Siméon Denis Poisson. Matemático francés, 1781-1840

Bioestadística (Clase 6):

Modelos Multinomial, Hipergeométrico y de Poisson

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Contenidos de la clase

Variables aleatorias discretas (repaso)

Distribución Multinomial

Distribución de Poisson

Distribución Hipergeométrica

Variables aleatorias discretas (repaso)

- Una variable aleatoria es una función que nos lleva del espacio Ω a $\mathbb{R}$
- Como estamos en $\mathbb{R}$ nos permite utilizar herramientas de cálculo
- La variable aleatoria es discreta cuanto toma un conjunto finito $\{0, 1, 2, \ldots, N\}$ o numerable $\{0, 1, 2, \ldots\}$ de valores.
- Nos da flexibilidad para considerar modelos de probabilidad no equiprobables.

Variable aleatoria uniforme

- Es una variable aleatoria X que toma los valores $1, \ldots N$.
- Asumimos que todos los valores son equiprobables.
- Tenemos

$$P(X = k) = \frac{1}{N}, \quad \text{para } k = 1, \ldots, N$$

Variable aleatoria de Bernoulli

- Nuestra variable toma dos valores:

$$X = \begin{cases} 0, & \text{que corresponde al fracaso} \\ 1, & \text{que corresponde al éxito} \end{cases}$$

de un experimento aleatorio con dos resultados posibles

- Los valores **no** son equiprobables. Para un $p \in (0,1)$:

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

Variable aleatoria Binomial

- La variable aleatoria binomial cuenta la cantidad de éxitos en una serie de n experimentos de Bernoulli independientes.
- Cada experimento tiene una probabilidad de éxito $p \in (0, 1)$, la variable binomial tiene parámetros n y p.
- Una variable binomial Y toma alguno de los valores $0, 1 \ldots, n$:

$$\boxed{P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n}$$

Variable geométrica

- Una variable aleatoria geométrica Y que cuenta el número de fracasos necesarios hasta el primer éxito es (por definición) de una serie arbitrariamente larga de experimentos de Bernoulli.
- La variable Y toma todos los valores naturales:

$$0, 1, \ldots, n, \ldots$$

- Si fracaso k veces: $P(Y = k) = P(\underbrace{0\ldots0}_{k}1) = (1-p)^k p.$

Distribución Multinomial

- Modela la repetición de n experimentos independientes con más de dos o más resultados posibles.
- En el caso de ser 2, tenemos la distribución Binomial.
- Supongamos que tenemos un experimento con tres resultados posibles, que etiquetamos A,B y C.
- Cada uno de los resultados tiene una probabilidad p, q, r que cumplen $p + q + r = 1$
- Nos preguntamos entonces por la probabilidad de tener x_1 resultados 0, x_2 resultados 1, x_3 resultados 2.
- Como son n experimentos tenemos $x_1 + x_2 + x_3 = n$

- Por ejemplo $n = 6$, $x_1 = 2$, $x_2 = 1$, $x_3 = 3$.
- Los resultados posibles entonces son tiras de la forma

$$A\ A\ B\ C\ C\ C$$

- Como son independientes, la probabilidad de esta tira es

$$P(A\ A\ B\ C\ C\ C) = p^2 q\, r^3.$$

- ¿Cuántas tiras disitintas tienen dos A, una B, y tres C?

- Para contarlas, primero ubico las A en los 6 lugares:

$$\_\ \underline{A}\ \_\ \underline{A}\ \_\ \_$$

- Tenemos $\binom{6}{2}$ formas de hacerlo
- Luego ubicamos la B en los 4 lugares restantes:

$$\_\ \underline{A}\ \underline{\color{red}{B}}\ \underline{A}\ \_\ \_$$

- Tenemos $\binom{4}{1}$ formas de hacerlo por cada una de las anteriores.
- Los restantes lugares se ocupan con las tres C:

$$\underline{\color{blue}{C}}\ \underline{A}\ \underline{\color{red}{B}}\ \underline{A}\ \underline{\color{blue}{C}}\ \underline{\color{blue}{C}}$$

- Si aplicamos la regla del producto, tenemos $\binom{6}{2}\binom{4}{1}$ formas diferentes de ubicar los resultados,

- Calculando

$$\binom{6}{2}\binom{4}{1} = \frac{6!}{2!4!} \times \frac{4!}{1!3!} = \frac{6!}{2!1!3!}$$

- El número que obtuvimos es el coeficiente de la probabilidad (es decir, la cantidad de tiras con dos A, una B y tres C), y se llama coeficiente multinomial, escribiéndose
$$\binom{6}{2,1,3} = \frac{6!}{2!1!3!}$$

- Nuestra conclusión es que
$$\begin{aligned} P(\text{dos A, una B, tres C}) &= P(x_1 = 2, x_2 = 1, x_3 = 3) \\ &= \frac{6!}{2!1!3!} p^2 q^1 r^3. \end{aligned}$$

Fórmula general

- Tenemos k resultados posibles de un experimento que se repite n veces
- Las probabilidades de los resultados de un experimento son $p_1, \ldots, p_k$, que verifican

$$p_1 + \cdots + p_k = 1.$$

- Nos interesa que de los n experimentos, el primer resultado salga x_1 veces, el segundo x_2 veces $\ldots$ el k-ésimo x_k veces, y se verifica

$$x_1 + \cdots + x_k = n.$$

- La probabilidad es

$$P((x_1, \ldots, x_k)) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}.$$

- Escrita de otra forma

$$P((x_1, \ldots, x_k)) = \frac{n!}{x_1! \cdots , x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

Ejemplo

- Tiramos 5 dados al jugar a la generala.
- ¿Cuál es la probabilidad de sacar un uno, un dos , y tres seis?

Solución

- Tenemos $n = 5$ experimentos con 6 resultados posibles cada uno.
- Cada uno de los 6 resultados tiene probabilidad 1/3.
- Queremos tiras de la forma

$$\underline{1}\ \underline{2}\ \underline{6}\ \underline{6}\ \underline{6}$$

- Para eso precisamos que no salgan ni tres, ni cuatros, ni cincos.
- Es decir, tenemos

$$x_1 = 1,\ x_2 = 1,\ x_3 = 0,\ x_4 = 0,\ x_5 = 0,\ x_6 = 3.$$

- Aplicando la probabilidad multinomial

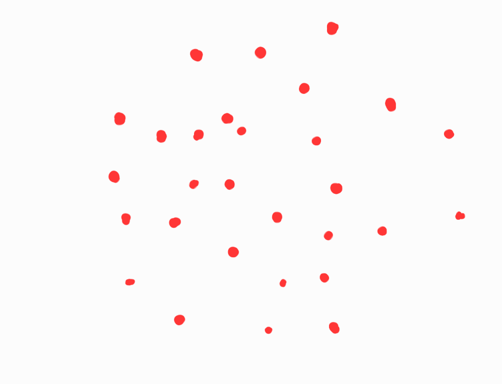
$$P(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 3)$$
$$= \binom{5}{1,1,0,0,0,3} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^0 \left(\frac{1}{6}\right)^3$$
$$= \frac{5!}{3!}\frac{1}{6^5} = 0{,}0026$$

Distribución de Poisson

Supongamos que estamos observando bacterias en una placa de Petri:

O elefantes en una foto aérea:

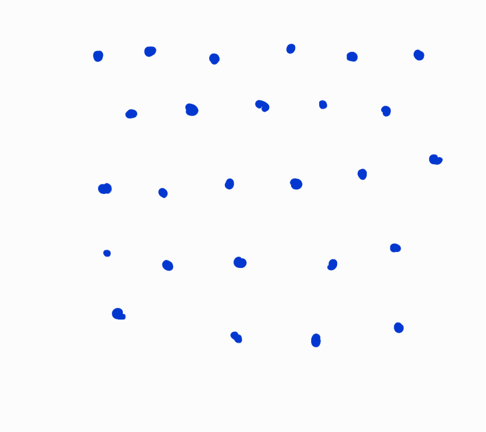
Digamos que tenemos una distribución de puntos en el plano:



Y nos interesa saber si los puntos entre sí

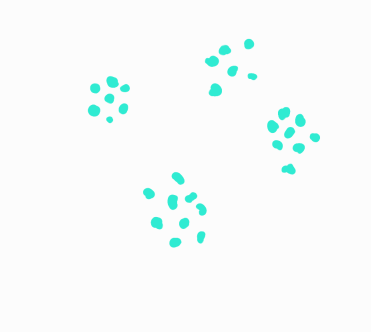
- Se atraen
- Se repelen
- Son indiferentes

Estos puntos se repelen:

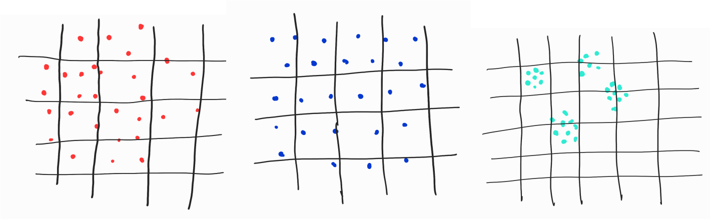


Podrían ser animales que compiten por recursos (comida, agua, etc.)

Estos puntos se atraen:



Para decidir que situación tenemos, cuadriculamos el plano:



# puntos	# cuadros
0	1
1	1
2	5
3	2
4	1
5	1
6	0
7	0
⋮	⋮

# Pts	# □
0	0
1	7
2	5
3	3
4	0
5	0
6	0
7	0
⋮	⋮

# Pts	# □
0	8
1	0
2	0
3	2
4	1
5	1
6	1
7	0
8	
⋮	

- De cantidades podemos pasar a frecuencias.
- De frecuencias a probabilidades
- La situación en que los puntos son indiferentes, dan las probabilidades de Poisson
- Una variable de Poisson X con intensidad λ toma los valores en los naturales $\{0, 1, \ldots, n, \ldots\}$ con probabilidades

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

- El λ representa la cantidad media de puntos por cuadro.

La distribución de Poisson se aplica en muchas situaciones:

- La cantidad de llamadas telefónicas a una central en un intervalo de tiempo
- La cantidad de estrellas de una cierta magnitud en el cielo
- La cantidad de clientes en un supermercado durante un lapso de tiempo
- La cantidad de tareas que recibe un procesador de computadora
- ⋮

Ejercicio

La cantidad de errores tipográficos en un texto sigue una distribución de Poisson, y son aproximadamente una por millón.

- El Quijote tiene 377 032 palabras y 1 687 570 caracteres
- Calcular la probabilidad de que en todo haya alguna errata.

Solución. El promedio de errores es

$$\lambda = 1,687570 \sim 1,7$$

Entonces

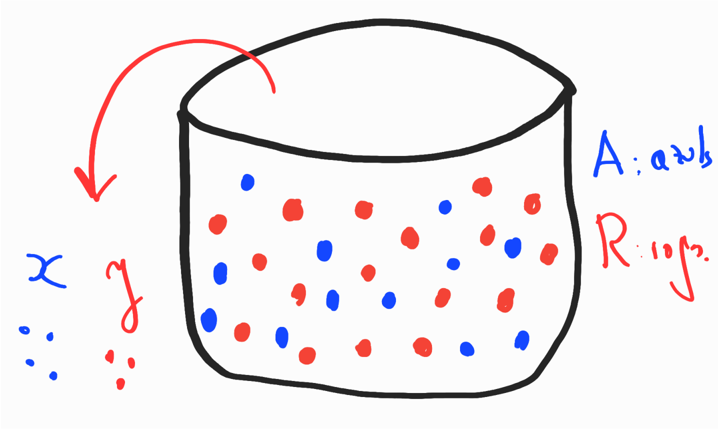
$$P(\text{cero error}) = P(X = 0) = e^{-\lambda} = e^{-1,7} = 0{,}18.$$

$$P(\text{alguna errata}) = 1 - P(X = 0) = 0{,}82.$$

Distribución Hipergeométrica

- Extraemos n bolas de una urna que tiene A bolas azules y R bolas rojas.

- ¿Cuál es la probabilidad de extraer exactamente x bolas azules, y bolas rojas?



- Extraemos $x + y$ bolas de un total de $A + R$
- Tenemos
$$\binom{A+R}{x+y}$$
extracciones distintas, cada una es un caso posible.
- Los favorables son cuando hay exactamente x azules, y rojas.
- Como hay un total de A azules, podríamos seleccionar
$$\binom{A}{x}$$
bolas azules, cada una conformaría un caso favorable

- A su vez, para cada configuración azul, preciso extraer y rojas entre las R posibles. Eso me da

$$\binom{R}{y}$$

posibilidades.

- Como cada posibilidad de las azules y las rojas se combina, aplicando la regla del producto, tengo

$$\binom{A}{x}\binom{R}{y}$$

casos favorables.

- La probabilidad buscada es la distribución hipergeométrica:

$$P((x,y)) = \frac{\binom{A}{x}\binom{R}{y}}{\binom{A+R}{x+y}}$$

Ejemplo Un mazo de 52 cartas tiene la mitad rojas y la mitad negras. ¿Cual es la probabilidad, al extraer 10 cartas, de obtener exactamente 5 rojas y 5 negras?

Solución

$$P((5,5)) = \frac{\binom{26}{5}\binom{26}{5}}{\binom{52}{10}} = 0{,}27.$$