Clase 12 de Bioestadística

Varianza de una variables aleatoria
Desigualdad de Chebishev

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Contenidos de la clase

Esperanza (repaso)

Definición: Llamamos esperanza matemática de una variable aleatoria X al número $E(X)$ definido como

$$E(X) = \sum_k x_k P(X = x_k)$$

si X es una variable discreta y

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

donde f es la densidad de X

Para calcular la esperanza matemática de una variable continua tenemos que conocer la la densidad de la variable aleatoria X.

Varianza de una variable aleatoria

- La esperanza matemática $E(X)$ de una variable aleatoria resume su posición
- Nos interesa ahora describir cuan lejos están los valores posibles que toma X de $E(X)$.
- Es decir, nos interesa resumir todos los desvíos $X - E(X)$ en un único número
- Para eso vamos a tomamos un promedio de los desvíos: para que no se eliminen los positivos con los negativos[1], tomamos las cantidades $(X - E(X))^2$

[1] El cuadrado “se come” los signos.

- Definimos entonces la varianza de una variable aleatoria X mediante

$$\boxed{Var(X) = E\left[(X - E(X))^2\right]}$$

- Como la varianza tiene unidades al cuadrado, definimos también el desvío estándar de X mediante

$$\boxed{\sigma_X = \sqrt{Var(X)} = \sqrt{E\left[(X - E(X))^2\right]}}$$

- Observación: tanto la varianza como el desvío estándar son siempre cantidades positivas.

Cálculo de la varianza

- Como definimos la varianza de X como la esperanza de $Y = (X - E(X))^2$, podemos usar la fórmula de esperanza.
- En el caso discreto tenemos entonces

  $$Var(X) = \sum_k (x_k - E(X))^2 P(X = x_k)$$

  si X es una variable discreta y

  $$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

  donde f es la densidad de X

Ejemplo: variable de Bernoulli

- Consideremos una variable X de Bernoulli, que toma los valores 0 y 1, con probabilidades

$$P(X = 0) = 1 - p, \qquad P(X = 0) = p.$$

- Habíamos calculado $E(X) = p$
- Entonces

$$\begin{aligned} Var(X) &= (0 - p)^2(1 - p) + (1 - p)^2 p \\ &= p^2 - p^3 + (1 - 2p + p^2)p \\ &= p^2 \cancel{- p^3} + p - 2p^2 + \cancel{p^3} = p - p^2 = \boxed{p(1 - p)} \end{aligned}$$

Ejemplo: el dado

- El caso del dado es similar pero con 6 resultados. Sabemos que $E(X) = 3{,}5$, y todas las probabilidades son iguales:

$$
\begin{aligned}
Var(X) &= \frac{1}{6}\left[(1 - 3{,}5)^2 + (2 - 3{,}5)^2 + \cdots + (6 - 3{,}5)^2\right] \\
&= 2{,}92
\end{aligned}
$$

Ejemplo: variable Binomial

- En el caso de una variable Binomial B con parámetros n y p, obtenemos
$$Var(B) = np(1-p)$$
- Vale la pena observar que una variable Binomial cuenta los éxitos en n ensayos de Bernoulli:
$$B = X_1 + \cdots + X_n$$
donde X_k es el resultado de cada ensayo, por lo que $Var(X_k) = p(1-p)$.
- El resultado es
$$Var(B) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$$

Otras variables discretas

- El cálculo de la varianza de variables discretas se hace mediante suma de series
- En el caso de una variable geométrica G con parámetro $p \in (0, 1)$ se obtiene

$$Var(G) = \frac{1 - p}{p^2}$$

- En el caso de una variable P de Poisson de parámetro $\lambda > 0$, se obtiene

$$Var(P) = \lambda.$$

Vale la pena observar que $E(G) = Var(G)$.

Varianza de variables continuas

- En este caso tenemos que utilizar la fórmula con la integral y la densidad.

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

- Comencemos con una variable U uniforme en $[0, 1]$. Sabemos que $E(X) = \frac{1}{2}$.

- Tenemos entonces

$$
\begin{aligned}
Var(U) &= \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)dx \\
&= \int_{0}^{1} \left(x - \frac{1}{2}\right)^2 1\, dx \\
&= \int_{0}^{1} \left(x^2 - x + \frac{1}{4}\right) dx \\
&= \left[\frac{x^3}{3}\right]_0^1 - \left[\frac{x^2}{2}\right]_0^1 + \frac{1}{4} \qquad = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} \\
&= \frac{4}{12} - \frac{3}{12} + \frac{3}{12} = \boxed{\frac{1}{12}}
\end{aligned}
$$

Propiedades de la varianza

Propiedad Sea X una variable aleatoria y dados a, b números reales consideremos la variable aleatoria

$$Y = aX + b$$

Entonces

$$E(Y) = aE(X) + b, \qquad Var(Y) = a^2 \; Var(X)$$

- Observemos que la esperanza de Y se traslada b y se multiplica por a.
- Mientras que la varianza se multiplica por b^2 pero no depende de a.

Demostración

- Vamos a demostrarlo en el caso discreto. El caso continuo es similar. Comenzamos con la esperanza:

$$
\begin{aligned}
E(aX+b) &= \sum_k (ax_k+b)P(X=x_k) \\
&= a \boxed{\sum_k x_k P(X=x_k)} + b \boxed{\sum_k P(X=x_k)} \\
&= a \boxed{E(X)} + b \boxed{1} = aE(X)+b.
\end{aligned}
$$

Ahora la varianza:

$$
\begin{aligned}
Var(aX+b) &= E\left[\{aX+b-E(aX+b)\}^2\right] \\
&= E\left[\{aX+\cancel{b}-aE(X)-\cancel{b}\}^2\right] \\
&= E\left[\{aX-aE(X)\}^2\right] \\
&= E\left[\{a(X-E(X))\}^2\right] \\
&= a^2 E\left[\{X-E(X)\}^2\right] = a^2\ Var(X).
\end{aligned}
$$

$\square$

Varianza de una variable uniforme

Para una variable uniforme se puede demostrar que:

$$V = a + (b - a)U \sim U[a, b] \Leftrightarrow U \sim U[0, 1]$$

Entonces

$$Var(V) = (b - a)^2 Var(U) = \frac{(b - a)^2}{12}.$$

Varianza de una variable normal

- Sea $Z \sim \mathcal{N}(0,1)$, es decir, una normal estándar.
- Comenzamos enunciando que

$$Var(Z) = \int_{-\infty}^{\infty} (x - 0)^2 \varphi(x) dx = 1$$

- Entonces si $X \sim \mathcal{N}(\mu, \sigma^2)$, por estandarización se puede escribir

$$X = \sigma Z + \mu$$

- Calculamos

$$Var(X) = Var(\sigma Z + \mu) = \sigma^2 Var(Z) = \sigma^2.$$

- Concluímos que si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces

$$E(X) = \mu, \qquad Var(X) = \sigma^2.$$

Desigualdad de Chebishev

Teorema. Sea X una variable aleatoria que toma valores no negativos, es decir, $X \geq 0$. Entonces, para cualquier $\varepsilon > 0$ tenemos

$$\boxed{P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}}$$

Demostración. Hacemos la demostración en el caso continuo.

$$\begin{aligned} E(X) &= \int_0^\infty xf(x)dx \geq \int_\varepsilon^\infty xf(x)dx \\ &\geq \int_\varepsilon^\infty \varepsilon f(x)dx = \varepsilon \int_\varepsilon^\infty f(x)dx \\ &= \varepsilon P(X \geq \varepsilon) \end{aligned}$$

Por último, si pasamos ε dividiendo tenemos la desigualdad buscada. $\square$

Desigualdad de Markov

Corolario: Sea Y una variable aleatoria y $Var(Y)$ su varianza. Entonces

$$\boxed{P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}}$$

Demostración:

- Observemos que

$$\boxed{P(|Y - E(Y)| \geq \varepsilon)} = P(|Y - E(Y)|^2 \geq \varepsilon^2)$$

- La variable $X = (Y - E(Y))^2$ es no negativa, entonces

$$P((Y - E(Y))^2 \geq \varepsilon^2) \boxed{\leq} \frac{E[(Y - E(Y))^2]}{\varepsilon^2} = \boxed{\frac{Var(Y)}{\varepsilon^2}}$$