

# Tema 1: Introducción a la Estadística Bayesiana

## Introducción

En general, se usan probabilidades de modo informal para expresar la información o la incertidumbre que se tiene acerca de observaciones de cantidades desconocidas. Sin embargo, el uso de probabilidades para expresar la información se puede hacer de modo formal.

Desde el punto de vista matemático se puede demostrar que con el Cálculo de Probabilidades se puede representar de modo numérico el conjunto de racional de creencias, de modo que existe una relación directa entre probabilidad e información y la regla de Bayes proporciona un modo natural de actualización de las creencias cuando aparece nueva información. Este proceso de aprendizaje inductivo por medio de la regla de Bayes es la base de la Inferencia Bayesiana.

De manera general, los métodos bayesianos son métodos de análisis de datos que se derivan de los principios de la inferencia bayesiana. Estos métodos, proporcionan

- Estimadores de los parámetros que tienen buenas propiedades estadísticas;
- Una descripción parsimoniosa (simple) de los datos observados;
- Estimación de los datos *missing* y predicciones de futuras observaciones;
- Una metodología computacional potente para la estimación, selección y validación de modelos.

La metodología bayesiana consta de tres pasos fundamentales:

1. Especificar un modelo de probabilidad que incluya algún tipo de conocimiento previo (*a priori*) sobre los parámetros del modelo dado.
2. Actualizar el conocimiento sobre los parámetros desconocidos condicionando este modelo de probabilidad a los datos observados.

3. Evaluar el ajuste del modelo a los datos y la sensibilidad de las conclusiones a cambios en los supuestos del modelo.

La diferencia fundamental entre la estadística clásica (*frecuentista*) y la bayesiana es el concepto de probabilidad. Para la estadística clásica es un concepto *objetivo*, que se encuentra en la naturaleza, mientras que para la estadística bayesiana se encuentra en el *observador*, siendo así un concepto subjetivo.

De este modo, en estadística clásica sólo se toma como fuente de información la muestra obtenidas suponiendo, para los desarrollos matemáticos, que se podría tomar un tamaño infinito para una muestra de manera hipotética.

En el caso bayesiano, sin embargo, además de la muestra también juega un papel fundamental la información previa o externa que se posee en relación a los fenómenos que se tratan de modelizar.

## Definiciones y Teoremas Básicos

El concepto básico en estadística bayesiana es el de *probabilidad condicional*:

Para dos sucesos  $A$  y  $B$ ,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \propto P(A \cap B)$$

Se puede aplicar esta definición también a variables discretas o continuas.

Desde el punto de vista bayesiano, en la práctica, todas las probabilidades son condicionales porque casi siempre existe algún conocimiento previo o cierta experiencia acerca de los sucesos.

### Ejemplo

En el caso judicial en USA de *Connecticut vs. Teal* se alegó un caso de posible discriminación racial en base a un test para una promoción de estudiantes dada.

Se observó que entre el grupo de personas que hacían un test, 48 eran negras ( $B$ ) y 259 eran blancas ( $W$ ). De este modo si se toma una persona al azar

$$P(B) = \frac{48}{307} = 0,16 \quad P(W) = \frac{259}{307} = 0,84$$

De las personas negras 26 aprobaron el test ( $P$ ) y el resto falló ( $F$ ), mientras que en las personas blancas 206 aprobaron y el resto, no. De este modo un total de 232 aprobaron.

De modo que

$$\begin{aligned} P(B|P) &= \frac{26}{232} = 0,11 \\ P(W|P) &= \frac{206}{232} = 0,89 \end{aligned}$$

Existe, por tanto, la tentación de asumir que existe una muy elevada discriminación racial...

Sin embargo lo que se debería calcular es la distribución condicionada contraria: asumiendo que si se es negro (o blanco) cuál es la probabilidad de aprobar:

$$\begin{aligned}P(P|B) &= \frac{26}{48} = 0,54 \\P(P|W) &= \frac{206}{259} = 0,80\end{aligned}$$

Y aún así, la probabilidades son bastante diferentes, aunque no tanto.

### Ley de la Probabilidad Total:

Para un suceso  $A$  y una partición  $B_1, \dots, B_k$ ,

$$P(A) = \sum_{i=1}^k P(A|B_i)P(B_i)$$

Se puede aplicar el teorema a variables discretas:

$$f(x) = \sum_y f(x|Y = y)P(Y = y)$$

o a variables continuas:

$$f(x) = \int f(x|y)f(y) dy.$$

### Ejemplo

En una fábrica de rodamientos se embalan en 4 cadenas de montaje;  $A_1$ ,  $A_2$ ,  $A_3$  y  $A_4$ . El 35 % de la producción total se embala en la cadena  $A_1$  y el 20 %, 24 % y 21 % en  $A_2$ ,  $A_3$  y  $A_4$  respectivamente. Los datos indican que no se embalan correctamente un porcentaje pequeño de las cajas; el 1 % de  $A_1$ , el 3 % de  $A_2$ , el 2.5 % de  $A_3$  y el 2 % de  $A_4$ . ¿Cuál es la probabilidad de que una caja elegida al azar de la producción total sea defectuosa?

Defino como  $D$  = defectuosa.

Luego,

$$\begin{aligned}P(D) &= \sum_{i=1}^4 P(D|A_i)P(A_i) = \\&= 0,01 \times 0,35 + 0,03 \times 0,20 + 0,025 \times 0,24 + \\&\quad + 0,02 \times 0,21 = \\&= 0,0197\end{aligned}$$

## Ejemplo

Supongamos que  $X|Y \sim \text{Pois}(Y)$ , una distribución Poisson, para  $x = 0, 1, 2, \dots$  para  $y > 0$ , donde  $Y \sim \text{Exp}(\beta)$ , una distribución exponencial

$$P(x|y) = \frac{y^x}{x!} e^{-y}$$

$$f(y) = \beta \exp(-\beta y)$$

Entonces, la distribución marginal de  $X$  es

$$\begin{aligned} P(x) &= \int_{-\infty}^{\infty} P(x|y) f(y) dy \\ &= \int_0^{\infty} \frac{y^x}{x!} e^{-y} \beta \exp[-\beta y] dy \\ &= \frac{\beta}{x!} \int_0^{\infty} y^x \exp[-(\beta + 1)y] dy \\ &= \frac{\beta}{x!} \int_0^{\infty} y^{(x+1)-1} \exp[-(\beta + 1)y] dy \end{aligned}$$

Para resolver la integral, se observa que el integrando está relacionado con una distribución gamma  $\text{Ga}(x + 1, \beta + 1)$ .

---

**NOTA:**

Si  $X \sim \text{Ga}(a, b)$  su función de densidad es

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp[-bx],$$

de este modo

$$\int_0^{\infty} \frac{b^a}{\Gamma(a)} x^{a-1} \exp[-bx] dx = 1 \implies \int_0^{\infty} x^{a-1} \exp[-bx] dx = \frac{\Gamma(a)}{b^a}$$

Alternativamente se puede calcular con Maxima:

```
assume (a>0, b>0);  
declare(a, noninteger);  
integrate((x^(a-1))*exp(-b*x), x, 0, inf);
```

---

Luego

$$\begin{aligned} P(x) &= \frac{\beta}{x!} \frac{\Gamma(x+1)}{(\beta+1)^{(x+1)}} \\ &= \frac{\beta}{x!} \frac{x!}{(\beta+1)^{(x+1)}} \\ &= \frac{\beta}{(\beta+1)^{(x+1)}} \end{aligned}$$

Si se denota como  $p = \beta/(1 + \beta)$ , entonces  $0 < p < 1$  y despejando  $\beta = \frac{p}{1-p}$ ,

$$\begin{aligned} P(x) &= p \left(1 + \frac{p}{1-p}\right)^{-x} = p \left(\frac{1}{1-p}\right)^{-x} = \\ &= p(1-p)^x, \end{aligned}$$

para  $x = 0, 1, 2, \dots$

Se observa que es una distribución **geométrica** con parámetro  $p$ .

**Ejemplo**

Si  $X|\theta \sim \text{Exp}(\theta)$  y  $\theta \sim \text{Ga}(\alpha, \beta)$ , la distribución marginal es

$$\begin{aligned} f(x) &= \int_0^{\infty} \theta e^{-\theta x} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \theta^\alpha e^{-(\beta+x)\theta} d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} \theta^{(\alpha+1)-1} e^{-(\beta+x)\theta} d\theta \end{aligned}$$

y el integrando está relacionado con otra distribución gamma,  $\text{Ga}(\alpha + 1, \beta + x)$ :

$$\int_0^\infty \theta^{(\alpha+1)-1} e^{-(\beta+x)\theta} d\theta = \frac{\Gamma(\alpha + 1)}{(\beta + x)^{\alpha+1}}.$$

Entonces,

$$\begin{aligned} f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + 1)}{(\beta + x)^{\alpha+1}} = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\alpha \Gamma(\alpha)}{(\beta + x)^{\alpha+1}} \\ &= \frac{\alpha \beta^\alpha}{(\beta + x)^{\alpha+1}}, \end{aligned}$$

donde se ha utilizado la propiedad básica de la función gamma,

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha).$$

No es una distribución estándar, pero si se define la v.a.  $Z = X + \beta$ , se puede ver que  $Z$  tiene una distribución de *Pareto*.

#### NOTA:

Ver, por ejemplo,

[http://en.wikipedia.org/wiki/Pareto\\_distribution](http://en.wikipedia.org/wiki/Pareto_distribution)

Para ello aplicamos el *teorema del cambio de variable*:

#### Teorema del cambio de variable

Sea  $X$  una v.a. con función de densidad  $f_X$  y sea  $g$  una función diferenciable, monótona con inversa diferenciable  $g^{-1}$ .

Entonces la función de densidad de  $U = g(X)$  es

$$f_U(u) = f_X [g^{-1}(u)] \left| \frac{d}{du} g^{-1}(u) \right|.$$

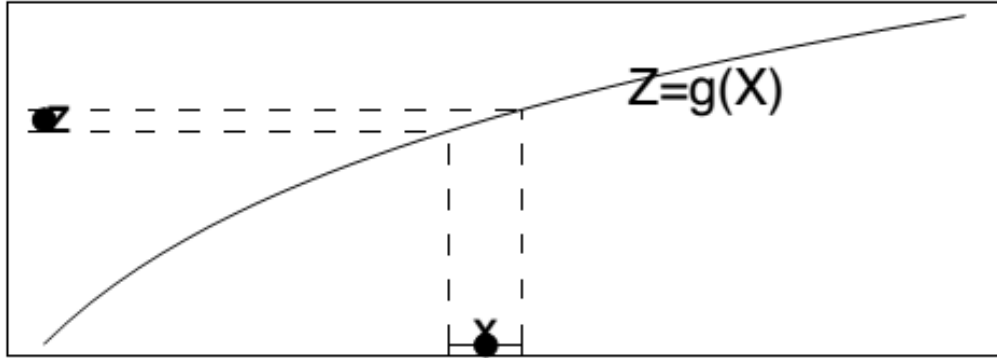
Se puede escribir de una manera más intuitiva como:

$$f_U(u) = f_X(x) \left| \frac{dx}{du} \right|.$$

Si  $g$  no es biyectiva, entonces se pueden encontrar subconjuntos de  $\mathbb{R}$  en los que  $g$  es biyectiva, y aplicar el teorema separadamente en cada subconjunto.

El resultado previo se puede explicar mediante una figura que muestra una serie de valores  $x, z$  que están relacionados mediante  $z = g(x)$ . Se muestra un pequeño intervalo alrededor de  $x$ : digamos  $\mathbf{I}_x$ . Se aplica mediante  $g$  a otro pequeño intervalo alrededor de  $z$ ; digamos  $\mathbf{I}_z$

La densidad correspondiente es



$$p_Z(z) \approx \frac{P(Z \in \mathbf{I}_z)}{\text{amplitud}(\mathbf{I}_z)} = \frac{P(X \in \mathbf{I}_x) \cdot \text{amplitud}(\mathbf{I}_x)}{\text{amplitud}(\mathbf{I}_x) \cdot \text{amplitud}(\mathbf{I}_z)} \approx p_X(x) \left| \frac{dx}{dz} \right|$$

La aproximación es exacta cuando las longitudes de  $\mathbf{I}_x$  y  $\mathbf{I}_z$  tienden a 0.

### Ejemplo

Supongamos que  $X$  es una variable aleatoria con función de densidad

$$p_X(x) = 2x \quad \text{si } x \in (0, 1)$$

Si  $Z = \frac{1}{X}$  ¿cuál es la densidad  $p_Z(z)$ ?

La transformación inversa es  $g^{-1}(Z) = 1/Z$  entonces, la función de densidad es

$$p_Z(z) = p_X(g^{-1}(z)) \left| \frac{dg^{-1}(z)}{dz} \right| = \frac{2}{z} \left| -\frac{1}{z^2} \right| = \frac{2}{z^3}$$

y los posibles valores de  $Z$  van de 1 a  $\infty$ .

Entonces,

$$p_Z(z) = \frac{2}{z^3} \quad \text{if } z \in (1, \infty)$$

### Ejemplo de la distribución de Pareto

$$\begin{aligned} f_Z(z) &= f_X(z - \beta) \cdot 1 \\ &= \alpha \beta^\alpha z^{-\alpha-1}, \end{aligned}$$

para  $Z > \beta$ . Luego  $Z \sim \text{PA}(\beta, \alpha)$ .

La distribución de *Pareto* se aplicó inicialmente a la modelización del reparto de riqueza. Es la llamada *ley 80-20* que afirma que el 20 % de la población posee el 80 % de la riqueza.

## El teorema de Bayes

Se tiene que, para los sucesos  $A_1, \dots, A_n$  y  $B$ ,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \propto P(B|A_i)P(A_i)$$

### Ejemplo

Los mellizos pueden ser de dos tipos: (M) monocigóticos (que proceden del mismo óvulo) o (D) dicigóticos (que proceden de óvulos diferentes). En general, los monocigóticos se parecen mucho y son del mismo género, mientras que los dicigóticos en ocasiones son bastante diferentes e incluso de géneros diferentes.

Así, asumiendo que los dos géneros son igualmente probables y denotando para cada par de mellizos (mujer u hombre) como  $mm$ ,  $hh$ ,  $mh$  entonces

$$\begin{aligned} P(mm|M) &= P(hh|M) = \frac{1}{2} & P(mh|M) &= 0 \\ P(mm|D) &= P(hh|D) = \frac{1}{4} & P(mh|D) &= \frac{1}{2} \end{aligned}$$

De lo que se deduce que

$$P(mm) = P(mm|M)P(M) + P(mm|D)P(D) = \frac{1}{2}P(M) + \frac{1}{4}(1 - P(M))$$

De modo que

$$P(M) = 4 \cdot P(mm) - 1$$

Y se obtiene un modo de estimar la proporción de de mellizos monocigóticos que hay en una población simplemente observando la distribución de géneros con respecto a todos los mellizos.

### Ejemplo

Supongamos un juego televisivo en el que tienes que elegir entre tres puertas cerradas,  $A$ ,  $B$  o  $C$ . Detrás de dos de las puertas hay una peineta y en la otra hay un coche, con igual probabilidad en los tres casos. Por tanto, la probabilidad de ganar el coche en cada una de las puertas es

$$p(A) = \frac{1}{3}, \quad p(B) = \frac{1}{3}, \quad p(C) = \frac{1}{3}.$$

Después de que hayas elegido una puerta, digamos  $A$ , antes de mostrarte lo que hay detrás de la puerta, el presentador (Wyoming) abre otra puerta, digamos  $B$ , que tiene una peineta. En este punto te ofrece la opción de cambiar de la puerta  $A$  a la puerta  $C$ . ¿Qué deberías hacer?



Intuitivamente parece que tú has elegido la puerta adecuada, pero que Wyoming te quiere liar... así, desde un punto de vista inocente la probabilidad de encontrar el coche entre las dos puertas que quedan es  $\frac{1}{2}$ . Pero esto es falso...

Asumimos que Wyoming va en tu contra (cobra de la productora de televisión) y calculamos cuál es la probabilidad de que el coche aparezca cuando él abre la puerta  $B$ , una vez que tú hayas abierto la puerta  $A$ :

- (i) La probabilidad de que Wyoming abra la puerta  $B$  dado que el coche está detrás de la puerta  $A$  es

$$p(B_{Wy}|A) = \frac{1}{2}$$

ya que le es indiferente abrir la puerta  $B$  o  $C$ .

- (ii) La probabilidad de que Wyoming abra la puerta  $B$  dado que el coche está detrás de la puerta  $B$  es

$$p(B_{Wy}|B) = 0$$

porque supones que no es *estúpido*.

- (iii) La probabilidad de que Wyoming abra la puerta  $B$  dado que el coche está detrás de la puerta  $C$  es

$$p(B_{Wy}|C) = 1$$

Aplicando la ley de la probabilidad total, se obtiene:

$$\begin{aligned} p(B_{Wy}) &= p(B_{Wy}|A) \cdot p(A) + p(B_{Wy}|B) \cdot p(B) + p(B_{Wy}|C) \cdot p(C) = \\ &= \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2} \end{aligned}$$

Finalmente, aplicando el teorema de Bayes, se tiene que

$$\begin{aligned} p(A|B_{Wy}) &= \frac{p(B_{Wy}|A) \cdot p(A)}{p(B_{Wy})} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3} \\ p(C|B_{Wy}) &= \frac{p(B_{Wy}|C) \cdot p(C)}{p(B_{Wy})} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} \end{aligned}$$

Luego es mucho **mejor** que elijas la puerta  $C$ .

Se puede aplicar el teorema de Bayes a variables discretas y continuas. En el caso de que la v.a.  $X$  sea continua se tiene

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)} = \frac{f(y|x)f(x)}{\int_{\mathbb{R}} f(y|x)f(x)dx},$$

como el denominador  $f(y)$  es independiente de  $x$ , entonces se puede escribir el teorema en la forma de proporcionalidad ( $\propto$ ):

$$f(x|y) \propto f(y|x)f(x).$$

Este resultado es útil para los cálculos porque implica que se pueden olvidar las constantes multiplicativas hasta el final de los cálculos en modelos complicados.

### Ejemplo

Retomando el ejemplo de la Poisson, se tenía que  $X|Y \sim \text{Pois}(Y)$  e  $Y \sim \text{Exp}(\beta)$ . Calculamos la distribución de  $Y|x$ , sabiendo que la distribución marginal de  $X$  era una geométrica:

$$\begin{aligned} f(y|x) &= \frac{P(x|y)f(y)}{P(x)} \propto P(x|y)f(y) \\ &\propto \frac{y^x e^{-y}}{x!} \beta e^{-\beta y} \\ &\propto y^x e^{-(\beta+1)y} \\ &\propto y^{(x+1)-1} e^{-(\beta+1)y} \end{aligned}$$

que es el **núcleo** de la densidad de una variable gamma:  $\text{Ga}(x + 1, \beta + 1)$ .

### Ejemplo

Volviendo al ejemplo donde  $X|\theta \sim \text{Exp}(\theta)$  y  $\theta \sim \text{Ga}(\alpha, \beta)$ , calculamos la distribución de  $\theta$  dada una observación  $x$  (se tenía que la distribución marginal de  $X$  era *Pareto*). De un modo directo:

$$\begin{aligned} f(\theta|x) &\propto f(x|\theta)f(\theta) \\ &\propto \theta e^{-\theta x} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &\propto \theta^{(\alpha+1)-1} e^{-(\beta+x)\theta} \end{aligned}$$

que está relacionado con una distribución gamma, es decir,  $\theta|x \sim \text{Ga}(\alpha + 1, \beta + x)$ .

## La media y varianza condicionadas.

Dadas dos variables  $X$  e  $Y$ , definimos la media y varianza de  $X$  cuando  $Y = y$  como

$$E[X|Y = y] = \int x f(x|y) dx$$

$$Var[X|Y = y] = \int (x - E[X|Y = y])^2 f(x|y) dx$$

El siguiente teorema nos proporciona la relación entre la esperanza y varianza marginal y la esperanza y varianza condicionadas.

### Teorema

Dadas dos variables  $X$  e  $Y$ , se tiene que

$$(i) E_x[X] = E_y[E_x[X|Y]]$$

$$(ii) Var_x[X] = E_y[Var_x[X|Y]] + Var_y[E_x[X|Y]]$$

*Demostración:*

(i) Se tenía que, en general,

$$E(g(x)) = \int g(x)f(x) dx$$

por ello, como  $E[X|Y]$  es una función de  $Y$ ,

$$\begin{aligned} E_y[E_x[X|Y]] &= \int E_x(X|y)f(y) dy \\ &= \int \left( \int x f(x|y) dx \right) f(y) dy \\ &= \int x \left( \int f(x|y)f(y) dy \right) dx \\ &= \int x \left( \int f(x, y) dy \right) dx \\ &= \int x f(x) dx = E_x[X] \end{aligned}$$

(ii) Se generaliza trivialmente el resultado anterior para una función cualquiera  $h$  de modo que

$$E[E[h(X)|Y]] = E[h(X)]$$

y en particular tomando como  $h$  la función cuadrado de un un valor,

$$E[E[X^2|Y]] = E[X^2]$$

Así

$$\begin{aligned} E_y[Var_x[X|Y]] &= E[E[X^2|Y] - (E[X|Y])^2] \\ &= E[E[X^2|Y]] - E[E[X|Y]^2] \\ &= E[X^2] - E[E[X|Y]^2] \end{aligned}$$

Mientras que

$$\begin{aligned} Var_y[E_x[X|Y]] &= E[E[X|Y]^2] - (E[E[X|Y]])^2 \\ &= E[E[X|Y]^2] - (E[X])^2 \end{aligned}$$

Sumando las expresiones previas se obtiene que

$$E_y[Var_x[X|Y]] + Var_y[E_x[X|Y]] = E[X^2] - (E[X])^2 = Var[X]$$

### Ejemplo

Volviendo al ejemplo de la Poisson, se tenía que  $X|Y \sim \text{Pois}(Y)$  e  $Y \sim \text{Exp}(\beta)$ . Supongamos que queremos calcular la media y varianza de  $X$  (y que no sabemos nada acerca de la distribución marginal de  $X$  que antes se mostró que sigue una distribución geométrica).

$$\begin{aligned} E_x[X] &= E_y[E_x[X|Y]] \\ &= E_y[Y] \quad \text{porque } X|Y \sim \text{Pois}(Y) \\ &= \frac{1}{\beta} \quad \text{la media de la exponencial} \end{aligned}$$

$$\begin{aligned}
\text{Var}_x[X] &= E_y[\text{Var}_x[X|Y]] + \text{Var}_y[E_x[X|Y]] \\
&= E_y[Y] + \text{Var}_y[Y] \quad \text{porque media = varianza} = Y \\
&= \frac{1}{\beta} + \frac{1}{\beta^2} \\
&= \frac{\beta + 1}{\beta^2}
\end{aligned}$$

Sustituyendo  $p = \frac{\beta}{1+\beta}$  y despejando  $\beta = \frac{p}{1-p}$ , se obtiene que

$$\begin{aligned}
E[X] &= \frac{1-p}{p} = \frac{q}{p} \\
\text{Var}[X] &= \frac{1-p}{p} + \left(\frac{1-p}{p}\right)^2 \\
&= \frac{1-p}{p^2} = \frac{q}{p^2},
\end{aligned}$$

que son los momentos que se obtienen directamente para la distribución geométrica en la notación habitual.

### Ejemplo

Retomando el ejemplo de la distribución de *Pareto*, donde  $X|\theta \sim \text{Exp}(\theta)$  y  $\theta \sim \text{Ga}(\alpha, \beta)$ , se tiene que

$$\begin{aligned}
E[X] &= E_\theta[E_x[X|\theta]] = E_\theta[1/\theta] \\
&= \int_0^\infty \frac{1}{\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^{(\alpha-1)-1} e^{-\beta\theta} d\theta
\end{aligned}$$

El integrando es el *núcleo* de una distribución gamma;  $\text{Ga}(\alpha - 1, \beta)$ . Entonces,

$$E[X] = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha - 1)}{\beta^{\alpha-1}} = \frac{\beta}{\alpha - 1},$$

es decir, la esperanza sólo existe si  $\alpha > 1$ .

Hemos visto anteriormente que  $Z = X + \beta \sim \text{PA}(\beta, \alpha)$ . De este modo, podemos calcular la media de  $X$  utilizando también la fórmula para la media de una distribución *Pareto*:

$$\begin{aligned}
E[X] &= E[Z] - \beta \\
&= \frac{\alpha\beta}{\alpha - 1} - \beta \quad [\text{para } \alpha > 1] \\
&= \frac{\beta}{\alpha - 1}.
\end{aligned}$$

## Intercambiabilidad y Teorema de *de Finetti*

Se realiza una encuesta nacional para averiguar si la población es o no feliz. Se denomina  $Y_i$  a la v.a. asociada:

$$Y_i = \begin{cases} 1 & \text{Si la persona } i \text{ dice que es feliz} \\ 0 & \text{Si la persona } i \text{ dice que no es feliz} \end{cases}$$

Supongamos que se toma una muestra de 10 personas.

Se trata de calcular las probabilidades de

$$\begin{aligned}
p(1, 0, 0, 1, 0, 1, 1, 0, 1, 1) &= ? \\
p(1, 0, 1, 0, 1, 1, 0, 1, 1, 0) &= ? \\
p(1, 1, 0, 0, 1, 1, 0, 0, 1, 1) &= ?
\end{aligned}$$

¿Hay algún argumento para asignarles el mismo valor de probabilidad? Se puede observar que cada secuencia contiene 6 *unos* y cuatro *ceros*.

### Definición

Sea  $p(y_1, \dots, y_n)$  la distribución conjunta de  $Y_1, \dots, Y_n$ .

Si  $p(y_1, \dots, y_n) = p(y_{\pi_1}, \dots, y_{\pi_n})$  para todas las permutaciones  $\pi$  de  $\{1, \dots, n\}$ , entonces  $Y_1, \dots, Y_n$  se dice que son **intercambiables** ■

Intuitivamente, se dice que  $Y_1, \dots, Y_n$  son intercambiables si las etiquetas de los subíndices no contienen información sobre el resultado obtenido.

Consideremos las siguientes probabilidades:

$$\begin{aligned}
P(Y_{10} = 1) &= a \\
P(Y_{10} = 1 | Y_1 = Y_2 = \dots = Y_8 = Y_9 = 1) &= b
\end{aligned}$$

Si  $a \neq b$  entonces  $Y_{10}$  **NO** es independiente de  $Y_1, \dots, Y_9$ .

Supongamos que se conoce el valor del parámetro  $\theta$  que es la tasa real de felicidad.  
 ¿Son razonables las siguientes probabilidades?

$$\begin{aligned}
 P(Y_{10} = 1|\theta) &\approx \theta \\
 P(Y_{10} = 1|Y_1 = Y_2 = \dots = Y_8 = Y_9 = 1, \theta) &\approx \theta \\
 P(Y_9 = 1|Y_1 = Y_2 = \dots = Y_8 = Y_{10} = 1, \theta) &\approx \theta
 \end{aligned}$$

Si lo anterior es razonable, se puede considerar que las  $Y_i$  son **condicionalmente independientes e idénticamente distribuidas dado  $\theta$**  (o al menos de manera aproximada con una muestra grande).

Así

$$\begin{aligned}
 P(Y_i = y_i|\theta, Y_j = y_j, j \neq i) &= \theta^{y_i}(1 - \theta)^{1-y_i} \\
 P(Y_1 = y_1, \dots, Y_{10} = y_{10}|\theta) &= \prod_{i=1}^{10} \theta^{y_i}(1 - \theta)^{1-y_i} = \\
 &= \theta^{\sum y_i}(1 - \theta)^{10 - \sum y_i}
 \end{aligned}$$

Por otro lado, se puede describir la información previa que se tiene sobre el parámetro  $\theta$  mediante la distribución  $p(\theta)$  denominada **distribución a priori**.

La distribución marginal es entonces

$$\begin{aligned}
 p(y_1, \dots, y_{10}) &= \int_0^1 p(y_1, \dots, y_{10}|\theta)p(\theta)d\theta = \\
 &= \int_0^1 \theta^{\sum y_i}(1 - \theta)^{10 - \sum y_i}p(\theta)d\theta
 \end{aligned}$$

Si se consideran las secuencias binarias del ejemplo inicial,

$$\begin{aligned}
 p(1, 0, 0, 1, 0, 1, 1, 0, 1, 1) &= \int_0^1 \theta^6(1 - \theta)^4p(\theta)d\theta \\
 p(1, 0, 1, 0, 1, 1, 0, 1, 1, 0) &= \int_0^1 \theta^6(1 - \theta)^4p(\theta)d\theta \\
 p(1, 1, 0, 0, 1, 1, 0, 0, 1, 1) &= \int_0^1 \theta^6(1 - \theta)^4p(\theta)d\theta
 \end{aligned}$$

Luego se puede afirmar que  $Y_1, \dots, Y_{10}$  son **intercambiables**.

**Teorema:**

Si  $\theta \sim p(\theta)$  e  $Y_1, \dots, Y_n$  son condicionalmente i.i.d. dado  $\theta$ , entonces  $Y_1, \dots, Y_n$  son intercambiables marginalmente (incondicionalmente respecto de  $\theta$ ) ■

La demostración es inmediata.

Y la interpretación es

$$\begin{aligned}
 Y_1, \dots, Y_n|\theta \text{ i.i.d.} &\implies Y_1, \dots, Y_n \text{ son intercambiables} \\
 \theta \sim p(\theta) &
 \end{aligned}$$

Aunque lo que resulta más interesante es la implicación contraria, lo que lleva al teorema de **de Finetti**:

$$\begin{array}{l} Y_1, \dots, Y_n | \theta \text{ i.i.d.} \\ \theta \sim p(\theta) \end{array} \iff Y_1, \dots, Y_n \text{ son intercambiables para todo } n$$

Se puede expresar como:

TEOREMA DE DE FINETTI

Sea  $Y_i \in \mathcal{Y}$  para todo  $i \in \{1, 2, \dots\}$ , supongamos que para todo  $n$  el modelo para  $Y_1, \dots, Y_n$  es intercambiable:

$$p(y_1, \dots, y_n) = p(y_{\pi_1}, \dots, y_{\pi_n})$$

para todas las permutaciones  $\pi$  de  $\{1, \dots, n\}$ .

Entonces, el modelo se puede escribir como

$$p(y_1, \dots, y_n) = \int \left\{ \prod_1^n p(y_i | \theta) \right\} p(\theta) d\theta$$

para algún parámetro  $\theta$ , alguna distribución a priori en  $\theta$  y algún modelo paramétrico  $p(y|\theta)$ .

La intercambiabilidad para todo  $n$  se cumple de manera natural cuando  $Y_1, \dots, Y_n$  es una muestra con reemplazamiento de una población finita, o una muestra sin reemplazamiento de una población infinita.