DISTRIBUCIONES GAUSSIANAS: NORMAL, BINORMAL Y POISSON

Bryson Jugo Harrison Prince[1], Cárdenas Fuentes Rivera Kennedy Ronald[2], Carrasco Patiño Ricardo Manlio[3], Ccohuanqui Antaquispe Renzo[4]

Universidad Nacional de Ingeniería – Facultad de Ciencias – Escuela Profesional de Matemática

Resumen – Se hablará de las distribuciones de variables aleatorias discretas las cuales nos sirven para describir la varianza de los resultados esperados en distintos experimentos. También estudiaremos el Teorema del Límite Central, el cual nos logrará relacionar las distribuciones estudiadas

I. INTRODUCCIÓN

Previamente a definir formalmente las distribuciones Gaussianas, explicaremos básicamente para qué sirven dichas distribuciones. Las distribuciones Gaussianas son distribuciones en variables aleatorias, las cuales nos da una idea de cómo poder esperar un resultado de un experimento con ciertos parámetros.

Algunas áreas en las que se pueden aplicar los experimentos son las siguiente: control de calidad, análisis de riesgos es una compañía, llamadas en una central telefónica, producción de alimentos, etc.

Para el presente trabajo desarrollaremos las definiciones formales de las Aproximaciones Gaussianas junto con algunas funciones y teoremas que sean necesarias para su correcto entendimiento. Adicionalmente a esto, se desarrollará un algoritmo en lenguaje R, el cual nos podrá ilustrar con ejemplos el correcto funcionamiento y aplicación de las Distribuciones Gaussianas a experimentos específicos, los cuales comprenderán algunas de las áreas anteriormente nombradas.

II. DEFINICIONES

II-A. Definición de distribuciones probabilísticas:

Una distribución de probabilidad la podemos concebir como una distribución teórica de frecuencia, es decir, es una distribución que describe como se espera que varíen los resultados. Dado que esta clase de distribuciones se ocupan de las expectativas son modelos de gran utilidad para hacer inferencias y tomar decisiones en condiciones de incertidumbre.

II-B. Distribución binomial:

Es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de *n* ensayos de Bernoulli independientes entre sí, con una probabilidad fija *p* de ocurrencia del éxito entre los ensayos.(para *n* = 1, la binomial se convierte en una distribución de Bernoulli).

La variable aleatoria ξ que representa el número de éxitos en una serie de "n" ensayos independientes, en cada uno de los cuales la probabilidad de éxito es constante e igual a "p", se denomina variable aleatoria binomial con parámetros (n,p).

La distribución de una variable aleatoria binomial se puede definir mediante la fórmula:

$$P(X=x)=C_n^x p^x (1-p)^{n-x} \ , \ \ x=0,1,\ldots,n$$

Y su función distribución es:

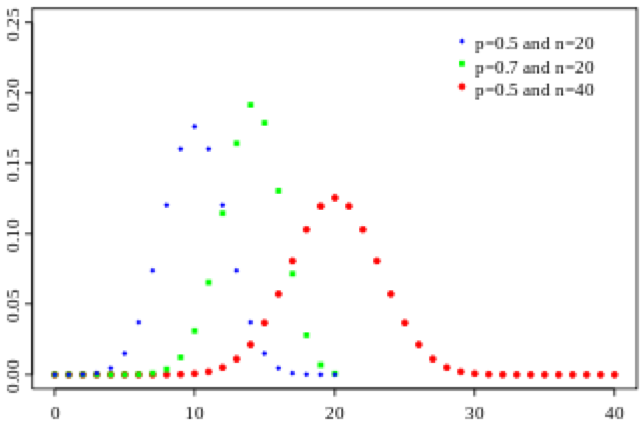
$$F(x)=\sum_{k=0}^{K<x} C_n^K p^k (1-p)^{n-k}$$

Ejemplo:

Supongamos que se lanza un dado (con 6 caras) 51 veces y queremos conocer la probabilidad de que el número 3 salga 20 veces.

En este caso tenemos que la probabilidad sería P(X=20):

$$P(X=51)=C_{20}^{51}\left(\frac{1}{6}\right)^{51}\left(1-\frac{1}{6}\right)^{51-20}=0.0000744\ldots$$



II-C. Distribución de Poisson:

La distribución de Poisson determina la probabilidad de ocurrencia de un resultado en el tiempo o en el espacio, esto es medir la probabilidad de ocurrencia de un evento sobre un intervalo de tiempo o espacio definido.

Consideremos que el experimento consiste en una sucesión de "n" ensayos independientes con probabilidad de éxito constante e igual a "p" en cada ensayo. Supongamos, además, que "p" es tan pequeño que un éxito es un suceso raro en cualquier ensayo, pero que el número "n" de ensayos es tan grande que la magnitud np=ƛ permanece constante de experimento en experimento.

Entonces:

$$P_x(k) = \mathrm{C}_{\mathrm{n}}^{\mathrm{k}}\mathrm{p}^{\mathrm{k}}(1-\mathrm{p})^{\mathrm{n-k}} = \frac{n(n-1)\ldots(n-k-1)}{k!}p^k(1-p)^{n-k}$$

$$\lambda = \mathrm{np} \;\rightarrow\; \mathrm{p} = \frac{\lambda}{n} \;\rightarrow\; p^k = \frac{\lambda^k}{n^k}$$

Distribuyendo 'n'

$$\frac{\lambda^k}{k!}\cdot\frac{n}{n}\cdot\frac{(n-1)}{n}\ldots\frac{(n-k)}{n}\cdot\left(1-\frac{\lambda}{n}\right)^{-k}\cdot\left(1-\frac{ƛ}{n}\right)^{n}$$

Cuando $n \rightarrow \infty$

$$\lim_{n\to\infty}\frac{n}{n}\cdot\frac{(n-1)}{n}\ldots\frac{(n-k)}{n}\cdot\left(1-\frac{\lambda}{n}\right)^{-k} = 1$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

Finalmente

$$P_x(k) = \frac{\lambda^k}{k!}\cdot e^{-\lambda}$$

Lo anterior motiva a formular la siguiente definición:

- **Definición:**

Se denomina variable aleatoria de Poisson con parámetro ƛ a toda variable aleatoria discreta con distribución

$P_x(k)=e^{-ƛ}\frac{ƛ^k}{k!}$, k=0,1,2,……..

Notación: ξ =Π[ƛ]

Esta definición es correcta, puesto que:

$$\sum_{\mathrm{k}=0}^{\infty} P_x(k) = e^{-ƛ}\sum_{\mathrm{k}=0}^{\infty}\frac{ƛ^k}{k!} = e^{-ƛ}.e^{ƛ} = 1$$

De esta manera, una variable aleatoria de Poisson representa el número de éxitos en una serie ilimitada de ensayos, suponiendo que la probabilidad de éxito en cada ensayo es muy pequeña.

Ejemplo:

La probabilidad de que haya un accidente en una compañía es de 0.02 por cada día de trabajo. Si se trabajan 300 días al año, ¿cuál es la probabilidad de tener 3 accidentes?

Para este caso, la probabilidad "p" es de

p = 0.02 → 2 % < 10 % → p < 0.10

El tamaño de muestra "n" vendría a estar dado

n = 300

p*n = 0.02 * 300 = 6 → p*n < 10

Gracias a estas condiciones, podemos aplicar el modelo de distribución de Poisson

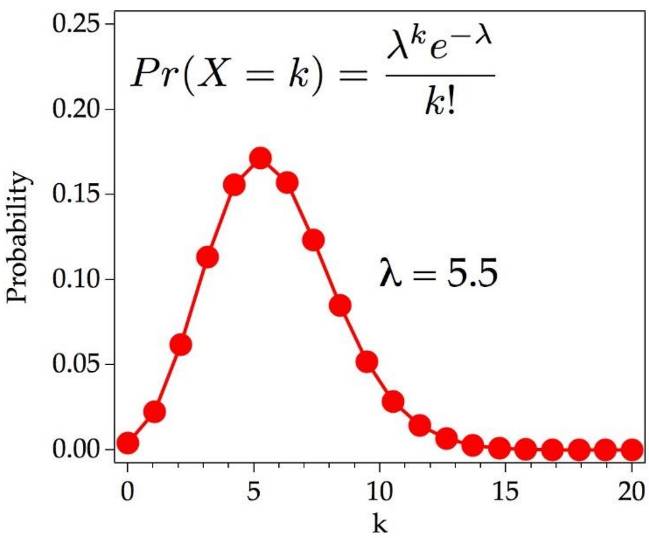
k = 3

λ = p*n = 6

$$P(X=3) = e^{-\lambda} \times \frac{\lambda^k}{k!}$$

$$P(X=3) = e^{-6} \times \frac{6^3}{3!}$$

$$P(X=3) = 0.0892$$

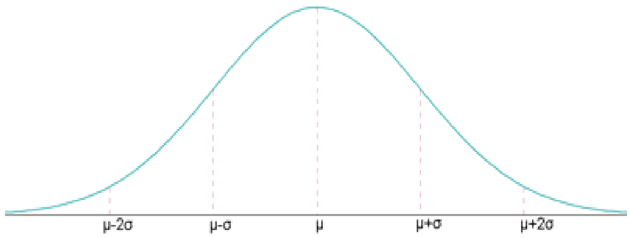


II-D. Distribución normal:

La distribución normal es la distribución de probabilidad más importante en Estadística, conocida por la cantidad de fenómenos que explica. Es la más frecuente utilizada en aplicaciones estadísticas, debido a su extensa utilización. Se le denomina Campana de Gauss ya que al representar su distribución probabilística tiene forma de campana.

Su importancia recae en que su aplicación es directa y permite observar muchas variables de interés, que pueden describirse fácilmente con este modelo, también sirve para acercarse a varias distribuciones de probabilidad discreta, entre estas la distribución de Poisson y la Binomial.

Muchas variables aleatorias siguen una Distribución Normal o cerca de la misma, pues su característica más resaltante es que la gran mayoría de distribución de probabilidad bien sea discreta o continua su característica tanto de la distribución como de la curva que lo representa son:



- La curva tiene forma de campana con un pico en el centro de distribución, por lo que la media aritmética, la moda y la mediana son iguales y se localizan en el pico.

- Es simétrica alrededor de su media, la mitad del área debajo se encuentra a la derecha de dicho punto central y la otra mitad está a la izquierda.

- Es asintótica, es decir la curva se acerca bastante al eje x y no la toca

II-E. Función de densidad

La expresión matemática que representa una función de densidad con estas características es la siguiente:

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Función de densidad que sólo depende de dos parámetros, la media μ, y la varianza de la distribución σ^2 . Si conocemos la media y la varianza de una variable X, podemos definir la distribución normal utilizando la notación: $X = N\,(\mu, \sigma^2)$

La Distribución Normal Estándar se cuando $\mu = 0 \; y \; \sigma = 1$

II-F. Función generatriz de momentos

La función generatriz de momentos se define como la esperanza de $e^{(tX)}$. Para una distribución normal, la función generatriz de momentos es:

$$M_x(t) = E[e^{tx}]$$
$$= \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} e^{tx}\, dx$$
$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

II-G. Función característica

La función característica se define como la esperanza de e^{itX}, donde i es la unidad imaginaria. De este modo, la función característica se obtiene reemplazando t por it en la función generatriz de momentos. Para una distribución normal, la función característica es

$$M_x(it) = E\left[e^{itx}\right] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} e^{itx}\, dx$$
$$= e^{i\mu t - \frac{\sigma^2 t^2}{2}}$$

II-H. Teorema del límite central

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas (discretas o continuas) de valor medio μ y varianza σ^2 , entonces la distribución de la variable

$$Z = \frac{\frac{(X_1 + X_2 + \cdots + X_n)}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}}$$

se aproxima a la de una normal estándar N(0,1), mejorándose la calidad de la aproximación a medida que 'n' aumenta.

Este resultado prueba que el estimador media muestral

$$\bar{X} = \frac{(X_1 + X_2 + \cdots + X_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

se distribuye aproximadamente como una variable

$$N(\mu, \frac{\sigma}{\sqrt{n}})$$

O de manera equivalente, que $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ se distribuye, aproximadamente como una variable N(0,1).

III. ESTADO DEL ARTE

- Tesis doctoral: “Análisis epistemológico de la variable aleatoria y comprensión de objetos matemáticos relacionados por estudiantes universitarios” – Autor: Blanca Ruíz Hernández. Universidad de Granada

- Distribuciones de probabilidad. Trabajo de investigación de la distribución de Poisson, la distribución exponencial y la distribución normal – Autor: Andrés Felipe Prieto Alarcón

- Tesis para el grado de Magister: “Tipificación de los errores que se presentan al identificar una variable aleatoria de distribución binomial en problemas contextualizados” – Autor: Marhori Vilca Álvarez. Pontificia Universidad Católica del Perú.

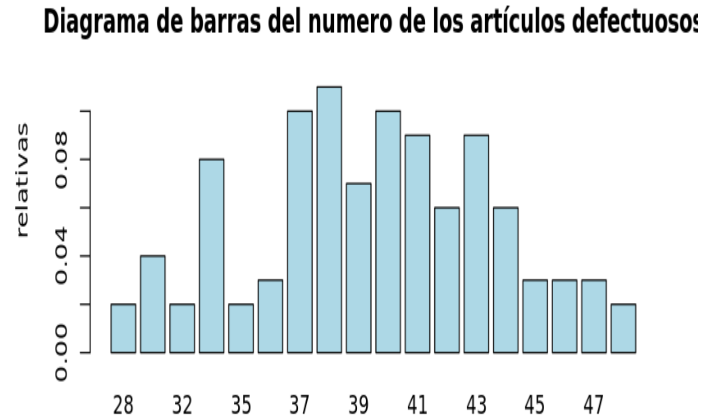
IV. DESARROLLO DEL EXPERIMENTO

- **Distribución binomial**

Para la distribución binomial, generaremos un ejemplo el cual está centrado al área de control de calidad.

Para un tema se debera elegir 100 articulos diarios con una probabilidad "p" de que dicho artículo sea defectuoso.

```
#Asignaremos el valor de "p" y "n"
p=0.4
n=100

#Asignaremos la cantidad de artículos a analizar
num.paq=100

#Aqui generaremos los 100 números aleatorios
bits.err=rbinom(num.paq,n,p)

frec.abs=table(bits.err) # Corresponde a las frecuencias absolutas
frec.rel=frec.abs/num.paq # Corresponde a las frecuencias relativas

#Dibujaremos una matriz de graficos de 2x1
par(mfrow=c(2,1))
```

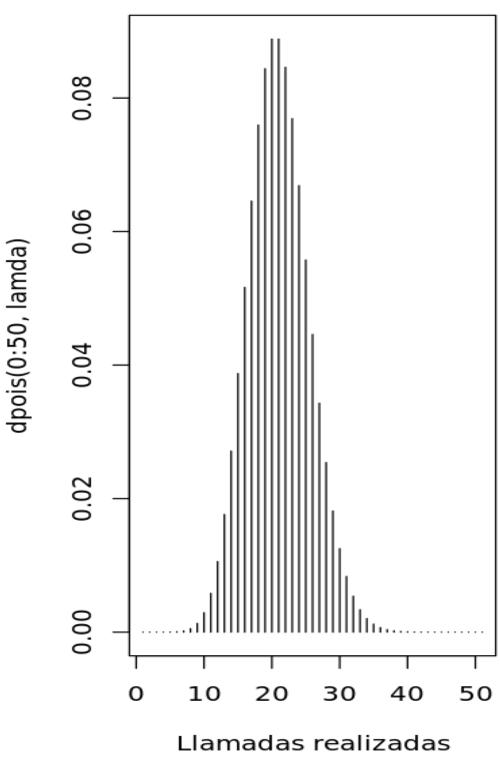


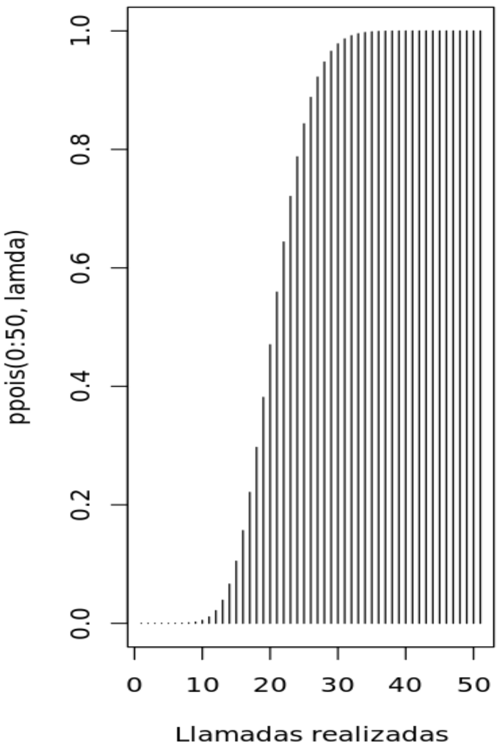
- **Distribución Poisson**

Para la distribución Poisson, el ejemplo se centrará en la cantidad de llamadas que recibe una central telefónica

Contabilizaremos el número de llamadas telefonicas en 1 hora, siendo lambda el número de llamadas.

```
#Asignamos el valor de 20 a lambda
lamda=20

#Observaremos las distintas probabilidades a partir de la función dpois
for (i in 0:50) cat(i,"\t",dpois(i,lamda),"\n") #el bucle for indica el número mínimo de interacciones, la función cat es para visualizar las prob.

#Dibujaremos una matriz de gráficos de 1x2
par(mfrow=c(1,2))

#colocaremos los datos en un grafico
plot(dpois(0:50,lamda),xlab="Llamadas realizadas",type="h") #xlab etiqueta al eje x,type="h" dibuja con lineas verticales

#Observaremos las distintas probabilidades a partir de la función acumulada ppois
for (i in 0:50) cat(i,"\t",ppois(i,lamda),"\n")
plot(ppois(0:50,lamda),xlab="Llamadas realizadas",type="h")
```






```
#Generaremos los valores desde el 0 hasta el 40
x=seq(0,40,by=1) #by=1 indica que la secuencia
aumentara de 1 en 1

#Generaremos un gráfico donde se indicarán los
valores de dbinom
plot(x,dbinom(x,40,0.4),pch=2,col=2) #El pch se
dibuja un símbolo especificado(2),col indica el color
especificado(2)

#Aproximando dbinom(x,n,p) y
dnorm(x,np,sqrt(npq)),q=1-p
np=40*0.4
u=sqrt(40*0.4*0.6) #u=sqrt(npq)

#Generamos el grafico formado por una línea continua
con los datos obtenidos
lines(x,dnorm(x,16,3.0983))

#Indicaremos la comparación de valores de dbinom y
dnorm
for (i in 0:40) cat(i,"\t",dbinom(i,40,0.4),"\t",
                    dnorm(i,16,3.0983),"\n")
```

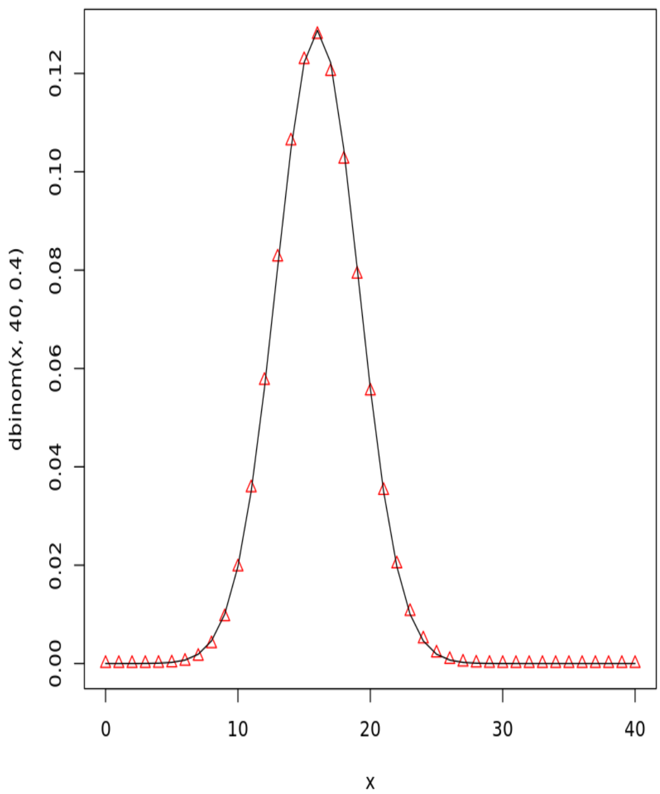
- **Distribución Normal**

Para la distribución Normal, nos apoyaremos del Teorema del Límite Central, el cual nos ayudará a aproximar las distribuciones antes vistas a una Distribución Normal

TEOREMA CENTRAL DEL LIMITE

- **Aproximación de D, Normal y D. Binomial**

Aproximación eficaz si "np > 10"
Asignaremos a los parámetros n (número de intentos) p(probabilidad)
n=40
p=0.4



- **Aproximación de D. Normal y D. Poisson**

Aproximación es eficaz si "λ>5"
λ=20

```
x=seq(0,40,by=1)

#Generaremos un gráfico donde se indicarán los valores de dpois
plot(x,dpois(x,20),pch=2,col=2)

#Aproximando dpois(x,lamda) y dnorm(x,lamda,sqrt(lambda))
v=sqrt(40) #v=sqrt(lambda)
lines(x,dnorm(x,20,4.4721))

#Indicaremos la comparación de valores de dpois y dnorm
for (i in 0:40)
cat(i,"\t",dpois(i,20),"\t",dnorm(i,20,4.4721),"\n")
```