Variables Aleatorias Discretas (continuación)

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Varianza y Covarianza

Recordamos de la clase pasada

- Una variable aleatoria es un número asociado al resultado de un experimento aleatorio. Por ejemplo, cuando efectuamos una medición. Lo modelamos por una función $X : \Omega \to \overline{\mathbb{R}}$.
- Hay distintos tipos de variables aleatorias. En esta clase seguiremos trabajando con las variables aleatorias discretas para las cuales $Im(X) = \{x_1, x_2, x_3, \ldots\}$ [el conjunto de valores que toma X] es finito o infinito numerable.
- En este caso, la distribución de X se especifica dadno las probabilidades

$$p_i = P\{X = x_i\}$$

Notemos las p_i forman un vector de probabilidades (posiblemente infinito)

$$0 \leq p_i \leq 1, \quad \sum_i p_i = 1$$

Recordamos de la clase pasada: Esperanza y Varianza

Para analizar la distribución de una variable aleatoria, frecuentemente resultan útiles dos parámetros

- La esperanza matemática o valor medio: es un parámetro de posición

$$E[X] = \sum_i p_i x_i \text{ siempre que } \sum_i p_i |x_i| < +\infty$$

- La varianza o vararianza: es un parámetro de dispersión

$$\begin{aligned} \text{Var}(X) &= E[(X - E(X))^2] \\ &= E[\varphi(X)] \\ &= \sum_i (x_i - \mu_X)^2 \cdot p_i \end{aligned}$$

donde $\mu_X = E[X]$ y $\varphi(x) = (x - \mu_X)^2$.

$$\sigma_X = \sqrt{\text{Var}(X)}$$

se demomina desviación estándar o desviación típica.

Desigualdades de Tchesbychev y de Markov (1)

Proposición (Desigualdad básica)

Sea X una variable aleatoria no negativa, entonces

$$P(X \geq \lambda) \leq \frac{1}{\lambda} E(X) \tag{1}$$

Demostración.

Sea $A = \{\omega \in \Omega : X(\omega) \geq \lambda\}$. Entonces $X \geq \lambda I_A$, en consecuencia:
$E[X] \geq \lambda E[I_A] = \lambda P(A)$ $\square$

Proposición (Desigualdad de Markov)

Si X es una variable aleatoria (discreta) y $p > 0$ entonces

$$P\{|X| \geq \lambda\} \leq \frac{1}{\lambda^p} E(|X|^p)$$

Demostración.

Si cambiamos X por $|X|^p$ en la desigualdad anterior tenemos que:

$$P\{|X| \geq \lambda\} = P\{|X|^p \geq \lambda^p\} \leq \frac{1}{\lambda^p} E(|X|^p)$$

$\square$

Proposición (desigualdad de Tchebyschev clásica)

Sea X una variable (discreta) entonces

$$P\{|X - E(X)| \geq \lambda\} \leq \frac{Var(X)}{\lambda^2}$$

Demostración.

Usamos la desigualdad anterior con $p = 2$ y cambiamos X por $X - E(X)$. $\square$

Intuitivamente, la desigualdad de Tchebschev dice que la varianza de la variable X nos da una estimación de la probabilidad de que X tome valores alejados de su esperanza. Si $\mathrm{Var}(X)$ es pequeña, entonces es poco probable que X tome un valor alejado de $E(X)$.

Covarianza

Definición

Sean X e Y dos variables aleatorias (discretas). Definimos la **convarianza** *o* **covariancia** *de X e Y por*

$$Cov(X,Y) = E[(X-\mu_X)\cdot(Y-\mu_Y)]$$

donde $\mu_X = E[X]$ y $\mu_Y = E[Y]$.

Digamos que $\text{Im}(X) = \{x_1, x_2, \ldots, x_i, \ldots\}$ e $\text{Im}(Y) = \{y_1, y_2, \ldots, y_j, \ldots\}$. y

$$p_{i,j} = P\{X = x_i, Y = y_j\}$$

denota la distribución conjunta de X e Y entonces:

$$\text{Cov}(X,Y) = E[\varphi(X,Y)] = \sum_{i,j} p_{i,j}(x_i - \mu_X)\cdot(y_j - \mu_Y)$$

donde

$$\varphi(x,y) = (x-\mu_X)\cdot(y-\mu_Y)$$

.

Variables aleatorias con segundo momento finito

Recordamos que la clase pasada introdujimos el espacio vectorial

$$L^2_d = \{X : \Omega \to \mathbb{R} \text{ variable aleatoria discreta con } E(|X|^2) < +\infty\}$$

Sus elementos se llaman variables aleatorias con segundo momento finito.
Notemos que:

$$X \in L^2_d \Rightarrow E(X) \text{ y } \mathrm{Var}(X) \text{ son finitas.}$$

$$\|X\|_2 = E(|X|^2)^{1/2}$$

es una norma en este espacio, que proviene del producto interno

$$\langle X, Y \rangle = E(X \cdot Y)$$

Entonces la desigualdad de Cauchy-Schwarz (aplicada a $X - \mu_X$ y $Y - \mu_Y$) nos da que

$$|Cov(X, Y)| \leq \sqrt{\mathrm{Var}(X)} \cdot \sqrt{\mathrm{Var}(Y)}$$

Varianza de una suma

Proposición

Si X e Y son variables aleatorias (discretas) con segundo momento finito:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Demostración.

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

$$= E[(X-E(X))^2] + E[(Y-E(Y))^2] + 2E[(X-E(X))(Y-E(Y))] =$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

$\square$

Observación: Si X e Y son variables aleatorias independientes entonces Cov(X,Y) $= 0$. Ya que entonces $X - E[X]$ e $Y - E[Y]$ son independientes, luego

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[X - \mu_X] \cdot E[Y - \mu_Y] = 0$$

La recíproca no es cierta, como muestra el siguiente ejemplo:
Ejemplo (Barry James, pag. 130) Sean X e Y dos variables aleatorias con valores -1, 0, 1 con la siguiente función de probabilidad conjunta:

	-1	0	1
-1	$\frac{1}{5}$	0	$\frac{1}{5}$
0	0	$\frac{1}{5}$	0
1	$\frac{1}{5}$	0	$\frac{1}{5}$

entonces $E[XY] = E[X] = E[Y] = 0$, pero X e Y no son independientes pues

$$P\{X = 0, Y = 0\} = \frac{1}{5} \neq \frac{1}{25} = \frac{1}{5} \cdot \frac{1}{5} = P\{X = 0\} \cdot P\{Y = 0\}$$

Variables no correlacionadas

Definición

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias discretas. Diremos que no están correlacionadas si $Cov(X_i, X_j) = 0$ para $i \neq j$.

Corolario

Si X e Y son variables aleatorias independientes, o más generalmente no están correlacionadas

$$Var(X + Y) = Var(X) + Var(Y)$$

Corolario

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias (discretas) con segundo momento finito, que no están correlacionadas, entonces

$$Var(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i)$$

Coeficiente de correlación

Notemos que decir que X e Y no están correlacionadas equivale a decir que $X - \mu_X$ e $Y - \mu_Y$ son ortogonales en L^2_d.

Esto nos sugiere considerar el número

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}$$

que sería geométricamente el coseno del ángulo entre $X - \mu_X$ y $Y - \mu_Y$, para medir que tan lejos están las variables X e Y de estar correlacionadas.

Como vimos antes la desigualdad de Cauchy-Schwarz nos dice que $0 \leq |\rho| \leq 1$, o sea $-1 \leq \rho \leq 1$.

Parte II

La ley débil de los grandes números

Bernoulli, siempre Bernoulli

Volvemos a los ensayos de Bernoulli, donde considerábamos un experimento aleatorio con dos resultados que convencionalmente se llaman

- éxito (1) con probabilidad p.
- fracaso(0) con probabilidad $q = 1 - p$.

Introdujimos las variables aleatorias de Bernoulli $X_i : \Omega^\infty \to \mathbb{R}$ dadas por

$$X_i = \begin{cases} 1 & \text{si la i-ésima realización del experimento es un éxito} \\ 0 & \text{si la i-ésima realización del experimento es un fracaso} \end{cases}$$

Las X_i son variables aleatorias discretas. También lo es el número de éxitos en n ensayos.

$$S_n = X_1 + X_2 + \ldots + X_n$$

Como ya vimos S_n tiene distribución binomial; $S_n \sim Bi(n, p)$

$$P\{S_n = k\} = b(k, n, p) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leq k \leq n$$

Pregunta: ¿Cuál es la varianza de S_n ?

La varianza de S_n

Notamos que las X_i son variables aleatorias independientes, y como calculamos la clase pasada

$$\mathrm{Var}(X_i) = pq$$

Entonces

$$\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = npq$$

Se deuce que si consideramos la frecuencia de éxitos en n ensayos

$$f_n = \frac{S_n}{n}$$

$$\mathrm{Var}(f_n) = \frac{\mathrm{Var}(S_n)}{n^2} = \frac{pq}{n}$$

El teorema de Bernoulli

Imaginemos que realizamos una sucesión ilimitada de ensayos de Bernoulli.. El siguiente teorema debido a Jacques Bernoulli, y publicado en 1713 en su libro *Ars Conjectandi*, formaliza la idea de que $f_n \to p$ cuando $n \to +\infty$.

Teorema (Teorema de J. Bernoulli)

Sea f_n la frecuencia de éxitos en los n primeros ensayos de una sucesión ilimitada de ensayos de Bernoulli. Entonces dado cualquier $\delta > 0$,

$$P\{|f_n - p| > \delta\} \to 0 \text{ conforme } \; n \to \infty$$

$$P\{|f_n - p| > \delta\} = \sum_{k: \left|\frac{k}{n} - p\right| > \delta} b(k, n, p)$$

Demostración del teorema de Bernoulli

Notemos que $E[f_n] = p$. Luego, por la desigualdad de Tchebyschev,

$$P\{|f_n - p| > \delta\} \leq \frac{\mathrm{Var}(f_n)}{\delta^2}$$

pero

$$\mathrm{Var}(f_n) = \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{pq}{n}$$

En consecuencia:

$$P\{|f_n - p| > \delta\} \leq \frac{pq}{n\delta^2} \to 0 \text{ cuando } n \to +\infty \tag{2}$$

Ley débil de los grandes números

Una generalización del teorema de Bernoulli (que se prueba con el mismo argumento) es la siguiente, conocida (al igual que a veces el teorema de Bernoulli) como la ley débil de los grandes números:

Teorema (Ley débil de los grandes números - caso de variancia finita)

Sean $X_1, X_2, \ldots, X_n, \ldots$ *una secuencia infinita de variables aleatorias independientes e idénticamente distribuidas, con*

$$E[X_i] = \mu$$

$$Var(X_i) = \sigma^2 < +\infty$$

Entonces si llamamos

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

y tomamos cualquier $\delta > 0$, *tenemos que*

$$P\{|\overline{X}_n - \mu| > \delta\} \to 0 \text{ cuando } n \to +\infty$$

Demostración

Por linealidad de la esperanza, $E[\overline{X}_n] = \mu$, y por otro lado

$$\text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$$

ya que las X_i son independientes. La desigualdad de Tchebyschev, dice entonces que:

$$P\{|\overline{X}_n - \mu| > \delta\} \leq \frac{\sigma^2}{n\delta^2} \to 0 \text{ cuando } n \to +\infty$$

Algunas observaciones sobre el teorema de Bernoulli

- Si bien la prueba del teorema de Bernoulli, resulta muy sencilla hoy en día, J. Bernoulli dice en su libro que estuvo pensando en este teorema durante más de 20 años, lo cuál muestra que el resultado no es para nada trivial.

- Como todo teorema matemático, el teorema de Bernoulli no afirma nada sobre la realidad, es solamente una afirmación sobre el modelo matemático

  (La cuestión de la validez práctica de un modelo matemático sólo se puede decidir sobre bases empíricas, es decir contrastándolo con la experiencia). Sin embargo, podemos interpretarlo como una muestra de la consistencia interna de nuestro modelo matemático.

- La ley débil de los grandes números recibe este nombre, porque, como veremos más adelante, existe otro teorema conocido como la ley fuerte de los grandes números, que afirma que en realidad $f_n \to p$ (o $\overline{X}_n \to \mu$) con probabilidad 1.

Otra manera de escribir la varianza

Proposición

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

Demostración.

Sea $\mu_X = E[X]$

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu_X)^2] \\
&= E[X^2 - 2\mu_X X + \mu^2] \\
&= E[X^2] - 2\mu_X E[X] + E[\mu_X^2] \\
&= E[X^2] - 2\mu_X^2 + \mu_X^2 \\
&= E[X^2] - \mu_X^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

$\square$

Parte III

El método de las funciones generatrices

Series de Potencias

Una serie de potencias es una expresión de la forma

$$\sum_{k=0}^{\infty} a_k z^k$$

Se pude considerar $z \in \mathbb{R}$ o $z \in \mathbb{C}$. Se prueba que existe un radio de convergencia $r \in [0, +\infty]$ tal que la serie converge en un intervalo/disco $|z| < r$, y diverge si $|z| > r$.

Muchas funciones admiten desarrollos en series de potencias. Ejemplos que vamos a usar:

$$\frac{1}{1-z} = \sum_{k=0}^{\infty} z^k \text{ si } |z| < 1$$

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

El método de las funciones generatrices

Definición

Sea $X : \Omega \to \mathbb{N}_0$ una variable aleatoria que toma valores enteros. Llamamos función generatriz de la distribución de probabilidades de X a

$$g_X(z) = \sum_{k=0}^{\infty} P\{X = k\} \cdot z^k \quad (z \in \mathbb{C})$$

suponiendo que esta serie tenga un radio de convergencia $r_X > 0$ (entonces convergerá absolutamente en $|z| < r_X$).
Notemos que si $0 < |z| < r_X$,

$$g_X(z) = E[z^X]$$

(Cuando $z = 0$ esta fórmula es problemática si X toma el valor 0, definimos $0^0 = 1$. Se tiene que $g_X(0) = P\{X = 0\}$)

Observación: En virtud de la unicidad del desarrollo en serie de potencias, la distribución de probabilidades de una variable aleatoria entera está unívocamente determinada por su función generatriz.

Ejemplos

Los siguientes ejemplos muestran que estamos generalizando las ideas que usamos la clase pasada para calcular esperanzas:

- Si S_n es el número de éxitos en n ensayos de Bernoulli (distribución binomial $\text{Bi}(n, p)$), tenemos

$$g_{S_n}(z) = \sum_{k=0}^{n} \binom{n}{k} z^k p^k q^{n-k} = (q + zp)^n$$

- Si T es el tiempo de espera del primer éxito en los ensayos de Bernoulli (distribución geométrica $\text{Ge}(p)$)

$$g_T(z) = \sum_{k=1}^{\infty} q^{k-1} p z^k = \sum_{j=0}^{\infty} q^j p z^{j+1} = \frac{pz}{1 - qz} \text{ si } |z| < r_T = \frac{1}{q}$$

Funciones generatrices y esperanza

Proposición

Si la serie que define la función generatriz g_X tiene radio de convergencia $r_X > 1$, entonces

$$E(X) = g_X'(1)$$

$$Var(X) = g_X''(1) + g_X'(1) - g_X'(1)^2$$

Demostración

Como las series de potencia pueden derivarse término a término en el interior de su disco de convergencia, tenemos que:

$$g_X'(z) = \sum_{k=1}^{\infty} kP\{X = k\} z^{k-1}$$

con convergencia absoluta si $|z| < r_X$. En particular si $z = 1$,

$$g_X'(1) = \sum_{k=1}^{\infty} kP\{X = k\} = E[X]$$

Volviendo a derivar tenemos que

$$g_X''(z) = \sum_{k=2}^{\infty} k(k-1)P\{X = k\} z^{k-2}$$

con convergencia absoluta si $|z| < r_X$, y haciendo $z = 1$,

$$g_X''(1) = \sum_{k=2}^{\infty} k(k-1)P\{X = k\} = E[X(X-1)] = E[X^2] - E[X]$$

Demostración (2)

Luego

$$Var(X) = E[X^2] - E[X]^2 = g_X''(1) + g_X'(1) - g_X'(1)^2$$

Un ejemplo: calcular la esperanza y la varianza de la distribución geométrica

$$T \sim \text{Ge}(p) \Rightarrow g_T(z) = \frac{pz}{1-qz} \Rightarrow g_T'(z) = \frac{p}{1-qz} + \frac{pqz}{(1-qz)^2}$$

Como vimos en la clase anterior, esto nos da que

$$E[T] = g_T'(1) = \frac{p}{1-q} + \frac{pq}{(1-q)^2} = 1 + \frac{q}{p} = 1 + \frac{1-p}{p} = \frac{1}{p}$$

Derivando una vez más

$$g_T''(z) = -\frac{2\,pq^2z}{(qz-1)^3} + \frac{2\,pq}{(qz-1)^2}$$

Entonces (¡hice las cuentas usando Sagemath!)

$$g_T''(1) = \frac{2\,(p-1)^2}{p^2} - \frac{2\,(p-1)}{p}$$

$$\text{Var}(T) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Funciones generatrices e independencia

Proposición

Si X e Y son variables aleatorias independientes, entonces:

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z)$$

para $|z| < \min(r_X, r_Y)$.

Demostración.

Como X e Y son independientes, z^X y z^Y son independientes. En consecuencia, si $0 < |z| < r_X$:

$$g_{X+Y}(z) = E[z^{X+Y}] = E[z^X \cdot z^Y] = E[z^X] \cdot E[z^Y] = g_X(z) \cdot g_Y(z)$$

Cuando $z = 0$,

$$g_{X+Y}(0) = P\{X + Y = 0\} = P\{X = 0, Y = 0\}$$

$$= P\{X = 0\} \cdot P\{Y = 0\} = g_X(0) \cdot g_Y(0)$$

La distribución de Pascal

Todavía en el esquema de ensayos de Bernoulli, llamemos T_r a la cantidad de ensayos que hay que realizar para tener r éxitos (de modo que $T = T_1$).

$$T_r = \min\{i > T_{r-1} \ : X_i = 1\}$$

donde las (X_i) son las variables de Bernoulli y $T_0 = 0$.
Notamos que

$$T_r = G_1 + G_2 + \ldots + G_r$$

donde G_i es cuántos ensayos debo realizar después del éxito $i - 1$ para obtener el siguiente. Claramente los G_i son independientes, y tienen distribución geométrica con parámetro p. entonces

$$g_{T_r}(z) = g_{G_1}(z) \cdot g_{G_2}(z) \cdots g_{G_r}(z) = \left(\frac{pz}{1 - qz}\right)^r = p^r z^r (1 - qz)^{-r}$$

siempre que $|z| < \frac{1}{q}$. A partir de esto podemos calcular la distribución de T_k.

El verdadero teorema del binomio de Newton

En los cursos de análisis, se ve el (verdadero) teorema del binomio de Newton, donde el exponente puede ser un número real cualquiera.

Teorema

Para todo α real, y x real con $|x| < 1$,

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = \lim_{n\to\infty} \sum_{k=0}^{n} \binom{\alpha}{k} x^k$$

donde los coeficientes binomiales generalizados se definen por

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!}$$

Para más información (y ejemplos), pueden ver `https://en.wikipedia.org/wiki/Binomial_series`.

La distribución de Pascal(2)

En nuestro caso,

$$g_{T_r}(z) = p^r z^r (1-qz)^{-r} = p^r z^r \sum_{j=0}^{\infty} \binom{-r}{j} (-qz)^j$$

$$= \sum_{j=0}^{\infty} \binom{-r}{j} p^r (-q)^j z^{j+r}$$

$$= \sum_{k=r}^{\infty} \binom{-r}{k-r} p^r (-q)^{k-r} z^k$$

Por lo que la unicidad del desarrollo en serie de potencias nos da

$$P\{T_r = k\} = \binom{-r}{k-r} p^r (-q)^{k-r}$$

La distribución de Pascal (3)

Ahora notemos que cuando $\alpha = -r$

$$\binom{\alpha}{j} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-j+1)}{j!}$$
$$= \frac{(-r)(-r-1)(-r-2)\cdots(-r-j+1)}{j!}$$
$$= (-1)^j \frac{r(r+1)(r+2)\cdots(r+j-1)}{j!} = (-1)^j \binom{r+j-1}{j}$$

Entonces concluimos que

$$P\{T_r = k\} = \binom{-r}{k-r} p^r (-q)^j$$
$$= \binom{k-1}{k-r} p^r q^{k-r}$$

La distribución de Pascal (4)

Finalmente usando la propiedad de simetría de los números combinatorios

$$\binom{n}{m} = \binom{n}{n-m} \quad 0 \leq m \leq n$$

$$\begin{aligned} P\{T_r = k\} &= \binom{k-1}{k-r} p^r q^{k-r} \\ &= \binom{k-1}{(k-1)-(k-r)} p^r q^{k-r} \\ &= \binom{k-1}{r-1} p^r q^{k-r} \end{aligned}$$

o sea

Dirstribución de Pascal

$$P\{T_r = k\} = \binom{k-1}{r-1} p^r q^{k-r} \quad \text{para } k \geq r$$

La esperanza y la varianza de la distribución de Pascal

Volviendo a la escritura

$$T_r = G_1 + G_2 + \ldots + G_r$$

vemos que

$$E[T_r] = E[G_1] + E[G_2] + \ldots + E[G_r] = r \cdot \frac{1}{p} = \frac{r}{p}$$

y como las G_i eran independientes

$$\text{Var}(T_r) = \text{Var}[G_1] + \text{Var}[G_2] + \ldots + \text{Var}[G_r] = r \cdot \frac{q}{p^2}$$