DISTRIBUCIONES DE PROBABILIDAD DE VARIABLES ALEATORIAS DISCRETAS

Uniforme
Binomial
de Poisson

Distribución Uniforme

Sea X una **variable aleatoria discreta** cuyos k valores posibles son $x_1, x_2, \ldots, x_k$, con igual probabilidad de ocurrencia, entonces la **función de masa de** X es:

$$p_X(x) = \begin{cases} 1/k & \text{para } x_1, x_2, \ldots, x_k \\ 0 & \text{en otro caso} \end{cases}$$

y la **distribución** se llama **uniforme**.

Ejemplos:

- Cuando un foco se selecciona al azar de una caja que contiene un foco de 40 W, uno de 60 W, uno de 75 W y otro de 100 W, cada uno de ellos tiene la misma probabilidad de ser elegido y se tiene una distribución uniforme cuya función de masa es:

$$p_X(x) = \begin{cases} 1/4 & \text{para } 40\text{ W}, 60\text{ W}, 75\text{ W y } 100\text{ W} \\ 0 & \text{en otro caso} \end{cases}$$

dado que existen $k = 4$ valores equiprobables de la variable aleatoria: $x_1 = 40$ W, ..., $x_4 = 100$ W.

- Cuando se lanza un dado obtenemos:

$$p_X(x) = \begin{cases} 1/6 & \text{para } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{en otro caso} \end{cases}$$

donde X es el número de puntos que muestra el dado.

- Si se selecciona "al azar" uno de los k enteros positivos; $1, 2, \ldots, k$, donde la frase "al azar" significa que los k enteros tienen la misma probabilidad de ser seleccionados, la función de masa es:

$$p_X(x) = \begin{cases} 1/k & \text{para } x = 1, 2, \ldots, k \\ 0 & \text{en otro caso} \end{cases}$$

Esta distribución se llama **distribución uniforme sobre los enteros** $\mathbf{1, 2, \ldots, k}$

No es posible seleccionar un entero al azar del conjunto de todos los enteros porque no es posible asignar la misma probabilidad a cada uno de los enteros positivos y obtener que la suma de las probabilidades sea igual a 1. En otras palabras *no se puede asignar una distribución uniforme a una sucesión infinita de valores posibles* pero se puede asignar una distribución de este tipo a cualquier sucesión finita de resultados igualmente probables.

Si seleccionamos al azar un alumno entre 100 y definimos la variable aleatoria X como el número con que se identifica a cada alumno, es decir a cada alumno se le hace corresponder un único número entre los enteros del 1 al 100, obtenemos:

$$p_X(x) = \begin{cases} 1/100 & \text{para } x = 1, 2, \ldots 100 \\ 0 & \text{en otro caso} \end{cases}$$

La representación gráfica de la distribución uniforme es por medio de un gráfico de bastones o de puntos o por un histograma del mismo modo que representábamos las distribuciones de frecuencias en estadística descriptiva. En el eje de abscisas representamos los valores que toma X y el eje de ordenadas corresponde a los valores de $p_X(x)$. En el histograma todos los rectángulos tienen la misma altura y los mismo sucede con los bastones.

Esperanza y varianza

Para la **distribución uniforme sobre los enteros**, con función de masa

$$p_X(x) = \begin{cases} 1/k & \text{para } x = 1, 2, \ldots, k \\ 0 & \text{en otro caso} \end{cases}$$

la **esperanza** y la **varianza** son: $E(X) = (k+1)/2$ $\quad VAR(X) = (k^2 - 1)/12$

p = éxito
q = fracaso = 1 - p

Distribución Binomial

: experimento aleatorio.

B(n, p)

n = muchos sucesivos e independ. e idénticos

Supóngase que una máquina produce un artículo defectuoso con probabilidad p $(0 < p < 1)$ y produce uno no defectuoso con probabilidad $q = 1 - p$ y que se examinan *independientemente* n artículos de los producidos por la máquina. Si X es la variable aleatoria definida como el *"número de artículos defectuosos"* obtenidos en este experimento entonces los valores posibles de X serán $0, 1, 2, \ldots n$ artículos defectuosos y para cada uno de ellos existe una probabilidad asociada.

La probabilidad $P(X = x)$ es la probabilidad de obtener cualquier sucesión ordenada de n artículos consistente de x artículos defectuosos y $n - x$ no defectuosos.

- p es la probabilidad de seleccionar **un** artículo defectuoso y como los artículos se seleccionan en forma independiente p^x *es la probabilidad de elegir* x *artículos defectuosos*. En efecto $p^x = p \cdot p \ldots p$ donde p se multiplica x veces empleando la regla multiplicativa para eventos independientes.

- q / $q = 1 - p$ es la probabilidad de seleccionar **un** artículo no defectuoso, entonces, si empleamos la regla multiplicativa para eventos independientes, *la probabilidad de seleccionar* $n - x$ *artículos no defectuosos* es:

$$q \cdot q \cdot \ldots \cdot q = q^{(n-x)} = (1-p)^{(n-x)}$$

- La *probabilidad de obtener una sucesión ordenada en particular de* n *artículos conteniendo* x *defectuosos y* $n - x$ *no defectuosos* es $p^x \cdot (1-p)^{(n-x)}$ empleando nuevamente la regla multiplicativa para eventos independientes.

- Consideremos ahora el *número de sucesiones diferentes posibles de* n *artículos con* x *defectuosos y* $n - x$ *no defectuosos*. Tenemos en cuenta que los artículos defectuosos no se distinguen entre sí y los mismo sucede con los no defectuosos entonces dos sucesiones son diferentes si se presentan en distinto orden los artículos defectuosos y no defectuosos. Resulta que el número buscado es el **número combinatorio n sobre x**

$$\frac{n!}{(n-x)!\,x!} = \binom{n}{x}$$

En efecto, si consideramos todas las formas de arreglar n objetos distintos resulta $n!$ y si separamos los n objetos en un lote de x objetos y otro de $n - x$ objetos se tiene que todas las formas de arreglar el grupo de x objetos es $x!$ y las formas de arreglar el otro lote de $(n - x)$ objetos es $(n - x)!$. Resulta que las formas de arreglar el lote de x objetos con el lote de $(n - x)$ objetos es $x!\,(n - x)!$. Si sucede que en cada grupo de objetos no podemos distinguir un objeto de

$\binom{n}{x} = {}_nC_x = \dfrac{n!}{x!\,(n-x)!}$

número combinatorio n sobre x

otro existe una única forma de arreglarlos, entonces los n objetos separados en dos grupos uno de x objetos indistinguibles y otro de $(n-x)$ objetos distintos a los del otro lote pero indistinguibles entre sí podrán arreglarse en $n!\,/\,x!\,(n-x)!$ formas.

- Se tienen $\binom{n}{x}$ sucesiones con x artículos defectuosos y $n-x$ <u>no</u> defectuosos y para cada una de ellas la probabilidad es $p^{x}\,.\,(1-p)^{(n-x)}$, entonces :

x defectuosos y (n-x) no defectuosos

$$P(X=x) = \underbrace{\binom{n}{x}}_{*}\, p^{x}\,.\,q^{(n-x)}$$

$1-p = q =$ probab. de fracaso

p = probab. de éxito.
n = tamaño de la muestra

La **función de masa** para X siguiente :

$$p_X(x) = \begin{cases} \binom{n}{x}\, p^{x}\,.\,q^{(n-x)} & \text{si } x = 0,1,2,\ldots,n \\ 0 & \text{en otro caso} \end{cases}$$

$q = 1-p$.

se llama **binomial** con **parámetros** n y p y se simboliza $X \sim b(n,p)$

El experimento sobre cuyo espacio muestral se define una variable aleatoria binomial **X** debe reunir las siguientes propiedades que lo definen como un **proceso de Bernoulli** :

- El experimento consiste en n **pruebas repetidas** . (población infinita sin reemplazo ó de una población finita con reemplazo)
- Cada prueba tiene **dos resultados posibles** que pueden clasificarse en **éxito o fracaso** .
- La probabilidad de éxito , representada por p permanece **constante** para todos los intentos .
- Los **intentos** repetidos son **independientes** , es decir , el resultado de cualquier prueba es independiente del resultado de cualquier otra prueba .

Este modelo se aplica a poblaciones finitas o a poblaciones conceptualmente infinitas como las piezas que producirá una máquina o las personas que contraen una cierta enfermedad , de las que tomamos elementos al azar con reemplazamiento . En esta situación el proceso generador debe ser **estable** , proporción de piezas defectuosas o de personas enfermas se mantiene constante al largo plazo y **sin memoria** , el resultado en cada momento es independiente de lo previamente ocurrido.

Ejemplo : Supongamos el proceso de seleccionar 3 artículos en forma independiente de un proceso que se supone produce un 10 % de piezas defectuosas y se los clasifica como D (Defectuosos) o N (No defectuosos)

- Un artículo defectuoso se considera un éxito.
- El número de éxitos es una variable aleatoria **X** que asume los valores enteros 0 , 1 , 2 y 3
- Los artículos se seleccionan en forma independiente de un proceso que supone un 10 % de piezas defectuosas. O sea $p = 0{,}10$ y $q = 1-p = 0{,}90$.

La variable aleatoria **X** : *"número de piezas defectuosas(éxitos)"* en $n = 3$ experimentos es una variable aleatoria binomial y sus valores se representan por $X \sim b(n=3,\, p=0{,}1)$

$$p_X(x) = \begin{cases} \binom{3}{x}\, 0{,}1^{x}\,.\,0{,}9^{(3-x)} & \text{si } x = 0,1,2,3. \\ 0 & \text{en otro caso} \end{cases}$$

Reemplazando podemos obtener la siguiente tabla :

x	0	1	2	3
p x (x)	0,729	0,243	0.027	0,001

De modo que si queremos obtener la probabilidad de obtener hasta dos artículos defectuosos, hacemos :

$P(X \leq 2) = p_X(0) + p_X(1) + p_X(2) = 1 - p_X(3) = 0{,}999$

Esta distribución de probabilidad coincide con la obtenida si consideramos los 8 resultados posibles de este experimento , la variable aleatoria **X** , *" número de piezas defectuosas "* y sus probabilidades asociadas :

S = { N N N , N N D , N D N , N D D , D N N , D N D , D D N , D D D }

x	puntos muestrales	p x (x)
0	(N , N , N)	$0{,}9^3 = \binom{3}{0} 0{,}1^0 \; 0{,}9^3$
1	(N , N , D) - (N , D , N) - (D , N , N)	$3 \cdot 0{,}1 \; 0{,}9^2 = \binom{3}{1} 0{,}1^1 \; 0{,}9^2$
2	(N , D , D) - (D , N , D) - (D , D , N)	$3 \cdot 0{,}1^2 \; 0{,}9 = \binom{3}{2} 0{,}1^2 \; 0{,}9$
3	(D , D , D)	$0{,}1^3 = \binom{3}{3} 0{,}1^3 \; 0{,}9^0$

La distribución de probabilidades puede representarse mediante un diagrama de barras o puntos o por un histograma.

Para el cálculo de probabilidades pueden emplearse tablas confeccionadas a tal fin .

Esperanza y varianza (p/ distribución binomial)

Dada una variable aleatoria con distribución binomial **X** ~ **b (n , p)** donde **n** es el número de intentos y **p** la probabilidad de éxito en un intento determinado , la esperanza y la varianza de **X** son:

$$E(X) = n\,p \qquad VAR(X) = n\,p\,q$$

Distribución Hipergeométrica

La distribución hipergeométrica es una distribución cercanamente relacionada a la distribución binomial pues la variable surge de la extracción de elementos de una determinada población a los que podemos clasificar en una de dos categorías mutuamente excluyentes denominadas éxito y fracaso.

La diferencia entre ambas distribuciones está en la forma en que se lleva a cabo el muestreo o extracción .

En la distribución binomial se requiere independencia entre los intentos y el muestreo debe realizarse con el reemplazo de cada artículo después de observarlo . De este modo la probabilidad no cambia de un intento a otro.

En la distribución hipergeométrica **no se requiere independencia** y se basa en un **muestreo** llevado a cabo **sin reemplazo** donde **la probabilidad cambia de un intento a otro**.

En la práctica la mayoría de los muestreos se efectúan sin reemplazo ya que seleccionar al individuo varias veces en la muestra carece de sentido . Sin embargo, cuando la muestra es pequeña en relación al tamaño de la población es natural considerar que estamos en presencia de un esquema binomial .

Supongamos , por ejemplo que se debe extraer una muestra de tamaño N = 100 de una población de 1.000.000 de habitantes donde se conoce que 600.000 son varones ; obtenemos que:

- La probabilidad de que el primer individuo seleccionado sea varón es de (600.000/1.000.000) = 0,6
- La probabilidad de que en la selección número 100 se elija un varón si en todas las elecciones anteriores se seleccionó un varón es de (599.900 / 999.900) ≅ 0,59996 .

Si comparamos estos cocientes , vemos se puede considerar que la probabilidad de seleccionar un varón no varía de repetición en repetición y podemos adaptar el experimento a un modelo de muestreo con reemplazo.

Consideremos el caso que consiste en seleccionar 5 personas para integrar una comisión de un grupo de 20 personas , entre las que hay 14 mujeres . Sea la variable aleatoria **X** : *" cantidad de mujeres que integran la comisión "* , entonces :

- La probabilidad de obtener la primera mujer será (14 / 20) = 0,70
- La probabilidad de elegir la segunda mujer será (13 / 19) = 0,684 .
- La probabilidad de elegir una tercera de (12 / 18) = 0,67 y así sucesivamente.

Vemos aquí que , la probabilidad condicionada de extraer mujeres de esa población , no se mantiene constante en cada repetición del experimento y no resulta posible , en consecuencia , aplicar el esquema binomial. Lo mismo sucede en el muestreo de aceptación , las pruebas electrónicas y el aseguramiento de calidad , en donde , en muchos casos , las piezas que se prueban se destruyen y no pueden reemplazarse en la muestra. Entonces el muestreo es sin reemplazo.

Dado este caso, se desea conocer la probabilidad de que a la comisión de 5 personas la integren 3 mujeres o , dicho en forma equivalente, la probabilidad de seleccionar 3 mujeres de las 14 disponibles y 2 varones entre los 6 disponibles para constituir la comisión de cinco miembros.
Se observa que :

- Se tienen $\binom{14}{3}$ combinaciones de 3 mujeres tomados del grupo de 14 .
- Se pueden seleccionar de $\binom{6}{2}$ maneras a 2 hombres de entre los 6 disponibles
- El número total de modos de seleccionar 3 mujeres y 2 hombres en 5 intentos es el producto $\binom{14}{3} \cdot \binom{6}{2}$
- El número total de maneras de seleccionar 5 personas de las 20 disponibles es $\binom{20}{5}$
- La probabilidad de seleccionar 5 personas (sin reemplazo) de las cuales 3 son mujeres y 2 son hombres es :

$$p_X(3) = \frac{\binom{14}{3}\binom{6}{2}}{\binom{20}{5}} = \frac{\text{Casos favorables}}{\text{Total de casos}}$$

En este problema **X** es la variable aleatoria que determina el número de éxitos de un experimento , en este caso es *" el número de mujeres "* que ~~se~~ pueden integrar una comisión de 5 personas ; los valores posibles o el recorrido de **X** resultan : REC **X** = { 0 , 1 , 2 , 3 , 4 , 5 } y el tamaño de la muestra es n = 5 .

El éxito es obtener x mujeres y el fracaso es obtener n – x hombres.

Escaneado con Cam

Parametros = N, m, k.

La selección se efectúa en un grupo de 20 personas, o sea que el tamaño de la población es N = 20.

La cantidad total de éxitos posibles es el número total de mujeres en la población y lo designamos con k, en este caso k = 14. La cantidad de fracasos posibles es el número de hombres del grupo, o sea N – k en nuestro caso N – k = 20 - 14 = 6

Podremos, entonces, obtener para **X**:

$$p_X(x) = P(X = x) = \frac{\binom{14}{x}\binom{20-14}{5-x}}{\binom{20}{5}} \quad \text{si} \quad x = 0, 1, 2, 3, 4, 5$$

Esta expresión corresponde a una distribución hipergeométrica de la VA **X** de parámetro N = 20, n = 5, k = 6 que se identifica con el símbolo **X ~ h (N , n , k)** = h (20, 5, 14).

total → muestra → éxitos

Si calculamos para este ejemplo, obtenemos la distribución de probabilidades que se indica en la tabla

X	0	1	2	3	4	5
h (20 , 5 , 14)	6 / 15.504 (≅ 0,0004)	210/15.504 (≅ 0,0135)	1.820/15.504 (≅ 0,1174)	5.460/15.504 (≅ 0,3522)	6.006/15.504 (≅ 0,3874)	2.002/15.504 (≅ 0,1291)

Como vemos, la distribución hipergeométrica permite calcular la probabilidad de seleccionar x éxitos de los k posibles artículos también considerados éxitos y n – x fracasos de los N – k posibles resultados o artículos también considerados fracasos, cuando una muestra de tamaño n se selecciona de N resultados o artículos totales.

La distribución de probabilidad de una variable aleatoria **hipergeométrica** X (*número de éxitos en una muestra aleatoria de tamaño* n *seleccionada de* N *artículos totales, de los cuales* k *son considerados como éxitos y* N – k *como fracasos*) es :

forma de tomar los éxitos

$$p_X(x) = \mathbf{h(N, n, k)} = \begin{cases} \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} & \text{si} \quad x = 0, 1, 2, \ldots, n \\ 0 & \text{en otro caso} \end{cases}$$

→ forma de tomar los fracasos.

→ cant. total.

mín {n, k}

Esperanza y varianza

La esperanza y la varianza de una variable hipergeométrica **X ~ h (N , n , k)**, están dados por :

$E(\mathbf{X}) = n\,(k / N)$ $\qquad$ $VAR(\mathbf{X}) = n\,(k/N)\,[(N-k)/N]\,[(N-n)/(N-1)]$

Donde si hacemos (k / N) = p proporción de éxitos de la población y (N – k) / N = q proporción de fracasos y reemplazamos obtenemos :

$E(\mathbf{X}) = n\,p$ $\qquad$ $VAR(\mathbf{X}) = n\,p\,q\;\mathbf{[(N-n)/(N-1)]}$

es lo mismo que en la binomial

Al factor **(N – n) / (N – 1)** se lo conoce como **factor de corrección para poblaciones finitas.**

Se observa que la esperanza es la misma que para la distribución binomial y la varianza se obtiene multiplicando la varianza de la distribución binomial por el factor de corrección (N – n) / (N – 1).

Distribución de Poisson → Hay tablas y p(la binomial también hay tablas)

La distribución de probabilidades de **Poisson** donde la variable aleatoria es discreta y mide el número de resultados que ocurren en un intervalo de tiempo t , o en una región específica t se define :

$$p_X(x) = \begin{cases} [e^{-\lambda t} (\lambda t)^x] / x! & \text{para } x = 0, 1, 2, \ldots \\ 0 & \text{en otro caso} \end{cases}$$

λ → es el número promedio de resultados por unidad de tiempo o región

El intervalo de tiempo puede ser de cualquier duración , por ejemplo , un minuto , un día , una semana , un mes o inclusive un año . La variable aleatoria **X** puede representar :

- el número de llamadas por hora que se reciben en una oficina .
- el número de automóviles que llegan por día a una casilla de peaje .
- el número de juegos pospuestos por lluvia durante la temporada de fútbol
- el número de partículas radiactivas que pasa a traves de un contador durante un intervalo de 3 milésimas de segundo.

La región específica puede ser un segmento de línea , un área , un volumen o , tal vez , un pedazo de material. En este caso X puede representar :

- el número de ratas de campo por acre
- el número de bacterias en un determinado cultivo
- el número de pasas de uva en un pan de navidad
- el número de errores de mecanografiado por página

El proceso que llevará a resultados que generen una variable aleatoria con distribución de Poisson se conoce como **proceso de Poisson** y satisface tres condiciones que surgen de la observación de un fenómeno concreto durante un período de tiempo fijo o en una región específica

Primera condición

El número de ocurrencias en un intervalo de tiempo o región específicos es independiente del número de ocurrencias en cualquier otro intervalo disjunto de tiempo o región disjunta del espacio. De esta manera se dice que el proceso de Poisson no tiene memoria.

Segunda condición

La **probabilidad** de **una** ocurrencia durante cualquier **intervalo** de tiempo **muy corto** (o región muy pequeña) debe ser aproximadamente **proporcional a la longitud** de ese intervalo(o al tamaño de la región) y no depende del número de ocurrencias fuera de ese intervalo o región . Vale decir que dependerá solo de la longitud del intervalo o del tamaño de la región, longitud y tamaño pequeños, y será, en consecuencia, un valor pequeño . La probabilidad de observar exactamente un éxito en el intervalo (pequeño) es estable

Tercera condición

La probabilidad de que haya dos o más ocurrencias en cualquier intervalo de tiempo muy pequeño , o región muy pequeña debe ser despreciable en comparación con la probabilidad de **una** ocurrencia .

La probab. va disminuyendo .

Estas tres condiciones se sintetizan diciendo que los eventos o resultados que poseen la propiedad de ser **aleatorios y raros** definen una variable aleatoria con la distribución de Poisson .

El hecho que un suceso sea raro , en la práctica significa que el valor esperado de la variable aleatoria debe ser pequeño en comparación con el número máximo de eventos que pueden ocurrir en la unidad de tiempo o espacio. La aleatoriedad significa que la aparición de un éxito debe ser independiente de los éxitos anteriores en una misma unidad de tiempo o espacio.

Veremos la aplicación de esta distribución para el siguiente *ejemplo* :

Se estudia la distribución de células de levadura en 400 cuadrículas de un aparato utilizado para este recuento llamado hemacitómetro . Después de haber contado en cada una de las 400 cuadrículas la cantidad de células de levadura , se confeccionó la siguiente tabla de distribución de frecuencias para la variable aleatoria *" cantidad de células por cuadrícula "* :

Cantidad de células por cuadrícula (X)	*Cantidad de cuadrículas observadas*
0	75
1	103
2	121
3	54
4	30
5	13
6	2
7	1
8	0
9	1
Total	400

Puede observarse que hay 75 cuadrículas que no presentan ninguna célula de levadura y que la mayor parte de las cuadrículas poseen 1 o 2 células. Por el contrario 17 cuadrículas contienen 5 o más células de levadura.

La esperanza matemática o promedio de esta distribución es 1,8 células por cuadrícula . Vemos que se trata de un valor pequeño , considerando que la cantidad de células que podrían existir en la cuadrícula es muy grande . Esto nos lleva a pensar que estamos ante un acontecimiento relativamente raro. También podemos esperar que la existencia de células de levadura en un cuadrado es independiente de la existencia de este tipo de células en otros cuadrados .Así, este experimento origina para la variable aleatoria discreta una función de masa igual a la de Poisson.

Calculemos la probabilidad de que se observen 2 células de levadura en una cuadrícula en particular.
El parámetro que caracteriza a esta distribución es

$\lambda = 1,8$ células / cuadrícula y $t = 1$ cuadrícula

$\underbrace{\mu}_{\text{Media}} = \lambda \; t = 1,8$ células

Luego $P(X=2) = (e^{-1,8} \;\; 1,8^{2}) \,/\, 2! = 0,2678$

Este valor puede obtenerse empleando la tabla correspondiente , en donde se entra con los valores del parámetro y de la variable aleatoria ;1,8 y 2 respectivamente en este caso.

Esperanza y varianza

La **esperanza** y la **varianza** de una variable de **Poisson** son : $E(\mathbf{X}) = VAR(\mathbf{X}) = \mu = \lambda \; t$

Ejemplos :

- El número promedio de partículas radiactivas que pasan a través de un contador durante un milisegundo en un experimento de laboratorio es 4 ¿Cuál es la probabilidad de que entren 6 partículas al contador en un milisegundo determinado ?

Se trata de una distribución de Poisson de parámetro $\mu = 4$ y debemos calcular $P(\mathbf{X} = 6)$.

Resulta : $P(\mathbf{X}=6) = (e^{-4} \;\; 4^{6}) \,/\, 6! = 0,1042$

- Se sabe que 10 es el número promedio de camiones tanque de aceite que llegan por día a una cierta ciudad portuaria . Las instalaciones del puerto pueden atender , cuanto mucho , a 15 camiones – tanque en un día . ¿ Cuál es la probabilidad de que en un determinado día se tengan que regresar los camiones tanque ?

Se trata de una distribución de Poisson con parámetro $\mu = 10$ para la variable aleatoria discreta **X** : *"cantidad de camiones que llegan por día "* para la que debemos calcular $P(X > 15)$

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} (e^{-10}\ 10^{x}) \ / \ x! = 1 - 0{,}9513 = 0{,}0487$$

La $P(X \leq 15)$ puede obtenerse de la tabla de distribución acumulada para Poisson en la que se ingresa con el parámetro y el último valor de la variable aleatoria ; $\mu = 10$ y $x = 15$ respectivamente en este caso.

La distribución de Poisson como límite de la distribución binomial

Para una variable aleatoria con distribución binomial si el número de pruebas es grande ($n \geq 20$) ~~en la~~ y p , la probabilidad de éxito cercana a cero ($p \leq 0{,}05$), la distribución de Poisson puede utilizarse , para aproximar la distribución binomial. En estos casos se toma el parámetro de Poisson como $\mu = n\,p$.

Ejemplo :

En un proceso de manufactura en el cual se producen piezas de vidrio , ocurren defectos o burbujas. Se sabe que , en promedio una de cada mil piezas tiene una o más burbujas . ¿ Cuál es la probabilidad de que en una muestra aleatoria de 8.000 piezas , menos de 7 de ellas tengan burbujas ?

Este es un experimento donde la variable aleatoria **X** : " número de piezas con burbujas " es binomial con $n = 8000$ y $p = 0{,}001$ donde la probabilidad solicitada se obtiene haciendo :

$$P(X < 7) = P(X \leq 6) = \sum_{x=0}^{6} \binom{8000}{x}\ 0{,}001^{x}\ 0{,}999^{\,n-x}$$

Dado que p se acerca a 0 y n es suficientemente grande se puede aproximar la distribución binomial a la de Poisson con parámetro $\mu = 8000\ 0{,}001 = 8$ y obtener la probabilidad requerida del siguiente modo :

$$P(X < 7) = \sum_{x=0}^{6} (e^{-8}\ 8^{x}) \ / \ x! = 0{,}3134$$

Se agregan como anexos las tablas que ayudan al cálculo de probabilidades de las distribuciones binomial y de Poisson .

DISTRIBUCIONES DE PROBABILIDAD DE VARIABLES ALEATORIAS CONTINUAS

Distribucion Uniforme (o Rectangular)

Sean a y b dos números reales tales que $a < b$ y sea el experimento que consiste en seleccionar un punto X del intervalo [a , b] de forma que ***la probabilidad de que X pertenezca a cualquier intervalo contenido en [a , b] es proporcional a la longitud de ese intervalo*** .

La constante de proporcionalidad es $1 / (b - a)$.

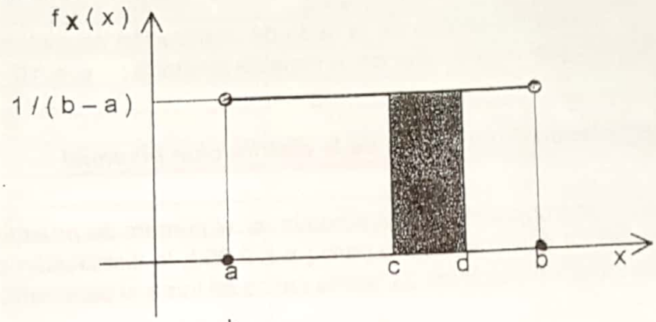
Esta distribución de la VA X continua se denomina **uniforme (o rectangular) sobre [a , b]**. Entonces :

La **función de densidad** de una variable aleatoria continua **X** definida como

$$f_X(x) = \begin{cases} 1/(b-a) & a < x < b \\ 0 & \text{en otro caso} \end{cases}$$

se denomina **uniforme (o rectangular) sobre [a , b]**

La representación gráfica de esta función resulta :



Además obtenemos : $\int_{-\infty}^{\infty} f_X(x)\, dx = \int_a^b [1/(b-a)]\, dx = 1$

$$P(c < X < d) = \int_c^d [1/(b-a)]\, dx = (d-c)/(b-a)$$

Puesto que la probabilidad que se seleccione uno de los puntos extremos **a** o **b** es cero, es irrelevante que en la expresión de $f_X(x)$ se indiquen intervalos abiertos , cerrados o semiabiertos .

Esperanza y varianza

Como la esperanza de una VA continua se definió como $\mu_X = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$, reemplazando , se obtiene:

$$\mu_X = \int_a^b [x/(b-a)]\, dx = (x^2/2)(b-a)\Big]_a^b = (b^2 - a^2)/2(b-a)$$

$$\boxed{\mu_X = (b+a)/2}$$

Teniendo en cuenta que la varianza de una variable aleatoria continua se definió :

$\sigma_X^2 = VAR(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx$, reemplazando e integrando, obtenemos

$$\boxed{\sigma_X^2 = (b-a)^2 / 12}$$

Distribucion Normal

Distribución normal estándar

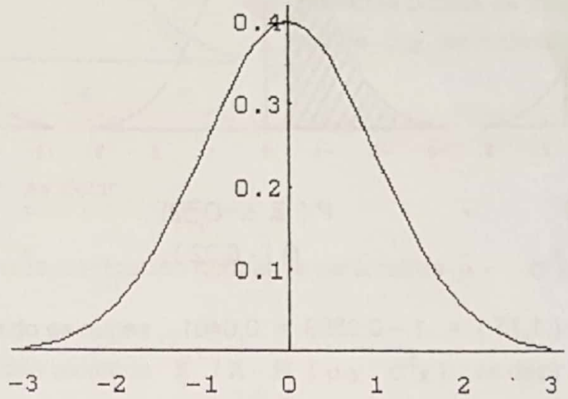
Una variable aleatoria continua **Z** tiene una **distribución normal estándar** si la función de densidad es

$$f_Z(z) = (1/\sqrt{2\pi})\, e^{-z^2/2} \qquad -\infty < z < \infty$$

y su gráfica recibe el nombre de **curva normal**.

La curva normal tiene forma de campana y tiene las siguientes propiedades :

- La moda que es el valor de la abscisa donde la curva tiene su máximo y ocurre en $z = 0$.
- La curva es simétrica con respecto al eje de ordenadas .
- Los puntos de inflexión están en $z = \pm 1$. Resulta cóncava hacia abajo en si $-1 < z < 1$ y cóncava hacia arriba si $z > 1$ o $z < -1$.
- La curva se acerca al eje de abscisas en forma asintótica en cualquiera de los dos sentidos alejándose del cero ($f_Z(z) \to 0$ si $z \to \infty$ o si $z \to -\infty$).
- El área total entre la curva y el eje z u eje de abscisas es 1 .



Una variable aleatoria Z que se distribuye según la normal estándar se simboliza $Z \sim N_E$.

La **esperanza** o **media** de esta distribución es 0 y es decir $E(Z) = \mu_Z = 0$.

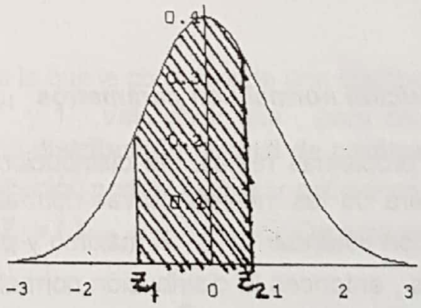
La **desviación estándar** es 1. Resulta $VAR(Z) = \sigma_Z^2 = 1 \Rightarrow \sigma_Z = \sqrt{VAR(Z)} = 1$

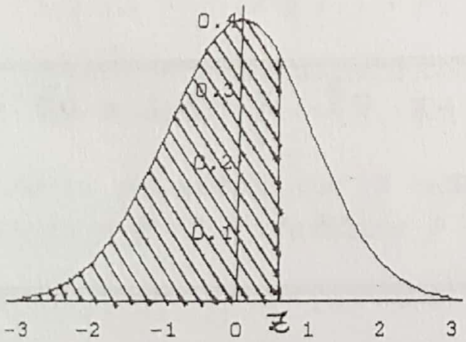
Estos valores, 0 y 1 son los **parámetros** de la distribución normal estándar.

Para calcular probabilidades para una variable aleatoria con distribución normal estándar , empleamos :

$$P(z_1 < Z < z_2) = \int_{z_1}^{z_2} (1/\sqrt{2\pi})\, e^{-z^2/2}\, dz$$

El resultado de esta integral definida coincide con el área de la región sombreada . . ⟶





Sea $\phi(z)$, la **FDA** de la normal estándar , entonces

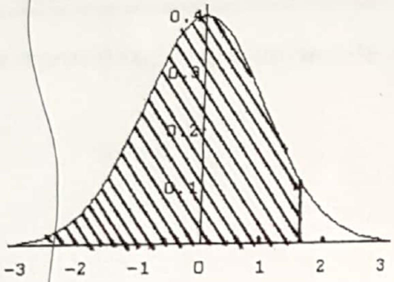
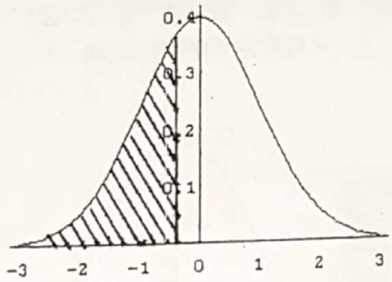
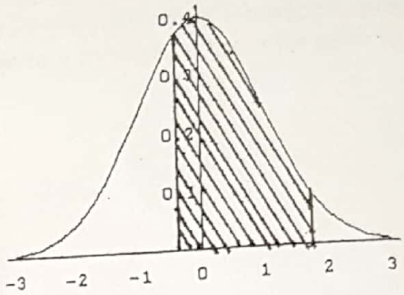
$$\phi(z) = P(Z \le z) = \int_{-\infty}^{z} (1/\sqrt{2\pi})\, e^{-u^2/2}\, du$$

⟵ . Esta probabilidad tiene un valor que coincide con el área sombreada.

Es habitual emplear tablas en donde figuran los valores de la función distribución acumulada (áreas bajo la curva normal estándar a la izquierda de z) para calcular probabilidades.

Si calculamos probabilidades usando esta tabla para una VA (Z) / $Z \sim N_E$, resulta :

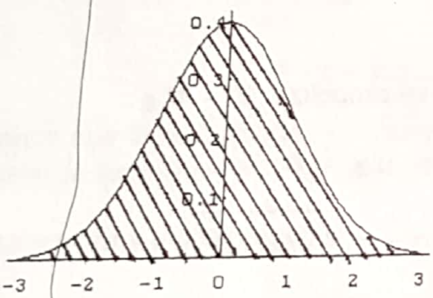
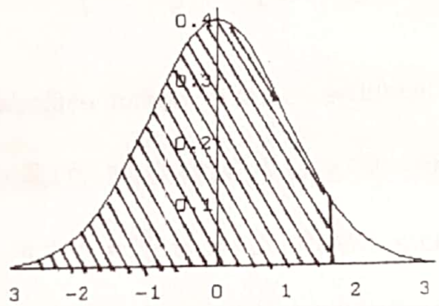
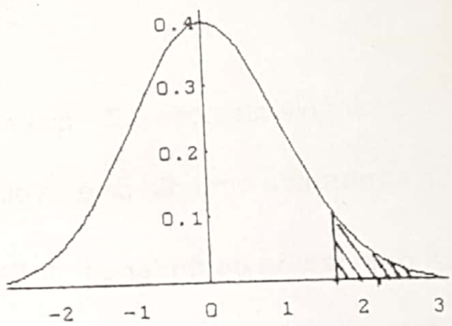
- $P(Z \le 1{,}75) = \phi(1{,}75) = 0{,}9599$ si entramos en la tabla con $z = 1{,}75$
- $P(-0{,}32 \le Z \le 1{,}75) = \phi(1{,}75) - \phi(-0{,}32) = 0{,}9599 - 0{,}3745 = 0{,}5854$
si entramos en la tabla con $z = 1{,}75$ y a este valor le restamos el que corresponde a $z = -0{,}32$

$P(Z \le 1{,}75)$ - $P(Z \le -0{,}32)$ = $P(-0{,}32 < Z \le 1{,}75)$

$\phi(1{,}75)$ - $\phi(-0{,}32)$ = $P(-0{,}32 < Z \le 1{,}75)$

- $P(Z > 1{,}75) = 1 - \phi(1{,}75) = 1 - 0{,}9599 = 0{,}0401$, según se observa a continuación :

1 - $P(Z \le 1{,}75)$ = $P(Z > 1{,}75)$

1 - $\phi(1{,}75)$ = $P(Z > 1{,}75)$

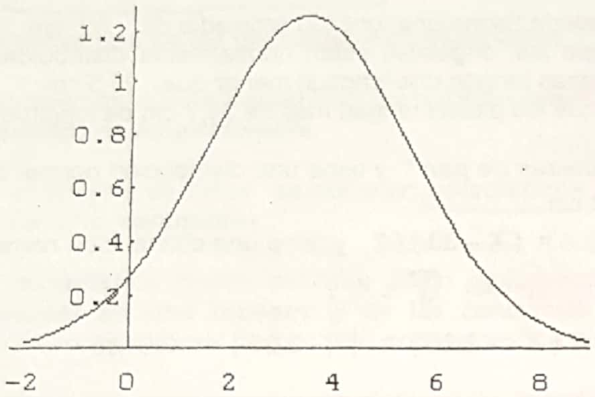
Distribución normal con parámetros μ_X , σ^2_X

En los problemas reales, las distribuciones de la variable aleatoria X pueden tomar la forma de una cualquiera de las infinitas curvas normales distintas que surgen al variar los valores de media μ_X y desviación estándar σ_X , el máximo y punto de inflexión de la curva respectivamente. En estos casos tenemos , entonces la distribución normal con parámetros $\mu_X / \mu_X \neq 0$ y $\sigma^2_X / \sigma^2_X \neq 1$ y que se indica $X \sim N(\mu_X, \sigma^2_X)$

Una variable aleatoria tiene **distribución normal con parámetros** μ_X , σ^2_X ($X \sim N(\mu_X, \sigma^2_X)$) si su **función de densidad** puede expresarse como :

$$f_X(x) = (1/\sigma_X \sqrt{2\pi})\ e^{-(½)[(x-\mu_X)/\sigma_X]^2} \qquad -\infty < x < \infty$$

La función definida tiene la representación gráfica siguiente para el caso de $\mu_X = 3{,}5$ y $\sigma_X = 2$:



La curva tiene las siguientes propiedades :

- La moda ocurre en $x = \mu_X$. La moda coincide con la media y con la mediana
- Es simétrica con respecto al eje vertical de ecuación $x = \mu_X$.
- Presenta puntos de inflexión en $x = \mu_X \pm \sigma_X$ donde σ_X es la desviación estándar

- El área bajo la curva es uno ; es decir:

$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$, $f_X(x)$ es la distribución normal de parámetros μ_X , σ^2_X ($X \sim N(\mu_X, \sigma^2_X)$)

Sea ξ la FDA para la variable aleatoria $X \;/ X \sim N(\mu_X, \sigma^2_X)$, es decir :

$$\xi(x) = P(X \le x) = \int_{-\infty}^{x} (1/\sigma_X\sqrt{2\pi})\, e^{-(½)[(t-\mu_X)/\sigma_X]^2}\, dt$$

Proponemos la **sustitución** $z = (t-\mu_X)/\sigma_X$ y obtenemos : $dz = dt/\sigma_X$

si $t = x \Rightarrow z = (x-\mu_X)/\sigma_X$ y si $t = -\infty \Rightarrow z = -\infty$

Reemplazamos en la expresión para la FDA , la función ξ y resulta :

$$P(X \le x) = \int_{-\infty}^{(x-\mu_X)/\sigma_X} (1/\sqrt{2\pi})\, e^{-(½)z^2}\, dz = P(Z \le (x-\mu_X)/\sigma_X) = \phi((x-\mu_X)/\sigma_X)$$

Así hemos definido una VA Z donde $Z = (X-\mu_X)/\sigma_X$ a la que le corresponde una distribución normal estándar , $Z \sim N_E$ donde los parámetros son 0 y 1 .Vale decir que , para calcular probabilidades de una variable aleatoria X cuando X tiene una **distribución normal de parámetros** $\mu_X \neq 0$, $\sigma^2_X \neq 1$ empleamos los valores conocidos para la distribución normal estándar definiendo esta nueva variable aleatoria llamada **variable estandarizada** $Z \;/ Z = (X-\mu_X)/\sigma_X$. De este modo , podemos calcular la probabilidad buscada haciendo :

$$P(X \le x_1) = P(Z \le z_1) = P((X-\mu_X)/\sigma_X \le (x_1-\mu_X)/\sigma_X)$$

e ingresando en la tabla de distribuciones acumuladas para la distribución normal estándar con el valor z_1 obtenido como $z_1 = (x_1-\mu_X)/\sigma_X$

Tenemos , por ejemplo , que X se distribuye normalmente con media $\mu_X = 3{,}5$ y desviación estándar $\sigma_X = 1$, para calcular $P(X < 3{,}9)$ hacemos :

$$P(X < 3{,}9) = P(Z < (3{,}9 - 3{,}5)/1) = P(Z < 0{,}4) = 0{,}6554$$

Este valor lo obtuvimos ingresando en la tabla de distribución acumulada con el valor 0,4

$N \sim (0,1)$; desviación media estandar con parametro medio 0 y desviación estandard 1

Ejemplos :

1) Las piezas de pan distribuidas por una cierta panadería tienen una longitud promedio de 30 cm y una desviación estándar de 2 cm . Suponiendo que las longitudes están normalmente distribuidas, obtener la probabilidades siguientes : A) que las piezas tengan una longitud menor que 25,5 cm . B) que la longitud esté entre 29,3 y 33,5 cm y C) que las piezas tengan más de 31,7 cm de longitud .

Tipificación

La variable aleatoria **X** es la ***" longitud de las piezas de pan "*** y tiene una distribución normal de media $\mu_X = 30$ cm y desviación estándar $\sigma_X = 2$ cm .

La variable estandarizada es $Z = (X - \mu_X) / \sigma_X = (X - 30) / 2$ y tiene una distribución normal estándar. (con media $\mu = 0$ y desviación típica $\sigma = 1$)

A) $P(X < 25{,}5) = P(Z < (25{,}5 - 30)/2) = P(Z < -2{,}25) = \phi(-2{,}25) = 0{,}0122$

En este caso buscamos en la tabla el valor correspondiente a $z = -2{,}25$. Este valor corresponde al área bajo la normal a la izquierda de $-2{,}25$.

B) $P(29{,}3 < X < 33{,}5) = P((29{,}3 - 30)/2 < Z < (33{,}5 - 30)/2) = P(-0{,}35 < Z < 1{,}75)$

$\Rightarrow P(29{,}3 < X < 33{,}5) = \phi(1{,}75) - \phi(-0{,}35) = 0{,}9599 - 0{,}3632$

En esta situación al valor correspondiente al área a la izquierda de 1,75 , le restamos el que se obtiene para el área a la izquierda de - 0,35 .

C) $P(X > 31{,}7) = P(Z > (31{,}7 - 30)/2) = P(Z > 0{,}85) = 1 - P(Z < 0{,}85) \Rightarrow$

$\Rightarrow P(X > 31{,}7) = 1 - \phi(0{,}85) = 1 - 0{,}8023 = 0{,}1977$

2) Dada una distribución normal con media $\mu_X = 40$ y desviación estándar $\sigma_X = 6$, encuentre el valor de x que tiene : A) El 45 % del área a su izquierda. B) El 14 % del área a la derecha

A) Se requiere un valor x_1 de **X** que tenga un área de 0,45 a su izquierda ; es decir

$$P(X < x_1) = 0{,}45 \quad \text{donde} \quad X \sim N(40, 6^2)$$

Para la variable estandarizada $Z = (X - \mu_X) / \sigma_X$ se obtiene $P(Z < z_1) = 0{,}45$ si $z_1 = -0{,}13$ buscando en la tabla el valor z_1 correspondiente a un valor de área a su izquierda igual o lo más próximo posible a 0,45 .

Como $z_1 = (x_1 - \mu_X) / \sigma_X$ entonces $x_1 = z_1 \sigma_X + \mu_X$

En nuestro caso el valor buscado es $x_1 = -0{,}13 \cdot 6 + 40 = 39{,}22$

B) Se debe calcular el valor x_2 tal que $P(X > x_2) = 0{,}14$

Sabemos que $P(X > x_2) = 1 - P(X < x_2) = 0{,}14$; entonces como $P(X < x_2) = 0{,}86$ el valor x_2 lo obtengo buscando en la tabla el z que corresponde a un área a su izquierda igual o lo más próximo posible a 0,86 .

Así obtenemos que $P(Z < z_2) = 0{,}86$ si $z_2 = 1{,}08$.

El valor de x_2 buscado es $x_2 = z_2 \sigma_X + \mu_X = 1{,}08 \; 6 + 40 = 46{,}48$