Johann Carl Friedrich Gauss (1777-1855, Alemania)

$$\boxed{1 + 2 + \cdots + 50 + 51 + \cdots + 99 + 100 = 101 \times 50 = 5050.}$$

Clase 8 de Bioestadística

Variables aleatorias exponenciales y normales

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Contenidos de la clase

Introducción

- Nos interesa modelar un experimento cuyo resultado es un número real.
- Identificamos el conjunto de valores I que puede tomar la variable, que pude ser
  - Todo $I = \mathbb{R} = \{x \colon -\infty < x < \infty\}$
  - Los reales no-negativos $I = \mathbb{R}^{+} = [0, \infty) = \{x \colon x \geq 0\}$
  - Un intervalo $I = [a, b] = \{a \leq x \leq b\}$

Estudiamos entonces variables aleatorias continuas

$$X : \Omega \to I.$$

Suponemos que hay un universo Ω en donde mediante un experimento se elige al azar un $\omega \in \Omega$ y que nosotros observamos

- $X(\omega)$ antes del experimento
- $X(\omega) = x \in I$ después del experimento

Lo que estudiamos a continuación es cómo se distribuyen las probabilidades en I

Variable uniforme en $[a, b]$

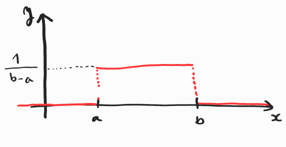
- Si X es tiene distribución uniforme en $[a, b]$, se verifica

$$\boxed{P(X \in [x, y]) = \frac{y - x}{b - a} \qquad a \leq x < y \leq b.}$$

- La densidad de la variable aleatoria X uniforme en $[a, b]$ es la función

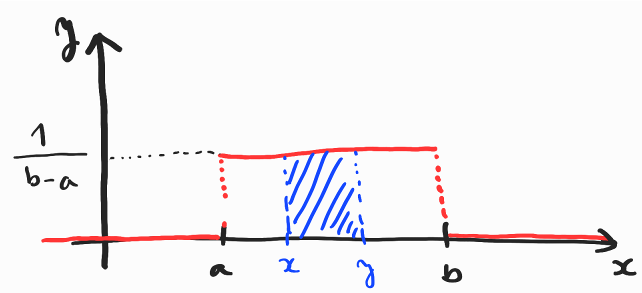
$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{cuando } a \leq x \leq b \\ 0, & \text{en otro caso} \end{cases}$$

se llama densidad de la variable aleatoria X, uniforme en $[a, b]$



- Las probabilidades se calculan como áreas:
- Para calcular áreas usamos integrales:

$$P(X \in [x, y]) = \int_x^y f(t)dt = \int_x^y \frac{1}{b-a}dt = \frac{y-x}{b-a}$$



Distribución de una variable aleatoria uniforme

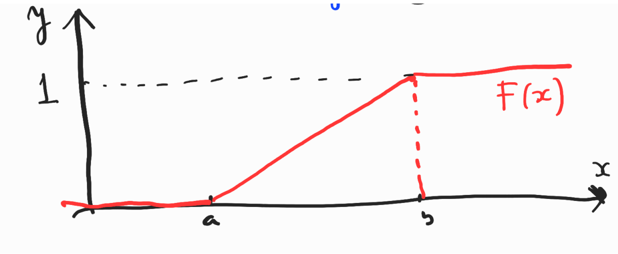
- Más en general nos interesa la siguiente función:

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

llamada función de distribución o también función de distribución de probabilidades e la variable aleatoria X

Obtenemos

$$F(x) = \begin{cases} 0, & \text{cuando } x < a \\ \frac{x-a}{b-a} & \text{cuando } a \leq x \leq b \\ 1, & \text{cuando } b \leq x \end{cases}$$



- También podemos obtener la densidad f a partir de la distribución F.
- Sabemos que se verifica

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

- Si derivamos con respecto a x obtenemos

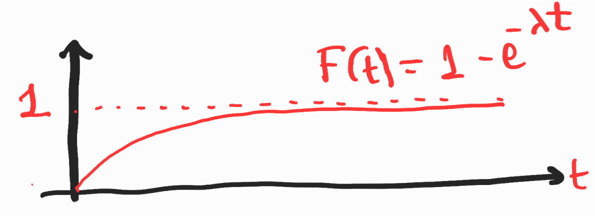
$$F'(x) = f(x)$$

- La densidad f y la distribución F tienen la misma información: se puede obtener una a partir de la otra

Variable aleatoria exponencial

- El segundo modelo continuo que consideramos es una variable aleatoria[1] T con distribución exponencial
- Es una variable que toma valores positivos: $I = [0, \infty)$
- Se puede definir a partir de su distribución:

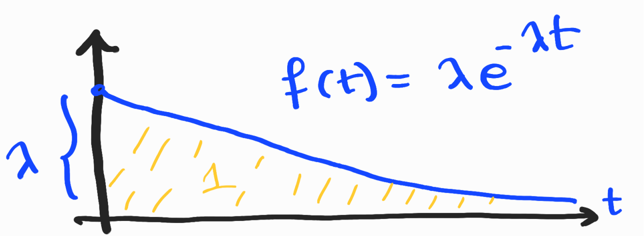
$$F(t) = P(T \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0)$$



[1] usamos la T porque podría tratarse de un tiempo aleatorio

Derivando tenemos la densidad exponencial:

$$f(t) = F'(t) = \lambda e^{-\lambda t} \quad (t \geq 0)$$



- El modelo exponencial se utiliza para medir duraciones de componentes que no envejecen
- Por ejemplo, que se rompen por motivos externos a su desgaste
- Se puede usar para modelar el tiempo que dura una lámpara, que se quema por un pico de corriente externo
- Eso es porque cumple la propiedad de pérdida de memoria.
- Para eso usamos la sobrevida de T, que es

$$P(T > t) = e^{-\lambda t}.$$

Pérdida de memoria

- La pérdida de memoria es una propiedad matemática:

$$\boxed{P(T > t + h \mid T > t) = P(T > h)}$$

- Primero verificamos la propiedad:

$$\begin{aligned} P(T > t + h \mid T > t) &= \frac{P(\{T > t + h\} \cap \{T > t\})}{P(T > t)} \\ &= \frac{P(\{T > t + h\})}{P(T > t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda(t+h)+\lambda t} \\ &= e^{-\lambda h} = P(T > h) \end{aligned}$$

Ahora la interpretamos:

$$P(T > t + h \mid T > t) = P(T > h)$$

- Si T no ocurrió en tiempo t, esperar h más (hasta $t + h$) es equivalente a esperar h a partir de 0
- Por eso no importa cuanto tiempo ocurrió, si la lámpara no se rompió, es como que fuese nueva

Simulación de una variable exponencial

En R los cuatro comandos de una variable exponencial son:

- `dexp` me da la densidad exponencial
- `pexp` me da la distribución de probabilidades exponencial
- `qexp` me da la función inversa F^{-1} de la distribución F
- `rexp` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

```
> rexp(3,1)
[1] 1.1611435 2.1357145 0.3589149
```

simula 3 variables exponenciales con parámetro $\lambda = 1$.

Aplicación: isótopos radiactivos

- El tiempo de desintegración de un isótopo radiactivo se modela mediante una v.a. exponencial de parámetro λ
- La vida media τ se define como $\tau = 1/\lambda$
- La vida media del Uranio 232 es de 70 años.

Problema: Cual es la probabilidad de que *n* átomos de Uranio (independientes) se desintegren todos antes de 100 años.

Solución.

- Sean T_k el tiempo de desintegración del átomo $k = 1, \ldots, n$.
- Cada T_k es una variable exponencial independiente de parámetro $\lambda = 1/70$.
- El tiempo se mide en años.

$$P(T_1 \leq 100) = 1 - e^{-\lambda 100} = 1 - e^{-100/70} = 0{,}76$$

- Como queremos que todos se desintegren antes de 100, queremos calcular

$$P(T_1 \leq 100, \cdots, T_n \leq 100)$$
$$= P(T_1 \leq 100) \cdots P(T_n \leq 100) = 0{,}76^n.$$

- En $=$ usamos la independencia.
- Si fuesen $n = 10$ átomos, tenemos

$$0{,}76^{10} = 0{,}06.$$

Ejemplo: las aguavivas inmortales

- Las aguavivas Turritopsis nutricula tiene un ciclo de vida en el que se revierte a pólipo después de llegar a su maduración sexual, presentándose como biológicamente inmortal: Turritopsis nutricula

- A pesar de esta remarcable habilidad, la mayoría de medusas Turritopsis suelen caer víctimas de las amenazas habituales de la vida del plancton, incluyendo ser comido por otros animales, o sucumbir a una enfermedad.

- ¿Se podría utilizar un modelo exponencial para modelar el tiempo de vida de estas aguavivas?

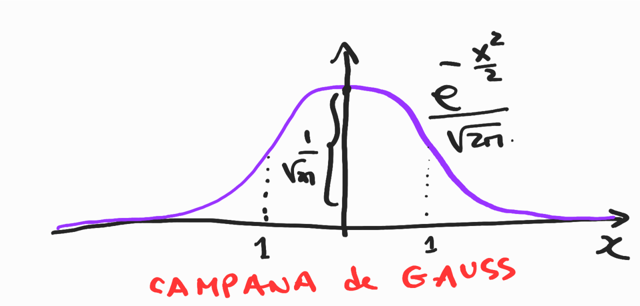
Variable aleatoria normal

Todo el mundo utiliza la variable aleatoria normal, para modelar errores de medición:

- los matemáticos porque piensan que es un hecho experimental,
- y los experimentadores porque suponen que es un teorema matemático.

Definición: Una variable aleatoria X es normal estándar o gaussiana si tiene densidad dada por

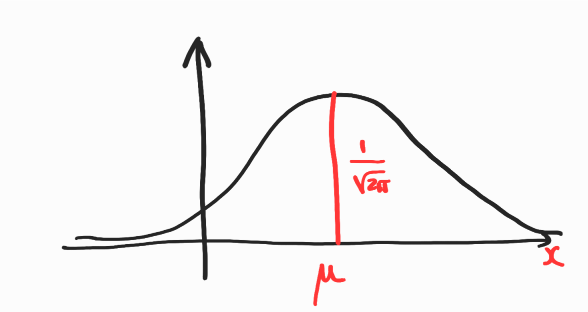
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$



Normal con parámetros μ,1

Nos interesa que el valor central sea un número arbitrario μ.
Obtenemos la fórmula

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

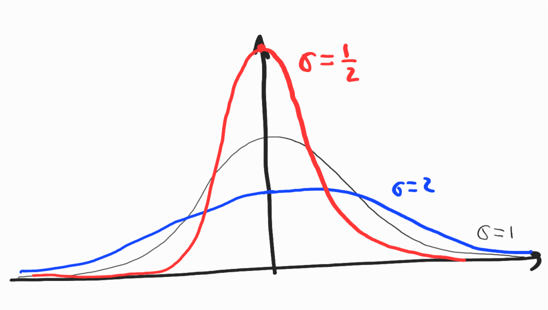


Tenemos $f(\mu) = \frac{1}{\sqrt{2\pi}}$

Normal con parámetros 0,σ

Nos interesa que la variable tenga dispersión arbitraria. Sea para eso $\sigma > 0$, y

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



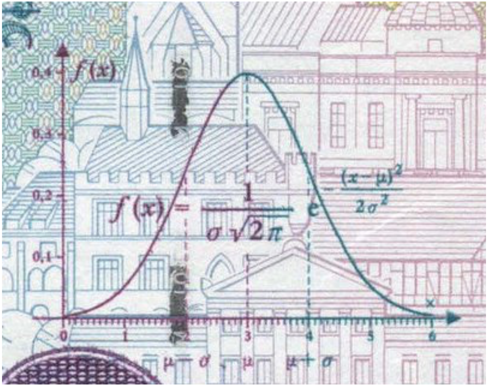
Tenemos $f(0) = \frac{1}{\sigma\sqrt{2\pi}}$

Normal con parámetros μ,σ

Si la posición y la dispersión son arbitrarias, tenemos la densidad

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

Figura: Zehn Deutsche Marks banknote


0,4
$f(x)$
0,3
0,2
0,1
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
0
1
2
3
4
5
6
x
$\mu - \sigma$
μ
$\mu + \sigma$

Áreas y probabilidades

$$\mathbf{P}(a - \sigma \leq X \leq a + \sigma) = 0{,}68$$
$$\mathbf{P}(a - 1{,}96\sigma \leq X \leq a + 1{,}96\sigma) = 0{,}95$$
$$\mathbf{P}(a - 3\sigma \leq X \leq a + 3\sigma) = 0{,}997$$

En el gráfico 3.7, el área de la figura delimitada por el gráfico de la función

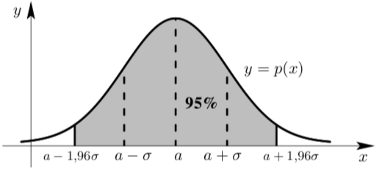


Figura 3.7: Gráfico de la densidad normal $p(x)$ con parámetros (a, σ). El área sombreada es el 95 % del área total

Simulación de una variable normal

En R los cuatro comandos de una variable normal son:

- `dnorm` me da la densidad normal
- `pnorm` me da la distribución de probabilidades normal
- `qnorm` me da la función inversa F^{-1} de la distribución F
- `rnorm` me numeros aleatorios (random) con distribución exponencial
- Por ejemplo:

```
> dnorm(0,mean=0,sd=1)
[1] 0.3989423
```

$$\frac{1}{\sqrt{2\pi}} \sim \frac{1}{2{,}5} = 0{,}4.$$

Distribución normal

La distribución normal es la integral de la densidad normal:

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