Distribuciones Probabilísticas

Curso de Estadística

TAE,2005

J.J. Gómez Cadenas

Distribución Binomial

Considerar **N** observaciones independientes tales que:

El resultado de cada experimento es "acierto" o "fallo"

La probabilidad de acierto de un experimento dado es **p**

El conjunto de observaciones **N** puede considerarse como una única medida que puede caracterizarse por la variable aleatoria **n** :

n = número de aciertos ($0 \leq n \leq N$)

El espacio de muestras S se define como el conjunto de posibles valores de **n** para **N** observaciones.

Si repetimos el experimento muchas veces con **N** observaciones cada vez, los valores resultantes de **n** ocurren con frecuencias relativas determinadas por la distribución binomial.

Derivación de la forma de la distribución binomial

Probabilidad de acierto para una observación dada →**p**

Probabilidad de fallo →1-**p**

Las observaciones son independientes →Probabilidad de acierto y/o fallo de una secuencia de observaciones (en un orden dado) es el producto de las observaciones individuales.

Ejemplo: *aafaf* $\rightarrow P=pp(1-p)p(1-p)=p^3(1-p)^2$

En general P, para una secuencia de n aciertos y (N-n) fallos $\rightarrow p^n(1-p)^{N-n}$

Puesto que el orden de acierto/fallo es irrelevante (estamos interesados sólo en el número total de aciertos **n**) calculamos el número de secuencias (permutaciones) con **n** éxitos en **N** observaciones:

$$\frac{N!}{n!(N-n)!}$$

La distribución binomial, es, entonces:

$$f(n;N,p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

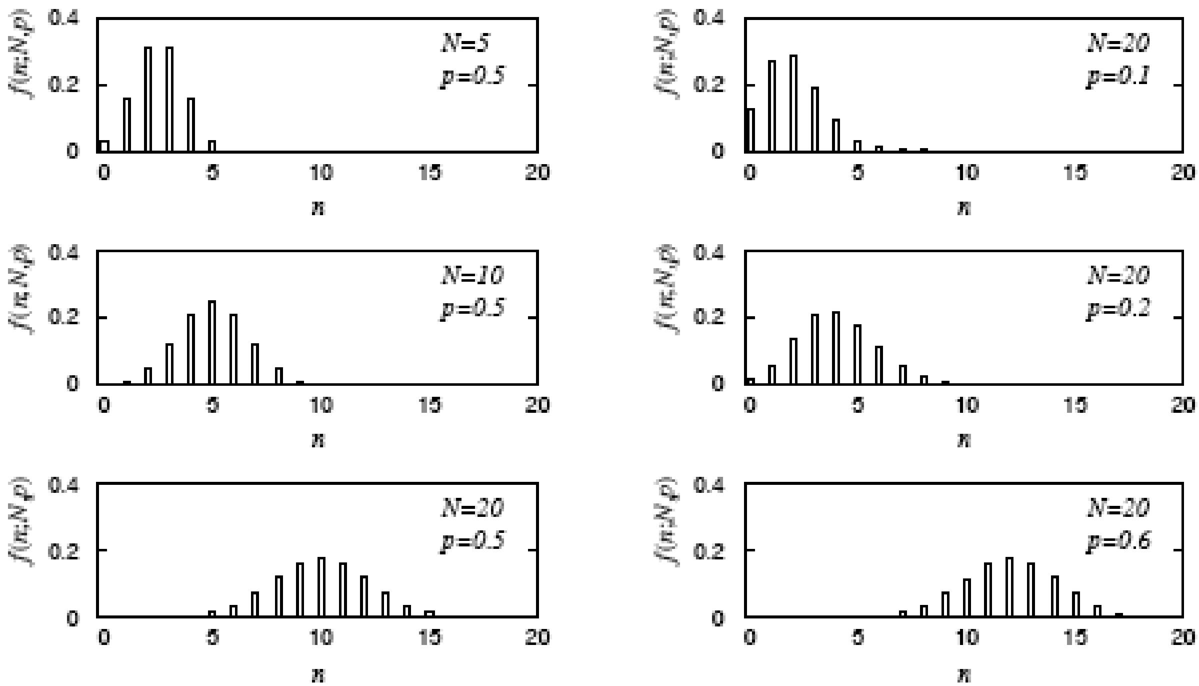
Donde, la notación *f(n; N,p)* indica que **n** es una variable aleatoria, mientras que **N** y **p** son parámetros.

Valor esperado y varianza (cálculo no trivial):

$$E[n] = \sum_{n=0}^{\infty} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = Np$$

$$V[n] = E[n^2] - (E[n])^2 = Np(1-p)$$

NB: *E[n]* y *V[n]* no son variables aleatorias sino constantes que dependen de los valores verdaderos (y posiblemente desconocidos) de **p** y **N**.



Ejemplo: Observamos N desintegraciones del τ.

El número **n** de estas observaciones correspondientes a un determinado canal (e.g, $\pi\nu$) sigue la distribución binomial, con **p** igual a la relación de semidesintegración del canal

Distribución Multinomial

Similar a la binomial, pero en lugar de 2 posibles resultados (acierto o fallo) hay **m** posibles resultados.

$$\vec{p} = (p_1, \ldots, p_m) \text{ tal que } \sum_{i=1}^{m} p_i = 1$$

Para N observaciones queremos calcular la probabilidad de observar:

n_1 con resultado 1

n_2 con resultado 2

…

n_m con resultado m

Ejemplo de multinomial: $n = (n_1, \ldots, n_m)$ representa un histograma con m bins y N entradas en total

Esta probabilidad sigue la distribución multinomial

$$f(\vec{n}; N, \vec{p}) = \frac{N!}{n_1! n_2! \ldots n_m!} p_1^{n_1} p_2^{n_2} \ldots p_m^{n_m}$$

Los n_i individuales se distribuyen binomialmente con parámetros N, p_i

Distribución de Poisson

Considerar la distribución binomial en el límite:

$N \rightarrow \infty$

$p \rightarrow 0$

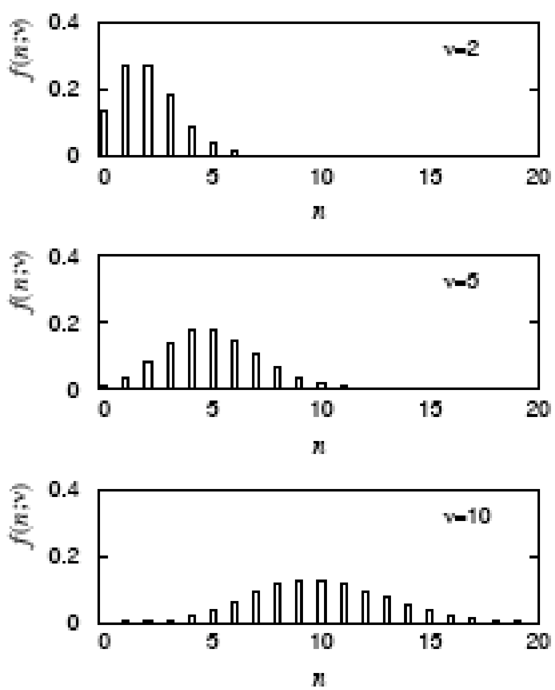
$E[n]=Np \rightarrow \nu$

Puede demostrarse que en este caso **n** sigue la distribución de Poisson

$$f(n;\nu) = \frac{\nu^n}{n!} e^{-\nu} \qquad (0 \leq n \leq \infty)$$

$$E[n] = \sum_{n=0}^{\infty} n \frac{\nu^n}{n!} e^{-\nu} = \nu$$

$$V[n] = \sum_{n=0}^{\infty} (n-\nu)^2 \frac{\nu^n}{n!} e^{-\nu} = \nu$$



Para ν grande Poisson tiende a Gauss

Ejemplos de Poisson: Número de desintegraciones de una cierta cantidad de material radioactivo en un tiempo fijo t, en el límite:

Número de desintegraciones posibles (e.g, número total de átomos) es muy grande

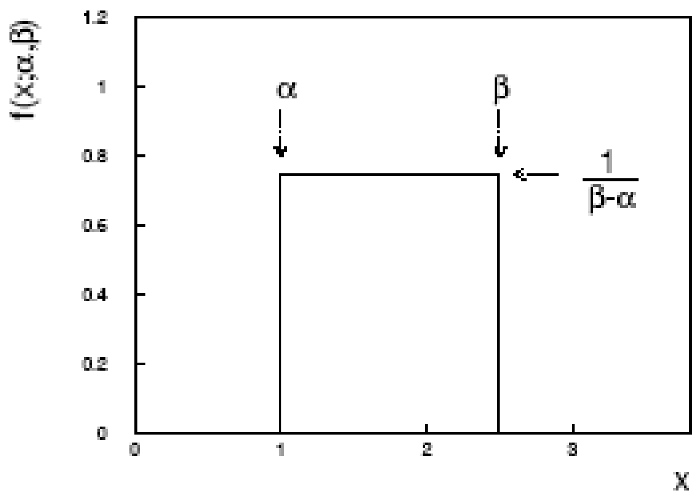
Probabilidad de una desintegración individual en t es muy pequeña.

Distribución uniforme

Considerar una variable continua aleatoria definida en todo R. La distribución uniforme es:

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta-\alpha} & \alpha\le x\le\beta\\ 0 & \text{en otro caso}\end{cases}$$

Es decir, podemos encontrar x con igual probabilidad entre α y β:



$$E[x]=\int_0^\beta \frac{x}{\beta-\alpha}dx=\frac{1}{2}(\alpha+\beta)$$

$$V[x]=\int_0^\beta [x-\frac{1}{2}(\alpha+\beta)]^2\frac{1}{\beta-\alpha}dx=\frac{1}{12}(\beta-\alpha)^2$$

Una propiedad importante de la distribución uniforme es la de que cualquier variable aleatoria continua *x* con pdf *f(x)* y distribución acumulativa *F(x)* puede transformarse en una nueva variable *y* distribuida uniforme entre 0 y 1 mediante el cambio de variable:

$$y = F(x)$$

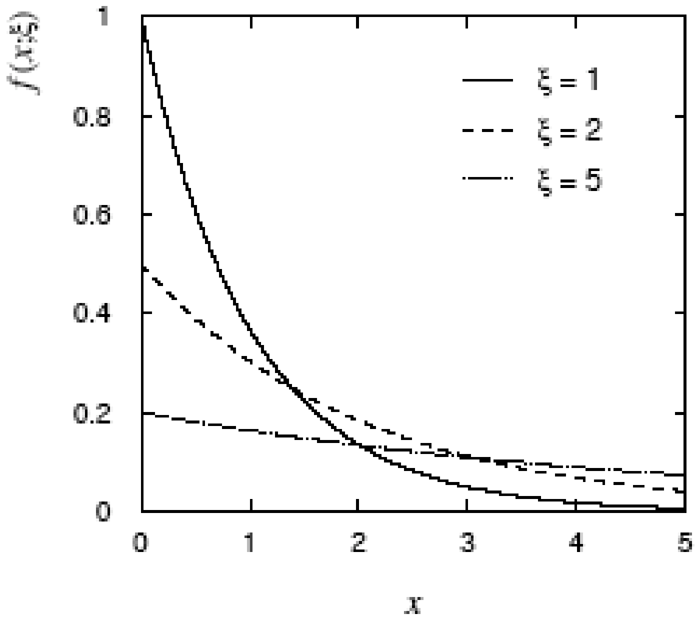
$$\frac{dy}{dx} = \frac{d}{dx}\int_{-\infty}^{x} f(x')dx' = f(x)$$

La pdf de y es:

$$g(y) = f(x)\left|\frac{dx}{dy}\right| = \left|\frac{dy}{dx}\right|\left|\frac{dx}{dy}\right| = 1 \quad 0 \le y \le 1$$

Usaremos esta propiedad de la distribución uniforme en el tema relacionado con técnicas de Monte Carlo.

Distribución exponencial



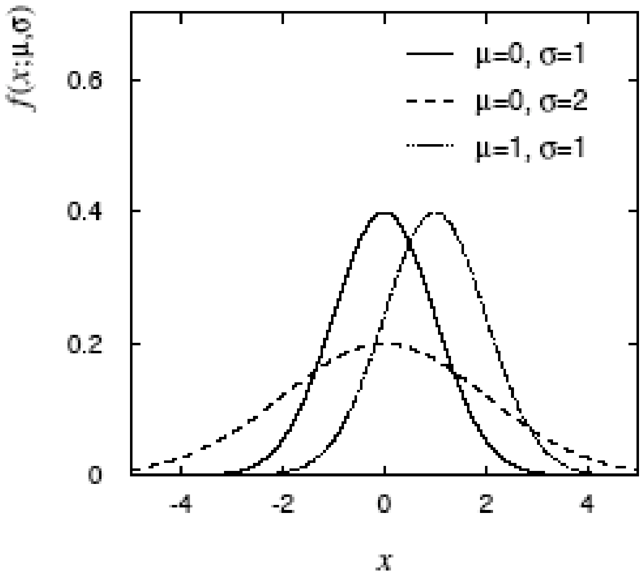
$$f(x;\xi) = \frac{1}{\xi} e^{-\frac{x}{\xi}}, \quad 0 \le x \le \infty$$

$$E[x] = \frac{1}{\xi} \int_0^\infty x e^{-\frac{x}{\xi}} dx = \xi$$

$$V[x] = \frac{1}{\xi} \int_0^\infty (x-\xi)^2 e^{-\frac{x}{\xi}} dx = \xi^2$$

Ejemplo: El tiempo de desintegración de una partícula inestable (medido en su sistema de referencia) sigue una distribución exponencial

Distribución de Gauss



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

$$E[x] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

$$V[x] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2$$

NB: A menudo μ y σ^2 se utlizan para denotar la media y la varianza de cualquier pdf, no sólo la pdf gausiana.

Distribución de Gauss estándar:

$$\mu = 0, \sigma = 1 \rightarrow \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

$$\Phi(x) = \int_{-\infty}^{x} \varphi(x')dx'$$

Si y es Gausiana con μ y σ^2 entonces $x=(y-\mu)/\sigma$ sigue $\varphi(x)$.

Teorema del límite central

Dadas n variables aleatorias independientes x_i, con:

medias μ_i , varianzas σ_i^2

pdf arbitraria

Su suma, en el límite de n grande es una variable aleatoria distribuida gausianamente

$$y_i = \sum_{i=1}^{n} x_i, \quad n \to \infty : y_i \text{ está distribuida gausianamente}$$

$$\mathrm{E}[y_i] = \sum_{i=1}^{n} \mu_i, \quad V[y_i] = \sum_{i=1}^{n} \sigma^2{}_i$$

El teorema del límite central supone la justificación formal para tratar los errores como variables aleatorias distribuidas gausianamente. Es aplicable siempre que el error total sea la suma de muchas contribuciones pequeñas.

Distribución de Gauss Multivariada

$$f(\vec{x};\vec{\mu},V)=\frac{1}{(2\pi)^{N/2}|V|^{1/2}}\exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T V^{-1}(\vec{x}-\vec{\mu})\right]$$

$$E[x_i]=\mu_i$$

$$\mathrm{cov}[x_i,x_j]=V_{ij}$$

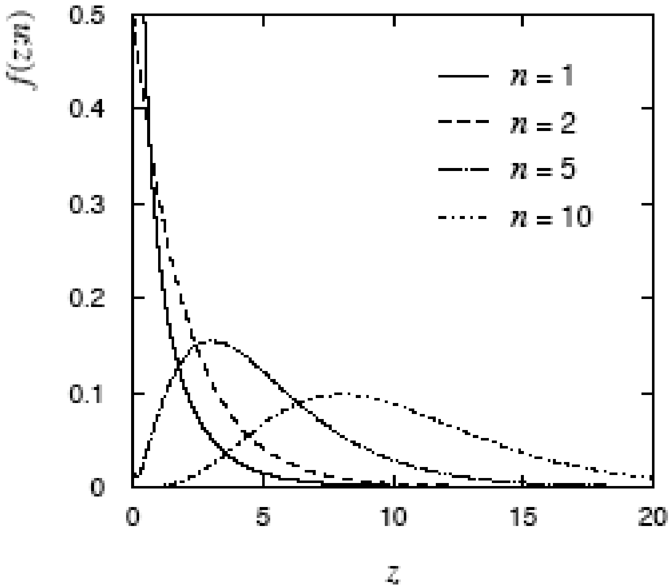
$$n=2$$

$$f(x_1,x_2;\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho}}$$

$$\times\exp\left\{-\frac{1}{2(1-\rho^2)}\left[(\frac{x_1-\mu_1}{\sigma_1})^2+(\frac{x_2-\mu_2}{\sigma_2})^2-2\rho(\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2})\right]\right\},$$

$\rho=\frac{\mathrm{cov}[x_1,x_2]}{(\sigma_1\sigma_2)}$ coeficiente de correlación

Distribución χ^2



$$f(z;n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2}, \quad n = 1, 2.,,,$$

$n \rightarrow$ número de grados de libertad

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \rightarrow \text{ función Gamma}$$

$$E[z] = \int_0^\infty z \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} dz = n$$

$$V[z] = \int_0^\infty (z-n)^2 \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} dz = 2n$$

Dadas N variables aleatorias independientes x_i, distribuidas gausianamente, con media μ_i y varianza σ_i^2, la variable:

$$z = \sum_{i=1}^{N} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

Sigue la distribución χ^2 para N grados de libertad.
Como veremos, estas variables son las que utilizamos para estimar la calidad de un ajuste, particularmente con el método de mínimos cuadrados