Tema 6
Algunas distribuciones importantes

Hugo S. Salinas

Distribución binomial

Se han estudiado numerosas distribuciones de probabilidad que "modelan" características asociadas a fenómenos que se presentan frecuentemente en la realidad.

Una distribución de probabilidad de variable aleatoria discreta que se usa frecuentemente es la **distribución binomial.**

A partir de una experiencia en la que ocurre uno de dos resultados posibles: **A** o **A'** (llamada experiencia de **Bernoulli**) y ante **n** repeticiones independientes de la misma, se define la variable binomial:

X : nº de veces que ocurre el resultado A en las n repeticiones.

Hipotesis de la distribucion binomial:

1.- Existen **n** repeticiones idénticas que conducen a uno de dos resultados**: éxito o fracaso** (A y A').

2.- La probabilidad de cada resultado permanece constante de repetición en repetición. La probabilidad de uno de estos resultados, llamado éxito, se designa por **p (p=**P(A)).

3.- Las repeticiones son independientes.

Un ejemplo de un experimento binomial es el de lanzar una moneda al aire varias veces. Sólo hay dos resultados posibles en cada tirada (o repetición de la experiencia) de la moneda: cara o sello. La probabilidad de obtener cara o sello se mantiene constante de tirada en tirada (0.5 para cada una) y las tiradas son independientes entre sí.

Distribución binomial cont.

Ejemplo.

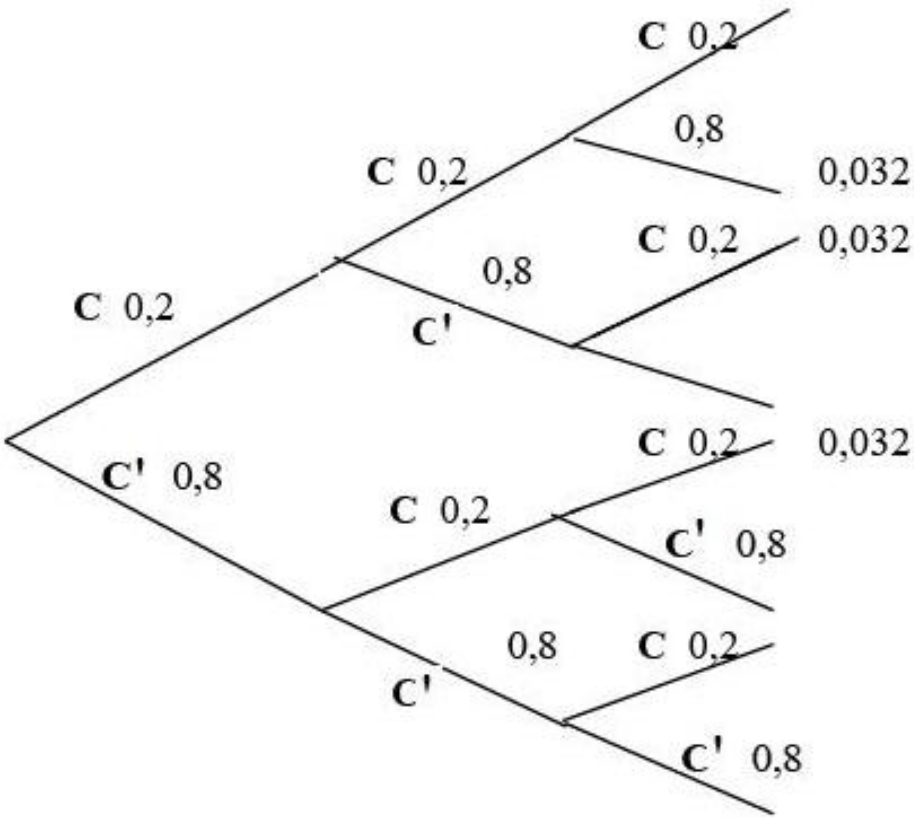
Suponga que 0.2 es la probabilidad de que una persona, que se conecta a un sitio específico en un centro comercial en la red internet, compre un artículo. Si el sitio tiene en un momento determinado tres personas que se han conectado, ¿cuál es la probabilidad de que exactamente dos personas compren un artículo?

Este es un ejemplo que se puede **modelar por binomial**, ya que:

- Sólo hay dos resultados posibles ante cada conexión a la red de una persona: compra (C) o no compra (C’).
- La probabilidad de comprar o no comprar se mantiene estable, con una probabilidad histórica de ocurrencia de 0.2 y 0.8 respectivamente.
- Las conexiones se consideran independientes entre sí.

Un método gráfico útil para visualizar esta situación es el **diagrama de árbol que aparece** en la siguiente figura:

Distribución binomial cont.



Cada rama en la figura representa la conexión de una persona (recordar que se consideran tres). La regla de la multiplicación se usa para calcular la probabilidad de cada manera independiente en la que pueden ocurrir dos compras. Observar que cada manera tiene la misma probabilidad (0.032). La regla de la suma se usa después para calcular la probabilidad final de que dos personas de tres que se conectaron al sitio realicen una compra (0.096).

Distribución binomial cont.

La primera parte de este cálculo indica el número de maneras (3) en las que puede ocurrir el resultado deseado (dos personas compran un artículo) .
El segundo cálculo (0.2 x 0.2 x 0.8) indica la probabilidad de lograr este resultado usando una de las trayectorias posibles en el diagrama.
El resultado final (0.096) es la probabilidad de que al conectarse tres personas al sitio, dos compren un artículo.
Este cálculo requiere la **fórmula binomial**:
En general, siendo X : nº de veces que ocurre el resultado A en las n repeticiones.

$$\boxed{P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}}$$

donde $P(X = x)$ = probabilidad de **x** éxitos en **n** repeticiones

n = número de repeticiones

$\binom{n}{x}$ = número de maneras de obtener exactamente **x** éxitos en **n** repetic.

p = probabilidad de éxito en cualquier repetición

$(1 - p)$ = probabilidad de fracaso en cualquier repetición

Distribución binomial cont.

En el ejemplo de la conexión en la red al sitio de un centro comercial, la probabilidad buscada se podría haber calculado de la siguiente forma :
Se define la variable X: número de personas que compran entre tres que se conectan.
X ~ Bi (3, 0.2). Lo que se lee: la variable aleatoria X se distribuye (~) según una distribución binomial (Bi) de parámetros 3 (n) y 0.2 (p). Entonces:

$$P(X = 2) = \binom{3}{2} 0.2^2 (1 - 0.2)^{3-2} = 0.096$$

Parámetros estadísticos

Se puede demostrar que:

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k} = n\, p$$

$$V(X) = E(X^2) - [E(X)]^2 = n \,.\, p \,.\, (1 - p)$$

Distribución hipergeométrica

La distribución binomial no se usa en situaciones en las que la población es finita, el muestreo se realiza sin reposición y el tamaño de la muestra es superior al 10% del tamaño de la población. En estos casos la distribución que se aplica es la **distribución hipergeométrica.** Sea **X: nº de éxitos en n repeticiones.**

$X \sim \text{Hip.}(N, r, n)$

$$\boxed{P(X = x) = \frac{C_{r,x} \; C_{N-r,n-x}}{C_{N,n}}}$$

donde:

N = tamaño de la población

n = tamaño de la muestra

r = número de éxitos en la población

x = número de éxitos en una muestra para los cuales se calcula la probabilidad

C = combinaciones

Nota: $C_{N,n}$ es una forma de indicar las combinaciones de N elementos de los cuales se seleccionan n de los mismos.

Distribución hipergeométrica cont.

Ejemplo. En una reunión hay ocho personas de las cuales 4 son miembros de un sindicato. Se seleccionan al azar tres personas para formar un comité. ¿Cuál es la probabilidad de que exactamente una de ellas sea miembro de un sindicato?

Como la población es pequeña y la muestra debe hacerse sin reemplazo, no resulta apropiada la distribución **binomial** de allí que se usa la distribución **hipergeométrica** con los siguientes parámetros:

$N = 8$, $n = 3$, $r = 4$, $x = 1$; es decir $X \sim$ Hip. (8,4,3)

X : nº de personas miembros del sindicato en una muestra de 3.

y la probabilidad pedida será:

$$P(X = 1) = \frac{C_{4,1}\ C_{4,2}}{C_{8,3}} = 0.429$$

PARAMETROS ESTADISTICOS

Se demuestra que:

$E(X) = n \cdot p$ siendo $p = r/N$

$$V(X) = n \cdot p \cdot (1-p)\ \frac{N-n}{N-1}$$

Distribución Poisson

La **distribución de Poisson** se usa para modelar situaciones en las que hay ocurrencias aleatorias de sucesos por unidad de espacio o tiempo y en donde se desea conocer la probabilidad de un número específico de éxitos.

Numerosos fenómenos discretos se representan mediante un **proceso de Poisson**.

Se dice que existe un proceso de Poisson si al considerar un experimento en el que se observa la aparición de sucesos puntuales en un intervalo continuo (de tiempo, longitud, área, etc.), en cualquier intervalo suficientemente pequeño del intervalo continuo, se verifica que:

1.- La probabilidad de observar exactamente un éxito en el intervalo es estable.

2.- La probabilidad de observar exactamente más de un éxito en el intervalo dado, es 0.

3.- La ocurrencia de un éxito en cualquier intervalo es estadísticamente independiente de aquella en cualquier otro intervalo.

Se define **X : número de éxitos en un intervalo**. La fórmula para la distribución de Poisson es:

$$\boxed{P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}} \qquad X \sim P_o(\lambda)$$

donde λ es el número promedio de ocurrencias por unidad de tiempo o espacio.
x es el número de ocurrencias cuya probabilidad se desea conocer
e base de los logaritmos naturales

Distribución Poisson cont.

La fórmula de la **distribución de Poisson** muestra que la misma describe una variable aleatoria discreta que puede tomar cualquier valor de una sucesión infinita (x=0, 1, 2, ………)

PARAMETROS ESTADISTICOS

Puede demostrarse que:

$$E(X) = \sum_{k=0}^{\infty} k \frac{e^{-\lambda} . \lambda^{k}}{k!} = \lambda$$

$$V(X) = E(X^2) - [E(X)]^2 = \lambda$$

Para comprender mejor el **proceso de Poisson**, supongamos que se analiza la variable número de clientes que llegan a un banco entre las 12 y 13 hrs. Cualquier llegada de un cliente es un evento discreto en un punto particular sobre el intervalo continuo de 1 hora. Durante tal intervalo de tiempo puede haber un promedio de 180 llegadas.

Para responder cuál es la probabilidad de que en un minuto lleguen exactamente dos clientes, es necesario conocer el número promedio de clientes en un minuto. Si el número promedio de llegadas en una hora es 180, en un minuto será igual a 3 (180 : 60), luego:

$$P(X = 2) = \frac{3^2 e^{-3}}{2!} = 0,224$$

Distribución Poisson cont.

siendo **X = nº de personas que llegan al banco en 1 minuto.**

Si hubiéramos estado interesados en el **nº de personas que llegan al banco en un intervalo de 4 minutos**, se debe considerar un promedio de llegadas igual a $4\lambda = 4 \cdot 3 = 12$.

En general, se escribe:

X_t : número de éxitos en un intervalo t

$X_t \sim Po(\lambda t)$

$$\boxed{P(X_t = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}}$$

Ejemplo. Una compañía telefónica observa que entran en promedio 3.2 llamadas por minuto en una línea determinada. Suponiendo que el número de llamadas se distribuye según un modelo de Poisson, se puede plantear el cálculo de las siguientes probabilidades para el intervalo de un minuto:

a) probabilidad de que entren exactamente 2 llamadas

b) probabilidad de que entren a lo sumo tres llamadas

c) probabilidad de que entren por lo menos tres llamadas

d) probabilidad de que entren entre 2 y 7 llamadas (incluidos estos valores)

Aproximación de la Poisson a la Binomial

En problemas de distribución **binomial** donde **n es grande** y la **probabilidad de éxito p es muy pequeña** (o muy grande), se puede demostrar que la distribución **binomial** se acerca a una distribución de **Poisson**.

Como regla empírica, el uso de la distribución de Poisson es apropiado cuando **n > 20** y **np ≤ 5** o **n(1 - p) ≤ 5**.

Es decir, si X es una variable distribuida binomial con parámetro p (con base en n repeticiones independientes del experimento). Esto es:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

y ***n →∞** , **np** = λ (constante) , o equivalente: cuando **n →∞**, **p → 0** tal que **np →** λ ;* bajo estas condiciones se tiene :

$$\lim_{n \to \infty} P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Es decir, la distribución de **Poisson con parámetro** λ .

Aproximación de la Poisson a la Binomial cont.

Ejemplo. Un 1% de los empleados de una fábrica se ausentan diariamente del trabajo. Si se eligen 70 empleados al azar, ¿cuál es la probabilidad de que sólo 1 esté ausente?
X = nº de empleados ausentes en un total de 70, luego X ~ Bi (70 , 0.01).

Entonces P (X = 1) =

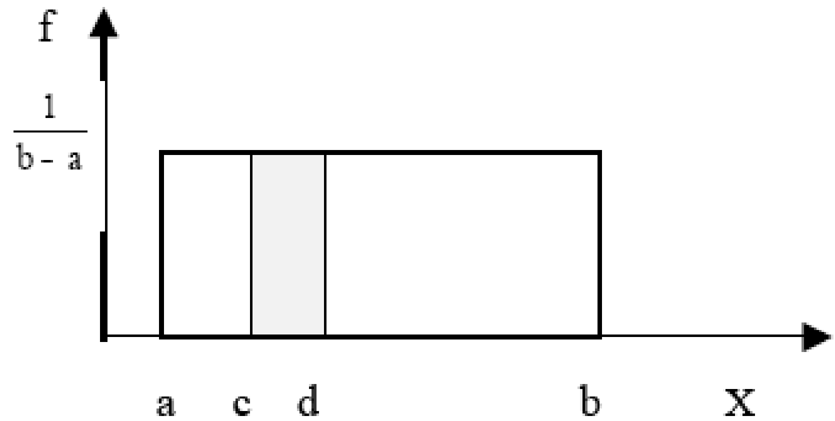
Como **n es grande** (70) y **p pequeño (0.01)** se puede utilizar la aproximación de Poisson de la siguiente forma: λ = np = 70 x 0.01 = 0.7

Entonces P(X=1)=

A continuación conoceremos algunas de las distribuciones de **variable aleatoria continua** de mayor aplicación:

DISTRIBUCIÓN UNIFORME
La **distribución uniforme** describe una variable aleatoria que tiene igual probabilidad de ocurrir en sub-intervalos de igual tamaño, dentro del campo de definición de la misma.
La siguiente figura muestra una distribución uniforme. El intervalo en el que está definida la variable aleatoria es (a, b). La curva de probabilidad tiene una altura uniforme en todos los puntos entre a y b.

Distribución Uniforme cont.



El área del rectángulo es igual a 1, de allí:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & si\ x \in [a,b] \\ 0 & en\ otro\ caso \end{cases}$$

La probabilidad de que la variable caiga entre cualesquiera dos puntos c y d (ver la figura) es igual a:

$$P(c \leq X \leq d) = \int_c^d \frac{1}{b-a}\, dx = \frac{d-c}{b-a}$$

Distribución Uniforme cont.

PARAMETROS ESTADISTICOS

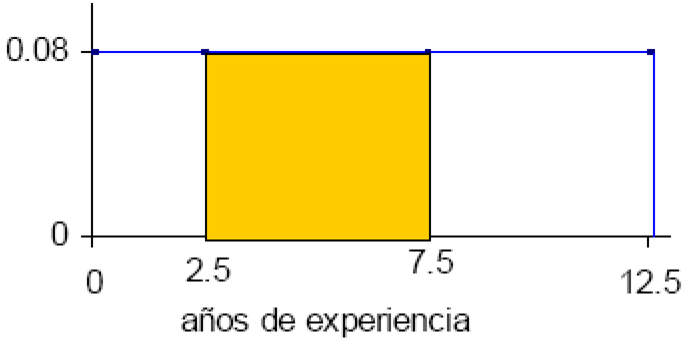
Se demuestra que:

$$E(X) = \int_a^b x\, f(x)\; dx = \frac{a+b}{2}$$

$$\sigma^2(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$$

Ejemplo. Se sabe que en una empresa, los años de experiencia de los trabajadores de planta tienen una distribución uniforme con un mínimo de 0 y un máximo de 12.5 años. Se elige un empleado al azar. Determine la probabilidad de que esta persona tenga entre 2.5 y 7.5 años de experiencia en la compañía.

La variable aleatoria uniforme se presenta escribiendo: **X : años de experiencia de un trabajador.**



Distribución Uniforme cont.

$$P(2.5 < X < 7.5) = \int_{2.5}^{7.5} \frac{1}{12.5}\,dx = \frac{7.5 - 2.5}{12.5 - 0} = \frac{5}{12.5} = 0.4$$

Hay un 40% de probabilidad de que un empleado de la compañía elegido al azar tenga entre 2.5 y 7.5 años de experiencia.

El promedio y la desviación estándar son:

$$E(x) = \frac{0 + 12.5}{2} = 6.25 \text{ años}$$

$$\sigma(x) = \sqrt{\frac{(12.5 - 0)^2}{12}} = 3.61 \text{ años}$$

La distribución de los años de experiencia en una empresa para los trabajadores de planta tiene un promedio de 6.25 años y una desviación estándar de 3.61 años

Distribución Exponencial

En el ejemplo de la página 10 se definió la variable aleatoria de Poisson: número de clientes que llegan a un banco entre las 12 y 13 hs. Otra variable de interés puede ser el **tiempo entre dos llegadas sucesivas. Interesa conocer su distribución.**

Sea la variable aleatoria T: tiempo entre dos llegadas **sucesivas a un banco.**

La distribución de T puede obtenerse a partir de conocer la distribución del número de clientes que llegan al banco. Por ejemplo: si interesa conocer la probabilidad de que el tiempo entre dos llegadas sucesivas sea mayor que t minutos, ésta podrá derivarse de la distribución de probabilidad de **X_t : nº de clientes que llegan al banco en t minutos.**

La clave para obtener esta relación es el siguiente concepto: el tiempo entre dos llegadas sucesivas es mayor que t, si y sólo si no hay llegadas en esos t minutos. Es decir:

$$P(T \geq t) = P(X_t = 0) = \frac{e^{-\lambda t}\,(\lambda\ t)^0}{0!} = e^{-\lambda t}$$

Distribución Exponencial cont.

Luego:

$$P(T \leq t) = 1 - P(T \geq t) = 1 - e^{-\lambda t}$$

$$F(t) = 1 - e^{-\lambda t}$$

$$\frac{dF(t)}{dt} = \lambda e^{-\lambda t} = f(t)$$

Se obtuvo la función de densidad de T, que es una variable aleatoria continua con **distribución exponencial. La obtención de la distribución de T depende sólo de la hipótesis** de que el número de llegadas sigue un proceso de Poisson.

Ejemplo. Interesa conocer la probabilidad de que el tiempo entre dos llegadas sucesivas a un banco sea mayor que 1 minuto. $X_1 \sim Po\ (\lambda = 3)$

$$P(T > 1) = P(X_1 = 0) = \frac{e^{-3}.3^0}{0!} = e^{-3} = 0.0498$$

Distribución Exponencial cont.

Formalizando se dice que una variable aleatoria continua **X tiene distribución exponencial si**:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

PARAMETROS ESTADISTICOS

Si la variable aleatoria X tiene una distribución exponencial con parámetro λ, entonces

$$E(X) = \frac{1}{\lambda} \qquad \text{y} \qquad V(X) = \frac{1}{\lambda^2}$$

Para el ejemplo planteado anteriormente, consideramos:

X : tiempo entre dos llegadas sucesivas a un banco. X ~ *Exp (3 minutos)*

$$P(X < 1) = 3\int_0^1 e^{-3x}\,dx = 1 - e^{-3} = 0.9502$$

Distribución Exponencial cont.

Ejercicios.

1.- El tiempo entre llegadas de taxis a una parada tiene una distribución exponencial con promedio 10 minutos. ¿Cuál es la probabilidad de que una persona que esté en el cruce tenga que esperar más de una hora para tomar un taxi?

2.- Demuestre que si $X \sim Exp(\lambda)$ entonces:

$$P (X > s + t \mid X > s) = P (X > t) \text{ para todo } s,t > 0. \text{ (Propiedad de no memoria)}$$