PUCE Virtual

Estadística Aplicada

Medición de Probabilidad variables Discretas 2

Clase 7

Ingeniería en ciberseguridad

La excelencia no se improvisa

1. INTRODUCCIÓN DE LA CLASE

Ahora que dominamos la distribución de probabilidad binomial, estamos listos para pasar a la siguiente distribución teórica discreta: la de Poisson. Esta distribución de probabilidad recibe su nombre de Simeon Poisson, un matemático francés que la desarrolló a principios del siglo XIX.

La distribución de Poisson es útil para calcular la probabilidad de que ocurra una cierta cantidad de eventos durante un período de tiempo o espacio específico. Por ejemplo, podríamos usar esta distribución para determinar la probabilidad de que 10 clientes entren a una tienda durante la próxima hora, o que ocurran 2 accidentes automovilísticos en una intersección concurrida este mes.

Clase 7:

Resultado o resultados de aprendizaje que serán abordados con el contenido de la clase: **Analizar información contextual sobre hábitat, infraestructura y movilidad, mediante técnicas estadísticas descriptivas e inferenciales, para la adecuada toma de decisiones**

Reto # 3

7) Medición de probabilidad variables discretas 2

7.1) Distribución Poisson

Para entender la distribución de Poisson debemos recordar los siguientes conceptos de la teoría de probabilidades.

FRECUENCIA RELATIVA

Este modelo utiliza los datos de las observaciones que se han registrado; con base a estos, calcula la frecuencia con la que se ha presentado dicho evento y, en base a esta frecuencia, determina la probabilidad.

Figura 1

Fórmula de frecuencia relativa

$$P(E) = \frac{Numero\ de\ veces\ que\ ha\ ocurrido\ el\ evento}{Número\ total\ de\ observaciones}$$

Este modelo es también llamado *a posteriori*, dado que la probabilidad del evento se calcula luego de ser recabada la información.

También hay que recordar que, al calcular este tipo de probabilidad, pueden presentarse algunos problemas. Por ejemplo, de la fórmula podemos deducir que, si las observaciones no incluyen una o más opciones, será imposible su cálculo de probabilidad; además, si se cuenta con pocas observaciones, sus resultados pueden ser engañosos.

Esta distribución, ideada por Simeon Poisson en 1840, se refiere a variables aleatorias de naturaleza discreta que tratan de inferir la frecuencia relativa de un evento sobre alguna unidad de tiempo o espacio. Por ejemplo, se utiliza para describir el número de llegadas de clientes por hora, el número de accidentes cada mes, el número de defectos en un enlace de fibra óptica por kilómetro, etcétera. Es decir, cada vez que se vea un enunciado en el que se pretende conocer el valor de una variable por tiempo o espacio, será una clave para revisar la distribución de Poisson.

Qué es un proceso de Poisson

De acuerdo con Donnelly R. (2019), Un proceso de Poisson tiene las siguientes características:

- ✓ El experimento consiste en contar el número de ocurrencias de un evento durante un período de tiempo, área, distancia o cualquier otro tipo de medición, este valor será dado por el argumento k.
- ✓ La probabilidad de ocurrencia del evento es constante para dos intervalos cualesquiera de tiempo o espacio. Por ejemplo, si seis clientes entran a la tienda durante la primera hora de actividad, esto no tendría ningún efecto en el número de clientes que llegarían durante la segunda hora, esto estaría calculado en la media como se menciona a continuación.
- ✓ La media de la distribución de Poisson tiene que ser la misma para cada intervalo de medición y se denota con el argumento λ.
- ✓ El número de ocurrencias durante un intervalo es independiente del número de ocurrencias en cualquier otro intervalo.
- ✓ Los intervalos no se superponen. Por ejemplo, al contar el número de clientes que entran a la tienda en períodos de una hora, los períodos de una hora no pueden superponerse entre sí. Podemos contar el número de clientes que llegan entre las 9 y las 10 a. m. y entre las 10 y las 11 a. m., y así sucesivamente, pero no podemos utilizar otro período de 9:30 a 10:30 a.m. porque se superpone con los otros intervalos.

FUNCIÓN DE PROBABILIDAD PMF

Recordando el concepto de la función PMF, esta calcula una probabilidad puntual; por ejemplo, la probabilidad de que entren 3 personas a una tienda asumiendo un lambda de x.

De acuerdo con Webster (2020), dado los supuestos mencionados arriba, la función de probabilidad estará dada por la siguiente fórmula

Figura 2

Función PMF para la distribución de Poisson

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

¡Note que al encontrar un k! implica que la variable debe ser discreta, dado que

no es posible calcular el factorial de un valor con decimales.

A diferencia de la distribución binomial, en la cual el experimento toma solo dos posibles resultados (el evento se presenta o no), en el proceso de Poisson se puede tener cualquier cantidad de resultados en la unidad de medida. Por ejemplo, la cantidad de clientes que ingresan a un banco físicamente durante la siguiente hora podría ser cero, uno, dos, tres, etc. La variable aleatoria para la distribución de Poisson sería la cantidad real de ocurrencias, en este caso, la cantidad de clientes que llegan durante la siguiente hora. Con esto en mente, entonces el CDF, que es la función de probabilidad acumulativa, se evalúa como la sumatoria de las probabilidades de todos los posibles valores de la variable discreta (x) que sean menores al valor k, dando como resultado la siguiente fórmula, por lo que es importante revisar la fórmula CDF de esta distribución.

Figura 4

Función CDF para distribución de Poisson

Note la sumatoria de todas las posibilidades de X menores a K

$$f(k, \lambda) = P(0 \leq X \leq k) = e^{-\lambda} \cdot \sum_{x=0}^{k} \frac{\lambda^x}{x!}$$

Funciones matemáticas Poisson

Describe las fórmulas de la distribución

Enlace

Figura 5

Media o valor esperado de aciertos (a), Varianza (b) y Desviación (c) de la distribución de Poisson

Creación de autor Alfonso Prado

(a)	(b)	(c)
$\mu = \lambda$	$\sigma^2 = \lambda$	$\sigma = \sqrt{\lambda}$

TEOREMA DEL LÍMITE CENTRAL EN POISSON

Por otro lado, es interesante mencionar cómo se aplica el teorema del límite central a la distribución de Poisson. Recordemos que el teorema establece que, a medida que n se vuelve más grande, la distribución de las medias muestrales se aproxima a una distribución normal con una media de **μ** y un error estándar de **σ/√n.**

En este caso, n serían todos los posibles valores que puede tomar la variable aleatoria y, por consiguiente, está relacionado con lambda. En otras palabras, si indicamos que $\lambda = 5$, los valores posibles que puede tomar la variable se ubican alrededor de la media 5; es decir, son unos pocos valores, pero si estamos trabajando con una lambda de 100, los valores de X aumentarán considerablemente, aunque las probabilidades de algunos puedan ser muy bajas.

Cuando lambda es pequeño, podemos observar la densidad de estos pocos valores; sin embargo, a medida que lambda aumenta, la cantidad de posibles valores se expande y su densidad va delineando la campana de Gauss.

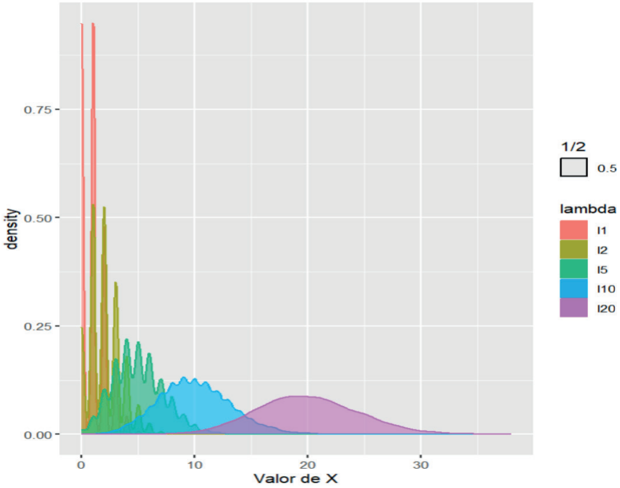
El siguiente código nos permite visualizar la aplicación del teorema del límite central.

```
#Primero vamos a crear varios datasets que cumple con una distribución de Poisson con
          distintos lambda

p1 <- rpois(n=10000 , lambda=1)

p2 <- rpois(n=10000 , lambda=2)

p3 <- rpois(n=10000 , lambda=5)

p4 <- rpois(n=10000 , lambda=10)

p5 <- rpois(n=10000 , lambda=20)


#Colocamos todos los datasets en el mismo dataframe

pos_df <- data.frame(l1=p1, l2=p2, l5=p3, l10=p4, l20=p5)

head(pos_df)


# Para poder visualizar mejor vamos a transponer las columnas en filas

#la función melt pasa valores que están en columnas a filas en una columna

#llamada valor en este caso

pos_melted <- melt( data=pos_df , variable.name=”lambda”, value.name=”valor”)

head(pos_melted)

class(pos_melted$lambda)


#Si la columna lambda fuera de tipo “character”, necesitaríamos convertirla en tipo factor  para
poder manejar el agrupamiento, para lo cual deberemos hacer una conversión si arriba salió tipo
factor estamos OK


head(pos_melted)

tail(pos_melted)

class(pos_melted$lambda)


#Visualizamos la  densidad
```

Figura 6

Demostración del teorema del límite central

Creación de autor Alfonso Prado



7.2) Funciones de distribución de Poisson

Al igual que la distribución normal, los prefijos usados para las funciones de distribución siguen siendo los mismos: “**d**” para densidad, “**p**” para probabilidad acumulada, “**q**” para obtener el inverso de **p** y “**r**” para crear un vector con determinado lambda, y el sufijo será siempre “pois.”

Figura 7

Funciones programáticas de la distribución Poisson

Creación de autor Alfonso Prado

Función de Distribución	Distrib. Poisson	
PDF	dpois()	dpois(x, lambda, log = FALSE)
CDF	ppois()	ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
Inverso CDF	qpois()	qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)

Función para obtener un set de datos de la distribución	Distrib. Poisson	
	rpois()	rpois(n, lambda)

Donde:

x Vector de cuantiles que (enteros positivos).

q vector de cuantiles.

p vector de probabilidades.

n número de valores aleatorios a devolver.

Lambda vector de medias (no negativas).

Funciones programáticas

Describe en detalle las funciones de la tabla

Enlace

Veamos ejemplos de uso de las funciones

Función rpois

Esta función retorna una cantidad dada de valores que cumplen con una media de distribución dada por lambda en un tiempo definido o en un espacio determinado.

```
#Generar una serie con la distribución de Poisson

#La función rpois() obtiene la serie

#la sintaxis es:

#rpois(n, lambda)

#Donde:

#n   es el número de valores deseados que deseamos

#lambda  media histórica por unidad de tiempo


#Ejemplo: Generar una serie de 10 elementos que representa la ocurrencia de un evento, cuya
        media por unidad de tiempo ha sido 10

p0 <- rpois(n=10 , lambda=10)

p0

[1]  9  6 14 11  5 13  8 11 18 11


mean(p0)

[1] 10.6

#Note: la media se aproxima a lambda en la distribución de Poisson, pero no es exactamente 10. Esto se debe a la pequeña cantidad de valores generados.#Intentando con un n más grande.

p0 <- rpois(n=1000 , lambda=5)

head(p0,n=15)

[1] 5 6 3 7 7 4 3 6 1 3 4 5 2 5 3

mean(p0)

[1] 5.0324

#Vemos que se acerca más
```

Función dpois

Esta función permite calcular la densidad de probabilidad para un valor puntual de x. En otras palabras, corresponde a la función de masa de probabilidad PMF. Si x es un vector, calculará la densidad para cada uno de los valores.

```
#Su sintaxis es:

#dpois(x, lambda, log = FALSE)

#Donde:

#x   Vector de cuantiles.

#lambda  es la media histórica

#log     TRUE su se desea obtener log(P[X]) FALSE si se desea obtener P[X]


#Ejemplo: Cuál es la probabilidad de hacer de exactamente 4 ventas en una semana si la tasa de ventas promedio es de 3 por semana?


dpois(4, lambda=3)

[1] 0.1680314


#Otro ejemplo: Una compañía constructora es responsable por la construcción de un  edificio, al terminar el mismo se han detectado 2 defectos por cada piso.  Para el nuevo contrato, la contratante desea poner una multa por defectos Le preguntan a usted cual es la probabilidad de tener 3 defectos por piso.


dpois(3,lambda=2)
```

Es importante mencionar que el valor de x debe estar expresado en las mismas unidades que lambda. En caso contrario, se debe convertir lambda para que ambos coincidan en unidades. La conversión puede hacerse con la siguiente fórmula:

```
lamda_nuevo=lambda_viejo*(unidades_nuevas)/(unidades_viejas )
```

Veamos un ejemplo:

```
#Un puente está calculado para soportar 60 toneladas métricas, si se considera que el vehículo más pesado es de 5000kg (5 toneladas) el puente puede resistir el paso de 12 vehiulos. El constructor pregunta cuál es la tasa de transito que tiene este puente le indican 10 vehículos al tiempo. Para estar seguro el constructor pregunta cuál es la probabilidad de que 12 o más vehículos circulen?
```

Respuesta:

```
ppois(11, lambda=10 , lower.tail=FALSE)

# Note el uso de  lower tail=FALSE dado que el enuncionado indica 12 o más

[1] 0.3
```

Función ppois

Esta función calcula la probabilidad acumulativa CDF para una distribución Poisson, tomo como argumentos el vector de cuantiles q, el lambda y el argumento de lower.tail que tiene igual significación que lo visto anteriormente.

Ejemplo

El número medio de automóviles que pasan por la intersección en una minuto determinada es $\lambda = 15$. Si queremos saber la probabilidad de que pasen exactamente 13 automóviles por ella en la próxima minuto.

```
dpois(x = 13, lambda = 15)

[1] 0.096
```

Pero más comúnmente las autoridades estarán interesadas en saber la probabilidad de que la cantidad de vehículos por minuto sea 20 o más, porque en dicha situación los mecanismos de control del tránsito ya no funcionarían .

```
ppois (x = 19, lambda = 15, lower.tail=FALSE)

[1] 0.12
```

Ejemplo con conversión de lambda

Si usted recibe llamadas al celular a una tasa constante 2 llamadas por hora. Si usted va al cine y se olvida de apagar su celular, ¿cuál es la probabilidad de que en una película de 1.5 horas, su teléfono timbre?

De este problema debemos notar dos cosas, la primera es que otra vez el lambda del enunciado no está en la misma unidad del lambda que se pregunta. Y la segunda es que el enunciado no indica si el teléfono sonará 1,2, o más veces, por lo tanto, concluimos que estamos ante una pregunta de probabilidad acumulativa o CDF.

```
#Primero realizamos la conversión al nuevo lambda

nuevo_lambda=2*1.5/1

#Aplicamos ppois

ppois(0 , lambda=nuevo_lambda, lower.tail=FALSE)

[1] 0.95
```

CÓMO VISUALIZAR LA DENSIDAD DE PROBABILIDAD

Entender las probabilidades es mucho más fácil en forma visual. Veamos un ejemplo: Una comercializadora tiene un promedio de ventas a 3 clientes por hora, queremos saber: cuál es la probabilidad de que en la próxima hora de consigan 0,1,2,3,4,5,6,7,9 o 10 ventas.

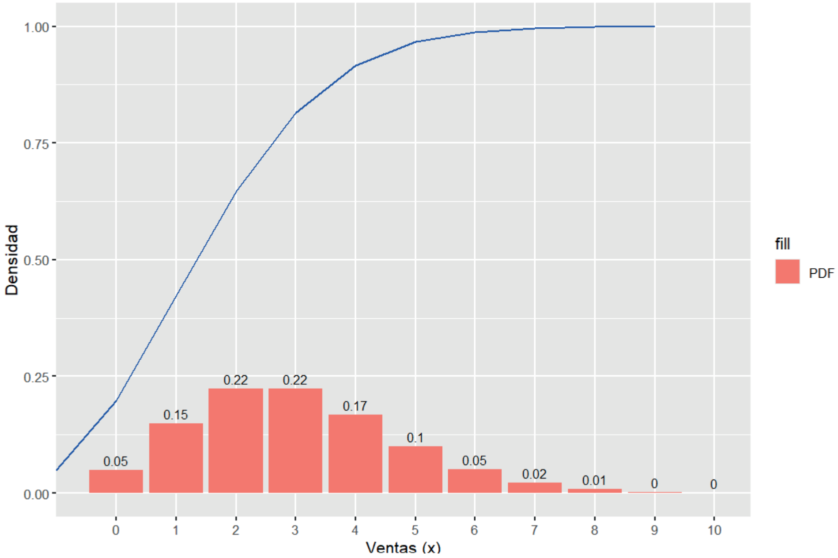
```
#opción para presentación de números con decimales o exponenciales
options(scipen = 999, digits = 2)
#creamos unos vectores
ventas <- 0:10
#Obtenemos las densidades
densidad <- dpois(x = ventas, lambda = 3)
#Obtenemos el CDF
prob <- ppois(q = ventas, lambda = 3, lower.tail = TRUE)


#pasamos los datos a dataframe
df <- data.frame(ventas, densidad, prob)
#Visualizamos
ggplot(df, aes(x = factor(ventas), y = densidad, fill=”PDF”)) +
  geom_col() +
  geom_text(
    aes(label = round(densidad,2), y = densidad + 0.01),
    position = position_dodge(0.9),
    size = 3,
    vjust = 0
  ) +
  labs(title = “PDF y CDF de Poisson “,
       x = “Ventas (x)”,
       y = “Densidad”) +
  geom_line(data = df, aes(x = ventas, y = prob) , color=”blue”)
#Comparemos con el gráfico
dpois( 1, lambda = 3) # da la probabilidad exacta de 1 venta
```

Y obtenemos el siguiente gráfico

Figura 8

PMF Y CDF en Distribución Poisson

Creación de autor Alfonso Prado



Función qpois

Como ya sabemos estas funciones q lo que nos indican es el inverso de las probabilidades, es decir, dado un vector p de probabilidades, encontrar los cuantiles debajo de los cuales se acumula las probabilidades p.

Por ejemplo:

```
# Por ejemplo, si tenemos un fenómeno Poisson con tasa promedio lambda de 55 eventos por minuto, podemos calcular los cuantiles que corresponden a las cuartiles 0.25, 0.50, 0.75:

p <- c(0.25, 0.50, 0.75)

qpois(p, lambda = 55)

[1] 50 55 60

#Indicaría que el primer cuartil estaría por debajo del valor 50, la media estaría por debajo del valor 55 y el tercer cuartil estaría ubicado debajo del valor 60
```

Referencias citadas en la Clase 7.

Webster A., (2000), Estadística aplicada a los negocios, Irwin Professional Publishing

Donnelly R., (2019) Business Statistics, Pearson Publishing

Definición de los términos citados en la Clase 7.

PMF	Una función de probabilidad o función de masa de probabilidad es una función que devuelve la probabilidad de que una variable aleatoria discreta sea exactamente igual a algún valor. Es una función que asocia a cada punto de su espacio muestral X la probabilidad de que esta lo asuma.
Proceso de Poisson	Sea X (t) el número de ocurrencias del evento con el tiempo (el proceso); entonces X (t) consiste en funciones de valores enteros no-decrecientes. La probabilidad de que sucedan exactamente k eventos en el tiempo t es: $P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

Profundización Clase 7

Video

Presenta casos de distribución Poisson

Enlace

La excelencia no se improvisa

síguenos

MDASC - M5 - A2