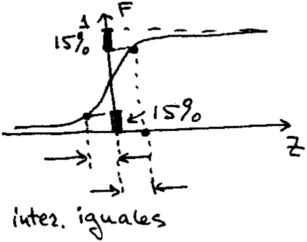

$$\int_{-\infty}^{z^*} f(z) = 0{,}15 \iff F(z^*) = 0{,}15$$

EN LA TABLA $F(z^*) = \underline{\underline{0.15}}$ ← NO APARECE!

$$\int_{-\infty}^{z^*} f(z) = 0{,}15 \quad \Longleftrightarrow \quad F(z^*) = 0{,}15$$

EN LA TABLA $F(z^*) = \underline{\underline{0.15}}$ ← NO APARECE!

F es <u>SIMÉTRICA</u>




$$F(\bar{z}) = 1 - 0{,}15 = 0{,}85$$

DE La Tabla

$$\bar{z} \simeq 1.03 \Rightarrow z^* = -1.03$$

$$\frac{x^* - 950}{10} = -1{,}03$$

$$\boxed{x^* = 950 - 10.3 = 939.7 \text{ mm}}$$

# Aproximación de variables binomiales

Supongamos que $X$ es una variable aleatoria binomial con $n \gg 1$:

$X \sim Bin(n, p)$ puede ser aproximada por $Y \sim N(np; \sqrt{np(1-p)})$

Variable binomial:

$$X \sim Bin(n,p)$$

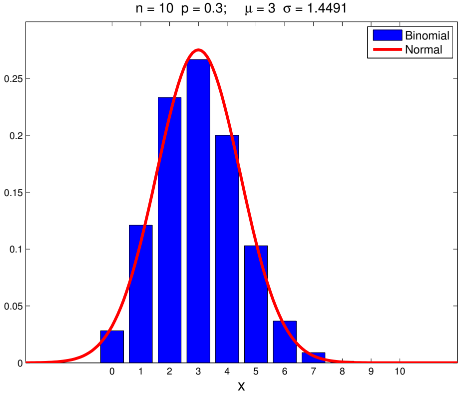
Fijamos $n = 10$ $p = 0{,}3$

$$\mu_X = np = 3$$

$$\sigma_X = \sqrt{np(1-p)} \approx 1{,}4491$$

Entonces $X$ aproximamos por

$$Y \sim N(3; 1{,}4491)$$




Cada valor $k$ de $X$ es la marca de clase de $Y$ de clase $(k-1/2, k+1/2]$

# Variables aleatorias múltiples

## Variables bidimensionales discretas

En un experimento aleatorio observamos **varias (dos) variables**:

$$X \quad \text{e} \quad Y$$

Las dos variables consideradas **juntas**: $(X, Y)$ forman una variable **bidimensional**.

**Variable discreta:**

$$X \in \{x_1, x_2, x_3, \ldots\} \qquad Y \in \{y_1, y_2, y_3, \ldots\}$$

**Probabilidad del par $(x_i, y_j)$:**

$$p_{ij} = P((X = x_i) \cap (Y = y_j))$$

## Suma de dos variables

**Definamos**:

$$S = X + Y$$

Ahora la probabilidad:

$$P(S = s_k) = \sum_{ij} P((X = x_i) \cap (Y = y_j) \text{ tal que } x_i + y_j = s_k)$$

**Esperanza** ($E[\cdot]$ es lineal)

$$\mu_S = E[S] = E[X + Y] = E[X] + E[Y] = \mu_X + \mu_Y$$

**Varianza**

$$\sigma_S^2 = E[(S - \mu_S)^2] = E[(X - \mu_X + Y - \mu_Y)^2] = E[(X - \mu_X)^2] +$$

$$E[(Y - \mu_Y)^2] + 2E[(X - \mu_X)(Y - \mu_Y)] = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

## Variables discretas independientes

**Variable suma**:

$$S = X + Y$$

**Covarianza:** (variables independientes)

$$\sigma_{XY} = 0$$

Por lo tanto:

**Media:**

$$\mu_S = \mu_X + \mu_Y$$

**Varianza:**

$$\sigma_S^2 = \sigma_X^2 + \sigma_Y^2$$

## Teorema central del límite

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con las medias $\{\mu_k\}_{k=1}^n$ y varianzas $\{\sigma_k^2\}_{k=1}^n$

Entonces la variable aleatoria

$$Z = \sum_{k=1}^{n} X_k$$

tiende a tener la distribución normal $(n \to \infty)$ $Z \sim N(\mu_Z, \sigma_Z)$ con

$$\mu_Z = \sum_{k=1}^{n} \mu_k \qquad \sigma_Z^2 = \sum_{k=1}^{n} \sigma_k^2$$

**Ejemplo (teorema central):** Antes ya hemos visto que:

$$Bin(n, p) \approx N(np, \sqrt{np(1-p)})$$

**Fundamento:** $X \sim Bin(n, p)$ es la suma de $n$ variables $Y \sim Bin(1, p)$

$$\mu_Y = 0*(1-p)+1*p = p$$

$$\sigma_Y^2 = p(1-p)$$

Del teorema:

$$\mu_X = np$$

$$\sigma_X^2 = np(1-p)$$