CO3121 - Introducción al Cálculo de Probabilidades (Semanas 8, 9, 10 y 11)

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

1 Densidad conjunta y densidades marginales

Un par ordenado (X, Y) de v.a. continuas, es un punto aleatorio en el plano. Experimentos aleatorios como el lanzamiento de un dardo u observaciones multivariadas en muestreo aleatorio pueden ser modelados por conjuntos ordenados de v.a.

Extendiendo el concepto de densidad de probabilidad que introducimos para una v.a. a un punto aleatorio en el plano; diremos que el par ordenado de v.a. (X, Y) es continuo si existe una función $f : R^2 \to (0, \infty)$ tal que para cualquier evento $D \subset R^2$ se cumple

$$P((X,Y) \in D) = \int\int_D f(x,y)dydx \qquad (1)$$

Por supuesto, supondremos

$$P((X,Y) \in R^2) = \int\int_{R^2} f(x,y)dydx = 1.$$

La función f es llamada densidad de probabilidad conjunta del vector aleatorio (X, Y).

A efecto de identificar la densidad conjunta de (X, Y), es suficiente verificar la ecuación (**??**) para conjuntos $D \in R^2$ que sean de la forma $A \times B$, es decir, f es la densidad conjunta de(X, Y) sii

$$P(X \in A, Y \in B) = \int_A \int_B f(x,y)dydx$$

para cualquier par de eventos $A, B \subset R$.

Denotemos por f_X la densidad de probabilidad de X. Si f es la densidad conjunta de (X, Y), la identidad

$$P(X \in A) = P(X \in A, Y \in R)$$

puede reescribirse por

$$\int_A f_X(x)dx = \int_A \int_R f(x,y)dydx$$

es decir,

$$f_X(x) = \int_R f(x,y)dy.$$

En este contexto, $f_X(x) = \int_R f(x,y)dy$ es llamada densidad marginal de X. De manera simétrica definimos la densidad marginal de Y por

$$f_Y(y) = \int_R f(x,y)dx.$$

2 Independencia de v.a. continuas

Anteriormente vimos que las v.a. X, Y son independientes sii para cualquier par de eventos A, B de la recta real

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Para el caso continuo que estamos considerando, la ecuación anterior se escribe

$$\int_A \int_B f(x,y)dydx = \int_A f_X(x)dx \int_B f_Y(y)dy,$$

lo que permite probar el siguiente

Teorema: X, Y son independientes si y solo si la densidad conjunta $f = f_X f_Y$.

De lo anterior se desprende que si la densidad conjunta admite la factorización

$$f(x,y) = g(x)h(y), \tag{2}$$

aun cuando g, h no sean las marginales de X, Y respectivamente, es suficiente para probar que X, Y son independientes.

Ejemplo: Consideremos el punto (X, Y) con distribución uniforme en el disco de radio r, $D_r = \{(x, y) : x^2 + y^2 \leq r^2\}$, el cual modela el resultado de lanzar un dardo a una diana de radio r por un lanzador sin puntería. Si $A \subset D_r$, entonces

$$P((X, Y) \in A) = \frac{\text{area}(A)}{\text{area}(D_r)} = \frac{1}{\pi r^2} \int \int_A dy dx \tag{3}$$

Luego, la densidad conjunta de (X, Y) es

$$\begin{aligned} f(x, y) &= \frac{1}{\pi r^2} \text{ si } (x, y) \in D_r \\ &= 0 \text{ si no} \end{aligned}$$

Son X, Y independientes?. Comunmente un novato aplicaría incorrectamente el criterio de factorización (**??**) sin darse cuenta que las regiones en las que esta definida por partes la densidad conjunta no pueden factorizarse. A veces ayuda usar en estos casos la función indicatriz de un conjunto A definida por

$$\begin{aligned} I_A(\omega) &= 1 \text{ si } \omega \in A \\ &= 0 \text{ si no} \end{aligned}$$

Así podemos escribir $f(x, y) = \frac{1}{\pi r^2} I_{D_r}(x, y)$ sin la posibilidad de incurrir en el error de creer que $I_{D_r}(x, y)$ puede ser factorizada como en (**??**).

3 Cambio de variable y aplicaciones

Para el punto (X,Y) distribuído uniformemente en el disco D_r, consideremos ahora las coordenadas polares $R = \sqrt{X^2 + Y^2}$ y Θ el ángulo entre el vector (X,Y) y el eje x. Usando la fórmula para el área de una rebanada del disco, (**??**) implica

$$P(0 < R < \rho, 0 < \Theta < \theta) = \frac{\rho^2 \theta}{r^2 2\pi} = \int_0^\rho \int_0^\theta \frac{2x}{r^2} \frac{1}{2\pi} dv du$$

De donde

$$f_{R,\Theta}(\rho, \theta) = \frac{\rho}{r^2} \frac{1}{\pi} I_{(0,r)}(\rho) I_{(0,2\pi)}(\theta)$$

y usando (**??**) vemos que R, Θ si son independientes. ¿Caules son las densidades marginales?.

El cambio a coordenas polares anterior, ejemplifica un problema general:

Dado un v.a. (X, Y) con densidad conjunta $f(x, y)$ y una transformacion $T(X, Y) = (U, V)$, ¿cual es la densidad conjunta del nuevo v.a. (U, V)?. Si la transformación es invertible

$$P((X, Y) \in A) = P((U, V) \in T(A))$$

Usando la fórmula de cambio de variable, la ecuación anterior se reescribe

$$\int\int_A f(x, y)dxdy = \int\int_{T(A)} f(x(u, v), y(u, v))|J(u, v)|dudv$$

donde $(x(u, v), y(u, v)) = T^{-1}(u, v)$ y J es el jacobiano de T^{-1}. Estas identidades permiten probar el siguiente

Teorema: Sea (X, Y) un v.a. con densidad conjunta $f(x, y)$, $D = \{(x, y) : f(x, y) > 0\}$ y $T : D \to S$ invertible, entonces la densidad conjunta del v.a. (U, V) es

$$f_{(U,V)}(u, v) = f(x(u, v), y(u, v))|J(u, v)|I_S(u, v)$$

Marginalizando, esta fórmula es de mucha utilidad para calcular densidades de v.a. que son funciones de puntos aleatorios

Ejemplo: Sean X, Y i.i.d con distribución exponencial de parámetro 1. ¿Como se distribuye $X/(X + Y)$?.

Consideremos las v.a. $V = X/(X + Y)$ y $U = X + Y$. La variable U es auxiliar a fin de poder invertir la transformación $T(x, y) = (x+y, x/(x+y))$. Usando el teorema anterior, podemos calcular la densidad conjunta de (U, V)

$$f_{(U,V)}(u, v) = e^{-u}|J(u, v)|I_{R^+}(u)I_{(0,1)}(v) = ue^{-u}I_{R^+}(u)I_{(0,1)}(v)$$

Marginalizando

$$f_V(v) = \int_0^\infty f_{(U,V)}(u, v)du = 1$$

para $v \in (0, 1)$. es decir $X/(X + Y)$ es uniforme en (0,1).

4 Propiedades de la esperanza

Para $g : R \times R \to R$, probamos que si X, Y son discretas

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) P(X = x, Y = y). \tag{4}$$

Tambien hemos visto como la esperanza para el caso univariado discreto, $\sum_x xP(X = x)$ tiene su versión para el caso continuo $\int xf(x)dx$. La intuición sugiere dar como versión continua de (**??**) a

$$E[g(X,Y)] = \int\int g(x,y) f_{X,Y}(x,y) dx dy$$

La prueba formal de este resultado escapa de los alcances del curso.

De manera identica a como lo hicimos para el caso discreto, usando ahora las densidades conjuntas y marginales puede probarse los siguientes resultados:

Proposición 1: $E[aX + bY] = aE[X] + bE[Y]$

Proposición 2: La covarianza definida por

$$Cov(X,Y) = E[(X - EX)(Y - EY)]$$

puede calcularse por

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

Proposición 3: $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + abCov(X,Y)$

Proposición 4: Si X, Y son independientes $E[XY] = E[X]E[Y]$ y en consecuencia $Var(X + Y) = Var(X) + Var(Y)$.

5 Densidad y esperanza condicional

Para el caso X, Y discretas, la probabilidad condicional de X dado $Y = y$ viene dada directamente por la fórmula de la probabilidad condicional

$$P(X \in A | Y = y) = \frac{P(X \in A, Y = y)}{P(Y = y)} = \sum_{x \in A} \frac{P(X = x, Y = y)}{P(Y = y)}$$

Para extender esta idea al caso continuo debemos proceder con cuidado, ya que si Y es continua $P(Y=y)=0$ para todo y y la ecuación anterior no tiene sentido. La idea es considerar eventos del tipo $\{y \leq Y \leq y+\delta\}$ para $\delta > 0$ y hacer luego $\delta \to 0$. Para lo que sigue $f(x,y)$ denota la densidad conjunta de (X,Y):

$$\begin{aligned} P(a \leq X \leq b | y \leq Y \leq y+\delta) &= \frac{P(a \leq X \leq b, y \leq Y \leq y+\delta)}{P(y \leq Y \leq y+\delta)} \\ &= \frac{\int_a^b \left(\int_y^{y+\delta} f(x,v)dv\right)dx}{\int_y^{y+\delta} f_Y(v)dv} \end{aligned}$$

Dividiendo y multiplicando por δ, y observando que

$$\frac{1}{\delta}\int_y^{y+\delta} f(x,v)dv \to f(x,y)$$

y

$$\frac{1}{\delta}\int_y^{y+\delta} f_Y(v)dv \to f_Y(y)$$

podemos verificar que

$$P(a \leq X \leq b | y \leq Y \leq y+\delta) \to \int_a^b \frac{f(x,y)}{f_Y(y)}dx$$

Usando la definición de densidad de probabilidad, el integrando, es decir $f(x,y)/f_Y(y)$, es la densidad condicional de X dado $\{y \leq Y \leq y+\delta\}$ cuando $\delta \to 0$, es decir dado $\{Y=y\}$. A la densidad condicional de X dado $Y=y$ la denotaremos por

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Teniendo una fórmula para la densidad condicional, la esperanza condicional de X dado $Y=y$ puede ser definida como la esperanza respecto a la densidad condicional, i.e.

$$E[X|Y=y] = \int x f_{X|Y}(x|y)dx$$

En el caso discreto vimos la utilidad de la esperanza condicional para el cálculo de valores esperados via formula de particionamiento

$$E[X] = \sum_y E[X|Y = y]P(Y = y)$$

Con las definiciones anteriores es fácil probar la versión continua de esta importante fórmula:

$$E[X] = \int E[X|Y = y]f_Y(y)dy$$

Ejemplo: Un distribución que sirve para ilustrar lo visto en las secciones anteriores es la normal bivariada. Sea

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

donde x, y son números reales y $|\rho| \leq 1$, la densidad conjunta de (X, Y). Verifique que

- X se distribuye $N(0, 1)$
- X dado $Y = y$ se distribuye $N(\rho y, 1 - \rho^2)$
- $E[X|Y = y] = \rho y$
- X, Y son independientes si y solo si son incorrelacionados. Recuerde que incorrelación no implica independencia, lo cual hace a la normal bivariada una distribución muy especial.
- Si X, Y son independientes, X/Y tiene distribución de Cauchy, es decir su densidad es $1/\pi(1 + x^2)$

6 Distribución del Máximo y del Mínimo

Sea $(X_1, X_2, \ldots, X_n)$ un vector aleatorio. Queremos determinar la distribución del mínimo:

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

Consideremos el evento $\{U_n > k\}$ para $k \in R$, es fácil chequear que

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

en consecuencia,

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

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,**

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

La función de distribución de $U_n(k)$ es

$$F_{U_n}(k) = P(U_n \leq k) = 1 - P(U_n > k) = 1 - P(X_1 > k)^n = 1 - [1 - F_X(k)]^n$$

y su función de densidad es

$$f_{U_n}(k) = n[1 - F_X(k)]^{n-1} f_X(k)$$

Para determinar la distribución del máximo:

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

consideremos el evento $\{V_n \leq k\}$ para $k \in R$. Es fácil chequear que

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

en consecuencia la función de distribución $F_{V_n}(k)$ será,

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

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

$$P(V_n \leq k) = P(X_1 \leq k)P(X_2 \leq k)\ldots P(X_n \leq k)$$

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

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

así

$$F_{V_n}(k) = P(V_n \leq k) = F_X(k)^n$$

con función de densidad

$$f_{V_n}(k) = n[F_X(k)]^{n-1} f_X(k)$$

Ejemplo Sean $X_1, X_2, \ldots, X_n$ v.a. i.i.d. exponenciales de parámetro $\lambda = 1$. Determine la distribución del mínimo. Usando el hecho que son i.i.d. tenemos

$$f_{U_n}(u) = n[1 - F_X(u)]^{n-1} f_X(u)$$

por tanto,

$$f_{U_n}(u) = n[1 - (1 - e^{-u}]^{n-1} e^{-u} = ne^{-nu}$$

Observando que la distribución del mínimo U_n es exponencial de parámetro $\lambda = n$.

7 Suma de Variables Aleatorias

Sean X, Y v.a. y $Z = X + Y$. Vamos a determinar la distribución de la v.a. Z. Consideremos ahora (X, Y) v.a. continuo con función de densidad conjunta. Si $Z = X + Y$ entonces,

$$\begin{aligned}
P(Z \leq z) &= P(X + Y \leq z) \\
&= \int\int_{\{x+y \leq z\}} f_{X,Y}(x, y)dydx \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y)dydx
\end{aligned}$$

haciendo un cambio de variables $u = x$ y $v = x + y$ donde $|J| = 1$, tenemos

$$\begin{aligned} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} f_{X,Y}(u, v-u) dv du \\ &= \int_{-\infty}^{z} \int_{-\infty}^{+\infty} f_{X,Y}(u, v-u) du dv \end{aligned}$$

observando que

$$P(Z \leq z) = \int_{-\infty}^{z} f_Z(v) dv$$

entonces

$$f_Z(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u, v-u) du \quad \forall v \in R$$

Teorema: Si X, Y v.a. independientes entonces $Z = X + Y$ tiene función de densidad:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$$

Se dice que la función de densidad $X + Y$ es la convolución de las funciones de densidad de X y Y.

Ejemplo: Sean X, Y v.a. independientes con distribuciones $Gamma(s, \lambda)$ y $Gamma(t, \lambda)$ respectivamente. Encuentre la distribución de $Z = X + Y$.

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$$

8 Tercer Problemario

1. Halle valor esperado y varianza de una variable aleatoria continua con distribucion
   - Uniforme(a, b)
   - Exponencial(λ)
   - Normal(μ, σ)
   - Gamma(α, β)

2. Suponga que X, Y son independientes y obtenga las siguientes fórmulas:

   - $f_{X+Y}(z) = \int f_X(u) f_Y(z-u) du$
   - $f_{XY}(u) = \int f_X(x) f_Y(u/x) |x|^{-1} dx$
   - $f_{X/Y}(v) = \int f_X(vy) f_Y(y) |y| dy$

3. Demuestre que si X, Y son exponenciales independientes con parámetros μ, λ respectivamente, entonces la distribución del mínimo es también exponencial y determine su parámetro.

4. Sean $X_1, \ldots X_n$ v.a.i.i.d con función de densidad común $f(x)$. Sean U, V el mínimo y el máximo de la muestra. Pruebe que la densidad conjunta de (U, V) es

$$n(n-1) f(u) f(v) (F(v) - F(u))^{n-2}, \quad \text{para} \quad u < v$$

   y calcule las marginales (aquí $F\prime = f$)

5. Considere el punto aleatorio (X, Y) con distribución normal bivariada y coordenadas independientes. Sea (R, Θ) las coordenadas polares del punto aleatorio. Identifique la distribución marginal de R^2 y Θ.

6. Un juego se llama *justo* cuando la esperanza de la ganancia de los participantes es cero.
   La flecha lanzada por un experto arquero, caerá a una distancia de R pies del centro de un blanco. Se pagan 5\$ como entrada para participar en un juego cuyas reglas son las siguientes: Si $R < 0.2$ pies, el participante recibe 50\$. Si $0.2 \leq R < 0.5$ pies, el participante recibe 10\$. Si $0.5 \leq R < 1$ pie, el participante no recibe ningún pago y, finalmente, si $R \geq 1$ pies el participante debe pagar x dólares a la casa (adicionales a los cancelados a la entrada, por supuesto). Si R tiene densidad $f(r) = r \exp(-\frac{1}{2} r^2)$, $r > 0$, ¿Cuanto debe valer x para que el juego sea justo?

7. (X, Y) tienen densidad conjunta

$$\begin{aligned} f(x,y) &= e^{-y} \quad \text{para} \quad 0 < x < y \\ &= 0 \quad \text{en otro caso} \end{aligned}$$

   encuentre $E[X|Y=y]$ y $E[Y|X=x]$

8. Sean X, Y v.a. independientes con distribución gamma de parámetros (n, β) y (m, β) respectivamente. Considere las variables

$$U = X + Y \qquad V = \frac{X}{X+Y}$$

Demuestre que U, V son independientes y calcule sus distribuciones. Deduzca la curiosa identidad válida para este caso

$$E\left[\frac{X}{X+Y}\right] = \frac{E[X]}{E[X] + E[Y]}$$

9. La variable aleatoria X tiene f.d.a. $F(x) = x^r$, $0 \le x \le 1$, siendo r un número natural. Dado $X = x$, la variable Y tiene una distribución $\text{Bin}(n, x)$.
(a) Hallar $E(Y)$.
(b) Hallar la f.d.p. de Y. Para esto puede necesitar la integral conocida como función β: Si i y j son números naturales, se tiene

$$\int_0^1 x^i (1-x)^j \, dx = \frac{i!\, j!}{(i+j+1)!}.$$

10. Se toma un punto (X, Y) al azar en el triángulo de vértices (0,0), (0,2) y (1,1). Probar que $E(Y \mid X = x)$ no depende de x. ¿Son X e Y independientes?

11. Debido a variabilidad en el proceso de producción, la *tasa de vida* Γ, de los amplificadores producidos por una fabrica, tiene una distribución $N(\mu, \sigma^2)$ con $\mu = 1.07 \times 10^4$ seg$^{1/2}$, $\sigma^2 = 1.1 \times 10^6$ seg. A su vez, el tiempo de vida T, de un amplificador con tasa de vida Γ, tiene distribución exponencial de parámetro $\lambda = \frac{1}{\Gamma^2}$ seg^{-1}. Hallar el tiempo de vida promedio (en meses) de los amplificadores producidos por esta fábrica.

12. Sea X v.a. con media μ, varianza σ^2 y cuarto momento central $\mu_4 = E((X-\mu)^4)$. Pruebe que $\mu_4 \ge \sigma^4$.
Ayuda: calcule la esperanza y la varianza de $Z = (X - \mu)^2$.

9 Ley Débil de Grandes Números

Para modelar un fenómeno aleatorio que depende del tiempo, podemos considerar sucesiones de v.a. $X_1, X_2, \ldots$ donde X_i modela el estado del fenómeno a tiempo i. El conjunto de índices puede representar unidades de tiempo, iteraciones de un proceso, etc. Es natural que en este tipo de situaciones, si queremos tener una información que no dependa del tiempo, nos preguntemos acerca del comportamiento de X_n cuando $n \to \infty$. Antes de formalizar esta idea observemos el siguiente

Ejemplo: Supongamos que lanzamos un dado n veces, n un nḿero muy grande, y observamos que el promedio de los números observados es $a_n = 3.500867$. Al comparar este valor con el valor esperado de una v.a. que modela el número observado al lanzar un dado, $\frac{1}{6}(1+2+\ldots+6) = 3.5$, podemos ver que estan muy cerca. Es natural conjeturar que

$$a_n \to 3.5 \quad \text{cuando} \quad n \to \infty$$

Varios resultados de la teoría de probabilidad establecen condiciones para probar conjeturas del tipo "si repetimos un experimento muchas veces entonces el promedio de los resultados se aproxima al valor esperado".

Teorema (Desigualdad de Markov) Sea X una v.a. y g una función positiva entonces,

$$P(g(X) \geq \epsilon) \leq \frac{E(g(X))}{\epsilon} \quad \forall \epsilon > 0$$

Prueba. Supongamos que X es continua, el caso discreto es simmilar.

$$\begin{aligned}
E(g(X)) &= \int_{\{x:g(x)\geq\epsilon\}} g(x)f(x)dx + \int_{\{x:g(x)<\epsilon\}} g(x)f(x)dx \\
&\geq \int_{\{x:g(x)\geq\epsilon\}} g(x)f(x)dx \\
&\geq \epsilon \int_{\{x:g(x)\geq\epsilon\}} f(x)dx \\
&= \epsilon P(g(X) \geq \epsilon)
\end{aligned}$$

Un caso particular de la desigualdad de Markov es

Corolario(Desigualdad de Chebyshev) Sea X una v.a. con $E(X) = \mu$ y $Var(X) < \infty$ entonces

$$P(|X - \mu| \geq \epsilon) \leq \frac{1}{\epsilon^2} Var(X)$$

Prueba

$$P(|X - \mu| \geq \epsilon) = P(|X - \mu|^2 \geq \epsilon^2) \leq \frac{1}{\epsilon^2} E((X - \mu)^2) = \frac{1}{\epsilon^2} Var(X)$$

Esta desigualdad es muy útil para *aproximar* cuán concentrada está una v.a. alrededor de su valor esperado.

Sean $X_1, X_2, \ldots, X_n$ v.a. i.i.d. con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$. Denotemos el promedio muestral por

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

Ya que

$$E(\overline{X}_n) = E(\frac{X_1 + X_2 + \ldots + X_n}{n}) = \frac{1}{n} n\mu$$

y

$$Var(\overline{X}_n) = \frac{1}{n^2} Var(X_1 + X_2 + \ldots + X_n) = \frac{\sigma^2}{n},$$

usando la desigualdad de Chebyshev se tiene que

$$P(|\overline{X}_n - \mu| \geq \epsilon) \leq \frac{1}{\epsilon^2} \frac{\sigma^2}{n}$$

Decimos que X_n **converge en probabilidad** a X si $\forall \epsilon > 0$,

$$P(|X_n - X| \geq \epsilon) \longrightarrow 0 \quad \text{cuando} \quad n \longrightarrow \infty$$

.

Teorema (Ley Débil de Grande Números) Sean $X_1, X_2, \ldots, X_n$ v.a. i.i.d. con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2 < \infty$, entonces el promedio muestral $\overline{X}_n$ converge en probabilidad a μ.

La Ley Débil de Grandes Números ofrece un potente método de estimación conocido como el **Método de Monte Carlo**. El siguiente ejemplo ilustra el la idea básica del método.

Ejemplo: Sea (X, Y) un vector aleatorio distribuído uniformemente en el rectángulo $B = [0, a] \times [0, b]$. Sea A un área contenida en B, así,

$$p = P((X, Y) \in A) = \frac{|A|}{ab}$$

Para estimar p, generamos puntos $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. $U[0, a] \times [0, b]$ y consideramos la v.a. Bernoulli

$$\begin{aligned} Z_i &= 1 \quad si \quad (X_i, Y_i) \in A \\ &= 0 \quad sino \end{aligned}$$

Es fácil verificar el promedio $\overline{Z}_n$ converge en probabilidad a p.

10 Función generatriz de momentos

Hemos visto que la esperanza y la varianza dan información acerca de la v.a. Para obtener esta información necesitamos calcular $E[X]$ y $E[X^2]$. El momento de orden k no es mas que la extensión de esta idea para tener mas información acerca de la variable.

Definición: El momento de k-ésimo orden de una v.a. X es $E[X^k]$ siempre y cuando $E[|X|^k] < \infty$.

Ejemplo: Si X tiene distribución Gamma(n, β) entonces el momento de orden k es

$$E[X^k] = \int_0^\infty x^k \frac{\beta^n}{\Gamma(n)} e^{-\beta x} dx = \frac{n(n+1)\ldots(n+k-1)}{\beta^k}$$

Ejemplo: Si X tiene distribución Cauchy entonces

$$E[X^k] = \int_{-\infty}^{\infty} x^k \frac{1}{\pi(1+x^2)} dx$$

En vista de que la integral no converge en valor absoluto para ningún $k \geq 1$ decimos que la distribución de Cauchy no tiene momentos.

Un útil criterio para garantizar la existencia de momentos es:

$$E[|X|^k] < \infty \ \text{ implica } \ E[|X|^r] < \infty \ \text{ para } \ 1 \leq r \leq k$$

En particular, cuando digamos que $Var(X) < \infty$, estaremos diciendo que los momentos de primer y segundo orden existen.

Si lo que estamos buscando es obtener información de un v.a.; formalmente de la distribución de una v.a.; a través de sus momentos, parece natural construir una función que reproduzca el valor de todos los momentos si es que existen:

Definición: La función generatriz de momentos de una v.a. X es la función

$$M_X(t) = E[e^{-tX}] \ \text{ para todo } t \ \text{ para el cual la esperanza existe}$$

Ejemplo: Si X tiene distribución Gamma(n, β) entonces

$$M_X(t) = \left(\frac{\beta}{\beta - t}\right)^n \ \text{ para } \ t \leq \beta$$

Ejemplo: Si X tiene distribución Cauchy entonces $M_X(t)$ solo está definida para $t = 0$.

Ejemplo: Si X tiene distribución Normal(μ, σ^2) entonces

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \ \text{ para todo } t \in R$$

Ejemplo: Si X tiene distribución Poisson(λ) entonces

$$M_X(t) = e^{-\lambda(1-e^t)} \ \text{ para todo } t \in R.$$

Volviendo a nuestro problema, ¿como reproducir los momentos de una v.a. a partir de su función generatriz?:

Teorema: Si M_X existe en un entorno de 0, entonces para cualquier $k \geq 1$

$$E[X^k] = \frac{d^k}{dt^k} M_X(0)$$

La prueba formal de este teorema se escapa del alcance de estas notas. La idea básica es la siguiente:

La esperanza es una sumatoria o una integral dependiendo cual sea el caso que estemos estudiando. Estos operadores (sumatoria e integral) conmutan, bajo condiciones de regularidad, con el operador derivada (¿puede el lector construir ejemplos concretos donde esto ocurre?). Así resulta que

$$\frac{d^k}{dt^k}M_X(t) = \frac{d^k}{dt^k}E[e^{tX}] = E[\frac{d^k}{dt^k}e^{tX}] = E[X^k e^{tX}]$$

evaluando la identidad anterior en $t = 0$ obtenemos el resultado del teorema.

Mas que reproducir los momentos de una distribución, la función generatriz provee una manera de caracterizar distribuciones de probabilidad.

Teorema: Si M_X existe en un entorno de 0, entonces hay una única distribución con función generatriz M_X. Además, bajo esta condición todos los momentos existen, no importa el orden, y el desarrollo de Taylor de la función generatriz es

$$M_X(t) = \sum_{k=0}^{\infty} E[X^k]\frac{t^k}{k!}$$

Del teorema anterior podemos observar que no basta conocer todos los momentos para caracterizar una distribución, es necesario que la serie

$$\sum_{k=0}^{\infty} E[X^k]\frac{t^k}{k!}$$

converja en un entorno de cero, tal como aparece en la hipótesis del teorema.

La prueba de este teorema se basa en propiedades de la transformada de Laplace y la omitiremos en estas notas, sin embargo su aplicación es de gran utilidad al conectarla con las siguientes propiedades:

- Para números a, b se tiene $M_{aX+b}(t) = E[e^{(aX+b)t}] = e^{bt}E[e^{atX}] = e^{tb}M_X(at)$
- Si X, Y son independientes entonces $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$
- Usando recursivamente la propiedad anterior, si $X_1, \ldots, X_n$ son independientes, entonces $M_{X_1+\ldots+X_n} = M_{X_1}\ldots M_{X_n}$

Ejemplo: Sean $X_1, \ldots, X_n$ v.a.i.i.d. exponenciales(β). ¿Cual es la distribución de la suma $S_n = X_1 + \ldots + X_n$?. Sabiendo que la función generatriz de una exponencial es

$$M_{X_1}(t) = \frac{\beta}{\beta - t} \quad \text{para} \quad t \leq \beta$$

podemos usar las propiedades enunciadas y verificar que

$$M_{S_n}(t) = \left(\frac{\beta}{\beta - t}\right)^n \quad \text{para} \quad t \leq \beta$$

que corresponde a la función generatriz de una Gamma(n, β). Usando el teorema anterior podemos concluir que la suma de n exponenciales i.i.d de parámetro β tiene distribución Gamma(n, β).

El lector puede usar esta técnica para probar los siguientes resultados:

- sumas de normales independientes es normal
- sumas de Poisson independientes es Poisson

en general, pruebe que si $(M_X(t))^{1/n}$ es la función generatriz de una v.a. Z, entonces sumas de n réplicas independientes de Z se distribuye como X.

11 Función característica

Como vimos, la función generatriz ofrece un poderoso método para identificar distribuciones de probabilidad. El problema es que la función generatriz no siempre existe, como en el caso de la Cauchy. La clase de distribuciones para la cual no existe es suficientemente grande como para necesitar una alternativa. Usando la fórmula de Moivre

$$e^{i\omega} = \cos\omega + i\mathrm{sen}\omega, \quad \text{con} \quad i = \sqrt{-1}$$

es fácil ver que

$$M_X(i\omega) = E[\cos(\omega X)] + iE[\mathrm{sen}(\omega X)]$$

Ya que las funciones cos y sen son acotadas, la función

$$\phi_X(\omega) = M_X(i\omega)$$

siempre existe y tiene propiedades equivalentes a las de la función generatriz. Por lo cual conviene introducir la siguiente definición

Definición. La función característica de una v.a. X es la función

$$\phi_X(\omega) = E[e^{i\omega X}]$$

En estas notas, suponemos que el lector no está familiarizado con el cálculo en variable compleja. Para calcular $\phi_X(\omega) = E[e^{i\omega X}]$, se puede usar la identidad

$$\phi_X(\omega) = M_X(t) \quad \text{para} \quad t = i\omega$$

cuando $M_X(t)$ existe.

Ejemplo: Si X tiene distribución Gamma(n, β) entonces

$$\phi_X(\omega) = \left(\frac{\beta}{\beta - i\omega}\right)^n.$$

Cuando $M_X(t)$ no existe, por ejemplo para el caso Cauchy, el uso de tablas para la transformada de fourier puede ser muy útil. Casi cualquier función característica ya ha sido tabulada. En particular, si X es Cauchy,

$$\phi_X(\omega) = \frac{1}{2}e^{-|\omega|}.$$

El hecho de que la función característica siempre exista simplifica el criterio de caracterización análogo al que estudiamos para la función generatriz de momentos.

Teorema: Las v.a. X, Y tienen la misma distribución si y solo si $\phi_X(\omega) = \phi_Y(\omega)$ para todo ω. Aún mas, si X es continua con densidad de probabilidad f, la siguiente fórmula puede ser usada para recobrar f a partir de la función característica

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega x}\phi_X(\omega)d\omega$$

Esta fórmula no es mas que la inversa de la transformada de fourier, observando que la función característica

$$\phi_X(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega x} f(x)dx$$

es la transformada de fourier de f.

Para terminar, enunciamos algunas propiedades que el lector puede probar sin dificultad, siguiendo las pruebas análogas para la función generatriz.

- Para números a, b se tiene
  $\phi_{aX+b}(\omega) = e^{i\omega b}\phi_X(a\omega)$
- Si X, Y son independientes entonces
  $\phi_{X+Y}(\omega) = \phi_X(t)\phi_Y(\omega)$
- Usando recursivamente la propiedad anterior, si $X_1, \ldots, X_n$ son independientes, entonces $\phi_{X_1+\ldots+X_n} = \phi_{X_1} \ldots \phi_{X_n}$

12 Convergencia en distribución

Cuando modelamos un fenómeno aleatorio que cambia en el tiempo, debemos considerar sucesiones de v.a. $X_1, X_2, \ldots$ donde X_i modela el estado del fenómeno a tiempo i. El conjunto de índices puede representar unidades de tiempo, iteraciones de un proceso, etc. Es natural que en este tipo de situaciones, si queremos tener una información que no dependa del tiempo, nos preguntemos acerca del comportamiento de X_n cuando $n \to \infty$. Por ejemplo: ¿cual es la distribución de X_n cuando $n \to \infty$?. Antes de formalizar esta idea observemos el siguiente ejemplo:

Consideremos la sucesión de v.a. $X_1, X_2, \ldots$ con $P(X_n = \frac{1}{n}) = 1$. Es decir, con probabilidad 1, $X_n = \frac{1}{n}$. Lo natural sería que $\{X_n\}$ converja a una v.a. X que tiene probabilidad 1 de valer 0. Veamos que pasa con las funciones de distribución de estas v.a.

$$\begin{aligned} F_n(x) = P(X_n \leq x) &= 0 \text{ si } x < \frac{1}{n} \\ &= 1 \text{ si } x \geq \frac{1}{n} \end{aligned}$$

$$\begin{aligned} F(x) = P(X \leq x) &= 0 \text{ si } x < 0 \\ &= 1 \text{ si } x \geq 0 \end{aligned}$$

Por otro lado

$$\begin{aligned}\lim_{n\to\infty} F_n(x) &= 0 \text{ si } x \leq 0 \\ &= 1 \text{ si } x > 0\end{aligned}$$

Es decir, con la excepción de 0, que es un punto de discontinuidad de F,

$$\lim_{n\to\infty} F_n(x) = F(X).$$

Este ejemplo muestra que los puntos de discontinuidad de la distribución límite deben ser ignorados, lo cual nos lleva a la siguiente

Definición: Consideremos la sucesión de v.a. $X_1, X_2, \ldots$ con funciones de distribución $F_1, F_2, \ldots$ respectivamente. Sea X una v.a. con función de distribución F. Diremos que X_n converge en distribución a X, o que $X_n \to X$ en distribución, sii

$$\lim_{n\to\infty} F_n(x) = F(X)$$

para todo x donde F sea continua.

Ejemplo: Considere el mínimo U_n de n variables i.i.d. uniformes en (0,1). Es fácil intuir que pasa con U_n cuando $n \to \infty$. ¿Que pasa con nU_n cuando $n \to \infty$?. Primero que nada, como $U_n \in (0,1)$, entonces $nU_n \in (0,n)$ y

$$P(nU_n \leq x) = P\left(U_n \leq \frac{x}{n}\right) \quad \text{para} \quad x \in (0,n)$$

Usando la definición del mínimo y la independencia, la probabilidad anterior puede escribirse como

$$1 - P(nU_n > x) = 1 - P\left(U_n > \frac{x}{n}\right) = 1 - \left(P(X > \frac{x}{n})\right)^n$$

con X uniforme en (0,1). Sustituyendo,

$$P(nU_n \leq x) = 1 - \left(1 - \frac{x}{n}\right)^n \quad \text{para} \quad x \in (0,n)$$

Usando el hecho de que si $na_n \to a$ entonces

$$\lim(1 - a_n)^n = e^a \tag{5}$$

se obtiene

$$\lim P(nU_n \leq x) = \lim 1 - \left(1 - \frac{x}{n}\right)^n = 1 - e^{-x} \quad \text{para} \quad x \in (0, \infty)$$

que corresponde a la distribución de una variable exponencial de parámetro 1. Es decir, si $X_1, X_2, \ldots$ son v.a.i.i.d. uniformes sobre (0,1), entonces

$$nU_n \to \exp(1) \quad \text{en distribucion.}$$

Cuando las variables toman valores 0, 1, 2, ...; la convergencia en distribución se reduce a la convergencia de la funciones de masa. El lector no tendrá dificultad en probar la siguiente

Proposición: Si $X, X_1, X_2, \ldots$ toman valores enteros y para todo k

$$\lim P(X_n = k) = P(X = k)$$

entonces $X_n \to X$ en distribución.

Ejemplo: Considere la sucesión de v.a. $\{X_n\}$ con X_n binomial(n, p_n). Supongamos que $p_n \to 0$ con $np_n \to \lambda > 0$. Es decir, para n grande, X_n es una binomial con muchas repeticiones de un experimento que tiene muy poco chance de éxito. Podemos probar

$$\lim P(X_n = 0) = e^{-\lambda}$$

$$\lim \frac{P(X_n = k+1)}{P(X_n = k)} = \lambda(k+1)$$

Usando recurrencia, vemos que

$$\lim P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

que es la función de masa de una Poisson(λ). Es decir,

$$X_n \to \text{Poisson}(\lambda) \quad \text{en distribucion.}$$

Los dos ejemplos anteriores muestran la dificultad de probar convergencia en distribución por definición. El siguiente teorema ofrece un método sencillo de hacerlo

Teorema de continuidad: Sean $X_1, X_2, \ldots$ v.a. y $\phi_1, \phi_2, \ldots$ sus respectivas funciones caracteísticas. Si

$$\lim \phi_n(\omega) =: \phi(\omega) \text{ para todo } \omega \in R$$

y $\phi(\omega)$ es continua en $\omega = 0$, entonces ϕ es la función característica de una v.a. X tal que $X_n \to X$ en distribución.

La prueba de este teorema exige de fuertes conocimientos de análisis matemático, sin embargo el teorema en si es facilmente intuible, a excepto de la condición técnica ϕ continua en cero.

Ejemplo: Usar el teorema para probar la convergencia a Poisson del ejemplo anterior.

La aplicación por excelencia del teorema de continuidad es el muy famoso

Teorema del Límite Central: Sean $X_1, X_2, \ldots$ v.a.i.i.d. con varianza finita σ^2. Denotemos por μ la esperanza común de las variables. Entonces

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}} \to N(0, \sigma^2) \text{ en distribucion}$$

13 Cuarto Problemario

1. 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.

2. En la fábrica del problema anterior, supóngase que los amplificadores con $\Gamma < 7.5 \times 10^3$ seg$^{-1/2}$ son rechazados por control de calidad.
   (a) Use la desigualdad de Chebyshev para estimar el % de amplificadores rechazados.
   (b) Calcule la misma probabilidad de la parte (a) usando la tabla de la distribución normal. Explique la discrepancia de los resultados.

3. A través de una encuesta se quiere estimar la fracción p de adultos de la población que se interesaría en un nuevo producto. Se interroga a n personas de la población, y se estima p como $\tilde{p} = X/n$, siendo

X el número de personas encuestadas que manifiestan interés en el producto. Utilizando el Teorema del Límite Central, y suponiendo que el verdadero valor de p es 0.35, encuentre, aproximadamente, el menor valor de n para el cual $\tilde{p}$ y p difieren en menos de 0.02, con probabilidad mayor que 0.9. ¿Como resolvería el problema en el caso (realista) en que p es desconocido?

4. Tomamos 50 números al azar (uniformemente) en el intervalo (1,3).
(a) Utilize la desigualdad de Chebyshev para estimar la probabilidad de que el promedio $\overline{X}$ de estos números se encuentre entre 1.9 y 2.1.
(b) Utilize el Teorema del Límite Central para aproximar la misma probabilidad de la parte (a) Según la aproximación que nos dá el T.L.C., ¿Cuánto debe ser ϵ para que $\overline{X}$ se encuentre en el intervalo $(1-\epsilon, 1+\epsilon)$ con probabilidad 0.95.

5. Use la función generatriz para calcular el momento de tercer orden de una Normal(μ, σ^2).

   2.- Sean $\{X_n\}$ v.a.i.i.d. y N una v.a. a valores enteros positivos independiente de $\{X_n\}$. Calcule la función generatriz de momentos de $X_1 + \ldots + X_N$ y deduzca la esperanza y varianza de esta v.a.

6. $\{X_n\}$ v.a. con

$$P\left(X_n = \frac{k}{n+1}\right) = \frac{1}{n} \quad \text{para} \quad k = 1, \ldots n$$

   Pruebe que $X_n \to$ U(0,1) en distribución.

7. Suponga que X_n tiene distribución Gamma$(n, 1)$. Calcule la función generatriz de momentos de $Z_n = (X_n - n)/\sqrt{n}$ y demuestre que

$$\lim M_{Z_n}(t) = e^{t^2/2}$$

   ¿Que concluye?.