Análisis de datos con Métodos Bayesianos

Instituto Bayesiano para la Investigación en Desarrollo (BayesGroup.org)

Rolando Gonzales
rgonzales@bayesgroup.org

5 de octubre de 2015

Espacio Muestral

El espacio muestral (S) son los posibles resultados de un experimento. Si el experimento es lanzar una moneda, el espacio muestral contiene dos resultados, cara y escudo,

$$S = \{C, E\}$$

Si el experimento es la reacción en el tiempo a cierto estímulo, el espacio muestral estará definido en $\mathbb{R}^+$,

$$S = \{0, \infty\}$$

Un evento es una colección de los posibles resultados de un experimento, i.e., cualquier subconjunto de S incluyendo S en sí mismo.

Probabilidad

Cuando se realiza un experimento, la realización es un resultado en el espacio muestral.

Para cada evento A del espacio muestral S, puede asociarse un número entre cero y uno que se llamará probabilidad de A, $\mathbb{P}(A)$.

Para una definición más precisa, es necesario definir primero el concepto de sigma álgebras.

Sigma álgebra (definición). *Una colección de subconjuntos de S se llama sigma álgebra (σ-algebra o campo de Borel), denotado por $\mathcal{B}$, si satisface las siguientes propiedades:*

1. $\emptyset \in \mathcal{B}$
2. *Si* $A \in \mathcal{B}$, *entonces* $A^c \in \mathcal{B}$
3. *Si* $A_1, A_2, \ldots \in \mathcal{B}$, *entonces* $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$

Probabilidad

Función de probabilidad (definición). *Dado un espacio muestral S y una sigma álgebra asociada $\mathcal{B}$, una función de probabilidad será aquella que satisfaga,*

1. $\mathbb{P}(A) \geq 0$ *para todo* $A \in \mathcal{B}$
2. $\mathbb{P}(S) = 1$
3. *Si* $A_1, A_2, \ldots \in \mathcal{B}$ *son disjuntos, entonces* $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Estas propiedades se denominan usualmente Axiomas de Probabilidad o Axiomas de Kolgomorov. La terna $(S, \sigma_{\mathcal{B}}, \mathbb{P})$ es un espacio de probabilidad en el que cada suceso $A \in \mathcal{B}$ recibe el nombre de probabilidad de A.

Variables aleatorias

La estadística analiza variables que fluctúan de una forma más o menos impredecible.

Variable aleatoria (definición). *Dado un espacio* $(S, \sigma_{\mathcal{B}}, \mathbb{P})$, *una variable aleatoria es una función del espacio muestral* S *a* $\mathbb{R}$, $X : S \to \mathbb{R}$.

Ejemplo: Para $S = \{C, E\}$, es posible definir una función $X(m)$ tal que,

$$X(m) = \begin{cases} 1 \text{ si m = C} \\ 0 \text{ si m = E} \end{cases}$$

Función de distribución acumulada

Asociada a una variable aleatoria X existe una función, denominada función de distribución acumulada de X.

Función de distribución acumulada (definición). *La función de distribución acumulada (cdf) de una variable aleatoria X, denotada por $F_X(X)$, se define como,*

$$F_X(X) = \mathbb{P}_X(X \leq x), \text{ para todo} x$$

Esta función satisface las siguientes propiedades,

1. $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1.$
2. $F(x)$ *es una función no decreciente de* x.
3. $\lim_{x\downarrow x_0} F(x) = F(x_0)$

Nótese que una variable aleatoria será continua si $F_X(x)$ es una función continua de x, y será discreta si $F_X(x)$ es una función en pasos de x.

Funciones de masa y densidad

Asociada a X y a su cdf F_X existe otra función, llamada función de densidad de probabilidad (pdf) o función de masa de probabilidad (pmf), refiriéndose al caso continuo y discreto, respectivamente. Ambas se refieren a probabilidades puntuales de variables aleatorias.

Función de masa de probabilidad (definición) . *La función de masa de probabilidad (pmf) $f_X(x)$ de una variable aleatoria discreta X e es* $f_X(x) = \mathbb{P}(X = x)$ para todo x

Función de densidad de probabilidad (definición) . *La función de densidad de probabilidad (pdf) $f_X(x)$ de una variable aleatoria continua X es* $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ para todo x *(nótese que $\frac{d}{dx}F_X(x) = f_X(x)$ y en el caso discreto, de manera similar, las probabilidades puntuales $f_X(x)$ se añaden para obtener $F_X(x)$)*

La expresión "X tiene una distribución $F_X(x)$", se escribe $X \sim F_X(x)$.

Momentos

La esperanza de una variable aleatoria es una medida de su tendencia central, que resume el valor esperado de esta variable.
Esperanza de una variable aleatoria (definición). *La esperanza de una variable aleatoria* $g(X)$*, denotada por* $\mathbb{E}(X)$*,es*

$$\mathbb{E}(X) \begin{cases} \sum_{x \in X} g(X)f(X) = \sum_{x \in X} g(X)\mathbb{P}(X = x) \text{ si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} g(X)f(X)dx \text{ si } X \text{ es continua,} \end{cases}$$

Sea una muestra $x_1, ..., x_n$. La esperanza $\mathbb{E}(X)$ está relacionada con el primer momento muestral (la media),

$$\overline{X}_n = \frac{1}{n} \sum\nolimits_{i=1}^{n} X_i$$

El segundo momento central es la varianza.

Varianza (definición). *La varianza de una variable aleatoria* X *es su segundo momento central,* $Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2$*. La raíz cuadrada positiva de* $Var(X)$ *es la desviación estándar de* X*.*

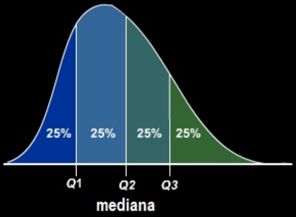
La varianza proporciona una medida del grado de dispersión de una distribución alrededor de la media. La varianza muestral de una muestra $x_1, ..., x_n$ se calcula con,

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)$$

Momentos superiores son el sesgo y la curtosis de una distribución. El sesgo mide el grado de simetría de una distribución, y la curtosis el grado de apuntamiento de la distribución. Dependiendo del valor de la curtosis, una distribución puede ser leptocúrtica (curtosis $>$ 3), mesocúrtica (curtosis $=$ 3) o platicúrtica (curtosis $<$ 3).

Fractiles

Fractiles (definición). *Sea X una variable aleatoria continua y sea $\alpha \in (0, 1)$. Si $q = q(X; \alpha)$ es aquel tal que $\mathbb{P}(X < q) = \alpha$ y $\mathbb{P}(X > q) = 1 - \alpha$, entonces q es llamado un fractil de X.*



Si se expresa las probabilidades en porcentaje, q será el 100α percentil de la distribución de X.

Distribuciones de probabilidad

Las distribuciones de probabilidad pueden ser discretas o continuas. Una variable aleatoria X tiene una distribución discreta si el rango de X, el espacio muestral, es contable (en la mayoría de las situaciones, se encuentra en $\mathbb{Z}$). Por ejemplo, una variable aleatoria X tiene una d*istribución uniforme discreta* si,

$$\mathbb{P}(X = x|N) = \frac{1}{N}, \quad x = 1, 2, ..., N$$

para $N \in \mathbb{Z}^+$. Esta distribución pone igual masa en todos los posibles resultados $1, 2, ..., N$.
En usual también trabajar con *variables continuas*, definidas en $\mathbb{R}$, o en $\mathbb{R}^{0,+}$ (como las variables de stock), por lo que se prestará atención a las distribuciones de variables continuas en esta sección, especialmente a las utilizadas usualmente en inferencia.

Distribución uniforme continua (definición). *La pdf de una distribución uniforme continúa es,*

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para } x < a \text{ o } x > b \end{cases}$$

por lo que si una variable aleatoria continua $X \sim \mathcal{U}(a,b)$, *su distribución queda definida por los parámetros* a *y* b. *Si* $a = 0$ *y* $b = 1$ *la distribución resultante* $\mathcal{U}(0,1)$ *se denomina distribución uniforme estándar.*

Distribución gaussiana (definición). *Una variable aleatoria continua X se distribuye normalmente (sigue una distribución Gaussiana o Gauss-Laplace) con media μ y varianza σ^2, denotado por $X \sim \mathcal{N}(\mu, \sigma^2)$, si la pdf de X es,*

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

Algunas propiedades útiles de $f(x)$ son,

1. $f(\mu + x) = f(\mu - x)$ (simetría alrededor de μ).
2. $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \cong 0.683$
3. $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \cong 0.954$
4. $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \cong 0.997$

Nótese que no existe una única distribución gaussiana, sino una familia de distribuciones determinada por diferentes valores de μ y σ^2.

Distribución gaussiana estándar (definición). *Una variable aleatoria continua* X *sigue una distribución normal estándar si* $\mu = 0$ *y* $\sigma^2 = 1$*, por lo que* $X \sim \mathcal{N}(0, 1)$*.*

Distribución log-normal (definición). *Una variable aleatoria cuyo logaritmo este normalmente distribuido tendrá una pdf,*

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

Distribución χ^2. *Si* $X_1, \ldots, X_k$ *son iid variables que siguen una distribución normal estándar* $\mathcal{N}(0,1)$, *y* $Q = \sum_{i=1}^{k} X_k^2$, *entonces* Q *sigue una distribución* χ^2 *con* k *grados de libertad,* $Q \sim \chi_k^2$. *La pdf de una distribución* χ^2 *es,*

$$f(x;\,k) = \frac{1}{2^{k/2}\Gamma(k/2)}\, x^{k/2-1}\, e^{-x/2}\, I_{\{x\geq 0\}}$$

Nótese que la pdf queda definida por un sólo parámetro: los grados de libertad k.

Distribución t de Student. *Sean dos variables aleatorias $Z \sim N(0,1)$ y $Y \sim \chi^2(k)$, la variable aleatoria continua T,*

$$T = \frac{Z}{\sqrt{\frac{Y}{k}}}$$

seguiría una distribución t-Student con una pdf,

$$f(T,k) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{T^2}{k}\right)^{-\frac{1}{2}(k+1)}$$

Una propiedad importante e interesante de la distribución t de Student es que puede ser usada para muestras pequeñas que impliquen la distirbución Gaussiana.
Sea F_k la cdf de $t(k)$ y Φ la cdf de una distribución $\mathcal{N}(0,1)$. Entonces, $\lim_{k\to\infty} F_k(t) = \mathcal{N}(0,1)$.De hecho para $k > 40$ se lográ una buena aproximación.