Ejercicio 6.5.4 · Doscientas muestras, con y sin reemplazo

Adaptado de: 2º final 2014-1, problema 6.

Las estaturas de 1000 estudiantes tienen distribución aproximadamente normal con media 174.5 cm y desviación estándar 6.9 cm. Si se extraen 200 muestras aleatorias de tamaño 25, determine:

- a) el número de muestras cuyo promedio cae entre 172.5 y 175.8 cm, con reemplazo;
- b) el número de muestras cuyo promedio cae por debajo de 172.5 cm, sin reemplazo.

Resolución.

1. a) Con reemplazo: $\sigma_{\bar{X}} = \frac{6.9}{\sqrt{25}} = 1.38$.
2. $z_a = \frac{172.5-174.5}{1.38} = -1.4493 \to -1.45$; $z_b = \frac{175.8-174.5}{1.38} = 0.9420 \to 0.94$.
3. $P(172.5 < \bar{X} < 175.8) = F_Z(0.94) - F_Z(-1.45) = 0.8264 - 0.0735 = 0.7529$.
4. Número esperado: $200(0.7529) = 150.58 \approx$ **151 muestras**.
5. b) Sin reemplazo de una población finita, $N = 1000$: $\sigma_{\bar{X}} = \frac{6.9}{5}\sqrt{\frac{1000-25}{1000-1}} = 1.38(0.98791) = 1.3633$.
6. $z = \frac{172.5-174.5}{1.3633} = -1.4670 \to -1.47$, así que $P(\bar{X} < 172.5) = F_Z(-1.47) = 1 - 0.9292 = 0.0708$.
7. Número esperado: $200(0.0708) = 14.16 \approx$ **14 muestras**.

Ejercicio 6.5.5 · Formularios de hipoteca: dos días independientes

Adaptado de: 2º final 2015-1, problema 6.

El tiempo que tarda un solicitante en llenar una forma para una hipoteca tiene distribución normal con media 10 minutos y desviación estándar 2 minutos. Un día llenan la forma 5 personas y al día siguiente 6. ¿Cuál es la probabilidad de que el tiempo promedio de la muestra de **cada** día sea a lo sumo 11 minutos?

Resolución.

1. $X \sim N(10,\ 2^2)$. La población es normal: $\bar{X}$ es normal con cualquier n.
2. Día 1, $n = 5$: $\sigma_{\bar{X}} = \frac{2}{\sqrt{5}} = 0.8944$; $z = \frac{11-10}{0.8944} = 1.1180 \to 1.12$; $P(\bar{X}_1 \le 11) = F_Z(1.12) = 0.8686$.
3. Día 2, $n = 6$: $\sigma_{\bar{X}} = \frac{2}{\sqrt{6}} = 0.8165$; $z = \frac{11-10}{0.8165} = 1.2247 \to 1.22$; $P(\bar{X}_2 \le 11) = F_Z(1.22) = 0.8888$.
4. Las muestras son independientes: $P = (0.8686)(0.8888) =$ **0.7720**.

Ejercicio 6.5.6 · Población uniforme discreta

Adaptado de: 2º final 2011-1, problema 5.

Dada la función de probabilidad de una población uniforme, $f_X(x) = \frac{1}{4}$ para $x = 0, 1, 2, 3$, calcule la probabilidad de que una muestra aleatoria de tamaño 36, seleccionada con reemplazo, dé un promedio muestral mayor que 1.45 pero menor que 1.75.

Resolución.

1. $\mu_X = \frac{1}{4}(0 + 1 + 2 + 3) = 1.5$.
2. $\sigma_X^2 = \frac{1}{4}\left[(0-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (3-1.5)^2\right] = \frac{1}{4}(5) = 1.25$.

- a) la probabilidad de que la resistencia de una muestra sea menor que 6250 kg/cm²;
- b) la probabilidad de que esté entre 5800 y 5900 kg/cm²;
- c) el valor de resistencia que excede al 95 % de las muestras.

Resolución.

1. $X \triangleq$ resistencia de una muestra [kg/cm²]; $X \sim N(6000,\ 100^2)$.
2. a) $z = \frac{6250-6000}{100} = 2.50$; $P(X < 6250) = F_Z(2.50) =$ **0.9938**.
3. b) $z_1 = \frac{5800-6000}{100} = -2.00$ y $z_2 = \frac{5900-6000}{100} = -1.00$; $P = F_Z(-1.00) - F_Z(-2.00) = 0.1587 - 0.0228 =$ **0.1359**.
4. c) Se busca x con $P(X < x) = 0.95$. En la tabla, 0.95 queda entre $z = 1.64$ y $z = 1.65$: $z_{0.95} = 1.645$.
5. $x = 6000 + 1.645(100) =$ **6164.5 kg/cm²** (con $z = 1.64$, 6164 kg/cm²).

Ejercicio 4.5.8 · Tiempo de viaje: normal y binomial «2 de 3»

Adaptado de: 1er final 2015-1, problema 4.

Un ingeniero tarda en promedio 24 minutos en ir de su casa a la oficina, con desviación estándar de 3.8 minutos; el tiempo de viaje es normal.

- a) Si sale a las 8:35 y el café se sirve en la oficina de 8:50 a 9:00, ¿cuál es la probabilidad de que se pierda el café?
- b) ¿Cuál es la probabilidad de que 2 de los 3 viajes siguientes tomen por lo menos media hora?

Resolución.

1. $X \triangleq$ duración de un viaje [min]; $X \sim N(24,\ 3.8^2)$.
2. a) Se pierde el café si llega a las 9:00 o después, es decir, si el viaje dura 25 min o más.
3. $z = \frac{25-24}{3.8} = 0.263 \approx 0.26$; $P(X \ge 25) = 1 - F_Z(0.26) = 1 - 0.6026 =$ **0.3974**.
4. b) Primero, la probabilidad de un viaje largo: $z = \frac{30-24}{3.8} = 1.58$, $p = P(X \ge 30) = 1 - 0.9429 = 0.0571$.
5. $Y \triangleq$ número de viajes largos entre los 3 siguientes; $Y \sim \text{Binomial}(n = 3,\ p = 0.0571)$.
6. $P(Y = 2) = \binom{3}{2}p^2q = 3(0.0571)^2(0.9429) =$ **0.0092**.

Ejercicio 4.5.9 · Examen de licencia: binomial exacta y aproximación normal

Adaptado de: 2º final 2016-1, problema 4.

El 70 % de los ingenieros civiles que presentan el examen para obtener la licencia de perito lo aprueba. Si 50 lo presentarán el próximo sábado, encuentre la probabilidad de que lo aprueben de 37 a 39.

Resolución.

1. $X \triangleq$ número de aprobados, de 50; $X \sim \text{Binomial}(n = 50,\ p = 0.7)$.
2. **Exacta:** $P(37 \le X \le 39) = f(37) + f(38) + f(39) = 0.1050 + 0.0838 + 0.0602 =$ **0.2490**.
3. **Aproximación normal:** $np = 35 > 5$ y $nq = 15 > 5$. $\mu = 35$, $\sigma^2 = npq = 10.5$, $\sigma = 3.2404$.
4. Con la corrección por continuidad: $P(37 \le X \le 39) \approx P(36.5 \le Y \le 39.5)$.
5. $z_1 = \frac{36.5-35}{3.2404} = 0.46$ y $z_2 = \frac{39.5-35}{3.2404} = 1.39$.
6. $P \approx F_Z(1.39) - F_Z(0.46) = 0.9177 - 0.6772 =$ **0.2405**, muy cerca del valor exacto.

6.3.6 Media con t de Student

1. Población normal, σ desconocida, se conoce s.
2. $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ con $\nu = n - 1$.
3. Lee la tabla t (área a la derecha), igual que la ji cuadrada.

6.4 Ejemplo resuelto

Adaptado de: 1er final 2014-1, problema 6.

Se lanza un dado. Sea X el número que queda hacia arriba.

- a) ¿Qué distribución tiene X?
- b) ¿Cuál es la media y la variancia de X?
- c) Si el dado se lanza 30 veces y se promedian los resultados, ¿cuál es la media y la variancia teórica de la media muestral?
- d) Al lanzar 30 veces el dado, ¿cuál es la probabilidad aproximada de que la media muestral sea mayor o igual que 4?

Datos. $X \in \{1, 2, 3, 4, 5, 6\}$; $n = 30$ lanzamientos.

Planteamiento. Primero los parámetros de la población (procedimiento 6.3.4); luego los de $\bar{X}$ y el TLC (procedimiento 6.3.1).

Resolución.

1. a) Las seis caras son igualmente probables: X tiene distribución **uniforme discreta**, $f_X(x) = \frac{1}{6}$ para $x = 1, \ldots, 6$.
2. b) $\mu_X = \frac{1+2+\cdots+6}{6} = \frac{21}{6} = \mathbf{3.5}$.
3. $E(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$, así que $\sigma_X^2 = \frac{91}{6} - 3.5^2 = \frac{35}{12} = \mathbf{2.9167}$.
4. c) $\mu_{\bar{X}} = \mu_X = \mathbf{3.5}$ y $\sigma_{\bar{X}}^2 = \frac{35/12}{30} = \frac{35}{360} = \mathbf{0.0972}$; de donde $\sigma_{\bar{X}} = 0.3118$.
5. d) $n = 30 \geq 30$: por el TLC, $\bar{X} \sim N(3.5,\ 0.0972)$ aproximadamente.
6. $z = \frac{4-3.5}{0.3118} = 1.6036 \to 1.60$.
7. «Mayor o igual» es el área a la derecha: $P(\bar{X} \geq 4) = 1 - F_Z(1.60) = 1 - 0.9452 = \mathbf{0.0548}$.
8. **Resultado:** la probabilidad de que el promedio de 30 lanzamientos sea al menos 4 es **0.0548** (con z sin redondear, 0.0544). Un solo lanzamiento da 4 o más con probabilidad $\frac{1}{2}$; el promedio de 30, apenas 5 % de las veces.

6.5 Ejercicios de examen (resueltos)

Ejercicio 6.5.1 · Examen de admisión: distribución desconocida

Adaptado de: 1er final 2016-2, problema 6.

Miles de estudiantes presentaron un examen de admisión. Sea X la calificación de un estudiante elegido al azar. Se desconoce la distribución de X, pero se estima que tiene media 68.3 y desviación estándar 8.4. Si se toma una muestra de 36 exámenes, ¿cuál es la probabilidad de que el promedio de las calificaciones sea superior a 70?

Resolución.

Encuentre la probabilidad de que tenga que disparar 7 veces para acertar 4.

Resultado: 0.0732.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.6

Adaptado de: 2º final 2017-1, problema 5.

El 2 % de los tornillos que fabrica una máquina tiene defectos. En un lote de 200 tornillos, ¿cuál es la probabilidad de que exactamente el 99 % sea no defectuoso? ¿Y de que lo sea el 99 % o más?

Resultado: 0.1458; 0.2351.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.7

Adaptado de: 2º final 2018-1, problema 5, inciso a.

En un proceso fotográfico, el tiempo de revelado de las copias es normal con $\mu = 18$ s y $\sigma = 0.75$ s. Determine la probabilidad de que el revelado de una copia tarde de 16 a 19.5 s.

Resultado: 0.9734.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.8

Adaptado de: 2º final 2011-1, problema 4.

En la noche de un sábado, uno de cada diez conductores maneja en estado de ebriedad. Si se inspeccionan al azar 400 conductores, ¿cuál es la probabilidad de que el número de conductores ebrios sea:

- a) menor que 32?
- b) mayor que 49?
- c) por lo menos 35, pero menor que 47?

Resultado (aproximación normal con corrección)**:** 0.0778; 0.0571; 0.6811.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 4.6.9

Adaptado de: 1er final 2015-1, problema 6.

El 25 % de las personas que entran a una librería compra algo, y las compras siguen una distribución binomial. Si en un día entraron 80 clientes, calcule la probabilidad de que se hagan al menos 28 compras.

Resultado: 0.0262 con la aproximación normal y corrección por continuidad (0.0295 exacta).

→ Resolución en el Apéndice E, p. 104.

Tabla	Entrada	Da el área…	Notación
t de Student	ν (fila) y α (columna)	a la **derecha** del valor	$P(T > t_{\alpha,\nu}) = \alpha$

6.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
«se toma una muestra de n…», «el promedio», «la media de la muestra»	$P(\bar{X}\ldots)$ con Z	6.3.1
«distribución normal» y una muestra	$\bar{X}$ normal con cualquier n	6.3.1
«se desconoce la distribución», $n \geq 30$	justificar con el TLC	6.3.1
«un solo» individuo, sin muestra	normal de la Unidad 4 (con σ, no $\sigma/\sqrt{n}$)	Unidad 4
la población se da como función de probabilidad (dado, $f_X(x) = \frac{1}{4}$)	μ_X y σ_X^2 primero, luego TLC	6.3.4
«de 200 muestras, ¿cuántas…?»	$M \cdot p$	6.3.3
«un día… y al día siguiente…», «ambas muestras»	producto de probabilidades	6.3.3
«sin reemplazo» y N conocida	factor de corrección	6.3.3
«peso total», «suma»	distribución de $\sum X_i$	6.3.3
«variancia de la muestra s^2», «desviación estándar muestral»	ji cuadrada	6.3.5
σ desconocida, se da s, población normal	t de Student	6.3.6

6.3 Procedimientos

6.3.1 Probabilidad para la media muestral

1. **Define:** «Sea X la [variable] de un individuo» y «sea $\bar{X}$ el promedio de la muestra de n».
2. **Datos:** μ_X, σ_X y n. Si se da la variancia ($\sigma^2 = 225$ o $N(100, 15^2)$), saca la raíz; si se da la desviación ($N(\mu = 106, \sigma = 12)$), úsala tal cual.
3. **Justifica** la normalidad: «la población es normal» o «$n \geq 30$, por el teorema del límite central».
4. **Modelo:** $\bar{X} \sim N\left(\mu_X,\ \frac{\sigma_X^2}{n}\right)$, con los números sustituidos.
5. **Error estándar:** $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$, con 4 decimales.
6. **Estandariza:** $z = \frac{\bar{x} - \mu_X}{\sigma_X/\sqrt{n}}$ y redondea a **2 decimales** para entrar a la tabla.
7. **Calcula** según la forma de la pregunta:

Se pide	Se escribe	Con la tabla
$P(\bar{X} < c)$	$P(Z < z)$	$F_Z(z)$

- Normal: $z = \frac{x-\mu}{\sigma}$ · $F_Z(-z) = 1 - F_Z(z)$ · $z_{0,95} = 1.645$
- Aproximación normal: $\mu = np$, $\sigma = \sqrt{npq}$ ($np > 5$, $nq > 5$), con ± 0.5

4.9 Procedimiento recomendado en el examen

1. La variable definida con palabras: «$X \triangleq$ número de… en…» o «$T \triangleq$ tiempo hasta…».
2. El modelo con sus parámetros **ya calculados**: $X \sim \text{Poisson}(12)$, $T \sim \text{Exponencial}\left(\frac{1}{16}\right)$, $X \sim N(24,\ 3.8^2)$.
3. La pregunta traducida a símbolos, con el complemento si es más corto y la condición si dice «si ya…» o «dado que…».
4. La fórmula escrita y luego sustituida; en la normal, la z con dos decimales y el área dibujada.
5. El resultado con 4 decimales y una frase: «el 28.35 % de las pantallas usará la garantía».

Si no se encuentra el camino, nombrar el modelo y escribir la fórmula con los números sustituidos ya vale buena parte del problema.

Unidad 6 · Distribuciones muestrales: teorema del límite central, media muestral y ji cuadrada

Objetivo (programa de la asignatura). El alumno identificará las distribuciones de algunos estadísticos que se utilizan en el muestreo.

Instrucciones de emergencia

- **Frecuencia:** aparece en 16 de los 18 finales de 2011-1 a 2018-1 (solo falta en el 2º final 2012-1 y en el 1er final 2015-1). Casi siempre es la media muestral con el teorema del límite central; la ji cuadrada salió dos veces. Vale de 10 a 15 puntos.
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 6.1 y la lectura de tablas de la sección 6.3.2 (20 min).
  2. Sigue el ejemplo 6.4 tapando la resolución.
  3. Resuelve los ejercicios 6.5 con cronómetro (10 min cada uno).

6.1 Conceptos y expresiones

Muestra aleatoria y estadístico. Una *muestra aleatoria* de tamaño n es un conjunto de variables $X_1, X_2, \dots, X_n$ independientes y con la misma distribución que la población. En el *muestreo aleatorio simple* cada elemento de la población tiene la misma probabilidad de ser elegido. Un *estadístico* es cualquier función de la muestra que no depende de parámetros desconocidos; los dos más usados son

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ (media muestral)}, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ (variancia muestral)}.$$

Un estadístico es una variable aleatoria: cambia de una muestra a otra. Su distribución se llama **distribución muestral**.

Parámetros de la media muestral. Si la población tiene media μ_X y variancia σ_X^2:

$$\mu_{\bar{X}} = \mu_X, \qquad \sigma^2_{\bar{X}} = \frac{\sigma_X^2}{n}, \qquad \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} \text{ (error estándar)}.$$

- El promedio tiene la misma media que un individuo, pero varía mucho menos: su desviación se divide entre $\sqrt{n}$.
- **Sin reemplazo de una población finita** de tamaño N se multiplica por el factor de corrección: $\sigma^2_{\bar{X}} = \frac{\sigma_X^2}{n}\cdot\frac{N-n}{N-1}$. Si el muestreo es con reemplazo, o N es muy grande, el factor se omite.

Forma de la distribución de $\bar{X}$.

La población es…	Tamaño de muestra	Entonces $\bar{X}$…
normal	cualquier n	es exactamente normal

- **Condicionales:** $f_{Y|x}(y \mid x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$, con $f_X(x) > 0$.
- **Región de definición.** En un *rectángulo* ($a \le x \le b$, $c \le y \le d$) todos los límites son números. En una región como $0 < y < x < 1$ el límite interior **depende** de la otra variable: para x fija, y va de 0 a x; para y fija, x va de y a 1.
- **Uniforme conjunta:** si f_{XY} es constante en una región, vale $\frac{1}{\text{área}}$ y $P(A) = \frac{\text{área de } A}{\text{área total}}$.

Valor esperado de una función de dos variables.

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{XY}(x,y) \qquad \text{o} \qquad E[g(X,Y)] = \iint g(x,y) f_{XY}(x,y)\,dy\,dx$$

- Con $g = x$ se obtiene $E(X)$, que también se calcula con la marginal: $E(X) = \sum x f_X(x)$ o $\int x f_X(x)\,dx$.
- **Valor esperado condicional:** $E(Y \mid X = x) = \sum_y y\, f_{Y|x}(y \mid x)$ o $\int y\, f_{Y|x}(y \mid x)\,dy$. Como función de x se llama *curva de regresión* de Y sobre X.

Independencia. X y Y son independientes si y solo si

$$f_{XY}(x,y) = f_X(x)\, f_Y(y) \quad \text{para todo } (x,y),$$

lo que equivale a $f_{X|y}(x \mid y) = f_X(x)$: conocer Y no cambia la distribución de X.

- Para concluir que **no** son independientes basta **un** par (x,y) donde falle la igualdad. Para concluir que **sí** lo son hay que revisar todos.
- En el caso continuo, si la región es un rectángulo y f_{XY} se factoriza como (función de x)·(función de y), son independientes. Si la región no es un rectángulo (por ejemplo $0 < y < x < 1$), son dependientes.

Covariancia y correlación.

$$\operatorname{Cov}(X,Y) = \sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X\mu_Y, \qquad \rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

- El signo de la covariancia indica la dirección de la relación lineal: positiva, las dos variables tienden a crecer juntas; negativa, cuando una crece la otra tiende a bajar.
- El coeficiente de correlación no tiene unidades y cumple $-1 \le \rho \le 1$. $|\rho|$ cercano a 1 indica una relación lineal fuerte; cercano a 0, débil o nula.
- Propiedades: $\operatorname{Cov}(X,X) = Var(X)$; $\operatorname{Cov}(aX+b,\ cY+d) = ac\operatorname{Cov}(X,Y)$.
- Si X y Y son independientes, $E(XY) = E(X)E(Y)$ y $\operatorname{Cov}(X,Y) = 0$. **El recíproco es falso:** puede haber covariancia cero con variables dependientes.

Combinaciones lineales.

$$E(aX+bY) = a\mu_X + b\mu_Y, \qquad Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\operatorname{Cov}(X,Y)$$

- Si son independientes: $Var(X \pm Y) = \sigma_X^2 + \sigma_Y^2$ (las variancias **se suman** también en la resta).
- Para n variables: $Var\left(\sum X_i\right) = \sum Var(X_i) + 2\sum_{i<j}\operatorname{Cov}(X_i, X_j)$. Esta fórmula es la base de la Unidad 6.

Distribución normal bivariada (programa, 5.5). Es la extensión de la normal a dos variables, con parámetros μ_X, μ_Y, σ_X, σ_Y y ρ. Sus marginales son normales y, solo en este modelo, $\rho = 0$ equivale a independencia. No aparece en los finales.

primera y Y la de la segunda, con densidad $f_{XY}(x,y) = \frac{3}{2}(x^2+y^2)$, $0 \le x \le 1$, $0 \le y \le 1$. Calcule:

- a) la probabilidad de que ninguna línea se use más de la mitad del tiempo;
- b) la probabilidad de que la primera línea esté ocupada más del 75 % del tiempo;
- c) la proporción del día que se espera esté en uso la primera línea.

Resultado: $\frac{1}{16} = 0.0625$; $\frac{53}{128} = 0.4141$; $E(X) = \frac{5}{8} = 0.625$.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 5.6.4

Adaptado de: 2º final 2016-1, problema 6.

Dos variables tienen densidad conjunta $f_{XY}(x,y) = \frac{1}{2}(x+y)$, $0 \le x \le 1$, $1 \le y \le 2$. Calcule:

- a) la probabilidad de que $Y > 1.5$ dado que $X < 1$;
- b) la probabilidad de que $X < 0.8$.

Resultado: $\frac{9}{16} = 0.5625$; 0.76.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 5.6.5

Adaptado de: 1er final 2013-1, problema 5.

X es la temperatura ambiente (°C) y Y el tiempo (min) que necesita el motor diésel de una camioneta para estar listo, con densidad conjunta $f_{XY}(x,y) = \frac{1}{6640}(4x+2y+1)$, $0 \le x \le 40$, $0 \le y \le 2$.

- a) Determine la probabilidad de que la temperatura sea mayor que 20 °C y se requiera por lo menos un minuto.
- b) ¿Son independientes X y Y? Justifique.

Resultado: $\frac{2480}{6640} = 0.3735$; no son independientes.

→ Resolución en el Apéndice E, p. 105.

Ejercicio propuesto 5.6.6

Adaptado de: 2º final 2015-1, problema 5.

X es el precio que paga el transportista por un barril de petróleo crudo y Y el que paga la refinería, con densidad conjunta $f_{XY}(x,y) = \frac{1}{200}$, $20 < x < y < 40$. Calcule:

- a) $E(X)$, $E(Y)$ y $E(XY)$;
- b) $\mathrm{Cov}(X,Y)$.

Resultado: $\frac{80}{3} = 26.667$; $\frac{100}{3} = 33.333$; 900; $\mathrm{Cov} = \frac{100}{9} = 11.111$.

→ Resolución en el Apéndice E, p. 105.

5.7 Errores frecuentes

1. **No comprobar que el total sea 1** (tabla) o que $\iint f = 1$. Si no da 1, hay un error de suma o falta despejar una constante.

5.3.3 Construir la tabla en extracciones sin reemplazo

Una caja tiene a objetos del tipo A, b del tipo B y c de otros tipos ($N = a+b+c$). Se extraen n sin reemplazo; $X \triangleq$ número de A y $Y \triangleq$ número de B. Generalizando la hipergeométrica de la Unidad 4:

$$f_{XY}(x,y) = \frac{\binom{a}{x}\binom{b}{y}\binom{c}{n-x-y}}{\binom{N}{n}}, \qquad x+y \le n$$

1. Lista los valores posibles de X y de Y; las celdas con $x+y>n$ valen 0.
2. Calcula cada celda con la fórmula (tecla nCr). Conviene dejar todas con el mismo denominador $\binom{N}{n}$.
3. Comprueba que los numeradores sumen $\binom{N}{n}$ y sigue con 5.3.1.

5.3.4 Densidad conjunta en un rectángulo

1. Dibuja el rectángulo y la región que pide la pregunta. Pasa las unidades al sistema de la densidad (15 minutos son $\frac{1}{4}$ h).
2. Si hay constante, plantea $\iint f_{XY} = 1$ y despéjala.
3. **Integral doble:** primero la interior, tratando la otra variable como constante; luego la exterior. El diferencial pegado a f es el de la integral interior.

   Ejemplo: $\int_0^1\int_0^2 xy\,dy\,dx = \int_0^1 x\left[\frac{y^2}{2}\right]_0^2 dx = \int_0^1 2x\,dx = 1$.
4. **Marginal:** integra la otra variable entre sus límites y escribe el dominio: «$f_X(x) = \frac{2}{3}(x+1)$, $0 \le x \le 1$».
5. **Condicional:** $f_{Y|x_0}(y) = \frac{f_{XY}(x_0,y)}{f_X(x_0)}$. Sustituye el valor x_0 **antes** de integrar en y.
6. Si f_{XY} es constante, usa áreas: base por altura entre el área total.

5.3.5 Regiones no rectangulares

1. Dibuja la región, por ejemplo el triángulo $0 < y < x < 1$, bajo la diagonal.
2. Para la marginal de X, fija x y recorre y de su límite inferior al superior: $f_X(x) = \int_0^x f_{XY}\,dy$.
3. Para la marginal de Y, fija y: $f_Y(y) = \int_y^1 f_{XY}\,dx$.
4. En las integrales dobles, los límites de la integral interior contienen la otra variable; los de la exterior, solo números.
5. Las variables son dependientes aunque f_{XY} parezca factorizable: el dominio de Y depende de X.

5.3.6 Valor esperado, covariancia, correlación y combinaciones lineales

1. $E(X)$ y $E(Y)$ con las marginales.
2. $E(XY)$ celda por celda: $\sum\sum xy\,f_{XY}(x,y)$ (las celdas con $x=0$ o $y=0$ no aportan), o $\iint xy\,f_{XY}$.
3. $\mathrm{Cov} = E(XY) - E(X)E(Y)$.
4. $\sigma_X^2 = E(X^2) - [E(X)]^2$ y σ_Y^2 igual; $\rho = \frac{\mathrm{Cov}}{\sigma_X\sigma_Y}$, que debe quedar entre −1 y 1.
5. Interpreta: signo (dirección) y magnitud (fuerza de la relación lineal).
6. $Var(aX+bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{Cov}$; con $a=b=1$ queda $Var(X+Y) = \sigma_X^2+\sigma_Y^2+2\,\mathrm{Cov}$.

2. a) $P(X<2,\ Y>1)=\frac{1}{9}\int_0^2\int_1^2 xy\,dy\,dx=\frac{1}{9}\int_0^2 x\left[\frac{y^2}{2}\right]_1^2 dx=\frac{1}{9}\int_0^2\frac{3}{2}x\,dx=\frac{1}{9}\cdot\frac{3}{2}\cdot 2=$ **0.3333**.
3. b) $f_Y(y)=\int_0^3\frac{xy}{9}\,dx=\frac{y}{9}\cdot\frac{9}{2}=\frac{y}{2}$, $0<y<2$.
4. $E(Y)=\int_0^2 y\cdot\frac{y}{2}\,dy=\left[\frac{y^3}{6}\right]_0^2=\frac{4}{3}=$ **1.3333**.
5. c) $f_Y(1)=\frac{1}{2}$, así que $f_{X|Y=1}(x)=\dfrac{x/9}{1/2}=\frac{2x}{9}$, $0<x<3$.
6. $P(X>2\mid Y=1)=\int_2^3\frac{2x}{9}\,dx=\frac{9-4}{9}=\frac{5}{9}=$ **0.5556**.
7. **Observación:** $\frac{xy}{9}=\frac{2x}{9}\cdot\frac{y}{2}$ en un rectángulo, así que X y Y son independientes y la condicional coincide con la marginal de X.

Ejercicio 5.5.7 · Región triangular y coeficiente de correlación

Adaptado de: 1er final 2015-1, problema 5.

La proporción X de alumnos y la proporción Y de alumnas que han concluido un examen tienen densidad conjunta $f_{XY}(x,y)=8xy$, $0<y<x<1$.

- a) Obtenga la función marginal de X.
- b) Calcule la probabilidad de que menos de $\frac{1}{8}$ de las alumnas hayan terminado, si se sabe que exactamente $\frac{1}{2}$ de los alumnos lo hicieron.
- c) Determine el coeficiente de correlación e interprételo, considerando $E(X)=\frac{4}{5}$, $Var(X)=\frac{2}{75}$ y $f_Y(y)=4y(1-y^2)$, $0<y<1$.

Resolución.

1. La región es el triángulo bajo la diagonal: para x fija, $0<y<x$.
2. a) $f_X(x)=\int_0^x 8xy\,dy=8x\cdot\frac{x^2}{2}=4x^3$, $0<x<1$.
3. b) $f_{Y|x}(y)=\dfrac{8xy}{4x^3}=\frac{2y}{x^2}$, $0<y<x$. Con $x=\frac{1}{2}$: $f_{Y|1/2}(y)=8y$, $0<y<\frac{1}{2}$.
4. $P\left(Y<\frac{1}{8}\mid X=\frac{1}{2}\right)=\int_0^{1/8}8y\,dy=4\left(\frac{1}{8}\right)^2=\frac{1}{16}=$ **0.0625**.
5. c) $E(XY)=\int_0^1\int_0^x 8x^2y^2\,dy\,dx=\int_0^1\frac{8}{3}x^5\,dx=\frac{4}{9}$.
6. $E(Y)=\int_0^1 4y^2(1-y^2)\,dy=\frac{4}{3}-\frac{4}{5}=\frac{8}{15}$ y $E(Y^2)=\int_0^1 4y^3(1-y^2)\,dy=1-\frac{2}{3}=\frac{1}{3}$, así que $Var(Y)=\frac{1}{3}-\frac{64}{225}=\frac{11}{225}$.
7. $\mathrm{Cov}(X,Y)=\frac{4}{9}-\frac{4}{5}\cdot\frac{8}{15}=\frac{100-96}{225}=\frac{4}{225}$.
8. $\rho=\dfrac{4/225}{\sqrt{\frac{2}{75}\cdot\frac{11}{225}}}=\frac{4}{\sqrt{66}}=$ **0.4924**.
9. **Resultado:** asociación lineal positiva moderada: cuando una mayor proporción de alumnos ha terminado, tiende a ser mayor la de alumnas.

Ejercicio 5.5.8 · Encuentro en un restaurante: uniforme conjunta

Adaptado de: 2º final 2013-1, problema 6.

Dos directivos acordaron encontrarse en un restaurante entre las 2:00 y las 3:00 p. m. X es la hora de llegada del directivo A y Y la del directivo B; son independientes y cada una

5.5 Ejercicios de examen (resueltos)

Ejercicio 5.5.1 · Repuestos de pluma: construir la tabla conjunta

Adaptado de: 2º final 2012-1, problema 6; repetido, con fusibles de tres marcas, en 1er final 2017-1, problema 4.

Se seleccionan al azar dos repuestos de una caja que contiene tres repuestos azules, dos rojos y tres verdes. Sea X el número de repuestos azules y Y el de rojos en la selección.

- a) Determine la distribución de probabilidad conjunta.
- b) Calcule $P(X+Y\le 1)$.
- c) ¿Son estadísticamente independientes?
- d) Obtenga la probabilidad de que la selección tenga un repuesto rojo y uno verde.

Resolución.

1. a) $f_{XY}(x,y)=\dfrac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$, con $\binom{8}{2}=28$ y $x+y\le 2$. Por ejemplo, $f_{XY}(1,0)=\frac{\binom{3}{1}\binom{2}{0}\binom{3}{1}}{28}=\frac{9}{28}$.
2. Tabla (numeradores sobre 28):

f_{XY}	$x=0$	$x=1$	$x=2$	$f_Y(y)$
$y=0$	3	9	3	15
$y=1$	6	6	0	12
$y=2$	1	0	0	1
$f_X(x)$	10	15	3	28

3. b) $X+Y\le 1$ son las celdas $(0,0)$, $(1,0)$ y $(0,1)$: $P=\frac{3+9+6}{28}=\frac{18}{28}=$ **0.6429**.
4. c) $f_{XY}(0,0)=\frac{3}{28}=0.1071$ y $f_X(0)f_Y(0)=\frac{10}{28}\cdot\frac{15}{28}=0.1913$: **no son independientes**. (Más rápido: $f_{XY}(2,1)=0$, pero $f_X(2)f_Y(1)>0$.)
5. d) Un rojo y un verde significa ningún azul y un rojo: $f_{XY}(0,1)=\frac{6}{28}=$ **0.2143**.

Ejercicio 5.5.2 · Defectos y fábrica: independencia y valor esperado

Adaptado de: 1er final 2018-1, problema 4.

Un producto se clasifica según su número de defectos, X, y la fábrica que lo produjo, Y. La función de probabilidad conjunta es:

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$
$y=1$	2/16	1/16	3/16	2/16
$y=2$	1/16	1/16	2/16	4/16

- a) Indique si X y Y son estadísticamente independientes y justifique.
- b) Obtenga el número de defectos que se espera tenga un producto.

Resolución.

4.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.4**.

Ejercicio propuesto 4.6.1

Adaptado de: 1er final 2016-1, problema 5.

Los barcos de carga llegan a un puerto según un proceso de Poisson, con un promedio de 8 por hora. ¿Cuál es la probabilidad de que lleguen exactamente 5 barcos en una hora?

Resultado: 0.0916.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.2

Adaptado de: 2º final 2014-1, problema 4.

El tiempo hasta la falla de una pantalla plana LCD es exponencial con media de tres años. La tienda hace válida la garantía durante el primer año de uso. ¿Qué porcentaje de pantallas hará uso de la garantía?

Resultado: $P(T \leq 1) = 0.2835$: el 28.35 %.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.3

Adaptado de: 2º final 2012-1, problema 5.

El tiempo de espera, en horas, entre la detección de dos conductores sucesivos a alta velocidad por un radar tiene función de distribución $F_X(x) = 1 - e^{-8x}$ para $x \geq 0$ (y 0 para $x < 0$). Determine:

- a) la probabilidad de esperar menos de 12 minutos;
- b) la probabilidad de esperar entre 12 y 18 minutos;
- c) la media y la variancia del tiempo de espera.

Resultado: 0.7981; 0.1112; $E(X) = 0.125$ h (7.5 min), $Var(X) = 0.0156$ h².

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.4

Adaptado de: 1er final 2018-1, problema 5.

El error en la longitud de un alambre puede variar de 1 a 4 mm, con la misma probabilidad en todo el intervalo. Obtenga la probabilidad de que un alambre tenga un error de más de 2 mm.

Resultado: $\frac{2}{3} = 0.6667$.

→ Resolución en el Apéndice E, p. 103.

Ejercicio propuesto 4.6.5

Adaptado de: 2º final 2016-2, problema 4.

La probabilidad de que una tiradora acierte en el objetivo en cualquier disparo es $\frac{1}{3}$.

1. X: calificación de un estudiante; $\bar{X}$: promedio de los 36 exámenes. $\mu_X = 68.3$, $\sigma_X = 8.4$, $n = 36$.
2. La distribución es desconocida, pero $n = 36 \geq 30$: por el TLC, $\bar{X} \sim N\left(68.3,\ \frac{8.4^2}{36}\right)$.
3. $\sigma_{\bar{X}} = \frac{8.4}{\sqrt{36}} = \frac{8.4}{6} = 1.4$.
4. $z = \frac{70-68.3}{1.4} = 1.2143 \to 1.21$.
5. $P(\bar{X} > 70) = 1 - F_Z(1.21) = 1 - 0.8869 = \mathbf{0.1131}$.
6. **Resultado:** en cerca del 11 % de las muestras de 36 exámenes el promedio pasa de 70.

Ejercicio 6.5.2 · Pruebas de inteligencia: probabilidad entre dos valores

Adaptado de: 2º final 2017-1, problema 6.

Varias pruebas de inteligencia dieron una puntuación con distribución normal de media 100 y desviación estándar 15. Si se selecciona al azar una muestra de 50 pruebas, ¿cuál es la probabilidad de que el puntaje promedio de la muestra esté entre 95 y 102 puntos?

Resolución.

1. $X \sim N(100,\ 15^2)$: puntaje de una prueba. La población es normal, así que $\bar{X} \sim N\left(100,\ \frac{15^2}{50}\right)$.
2. $\sigma_{\bar{X}} = \frac{15}{\sqrt{50}} = \frac{15}{7.0711} = 2.1213$.
3. $z_a = \frac{95-100}{2.1213} = -2.3570 \to -2.36$; $z_b = \frac{102-100}{2.1213} = 0.9428 \to 0.94$.
4. $F_Z(0.94) = 0.8264$ y $F_Z(-2.36) = 1 - 0.9909 = 0.0091$.
5. $P(95 \leq \bar{X} \leq 102) = 0.8264 - 0.0091 = \mathbf{0.8173}$.

Ejercicio 6.5.3 · Revelado fotográfico: un individuo contra el promedio

Adaptado de: 2º final 2018-1, problema 5.

En un proceso fotográfico, el tiempo de revelado de las copias tiene distribución normal con $\mu_X = 18$ [s] y $\sigma_X = 0.75$ [s].

- a) Determine la probabilidad de que el revelado de una copia tarde de 16 a 19.5 [s].
- b) Si se toma una muestra aleatoria de 10 tiempos de revelado, defina los parámetros de la distribución del tiempo promedio de la muestra.
- c) Obtenga la probabilidad de que el tiempo promedio de la muestra sea mayor que 18.5 [s].

Resolución.

1. a) Es **una** copia: se usa σ_X, no $\sigma_{\bar{X}}$. $z_a = \frac{16-18}{0.75} = -2.67$ y $z_b = \frac{19.5-18}{0.75} = 2$.
2. $P(16 \leq X \leq 19.5) = F_Z(2) - F_Z(-2.67) = 0.9772 - 0.0038 = \mathbf{0.9734}$.
3. b) La población es normal, así que con $n = 10$ (aunque sea menor que 30) $\bar{X}$ es normal: $\mu_{\bar{X}} = \mathbf{18\ s}$ y $\sigma_{\bar{X}} = \frac{0.75}{\sqrt{10}} = \mathbf{0.2372\ s}$ ($\sigma^2_{\bar{X}} = 0.05625$ s²).
4. c) $z = \frac{18.5-18}{0.2372} = 2.1082 \to 2.11$.
5. $P(\bar{X} > 18.5) = 1 - F_Z(2.11) = 1 - 0.9826 = \mathbf{0.0174}$.
6. **Resultado:** una copia tarda entre 16 y 19.5 s con probabilidad 0.9734; el promedio de 10 copias pasa de 18.5 s con probabilidad 0.0174.

Ejercicio propuesto 4.6.10

Adaptado de: 1er final 2016-2, problema 3.

En un lote de diez cepillos dentales, cuatro tienen defecto en el acabado. Si se toman tres cepillos al azar, determine:

- a) la función de probabilidad del número de cepillos defectuosos;
- b) su media y su desviación estándar.

Resultado: $f_X(0,1,2,3) = \frac{1}{6}, \frac{1}{2}, \frac{3}{10}, \frac{1}{30}$; $E(X) = 1.2$; $\sigma_X = 0.7483$.

→ Resolución en el Apéndice E, p. 104.

4.7 Errores frecuentes

1. **No reescalar el parámetro de Poisson.** «6 por minuto» y la pregunta habla de 2 minutos: el parámetro es 12, no 6.
2. **Usar la media como parámetro de la exponencial.** «Media de 3 años» significa $\lambda = \frac{1}{3}$, no $\lambda = 3$.
3. **Tasa en unidades distintas a las de la pregunta.** 50 llamadas al día en una jornada de 8 h son 6.25 por hora; si la pregunta está en horas, λ debe estar por hora.
4. **Ignorar una condición en la exponencial.** «Si ya lleva 5 años» pide $P(T < 25 \mid T > 5) = P(T < 20)$, no $P(5 < T < 25)$.
5. **Sumar muchos términos en lugar de usar el complemento.** «Al menos 2» es $1 - f(0) - f(1)$.
6. **Confundir $\geq$ con $>$ en los modelos discretos.** «Al menos 2» incluye el 2; «más de 2» no. «Menor que 32» es $X \leq 31$. En los continuos da igual.
7. **Confundir binomial con Pascal.** Si el último ensayo es el éxito buscado («el duodécimo es el segundo en suscribirse»), es Pascal y se usa $\binom{x-1}{r-1}$, no $\binom{x}{r}$.
8. **Usar binomial cuando no hay reemplazo** y el lote es pequeño: es hipergeométrica.
9. **Tomar σ^2 por σ.** En $N(\mu, \sigma^2)$ el segundo número es la variancia; saca la raíz antes de estandarizar.
10. **Leer mal la tabla.** La tabla da el área a la **izquierda**; para «mayor que», $1 - F_Z(z)$. Una probabilidad negativa o mayor que 1 indica un área volteada.
11. **Olvidar o aplicar mal la corrección por continuidad.** Úsala según la tabla de 4.3.5 y estandariza siempre con $\mu = np$.

4.8 Formulario de emergencia de la unidad

- Bernoulli: $E = p$, $Var = pq$ · Binomial: $\binom{n}{x}p^x q^{n-x}$, $E = np$, $Var = npq$
- Geométrica: $q^{x-1}p$, $E = \frac{1}{p}$, $Var = \frac{q}{p^2}$ · Pascal: $\binom{x-1}{r-1}p^r q^{x-r}$, $E = \frac{r}{p}$, $Var = \frac{rq}{p^2}$
- Hipergeométrica: $\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $E = n\frac{r}{N}$, $Var = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$
- Poisson: $\frac{(\lambda t)^x e^{-\lambda t}}{x!}$, $E = Var = \lambda t$ · Binomial $\approx$ Poisson(np) si $n \geq 20$, $p \leq 0.05$
- Uniforme: $\frac{1}{b-a}$, $P = \frac{d-c}{b-a}$, $E = \frac{a+b}{2}$, $Var = \frac{(b-a)^2}{12}$
- Exponencial: $F = 1 - e^{-\lambda t}$, $P(T > t) = e^{-\lambda t}$, $E = \frac{1}{\lambda}$, $Var = \frac{1}{\lambda^2}$ · $P(T > s + t \mid T > s) = P(T > t)$

Se pide	Se escribe	Con la tabla
$P(\bar{X} > c)$	$P(Z > z)$	$1 - F_Z(z)$
$P(a < \bar{X} < b)$	$P(z_a < Z < z_b)$	$F_Z(z_b) - F_Z(z_a)$

8. **Conclusión** con palabras: «La probabilidad de que el promedio… sea… es 0.____».

6.3.2 Lectura de tablas

- **Normal estándar.** La fila da el entero y el primer decimal; la columna, el segundo decimal. Ejemplo: $F_Z(1.21)$ está en la fila 1.2, columna 0.01: **0.8869**.
  - Si la tabla solo trae valores positivos: $F_Z(-z) = 1 - F_Z(z)$. Ejemplo: $F_Z(-1.37) = 1 - 0.9147 = 0.0853$.
  - Comprobación: si $z < 0$, $F_Z(z) < 0.5$; si $z > 0$, $F_Z(z) > 0.5$.
- **Ji cuadrada.** Se busca la fila $\nu = n - 1$; cada columna es el área a la derecha α.
  - Para $P(\chi^2 > c)$: se localiza c en la fila ν y se lee el encabezado de su columna. Ejemplo: con $\nu = 24$, $\chi^2_{0.05,24} = 36.415$, así que $P(\chi^2 > 36.4) \approx 0.05$.
  - Para $P(\chi^2 < c)$: es $1 - \alpha$.
  - Para $P(a < \chi^2 < b)$: $\alpha_a - \alpha_b$ (área derecha de a menos área derecha de b). Ejemplo con $\nu = 9$: $\chi^2_{0.99,9} = 2.088$ y $\chi^2_{0.05,9} = 16.919$, así que $P(2.088 < \chi^2 < 16.919) = 0.99 - 0.05 = 0.94$.
  - Si c cae entre dos columnas, se reporta la probabilidad aproximada o se interpola.
- **t de Student.** Igual que la ji cuadrada, con área a la derecha. Ejemplo: $t_{0.025,9} = 2.262$, es decir, $P(T > 2.262) = 0.025$ con 9 grados de libertad. Por simetría, $P(T < -t) = P(T > t)$.

6.3.3 Variantes del problema de la media muestral

- **«¿Cuántas de M muestras?»:** calcula la probabilidad p para una muestra y multiplica: $M \cdot p$. Redondea al entero más cercano.
- **Dos muestras independientes** («el primer día…, el segundo día…»): calcula cada probabilidad con su propia n y multiplica, $P(A \cap B) = P(A)\,P(B)$.
- **Sin reemplazo, población finita:** $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
- **Suma o total:** $\sum X_i \sim N(n\mu,\ n\sigma^2)$ y $z = \frac{t - n\mu}{\sigma\sqrt{n}}$.

6.3.4 Población dada por su distribución

1. Identifica el modelo (por ejemplo, uniforme discreta: $f_X(x) = \frac{1}{k}$).
2. Calcula $\mu_X = \sum x\, f_X(x)$ y $\sigma^2_X = \sum x^2 f_X(x) - \mu^2_X$.
3. Con $n \geq 30$ aplica el TLC y sigue el procedimiento 6.3.1.

6.3.5 Variancia muestral con ji cuadrada

1. Comprueba que la población sea **normal** (requisito) y anota σ^2, n y $\nu = n - 1$.
2. Transforma cada valor de S^2: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$. Si el enunciado da valores de S, elévalos al cuadrado primero.
3. Lee la tabla en la fila ν (área a la derecha) y combina las áreas como en 6.3.2.

Unidad 5 · Variables aleatorias conjuntas: marginales, condicionales, independencia, covariancia y correlación

Objetivo (programa de la asignatura). El alumno conocerá el concepto de variable aleatoria conjunta y podrá analizar el comportamiento probabilista, conjunta e individualmente, de las variables a través de su distribución, e identificará relaciones de dependencia entre dichas variables.

Instrucciones de emergencia

- **Frecuencia:** aparece en los 18 finales de 2011-1 a 2018-1. Vale de 15 a 20 puntos. Lo más común: una tabla discreta (independencia, condicional, valor esperado) o una densidad en un rectángulo (probabilidad de una región, marginal).
- **Tiempo:** de 2 a 3 horas para esta unidad.
- **Ruta corta:**
  1. Lee la sección 5.1 (20 min).
  2. Sigue el ejemplo 5.4 tapando la resolución.
  3. Resuelve los ejercicios 5.5 con cronómetro (12 min cada uno).

5.1 Conceptos y expresiones

Variables aleatorias conjuntas. Son dos (o más) variables, X y Y, definidas sobre el mismo experimento. Su *distribución conjunta* asigna probabilidad a cada **pareja** de valores. Con ella se responde cómo se comportan las variables juntas, cada una por separado (*marginales*), una cuando se conoce la otra (*condicionales*) y si están relacionadas (*independencia*, *covariancia*, *correlación*). Es la Unidad 3 con dos variables: sumar en el caso discreto, integrar en el continuo.

Caso discreto. La función de probabilidad conjunta es $f_{XY}(x,y) = P(X = x,\ Y = y)$, y casi siempre se da como tabla.

- Propiedades: $0 \le f_{XY}(x,y) \le 1$ y $\sum_x \sum_y f_{XY}(x,y) = 1$.
- Probabilidad de un evento: $P\big((X,Y) \in A\big) = \sum_{(x,y)\in A} f_{XY}(x,y)$, la suma de las celdas que cumplen la condición.
- **Marginales:** $f_X(x) = \sum_y f_{XY}(x,y)$ (total de cada renglón o columna de x) y $f_Y(y) = \sum_x f_{XY}(x,y)$.
- **Condicionales:** $f_{X|y}(x \mid y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$, con $f_Y(y) > 0$; análogamente $f_{Y|x}(y \mid x) = \dfrac{f_{XY}(x,y)}{f_X(x)}$.

Caso continuo. La función de densidad conjunta $f_{XY}(x,y)$ cumple $f_{XY} \ge 0$ y $\iint f_{XY}\,dx\,dy = 1$.

- Probabilidad de una región: $P\big((X,Y) \in A\big) = \iint_A f_{XY}(x,y)\,dy\,dx$.
- **Marginales:** $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$ y $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$: se integra **la otra** variable en todo su recorrido. Cada marginal se escribe con su dominio.

La población es…	Tamaño de muestra	Entonces $\bar{X}$…
de distribución desconocida o no normal	$n \ge 30$	es aproximadamente normal (teorema del límite central)
de distribución desconocida	$n < 30$	no puede calcularse la probabilidad con la información dada

Teorema del límite central (TLC). Si $X_1, \dots, X_n$ es una muestra aleatoria de una población con media μ y variancia σ^2 finitas, la variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tiende a la normal estándar cuando n crece. En la práctica se aplica con $n \ge 30$:

$$\bar{X} \sim N\left(\mu_{\bar{X}} = \mu,\ \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}\right) \text{ (aproximadamente).}$$

- Para la **suma** o total de la muestra: $\sum X_i \sim N(n\mu,\ n\sigma^2)$, de modo que $z = \frac{t - n\mu}{\sigma\sqrt{n}}$.
- El programa menciona además la generación de números aleatorios normales con el TLC (sumar muchas uniformes); no se ha evaluado en los finales.

Distribución ji cuadrada. Si la población es **normal** con variancia σ^2,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(\nu)}, \qquad \nu = n - 1 \text{ grados de libertad.}$$

- Solo toma valores positivos y es asimétrica a la derecha; su media es ν.
- Sirve para calcular probabilidades sobre la variancia muestral S^2 (o sobre la desviación S, elevando al cuadrado).

Distribución t de Student. Si la población es normal y σ es **desconocida**, se sustituye por S:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

- Es simétrica respecto a 0, como Z, pero con colas más pesadas; con $n \ge 30$ es casi igual a Z.
- En los finales de 2011-1 a 2018-1 siempre se da σ, así que se usa Z. La t solo se lee en tablas.

Tablas. En el examen se permiten las tablas de la normal estándar, de la t y de la ji cuadrada. Lo importante es saber **qué área da cada una**:

Tabla	Entrada	Da el área…	Notación
Normal estándar	z con dos decimales	a la **izquierda** de z	$F_Z(z) = P(Z \le z)$
Ji cuadrada	ν (fila) y α (columna)	a la **derecha** del valor	$P(\chi^2 > \chi^2_{\alpha,\nu}) = \alpha$

5.2 Cómo reconocer el problema

Si el enunciado dice…	Se pide…	Procedimiento
una tabla con x en un eje y y en el otro	trabajar con la tabla conjunta	5.3.1
«función marginal», «número esperado de defectos»	sumar o integrar la otra variable	5.3.1, 5.3.3
«dado que», «si se sabe que», «si Y es igual a…»	condicional	5.3.1, 5.3.3
«¿son estadísticamente independientes?»	comparar f_{XY} con $f_X f_Y$	5.3.2
«se eligen dos al azar de una caja con a de un tipo, b de otro…»	construir la tabla	5.3.3
una fórmula $f_{XY}(x, y)$ en un rectángulo; «la proporción del día», «el tiempo de mantenimiento»	integral doble	5.3.4
región $0 < y < x < 1$ o $20 < x < y < 40$	límites variables	5.3.5
«distribuidas uniformemente», f constante	áreas	5.3.4
«covariancia», «coeficiente de correlación», «interprételo»	Cov y ρ	5.3.6
«variancia de la suma», «de la diferencia»	combinación lineal	5.3.6

5.3 Procedimientos

5.3.1 Tabla discreta: marginales, probabilidades y condicionales

1. Agrega a la tabla una columna con los totales de cada renglón y un renglón con los totales de cada columna: son las dos marginales.
2. Comprueba que el total general sea 1. Si hay una constante k, despéjala de $\sum \text{celdas} = 1$.
3. **Probabilidad de un evento:** marca las celdas que cumplen la condición y súmalas.
4. **Condicional de un valor:** $P(X = x \mid Y = y) = \dfrac{\text{celda}}{\text{total de su renglón o columna } y}$.
5. **Condicional de eventos:** $P(A \mid B) = \dfrac{\sum \text{celdas en } A \cap B}{\sum \text{celdas en } B}$. Lo que sigue a «dado que» va en el denominador.

5.3.2 Independencia

1. Calcula las marginales.
2. Compara una celda con el producto de sus marginales: $f_{XY}(x, y)$ contra $f_X(x) f_Y(y)$.
3. Empieza por una celda con **0**: si sus marginales son positivas, el producto no es 0 y la independencia falla de inmediato.
4. Si encuentras una celda que no cumple, escribe la desigualdad con números y concluye «no son independientes».
5. Si la celda revisada sí cumple, **no concluyas nada** y sigue con otra; para afirmar la independencia hay que revisar todas.
6. En el caso continuo, factoriza f_{XY} o compara f_{XY} con el producto de las marginales ya calculadas.

2. **Concluir independencia con una sola celda que sí cumple.** Una celda que cumple no prueba nada; para negar basta una que falle.
3. **Creer que covariancia cero implica independencia.** Es falso (ejercicio 5.5.3).
4. **Voltear la condicional.** Lo que sigue a «dado que» o «si se sabe que» va en el denominador. Un resultado mayor que 1 indica que está volteada.
5. **Usar las celdas en lugar de la marginal** para $E(X)$: $E(X)$ usa f_X; $E(XY)$ usa las celdas.
6. **Integrar la variable equivocada en la marginal.** Para f_X se integra y, en **todo** su recorrido.
7. **Límites fijos en una región triangular.** En $0 < y < x < 1$, y va de 0 a x, no de 0 a 1.
8. **No sustituir el valor dado en la condicional continua,** o no dividir entre la marginal evaluada en ese valor.
9. **Escribir la marginal sin dominio** («$0 < y < 2$»): se pierde parte del puntaje.
10. **Mezclar unidades:** si la densidad está en horas, 15 minutos son $\frac{1}{4}$, no 15.
11. **Restar variancias:** $Var(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\,\text{Cov}$; con independencia, las variancias se suman.

5.8 Formulario de emergencia de la unidad

- Marginales: $f_X(x) = \sum_y f_{XY}$ o $\int f_{XY}\,dy \cdot f_Y(y) = \sum_x f_{XY}$ o $\int f_{XY}\,dx$
- Condicional: $f_{Y|x}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)} \cdot P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- Independencia: $f_{XY}(x, y) = f_X(x) f_Y(y)$ para todo (x, y)
- $E[g(X, Y)] = \sum\sum g\, f_{XY}$ o $\iint g\, f_{XY} \cdot E(Y \mid X = x) = \int y\, f_{Y|x}\,dy$
- $\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y \cdot \rho = \frac{\text{Cov}}{\sigma_X \sigma_Y}$, $-1 \le \rho \le 1$
- $Var(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\text{Cov} \cdot$ independientes $\Rightarrow \text{Cov} = 0$ (no al revés)
- Sin reemplazo: $f_{XY} = \frac{\binom{a}{x}\binom{b}{y}\binom{c}{n-x-y}}{\binom{N}{n}} \cdot$ uniforme conjunta: $P = \frac{\text{área}}{\text{área total}}$

5.9 Procedimiento recomendado en el examen

1. Las variables definidas con palabras: «$X \triangleq \ldots$, $Y \triangleq \ldots$».
2. Tabla: las marginales agregadas y el total igual a 1. Densidad: el dibujo de la región y $\iint f = 1$ escrito.
3. La fórmula que se va a usar (condicional, independencia, E, Cov) escrita antes de sustituir.
4. Independencia: la celda o el punto que falla, con números: «$0.10 \neq (0.16)(0.24) = 0.0384$».
5. Probabilidad: las celdas que se suman o los límites de la región, explícitos.
6. El resultado con 4 decimales y una frase: «se esperan 1.875 defectos por producto».

Si no se encuentra el camino, las marginales bien calculadas y la integral planteada con los límites correctos ya valen buena parte del problema.

5.4 Ejemplo resuelto

Adaptado de: 2º final 2011-1, problema 3.

Las variables aleatorias X y Y tienen la función de probabilidad conjunta

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$
$y=0$	0.10	0.08	0.06
$y=1$	0.04	0.20	0.14
$y=2$	0.02	0.06	0.30

- a) Calcule $P(X \leq 1,\ Y \geq 1)$.
- b) Determine $P(X \leq 1 \mid Y \geq 1)$.
- c) Obtenga la covariancia entre las variables.
- d) ¿Son estadísticamente independientes?

Datos. Tabla conjunta de 3×3; X en las columnas y Y en los renglones.
Planteamiento. Marginales por suma de renglones y columnas; probabilidad y condicional por suma de celdas; Cov $= E(XY) - E(X)E(Y)$; independencia con una celda.

$f_{XY}(x,y)$	$x=0$	$x=1$	$x=2$	$f_Y(y)$
$y=0$	0.10	0.08	0.06	0.24
$y=1$	0.04	0.20	0.14	0.38
$y=2$	0.02	0.06	0.30	0.38
$f_X(x)$	0.16	0.34	0.50	1

Resolución.

1. a) Celdas con $x \in \{0,1\}$ y $y \in \{1,2\}$: $P(X \leq 1,\ Y \geq 1) = 0.04 + 0.20 + 0.02 + 0.06 =$ **0.32**.
2. b) $P(Y \geq 1) = 0.38 + 0.38 = 0.76$, así que $P(X \leq 1 \mid Y \geq 1) = \frac{0.32}{0.76} =$ **0.4211**.
3. c) $E(X) = 1(0.34) + 2(0.50) = 1.34$ y $E(Y) = 1(0.38) + 2(0.38) = 1.14$.
4. $E(XY) = (1)(1)(0.20) + (1)(2)(0.06) + (2)(1)(0.14) + (2)(2)(0.30) = 1.80$.
5. $\text{Cov}(X,Y) = 1.80 - (1.34)(1.14) =$ **0.2724**: hay asociación lineal positiva.
6. d) Celda $(0,0)$: $f_{XY}(0,0) = 0.10$, mientras que $f_X(0)f_Y(0) = (0.16)(0.24) = 0.0384$. Como $0.10 \neq 0.0384$, **no son independientes** (también se deduce de Cov $\neq 0$).
7. **Resultado:** 0.32; 0.4211; Cov = 0.2724; variables dependientes, con relación lineal positiva.

Observación (no la pide el examen; sirve para practicar 5.3.6). $E(X^2) = 0.34 + 4(0.50) = 2.34$, $\sigma_X^2 = 2.34 - 1.34^2 = 0.5444$; $E(Y^2) = 0.38 + 4(0.38) = 1.90$, $\sigma_Y^2 = 1.90 - 1.14^2 = 0.6004$. Entonces $\rho = \frac{0.2724}{\sqrt{0.5444(0.6004)}} = 0.4765$ (relación lineal positiva moderada) y $Var(X+Y) = 0.5444 + 0.6004 + 2(0.2724) = 1.6896$.

tiene distribución uniforme en $[2,3]$.

- a) Obtenga la función de densidad conjunta de X y Y.
- b) ¿Cuál es la probabilidad de que ambos lleguen entre las 2:00 y las 2:30?
- c) Si B llega a las 2:00 y espera 15 minutos antes de irse, ¿cuál es la probabilidad de que coman juntos?

Resolución.

1. a) $f_X(x) = 1$ en $[2,3]$ y $f_Y(y) = 1$ en $[2,3]$. Por independencia, $f_{XY}(x,y) = f_X(x)f_Y(y) =$ **1,** $2 \leq x \leq 3$**,** $2 \leq y \leq 3$.
2. b) La densidad es constante: $P(X < 2.5,\ Y < 2.5) = \frac{(0.5)(0.5)}{1} =$ **0.25**.
3. c) Comen juntos si A llega a más tardar a las 2:15, es decir, $X \leq 2.25$ dado $Y = 2$.
4. $f_{X|Y=2}(x) = \frac{f_{XY}(x,2)}{f_Y(2)} = 1$, $2 \leq x \leq 3$ (por la independencia, es la marginal de X).
5. $P(X \leq 2.25 \mid Y = 2) = \int_2^{2.25} dx =$ **0.25**.

5.6 Ejercicios propuestos

Resuélvelos sin ver la solución. Las resoluciones completas están en el **Apéndice E, sección E.5**.

Ejercicio propuesto 5.6.1

Adaptado de: 1er final 2014-1, problema 5; repetido en 2º final 2014-1, problema 5.

Un consultorio tiene dos líneas telefónicas. En un día elegido al azar, X es la proporción del tiempo que se usa la línea 1 y Y la de la línea 2, con densidad conjunta $f_{XY}(x,y) = \frac{2}{3}(x+2y)$, $0 \leq x \leq 1$, $0 \leq y \leq 1$. ¿Cuál es la probabilidad de que la línea 2 esté libre durante el 80 % del día?

Resultado: $P(Y \leq 0.2) = \frac{7}{75} = 0.0933$.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 5.6.2

Adaptado de: 2º final 2016-2, problema 5.

X es el diámetro de los puntos de soldadura (cm) y Y la resistencia a la tensión (ohms) de un dispositivo, con densidad conjunta $f_{XY}(x,y) = \frac{1}{500}$, $0 \leq x < 0.25$, $0 \leq y < 2000$. Calcule la probabilidad de que:

- a) el diámetro esté entre 0.1 y 0.2 cm y, al mismo tiempo, la resistencia entre 100 y 200 ohms;
- b) la resistencia sea menor que 1000 ohms.

Resultado: 0.02; 0.5.

→ Resolución en el Apéndice E, p. 104.

Ejercicio propuesto 5.6.3

Adaptado de: 1er final 2012-1, problema 5; la misma densidad, con la pregunta del inciso c, en 2º final 2017-1, problema 4.

Una instalación telefónica opera con dos líneas. X es la proporción del día que se usa la

1. Marginales: $f_X(0,1,2,3) = \frac{3}{16}, \frac{2}{16}, \frac{5}{16}, \frac{6}{16}$; $f_Y(1) = f_Y(2) = \frac{8}{16}$. Total: 1.
2. a) Celda $(0,1)$: $f_{XY}(0,1) = \frac{2}{16} = 0.125$, mientras que $f_X(0)f_Y(1) = \frac{3}{16} \cdot \frac{8}{16} = 0.09375$. Son distintos: **no son independientes**.
3. b) $E(X) = 0 \cdot \frac{3}{16} + 1 \cdot \frac{2}{16} + 2 \cdot \frac{5}{16} + 3 \cdot \frac{6}{16} = \frac{30}{16} =$ **1.875 defectos**.

Ejercicio 5.5.3 · Covariancia cero sin independencia

Adaptado de: 1er final 2016-1, problema 4.

Las variables X y Y tienen la función de probabilidad conjunta:

$f_{XY}(x,y)$	$y=-1$	$y=0$	$y=1$
$x=-1$	1/5	0	1/5
$x=0$	0	1/5	0
$x=1$	1/5	0	1/5

Determine si X y Y son estadísticamente independientes y justifique.

Resolución.

1. Marginales: $f_X(-1,0,1) = \frac{2}{5}, \frac{1}{5}, \frac{2}{5}$ y $f_Y(-1,0,1) = \frac{2}{5}, \frac{1}{5}, \frac{2}{5}$.
2. Celda $(-1,-1)$: $f_{XY}(-1,-1) = 0.2$, mientras que $f_X(-1)f_Y(-1) = \frac{2}{5} \cdot \frac{2}{5} = 0.16$. **No son independientes.**
3. Más rápido: $f_{XY}(-1,0) = 0$, pero $f_X(-1)f_Y(0) = \frac{2}{25} \neq 0$.
4. **Observación:** $E(X) = E(Y) = 0$ y $E(XY) = \frac{1}{5}(1-1-1+1) = 0$, así que $\mathrm{Cov}(X,Y) = 0$. Las variables tienen covariancia cero **y aun así** son dependientes: $X^2 = Y^2$ en todas las celdas con probabilidad positiva, una relación que no es lineal.

Ejercicio 5.5.4 · Tabla con una celda engañosa y condicional de eventos

Adaptado de: 1er final 2016-2, problema 5.

Las variables X y Y tienen la distribución conjunta:

$f_{XY}(x,y)$	$y=-1$	$y=1$
$x=1$	1/6	1/3
$x=2$	1/12	1/4
$x=3$	1/12	1/12

- a) Determine si X y Y son estadísticamente independientes.
- b) Calcule $P(X \geq 2 \mid Y \leq 0)$.

Resolución.

1. Marginales: $f_X(1,2,3) = \frac{6}{12}, \frac{4}{12}, \frac{2}{12}$; $f_Y(-1) = \frac{4}{12}$ y $f_Y(1) = \frac{8}{12}$. Total: 1.
2. a) La celda $(1,-1)$ **sí cumple**: $\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}$. No se concluye nada; se revisa otra.
3. Celda $(2,-1)$: $f_{XY}(2,-1) = \frac{1}{12} = 0.0833$, mientras que $f_X(2)f_Y(-1) = \frac{4}{12} \cdot \frac{4}{12} = \frac{1}{9} = 0.1111$. **No son independientes.**

4. b) $Y \leq 0$ es solo la columna $y=-1$: $P(Y \leq 0) = \frac{4}{12}$.
5. $P(X \geq 2,\ Y \leq 0) = f_{XY}(2,-1) + f_{XY}(3,-1) = \frac{1}{12} + \frac{1}{12} = \frac{2}{12}$.
6. $P(X \geq 2 \mid Y \leq 0) = \frac{2/12}{4/12} =$ **0.5**.

Ejercicio 5.5.5 · Mantenimiento mecánico y eléctrico: densidad en un rectángulo

Adaptado de: 1er final 2011-1, problema 5.

El tiempo semanal de mantenimiento de una máquina depende de X, la duración del mantenimiento mecánico, y Y, la del eléctrico, ambas en horas, con densidad conjunta

$$f_{XY}(x,y) = \frac{2}{3}(x+2y), \qquad 0 \leq x \leq 1,\ 0 \leq y \leq 1.$$

- a) Calcule la probabilidad de que en una semana el mantenimiento mecánico dure menos de 15 minutos y el eléctrico más de 30 minutos.
- b) Determine la función de densidad marginal de X.
- c) Obtenga la función de densidad condicional de Y dado $X = x_0$.
- d) Calcule la probabilidad de que el mantenimiento eléctrico dure menos de 15 minutos, dado que el mecánico duró 30 minutos.

Resolución.

1. Las unidades son horas: 15 min $= \frac{1}{4}$ h y 30 min $= \frac{1}{2}$ h.
2. a) $P\left(X < \frac{1}{4},\ Y > \frac{1}{2}\right) = \int_{1/2}^{1}\int_{0}^{1/4} \frac{2}{3}(x+2y)\,dx\,dy = \frac{2}{3}\int_{1/2}^{1}\left(\frac{1}{32} + \frac{y}{2}\right)dy$.
3. $= \frac{2}{3}\left[\frac{y}{32} + \frac{y^2}{4}\right]_{1/2}^{1} = \frac{2}{3}\left(\frac{1}{64} + \frac{3}{16}\right) = \frac{13}{96} =$ **0.1354**.
4. b) $f_X(x) = \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}\left[xy + y^2\right]_0^1 = \frac{2}{3}(x+1)$, $0 \leq x \leq 1$.
5. c) $f_{Y|x_0}(y) = \dfrac{\frac{2}{3}(x_0+2y)}{\frac{2}{3}(x_0+1)} = \dfrac{x_0+2y}{x_0+1}$, $0 \leq y \leq 1$.
6. d) Con $x_0 = \frac{1}{2}$: $f_{Y|1/2}(y) = \frac{\frac{1}{2}+2y}{\frac{3}{2}} = \frac{1+4y}{3}$, $0 \leq y \leq 1$.
7. $P\left(Y < \frac{1}{4} \mid X = \frac{1}{2}\right) = \int_0^{1/4} \frac{1+4y}{3}\,dy = \frac{1}{3}\left[y + 2y^2\right]_0^{1/4} = \frac{1}{3}\left(\frac{1}{4} + \frac{1}{8}\right) =$ **0.125**.

Ejercicio 5.5.6 · Densidad $\frac{xy}{9}$: región, valor esperado y condicional

Adaptado de: 2º final 2018-1, problema 3.

Dada la función de densidad conjunta $f_{XY}(x,y) = \frac{xy}{9}$, $0 < x < 3$, $0 < y < 2$:

- a) Obtenga la probabilidad de que X sea menor que 2 y, al mismo tiempo, Y sea mayor que 1.
- b) Obtenga el valor esperado de Y.
- c) Obtenga la probabilidad de que X sea mayor que 2, si Y es igual a 1.

Resolución.

1. Comprobación: $\int_0^3\int_0^2 \frac{xy}{9}\,dy\,dx = \frac{1}{9} \cdot \frac{9}{2} \cdot 2 = 1$.