Repaso de Probabilidades

Introducción

En este capítulo, se recordarán algunos aspectos específicos del curso de probabilidades, en especial, lo relacionado a variables aleatorias, funciones de distribución, esperanza, varianza, y algunos teoremas de importancia para los capítulos siguientes.

Variable Aleatoria

Una variable aleatoria es una función que asocia un suceso con un número real o:

$$\begin{aligned} X \ : \ \Omega &\longrightarrow A \\ \omega &\longrightarrow X(\omega) \end{aligned}$$

El conjunto A puede ser $\mathbb{R}$ o $\mathbb{Z}$. Teóricamente se asocia una determinada probabilidad a cada posible valor de la variable aleatoria X. El conjunto de valores que toma la variable junto con sus probabilidades asociadas constituyen una *distribución de probabilidad*. Si el conjunto de valores que toma la variable aleatoria es discreto, entonces se dice que la v.a. es *discreta*, en otro caso se dirá *continua*. Se tiene para cada caso[1]:

Caso discreto. En este caso la masa total de probabilidad se distribuye en un conjunto discreto de valores de la recta real. La función de distribución en este caso se determina como:

[1] Los resultados siguientes se extienden para más de una variable, en cuyo caso se habla de distribución o densidad *conjunta*

$$p(x) = P(X = x) \geq 0 \qquad x \in \Omega_x$$
$$\sum_{x_i \in \Omega_x} p(x_i) = 1$$

Caso continuo. En este caso, la masa de probabilidad se reparte mediante una *función de densidad de probabilidad* (f.d.p.), tal que:

$$f(x) \geq 0 \quad \forall x \in \Omega_x$$
$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

Además, se cumple que $f(x) = \frac{d}{dx}F(x)$, excepto quizá para un número finito de valores de x.

Ya sea el caso continuo o discreto, la *función de distribución acumulada* (F.d.a.) se define como:

$$F(x) = P(X \leq x)$$

Algunas de sus propiedades:

- $0 \leq F(x) \leq 1$
- Si $x \geq y$ entonces $F(x) \geq F(y)$
- $F(+\infty) = 1$
- $F(-\infty) = 0$

Esperanza

La esperanza es un promedio ponderado de los valores de una variable aleatoria, donde los "pesos" corresponden a las probabilidades asociadas a dichos valores. Dicho de otra forma, la esperanza es el "centro de masa" de una variable aleatoria. Se define la esperanza de una v.a. X como:

$$E(X) = \begin{cases} \displaystyle\sum_{x \in \Omega_x} x p(x) & \text{caso discreto} \\ \\ \displaystyle\int_{\Omega_x} x f(x)\, dx & \text{caso continuo} \end{cases}$$

Propiedades:

Sean X e Y dos v.a.s, y a, b dos constantes, entonces:

- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- En general:

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$

- $E(a) = a$
- La esperanza es una constante, i.e.: $E(E(X)) = E(X)$

Momentos de orden k de una variable aleatoria

Los momentos de una v.a. se definen como:

$$\mu_k = E(X^k) = \begin{cases} \displaystyle\sum_{x \in \Omega_x} x^k p(x) & \text{caso discreto} \\ \\ \displaystyle\int_{\Omega_x} x^k f(x)\, dx & \text{caso continuo} \end{cases}$$

Nótese que si $k = 1$, se tiene el momento de orden 1, es decir, la esperanza. Existen otra clase de momentos, llamados *momentos centrales*, ya que se calculan respecto del valor esperado:

$$m_k = E[(X-\mu)^k] = \begin{cases} \displaystyle\sum_{x\in\Omega_x} (x-\mu)^k p(x) & \text{caso discreto} \\ \displaystyle\int_{\Omega_x} (x-\mu)^k f(x)\,dx & \text{caso continuo} \end{cases}$$

Para el caso $k = 2$ se tiene la *varianza* de la variable aleatoria X.

Otras medidas de tendencia central

Al igual que la esperanza, existen otras medidas de tendencia central. Se define la *mediana* como el valor m, tal que divide una distribución por la mitad[2], es decir, la probabilidad acumulada a la izquierda y a la derecha de m es de 0,5. Además, la mediana no se ve afectada por los extremos a diferencia del valor esperado. La *moda* es el valor de mayor ocurrencia de una v.a., y corresponde al valor que maximiza la función de densidad de la misma.

Varianza

Se define la varianza de una v.a. X como:

$$Var(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$$

La varianza es una medida de *dispersión* de los datos, es decir,que tan lejos están unos de otros. Nótese que la unidad de medida de la varianza es el cuadrado de la unidad de X. A la raíz cuadrada de la varianza se le denomina *desviación estándar*.

Algunas propiedades de la varianza son:

Sean X, Y dos v.a.s, y a, b dos constantes:

- $Var(aX + b) = a^2 Var(X)$
- Para cualquier constante c: $Var(c) = 0$

[2]Sin emabrgo, hay que tener ojo cuidado con esta definición en el caso de v.a.s discretas

- Sean $X_1, \ldots, X_n$ variables aleatorias, entonces:

$$Var(\sum_{i=1}^{n} \alpha_i X_i) = \sum_{i=1}^{n} \alpha_i^2 Var(X_i) + \sum_i \sum_j \alpha_i \alpha_j Cov(X_i, X_j) \qquad (i \neq j)$$

Donde se define la *covarianza* $Cov(\cdot, \cdot)$ como:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

Nótese que $Cov(X, X) = E[(X - E(X))^2] = E(X^2) - E^2(X) = Var(X)$.

Función Generatriz de Momentos

Se define la función generatriz de momentos (f.g.m.) de una v.a. X como:

$$\varphi_X(t) = E(e^{tX}) = \begin{cases} \displaystyle\sum_{x \in \Omega_x} e^{tx} p(x) & \text{caso discreto} \\ \displaystyle\int_{\Omega_x} e^{tx} f(x)\, dx & \text{caso continuo} \end{cases}$$

Propiedades:

- $\varphi_X^{(n)}(t = 0) = E(X^n)$
- Si $Y = aX + b$, entonces $\varphi_Y(t) = e^{bt} \varphi_X(at)$
- Si $\varphi_X(t) = \varphi_Y(t) \Rightarrow F_X = F_Y$

Función de una variable aleatoria

Sea X una v.a. continua, y sea $Y = h(X)$, luego la función de densidad de Y se puede encontrar como:

$$f_Y(y) = f_X(h^{-1}(y)) |\frac{dH^{-1}(y)}{dy}|$$

En el caso multivariado, se tiene que:

$$f_Y(y_1,\ldots,y_n) = f_X(x_1(y),\ldots,x_n(y)) \begin{vmatrix} \frac{\partial x_1(y)}{\partial y_1} & \cdots & \frac{\partial x_1(y)}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n(y)}{\partial y_1} & \cdots & \frac{\partial x_n(y)}{\partial y_n} \end{vmatrix}$$

Donde el Jacobiano corresponde al valor absoluto del determinante presentado.

Esperanza y varianza

Sea $Y = h(X)$ una función de una v.a., luego:

$$E(h(X)) = \begin{cases} \displaystyle\sum_{x\in\Omega_x} h(x)p(x) & \text{caso discreto} \\ \displaystyle\int_{\Omega_x} h(x)f(x)\,dx & \text{caso continuo} \end{cases}$$

La varianza se obtiene a partir de la definición: $Var(X) = E(h^2(X)) - E^2(h(X))$.

Distribuciones marginales e independencia

Sean X e Y variables aleatorias con densidad con densidad conjunta $f(x,y)$, entonces la expresión para la densidad marginal de X es:

$$f_X(x) = \begin{cases} \displaystyle\sum_{y} f(x,y) & \text{caso discreto} \\ \displaystyle\int_{y} f(x,s)\,ds & \text{caso continuo} \end{cases}$$

El resultado es análogo para la densidad marginal de Y.

Independencia

Dos variables aleatorias X e Y son estadísticamente independientes si y solo si la densidad conjunta de ambas es igual al producto de sus densidades marginales:

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

O de otra forma:

$$F(x,y) = F_X(x) \cdot F_Y(y) \Leftrightarrow P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

Si X e Y son independientes se tiene que:

- $E(X \cdot Y) = E(X) \cdot E(Y)$
- $Cov(X,Y) = 0$, pero no necesariamente a la inversa.
- Si $Z = X+Y$, entonces $\varphi_Z(t) = \varphi_X(t) \cdot \varphi_Y(t)$. En general, si $X_1, \ldots, X_n$ son v.a.s independientes, tal que $Z = \sum_{i=1}^{n} X_i$ entonces:
$$\varphi_Z(t) = \prod_{i=1}^{n} \varphi_{X_i}(t)$$

Distribuciones condicionales

Sean X e Y dos v.a.s con densidad conjunta $f(x,y)$, entonces la densidad condicional de Y sobre cada valor de X es:

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

Análogamente para X:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Es fácil observar que si X e Y son idependientes, entonces $f(y|x) = f_Y(y)$ y $f(x|y) = f_X(x)$. Las expresiones para la esperanza y varianza de una distribución condicional son fáciles de obtener usando las definiciones correspondientes:

$$E[Y|X] = \begin{cases} \displaystyle\sum_y yf(y|x) & \text{caso discreto} \\ \\ \displaystyle\int_y yf(y|x)\,dy & \text{caso continuo} \end{cases}$$

En el caso de la varianza se tiene que: $Var[Y|X] = E[Y^2|X] - E^2[Y|X]$, y el desarrollo es análogo para X condicionada a Y.

Algunas propiedades útiles vinculadas a la condicionalidad:

- $E(Y) = E_x[E[Y|X]]$ donde $E_x[\cdot]$ representa la esperanza sobre los valores de X.
- $Cov(X,Y) = Cov[X, E[Y|X]]$
- $Var(Y) = Var_x[E[Y|X]] + E_x[Var[Y|X]]$ donde $Var_x[\cdot]$ representa la varianza sobre los valores de X.

Ley de los grandes números

Suponga que $X_1, X_2, \ldots, X_n$ es una secuencia de variables aleatorias i.i.d. con media μ y varianza σ^2, y sea $S_n = X_1 + X_2 + \ldots + X_n$, de modo que $E(S_n) = n\mu$ y $Var(S_n) = n\sigma^2$, entonces:

- La ley débil de los grandes números muestra que para $\varepsilon > 0$:

$$P(|S_n/n - \mu| > \varepsilon) \to 0 \quad \text{cuando} \quad n \to \infty$$

- La ley fuerte de los grandes números muestra que:

$$P(S_n/n \to \mu) \to 1$$

Teorema Central del Límite

Suponga que $X_1, X_2, \ldots, X_n$ es una secuencia de variables aleatorias i.i.d. con media μ y varianza σ^2, y sea $T_n = X_1 + X_2 + \ldots + X_n$, de modo que $E(T_n) = n\mu$ y $Var(T_n) = n\sigma^2$, entonces para un número n "suficientemente grande"[3]:

$$\frac{T_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

Este resultado es muy importante en Estadística ya que para un tamaño de muestra grande, la suma(o promedio) de variables aleatorias i.i.d. sigue una distribución Normal, sin importar la distribución de origen de las variables de la suma.

[3] Típicamente para una muestra con $n \geq 30$