Estadística y Optimización

Modelos de variables aleatorias discretas

Variable aleatoria uniforme discreta: $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$

- **Modelos equiprobables discretos**: el resultado de lanzar una moneda equilibrada, un dado equilibrado, ...

- Para un conjunto de n elementos $\{x_1, \ldots, x_n\}$ con misma probabilidad:

$$\boxed{P(X = x_k) = \frac{1}{n}, \quad k = 1, \ldots, n}$$

- **Esperanza y varianza:**

$$\boxed{E[X] = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}, \qquad V[X] = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \overline{x}^2}$$

Variable aleatoria de Bernoulli: $X \sim \mathcal{B}er(p)$

- Si tenemos un experimento y un **suceso "éxito"** E con

$$\boxed{P(E) = p \in [0,1] \quad \text{y} \quad P\left(\overline{E}\right) = q = 1 - p}$$

Notación: $q = 1 - p$.

- **Experimentos binarios**: Si/No, Éxito/fracaso,..
  - Lanzar una moneda y $E = \{$"salga cara"$\}$.
  - Lanzar dos dados y $E = \{$"suma más de 10"$\}$.
  - Escoge una persona al azar y $E = \{$"tiene una cierta enfermedad"$\}$.

- **Función de masa de probabilidad:** $X \sim \mathcal{B}er(p)$ toma valores

$$X = \begin{cases} 1, & \text{si } E \\ 0, & \text{si } \overline{E}. \end{cases}$$

x_i	0	1
$P(X = x_i)$	q	p

- **Esperanza y varianza:**

$$\boxed{E[X] = p} \qquad \boxed{V[X] = pq.}$$

- **Función generadora de momentos y función característica:**

$$\boxed{M(t) = q + pe^t} \qquad \boxed{\varphi(t) = q + pe^{it}}$$

Variable aleatoria binomial: $X \sim \mathcal{B}(n,p)$

- Contar $X =$ núm. de éxitos en n pruebas de Bernoulli independientes de mismo parámetro p:

$$X = X_1 + \cdots + X_n \quad \text{con} \quad X_i = \begin{cases} 1, & \text{con prob. } p \\ 0, & \text{con prob. } q = 1-p. \end{cases}$$

- **Ejemplo:** Tiramos un dado 5 veces, ¿cuál es la probabilidad de obtener 2 veces un 3?
  - $E = \{\text{“obtener un 3”}\}$ con $P(E) = 1/6$, por lo que $X \sim \mathcal{B}(5, 1/6)$.
  - Tenemos que contar las combinaciones de 2 éxitos entre 5 lanzadas:

$$(E,E,\overline{E},\overline{E},\overline{E}),\ (E,\overline{E},E,\overline{E},\overline{E}),\ \ldots,\ (\overline{E},\overline{E},\overline{E},E,E)$$

$$\begin{aligned} P(X=2) &= (\text{núm. de formas de obtener 2 éxitos sobre 5 intentos}) \times P(E)^2 \times P\left(\overline{E}\right)^3 \\ &= \binom{5}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^3 = 0.16. \end{aligned}$$

- **Función de masa de probabilidad:** $X \sim \mathcal{B}(n,p)$ toma valores en $k = 0, \ldots, n$ con:

$$\boxed{P(X=k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, \ldots, n}$$

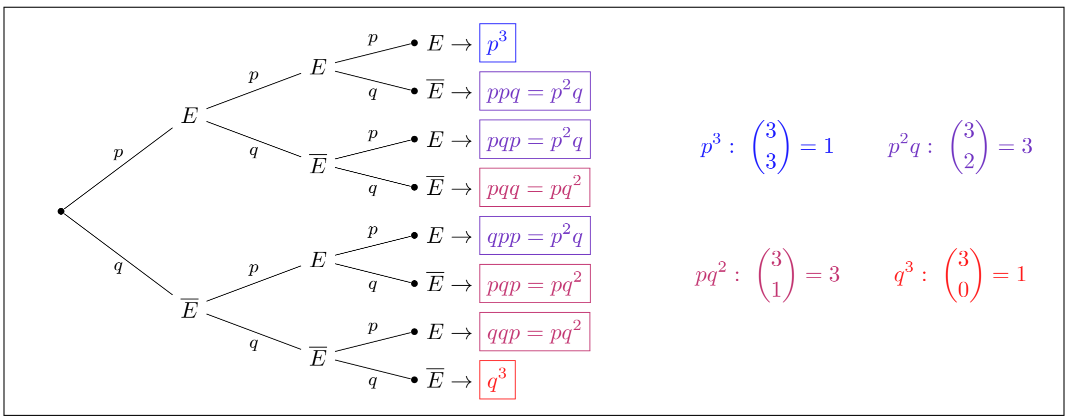


Figura 1: Las distintas combinaciones de k éxitos en una $\mathcal{B}(3,p)$.

- **Esperanza y varianza:** Estamos sumando n variables independientes $X_i \sim \mathcal{B}er(p)$:

$$\boxed{E[X] = np} \qquad \boxed{V[X] = npq.}$$

- **Función generadora de momentos y función característica:** Por la misma razón,

$$\boxed{M(t) = (q + pe^t)^n} \qquad \boxed{\varphi(t) = (q + pe^{it})^n}$$

- **Propiedad de la suma:** Si $X \sim \mathcal{B}(n,p)$ y $Y \sim \mathcal{B}(m,p)$ son independientes, entonces

$$X + Y \sim \mathcal{B}(n+m, p).$$

- **Problema:** Para valores de n grandes y/o p muy pequeños, las probabilidades $P(X=k)$ son complicadas de calcular en la práctica.

Variable aleatoria geométrica: $X \sim \mathcal{G}(p)$

- Contar $X =$ núm. de pruebas de Bernoulli $X_i \sim \mathcal{B}er(p)$ independientes hasta obtener el primer éxito E:

$$\underbrace{\overline{E}, \overline{E}, \cdots, \overline{E}}_{k-1 \text{ fracasos}}, E \longleftarrow \text{éxito en la } k\text{-ésima prueba}$$

- **Función de masa de probabilidad:** $X \sim \mathcal{G}(p)$ toma valores $k = 1, 2, 3, \ldots$ con probabilidad:

$$\boxed{P(X = k) = q^{k-1} p, \qquad k = 1, 2, 3, \ldots}$$

- **Función de distribución:**

$$F(k) = P(X \leq k) = 1 - \underbrace{P(X > k)}_{\text{prob. fracasar } k \text{ veces}} = 1 - q^k.$$

- **Función generadora de momentos y función característica:** Calculando $M(t) = E[e^{tX}]$ y $\varphi(t) = E[e^{itX}]$:

$$\boxed{M(t) = \frac{p}{e^{-t} - q}} \qquad \boxed{\varphi(t) = \frac{p}{e^{-it} - q}}$$

Nota: $M(t)$ está definida para $|qe^t| < 1$, o equivalentemente $t < \log(1/q)$.

- **Esperanza y varianza:** Basta hallar las derivadas $M'(0)$ y $M''(0)$:

$$\boxed{E[X] = \frac{1}{p}} \qquad \boxed{V[X] = \frac{q}{p^2}.}$$

- **Ejemplo:** Vistos en clase

a) $X =$ núm. de lanzamientos de una moneda hasta que sale cara, $X \sim \mathcal{G}(1/2)$:

$$P(X = k) = \frac{1}{2^{k-1}} \times \frac{1}{2} = \frac{1}{2^k}.$$

$$E[X] = \frac{1}{1/2} = 2, \quad V[X] = \frac{1 - 1/2}{(1/2)^2} = 2, \quad M(t) = \frac{1/2}{e^{-t} - 1/2} = \frac{1}{2e^{-t} - 1}, \quad \varphi(t) = \frac{1}{2e^{-it} - 1}.$$

b) $Y =$ núm. de lanzamientos de un dado hasta que sale un 5, $Y \sim \mathcal{G}(1/6)$:

$$P(Y = k) = \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6} = \frac{5^{k-1}}{6^k}.$$

$$E[Y] = \frac{1}{1/6} = 6, \quad V[Y] = \frac{5/6}{(1/6)^2} = 30, \quad M(t) = \frac{1/6}{e^{-t} - 5/6} = \frac{1}{6e^{-t} - 5}, \quad \varphi(t) = \frac{1}{6e^{-it} - 5}.$$

- **Propiedad de "Falta de memoria":** Si $X \sim \mathcal{G}(p)$, para todo $m, k \in \mathbb{N}$ se tiene:

$$P(X = m + k \mid X > m) = P(X = k).$$

Interpretación: no importa que hayamos observado m fracasos anteriormente...¡la probabilidad de obtener un éxito en la siguiente k-ésima vez es la misma!

Variable aleatoria de Poisson: $X \sim \mathcal{P}(\lambda)$

- Contar X = núm. de veces que ocurre un suceso *en un intervalo* (de tiempo/espacio/unidad de observación), asumiendo:

  a) los sucesos se producen de forma independiente (no "tiene memoria").

  b) este suceso ocurre en promedio $\lambda > 0$ veces por unidad de observación (por hora, por km,..), siendo este promedio constante.

- **Ejemplos:**

  - Coches que pasan a través de cierto punto durante una hora.
  - Errores ortográficos en una página.
  - Número de llamadas telefónicas en una central telefónica por minuto.
  - Número de servidores web accedidos por minuto.
  - Aparición de animales por unidad de longitud de ruta.
  - Demanda de bicis en una estación BiciMad durante una determinada hora.

- **Función de masa de probabilidad:** $X \sim \mathcal{P}(\lambda)$ toma valores $k = 0, 1, 2, 3, \ldots$ con probabilidad:

$$\boxed{P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \qquad k = 1, 2, 3, \ldots}$$

- **Función generadora de momentos y función característica:** Calculando

$$\boxed{M(t) = e^{\lambda(e^t - 1)}} \qquad \boxed{\varphi(t) = e^{\lambda(e^{it} - 1)}}$$

- **Esperanza y varianza:** Basta hallar las derivadas:

$$\boxed{E[X] = V[X] = \lambda.}$$

- **Ejemplo:** En la centralita de emergencias del 112, tenemos la variable X = "núm. de llamadas de llamadas en una hora", con $X \sim \mathcal{P}(2)$:

  a) Probabilidad de no recibir ningún aviso en una hora: como $\lambda = 2$, tenemos

$$P(X = 0) = e^{-2} \frac{2^0}{0!} = 0.1353.$$

  b) Probabilidad de recibir algún aviso en media hora: vamos a tener un promedio de $\lambda = 2 \times 1/2 = 1$ llamadas por media hora, por lo que $Y \sim \mathcal{P}(1)$ y

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-1} = 0.6321.$$

  c) ¿Cuantas llamadas se esperan recibir en 8 horas? El promedio en 8 horas será de $\lambda = 2 \times 8 = 16$, por lo que $Z \sim \mathcal{P}(16)$ y $E[Z] = 16$.

- **Propiedad de la suma:** Si $X \sim \mathcal{P}(\lambda_1)$ y $Y \sim \mathcal{P}(\lambda_2)$ son <u>independientes</u>, entonces

$$X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2).$$

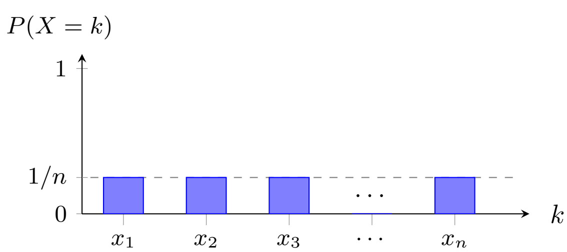
☞ **Aproximación por Poisson de $\mathcal{B}(n,p)$ con n grande y p pequeña:** Tomando $\boxed{\lambda = np}$, podemos aproximar $X \sim \mathcal{B}(n,p)$ por una distribución de Poisson $Y \sim \mathcal{P}(\lambda)$:

$$P(X = k) \approx P(Y = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

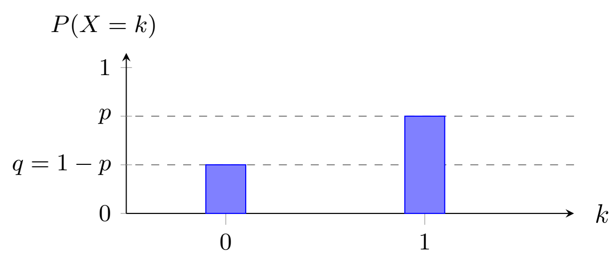
☞ **Ejemplo:** Número de erratas para un texto de 3500 letras. Supongamos que la probabilidad de que una letra esté mal es de $p = 0.001$. ¿Cuál es la probabilidad de que haya a los sumo 3 erratas? Podemos aproximar por una Poisson con $\lambda = 3500 \times 0.001 = 3.5$:

$$P(X \leq 3) = e^{-3.5}\left(\frac{3.5^0}{0!} + \frac{3.5^1}{1!} + \frac{3.5^2}{2!} + \frac{3.5^3}{3!}\right) = 0.5366.$$

Apéndice: gráficas de funciones de masa de probabilidad



(a) Uniforme: $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$



(b) Bernoulli: $X \sim \mathcal{B}er(p)$

Figura 2: Variables más simples: uniforme y Bernoulli

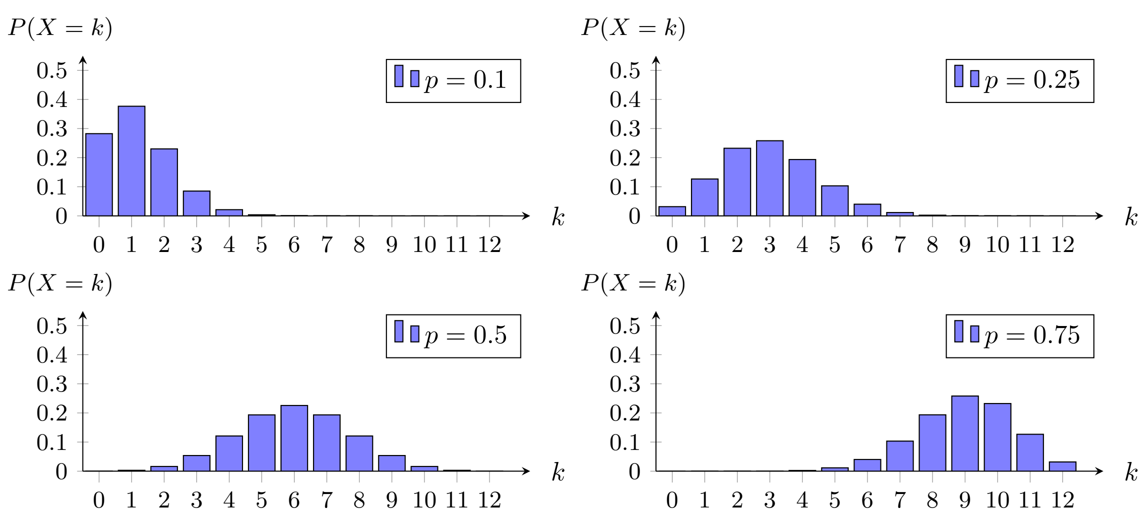


Figura 3: Binomial $X \sim \mathcal{B}(12, p)$ con valores $p = 0.1, 0.25, 0.5$ y 0.75.

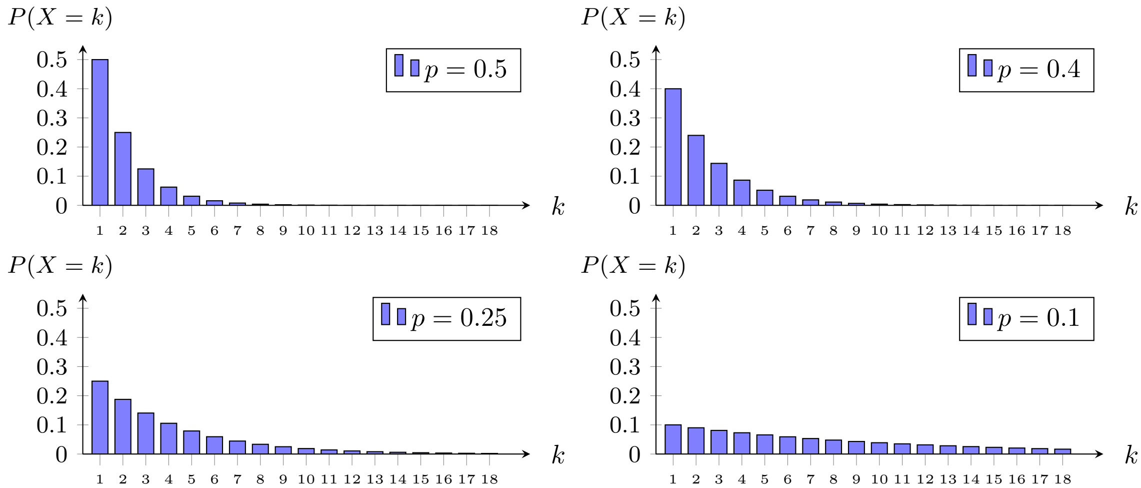


Figura 4: Geométrica $X \sim \mathcal{G}(p)$ con parámetros $p = 0.5, 0.4, 0.25$ y 0.1.

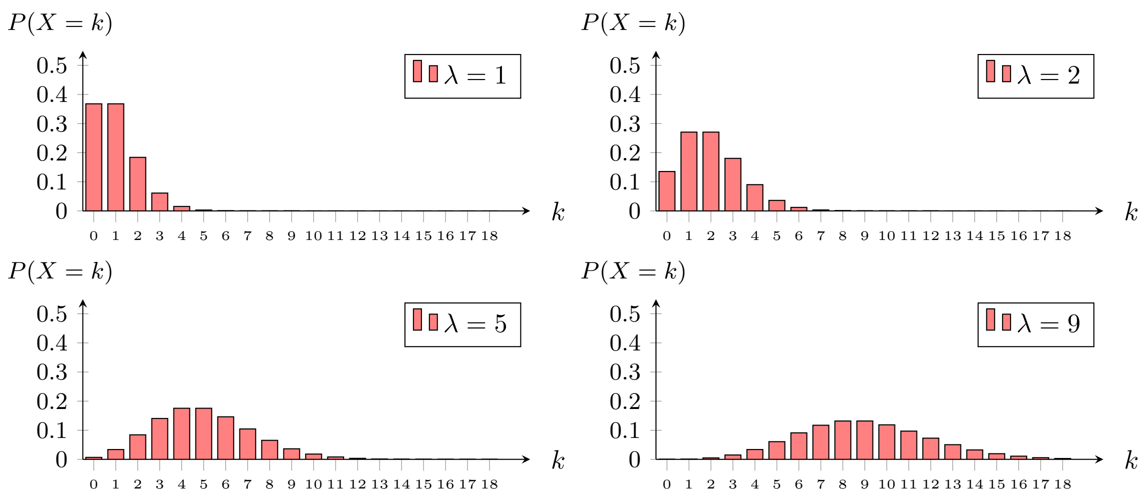


Figura 5: Poisson $X \sim \mathcal{P}(\lambda)$ con parámetros $\lambda = 1, 2, 5$ y 9.

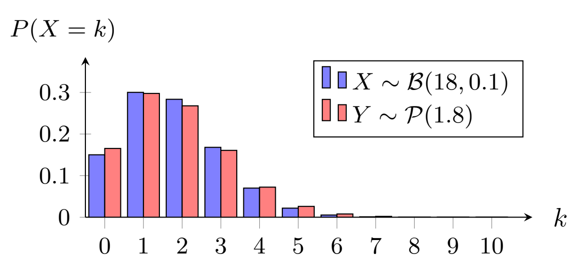


Figura 6: Aproximación de la Binomial $X \sim \mathcal{B}(18, 0.1)$ por una Poisson con $\lambda = 18 \times 0.1 = 1.8$.