Intro a la Probabilidad y estadistica

Martes y Jueves Aula B17

Dra Ana Georgina Flesia

Degenerada: $X \equiv \delta_c$

Una variable aleatoria X tiene **distribución degenerada en el punto c** si tiene toda la probabilidad concentrada en dicho punto

Función de masa:

$$P(X = c) = 1, \qquad P(X \neq c) = 0$$

$$E[X] = c, \qquad V(X) = c^2 - c^2 = 0$$

Uniforme discreta: $X \equiv Unif\{x_1, \ldots, x_n\}$

Una variable aleatoria X tiene una **distribución uniforme discreta** sobre el conjunto de números $\{x_1, \ldots, x_n\}$ si la probabilidad de que tome cualquiera de estos valores es la misma

Función de masa:

$$P(X = x_i) = \begin{cases} \dfrac{1}{n}, & \text{si } i = 1, \ldots, n \\ 0, & \text{en otro caso} \end{cases}$$

$$E[X] = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad V(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

Pruebas de Bernouilli

Experimento aleatorio con dos resultados posibles, "éxito", E y "fracaso", F, con probabilidades $P(E) = p$ y $P(F) = 1 - p = q$.

- "Observar un chip al azar de los producidos en una fábrica" $\rightarrow$ $E =$"el chip no es defectuoso" y $F =$"el chip es defectuoso".
- "Lanzar una moneda dos veces" $\rightarrow$ nos puede interesar sacar el mismo resultado las dos veces $\rightarrow$ $E = \{CC, ++\}$ y $F = \{C+, +C\}$.
- "Elegir al azar un número entre 0 y 1" $\rightarrow$ podemos considerar $\rightarrow$ $E = [0, 0{,}3]$ y $F = (0{,}3, 1]$.

Bernuilli: $X \equiv Bern(p)$

La variable aleatoria X obtenida al realizar una prueba de Bernouilli con $P(E) = p$ tiene distribución $Bern(p)$

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

Función de masa:

$$P(X = x) = p^x(1 - p)^{1-x} = p^x q^{1-x},\ x = 0, 1$$

$$E[X] = p, \qquad V(X) = p(1 - p) = pq$$

- **SUMA:** la suma de v.a.'s independientes $Bern(p)$ da lugar a otro tipo de v.a. denominada Binomial, de la que la Bernouilli es un caso particular.

Binomial

Sea I_A la variable indicadora de un evento particular.

$$I_A = \begin{cases} 1 & \text{si } A \text{ ocurre} \\ 0 & \text{si } A \text{ no ocurre} \end{cases}$$

Entonces I_A es una variable Bernoulli de parámetro $p = P(A)$.
La esperanza de una variable Bernoulli es p y su varianza es $p(1-p)$.

Binomial: $X \equiv B(n,p)$

- se genera por la repetición de n pruebas de Bernouilli independientes
- $X=$"número de éxitos obtenidos en las n pruebas de Bernouilli"
- puede escribirse como la suma de n v.a. $Bern(p)$, independientes:

$$X = X_1 + ... + X_n,$$

con

$$X_i = \begin{cases} 1, & \text{si obtenemos } E \text{ en la } i\text{-ésima prueba} \\ 0, & \text{si obtenemos } F \text{ en la } i\text{-ésima prueba} \end{cases} \quad (X_i \equiv Bern(p), i = 1, ..., n)$$

Función de masa:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}, x = 0, 1, ..., n$$

$$E[X] = np, \qquad V(X) = npq$$

Binomial:Ejemplo

En 1693, Samuel Pepys (a quien hoy se recuerda mejor por su diario) escribió una carta a Isaac Newton preguntándole acerca de una apuesta que Pepys planeaba hacer. Pepys quería saber cuál de los siguientes eventos tenía la mayor probabilidad de ocurrir.

1. $A = \{$Se lanzan 6 dados y al menos 1 es un seis$\}$
2. $B = \{$Se lanzan 12 dados y al menos 2 son seis$\}$
3. $C = \{$Se lanzan 18 dados y al menos 3 son seis$\}$

Pepys pensó que el tercero tenía la mayor probabilidad, pero Newton no estaba de acuerdo.

Binomial: $P(A)$

La probabilidad de l evento A es sencilla de calcular.

$$\begin{aligned} P(\text{al menos un seis en seis tiradas}) &= 1 - P(\text{0 seis en seis tiradas}) \\ &= 1 - \frac{5^6}{6^6} \\ &\approx .665 \end{aligned}$$

Sin embargo, las probabilidades de los otros dos eventos son más complicadas de calcular.
Por ejemplo, no es obvio cómo contar el número de formas de obtener exactamente 2 seis en 18 tiradas de dados.

Binomial: $P(B)$

1. Definamos como éxito E obtener un seis en una tirada de dados y fracaso F no observar un seis en la misma tirada. La probabilidad de éxito es $p = 1/6$ y la de fracaso $q = 5/6$.
2. Este es un experimento Bernoulli.
3. Si lo repetimos 12 veces en forma independiente (tiramos 12 dados o tiramos un dado 12 veces en forma consecutiva), y definimos a X como el número de éxitos en los 12 experimentos, esta variable tiene distribución binomial de parámetros $n = 12$ y $p = 1/6$.
4. $P(B) = P(X \geq 2)$

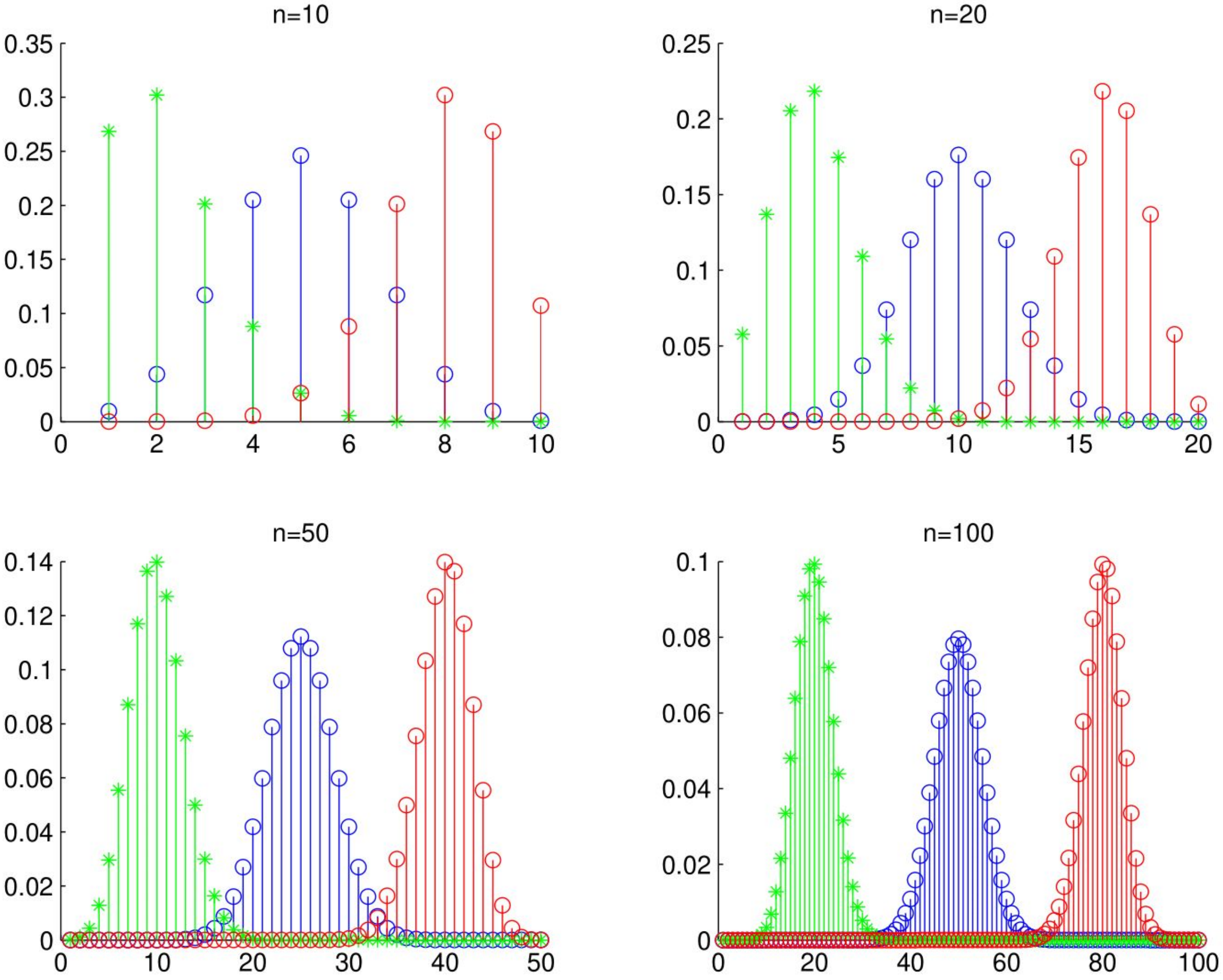
Binomial: $P(B)$

$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X \leq 1) \\
&= 1 - [P(X = 0) + P(X = 1)] \\
&= 1 - \binom{12}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{12} - \binom{12}{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{11} \\
&= 1 - \left(\frac{5}{6}\right)^{12} - 12\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^{11} \\
&=\approx .6187.
\end{aligned}
$$

Binomial: $P(C)$

1. Sea Y el número de seis en 18 tiradas de un dado. Y es una variable binomial con parámetros $p = 1/6$ y $n = 18$.

$$
\begin{aligned}
P(Y \geq 3) &= 1 - P(Y \leq 2) \\
&= 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)] \\
&= 1 - \binom{18}{0}\left(\frac{1}{6}\right)^{0}\left(\frac{5}{6}\right)^{18} - \binom{18}{1}\left(\frac{1}{6}\right)^{1}\left(\frac{5}{6}\right)^{17} \\
&- \binom{18}{2}\left(\frac{1}{6}\right)^{2}\left(\frac{5}{6}\right)^{16} \\
&\approx 0.5973.
\end{aligned}
$$



n=10
n=20
n=50
n=100

Poisson: $X \equiv \mathbb{P}(\lambda)$

- expresa la probabilidad de que ocurran un número x de sucesos en un tiempo fijo, si ocurren con una tasa media conocida, λ, y son independientes del tiempo transcurrido desde el último suceso
- se aplica a sucesos con probabilidad muy baja de ocurrir (sucesos "raros")

Función de masa:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

$$E[X] = V[X] = \lambda$$

- **SUMA:** si $X_i \equiv \mathbb{P}(\lambda_\mathbf{i})$ son independientes, $\Rightarrow$ la v.a. $\sum_i X_i \equiv \mathbb{P}(\sum_\mathbf{i} \lambda_\mathbf{i})$

Poisson

1. Estás en una habitación con n personas.
2. Cada persona en la sala ha contribuido con \$1000 a una colecta.
3. El dinero recolectado se redistribuirá de vuelta a las personas en la sala, de la siguiente manera: cada billete de mil pesos tiene la misma probabilidad de ir a cualquiera de los n personas, independientemente de los otros billetes recolectados.
4. Esto significa que algunas personas podrían recibir más de \$1000, mientras que otras se quedan con nada.

Poisson

1. A medida que n tiende a infinito ¿Cuál es la probabilidad de que termines sin dinero?
2. Hay dos corrientes de pensamiento comunes:
   2.1 A medida que n tiende a infinito , el número de billetes en la colecta aumenta hasta el infinito, así que parece que la probabilidad de que termines con al menos uno de esos billetes debería acercarse a 1, es decir, la probabilidad de que termines sin dinero se aproxima a 0.
   2.2 La probabilidad de que ganes un billete de 1000 es $1/n$ y medida que n tiende a infinito, esa probabilidad disminuye a 0, por lo que parece que la probabilidad de que termines sin dinero se aproxima a 1.
3. ¿Cuál escuela de pensamiento es la correcta? Resulta que ambos argumentos están equivocados. Veamos por qué.

Poisson

1. Vamos a definir la variable X como el número de billetes que recibes de la colecta de n billetes. La variable X tiene distribución binomial de parámetros $(n, 1/n)$, por lo cual la probabilidad de no recibir ningún billete es

$$\begin{aligned} P(X=0) &= \binom{n}{0}\left(\frac{1}{n}\right)^0\left(1-\frac{1}{n}\right)^n \\ &= \left(1-\frac{1}{n}\right)^n. \end{aligned}$$

2. Cual es el límite cuando n tiende a infinito? Es un límite famoso en matemática, un límite notable.

Poisson

1. Para calcularlo, se toma el logaritmo (natural) para bajar la n del exponente y aplica la regla de L'Hôpital:

$$\begin{aligned}
\lim_{n\to\infty} \log\left(1-\frac{1}{n}\right)^n &= \lim_{n\to\infty} n\log\left(1-\frac{1}{n}\right) \\
&= \lim_{x\to 0} \frac{\log(1-x)}{x} \\
&= \lim_{x\to 0} \frac{\frac{d}{dx}\log(1-x)}{\frac{d}{dx}x} \\
&= \lim_{x\to 0} \frac{-\frac{1}{1-x}}{1} \\
&= -1.
\end{aligned}$$

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = e^{-1}.$$

2. La probabilidad de no recibir ningún dinero es $1/e \approx 0.368$ cuando $n \to \infty$

Poisson

1. Si X es una variable Binomial de parámetros $(n, p = \mu/n)$, la función de masa de X converge a la función de masa de una variable Poisson con parámetro μ cuando n tiende a infinito.

$$
\begin{aligned}
f(x) &= \binom{n}{x} (\frac{\mu}{n})^x (1 - \frac{\mu}{n})^{n-x} \\
&= \frac{n!}{x!(n-x)!} \frac{\mu^x}{n^x} (1 - \frac{\mu}{n})^{n-x} \\
&= \frac{n!}{(n-x)! n^x} \cdot \frac{\mu^x}{x!} \cdot (1 - \frac{\mu}{n})^{n-x} \\
&\to 1 \cdot \frac{\mu^x}{x!} \cdot e^{-\mu}
\end{aligned}
$$

Ejemplo Poisson

1. El número de errores en un artículo de investigación cuando llega al editor final es una variable con distribución Poisson con $\mu = 4.6$. ¿Cual es la probabilidad de que haya 2 o mas errores en un artículo seleccionado al azar?

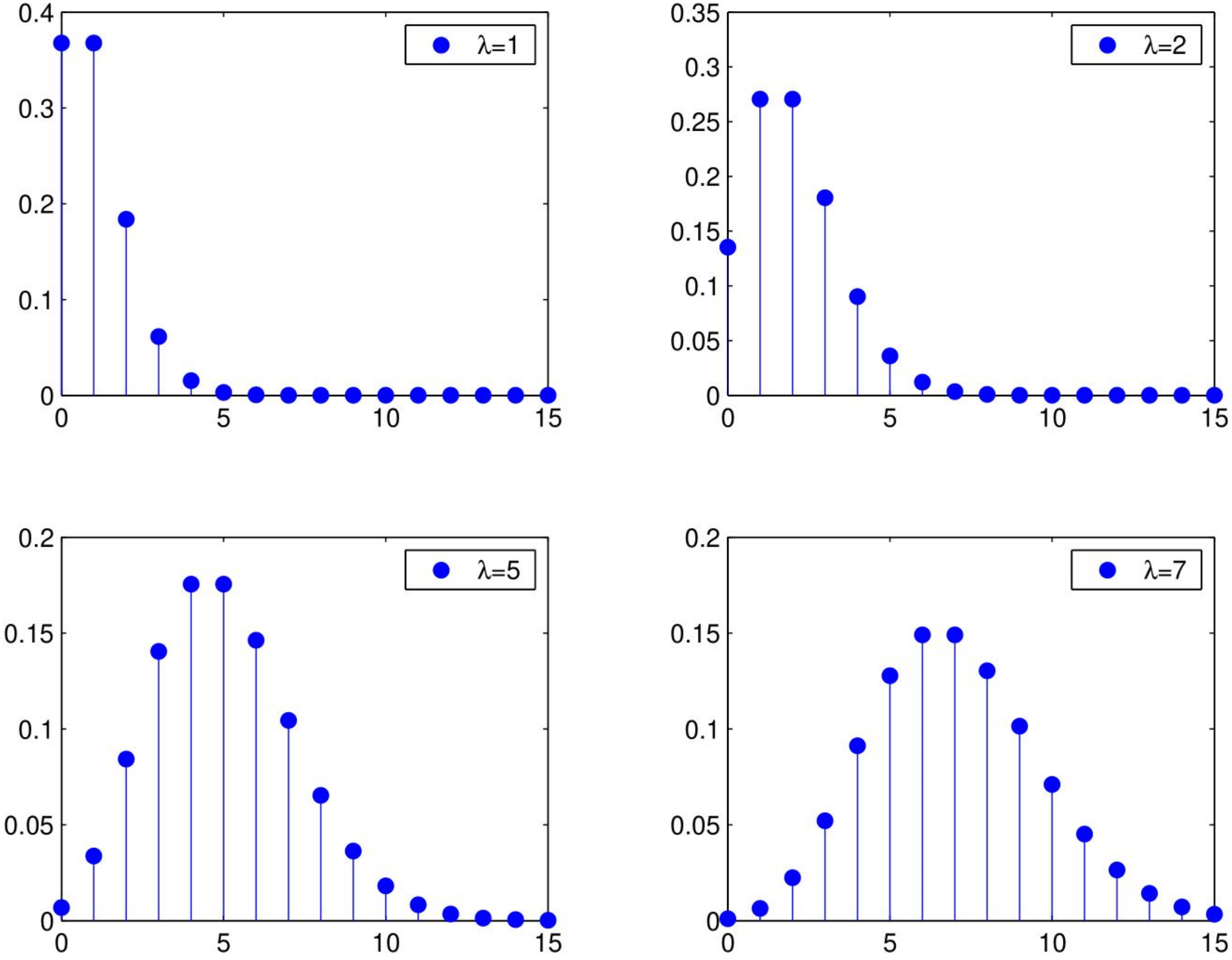
$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) \\
&= 1 - f(0) - f(1) \\
&= 1 - e^{-4.6}\frac{4.6^0}{0!} - e^{-4.6}\frac{4.6^1}{1!} \\
&= 1 - e^{-4.6} - 4.6e^{-4.6} \\
&\approx .944.
\end{aligned}
$$

Ejemplo Poisson

1. ¿Por qué podría ser razonable utilizar la distribución de Poisson como modelo para el número de errores tipográficos?
2. Cada artículo científico tiene muchas palabras (por ejemplo, $n = 1000$).
3. Hay una pequeña probabilidad de que cada palabra tenga un error tipográfico (por ejemplo, $p = 0.0046$).
4. Si los errores tipográficos son independientes entre palabras, entonces el número de errores tipográficos sigue una distribución binomial.
5. Dado que n es grande y p es pequeño, esta distribución binomial se puede aproximar a una distribución Poisson ($\mu = np = 4.6$).

Ejemplo Poisson

1. Sin embargo, todo esto es conjetura.
2. No se nos dice cuántas palabras tiene el artículo de opinión, ni la probabilidad de que cada palabra tenga un error tipográfico.
3. Simplemente asumimos que el número de errores tipográficos sigue una distribución de Poisson.
4. En la práctica, el modelo de Poisson se utiliza a menudo para datos de conteo, incluso cuando no hay un modelo binomial subyacente.


λ=1
λ=2
λ=5
λ=7
0.4
0.3
0.2
0.1
0
0.35
0.25
0.15
0.05
0
5
10
15

Geométrica $X \equiv Geom(p)$

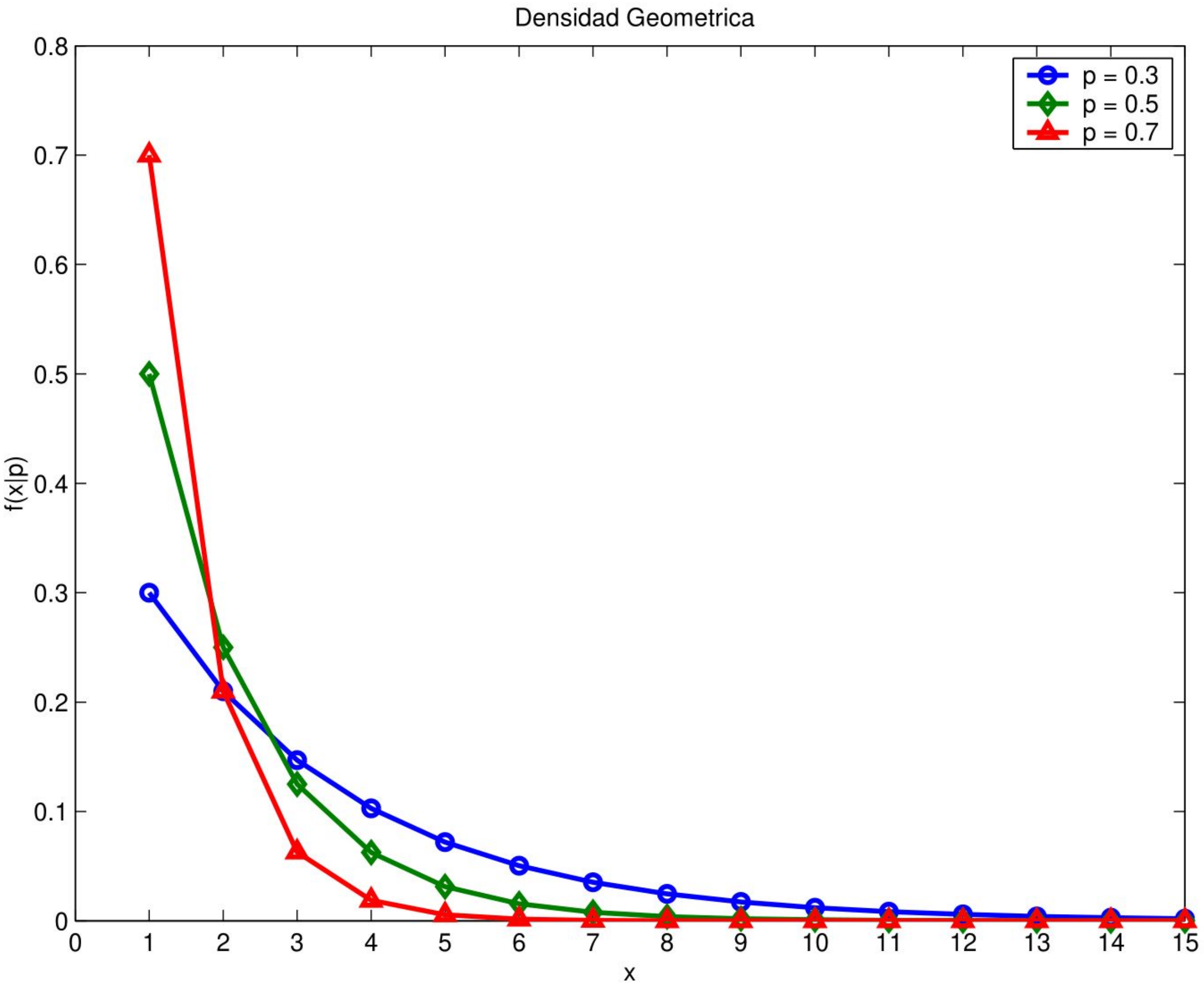
- se genera por la repetición de pruebas de Bernouilli con $P(E) = p$, independientes hasta que obtenemos el primer éxito
- $X=$"número de fracasos hasta el primer éxito"

Función de masa:

$$P(X = x) = (1 - p)^x p = q^x p, \qquad x = 0, 1, ...$$

$$E[X] = \frac{q}{p}, \qquad V(X) = \frac{q}{p^2}$$

- **SUMA:** la suma de v.a.'s independientes $Geom(p)$ da lugar a otro tipo de v.a. denominada Binomial Negativa, de la que la Geométrica es un caso particular.


Densidad Geometrica
p = 0.3
p = 0.5
p = 0.7
f(x|p)
x
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Binomial Negativa $X \equiv BN(r, p)$

- se genera por la repetición de pruebas de Bernouilli con $P(E) = p$, independientes hasta que obtenemos el r-ésimo éxito
- X="número de fracasos hasta el r-ésimo éxito"
- puede escribirse como suma de r v.a. $Geom(p)$, independientes, $X = X_1 + ... + X_r$, con X_i="número de fracasos comprendidos entre el $(i-1)$-ésimo éxito y el i-ésimo éxito', $(X_i \equiv Geom(p),\ i = 1, ..., r)$

Función de masa

$$P(X = x) = \binom{x+r-1}{x}(1-p)^x p^r = \binom{x+r-1}{x} q^x p^r, \qquad x = 0, 1, ...$$

$$E[X] = r\frac{q}{p}, \qquad V(X) = r\frac{q}{p^2}$$

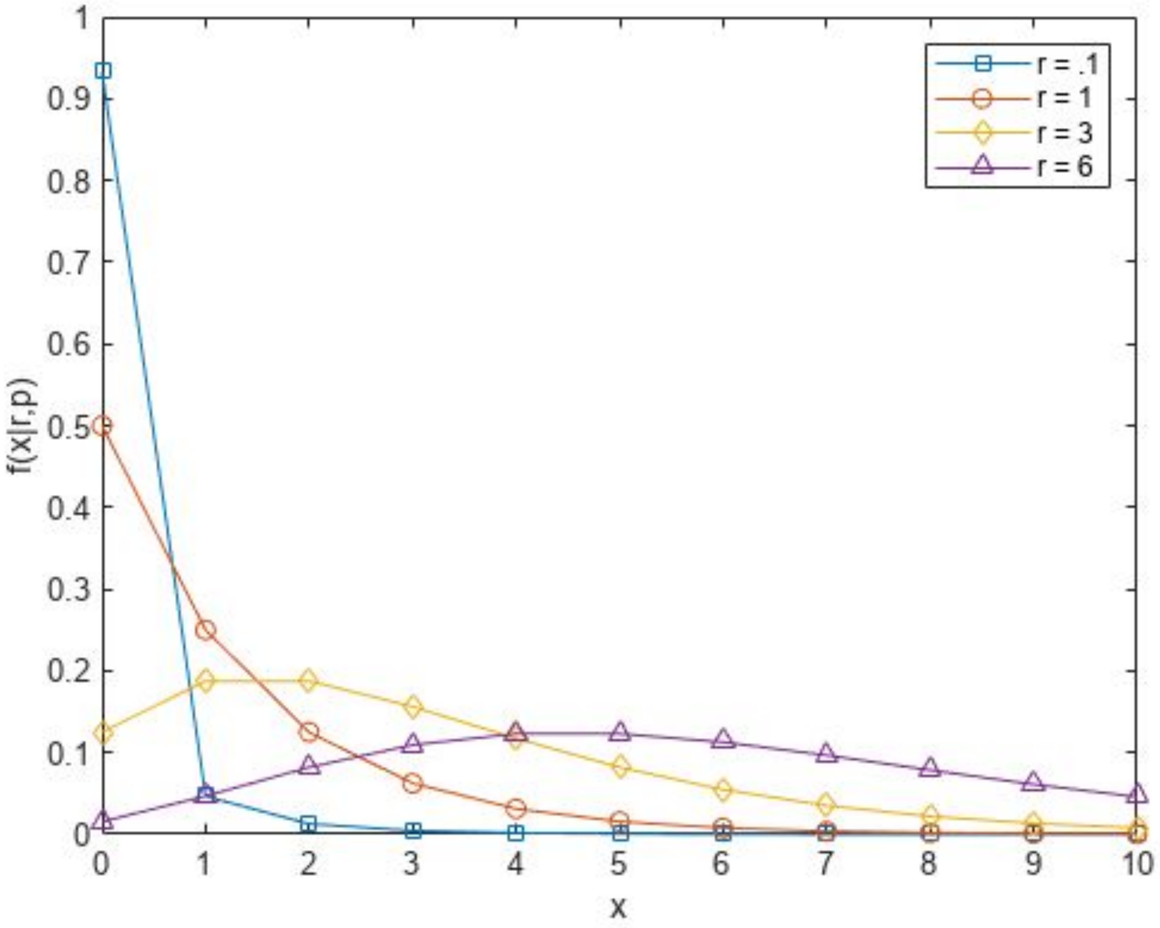
Binomial Negativa $X \equiv BN(r, p)$

- se genera por la repetición de pruebas de Bernouilli con $P(E) = p$, independientes hasta que obtenemos el r-ésimo éxito
- X="número de fracasos hasta el r-ésimo éxito"
- puede escribirse como suma de r v.a. $Geom(p)$, independientes, $X = X_1 + ... + X_r$, con X_i="número de fracasos comprendidos entre el $(i-1)$-ésimo éxito y el i-ésimo éxito', ($X_i \equiv Geom(p)$, $i = 1, ..., r$)

Función de masa

$$P(X = x) = \binom{x+r-1}{x}(1-p)^x p^r = \binom{x+r-1}{x} q^x p^r, \qquad x = 0, 1, ...$$

$$E[X] = r\frac{q}{p}, \qquad V(X) = r\frac{q}{p^2}$$



p=0.5