CO3121 - Introducción al Cálculo de Probabilidades (Semanas 4, 5 y 6)

Raúl Jiménez, Haydee Lugo y Adolfo Quiroz
Departamento de Cómputo Científico y Estadística
Universidad Simón Bolívar

1 Variables Aleatorias Discretas

Dado un espacio de probabilidad (Ω, F, P), estamos interesados en situaciones que envuelven alguna función $X : \Omega \to R$. Por ejemplo, consideremos el experimento de lanzar un dado, $\Omega = \{1, 2, 3, 4, 5, 6\}$, y supongamos que apostamos al resultado del experimento de tal manera que nuestra ganancia es:

$$-1 \text{ si el resultado es } 1, 2 \text{ o } 3$$

$$0 \text{ si el resultado es } 4$$

$$2 \text{ si el resultado es } 5 \text{ o } 6$$

donde ganancias negativas son pérdidas positivas. Si el resultado es ω la ganancia es $X(\omega)$ donde $X : \Omega \to R$ y esta definida por:

$$X(1) = X(2) = X(3) = -1$$

$$X(4) = 0$$

$$X(5) = X(6) = 2$$

La aplicación X es un ejemplo de variable aleatoria discreta.

Definición. Una **variable aleatoria discreta** X en el espacio de probabilidad (Ω, F, P) es una función $X : \Omega \to R$ tal que la imagen Ω bajo X, $ImX = X(\Omega)$, es un subconjunto contable de R.

El término discreto se refiere a la condición de que X toma solamente valores en un subconjunto contable de números reales.

Función de masa. La probabilidad o función de masa de la variable aleatoria discreta X es la función $P_X : R \to [0, 1]$ definida por

$$P_X(x) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

Es decir, $P_X(x)$ es la probabilidad de que X tome el valor x. Por simplificar la notación, comunmente se usa la identidad

$$P(X \in F) = P(\{\omega \in \Omega : X(\omega) \in F\}).$$

En particular, $P_X(x) = P(X = x)$. Por definición, ImX es un conjunto contable para cualquier variable aleatoria discreta X. Así que

$$P(X = x) = 0 \quad si \quad x \notin ImX.$$

Además,

$$\sum P_X(x) = P(\bigcup_{x \in ImX} \{\omega \in \Omega : X(\omega) = x\}) = P(\Omega) = 1$$

y escribimos

$$\sum P_X(x) = 1$$

2 Modelos Clásicos

Modelo Bernoulli. Decimos que X es una variable aleatoria con distribución Bernoulli de parámetro p, $0 \leq p \leq 1$, si la imagen de X es $\{0, 1\}$ y

$$P(X = 0) = 1 - p$$
$$P(X = 1) = p$$

Modelo Binomial. Decimos que X es una variable aleatoria con distribución Binomial de parámetros n y p, $n \in N$ y $p \in [0, 1]$, si la imagen de X es $\{0,1,\ldots,n\}$ y

$$P(X=k)=\binom{n}{k}p^k q^{n-k} \qquad\qquad k=0,1,\ldots,n$$

Modelo Geométrico. Decimos que X es una variable aleatoria con distribución Geométrica de parámetro p, $0 \leq p \leq 1$, si la imagen de X es $\{1,2,3,\ldots\}$ y

$$P(X=k)=(1-p)^{k-1}p \qquad k=\{1,2,3,\ldots\}$$

Modelo Hipergeométrico. Decimos que X es una variable aleatoria con distribución Hipergeométrica de parámetros N, A y n, $A \leq N$ y $n \leq n$, si la imagen de X es $\{0,1,\ldots,n\}$ y

$$P(X=k)=\frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}} \qquad k=\{0,1,\ldots,inf\{A,n\}\}$$

Modelo Poisson. Decimos que X es una variable aleatoria discreta con distribución Poisson de parámetro λ, si la imagen de X es $\{0,1,2\ldots\}$ y

$$P(X=k)=\frac{1}{k!}\lambda^k e^{-\lambda} \qquad k=\{0,1,2,\ldots\}$$

3 Vectores Aleatorios Discretos

Sean X e Y variables aleatorias discretas en un espacio de probabilidad. El vector (X,Y) es llamado vector aleatorio. Note que (X,Y) toma valores en R^2. La función de masa probabilidad **conjunta** de X e Y es la función $P_{X,Y}: R^2 \to [0,1]$ definida por

$$P_{X,Y}(x,y)=P(\{\omega\in\Omega : X(\omega)=x \text{ y } Y(\omega)=y\})$$

es decir, $P_{X,Y}(x,y)$ es la probabilidad de que $X=x$ e $Y=y$. Al igual que para el caso univariado, simplificamos la notación usando la identidad

$$P(X\in A, Y\in B)=P(\{\omega\in\Omega : X(\omega)\in A \text{ y } Y(\omega)\in B\})$$

Asi, $P_{X,Y}(x,y) = P(X = x, Y = y)$. Note que, si $x \notin ImX$ o $y \notin ImY$ entonces $P_{X,Y}(x,y) = 0)$. Adicionalmente,

$$\sum_{x \in ImX} \sum_{y \in ImY} P_{X,Y}(x,y) = 1$$

Las funciones de probabilidad P_X y P_Y de X e Y las podemos obtener de $P_{X,Y}$ usando el siguiente hecho:

$$\begin{aligned} P_X(x) = P(X = x) &= \sum_{y \in ImY} P(X = x, Y = y) \\ &= \sum_y P_{X,Y}(x,y). \end{aligned}$$

De igual forma,

$$P_Y(y) = \sum_x P_{X,Y}(x,y)$$

En este contexto, P_X y P_Y son llamadas funciones de probabilidad marginales de X e Y respectivamente.

Podemos generalizar las definiciones anteriores al caso n-dimensional al considerar vectores aleatorios $X = (X_1, X_2, \ldots, X_n)$, donde X_i, para $i = 1, 2 \ldots, n$, son v.a. discretas. La función de probabilidad conjunta del v.a. X se define de la manera natural,

$$P_X(x) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

4 Independencia de v.a. discretas

Recordemos que dos eventos A y B son independientes si

$$P(A \cap B) = P(A)P(B).$$

La noción de independencia se extiende a variables aleatorias de la manera natural. Diremos que X e Y son independientes si los eventos

$$\{\omega \in \Omega : X(\omega) \in D\} \text{ y } \{\omega \in \Omega : Y(\omega) \in F\}$$

son independientes para cualquier $D \subset ImX$ y $F \subset ImY$. Note que si X e Y son v.a. discretas e independientes, la función de probabilidad conjunta de X e Y se escribe como el producto de las funciones marginales de X e Y, esto es,

$$P(X = x, Y = y) = P(X = x)P(Y = y) \ \ \forall x, y \in R$$

Teorema de Factorización. Sean X e Y v.a. discretas. X e Y son independientes si y solo si existen funciones $f, g : R \to R$ tales que la función de probabilidad conjunta de X e Y satisface,

$$P_{X,Y}(x, y) = P(X = x, Y = y) = f(x)g(y) \quad \forall x, y \in R$$

Ejemplo. Sean X, Y v.a. discretas a valores enteros no-negativos con función de masa conjunta:

$$P_{X,Y}(i, j) = P(X = i, Y = j) = \frac{1}{i!j!}\lambda^i\mu^j e^{-(\lambda+\mu)} \ \ i, j = 0, 1, \ldots$$

factorizando tenemos que

$$P_{X,Y}(i, j) = (\frac{1}{i!}\lambda^i)(\frac{1}{j!}\mu^j e^{-(\lambda+\mu)}) \ i, j = 0, 1, \ldots$$

Por el Teorema de Factorización, X e Y son independientes.

5 Máximo y Mínimo de v.a. discretas i.i.d

Sean $X_1, X_2, \ldots, X_n$ v.a. discretas. Para determinar función de masa de probabilidad del mínimo

$$U_n = min\{X_1, X_2, \ldots, X_n\},$$

observemos que el evento $\{U_n > k\}$, $k \in R$, se puede reescribir por

$$\{U_n > k\} = \{X_1 > k, X_2 > k, \ldots, X_n > k\}.$$

Es decir,

$$P(\{U_n > k\}) = P(\{X_1 > k, X_2 > k, \ldots, X_n > k\})$$

Si $X_1, X_2, \ldots, X_n$ son v.a. independientes,

$$P(U_n > k) = P(X_1 > k)P(X_2 > k)\ldots P(X_n > k)$$

Si $X_1, X_2, \ldots, X_n$ son v.a. **independientes e idénticamente distribuídas (i.i.d.),**

$$P(U_n > k) = P(X_1 > k)^n$$

Por tanto, la función de probabilidad de U_n la podemos obtener de la identidad

$$P(U_n = k) = P(U_n > k-1) - P(U_n > k)$$

Ejemplo. Pruebe que el mínimo de v.a. Geométricas i.i.d es también una v.a. Geométrica.

Consideremos ahora el máximo

$$V_n = max\{X_1, X_2, \ldots, X_n\}.$$

Es fácil chequear que

$$\{V_n \leq k\} = \{X_1 \leq k, X_2 \leq k, \ldots, X_n \leq k\}$$

Usando, los mismos argumentos que antes, si $X_1, X_2, \ldots, X_n$ son i.i.d.

$$P(V_n \leq k) = P(X_1 \leq k)^n.$$

Así, la función de probabilidad de V_n la podemos obtener usando la identidad

$$P(V_n = k) = P(V_n \leq k) - P(V_n \leq k-1)$$

6 Suma de variables aleatorias independientes

Sean X, Y v.a. discretas e independientes. Consideremos la variable aleatoria $Z = X + Y$. Note que si $Z = z$ y $X = x$ entonces necesariamente $Y = z - x$. Así,

$$\begin{aligned} P(Z = z) &= P(\cup_{x \in ImX}\{X = x, Y = z - x\}) \\ &= \sum_{x \in ImX} P(X = x, Y = z - x) \end{aligned}$$

Teorema: Si X, Y v.a. discretas e independientes entonces $Z = X + Y$ tiene función de masa de probabilidad

$$P(Z = z) = \sum_{x \in ImX} P(X = x)P(Y = z - x).$$

En particular, si $ImX = ImY = \{0, 1, 2, \ldots\}$

$$P(Z = k) = \sum_{x=0}^{k} P(X = x)P(Y = k - x)$$

para cualquier k entero no negativo. Se dice que la función de probabilidad $X + Y$ es la convolución de las funciones de probabilidad de X e Y.

Ejemplo. Sean X, Y v.a. independientes con distribución Poisson de parámetros λ y μ respectivamente. Determine la distribución de $X + Y$.

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^{z} P(X = x)P(Y = z - x) \\ &= \sum_{x=0}^{z} \left(\frac{1}{x!}\lambda^x e^{-\lambda}\right)\left(\frac{1}{(z-x)!}\mu^{z-x}e^{-\mu}\right) \\ &= \frac{1}{z!}(\lambda + \mu)^z e^{-(\lambda+\mu)} \end{aligned}$$

7 Esperanza

Consideremos un dado justo. Si este es lanzado un número grande de veces, cada posible resultado $1, 2, \ldots, 6$ podría aparecer en un sexto de las veces y el promedio, o valor medio, de los números observados sería aproximadamente

$$1(1/6) + 2(1/6) + \ldots + 6(1/6) = 3,5$$

Si lanzamos una moneda justa cien veces. ¿Cuantas caras esperamos observar?

Definición Si X es una v.a. discreta, la esperanza de X, llamada también valor esperado X, es

$$E(X) = \sum_{x \in ImX} xP(X = x) = \sum_{x} xP_X(x).$$

Teorema Si X es una v.a. discreta y $g : R \to R$ entonces

$$E(g(X)) = \sum_{x \in ImX} g(x)P(X = x)$$

Prueba

$$\begin{aligned} E(Y) &= \sum_{y \in g(ImX)} yP(Y = y) \\ &= \sum_{y \in g(ImX)} y \sum_{x \in ImX, g(x)=y} P(X = x) \\ &= \sum_{x \in ImX} g(x)P(X = x) \end{aligned}$$

Otra cantidad importante asociada con X es su varianza, la cual nos indica el grado de dispersión que tiene X respecto a su esperanza $E(X)$. Formalmente, la varianza de una v.a. X se define como el valor esperado de $g(X) = [X - E(X)]^2$. La varianza es escrita como:

$$\begin{aligned} Var(X) &= E([X - E(X)]^2) \\ &= \sum (x - \mu)^2 P(X = x) \end{aligned} \tag{1}$$

donde $\mu = E(X)$.

Es fácil probar que

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

8 Esperanza en el caso multivariado

Sea (X, Y) vector aleatorio discreto y $g : R^2 \to R$ entonces

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)P(X = x, Y = y) \tag{2}$$

Proposición 1. Sea $Z = g(X, Y) = aX + bY$ donde $a, b \in R$, entonces

$$E(Z) = E(aX + bY) = aE(X) + bE(Y)$$

La **covarianza** entre X e Y es el valor esperado

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

donde $\mu_X = E(X)$ y $\mu_Y = E(Y)$. La covarianza es usada como una medida que indica el grado de dependencia lineal que existe entre las variables X e Y. Es fácil probar que

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Proposición 2.

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + abCov(X,Y)$$

Proposición 3.

Si X, Y v.a. independientes entonces

- $E(XY) = E(X)E(Y)$
- $Var(X + Y) = Var(X) + Var(Y)$

9 Esperanza Condicionada a un Evento

Sea X una v.a. discreta y B un evento con $P(B) > 0$. La Esperanza Condicional de X dado el evento B es denotada por $E(X/B)$ y definida por

$$E(X|B) = \sum_x xP(X = x|B)$$

Teorema Si X es una v.a. discreta y $\{B_1, B_2, \ldots\}$ es una partición del espacio muestral Ω tal que $P(B_i) > 0$ para cada i, entonces

$$E(X) = \sum_i E(X|B_i)P(B_i)$$

Prueba

$$\begin{aligned}
E(X) &= \sum_i E(X|B_i)P(B_i) \\
&= \sum_i \sum_x xP(X = x|B)P(B_i) \\
&= \sum_i \sum_x xP(\{X = x\} \cap B_i) \\
&= \sum_x xP(\{X = x\} \cap (\cup B_i)) \\
&= \sum_x xP(X = x)
\end{aligned}$$

Ejemplo. Una moneda es lanzada repetidamente y sea p la probabilidad de obtener cara en cada lanzamiento. Encuentre la longitud esperada de la corrida inicial, esto es el número de lados iguales consecutivos hasta obtener el lado contrario de la moneda.

Sea H el evento el primer lanzamiento es cara y H^c el evento el primer lanzamiento es sello. El par H, H^c forman una partición del espacio muestral. Si X es la longitud de la corrida inicial. Es fácil verificar que:

$$P(X = k|H) = p^{k-1}q \qquad k = 1, 2, \ldots$$

ya que si H ocurre entonces $X = k$ ocurre sii el primer lanzamiento esta seguido por exactamente $k - 1$ caras y después un sello. Similarmente,

$$P(X = k|H^c) = q^{k-1}p \quad k = 1, 2, \ldots$$

Así,

$$E(X|H) = \sum kp^{k-1}q = \frac{q}{(1-p)^2} = \frac{1}{q}$$

similarmente

$$E(X|H^c) = \frac{1}{p}$$

usando el teorema anterior tenemos que

$$E(X) = E(X|H)P(H) + E(X|H^c)P(H^c) = \frac{1}{q}p + \frac{1}{p}q = \frac{1}{pq} - 2$$

10 Segundo Problemario

1. Calcular el valor esperado de X, cuando X tiene distribución:

   - binomial de parámetros n y p
   - geométrica de parámetro p
   - hipergeométrica
   - Poisson de parámetro λ

2. Calcular la varianza de X, cuando X tiene distribución:

   - binomial de parámetros n y p
   - geométrica de parámetro p
   - Poisson de parámetro λ

3. Si X se distribuye Poisson de parámetro λ, pruebe que

$$E[X(X-1)(X-2)\dots(X-k)] = \lambda^{k+1}$$

4. Si X tiene distribución geométrica, pruebe la propiedad de pérdida de memoria

$$P(X > m+n|X > m) = P(X > n)$$

5. Sea N una v.a. a valores enteros no negativos. Verifique que

$$E[N] = \sum_{k \geq 0} P(N > k)$$

Un dado tiene dos cara azules, dos rojas y dos verdes. Se lanza repetidamente. Encuentre la probabilidad de que no todos los colores aparezcan en los primeros k lanzamientos. Deduzca que si N es la v.a. que toma el valor n si el tercer color aparece en el n-ésimo lanzamiento por primera vez, entonces $E[N] = 11/2$.

6. Suponga que $P(X = i, Y = j) = \lambda^{1+i+j}$, para $i, j = 0, 1, 2$. Pruebe que

$$E[XY] = \lambda^3 + 4\lambda^4 + 4\lambda^5$$

7. Sean X, Y v.a. i.i.d. con $P(X = k) = pq^k$, $k \geq 0$. Demuestre que para $k = 0, \dots n$

$$P(X = k | X + Y = n) = \frac{1}{n + 1}$$

Sugerencia: Use Bayes y la fórmula de convolución.

8. Se lanza un dado n veces. Sea U_n el mínimo valor observado y V_n el máximo valor observado. Calcular $P(U_n = 1)$, $P(V_n = 6)$.

9. Existen c diferentes tipos de barajitas y cada una tiene el mismo chance de ser adquirida. Sea Y_i el número adicional de barajitas coleccionadas después de obtener i tipos de barajitas antes de obtener un nuevo tipo. Demuestre que Y_i tiene distribución geométrica con parámetro $(c-i)/c$. Calcule el número esperado de barajitas que necesitas adquirir antes de llenar el álbum.

10. Sean X, Y v.a. independientes con distribución geométrica con parámetro p y r respectivamente. Pruebe que $\min(X, Y)$ tiene distribución geométrica con parámetro $p + r - pr$

11. Sean X, Y v.a. independientes con distribución Poisson de parámetro λ y μ respectivamente. Use el hhecho de que $X + Y$ es Poisson para calcular $P(X = k | X + Y = n)$ para $k = 0, \dots, n$. Demuestre que

$$E[X | X + Y = n] = n\lambda/(\lambda + \mu)$$

Sugerencia: Use Bayes para la primera parte

12. Sea N el número de lanzamientos de una moneda hasta que se repita el resultado del primer lanzamiento. Condicionando en el primer lanzamiento, calcule $E[N]$.

13. La función generatriz de probabilidades de una variable aleatoria discreta X está definida por la serie de potencias

$$g(s) = E\left[s^X\right] = \sum_{k=0}^{\infty} s^k P(X = k), \quad |s| < 1.$$

Calcule las funciones generatrices de probabilidades de las siguientes distribuciones: Bernoulli, binomial y Poisson.

14. Dos distribuciones de probabilidades distintas tienen diferentes funciones generatrices de probabilidades. Use la identificación de las funciones generatrices para probar los siguientes resultados:

    - sumas de Bernoulli i.i.d es binomial.
    - la suma de binomiales independientes con el mismo parámetros p es tambien binomial.
    - sumas de Poisson independientes es Poisson.

15. Considere que el número de veces que una moneda es lanzada es una v.a. Poisson. Sea X el número de caras y Y el número de sellos. Verifique que X, Y son independientes. Sugerencia: Use la fórmula de particionamiento para calcular la masa de probabilidad de X.

16. Se lanza un dado n veces. Sea U_n el mínimo valor observado y V_n el máximo valor observado. Calcular $P(U_n = 1)$, $P(V_n = 6)$.

17. Una línea aérea cubre la ruta Caracas-Los Roques, con un avión que tiene 50 plazas. La politica de la aerolinea es aceptar 55 reservaciones para este vuelo. Se supone que todos los pasajeros actúan en forma independiente y que la probabilidad de que un pasajero se presente es p. El precio del pasaje es G, pero si un pasajero se presenta y no puede ser embarcado, se le reintegra su dinero mas una compensación, H.

- Calcule la esperanza del número de pasajeros que se presentan a abordar.
- Calcule la esperanza del número de pasajeros que acuden y no pueden ser embarcados.
- Calcule la ganancia esperada por la aereolínea.

18. El costo de producción de cierta máquina que se fabrica por encargo, es, para nuestra empresa, de 4.3×10^8 Bs. por máquina, cuando se producen menos de cinco unidades. Si se producen de cinco a nueve unidades, inclusive, el costo por máquina baja a 4.0×10^8 Bs. por máquina, y cuando se producen diez o más unidades el costo por unidad baja a 3.5×10^8 Bs. La demanda de estas máquinas fluctúa según una distribución Poiss(8). Hallar el precio de venta unitario, para que la ganancia esperada por máquina sea de 50 millones de Bs. Si vendemos a ese precio, ¿Cual es la probabilidad de que nuestra empresa pierda dinero?

19. El número de llamadas que llegan a la central telefónica de Sartenejas en un minuto, es, en promedio, 10^2. La central puede manejar un máximo de 10^3 llamadas, colapsando si recibe mas de este número de llamadas en un minuto. Usar la desigualdad de Chebyshev para estimar la probabilidad de que la central colapse en un minuto dado.