Ingeniería Matemática

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Probabilidad y Estadística MA 3403,30/11/15, Prof. R. Gouet.

Solución Control #3

1. Sean $X_1, \dots, X_n$ v.a. iid normales, de parámetros $\mu = 0, \sigma^2 = 1$. Muestre que la va $X_1^2 + \dots + X_n^2$, $(n \in \mathbb{N}, n \geq 1)$ tiene ley gama de parámetros $\lambda = 1/2$ y $p = n/2$ (χ^2 (chi cuadrado) con n grados de libertad). Ind: demuestre el resultado por inducción sobre n, para lo cual requiere poco más que un cambio de variable en $\mathbb{R}^2$. Alternativamente, si prefiere, puede usar la función generadora de momentos.

   Sol: Usamos la función generadora de momentos (fgm). Calculamos para la ley Gamma de parámetros λ, p. Sea

$$\psi(t;\lambda,p) := \int_0^\infty e^{tx}\frac{\lambda e^{-\lambda x}(\lambda x)^{p-1}}{\Gamma(p)}dx = \frac{\lambda^p}{(\lambda-t)^p}\int_0^\infty \frac{(\lambda-t)^p e^{-(\lambda-t)x}x^{p-1}}{\Gamma(p)}dx = \frac{1}{(1-t/\lambda)^p},\ t<\lambda.$$

   También calculamos la fgm de X^2, donde X es $N(0,1)$. Sea

$$\phi(t) = E(e^{tX^2}) = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{tx^2-x^2/2}dx = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2(1-2t)}dx = \frac{1}{(1-2t)^{1/2}}.$$

   Al comparar con la fgm de la ley Gamma, vemos que X^2 es $\Gamma(\frac{1}{2},\frac{1}{2})$. Esto indica que S_1 $\chi^2(1)$ y tenemos el primer paso de la inducción.

   Consideremos ahora S_{n+1} y supongamos que S_n es $\chi^2(n)$. Entonces, dado que $S_{n+1} = S_n + X_{n+1}^2$ y que S_n y X_{n+1} son independientes, la fgm de S_{n+1} se obtiene multiplicando la fgm de S_n con la de ${}^2_{n+1}$. Entonces, designado por $\varphi_n(t)$ la fgm de S_n tenemos la ecuación

$$\varphi_{n+1}(t) = \varphi_n(t)\phi(t) = \psi(t;\frac{1}{2},\frac{n}{2})\phi(t) = \frac{1}{(1-2t)^{n/2}}\frac{1}{(1-2t)^{1/2}} = \frac{1}{(1-2t)^{(n+1)/2}},$$

   lo cual muestra que S_{n+1} es $\chi^2(n+1)$. Con esto se completa el argumento inductivo.

2. Se escogen dos puntos P_1, P_2 al azar en el perímetro de un un cuadrado unitario.

   a) (2 pts.) Calcule la probabilidad de que P_1, P_2 estén: (i) en el mismo lado; (ii) en lados adyacentes; (iii) en lados opuestos del cuadrado.

   Sol: Primero numeramos los 4 lados del cuadrado, comenzando por la base (1), luego el lado vertical derecho (2); el horizontal superior (3) y el vertical izquierdo (4). Definimos las va $L_i, i = 1,2$, donde L_i es el lado donde cae P_i. Debido a que P_1, P_2 son escogidos independientemente, las va L_i son independientes. Por otra parte, como los P_i son puntos al azar, tienen igual probabilidad de caer en cualquiera de los lados. Por lo tanto $P(L_i = j) = 1/4$, para cualquier $i = 1,2; j = 1,\dots,4$. Además, $P(L_1 = j_1, L_2 = j_2) = 1/16$, para cualquier par $j_1, j_2 = 1,\dots,4$. Ahora consideramos los sucesos M, A, O que indican que P_1, P_2 caen respectivamente en el mismo lado, en lados adyacentes y en lados opuestos. Vemos que A es equivalente a $L_1 = L_2$, lo cual tiene probabilidad $P(M) = 1/4$. El suceso lados adyacentes equivale a que $\{L_2 = L_1 + 1 \mod 4\} \cup \{L_1 = L_2 + 1 \mod 4\}$ o bien, que (L_1, L_2) tome valores en

$$\{(1,2),(2,1),(2,3),(3,2),(3,4),(4,3),(4,1),(1,4)\}.$$

   De lo anterior se concluye que $P(A) = 1/2$. Finalmente, los pares (L_1, L_2) para el suceso O son $\{(1,3),(3,1),(2,4),(4,2)\}$ y entonces, $P(O) = 1/4$.

b) (4 pts.) Sea Y la distancia (euclidiana en $\mathbb{R}^2$) entre P_1 y P_2. Calcule la densidad de probabilidad de Y. Ind: condicione en la ubicación relativa de los puntos.

Condicionamos en (L_1, L_2), es decir, consideramos las funciones de distribución condicionales $F_Y(y|j_1, j_2) = P(Y \leq y|L_1 = j_1, L_2 = j_2)$, para todas las combinaciones posibles de j_1, j_2, donde $Y = d(P_1, P_2)$ es la distancia entre P_1 y P_2. Luego, por la regla de probabilidades totales, obtenemos $F_Y(y)$ sumando sobre los pares j_1, j_2. Tenemos

$$F_Y(y) = \sum_{j_1,j_2=1}^{4} F_Y(y|j_1, j_2)P(L_1 = j_1, L_2 = j_2) = \frac{1}{16}\sum_{j_1,j_2=1}^{4} F_Y(y|j_1, j_2).$$

La suma anterior tiene solo 3 términos distintos, que corresponden a los sucesos M, A, O. Entonces podemos escribir

$$F_Y(y) = F_Y(y|1, 1)\frac{1}{4} + F_Y(y|4, 1)\frac{1}{2} + F_Y(y|1, 3)\frac{1}{4}.$$

Examinamos los 3 casos: 1) P_1, P_2 caen el mismo lado (suceso M) entonces sus posiciones están descritas por dos va uniformes en $[0, 1]$ independientes, digamos U, V, y la distancia es simplemente $Y = |U - V|$. Por lo tanto,

$$F_Y(y|1, 1) = P(|U - V| \leq y).$$

En el caso 2) los puntos caen en lados adyacentes y nuevamente sus posiciones están caracterizadas por va U, V iid uniformes en $[0, 1]$. Sin embargo, ahora la distancia se escribe como $Y = \sqrt{U^2 + V^2}$. Por lo tanto,

$$F_Y(y|4, 1) = P(\sqrt{U^2 + V^2} \leq y).$$

Finalmente, en el caso 3) los puntos están en lados opuestos y nuevamente descritos por las va U, V. Su distancia ahora es $Y = \sqrt{1 + (U - V)^2}$ y

$$F_Y(y|1, 3) = P(\sqrt{1 + (U - V)^2} \leq y).$$

Ahora debemos calcular las tres probabilidades con las uniformes U, V. Para ello notamos que la probabilidad $P(|U - V| \leq y)$ es el área de la región $\{(u, v) \in [0, 1]^2||u - v| \leq y\}$. Dicha área es y^2, para $y \in [0, 1]$. En el caso 2) consideramos la región $\{(u, v) \in [0, 1]^2|\sqrt{u^2 + v^2} \leq y\}$, cuya área es $\pi y^2/4$, para $y \in [0, 1]$ y es igual a

$$a(y) := \int_{\sqrt{y^2-1}}^{1} (1 - \sqrt{y^2 - u^2})du.$$

Finalmente, el área en el caso 3) calculamos el área de $\{(u, v) \in [0, 1]^2|\sqrt{1 + (u - v)^2} \leq y\}$, que vale 0 cuando $y \in [0, 1]$. Para valores $y \in (1, \sqrt{2})$ la región es equivalente a $\{(u, v) \in [0, 1]^2||u - v| \leq \sqrt{y^2 - 1}\}$, cuya área es $\sqrt{y^2 - 1}$. Para valores de y mayores que los señalados, el área es 1 en todos los casos, es decir, las distribuciones condicionales valen 1 en ese rango y, al derivar para obtener la densidad, obtenemos 0. Derivando F_Y obtenemos finalmente

$$f_Y(y) = 2y\mathbf{1}_{[0,1]}(y)\frac{1}{4} + \pi\frac{y}{2}\mathbf{1}_{[0,1]}(y) + a'(y)\mathbf{1}_{[1,\sqrt{2}]}(y))\frac{1}{2} + \frac{y}{\sqrt{y^2 - 1}}\mathbf{1}_{[1,\sqrt{2}]}(y)\frac{1}{4}.$$

$$= \frac{1}{2}(1 + \pi/2)y\mathbf{1}_{[0,1]}(y) + (a'(y)/2 + \frac{y}{4\sqrt{y^2 - 1}})\mathbf{1}_{[1,\sqrt{2}]}(y).$$

3. Una fábrica produce un 5 % de artículos defectuosos. De qué tamaño se debe tomar una muestra aleatoria de artículos de manera que la proporción de defectuosos en la muestra sea inferior al 6 % con probabilidad al menos 0.99? Ind: Aplique el TCL para aproximar la probabilidad y despeje n de la desigualdad resultante.

   Sol: En n artículos tenemos $\sum_{i=1}^n X_i$ defectuosos, donde las X_i son va iid Bernoulli con parámetro $p = 5/100$. La probabilidad de tener menos de 6 % de defectuosos es

$$P(\sum_{i=1}^n X_i/n < 6/100) = P(\sum_{i=1}^n X_i - 5n/100 < n/100).$$

   La varianza del número de defectuosos es $np(1-p) = 5 \cdot 95n/10000$. La desviación standard es $\sqrt{475n}/100$ y la incorporamos en la expresión de la probabilidad para invocar el TCL.

$$P(\sum_{i=1}^n X_i - 5n/100 < 6n/100) = P\left(\frac{\sum_{i=1}^n X_i - 5n/100}{\sqrt{475n}/100} < \frac{n/100}{\sqrt{475n}/100}\right) \sim \Phi\left(\frac{n/100}{\sqrt{475n}/100}\right) > 0{,}99.$$

   La inecuación anterior equivale a

$$\frac{\sqrt{n}}{\sqrt{475}} > \Phi^{-1}(0{,}99) \Leftrightarrow n > 475\Phi^{-1}(0{,}99)^2 \sim 2592.$$