Variables Aleatorias Discretas: Esperanza

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

Variables Aleatorias Discretas

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 $X : \Omega \to \overline{\mathbb{R}}$.
- Por razones técnicas, hay que pedir que $X^{-1}(I)$ sea un evento para cada intervalo $I \subset \overline{\mathbb{R}}$, o sea que podamos calcular $P\{X \in I\} = P(X^{-1}(I))$.
- Hay distintos tipos de variables aleatorias. En esta clase focalizaremos en 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)

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

Un ejemplo

Tiramos dos dados. Nuestro espacio muestral es:

$$\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in D\}$$

donde $D = \{1, 2, 3, 4, 5, 6\}$. Consideramos la suma S de los puntos obbtenidos

$$S : \Omega \to \mathbb{N}, \quad S(\omega) = \omega_1 + \omega_2$$

Es una variable aleatoria discreta. ¿Cuál es la distribución de probabilidades de S?

x_i	p_i
2	$1/36 \quad = 0{,}028$
3	$2/36 \quad = 0{,}056$
4	$3/36 \quad = 0{,}083$
5	$4/36 \quad = 0{,}111$
6	$5/36 \quad = 0{,}139$
7	$6/36 \quad = 0{,}167$
8	$5/36 \quad = 0{,}139$
9	$4/36 \quad = 0{,}111$
10	$3/36 \quad = 0{,}083$
11	$2/36 \quad = 0{,}056$
12	$1/36 \quad = 0{,}028$

Parte II

La esperanza matemática

Esperanza matemática

Definición

Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria discreta. Diremos que X es integrable (o que tiene esperanza finita) si la serie

$$\sum_i p_i x_i$$

es absolutamente convergente, es decir si:

$$\sum_i p_i |x_i| < +\infty$$

En este caso definimos, la esperanza de X (también llamada valor medio) o valor esperado) como el valor de dicha suma.

$$E[X] = \sum_i p_i x_i$$

Nota: si $X : \Omega \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, entonces pedimos que X sea finita con probabilidad 1, o equivalentemente

Esperanza de variable aleatorias con imagen finita

- Una variable aleatoria cuya imagen es finita siempre es integrable, ya que la suma
$$E[X] = \sum_i p_i x_i$$
es finita en este caso.

Ejemplo Supongamos que arrojamos un dado ¿cuál es la esperanza del valor obtenido X ?

$$E[X] = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3,5$$

Ley de los grandes números (idea intuitiva)

Supongamos que tenemos una variable aleatoria discreta X y repetimos el experimento aleatorio asociado a X una cantidad de veces n (en condiciones independientes). Llamemos $X_1, X_2, \ldots X_n$ a los valores obtenidos para nuestra variable y consideramos el promedio de dichos valores

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

Llamemos n_i a la cantidad de veces que obtenemos uno de los valores x_i posibles para x_i, y $f_i = n_i/n$ a la correspondiente frecuencia.
Entonces

$$S_n = X_1 + X_2 + \ldots X_n = \sum_i n_i \cdot x_i$$

y en consecuencia

$$\overline{X}_n = \sum_i f_i \cdot x_i$$

Ahora si n es grande, esperamos que $f_i \approx p_i$ (ley de los grandes números) y por lo tanto que $\overline{X}_n \approx E[X]$. Matemáticamente nos gustaría decir que $\overline{X}_n \to E[X]$ cuando $n \to \infty$ en algún sentido. Formalizamos esto más adelante.

Un ejemplo: La ruleta

Supongamos que jugamos un peso a la ruleta y apostamos a un color (por ej. negro). Sea X nuestra ganancia (o pérdida) ¿cuánto debemos esperar ganar (o perder) ?
Aquí

$$X = \begin{cases} 1 & \text{si sale negro} \quad (\text{con probabilidad } \frac{18}{37}) \\ -1 & \text{si sale rojo o cero} \quad (\text{con probabilidad } \frac{19}{37}) \end{cases}$$

En consecuencia:

$$E[X] = \frac{18}{37} - \frac{19}{37} = \frac{-1}{37} = -0,027\ldots$$

Así pues, al jugar a la ruleta, debemos esperar perder un 27 por mil.

Variable indicadora de un evento

Sea A un evento, consideramos la función $I_A : \Omega \to \mathbb{R}$ definida por

$$I_A(\omega) = \begin{cases} 1 & \text{si} \quad \omega \in A \\ 0 & \text{si} \quad \omega \notin A \end{cases}$$

Intuitivamente I_A vale 1 cuando el evento A ocurre, y 0 sino. Se denomina el indicador del evento A. (En la teoría de la medida, esta función se llama la función característica del conjunto A y se suele denotar por χ_A, pero en la teoría de probabilidades la expresión "función característica" tiene un significado diferente).

I_A es una variable aleatoria discreta pues su imagen consta de dos valores (0 y 1) y sus pre-imágenes son $X^{-1}(0) = \Omega - A$ y $X^{-1}(1) = A$, que son eventos.

La esperanza de I_A es:

$$E[I_A] = 0 \cdot P(\Omega - A) + 1 \cdot P(A) = P(A)$$

Es decir, la esperanza del indicador de un evento, coincide con su probabilidad.

Esperanzas en la computadora

Consideremos una variable aleatoria discreta con 3 valores $1, 2, 3$ tal que

$$P\{X = 1\} = \frac{1}{2}, P\{X = 2\} = P\{X = 3\} = \frac{1}{4}$$

$$E[X] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{7}{4} = 1{,}75$$

Calculando la esperanza de una disstribución de probabilidad discreta

```python
import scipy.stats
xk = (1, 2, 3)
pk = (0.5, 0.25, 0.25)
distribucion = scipy.stats.rv_discrete(values=(xk, pk))
print(distribucion.mean())
```

Un ejemplo de una variable aleatoria que toma infinitos valores y que puede valor infinito

En la clase pasada, consideramos en el esquema de infinitos ensayos de Bernoulli cuántos ensayos debemos realizar hasta obtener el primer éxito. Vimos que esto define una variable aleatoria $T : \Omega^\infty \to \mathbb{N} \cup \{+\infty\}$ con la distribución geométrica de parámetro p.

$$P\{T = k\} = q^{k-1} \cdot p$$

$$P\{T = +\infty\} = 0$$

donde recordamos que $q = 1 - p$.
Por ejemplo: tiramos infinita veces una moneda (con $p = q = 1/2$) ¿cuánto debemos esperar hasta que salga la primer cara?
¿Cuál es la esperanza de T?

$$E[T] = \sum_{k=1}^{\infty} k\, q^{k-1} \cdot p$$

Calculando la esperanza de la distribución geométrica

Para sumar esta serie, debemos recordar que la distribución geométrica se llama así porque está emparentada con la serie geométrica

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \quad |x| < 1$$

Si derivamos esta serie, tenemos:

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1} \quad |x| < 1$$

Entonces:

$$E[T] = p \cdot \sum_{k=1}^{\infty} k\ q^{k-1} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Para su tabla de esperanzas

$$T \sim \mathrm{Ge}(p) \Rightarrow E[T] = \frac{1}{p}$$

Esperanzas infinitas

A veces resulta conveniente admitir esperanzas infinitas. Si $X \geq 0$ diremos que $E[X] = +\infty$ si

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

diverge.
Si X es una variable aleatoria discreta cualquiera, escribimos

$$X = X^+ - X^-$$

donde

$$X^+ = \begin{cases} X & \text{si} \quad X \geq 0 \\ 0 & \text{si} \quad X < 0 \end{cases}$$

y

$$X^- = \begin{cases} -X & \text{si} \quad X < 0 \\ 0 & \text{si} \quad X \geq 0 \end{cases}$$

Notamos que X^+ y X^- son variables aleatorias no negativas.
Decimos que $E[X] = +\infty$ si $E[X^+] = +\infty$ y $E[X^-] < \infty$. Similarmente diremos que $E[X] = -\infty$ si $E[X^-] = +\infty$ y $E[X^+] < \infty$. Si $E[X^+]$ y $E[X^-]$ son ambas infinitas, $E[X]$ no está definida.

Linealidad de la esperanza

Proposición (linealidad de la esperanza)

1. *Si* $X, Y : \Omega \to \mathbb{R}$ *son variables aleatorias discretas con esperanza finita, entonces*
$$E[X+Y] = E[X] + E[Y]$$
2. *Si* $X : \Omega \to \mathbb{R}$ *es una variable aleatoria discreta con esperanza finita, entonces:*
$$E[\lambda X] = \lambda E[X]$$

Demostración

Sean (x_i) los valores que toma X, e (y_j) los valores que toma Y: entonces

$$E[X] = \sum_i x_i P\{X = x_i\} = \sum_{i,j} x_i P\{X = x_i, Y = y_j\}$$

ya que

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\} \text{ (unión disjunta)}$$

y el reordenamiento de la serie está justificado por la convergencia absoluta, de la serie:

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

Similarmente,

$$E[Y] = \sum_j y_j P\{Y = y_j\} = \sum_{i,j} y_j P\{X = x_i, Y = y_j\}$$

$$E[X] + E[Y] = \sum_{i,j} (x_i + y_j) P\{X = x_i, Y = y_j\}$$

$p_{ij} = P\{X = x_i, Y = y_j\}$ se llaman la distribución conjunta de X e Y.

Demostración (2)

Sea $Z = X + Y$ y sean $z_1, z_2, \ldots, z_k, \ldots$ los valores de Z. Entonces los z_k son exactamente los valores $x_i + y_j$ (pero estos últimos pueden repetirse). Entonces,

$$E[Z] = \sum_k z_k P\{Z = z_k\} = \sum_k \sum_{i,j: x_i + y_j = z_k} z_k P\{X = x_i, Y = y_j\}$$

pues

$$\{Z = z_k\} = \bigcup_{i,j: x_i + y_j = z_k} \{X = x_i, Y = y_j\} \text{ (unión disjunta)}$$

Deducimos que

$$E[Z] = \sum_k (x_i + y_j) P\{X = x_i, Y = y_j\} = E[X] + E[Y]$$

Esto completa la prueba de la primera afirmación. En cuanto a la segunda afirmación, λX es una variable aleatoria discreta que toma los valores λx_i, por lo tanto:

$$E[\lambda X] = \sum_i \lambda x_i P\{\lambda X = \lambda x_i\} = \lambda \sum_i x_i P\{X = x_i\} = \lambda E[X]$$

Un ejemplo que ya vimos: el esquema de ensayos de Bernouli

Consideramos 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; $S_n \sim Bi(n, p)$

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

¿Cómo calcular la esperanza de la distribución binomial?: Método 1

Una forma de calcular la esperanza de S_n sería usar directamente la definición

$$E[S_n] = \sum_{k=0}^{n} kP\{S_n = k\} = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

Para hacer esto, usamos un truco parecido al que usamos para la distribución geométrica:

$$(px + q)^n = \sum_{k=0}^{n} \binom{n}{k} x^k p^k q^{n-k}$$

derivando

$$np(px + q)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} kx^{k-1} p^k q^{n-k}$$

Especializando en $x = 1$, como $p + q = 1$, vemos que

$$E[S_n] = np$$

¿Cómo calcular la esperanza de la distribución binomial?: Método 2

Como

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

donde las X_i son las variables aleatorias de Bernoulli, entonces por la linealidad de la esperanza

$$E[S_n] = E[X_1] + E[X_2] + \ldots + E[X_n] = np$$

Para su tabla de esperanzas

$$S_n \sim \mathrm{Bi}(n, p) \Rightarrow E[S_n] = np$$

Fíjense que si $f_n = \frac{S_n}{n}$ representa la frecuencia de éxitos en n ensayos, entonces

$$E[f_n] = p$$

Monotonía de la esperanza

Proposición (Monotonía de la esperanza)

1. *Si X es una variable aleatoria con esperanza finita y $X \geq 0$ con probabilidad 1, entonces $E[X] \geq 0$.*
2. *Sean X e Y variables aleatorias con esperanza finita. Entonces, si $X \leq Y$ con probabilidad 1, tenemos que $E[X] \leq E[Y]$*
3. *Si X es una variable aleatoria acotada, entonces:*

$$\inf_{\Omega} X \leq E[X] \leq \sup_{\Omega} X$$

.

4. *Si X es una variable aleatoria discreta con esperanza finita, entonces:*

$$|E[X]| \leq E[|X|]$$

Esperanza de una función de una variable aleatoria

Proposición

Sean X una variable aleatoria discreta y $\varphi : \mathbb{R} \to \mathbb{R}$. Entonces

$$E[\varphi(X)] = \sum_i \varphi(x_i) P\{X = x_i\}$$

siempre que esta serie sea absolutamente convergente.

Demostración.

Sea $Y = \varphi(X)$, y sean (y_j) los valores de Y, entonces:

$$\begin{aligned} E[Y] &= \sum_j y_j P\{Y = y_j\} = \sum_j y_j \sum_{i:\varphi(x_i)=y_j} P\{X = x_i, Y = y_j\} \\ &= \sum_i \varphi(x_i) P\{X = x_i\} \end{aligned}$$

(El reordenamiento se justifica usando la convergencia absoluta de la serie.) □

Un ejemplo

Supongamos que $S_n \sim \mathrm{Bi}(n,p)$ con $n \geq 2$,y que queremos calcular $E[S_n^2]$. Según la fórmula anterior con $\varphi(x) = x^2$

$$E[S_n^2] = \sum_{k=0}^{n} k^2 P\{S_n = k\} = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k}.$$

Recordamos que

$$(px+q)^n = \sum_{k=0}^{n} \binom{n}{k} x^k p^k q^{n-k}$$

Derivando dos veces

$$n(n-1)p(px+q)^{n-2} = \sum_{k=2}^{n} \binom{n}{k} k(k-1) x^{k-2} p^k q^{n-k}$$

y especializando en $x = 1$, tenemos eligiendo $\varphi(x) = x(x-1)$

$$n(n-1)p = \sum_{k=2}^{n} \binom{n}{k} k(k-1) p^k q^{n-k} = E[S_n(S_n - 1)]$$

Un ejemplo (2)

pero por la linealidad de la esperanza

$$E[S_n(S_n - 1)] = E[S_n^2 - S_n] = E[S_n^2] - E(S_n)$$

y ya vimos que $E[S_n] = np$ entonces ,

$$E[S_n^2] = E[S_n(S_n - 1)] + E[S_n] = n(n-1)p + np = n^2 p$$

Vectores aleatorios

Esta propiedad se puede generalizar a funciones de **vectores aleatorios**. Este concepto es una generalización natural del de variable aleatoria discreta:

Definición

Un vector aleatorio discreto n-dimensional es una función $X : \Omega \to \mathbb{R}^n$ *tal que* $Im(X)$ *sea finita o infinita numerable, y* $P\{X = x\}$ *sea un evento para todo* $x \in \mathbb{R}^n$. *Dar un vector aleatorio discreto* $X = (X_1, X_2, \ldots, X_n)$ *es equivalente a dar n variables aleatorias discretas* $x_1, x_2, \ldots, x_n$

Con esta terminología tenemos [con la misma demostración de antes]:

Proposición

Sean X *un vector aleatorio n-dimensional y* $\varphi : \mathbb{R}^n \to \mathbb{R}$, *entonces*

$$E[\varphi(X)] = \sum_i \varphi(x_i) P\{X = x_i\}$$

donde x_i *recorre la imagen de* X, *siempre que esta serie sea absolutamente convergente.*

Variables Aleatorias discretas independientes

Definición

Sean X e Y dos variables aleatorias <u>*discretas*</u> *definidas en un mismo espacio muestral. Diremos que son* **independientes**, *si para cada x_i, y_j los eventos $\{X = x_i\}$ e $\{Y = y_j\}$ son independientes, es decir de acuerdo a la definición de eventos independientes si,*

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

Observación: Remarcamos que esta definición solamente se aplica a variables discretas, cuando generalicemos esta noción a variables aleatorias no discretas, nos veremos en la necesidad de adoptar una definición diferente.

Funciones de variables independientes

Proposición

Si X e Y son variables aleatorias discretas independientes, y $f, g : \mathbb{R} \to \mathbb{R}$ son funciones, entonces $Z = f(X)$ y $W = g(Y)$ también son variables aleatorias discretas independientes.

Demostración.

Calculemos la distribución conjunta de Z y W:

$$P\{Z = z, W = w\} = \sum_{x,y: f(x)=z, g(y)=w} P\{X = x, Y = y\}$$

$$= \sum_{x,y: f(x)=z, g(y)=w} P\{X = x\} P\{Y = y\}$$

$$= \left(\sum_{x: f(x)=z} P\{X = x\} \right) \left(\sum_{y: g(y)=w} P\{Y = y\} \right) = P\{Z = z\} P\{W = w\}$$

$\square$

Esperanza e independencia

Proposición

Si X e Y son variables aleatorias discretas independientes con esperanza finita, entonces:

$$E(XY) = E(X)E(Y)$$

Demostración.

$$E[XY] = \sum_{i,j} x_i y_i P\{X = x_i, Y = y_j\} = \sum_{i,j} x_i y_i P\{X = x_i\} P\{Y = y_j\}$$

$$\left(\sum_i x_i P\{X = x_i\} \right) \left(\sum_j y_j P\{Y = y_j\} \right) = E[X]E[Y]$$

Observación: En el caso en que X e Y toman infinitos valores, la aplicación de la propiedad distributiva, está justificada por el hecho de que las series que intervienen son absolutamente convergentes, por hipótesis. $\square$

Un ejemplo

En el esquema de ensayos de Bernoulli las variables X_i las variables aleatorias de Bernoulli $X_i : \Omega^\infty \to \mathbb{R}$ dadas por

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

son variables independientes. Entonces

$$E[X_i X_j] = E[X_i]E[X_j] = p \cdot p = p^2$$

Funciones convexas

Definición

Sea $I = [a, b] \subset \overline{\mathbb{R}}$ *un intervalo y* $f : I \to \mathbb{R}$ *una función. Diremos que* f *es convexa, si dados* $x_1, x_2 \in I$ *y* $\alpha \in [0, 1]$, *se verifica que:*

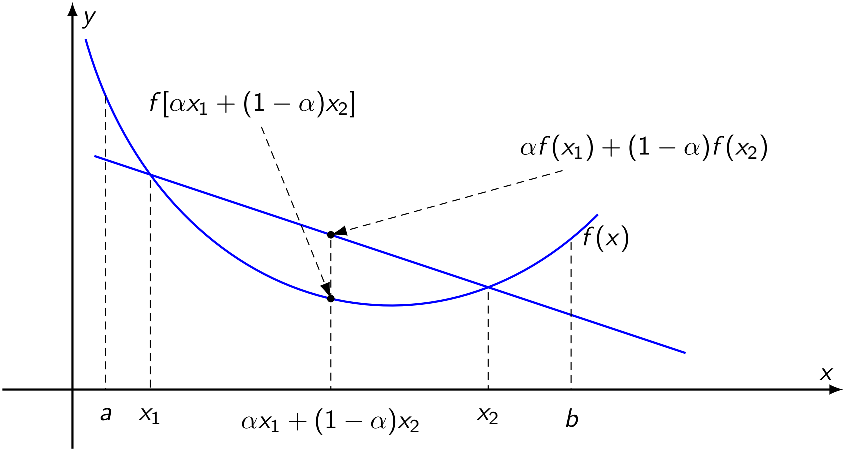
$$f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)$$

Observación: *Si* f *es de clase* C^2, *entonces* f *es convexa, si y sólo si* $f''(x) \geq 0$.

Observación: *Una función convexa en* $\mathbb{R}$ *es necesariamente continua. Además es posible probar que su derivada* $f'(x)$ *existe salvo quizás para un conjunto a lo sumo numerable de valores de* x, *y que* f' *es creciente.*

Ejemplos: $f(x) = |x|^p$si $p \geq 1, f(x) = e^x, f(x) = |x|$

Funciones convexas: Gráfico de la definición



Combinaciones convexas

Definición

Una **combinación convexa** *de los* x_i *es una combinación lineal*

$$\sum_{i=1}^{n} \alpha_i x_i$$

en la que $0 \leq \alpha_i$ *y* $\sum_{i=1}^{n} \alpha_i = 1$.

Ejercicio (de inducción)

Si $f : \mathbb{R} \to \mathbb{R}$ es una función convexa y $\sum_{i=1}^{n} \alpha_i x_i$ es una combinación convexa, entonces:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

Desigualdad de Jensen

Proposición (Desigualdad de Jensen)

Si $g : \mathbb{R} \to \mathbb{R}$ *es una función convexa, entonces:*

$$g(E[X]) \leq E[g(X)])$$

en los siguientes casos: si X *es no negativa y* $g(x) \geq 0$ *para* $x \geq 0$, *o si* X *y* g *son arbitrarias y* $E(|g(X)|) < \infty$.

Demostración

Hagamos la demostración primero, en el caso que X toma sólo finitos valores. Sea $p_i = P\{X = x_i\}$. Entonces

$$E[X] = \sum_{i=1}^{n} p_i x_i$$

es una combinación convexa de los valores de X. Como X es una función convexa,

$$g(E[X]) = g\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i g(x_i) = E[g(X)]$$

Si X toma un número numerable de valores, x_i con probabilidades p_i, entonces hacemos lo siguiente: para cada $n \in \mathbb{N}$ definamos,

$$s_n = \sum_{i=1}^{n} p_i$$

y notamos que

$$\sum_{i=1}^{n} \frac{p_i}{s_n} x_i$$

es una combinación convexa.

Demostración (2)

Entonces, como g es convexa:

$$g\left(\sum_{i=1}^{n} \frac{p_i}{s_n} x_i\right) \leq \sum_{i=1}^{n} \frac{p_i}{s_n} g(x_i)$$

Cuando $n \to +\infty$, tenemos que $s_n \to 1$. Entonces, utilizando la continuidad de g, obtenemos que:

$$g(E[X]) = g\left(\sum_{i=1}^{\infty} p_i x_i\right) \leq \sum_{i=1}^{\infty} p_i g(x_i) = E[g(X)]$$

Parte III

Momentos y Varianza

Momentos

Definición

Sea X una variable aleatoria (discreta). Definimos el k-ésimo momento de X entorno de b como $E[(X-b)^k]$. El k-ésimo momento absoluto entorno de b se define como $E[|X-b|^k]$.

Ejemplo: Antes calculamos que si n es el número de éxitos en n ensayo de Bernulli $E[S_n^2] = E[|S_n|^2] = n^2 p$. Esto da su segundo momento en torno al origen.

Algunas observaciones:

1. Si $E[|X|^t] < \infty$ y $0 \leq s \leq t$, entonces $E[|X|^s] < +\infty$. En efecto según la desigualdad de Jensen,
$$(E[|X|^s])^p \leq E[|X|^t]$$
donde $p = \frac{t}{s} \geq 1$. Es más, vemos que:
2. $E[|X|^p]^{1/p}$ es una función creciente de p.
3. Si $E[|X|^p] < +\infty$ y $E[|Y|^p] < +\infty$ entonces $E[|X+Y|^p]^{1/p} < +\infty$

Demostración.

$$|X+Y|^p \leq (|X| + |Y|)^p \leq (2 \max |X|, |Y|)^p$$
$$\leq 2^p \max(|X|^p, |Y|^p) \leq 2^p(|X|^p + |Y|^p)$$

Por lo tanto,
$$E[|X+Y|^p \leq 2^p(E[|X|^p] + E[|Y|^p]) < +\infty$$

$\square$

Algunas observaciones (2)

1. En consecuencia, el conjunto

$$L_d^p(\Omega,\mathcal{E},P) = \{X : \Omega \to \overline{\mathbb{R}} \text{ variable aleatoria discreta} : E[|X|^p] < +\infty\}$$

   (siendo $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$) es un espacio vectorial.

2. Si $p \geq 1$, es posible probar que

$$\|X\|_p = E[|X|^p]^{\frac{1}{p}}$$

   es una norma en dicho espacio.

En lo sucesivo, nos van a interesar especialmente dos clases L^p:

$$L_d^1(\Omega) = \{X : \Omega \to \overline{\mathbb{R}} : \text{ variable aleatoria (discreta) con esperanza finita}\}$$

$$L_d^2(\Omega) = \{X : \Omega \to \overline{\mathbb{R}} : \text{ variable aleatoria (discreta) con segundo momento finito}\}$$

Un ejemplo

Notemos que $L^2_d \subset L^1_d$ por lo anterior. Veamos un ejemplo de una variable aleatoria que está en L^1_d pero no en L^2_d: Consideramos un espacio muestral numerable

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$$

en el que

$$P\{\omega_n\} = \frac{1}{n(n+1)}$$

Verifiquemos que esta asignación efectivamente define una distribución de probabilidades en Ω:

$$\sum_{n=1}^{\infty} P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \sum_{n=1}^{\infty} \left[\frac{1}{n} - \frac{1}{n+1}\right] = 1$$

(serie telescópica). Definamos la variable aleatoria $X : \Omega \to \mathbb{R}$, dada por $X(\omega_n) = \sqrt{n}$. Entonces,

$$E(X) = \sum_{n=1} X(\omega_n) P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{\sqrt{n}}{n(n+1)} \leq \sum_{n=1}^{\infty} \frac{1}{n^{3/2}} < +\infty$$

Un ejemplo (2)

pero

$$E(X^2) = \sum_{n=1} X(\omega_n)^2 P\{\omega_n\} = \sum_{n=1}^{\infty} \frac{n}{n(n+1)} = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty$$

Varianza

Definición

El segundo momento de X entorno de su media se llama la **varianza** *(o* **variancia**[a]*) de X, es decir:*

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

Por lo anterior $Var(X) < +\infty$ si y sólo si el segundo momento de X es finito, es decir si $X \in L^2_d$.

[a] Según el diccionario de la RAE, ambas grafías son aceptables.

Como veremos la próxima clase, la varianza mide cuánto puede alejarse X de su esperanza o valor medio. Si $var(X)$ es pequeña significa que la distribución de X está muy concentrada cerca de su esperanza. Es un parámetro de dispersión.

Varianza de una indicadora

Sea A un evento con probabilidad p, e I_A su indicador. Calculemos su varianza. Ya vimos que:

$$E[I_A] = P(A) = p$$

En consecuencia:

$$\text{Var}(I_A) = E[(I_A - p)^2]$$

La distribución de probabilidades de $(I_A - p)^2$ es:

$$(I_A - p)^2 = \begin{cases} (1-p)^2 & \text{si ocurre } A \quad \text{(con probabilidad } p\text{)} \\ p^2 & \text{si no ocurre } A \quad \text{(con probabilidad } q = 1-p\text{)} \end{cases}$$

En consecuencia,

$$\text{Var}(I_A) = (1-p)^2 p + p^2(1-p) = p - p^2 = pq$$

Propiedades de la varianza

Proposición

1. *Si* $X = c$ *es constante, entonces* $Var(X) = 0$.
2. $Var(aX + b) = a^2 Var(X)$.

Demostración.

Si $X = c$ es constante, $E[X] = c$ luego $Var(X) = E[0] = 0$.

$$E[aX + b] = aE[X] + E[b] = a \cdot E[X] + b$$

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

$\square$