Distribuciones de Probabilidad Variable Aleatoria Discreta (VAD)

ESTADÍSTICA PARA CIENCIA DE DATOS

PROF. ESTEBAN BALLESTERO

Distribución de probabilidad Binomial

Ejemplos:

- Cada artículo que sale de una línea de producción de manufactura puede estar defectuoso o en buen estado
- Cada tiro en una serie de disparos puede dar o no en el blanco
- Un trabajador puede ser empleado o desempleado
- El resultado de una llamada de ventas puede resultar que el cliente compre o no el producto

Distribución de Probabilidad Binomial: características

- El resultado de cada ensayo dentro de un experimento se clasifica dentro de dos categorías mutuamente excluyentes: éxito o fracaso
- La variable aleatoria cuenta el número de éxitos en un número fijo de ensayos
- La probabilidad de éxito permanece igual en cada ensayo.
- Los ensayos son independientes, lo que significa que la ocurrencia de uno de ellos no afecta el resultado de cualquier otro

Nota

- Es necesario advertir que un éxito no es necesariamente lo mejor, tal como se usa coloquialmente la palabra.

- El éxito debe entenderse como uno de los resultados posibles de un ensayo en un experimento

Definición: DPB

Se dice que una variable aleatoria X tiene una distribución binomial basada en n ensayos con probabilidad de éxito p si y solo si

$$p(x) = \binom{n}{x} p^x q^{n-x}, x = 0,1,2,\ldots,n;\ 0 \leq p \leq 1$$

Donde:

n: Numero de ensayos

x: número de éxitos

p: probabilidad de éxito de cada ensayo

q: probabilidad de fracaso

Ejemplo1

Suponga que un lote de 5000 fusibles tiene 5% de defectuosos. Si se toma una muestra de 5 fusibles, encuentre la probabilidad de encontrar por lo menos uno defectuoso

Ejemplo 2

La experiencia ha demostrado que el 30% de las personas que padecen cierta enfermedad se recupera. Un empresa farmacéutica desarrolló un nuevo medicamento y lo administró a 10 enfermos elegidos aleatoriamente. Nueve se recuperaron de inmediato. Suponga que el medicamento es ineficaz, ¿cuál es la probabilidad de que por lo menos 9 de las 10 personas que toman medicamento se recuperen?

Ejemplo 3 (binomial acumulada)

Una librería tiene 4 fotocopiadoras de las cuales necesita que dos funcionen para trabajar normalmente. Cada fotocopiadora pasa en promedio 8% del tiempo fuera de servicio. ¿Qué porcentaje del tiempo trabaja normalmente la librería?.

Ejemplo 4

Considere la situación del Ejemplo 1, donde ahora se tomará una muestra de tamaño 20 fusibles. Determine la probabilidad de detectar por lo menos cuatro con defecto.

Caso:

Un comprador recibe un embarque que contiene 20 computadoras personales y necesita muestrear tres de ellas para ver si funcionan antes de aceptar el embarque. Con este fin elige las PC más cercanas, las prueba y después decide si están o no defectuosas. El comprador ignora que dos de las 20 computadoras están defectuosas. ¿Es este un experimento binomial?.

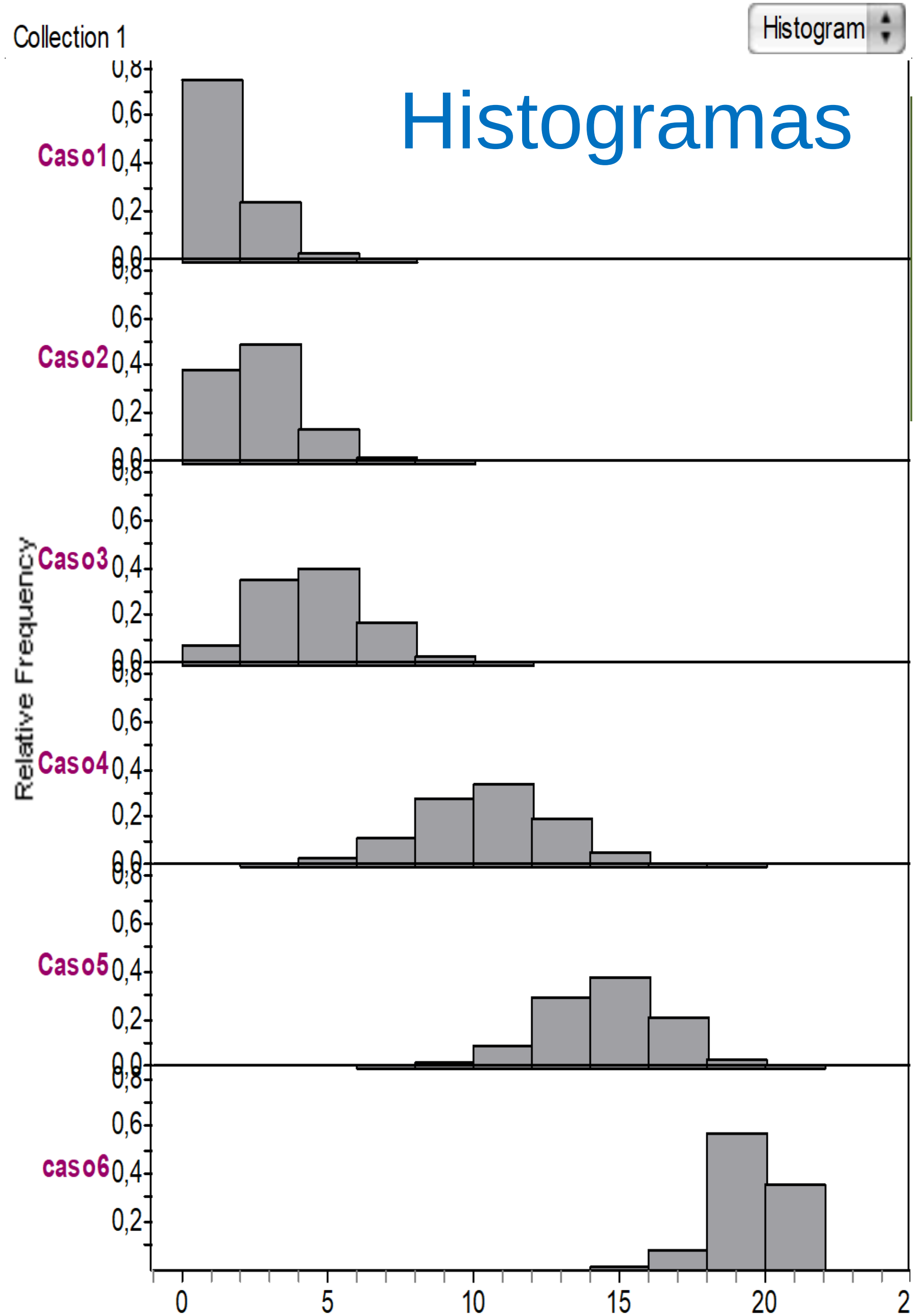
Representaciones Gráficas

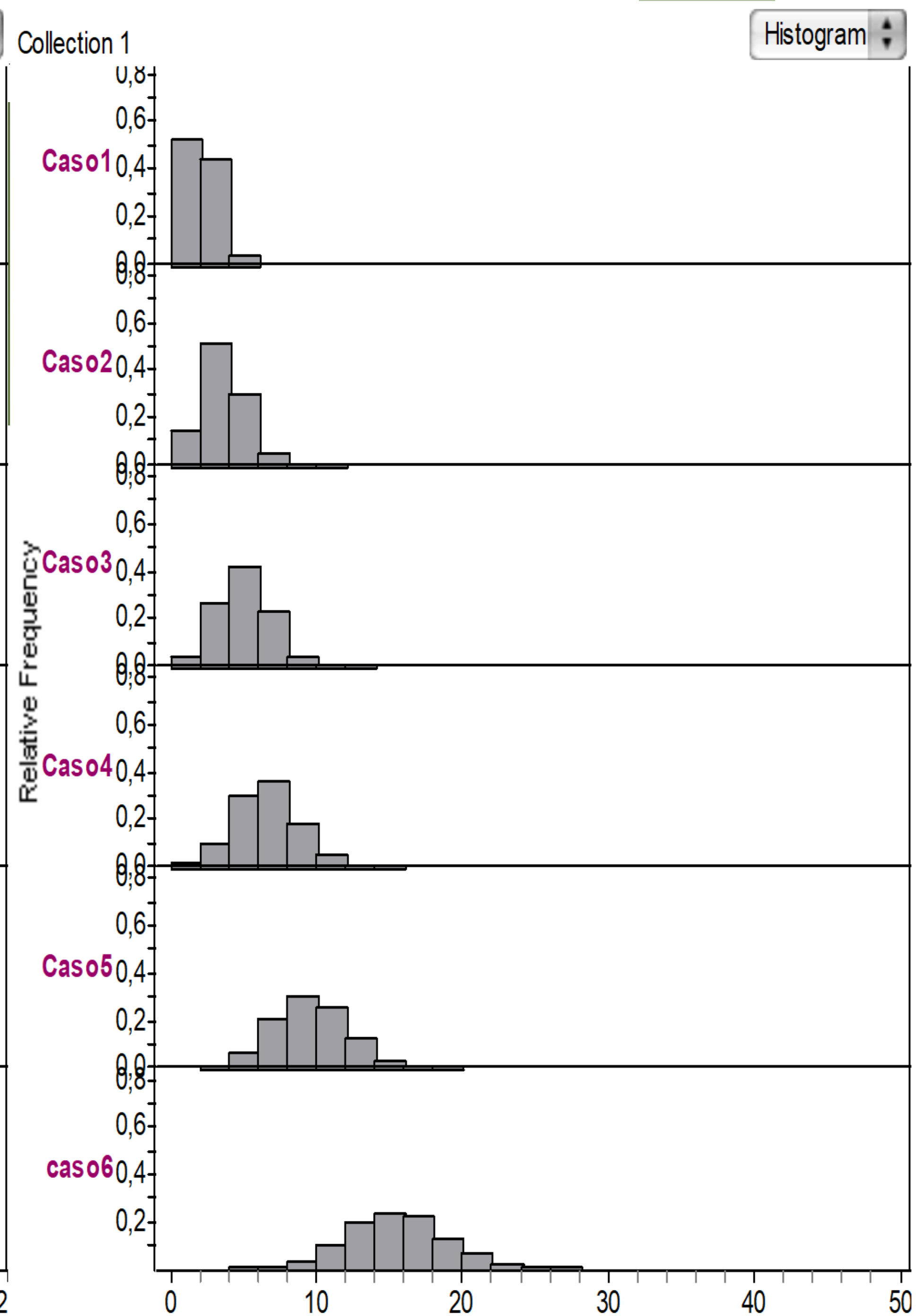
Seguidamente se le presenta los histogramas de distribuciones binomiales para diferentes valores de p y n. El primer bloque se obtuvo dejando fijo n y variando p y el segundo bloque se dejó fijo p y se varió n.

	$Caso1$	$Caso2$	$Caso3$	$Caso4$	$Caso5$	$Caso6$
p	0,05	0,1	0,2	0,5	0,7	0,95
n	20	20	20	20	20	20

	$Caso1$	$Caso2$	$Caso3$	$Caso4$	$Caso5$	$Caso6$
p	0,3	0,3	0,3	0,3	0,3	0,3
n	5	10	15	20	30	50

Histogramas





Ejercicio:

Analice el gráfico en cada caso y conteste ¿Cómo son las distribuciones en cada uno de los casos?, ¿qué ocurre con la distribución binomial si se mantiene fijo el número de ensayos y variamos la probabilidad de éxito?, ¿qué ocurre con la distribución binomial si se mantiene fijo el la probabilidad de éxito y variamos número de ensayos?

Otros estadísticos en DB

Si X es una variable aleatoria binomial basada en ***n*** ensayos y con una probabilidad de éxitos p, entonces:

$$\mu = E(X) = np \quad y \quad \sigma^2 = npq$$

Ejemplo 5

1. Al tirar una tachuela, la probabilidad de que esta caiga punta arriba es 0,65.Si se tiran 20 tachuelas, ¿cuál es el valor esperado de puntas hacia arriba y la desviación estándar?.
2. Un examen consta de 20 preguntas de selección única con 4 opciones por pregunta. Si se contesta aleatoriamente:
   a. ¿Cuál es la probabilidad de que la nota sea mayor o igual que 85?
   b. ¿Cuál es la nota esperada y la varianza?

Ejercicio:

- Calcule la desviación estándar para cada uno de los casos estudiados en el ejercicio anterior, tanto para el bloque 1 como para el bloque 2
- Compare las desviaciones estándar y realice conclusiones sobre lo que ocurre con la variabilidad, indicando para que valores de p la distribución es más variable en el bloque 1 y para el bloque 2, explique para cuáles valores de **n** la distribución se vuelve más variable.

Distribución de probabilidad Hipergeométrica

Binomial	Hipergeométrica
• La probabilidad de éxito debe permanecer siempre igual en cada ensayo • Enfocado hacia experimento con reemplazo • Poblaciones grandes o infinitas	• La probabilidad de éxito cambia en cada ensayo • Enfocada hacia experimentos sin reemplazo • Poblaciones finitas y pequeñas

D. Hipergeométrica

Se utiliza:

1. Si se selecciona una muestra de una población finita sin reemplazo
2. Si el tamaño de la muestra ***n***, es más de un 5% del tamaño de la población N

D. Hipergeométrica: definición

Se dice que una variable aleatoria discreta X tiene una distribución Hipergeométrica si y solo si:

$$p(x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$$

Donde:

N: Tamaño de la población

S: Número de éxitos de la población

x: Número de éxitos de interés

n: Tamaño de la muestra o número de ensayos

Ejemplo 6

Suponga que durante la semana se fabricaron 50 estaciones de juegos para videos. Cuarenta de ellas funcionaron perfectamente y diez tuvieron al menos un defecto. Se seleccionó al azar una muestra de 5. Utilizando la fórmula hipergeométrica, ¿cuál es la probabilidad de que 4 de los 5 funcionaran correctamente?

Ejemplo 7

De un grupo de 20 ingenieros con doctorados, se eligen 10 aleatoriamente con el fin de contratarlos. ¿Cuál es la probabilidad de que entre los 10 seleccionados estén los 5 mejores del grupo de 20?

Valor esperado y varianza

Si X es una variable aleatoria con una distribución hipergeométrica, entonces se cumple que:

$$\mu = E(X) = \frac{nS}{N} \quad y \quad \sigma^2 = V(X) = \frac{nS(N-S)(N-n)}{N^2(N-1)}$$

Ejercicio

Calcule la media y la varianza para los ejemplos 6 y 7

Ejemplo 8

Un producto industrial se envía en lotes de 20 unidades. Efectuar pruebas para determinar si un artículo tiene defectos es costoso, así el fabricante toma muestras de su producción en lugar de probar el 100%. Un plan de muestreo elaborado para reducir al mínimo la cantidad de artículos defectuosos que se envían a los consumidores requiere que se muestreen 5 artículos de cada lote y que se rechace el lote completo si se encuentra más de un artículo defectuoso.

Ejemplo 8

(Si el lote es rechazado se prueba cada artículo del lote). Si un lote tiene 4 artículos defectuosos, ¿cuál es la probabilidad de que el lote sea rechazado?, ¿cuál es el número esperado de artículos que tienen defecto en la muestra de tamaño 5 y cuál es su varianza?

Distribución de probabilidad Geométrica

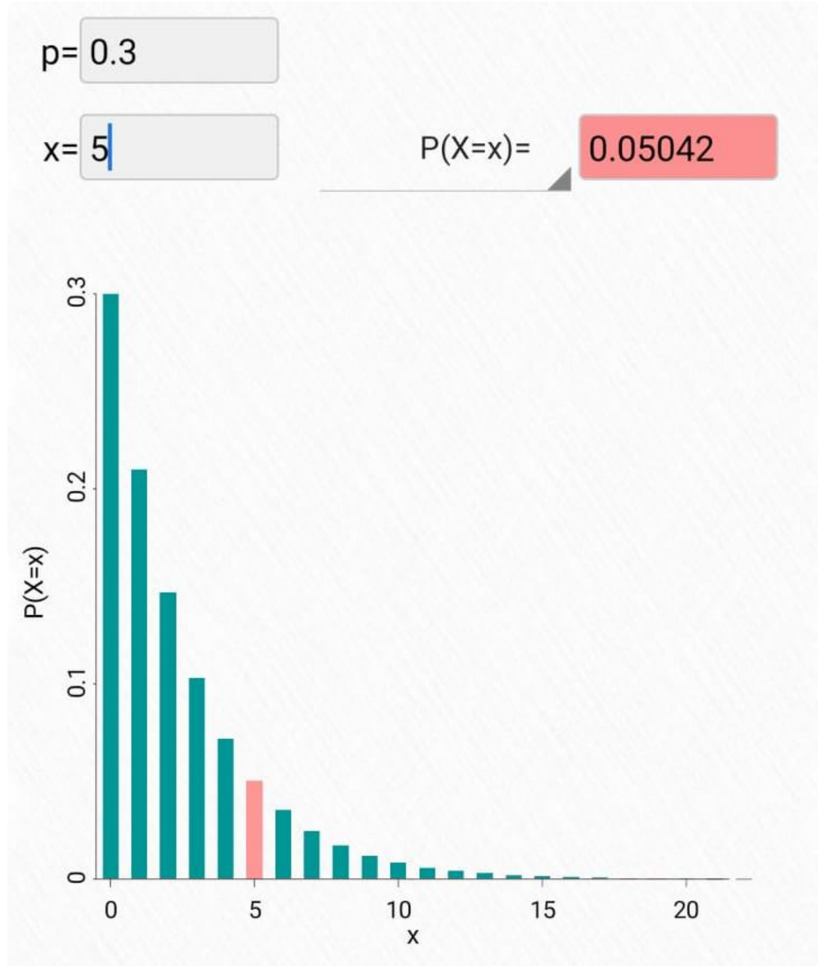
Ejemplo 9:

Se lanza un dado hasta obtener un 6, ¿cuál es la probabilidad de realizar al menos cinco lanzamientos?

DPG

Si se tiene un urna con éxitos y fracasos. Suponga que se empieza a realizar extracciones con reposición hasta obtener el primer éxito. Sea Y la variable que indica el número de extracciones realizadas hasta obtener el primer éxito, entonces se dice que Y sigue una distribución geométrica.

Nota: debe considerarse que las pruebas repetidas que se realicen, son independientes



DPG: fórmula, media y varianza

$$g(y;p) = p(y) = pq^{y-1}, con\ y = 1, 2, 3, \ldots$$

Donde:

p: probabilidad de éxito

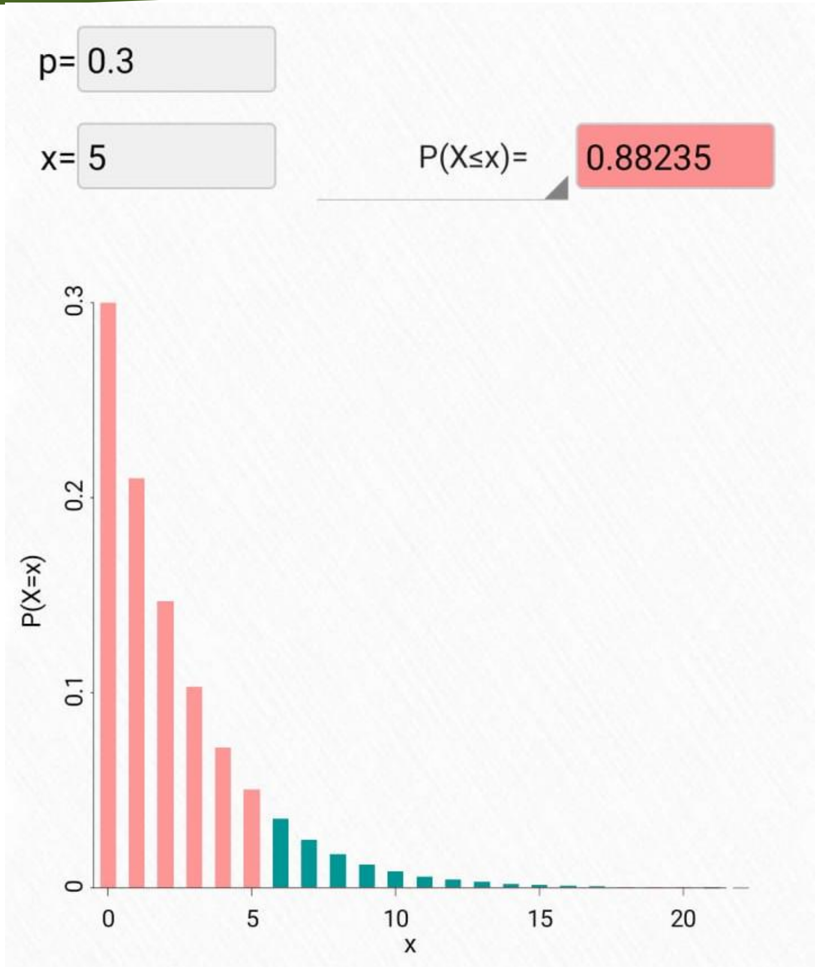
$q = 1 - p$: probabilidad e fracaso

y: número de la prueba en la que ocurre el primer éxito

$$\mu = \frac{1}{p},\ \sigma^2 = \frac{1-p}{p^2}$$

DPG: Acumulada

$$p(y \leq k) = 1 - q^k$$



Ejemplo 10:

Se sabe que en cierto proceso de fabricación uno de cada 100 artículos en promedio resulta defectuoso, ¿cuál es la probabilidad de que el quinto artículo que se inspecciona en un grupo de 100 sea el primero defectuoso que se encuentre?

Ejemplo 11

Una tri-jugada para un dado consiste en lanzar un dado hasta obtener un múltiplo de 3 o detenerse: se gana el juego si se obtiene un múltiplo de 3 antes del quinto lanzamiento, o en caso contrario, se pierde. Don Edgar realiza 10 tri-jugadas con un dado, ¿cuál es la probabilidad de que gane por lo menos 8 de estas?

Ejemplo 12

Un concurso de TV tiene 10 participantes que tratarán de ganarse un automóvil por medio del juego BOLA BLANCA, de la siguiente manera: si el primer participante gana el juego, obtiene el automóvil y se acaba el concurso, caso contrario para el segundo participante, si gana el juego obtiene el automóvil y se acaba el concurso; sino pasa a jugar el tercer participante y así sucesivamente. **¿Cómo se juega la Bola Blanca?**

Se tiene una canasta con cinco bolas blancas y 7 rojas, el jugador debe extraer una bola y regresarla a la canasta, si obtiene una bola blanca en a lo sumo la tercera extracción, gana el juego, sino pierde.

Ejemplo 12

1. Halle la función de distribución de probabilidad para el número de extracciones de la canasta hasta obtener una bola blanca R/ $p(y) = \left(\frac{7}{12}\right)^{y-1} \frac{5}{12}$
2. Determine la probabilidad de ganar bola blanca R/ $\frac{1385}{1728}$
3. ¿Cuál es la probabilidad de que el tercer participante se gane el automóvil? R/ 0.0316
4. ¿Cuál es la probabilidad de que uno de los primero cuatro participantes se gane el automóvil? R/ 0.9984
5. Determine la probabilidad de que ningún participante se gane el automóvil. R/ 9.49519×10^{-8}

Algunas consideraciones

- Es posible calcular las distribuciones de probabilidad binomial para probabilidades de éxito menores a 0.05, pero los cálculos tal vez consuman demasiado tiempo, en especial para n grande (por ejemplo 100 o más).

- La distribución de probabilidad tendrá cada vez un mayor sesgo a medida que la probabilidad de éxito disminuya

Algunas consideraciones

¿qué ocurriría con la distribución binomial cuando n es muy grande?. Realicemos la deducción.

Algunas consideraciones

- Supongamos que se desea determinar la distribución de probabilidad del número de accidentes de tránsito en un cruce específico, durante una semana.
- Este período de una semana se puede dividir en subintervalos
- Esto se puede hacer de varias formas, y aunque, no se conozca $\boldsymbol{n}$ ni $\boldsymbol{p}$, parece razonable suponer que al aumentar $\boldsymbol{n}$, $\boldsymbol{p}$ disminuye.

Distribución de probabilidad de Poisson

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}, con x = 1,2,3, \ldots, \mu > 0$$

λ : $\quad valor\ esperado$

e : $\quad constante$

x : $\quad número\ de\ éxitos$

Distribución de probabilidad de Poisson

- Es conocida como la forma de límite de la distribución binomial cuando la probabilidad de éxito es muy pequeña y n es grande.
- Se le conoce con el nombre de "**Ley de los eventos improbables**", haciendo referencia a la baja probabilidad de éxito.

DPP

- Se usa para determinar la probabilidad de la ocurrencia de un número determinado de eventos cuando estos ocurren en un continuo de espacio o tiempo.

Experimento de Bernoulli

- Es un proceso de muestreo en el que en cada ensayo solo puede presentarse dos resultados u observaciones mutuamente excluyentes: éxito o fracaso

Proceso de Poisson

- Es similar al proceso de Bernoulli, con la diferencia de que los eventos ocurren a los largo de un continuo (por ejemplo, un intervalo de tiempo) y donde los ensayos no son explícitos.

Ejemplos de procesos de Poisson

- Entrada de llamadas en un conmutador telefónico
- Errores en los registros de datos
- Número de marcas y otras imperfecciones en partes automotrices recién pintadas
- Cantidad de clientes en espera

Otras consideraciones

- Para calcular una probabilidad de Poisson solamente se necesita un valor: la media o valor esperado de eventos a largo plazo en un lapso específico o en una dimensión de espacio que interese.
- En las distribuciones de Poisson se cumple que:

$$\mu = E(X) = V(X) = \sigma^2$$

Ejemplo 1

- Suponga que se diseña un sistema aleatorio de patrullaje policíaco, de modo que un patrullero pueda visitar un sector de vigilancia $X=1,2,3,\ldots,$ veces en períodos de media hora, visitando cada sector un promedio de una vez cada período. Suponga que X posee aproximadamente una distribución de Poisson. Calcule la probabilidad de que el vigilante pase por alto un sector en un período de media hora. ¿Cuál es la probabilidad de que lo visite una vez, dos veces y por lo menos una vez?.

Ejemplo 2

- Los almácigos de cierta clase de árbol se encuentran dispersos al azar en una vasta área, con una densidad promedio de aproximadamente cinco por metro cuadrado. Si un arboricultor escoge aleatoriamente 10 regiones muestrales de un metro cuadrado en esa área, encuentre la probabilidad de que ninguna de las regiones contenga árboles.

Ejemplo 3

- Suponga que *Y* posee una distribución binomial con ***n=20*** y ***p=0,1***. Determine el valor exacto de $P(Y \leq 3)$. Aproxime esta probabilidad mediante la probabilidad correspondiente dada por la distribución de Poisson. ¿Qué ocurriría para ***n*** más pequeño?

Ejemplo 4

- El número promedio de accidentes que ocurren en una fábrica es de tres por mes. En el último mes ocurrieron seis accidentes, ¿parece muy improbable esta cantidad si la media permanece igual a 3?, ¿indica esto un incremento en la media?.

Ejemplo 5

- La contaminación constituye un problema en la fabricación de discos de almacenamiento óptico. El número de partículas de contaminación que ocurre en un disco óptico tiene una distribución de Poisson y el número promedio de partículas por centímetro cuadrado de superficie del disco es 0,1. El área de un disco bajo estudio es 100 centímetros cuadrados. Determine la probabilidad de que ocurran 12 partículas en el área del disco bajo estudio.

Ejemplo 6

- Para el caso del alambre delgado de cobre, suponga que el número de imperfecciones sigue una distribución de Poisson con una media de 2,3 imperfecciones por milímetro. Determine la probabilidad de encontrar exactamente 2 imperfecciones en un milímetro de alambre. Además, determine la probabilidad de encontrar 10 imperfecciones en 5 milímetros de alambre.