3. VARIABLES ALEATORIAS

Repasar primero las notaciones de función, imagen y preimagen.

Introducción

Ejemplo 1: tiro dos veces una moneda. Me interesa el nro de veces que sale cara.

Ej. 2: idem. Me interesa la resta entre el nro de caras y el nro de cecas.

Ej. 3: elijo un alumno al azar entre todos los alumnos de 1er año de la fac. y le pregunto su nota en Matematicas del CBC.

Ej. 4: Elijo un hombre adulto al azar en Buenos Aires y observo su peso y su estatura.

Como se puede plantear modelos probabilísticos para estos ejemplos?

Ej. 1 Podriamos considerar S={0,1,2}, pero tambien podemos dejar
S={CC,CS,SC,SS} describiendo todos los resultados de tirar dos monedas y luego asociar

$$CC \rightarrow 2, \quad CS \rightarrow 1, \text{ etc.}$$

Se puede considerar S={CC,CS,SC,SS} y una función X="nro de caras" que a cada elemento de S le asigna un nro:

$$\begin{array}{ccc} & X & \\ CC & \rightarrow & 2 \\ CS & \rightarrow & 1 \\ SC & \rightarrow & 1 \\ SS & \rightarrow & 0 \end{array}$$

Ej. 2, podemos considerar el mismo S y otra función.

Definición de variable aleatoria:
Sea S un espacio muestral Una variable aleatoria es una función de S en los números reales.

Generalmente las variables aleatorias se las denota con las ultimas letras del alfabeto en mayuscula (X, Y, etc.) Por definición si X es una v.a.

$$\mathbf{X: S \rightarrow R}$$

En el ej. 1: Qué entendemos por P(X=1)?

En general si $B \subseteq R$,

$$\mathbf{P(X \in B) = P\{s \in S|\ X(s) \in B) = P(X^{-1}(B))}$$

Ejemplo: cuanto vale $P(X \leq 1)$en el ejemplo 1?

Imagen de una variable aleatoria:
Usaremos la notación I_X para la imagen de la variable aleatoria (conjunto de valores que toma).

3.1. Variables aleatorias discretas

Definición de variable aleatoria discreta:
Una variable aleatoria es discreta si I_X es un conjunto finito o infinito numerable.

Dar ejemplos de variables aleatorias discretas
- que tomen un numero finito de valores
- que tomen un numero infinito numerable de valores

Función de probabilidad puntual de una v.a. discreta.

Sea X v.a. discreta. Se llama f.p.p. a la función

$$p(x) = P(X=x) \quad \text{para todo } x \in R$$

En el ejemplo 1:

$$p(0)=1/4$$
$$p(1)=1/2$$
$$p(2)=1/4$$
$$p(x)=0 \text{ para todo } x \notin \{0,1,2\}$$

Propiedades de las funciones de probabilidad puntual
Son funciones $p:R \rightarrow R$ que cuplen dos propiedades:

a) $p(x) \geq 0$ para todo $x \in R$

b) $\sum_{x \in I_X} p(x) = 1$

Nota: la sumatoria también puede extenderse al conjunto $\{x | p(x)>0\}$

Estas dos propiedades a) y b) caracterizan a las funciones de probabilidad puntual.

Comentario:
Para dar una función de probabilidad puntual, a veces se dá una fórmula, otras una tabla:

x	0	1	2
p(x)	1/4	1/2	1/4

y también puede mostrarse con un gráfico (con líneas o con barras como el histograma).

Comentario:
Si conozco la f.p.p. de una v.a. X, como puedo calcular P(B) para cualquier $B \subset R$?

Otros ejemplos:
- Elijo un alumno al azar entre todos los alumnos de 1er año de la facultad. Sea X su nota en Matematicas del CBC. Como se podria conocer la f.p.p. de esta v.a.?

- Se tira un dado hasta que sale as por primera vez. Sea X=nro de tiros. Es X discreta? Cuánto vale I_X? Cual es la f.p.p. de X?

Comentario:
Conocer la distribución de una v.a. quiere decir poder calcular
$P(X \in B)$ para cualquier $B \subset R$.
Observar que si conozco p(x) conozco la distribución, ya que

$$P(X \in B) = \sum_{x \in B \cap I_X} p(x) \qquad (5)$$

Observar que variables aleatorias distintas pueden tener la misma distribucion. Por ej. la variable X=nro de caras al tiras dos monedas, tiene la misma distribución que el nro de cecas.

<u>Función de distribución o función de distribución acumulada de una v.a.</u> <u>Definición</u>

Sea X una v.a. Su función de distribución (acumulada) es la función
$$F(x) = P(X <= x) = P(X \in (-\infty, x]) \quad (\text{definida } \forall\ x \in R)$$

La definición anterior vale para cualquier v.a., sea discreta o no.
Para el caso particular de las v.a. discretas, se puede calcular sumando la f.p.p. para todos los valores $\leq x$:

$$F(x) = \sum_{u \in I_X,\ u \leq x} p(u) \qquad (6)$$

Ejemplo 1: Calcular y graficar la función de distribución de X = nro de caras al tirar dos veces una moneda.

Ejemplo 2: Como sería la función de distribución de X = nro de veces que tengo que tirar un dado hasta que salga as por primera vez?

<u>Propiedades de las funciones de distribución</u>.

Las funciones de distribución son funciones reales ($F: R \to R$) que cumplen:

1) Son monótonas no decrecientes.
2) $\lim_{x \to +\infty} F(x) = 1$

3) $\lim_{x \to -\infty} F(x) = 0$

4) Son continuas a derecha (o sea $\lim_{x \to x_0^+} F(x) = F(x_0)$)

Es evidente de los ejemplos que las funciones de distribucion no tienen por que ser continuas ya que $\lim_{x \to x_0^-} F(x) = F(x_o^-)$ no tiene por que coincidir con $F(x_o)$

(la funcion de distribucion puede dar saltos).

Si la funcion de distribución da un salto en el punto x_o, entonces la altura del salto es precisamente el valor de la probabilidad en el punto:

$$\mathbf{P(X = x_o)} = F(x_0) - F(x_o^-) \ \textbf{= altura del salto} \qquad (7)$$

Propiedad importante:

Si conozco la función de distribución de una v.a. puedo calcular cualquier $P(X \in B)$ para cualquier $B \subset R$.

a) En particular es facil calcular la probabilidad de un intervalo de la forma (a,b] (asi abierto a izquierda y cerrado a derecha), ya que es facil demostrar que:

$$\mathbf{P(a < X \leq b) = F(b) - F(a)} \qquad (8)$$

Demostración:

$$(a,b] = (-\infty,b] - (-\infty,a]$$

luego:

$$P(X \in (a,b]) = P(X \in (-\infty,b]) - P(X \in (-\infty,a])$$

o sea

$$P(a < X \leq b) = F(b) - F(a)$$

Pero como se puede calcular $P(a \leq X \leq b)$?
Usando (8) y (7) puede calcularse.

Ejercicio: escribir expresiones para calcular:

$$P(a \leq X \leq b)$$
$$P(a < X < b)$$
$$P(a \leq X < b)$$
$$P(a \leq X)$$
$$P(a < X)$$
$$P(X < b)$$

conocida la Función de distribución acumulada.

3.2. Esperanza de una variable aleatoria discreta

La esperanza se puede definir para cualquier tipo de variable aleatoria. Pero por ahora veremos el caso particular de v.a. discreta.

Definición de esperanza para una v.a. discreta.

Sea X una v.a. discreta, I_X su imagen, p su f.p.p.
Se define

$$E(X) = \sum_{x \in I_X} x\, p(x)$$

E(X) se lee 'esperanza de X'.

Ejemplo 1: Sea X el nro. de caras al tirar dos veces una moneda equilibrada. Hemos visto que $I_X = \{0,1,2\}$ y que

x	0	1	2
p(x)	1/4	1/2	1/4

Entonces $E(X) = \sum_{x=0}^{2} x\, p(x) = 0*(1/4) + 1*(1/2) + 2*(1/4) = 1$

Ejemplo 2: Sea X el nro. de ases al tirar dos veces un dado equilibrado. Entonces, igual que en el ej. 1 es $I_X = \{0,1,2\}$, mientras que la f.p.p. es

x	0	1	2	3
p(x)	25/36	10/36	1/4	1/36

Entonces $E(X) = 0*(25/36) + 1*(10/36) + 2*(1/36) = 12/36 = 1/3$

Se ve en este ejemplo (no en el anterior) que E(X) no tiene por qué coincidir con ninguno de los valores que toma la v.a.

E(X) se puede interpretar como un promedio ponderado de los distintos valores que toma la variable, dandole mayor peso a los valores más probables. La definición coincide con la de centro de gravedad en física.

Vamos a ver ahora otras dos interpretaciones intuitivas de E(X).

Una interpretación intuitiva de la esperanza: como media de una población.

Esta interpretación no siempre tiene sentido sólo en algunos ejemplos (en los que el experimento aleatorio consiste en extraer al azar un individuo de una población finita). Consideremos el siguiente ejemplo. Supongamos que tenemos registrados todas las notas en Matemáticas del examen de ingreso de los alumnos de 1er. año. Estas notas van del 4 a 10 y son las siguientes:

6, 8, 9, 8, 4, 10, 6, etc.

Llamemos N al nro. de alumnos de 1er. año. Entonces la media (poblacional) de las notas es:

$$\mu = \frac{6+8+9+8+4+......}{N}$$

El titulo de la “Interpretación” dice que la E(X) es la media poblacional μ. Como podemos definir una v.a. con las notas?

Experimento aleatorio: entre todos los alumnos de 1er. año selecciono uno al azar (o sea de la población elijo una “muestra” de un solo alumno).
Luego observo X = nota del alumno seleccionado.

Quién S? Es X una v.a.? Cuál es su imagen? Cuál es su f.p.p.?
Cuánto vale E(X)?

Esta interpretación no sólo vale para este ejemplo, sino para cualquier ejemplo en que hay una población de un nro. finito de individuos y para cada individuo se puede observar una variable discreta. Si se selecciona al azar un individuo y se considera la v.a. X = ’valor de la variable en el individuo elegido al azar’ entonces E(X) coincide con la media poblacional.

Pero esta interpretación no siempre tiene sentido, por ejemplo no lo tiene en los ejemplos 1 y 2.

<u>Otra interpretación intuitiva de la esperanza.</u>

Esta interpretación **vale para cualquier v.a.** Sin embargo, sólo para simplificar, vamos a pensar en una v.a. discreta que sólo toma un nro. finito de valores.

Sea $I_X = \{x_1, x_2, \dots, x_k\}$ la imagen

y

x	x_1	x_2	...	x_k
p(x)	$p(x_1)$	$p(x_2)$		$p(x_k)$

la tabla que dá la f.p.p.

Supongamos que se repite n veces (en las mismas condiciones) el experimento aleatorio. Cada vez que se repite el experimento se observa el valor que toma la variable X. Se anotan esos valores, por ejemplo:

$$x_3 \quad x_5 \quad x_2 \quad x_5 \quad x_3 \quad x_4 \quad \text{etc}$$

y se calcula el promedio de todos los valores que salieron

$$\frac{x_3 + x_5 + x_2 + x_5 + x_3 + x_4 + \dots}{n} \qquad (9)$$

Llamando n_i = nro. de veces que sale x_i (i=1,...,k), reordenando los términos (9) puede escribirse así

$$\frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \ldots + \frac{n_k}{n} x_k \qquad (10)$$

Qué ocurre con el cociente $\frac{n_i}{n}$ cuando n es grande?

Por la idea intuitiva de probabilidad, sabemos que

$$\lim_{n\to\infty} \frac{n_i}{n} = P(X = x_i) = p(x_i)$$

Luego la expresión (10) cuando $n\to\infty$ tiende a

$$p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_k)x_k = E(X)$$

O sea esta **interpretación de la esperanza** se puede enunciar así**:**

"Si se repite n veces el experimento aleatorio y se calcula el promedio de todos los valores que toma la v.a., este promedio tiende a E(X) cuando n tiende a infinito."

Es lo que hicimos recién una demostración rigurosa de la afirmación anterior?

Qué quiere decir esta interpretación en el ejemplo 1 (nro de caras al tirar 2 veces una moneda)?

Qué quiere decir en el ejemplo de elegir un alumno al azar y observar su nota?

Otro ejemplo:
Juguemos 1$ en la ruleta a colorado. Sea X=ganancia.
Calcular E(X) y ver que quiere decir aquí la interpretación intuitiva.
Es buen negocio jugar mucho a la ruleta?

Funciones de variables aleatorias.

- Si X es una v.a. Qué es X^2?

- En general si $g:R\to R$, g(X) es una v.a. (es la composición de la función X con la función g).

$$\mathbf{S} \xrightarrow{X} \mathbf{R} \xrightarrow{g} \mathbf{R}$$

- Ejemplo: Sea X el nro. de caras al tirar dos veces una moneda. Sea $Y=X^2$
  a) Calcular la f.p.p. de Y (para distinguir llamaremos p_X y p_Y). b) Calcular E(Y).
  c) Luego observar que:

$$E(Y)=\sum_{y\in I_Y} y\,p_Y(y)=\sum_{x\in I_X} x^2\,p_X(x)$$

En general:

Teorema: Sea X una v.a. discreta, I_X su imagen, p_X su f.p.p. Sea $g:R\rightarrow R$. Entonces:

$$E(g(X))=\sum_{x\in I_X} g(x)p_X(x) \qquad (11)$$

Observar que este teorema tiene la utilidad de que permite calcular la esperanza de la v.a. g(X), conocida la función de probabilidad puntual de la v.a. X, sin necesidad de calcular primero la f.p.p. de la v.a. g(X).

Propiedades de la Esperanza (estas propiedades valen no sólo para variables aleatorias discretas, sino para cualquier variable aleatoria).

a) Sea X una v.a., a y b números reales. Entonces

$$\mathbf{E(a\,X + b) = a\,E(X)+ b}$$

Casos particulares:

$E(aX) = a\,E(X)$

$E(X+b) = E(X) + b$

b) Si X e Y son dos variables aleatorias entonces

$$\mathbf{E(X + Y) = E(X) + E(Y)}$$

b*) Generalizando b) si $X_1, X_2, \ldots X_n$ son vs. as. entonces

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Vamos a definir varianza de una v.a.

Notaciones para medias y varianzas de un conjunto de datos.

En estadística descriptiva, definimos media y varianza. Muchas veces el investigador considera que los datos observados son una muestra de una población y desea, a partir de los datos de la muestra observada, hacer inferencias sobre la población. La inferencia estadistica estudia técnicas para hacer estas inferencias. Cuando queremos hacer inferencias, conviene distinguir con notaciones distintas a la media y varianza de la muestra y a la media y varianza de la poblacion. Las notaciones usuales son:

	Población	Muestra
Media	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Varianza	$\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	$s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

Luego hemos definido esperanza (también llamada media) de una v.a. discreta:

$$E(X) = \sum_{x \in I_X} x\ p(x)$$

Como se define Varianza de una variable aleatoria?

Varianza de una variable aleatoria

La varianza se puede definir para cualquier tipo de variable aleatoria. Pero por ahora veremos el caso particular de variables aleatorias discretas.

Definición de varianza para una v.a. discreta.

Sea X una v.a. discreta, I_X su imagen, p su f.p.p.
Se define

$$Var(X) = \sum_{x \in I_X} (x - E(X))^2\ p(x)$$

Var(X) se lee 'varianza de X'.

Otra forma de definir varianza de una variable aleatoria:
Usando el teorema que da la E(g(X)) se puede escribir:

$$\mathbf{Var(X) = E(X\text{-}E(X))^2} \qquad (12)$$

Luego veremos que esta expresion se usa también para definir varianza para variables aleatorias que no son discretas. La expresión (12) es la definicion más general de Var(X)

Comentario: cuando la variable aleatoria se define extrayendo al azar un individuo de una población finita, como en el ej de la nota de una alumno elegido al azar) la varianza de la v.a. es igual a al varianza poblacional.

Interpretación de la varianza como medida de dispersión

Ejemplo 1: X = nro. de caras al tirar dos monedas. Calcular Var(X).
(respuesta: 0.5)

Ejemplo 2: Si una v.a. Y tiene f.p.p.

y	0.9	1	1.1
$P_Y(y)$	1/4	1/2	1/4

Entonces, comparando con la v.a. X del ej. 1 se observa que Y tiene la misma esperanza (E(Y)=E(X)=1),

Var(Y) es <, = o > que Var(X) ?

Ejemplo 3: Sea Z una v.a. con f.p.p.

z	0	1	2
$P_Z(z)$	0.01	0.98	0.01

Que también tiene esperanza 1,

Var(Z) es < = o > que Var(X) ?

Otra expresión para Var(X).

$$\mathbf{Var(X) = E(X\text{-}E(X))^2 = E(X^2) - (E(X))^2} \qquad (13)$$

Va

La primera igualdad es la definición de esperanza, la segunda igualdad se demuestra facilmente usando las propiedades a) y b) de la Esperanza.

De (11) y (13), para el caso de las variables discretas, hay dos formas de calcular la varianza:

$$Var(X) = \sum_{x \in I_X} (x - E(X))^2 \, p(x) = \sum_{x \in I_{Xx}} x^2 \, p(x) - (E(X))^2$$

La primera expresión (la definición) se entiende más (se entiende que es una medida de dispersión) la segunda era más simple para calcular, cuando las cuentas se hacían a mano.

Ejemplo: En el ejemplo 1, calcular Var(X) con esta nueva expresión y comprobar que da lo mismo.

Antes de enunciar propiedades análogas para la varianza, necesitamos definir el concepto de variables aleatorias independientes.

Definición: Dos variables aleatorias X e Y son **independientes** si

$$P((X\in A) \cap (Y \in A)) = P(X\in A)\, P(Y\in B)$$

para todo $A\subset R$, $B\subset R$.

Ejemplos:
i) Si elijo una persona al azar y observo X=peso, Y=estatura, son variables aleatorias independientes?
ii) Si juego dos veces en la ruleta y X=nro que sale en la 1a tirada, Y=nro que sale en la segunda tirada, son independientes?
iii) Dar otros ejemplos de vs. as. que (intuitivamente) sean independientes y no lo sean.

Recordemos las propiedades de esperanza que enunciamos:

a) $E(a X + b) = a E(X)+ b$
b) $E(X + Y) = E(X) + E(Y)$
b*) $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

Propiedades de la varianza:

a) $\mathbf{Var(a X + b) = a^2 Var(X)}$

y sus casos particulares

$$\mathbf{Var(aX) = a^2 Var(X)}$$
$$\mathbf{Var(X+b) = Var(X)}$$

No vale en general para varianzas una expresión similar a b) o b*). Pero sí vale para el caso particular de que las variables aleatorias sean independientes.

b) Si X e Y son dos variables aleatorias independientes entonces (no demostramos por ahora esta propiedad)

$$\mathbf{Var(X + Y) = Var(X) + Var(Y)}$$

b*) (generalización de b) Si $X_1, X_2,... X_n$ son vs. as. independientes entonces

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Recordemos que
conozco la f.p.p. de un v.a. discreta $\Leftrightarrow$ conozco la distribución de esa variable

Algunas distribuciones discretas usadas como modelos.
Vamos a ver ahora algunas f.p.p. que son las "más usadas":

Las distribuciones:
i) Binomial (y su caso particular la distribución de Bernoulli)
ii) de Poisson.

<u>Distribución de Bernoulli</u>:
Un caso particularmente simple, pero que a pesar de ello se usa muchas veces es el de las variables aleatorias que solo toman dos valores 0 y 1. De una v.a. así se dice que tiene distribución de Bernoulli. Si llamamos p a la P(X=1), su función de probabilidad es de la forma:

x	0	1
p(x)	1-p	p

Su esperanza y varianza valen: E(X) = p, Var(X) = p(1-p) (demostrarlo).

<u>Distribución Binomial</u>

Ejemplo previo nro. 1: Se tira 5 veces un dado equilibrado. X es el número de veces que sale as.
Ej. previo nro. 2: Un jugador juega 10 veces a "primera docena" en la ruleta. Sea X el número de veces que gana.

En estos dos ejemplos y en muchos otros el experimento aleatorio cumple las siguientes condiciones:
1) El experimento aleatorio consiste en repetir n veces una prueba (el número n es fijo)
2) En cada prueba se observa si sale o no un evento. Llamemos E al evento que nos interesa y F a su complemento. (La letra E viene de "éxito" y la F de "fracaso").
3) Las n repeticiones se hacen en iguales condiciones. O sea que P(E) es la misma en todas las pruebas. Llamemos p=P(E).
4) Las repeticiones se realizan en forma independiente. O sea el resultado de una prueba no influye sobre las otras.

Sea X la variable aleatoria:

X = número de veces que sale E en las n pruebas.

Nos va a interesar calcular la f.p.p. de X.

-Definición: Un experimento aleatorio que cumple las cuatro condiciones enunciadas se llama **Experimento Binomial**. Cada prueba se llama **Ensayo o Prueba de Bernoulli.**

- Veamos que los experimentos previos 1 y 2 son experimentos binomiales. Quién es n, E y p en cada uno?

Calculemos ahora la f.p.p. de X.
Se deduce que

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x \in \{0,1,2,...,n\} \\ 0 & \text{en caso contrario} \end{cases} \qquad (14)$$

<u>Definición</u>: Si una v.a. X tiene la función (14) como f.p.p. entonces se dice que **X tiene distribución Binomial** con parámetros n y p. Se escribe abreviando:

$$\mathbf{X \sim Bin\ (n,p)}$$

Ejemplos: En el ejemplo 1, X ~ Bin (5, 1/6)
en el ejemplo 2, X ~ Bin (10, 12/37)

Si deseamos en el ej. 1, calcular la probabilidad de que salga as tres veces, no necesitamos pensar especialmente en este ejemplo, sino simplemente aplicar la expresión (14)

$$P(X=3) = p(3) = \binom{5}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2$$

Esta expresión puede evaluarse con una calculadora o con un programa (antiguamente se usaban tablas).

Con el programa Statistix se puede calcular ingresando a "Statistis", "Probability Functions" y elegir la función "Binomial(x,n,p)".
El programa no calcula p(x), sino la función de distribución acumulada F(x)=P(X≤x). Por ejemplo si ingreso n=5, p=0.1666667 (esto es 1/6), x=3, el resultado es:

Binomial(3,5,0.1666667) = 0.99666

y si luego pongo x=2 obtengo

Binomial(2,5,0.1666667) = 0.96451

Como se puede a partir de estos resultados obtener P(X=3)?

Otro ejemplo de aplicación de la distribución binomial:

Un bolillero contiene N bolitas, de las cuales B son blancas. Se extraen n bolitas (sin o con?) reposición y se cuenta X= número de bolitas blancas extraídas.
Qué distribución tiene X?

Comentarios:

- Si las extracciones son **sin reposición**, X no es binomial, por que? Pero si N es "grande" en relación a n, las extracciones con o sin reposición dan resultados similares, y la distribución binomial es una buena aproximación a la distribución de la variable X. Como receta práctica si **N/n>20 la aproximación binomial es muy buena.**

- Hay muchos problemas prácticos similares a la extracción de bolitas. Son todos los problemas en los que se eligen n individuos al azar de una población de individuos de los cuales algunos tienen cierta característica y otros no. Si se cuenta el número de individuos con la característica entre los n seleccionados, ese número se puede considerar una observación de un v.a. con distribución binomial.

Comentario:

$$X \sim Bin\ (1,p) \Leftrightarrow X \sim Bernoulli(p)$$

<u>Esperanza y varianza de una v.a. con distribución binomial</u>.

Sea X ~ Bin(n,p). Entonces

$$\mathbf{E(X) = np}$$

$$\mathbf{Var(X) = np(1-p)}$$

Demostrarlo.

Distribución de Poisson

<u>Definición</u>: Se dice que una v.a. **X tiene distribución de Poisson** con parámetro λ si su función de probabilidad puntual es:

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{si } x \in \{0,1,2,...\} \qquad (15)$$

"X tiene distribución de Poisson con parámetro λ" se abrevia así: **X** ~ ***P(λ)***

Comentario: La binomial fue derivada a partir de un experimento aleatorio, que consistía en repetir una prueba. No existe un experimento simple que permita deducir la f.p.p. de Poisson. Como se le ocurrió a Poisson esta f.p.p.? La obtuvo como límite de la f.p.p. binomial, como veremos luego.

Para que sirve? Sirve como modelo aproximado para las f.p.p. de variables aleatorias del siguiente tipo:

a) Número de pulsos que cuenta un medidor de radiación en un intervalo de longitud fijada (digamos de 3 minutos). La distribución de Poisson dá una muy buena aproximación a la distribución de esta variable y por lo tanto a la variabilidad de la medición.
b) Número de accidentes que ocurren en una semana en la ruta Bs.As.-Córdoba.
c) Número de llamadas telefónicas que llegan a un conmutador entre las 12 y las 12 y 15.
d) Número de clientes que entran a un negocio entre las 14 y las 15.
e) Número de parásitos que cuenta un laboratorista en un volumen (fijo) de sangre. (Me refiero a la variabilidad de la medición, o sea lo que tiene distribución de Poisson son repetidas mediciones para la misma sangre, no para sangre de distintos individuos).

Luego discutiremos por qué o bajo qué suposiciones este tipo de variables aleatorias tienen una f.p.p. que se puede aproximar por la de Poisson.

Demostrar que (15) cumple las propiedades de una f.p.p.

<u>Esperanza y Varianza</u>:

Si X ~ ***P(λ)*** entonces **E(X) = λ** y **Var(X) = λ**

Demostrarlo.

Nos habíamos preguntado como había surgido la f.p.p. llamada de Poisson? Acá está la respuesta: se dedujo como límite de la binomial.
Se puede demostrar que la f.p.p. binomial se parece a la de Poisson si n es "grande" y p "pequeño":

$$\binom{n}{x} p^x (1-p)^{n-x} \cong \frac{\lambda^x}{x!} e^{-\lambda} \qquad \text{con } \lambda = np$$

Receta: la aproximación es buena cuando $n \geq 100$, $p \leq 0.01$ y $np \leq 20$.

Pero aún para n no tan grandes ni p tan pequeños la aproximación puede ser bastante satisfactoria, como se ve en el siguiente ejemplo.

Ejemplo: Sea X~Bin(n=20,p=0.05).

$$P(X=2) = \binom{20}{2} 0.05^2 0.95^{18} = 0.189$$

Aproximando la f.p.p. binomial por la de Poisson con $\lambda=np=20*0.05=1$:

$$P(X=2) \cong \frac{1^2}{2!} e^{-1} = 0.184$$

Otra forma de deducir la distribución de Poisson: "Proceso de Poisson".

Otra forma de deducir la distribución de Poisson es considerar eventos que ocurren al azar en cualquier momento del tiempo. Un ejemplo clásico es el de los pulsos que cuenta un medidor de radiactividad. Si se hacen tres suposiciones sobre la forma en la que ocurren los eventos, se puede demostrar que el número de eventos en un intervalo de tiempo de longitud fija es una v.a. con distribución de Poisson.

Proceso de Poisson:
Ocurren eventos al azar en cualquier instante del tiempo, de modo que se cumplen las siguientes suposiciones:
1. Existe un parámetro $\alpha>0$ tal que la probabilidad de que ocurra un evento en un intervalo de longitud Δt (con Δt "pequeño") es aproximadamente $\alpha.\Delta t$ (o sea es proporcional a la longitud del intervalo).
Escrita rigurosamente la suposición es:

$$P(\text{ocurra un intervalo entre t y } t+\Delta t)= \alpha.\Delta t + o(\Delta t)$$

Comentario: la notación o(h) significa que es una función de h que tiende a cero más rápido que h, en el sentido siguiente:

$$\lim_{h\to 0} \frac{o(h)}{h} = 0$$

2. La probabilidad de que ocurran dos o más eventos en un intervalo de longitud "pequeña" Δt es despreciable. Rigurosamente

$$P(\text{ocurra dos o más eventos entre } t \text{ y } t+\Delta t) = o(\Delta t)$$

3. Sean $t_1<t_2<t_3<t_4$. Entonces el número de eventos que ocurren en el intervalo $[t_3,t_4]$ es independiente del número de eventos que ocurren en el intervalo $[t_1,t_2]$.

Proposición (no la demostraremos):
Si se cumplen las tres suposiciones del proceso de Poisson, se puede demostrar que si $[t_1,t_2]$ es cualquier intervalo y llamo

X = número de eventos en el intervalo $[t_1,t_2]$

entonces $\mathbf{X \sim \mathit{P}(\alpha\,(t2-t1))}$

Observación: De la proposición anterior y del hecho de que el parámetro de la Poisson es su esperanza, deducimos que, bajo las suposiciones 1 a 3, el valor esperado de eventos en un intervalo es α por la longitud del intervalo. Por lo tanto el significado intutivo de α es "el valor medio de eventos que ocurren en una unidad de tiempo". Obviamente el valor de α depende de la unidad en la que midamos el tiempo.

Ejemplo: Los clientes llegan a un negocio siguiendo un proceso de Poisson con una media de 12 clientes por hora. Cuál es la probabilidad de que lleguen por lo menos dos clientes en un intervalo de 15 minutos?
Respuesta: 0.801 (se puede obtener con el Statistix).
En que forma tienen que llegar los clientes para que se cumplan aproximadamente las suposiciones del proceso de Poisson?

Extensiones del proceso de Poisson:
Hemos definimos el proceso de Poisson en la forma clásica en la que los eventos ocurren en distintos instantes del tiempo. Pero la misma idea puede ser aplicada también a eventos que se pueden observar en distintos puntos de una región de dos o tres dimensiones (ejemplo: arboles en un campo, o parásitos en una gota de sangre). Se pueden formular suposiciones similares a las anteriores 1 a 3 y demostrar que si

X= número de eventos en un subconjunto A de la región

entonces

$$\mathbf{X \sim \mathit{P}(\alpha * volumen(A))}$$

(reemplazando volumen por area si la region es en dos dimensiones). El significado intuitivo de **α es en estos casos** "el valor medio de eventos que ocurren en una unidad de volumen (o en una unidad de área)".