Universidad Simón Bolívar
Curso de Probabilidades
Enero-Marzo 2010

Problemario III

1. La variable X toma los valores 1, 2 y 3 con la misma probabilidad. Una vez observado el valor de X, digamos, $X = x$, Y se genera con una distribución de Poisson de parámetro $\lambda = x$.
(a) Hallar la función de probabilidad conjunta $p(x, y)$ para todo $x \in \{1, 2, 3\}$, $y \in \{1, 2, \ldots\}$.
(b) Hallar $\Pr(X + Y \geq 4)$.
(c) Hallar $\Pr(Y > 2 \mid X = 2)$.

2. La flecha lanzada por un arquero, caerá en un punto del blanco, cuyas coordenadas X e Y (en pies) son N(0,1) independientes. El punto $(0, 0)$ coincide con el centro del blanco. Llamemos R a la distancia de (X, Y) al centro del blanco. El arquero paga C\$ 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 recibe un dolar y, finalmente, si $R \geq 1$ el arquero no recibe ningún pago. ¿Cuanto debe valer C para que el valor esperado de la ganancia del arquero sea -1.0 dólares?

3. En condiciones de descuido severo y humedad excesiva, sobre la superficie de un CD, de radio ρ cms, aparece una mancha producida por un hongo. La mancha daña el contenido del CD, si su distancia R al centro del CD cumple $\rho_1 < R < \rho_2$, siendo ρ_1 y ρ_2 radios comprendidos entre 0 y ρ, que definen la región sobre la cual se guarda información. Suponga, por simplicidad, que el disco no tiene un agujerito en el centro. Si la mancha aparece al azar (con distribución uniforme) sobre la superficie del disco, (a) Hallar la probabilidad de que la información en el mismo resulte dañada. (b) Hallar la densidad de probabilidad de la coordenada X del punto sobre el disco donde aparece la mancha (tome el centro del disco como origen en su sistema de coordenadas). (c) Hallar la densidad condicional $f(y|x)$ para la coordenada Y cuando se conoce la coordenada X del punto donde aparece la mancha.

4. En condiciones ideales, la temperatura T, presión S y volumen V, de un gas, están relacionados mediante $T = \alpha\, S\, V$, siendo α una constante de proporcionalidad. Si S y V son variables aleatorias con densidad conjunta:

$$f(s, v) = \begin{cases} \frac{1}{s^2 v^2} & \text{si } s \geq 1,\ v \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

(a) Hallar $\Pr(T/\alpha \leq 2)$.
(b) Hallar la función de distribución acumulativa y la densidad de la variable T/α.
(c) Hallar la densidad marginal de S.
(d) ¿Son S y V independientes?

5. El par (X, Y) tiene distribución Unif($\mathcal{R}$), siendo $\mathcal{R}$ el triángulo de vértices (0,0), (1,0) y (0,2).
(a) Hallar $\Pr(Y > X)$.
(b) Hallar la densidad marginal $f_Y(y)$.
(c) Hallar la densidad condicional $f_{X|Y}(x|y)$.

6. Dos números son seleccionados independientemente y al azar en el intervalo $[0, 1]$. Sabemos que el menor de los dos es menor que 1/3. ¿Cuál es la probabilidad de que el mayor sea mayor que 2/3?

7. Se elige un punto al azar, uniformemente, en la bola unitaria 3-dimensional: $\{(x, y, z) \in \boldsymbol{R}^3, x^2 + y^2 + z^2 \leq 1\}$. Hallar la f.d.a. y la densidad de su distancia al origen.

8. La variable aleatoria X tiene f.d.a. $F(x) = x^r$, $0 \leq x \leq 1$, siendo r un número natural. Dado $X = x$, la variable Y tiene una distribución $\mathrm{Bin}(n, x)$. Hallar $\boldsymbol{E}(Y)$.

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

10. Los tiempos de vida, $X_1, \ldots, X_n$ de n circuitos integrados que forman parte de un módulo, son variables aleatorias independientes con distribución exponencial de parámetro $\lambda = 1.12 \times 10^{-12}\mathrm{seg}^{-1}$. Hallar la distribución de la variable correspondiente al tiempo en que, por vez primera, se daña uno de los circuitos.

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

12. Se eligen, de manera independiente, 5 números con distribución $U[a, b]$.
(a) Hallar la probabilidad de que todos ellos sean mayores que $\frac{a+b}{2}$.
(b) Hallar la probabilidad de que el máximo de los números elegidos no pase de $(a + 9\,b)/10$.

13. Debido a variabilidad en el proceso de producción, el *coeficiente de desgaste* Υ, de los amplificadores producidos por una fabrica, tiene una distribución

Unif$(0.1, 1.0)$. A su vez, el tiempo de vida T, en años, de un amplificador con coeficiente de desgaste Υ, tiene distribución exponencial de parámetro Υ. Hallar el tiempo de vida promedio de los amplificadores producidos.

14. Considere el punto aleatorio (X, Y) con coordenadas independientes N(0,1). Sean (R, Θ) las coordenadas polares del punto aleatorio (X, Y). Hallar la densidad conjunta del par (R, Θ) y las marginales de R^2 y Θ.

15. Si las variables aleatorias X e Y cumplen $\sigma_X^2 = \mathrm{Var\,X} = 1$ y $\sigma_Y^2 = \mathrm{Var\,Y} = 2$,
(a) ¿Cuanto debe valer $\rho_{X,Y}$ para tener $\mathrm{Var(X + Y)} = 5$?
(b) Pruebe que $\mathrm{Var(X + Y)}$ no puede ser igual a 6.
(c) ¿Cuál es el rango de valores posibles para $\mathrm{Var(X + Y)}$?

16. Sean X, Y v.a. independientes con distribución gamma de parámetros (n, θ) y (m, θ) respectivamente, para n y m enteros positivos. Considere las variables
$$U = X + Y \qquad V = \frac{X}{X + Y}$$
Demuestre que U, V son independientes y calcule sus densidades marginales.

17. 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 L'imite 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. ¿Cómo resolvería el problema en el caso realista en que p es desconocido?

18. Un dado es lanzado 12.000 veces. Use el teorema del límite central para hallar, aproximadamente, $\Pr(1900 < S < 2200)$, siendo S el número total de "unos" obtenidos en los lanzamientos.

19. Tomamos 100 números al azar (uniformemente) en el intervalo (0,2). Utilize el Teorema del Límite Central para estimar la probabilidad de que el promedio $\overline{X}$ de estos números se encuentre entre 0.9 y 1.1. 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?

Soluciones

Para algunos problemas se incluyen soluciones completas; para otros solo el resultado.

1. (a) Del enunciado tenemos que la distribución condicional de Y, dado $X = x$, es Poisson(x). Es decir

$$p_{Y|X}(y|x) = \frac{\mathrm{e}^{-x}x^y}{y!}, \quad \text{para } y = 0, 1, \ldots$$

así,

$$p(x, y) = p_X(x)p_{Y|X}(y|x) = \frac{\mathrm{e}^{-x}x^y}{3y!},$$

para $x \in \{1, 2, 3\}$, $y \in \{0, 1, 2, \ldots\}$.

(b) Descomponemos el evento en tres eventos disjuntos:

$$\Pr(X + Y \geq 4) = \Pr(X = 1, Y \geq 3) + \Pr(X = 2, Y \geq 2) + \Pr(X = 3, Y \geq 1).$$

Ahora bien, $\Pr(X = 1, Y \geq 3) = \Pr(X = 1)\Pr(Y \geq 3|X = 1) =$
$\frac{1}{3}(1 - \mathrm{e}^{-1}(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!})) \sim 0.02677$.
Similarmente, $\Pr(X = 2, Y \geq 2) = (1/3)(1 - \Pr(Y \leq 1|X = 2)) = 0.198$ y
$\Pr(X = 3, Y \geq 1) = (1/3)(1 - \Pr(Y = 0|X = 3)) = 0.3167$. Así,

$$\Pr(X + Y \geq 4) = 0.02677 + 0.198 + .03167 = 0.5415$$

(c) $Pr(Y \geq 2|X = 2) = 1 - \Pr(Y \leq 1|X = 2) = 1 - \mathrm{e}^{-2}(\frac{2^0}{0!} + \frac{2^1}{1!}) = 0.594$

2. Buscamos primeramente la función de distribución acumulativa (f.d.a.) de la variable R. Claramente, Rango$(R) = [0, \infty)$ y, para $r > 0$, usando que las variables X e Y son independientes, tenemos

$$F_R(r) = \Pr(R \leq r) = \int\int_{\{x^2+y^2 \leq r^2\}} \frac{1}{2\pi}\mathrm{e}^{-(1/2)(x^2+y^2)}\mathrm{d}y\mathrm{d}x =$$
$$\overset{x=\rho\cos(\theta),\, y=\rho\mathrm{sen}(\theta)}{=} \int_0^r \int_0^{2\pi} \frac{1}{2\pi}\rho\mathrm{e}^{-(1/2)\rho^2}\mathrm{d}\theta\mathrm{d}\rho = 1 - \mathrm{e}^{-r^2/2}.$$

Con la f.d.a. y considerando que R es una variable continua, podemos calcular
$\Pr(R < 0.2) = 1 - \mathrm{e}^{-0.2^2/2} = 0.0198$;
$\Pr(0.2 \leq R < 0.5) = \mathrm{e}^{-0.2^2/2} - \mathrm{e}^{-0.5^2/2} = 0.0977$;
$\Pr(0.5 \leq R < 1) = \mathrm{e}^{-0.5^2/2} - \mathrm{e}^{-1^2/2} = 0.276$ y

$\Pr(R \geq 1) = 1-0.0198-0.0977-0.276 = 0.6065$. Con lo que podemos plantear la esperanza pedida:

$$-1 = (50 - C) \times 0.0198 + (10 - C) \times 0.0977 + (1 - C) \times 0.276 - C \times 0.6065,$$

de donde $C = 1 + 2.243 = 3.243$ dólares.

3. (a) $\frac{\rho_2^2-\rho_1^2}{\rho^2}$ (b) $f_X(x) = \frac{2\sqrt{\rho^2-x^2}}{\pi\rho^2}$, $-\rho \leq x \leq \rho$
(c) Dado $X = x$, $Y \sim$ Unif$(-\sqrt{\rho^2 - x^2}, \sqrt{\rho^2 - x^2})$.

4. (a) $\Pr(T/\alpha \leq 2) = \Pr(SV \leq 2) = \int_1^2 \int_1^{2/s} \frac{1}{s^2v^2}\mathrm{d}v\mathrm{d}s = 0.5(1 - \ln 2)$. (Hacer un dibujo para entender los límites de integración).
(b) Observamos que Rango$(T/\alpha) = [1, \infty)$. Para $t \in [1, \infty)$, tenemos

$$F_{T/\alpha}(t) = \Pr(T/\alpha \leq t) = \Pr(SV \leq t) = \int_1^t \int_1^{t/s} \frac{1}{s^2v^2}\mathrm{d}v\mathrm{d}s = 1 - \frac{1}{t}(1 + \ln(t)),$$

de donde, al derivar, obtenemos $f_{T/\alpha}(t) = \ln(t)/t^2$ para $t \geq 1$.
(c) $f_S(s) = \int_1^\infty \frac{1}{s^2v^2}\mathrm{d}v = \frac{1}{s^2}$, para $s \geq 1$.
(d) El rango de valores del par (S, V) es el "rectángulo" de lados paralelos a los ejes: $S \geq 1, V \geq 1$, y en todo ese rango la densidad conjunta es factorizable como $\frac{1}{s^2} \times \frac{1}{v^2} = g(s)h(v)$. Por tanto las variables S y V son independientes.

5. El triángulo considerado tiene área 1, por lo que la densidad uniforme vale 1 dentro de $\mathcal{R}$. El segmento que une los puntos (1,0) y (0,2) pertenece a la recta $y = 2 - 2x$, que se intersecta con $y = x$ en el punto (2/3,2/3). Por lo tanto

$$\Pr(Y > X) = \int_0^{2/3} \int_x^{2-2x} \mathrm{d}y\mathrm{d}x = \frac{2}{3}.$$

(b) Para un valor fijo de y entre 0 y 2, x dentro de $\mathcal{R}$ varía entre 0 y $(2 - y)/2$ (Hacer un dibujo para entender los límites). Por tanto

$$f_Y(y) = \int_0^{(2-y)/2} \mathrm{d}x = \frac{2 - y}{2}, \text{ para } y \in [0, 2]$$

(c) $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{2-y}$, $x \in [0, (2 - y)/2]$. Por tanto la distribución condicional de X dado Y es uniforme en su rango.

6. Por la independencia entre las variables, la densidad conjunta de (X, Y) vale $1 \times 1 = 1$ en el cuadrado unitario, C. Por lo tanto el par (X, Y) tienen distribución uniforme en el cuadrado. El evento E_1 ="El menor de los dos es menor que 1/3" es la unión de las franjas $x < 1/3$ y $y < 1/3$ en C. Esta unión tiene área (y por tanto probabilidad) 1/3 + 1/3 - 1/9=5/9. El evento E_2 ="El

mayor de los dos es mayor que 2/3” es la unión de las franjas $x > 2/3$ y $y > 2/3$. La intersección de las franjas, $E_1 \cap E_2$ es la región:
$\{x < 1/3, y > 2/3\} \cup \{x > 2/3, y < 1/3\}$, con área 2/9. Entonces

$$\Pr(E_2|E_1) = (2/9)/(5/9) = 2/5.$$

7. $F_R(r) = r^3$, para $r \in [0,1]$ y $f_R(r) = 3\,r^2$, en ese mismo rango.

8. El enunciado nos da la distribución condicional de Y dado $X = x$. Como esta es $\mathrm{Bin}(n,x)$, tenemos que $\boldsymbol{E}\,(Y|X = x) = nx$. Por otro lado, la densidad de la variable X es (al derivar) $f_X(x) = rx^{r-1}$ para $x \in [0,1]$. Usamos ahora la fórmula de esperanzas iteradas (también llamada propiedad torre)
$$\boldsymbol{E}\,(Y) = \boldsymbol{E}_X\,(\boldsymbol{E}\,(Y|X = x)) = \boldsymbol{E}_X(nx) = \int_0^1 nxrx^{r-1}\mathrm{d}x = \frac{n\,r}{r+1}.$$

9. El triángulo T considerado está limitado por el eje y, la recta $y = x$ y la recta $y = 2 - x$ y tiene área 1, por lo que la densidad uniforme vale 1 en T. $\mathrm{Rango}(X) = [0,1]$ y $\mathrm{Rango}(Y) = [0,2]$. Para un valor fijo de $x \in [0,1]$, y va de x a $2 - x$ (hacer un dibujo). Por tanto:
$$f_X(x) = \int_x^{2-x} \mathrm{d}y = 2 - 2x, \text{ para } x \in [0,1].$$
Luego, para cada $x \in (0,1)$
$$f_{Y|X}(y|x) = \frac{1}{f_X(x)} = \frac{1}{2-2x} \text{ para } y \in (x, 2-x).$$
Observe que como la densidad condicional de Y no depende de la variable y, la distribución condicional es uniforme, en el intervalo correspondiente. Entonces,
$$\boldsymbol{E}\,(Y|X = x) = \int_x^{2-x} \frac{1}{2-2x}\mathrm{d}y = 1,$$
que no depende de x, pero las variables X e Y son dependientes, pues $\mathrm{Rango}(X,Y)$ no es un rectángulo de lados paralelos a los ejes.

10. La variable Y a estudiar es el mínimo de los X_i. Buscamos primero,
$$1 - F_Y(y) = \Pr(Y > y) = \Pr(\min\{X_1, X_2, \ldots, X_n\} > y)$$
Observamos que el mínimo es mayor que y si, y solo si, todos los X_i son mayores que y. Tomando en cuenta la independencia de los X_i obtenemos
$$\begin{aligned} 1 - F_Y(y) = &\quad \Pr(Y > y) = \Pr(X_1 > y, X_2 > y, \ldots, X_n > y) \overset{\text{indep.}}{=} \\ = &\quad \textstyle\prod_{i=1}^n \Pr(X_i > y) = (\text{haciendo la integral})(\mathrm{e}^{-\lambda y})^n = \mathrm{e}^{-n\lambda y} \end{aligned}$$
De esta forma, $F_Y(y) = 1 - \mathrm{e}^{-n\lambda y}$ y, al derivar, $f_Y(y) = n\lambda\mathrm{e}^{-n\lambda y}$, con lo que concluimos que $Y \sim \exp(n\lambda)$.

11. El procedimiento de la demostración es basicamente el mismo del problema anterior.

12. (a) $1/2^5$
(b) La probabilidad de que uno de los números no pase de $(a+9\,b)/10$ la calculamos como "longitud favorable" sobre "longitud total":

$$\Pr(X_i \leq (a+9\,b)/10) = \frac{(a+9\,b)/10 - a}{b-a} = \frac{9}{10}$$

Para que el máximo no pase de la cantidad indicada, todos los X_i deben quedar por debajo de $(a+9\,b)/10$, por lo que la probabilidad pedida es $(0.9)^5$=0.59

13. Del enunciado tenemos que la distribución condicional de T dado $\Upsilon = \upsilon$ es $\exp(\upsilon)$, por lo que (de las propiedades de la distribución exponencial), $\boldsymbol{E}\,(T|\Upsilon=\upsilon) = 1/\upsilon$. Aplicamos ahora la fórmula de esperanzas iteradas:

$$\boldsymbol{E}\,(T) = \boldsymbol{E}_{\Upsilon}\boldsymbol{E}\,(T|\Upsilon=\upsilon) = \boldsymbol{E}_{\Upsilon}\left(\frac{1}{\upsilon}\right) = \frac{1}{0.9}\int_{0.1}^{1}\frac{1}{\upsilon}\,\mathrm{d}\upsilon = 2.558$$

El tiempo de vida promedio de los amplificadores producidos es 2.558 años.

14. Por independencia, la densidad conjunta del par (X,Y) es $f(x,y) = \frac{1}{2\pi}\mathrm{e}^{-(1/2)(x^2+y^2)}$ para $x,y\in\boldsymbol{R}$. Las variables (X,Y), en función de las coordenadas polares, vienen dadas por $x = h_1(r,\theta) = r\cos(\theta)$; $y = h_2(r,\theta) = r\mathrm{sen}(\theta)$. El Jacobiano de la transformación $h = (h_1,h_2)$ tiene determinante r. Entonces, aplicando el Teorema de cambio de variables para densidades conjuntas, tenemos

$$f_{R,\Theta}(r,\theta) = \frac{1}{2\pi}r\mathrm{e}^{-\frac{1}{2}(r^2\cos^2(\theta)+r^2\mathrm{sen}^2(\theta))} = \frac{1}{2\pi}r\mathrm{e}^{-\frac{1}{2}r^2} \text{ para } r>0, \theta\in[0,2\pi).$$

Como esta densidad conjunta es factorizable, en $s_1(\theta) = \frac{1}{2\pi}$ y $s_2(r) = r\mathrm{e}^{-\frac{1}{2}r^2}$ $\forall r>0, \forall\theta\in[0,2\pi)$ vemos que las coordenadas polares (R,Θ) son independientes. Asimismo, vemos que la distribución de Θ es uniforme en su rango de valores.

15. (a) Tenemos que $\mathrm{Var}(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\mathrm{Cov}(X,Y)$. De la fórmula de correlación, despejamos $\mathrm{Cov}(X,Y) = \rho\,\sigma_X\sigma_Y$. Al sustituir los datos, obtenemos

$$5 = 1 + 2 + 2\rho\sqrt{2}$$

de donde $\rho = \sqrt{2}/2$.
(b) Si suponemos ahora que $\mathrm{Var}(X+Y) = 6$, haciendo el mismo cálculo, resulta $\rho = 3/(2\sqrt{2}) > 1$, lo cual no puede ocurrir.

(c) Como $-1 \leq \rho \leq 1$, sustituyendo esta desigualdad en la fórmula de $\mathrm{Var}(X+Y)$ obtenemos

$$\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y \leq \mathrm{Var}(X+Y) \leq \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y,$$

con lo cual $3 - 2\sqrt{2} \leq \mathrm{Var}(X+Y) \leq 3 + 2\sqrt{2}$.

16. Como X y Y viven en $(0,\infty)$, resulta $\mathrm{Rango}(U) = (0,\infty)$, mientras que V cae en (0,1) (porque $X < X+Y$). Fijado $U = u$, el par (X,Y) se mueve en el segmento de la recta $x+y=u$ entre los puntos $(u,0)$ y $(0,u)$ (sin incluirlos). Vemos que, en este segmento, el valor de V varía entre 0 y 1. En otras palabras, el Rango de V no está afectado por el valor de U. Es decir, $\mathrm{Rango}(U,V) = (0,\infty)\times(0,1)$. La transformación $(X,Y) \rightsquigarrow (U,V)$ es invertible de $(0,\infty)\times(0,\infty)$ a $(0,\infty)\times(0,1)$ (esto lo sabemos pues podemos calcular la inversa de manera única) con inversa $X = UV$, $Y = U(1-V)$. Y el Jacobiano de esta transformación inversa es

$$\begin{pmatrix} v & u \\ 1-v & -u \end{pmatrix}$$

cuyo determinante es $-u$. Por independencia, obtenemos la conjunta de (X,Y) multiplicando las marginales:

$$f_{X,Y}(x,y) = \frac{1}{\Gamma(n)\Gamma(m)\theta^{n+m}} x^{n-1} y^{m-1} \mathrm{e}^{-(x+y)/\theta}, \text{ para } x>0,\ y>0.$$

Aplicamos ahora el Teorema de Cambio de Variables para obtener

$$f_{U,V}(u,v) = \frac{1}{\Gamma(n)\Gamma(m)\theta^{n+m}} (uv)^{n-1} (u(1-v))^{m-1} \mathrm{e}^{-u/\theta} u, \text{ para } u>0, 0<v<1.$$

Esta densidad conjunta puede factorizarse como $s_1(u)s_2(v)$ siendo $s_1(u) = c_1 u^{n+m-1}\mathrm{e}^{-u/\theta}$ y $s_2(v) = c_2 v^{n-1}(1-v)^{m-1}$ para constantes c_1, c_2. Por lo tanto U y V son independientes y la densidad conjunta es el producto de las marginales. Identificamos $s_1(u)$ como una densidad $\mathrm{Gamma}(n+m,\theta)$, por lo que, para que s_1 sea la densidad marginal correspondiente, debemos tomar

$$c_1 = \frac{1}{\Gamma(n+m)\theta^{n+m}}.$$

Finalmente, despejamos c_2 de la ecuación $f_{U,V}(u,v) = s_1(u)s_2(v)$ y obtenemos

$$s_2(v) = \frac{\Gamma(n+m)}{\Gamma(n)\Gamma(m)} v^{n-1}(1-v)^{m-1}, \text{ para } 0<v<1,$$

como la densidad marginal de V. Esta densidad corresponde a la distribución $\mathrm{Beta}(n,m)$.

17. Llamemos X_i a la siguiente variable

$$X_i = \begin{cases} 1 & \text{Si la i-ésima persona encuestada se interesa en el producto} \\ 0 & \text{en caso contrario.} \end{cases}$$

Las X_i tienen distribución Bern(p) y, si la población total es suficientemente grande, podemos suponer que son, aproximadamente, independientes (trabajamos como si el muestreo fuese con remplazo). A partir de la definición de las X_i, obtenemos facilmente, $\boldsymbol{E}(X_i) = p$ y $\text{Var}(X_i) = p(1-p)$. Tenemos, además, que $\tilde{p} = (X_1 + \ldots X_n)/n$. Por tanto, el Teorema del Límite Central, nos dice que

$$Z = \frac{\sqrt{n}(\tilde{p} - p)}{\sqrt{p(1-p)}} \sim N(0,1), \text{ aproximadamente.}$$

Suponiendo que $p = 0.35$,

$$\Pr(|\tilde{p} - p| < 0.02) = \Pr\left(\frac{\sqrt{n}|\tilde{p} - p|}{\sqrt{p(1-p)}} < \frac{\sqrt{n}0.02}{\sqrt{0.2275}}\right) \simeq 2\Phi\left(\frac{\sqrt{n}0.02}{\sqrt{0.2275}}\right) - 1,$$

siendo Φ la f.d.a. de la $N(0,1)$. (Recuerde que por la simetría de la densidad normal, se tiene, para $x > 0$, $\Phi(x) - \Phi(-x) = 2\Phi(x) - 1$).
Queremos que la probabilidad de que $\tilde{p}$ y p difieran en menos de 0.02 sea al menos 0.9. Usando la aproximación dada por el teorema, buscamos que $2\Phi\left(\frac{\sqrt{n}0.02}{\sqrt{0.2275}}\right) - 1 \geq 0.9$, es decir $\Phi\left(\frac{\sqrt{n}0.02}{\sqrt{0.2275}}\right) \geq 0.95$ De la tabla obtenemos que $\Phi(1.65) \simeq 0.95$, por lo que, la condición sobre n resulta

$$\frac{\sqrt{n}0.02}{\sqrt{0.2275}} \geq 1.65 \Leftrightarrow n \geq 0.2275(1.65/0.02)^2 = 1548.4$$

Luego, aproximadamente, se requiere una muestra de 1549 personas para lograr el margen de error deseado.

En el caso en que p es desconocido, se puede hacer todo el análisis, hasta la última ecuación, arrastrando el factor desconocido $p(1-p)$ (que arriba era 0.2275). Se llega a la condición $n \geq p(1-p)(1.65/0.02)^2$. Siendo conservadores, consideramos el peor caso, que corresponde al p que maximiza el tamaño de la muestra requerida, el cual es $p = 1/2$. (Allí alcanza su máximo la parábola $p(1-p)$). Resulta entonces $n \geq (1/4)(1.65/0.02)^2 \simeq 1701.6$, y el tamaño de la muestra necesaria queda en 1702.

18. Usamos ahora la versión "suma" del T.L.C. El número de "unos" obtenidos en 12000 lanzamientos es una variable Bin(12000,1/6). Es a la vez, la suma de 12000 Bernoullis independientes de parámetro 1/6, que podemos llamar, como

en el problema anterior, X_i's. Cada una de las X_i's tiene esperanza 1/6 y varianza (1/6)*(5/6). Su suma Binomial, tiene esperanza $\mu = 12000/6 = 2000$ y varianza $\sigma^2 = 12000 \times 5/36 \approx 1667$. Tenemos entonces que

$$Z = \frac{S - 2000}{\sqrt{1667}} \sim \mathrm{N}(0, 1), \text{ aproximadamente.}$$

Entonces,

$$\begin{aligned} \Pr(1900 < S < 2200) &= \Pr(\frac{1900 - 2000}{40.83} < Z < \frac{2200 - 2000}{40.83}) = \\ \Pr(-2.449 < Z < 4.898) &= \Phi(4.898) - \Phi(-2.449) = \Phi(4.898) + \Phi(2.449) - 1, \end{aligned}$$

haciendo la sustitución $\Phi(-2.449) = 1 - \Phi(2.449)$. $\Phi(4.898)$ no aparece en la tabla y por estar a la derecha de 3.5, aproximamos $\Phi(4.898)$ a 1 (en realidad vale 0.9999995). De la tabla obtenemos $\Phi(2.449) \approx 0.9928$, por lo que resulta $\Pr(1.900 < S < 2.200) \approx 0.9928$. Es muy probable que la cantidad de "unos" obtenidos no se salga del intervalo indicado.

19. (a) Para la Unif(0,2), $\mu = 1$ y $\sigma^2 = (2 - 0)^2/12 = 1/3$. El T.L.C. nos dice, en este caso que

$$Z = \sqrt{3n}(\overline{X} - 1) \sim \mathrm{N}(0, 1), \text{ aproximadamente.}$$

Así,

$$\begin{aligned} \Pr(0.9 < \overline{X} < 1.1) &= \Pr(\sqrt{300}(0.9 - 1) < Z < \sqrt{300}(1.1 - 1)) = \\ \Pr(-1.732 < Z < 1.732) &= 2\Phi(1.732) - 1 = 0.9583, \end{aligned}$$

(b) Procedemos de manera similar a la parte (a), pero igualando la probabilidad de caer en el intervalo indicado a 0.95:

$$\begin{aligned} \Pr(1 - \epsilon < \overline{X} < 1 + \epsilon) &= \Pr(-\sqrt{300}\epsilon < Z < \sqrt{300}\epsilon) = \\ \Pr(-17.32\epsilon < Z < 17.32\epsilon) &= 2\Phi(17.32\epsilon) - 1 = 0.95. \end{aligned}$$

Resulta, entonces $\Phi(17.32\epsilon) = 0.975$, por lo que, de la tabla, obtenemos $17.32\epsilon \approx 1.96$, resultando $\epsilon \approx 0.1131$.