14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Clase 5

Distribuciones Especiales, Media Muestral y TLC

Teoría

Módulos oficiales del MicroMaster:
M5 – Special Distributions, The Sample Mean, the Central Limit Theorem and Estimation

B.Sc. Cristian Lazo Quispe
clazoq@uni.pe
linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

11 de agosto de 2026

Índice

1. De qué trata esta clase

Hasta ahora describimos variables aleatorias *en general*: con su pmf, su pdf, su esperanza y su varianza. Pero unas pocas distribuciones aparecen *una y otra vez* en la práctica, porque modelan situaciones muy comunes: contar éxitos en encuestas, tiempos de espera, conteos de eventos raros, errores de medición. Son las **distribuciones especiales (special distributions)**: tienen nombre propio, fórmulas conocidas de $\mathbb{E}[X]$ y $\mathrm{Var}(X)$, y se relacionan entre sí de maneras útiles.

Luego damos el salto de la *probabilidad* a la *estadística*. Cuando tomamos una muestra y calculamos su promedio, ese promedio —la **media muestral (sample mean)** $\bar{X}$— es él mismo una variable aleatoria. Estudiar cómo se distribuye nos lleva al resultado más importante de todo el curso: el **Teorema del Límite Central (Central Limit Theorem, TLC/CLT)**, que explica por qué la campana de Gauss aparece en todas partes. Terminamos con la idea de **estimación (estimation)**: usar los datos para adivinar un parámetro desconocido de la población.

Un catálogo de distribuciones especiales es como tener *plantillas* listas: en vez de deducir todo desde cero, reconoces "esto es un conteo de éxitos en n intentos $\Rightarrow$ Binomial" o "esto es un tiempo de espera $\Rightarrow$ Exponencial", y ya sabes su media, su varianza y cómo calcular probabilidades. El TLC es la joya: dice que el *promedio* de muchas observaciones se vuelve normal, sin importar de qué distribución salgan.

2. Distribuciones discretas especiales

Recordemos la notación: p es la probabilidad de *éxito (success)*, $q = 1 - p$ la de *fracaso (failure)*, y el **soporte (support)** de X es el conjunto de valores con probabilidad positiva.

2.1. Bernoulli

Definición 2.1 (Distribución Bernoulli)**.** $X \sim \text{Bernoulli}(p)$ modela *un* ensayo con dos resultados: éxito ($X = 1$) con probabilidad p, fracaso ($X = 0$) con probabilidad $q = 1 - p$. Su pmf es

$$f_X(x) = p^x(1-p)^{1-x}, \qquad x \in \{0, 1\}.$$

$$\mathbb{E}[X] = p, \qquad \mathrm{Var}(X) = p(1-p) = pq.$$

Es el ladrillo básico: cualquier "sí/no", "votó/no votó", "aprobó/no aprobó". La esperanza p es, simplemente, la proporción de éxitos a largo plazo. La varianza $p(1-p)$ es máxima en $p = 0.5$ (incertidumbre máxima) y cae a 0 en los extremos $p = 0$ o $p = 1$ (resultado casi seguro).

Ejemplo 2.1 (Indicador de pobreza)**.** Sea $X = 1$ si un hogar peruano elegido al azar está por debajo de la línea de pobreza, y $X = 0$ si no. Si la tasa de pobreza es $p = 0.27$, entonces $X \sim \text{Bernoulli}(0.27)$, con $\mathbb{E}[X] = 0.27$ (la proporción poblacional) y $\text{Var}(X) = 0.27 \cdot 0.73 = 0.1971$.

2.2. Binomial

Definición 2.2 (Distribución Binomial)**.** Si $X_1, \ldots, X_n$ son Bernoulli(p) **independientes e idénticamente distribuidas (iid)**, su suma $X = \sum_{k=1}^{n} X_k$ es **Binomial**, $X \sim \text{Bin}(n, p)$: cuenta el *número de éxitos en n ensayos independientes*. Su pmf es

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n.$$

Como suma de n Bernoulli iid,

$$\mathbb{E}[X] = np, \qquad \text{Var}(X) = np(1-p) = npq.$$

La Bernoulli es el caso $n = 1$. El factor $\binom{n}{x}$ cuenta de cuántas formas pueden ocurrir x éxitos entre los n ensayos; $p^x(1-p)^{n-x}$ es la probabilidad de *una* secuencia concreta con x éxitos. La media np es muy intuitiva: si lanzas una moneda justa ($p = 0.5$) $n = 100$ veces, esperas ≈ 50 caras.

Ejemplo 2.2 (Encuesta de aprobación)**.** En un distrito, el 40 % de los votantes aprueba a su alcalde ($p = 0.4$). Encuestamos a $n = 10$ personas al azar. El número X que aprueba es $\text{Bin}(10, 0.4)$, con

$$\mathbb{E}[X] = 10(0.4) = 4, \qquad \text{Var}(X) = 10(0.4)(0.6) = 2.4, \quad \text{SD}(X) = \sqrt{2.4} \approx 1.55.$$

La probabilidad de que exactamente 3 aprueben es $f_X(3) = \binom{10}{3}(0.4)^3(0.6)^7 \approx 0.215$.

2.3. Geométrica

Definición 2.3 (Distribución Geométrica)**.** $X \sim \text{Geom}(p)$ cuenta el *número de ensayos hasta el primer éxito* (incluyéndolo), en ensayos Bernoulli(p) independientes. Su pmf es

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

$$\mathbb{E}[X] = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2}.$$

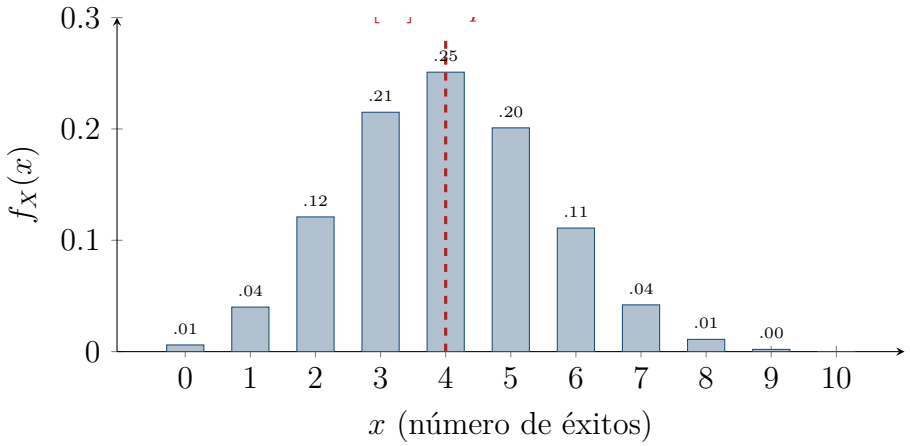

Figura 1: pmf de una Bin$(10, 0.4)$. La línea roja marca la media $\mathbb{E}[X] = np = 4$. Note la forma acampanada (un anticipo del TLC: la Binomial es suma de Bernoulli iid).

Para tener el primer éxito en el ensayo k, hacen falta $k-1$ fracasos seguidos (cada uno con prob. $1-p$) y luego un éxito (prob. p). La media $1/p$ es muy natural: si la probabilidad de éxito es $p = 0.2$, en promedio necesitas $1/0.2 = 5$ intentos. La geométrica tiene **falta de memoria**: si ya fallaste 20 veces, la cuenta "se reinicia".

Ejemplo 2.3 (Postulaciones a un empleo)**.** Una persona envía postulaciones independientes; cada una tiene probabilidad $p = 0.1$ de conseguir entrevista. Sea X el número de postulaciones hasta la primera entrevista. Entonces $X \sim \text{Geom}(0.1)$, con $\mathbb{E}[X] = 1/0.1 = 10$ postulaciones en promedio, y $\text{Var}(X) = (0.9)/(0.1)^2 = 90$. La probabilidad de lograrlo justo en el tercer intento es $f_X(3) = (0.9)^2(0.1) = 0.081$.

2.4. Binomial negativa

Definición 2.4 (Distribución Binomial negativa)**.** $X \sim \text{NegBin}(r, p)$ cuenta el *número de ensayos hasta el r-ésimo éxito*. Su pmf es

$$f_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \qquad k = r,\, r+1,\, r+2, \ldots$$

$$\mathbb{E}[X] = \frac{r}{p}, \qquad \text{Var}(X) = \frac{r(1-p)}{p^2}.$$

Es la generalización de la geométrica: la geométrica es el caso $r = 1$. De hecho, una $\text{NegBin}(r, p)$ es la *suma de r geométricas(p) independientes*, lo que explica de inmediato por qué $\mathbb{E}[X] = r \cdot (1/p) = r/p$ (linealidad de la esperanza) y $\text{Var}(X) = r \cdot (1-p)/p^2$ (varianza de una suma de independientes).

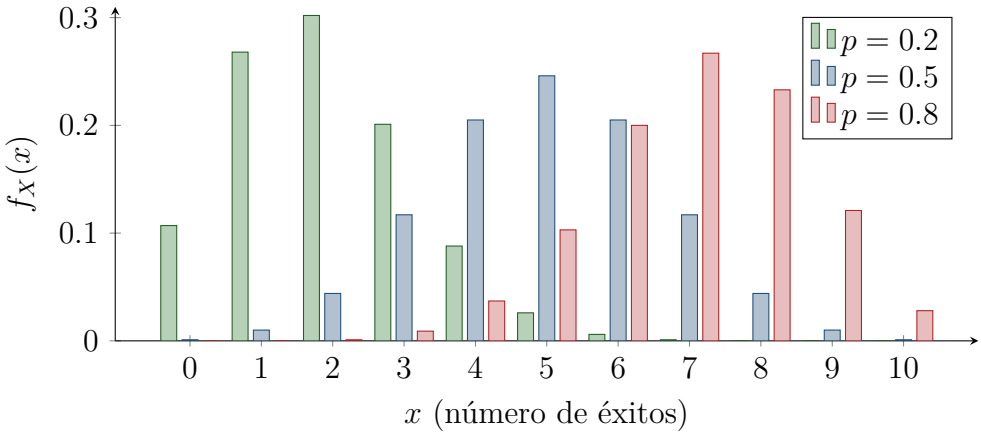

Figura 2: Efecto del parámetro p en una $\text{Bin}(10, p)$. Con $p = 0.2$ (verde) la masa se concentra a la izquierda (pocos éxitos); con $p = 0.5$ (azul) es *simétrica* y centrada en 5; con $p = 0.8$ (rojo) se corre a la derecha. En los tres casos la media está en np y la dispersión es máxima cerca de $p = 0.5$.

Ejemplo 2.4 (Encuestador puerta a puerta)**.** Un encuestador necesita $r = 3$ entrevistas completas. Cada hogar acepta responder con probabilidad $p = 0.25$, de forma independiente. Sea X el número de hogares que visita hasta completar las 3 entrevistas. Entonces $X \sim \text{NegBin}(3, 0.25)$, con

$$\mathbb{E}[X] = \frac{3}{0.25} = 12 \text{ hogares}, \qquad \text{Var}(X) = \frac{3(0.75)}{0.25^2} = \frac{2.25}{0.0625} = 36.$$

2.5. Hipergeométrica

Definición 2.5 (Distribución Hipergeométrica)**.** Suponga una población finita con A éxitos y B fracasos ($N = A + B$). Si extraemos una muestra de tamaño n **sin reemplazo**, el número de éxitos X es **Hipergeométrica**:

$$f_X(x) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}, \qquad x = \text{máx}(0,\, n-B), \ldots, \text{mín}(n, A).$$

Con $p = \dfrac{A}{A+B}$,

$$\mathbb{E}[X] = n\,\frac{A}{A+B} = np, \qquad \text{Var}(X) = n\,\frac{A}{A+B}\cdot\frac{B}{A+B}\cdot\frac{A+B-n}{A+B-1}.$$

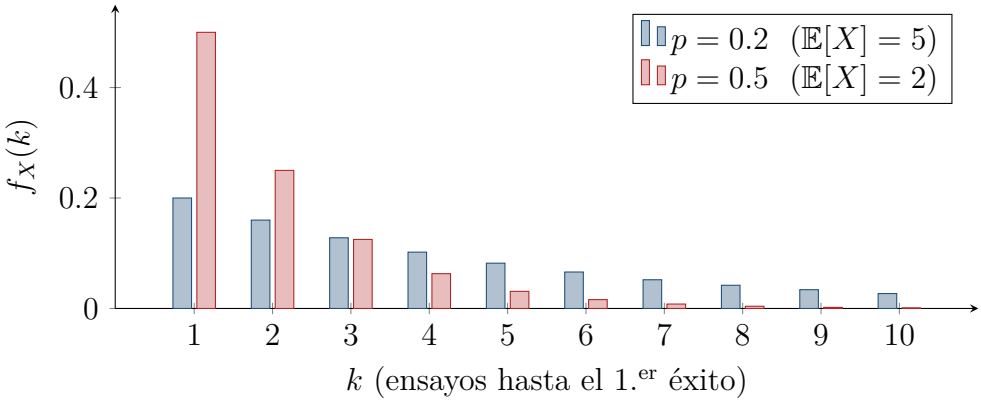

Figura 3: pmf de la Geométrica: *decaimiento* geométrico. Cada barra es $(1-p)$ veces la anterior, así que la probabilidad cae más rápido cuando p es grande (rojo, $p = 0.5$: la mitad cada paso) y más lento cuando p es pequeño (azul, $p = 0.2$). El éxito *más probable* siempre es en el primer intento ($k = 1$).

La diferencia clave con la Binomial es el *sin reemplazo*: cada extracción cambia la composición de lo que queda, así que los ensayos **no** son independientes. El último factor $\frac{A+B-n}{A+B-1}$ (la *corrección por población finita*) reduce la varianza respecto de la Binomial. Cuando la población $N = A + B$ es mucho mayor que la muestra n, ese factor $\to 1$ y la Hipergeométrica se aproxima a una $\text{Bin}(n, p)$.

Ejemplo 2.5 (Auditoría de expedientes)**.** Un lote tiene $N = 20$ expedientes, de los cuales $A = 5$ tienen un error y $B = 15$ están correctos. Un auditor revisa $n = 4$ al azar (sin reemplazo). El número X de expedientes con error es Hipergeométrico, con $\mathbb{E}[X] = 4 \cdot \frac{5}{20} = 1$. La probabilidad de no hallar ningún error es

$$f_X(0) = \frac{\binom{5}{0}\binom{15}{4}}{\binom{20}{4}} = \frac{1 \cdot 1365}{4845} \approx 0.282.$$

2.6. Poisson

Definición 2.6 (Distribución de Poisson)**.** $X \sim \text{Pois}(\lambda)$ modela el *número de eventos que ocurren en un intervalo fijo* (de tiempo o espacio) cuando los eventos son independientes y ocurren a una tasa media constante $\lambda > 0$. Su pmf es

$$f_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2, 3, \ldots$$

$$\mathbb{E}[X] = \lambda, \qquad \text{Var}(X) = \lambda.$$

La Poisson es la distribución de los *conteos*: goles en un partido, llamadas a un call center por hora, accidentes en una vía por mes, errores tipográficos por página. Una peculiaridad: **la media y la varianza son iguales** ($= \lambda$). El único parámetro λ controla las dos cosas a la vez; si los datos tienen más dispersión que media ("sobredispersión"), la Poisson no basta y se usa la Binomial negativa.

Relación Binomial → Poisson (eventos raros)

Cuando hay *muchos* ensayos pero la probabilidad de éxito es *pequeña*, la Binomial se aproxima a una Poisson. Formalmente, si $n \to \infty$ y $p \to 0$ manteniendo $np = \lambda$ fijo,

$$\binom{n}{x} p^x (1-p)^{n-x} \longrightarrow \frac{\lambda^x e^{-\lambda}}{x!}.$$

Por eso la Poisson se llama la "ley de los **eventos raros (law of rare events)**": muchas oportunidades, cada una poco probable, dan un conteo total tipo Poisson con $\lambda = np$.

Ejemplo 2.6 (Accidentes de tránsito)**.** En un cruce de Lima ocurren en promedio $\lambda = 2$ accidentes por semana, de forma independiente. Modelamos el número semanal $X \sim \text{Pois}(2)$, con $\mathbb{E}[X] = \text{Var}(X) = 2$. La probabilidad de una semana *sin* accidentes es

$$f_X(0) = \frac{2^0 e^{-2}}{0!} = e^{-2} \approx 0.135,$$

y la de exactamente 3 accidentes es $f_X(3) = \dfrac{2^3 e^{-2}}{3!} = \dfrac{8e^{-2}}{6} \approx 0.180.$

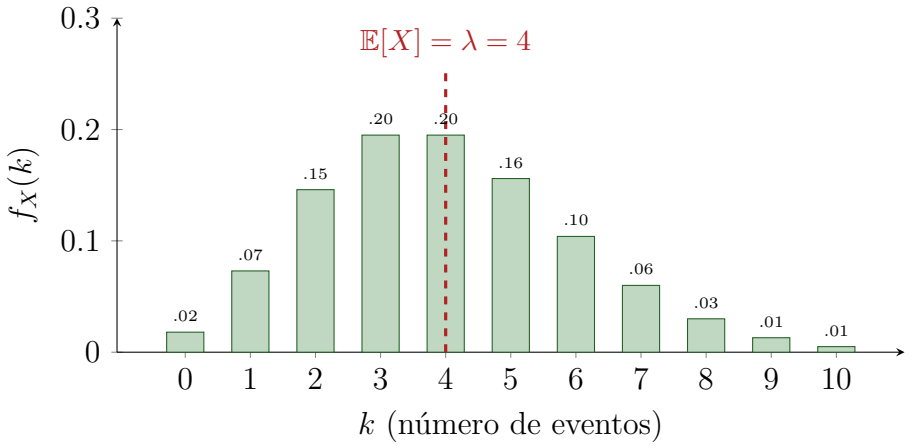

Figura 4: pmf de una Pois(4). La media y la varianza valen ambas $\lambda = 4$. La distribución es asimétrica (no puede ser negativa) y se vuelve más simétrica al crecer λ.

Ejemplo 2.7 (Llegadas a una oficina del Estado (Poisson) con su gráfico)**.** A una ventanilla de una oficina pública llegan, en promedio, $\lambda = 5$ personas cada 10 minutos, de forma

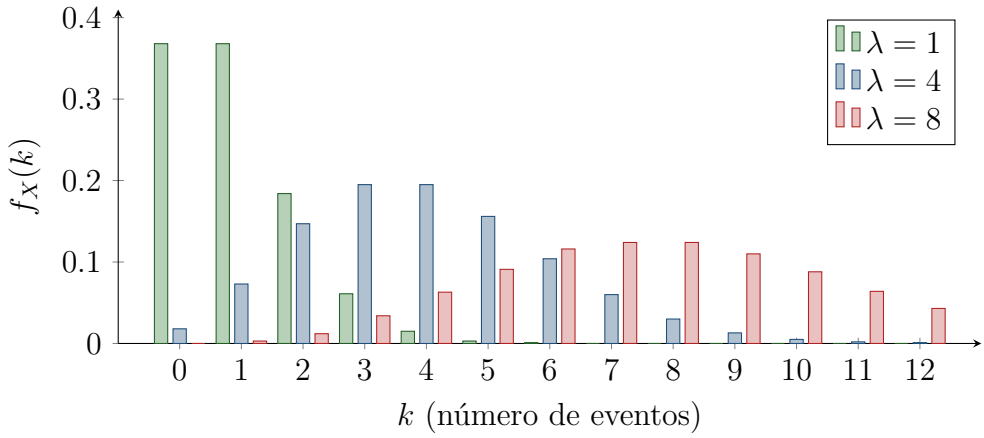

Figura 5: Efecto de λ en la Poisson. Con λ pequeño (verde, $\lambda = 1$) la masa se aprieta cerca de 0 y es muy asimétrica; al crecer λ (azul 4, rojo 8) la distribución se *desplaza* a la derecha, se *ensancha* (recuerde Var $= \lambda$) y se vuelve cada vez más *simétrica* —acercándose a una campana normal, otro anticipo del TLC.

independiente. Modelamos el número X de llegadas en 10 minutos como $X \sim \text{Pois}(5)$, con $\mathbb{E}[X] = \text{Var}(X) = 5$. La probabilidad de que lleguen *a lo sumo* 3 personas en ese lapso es

$$\mathbb{P}(X \le 3) = \sum_{k=0}^{3} \frac{5^k e^{-5}}{k!} = e^{-5}\left(1 + 5 + \tfrac{25}{2} + \tfrac{125}{6}\right) \approx 0.265.$$

En la figura 6 se sombrean en rojo justamente esas barras ($k = 0, 1, 2, 3$).

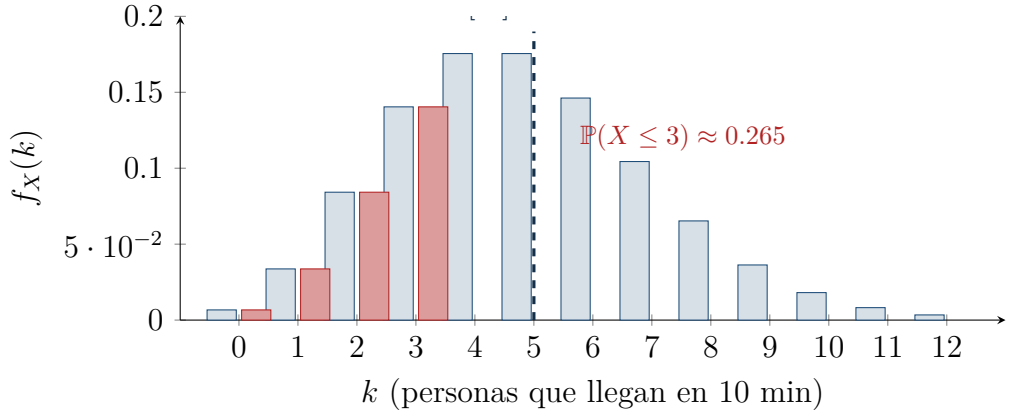

Figura 6: pmf de Pois(5) (llegadas en 10 min). Las barras rojas ($k = 0, \ldots, 3$) suman $\mathbb{P}(X \le 3) \approx 0.265$. La línea marca la media $\lambda = 5$.

Algunos textos definen la geométrica como el *número de fracasos antes del primer éxito* ($k = 0, 1, 2, \ldots$, con $\mathbb{E}[X] = q/p$), y otros como el *número de ensayos hasta el primer éxito*

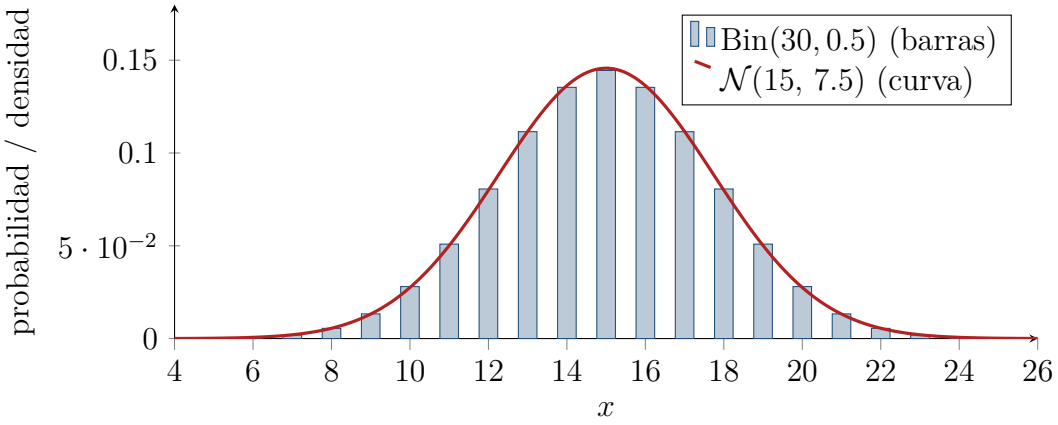

Figura 7: **Aproximación Normal a la Binomial.** Las barras son la pmf de una $\text{Bin}(30, 0.5)$; la curva roja es la $\mathcal{N}(np, np(1-p)) = \mathcal{N}(15, 7.5)$. Cuando n es grande (y p no muy extremo), la campana normal calca la forma de la Binomial: por eso un conteo grande se puede aproximar con la Normal (con *corrección por continuidad* si se quiere precisión).

($k = 1, 2, \ldots$, con $\mathbb{E}[X] = 1/p$). Aquí usamos la segunda (consistente con el cheatsheet del curso). Lo importante es ser coherente: ¡siempre revisa qué cuenta exactamente la variable!

3. Distribuciones continuas especiales

3.1. Uniforme continua

Definición 3.1 (Distribución Uniforme continua)**.** $X \sim \text{Unif}[a, b]$ reparte la probabilidad *por igual* sobre el intervalo $[a, b]$: la probabilidad de caer en un sub-intervalo depende solo de su *longitud*. Su pdf es

$$f_X(x) = \frac{1}{b-a}, \quad a \le x \le b \quad \text{(y 0 fuera)}.$$

$$\mathbb{E}[X] = \frac{a+b}{2}, \qquad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

La densidad es "plana": ningún valor del intervalo es más probable que otro. La media es el punto medio $\frac{a+b}{2}$ por simetría. La uniforme estándar $\text{Unif}[0, 1]$ es el motor de los **generadores de números pseudo-aleatorios**: casi toda simulación en una computadora empieza por sortear un $U \sim \text{Unif}[0, 1]$.

Ejemplo 3.1 (Asignación a tratamiento)**.** Para un experimento, a cada participante le asignamos un número $U \sim \text{Unif}[0, 1]$ y lo ponemos en el grupo de tratamiento si $U < 0.5$.

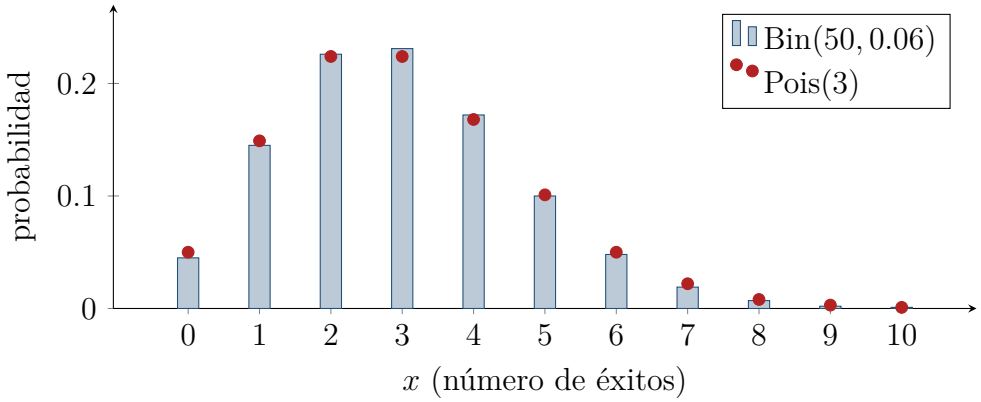

Figura 8: **Aproximación Poisson a la Binomial** (eventos raros). Las barras son $\text{Bin}(50, 0.06)$ y los puntos rojos, $\text{Pois}(\lambda)$ con $\lambda = np = 3$. Con muchos ensayos ($n = 50$) y éxito poco probable ($p = 0.06$), ambas casi coinciden: ese es el sentido de la "ley de los eventos raros".

Como $\mathbb{P}(U < 0.5) = 0.5$, en promedio la mitad recibe el tratamiento. Aquí $\mathbb{E}[U] = 0.5$ y $\text{Var}(U) = \frac{1}{12} \approx 0.083$.

3.2. Exponencial

Definición 3.2 (Distribución Exponencial)**.** $X \sim \text{Exp}(\lambda)$ modela el *tiempo de espera entre dos eventos* de un proceso de Poisson con tasa $\lambda > 0$. Su pdf es

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \quad (\text{y } 0 \text{ si } x < 0),$$

y su cdf es $F_X(x) = 1 - e^{-\lambda x}$ para $x \geq 0$. Además

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \text{Var}(X) = \frac{1}{\lambda^2}.$$

Falta de memoria (memorylessness)

La exponencial es la *única* distribución continua sin memoria:

$$\mathbb{P}(X \geq t + h \mid X \geq t) = \mathbb{P}(X \geq h) \qquad \text{para todo } t, h \geq 0.$$

En palabras: haber esperado ya t unidades **no** cambia la distribución del tiempo que falta. Una bombilla "sin memoria" que llevó 1000 horas encendida es, en cuanto a su vida restante, igual a una nueva. (Prueba: $\mathbb{P}(X \geq t) = e^{-\lambda t}$, así que el cociente $e^{-\lambda(t+h)}/e^{-\lambda t} = e^{-\lambda h}$.)

Es la versión continua de la geométrica (que también es "sin memoria", pero en ensayos discretos). El parámetro λ es la *tasa* de eventos por unidad de tiempo; el tiempo medio

de espera es su inverso $1/\lambda$. Tasa alta $\Rightarrow$ esperas cortas.

Ejemplo 3.2 (Tiempo hasta el próximo cliente)**.** A una bodega llegan clientes a una tasa de $\lambda = 3$ por hora. El tiempo X (en horas) hasta el próximo cliente es Exp(3), con $\mathbb{E}[X] = 1/3$ h $= 20$ min de espera media. La probabilidad de esperar más de media hora es

$$\mathbb{P}(X > 0.5) = e^{-3 \cdot 0.5} = e^{-1.5} \approx 0.223.$$

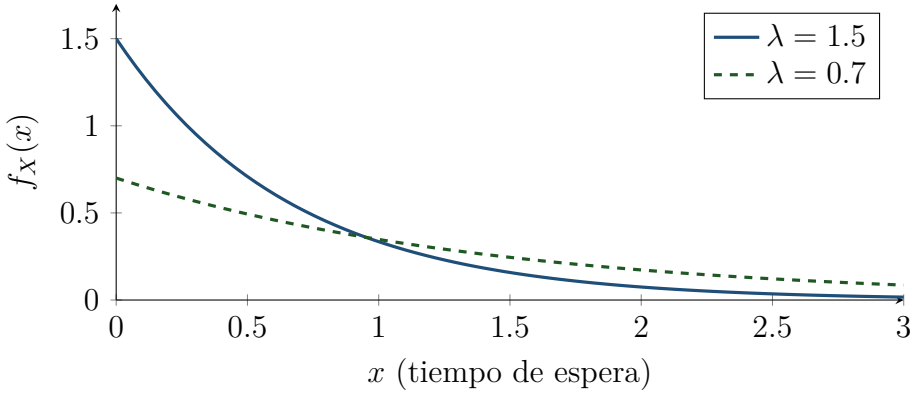

Figura 9: pdf de la Exponencial para dos tasas. Mayor λ (azul) concentra la masa cerca de 0 (esperas cortas); menor λ (verde) tiene cola más larga. La densidad siempre *decae* de forma exponencial.

Ejemplo 3.3 (Tiempo de espera en un paradero (Exponencial) y $\mathbb{P}(X > t)$)**.** En un paradero de Lima, el tiempo X (en minutos) hasta el próximo bus sigue una Exp(λ) con tasa $\lambda = 1/8$ por minuto (es decir, en promedio $\mathbb{E}[X] = 1/\lambda = 8$ min). La probabilidad de *esperar más de* 10 *minutos* usa la "cola" de la cdf:

$$\mathbb{P}(X > 10) = e^{-\lambda \cdot 10} = e^{-10/8} = e^{-1.25} \approx 0.287.$$

Esa probabilidad es exactamente el área sombreada bajo la cola en la figura 10.

3.3. Normal (Gaussiana)

Definición 3.3 (Distribución Normal)**.** $X \sim \mathcal{N}(\mu, \sigma^2)$ es la clásica "campana de Gauss", con parámetros μ (centro) y $\sigma^2 > 0$ (dispersión). Su pdf es

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}(X) = \sigma^2.$$

La **normal estándar (standard normal)** es $Z \sim \mathcal{N}(0, 1)$; su pdf se denota φ y su cdf Φ.

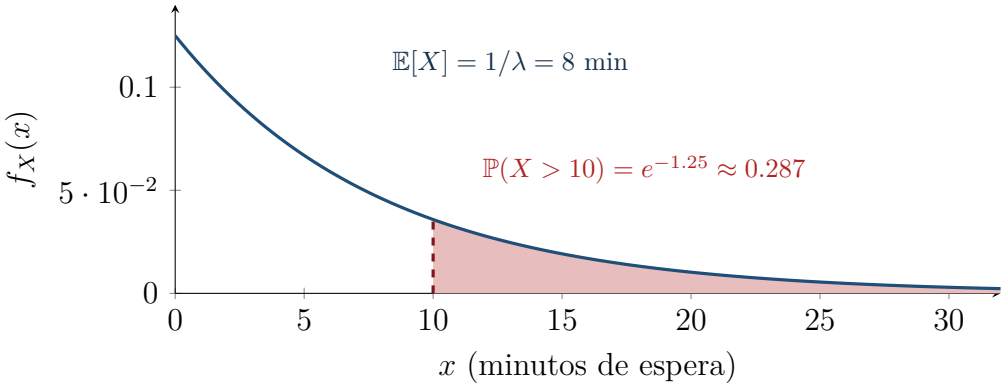

Figura 10: pdf de Exp(1/8) (espera del bus, media 8 min). El área roja a la derecha de $x = 10$ es $\mathbb{P}(X > 10) = e^{-\lambda t} \approx 0.287$: la probabilidad de esperar más de 10 minutos.

Estandarización y regla 68–95–99.7

Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Esto permite calcular cualquier probabilidad usando una única tabla (o `pnorm` en R). Para la normal vale la **regla empírica**:

$$\mathbb{P}(|X - \mu| \le \sigma) \approx 68.2\,\%, \quad \mathbb{P}(|X - \mu| \le 2\sigma) \approx 95\,\%, \quad \mathbb{P}(|X - \mu| \le 3\sigma) \approx 99.7\,\%.$$

Además, *toda combinación lineal de normales independientes es normal*: si $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ indep., entonces $\sum X_i \sim \mathcal{N}\Big(\sum \mu_i, \sum \sigma_i^2\Big)$.

La normal es simétrica, unimodal, con colas finas, y soporte en toda la recta. Aparece naturalmente como "error" de muchas pequeñas causas que se suman (estatura, errores de medición, puntajes). El *porqué* de su omnipresencia es, justamente, el Teorema del Límite Central que veremos enseguida.

Ejemplo 3.4 (Puntajes estandarizados)**.** Los puntajes de una prueba nacional se modelan como $X \sim \mathcal{N}(500, 100^2)$ ($\mu = 500$, $\sigma = 100$). ¿Qué proporción de alumnos saca más de 650? Estandarizamos:

$$Z = \frac{650 - 500}{100} = 1.5, \qquad \mathbb{P}(X > 650) = \mathbb{P}(Z > 1.5) = 1 - \Phi(1.5) \approx 1 - 0.933 = 0.067.$$

Es decir, alrededor del 6.7 % de los alumnos supera los 650 puntos.

Ejemplo 3.5 (Aproximación Normal a un conteo grande, con región sombreada)**.** Una ONG entrega $n = 500$ bonos; cada beneficiario los usa correctamente con probabilidad $p = 0.7$, de forma independiente. Sea $X \sim \text{Bin}(500, 0.7)$ el número que los usa bien. Como n es grande, aproximamos por la Normal:

$$\mu = np = 350, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{500 \cdot 0.7 \cdot 0.3} = \sqrt{105} \approx 10.25.$$

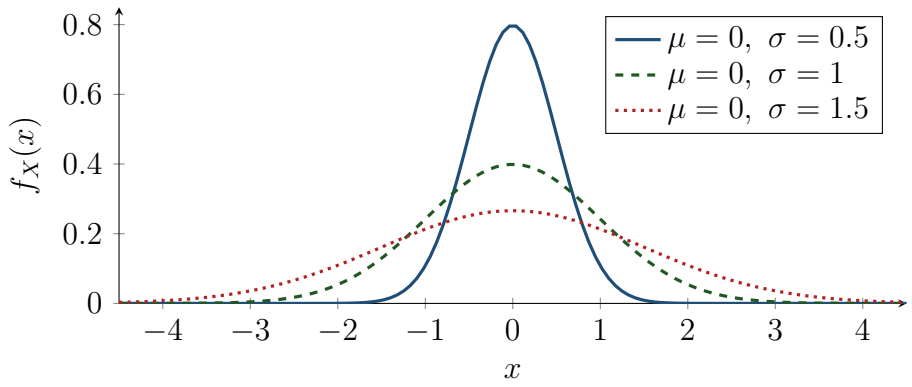

Figura 11: pdf de la Normal con la misma media $\mu = 0$ y distintas σ. Menor σ (azul) da una campana más alta y angosta; mayor σ (roja), más baja y ancha. El área bajo cada curva es siempre 1.

¿Probabilidad de que *más de* 370 los usen bien?

$$Z = \frac{370 - 350}{10.25} \approx 1.95, \qquad \mathbb{P}(X > 370) \approx 1 - \Phi(1.95) \approx 0.026.$$

El área roja de la cola en la figura 12 es justamente ese $\approx 2.6\,\%$.

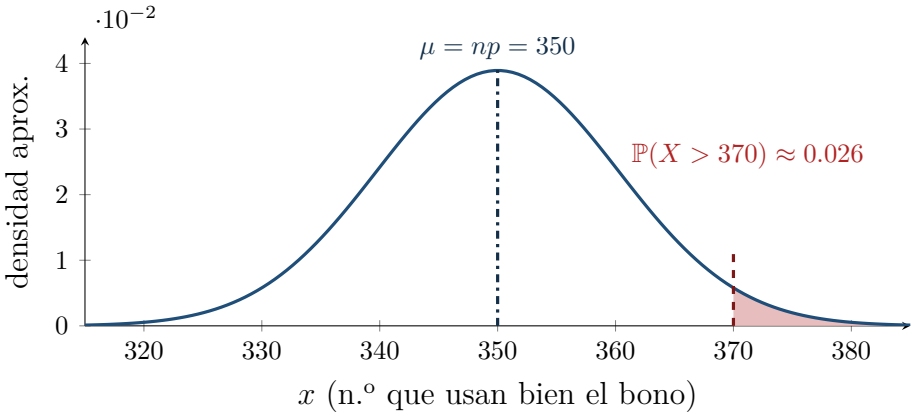

Figura 12: Aproximación Normal a Bin$(500, 0.7)$: la campana $\mathcal{N}(350,\ 105)$. El área roja a la derecha de 370 aproxima $\mathbb{P}(X > 370) \approx 0.026$. Así se calculan probabilidades de conteos grandes sin sumar muchas barras binomiales.

¿Cuándo se usa cada una en la vida real?

Una brújula rápida para reconocer la distribución a partir del problema:

- **Poisson** = *conteos de eventos raros* en un intervalo fijo: llamadas a un call center por hora, accidentes por mes, casos de una enfermedad rara por año. Pista: "¿cuántas veces ocurrió algo?" sin un máximo natural claro.
- **Exponencial** = *tiempos de espera* entre eventos: minutos hasta el próximo cliente,

días hasta la próxima falla de una máquina. Pista: "¿cuánto tiempo hasta que pase algo?".

- **Normal** = *promedios, sumas y errores de medición*: estaturas, puntajes, ruido de un sensor, y —por el TLC— casi cualquier *promedio* de muchos datos. Pista: una magnitud simétrica formada por muchas pequeñas causas que se suman.
- (Bonus) **Binomial** = *número de éxitos en n intentos* fijos (cuántos de 50 encuestados aprueban); **Geométrica** = *cuántos intentos hasta el primer éxito.*

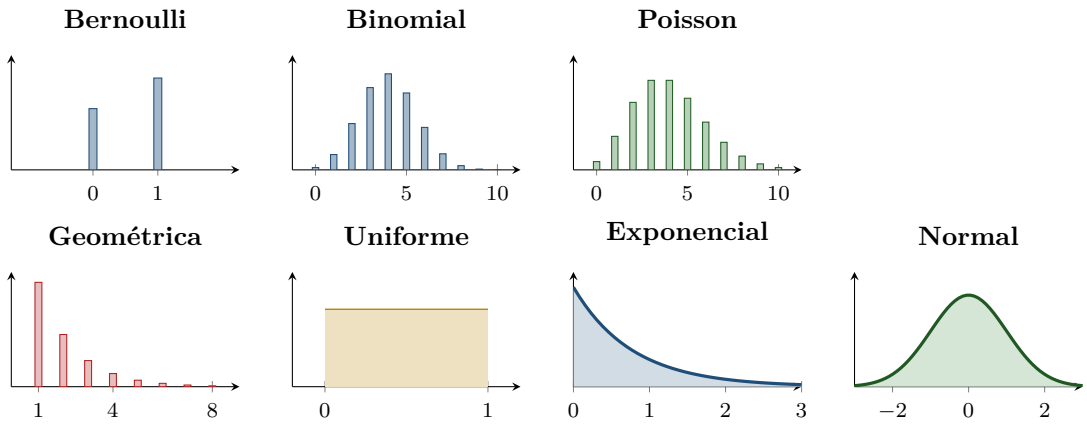

Figura 13: **Galería "¿qué distribución uso?"** **Bernoulli**: un sí/no. **Binomial**: # éxitos en n intentos. **Poisson**: conteo de eventos en un intervalo. **Geométrica**: # intentos hasta el primer éxito (decae). **Uniforme**: todo el rango igual de probable. **Exponencial**: tiempo de espera (decae). **Normal**: campana simétrica (sumas, promedios, errores).

4. Sumas de variables aleatorias independientes

Antes de promediar (la media muestral $\bar{X}$ es una suma escalada), conviene saber *qué familia* resulta al **sumar** variables independientes y cómo se comportan la esperanza y la varianza de una suma. Estas son exactamente las herramientas que usan los problemas de esta clase.

Propiedades reproductivas: sumas que "no cambian de familia"

Si $X_1, \ldots, X_k$ son **independientes**, la suma $S = \sum_i X_i$ a menudo pertenece a la misma familia (esto se llama *propiedad reproductiva* o cierre bajo la suma):

Sumandos independientes	**Condición**	**Suma** $S = \sum_i X_i$
$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$	—	$\mathcal{N}(\sum_i \mu_i,\ \sum_i \sigma_i^2)$
$X_i \sim \text{Poisson}(\lambda_i)$	—	$\text{Poisson}(\sum_i \lambda_i)$
$X_i \sim \text{Bernoulli}(p)$	mismo p	$\text{Binomial}(k, p)$
$X_i \sim \text{Binomial}(n_i, p)$	mismo p	$\text{Binomial}(\sum_i n_i,\ p)$
$X_i \sim \text{Geom}(p)$	mismo p	$\text{BinNeg}(k, p)$
$X_i \sim \text{Exp}(\lambda)$	i.i.d.	$\text{Gamma/Erlang}(k, \lambda)$
$X_i \sim \text{Gamma}(\alpha_i, \lambda)$	misma tasa λ	$\text{Gamma}(\sum_i \alpha_i,\ \lambda)$
$X_i \sim \chi^2(d_i)$	—	$\chi^2(\sum_i d_i)$

Ojo: la *Uniforme* **no** es reproductiva (la suma de dos uniformes ya es triangular, no uniforme); ese es justo el primer paso del TLC.

Suma "cruzada" (combinación lineal) y normalidad

Para constantes a, b, la **combinación lineal** $aX + bY$ es la versión con pesos de una suma. Toda combinación lineal de **normales independientes sigue siendo normal**:

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2),\ Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \text{ indep.} \ \Rightarrow\ aX + bY \sim \mathcal{N}\Big(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2\Big).$$

En particular, la *diferencia* $X - Y$ **suma** las varianzas: $\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2$ (no las resta).

Cómo se calcula $\mathbb{E}[XY]$ y la covarianza

La esperanza de un **producto** se calcula con la distribución conjunta:

$$\mathbb{E}[XY] = \sum_x \sum_y x\,y\,p(x,y) \quad \text{(discreto)}, \qquad \mathbb{E}[XY] = \iint x\,y\,f(x,y)\,dx\,dy \quad \text{(continuo)}.$$

La **covarianza** mide cuánto se mueven juntas X e Y:

$$\boxed{\text{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]}.$$

Si X e Y son **independientes**, la conjunta se factoriza $f(x,y) = f_X(x)f_Y(y)$ y entonces

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \quad \Longrightarrow \quad \text{Cov}(X,Y) = 0.$$

(La recíproca no siempre vale: covarianza 0 no implica independencia.)

Varianza de una suma: de dónde sale el término cruzado

Para dos variables cualesquiera,

$$\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)+2\ \text{Cov}(X,Y), \qquad \text{Var}(aX+bY) = a^2\,\text{Var}(X)+b^2\,\text{Var}(Y)+2ab\ \text{Cov}(X,Y)$$

Con **independencia** el término cruzado $2\,\mathrm{Cov}(X,Y)$ **desaparece**, y la varianza de la suma es la suma de varianzas. Esto es *exactamente* lo que permite escribir $\mathrm{Var}(\bar{X}) = \sigma^2/n$ en la próxima sección.

Ejemplo 4.1 (Suma de exponenciales → Gamma/Erlang)**.** Un trámite tiene $k = 4$ etapas independientes, cada una con duración $\mathrm{Exp}(\lambda)$ de media $1/\lambda = 15$ min (es decir $\lambda = 1/15$). El tiempo total es $S = X_1 + \cdots + X_4 \sim \mathrm{Gamma}(k{=}4,\ \lambda{=}1/15)$, con

$$\mathbb{E}[S] = \frac{k}{\lambda} = 4 \cdot 15 = 60 \text{ min}, \qquad \mathrm{Var}(S) = \frac{k}{\lambda^2} = 4 \cdot 15^2 = 900 \ (\mathrm{SD} = 30 \text{ min}).$$

La suma usa la propiedad reproductiva; la media y la varianza usan que las etapas son independientes (sin términos de covarianza).

5. La media muestral $\bar{X}$

Aquí empieza la *estadística*. Tomamos una **muestra aleatoria (random sample)**: variables $X_1, \ldots, X_n$ iid, cada una con media $\mu = \mathbb{E}[X_i]$ y varianza $\sigma^2 = \mathrm{Var}(X_i)$. El resumen más usado es su promedio.

Definición 5.1 (Media muestral)**.** La **media muestral (sample mean)** de una muestra de tamaño n es

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Como es función de variables aleatorias, $\bar{X}$ es *ella misma* una variable aleatoria, con su propia distribución (la *distribución muestral*).

Media y varianza de $\bar{X}$

Si $X_1, \ldots, X_n$ son iid con $\mathbb{E}[X_i] = \mu$ y $\mathrm{Var}(X_i) = \sigma^2$, entonces

$$\boxed{\mathbb{E}[\bar{X}] = \mu} \qquad \boxed{\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}}$$

La raíz de la varianza, $\mathrm{SD}(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$, se llama **error estándar (standard error)** de la media.

A continuación, la deducción (usa solo linealidad de $\mathbb{E}$ y propiedades de Var):

$$\mathbb{E}[\bar{X}] = \mathbb{E}\Big[\tfrac{1}{n}\textstyle\sum_i X_i\Big] = \tfrac{1}{n}\textstyle\sum_i \mathbb{E}[X_i] = \tfrac{1}{n}\,(n\mu) = \mu,$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\Big[\tfrac{1}{n}\textstyle\sum_i X_i\Big] = \tfrac{1}{n^2}\textstyle\sum_i \mathrm{Var}(X_i) = \tfrac{1}{n^2}\,(n\sigma^2) = \tfrac{\sigma^2}{n}.$$

El cálculo de $\mathbb{E}[\bar{X}] = \mu$ vale *siempre* (la esperanza de una suma es la suma de esperanzas, con o sin independencia). En cambio $\mathrm{Var}(\bar{X}) = \sigma^2/n$ usa que las X_i son independientes (para que la varianza de la suma sea la suma de varianzas, sin términos de covarianza).

Promediar *no* cambia el centro ($\mathbb{E}[\bar{X}] = \mu$): la media muestral apunta al lugar correcto. Pero *sí* reduce la dispersión: la varianza se divide entre n, y el error estándar cae como $\sigma/\sqrt{n}$. Por eso encuestas más grandes dan estimaciones más precisas... pero con **rendimientos decrecientes**: para reducir el error a la mitad hay que *cuadruplicar* el tamaño de muestra ($\sqrt{4} = 2$).

Ejemplo 5.1 (Ingreso promedio de una muestra)**.** El ingreso mensual de los hogares de una ciudad tiene $\mu = 2000$ soles y $\sigma = 600$ soles. Tomamos una muestra aleatoria de $n = 100$ hogares y calculamos el promedio $\bar{X}$. Entonces

$$\mathbb{E}[\bar{X}] = 2000, \qquad \mathrm{Var}(\bar{X}) = \frac{600^2}{100} = 3600, \qquad \mathrm{SE}(\bar{X}) = \frac{600}{\sqrt{100}} = 60 \text{ soles}.$$

Aunque los hogares individuales varían mucho ($\sigma = 600$), el *promedio* de 100 de ellos varía poco (error estándar de solo 60 soles).

La figura 14 muestra la distribución de $\bar{X}$ para distintos tamaños de muestra n (con $\mu = 2000$, $\sigma = 600$, como en el ejemplo). Todas están *centradas en el mismo* μ, pero al crecer n la campana se vuelve cada vez más *alta y angosta*: el error estándar $\mathrm{SE} = \sigma/\sqrt{n}$ se encoge ($120 \to 85 \to 60 \to 30$ soles). Esa es la imagen exacta de "más datos $\Rightarrow$ estimación más precisa". Ojo al $\sqrt{n}$: pasar de $n = 25$ a $n = 100$ ($4\times$) reduce el SE a la mitad, no a la cuarta parte.

6. Teorema del Límite Central (TLC)

Sabemos el *centro* y la *dispersión* de $\bar{X}$. ¿Pero qué *forma* tiene su distribución? La respuesta es asombrosa: para n grande, casi siempre es normal.

Teorema 6.1 (Teorema del Límite Central (Central Limit Theorem))**.** *Sean $X_1, \ldots, X_n$ una muestra aleatoria (iid) de* cualquier *distribución con media μ y varianza σ^2* finitas*. Entonces, al crecer n, la versión estandarizada de la media muestral converge a una normal estándar:*

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{} \mathcal{N}(0,1), \qquad \textit{es decir} \qquad \mathbb{P}(Z_n \le x) \to \Phi(x) \quad \forall x.$$

En la práctica, para n razonablemente grande,

$$\boxed{\bar{X} \;\approx\; \mathcal{N}\Big(\mu,\ \tfrac{\sigma^2}{n}\Big)}$$

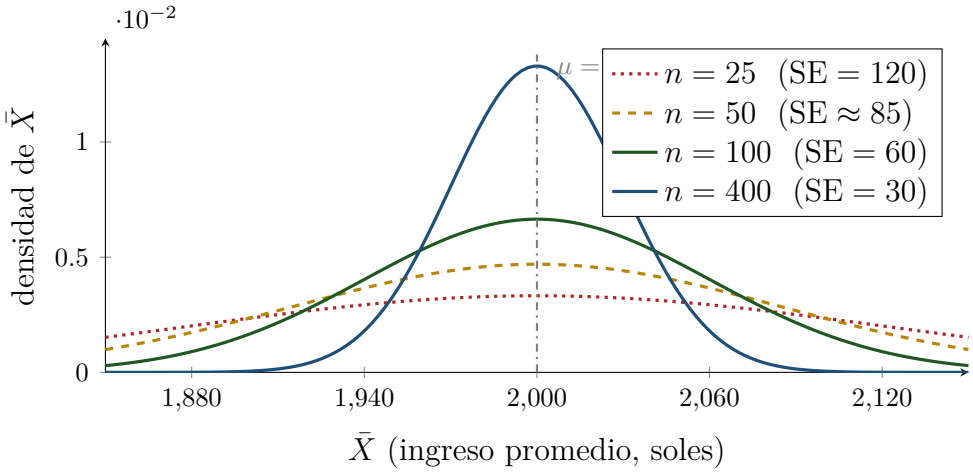

Figura 14: Distribución de $\bar{X}$ (ingresos, $\mu = 2000$, $\sigma = 600$) al crecer n. Todas comparten el centro μ, pero el error estándar $\text{SE} = \sigma/\sqrt{n}$ encoge la campana: más muestra $\Rightarrow$ promedio más concentrado. Como crece $\sqrt{n}$, los retornos son decrecientes.

Lo notable: **no importa de qué distribución salen los datos** —puede ser uniforme, exponencial, sesgada, casi degenerada—. Mientras tengan media y varianza finitas, el *promedio* de muchas de ellas se vuelve normal. Por eso la campana de Gauss aparece por todos lados: tantos fenómenos del mundo son, en el fondo, *promedios* o *sumas* de muchos efectos pequeños. El TLC es la razón última de por qué la normal es tan central en estadística.

La media muestral $\bar{X}$ *suma* muchas contribuciones pequeñas e independientes y luego divide entre n. Cada dato trae su propia "idiosincrasia": unos salen por encima de μ, otros por debajo. Al promediar, esas desviaciones **se cancelan** entre sí —no exactamente, pero sí en su mayor parte—. Lo que *no* se cancela es el efecto neto y acumulado de muchos pequeños empujones aleatorios, y resulta que esa acumulación *siempre* toma la misma forma: una **campana**. La clave es que ningún dato manda sobre el resto (todos pesan $1/n$, y ese peso $\to 0$); por eso la forma original de los datos —plana, sesgada, con dos jorobas— se *borra* y emerge la normal. Es un fenómeno de "promediado": los detalles de cada pieza dejan de importar y solo sobrevive su tendencia central más un ruido simétrico en forma de campana.

Morphing: cómo cambia la forma de $\bar{X}$ al crecer n (el dado)

Imagina promediar tiradas de un **dado** (uniforme discreta, "plana", de 1 a 6):

- $n = 1$: $\bar{X}$ *es* un dado $\Rightarrow$ distribución **plana** (cada cara igual de probable). Nada de campana todavía.

- $n = 2$: el promedio de dos dados ya forma un **triángulo** (el 3.5 central es el más probable; los extremos 1 y 6 casi no salen).
- $n = 5$: la distribución ya se **acampana** visiblemente.
- $n = 30$: es **casi indistinguible** de una Normal.

Misma idea en la figura 16: al crecer n, $\bar{X}$ no solo se concentra, sino que *cambia de forma* hasta volverse normal. El TLC dice que esto pasa *sea cual sea* la distribución de partida (ver la figura 15 con un padre muy asimétrico).

Recursos en video: ver el TLC en movimiento

Dos animaciones muy recomendadas para *visualizar* cómo, al sumar o promediar muchas variables, emerge la campana normal:

- ***But what is the Central Limit Theorem?*** — 3Blue1Brown: youtu.be/zeJD6dqJ5lo. La animación del *morphing* de varias distribuciones hacia $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.
- ***Bunnies, Dragons and the "Normal" World*** — The New York Times: youtube.com/watch?v=jvoxEYmQHNM. Intuición visual: muchos efectos pequeños e independientes se suman en una campana.

El "ancho" típico de $\bar{X}$ es el error estándar $\text{SE} = \sigma/\sqrt{n}$. Como aparece $\sqrt{n}$ y no n, la precisión mejora *lentamente*: para **duplicar** la precisión (reducir SE a la mitad) hace falta **4**× datos, porque $\sqrt{4} = 2$. Para 10× más precisión, 100× datos. Esta es la "ley de los grandes presupuestos" de las encuestas: pasar de 1000 a 4000 encuestados solo recorta el error a la mitad.

- **El TLC NO vuelve normales los datos.** Vuelve normal la distribución del *promedio* $\bar{X}$ (o de la *suma* S_n). Si tu población es sesgada, los *datos individuales* siguen siendo igual de sesgados aunque tomes n enorme; lo que se normaliza es $\bar{X}$.
- **No confundas σ con SE.** σ mide la dispersión de *un dato*; $\text{SE} = \sigma/\sqrt{n}$ mide la dispersión del *promedio*. Son números muy distintos: SE es mucho más pequeño (y se encoge con n, mientras σ no cambia).
- **"n grande" depende de la asimetría.** La regla práctica $n \approx 30$ vale para poblaciones razonablemente simétricas. Con mucha asimetría o colas pesadas se necesita un n mayor. Y si la **varianza es infinita** (colas extremas, tipo ciertas Pareto/Cauchy), el TLC *no* aplica: la hipótesis de varianza finita es indispensable.

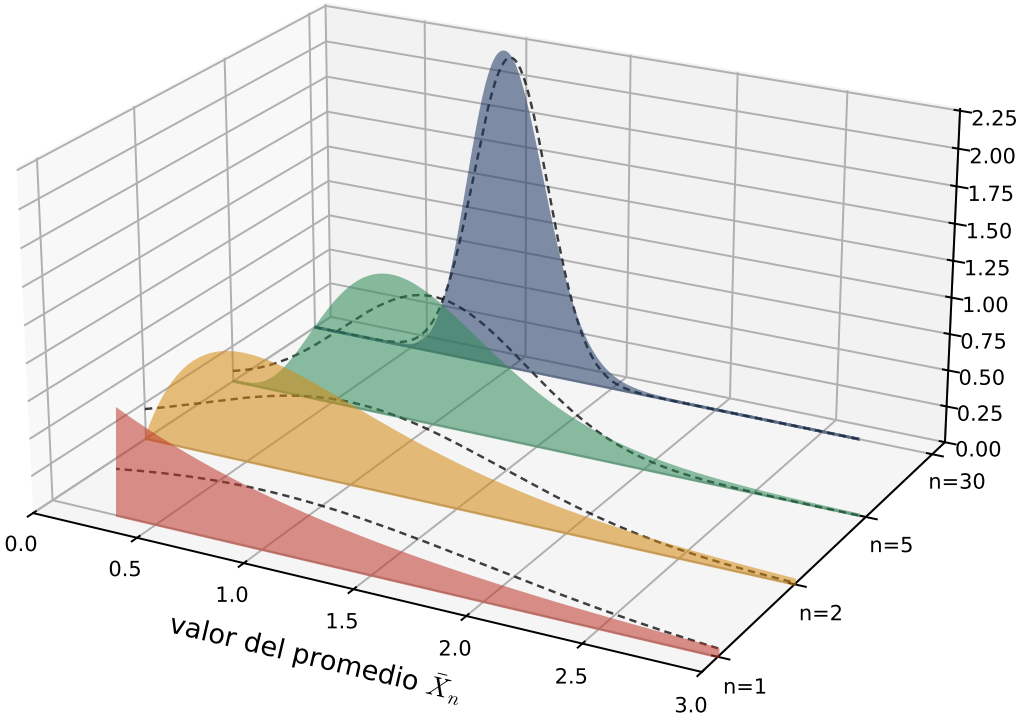

Figura 15: El TLC en 3D, partiendo de una distribución **muy asimétrica** (exponencial, media 1). Cada "loncha" es la densidad del promedio $\bar{X}_n$ para un tamaño de muestra n. Con $n = 1$ la forma es la exponencial original (sesgada); al crecer n, $\bar{X}_n$ se vuelve **más angosta y más simétrica**, hasta calzar con la campana Normal $N(1,\, 1/n)$ (línea punteada) en $n = 30$. La forma original "se borra": es la esencia del Teorema del Límite Central.

Cómo usar el TLC para calcular probabilidades

Para aproximar una probabilidad sobre $\bar{X}$ (o sobre la suma $S_n = \sum X_i$), se *estandariza* y se usa la tabla normal:

$$\mathbb{P}(\bar{X} \le a) \approx \Phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right), \qquad \mathbb{P}(a \le \bar{X} \le b) \approx \Phi\left(\tfrac{b-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\tfrac{a-\mu}{\sigma/\sqrt{n}}\right).$$

Ejemplo 6.1 (Aproximación normal de un promedio)**.** Retomemos los ingresos con $\mu = 2000$, $\sigma = 600$ y muestra $n = 100$, de modo que $\bar{X} \approx \mathcal{N}(2000,\, 60^2)$. ¿Cuál es la probabilidad de que el promedio muestral supere los 2100 soles?

$$Z = \frac{2100 - 2000}{600/\sqrt{100}} = \frac{100}{60} \approx 1.67, \qquad \mathbb{P}(\bar{X} > 2100) \approx \mathbb{P}(Z > 1.67) = 1 - \Phi(1.67) \approx 0.048.$$

Hay solo un 4.8 % de probabilidad de observar un promedio tan alto, ¡aunque la población de ingresos sea muy asimétrica! Eso es el TLC en acción.

Ejemplo 6.2 (Un $\mathbb{P}(\bar{X} > c)$ paso a paso, con interpretación)**.** El tiempo de atención por cliente en una ventanilla tiene media $\mu = 4$ min y desviación $\sigma = 3$ min (distribución muy

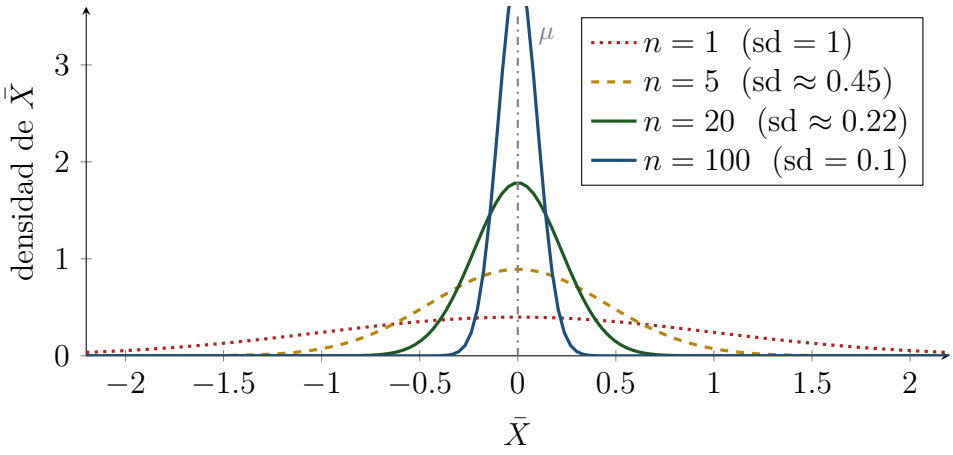

Figura 16: Distribución de $\bar{X}$ (con $\mu = 0$, $\sigma = 1$) para n creciente. Todas están centradas en μ, pero se vuelven *más altas y angostas* (más concentradas) al crecer n, porque $\mathrm{Var}(\bar{X}) = \sigma^2/n \to 0$. El TLC garantiza, además, que la forma tiende a la campana normal sin importar la distribución original.

sesgada: algunos trámites se eternizan). Atendemos $n = 36$ clientes. ¿Probabilidad de que el *tiempo promedio* por cliente supere los 5 min?

1. **Identifica la distribución de $\bar{X}$.** Por el TLC, con $n = 36$,

$$\bar{X} \approx \mathcal{N}\Big(\mu,\ \tfrac{\sigma^2}{n}\Big) = \mathcal{N}\Big(4,\ \tfrac{9}{36}\Big) = \mathcal{N}(4,\ 0.25), \qquad \mathrm{SE} = \tfrac{\sigma}{\sqrt{n}} = \tfrac{3}{6} = 0.5 \text{ min}.$$

2. **Estandariza usando** SE **(no σ).**

$$Z = \frac{\bar{X} - \mu}{\mathrm{SE}} = \frac{5 - 4}{0.5} = 2.0.$$

3. **Lee la tabla normal.** $\mathbb{P}(\bar{X} > 5) \approx \mathbb{P}(Z > 2) = 1 - \Phi(2) \approx 1 - 0.977 = 0.023$.

Interpretación en palabras: hay solo $\approx 2.3\,\%$ de probabilidad de que, en un día de 36 clientes, el tiempo *medio* por cliente pase de 5 min. Ojo: *un* cliente individual con 9 min es perfectamente normal (la población es sesgada), pero que el *promedio* de 36 se desvíe 1 min completo por encima de μ es raro, porque el promedio varía poco (SE $= 0.5$, seis veces menor que $\sigma = 3$).

Ejemplo 6.3 (Aproximación Normal a una proporción (una encuesta))**.** En una encuesta preguntamos a $n = 400$ personas si aprueban una medida; la proporción verdadera es $p = 0.55$. Sea $\hat{p}$ la proporción muestral que aprueba (¡un *promedio* de Bernoulli!: $\hat{p} = \bar{X}$ con datos 0/1). Como cada dato es Bernoulli(p) con $\mu = p$ y $\sigma^2 = p(1-p)$, el TLC da

$$\hat{p} \ \approx\ \mathcal{N}\Big(p,\ \tfrac{p(1-p)}{n}\Big), \qquad \mathrm{SE}(\hat{p}) = \sqrt{\tfrac{p(1-p)}{n}} = \sqrt{\tfrac{0.55 \cdot 0.45}{400}} \approx 0.0249.$$

¿Probabilidad de que la encuesta muestre *menos* del $50\,\%$ de aprobación (un "falso empate")?

$$Z = \frac{0.50 - 0.55}{0.0249} \approx -2.01, \qquad \mathbb{P}(\hat{p} < 0.50) \approx \Phi(-2.01) \approx 0.022.$$

Interpretación: aunque la verdad sea 55 %, hay $\approx 2.2\,\%$ de probabilidad de que *esta* encuesta caiga por debajo del 50 % solo por azar muestral. Así se construyen los *márgenes de error* de las encuestas: con $n = 400$, el margen ($\approx 2 \times \text{SE}$) ronda los ± 5 puntos.

El TLC es asintótico: la aproximación mejora con n. ¿Cuán grande basta? Si la distribución original ya es bastante simétrica, $n \approx 30$ suele alcanzar. Si es muy sesgada (ingresos, tiempos de espera), puede hacer falta un n mayor. Y el TLC *exige* varianza finita: distribuciones de colas muy pesadas (sin varianza, como ciertas Pareto) no lo cumplen.

7. Introducción a la estimación

Por primera vez en el curso pasamos de la probabilidad a la **estadística**: el estudio de la *estimación* y la *inferencia*. Aquí vemos la idea base de la estimación, que la Clase 6 desarrollará.

Definición 7.1 (Parámetro, estimador y estimación)**.** Un **parámetro (parameter)** es una constante (desconocida) que indexa una familia de distribuciones: μ y σ^2 de la normal, λ de la Poisson, p de la Binomial, a, b de la uniforme. Se suele denotar genéricamente θ.

Un **estimador (estimator)** $\hat{\theta}$ es una *función de la muestra aleatoria* $X_1, \ldots, X_n$ (por tanto, una variable aleatoria). Al evaluarlo en los datos concretos obtenemos un número, la **estimación (estimate)**.

El parámetro es la verdad *fija pero desconocida* de la población; el estimador es nuestra *receta* para adivinarla a partir de la muestra; la estimación es el *número* que sale al aplicar la receta a los datos. La media muestral $\bar{X}$ es el estimador natural de la media poblacional μ: una regla que toma los datos y devuelve un número cercano (esperamos) a μ.

$\bar{X}$ como estimador de μ y su distribución muestral

La media muestral $\bar{X} = \hat{\mu}$ es un estimador de μ. Como es aleatoria, tiene una **distribución muestral (sampling distribution)**: distintas muestras dan distintos $\bar{X}$. Ya sabemos sus dos primeros momentos: $\mathbb{E}[\bar{X}] = \mu$ y $\text{Var}(\bar{X}) = \sigma^2/n$, y por el TLC su forma es aproximadamente normal para n grande.

Definición 7.2 (Sesgo y estimador insesgado)**.** El **sesgo (bias)** de un estimador $\hat{\theta}$ de θ es $\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$. Decimos que $\hat{\theta}$ es **insesgado (unbiased)** si $\mathbb{E}[\hat{\theta}] = \theta$ para todo valor posible del parámetro: en promedio, da en el blanco.

Ejemplo 7.1 ($\bar{X}$ es insesgado para μ)**.** Como $\mathbb{E}[\bar{X}] = \mu$, la media muestral es un estimador **insesgado** de la media poblacional: $\text{Bias}(\bar{X}) = \mathbb{E}[\bar{X}] - \mu = \mu - \mu = 0$. No quiere decir que

una muestra acierte exactamente, sino que el estimador no se inclina sistemáticamente ni hacia arriba ni hacia abajo.

Ejemplo 7.2 (Estimar el máximo de una uniforme)**.** Suponga $X \sim \text{Unif}[0, \theta]$ con θ desconocido (p.ej. el monto máximo de un beneficio que no conocemos). Dos recetas razonables:

- $\hat{\theta}_1 = 2\bar{X}$ (el doble del promedio: como $\mathbb{E}[X] = \theta/2$, se tiene $\mathbb{E}[2\bar{X}] = \theta$, *insesgado*).
- $\hat{\theta}_2 = \text{máx}\{X_1, \ldots, X_n\}$ (el máximo muestral: mejora al crecer n, pero tiende a *subestimar* θ, es decir es *sesgado* hacia abajo).

La Clase 6 dará criterios (sesgo, varianza, error cuadrático medio) para elegir entre estimadores como estos.

Imagina que cada muestra es un *dardo* lanzado a una diana cuyo centro es el verdadero θ. Un buen estimador combina dos virtudes:

- **Poco sesgo (apuntar bien):** en promedio, los dardos caen *centrados* en el blanco. Sesgo = apuntar sistemáticamente desviado (siempre a la izquierda, por ejemplo). Formalmente, Bias $= \mathbb{E}[\hat{\theta}] - \theta$.
- **Poca varianza (pulso firme):** los dardos caen *juntos* entre sí, no dispersos por toda la diana. Mucha varianza = pulso tembloroso.

Las cuatro combinaciones: *insesgado + poca varianza* (dardos agrupados en el centro: el ideal); *insesgado + mucha varianza* (centrados pero dispersos); *sesgado + poca varianza* (agrupados pero en la zona equivocada); *sesgado + mucha varianza* (lo peor). El criterio que junta ambas cosas en un solo número es el *error cuadrático medio*: MSE = sesgo^2 + varianza, es decir, "qué tan lejos del centro caen los dardos en promedio".

Nota. Insesgado no siempre es "mejor": a veces conviene un estimador con un pequeño sesgo pero mucha menos varianza. El criterio que combina ambos —el *error cuadrático medio* (MSE = sesgo^2 + varianza)— y los métodos para construir estimadores (máxima verosimilitud, momentos) son el tema central de la próxima clase.

8. En R: muestrear de las distribuciones y verlas

R tiene una familia de funciones por distribución: `r*` (muestrear), `d*` (pmf/pdf), `p*` (cdf) y `q*` (cuantiles). Aquí muestreamos de cada distribución y comparamos el histograma simulado con la pmf/pdf *teórica*; deberían coincidir cada vez mejor cuanto más grande sea la muestra.

Discretas: histograma simulado vs pmf teorica

```
set.seed(2026)

# --- Binomial(10, 0.4): contar exitos en 10 ensayos ---
x  <- rbinom(100000, size = 10, prob = 0.4)   # muestra grande
tab <- table(x) / length(x)                   # frecuencias relativas (~ pmf)
```

```
k  <- 0:10
plot(k, dbinom(k, 10, 0.4), type = "h", lwd = 3, col = "blue",
     ylab = "P(X=k)", main = "Binomial(10, 0.4): teorica vs simulada")
points(as.numeric(names(tab)), as.numeric(tab), pch = 19, col = "red")
legend("topright", c("teorica dbinom", "simulada"),
       col = c("blue","red"), pch = c(NA,19), lwd = c(3,NA))

# --- Poisson(3): conteo de eventos raros ---
y <- rpois(100000, lambda = 3)
barplot(table(y)/length(y), col = "lightgreen",
        main = "Poisson(3): proporciones simuladas")
points(0:12 + 0.7, dpois(0:12, 3), pch = 19, col = "red")  # pmf teorica
```

Continuas: histograma simulado vs pdf teorica

```
set.seed(2026)

# --- Exponencial(rate = 1.5): tiempos de espera ---
te <- rexp(100000, rate = 1.5)
hist(te, breaks = 60, freq = FALSE, col = "lightblue",
     main = "Exponencial(1.5): histograma vs pdf", xlab = "x")
curve(dexp(x, rate = 1.5), add = TRUE, lwd = 2, col = "red")   # pdf teorica

# --- Normal(mu = 70, sd = 8): notas de un examen ---
no <- rnorm(100000, mean = 70, sd = 8)
hist(no, breaks = 60, freq = FALSE, col = "lightyellow",
     main = "Normal(70, 8^2): histograma vs pdf", xlab = "nota")
curve(dnorm(x, mean = 70, sd = 8), add = TRUE, lwd = 2, col = "red")
```

Las barras/histogramas simulados se *superponen* casi perfectamente a las curvas teóricas: muestrear muchas veces "revela" la distribución que está detrás. Cambia el tamaño de la muestra (de 100 a 100000) para *ver* cómo la simulación se acerca a la teoría.

Estimar E[X

```
 y Var por simulacion y comparar con la teoria]
set.seed(2026)
N <- 200000

# Poisson(3): teorico E[X] = Var(X) = lambda = 3
y <- rpois(N, lambda = 3)
c(media_sim = mean(y),  media_teo = 3)
c(var_sim   = var(y),   var_teo   = 3)

# Exponencial(rate = 1.5): E[X] = 1/lambda, Var = 1/lambda^2
```

```
e <- rexp(N, rate = 1.5)
c(media_sim = mean(e),  media_teo = 1/1.5)
c(var_sim   = var(e),   var_teo   = 1/1.5^2)

# Binomial(10, 0.4): E[X] = np = 4, Var = np(1-p) = 2.4
b <- rbinom(N, size = 10, prob = 0.4)
c(media_sim = mean(b),  media_teo = 4)
c(var_sim   = var(b),   var_teo   = 2.4)
```

Con N grande, `mean()` y `var()` de la muestra se acercan a los valores teóricos $\mathbb{E}[X]$ y $\mathrm{Var}(X)$: la *ley de los grandes números* en acción. Es una forma excelente de *verificar* a mano cualquier fórmula de esperanza o varianza de la clase.

9. En R: simular el Teorema del Límite Central

El TLC se "ve" muy bien por simulación: tomamos muchas muestras de una distribución *nada normal* (aquí una exponencial, muy sesgada), calculamos la media de cada una, y graficamos el histograma de esas medias. Debería parecerse a una campana.

Simulacion del TLC: medias de muestras exponenciales

```
set.seed(2026)
lambda <- 1            # Exp(1): media 1, sd 1, MUY sesgada
n      <- 30           # tamano de cada muestra
B      <- 10000        # numero de muestras (repeticiones)

# Para cada repeticion: sacar n datos Exp(1) y promediarlos
medias <- replicate(B, mean(rexp(n, rate = lambda)))

mean(medias)           # ~ 1   (= mu, la media de Exp(1))
sd(medias)             # ~ 1/sqrt(n) = 0.1826  (error estandar teorico)

# El histograma de las medias se ve aproximadamente NORMAL,
# aunque la Exp(1) original es fuertemente asimetrica:
hist(medias, breaks = 40, freq = FALSE,
     main = "TLC: medias de muestras Exp(1)",
     xlab = expression(bar(X)))
curve(dnorm(x, mean = 1, sd = 1/sqrt(n)),
      add = TRUE, lwd = 2, col = "blue")   # normal teorica superpuesta

# Comparacion: una sola observacion Exp(1) NO es normal
hist(rexp(B, rate = lambda), breaks = 40, freq = FALSE,
     main = "Una sola Exp(1): sesgada", xlab = "x")
```

La media de las 10 000 medias se acerca a $\mu = 1$ y su desviación, a $\sigma/\sqrt{n} = 1/\sqrt{30} \approx 0.183$.

El histograma de las medias luce acampanado (normal), *aunque* cada muestra venga de una exponencial muy sesgada: ese es exactamente el mensaje del TLC. Prueba a cambiar `rexp` por `runif` o a variar `n` para ver cómo mejora la aproximación.

El siguiente bloque hace lo más ilustrativo: toma **tres distribuciones madre muy distintas** (uniforme plana, exponencial sesgada y Bernoulli con dos picos) y, para cada una, muestra cómo el histograma de $\bar{X}$ se acerca a la campana al crecer n.

TLC con distribuciones madre distintas y n creciente

```
set.seed(2026)
B <- 20000                        # numero de muestras por escenario

# funcion: histograma de las medias de B muestras de tamano n
tlc <- function(rgen, n, mu, sigma, titulo) {
  medias <- replicate(B, mean(rgen(n)))
  hist(medias, breaks = 40, freq = FALSE, col = "lightgray", border = "white",
       main = paste0(titulo, "  (n = ", n, ")"), xlab = expression(bar(X)))
  curve(dnorm(x, mean = mu, sd = sigma/sqrt(n)),  # normal teorica del TLC
        add = TRUE, lwd = 2, col = "blue")
}

par(mfrow = c(3, 3))              # 3 distribuciones x 3 tamanos n

# 1) Uniforme(0,1): mu = 0.5, sigma = 1/sqrt(12)
for (n in c(1, 5, 30))
  tlc(function(n) runif(n), n, 0.5, 1/sqrt(12), "Uniforme(0,1)")

# 2) Exponencial(1): mu = 1, sigma = 1 (MUY sesgada)
for (n in c(1, 5, 30))
  tlc(function(n) rexp(n, 1), n, 1, 1, "Exp(1)")

# 3) Bernoulli(0.3): mu = 0.3, sigma = sqrt(0.3*0.7) (dos picos)
for (n in c(1, 5, 30))
  tlc(function(n) rbinom(n, 1, 0.3), n, 0.3, sqrt(0.3*0.7), "Bernoulli(0.3)")

par(mfrow = c(1, 1))
```

La primera columna ($n = 1$) muestra la *forma original* de cada distribución madre (plana, sesgada, dos picos). A medida que n crece (5, 30), el histograma de $\bar{X}$ pierde esa forma y se ajusta a la curva normal azul: **el TLC funciona sin importar de dónde salgan los datos**. La exponencial (muy sesgada) tarda un poco más en "acampanarse" que la uniforme.

Resumen de la clase

Lo esencial de la Clase 5

- **Discretas:** Bernoulli(p) [$\mathbb{E} = p$, Var $= pq$]; Binomial(n, p) [np, npq]; Geométrica(p) [$1/p$, $(1-p)/p^2$]; Bin. negativa(r, p) [r/p, $r(1-p)/p^2$]; Hipergeom. [np, con corrección finita]; Poisson(λ) [λ, λ].
- **Binomial $\to$ Poisson:** eventos raros, $n \to \infty$, $p \to 0$, $\lambda = np$.
- **Continuas:** Uniforme$[a, b]$ [$\frac{a+b}{2}$, $\frac{(b-a)^2}{12}$]; Exponencial(λ) [$1/\lambda$, $1/\lambda^2$, *sin memoria*]; Normal(μ, σ^2) [μ, σ^2; estandarizar con $Z = \frac{X-\mu}{\sigma}$].
- **Media muestral:** $\bar{X} = \frac{1}{n}\sum X_i$, con $\mathbb{E}[\bar{X}] = \mu$ y $\text{Var}(\bar{X}) = \sigma^2/n$; error estándar $\sigma/\sqrt{n}$.
- **TLC:** para n grande, $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$ *sea cual sea* la distribución original; estandarizar $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.
- **Estimación:** parámetro θ (fijo, desconocido) vs. estimador $\hat{\theta}$ (función de la muestra, aleatorio). $\bar{X}$ es estimador *insesgado* de μ ($\mathbb{E}[\bar{X}] = \mu$). Sesgo $= \mathbb{E}[\hat{\theta}] - \theta$.

Fuentes y atribución

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.ª ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.ª ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: `gevorg-hunanyan/probability`; `hanmochen/Probability-Theory-2020`; `mynameisjanus/14310xDataAnalysis`.