Estadística. Tema 3

Esperanzas

Objetivos

1. Medidas características distribución de VA.
2. Media y varianza.
3. Correlación lineal vs. esperanza condicional.

Bibliografía

Dekking et al. (2005). Capítulos 7 y 10.

Wasserman (2005). Capítulo 3.

Wooldridge (2006). Apéndice B.

3.1. Esperanza. Propiedades

Existen muchas medidas características de las distribuciones de probabilidad. Las más frecuentes se basan en el concepto de esperanza, aunque pueden tener significados diferentes (posición central, dispersión, asimetría, etc.

Definición. El **valor esperado**, o **esperanza** o media o primer momento de una VA X se define como

$$E(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{si } X \text{ es discreta} \\ \int x f(x)\, dx & \text{si } X \text{ es continua} \end{cases}$$

Notación

$$E(X) = \mathbb{E}X = \int x dF(x) = \mu = \mu_X.$$

Interpretación. Existencia.

Ejemplos.

■ Esperanza de una función de una VA: $Y = r(X)$,

$$E(Y) = E(r(X)) = \int r(x)\, dF_X(x).$$

Si $Y = I_A(X) = 1$ si $X \in A; = 0$ si $X \notin A$, entonces

$$E(I_A(X)) = \int I_A(x)\, dF_X(x) = \int_A dF_X(x) = \Pr(X \in A).$$

Caso multivariante: $Z = r(X, Y)$

$$E(Z) = E(r(X, Y)) = \int r(x, y)\, dF_{X,Y}(x, y).$$

PROPIEDADES

Teorema: si $X_1, \ldots, X_n$ son VA y $a_1, \ldots, a_n$ son constantes, entonces

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right).$$

Ejemplo: Binomial.

Teorema: si $X_1, \ldots, X_n$ son VA independientes, entonces

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E\left(X_i\right).$$

3.2. Varianza y covarianza. Correlación

Sea X es una VA con media μ. La varianza de X, $\sigma^2 = \sigma_X^2 = \mathbb{V}(X) = V(X)$ se define como

$$\sigma^2 = E(X-\mu)^2 = \int (x-\mu)^2 \, dF(x).$$

La desviación estándar se define como $sd(X) = \sqrt{V(X)} = \sigma = \sigma_X$.

Teorema. Propiedades de la varianza:

1. $V(X) = E(X^2) - \mu^2$.
2. Si a y b son constantes, entonces $V(aX+b) = a^2 V(X)$.
3. Si $X_1, \ldots, X_n$ son VA independientes y $a_1, \ldots, a_n$ son constantes, entonces

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i).$$

Ejemplos: Binomial.

■ Si $X_1, \ldots, X_n$ son VA, entonces la **media muestral** se define como

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

y la **varianza muestral** es

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

Si $X_1, \ldots, X_n$ son VA con $\mu = E(X_i)$, $\sigma^2 = V(X_i)$, entonces

$$E\left(\bar{X}_n\right) = \mu, \;\; V\left(\bar{X}_n\right) = \frac{\sigma^2}{n} \quad \text{y} \quad E\left(S_n^2\right) = \sigma^2.$$

Covarianza entre dos VA X e Y, con medias μ_X y μ_Y y desviaciones estándar σ_X y σ_Y :

$$Cov\,(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right).$$

Correlación (coeficiente de correlación):

$$\rho = \rho_{X,Y} = \rho\,(X,Y) = \frac{Cov\,(X,Y)}{\sigma_X \sigma_Y}.$$

Propiedades:

$$\begin{aligned} Cov\,(X,Y) &= E\,(XY) - E\,(X)\,E\,(Y) \\ -1 &\leq \rho\,(X,Y) \leq 1. \end{aligned}$$

- Si $Y = aX + b$, a y b constantes, entonces

$$\begin{aligned} \rho\,(X,Y) &= 1 \quad \text{si} \quad a > 0 \\ \rho\,(X,Y) &= -1 \quad \text{si} \quad a < 0. \end{aligned}$$

- Si X e Y son independientes, entonces $Cov\,(X,Y) = \rho = 0$. Lo contrario no es verdad en general.

$$\begin{aligned} Cov\,(aX+b, cY+d) &= ac \cdot Cov\,(X,Y) \\ \rho\,(aX+b, cY+d) &= \rho\,(X,Y) \quad \text{si } a \text{ y } c \text{ tienen el mismo signo.} \\ V\,(X+Y) &= V\,(X) + V\,(Y) + 2Cov\,(X,Y) \\ V\,(X-Y) &= V\,(X) + V\,(Y) - 2Cov\,(X,Y) \\ V\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i^2 V\,(X_i) + 2\sum_{i=1}^{n}\sum_{i<j} a_i a_j Cov\,(X_i, X_j). \end{aligned}$$

Ejemplo. En una población con un número muy elevado de individuos se considera el experimento de seleccionar uno al azar y observar dos variables,

$$\begin{aligned} Y &\rightarrow \text{"Número de veces que ha estado desempleado"} \\ X &\rightarrow \text{"Número de titulaciones académicas"} \end{aligned}$$

donde las probabilidades de cada combinación de valores viene dada por la siguiente tabla,

$Y \backslash X$	1	2	3	
1	0,4	0,2	0,05	0,65
2	0,2	0,1	0	0,3
3	0,05	0	0	0,05
	0,65	0,3	0,05	1

$$\begin{aligned} E[Y] &= E[X] = 1 \cdot 0{,}65 + 2 \cdot 0{,}3 + 3 \cdot 0{,}05 = 1{,}4 \\ E\left[Y^2\right] &= E\left[X^2\right] = 1^2 \cdot 0{,}65 + 2^2 \cdot 0{,}3 + 3^2 \cdot 0{,}05 = 2{,}3 \\ V[Y] &= V[X] = 2{,}3 - 1{,}4^2 = 0{,}34 \\ E[YX] &= 1 \cdot 1 \cdot 0{,}4 + 1 \cdot 2 \cdot 0{,}2 + 1 \cdot 3 \cdot 0{,}05 + 2 \cdot 1 \cdot 0{,}2 + 2 \cdot 2 \cdot 0{,}1 + 3 \cdot 1 \cdot 0{,}05 \\ &= 1{,}9 \\ C[Y, X] &= 1{,}9 - 1{,}4^2 = -0{,}06 \end{aligned}$$

3.3. Esperanza y varianza condicional. Predicción

Esperanza condicional:

$$E(Y|X=x) = \int y dF_{Y|X=x}(y) = \begin{cases} \sum_y y f_{Y|X=x}(y) & \text{si } Y \text{ es discreta} \\ \int_y y f_{Y|X=x}(y)\,dy & \text{si } Y \text{ es continua} \end{cases}$$

$E(Y|X=x)$ es una función de x, denominada **función de regresión**, mientras que $E(Y|X)$ es una variable aleatoria, con la misma distribución de probabilidad que X, pero con valores $E(Y|X=x)$.

En general:

$$E(r(Y)|X=x) = \int r(y)\,dF_{Y|X=x}(y) = \begin{cases} \sum_y r(y) f_{Y|X=x}(y) & \text{si } Y \text{ es discreta} \\ \int_y r(y) f_{Y|X=x}(y)\,dy & \text{si } Y \text{ es continua} \end{cases}$$

Varianza condicional: se define como la varianza de las distribuciones condicionadas,

$$\begin{aligned} V(Y|X=x) &= \sigma^2(x) \\ &= E(Y - E(Y|X=x)|X=x)^2 \\ &= \int (y - E(Y|X=x))^2\,dF_{Y|X=x}(y) \\ &= E\left(Y^2|X=x\right) - E(Y|X=x)^2 . \end{aligned}$$

Ejemplo. Las probabilidades de Y dado X están dadas por

Y	$\Pr\{Y=y_i\|X=x_1\}$	$\Pr\{Y=y_i\|X=x_2\}$	$\cdots$	$\Pr\{Y=y_i\|X=x_s\}$
y_1	$f_{11}=p_{11}/p_{x1}$	$f_{12}=p_{12}/p_{x2}$	$\cdots$	$f_{1s}=p_{1s}/p_{xs}$
y_2	$f_{21}=p_{21}/p_{x1}$	$f_{22}=p_{22}/p_{x2}$		$f_{2s}=p_{2s}/p_{xs}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_m	$f_{m1}=p_{m1}/p_{x1}$	$f_{m2}=p_{m2}/p_{x2}$	$\cdots$	$f_{ms}=p_{ms}/p_{xs}$
	1	1	$\cdots$	1

y la esperanza condicional de Y dado $X=x_i$ está dada por

$$E[Y|X=x_i] = \sum_{j=1}^{m} y_j f_{ji}.$$

Ejemplo:

Y	$\Pr\{Y=y_i \mid X=1\}$	$\Pr\{Y=y_i \mid X=2\}$	$\Pr\{Y=y_i \mid X=3\}$
1	$f_{11}=0,4/0,65=0{,}615$	$f_{12}=0,2/0,3=0,667$	$f_{13}=0,05/0,05=1$
2	$f_{21}=0,2/0,65=0{,}308$	$f_{22}=0,1/0,3=0,333$	$f_{23}=0$
3	$f_{31}=0,05/0,65=0{,}077$	$f_{32}=0$	$f_{33}=0$
	1	1	1

Por tanto

$$\begin{aligned}
E\left[Y|X=1\right] &= 1\cdot 0{,}615+2\cdot 0{,}308+3\cdot 0{,}077=1{,}462\\
E\left[Y|X=2\right] &= 1\cdot 0{,}667+2\cdot 0{,}333+3\cdot 0=1{,}333\\
E\left[Y|X=3\right] &= 1\cdot 1+2\cdot 0+3\cdot 0=1
\end{aligned}$$

y la función de regresión es

$$\begin{aligned}
E\left[Y^2|X=1\right] &= 1^2\cdot 0{,}615+2^2\cdot 0{,}308+3^2\cdot 0{,}077=2{,}54\\
E\left[Y^2|X=2\right] &= 1^2\cdot 0{,}667+2^2\cdot 0{,}333+3^2\cdot 0=2\\
E\left[Y^2|X=3\right] &= 1^2\cdot 1+2^2\cdot 0+3^2\cdot 0=1
\end{aligned}$$

y entonces

$$\begin{aligned}
V\left[Y|X=1\right] &= 2{,}54-1{,}462^2=0{,}4026\\
V\left[Y|X=2\right] &= 2-1{,}333^2=0{,}2231\\
V\left[Y|X=3\right] &= 1-1^2=0
\end{aligned}$$

PROPIEDADES DE LA ESPERANZA CONDICIONAL

1. $g(X) = E[Y|X]$ es una v.a. (una función de X) con función de probabilidad

$g(X) = E[Y\|X]$	$\Pr\{E[Y\|X] = g(x_i)\}$
$g(x_1)$	p_{x1}
$g(x_s)$	p_{x2}
$\vdots$	
$g(x_s)$	p_{xs}
	1

2. Ley de Esperanzas iteradas

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

porque

$$\begin{aligned} E[E[Y|X]] &= \sum_{i=1}^{s} \overbrace{\left[\sum_{j=1}^{m} y_j f_{ji}\right]}^{=E[Y|X=x_i]} p_{xi} \\ &= \sum_{j=1}^{m} y_j \overbrace{\left[\sum_{i=1}^{s} f_{ji} p_{xi}\right]}^{=p_{yj}} \qquad\qquad f_{ji} = \frac{p_{ji}}{p_{xi}} \\ &= E[Y] \end{aligned}$$

3. Para cualquier función $m(x)$,

$$E[Ym(X)|X = x_i] = m(x_i) E[Y|X = x_i]$$

y por tanto $E[Ym(X)|X] = m(X) E[Y|X]$.

4. Esperanza condicional por etapas:

$$E[Y|X] = E[E[Y|X, Z]|X].$$

5. Si X e Y son independientes, entonces

$$E[Y|X] = E[Y].$$

6. Si $E[Y|X] = E[Y]$ entonces
$$Cov(Y,X) = 0.$$
Prueba:
$$\begin{aligned} Cov(Y,X) &= E[XY] - E[Y]E[X] \\ &= E[E[XY|X]] - E[Y]E[X] \\ &= E[XE[Y|X]] - E[Y]E[X] \\ &= E[XE[Y]] - E[Y]E[X] \\ &= E[X]E[Y] - E[Y]E[X] = 0. \end{aligned}$$
El recíproco de estas dos últimas propiedades no tiene porqué ser verdad en general.

7. **Predicción:**

La función de regresión es la que, de entre todas las posibles (funciones de X), minimiza la varianza del error de predicción,
$$g(x) = E[Y|X=x] = \arg\min_{m(x)} E\left[(Y - m(X))^2 |X = x\right] \quad \text{todo } x,$$
$$\Rightarrow \quad g \ \text{ minimiza en } m \quad E\left[(Y - m(X))^2\right] = \text{ Varianza del error.}$$

8. Descomposición de la Varianza:
$$V(Y) = E[V(Y|X)] + V[E(Y|X)].$$

► Ejemplo:

$$\begin{aligned} E\left(E\left[Y|X\right]\right) &= E\left[Y|X=1\right]*\Pr\left[X=1\right]+E\left[Y|X=2\right]*\Pr\left[X=2\right]+E\left[Y|X=3\right]*\Pr\left[X=3\right] \\ &= 1{,}462*0,65+1{,}333*0,3+1*0,05 \\ &= 1.\,400\; = E\left(Y\right) \end{aligned}$$

y de igual forma

$$\begin{aligned} E\left(E\left[Y|X\right]^2\right) &= E\left[Y|X=1\right]*\Pr\left[X=1\right]+E\left[Y|X=2\right]*\Pr\left[X=2\right]+E\left[Y|X=3\right]*\Pr\left[X=3\right] \\ &= 1{,}462^2*0{,}65+1{,}333^2*0{,}3+1^2*0{,}05 \\ &= 1.\,972\,4 \end{aligned}$$

y por tanto

$$\begin{aligned} V\left(E\left[Y|X\right]\right) &= E\left(E\left[Y|X\right]^2\right)+E\left(E\left[Y|X\right]\right)^2 \\ &= 1.\,972\,4-1.\,400^2 \\ &= 0{,}012\,4. \end{aligned}$$

Por otro lado,

$$\begin{aligned} E\left(V\left[Y|X\right]\right) &= V\left[Y|X=1\right]*\Pr\left[X=1\right]+V\left[Y|X=1\right]*\Pr\left[X=3\right]+V\left[Y|X=1\right]*\Pr\left[X=3\right] \\ &= 0{,}4026*0,65+0{,}2231*0,3+0*0,05 \\ &= 0{,}328\,6 \end{aligned}$$

por lo que

$$V\left(Y\right)=0{,}012\,4+0{,}328\,6=0{,}341\,.$$