PARTE PRIMERA: DISTRIBUCIONES DE PROBABILIDAD

TEMA I: ESTUDIO DE ALGUNAS VARIABLES ALEATORIAS

I.1.- Variables aleatorias discretas

I.1.1.- Introducción

El objetivo de este apartado es abordar el estudio de algunas distribuciones de probabilidad de variables aleatorias discretas, concretamente las siguientes distribuciones:

- Distribución Uniforme
- Distribución Binomial
- Distribución de Poisson

Cuando nos planteamos estudiar estas distribuciones de probabilidad, lo hacemos partiendo de la base que su estudio nos permitirá simplificar el tratamiento estadístico de muchos fenómenos reales. De esta manera, si nosotros nos encontramos con un fenómeno real tal y como puede ser realizar una inversión o no. Este es un fenómeno que tiene dos posibles valores, invertir, no invertir. Bien, veremos que este tipo de fenómenos los podemos estudiar como una variable o distribución de Bernuille. Si nosotros hemos estudiado esta variable tendremos perfectamente identificados tanto la media como la varianza como su función de cuantía, etc... Es decir, conocemos el comportamiento probabilístico de este fenómeno. Si nos ponemos a pensar en fenómenos económicos reales, veremos que existen muchos que se pueden ajustar a un comportamiento de este tipo. Todos ellos están estudiados simultáneamente mediante la distribución de Bernuilli o la generalización binomial.

Por tanto, cuando estudiamos la distribución binomial, estamos estudiando miles de posibles distribuciones. Lo mismo pasará con el resto de distribuciones que analizaremos.

Para abordar el estudio de estas distribuciones el alumno deberá repasar los siguientes conceptos:

- Variable aleatoria
- Variable aleatoria discreta
- Función de distribución y propiedades de la misma
- Función de cuantía de una variable aleatoria discreta
- El operador esperanza matemática.
- Media y varianza de una variable aleatoria

El estudio de este tema servirá al alumno para:

- Conocer y describir las características de cada una de las funciones de distribución indicadas.
- Determinar qué función de distribución utilizar para cada situación concreta.
- Identificar que fenómenos reales se pueden ajustar a cada una de las distribuciones estudiadas.
- Trabajar de forma abstracta con fenómenos económicos.

I.1.2.- Distribución uniforme discreta

Decimos que una variable aleatoria discreta (X) tiene distribución uniforme cuando la probabilidad en todos los puntos de masa probabilística es la misma; es decir, cuando todos los posibles valores que puede adoptar la variable (x_1, x_2,...,x_k) tienen la misma probabilidad.

Pongamos el socorrido pero útil caso del lanzamiento de un dado. Si definimos una variable aleatoria (X) como el número resultante tras su lanzamiento, los valores que puede tomar esa variable aleatoria son {1, 2, 3, 4, 5, 6}. Pues bien, esa variable aleatoria tiene distribución uniforme si, como es el caso, la probabilidad es la misma para cada uno de los resultados posibles.

I.1.2.1.- Función de cuantía. Representación gráfica.

En vista de lo dicho, la función de cuantía de una variable aleatoria discreta con distribución uniforme será:

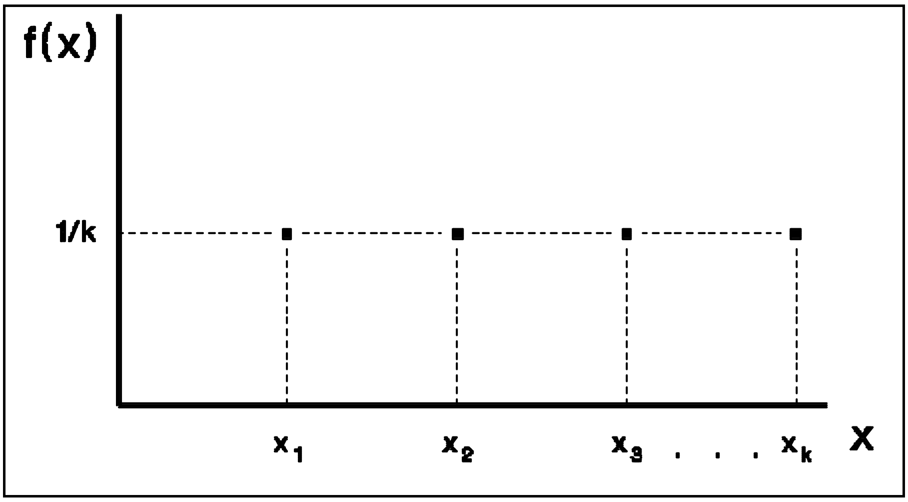
$$f(X) = \left\{ \begin{array}{ll} 0 \;\; si\; x \neq x_i & i = 1,2,...,k \\ \dfrac{1}{k} \;\; si\; x = x_i & i = 1,2,...k \end{array} \right\}$$

En nuestro sencillo ejemplo del lanzamiento de un dado, la función de cuantía, es decir, la probabilidad de que salga un resultado determinado será:

$$f(x) = \begin{cases} 1/6 \quad si \quad X = x_i \quad (i = 1,2,3,4,5,6) \\ 0 \text{ en otro caso} \end{cases}$$

La representación gráfica de la función de cuantía es muy sencilla e inmediata.

Suponiendo que $x_1 < x_2 < x_3 < .... < x_k$



I.1.2.2 .- Función de distribución. Representación gráfica.

Recordemos el concepto de función de distribución: la función de distribución mide la probabilidad de que la variable adopte valores iguales o inferiores a uno dado. Por tanto

$$F(x) = P\ (X \leq x_i) \quad (i= 1,\ 2,\ 3, \ldots, k)$$

En su expresión analítica, la función de distribución vendrá dada como

$$F(x)= \sum_{r=1}^{i} f(x_r)$$

Apliquemos esta fórmula a nuestro ejemplo del dado. Para resultados teóricos inferiores al valor 1, la función de cuantía, y la de distribución valen 0; para resultados iguales a 1 la función de cuantía y la de distribución valen 1/6; cuando el resultado es 2, la función de cuantía vale 1/6, pero la de distribución, que es la probabilidad de que la variable adopte resultados iguales o inferiores a 2, vale 2/6, etc....

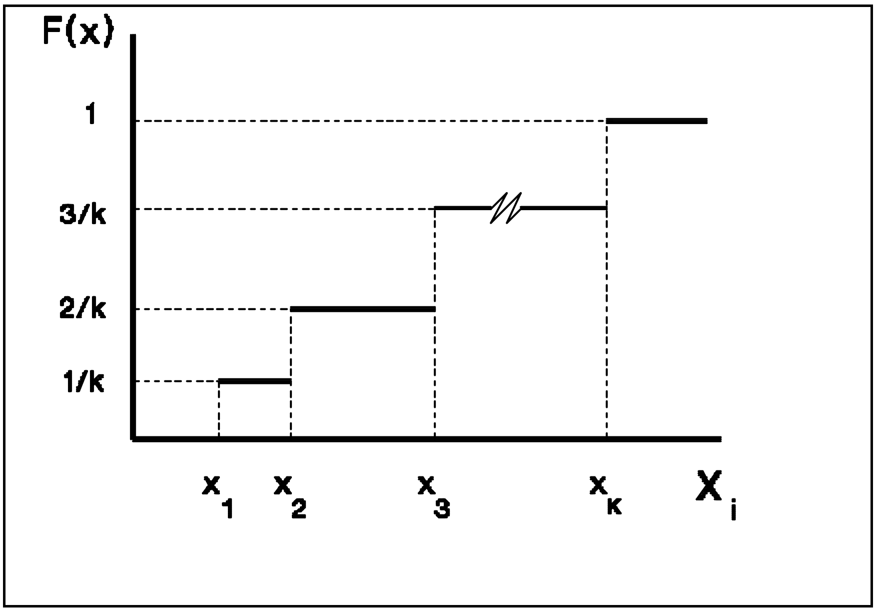
Si generalizamos este razonamiento obtendremos la expresión de la función de distribución, que será:

$$F(x)=\begin{cases} 0 & si\ x < x_1 \\ \frac{1}{k} & si\ x_1 \leq x < x_2 \\ \frac{2}{k} & si\ x_2 \leq x < x_3 \\ & . \\ & . \\ & . \\ 1 & si\ x \geq x_k \end{cases}$$

Es decir,

$$F(X)=\begin{cases} 0 & si\ x < x_1 \\ \frac{i}{k} & si\ x_i \leq x < x_{i+1} \\ 1 & si\ x \geq x_k \end{cases}$$

La representación gráfica de la función de distribución es, también, sencilla e inmediata:



I.1.2.3.- Media y varianza.

La media de esta distribución puede ser obtenida como una media aritmética de los valores que toma la variable (x_1, x_2, ..., x_k).

$$m_x = E(X) = \sum_{i=1}^{k} x_i\, f(x_i) = \frac{1}{k}\sum_{i=1}^{k} x_i = a_1$$

La varianza se obtiene de la forma ya conocida; es decir, como la varianza de esos mismos valores.Expresada en términos de momentos, la varianza será:

$$s_x^2 = V(X) = \frac{1}{k}(x_i - \overline{X})^2 = E(X^2) - [E(X)]^2 = a_2 - a_1^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \left(\frac{1}{k}\sum_{i=1}^{k} x_i\right)^2$$

I.1.3.- Distribucion binomial

Una buena parte de los fenómenos que ocurren en la vida real pueden ser estudiados como una variable aleatoria discreta con distribución binomial, por lo que su estudio puede ser de gran utilidad práctica.

Pero antes de pasar a la consideración de la Binomial, es conveniente que nos detengamos un momento a observar el denominado proceso de Bernouilli, fundamento de la distribución Binomial, de Poisson y de otras que no veremos en este curso.

I.1.3.1.- Proceso de bernouilli

Para comprender el proceso de Bernouilli pensemos, por ejemplo, en situaciones en las que sólo hay dos posibles resultados mutuamente excluyentes (verdadero/falso, en un test; defectuoso/no defectuoso, en los artículos que salen de una fábrica; aprobado/suspendido, en los resultados de un examen,etc....). Decimos que son mutuamente excluyentes porque no pueden darse simultáneamente (un examen no puede estar aprobado y suspendido al mismo tiempo; una respuesta no puede ser simultáneamente verdadera o falsa, etc...). Una manera común de designar estos dos resultados es como Exito (E) o Fracaso (F).

Una segunda característica de los fenómenos que siguen el denominado Proceso de Bernouilli es que las pruebas de las que se obtienen los éxitos o los fracasos son independiente. Así, el hecho de que un artículo salga defectuoso en una línea de producción no tiene que ver con el resultado obtenido en el siguiente artículo que examinamos.

Por último, una tercera característica de este Proceso es que las probabilidades de Exito o Fracaso son constantes.

Los fenómenos que en la vida real cumplen estas tres características pueden ser considerados como Procesos de Bernouilli.
Llamemos p a la probabilidad de éxito: $P(E) = p$

y llamemos q a la probabilidad de fracaso: $P(F) = q$

Definamos ahora una variable aleatoria, tal que

$x_i = 1$ si el resultado es éxito
$x_i = 0$ si el resultado es fracaso.

entonces

$$P(E) = P(X=1) = p$$
$$P(F) = P(X=0) = q$$

Tal como hemos definido las probabilidades es fácil concluir que

$$q = 1-p$$

Calculemos ahora la media de esa variable aleatoria:

$$m_x = E(X) = \sum_{i=1}^{2} x_i f(x_i) = 1 * p + 0 * q = p$$

y, calculemosla Varianza

$$s_x^2 = V(X) = \sum_{i=1}^{2} (x_i - m_x)^2 f(x_i) = \sum_{i=1}^{2} (x_i - p)^2 f(x_i)$$

$$s_x^2 = (1-p)^2 p + (0-p)^2 q = (1 + p^2 - 2p)p + p^2 q = p + p^3 - 2p^2 + p^2(1-p)$$

$$s_x^2 = p + p^3 - 2p^2 + p^2 - p^3 = p - p^2 = (1-p)p = pq$$

Por tanto la media de una variable de Bernuilli vale p y su varianza p*q.

Supongamos que estamos estudiando si una familia de Las Palmas de Gran Canaria tiene radio o no. En este caso nos encontramos con una distribución de Berniulli. Sin embargo, en Las Palmas de Gran Canaria hay más de una familia, por tanto, vamos a denotar por x_i a la familia i-ésima. Bajo este esquema de trabajo, llamaremos sucesión de Bernoulli a aquella serie que viene dada por $(x_1, x_2, ..., x_n)$, en donde cada x_i indica si la familia i tiene radio, en este caso tomará el valor 1, o no tiene rario, en este caso tomará el valor 0. Por otra parte, el que la familia i tenga rardio no afecta a que la familia j la tenga o no, es decir, x_i es independiente de x_j. Y, además,

$$m_{x_i} = E(x_i) = p$$

$$s_{x_i}^2 = V(x_i) = pq$$

I.1.3.2 .- Distribución binomial

Una vez visto el proceso de Bernouilli y la sucesión de Bernuilli estamos en disposición de abordar el estudio de la distribución Binomial.

Sea un experimento aleatorio en el que pueden obtenerse dos resultados posibles, mutuamente excluyentes, con probabilidades constantes en el que p es la probabilidad de éxito.

Supongamos que se realizan n pruebas independientes (es decir, se dan las condiciones de Bernouilli) y tenemos una sucesión de Bernuilli de tamaño n. Sea X la variable definida como el número de éxitos resultantes en la sucesión de Bernuilli. X diremos que se distribuye como una distribución binomial. Pensemos, por ejemplo, en un agente de seguros que tiene como posibles resultados de su gestión hacer un seguro (éxito) o no hacerlo (fracaso); o en un test que debemos hacer para acceder a un master con posibles respuestas correctas o incorrectas; o en la posibilidad de que acepten (éxito) o no acepten (fracaso) un grupo de personas una invitación a cenar; o una máquina etiquetadora de botellas que pone mal o bien las etiquetas; etc... Todas ellas, si son una sucesión de Bernuilli de tamaño n, se distribuyen como una distribución binomial de parámetros n y p, y lo denotaremos como

$$X \sim B(n,p)$$

En donde n es el tamaño de la sucesión de bernuilli, el número de veces que se repite el experimento (número de familias de Las Palmas de Gran Canaria, etc..), y p es la probabilidad del éxito.

La expresión formal de la función de cuantía de una distribución binomial es

$$f(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & si\ x = x_i \\ 0 & si\ x \neq x_i \end{cases}$$

$$donde\ ,\ p = probabilidad\ del\ éxito.$$

$$q = probabilidad\ del\ fracaso.$$

¿Cómo llegamos a esa expresión de la función de cuantía?

Volvamos ejemplo de la máquina etiquetadora de botellas y supongamos que queremos estudiar el resultado "poner bien la etiqueta". Esa será la variable aleatoria a estudiar, por lo que ese será el "éxito" de la distribución binomial y su probabilidad será p.
Tomemos una muestra de 6 botellas con etiqueta. ¿Cuál es la probabilidad de que una sola etiqueta este correctamente colocada?

Se nos pide la probabilidad del suceso:

A= (100000) U (010000) U (001000) U (000100) U (000010) U (000001)

donde el 1 denota el "éxito", es decir una etiqueta bien puesta y el 0 denota el "fracaso", es decir, una etiqueta mal puesta.

Al ser los sucesos disjuntos, la probabilidad de la unión es igual a la suma de las probabilidades. Por tanto:

P(A) = P(100000) + P(010000) + P(001000) + P(000100) + P(000010) + P(000001).

Cojamos el primer suceso. Por ser los sucesos independientes:

P(100000) = P(1)*P(0)*P(0)*P(0)*P(0)*P(0)

recordemos que la probabilidad de éxito la denotábamos p (P(1)=p) y la de fracaso q (P(0)=q), por tanto:

P(100000) = p * q * q * q * q * q = p * q^5

A este mismo resultado llegaríamos con el segundo y restantes sucesos, por lo que

$$f\ (x=1) = P(1) = 6 * p * q^5$$

Esta es la respuesta a la pregunta ¿Cuál es la probabilidad de que una sola sea correcta?.

El razonamiento será similar para la probabilidad de que dos resulten correctas. En este caso habrá tantos sucesos como formas posibles de ubicar dos "éxitos" (1 en nuestro razonamiento) en 6 posiciones. Será, por tanto, combinaciones de seis elementos tomados de dos en dos.

Si generalizamos el razonamiento llegaremos a la formulación de la función de cuantía para la distribución Binomial.

El alumno deberá practicar el manejo de la distribución Binomial haciendo uso del paquete STATG, comprobando la forma que adopta la función de cuantía y los cambios que se producen al variar el número de experimentos y la probabilidad de éxito.

La función de distribución de la Binomial adopta la forma:

$$F(x) = P(X \le x) = \sum_{i \le x} f(i) = \sum_{i \le x} \binom{n}{i} p^i q^{n-i}$$

Esta expresión, ciertamente, no es muy manejable, por lo que utilizaremos el paquete informático STATG para su cálculo.

Hemos visto ya la media y la varianza en el proceso de Bernouilli, por lo que la particularización a la distribución Binomial es sencilla.

$$Sea \;\; x_i = \left\{ \begin{array}{l} 1 \quad si \; éxito \\ \\ 0 \quad si \;\; fracaso \end{array} \right\}$$

$$donde, \quad m_{xi} = E(x_i) = p$$

$$s^2_{x_i} = V(x_i) = pq$$

$$Por \; tanto, X = n” \; de \; éxitos = x_1 + x_2 + ... + x_n \, ,$$

$$siendo \; los \; x_i \; independientes.$$

$$- MEDIA : m_x = E(X) = E(x_1 + x_2 + ... + x_n) = \sum_{i=1}^{n} E[x_i] = \sum_{i=1}^{n} p = np$$

$$- VARIANZA : s^2_x = V(X) = V[x_1 + x_2 + ... + x_n] = \sum_{i=1}^{n} V(x_i) \Rightarrow$$

$$\Rightarrow s^2_x = \sum_{i=1}^{n} pq = npq$$

En definitiva, en la distribución Binomial conociendo n y p queda perfectamente definido el comportamiento probabilístico de X.

Veamos un ejemplo de Distribución Binomial. Un reciente estudio de la Asociación Americana de Conductores de Autopista ha revelado que el 60% de los conductores norteamericanos usa regularmente el cinturón de seguridad. Se selecciona una muestra de 10 conductores en una autopista del estado de Oklahoma.

a) ¿Cuál es la probabilidad de que exactamente siete de ellos lleven el cinturón de seguridad?
b) ¿Cuál la probabilidad de que al menos siete de los conductores lleven el cinturón de seguridad?

Resolución:

Debemos determinar primero de qué tipo de distribución se trata. Veamos:

* Solamente hay dos posibles resultados en cada una de las comprobaciones que se hacen a los conductores: llevan el cinturón de seguridad (resultado que denominaremos "éxito") o no lo llevan ("fracaso").

* La probabilidad de "éxito" (llevar el cinturón) es la misma e invariable : 60%.

* Las pruebas son independientes: si el cuarto conductor que es parado no lleva el cinturón de seguridad, eso no condiciona el resultado de la comprobación para el quinto conductor que sea parado.

Cumple, por tanto, las condiciones del Proceso de Bernouilli, en el cual definimos una variable aleatoria que es "número de conductores que llevan el cinturón", es decir, "número de éxitos". Se trata, por tanto, de una distribución Binomial con n=10 y p=0.6.

a) Para calcular la probabilidad de que exactamente siete conductores lleven puesto el cinturón de seguridad debemos hacer uso de la función de cuantía, cuya expresión genérica es

$$f(X=x)=\begin{cases}\begin{pmatrix}n\\x\end{pmatrix}p^x q^{n-x} & si\ x = x_i\\ 0 & si\ x \neq x_i\end{cases}$$

$$donde,\ p = probabilidad\ del\ Çxito.$$

$$q = probabilidad\ del\ fracaso.$$

$$f(X=7)=\begin{pmatrix}10\\7\end{pmatrix}0{,}6^7 * 0{,}4^3 =$$

$$\frac{10!}{7!\,(10-7)!}\,0{,}6^7 * 0{,}4^3 = 0{,}215$$

Esta expresión, aplicada al problema que nos ocupa será:

b) La probabilidad de que como máximo siete conductores lleven cinturón es P(X≤7), es decir, será la función de distribución para x_i = 7.

Recordemos que la función de distribución tiene una expresión que puede dificultar la realización de los cálculos numéricos, por lo que utilizamos una vía que nos facilite los cálculos. Así:

P(X≤7) = 1 - P(X=10) - P(X=9) - P(X=8)

= 1 - f(X=10) - f(X=9) - f(X=8)

Las funciones de cuantía las calculamos de la forma ya conocida y el resultado es:

$P(X\leq 7) = 1 - 0{,}006 - 0{,}04 - 0.121 = 0.833$

Otra vía alternativa de cálculo sería:

$$P(X\leq 7) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6) + P(X=7)$$

I.1.4.- Distribución de Poisson

I.1.4.1.- Definición

La distribución de Poisson se puede entender como un caso particular de la Binomial que utilizamos para determinadas distribuciones en las que el cálculo de la probabilidad es engorroso debido bien a que el número de pruebas es excesivamente elevado o bien a que la probabilidad de éxito es excesivamente baja; en ambos casos la media (n*p) es muy pequeña en relación al número de pruebas (n). En estos casos se puede demostrar que la distribución binomial converge, tiende a comportarse, como una distribución de Poisson.

Como regla práctica entenderemos que es aplicable la distribución de Poisson en aquellas binomiales cuya media tenga un valor inferior a 5 y el número de pruebas sea superior a 30.

Algunos ejemplos de fenómenos que se ajustan a una distribución de Poisson son los siguientes:

* el número de accidentes de tráfico en una ciudad durante una semana.

* el número de emergencias que llegan a un servicio de urgencia hospitalaria.
* el número de robos denunciados en un mes en la ciudad de Madrid.
* el número de llamadas telefónicas que llegan a la centralita de una gran empresa en hora punta.

Todos estos casos pueden ser caracterizados como el número de sucesos de un determinado evento en un período de tiempo. Cada uno de estos eventos toma dos posibles valores, éxito o fracaso, equivalente a recibir una llamada o no recibirla, accidentarse o no accidentarse, etc... Además, la probabilidad del éxito es muy pequeña. Es decir, la probabilidad de tener un accidente es muy baja, la probabilidad de que una persona llama a la centralita de la gran empresa es mínima. Y, no debemos de olvidar, que n, número de personas que hay en una ciudad, número de personas que circulan en coche, etc... es muy grande. Todo esto nos lleva a que una buena parte de fenómenos de esta naturaleza siguen una distribución del tipo Poisson. Es por ello que esta distribución, también denominada "de los sucesos raros" es particularmente útil para resolver problemas de colas y líneas de espera, temas ambos que se podrán estudiar en otras asignaturas del área de métodos cuantitativos en economía.

Por tratarse de un caso particular de la Binomial, la distribución de Poisson cumple los requisitos del Proceso de Bernouilli y la variable aleatoria (X) sigue definiéndose como el número de éxitos en una sucesión de Bernouilli.

Se puede demostrar que la forma que adopta la función de cuantía para una variable aleatoria (X) con una distribución de Poisson de parámetro 8 es:

$$f(x) = P(X = x) = \frac{e^{-l}\, l^{x}}{x!}$$

donde 8 = n*p, x! Es el factorial de x y e es el número e.

La forma que adopta la función de distribución es

$$F(x) = \sum_{r \leq x} P(X = r) = \sum_{r \leq x} \frac{e^{-l}\, l^{r}}{r!}$$

En base a la función de cuantía podemos calcular la media y la varianza de la distribución de Poisson. La determinación de la media es casi inmediata de forma inmediata

$$m_x = E(X) = a_1 = \sum_{x=0}^{\infty} x \left(\frac{e^{-l}\, l^{x}}{x!} \right) = e^{-l} \sum_{x=0}^{\infty} x \frac{l^{x}}{x!} = e^{-l} \sum_{x=1}^{\infty} x \frac{l^{x}}{x!}$$

$$m_x = E(X) = e^{-l} \sum_{x=1}^{\infty} \frac{l^{x}}{(x-1)!}$$

$$Si\ hacemos\ y = (x-1)\ nos\ quedar: E(X) = e^{-l}\, l \sum_{y=0}^{\infty} \frac{l^{y}}{y!} = e^{-l}\, l\, e^{l} = l$$

Para el cálculo de la varianza, determinaremos primero el momento no centrado de orden 2, $"_2$.

$$\alpha_2 = E(X^2) = \sum_{x=0}^{\infty} x^2 \left(\frac{e^{-\lambda}\lambda^x}{x!}\right) = e^{-\lambda}\sum_{x=0}^{\infty} x^2 \frac{\lambda^x}{x!}$$

$$\alpha_2 = e^{-\lambda}\left[1^2\frac{\lambda}{1!}+2^2\frac{\lambda^2}{2!}+\dots\right] = e^{-\lambda}\left[\lambda+2\frac{\lambda^2}{1!}+3\frac{\lambda^3}{2!}+\dots\right]$$

$$\alpha_2 = e^{-\lambda}\lambda\left[1+(1+1)\frac{\lambda}{1!}+(1+2)\frac{\lambda^2}{2!}+(1+3)\frac{\lambda^3}{3!}+\dots\right]$$

Aplicando el desarrollo de Taylor llegamos a la seguinete expresión para el momento no centrado de orden 2

$$\alpha_2 = e^{-\lambda}\lambda\left[1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\dots+\lambda+\frac{\lambda^2}{1!}+\frac{\lambda^3}{2!}+\dots\right] = e^{-\lambda}\lambda\left[e^{\lambda}+\lambda\left(1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\dots\right)\right]$$

Por tanto, $\alpha_2 = e^{-\lambda}\lambda\,[e^{\lambda}+\lambda e^{\lambda}] = \lambda(1+\lambda) = \lambda+\lambda^2$

Y la varianza se calcula como

$$\sigma_x^2 = \alpha_2 - \alpha_1^2$$

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

Por tanto, la media y la varianza de esta distribución son idénticas e iguales a 8=np.

Veamos un ejemplo de distribución de Poisson. Un analista de empresas ha pronosticado que el 3.5% de las pequeñas empresas irán a la bancarrota en 1995. Para una muestra de 100 pequeñas empresas, estime la probabilidad de que al menos tres de ellas entren en bancarrota, suponiendo que la predicción del experto es correcta.

Resolución : Veamos primero de qué tipo de distribución se trata.

* Cumple los requisitos de un Proceso de Bernouilli, puesto que hay dos resultados posibles (bancarrota o no bancarrota); la probabilidad es constante (3.5% predicho por el experto); y las pruebas son independientes.

* Puede ser considerada una Binomial, puesto que definimos una variable aleatoria que es el número de empresas que entran en bancarrota ("éxito").

* Pero cumple las condiciones para ser analizada como una distribución de Poisson, puesto que n es muy grande (n=100) y p muy pequeña (p=0,035), por lo que la media es muy pequeña en relación a n (n*p = 3,5)

Usaremos la distribución de Poisson con media 3,5 para aproximar nuestra distribución.

La probabilidad de que al menos 3 de las 100 empresas entren en bancarrota es igual a la probabilidad de que entren 3 empresas, más la de que entren cuatro, mas cinco, y así hasta 100:

P(3) + P(4) + P(5) +.........+ P(100) = 1 - P(0) - P(1) - P(2).

Para hallar las probabilidades necesarias para resolver el problema debemos hacer uso de la función de cuantía, cuya expresión genérica es:

$$f(x) = P(X = x) = \frac{e^{-l}\, l^{x}}{x!}$$

Aplicada a nuestro ejemplo, y para el caso P(0):

$$f(0) = P(X = 0) = \frac{e^{-3,5}\, 3,5^{0}}{0!} =$$
$$\frac{(0,30197)*(1)}{1} = 0,302$$

Si aplicamos la fórmula al resto de casos, la resolución del problema será:

1 - 0,302 - 0,1057 - 0,1850 = 0.679

EJERCICIOS PROPUESTOS(estos ejercicios deben ser, además, practicados en EXCEL).

1.- Para acceder a un Master de postgrado se realiza un examen tipo test a los solicitantes. El examen consta de cinco preguntas y cada una tiene cuatro posibles resultados. Un alumno no conoce las respuestas, pero decide contestar en función del siguiente juego: tira un dado, si sale un 1 elige el primer resultado; si sale un 2, elige el segundo resultado, y así sucesivamente; si sale un 5 ó un 6, tira el dado de nuevo. Se pide determinar:

a) cuál es la probabilidad de éxito y cuál la de fracaso

b) cuál es la probabilidad de que acierte una respuesta

c) cuál es la probabilidad de acertar dos o tres respuestas.

2.- El número medio de solicitudes de préstamo que recibe una entidad bancaria es 5 por día. Suponiendo que las solicitudes de préstamo sigan una distribución de Poisson, calcular la probabilidad de que en un día se reciban exactamente dos solicitudes.

3.- La proporción de individuos en una población con renta superior a veinte millones de ptas. es de 0,005%. Determinar la probabilidad de que entre 5.000 individuos consultados haya dos con ese nivel de renta, supuesto que todos los consultados respondan.

4.- Se lanza una moneda al aire diez veces. Determinar la probabilidad de que salgan siete caras.

5.- Un agente de seguros debe visitar cinco posibles clientes en una tarde, sabiendo que la probabilidad de que le contraten un seguro es 0,4 en cada visita.

a) Determinar la probabilidad que tiene de contratar tres seguros esa tarde.

b) Determinar el número esperado de seguros a contratar esa tarde.

c) Determinar la varianza de la distribución

d) Determinar la probabilidad de que venda entre dos y cuatro seguros.

e) Determinar la probabilidad de que no se vaya a su casa sin contratar al menos un seguro.

6.- El 80% de las bolas contenidas en una urna son de color blanco, siendo el 20% restante de color encarnado. Determinar la probabilidad de que al efectuar tres extracciones sucesivas con reemplazamiento dos de las bolas extraídas sean de color blanco y una de color encarnado.

7.- He enviado invitaciones a cenar a 20 amigos míos que no se conocen entre sí, sabiendo que la probabilidad de que individualmente acepten es del 90%.)Cuál es la probabilidad de que acepten como mucho 17 de mis amigos?.

8.- El número medio de automóviles que llega a una estación de suministro de gasolina es de 210 a la hora. Si dicha estación puede atender a un máximo de 10 automóviles por minuto, determinar la probabilidad de que en un minuto dado lleguen a la estación de suministro más automóviles de los que pueden ser atendidos.

9.- En una determinada zona geográfica se pretende introducir un nuevo producto del que es razonable esperar sea demandado por el 0,4% de sus habitantes. Determinar la probabilidad de que, consultados 1.000 de ellos, dicho producto sea demandado:

a) por tres o mas personas
b) por cinco personas o menos.

10.- Un día de verano muy caluroso el 10% de los empleados en una gran oficina bancaria no acude a trabajar. La dirección de la empresa decide hacer un estudio en profundidad sobre el absentismo, para lo que selecciona aleatoriamente a diez de los empleados de la oficina.

a))cuál es la variable aleatoria de este problema?

b))esa variable aleatoria es discreta o continua?

c))cuál es la probabilidad de que se seleccionen diez empleados en un día caluroso y ninguno de ellos sea absentista?.

11.- En un reciente estudio se comprobó que el 90% de los hogares norteamericanos tenían TV en color. En una muestra de 9 hogares, cuál es la probabilidad de que:

a) los nueve tengan TV en color

b) menos de cinco tengan TV en color

c) Más de cinco tengan TV en color

d) Al menos siete tengan TV en color.

12.- Uno de cada cinco fines de semana tengo dolor de muelas, por lo que no puedo realizar las excursiones que tanto me apetecen. Usando el programa STATG, determine:

a) La probabilidad de que en los próximos siete fines de semana no tenga dolor de muelas.

b) La probabilidad de que tenga dolor de muelas exactamente uno de los fines de semana.

c) La probabilidad de que tenga el dolor exactamente tres fines de semana.

d) La probabilidad de que tenga dolor de muelas 4 ó 5 fines de semana.

13.- Basado en un reciente experimento se ha detectado que el 5% de los tornillos fabricados por una fresadora Carter-Bell automática de gran velocidad, son defectuosos. Si seleccionamos seis tornillos determinar la probabilidad de que:

a) ninguno salga defectuoso

b) tres salgan defectuosos

c) resulten defectuosos menos de dos tornillos

d) los seis salgan defectuosos.

<u>Soluciones a los ejercicios propuestos</u>

1.- a) p=0,25 ; q=0,75
    b) 0,39551
    c) 0,35156

2.- 0,0842

3.- 0,02433

4.- 0,11718

5.- a) 0,23
    b) 2
    c) 1,2
    d) 0,6524
    e) 0,922

6.- 0,384

7.- 0,323

8.- 0,0092

9.- a) 0,000092
    b) 0,9999

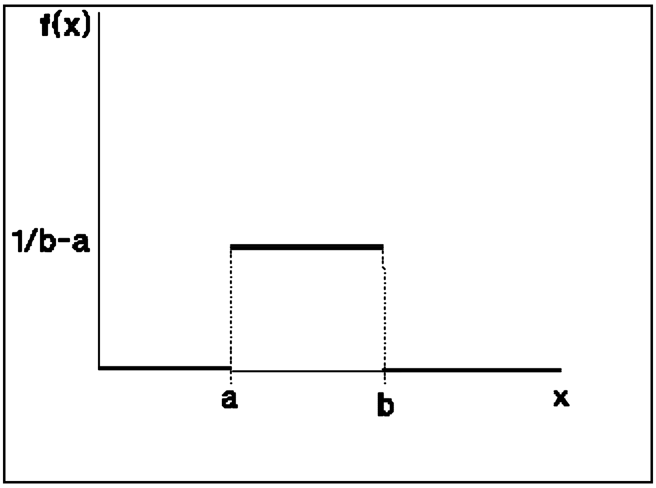
I.2.- Variables aleatorias continuas

I.2.1.- La distribución uniforme

Diremos que la variable aleatoria X se distribuye **UNIFORMEMENTE** en un intervalo [a,b] y lo representamos como X - U(a,b) cuando su función de densidad es:

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

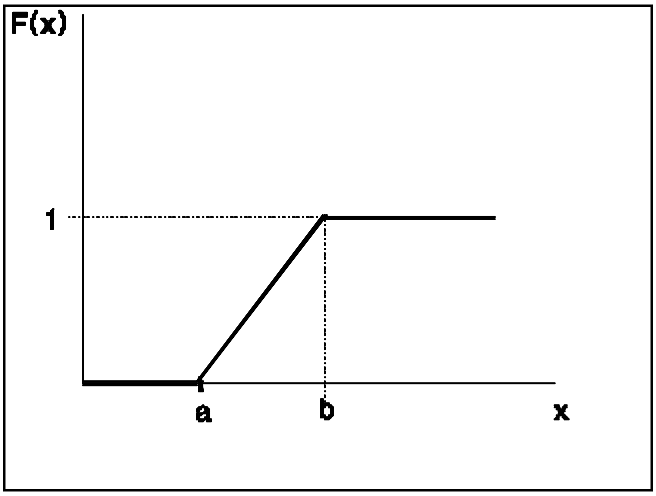
Cuya representación gráfica es:



$$F(x)=\begin{cases} 0 & si \;\; x < a \\ \dfrac{x-a}{b-a} & si \; a \le x \le b \\ 1 & si \; x > b \end{cases}$$

La función de distribución será:

Y su representación gráfica:



El cálculo de la media de esta variable es inmediato

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x\, dx =$$

$$= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \left[\frac{b^2-a^2}{2} \right] =$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

La varianza la calculamos a través de los momentos no

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x^2\, dx =$$

$$= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = = \frac{1}{b-a} \left[\frac{b^3-a^3}{3} \right] =$$

$$= \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

centrados de orden 1 y 2

I.2.2 LA DISTRIBUCION NORMAL

I.2.2.1 INTRODUCCION

La distribución normal es la más común entre todas las distribuciones de probabilidad utilizadas en Estadística y tiene importantes aplicaciones en la modelización de variables estadísticas asociadas a los elementos de una población. Por ejemplo, las medidas físicas del cuerpo humano en una población, las características psíquicas medidas por test de inteligencia o personalidad, las medidas de calidad en muchos procesos industriales, los errores de las observaciones astronómicas; siguen distribuciones normales.

Una justificación de la frecuente aparición de la distribución normal es el **teorema central del límite** que establece que cuando los resultados de un experimento sean debidos a un conjunto muy grande de causas independientes, que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto, es esperable que los resultados sigan una distribución normal.

Diremos que una variable aleatoria sigue una distribución

$$X \rightarrow N(m, s^2)$$

normal con parámetros μ y σ^2 y se representa:
cuando su función de densidad es de la forma:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\forall\, x \in \Re \text{ siendo } \sigma \succ 0 \text{ y } \mu \in \Re$$

en donde $\Re$ es el conjunto de los números reales.

Como se puede observar, la familia de distribuciones normales depende de los parámetros μ y σ^2, que coinciden con la media y la varianza respectivamente.

I.2.2.2 LA DISTRIBUCION NORMAL TIPIFICADA O STANDARD

I.2.2.2.1 Función de densidad. Propiedades

Un caso particular de distribución normal es aquella en la cual la media vale cero y la varianza 1. En este caso decimos que la variable X se distribuye como una variable normal tipificada o standard y la denotamos por

$$X \to N(0,1)$$

En este caso, su función de densidad toma la forma:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \forall\, x \in \Re$$

PROPIEDADES

(1) Es simétrica respecto de x = 0, ya que:

$$f(-x) = f(x)$$

(2) Alcanza un máximo en x = 0 y vale

$$f(0) = \frac{1}{\sqrt{2\pi}}$$

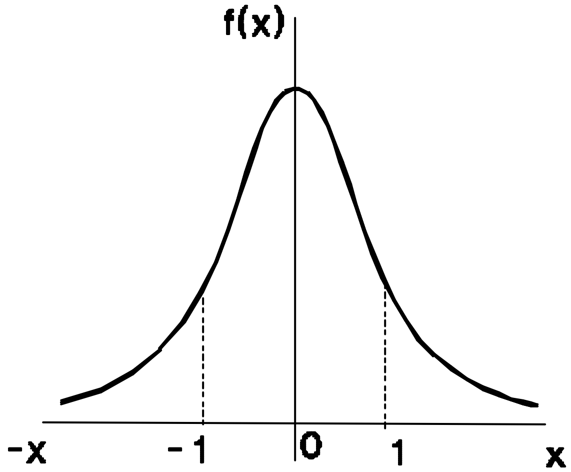
(3) Es creciente para x < 0 y decreciente para x > 0.

(4) Los puntos de abcisas 1 y -1 son de inflexión de la función.

(5) La recta y = 0 es asíntota de la función, pues

$$\lim_{x\to\infty} f(x)=\lim_{x\to-\infty} f(x)=0$$

Su representación gráfica es:



I.2.2.2.2 Función de distribución. Propiedades

La expresión de la función de distribución es:

$$\Phi(x)=\int_{-\infty}^{x} f(x)dx=\int_{-\infty}^{x}\frac{1}{\sqrt{2p}}e^{-\frac{x^2}{2}}dx\ \ \forall\ x\in \_$$

F(x) verifica:

(1) F(-x) = 1 - F(x)

Demostración:

$$\Phi(-x)=\int_{-\infty}^{-x} f(x)dx=\int_{x}^{\infty} f(x)dx=$$

$$=P(X>x)=1-\Phi(x)$$

Como f(x) = f(-x); entonces:

y P(X > x) = F(-x)

(2) Las rectas y = 0 e y = 1 son asíntotas de la función F(x), pues se cumple:

$$\lim_{x \to -\infty} \Phi(x) = 0$$

$$\lim_{x \to \infty} \Phi(x) = 1$$

y además como f(x) > 0, entonces nunca ocurre que:

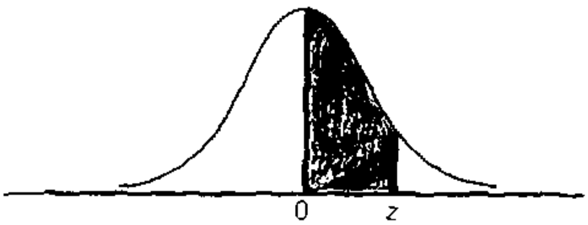
F(x) = 0; ni F(x) = 1

Un elemento importante cuando se trabaja con distribuciones normales es el uso de las tablas. Las tablas estadísticas son un instrumento que nos facilita el cálculo de probabilidades para distintos sucesos. La mejor forma para trabajar con tablas estadísticas es usarlas. Veamos un ejemplo para el caso de cálculo de probabilidades mediante el uso de la tabla de la normal tipificada.

En la tabla 1 se muestra un tipo posible de tabla de la N(0,1). En esta tabla se nos da la probabilidad de que la variable tome valores en el intervalo (0,z). La forma de usarla es la siguiente. Sea z=1.25, la probabilidad de que a variable tome valores en el intervalo (0,1.25) es igual a 0.3944. La obtención de esta probabilidad es inmediata. En la primera columna buscamos la parte entera de z y el primer decimal, es decir, en nuestro caso buscamos 1.2, y en la primera fila buscamos las centenas de z, es decir, el 0.05. El punto de corte de la fila que empieza con 1.2 y la columna que comienza con 0.05 nos da la probabilidad de que la variable N(0,1) tome valores entre 0 y 1.25.

Tabla 1. Normal tipificada.

TABLE 4 Normal curve areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0174	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald. *Statistical Tables and Formulas* (New York: John Wiley & Sons, 1952). Reproduced by permission of A. Hald and the publisher.

Veamos como calculamos $P(|x|<1)$con el uso de esta tabla. Esta probabilidad es la probabilidad de que la variable tome valores entre –1 y 1. Dado que sabemos que la variable normal es simétrica, la probabilidad entre 0 y 1 es la misma que la que hay entre –1 y 0. La probabilidad entre 0 y 1 la calculamos directamente de la tabla 1 buscando el cruce entre la fila 1.0 y la columna 0.00. Por tanto vale 0.3413. En consecuencia, la probabilidad pedida será igual a:

$$P(|x|<1)=P(-1,0)+P(0,1)=2*P(0,1)=2*0.3413=0.6826$$

Obsérvese que a partir de la tabla 1 podemos calcular la función de distribución de cualquier punto del espacio de los números reales. Para ello, lo único que tenemos que hacer es:

a.- Si z es mayor que 0, a la probabilidad que nos da la tabla le sumamos 0.5

b.- Si z es menor que cero, al 0.5 le restamos el valor que nos da la tabla para el correspondiente z positivo.

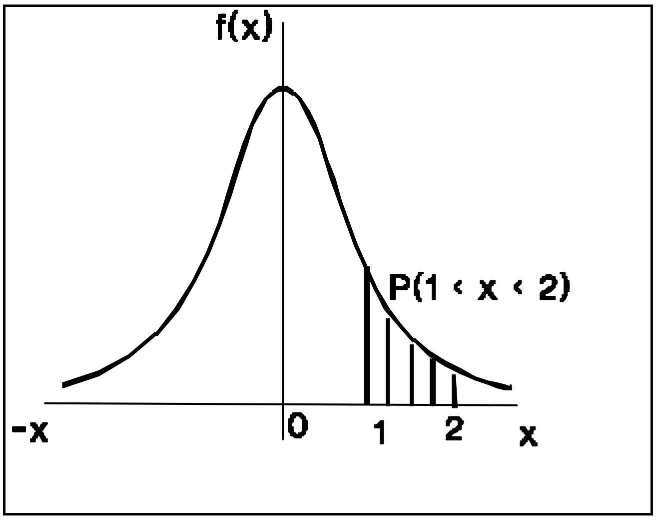
Veamos un ejemplo: La función de distribución en el punto 1.35 será igual a 0.5+0.4115=0.9115. El valor 0.4115 es el que nos da la tabla de la normal.

Si lo que queremos calcular es la función de distribución en el punto –1.35, esta la obtendremos como 0.5-P(0,1.35)=0.5-0.4115= 0.0885

Si queremos calcular la probabilidad de que x esté comprendida en un intervalo, por ejemplo: $P(1 < X < 2)$, tendríamos:

$$P(1 < X < 2) = F(2) - F(1) = 0.97725 - 0.8413 = 0.13595$$

Cuya representación gráfica sería:



Es por ello que no existe una única tabla de la distribución N(0,1). Todas ellas nos dan la misma información pero de distinta forma. El alumno debe manejase con cualquiera de ellas. La tabla 2 nos muestra otra posible forma de la tabla de la normal reducida.

Hasta ahora hemos calculado la probabilidad que hay entre dos valores posibles de la variable, sin embargo otras veces tendremos que calcular el valor x_1 conocida su probabilidad. Es decir, dado $F(x_1)$,por ejemplo $F(x_1) = 0.90$, que lo representamos como $x_{0.90}$ tendremos que buscar en las tablas el valor de x_1 (la abcisa) que nos da una probabilidad (área) de 0.90.

En este caso vemos que no existe en las tablas una probabilidad exactamente igual a 0.90, lo que tenemos es:

A z=1.28 le corresponde una función de distribución de 0.8997

A $z=1.29$ le corresponde una función de distribución de 0.90147

Para calcular el valor exacto de x_1 tenemos que interpolar y lo hacemos empleando el siguiente fases

Fase 1.-Calculamos las diferencias siguientes:

Diferencia de abcisas	Diferencia de probabilidad	
1.29	0.90147	0.90147
-1.28	-0.8997	-0.90
0.01	0.00177	0.00147

Fase 2.-Hacemos la siguiente proporción:

Diferencia de abcisas	Diferencia de probabilidad	
0.01	0.00177	x = 0.0083051
x	0.00147	

Fase 3.-$x_1 = 1.29 - 0.0083051 = 1.28169$; $0.90 = F(1.282)$

I.2.2.3 DISTRIBUCION NORMAL GENERAL

Ya hemos visto que una variable aleatoria x sigue una distribución Normal (μ,σ^2) cuando su función de densidad es:

$$f(x)=\frac{1}{\sqrt{2ps}}e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

$$\forall\, x\in \Re \; siendo \; s \succ 0 \; y \; m\in \Re$$

PROPIEDADES: La función

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

$$\forall\, x \in \Re \; siendo \; \sigma \succ 0 \; y \; \mu \in \Re$$

cumple las siguientes propiedades:

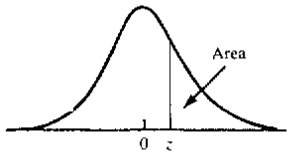
(1) Es simétrica respecto de x = µ pues f(µ-x) = f(µ+x).

(2) Alcanza un máximo en x = µ y vale

$$f(\mu) = \frac{1}{\sqrt{2\pi}\,\sigma}$$

Tabla 2:Distribución Normal

TABLA 4 Áreas bajo la curva normal
Probabilidad normal estándar de cola superior
(para valores negativos de z, las áreas se obtienen por simetría)



Segundo decimal de z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

De: R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

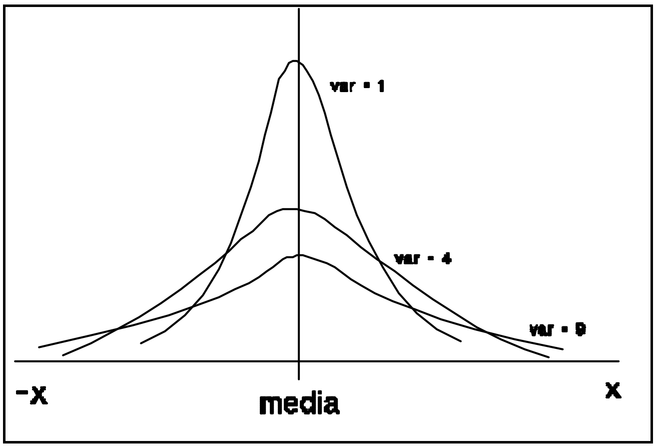
(3) Es creciente para x < μ, y decreciente para x > μ

(4) Los puntos de abcisas (μ-σ) y (μ+σ) son de inflexión.

(5) La recta y = 0 es asíntota de f(x) pues:

$$\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$$

En el siguiente gráfico se puede apreciar la diferencia entre algunas distribuciones normales con la misma media y distinta varianza.



I.2.2.3.1 Función de distribución

La función de distribución de una normal general toma la forma

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

que, como puede verse a simple vista, es difícil de manejar.

Sin embargo, no será necesario trabajar con esta función para el cálculo de las probabilidades de una distribución normal general. Los valores de esta función pueden ser obtenidos a través de la función de distribución normal tipificada mediante el proceso de tipificación de una variable normal, el cual se prueba mediante los siguientes resultados:

(1) Sea X - N (μ,σ^2). Entonces la variable Z = $(X-\mu)/\sigma$ - N (0,1) y se cumple:

$$F(x) = \Phi(\frac{x - m}{s}); \ \forall \, x \in \Re$$

(2) Sea Z - N(0,1), entonces la variable X = $(\sigma Z + \mu)$ se distribuye como una $N(\mu,\sigma^2)$ con $\sigma^2 > 0$, siendo su función de densidad:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2p}\, s} e^{-\frac{1}{2}(\frac{x-m}{s})^2} dx$$

Demostración de (1):

Calculamos la función de distribución de

$$Z = \frac{X - m}{s} \quad \forall \;\; Z \in \Re$$

$$F(z) = P\left(Z \leq z\right) = P\left(\frac{X - m}{s} \leq z\right) = P\left(X \leq sz + m\right) = F\left(sz + m\right)$$

$$luego \;\; F(z) = F\left(sz + m\right);\; y\ derivando :$$

$$f(z) = s \;\; f(\,sz + m\,) = s\,\frac{1}{\sqrt{2p}\;s}\,e^{-\frac{1}{2}\left(\frac{sz+m-m}{s}\right)^2} = \frac{1}{\sqrt{2p}}\,e^{-\frac{1}{2}z^2};$$

$$luego :\; f(z) = \frac{1}{\sqrt{2p}}\,e^{\frac{z^2}{2}}; \quad \forall\, z \in \Re$$

$$entonces : Z = \frac{X - m}{s} \longrightarrow \;\; N\,(0,1).$$

$$F(x) = \Phi\left(\frac{x - m}{s}\right); \;\; \forall\, x \in \_ \;\; si \;\; X \;\_\; N(\,m, s^2\,)$$

De forma que:

$$F(x) = P\left(X \leq x\right) = P\left(\frac{X - m}{s} \leq \frac{x - m}{s}\right) = P\,(Z \leq \frac{x - m}{s});$$

$$pero\ Z = \frac{X - m}{s}\;\_\; N(0,1)\ como\ hemos\ visto, luego :$$

$$F(x) = P\left(Z \leq \frac{x - m}{s}\right) = \Phi\left(\frac{x - m}{s}\right)$$

Se propone como ejercicio demostrar el segundo resultado.

Ejemplo: Si X se distribuye como N(20,4), calcular las siguientes probabilidades:

1) P(X < 15).

2) P(4X -5 > 80).

Solución:

1) Si X - N(20,4)

$$Z = \frac{X - 20}{2} \_ \ N(0,1) \quad y\ como \quad P(X < x) = P\left(\frac{X - m}{s} < \frac{x - m}{s}\right) =$$

$$= P\left(Z < \frac{x - m}{s}\right) = \Phi\left(\frac{x - m}{s}\right); luego,\ en\ este\ caso\ :$$

$$P(X < 15) = P\left(\frac{X - 20}{2} < \frac{15 - 20}{2}\right) = \Phi\ (-2.5) = 1 - \Phi\ (2.5) =$$

$$= 1 - 0.99379 = 0.00621; \quad P\ (X < 15) = 0.6\%$$

2) P(4X - 5 > 80) = P(X > 21.25) = 1 - P(X < 21.25). Tipificando:

$$P(X < 21.25) = \Phi\left(\frac{21.25 - 20}{2}\right) = \Phi\,(0.625) = 0.73405$$

$$[interpolando\ entre\ \Phi\,(0.620)\ y\ \Phi\,(0.630)]$$

resultando que P(4X - 5 > 80) = 1 - 0.73405 = 0.26595 = 26.6%

I.2.2.4 TEOREMA DE LA ADICIÓN

Cualquier combinación lineal de variables normales e independientes sigue una distribución normal.

Si X_1, X_2, ..., X_n son independientes, tal que X_i se distribuye como $N(\mu_i, \sigma^2_i)$;(con $\sigma^2_i > 0$ para todo i = 1, 2,..., n), y dados a_1, a_2, ..., a_n, b números reales tal que a_i es distinto de cero para todo i = (1, 2, ..., n), se cumple:

$$\text{Llamando } X = \sum_{i=1}^{n} a_i X_i + b;$$

$$X \rightarrow N\left(\sum_{i=1}^{n} a_i \mu_i + b, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

La demostración a este teorema se realiza a través de la función generatriz. Dado que en este curso esta función no se ha estudiado, no se demostrará este teorema.

Corolario: Si X_1, X_2, ..., X_n son independientes y con la misma distribución normal: X_i se distribuye como $N(\mu,\sigma^2)$; para i =

$$(1)\ \sum_{i=1}^{n} X_i \rightarrow N(n\mu, n\sigma^2)$$

(2) La variable "media muestral" sigue una distribución normal:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \_ N(\mu, \frac{\sigma^2}{n})$$

(1, 2, ..., n) y $\sigma^2 > 0$, entonces:

I.2.3 TEOREMA CENTRAL DEL LIMITE

La importancia de la distribución Normal se basa en el TEOREMA CENTRAL DEL LIMITE, cuyo enunciado puede establecerse de la siguiente manera:

" Sea x_1, x_2, ..., x_n un conjunto de variables aleatorias independientes (cualquiera que sea su distribución, discreta o contínua), y si "n" es lo suficientemente grande, la suma de las variables x_i se distribuye como una distribución NORMAL".

En la práctica n es grande cuando supera el valor 30.

La Normal formada tendrá como media la suma de las medias y como varianza la suma de las varianzas.

I.2.5 La distribución exponencial

Es un caso particular de la familia de distribuciones Gamma descritas en el anexo I.2 que, debido a sus propiedades, merece un estudio a parte.

Decimos que una variable aleatoria X sigue una distribución **Exponencial** con parámetro λ ($\lambda > 0$) y la denotamos como

$$X \rightarrow EXP(\lambda)$$

cuando X es una variable cuya función de densidad es de la forma.
Y como función de distribución tiene la siguiente expresión:

$$f(x) = \left\{ \begin{array}{ll} l\,e^{-lx} & para\ \ x > 0 \\ 0 & para\ x \leq 0 \end{array} \right\}$$

$$si\ x > 0 \rightarrow F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{x} \lambda e^{-\lambda x}\,dx = -\left[e^{-\lambda x} \right]_{0}^{x} = 1 - e^{-\lambda x}$$

$$luego: \quad F(x) = \begin{cases} 0 & si\ x \leq 0 \\ 1 - e^{-\lambda x} & si\ x > 0 \end{cases}$$

Además, se puede demostrar que la Esperanza y Varianza de una variable exponencial de parámetro λ son iguales a:

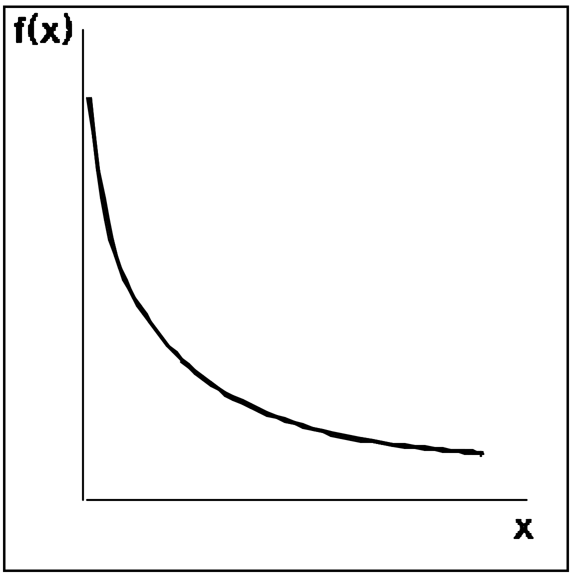
$$\mathbf{E(x) = 1/\lambda}$$

$$\mathbf{Var(x) = 1/\lambda^2}$$

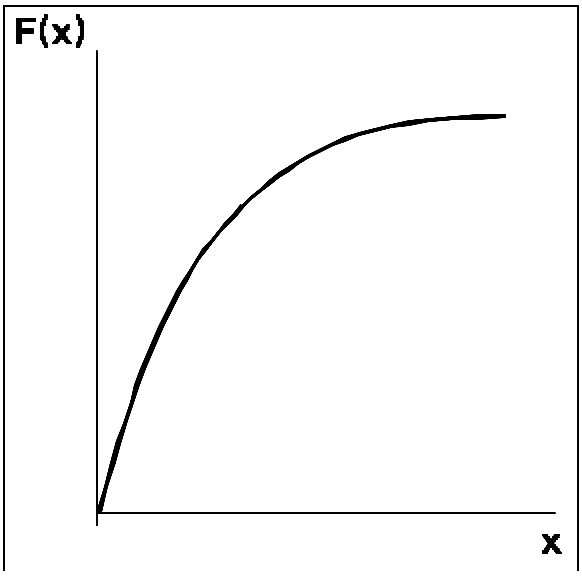
Por lo tanto la desviación típica de una distribución exponencial es igual a su media.

$$\mathbf{s_x = 1/\lambda}$$

La representación gráfica de la función de densidad es:



Y la gráfica de la función de distribución:



Una **característica** importante de la distribución exponencial es la propiedad de que **no tiene memoria**: si X tiene una distribución exponencial, entonces

$$P\left(X > b+c \,/\, X > b\right) = P\left(X > c\right)$$

para b y c no negativos cualesquiera. Para probar esta

$$P\left(X > x\right) = 1 - P\left(X \le x\right) = 1 - F(x) = e^{-l x}$$

$$(para\ x \rightarrow 0)$$

afirmación observamos que
De la definición de probabilidad condicionada resulta que:

$$P(X > b + c/X > b) = \frac{P(X > b + c, X > b)}{P(X > b)} = \frac{P(X > b + c)}{P(X > b)} =$$

$$= \frac{e^{-(b+c)\,l}}{e^{-b\,l}} = e^{-c\,l} = P(X > c) \quad c.q.d$$

Como consecuencia de esta característica, el uso de distribuciones exponenciales es apropiado para distribuciones de tiempo de vida cuando no hay deterioro con la edad.

EJERCICIO: Sabiendo que la clase de Estadística comienza a las 10 horas y que el profesor que imparte dicha asignatura emplea en el trayecto de su casa a la Facultad entre 15 y 20 minutos, calcular a qué hora debe salir de casa para que pueda llegar puntualmente a dar su clase con una probabilidad del 99%.

Solución:

Si llamamos X al tiempo que tarda en llegar a la Facultad,

X - U(15,20)

Entonces:

$$f(x) = \begin{cases} \dfrac{1}{20-15} & si\ 15 \le x \le 20 \\ 0 & en\ otro\ caso \end{cases}$$

Lo que nos piden es:

$$P\left(X \le x\right) = 0.99 = \frac{X - 15}{20 - 15} = \frac{X - 15}{5} = 0.99$$

$$X - 15 = 4.95;\ \ X = 4.95 + 15 = 19.95$$

Tiene que salir de casa a las 9h 40m 3s

<u>EJERCICIO:</u> Si en nuestro país la altura de los jóvenes en edad militar sigue una distribución normal con media 173 cm. y desviación típica 10 cm, calcular:

(1) Qué porcentaje de jóvenes sería rechazado del servicio militar si se establece que no irán a la "mili" aquellos cuya altura sea inferior o superior a la media en 20 cm.
(2) Suponiendo que un determinado año el Ministerio de Defensa decide que únicamente harán el Servicio Militar el 75% del censo de jóvenes, inicialmente útiles,)qué intervalo de altura tendrá que elegir?

Solución:
(1) Llamando X a la altura de los individuos en edad militar:

X - N(173,100)

y nos piden:

P(0x - 1730 > 20) = 1 - P(0x - 1730) # 20) =
= 1 - P(153 # x # 193) = 1 - (F(193) - F(153))=
(tipificando):

= 1 - (F(2) - F(-2)) = 1 - F(2) + F(-2) = 1 - F(2) + 1 - F(2) =
= 2 - 2F(2) = 2 - 2 x 0.97725 = 0.0455 = 4.55%

(2) Si hace el Servicio Militar el 75%, se rechaza el 25%,

luego tenemos que calcular:

$$P\left(|X-173|>a\right)=1-P\left(-a\leq X-173\leq a\right)=$$

$$1-P\left(-\frac{a}{10}\leq\frac{X-173}{10}\leq\frac{a}{10}\right)=1-P\left(-\frac{a}{10}\leq Z\leq\frac{a}{10}\right)$$

$$siendo\ Z=\frac{X-173}{10}\rightarrow N(0,1)$$

$$luego\ P\left(|X-173|>a\right)=1-\left\{\Phi(\frac{a}{10})-\Phi(-\frac{a}{10})\right\}=$$

$$=2-2\Phi\left(\frac{a}{10}\right)=0.25\rightarrow\Phi\left(\frac{a}{10}\right)=0.875$$

$$según\ las\ tablas:$$

$$\Phi(0.8749)=1.15$$

$$\Phi(0.8770)=1.16$$

$$Interpolando:\frac{a}{10}=1.1505, luego\ a=11.505.$$

P(0X - 1730 > a) = 0.25

Para admitir únicamente al 75%, hay que establecer un intervalo de 11.5 cm sobre la media.

EJECICIO 3: Una tienda vende camisetas de dos marcas distintas (Top y Body). Las ventas de cada una de estas marcas siguen distribuciones normales con media de 2000 unidades y desviación típica de 100 unidades para la marca Top, y 2100 unidades de media con desviación típica de 110 unidades para la marca Body.

Al hacer el pedido para la próxima temporada, el dueño de la tienda quiere conocer la probabilidad de que las ventas de la marca Body superen en más de 150 unidades a las de la marca Top.

Solución:

Llamando T ventas de la marca Top: $T \rightarrow N(2000, 100^2)$

Llamando B ventas de la marca Body: $B \rightarrow N(2100, 110^2)$

Se pide: $P(B - T > 150)$; llamando $B - T = X$

$$X \rightarrow N \left\{ \begin{array}{l} m_x = 2100 - 2000 = 100 \\ \\ s_x^2 = 100^2 + 110^2 = 22100; \; s_x = 148.67 \end{array} \right\}$$

$$X \rightarrow N(100,148.67^2)$$

$$Tipificando : si \; X \rightarrow N(100,148.67^2) \; entonces :$$

$$Z = \frac{X - 100}{148.67} \rightarrow N(0,1)$$

$$luego : P(X > 150) = P\left(Z > \frac{150 - 100}{148.67}\right) = P(Z > 0.34) =$$

$$= 1 - P(Z < 0.34) = 1 - 0.6331 = 0.3669 = 37\%$$

EJERCICIO 4: El coeficiente intelectual de los alumnos de C.O.U en un determinado colegio sigue una distribución Normal con media 100 y desviación típica 5:

(1) Elegidos al azar 10 alumnos para participar en un concurso regional, ¿cuál es la probabilidad de que el coeficiente intelectual, por término medio, de estos 10 alumnos no difiera de la media de su curso en más de 2 puntos?
(2) ¿Cuántos alumnos tendrían que ir al concurso para que su coeficiente intelectual medio no difiera del de su curso en más de 2 puntos con una probabilidad del 97.5%?

Solución:
Llamando X_i al coeficiente intelectual del alumno i-ésimo para

i = 1, 2, ...,10, $X_i \rightarrow N(100,5^2)$; suponemos que las variables son independientes.

El coeficiente medio viene dado por:

$$\overline{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i \_ N(100,\frac{5^2}{10})$$

La probabilidad pedida es:

$$P(|\overline{X}_{10} - 100 |\leq 2) = P(-2 \leq \overline{X}_{10} - 100 \leq 2) =$$

$$\left(Tipificando\ Z = \frac{\overline{X}_{10} - 100}{5/\sqrt{10}} = \frac{\sqrt{10}(\overline{X}_{10} - 100)}{5} \right)$$

$$= P(-2\frac{\sqrt{10}}{5} \leq \frac{\sqrt{10}(\overline{X}_{10} - 100)}{5} \leq 2\frac{\sqrt{10}}{5}) = P(-1.26 \leq z \leq 1.26) =$$

$$= \Phi(1.26) - \Phi(-1.26) = 2\Phi(1.26) - 1 = 2 * 0.8962 - 1 = 0.7924 = 79\%$$

(2) Tenemos que encontrar el valor de **n** para el cual se cumple:

$$P\left(|\overline{X}_n - 100| \leq 2\right) \geq 0.975$$

$$Como\ \overline{X}_n \_ N\left(100, \frac{5^2}{n}\right), entonces\ Z = \frac{\sqrt{n}(\overline{X}_n - 100)}{5} \_ N(0,1)$$

$$por\ lo\ tanto: P\left(|\overline{X}_{10} - 100| \leq 2\right) = P\left(-\frac{2\sqrt{n}}{5} \leq Z \leq \frac{2\sqrt{n}}{5}\right) =$$

$$= 2\Phi\left(\frac{2\sqrt{n}}{5}\right) - 1 \geq 0.975 \rightarrow \Phi\left(\frac{2\sqrt{n}}{5}\right) \geq 0.9875$$

$$Interpolando: \frac{2\sqrt{n}}{5} \geq 2.24152; de\ donde\ n \geq 31.40$$

Tienen que ir al concurso más de 31 alumnos.

EJERCICIO 6: Un sistema eléctrico es alimentado por siete baterias, las cuales funcionan de forma independiente, siendo el tiempo de vida para cada bateria una variable aleatoria exponencial con $\lambda = 0.0001$ (el tiempo se mide en horas). El sistema deja de funcionar cuando se paran 4 o más baterias. ¿Cuál es la probabilidad de que el sistema funcione más de 10.000 horas.

Solución:

Tenemos que calcular P(Y ≥ 4). Consideramos la variable Y = nº de baterías que funcionan más de 10.000 horas.

$$Definimos\ Y_i = \begin{cases} 1 & si\ la\ bateria\ funciona\ al\ cabo\ de\ 10.000h \\ & para\ i = \{1,2,3,4,5,6,7\} \\ 0 & si\ no\ es\ asi \end{cases}$$

Definimos X_i: "tiempo de vida de la bateria i" para i=1,...,7

X_i - Exp(0.0001) siendo las variables X_i independientes.

De forma que: $P(Y_i = 1) = P(X_i > 10000) = 1 - F(10000) =$
$= e^{-0.0001*10000} = e^{-1}$

luego $P(Y_i = 1) = e^{-1} = 0.3679$ y $P(Y_i = 0) = 1 - e^{-1} = 0.6321$

Y_i - b(0.3679) y son independientes, pues X_i son independientes, y se cumple que $Y = Y_1 + Y_2 +...+Y_7$.

Por tanto Y - B(7, 0.3679) y entonces:

$$P(Y \geq 4) = \sum_{i=4}^{7} \binom{7}{i} 0.3679^i * 0.6321^{7-i} = 0.230371 = 23\%$$

Que es la probabilidad de que el sistema pueda funcionar más de 10.000 horas.

EJERCICIO 7: Si los autobuses con destino a la Universidad salen cada 15 minutos, por término medio, y suponemos que los tiempos entre las salidas de los distintos autobuses son independientes y siguen una distribución exponencial, ¿cuál sería la probabilidad de que un estudiante tuviera que esperar más de 20 minutos?

Solución:

Si T - Exp(λ) entonces $E(T) = 1/\lambda$ y $\lambda = 1/15$

Llamamos X_{20}: "número de autobuses que llegan en 20 minutos"

$$X_{20} \text{ - } P(1/15*20) = P(4/3)$$

Tenemos que calcular:

$$P(X = 0) = \frac{(4/3)^0}{0!} e^{-\frac{4}{3}} = e^{-\frac{4}{3}} \_ 0.2636 = 26\%$$

Que es la probabilidad de que en 20 minutos no llegue ningún autobús.