Variables aleatorias continuas

Hemos definido que una variable aleatoria X es discreta si I_X es un conjunto finito o infinito numerable.
En la práctica las variables aleatorias discretas sirven como modelos para hacer inferencias estadísticas cuando los datos que se recogen son discretos ("número de algo").
Pero cuando los datos que se registran son continuos (por ejemplo estatura o glucemia de una persona, porcentaje de hierro en un mineral, tiempo desde que un enfermo comienza un tratamiento hasta que se observa una mejoría o un empeoramiento, o tiempo de duración de una lamparita), se usan otro tipo de variables aleatorias como modelo probabilístico: las variables aleatorias continuas.

Definición de variable aleatoria continua:

Una variable aleatoria X es (absolutamente) continua si existe una función $f:R\to R^+$ tal que

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Nota: la función f se llama función de densidad de la v.a. X

Comentarios:

1) **Si X es v.a. continua $\Rightarrow$ F es continua**

2) Si **F es continua** $\Rightarrow$ **$P(X=x_0)=0$** para todo $x_0 \in R$

3) **Si X es v.a. continua $\Rightarrow$**

$$P(a \le X \le b) = F(b) - F(a) = \int_a^b f(x)\,dx$$

para todo $a\le b$

Propiedades de las funciones de densidad

Son funciones $f:R\to R$ que cuplen dos propiedades:

a) $f(x)\ge 0$ para todo $x\in R$

b) $\int_{-\infty}^{+\infty} f(x)dx = 1$

<u>Interpretación intuitiva de la función de densidad</u>:

1) Ver la relación entre la función de densidad y el histograma.

2) Significado intuitivo del valor de la función de ensidad en un punto:

$$\mathbf{P(x_0 - h \leq X \leq x_0 + h) \cong f(x_0) 2h}$$

Comentario: ¿Que significa intuitivamente que $f(x_1)<f(x_2)$?

<u>Relación entre la función de densidad y la función de distribución</u>.

- Conocida la función de densidad de una v.a. continua, ¿se puede calcular su función de distribución?

- Recíprocamente si se conoce la función de distribución de una v.a. continua puede calcularse su función de densidad derivando:

$$\mathbf{f(x) = F'(x) \quad \text{para todo } x \in R \text{ donde F es derivable.}}$$

Ejemplos de funciones de densidad:

Ejemplo 1: $f(x) = ax\, I_{[0,1]}(x)$
a) Cuanto vale la constante a?
b) Calcular $P(X>0.5)$

Ejemplo 2: $f(x) = e^{-x}\, I_{[0,+\infty]}(x)$
Calcular la función de distribución.

¿Existen variables que no son discretas ni continuas?

<u>Esperanza</u>. Hemos definido esperanza para una v.a. X discreta: $E(X) = \sum_{x \in I_X} x\, p(x)$

donde p(x) es la función de probabilidad puntual de X.

Análogamente se define esperanza para variables continuas:

<u>Definición de esperanza para una v.a. continua</u>: Sea X una v.a. continua, f su función de densidad. Se define

$$\mathbf{E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx}$$

Nota: la esperanza está definida cuando $\int_{-\infty}^{+\infty} |x|\, f(x)\, dx < \infty$

Ejemplo: Calcular E(X) en el ejemplo 2.
Ejercicio: Calcular E(X) en el ejemplo 1.

Esperanza de una función de una variable aleatoria

Teorema: Sea X una v.a., g una función $g:R \to R$.

a) Si X es discreta con f.p.p. p:

$$\mathbf{E(g(X)) = \sum_{x \in I_X} g(x)\, p(x)}$$

b) Si X es continua con función de densidad f:

$$\mathbf{E(g(X)) = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx}$$

Propiedades de la esperanza. Dijimos que:

a) **E(a X + b) = a E(X)+ b**
b) **E(X + Y) = E(X) + E(Y)**

y que estas propiedades valían para cualquier variable aleatoria (discretas, continuas, cualquiera).

Ejercicio: Hemos demostrado la propiedad a) para el caso de variables aleatorias discretas, demostrarlo también para el caso de variables aleatorias continuas (sugerencia: usar el teorema anterior sobre E(g(X)).

Hemos dicho que la siguiente es la definición de varianza para cualquier variable aleatoria:

Definición de varianza

Sea X una v.a. se define

$$\mathbf{Var(X) = E(X-E(X))^2}$$

Cálculo de la varianza.

De la definición general de Var(X) y del teorema que permite calcular E(g(X)) regulta lo siguiente:
a) Para una v.a. **discreta.** Si X es discreta con f.p.p. p(x) es

$$\mathbf{Var(X) = \sum_{x \in I_X} (x - E(X))^2\, p(x)}$$

b) Para una v.a. **continua.** Si X es continua con función de densidad f:

$$\mathbf{Var(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 \; f(x)\, dx}$$

También vimos la siguiente proposición:

<u>Proposición</u>: $\mathbf{Var(X) = E(X^2) - (E(X))^2}$

- Esta proposición ya la demostramos. La demostración que hemos hecho vale para cualquier v.a. (no solamente para el caso discreto).

- En general es más fácil calcular la varianza de una v.a. discreta o continua usando esta proposición que usando la definición de varianza.

Para el caso continuo, usando esta proposición, como se calcula la Varianza?

a) calculando $E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$

b) calculando $E(X^2) = \int_{-\infty}^{+\infty} x^2\, f(x)\, dx$

y luego restando $E(X^2) - (E(X))^2$

Ejemplo: Calcular Var(X) en el ejemplo 2.

Percentiles y cuantiles de una variable aleatoria continua:

En estadística descriptiva dijimos:

"¿Qué es el percentil 10? El número que deja aproximadamente el 10% de los datos “a izquierda”. O lo que es lo mismo el 10% de los datos son menores que el Percentil 10. Se lo denota P_{10}."

Si X es una variable aleatoria continua, su percentil 10 es el valor que deja un área de 0.10 "a izquierda" en la función de densidad de la variable.

En general se llama cuantil α al valor x_α tal que

$$P(X \le x_\alpha) = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha$$

Comentario: la notación x_α no es uniformemente usada en todos los libros.

Algunas distribuciones continuas usadas como modelos: distribuciones uniforme, normal, exponencial y otras

Distribución uniforme.

(no es muy usada como modelo, pero la presento primero porque es la más simple).
Supongamos que se elige al azar un número real en el intervalo [0,1] ¿Como se puede realizar el experimento?

Sea X= número elegido

Cuánto vale $P(X=0.4)$?
Cuánto vale $P(0 \leq X \leq 1/4)$? $P(1/4 \leq X \leq 1/2)$?

¿Cuál es la función de densidad de X?

Distribución uniforme en el intervalo [a,b]

Se dice que $X \sim U[a,b]$ si su función de densidad es

$$\mathbf{f(x) = \frac{1}{b-a} I_{[a,b]}(x)}$$

- Si $X \sim U[a,b]$, cuánto vale $E(X)$?
- Demostrar que $Var(X) = (b-a)^2/12$

Antes de definir distribución normal o de Gauss voy a hacer dos comentarios.

Comentario 1: Sea X una v.a. con cualquier distribución, con esperanza y varianza finitas. Llamemos μ a su esperanza y σ^2 a su varianza. ¿Se puede definir una nueva v.a. Y de la forma $Y=aX+b$ (a, b constantes) de modo que $E(Y)=0$, $Var(Y)=1$? Esta nueva variable así obtenida se suele llamar "variable estandarizada".

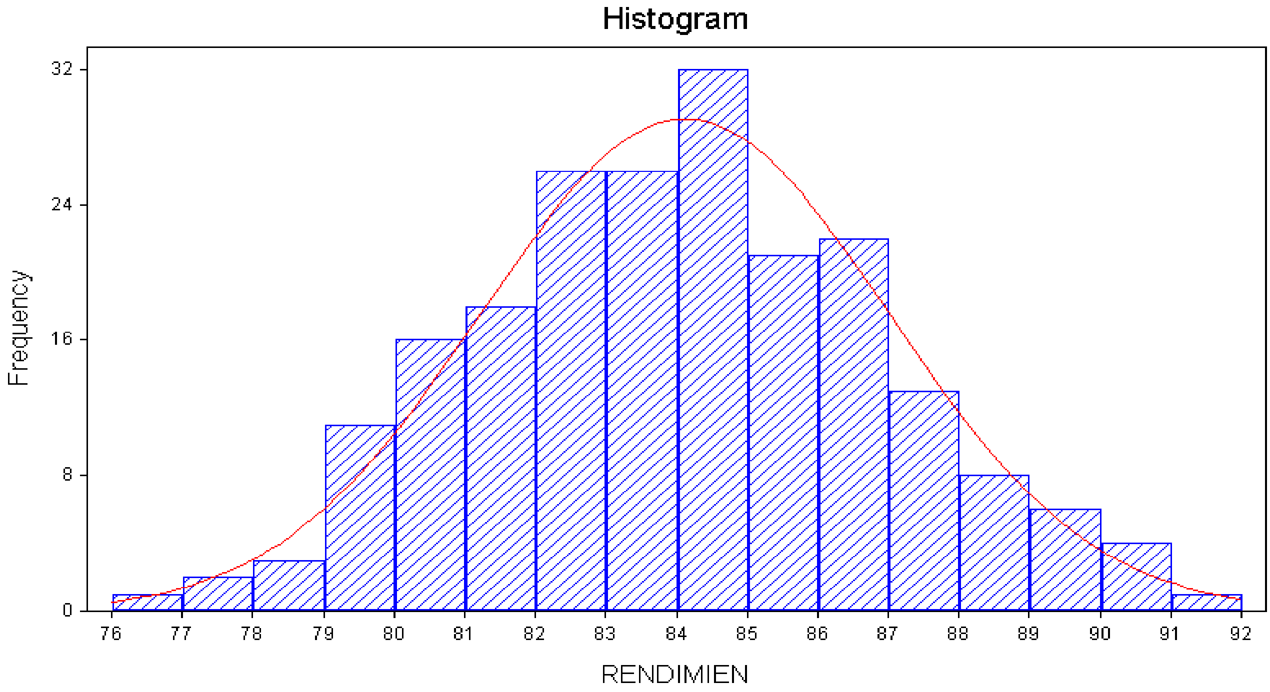
Comentario 2:

$$\int e^{ax} dx = ? \;;\quad \int e^{x^2} dx = ? \;;\quad \int e^{ax^2} dx = ? \;;\quad \int x e^{ax^2} dx = ?$$

Se puede demostrar que

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

Comentario 3: En la clase de estadística descriptiva mostramos el histograma de los datos de rendimiento (%) en la producción de 210 lotes consecutivos de un producto químico. Si marcamos en el Statistix "Display Normal Curve", obtenemos una curva (curva normal o de Gauss) superpuesta al histograma:



Cuando, como en el caso del rendimiento, una variable tiene un histograma parecido a una curva "normal", diremos que la variable tiene distribución aproximadamente normal o aproximadamente gaussiana. Como se definen las curvas normales o de Gauss?

Distribución Normal o de Gauss

Daremos primero un caso particular: la distribución normal standard.

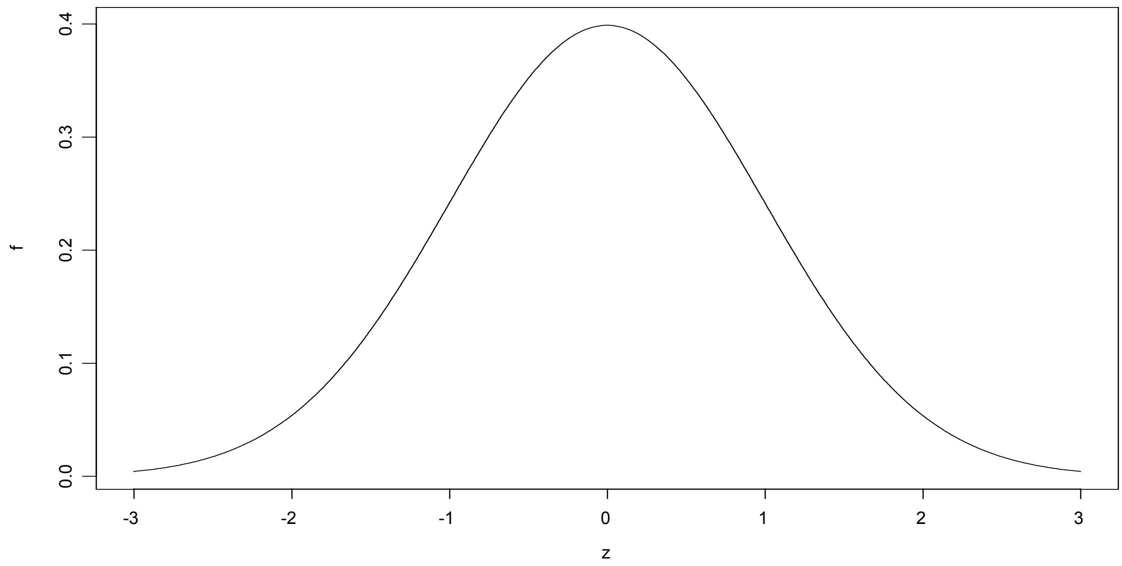
Definición:

Se dice que una v.a. Z tiene distribución Normal standard si su función de densidad es:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}$$

Es una función de densidad?

El gráfico de esta función tiene la siguiente forma:



Se puede demostrar que

a) f(z) es una función par.
b) Es creciente para z<0 y decreciente para z>0.
c) Tiene puntos de inflexión en –1 y 1
d) $\lim_{x\to+\infty} f(z) = 0 \qquad \lim_{x\to-\infty} f(z) = 0$

También se puede demostrar que:

Si Z tiene distribución normal standard entonces:

E(Z)=0 y Var(Z)=1

Por eso la normal standard se la llama también Normal(0,1) (se abrevia N(0,1))

Llamemos **Φ a la función de distribución acumulada de la N(0,1)**:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Los programas estadísticos calculan esta función (también hay tablas), y usándola se puede calcular cualquier P(a≤Z≤b):

si Z ~ N(0,1) entonces P(a≤Z≤b) = Φ(b) - Φ(a)

Usando la tabla o el programa, se puede verificar fácilmente que si Z ~ N(0,1)

P(-1 ≤ Z ≤ 1) = área bajo la curva normal standard entre -1 y 1 = 0.683 (≅ 0.68 o 68%)
P(-2 ≤ Z ≤ 2) = área entre -2 y 2 = 0.955 (≅ 95%)
P(-1.96 ≤ Z ≤ 1.96) = área entre -1.96 y 1.96 = 0.95 (= 95%)
P(-3 ≤ Z ≤ 3) = área entre -3 y 3 = 0.997

La curva normal graficada por el Statistix superpuesta al histograma de los 210 rendimientos del producto químico no es la normal standard.
Puede observarse que esta curva normal toma su valor máximo no en el cero sino en la media de los rendimientos de los 210 lotes observados (que es 84.1).
Hay muchas curvas normales: una para cada media y cada DS. La normal standard tiene media = 0 y DS = 1

Definición:

Se dice que $X \sim N(\mu,\sigma^2)$ si su función de densidad es:

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

Qué forma tiene el gráfico de esta función? Donde toma su valor máximo?

La familia de distribuciones $N(\mu,\sigma^2)$ se ha usado como modelo aproximado para la distribución de muchas variables, como por ejemplo:

a) algunas variables biológicas como la estatura,
b) errores de medición,
c) promedios muestrales (esto lo veremos luego)

Se puede demostrar la siguiente proposición:

Proposición:

$$\text{Si } X \sim N(\mu,\sigma^2) \text{ y } Z = \frac{X-\mu}{\sigma} \quad \Rightarrow \quad Z \sim N(0,1)$$

Como consecuencia de esta proposición:

1) No es necesario calcular integrales para demostrar que

$$X \sim N(\mu,\sigma^2) \quad \Rightarrow \quad E(X)=\mu \ , \quad Var(X)=\sigma^2$$

2) Si $X \sim N(\mu,\sigma^2)$, como se calcula $P(a \le X \le b)$?

Ejemplo: Sea X la estatura (en cm.) de un hombre elegido al azar en cierta ciudad. Se supone que X ~ N(172,64).
Calcular $P(170 \le X \le 180)$
Rta.: 0.44

Comentario: Usando la tabla de la función de distribución N(0,1) vimos que:

a) Si $Z \sim N(0,1)$ entonces:
$P(-1 \leq Z \leq 1) = 0.683 \cong 0.68$
$P(-2 \leq Z \leq 2) = 0.954 \cong 0.95$
$P(-3 \leq Z \leq 3) = 0.997$

b) Escribir expresiones similares a las de a) para una v.a. $X \sim N(\mu,\sigma^2)$

<u>Distribución exponencial</u> (definición): Se dice que una variable aleatoria continua X tiene distribución exponencial con parámetro λ (notación: $X \sim \exp(\lambda)$), si su función de densidad es de la forma:

$$f(x) = \lambda\, e^{-\lambda x}\; I_{[0,+\infty]}(x)$$

- Es evidente que el parámetro λ tiene que ser >0 para que sea una función de densidad, verdad?
- Verificar que f(x) cumple las propiedades de una función de densidad.
- La distribución exponencial se usa: a) como modelo aproximado para la duración de piezas que no tienen desgaste; b) para el tiempo que transcurre entre un evento y el siguiente en un proceso de Poisson.
- Demostrar que si $X \sim \exp(\lambda)$ entonces $F(x) = 1 - e^{-\lambda x}$ para $x \geq 0$.
- Demostrar que si $X \sim \exp(\lambda)$, entonces $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$

<u>Distribución Gamma, Weibull y otras distribuciones</u>

Distribución Gamma (definición): Se dice que $X \sim \Gamma(\alpha,\lambda)$ si su función de densidad es de la forma:

$$f(x) = cte.\, x^{\alpha-1}\, e^{-\lambda x}\; I_{[0,+\infty)}(x)$$

donde $\alpha>0$, $\lambda>0$ y la cte multiplicativa es tal que $\int f(x) = 1$
Se puede demostrar (no lo haremos) que $E(X) = \alpha/\lambda$ y $Var(X) = \alpha/\lambda^2$

Distribución de Weibull (definición): Se dice que $X \sim Weibull(\alpha,\lambda)$ si su función de densidad es de la forma:

$$f(x) = \alpha\lambda\, x^{\alpha-1} e^{-\lambda x^{\alpha}}\; I_{[0,+\infty]}(x)$$

donde $\alpha>0$ y $\lambda>0$

Tanto la familia de distribuciones Gamma como la de Weibull se han usado como modelo para tiempos de duración en estudios de confiabilidad en la industria. Es fácil ver (hacerlo) que ambas familias incluyen a la familia de distribuciones exponencial.

Hay muchas otras familias de distribuciones que han sido propuestas como modelo para distintos problemas prácticos.

Pero (**por suerte!**) para poder aplicar algunos de los métodos de inferencia estadística que estudiaremos no se requiere conocer previamente la distribución de los datos.