Estadística

Clase 3

Rodrigo Del Rosso

29 de Abril de 2022

Variables Aleatorias

UNIVERSIDAD AUSTRAL | INGENIERÍA

Supongamos que se realiza el siguiente experimento aleatorio:

- ***Arrojar tres veces al aire una moneda equilibrada***

El espacio muestral de este experimento aleatorio esta compuesto por los siguientes resultados,

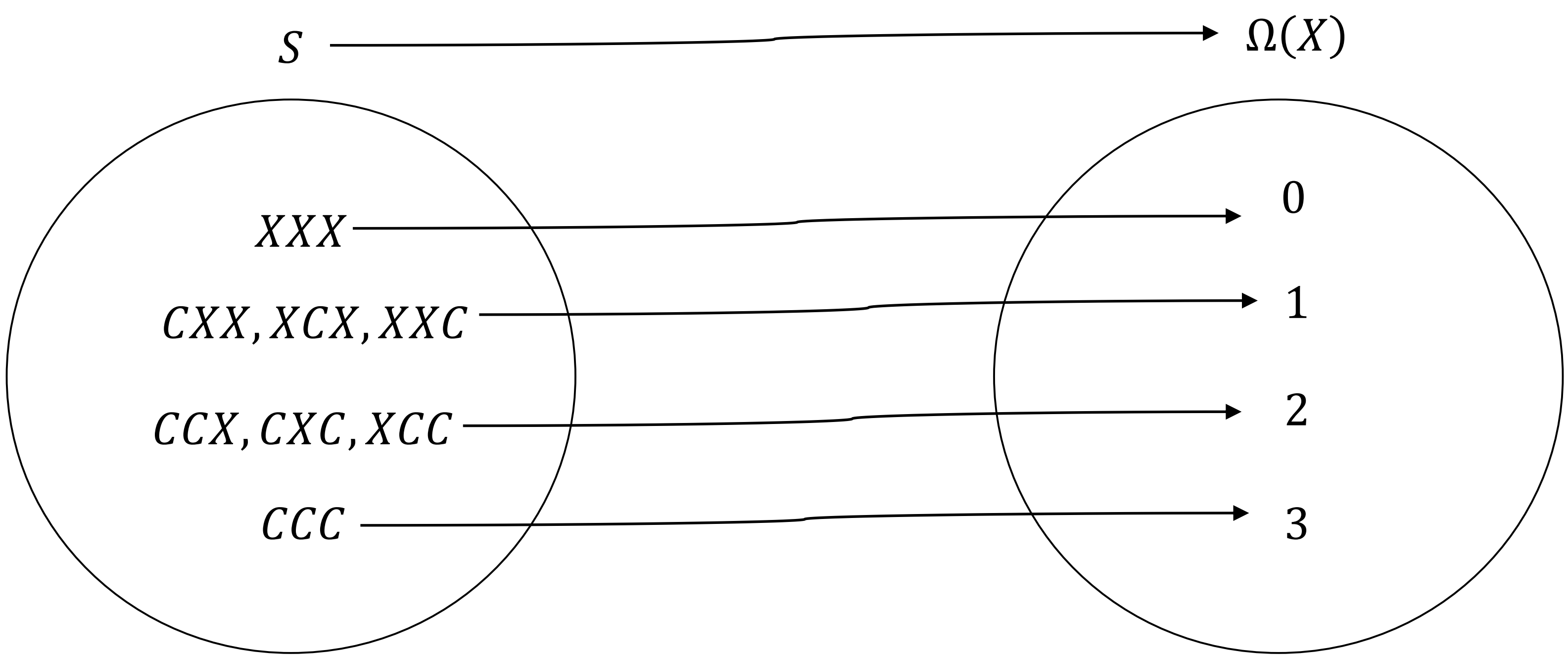
$$S = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$$

Y supongamos que queremos cuantificar la cantidad de caras que salen al realizar este experimento aleatorio, es decir estamos interesados en determinar cuantas veces sale cara,

$$X = Cantidad\ de\ Caras\ que\ salen\ al\ lanzar\ 3\ veces\ una\ moneda\ al\ aire$$

Por lo tanto, nos queda definido una variable que asume, en este ejemplo particular, los siguientes resultados,

$$\Omega(X) = \{0,1,2,3\}$$



Una variable aleatoria es una función o regla bien definida que asigna a cada elemento del espacio muestral S un número real dentro del dominio $\Omega(X)$

$$\boldsymbol{X: S \to \Omega(X)}$$

En función del dominio del Espacio Muestral las variables aleatorias se pueden clasificar en:

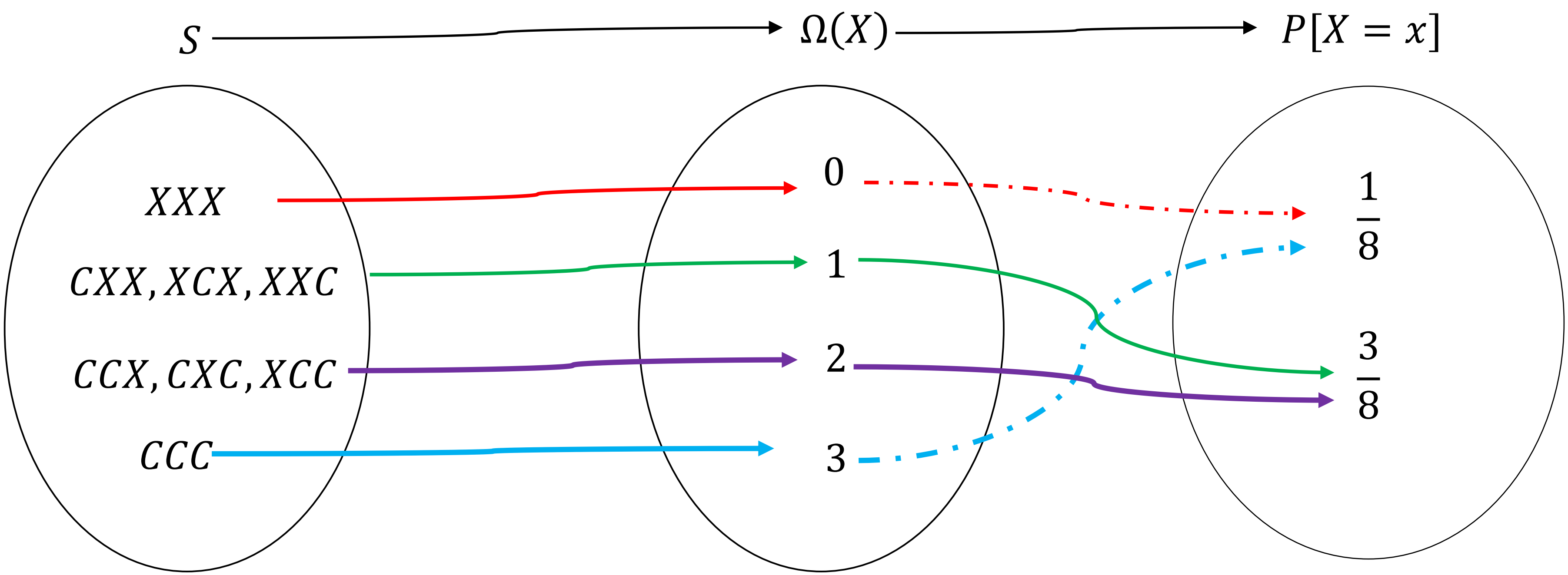
- **Variable Aleatoria Discreta**
  - Espacio Muestral Finito
  - Espacio Muestral Infinito Numerable
- **Variable Aleatoria Continua**
  - Espacio Muestral Infinito No Numerable

El ejemplo precedente es de variable discreta.

El Espacio Muestral es finito, el experimento aleatorio asume una cantidad finita de resultados.

Variable Aleatoria Discreta

A partir de los valores que asuma la variable aleatoria y los resultados posibles del Experimento Aleatorio, podemos definir probabilidades de ocurrencia asociadas con cada evento aleatorio.



A cada valor que asume la variable aleatoria se le asignó un valor de probabilidad. Se entiende que cada valor que asume X representan eventos mutuamente excluyentes que forman un sistema exhaustivo.

En el gráfico anterior apareció una nueva función $P[X = x]$ la cual realiza el siguiente mapeo,

$$P[X = x]: \Omega(X) \to (0,1)$$

Es decir que a cada valor posible que asume la variable esta función le asigna un valor de probabilidad entre 0 y 1.

A esta función se la denomina **"Función de Probabilidad"** o "Función Masa de Probabilidad". En esta curso utilizaremos la primera forma de llamarla.

Surge la siguiente pregunta,

¿Cualquier función puede ser considera "Función de Probabilidad"?

Para que una función pueda ser considerada Función de Probabilidad debe cumplir con dos condiciones en forma simultánea,

- **Condición de Cierre**

Esta asociada en que los valores que asume la variable aleatoria son excluyentes y forma un sistema exhaustivo. En términos formales,

$$\sum_{x=a}^{b} P[X = x] \qquad \Omega(X) = [a, \dots, b]$$

- **Condición de no negatividad**

Esta asociado en que la imagen de esta función sea positiva. Es decir, se excluye la posibilidad del conjunto vacío.

$$P[X = x] > 0 \;\; para\; todo\; x \in \Omega(X)$$

Los valores del tercer conjunto pueden agruparse en una **tabla de probabilidades**, también conocida como **distribución de probabilidades**. Este concepto es similar al de tabla de frecuencias. Es decir,

X	$P[X=x]$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
Total	1

Es decir, que a cada valor posible que asuma la variable se le asocia un valor de probabilidad. Se cumple la siguiente dupla $(X, P[X=x])$

Ahora bien, con esta distribución de probabilidades podría interesarnos calcular las siguientes medidas,

- Función de Distribución
- Función de Supervivencia
- Esperanza Matemática
- Varianza
- Desvío Estándar
- Mediana
- Modo
- Percentiles
- Etc

Función de Distribución (de Probabilidad)

Esta función es análoga a la Frecuencia Relativa Acumulada en Estadística Descriptiva. Indica la probabilidad acumulada desde el mínimo valor que asume la variable hasta un valor específico. Formalmente,

$$F_X(x) = F[X = x] = P[X \leq x]$$

Cuando trabajamos con una distribución de probabilidad tabular (tabla) debemos agregar una columna más.

X	$P[X = x]$	$F[X = x]$
0	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{3}{8}$	$\frac{4}{8}$
2	$\frac{3}{8}$	$\frac{7}{8}$
3	$\frac{1}{8}$	1
Total	1	

De la tabla anterior se observa que la probabilidad acumulada a 2 es $\frac{7}{8}$, es decir que hay una probabilidad de $\frac{7}{8}$ de que la variable asuma un valor de como máximo igual a 2.

$$F_X(2) = F[X = 2] = P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] = \frac{7}{8}$$

La Función de Distribución nos servirá además para calcular Percentiles entre otras cuestiones.

Por ejemplo si quisiéramos la probabilidad de que la variable aleatoria se encuentre entre 1 y 3 podría plantearse lo siguiente,

$$P[1 \leq X \leq 3] = P[X = 1] + P[X = 2] + P[X = 3] = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8}$$

O formalmente plantearlo mediante la Función de Distribución,

$$P[1 \leq X \leq 3] = F[X = 3] - F[X = 0] = 1 - \frac{1}{8} = \frac{7}{8}$$

Función de Supervivencia (de Probabilidad)

Esta función es análoga al Complemento de la Frecuencia Relativa Acumulada en Estadística Descriptiva. Indica la probabilidad acumulada a partir del valor siguiente evaluado en la función. Formalmente,

$$S_X(x) = S[X = x] = P[X > x]$$

$\boldsymbol{X}$	$\boldsymbol{P[X = x]}$	$\boldsymbol{F[X = x]}$	$\boldsymbol{S[X = x]}$
$\mathbf{0}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{7}{8}$
$\mathbf{1}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{4}{8}$
$\mathbf{2}$	$\frac{3}{8}$	$\frac{7}{8}$	$\frac{1}{8}$
$\mathbf{3}$	$\frac{1}{8}$	$\mathbf{1}$	$\mathbf{0}$
Total	$\mathbf{1}$		

$$S_X(x) = 1 - F_X(x) \Rightarrow P[X \leq x] + P[X > x] = 1$$

Esperanza Matemática

Este concepto está vinculado fuertemente a los inicios de los juegos de azar.

Supongamos que entre dos personas juegan el siguiente juego,

- Cada persona tira una moneda al azar respetando el turno
- Si sale "Cara" la persona "A" le paga a la persona "B" la suma fija de $ 1
- Si sale "Ceca" la persona "A" recibirá de la persona "B" la misma suma.

Ahora bien, supongamos que se lanzaron 5 veces la moneda y los resultados son los siguientes,

$$Persona\ A\!: \quad \$\,1 \quad -\$\,1 \quad \$\,1 \quad \$\,1 \quad \$\,1$$

$$Persona\ B\!: \quad -\$\,1 \quad \$\,1 \quad -\ \$\,1 \quad -\ \$\,1 \quad -\ \$\,1$$

Pareciera que la persona "A" de las 5 tiradas ha recibido una suma acumulada de $ 4.

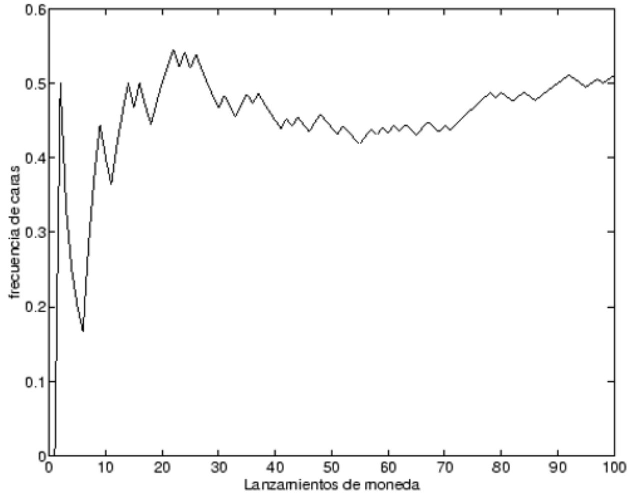
¿Qué ocurriría si se repite este juego 100?

¿Qué ocurriría si se repite este juego 10.000?

Y si, se repite un millón de veces?

Lo que se observaría es que de 100 tiradas en aproximadamente la mitad de las veces la moneda tendería a las dos caras. Es decir, habría aproximadamente un 50% de probabilidad de que salga tanto cara como cruz.

Hint: **Principio de Estabilidad de la Frecuencia Relativa**



La Esperanza Matemática o Valor Esperado de la Variable Aleatoria es un concepto vinculado en un marco ideal de repetición al infinito del experimento aleatorio.

Se dice que **la Esperanza Matemática "es aquel valor que se espera obtener si se repite el experimento una gran cantidad de veces".** Ahora bien, **¿Qué se entiende por "gran cantidad de veces"?**

Teniendo en cuenta lo precedente se dice que la Esperanza Matemática está asociado con el precio justo que los apostadores ganarían (perderían) al entrar en un juego de esta naturaleza. El valor esperado de este juego es de cero.

$$Expected\ Value = \$\,1 * \frac{1}{2} + (\$\,1) * \frac{1}{2} = 0$$

Puede entenderse como aquel valor que se esperaría ganar o perder si se apuesta una infinidad de veces.

Tarea: Buscar en Internet sobre la Ruina del Apostador y su vinculación con las apuestas en los Casinos.

Formalmente el valor esperado o Esperanza Matemática puede definirse de la forma siguiente,

$$E[X] = \sum_{x=a}^{b} xP[X = x] \qquad \Omega(X) = [a, \dots, b]$$

La Esperanza Matemática resulta de ponderar cada valor posible que pueda asumir la variable por su correspondiente probabilidad y sumarla en todo su dominio. Esta fórmula es similar a la media aritmética simple cuando la calculábamos con la frecuencia relativa simple. En términos formales,

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i * f_i}{n} = \sum_{i=1}^{n} X_i * fr_i$$

Y por el principio de estabilidad de la frecuencia relativa podría pensarse de la forma siguiente,

$$\bar{X} \to E[X]$$

En nuestro ejemplo deberíamos incorporar una columna más para representar el producto de cada valor posible y su función de probabilidad. Es decir,

X	$P[X=x]$	$F[X=x]$	$S[X=x]$	$xP[X=x]$
0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{7}{8}$	0
1	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{4}{8}$	$\frac{3}{8}$
2	$\frac{3}{8}$	$\frac{7}{8}$	$\frac{1}{8}$	$\frac{6}{8}$
3	$\frac{1}{8}$	1	0	$\frac{3}{8}$
Total	1			$\frac{12}{8}=1,5$

La Esperanza Matemática de este experimento aleatorio que consistió en arroja al aire tres veces una moneda al azar es de $E[X]=1{,}5\ caras$.

Ahora bien, **¿es posible que la cantidad de caras no sea un valor de los posibles que pueda asumir la variable?**

Varianza

Para calcular la varianza con una distribución de probabilidad se debe calcular de la forma siguiente,

$$Var[X] = E[(X - E[X])^2] = \sum_{x=a}^{b} (x - E[X])^2 P[X = x] \qquad \Omega(X) = [a, \dots, b]$$

Si se resuelve el Binomio de Newton en la expresión anterior se llega a la siguiente expresión,

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

$$Var[X] = \sum_{x=a}^{b} x^2 P[X = x] - \left(\sum_{x=a}^{b} x P[X = x] \right)^2$$

El segundo término lo tenemos, dado que es la Esperanza Matemática. El primer término se puede calcular con la tabla de probabilidad agregando una columna más.

Tarea: Calcular Varianza, Desvío Estándar y Coeficiente de Variación.

El desvío estándar y coeficiente de variación se calculan de la misma forma como vimos con Estadística Descriptiva,

$$Var[X] = \sum_{x=a}^{b} x^2 P[X = x] - \left(\sum_{x=a}^{b} xP[X = x] \right)^2$$

$$SD[X] = \sqrt{Var[X]} = \sqrt{\sum_{x=a}^{b} x^2 P[X = x] - \left(\sum_{x=a}^{b} xP[X = x] \right)^2}$$

$$CV[X] = \frac{SD[X]}{E[X]}$$

Modo

Esta medida se asocia con el valor más probable. A diferencia de Estadística Descriptiva, que se asociaba con aquel valor de variable más frecuente. En el ejemplo considerado no hay un valor más frecuente. Por lo tanto no hay modo.

Mediana

Para determinarla se emplea el mismo concepto que en Estadística Descriptiva. Es el primer valor de variable que supera el 50% de probabilidad. Por lo tanto, alcanza con observar la columna asociada a la Función de Distribución.

En la situación de coincidir el ORP (Orden Relativo Percentilar) del 50% con la Función de Distribución, se procederá de igual manera que como en Estadística Descriptiva, tomando como valor de Mediana la semi suma entre ese valor de variable y el siguiente inmediato en la tabla.

Percentiles

El procedimiento es similar a Estadística Descriptiva.

Variable Aleatoria Continua

Asociada a cada variable aleatoria continua se dispondrá de una **función de densidad** (de probabilidad) que será "similar" a la función de probabilidad en variable discreta.

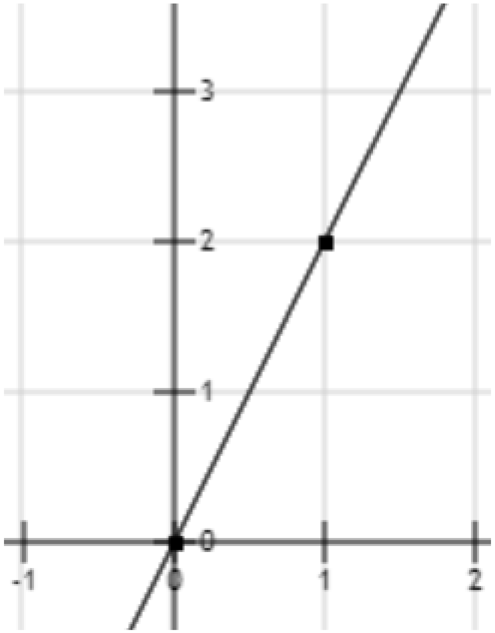
Supongamos que X es una variable continua con dominio continuo en el intervalo $[0,1]$ y que posee la siguiente función de densidad,

$$\Omega(X) = [0,1]$$

$$f_X(x) = 2x$$

$$f_X(0) = 2 * 0 = 0$$

$$f_X(1) = 2 * 1 = 2$$



Entonces la imagen de esta función no nos devuelve un valor de probabilidad. (Diferencia importante con una variable discreta!!!!)

¿Cómo podemos calcular probabilidades? ¿Qué se les ocurre?

Tener presente que una función para ser considerada función de densidad debe cumplir con las mismas condiciones que para una variable discreta,

- **Condición de Cierre**

Esta asociada en que los valores que asume la variable aleatoria son excluyentes y forma un sistema exhaustivo. En términos formales,

$$\int_a^b f_X(x)\,dx = 1$$

- **Condición de no negatividad**

Esta asociado en que la imagen de esta función sea positiva. Es decir, se excluye la posibilidad del conjunto vacío.

$$f_X(x) > 0 \quad para\, todo\; x \in \Omega(X)$$

¿Qué ocurre en nuestro ejemplo?

$$\int_0^1 2x\,dx = 2\frac{x^2}{2}\Big|_0^1 = 1^2 - 0^2 = 1$$

La función de densidad $f_X(x) = 2x$ cumple con la condición de cierre y como vimos antes la imagen de esta función en todo su dominio es positiva. Por lo tanto $f_X(x) = 2x$ es una función de densidad.

Ahora bien, si quisiéramos calcular la siguiente probabilidad podríamos emplear el mismo procedimiento que utilizamos para determinar la condición de cierre. En términos formales,

$$P(0{,}25 \leq X \leq 0{,}50) = \int_{0{,}25}^{0{,}50} 2x\,dx = 2\frac{x^2}{2}\Big|_{0{,}25}^{0{,}50} = 0{,}50^2 - 0{,}25^2 = 0{,}1875$$

La probabilidad de que la variable continua se encuentre entre 0,25 y 0,50 es de 0,1875.

Ahora bien, se podría haber arribado al mismo resultado si hubiésemos empleado el concepto de Función de Distribución en una variable continua.

$$F_X(x) = P(X \leq x) = \int_a^x f_X(s)\, ds$$

En nuestro ejemplo, podemos obtener una expresión que represente la probabilidad acumulada a un número determinado. Es decir,

$$F_X(x) = P(X \leq x) = \int_0^x 2s\, ds = x^2$$

Es decir que la función de distribución para nuestro ejemplo es x^2. Veamos si lo aplicamos al cálculo de,

$$P(0{,}25 \leq X \leq 0{,}50) = P(X \leq 0{,}50) - P(X \leq 0{,}25) = 0{,}50^2 - 0{,}25^2 = 0{,}1875$$

Esperanza Matemática

En el caso continuo se calcula con una integral definida entre el límite inferior y superior que asume la variable.

En términos formales,

$$E[X] = \int_a^b x f_X(x)\, dx$$

En nuestro ejemplo,

$$E[X] = \int_0^1 x * 2x\, dx = \int_0^1 2x^2 dx = 2\frac{x^3}{3}\bigg|\begin{matrix}1\\0\end{matrix} = \frac{2}{3}(1^3 - 0^2) = \frac{2}{3}$$

Es decir que el valor esperado en este ejemplo es de $\frac{2}{3}$.

Varianza

En el caso continuo se calcula con una integral definida entre el límite inferior y superior que asume la variable. En términos formales,

$$Var[X] = E[(X - E[X])^2] = \int_a^b (x - E[X])^2 f_X(x)\, dx$$

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

$$Var[X] = \int_a^b x^2 f_X(x)\, dx - \left(\int_a^b x f_X(x)\, dx\right)^2$$

El segundo término lo tenemos, dado que es la Esperanza Matemática.

El primer término se puede calcular con la tabla de probabilidad agregando una columna más.

Tarea: Calcular Varianza, Desvío Estándar y Coeficiente de Variación.

Modo

Esta medida se asocia con el valor más probable. En el caso continuo se asocia al máximo relativo de una función.

Por lo tanto, para obtener el modo de una función de densidad es análogo a encontrar el máximo de dicha función, tal como han visto en Análisis Matemático.

Condición de Primer Orden (CPO)

$$f'_X(x) = 0 \Rightarrow salen\ puntos\ críticos\ x^*$$

Condición de Segundo Orden (CSO)

$$f'_X(x^*) < 0 \Rightarrow \ x^*\ es\ un\ máximo\ relativo\ de\ la\ función\ f_X(x)$$

En nuestro ejemplo $f_X(x) = 2x$ es una función lineal y por lo tanto no tiene máximo relativo, y en esta situación no disponemos de Modo.

Mediana

Para determinarla se emplea la Función de Distribución para obtener aquel valor de la variable que acumula 50% de probabilidad.

$$F_X(Me) = P(X \leq Me) = 0{,}50$$

$$F_X(x) = x^2$$

Entonces,

$$Me(X)^2 = 0{,}50 \Rightarrow \sqrt{0{,}50} \cong 0{,}7071$$

Procedimiento similar si quisiéramos por ejemplo el Percentil 25,

$$P_{25}^2 = 0{,}25 \Rightarrow \sqrt{0{,}25} = 0{,}50$$

Tarea: Obtener el Rango Intercuartilar de esta variable continua

Propiedades de la Media Aritmética y Varianza

Esperanza Matemática

1. El valor esperado de una constante es la constante misma.

$$Si\ X = k \Rightarrow E[X] = k$$

2. El valor esperado de una constante multiplicada por una variable es la constante multiplicada por el valor esperado de la variable.

$$Si\ Y = kX \Rightarrow E[Y] = kE[X]$$

3. El valor esperado de una transformación afín es

$$Si\ Y = a + bX \Rightarrow E[Y] = a + bE[X]$$

4. El valor esperado de una suma de variables es la suma del valor esperado de las variables

$$\text{Si}\ W = X + Y \Rightarrow E[W] = E[X] + E[Y]$$

Este resultado es extensible a más de 2 variables aleatorias.

Varianza

1. La varianza de una constante es cero

Si $X = k \Rightarrow Var(X) = 0$

2. La varianza de una constante por una variable aleatoria es la constante al cuadrado por la varianza de la variable aleatoria

Si $Y = kX \Rightarrow Var(Y) = k^2 Var(X)$

3. La varianza de la transformación afín

Si $Y = a + bX \Rightarrow Var(Y) = b^2 Var(X)$

4. La varianza de una suma de variables aleatorias

Si $W = X + Y \Rightarrow Var(W) = Var(X) + Var(Y) \pm 2\,\gamma(X, Y)$

Varianza:

4. La varianza de una suma de variables aleatorias

Si $W = X + Y \Rightarrow Var(W) = Var(X) + Var(Y) \pm 2\,\gamma(X,Y)$

Si $X \perp Y \;\Rightarrow \gamma(X,Y) = 0 \;\; (\Rightarrow \rho(X;Y) = 0) \Rightarrow Var(W) = Var(X) + Var(Y)$

$$\gamma(X,Y) = E\{[X - E(X)][Y - E(Y)]\} = E[XY] - E[X]E[Y]$$

$$\rho(X;Y) = \frac{\gamma(X,Y)}{\sigma_X * \sigma_Y}$$

Independencia ⇒ Incorrelación

Varianza

4. La varianza de una suma de variables aleatorias

Si $W = aX + bY \Rightarrow Var(W) = a^2Var(X) + b^2Var(Y) \pm 2ab\,\gamma(X,Y)$

$$Var(W) = a^2Var(X) + b^2Var(Y) \pm 2\ a\ b * \rho(X;Y) * \sigma_X * \sigma_Y$$

$$\rho(X;Y) = \frac{\gamma(X,Y)}{\sigma_X * \sigma_Y}$$

Distribuciones Discretas

Bernoulli

Esta distribución de probabilidad para variable discreta cuantifica la ocurrencia de un determinado atributo. Por ejemplo, supongamos que queremos determinar si una persona se ha contagiado con la enfermedad COVID-19 entonces nuestra variable aleatoria queda definida de la forma siguiente,

$$X = Persona\ infectada\ con\ COVID-19$$

La variable se puede generalizar de la forma siguiente,

$$X: Elemento\ con\ un\ determinado\ atributo$$

En notación compacta $X \sim B(1;p)$. Se lee X se distribuye como una Bernoulli con probabilidad de éxito igual a p. Los posibles valores que puede asumir son dos,

$$\Omega(X) = \{0,1\}$$

Si la variable asume el valor $X = 0$ significa que la persona no presenta el atributo, que es que fue infectada por tal enfermedad. En caso contrario, si la variable asume el valor $X = 1$ significa que la persona presenta el atributo, es decir que fue infectada por tal enfermedad.

A partir de la definición anterior y entendiendo el recorrido de la Variable Bernoulli se entiende que se trata de la cuantificación de un experimento aleatorio, en el cual se observa una persona y se determina si presenta o no el atributo deseado.

Ahora bien supongamos que la probabilidad de infectarse con dicha enfermedad es un valor constante e igual a 30%, esto significa que $P[X=1]=0{,}30$ y por complemento la probabilidad de no infectarse es 70%, en términos formales $P[X=0]=0{,}70$.

Lo anterior puede ser descripto mediante una tabla de probabilidades de la forma siguiente,

$\boldsymbol{X}$	$\boldsymbol{P[X=x]}$	$\boldsymbol{F[X=x]}$	$\boldsymbol{S[X=x]}$	$xP[X=x]$
0	$0{,}70$	$0{,}70$	$0{,}30$	0
1	$0{,}30$	1	0	$0{,}30$
	1			$0{,}30$

De la tabla anterior se puede expresar la función de probabilidad de la forma siguiente,

$$P[X = x] = p^x q^{1-x} \qquad (x = 0,1)$$

La Función de Distribución se forma de la forma siguiente,

$$F[X = x] = \sum_{j=0}^{1} p^j q^{1-j}$$

Asimismo se observa que la Esperanza Matemática coincide con la probabilidad de ocurrencia del atributo bajo análisis. Formalmente,

$$E[X] = \sum_{x=0}^{1} x p^x q^{1-x} = 0 * (1-p) + 1 * p = p$$

En nuestro ejemplo,

$$E[X] = p = 0,30$$

La Varianza de una Variable Bernoulli se determina de la forma siguiente,

$$Var[X] = \sum_{x=0}^{1} (x - E[X])^2 p^x q^{n-x} = pq$$

El interés de esta variable aleatoria discreta radica en la generalización de la repetición de n ensayos del mismo experimento, la cual se conoce como Variable Binomial.

Para próxima clase pensar otras aplicaciones de ejemplos de la vida real en la cual se podría aplicar esta distribución de probabilidad.

Binomial

Esta distribución cuantifica la ocurrencia de un determinado atributo en la repetición de un experimento una cierta cantidad de veces en condiciones de independencia. Por ejemplo, supongamos que queremos determinar si dos personas se han contagiado con la enfermedad COVID-19 entonces nuestra variable aleatoria queda definida de la forma siguiente,

$$X = Cantidad\ de\ Personas\ infectadas\ con\ COVID - 19$$

Los posibles valores que puede asumir esta variable (con 2 observaciones) son,

$$\Omega(X) = \{0,1,2\}$$

Si la variable asume el valor $X = 0$ significa que ninguna de las dos personas se han contagiado con tal enfermedad. Si la variable asume el valor $X = 1$ significa que al menos una de las dos personas se han contagiado con tal enfermedad. Puede ser que se haya contagiado la primera o la segunda, pero no los dos en forma conjunta. Por último, si el valor $X = 2$ significa que las dos personas observadas se han contagiado.

A partir de la definición anterior y entendiendo el recorrido de esta Variable Binomial se entiende que se trata de la cuantificación de un experimento aleatorio, en el cual se observan dos personas y se determina si presentan o no el atributo deseado.

En las siguientes dos tablas se exhiben en forma individual la distribución de probabilidad de la variable Bernoulli que representa la afectación de cada persona,

X_1	$P[X_1 = x]$
0	0,70
1	0,30
	1

X_2	$P[X_2 = x]$
0	0,70
1	0,30
	1

$X = X_1 + X_2$	$P[X = x]$
0	$0,70 * 0,70 = 0,49$
1	$0,70 * 0,30 + 0,30 * 0,70 = 0,42$
2	$0,30 * 0,30 = 0,09$

Se dice que X es una Binomial

$$X \sim B(2; 0,30)$$

El procedimiento anterior se conoce como ***"Convolución"*** de variables aleatorias discretas.

El objetivo fue determinar una nueva distribución de probabilidad mediante la suma de variables bernoulli independientes.

¿Cómo podemos hacer cuando en vez de observar 2 personas queremos observar una muestra más grande?

Supongamos que contamos con una Binomial con los siguientes valores de parámetros,

$$\boldsymbol{X \sim B(5; 0,30)}$$

Y se desea calcular la probabilidad de encontrar 3 personas **infectadas** de las 5 observadas, bajo el supuesto de que la probabilidad de encontrar una persona infectada es constante e igual para cada integrante de la muestra y que si una persona esta infectada no influye que otra se infecte (Aclaración: Es una suposición!!!!!). Es decir, me solicitan calcular,

$$\boldsymbol{P[X = 3]} = ?$$

Es decir, me solicitan calcular,

$$\boldsymbol{P}[\boldsymbol{X} = \boldsymbol{3}] = ?$$

¿Qué primer acercamiento a la determinación del valor de probabilidad podemos hacer sin emplear convolución?

Podemos pensar del siguiente modo,

$$\boldsymbol{n} = \boldsymbol{5} \quad \begin{array}{l} \nearrow \quad \boldsymbol{x} = \boldsymbol{3} \\ \searrow \quad \boldsymbol{n} - \boldsymbol{x} = \boldsymbol{2} \end{array}$$

Es decir, de las 5 personas existen 3 que poseen el atributo deseado y el resto no lo posee y mediante probabilidades se puede formalizar de la forma siguiente,

$$0{,}30 * 0{,}30 * 0{,}30 * 0{,}70 * 0{,}70 = 0{,}30^3 0{,}70^2$$

¿Qué le falta a la anterior expresión?, ¿Qué considera implícitamente?

La siguiente expresión está considerando únicamente un ordenamiento y es que las tres primeras personas observadas estén infectadas y las últimas dos no lo estén,

$$0{,}30 * 0{,}30 * 0{,}30 * 0{,}70 * 0{,}70 = 0{,}30^3 0{,}70^2$$

Ahora bien con esta información, ***¿Cuántos ordenamientos existen?***

La probabilidad anterior debe ir multiplicada por un número combinatorio que indique las combinaciones de dado una cantidad de ensayos (o repeticiones) encontrar una cantidad de elementos con atributo. Por ejemplo, supongamos que contamos con 3 profesoras y queremos determinar la cantidad de combinaciones seleccionando 2 de ellas. Al pensar por un momento determinamos que existen 3 combinaciones (Pro1-Pro2, Pro1-Pro3, Pro2-Pro3). Formalmente con un número combinatorio,

$$\binom{3}{2} = 3$$

¿Cómo se calcula un número combinatorio? (genéricamente),

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

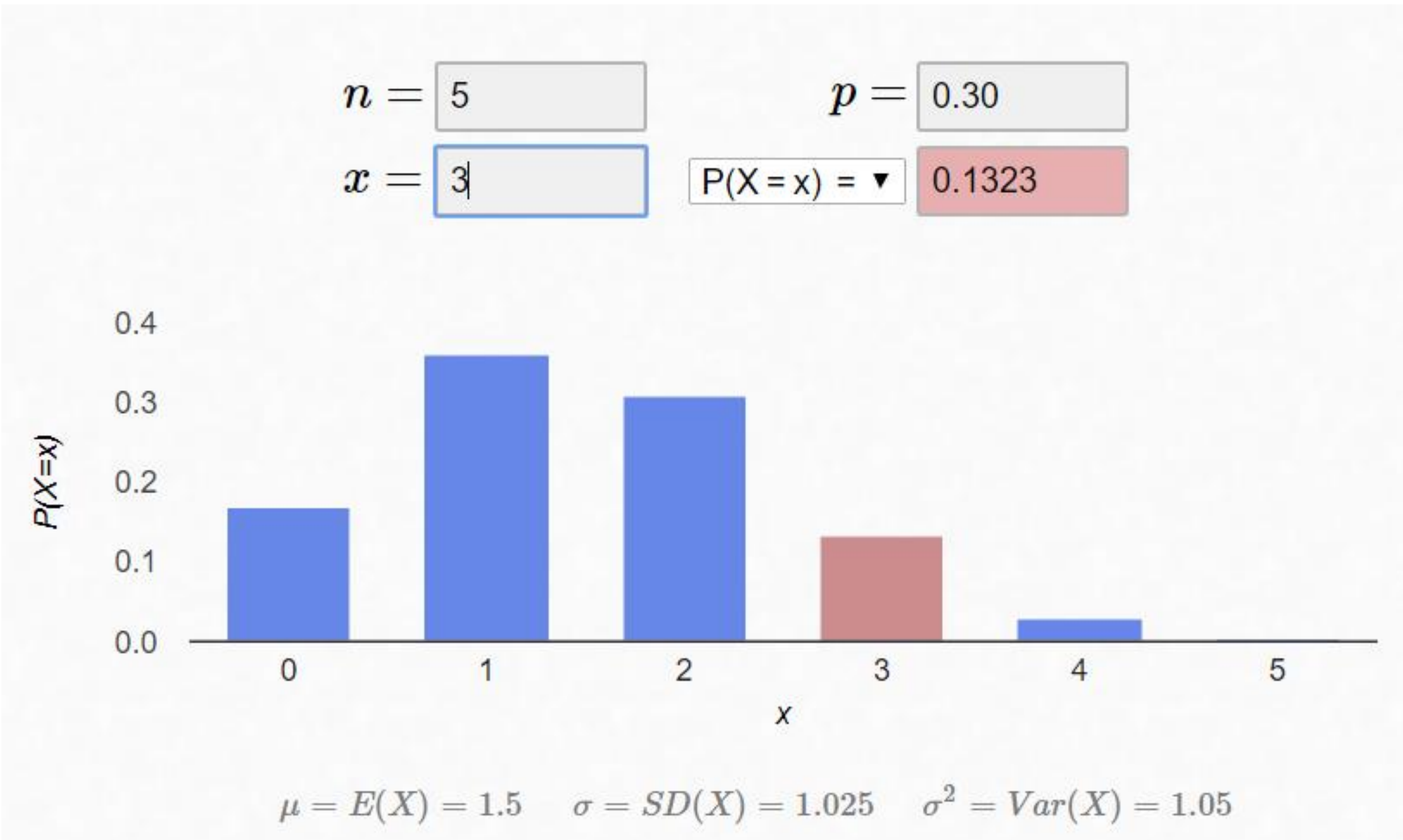
En nuestro ejemplo, ***¿Qué número combinatorio debemos incluir en el cálculo?***

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5*4*3!}{3!\,2!} = \frac{5*4}{2*1} = 10$$

Es decir, que el número que debe anteceder a $0{,}30^3 0{,}70^2$ es 10. Formalmente,

$$P[X=3] = \binom{5}{3} 0{,}30^3 0{,}70^2 = 10 * 0{,}30^3 * 0{,}70^2 = 0{,}1323$$

La probabilidad de que en una muestra de 5 personas se encuentren 3 con un determinado atributo es de 0,1323. El gráfico (bastones) siguiente es la distribución de probabilidad que se forma para todo el recorrido de la variable,


$n =$ 5
$p =$ 0.30
$x =$ 3
P(X = x) =
0.1323
0.4
0.3
0.2
0.1
0.0
P(X=x)
0
1
2
3
4
5
x
$\mu = E(X) = 1.5$ $\sigma = SD(X) = 1.025$ $\sigma^2 = Var(X) = 1.05$

TRIÁNGULO DE PASCAL	POTENCIA DE UNA SUMA
1	$(a+b)^0 = 1$
1 1	$(a+b)^1 = 1a + 1b$
1 2 1	$(a+b)^2 = 1a^2 + 2ab + 1b^2$
1 3 3 1	$(a+b)^3 = 1a^3 + 3a^2b + 3ab^2 + 1b^3$
1 4 6 4 1	$(a+b)^4 = 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4$
1 5 10 10 5 1	$(a+b)^5 = 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n$$

https://matematicascercanas.com/2018/10/22/triangulo-pascal-binomio-newton/

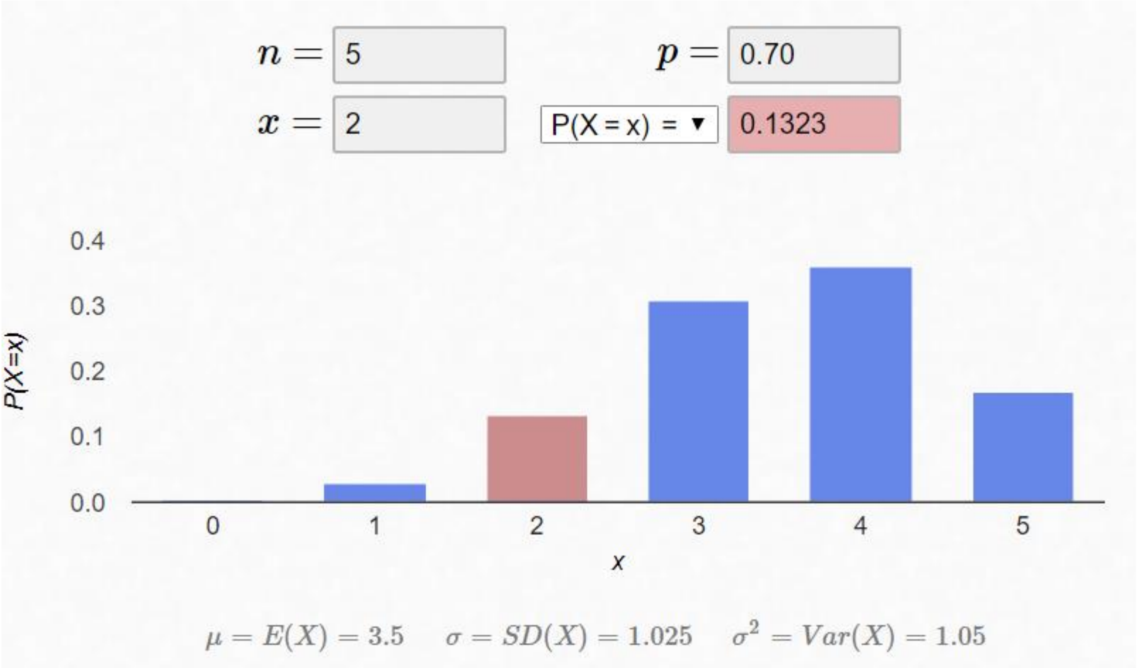
Ahora bien, supongamos que contamos con una Binomial con los siguientes valores de parámetros,

$$\boldsymbol{X \sim B(5; 0,70)}$$

Y se desea calcular la probabilidad de encontrar 2 personas **no infectadas** de las 5 observadas, es decir

$$\boldsymbol{P[X = 2] = ?}$$

¿Cuánto da esta probabilidad?



En notación compacta nuestra variable se escribe de la forma siguiente,

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

$$E[X] = np$$

Veamos en los ejemplos,

Ejemplo 1

$$X \sim B(5; 0{,}30) \Rightarrow E[X] = 5 * 0{,}30 = 1{,}5\ personas\ infectadas$$

Ejemplo 2

$$X \sim B(5; 0{,}70) \Rightarrow E[X] = 5 * 0{,}70 = 3{,}5\ personas\ \boldsymbol{no}\ infectadas$$

En notación compacta nuestra variable se escribe de la forma siguiente,

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

$$Var[X] = npq$$

Veamos en los ejemplos,

Ejemplo 1

$$X \sim B(5;0{,}30) \Rightarrow Var[X] = 5 * 0{,}30 * 0{,}70 = 1{,}05$$

Ejemplo 2

$$X \sim B(5;0{,}70) \Rightarrow Var[X] = 5 * 0{,}30 * 0{,}70 = 1{,}05$$

Sea un fenómeno binomial repetible no determinístico que admite dos resultados posibles, E (“éxito”) y $\bar{E}$ (“fracaso”) y sean $P(E) = p$ y $P(\bar{E}) = q = 1 - p$ sus respectivas probabilidades de ocurrencia. La variable aleatoria

$$X\text{: } Cantidad\ de\ elementos\ con\ un\ determinado\ atributo$$

En notación compacta,

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

Función de Probabilidad Puntual

$$P[X = x] = \binom{n}{x} p^x q^{n-x} \qquad (x = 0,1,2, \dots, n)$$

Función de Distribución (Acumulada)

$$F[X = x] = \sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j}$$

Función de Supervivencia

$$S[X = x] = 1 - \underbrace{F[X = x]}_{P[X \le x]} = 1 - \sum_{j=0}^{x} \binom{n}{j} p^j q^{n-j} = P[X \ge x + 1]$$

Esperanza Matemática

$$E[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = np$$

Varianza

$$Var[X] = \sum_{x=0}^{n} (x - E[X])^2 \binom{n}{x} p^x q^{n-x} = npq$$

Modo

Es aquel valor de la variable más probable. Se corresponde con la parte entera del producto entre la cantidad de valores que puede asumir la variable y la probabilidad de ocurrencia del atributo. Es decir,

$$x_{Mo} = \lfloor (n+1)p \rfloor$$

Mediana

$$F[X = x_{Me}] = \sum_{j=0}^{x_{Me}} \binom{n}{j} p^j q^{n-j} = 0{,}50$$

Coeficiente de Asimetría

$$AS[X] = \frac{1-2p}{\sqrt{npq}}$$

Coeficiente de Curtosis

$$K[X] = \frac{1-6pq}{npq}$$

Fórmula de Panjer

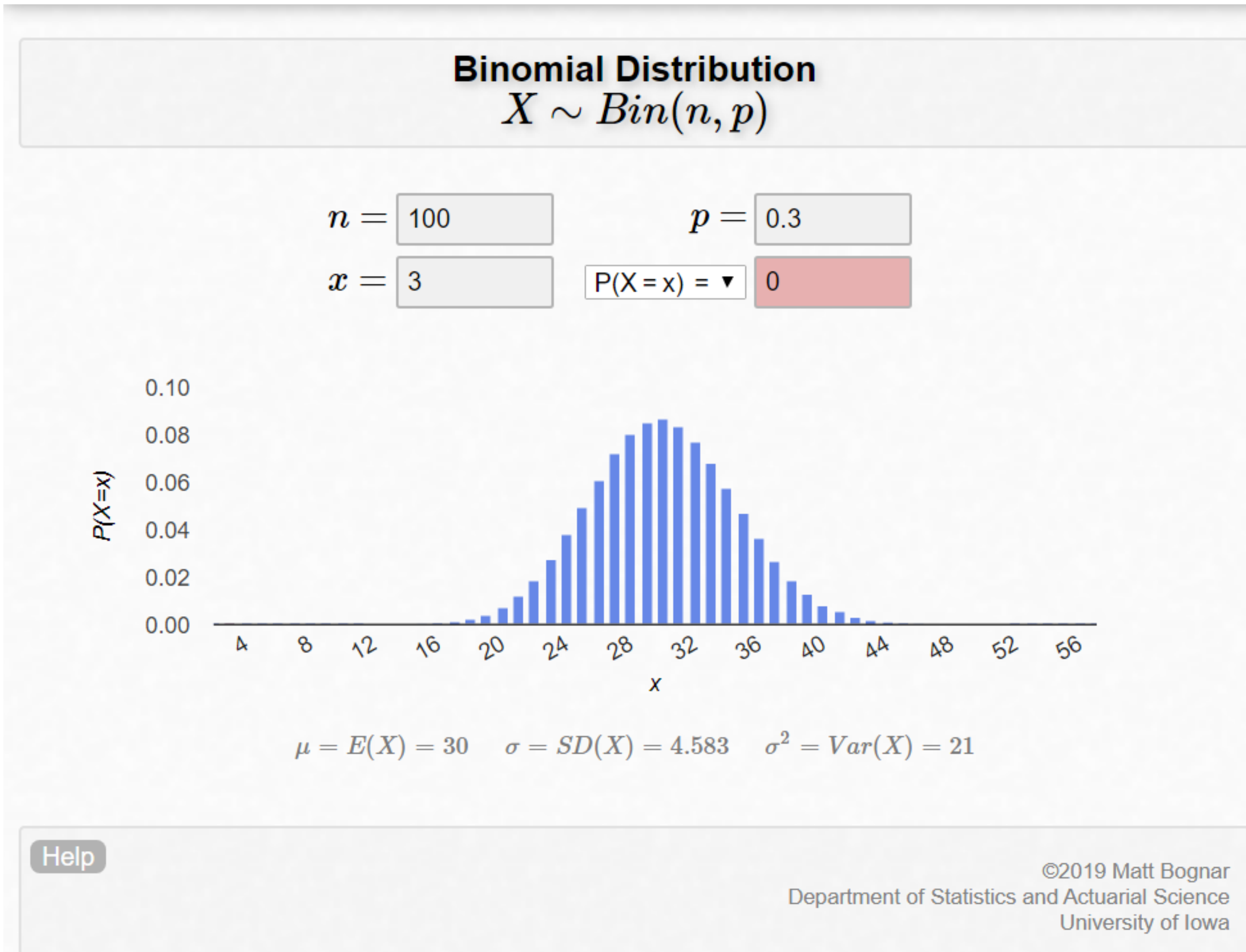
$$\frac{P[X=k]}{P[X=k-1]} = -\frac{p}{q} + \left(\frac{n+1}{k}\right)\frac{p}{q} = \frac{p}{q}\left(\frac{n+1}{k} - 1\right) = \frac{p}{q}\left(\frac{n+1-k}{k}\right)$$

https://stattrek.com/online-calculator/binomial.aspx

- Enter a value in each of the first three text boxes (the unshaded boxes).
- Click the **Calculate** button.
- The Calculator will compute Binomial and Cumulative Probabilities.

Probability of success on a single trial	
Number of trials	
Number of successes (x)	
Binomial probability: $P(X = x)$	
Cumulative probability: $P(X < x)$	
Cumulative probability: $P(X \leq x)$	
Cumulative probability: $P(X > x)$	
Cumulative probability: $P(X \geq x)$	

https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html



Poisson

Ahora bien supongamos que la cantidad de personas infectadas con COVID-19 que se infectan en promedio por semana es de 2.

Podría ser interesante plantearse ***¿Cuál es la probabilidad de que se infecten exactamente 7 personas en 2 semanas?***

A diferencia de las variables anteriores en la pregunta se considera la idea un espacio temporal (2 semanas). *Es decir, lo que interesa es determinar la cantidad de personas infectadas en 2 semanas. Si en una semana tenemos en promedio 2 infectados, entonces en 2 semanas habrá en promedio 4 infectados.*

Esta Variable se distribuye como Poisson. En términos formales,

$$X: Cantidad\ de\ personas\ infectadas\ con\ COVID-19\ en\ 2\ semanas$$

En notación compacta, $X \sim P(\lambda)$ donde λ es el único parámetro que representa el promedio de ocurrencia en el límite o espacio temporal considerado.

En nuestro ejemplo $\lambda = 4 = 2\ infectados/semana * 2\ semanas$

La pregunta a responder es: ***¿Cuál es la probabilidad de que se infecten exactamente 7 personas en 2 semanas?***

$$P[X = 7] = ?$$

La distribución de probabilidad se obtiene mediante la resolución de Ecuaciones Diferenciales (tiempo) y Ecuaciones en Diferencias (variables). Para un mayor detalle ver en el libro de Canavos, en el apéndice al final del capítulo 4. La función de probabilidad es la siguiente,

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

En nuestro ejemplo,

$$P[X = 7] = \frac{e^{-4}4^{7}}{7!} = 0{,}05954$$

La probabilidad de que se infecten exactamente 7 personas en el lapso de 2 semanas es de 0,05954.

Si la pregunta fuera: ***¿Cuál es la probabilidad de que se infecten exactamente 10 personas en 3 semanas?***

Lo que deberíamos realizar es adaptar el valor del parámetro lambda mediante regla de 3 simples

$$\lambda = 2\,\frac{infectados}{semana} * 3\ semanas = 6\ infectados$$

Ahora podemos responder la pregunta,

$$P[X = 10] = \frac{e^{-6}6^{10}}{10!} = 0{,}0413$$

La probabilidad de que se infecten exactamente 10 personas en el lapso de 3 semanas es de 0,0413.

UNIVERSIDAD AUSTRAL | INGENIERÍA

Si la pregunta fuera: ***¿Cuál es la probabilidad de que se infecten como mínimo 10 personas en 2 semanas?*** **Ahora el valor del parámetro lambda es 4**

$$\lambda = 2\frac{infectados}{semana} * 2\, semanas = 4\, infectados$$

Ahora podemos responder la pregunta $P[X \geq 10] = 1 - F[X = 9] = 0{,}00813$

La probabilidad de que se infecten exactamente 10 personas en el lapso de 2 semanas es de 0,0413.

La Variable de Poisson está definida para los enteros positivos.

Es decir $\boldsymbol{\Omega(X) = \{0, 1, \ldots, \infty\}}$

- Enter a value in BOTH of the first two text boxes.
- Click the **Calculate** button.
- The Calculator will compute the Poisson and Cumulative Probabilities.

Poisson random variable (x)	10
Average rate of success	4
Poisson Probability: P(X = 10)	0.00529
Cumulative Probability: P(X < 10)	0.99187
Cumulative Probability: P(X ≤ 10)	0.99716
Cumulative Probability: P(X > 10)	0.00284
Cumulative Probability: P(X ≥ 10)	0.00813

https://stattrek.com/online-calculator/poisson.aspx

Si la pregunta fuera: ***¿Cuál es la probabilidad de que se infecten como máximo 10 personas en 2 semanas?*** **El valor del parámetro lambda es 4**

$$P[X \leq 10] = F[X = 10] = 0{,}99716$$

¿Cuál es la probabilidad de que se infecten entre 7 y 10 personas (inclusive) en 2 semanas? **El valor del parámetro lambda es 4**

$$P[7 \leq X \leq 10] = F[X = 10] - F[X = 6] = 0{,}99716 - 0{,}88933 = 0{,}10783$$

Es decir la probabilidad que en un lapso de 2 semanas se infecten entre 7 y 10 personas es de 0,10783 respectivamente.

Poisson random variable (x)	6
Average rate of success	4
Poisson Probability: P(X = 6)	0.10420
Cumulative Probability: P(X < 6)	0.78513
Cumulative Probability: P(X ≤ 6)	0.88933
Cumulative Probability: P(X > 6)	0.11067
Cumulative Probability: P(X ≥ 6)	0.21487

La variable aleatoria cuantifica,

$$X: Cantidad\ de\ elementos\ con\ un\ determinado\ atributo\ en\ un\ espacio\ temporal$$

En notación compacta,

$$X \sim P(\lambda) \qquad \Omega(X) = \{0,1,\dots,\infty\}$$

Función de Probabilidad Puntual

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

Función de Distribución Acumulada

$$F[X = x] = \sum_{j=0}^{x} \frac{e^{-\lambda}\lambda^{j}}{j!}$$

Función de Supervivencia

$$S[X=x]=\sum_{j=x+1}^{\infty}\frac{e^{-\lambda}\lambda^{j}}{j!}$$

Esperanza Matemática

$$E[X]=\sum_{x=0}^{\infty}x\frac{e^{-\lambda}\lambda^{x}}{x!}=\lambda$$

Varianza

$$Var[X]=\sum_{x=0}^{n}(x-E[X])^{2}\frac{e^{-\lambda}\lambda^{x}}{x!}=\lambda$$

Fórmula de Panjer

$$P[X=k]=\frac{\lambda}{k}*P[X=k-1]\Rightarrow\frac{P[X=k]}{P[X=k-1]}=\frac{\lambda}{k}$$

Distribuciones Continuas

Normal

Sea X una variable aleatoria que se distribuye como una Normal con media igual a μ y Desvío Estándar igual a σ. En notación compacta,

$$X \sim N(\mu; \sigma)$$

La función de densidad que rige el comportamiento de esta variable es,

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

$$E[X] = \mu = Me = Mo$$

$$Var[X] = \sigma^2$$

Tal como se ha visto previamente si se desea calcular una probabilidad deberá integrarse la función de densidad, por ejemplo $X \sim N(10;2)$ y se desea saber la probabilidad de que la variable asuma un valor entre 8 y 12 respectivamente.
En términos formales,

$$P[8 \leq X \leq 12] = \int_{8}^{12} f_X(x)\,dx$$

$$P[8 \leq X \leq 12] = \int_{8}^{12} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$$

Para resolver dicha integral en la práctica se realiza una sustitución de variable, que se denomina “variable estandarizada” o “tipificada”,

$$Z = \frac{x-\mu}{\sigma} \sim N(0;1)$$

$$E[Z] = 0 \quad Var[Z] = 1$$

Es una transformación de la variable X, a cada valor se le resta la media y se lo divide por el desvío.

$$P[8 \le X \le 12] = \int_{8}^{12} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$Z = \frac{x-\mu}{\sigma} \Rightarrow dZ = \frac{dx}{\sigma} \Rightarrow dZ\sigma = dx$$

$$P[8 \le X \le 12] = \int_{\frac{8-10}{2}}^{\frac{12-10}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}Z^2} dZ\sigma = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} dZ = P[-1 \le Z \le 1]$$

El problema inicial de encontrar la probabilidad de que X se encuentre entre 8 y 12 es equivalente a que Z se encuentre entre -1 y 1 respectivamente, el cual es más fácil de resolver.

$$P[-1 \le Z \le 1] = F[Z=1] - F[Z=-1] = 0{,}841 - 0{,}159 = 0{,}6826$$

https://stattrek.com/online-calculator/normal.aspx

Función de Distribución Acumulada

$$P[X \leq x] = F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2} ds = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = F_Z(z)$$

Existe una tabla en la cual figura para cada valor de Z cual es el valor de probabilidad acumulada a dicho valor.

¿Cuánto da la probabilidad de que la variable X se encuentre entre 6 y 14?

$$P[6 \leq X \leq 14] = F[X = 14] - F[X = 6] = F\left[Z = \frac{14-10}{2}\right] - F\left[Z = \frac{6-10}{2}\right]$$

$$P[6 \leq X \leq 14] = F[Z = 2] - F[Z = -2] = 0{,}9545$$

¿Qué representa Z? Es la cantidad de desvíos respecto al valor esperado. Si Z = 1 significa que estamos posicionados a 1 desvío de X a la derecha del valor esperado.

¿Cuánto da la probabilidad de que la variable X se encuentre entre 6 y 14?

$$P[4 \leq X \leq 16] = F[X = 16] - F[X = 4] = F\left[Z = \frac{16-10}{2}\right] - F\left[Z = \frac{4-10}{2}\right]$$

$$P[4 \leq X \leq 16] = F[Z = 3] - F[Z = -3] = 0{,}9973$$

Ahora bien que pasaría *si* $X \sim N(15; 3)$, cuanto da la siguiente probabilidad,

$$P[12 \leq X \leq 18] = ?$$

$$P[9 \leq X \leq 15] = ?$$

$$P[6 \leq X \leq 21] = ?$$

Ahora bien que pasaría *si* $X \sim N(15;3)$, cuanto da la siguiente probabilidad,

$$P[12 \leq X \leq 18] = 0{,}6628$$

$$P[9 \leq X \leq 15] = 0{,}9545$$

$$P[6 \leq X \leq 21] = 0{,}9973$$

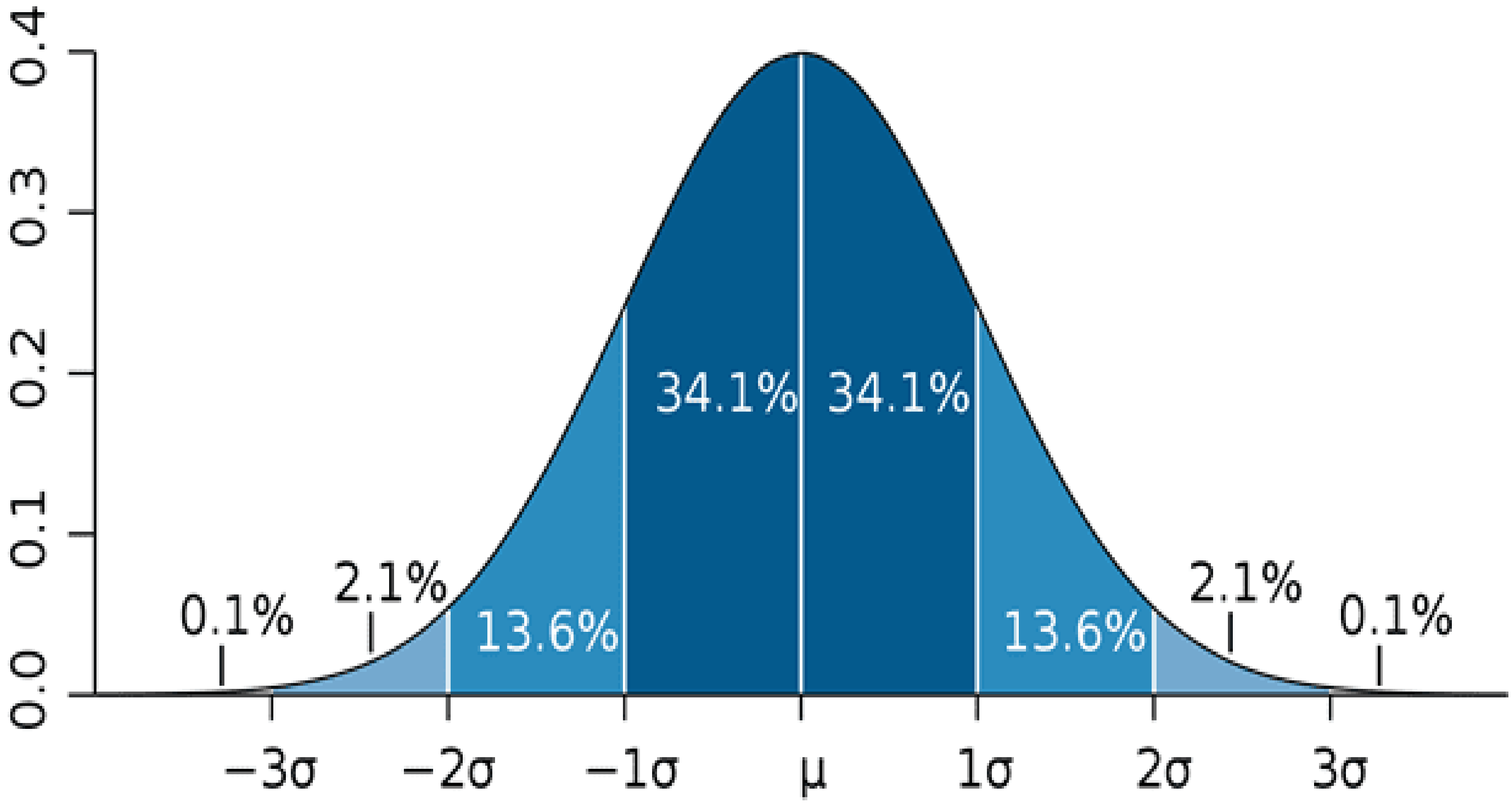
Es decir, las probabilidades son iguales que las calculadas con $X \sim N(10;2)$

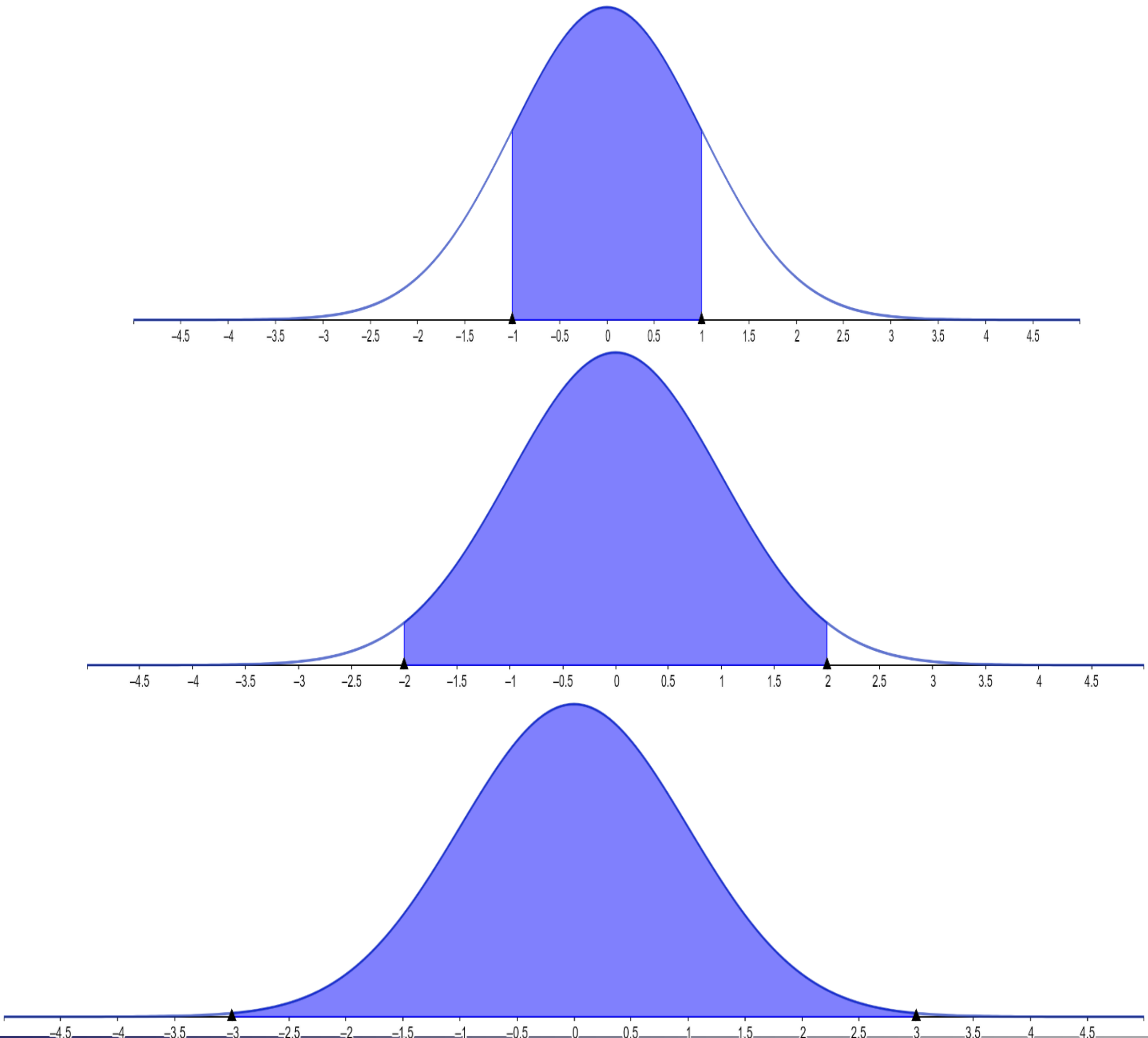
¿Y esto a qué se debe? A que se está calculando la probabilidad a la misma distancia del valor esperado. Formalmente,

$$P[\mu - \sigma \leq X \leq \mu + \sigma] = 0{,}6628 \rightarrow \pm 1\ Desvío$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 0{,}9545 \rightarrow \pm 2\ Desvíos$$

$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 0{,}9973 \rightarrow \pm 3\ Desvíos$$


0.4
0.3
0.2
0.1
0.0
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
0.1%
0.1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ



https://www.geogebra.org/classic#probability

¿Qué pasa si ahora queremos encontrar un percentil?

Supongamos que $X \sim N(10; 2)$, ¿Cuál es aquel valor de variable tal que se acumulada a izquierda el 25% de probabilidad? (Percentil 25)

Sabemos lo siguiente,

$$F[X = x_{25}] = 0{,}25$$

$$Z_{0,25} = \frac{x_{25} - 10}{2}$$

Ahora bien, por la estandarización de la variable se sabe cuanto vale $Z_{0,25}$ (vemos en el Excel y en cualquier tabla estadística)

$$Z_{0,25} = -0{,}674490 \Rightarrow x_{25} = -0{,}674490 * 2 + 10 = 8{,}65102$$

¿Qué pasa si desconocemos el valor de un parámetro?

Supongamos que $X \sim N(10;?)$, y qué la probabilidad de que sea menor a $8{,}65102$ es 25%.

$$F[X = x_{25}] = 0{,}25$$

$$-0{,}674490 = \frac{8{,}65102 - 10}{\sigma}$$

Ahora bien, por la estandarización de la variable se puede obtener el valor del desvío,

$$\sigma = \frac{8{,}65102 - 10}{-0{,}674490} = 2$$

Lognormal

$$X \sim N(\mu;\sigma)$$

$$Y = e^{X} \sim LN(\mu;\sigma) \Rightarrow ln\,Y = X \sim N(\mu;\sigma)$$

Función de Densidad

$$f_Y(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln y-\mu}{\sigma}\right)^2}$$

Esperanza Matemática

$$E[Y] = e^{\mu+\frac{\sigma^2}{2}}$$

Varianza

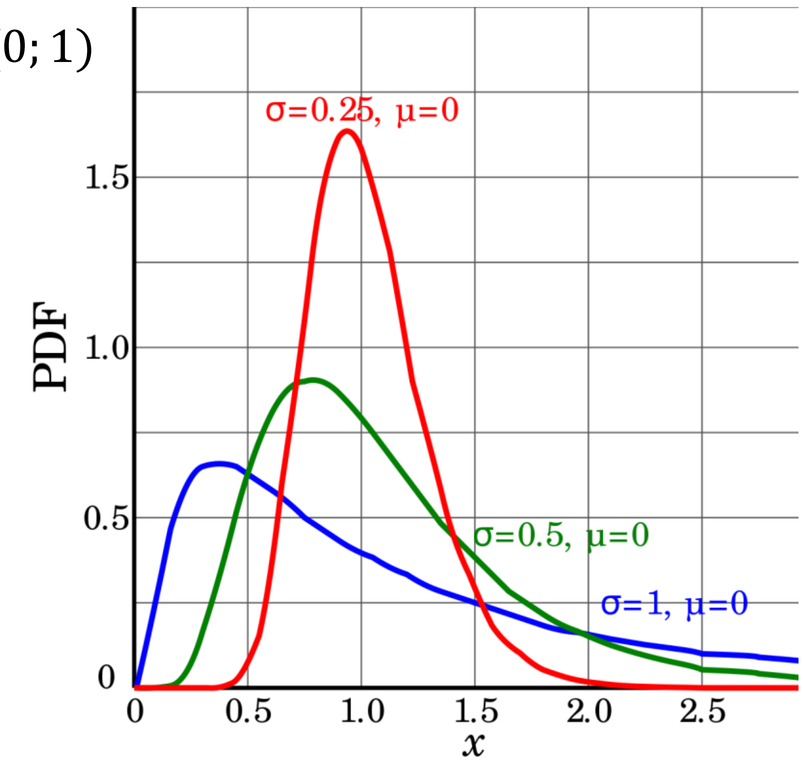
$$Var[Y] = e^{2\mu+2\sigma^2}\left(e^{\sigma^2} - 1\right)$$

Momento absoluto de orden "s"

$$E[Y^s] = e^{\mu s + \frac{(s\sigma)^2}{2}}$$

Variable Estándar o Tipificada

$$Z = \frac{\ln y - \mu}{\sigma} \sim N(0;1)$$



Ejemplo

Supongamos que el rendimiento diario del precio de la Acción YPF se distribuye lognormalmente con un rendimiento esperado de 5% diario y un desvío de 3% diario.

$$Y \sim LN(5;3) \Rightarrow ln\,Y = X \sim N(3;5)$$

Calcular la probabilidad de que el rendimiento de la acción se encuentre entre 10% y 30% respectivamente.

$$P[10 \leq Y \leq 30] = P[\ln 10 \leq X \leq \ln 30] = P\left[\frac{\ln 10 - 5}{3} \leq Z \leq \frac{\ln 30 - 5}{3}\right]$$

$$P\left[\frac{\ln 10 - 5}{3} \leq Z \leq \frac{\ln 30 - 5}{3}\right] = F[Z = -0{,}53] - F[Z = -0{,}90] = 0{,}297040 - 0{,}184289$$

$$P[10 \leq Y \leq 30] = 0{,}112751$$

Uniforme

Se dice que una variable se distribuye como una Uniforme Continua,

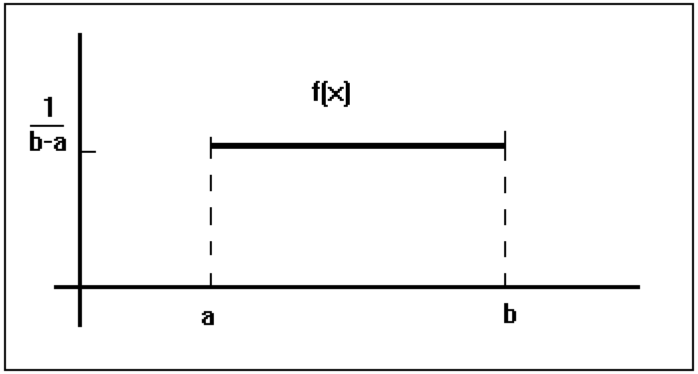
$$X \sim Un(\alpha;\beta) \quad \alpha < \beta$$

$\alpha: Límite\ Inferior$

$\beta: Límite\ Superior$

Donde la función de densidad es una constante y se determina de la forma siguiente,

$$f_X(x) = \frac{1}{\beta - \alpha}$$



Supongamos que $X \sim Un(100; 200)$, es decir,

$$f_X(x) = \frac{1}{\beta - \alpha} = \frac{1}{200 - 100} = \frac{1}{100}$$

Y se desea calcular la siguiente probabilidad,

$$P[120 \leq X \leq 140]$$

Para calcula esta probabilidad debemos integrar la función de densidad entre ambos valores de la variable.

$$P[120 \leq X \leq 140] = \int_{120}^{140} f_X(x)dx$$

$$P[120 \leq X \leq 140] = \int_{120}^{140} \frac{1}{100} dx = \frac{x}{100}\Bigg|_{120}^{140} = \frac{140 - 120}{100} = \frac{20}{100} = 0{,}20$$

Es decir, la probabilidad de que la variable uniforme X asuma un valor entre 120 y 140 es de 0,20.

Ahora bien, supongamos que deseamos calcular la siguiente probabilidad,

$$P[160 \leq X \leq 180]$$

¿Cuánto da esta probabilidad? (pensarla sin hacer cálculo)

¿Porqué?

¿Qué particularidad encuentra en el carácter uniforme de la variable?

La probabilidad se calcula tal como hemos visto antes,

$$P[160 \leq X \leq 180] = \int_{160}^{180} f_X(x)dx$$

$$\int_{160}^{180} \frac{1}{100}dx = \frac{x}{100}\Big|\begin{matrix}180\\160\end{matrix} = \frac{180-160}{100} = \frac{20}{100} = 0{,}20$$

La probabilidad es la misma y esto se debe porque a intervalos de igual amplitud la probabilidad es la misma, de ahí deviene su "uniformidad".

Ahora bien, ¿Existe alguna otra forma de calcular probabilidades acumuladas sin necesidad de integrar?

Si, tal como hemos visto… mediante la Función de Distribución. Por lo tanto, se procederá a calcular la expresión genérica de esta función para esta distribución de probabilidad.

Función de Distribución Acumulada

Esta función provee la probabilidad acumulada desde el mínimo valor que asume la variable hasta un valor específico. En términos formales,

$$F_X(x) = P[X \leq x] = \int_{\alpha}^{x} f_X(s)ds$$

$$F_X(x) = \int_{\alpha}^{x} \frac{1}{\beta - \alpha} ds = \frac{x - \alpha}{\beta - \alpha}$$

Ahora bien, probemos si es posible calcular la probabilidad entre 120 y 140 con esta expresión genérica, $X \sim Un(100; 200)$

$$P[120 \leq X \leq 140] = F[X = 140] - F[X = 120] = \frac{140 - 100}{100} - \frac{120 - 100}{100} = 0{,}20$$

El resultado es el mismo que integrando entre dichos valores.

Función de Supervivencia

De igual forma se puede obtener una expresión genérica para la Función de Supervivencia. En términos formales,

$$S_X(x) = P[X > x] = 1 - \frac{x-\alpha}{\beta-\alpha} = \frac{\beta - x}{\beta - \alpha}$$

Por ejemplo, $X \sim Un(100; 200)$

$$S_X(x) = \frac{\beta - x}{\beta - \alpha} = \frac{200 - x}{100}$$

$$P[X > 120] = \frac{200 - 120}{100} = \frac{80}{100} = 0{,}80$$

Es decir, la probabilidad de que la variable asuma un valor superior a 120 es 0,80.

Esperanza Matemática

$$E[X] = \int_{\alpha}^{\beta} x\, f_X(x)dx$$

$$E[X] = \int_{\alpha}^{\beta} x \frac{1}{\beta - \alpha} dx = \frac{1}{\beta - \alpha} \frac{x^2}{2} \Big|_{\alpha}^{\beta} = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)}$$

$$E[X] = \frac{(\beta + \alpha)}{2}$$

La Esperanza Matemática en el ejemplo $X \sim Un(100; 200)$

$$E[X] = \frac{(200 + 100)}{2} = 150$$

Varianza

$$Var[X] = \int_{\alpha}^{\beta} (x - E[X])^2 \, f_X(x) dx$$

Si se resuelve la integral se obtiene la siguiente expresión,

$$Var[X] = \frac{(\beta - \alpha)^2}{12}$$

Mediana

$$Me[X] = \frac{(\beta + \alpha)}{2}$$

Modo

No tiene modo esta variable aleatoria por ser una constante.

Percentiles

$$F_X(x) = \frac{x_k - \alpha}{\beta - \alpha} = k\% = \frac{k}{100}$$

$$x_k = \frac{k}{100}(\beta - \alpha) + \alpha$$

$$x_k = \alpha + \frac{k}{100}(\beta - \alpha)$$

Por ejemplo, $X \sim Un(100; 200)$

$$x_{75} = 100 + \frac{75}{100}(200 - 100) = 175$$

El percentil 75 resulta ser 175 para esta variable aleatoria.

Exponencial

Se dice que una variable se distribuye como una Exponencial,

$$X \sim Ex(\beta) \qquad \beta > 0$$

Donde β es el único parámetro de esta distribución y representa un parámetro de forma (leer Canavos sobre los distintos tipos de parámetros)

La función de densidad para esta variable se deriva de una Distribución Gamma (es un caso particular) cuando $\alpha = 1$. En términos formales,

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$$

Supongamos que $X \sim Ex(95)$, es decir,

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} = \frac{1}{95} e^{-\frac{x}{95}}$$

Supongamos que deseamos calcular la siguiente probabilidad,

$$P[100 \leq X \leq 200] = \int_{100}^{200} f_X(x)dx$$

Donde,

$$f_X(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}} = \frac{1}{95}e^{-\frac{x}{95}}$$

Por lo tanto,

$$P[100 \leq X \leq 200] = \int_{100}^{200} \frac{1}{95}e^{-\frac{x}{95}}dx$$

$$P[100 \leq X \leq 200] = \frac{1}{95}\int_{100}^{200} e^{-\frac{x}{95}}dx$$

Sustitución $u = \frac{x}{95}$

$$P[100 \leq X \leq 200] = \frac{1}{95}\int_{100}^{200} e^{-\frac{x}{95}} dx$$

Sustitución $u = \frac{x}{95} \Rightarrow du = \frac{dx}{95} \Rightarrow 95du = dx$

$$P[100 \leq X \leq 200] = \frac{1}{95}\int_{\frac{100}{95}}^{\frac{200}{95}} e^{-u} 95du = \int_{\frac{100}{95}}^{\frac{200}{95}} e^{-u} du$$

$$P[100 \leq X \leq 200] = -\left(e^{-\frac{200}{95}} - e^{-\frac{100}{95}}\right) = -(0{,}121813613 - 0{,}34901807)$$

$$P[100 \leq X \leq 200] = 0{,}227204457$$

La probabilidad de que la variable exponencial asuma un valor entre 100 y 200 es de 0,227204457.

Función de Distribución

De igual forma que con la Variable Uniforme se puede obtener una expresión genérica para calcular probabilidades acumuladas con esta distribución de probabilidad. En términos formales,

$$F_X(x) = P[X \leq x] = \int_0^x f_X(s)ds$$

$$F_X(x) = P[X \leq x] = \int_0^x \frac{1}{\beta} e^{-\frac{s}{\beta}} ds$$

Con la misma sustitución se obtiene la siguiente expresión,

$$F_X(x) = P[X \leq x] = 1 - e^{-\frac{x}{\beta}}$$

Supongamos que deseamos calcular la misma probabilidad que antes pero mediante la Función de Distribución,

$$P[100 \leq X \leq 200] = F[X = 200] - F[X = 100]$$

$$F[X = x] = 1 - e^{-\frac{x}{\beta}}$$

$$P[100 \leq X \leq 200] = 1 - e^{-\frac{200}{95}} - \left(1 - e^{-\frac{100}{95}}\right)$$

$$P[100 \leq X \leq 200] = e^{-\frac{100}{95}} - e^{-\frac{200}{95}}$$

$$P[100 \leq X \leq 200] = 0{,}227204457$$

La probabilidad de que la variable exponencial asuma un valor entre 100 y 200 es de 0,227204457.

Función de Supervivencia

La función de Supervivencia para esta distribución de probabilidad es,

$$P[X > x] = S_X(x) = e^{-\frac{x}{\beta}}$$

Por ejemplo, para nuestro ejemplo $X \sim Ex(95)$

$$P[X > 100] = e^{-\frac{100}{95}} = 0{,}34901807$$

$$P[X > 200] = e^{-\frac{200}{95}} = 0{,}121813613$$

$$P[X > 200/X > 100] = \frac{P[X > 200]}{P[X > 100]} = e^{-\frac{(200-100)}{95}} = e^{-\frac{100}{95}} = 0{,}34901807$$

<u>Propiedad:</u> Pérdida de Memoria

$$P[X > x_2/X > x_1] = \frac{P[X > x_2]}{P[X > x_1]} = \frac{e^{-\frac{x_2}{\beta}}}{e^{-\frac{x_1}{\beta}}} = e^{-\frac{x_2}{\beta}+\frac{x_1}{\beta}} = e^{-\frac{(x_2-x_1)}{\beta}} \qquad (x_2 > x_1)$$

UNIVERSIDAD AUSTRAL | INGENIERÍA

Esperanza Matemática

$$E[X] = \beta$$

Varianza

$$Var[X] = \beta^2$$

Percentiles

$$P_k = -\beta * l\,n\left(1 - \frac{k}{100}\right)$$

Mediana

$$Me[X] = -\beta * l\,n\left(1 - \frac{50}{100}\right) = -\beta * l\,n(0{,}50)$$

Aproximaciones

Las aproximaciones entre Variables Aleatorias Discretas y Continuas que veremos serán las siguientes,

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

$$\Downarrow$$

$$X \sim P(\lambda)$$

$$\Downarrow$$

$$X \sim N(\mu; \sigma)$$

De Binomial a Normal

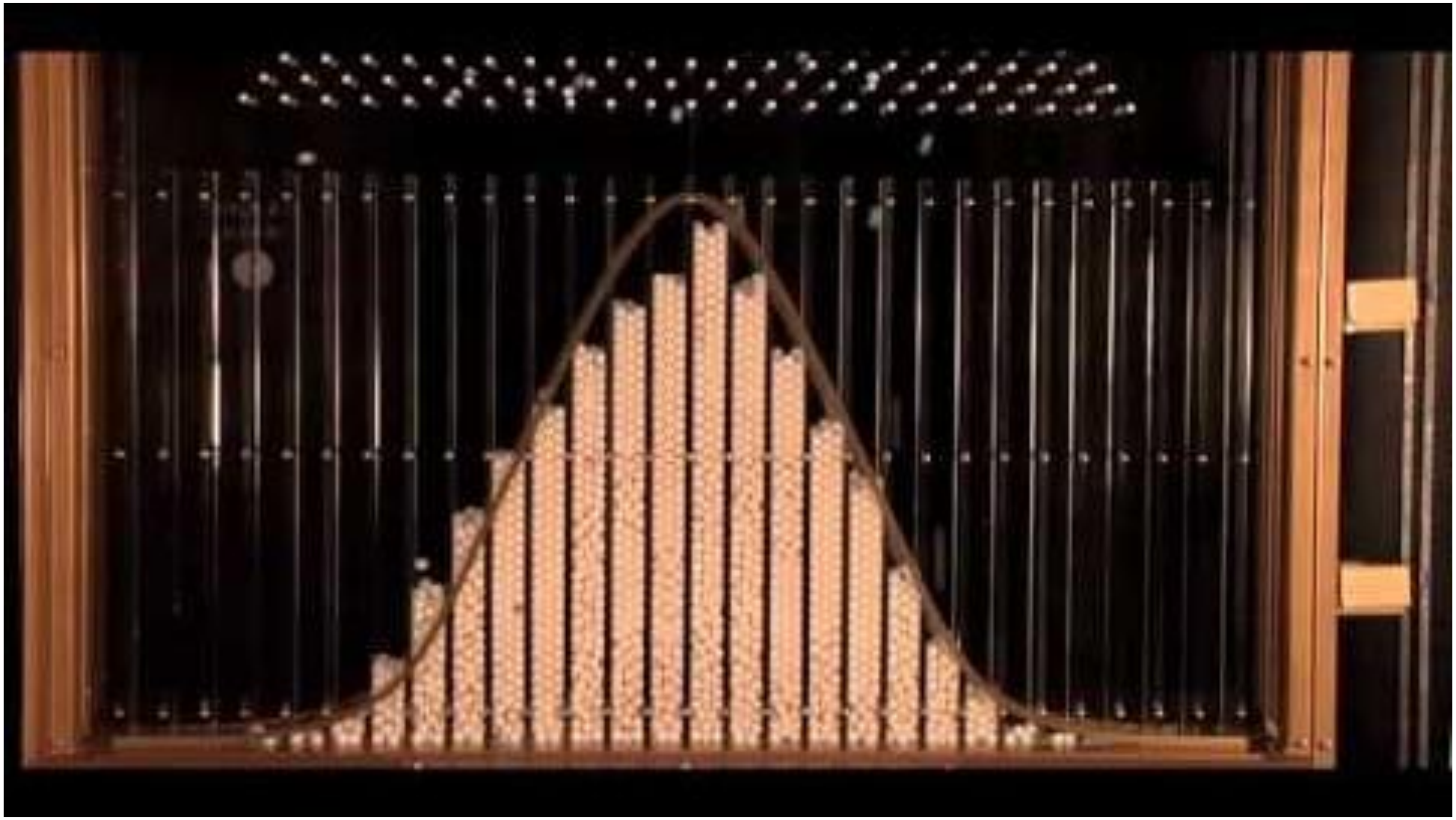
$$X \sim B(n;p) \cong N(np;\sqrt{npq})$$

Condición (deben verificarse en forma simultánea)

$$n \to +\infty \quad (n > 20)$$

$$p \to 0{,}50 \quad (0{,}05 < p < 0{,}95)$$

Corrección por continuidad (sumar ó restar 0,50 al valor que asume la variable)



https://www.youtube.com/watch?v=1DTRzPRfu6s

Suma de Variables Aleatorias Normales e Independientes

Caso General

$$S = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$$

Donde,

$$X_i \sim N(\mu_i; \sigma_i) \qquad c_i \in \mathbb{R}$$

Esperanza Matemática

$$E[S] = \mu_S = \sum_{i=1}^{n} c_i \mu_i$$

Varianza

$$Var[S] = \sigma_S^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2$$

Variable Estándar o Tipificada

$$Z = \frac{S - \sum_{i=1}^{n} c_i \mu_i}{\sqrt{\sum_{i=1}^{n} c_i^2 \sigma_i^2}} \sim N(0; 1)$$

Ejemplo

$X_1 \sim N(5;2)$

$X_2 \sim N(4;1)$

$X_3 \sim N(3;3)$

$$S = 3X_1 + 2X_2 + X_3 \quad \Rightarrow S \sim N(\mu_S;\sigma_S)$$

Donde

$$\mu_S = \sum_{i=1}^{n} c_i\mu_i \Rightarrow \mu_S = 3*5+2*4+1*3 = 26$$

$$\sigma_S^2 = \sum_{i=1}^{n} c_i^2\sigma_i^2 \Rightarrow \sigma_S^2 = 3^2*2^2+2^2*1^2+1^2*3^2 = 49 \Rightarrow \sigma_S = \sqrt{49} = 7$$

$S \sim N(26;7)$

$$S \sim N(26; 7)$$

Entonces la suma de variables normales e independientes da como resultado una nueva variable S que se distribuye normalmente con una media y desvío estándar que se forma a partir de aplicar las propiedades de Esperanza Matemática y Varianza.

Supongamos que se desea saber cual es la probabilidad de que S sea menor e igual a 26. Es decir,

$$P[S \leq 26] = ?$$

¿Qué sabemos de S?

¿Cuánto da esta probabilidad?

¿Por qué?

Supongamos que se desea saber cual es la probabilidad de que S sea menor e igual a 26. Esta probabilidad es igual a 0,50 porque S es normal y estamos interesados en determinar cual es la probabilidad acumulada a la mediana de la variable.

$$P[S \leq 26] = 0{,}50$$

Es decir que se cumple todo lo visto para una variable normal, entre esas cosas la relación entre el valor esperado y la cantidad de desvíos estándar. En términos formales,

$$P[\mu - \sigma \leq X \leq \mu + \sigma] = 0{,}6628 \;\; \rightarrow \;\; \pm 1 \; Desvío$$
$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 0{,}9545 \rightarrow \;\; \pm 2 \; Desvíos$$
$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 0{,}9973 \rightarrow \;\; \pm 3 \; Desvíos$$

En este ejemplo $S \sim N(26; 7)$ se cumple que,

$$P[19 \leq X \leq 33] = 0{,}6628 \;\; \rightarrow \;\; \pm 1 \; Desvío$$
$$P[12 \leq X \leq 40] = 0{,}9545 \rightarrow \;\; \pm 2 \; Desvíos$$
$$P[5 \leq X \leq 47] = 0{,}9973 \rightarrow \;\; \pm 3 \; Desvíos$$

Ahora bien supongamos que se desea saber cual es la probabilidad de que S se encuentre entre 22 y 35 inclusive. Es decir,

$$P[22 \leq S \leq 35] = ?$$

¿Qué podemos hacer para calcular esta probabilidad?

$$P[22 \leq S \leq 35] = F[S = 35] - F[S = 22]$$

$$P[22 \leq S \leq 35] = F\left[Z = \frac{35 - 26}{7}\right] - F\left[Z = \frac{22 - 26}{7}\right]$$

$$P[22 \leq S \leq 35] = F[Z \cong 1{,}29] - F[Z \cong -0{,}57]$$

$$P[22 \leq S \leq 35] = 0{,}9007286 - 0{,}28385458 = 0{,}61687402$$

Es decir la probabilidad de que S se encuentre entre 22 y 35 inclusive es de 0,61687402.

Caso I– $c_i = 1 \ \forall i$ – *Todas las constantes son iguales a 1*

$$S = X_1 + X_2 + \cdots + X_n$$

Donde,

$$X_i \sim N(\mu_i; \sigma_i)$$

Esperanza Matemática

$$E[S] = \mu_S = \sum_{i=1}^{n} \mu_i$$

Varianza

$$Var[S] = \sigma_S^2 = \sum_{i=1}^{n} \sigma_i^2$$

Variable Estándar o Tipificada

$$Z = \frac{S - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \sim N(0; 1)$$

Ejemplo

$X_1 \sim N(5; 2)$

$X_2 \sim N(4; 1)$

$X_3 \sim N(3; 3)$

$$S = X_1 + X_2 + X_3 \;\; \Rightarrow S \sim N(\mu_S; \sigma_S)$$

Donde

$$\mu_S = \sum_{i=1}^{n} \mu_i \Rightarrow \mu_S = 5 + 4 + 3 = 12$$

$$\sigma_S^2 = \sum_{i=1}^{n} \sigma_i^2 \Rightarrow \sigma_S^2 = 2^2 + 1^2 + 3^2 = 14 \Rightarrow \sigma_S = \sqrt{14} \cong 3{,}74$$

$S \sim N(12; 3{,}74)$

Ahora bien supongamos que se desea saber cual es la probabilidad de que S se encuentre entre 22 y 35 inclusive. Es decir,

$$P[22 \leq S \leq 35] = ?$$

¿Tiene sentido calcular esta probabilidad con esta nueva variable normal?

$$P[22 \leq S \leq 35] = F[S = 35] - F[S = 22]$$

$$P[22 \leq S \leq 35] = F\left[Z = \frac{35 - 12}{3{,}74}\right] - F\left[Z = \frac{22 - 12}{3{,}74}\right]$$

$$P[22 \leq S \leq 35] = F[Z \cong 6{,}15] - F[Z \cong 2{,}67]$$

$$P[22 \leq S \leq 35] = 1 - 0{,}99623684 = 0{,}00376316$$

Es decir la probabilidad de que S se encuentre entre 22 y 35 inclusive es de 0,00376316.

Caso II – *Todas las variables están identicamente distribuidas*

$$S = X_1 + X_2 + \cdots + X_n$$

Donde,

$$X \sim N(\mu; \sigma)$$

Esperanza Matemática

$$E[S] = \mu_S = \sum_{i=1}^{n} \mu_i = n\mu$$

Varianza

$$Var[S] = \sigma_S^2 = \sum_{i=1}^{n} \sigma_i^2 = n\sigma^2$$

Variable Estándar o Tipificada

$$Z = \frac{S - n\mu}{\sqrt{n\sigma^2}} = \frac{S - n\mu}{\sqrt{n}\sigma} \sim N(0; 1)$$

Ejemplo

$X_1 \sim N(5; 2)$

$X_2 \sim N(5; 2)$

$X_3 \sim N(5; 2)$

$$S = X_1 + X_2 + X_3 \quad \Rightarrow S \sim N(\mu_S; \sigma_S)$$

Donde

$$\mu_S = \sum_{i=1}^{n} \mu_i = n\mu \Rightarrow \mu_S = 3 * 5 = 15$$

$$\sigma_S^2 = \sum_{i=1}^{n} \sigma_i^2 = n\sigma^2 \Rightarrow \sigma_S^2 = 3 * 2^2 = 12 \Rightarrow \sigma_S = \sqrt{12} \cong 3{,}46$$

$S \sim N(15; 3{,}46)$

Ahora bien supongamos que se desea saber cual es la probabilidad de que S se encuentre entre 22 y 35 inclusive. Es decir,

$$P[22 \leq S \leq 35] = ?$$

¿Tiene sentido calcular esta probabilidad con esta nueva variable normal?

$$P[22 \leq S \leq 35] = F[S = 35] - F[S = 22]$$

$$P[22 \leq S \leq 35] = F\left[Z = \frac{35 - 15}{3{,}46}\right] - F\left[Z = \frac{22 - 15}{3{,}46}\right]$$

$$P[22 \leq S \leq 35] = F[Z \cong 5{,}77] - F[Z \cong 2{,}02]$$

$$P[22 \leq S \leq 35] = 1 - 0{,}97834593 = 0{,}02165407$$

Es decir la probabilidad de que S se encuentre entre 22 y 35 inclusive es de 0,02165407.

Caso III – *Diferencia de Variables Aleatorias*

$$S = X_1 - X_2$$

Donde,

$$X_i \sim N(\mu_i; \sigma_i)$$

Esperanza Matemática

$$E[S] = \mu_S = \mu_1 - \mu_2$$

Varianza

$$Var[S] = \sigma_S^2 = \sigma_1^2 + \sigma_2^2$$

Variable Estándar o Tipificada

$$Z = \frac{S - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(0; 1)$$

Ejemplo

$X_1 \sim N(5; 2)$

$X_2 \sim N(4; 3)$

$S = X_1 - X_2 \Rightarrow S \sim N(\mu_S; \sigma_S)$

Donde

$\mu_S = \mu_1 - \mu_2 \Rightarrow \mu_S = 5 - 4 = 1$

$\sigma_S^2 = \sigma_S^2 = \sigma_1^2 + \sigma_2^2 \Rightarrow \sigma_S^2 = 2^2 + 3^2 = 13 \Rightarrow \sigma_S = \sqrt{13} \cong 3{,}61$

$S \sim N(1; 3{,}61)$

Ahora bien supongamos que se desea saber cual es la probabilidad de que S se encuentre entre 0,50 y 2 inclusive. Es decir,

$$P[0{,}50 \leq S \leq 2] = ?$$

¿Tiene sentido calcular esta probabilidad con esta nueva variable normal?

$$P[0{,}50 \leq S \leq 2] = F[S = 2] - F[S = 0{,}50]$$

$$P[0{,}50 \leq S \leq 2] = F\left[Z = \frac{2-1}{3{,}61}\right] - F\left[Z = \frac{0{,}50-1}{3{,}61}\right]$$

$$P[0{,}50 \leq S \leq 2] = F[Z \cong 0{,}28] - F[Z \cong -0{,}14]$$

$$P[0{,}50 \leq S \leq 2] = 0{,}60924435 - 0{,}44485347 = 0{,}16439088$$

Es decir la probabilidad de que S se encuentre entre 0,50 y 2 inclusive es de 0,16439088.

Teorema Central del Límite

En el resultado anterior, veíamos que la suma de variables aleatorias normales es otra variable aleatoria normal. Sin embargo, la normalidad de una suma de variables no se limita solo a las variables normales.

El teorema central del límite es un resultado matemático que garantiza que, si sumamos variables cualesquiera (no necesariamente normales), la variable suma también seguirá una distribución normal (esto siempre que se cumplan algunas condiciones básicas).

Así, cuando un dato o resultado es la suma de contribuciones independientes, de igual magnitud y "con un tamaño típico", este resultado corresponderá a una Distribución Normal siempre que el número de contribuciones (el número de sumandos) sea un número considerable (no pequeño).

Con un tamaño típico se quiere garantizar que las contribuciones tienen que "estar controladas", esto es, las contribuciones extremas tienen que estar controladas por una probabilidad muy pequeña (En jerga matemática las contribuciones tiene que tener varianza finita).

Este teorema asegura, de manera esquemática, que, cuando sumamos un número grande de variables, la variable resultante sigue una distribución normal.

De manera general si contamos con la siguiente sucesión de variables aleatorias,

$$X_1, X_2, \dots, X_n$$

Y de cada variable se conoce el valor esperado y su varianza, en términos formales,

$$\mu_i = E[X_i] \qquad \sigma_i^2 = Var[X_i]$$

Se verifica que la variable suma $S = X_1 + X_2 + \cdots + X_n$ si n es un número tendiendo a infinito) se puede aproximar por una variable normal, de media la suma de las medias y varianza la suma de varianzas, es decir

$$S \approx N\left(\sum_{i=1}^{n} \mu_i\,, \sqrt{\sum_{i=1}^{n} \sigma_i^2}\right) \Rightarrow Z = \frac{S - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \to N(0;1)$$

En el curso utilizaremos este TCL cuando $n > 30$. (es una suposición demasiado fuerte que no se cumple pero que en los libros básicos se afirma como cantidad de observaciones necesarias para aplicarlo)

En el caso de sumar variables aleatorias normales, la aproximación anterior no es tal, sino que es una distribución exacta, como hemos visto anteriormente.

Si, en vez de sumar variables, realizamos la media aritmética de las mismas, también podemos utilizar el teorema central del límite (puesto que la media aritmética es sumar y luego dividir por una constante).

Este teorema (del que damos únicamente una idea general, sin establecer las hipótesis matemáticas reales) establece la importancia de la distribución normal.

Su resultado es que, cuando se suma un número grande de variables aleatorias, la variable resultante es una variable con distribución aproximadamente igual a la distribución normal. Incluso, el término número grande(porque matemáticamente el teorema se establece cuando n tiende a infinito) no lo es tanto, porque, en la práctica, con tener que n sea un número mayor o igual a 30 , la aproximación ya proporciona buenas resultados.

Ejemplo

$$X_1 \sim ? \Rightarrow \begin{cases} \mu_1 = E[X_1] = 5 \\ \sigma_1^2 = Var[X_1] = 4 \end{cases}$$

$$X_2 \sim ? \Rightarrow \begin{cases} \mu_2 = E[X_2] = 6 \\ \sigma_2^2 = Var[X_2] = 3 \end{cases}$$

$$X_3 \sim ? \Rightarrow \begin{cases} \mu_2 = E[X_2] = 4 \\ \sigma_2^2 = Var[X_2] = 2 \end{cases}$$

$$S = 20X_1 + 30X_2 + 50X_3 \Rightarrow S \sim ?$$

$$\mu_S = 20\mu_1 + 30\mu_2 + 50\mu_3 \Rightarrow \mu_S = 20 * 5 + 30 * 6 + 50 * 4 = 480$$

$$\sigma_S^2 = 4\sigma_1^2 + 3\sigma_2^2 + 2\sigma_3^2 \Rightarrow \sigma_S^2 = 20 * 4 + 30 * 3 + 50 * 2 = 270 \Rightarrow \sigma_S = \sqrt{270} \cong 16{,}46$$

$$S \approx N(480{,}16{,}46) \Rightarrow Z = \frac{S - 480}{16{,}46} \rightarrow N(0;1)$$

Ahora bien supongamos que se desea saber cual es la probabilidad de que S se encuentre entre 460 y 480 inclusive. Es decir,

$$P[460 \leq S \leq 480] \approx F\left[Z = \frac{480 - 480}{16,46}\right] - F\left[Z = \frac{460 - 480}{16,46}\right]$$

$$P[460 \leq S \leq 480] \approx F[Z \cong 0] - F[Z \cong -1,22]$$

$$P[460 \leq S \leq 480] \approx 0,50 - 0,11123244 \ = 0,38876756$$

Es decir la probabilidad de que S se encuentre entre 0,50 y 2 inclusive es de 0,38876756.

Además, el teorema es cierto independientemente de la distribución que sigan las variables que se sumen (no importa si son exponenciales, binomiales, etc.). Lo único que se necesita es saber su media y su varianza.


TO BE
CONTINUED....

Preguntas, Sugerencias y Comentarios

RDelRosso-ext@austral.edu.ar

¡Muchas Gracias!