5. DISTRIBUCIONES DE PROBABILIDADES

Dr. Edgar Acuña

http://academic.uprm.edu/eacuna

UNIVERSIDAD DE PUERTO RICO

RECINTO UNIVERSITARIO DE MAYAGUEZ

DISTRIBUCIONES DE PROBABILIDADES

Se introducirá el concepto de **variable aleatoria**, cuya importancia radica en introducir modelos matemáticos en el cálculo de probabilidades. Luego, se considerarán las principales distribuciones de probabilidades de variables aleatorias discretas y continuas con su media y varianza respectiva.

5.1 Variables Aleatorias

Una variable aleatoria es aquella que asume sus valores de acuerdo a los resultados de un experimento aleatorio. Usualmente se representa por las últimas letras del alfabeto: X, Y o Z. Una variable aleatoria X es una función cuyo dominio es la colección de eventos del espacio muestral S y cuyo rango R_x, es un subconjunto de los números reales.

Ejemplos de variables aleatorias:

- X: La suma que aparece al lanzar un par de dados.
- Y: El número de caras que aparecen al lanzar una moneda tres veces.
- Z: El número de errores que se encuentran en la página de un libro.
- T: El tiempo de vida de la componente de un sistema
- W: El tiempo de espera para ser atendido en un banco

Ejemplo 5.1

De una caja que contiene 5 bolas numeradas del 1 al 5 se extraen 3 bolas una por una y sin reposición. Entonces **X**: El mayor de los tres números sacados, es una variable aleatoria.

El espacio muestral es:

S = {(1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), (3,4,5)}

y la variable aleatoria X asume los valores: 3, 4 y 5. Por ejemplo,

$$X\left(2,3,4\right) = 4$$

Si el rango de valores R_x de la variable aleatoria X es finito o infinito enumerable entonces se dice que es una **variable aleatoria discreta**. Si su rango de valores R_x es infinito no enumerable entonces se dice que es una **variable aleatoria continua**.

5.1.1. Función de probabilidad de una variable aleatoria discreta

Si X es una variable aleatoria discreta con rango de valores R_x entonces, su función de probabilidad se define por:

$p(x) = P[X = x]$, para todo x R_x

y tiene las siguientes propiedades:

- $p(x) > 0$ y
- $\Sigma\, p(x) = 1, \quad x \in R_x$

Cuando Rx no contiene muchos valores es más conveniente expresar p(x) en una tabla de valores, la cual es llamada tabla de función de probabilidad.

Ejemplo 5.2.

De un lote que contiene 10 artículos, de los cuales 4 son dañados se extraen al azar y sin reposición 3. Se define la variable aleatoria X: Número de artículos dañados que hay en la muestra. Hallar la función de probabilidad de X.

Solución: En este caso el rango de valores de X es Rx = {0, 1, 2, 3} y en particular.

p(2) = Prob(sacar 2 dañados) = $\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}$, y en general p(x) = $\frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$, para x = 0,1,2,3.

Las combinaciones se muestran en la siguiente tabla de función de probabilidad:

X	p(x)
0	1/6
1	½
2	3/10
3	1/30

5.1.2. Función de distribución acumulativa

Sea X una variable aleatoria discreta con función de probabilidad p(x) y rango de valores R_x, entonces su función de distribución acumulativa se define por:

$$F(t) = P(X \leq t) = \sum_{x \leq t} p(x)$$

t es cualquier número real. En particular, si t es un valor que está en Rx , el cual consiste de enteros no negativos, entonces:

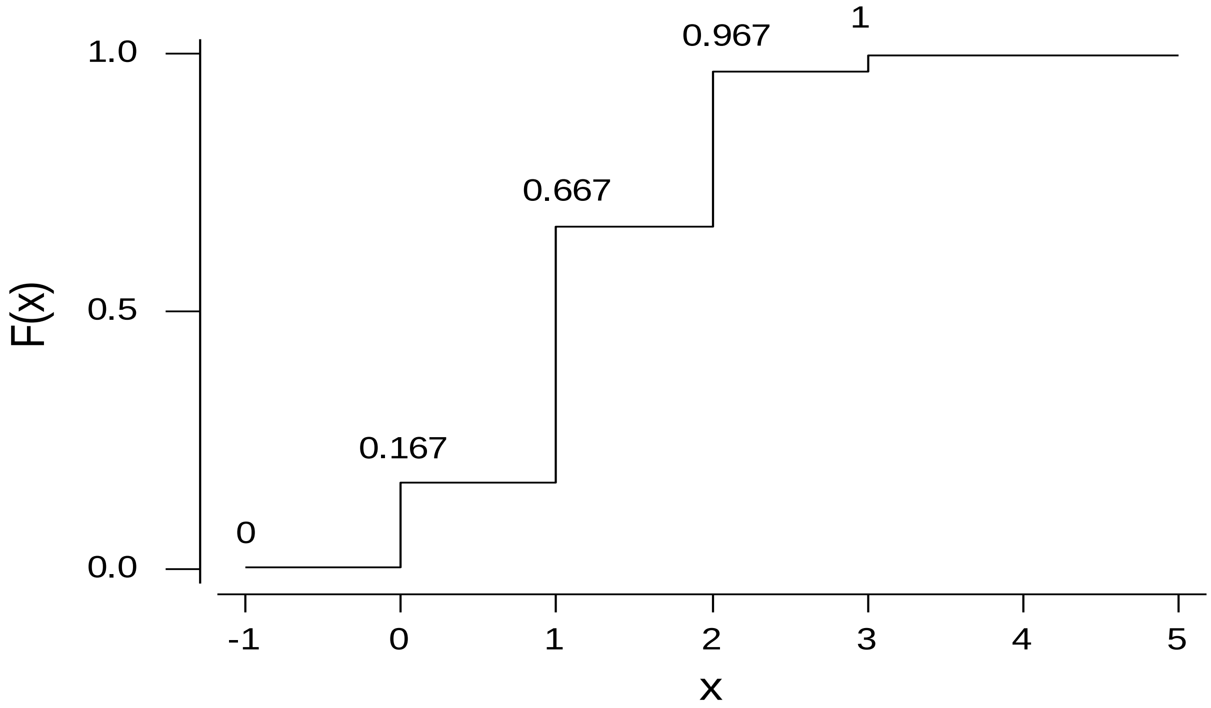
F(t) = p(0) + p(1) + p(2) + p(3) +…+ p(t)

Ejemplo 5.3. Hallar la función de distribución acumulativa para el Ejemplo anterior.

Solución:

x	F(x)
0	1/6
1	4/6
2	29/30
3	1

Grafica de la Distribucion Acumulada



La gráfica de una función de distribución acumulativa es creciente y del tipo escalonado, con saltos en los puntos que están en el rango de valores y cuya magnitud es igual al valor de la función de probabilidad en dicho punto. Más formalmente tiene la siguiente propiedad:

Propiedad. *La relación entre la función de distribución de probabilidad y la función de distribución acumulativa está dada por:*

$$p(x) = F(x) - F(x-1)$$

para todo valor de x en el rango de valores de la variable aleatoria.

5.1.3 Valor Esperado y Varianza de una Variable Aleatoria Discreta

Sea X una variable aleatoria discreta con función de probabilidad $p(x)$ y rango de valores Rx, entonces su Valor Esperado o Media se define como el número:

$$\mu = E(X) = \sum_x xp(x)$$

La suma es sobre todos los valores *x* que están en R_x.

La **Varianza** de una variable aleatoria discreta x con función de probabilidad p(x) y media μ se define por:

$$\sigma^2 = \sum (x-\mu)^2 p(x)$$

Donde la suma es sobre todos los valores del rango de X.

Ejemplo 5.4. Hallar la media y varianza del número de artículos dañados del Ejemplo 5.4.

Solución:

x	p(x)	Xp(x)	X-μ	$(x-u)^2p(x)$
0	1/6	0	-1.2	.24
1	½	½	-0.2	.02
2	3/10	6/10	0.8	.192
3	1/30	1/10	1.8	.121
		μ=1.2		σ^2=0.573

Otra formas del calcular la varianza es $\sigma^2 = \sum x^2p(x)-\mu^2$.

Principales distribuciones Discretas

- Distribucion Binomial
- Distribucion Hipergeometrica
- Distribucion Geometrica
- Distribucion Binomial Negativa o de Pascal
- Distribucion Poisson

5.2 La Distribución Binomial.

Un experimento es llamado de Bernoulli, si satisface las siguientes características:

- En cada repetición puede ocurrir sólo una de dos maneras, una de ellas es llamada *Exito* y la otra *Fracaso*.
- La probabilidad de *Exito*, representada por *p*, debe permanecer constante cuando el experimento es repetido muchas veces.
- Las repeticiones de los experimentos deben ser independientes entre sí.

Ejemplos:

- Observar las veces que sale 6 al lanzar varias veces un dado, en este caso la probabilidad de éxito es 1/6.
- Contar el número de pacientes que sobreviven a una operación de corazón abierto.
- Contar el numero de solicitantes de empleo a los que se les hace una oferta.

5.2 La Distribución Binomial.

Una variable aleatoria X se dice que tiene una distribución Binomial si define como el numero de Exitos que ocurren en n repeticiones de un experimento de Bernoulli

La función de probabilidad de una binomial es:

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

para $x = 0, 1, \ldots, n$.

El valor de $p(x)$ para diversos valores de n y p aparece en tablas de todo texto básico de Estadística.

Calculos en Python para la funcion de distribucion Binomial

La librería scipy.stats de **Python** permite calcular probabilidades binomiales, usando **from scipy.stats import binom**

1-La función de probabilidad (***Probability***) para cualquier valor de *n* y *p*.

usando el comando **binom.pmf**

2-La función de distribución acumalada (***Cumulative probability***) usando la funcion **binom.cdf**

3-Los percentiles ***(Inverse cumulative probability***) usando la funcion **binom.ppf**

4-Datos simulados de una binomial usando la funcion **binom.rvs**

La media (o valor Esperado) de una Binomial con parametros n y p es np y la varianza es npq.

Ejemplo 5.7

Expresar en una tabla de valores la función de probabilidad y la función de distribución acumulada de la variable aleatoria X: Número de preguntas bien contestadas por un estudiante que responde al azar un examen tipo selección múltiple que consiste de 10 preguntas, cada una con 4 alternativas de las cuales sólo una es correcta.

Lab 12

Distribucion Hipergeometrica

La distribucion hipergeometrica se aplica cuando de una poblacion que tiene dos tipos de objetos: m de un tipo I y n de un tipo II se extrae sin REEMPLAZO una muestra aleatorio de tamano de k y se define la v.a.
X: Numero de objetos del tipo I que hay en la muestra.
Su funcion de probabilidad esta dada por

$$P[X=i]=\frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}}$$

Para i=max(0,k-n),…,min(k,m). Cuando la muestra es pequena comparada con la poblacion las probabilidades de la hipergeometrica se acercan bastante a las probabilidades binomials.
En Python se usan las funciones hyper**geom.pmf, hypergeom.cdf, hypergeom.ppf y hypergeom.rvs**. Los parametros de la hipergeometrica son m, n y k. En Python prefieren usar M=m+n como un parametro, los otros serian m y k.

Distribucion Geometrica

La distribucion Geometrica se aplica cuando se repiten los experimentos de Bernoulli hasta que ocurra el primer exito y se define la v.a. X como el numero de fracasos que ocurren en el experimento hasta que ocurre primer exito.
La funcion de probabilidad esta dada por:

$P[X=k]=q^k p$

$k=0,1,2,3,4$
Tambien se puede definir la variable geometrica como el numero de repeticiones hasta que ocurre el prmer exito. En ese caso k comienza en 1. y en la formula la potencia es k-1. La libreria scipy.stats de Python lo define de esta forma. Se usan las funciones **geom.pmf, geom.cdf,geom.ppf y geom.rvs**. El parametro de la geometrica es la probabilidad de exito, p.

Distribucion Binomial Negativa o de Pascal

La distribucion Binomial negativa se aplica cuando se repiten los experimentos de Bernoulli hasta que ocurra el r-esimo exito y se define la v.a. X como el numero de fracasos que ocurren en el experimento hasta que ocurra el r-esimo exito.
La function de probabilidad esta dada por

$$P[X = k] = \binom{k+r-1}{k} q^k p^r$$

Donde k=0,1,2,3……

Tambien se puede definir la variable aleatoria Binomial Negativa como el numero de repeticiones hasta que ocurra el r-esimo exito. En este caso k comienza en 1 y en la formula la potencia de q es k-1.
En Python se usa las funciones **nbinom.pmf** para obtener las probabibilidades de una Binomial Negativa, **nbimon.cdf** para obtener las probabilidades acumuladas, **nbinom.qqf** para hallar los quantles y **nbinom.rvs** para simular datos de una **BN.**

Lab 13

5.4 La distribucion Poisson

Se dice que una variable aleatoria X tiene una distribucion Poisson con parametro $\lambda>0$ si su funcion de probabilidad esta dado por

$$P[X = k] = \frac{e^{-\lambda}\lambda^k}{k!}$$

Para k=0,1,2,3,….. Notar que $\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = e^0 = 1$

usando el desarrollo en series de la funcion exponencial.

Se puede mostrar que tanto el esperado como la varianza de una distribucion Poisson es λ
En Python se usa las funciones **poisson.pmf** para obtener las probabibilidades de una Poisson, **poisson.cdf** para obtener las probabilidades acumuladas, **poisson.qqf** para halar los quantles y **poisson.rvs** para simular datos de una Poisson.

Ejemplos de variables aleatorias Poisson

Una variable Poisson en la practica cuenta el numero de ocurrencia de eventos que tienen una probabilidad pequena de ocurrencia.

X: Numero de errores por pagina de un libro
X: Numero de llamadas que entran a un cuadro telefonico en un intervalo de tiempo dado
X: Numero de personas que hay en la fila de un banco en un intervalo de tiempo dado
X:Numero de accidentes que ocurren en una interseccion semanalmente.
X: Numero de ocurrencia de terremotos por ano

Ejemplo 5.8

El numero promedio de accidentes en una interseccion es 2 por semana. Asumiendo que el numero de accidentes por semana sigue una distribucion Poisson

a)Hallar la probabilidad de que no haya accidentes en una semana cualquiera

b)Hallar la probablidad de que ocurran a lo mas tres accidentes en un periodo de dos semanas.

Solucion: Sea X:numero de accidentes por semana en la interseccion. X es Poisson(λ=2)

a) $P(X=0)=e^{-2}2^0/0!=e^{-2}$=poisson.pmf(0,2)

b) Sea Y=numero de accidentes en un periodo de dos semanas. Y es una Poisson con $\lambda=2(2)=4$

$P(Y<=3)=P(Y=0)+P(Y=1)+P(Y=2)+P(Y=3)=e^{-4}+4e^{-4}+8e^{-4}+64^{e-4}/6=.4334$

=poisson.cdf(3,4)

Ejemplo 5.9

El numero de errores en un libro se distribuye como una variable aleatoria Poisson con λ=.5 errores por pagina

a) Cual es la probabilidad de que en una pagina elegida al azar hayan dos errores?
b) Cual es la probabilidad de que en un capitulo del libro, que tiene 30 paginas, hayan 5 errores?
c) Cual es la probabilidad de que en un capitulo del libro, que tiene 30 paginas, hayan por lo menos 10 errores?

Solucion: a) Sea X:numero de errores por pagina, X es Poisson(λ=.5)

$P(X=2)=e^{-.5}.5^2/2!=.0758$=poisson.pmf(2,.5)

b) Sea Y=numero de errores en un capitulo de 30 paginas. Y es una Poisson con λ=30(.5)=15

$P(Y=5)=e^{-15}15^5/5!=.00193$=poisson.pmf(5,15)

c) $P(Y>=10)=1-P(Y<=9)=1-\sum_{k=0}^{9}\frac{e^{-15}15^k}{k!}=1-.0698=.9302$=1-poisson.cdf(9,15)

Aproximacion Poisson a la Binomial

Las probabilidades de una binomial con n grande pueden ser aproximadas usando una Poisson con parametro λ=np. Es decir,

$$\lim_{n\to\infty}\binom{n}{k}p^k(1-p)^{n-k}=\frac{e^{-\lambda}(\lambda)^k}{k!} \qquad \text{donde} \quad p=\frac{\lambda}{n}$$

Comparacion de probabilidades binomiales B(10,.1) y las probabilidades Poisson(λ=1)

x	bin	poiss
0	0.348678	0.367879
1	0.387420	0.367879
2	0.193710	0.183940
3	0.057396	0.061313
4	0.011160	0.015328
5	0.001488	0.003066
6	0.000138	0.000511
7	0.000009	0.000073
8	0.000000	0.000009
9	0.000000	0.000001
10	0.000000	0.000000

Ejemplo 5.10

Estimar la probabilidad de que en un grupo de 400 personas por lo menos 3 celebren sus cumpleanos el 4 de julio. Asumir que hay 365 dias en el ano y que cada uno de ellos es igualmente probable que sea el dia de cumpleanos de una persona dada.

Solucion:
Sea X: el numero de personas que cumplen anos el 4 de julio. Propiamente X es una binomial con n=400 y p=1/365. Asi que habria que calcular
P(X>=3)=1-P(X<=2)=1-.9014=.0986.=1-binom.cdf(2,400,1/365)
Pero, un calculo mas facil , es usando la aproximacion Poisson, con λ =400/365=1.0958
$P(X>=3)=1-e^{-1.09}-1.09e^{-1.09}-1.09^{2}e^{-1.09}/2=1-.9012=.0988=1-$ poisson.cdf(2,400/365)

Lab 14

Distribuciones continuas

Una variable aleatoria X es continua si su rango de valores cualquuier intervalo de la recta real.
La probabilidad de X asuma valores en un intervalo dado esta determinado por una funcion real no negativa llamada funcion de densidad f(x).
Mas especificamente,

$P(a<X<b)=\int_a^b f(x)dx$

que es lo mismo que el area bajo la curva f(x) desde x=a hasta x=b.

La funcion de distribucion acumulada de la v.ac.X se define por
$F(t)=P(X<=t)=\int_{-\infty}^t f(u)du$

Notar, que $P(a<X<b)=F(b)-F(a)$

5.5. La distribucion Uniforme o Rectangular

Una variable aleatoria continua X se dice que se distribuye uniformememte en el intervalo (a,b) si su funcion de densidad esta dada por

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

$$f(x) = 0 \quad e.o.c(en-otro-caso)$$

El caso mas sencillo es cuando a=0 y b=1, se representa por U(0,1)
Esta distribucion es muy usada en la generacion aleatoria de todas las distribuciones conocidas en probabilidades.
En Python existen las funciones **uniform.pdf**, que da el valor de la funcion de densidad en un punto, **uniform.cdf** que da la acumulada, **uniform.ppf** que da el quantile y **uniform.rvs** que genera valores aleatorios distribuidos uniformemente. Hay que usar los parametros loc=a y scale=b-a.

Valor esperado y Varianza de una variable aleatoria Uniforme

Si $X \sim U(a,b)$ entonces su media es $E(X)=(a+b)/2$, es el punto medio del intervalo.

Su varianza esta dada por $Var(X)=(b-a)^2/12$

Ejemplo 5.11

Un punto se elige al azar sobre el segmento de linea [0,2] . Cual es la probabilidad de que el punto escogido quede entre 1 y 3/2

Solucion:

Sea X:numero elegido al azar entre 0 y 2. Entonces la densidad de X esta dada por

$$f(x)=1/2 \quad \text{si } 0<x<2$$
$$=0 \quad \text{e.o.c}$$

Hay que hallar

$$P(1 < x < 3/2) = \int_{1}^{3/2} \frac{1}{2}\,dx = \frac{1/2}{2} = 1/4$$

Esto es lo mismo que uniform.cdf(3/2,0,2)-uniform.cdf(1.0,2)

5.6 Generacion Aleatoria de una Binomial

Procedimiento:

1- Generar una variable aleatoria Bernoulli X con probabilidad de exito p. Usando los siguientes pasos:

1a Generar una una uniforme U en el inntervalo (0,1)

1b si U<=p entonces X=1 de lo contrario X=0

2- Generar una variable aleatoria Binomial Y con parametros n y p. usando los siguientes pasos:

2a Generar n variable aleatorias Bernoulli X1, ….Xn

2b. Y=X1+x2+………+Xn

Generacion Aleatoria de una Geometrica y Binomial Negativa

Procedimiento para la geometrica

Generar variables aleatoria Bernoulli X1, X2,… con probabilidad de exito p y anotar el primer indice I para el cual X_I=1. Entonces la variable Geometrica Y=I.

Procedimiento para la binomial Negativa

Hay que usar el hecho que que una binomial negativa con parametros r y p.es igual a la suma de r geometricas X1+X2+…..+Xr
Es decir se generan r geometricas y leugo se suman y eso da una variable aleatoria Binomial negativa r con p.

Lab 15

5.7. La distribucion Exponencial

Una variable aleatoria se dice que tiene una distribucion exponencial con parametro θ si su funcion de densidad esta dada por

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \qquad x > 0$$

Notar que

$$\int_0^\infty f(x)dx = \int_0^\infty \frac{1}{\theta} e^{-x/\theta} dx = -e^{-x/\theta} \Big|_0^\infty = 0 + 1 = 1$$

θ es llamado el parametro de escala.Muchas veces por conveniencia se usa λ=1/θ, e
En Python, la funcion expon.pdf(x, θ) da la densidad de una exponencial, en tanto que expon.cdf, da el area acumulada hasta un valor dado, expon.ppf, da el quantil y expon.rvs(θ,n), genera aleatoriamente n valores de una exponencial con parametro de escala θ.

Ejemplo 5.13

Suponga que el tiempo para que falle un componente esta distribuido exponencialmente con parametro 100.
a) Cual es la probabilidad de que un componente dure entre 50 y 100 horas?
b) Cual es la probabilidad de que un componente dure mas de 120 horas?
Solucion:
Sea X:el tiempo de duracion del componente. Luego

$$f(x) = \frac{1}{100} e^{-x/100}$$

$$a) P(50 < X < 100) = \int_{50}^{100} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{50}^{100} = e^{-.5} - e^{-1}$$

=expon.cdf(100,1/100)-expon.cdf(50,1/100)=.2386

$$b) P(X > 120) = \int_{120}^{\infty} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{120}^{\infty} = e^{-1.2}$$

=1-expon.cdf(120,1/100)=.3011

Media y Varianza de una Exponencial

Si $X \sim exp(\theta)$ entonces

a) $F(t)=1-exp(t/\theta)$ para $t>0$ y $F(t)=0$ e.o.c

b) $E(X)= \theta$ y

c) $Var(X)= \theta^2$

d) Propiedad de falta de memoria: Si $X \sim exp(\theta)$ entonces $P(X>s+t/X>t)=P(X>s)$

e) Relacion entre la Exponencial y la Poisson: Si el tiempo de inter-llegadas de los clientes a un servicio es exponencial con parametro λ entonces el numero de clientes en el sistema en un instante t es Poisson con parametro λt

5.3 La Distribución Normal

Es llamada tambien Distribución Gaussiana en honor a K. Gauss, es una distribución del tipo continuo. Su comportamiento es reflejado por la Curva Normal que es la gráfica de la siguiente ecuación.

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

Donde la media μ y la desviación estándar σ son los parámetros de la distribución.

Si una variable aleatoria X tiene una distribución Normal, para calcular la probabilidad de que X caiga entre dos valores a y b entonces, hallar el área debajo de la curva entre a y b, y se puede hacer usando Integración.

Calculos de la distribucion Normal en Python

En **Python** se pueden calcular la función de densidad (***norm.pdf***), la función de distribución acumulada (***norm.cdf***) y los percentiles ***(norm.ppf***) de la distribución Normal para cualquier valor de la media μ y desviación estándar σ. No se requiere transformación a una normal estándar. Las funciones usan los parámetros loc=media y scale=desviación estándar.

La normal mas conocida es la que tiene media=0 y desviación estándar=1 y es llamada la Distribucion Normal Estandar.

Estandarización de una Normal

Dada una variable aleatoria X distribuida Normalmente con media μ y desviación estándar σ entonces puede ser convertida a una normal estándar mediante el proceso de estandarización, definido por $Z = (X - \mu)/\sigma$, donde $X \sim N(\mu, \sigma^2)$. Además si Xp y Zp representen sus respectivos percentiles entonces:

$$Xp = \mu + \sigma Zp$$

Fórmulas para calcular área debajo de la curva normal

F representa la distribución acumulada de la distribución Normal, es decir el área acumulada a la izquierda del valor dado

a) $P(X < a) = F(a)$=norm.cdf(a, μ, σ)

b) $P(a < X < b) = F(b) - F(a,)$=norm.cdf(b, μ, σ)-norm.cdf(a, μ, σ)

c) $P(X > b) = 1 - F(b)$=1-norm.cdf(b, μ, σ)

Ejemplo 5.17

Si X es una población Normal con media $\mu = 70$ y $\sigma = 10$. Hallar las siguientes probabilidades:

a) P (X < 60)

b) P (X > 95)

c) P (50 < X < 80)

Solución:

Usando **Python** con **mean** = 70 y **standard deviation** = 10, se tiene que:

a) P (X < 60) = F (60) = norm.cdf(60,70,10)=.1587

b) P (X > 95) = 1 – F (95) =1-norm.cdf(95,70,10)= 1 - .9938 = .0062

c) P (50 < X < 80) = F (80) – F (50) = norm.cdf(80,70,10)-norm.cdf(50,70,10)=.8413 - .0228 = .8185

Laboratorio 16

5.4 Cotejando si hay Normalidad

Cuando se trata de sacar conclusiones acerca de la población usando los datos de la muestra, se asume generalmente que los datos de la población se distribuyen de forma normal. Como no se conocen todos los elementos de la población, se deben usar los datos de la muestra para verificar si efectivamente la población es Normal. Existen varias pruebas estadisticas para verificar Normalidad.

En **Python**, La librería stats de scipy contiene una función **probplot** que permite hacer el llamado plot de normalidad para determinar si hay normalidad en los datos.

Si los puntos caen cerca de la línea diagonal, entonces se dice que hay **Normalidad**.

Lab 16. Aproximacion Normal a la Binomial

Si x es una binomial con parametros n y p. Entonces se puede mostrar que La variable $\frac{X-np}{\sqrt{npq}}$ se distribuye aproximadamente como una normal estandar cuando n>30. Si p es cerca de .5 la aprixmacion es mejor.

Como X es discreta y la Normal es continua se acostumbra a aplicara una factor de correcion de ½ antes de usar la aproximacion normal