Fundamentos de Econometría

Ignacio Lobato

ITAM

Distribuciones de Probabilidad Univariadas

- Los distintos eventos que pueden ocurrir son identificados en su totalidad y a estos se les asigna un valor mediante una variable aleatoria X (X es una variable aleatoria). .
- X al ser una variable aleatoria es caracterizada por su dominio y su densidad. .
- En particular distinguiremos dos casos: El discreto y el continuo

El Caso Discreto

- En el caso discreto el nro. de resultados posibles es finito o numerable (contable). Se compila una lista de valores de x ordenadas: x_1, x_2, x_3.

.

- Convencionalmente se ordenan los valores de menor a mayor y se asignan probabilidades con la función f(x), con las siguientes propiedades:
  - $f(x) \geq 0$ en todo punto
  - $f(x) = 0$ excepto en los puntos de masa x_1, x_2...
  - $\sum_i (f(x_i)) = 1$

.

- Existen varios ejemplos de funciones de distribución probabilística discretas.

El Caso Discreto

- Distribución de Bernoulli

$$f(x) = p^x(1-p)^{(1-x)}$$

para x=0,1

- Distribución Uniforme Discreta con parámetro N (entero positivo)

$$f(x) = 1/N$$

para x=1,2,...N

- Distribución Binomial

$$f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

para x=1,2,...N

El Caso Continuo

- Hay un continuo de valores que la variable aleatoria X puede tomar, aquí las probabilidades ya no se calculan puntualmente sino para un intervalo.
- La asignación de probabilidades se realiza a través de la función de densidad que satisface lo siguiente:

$$f(x) \geq 0$$

en todo punto y

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

- Así para cualquier par de números a,b con a$\leq b$

$$Pr(a \leq x \leq b) = \int_{a}^{b} f(x)\, dx$$

El Caso Continuo

- De esta manera definimos la función de distribución acumulada (cdf) como:

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

Es decir, F(x) denota

$$Pr(X \leq x) \tag{1}$$

- La cdf presenta las siguientes propiedades:
  1) $F(-\infty) = 0$
  2) F(.) es monotónicamente no decreciente.
  3) F(.) es diferenciable en cualquier punto por su definición.
- Note lo siguiente:

$$Pr(a \leq X \leq b) = \int_{a}^{b} f(x) = \int_{-\infty}^{b} f(x) - \int_{-\infty}^{a} f(x) = F(b) - F(a)$$

El Caso Continuo

- Algunos ejemplos de funciones de distribución continua:

  1) Rectangular o Uniformemente Continua en el intervalo [a, b]. La pdf es:

  $$f(x) = 1/(b-a)$$

  para $a \leq x \leq b$

  De esta manera tendríamos que:

  $$F(x) = \begin{cases} 0 \text{ si } x < a \\ \text{(x-a)/(b-a) si } a \leq x \leq b \\ 1 \text{ si } b < x \end{cases}$$

  2) Exponencial con parámetro $\lambda > 0$. La pdf es $f(x) = \lambda e^{-\lambda x}$ para $x > 0$:

  Entonces tendríamos que:

  $$F(x) = \begin{cases} 0 \text{ si x} \leq 0 \\ 1 - e^{-\lambda x} \; si \; x > 0 \end{cases}$$

El Caso Continuo

3) Normal Estándar. La pdf es $\phi(x)$ y denotamos la cdf por $\Phi(x)$ donde:

$$\phi(x) = (2\pi)^{-0,5} exp(-x^2/2)$$

4) Logística Estándar. La pdf es:

$$f(x) = e^x/(1+e^x)^2$$

De esta manera:

$$F(x) = e^x/(1+e^x)$$

5) Potencia en el intervalo [0, 1] con parámetro $\theta > 0$. La pdf es:

$$f(x) = \theta x^{\theta-1}$$

para $0 \leq x \leq 1$ con densidad cero en cualquier otro punto. De esta manera, tenemos que:

$$F(x) = \begin{cases} 0 \text{ si } x < 0 \\ x^\theta \text{ si } 0 \leq x \leq 1 \\ 1 \text{ si } x > 1 \end{cases}$$

Esperanzas en el caso Univariado

- La media en una distribución de probabilidad es llamada La Esperanza o Valor Esperado.

- Supongamos que la variable aleatoria X tiene una función de distribución o de densidad f(x), $X \sim f(x)$.

- Luego la esperanza de X es definida como:

$$E(X) = \begin{cases} \sum_i x_i f(x_i) \text{ en el caso discreto} \\ \int_{-\infty}^{\infty} x f(x) dx \text{ en el caso continuo} \end{cases}$$

- De esta manera, dada Z=h(x), podemos calcular la esperanza de dicha variable aleatoria utilizando la función de distribución de X.

$$E(Z) = \begin{cases} \sum_i h(x_i) f(x_i) \text{ en el caso discreto} \\ \int_{-\infty}^{\infty} h(x) f(x) dx \text{ en el caso continuo} \end{cases}$$

Momentos Centrados y No Centrados

- Sea $X^* = X - E(X) = X - \mu$.
- $E(X^r) = \mu_r^{'}$ es el momento r-ésimo no centrado de X.
- $E(X^{*r}) = \mu_r$ es el momento r-ésimo centrado en la media de X
- Cada momento da información sobre la distribución de X. Por ejemplo tendremos que si r=1, $E(X) = \mu$ y $E(X^*) = 0$.
- Si tomamos r=2, tendremos $E(X^2)$ y $E(X^{*2}) = \sigma^2$ es la varianza de X.

Teoremas Usando Esperanzas

- **T1:** Sea X una variable aleatoria y a,b dos constantes cualesquiera, tales que Z=a+bX. Entonces se cumple que:

$$E(Z) = a + bE(X)$$

y

$$V(Z) = b^2 V(X)$$

- **T2:** La Varianza de una variable aleatoria se puede calcular como:

$$V(X) = E(X^2) - E^2(X)$$

Teoremas Usando Esperanzas

- **T3:** (Error Cuadrático Medio): Sea c cualquier constante. Luego el error cuadrático medio de una variable X respecto a c es:

$$E(X-c)^2 = \sigma^2 + (c-\mu)^2$$

Demostración:

Podemos escribir (X-c) = (X-μ) $-$ $(c-\mu) = X^* - (c-\mu)$.

Luego:

$$(X-c)^2 = X^{*2} + (c-\mu)^2 - 2(c-\mu)X^*$$

Usando T1:

$$E(X-c)^2 = E(X^{*2}) + (c-\mu)^2 - 2(c-\mu)E(X^*)$$

Teoremas Usando Esperanzas

Como $E(X^{*2}) = \sigma^2$ y $E(X^*) = 0$, tenemos que:

$$E(X-c)^2 = \sigma^2 + (c-\mu)^2$$

- **T4 (Error Cuadrático Mínimo):** El valor que minimiza $E(X-c)^2$ es μ.

  Demostración: Utilizando el T3, tenemos que:

  $$E(X-c)^2 = \sigma^2 + (c-\mu)^2$$

  Pero $(c-\mu)^2 \geq 0$ con igualdad sí y solo sí c$=\mu$

Predicción

- Suponemos que la distribución de la variable aleatoria X es conocida (f(x) conocida) y queremos predecir el valor de X a partir de una constante c como predictor.
- Suponemos que nuestro criterio para evaluar la calidad del predictor es el cuadrático medio a través de su minimización.
- Es decir queremos minimizar $E(U^2)$ con $U = (X - c)$.
- Del T4, se sabe que c^* es μ y a partir de ello, $E(U^2) = \sigma^2$.
- A partir de ello, sea $\epsilon = X - \mu$, de lo cual se observa que $E(\epsilon) = 0$ y $E(\epsilon^2) = \sigma^2$.
- Cuando la esperanza del error de predicción de un predictor es cero, decimos que dicho predictor es insesgado (μ es insesgado).

- **Desigualdad de Markov:** Si Y es una variable aleatoria no-negativa; es decir, $Pr(Y < 0) = 0$, y k es cualquier constante positiva, luego $Pr(Y \geq k) \leq E(Y)/k$.

  Demostración (para el caso continuo):

$$E(Y) = \int_0^\infty yf(y)dy = \int_0^k yf(y)dy + \int_k^\infty yf(y)dy = a + b$$

  Ahora como $a \geq 0$, $E(Y) \geq b$. También $b \geq k \int_k^\infty f(y)dy = kPr(Y \geq k)$, tenemos que E(Y)$\geq kPr(Y \geq k)$.

Esperanzas y Probabilidades

- **Desigualdad de Chebychev 1:** Si X es una variable aleatoria, c es cualquier constante, y d es una constante positiva, luego:

$$Pr(|X - c| \geq d) \leq E(x - c)^2/d^2$$

Demostración:

Sea $Y = (X - c)^2$ una variable aleatoria no negativa. Así:

$|X - c| \geq d \leftrightarrow Y \geq d^2$

De esta manera, haciendo $k = d^2$ y aplicando la desigualdad de Markov tenemos que:

$$E(Y) \geq d^2 Pr(Y \geq d^2)$$

- **Desigualdad de Chebychev 2:** Si X es una variable aleatoria con $E(X) = \mu$ y varianza $V(X) = \sigma^2$, y d es cualquier constante positiva, entonces:

$$Pr(|X - \mu| \geq d) \leq \sigma^2/d^2$$

Demostración: Aplicar la Desigualdad de Chebychev 1 con $c = \mu$.

- **Desigualdad de Jensen :** Si $Y = h(X)$ es cóncava y $E(X) = \mu$, entonces: $E(Y) \leq h(\mu)$.

Distribuciones de Probabilidad Bivariadas

- Para estudiar las relaciones entre variables necesitamos más de una variable en las distribuciones de probabilidad. Se detalla aquí el caso bivariado.
- Los resultados son distinguidos por los valores de un par de variables aleatorias X y Y. Nos referiremos a (X,Y) como un vector aleatorio.
- Al igual que en el caso univariado, tendremos el caso discreto y el caso continuo.

Distribuciones de Probabilidad Bivariadas (Caso Discreto)

- Los valores que toman X y Y son finitos o numerables. Así podemos distinguir entre los distintos pares que denotan los valores tomados por X y Y.
- Los puntos enlistados son llamados puntos de masa. Asimismo, la función f(x,y) es llamada la función de masa de probabilidad conjunta y satisface las siguientes propiedades:

  \- $f(x, y) \geq 0$ en todo punto y $f(x, y) > 0$ solo en los puntos de masa.
  \- $\sum_i \sum_j f(x_i, y_i) = 1$

- De esta manera Pr(X=x, Y=y)=f(x,y).

Distribuciones de Probabilidad Bivariadas (Caso Continuo)

- Hay un continuo de valores posibles que pueden tomar X y Y. Así hay un continuo bidimensional de valores que puede tomar el par (X,Y).
- La función f(x,y) es llamada la función de densidad de probabilidad conjunta, la cual satisface las siguientes propiedades:

  - $f(x, y) \geq 0$ en cualquier punto.
  -$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$

- De esta manera, tenemos que para cualquier a$\leq b, c \leq d$:

$$Pr(a \leq X \leq b, c \leq Y \leq d) = \int_{a}^{b} \int_{c}^{d} f(x, y) dydx$$

Distribuciones Marginales (Caso Discreto)

- Sean dos variables aleatorias X y Y. La distribución marginal de X denota la probabilidad de que X tome un determinado valor independientemente de los valores que tome Y.
- Sea $A = \{X = x\}$ y $A_j = \{X = x, Y = y_j\}$ para $j = 1, 2...$ Nótese que $A = \bigcup_j A_j$ que es una unión disjunta de eventos.
- Así calculamos:
  $Pr(X = x) = Pr(A) = \sum_j Pr(A_j) = \sum_j f(x, y_j) = f_1(x)$
- Esta nueva función $f_1(x)$ es llamada la función de distribución marginal de X. Note que $f_1(x) \geq 0$ en cualquier punto y $f_1(x) > 0$ solo en los puntos de masa o enlistados.
- Por otro lado, se tiene que $\sum_i f_1(x_i) = \sum_i [\sum_j f(x_i, y_j)] = f_1(x)$
- Lo mismo aplicaría en el caso de la variable aleatoria Y.

Distribuciones Marginales (Caso Continuo)

- Sea $A = \{a \leq X \leq b\} = \{a \leq X \leq b, -\infty \leq Y \leq \infty\}$.
- Luego $Pr(a \leq X \leq b) = Pr(A) = \int_a^b \int_{-\infty}^{\infty} f(x, y) dydx = \int_a^b f_1(x) dx$.
- Es decir: $f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$ a la cual se le llama la función de densidad marginal de X.
- Nótese que $f_1(x) \geq 0$ en cualquier punto y que $\int_{-\infty}^{\infty} f_1(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dydx = 1$. Por lo tanto, $f_1(x)$ actúa como una función de densidad univariada.
- Lo mismo aplicaría para el caso de la variable aleatoria Y.

Distribuciones Condicionales (Caso Discreto)

- Sea $A = \{X = x\}$, $B = \{Y = y\}$.
- Por probabilidades se sabe que $Pr(A \backslash B) = Pr(A \bigcap B)/Pr(B)$ siempre que $Pr(B) > 0$.
- Entonces en el caso discreto tendríamos que $Pr(A \bigcap B) = f(x, y)$ y $Pr(B) = f_2(y)$.
- Entonces $Pr(A/B) = f(x, y)/f_2(y) = g_1(x/y)$.
- $g_1(x \backslash y)$ es llamada la función de distribución (masa) condicional de X dada Y=y.
- Note que $\sum_i g_1(x_i/y) = \sum_i [f(x_i, y)/f_2(y)] = [\sum_i f(x_i, y)]/f_2(y) = f_2(y)/f_2(y) = 1$.

Distribuciones Condicionales (Caso Continuo)

- Análogamente al caso anterior, para cada y tal que $f_2(y) \neq 0$, definimos la función: $g_1(x/y) = f(x,y)/f_2(y)$, dejando indefinido $g_1(./y)$ indefinido en otros puntos, esta función es llamada la función de densidad de X dado Y=y.

- Se puede ver que $f(x/y) \geq 0$ en cualquier punto y que:

$$\int_{-\infty}^{\infty} g_1(x/y)dx = \int_{-\infty}^{\infty} [f(x,y)/f_2(y)]dx = [\int_{-\infty}^{\infty} f(x,y)dx]/f_2(y)$$

$$= f_2(y)/f_2(y) = 1$$

Distribuciones Condicionales con un Nro. finito de Variables

- Sean $X_1, X_2, X_3 ... X_n$ una colección de n variables aleatorias con $n \in \mathbb{N}$.
- Podemos expresar las funciones de densidad conjunta en términos de las densidades condicionales de la siguienta manera:

$$f(x_1, x_2, x_3 ..., x_n) = f(x_1/x_2, x_3, ..., x_n) * f(x_2, x_3, ..., x_n)$$

$$= f(x_1/x_2, x_3, ..., x_n) * f(x_2/x_3, x_4, ..., x_n) * f(x_3, x_4, ..., x_n)$$

$$= f(x_1/x_2, x_3, ..., x_n) * f(x_2/x_3, x_4, ..., x_n) * f(x_3/x_4, x_5, ..., x_n) * f(x_4, ..., x_n)$$

$$= f(x_1/x_2, ..., x_n) * f(x_2/x_3, ..., x_n) * ... * f(x_{n-1}/x_n) * f(x_n)$$

- Es decir, podemos expresar la función de densidad conjunta como un producto sucesivo de densidades condicionales.