Distribuciones de probabilidad discretas

* Experimento Bernoulli:

- Es un experimento con elementos de éxito o fracaso.
- La probabilidad de éxito se denota por p. La de fracaso por q, o bien, 1−p.

Distribución de Bernoulli

$$f(x) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases} \qquad X \sim Ber(p)$$

(en sí no es una distribución como tal)

$$E(X) = \sum_{x=0}^{1} x \cdot P(X=x) = 0 \cdot \cancelto{(1-p)}{P(X=0)} + 1 \cdot \cancelto{p}{P(X=1)} = p.$$

$$V(X) = ?$$

Función generatriz de momentos:

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{1} e^{tx} \cdot P(X=x) = \underbrace{e^0}_{1} \cdot \cancelto{(1-p)}{P(X=0)} + e^t \cdot \cancelto{p}{P(X=1)} = (1-p) + pe^t$$

$$\mathbb{E}(X) = \alpha_1 = \left.\frac{dM_X}{dt}\right|_{t=0} = \left. pe^t \right|_{t=0} = p$$

$$\mathbb{E}(X^2) = \alpha_2 = \left.\frac{d^2M_X}{dt^2}\right|_{t=0} = \left. pe^t \right|_{t=0} = p$$

$$\therefore V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1-p) = pq$$

* Experimento Binomial:

- Consta de n "miniexperimentos" llamados ensayos, donde n se fija previo al experimento
- Dos resultados posibles: éxito o fracaso.
- Los ensayos son independientes. (El resultado en un ensayo cualquiera no influye en el resultado de otro)
- La probabilidad de éxito (p), se mantiene ctte. entre ensayos.

REGLA	Considérese muestreo sin reemplazo de una población dicotómica de tamaño N. Si el tamaño de la muestra (número de ensayos) n es cuando mucho 5% del tamaño de la población, el experimento puede ser analizado como si fuera exactamente un experimento binomial.

* Distribución Binomial:

- Nace de la "distribución" de Bernoulli.
- Se aplica para experimentos binomiales.

Función de masa:

$$X \sim bin(x; n, p) \Rightarrow P(X = x) = \binom{n}{x} \cdot p^x \cdot \underbrace{q}_{1-p}{}^{n-x} \quad ; \quad x \in \bigcup_{i=0}^{n} \{i\}$$

Esperanza y varianza → Función generatriz de momentos.

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x q^{n-x} = \dots = \text{(completar)}$$

$$\mathbb{E}(X) = \alpha_1 = \left.\frac{dM_x}{dt}\right|_{t=0} = \dots = np \; ; \quad V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \alpha_1 - \alpha_2^2 = \dots = npq$$

La probabilidad de que un paciente se recupere de una rara enfermedad sanguínea es de 0.4. Si se sabe que 15 personas contrajeron la enfermedad, ¿cuál es la probabilidad de que *a*) sobrevivan al menos 10, *b*) sobrevivan de 3 a 8, y *c*) sobrevivan exactamente 5?

$p = 0.4$

$n = 15$

$X \sim bin(15, 0.4)$; X: Nro pacientes que se recuperan

a) $$\mathbb{P}(X \geq 10) = \sum_{x=10}^{15} \mathbb{P}(X = x) = \sum_{x=10}^{15} \binom{15}{x} \cdot (0.4)^x \cdot (0.6)^{15-x}$$

b) $$\mathbb{P}(3 \leq X \leq 8) = \sum_{x=3}^{8} \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

c) $$\mathbb{P}(X = 5) = \binom{15}{5} (0.4)^5 \cdot (0.6)^{10}$$

Función de distribución acumulativa para $X \sim bin(n,p)$:

$$P(X \leq x) = Bin(x; n, p) = \sum_{y=0}^{x} bin(y; n, p) \quad ; \quad x \in \bigcup_{i=0}^{n} \{i\}$$

Suponga que 20% de todos los ejemplares de un libro de texto particular no pasan una prueba de resistencia de encuadernación. Sea X el número entre 15 ejemplares seleccionados al azar que no pasan la prueba. Entonces X tiene una distribución binomial con $n = 15$ y $p = 0.2$.

1. La probabilidad de que cuando mucho 8 no pasen la prueba es
2. La probabilidad de que exactamente 8 fallen es
3. La probabilidad de que por lo menos 8 fallen es
4. Finalmente, la probabilidad de que entre 4 y 7, inclusive, fallen es

1. $$P(X \leq 8) = Bin(8; 15, 0.2) = \sum_{y=0}^{8} bin(y; 15, 0.2)$$

2. $$P(X = 8) = P(X \leq 8) - P(X \leq 7) = Bin(8; 15, 0.2) - Bin(7; 15, 0.2)$$

3. $$P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7) = 1 - Bin(7; 15, 0.2)$$

4. $$P(4 \leq X \leq 7) = P(X \leq 7) - P(X \leq 3) = Bin(7; 15, 0.2) - Bin(3; 15, 0.2)$$

* Distribución Multinomial:

$E_1, E_2, \dots, E_K$

- Es una generalización de la binomial.
- Ya no hay solo 2 resultados posibles (hay K resultados posibles)
- Cada resultado tiene su propia probabilidad:

$$P(E_1) = p_1$$
$$P(E_2) = p_2$$
$$\vdots$$
$$P(E_K) = p_K$$

- Se hacen n ensayos independientes.
- Definimos las variables aleatorias $X_1, X_2, \dots, X_K$, de modo que X_i es el número de veces que ocurre E_i, $\forall\ i \in \mathbb{N}_0 / 0 \leq i \leq K$.

Se debe cumplir que $\sum_{i=1}^{K} p_i = 1$, y $\sum_{i=1}^{K} x_i = n$

La función de masa es:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_K = x_K; p_1, p_2, \dots, p_K, n) = \binom{n}{x_1, x_2, \dots, x_K} p_1^{x_1} p_2^{x_2} \dots p_3^{x_K}$$

* Hipergeométrica:

- Con reemplazo → Independencia → Binomial/Multinomial
- Sin reemplazo → Dependencia → Hipergeométrica

Consideraciones para la hipergeométrica:

- Población de N elementos (N finito). Se extraen n sin reemplazo.
- Cada elemento puede ser caracterizado de 2 formas, y hay K elementos que son de la categoría 1 ("éxitos")

Sea X el número de "éxitos" (elementos de la categoría 1) en la muestra, entonces:

$$P(X=x) = h(x; n, K, N) = \frac{\binom{K}{x}\cdot\binom{N-K}{n-x}}{\binom{N}{n}}$$

$$E(X) = \frac{nK}{N} \; ; \quad V(X) = \frac{N-n}{N-1}\cdot n\cdot\frac{K}{N}\cdot\left(1-\frac{K}{N}\right)$$

Lotes con 40 componentes cada uno que contengan 3 o más defectuosos se consideran inaceptables. El procedimiento para obtener muestras del lote consiste en seleccionar 5 componentes al azar y rechazar el lote si se encuentra un componente defectuoso. ¿Cuál es la probabilidad de, que en la muestra, se encuentre exactamente un componente defectuoso, si en todo el lote hay 3 defectuosos?

$N = 40$ (total) ; $n = 5$ (muestra de componentes) ; $K = 3$ (defectuosos)

X : Nro de componentes defectuosos que se obtienen de la muestra.

$$X \sim h(x; n, K, N) \rightarrow P(X=1) = h(1; 5, 3, 40) = \frac{\binom{3}{1}\binom{40-3}{5-1}}{\binom{40}{5}}$$

Se capturaron, etiquetaron y liberaron cinco individuos de una población de animales que se piensa están al borde de la extinción en una región para que se mezclen con la población. Después de haber tenido la oportunidad de mezclarse, se selecciona una muestra aleatoria de 10 de estos animales. Sea X = el número de animales etiquetados en la segunda muestra. Si en realidad hay 25 animales de este tipo en la región, ¿cuál es la probabilidad de que a) $X = 2$? b) ¿$X \leq 2$?

$n = 10$ (muestra) ; $N = 25$ (total de animales) ; $K = 5$ (animales etiquetados) $\rightarrow X \sim h(x; n, K, N)$

a) $$P(X=2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{25-5}{10-2}}{\binom{25}{10}} = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}}$$

b) $$P(X \leq 2) = H(2; 10, 5, 25) = \sum_{y=0}^{2} h(y; 10, 5, 25) = \sum_{y=0}^{2} \frac{\binom{5}{y}\binom{20}{10-y}}{\binom{25}{10}}$$

* Hipergeométrica Multinomial:

Hay 30 estudiantes en Física I, se sabe que 12 aprobarán. 5 de ellos con susti y/o examen, y 7 aprobarán por eximición inmediata. ¿Cuál es la probabilidad de que al seleccionar 5 estudiantes al azar, 2 vayan a aprobar sin mayor inconveniente, 1 apruebe yéndose a susti y/o examen, y los otros 2 reprueben?

$$P(X_1=2, X_2=1, X_3=2; n=5) = \frac{\binom{7}{2}\binom{5}{1}\binom{18}{2}}{\binom{30}{5}}$$

* Binomial negativa y geométrica:

- Estas distribuciones se utilizan cuando se deben hacer experimentos hasta que se cumpla una acción.
- Estos experimentos se conocen como experimentos negativos.
- Al proceso se le llama proceso binomial negativo.
- Variable aleatoria binomial negativa: Número de ensayos necesarios para generar K éxitos.

Binomial Negativa:

- El experimento es una serie de ensayos independientes (al igual que en la binomial).
- Cada ensayo puede resultar en éxito o fracaso (al igual que en la binomial).
- La probabilidad de éxito (p), es cte. entre ensayos (al igual que en la binomial).
- El nº de ensayos (x) es desconocido (a diferencia de la binomial), ya que el experimento sigue hasta que ocurran K éxitos.

Deducción de función de masa:

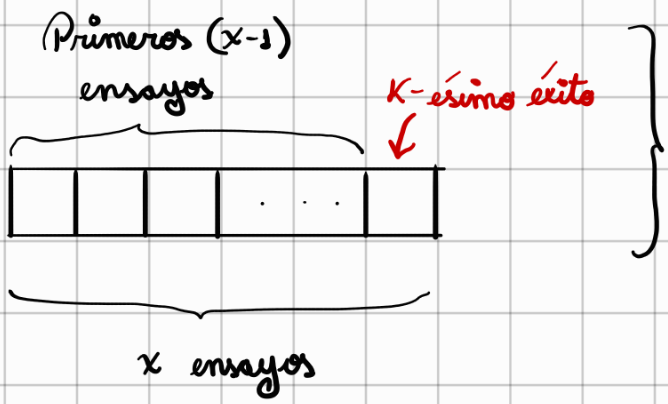
Supongamos que estamos jugando con una computadora. La compu genera aleatoriamente un nro. del 1 al 20 cada vez. Nosotros debemos tratar de adivinar el nro en cada ensayo. El juego termina luego de que logre adivinar el nro generado por la computadora por 3^{ra} vez, y el puntaje se calcula como:

$$p(x) = \frac{100}{\sqrt{x-2}}$$, en donde x es el nro. de rondas que fueron necesarias para que el juego termine.

Hallar la probabilidad de que el juego termine luego de x rondas. $(x \geq 3)$, y el pje. esperado

$p = \frac{1}{20}$ (cte. entre ensayos).

$K = 3$ (al 3^{er} éxito termina)



$\Rightarrow$ **Análisis:**
Para que en el n-ésimo ensayo haya ocurrido el K-ésimo éxito, en los primeros $(x-1)$ ensayos, deben haber ocurrido $(K-1)$ éxitos

Entonces, la función de masa para $X \sim b^*(x; K, p)$ es:

$$f(x) = \underbrace{\binom{x-1}{K-1}}_{\text{Formas de escoger los (K-1) éxitos de los (x-1) primeros ensayos}} \cdot \underbrace{p^K}_{\text{K fueron éxito. Al ser independientes, queda } p^K} \cdot \underbrace{(1-p)^{x-K}}_{\text{(x-K) fueron fracaso.}} \quad ; \quad x \geq K \ / \ x, K \in \mathbb{N} - \{0\}$$

Entonces:

$$\mathbb{P}(X=x) = \binom{x-1}{3-1}\cdot\left(\frac{1}{20}\right)^3\cdot\left(\frac{19}{20}\right)^{x-3} = \binom{x-1}{2}\cdot\left(\frac{1}{20}\right)^3\cdot\left(\frac{19}{20}\right)^{x-3}$$

Esperanza del puntaje:

Sea $P = p(X)$; $X \sim b^*(x; 3, 0.05)$

$$\Rightarrow \mathbb{E}(P) = \sum_{\substack{x \geq 3 \\ x \in \mathbb{N}}} p(x)\cdot b^*(x; 3, 0.05)$$

Pero la binomial negativa también se puede plantear de otra forma.

Recordemos que x representaba el número de ensayos necesarios para llegar al k-ésimo éxito. Por lo que si tomamos un y, que representa el nro de fracasos que se necesitaron para llegar al k-ésimo éxito, tendríamos que:

$$\underbrace{x}_{\text{total}} = \underbrace{y}_{\text{fracasos}} + \underbrace{k}_{\text{éxitos}}.$$

$$\Rightarrow f(x) = \binom{x-1}{k-1}\cdot p^k\cdot(1-p)^{x-k} = \binom{y+k-1}{k-1}\cdot p^k\cdot(1-p)^{y+\cancel{k}-\cancel{k}}$$

Pero usualmente, a k, ahora se le llama r, por lo que:

$$Y \sim nb(y; r, p) \Leftrightarrow P(Y=y) = \binom{y+r-1}{r-1} p^r (1-p)^y \; ; \; y \geq 0, r \geq 1$$

Geométrica: Caso específico de la binomial negativa ($k=1$, según la 1^{ra} interpretación, ó $r=1$ según la 2^{da})

① sea:

función de masa distr. geométrica

* Si $X \sim geom(x;p) \Rightarrow X \sim b^*(x;\overset{k}{\overset{\shortparallel}{1}},p)$

$$\Rightarrow P(X=x) = \binom{x-1}{1-1}\cdot p^1\cdot(1-p)^{x-1} = \cancel{\binom{x-1}{0}}^{1}\, p\,(1-p)^{x-1} = p(1-p)^{x-1}$$

* Si $Y \sim geom(y;p) \Rightarrow Y \sim nb(y;\overset{r}{\overset{\shortparallel}{1}},p)$

$$\Rightarrow P(Y=y) = \binom{y+\cancel{1}-\cancel{1}}{1-1}\cdot p^1\cdot(1-p)^y = \cancel{\binom{y}{0}}^{1}\cdot p(1-p)^y = p(1-p)^y$$

En la serie de campeonato de la NBA (National Basketball Association), el equipo que gane 4 de 7 juegos será el ganador. Suponga que los equipos *A* y *B* se enfrentan en los juegos de campeonato y que el equipo *A* tiene una probabilidad de 0.55 de ganarle al equipo *B*.

a) ¿Cuál es la probabilidad de que el equipo *A* gane la serie en 6 juegos?

b) ¿Cuál es la probabilidad de que el equipo *A* gane la serie?

c) Si ambos equipos se enfrentaran en la eliminatoria de una serie regional y el triunfador fuera el que ganara 3 de 5 juegos, ¿cuál es la probabilidad de que el equipo *A* gane la serie?

Suposiciones:

* Apenas un equipo logre ganar por 4ta vez, termina el partido.

$X \sim b^*(x;k,p)$, con $k=4$ y $p=p_A=0.55$; X: Duración en juegos del partido.

a) $P(X=6) = b^*(6;4,0.55) = \binom{5}{3}\cdot(0.55)^4\cdot(0.45)^2$

b) $P(A \text{ gane}) = \sum_{x=4}^{7} b^*(x;4,0.55) = \sum_{x=4}^{7}\binom{x-1}{3}\cdot(0.55)^4\cdot(0.45)^{x-4}$

c) $\displaystyle\sum_{x=3}^{5} b^*(x;3,0.55)$

* Poisson:

- Se asocia a experimentos que ocurren en un intervalo de tiempo Δt.
- Δt puede ser de cualquier duración
- Los resultados que ocurren en un Δt son independientes de lo que haya ocurrido en otro Δt.
- No tiene memoria
- La probabilidad de que ocurra un evento depende del largo del intervalo. (Escuchar a Milton hablar en una clase online de Mate para la ciencia es más probable que ocurra de 14:45 a 15:20, a que ocurra de 15:30 a 15:35)
- Insignificante la probabilidad de que ocurra un evento de otro tipo en el mismo Δt.

En Poisson, X representa el nº de resultados que ocurren en un Δt dado:

$$X \sim \text{poisson}(x;\lambda t) \longleftrightarrow \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad ; \quad x \in \mathbb{N}_0, \ \lambda t > 0$$

λ: Representa $\overline{v}$ con que ocurre un evento.

$$\boxed{\mathbb{E}(X) = V(X) = \lambda t}$$

$$e^{\alpha} = 1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \dots = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!}$$

$$\mathbb{E}(X) = \sum_{x\in\mathbb{N}_0} \frac{x e^{-\lambda t}(\lambda t)^x}{x!} = e^{-\lambda t}\sum_{x\in\mathbb{N}_0}\frac{x(\lambda t)^x}{x!} = e^{-\lambda t}\left(\cancel{\frac{0\cdot(\lambda t)^0}{0!}}^{\,0} + \sum_{x\in\mathbb{N}-\{0\}} \frac{\cancel{x}(\lambda t)^x}{\cancel{x}(x-1)!}\right)$$

$$= e^{-\lambda t}\sum_{x\in\mathbb{N}-\{0\}} \frac{\lambda t\cdot(\lambda t)^{x-1}}{(x-1)!} = \lambda t e^{-\lambda t}\sum_{x\in\mathbb{N}-\{0\}}\frac{(\lambda t)^{x-1}}{(x-1)!} = \lambda t e^{-\lambda t}\underbrace{\sum_{u=0}^{\infty}\frac{(\lambda t)^u}{u!}}_{\Rightarrow \lambda t}$$

$x-1=u$

$= \lambda t \cancel{e^{-\lambda t}} \cdot \cancel{e^{\lambda t}} = \lambda t$ (Para la varianza es lo mismo)

Sea X el número de criaturas de un tipo particular capturadas en una trampa durante un periodo determinado. Suponga que X tiene una distribución de Poisson con $\lambda = 4.5$, así que en promedio las trampas contendrán 4.5 criaturas [El artículo "Dispersal Dynamics of the Bivalve *Gemma Gemma* in a Patchy Environment (*Ecological Monographs*, 1995: 1–20) sugiere este modelo: el molusco bivalvo *Gemma gemma* es una pequeña almeja.] La probabilidad de que una trampa contenga exactamente cinco criaturas es

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = 0.1708$$

La probabilidad de que una trampa contenga cuando mucho cinco criaturas es

$$p(X \leq 5) = \sum_{x=0}^{5} \text{poisson}(x; 4.5) = \boxed{\sum_{x=0}^{5} \frac{e^{-4.5} \cdot (4.5)^x}{x!}}$$

La Poisson es un límite de la distribución binomial tomando $n \to \infty$ y $p \to 0$

El parámetro λ en tal caso sería np.

$$\lim_{\substack{n \to +\infty \\ p \to 0}} \underbrace{\binom{n}{x} \cdot p^x (1-p)^{n-x}}_{b(x;n,p)} = \text{poisson}(x; np) = \frac{e^{-np}(np)^x}{x!}$$ (Demostrar ...)

Regla segura de aproximación: $n > 50$; $np < 5$.

Si un editor de libros no técnicos hace todo lo posible porque sus libros estén libres de errores tipográficos, de modo que la probabilidad de que cualquier página dada contenga por lo menos uno de esos errores es de 0.005 y los errores son independientes de una página a otra, ¿cuál es la probabilidad de que una de sus novelas de 400 páginas contenga exactamente una página con errores? ¿Cuándo mucho tres páginas con errores?

* Como binomial (exacto):

X: Páginas con errores. ; $n = 400$; $p = 0.005$

$$P(X = 1) = \binom{400}{1} \cdot (0.005)^1 \cdot (0.995)^{399} \approx 0.2707$$

$$P(X \leq 3) = \sum_{x=0}^{3} \binom{400}{x} \cdot (0.005)^x \cdot (0.995)^{400-x} \approx 0.8576$$

* Como Poisson (aprox.):

X: Páginas con errores ; $\lambda t = (0.005)(400) = 2$

$$P(X=1) \approx \text{poisson}(1;2) = \frac{e^{-2} \cdot 2^1}{1!} = 2e^{-2} \approx 0.2707$$

$$P(X \leq 3) \approx \sum_{x=0}^{3} \frac{e^{-2} \cdot 2^x}{x!} \approx 0.8571$$

Distribuciones de probabilidad continuas

* Uniforme

$$X \sim U[a,b] \iff f(x; A, B) = \begin{cases} \frac{1}{B-A} ; & A \leq x \leq B \\ 0 ; & \text{c.c.} \end{cases}$$