

CLASES DE ESTADÍSTICA II

CLASE 5) UNA TRANSFORMACIÓN DE DOS VARIABLES.

Sea $Z = g(X, Y)$ una función de las variables aleatorias X e Y , tales que se conoce la función de densidad conjunta entre estas variables.

¿Cuáles son las características aleatorias de la nueva variable?

Esta pregunta se va a responder en esta clase. El orden de presentación de estas respuestas será el siguiente

1. Curvas de nivel.
2. Regiones de interés en R^2 .
3. Cálculo de la función de distribución acumulativa de Z .
4. Cálculo de la función de densidad de Z .

1. Curvas de nivel.

La representación de $Z = g(X, Y)$ es el de una función en R^3 cuyo dominio está en R^2 . Entendemos por curva de nivel a la intersección de un plano paralelo al plano XY con la superficie $g(X, Y)$. Estas intersecciones se proyectan sobre el plano XY para diferentes valores de k , siendo $k = g(X, Y)$.

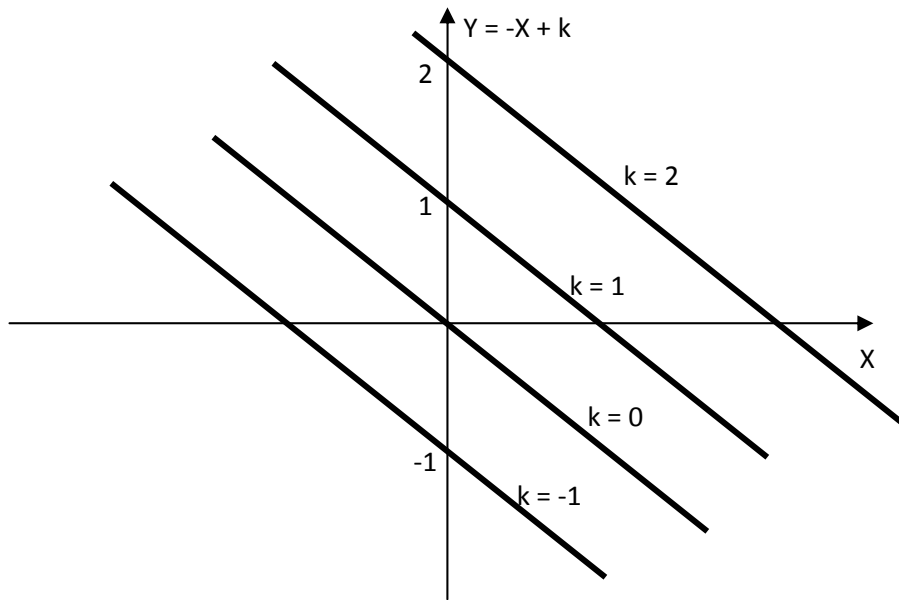
Como ejemplos de curvas de nivel veamos las funciones siguientes:

Ejemplo 5.1) Sean las siguientes funciones $g(X, Y)$. Hallar las curvas de nivel en cada caso.

- a) $g(X, Y) = aX + bY$.
- b) $g(X, Y) = (aX + b)(cY + d)$.
- c) $g(X, Y) = XY$.
- d) $g(X, Y) = \frac{Y}{X}$.
- e) $g(X, Y) = \min(X, Y)$.
- f) $g(X, Y) = \max(X, Y)$.

a) $g(X, Y) = aX + bY$.

En este caso, $g(X, Y) = k$ genera **rectas** paralelas de la forma $Y = -\frac{a}{b}X + \frac{k}{b}$. En la gráfica se presentan las curvas de nivel para el caso específico $a = b = 1$. Observe el sentido en el cual crece el parámetro k .

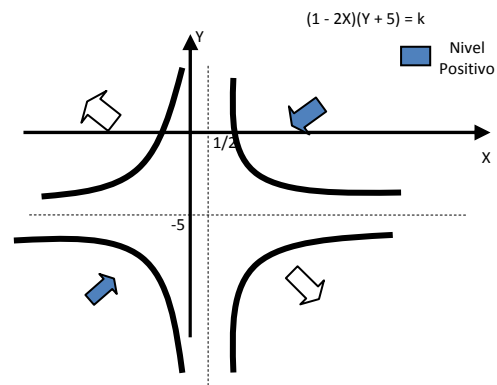
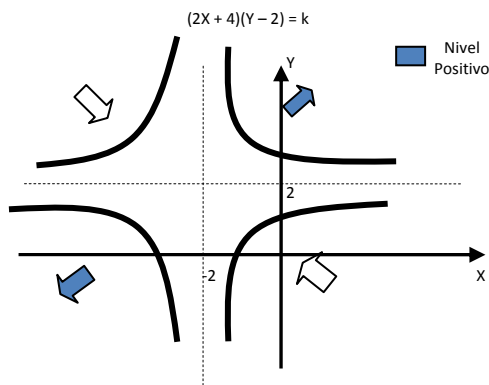


b) $g(X, Y) = (aX + b)(cY + d)$.

En este caso las curvas de nivel serán **hipérbolas equiláteras** de dos ramas de la forma

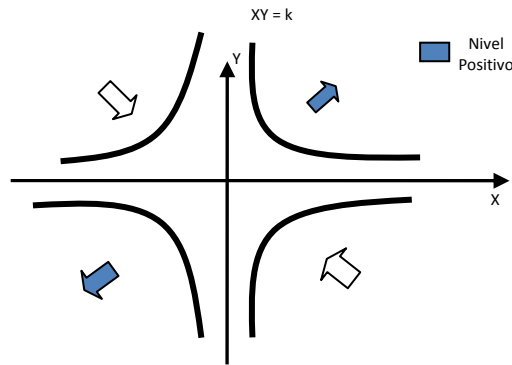
$$(aX + b)(cY + d) = k$$

El centro de las hipérbolas será el punto $C = (c_1, c_2) = \left(-\frac{b}{a}, -\frac{d}{c}\right)$. Una curva de nivel está representada por dos ramas simétricas y opuestas por el vértice donde las curvas en los cuadrantes I y III son del mismo signo; igual ocurre con las curvas en los cuadrantes II y IV, pero con el signo opuesto al de los otros cuadrantes. Los niveles negativos crecen acercándose al centro y los niveles positivos crecen alejándose del centro. En las gráficas siguientes se muestran dos ejemplos. Se destaca en cada caso el sentido de crecimiento con las flechas.



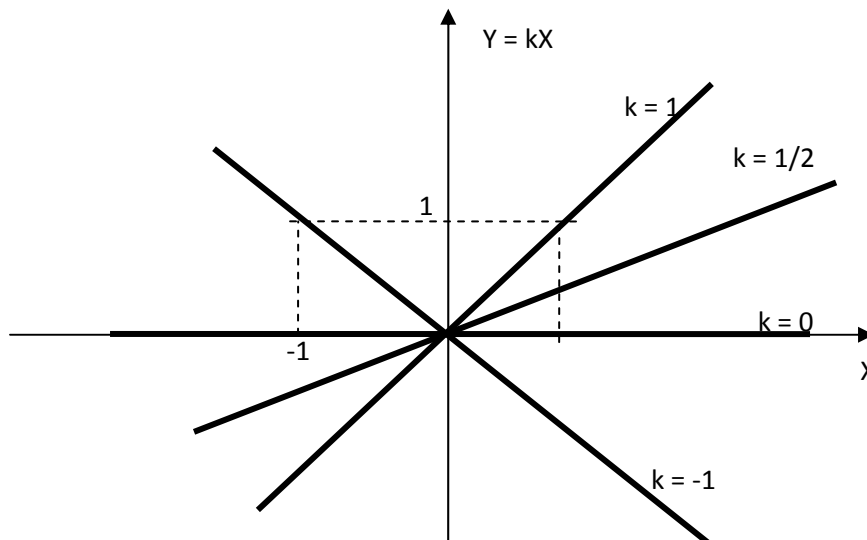
c) $g(X, Y) = XY$.

Este es un caso particular del ejemplo b). Aquí las hipérbolas tienen como centro el origen de coordenadas. Nuevamente, cada curva de nivel tiene dos ramas o las de los cuadrantes I y III o las de los cuadrantes II y IV. La curva de nivel en los cuadrantes I y III crece alejándose del origen.



d) $g(X, Y) = \frac{Y}{X}$.

Aquí las curvas de nivel son rectas de la forma $Y = kX$, que pasan por el origen de coordenadas con diferentes pendientes iguales a k . Nuevamente note el sentido de crecimiento de k .

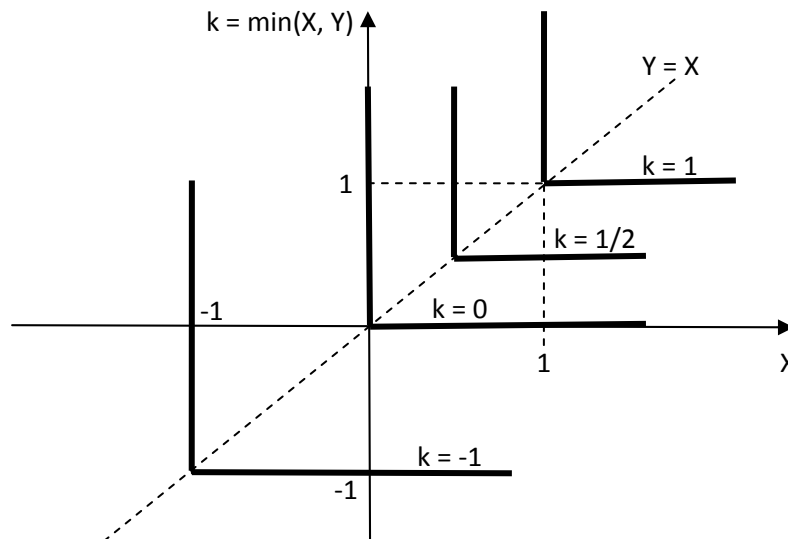


e) $g(X, Y) = \min(X, Y)$.

La función Mínimo requiere de una definición por partes ya que toma distintos valores dependiendo de la región considerada. La frontera que separa las dos posibles regiones es $Y = X$, es decir,

$$k = \min(X, Y) = \begin{cases} X, & \text{si } X < Y \\ Y, & \text{si } X > Y \end{cases}$$

En la gráfica siguiente se presentan distintas curvas de nivel $k = \min(X, Y)$. Nótese el sentido de crecimiento de k .

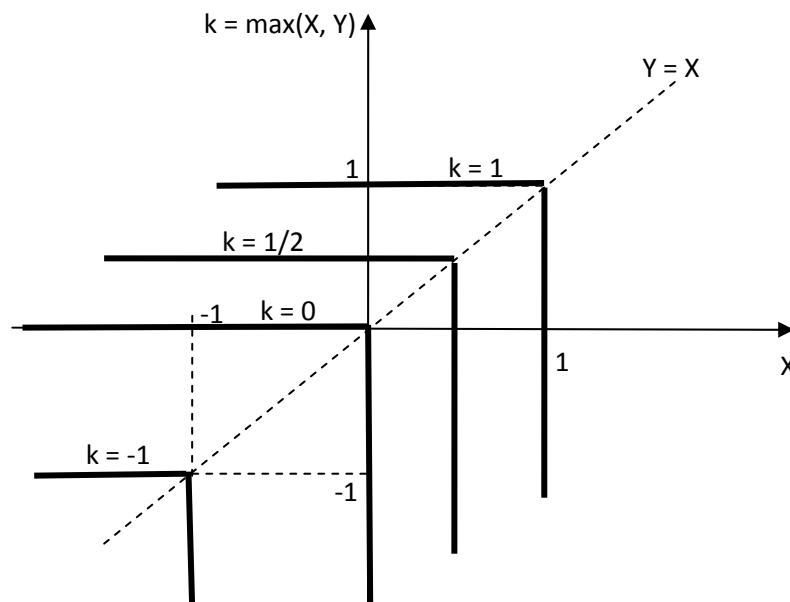


f) $g(X, Y) = \max(X, Y)$.

La función Máximo también requiere de una definición por partes. Nuevamente, $Y = X$ es la frontera que separa las dos posibles regiones.

$$k = \max(X, Y) = \begin{cases} Y, & \text{si } X < Y \\ X, & \text{si } X > Y \end{cases}$$

En la gráfica siguiente se presentan distintas curvas de nivel $k = \max(X, Y)$. Nótese el sentido de crecimiento de k .



IOIO

2. Regiones de interés en R^2 .

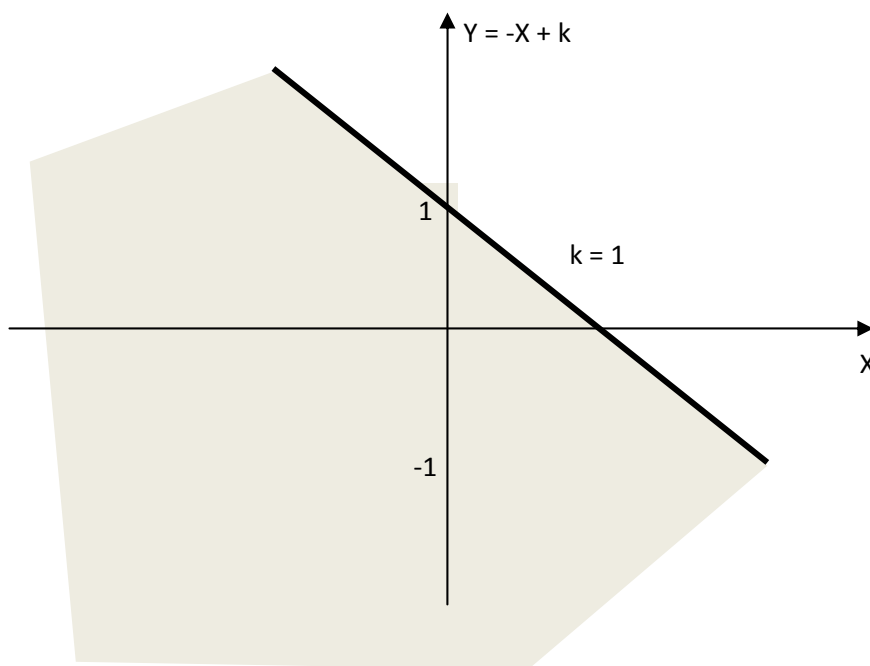
Cada curva de nivel divide el plano XY en dos regiones que denominaremos, región interior y región exterior, con las características siguientes:

$$R_I(Z) = \{ \forall Z \in R / Z \leq g(X, Y) \} \quad R_E(Z) = \{ \forall Z \in R / Z > g(X, Y) \}$$

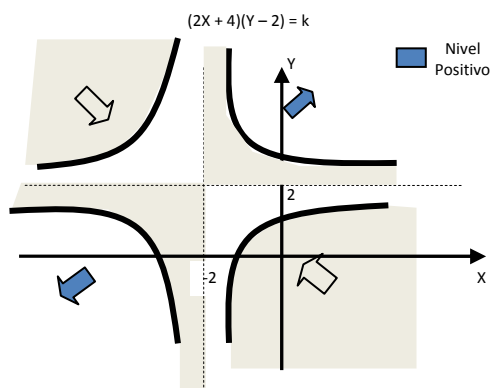
Ejemplo 5.2) Hallar la región interior para las curvas de nivel de las funciones $Z = g(X, Y)$ del ejemplo anterior.

En cada caso la región interior se destaca como la región sombreada.

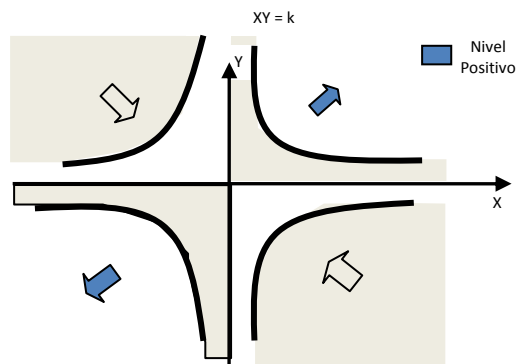
a) $g(X, Y) = aX + bY$.



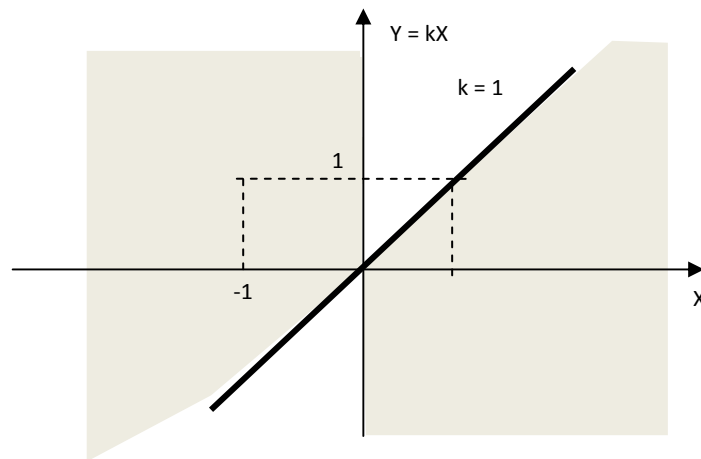
b) $g(X, Y) = (aX + b)(cY + d)$.



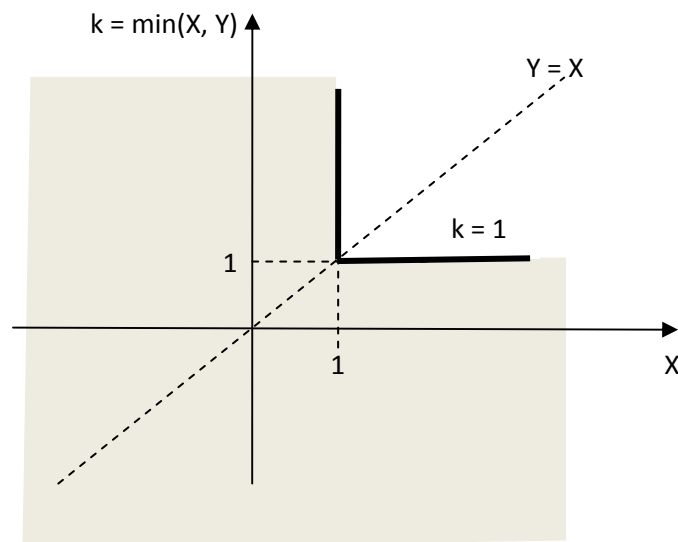
c) $g(X, Y) = XY$.



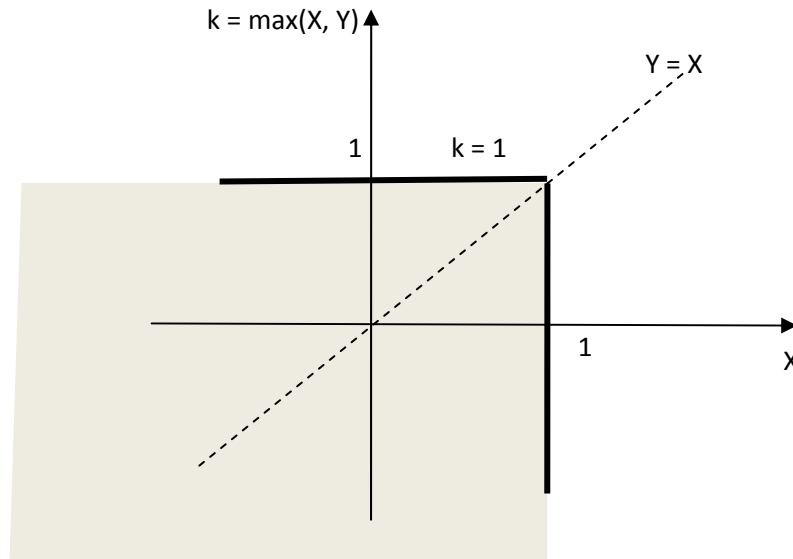
d) $g(X, Y) = \frac{Y}{X}$.



e) $g(X, Y) = \min(X, Y)$.



f) $g(X, Y) = \max(X, Y)$.



IOIO

3. Cálculo de la función de distribución acumulativa de Z.

Una vez que se conocen las curvas de nivel y se ha determinado cuál es la región interior, el cálculo de la función de distribución acumulativa de Z es muy sencillo. Sea $f_{XY}(x, y)$ la función de densidad conjunta de X e Y, entonces

$$F_Z(z) = P\{Z \leq z\} = P\{g(X, Y) \leq z\} = P\{(x, y) \in R_I(z)\} = \iint_{R_I(z)} f_{XY}(x, y) dx dy$$

Ejemplo 5.3) a) Hallar la función de distribución acumulativa de $Z = X + Y$. b) Sean X e Y ambas uniformes continuas en (0, 1) e independientes; hallar la función de distribución de su suma.

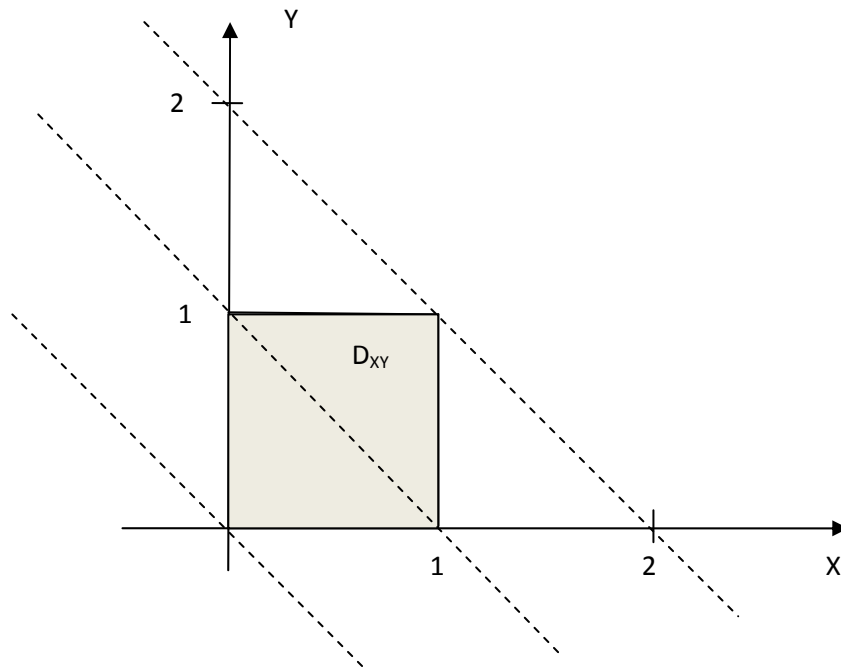
Parte a)

$$F_Z(z) = \iint_{R_I(z)} f_{XY}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x, y) dx dy$$

Parte b) Ya que X e Y son uniformes continuas en (0, 1) e independientes se puede escribir

$$f_{XY}(x, y) = \begin{cases} 1; & (x, y) \in D_{XY} \\ 0; & \text{en otros casos} \end{cases}$$

Donde la región D_{XY} es la región sombreada en la figura siguiente, allí se destacan varias curvas de nivel de $Z = X + Y$ (para $z = 0, 1$ y 2). Nótese que en distintos casos la región resultante de interceptar $R_I(z)$ y D_{XY} cambia por lo que el resultado final cambia.



Entonces,

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x,y) dx dy = \begin{cases} 0; & \text{si } z < 0 \\ \int_0^z \int_0^{z-y} dx dy; & \text{si } 0 \leq z < 1 \\ 1 - \int_{z-1}^1 \int_{z-y}^1 dx dy; & \text{si } 1 \leq z < 2 \\ 1; & \text{si } z \geq 2 \end{cases}$$

$$F_Z(z) = \begin{cases} 0; & \text{si } z < 0 \\ \frac{z^2}{2}; & \text{si } 0 \leq z < 1 \\ 1 - \frac{(2-z)^2}{2}; & \text{si } 1 \leq z < 2 \\ 1; & \text{si } z \geq 2 \end{cases}$$

□□□

4. Cálculo de la función de densidad de Z.

Para el cálculo de la función de densidad de Z conocida la función de distribución de Z solo hay que tomar la derivada respecto de Z de la función de distribución, es decir,

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_{R_I(z)} f_{XY}(x, y) dx dy$$

Ahora bien, para tomar esta derivada hay que tener en cuenta que la variable Z está presente en los límites de integración en ambas integrales por lo que es conveniente recordar la expresión siguiente

$$\frac{d}{dz} \left(\int_{a(z)}^{b(z)} C(x, z) dx \right) = C(b(z), z) \frac{d}{dz} b(z) - C(a(z), z) \frac{d}{dz} a(z) + \int_{a(z)}^{b(z)} \frac{\partial}{\partial z} (C(x, z)) dx$$

Ejemplo 5.4) a) Hallar la función de densidad de $Z = X + Y$ cuando ambos sumandos son independientes. b) Sean X e Y ambas uniformes continuas en (0, 1) e independientes; hallar la función de densidad de su suma.

Parte a)

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_{R_I(z)} f_{XY}(x, y) dx dy = \frac{d}{dz} \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x, y) dx dy \\ f_Z(z) &= \int_{-\infty}^{\infty} \frac{d}{dz} \int_{-\infty}^{z-y} f_{XY}(x, y) dx dy = \int_{-\infty}^{\infty} f_{XY}(z-y, y) dy = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \end{aligned}$$

Esta última integral es la definición de convolución. Lo que lleva a emitir un resultado de gran utilidad en aplicaciones:

“La función de densidad de la suma de variables aleatorias independientes es igual a la convolución de las funciones de densidad marginal de cada sumando”

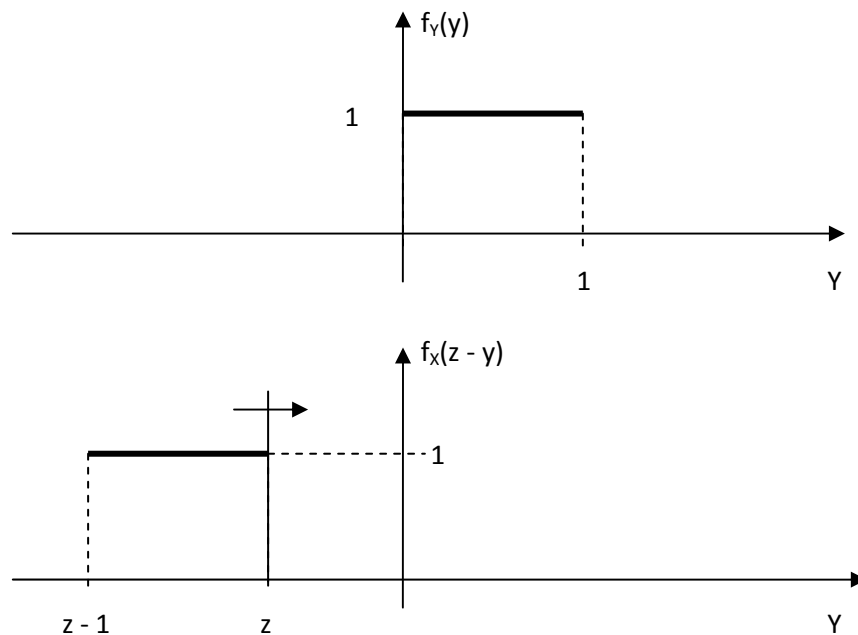
Parte b) Del ejercicio anterior se obtuvo

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \begin{cases} 0; & \text{si } z < 0 \\ \frac{z^2}{2}; & \text{si } 0 \leq z < 1 \\ 1 - \frac{(2-z)^2}{2}; & \text{si } 1 \leq z < 2 \\ 1; & \text{si } z \geq 2 \end{cases} = \begin{cases} 0; & \text{si } z < 0 \\ z; & \text{si } 0 \leq z < 1 \\ 2-z; & \text{si } 1 \leq z < 2 \\ 0; & \text{si } z \geq 2 \end{cases}$$

Pensando en el resultado general de la parte a) de este problema vamos a resolver la integral de convolución con base en un método gráfico que se explicará con las funciones uniformes de este ejemplo. La expresión de la integral de convolución es

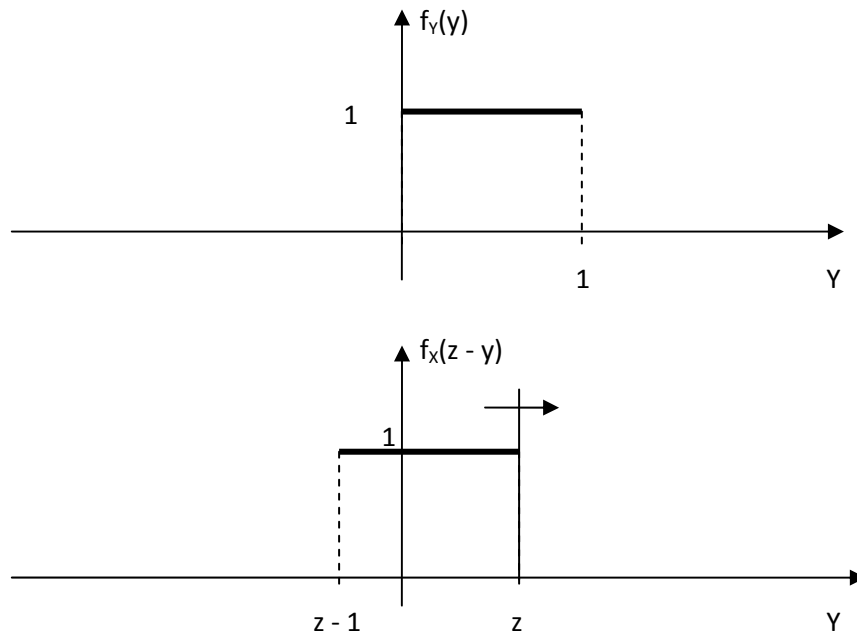
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy = f_X(x) \otimes f_Y(y)$$

Para realizar la convolución es de mucha ayuda graficar las funciones que se multiplican en el integrando. Una de ellas se describirá tal cual está dada y la otra debe pasar por dos procesos consecutivos, el primero de ellos es un rebatimiento respecto al eje vertical y luego un desplazamiento de magnitud z . Las gráficas resultantes serán



Nótese que para el valor actual de z ($z < 0$) el producto de las dos funciones es cero por lo que la función de densidad vale cero. La idea es ir dándole valores a z y verificando el intervalo donde ambas funciones toman valores distintos de cero. a este intervalo se le llamará intervalo de solapamiento. Cuando $z = 0$ se inicia el solapamiento.

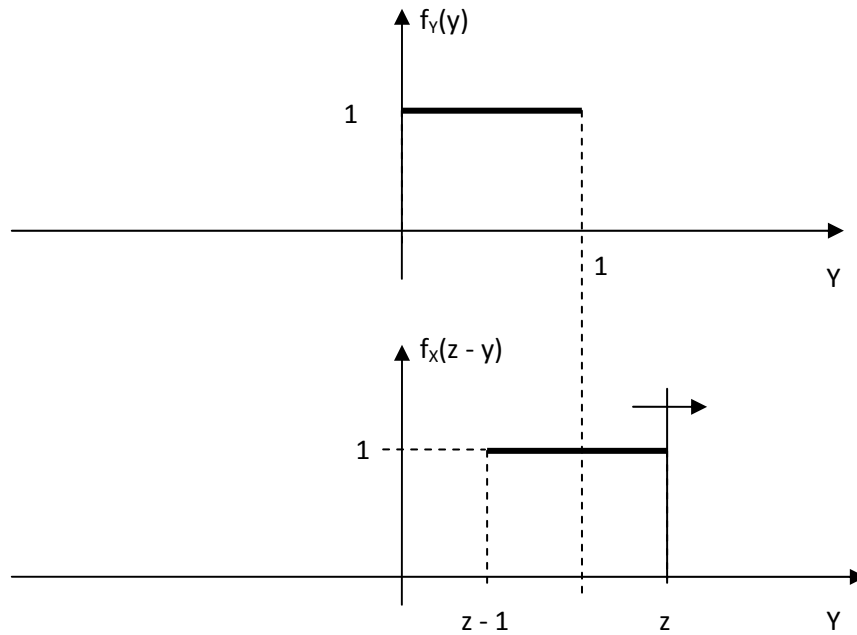
Sea $0 < z < 1$, la gráfica de ambas funciones será



El intervalo de solapamiento va desde $y = 0$ hasta $y = z$, por tanto,

$$f_Z(z) = \int_0^z dy = z$$

Sea $1 < z < 2$, la gráfica de ambas funciones será



El intervalo de solapamiento va desde $y = z - 1$ hasta $y = 1$, por tanto,

$$f_Z(z) = \int_{z-1}^1 dy = 2 - z$$

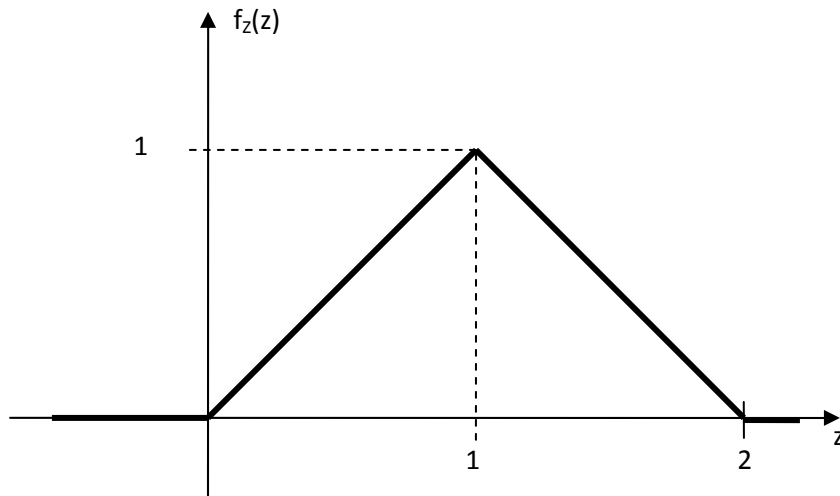
Finalmente, cuando $z > 2$, ya no existe solapamiento

$$f_Z(z) = 0$$

En resumen,

$$f_Z(z) = \begin{cases} 0; & \text{si } z < 0 \\ z; & \text{si } 0 \leq z < 1 \\ 2 - z; & \text{si } 1 \leq z < 2 \\ 0; & \text{si } z \geq 2 \end{cases}$$

Es interesante, observar la gráfica resultante de $f_Z(z)$



La variable Z resulta ser una triangular simétrica. Este ejemplo permite demostrar la aseveración hecha en la primera clase y que se transcribe a continuación.

“Si X_1 y X_2 son dos variables aleatorias independientes e idénticamente distribuidas según una distribución uniforme en (a, b) , entonces la variable $Y = X_1 + X_2$ tiene una distribución de tipo triangular simétrica en el intervalo $(2a, 2b)$ ”

IOIO

Como consecuencia de las conclusiones logradas en este ejemplo 40, el lector puede realizar la convolución correspondiente cuando los sumandos independientes tienen las distribuciones siguientes: (Trate de identificar, en cada caso, el tipo de modelo probabilístico que corresponda al resultado obtenido)

- i) X e Y son Exponenciales con igual parámetro λ .
- ii) X e Y son Exponenciales con distinto parámetro λ .
- iii) X e Y son Normales con iguales parámetros (μ, σ^2) .
- iv) X e Y son Normales con distintos parámetros (μ, σ^2) .
- v) X e Y son Gamma con igual parámetro α y distinto parámetro r .

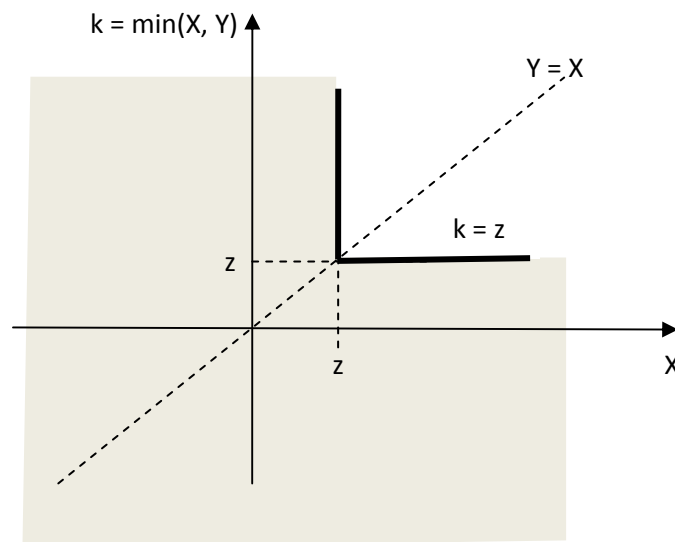
El resultado del ejemplo 5.4 permite hacer una generalización a más de dos variables que deriva en unos resultados de aplicación práctica de mucho interés, como son los siguientes:

- La suma de n variables Bernoulli de igual parámetro p e independientes es una variable Binomial con parámetros n y p .
- La suma de r variables Geométricas de igual parámetro p e independientes es una variable Binomial Negativa con parámetros r y p .
- La suma de n variables Binomiales Negativas con igual parámetro p y distintos parámetros r_i , $i = 1, \dots, n$ e independientes, es una variable Binomial Negativa con parámetro p y $(r = \sum_{i=1}^n r_i)$
- La suma de n variables Poisson con parámetros λ_i , $i = 1, \dots, n$, respectivamente, definidas en el mismo intervalo t , es una variable Poisson con parámetros $(\lambda = \sum_{i=1}^n \lambda_i)$ y t .
- La suma de n variables Exponenciales con igual parámetro λ e independientes es una variable Erlang con parámetros (λ, n) .
- La suma de n variables Chi-Cuadrado con r_i , $i = 1, \dots, n$ grados de libertad e independientes es una variable Chi-Cuadrado con grados de libertad la suma de los grados de libertad de cada Chi-Cuadrado.
- La suma de n variables Gamma con parámetro α común y parámetro r distinto e independientes es una variable Gamma con parámetros, α y, r igual a la suma de los parámetros r de cada sumando.
- La suma de n variables normales con parámetros (μ_i, σ_i^2) , $i = 1, \dots, n$, e independientes es una variable normal con parámetros $(\mu = \sum_{i=1}^n \mu_i, \sigma^2 = \sum_{i=1}^n \sigma_i^2)$.

Está de más indicarle al lector que tiene que demostrar estas aseveraciones. La mejor manera de hacerlo es a través de la relación entre la función generadora de momentos n -dimensional y las funciones generadoras de momentos marginales. **¿La recuerda Usted?**

Ejemplo 5.5) a) Hallar la función de densidad de $Z = \min(X, Y)$ cuando ambas variables son independientes.

Recordando la región interior para $Z = \min(X, Y)$



Entonces,

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_{R_I(z)} f_{XY}(x, y) dx dy = \frac{d}{dz} \iint_{R_I(z)} f_X(x) f_Y(y) dx dy$$

$$f_Z(z) = \frac{d}{dz} \left(\int_{-\infty}^{\infty} f_Y(y) \left[\int_{-\infty}^z f_X(x) dx \right] dy + \int_{-\infty}^z f_Y(y) \left[\int_z^{\infty} f_X(x) dx \right] dy \right)$$

La expresión entre corchetes en cada integral no depende del diferencial correspondiente por lo que puede escribirse fuera de la integral correspondiente

$$f_Z(z) = \frac{d}{dz} \left(\left[\int_{-\infty}^z f_X(x) dx \right] \int_{-\infty}^{\infty} f_Y(y) dy + \left[\int_z^{\infty} f_X(x) dx \right] \int_{-\infty}^z f_Y(y) dy \right)$$

Sustituyendo las integrales por las definiciones correspondientes

$$f_Z(z) = \frac{d}{dz} \left(F_X(z) + (1 - F_X(z)) F_Y(z) \right)$$

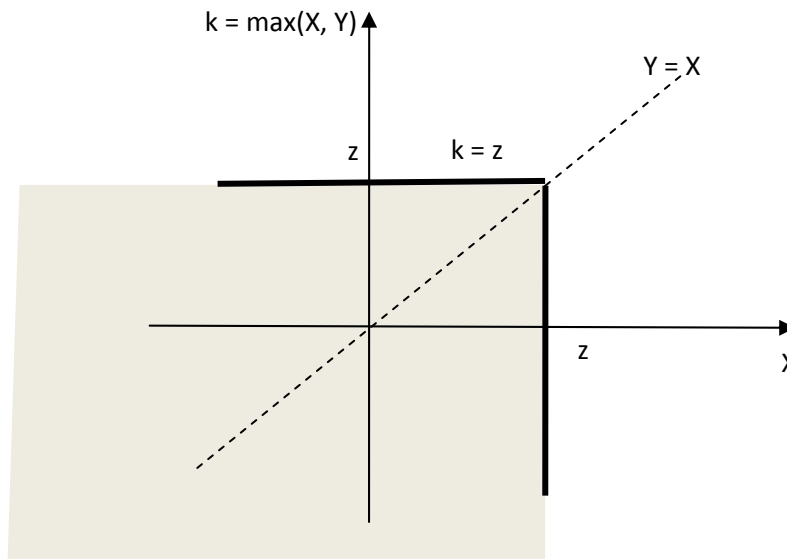
Finalmente, al tomar la derivada respecto de z, se obtiene

$$f_Z(z) = (1 - F_X(z)) f_Y(z) + (1 - F_Y(z)) f_X(z)$$

IOIO

Ejemplo 5.6) a) Hallar la función de densidad de $Z = \max(X, Y)$ cuando ambas variables son independientes.

Recordando la región interior para $Z = \max(X, Y)$



Entonces,

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_{R_I(z)} f_{XY}(x, y) dx dy = \frac{d}{dz} \iint_{R_I(z)} f_X(x) f_Y(y) dx dy$$

$$f_Z(z) = \frac{d}{dz} \left(\int_{-\infty}^z f_Y(y) \left[\int_{-\infty}^z f_X(x) dx \right] dy \right)$$

La expresión entre corchetes en cada integral no depende del diferencial correspondiente por lo que puede escribirse fuera de la integral correspondiente

$$f_Z(z) = \frac{d}{dz} \left(\left[\int_{-\infty}^z f_X(x) dx \right] \int_{-\infty}^z f_Y(y) dy \right)$$

Sustituyendo las integrales por las definiciones correspondientes

$$f_Z(z) = \frac{d}{dz} (F_X(z) F_Y(z))$$

Finalmente, al tomar la derivada respecto de z, se obtiene

$$f_Z(z) = F_X(z) f_Y(z) + F_Y(z) f_X(z)$$

IOIO

Ejemplo 5.7) a) Hallar la función de densidad del máximo y el mínimo de dos variables exponenciales independientes con parámetros λ_1 y λ_2 , respectivamente.

Sean X e Y las variables exponenciales del enunciado, entonces,

$$f_X(x) = \lambda_1 e^{-\lambda_1 x} u(x) \rightarrow F_X(x) = (1 - e^{-\lambda_1 x}) u(x)$$

$$f_Y(y) = \lambda_2 e^{-\lambda_2 y} u(y) \rightarrow F_Y(y) = (1 - e^{-\lambda_2 y}) u(y)$$

Sean $Z = \min(X, Y)$ y $W = \max(X, Y)$ entonces, de los resultados de los ejemplos 41 y 42 se tiene que

$$f_Z(z) = e^{-\lambda_1 z} \lambda_2 e^{-\lambda_2 z} u(z) + e^{-\lambda_2 z} \lambda_1 e^{-\lambda_1 z} u(z) = (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)z} u(z)$$

Nótese que Z también es Exponencial pero con parámetro igual a la suma de los parámetros.

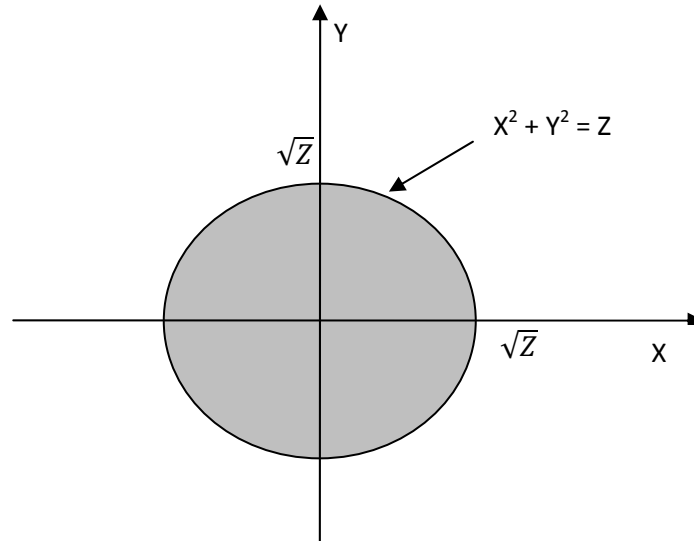
$$f_W(w) = (1 - e^{-\lambda_1 w}) \lambda_2 e^{-\lambda_2 w} u(w) + (1 - e^{-\lambda_2 w}) \lambda_1 e^{-\lambda_1 w} u(w)$$

$$f_W(w) = (\lambda_1 e^{-\lambda_1 w} + \lambda_2 e^{-\lambda_2 w} - (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)w}) u(w)$$

IOIO

Ejemplo 5.8) a) Hallar la función de densidad de $Z = X^2 + Y^2$, donde X e Y son normales estándar e independientes.

Al plantear las curvas de nivel se consiguen circunferencias centradas en el origen de radio \sqrt{Z} y al buscar la región interior se consigue la región de la figura siguiente



Entonces la función de densidad de Z será

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} \iint_{R_I(z)} f_{XY}(x, y) dx dy = \frac{d}{dz} \int_{-\sqrt{z}}^{\sqrt{z}} \int_{-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x, y) dx dy$$

Para resolver esta integral es conveniente hacer un cambio de variable de la forma

$$x = r \cos \theta \quad y = r \sin \theta$$

Obteniendo, (solo para valores de $z > 0$)

$$f_Z(z) = \frac{d}{dz} \left(\int_0^{\sqrt{z}} \int_0^{2\pi} r f_{XY}(r \cos \theta, r \sin \theta) d\theta dr \right) = \frac{d}{dz} \left(\int_0^{\sqrt{z}} \int_0^{2\pi} \frac{r}{2\pi} e^{-\frac{r^2}{2}} d\theta dr \right)$$

Resolviendo las integrales y tomando la derivada se obtiene, (solo para valores de $z > 0$)

$$f_Z(z) = \frac{1}{2} e^{-\frac{z}{2}}$$

Nótese que Z sigue una distribución Exponencial con parámetro $\lambda = \frac{1}{2}$ o también una Chi-Cuadrado con dos grados de libertad.

IOIO

TEOREMA CENTRAL DEL LÍMITE

Este teorema estudia el comportamiento de la suma de variables aleatorias cuando crece el número de sumandos. Este comportamiento, para un número de sumandos grande, se aproxima a una variable aleatoria normal. En la literatura se presenta este teorema de diversas maneras.

Teorema Central del Límite: Sean n variables aleatorias X_1, X_2, \dots, X_n independientes y con valores esperados $\mu_i, i = 1, 2, \dots, n$ y varianzas $\sigma_i^2, i = 1, 2, \dots, n$, respectivamente. Sean b, a_1, a_2, \dots, a_n , constantes reales arbitrarias. Sea otra variable aleatoria Y definida como

$$Y = b + \sum_{i=1}^n a_i X_i$$

Entonces, para un n lo suficientemente grande, la variable Y tiene una distribución aproximadamente NORMAL con parámetros

$$\mu_Y = b + \sum_{i=1}^n a_i \mu_i \quad \sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

De acuerdo con los valores de las constantes b, a_1, a_2, \dots, a_n , el teorema se puede escribir de otras maneras.

Versión 2: Sean n variables aleatorias X_1, X_2, \dots, X_n independientes y con valores esperados $\mu_i, i = 1, 2, \dots, n$ y varianzas $\sigma_i^2, i = 1, 2, \dots, n$, respectivamente. Sean S la suma y \bar{X} el promedio de estas variables. Entonces

$$S = \sum_{i=1}^n X_i \approx N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{S}{n} \approx N\left(\frac{1}{n} \sum_{i=1}^n \mu_i, \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2\right)$$

Versión 3: Sean n variables aleatorias X_1, X_2, \dots, X_n independientes e idénticamente distribuidas con valor esperado y varianzas comunes e iguales a μ y σ^2 , respectivamente. Sean S la suma y \bar{X} el promedio de estas variables. Entonces

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

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{S}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Nota: en las expresiones anteriores el símbolo \approx quiere decir que “está aproximadamente distribuida como”.

Es interesante observar el comportamiento de la variable X en la versión 3 del Teorema Central del Límite cuando n tiende a infinito. Se comporta como una normal donde su valor esperado es igual al valor esperado de cada X_i pero su varianza tiende a cero. !!!

Ejemplo 5.9) Se colocan 12 resistencias en serie extraídas de una caja donde hay un gran lote de ellas con igual valor nominal y tolerancia. De los datos del fabricante se sabe que el valor nominal es de 1 K Ω y que están distribuidas uniformemente en un intervalo de 10% alrededor del valor nominal. Calcular de forma aproximada la probabilidad de que el valor de la resistencia total en serie difiera de su valor nominal en menos del 1%.

Para cada resistencia en la conexión en serie, su valor esperado es el nominal (1 K Ω) y su varianza viene dada por

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(1,1 \text{ K}\Omega - 0,9 \text{ K}\Omega)^2}{12} = \frac{0,04}{12} \text{ K}\Omega^2$$

Por tanto, la resistencia total en serie será

$$R = \sum_{i=1}^{12} R_i$$

Por tanto R tiene una distribución aproximadamente normal con parámetros

$$E(R) = 12E(R_i) = 12\text{K}\Omega; \quad V(R) = 12V(R_i) = 0,04 \text{ K}\Omega^2$$

En consecuencia la probabilidad solicitada, apoyándose en el Teorema Central del Límite será

$$P\{12 - 0,12 < R < 12 + 0,12\} \cong P\{-0,6 < Z < 0,6\} = \Phi(0,6) - \Phi(-0,6) = 2\Phi(0,6) - 1$$

De una tabla de áreas bajo la normal estándar se tiene que $\Phi(0,6) = 0,72575$, entonces

$$P\{12 - 0,12 < R < 12 + 0,12\} \cong 2 \times 0,72575 - 1 = 0,4515$$

IO IO

Ejemplo 5.10) En la fabricación de un artículo se toman en cuenta tres costos: un costo fijo de producción de 12 unidades monetarias (um), un costo debido a insumos de 4 um/Kg y un costo debido al número de operarios utilizados de 8 um/operario. Se tienen disponibles para la producción un monto de 2600 um. Si se sabe que cada unidad producida consume una cantidad aleatoria de insumos X y necesita un número de operarios también aleatorio Y , ¿Cuántas unidades deben planificarse producir para tener una probabilidad de 0,9 de tener éxito, es decir, de que el costo total no exceda el monto disponible?. Las variables X e Y tienen distribuciones dadas por las expresiones siguientes

$$f_X(x) = \begin{cases} \frac{x^2}{19}; & 2 \leq x < 3 \\ \frac{5}{12}; & 3 \leq x < 4 \\ \frac{x^3 + 2x + 1}{409}; & 4 \leq x < 5 \\ 0; & \text{en otros casos} \end{cases} \quad F_Y(y) = \begin{cases} 0; & y < 1 \\ 0,60; & 1 \leq y < 2 \\ 0,80; & 2 \leq y < 3 \\ 0,95; & 3 \leq y < 4 \\ 1; & y \geq 4 \end{cases}$$

Nótese que la variable X es continua mientras que la variable Y es discreta.

Si definimos como C el costo total de la producción de las n unidades, X_i el insumo en Kg que utiliza la i-ésima unidad producida y por Y_i el número de operarios utilizados en la i-ésima unidad producida, se tiene que

$$C = 12 + \sum_{i=1}^n 4X_i + \sum_{i=1}^n 8Y_i$$

Donde n es la incógnita. Por el Teorema Central del Límite, C tiene una distribución aproximadamente normal con parámetros dados por

$$E(C) = 12 + 4nE(X_i) + 8nE(Y_i); \quad V(C) = 16nV(X_i) + 64nV(Y_i)$$

De las distribuciones dadas para X e Y se pueden calcular sus valores esperados y varianzas, obteniendo

$$E(X_i) = 3,4514; \quad V(X_i) = 0,6466; \quad E(Y_i) = 1,65; \quad V(Y_i) = 0,8275$$

En consecuencia,

$$E(C) = 12 + 27,006n; \quad V(C) = 63,306n$$

El valor de n se consigue al considerar que $P\{C < 2600\} = 0,90$; es decir, que 2600 es el percentil 90 de C. En una tabla de áreas bajo la normal estándar se puede leer que el percentil 90 de Z es 1,2816, por tanto

$$2600 = 12 + 27,006n + 1,2816\sqrt{63,306n} \rightarrow n = 92,205 \rightarrow n = 92$$

IO IO

CORRECCIÓN POR CONTINUIDAD AL USAR EL TEOREMA CENTRAL DEL LÍMITE

En el caso en el cual las variables X_i son de tipo discreto la variable Y definida en el Teorema Central del Límite también será de tipo discreto. Ahora bien, la variable Y se aproxima por una Normal que es una variable de tipo continuo. Este hecho incorpora un error que se corrige al aplicar la llamada **Corrección por Continuidad** que se define a continuación.

Versión de Corrección por Continuidad del TCL:

Sean n variables aleatorias DISCRETAS X_1, X_2, \dots, X_n independientes y con valores esperados $\mu_i, i = 1, 2, \dots, n$ y varianzas $\sigma_i^2, i = 1, 2, \dots, n$, respectivamente. Sean b, a_1, a_2, \dots, a_n , constantes reales arbitrarias. Sea otra variable aleatoria Y DISCRETA definida como

$$Y = b + \sum_{i=1}^n a_i X_i$$

Entonces, para un n lo suficientemente grande, la variable Y se aproxima por N , una distribución NORMAL con parámetros

$$\mu_N = b + \sum_{i=1}^n a_i \mu_i \quad \sigma_N^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

Para calcular probabilidades asociadas con esta variable Y se calculan las probabilidades correspondientes en términos de la variable N , es decir,

- i) $P\{Y = y\} = P\{y - 0,5 \leq N \leq y + 0,5\}$
- ii) $P\{Y \leq a\} = P\{N \leq a + 0,5\}$
- iii) $P\{Y \geq b\} = P\{N \geq b - 0,5\}$
- iv) $P\{a \leq Y \leq b\} = P\{a - 0,5 \leq N \leq b + 0,5\}$

Ejemplo 5.11) Se realiza un muestreo en un proceso de producción en el cual se conoce que el 20% de los artículos producidos es defectuoso. Cada hora se realiza la revisión de 100 unidades para determinar si se acepta o no el lote de producción de esa hora. Se define como X la variable aleatoria número de artículos defectuosos en la muestra de tamaño 100. Calcular las siguientes probabilidades, a) $P\{X = 15\}$, b) $P\{X \leq 15\}$, c) $P\{X < 18\}$, d) $P\{X \geq 22\}$, e) $P\{18 < X < 21\}$.

Nótese que X es discreta y sigue una distribución Binomial con parámetros $n = 100$ y $p = 0,2$. Mientras que la revisión de un artículo i es una variable Y_i , con $i = 1, \dots, 100$, entonces Y_i sigue una distribución Bernoulli con parámetro $p = 0,2$. Entonces,

$$X = \sum_{i=1}^{100} Y_i$$

Por el Teorema Central del Límite, X tiene una distribución aproximadamente normal con parámetros dados por

$$E(X) = np = 20; \quad V(X) = np(1 - p) = 16 \rightarrow \sigma_X = 4$$

Pero X mantiene su carácter de variable discreta por lo que se debe usar la corrección por continuidad para el cálculo de las probabilidades de X por medio de la normal equivalente.

Parte a)

$$P\{X = 15\} = P\{14,5 \leq N \leq 15,5\} = P\left\{\frac{14,5 - 20}{4} \leq Z \leq \frac{15,5 - 20}{4}\right\} = \Phi(-1,125) - \Phi(-1,375)$$

$$P\{X = 15\} \cong 0,04573$$

Parte b)

$$P\{X \leq 15\} = P\{N \leq 15,5\} = P\left\{Z \leq \frac{15,5 - 20}{4}\right\} = \Phi(-1,125)$$

$$P\{X \leq 15\} \cong 0,1303$$

Parte c)

$$P\{X < 18\} = P\{N \leq 17,5\} = P\left\{Z \leq \frac{17,5 - 20}{4}\right\} = \Phi(-0,625)$$

$$P\{X < 18\} \cong 0,26599$$

Parte d)

$$P\{X \geq 22\} = P\{N \geq 21,5\} = P\left\{Z \geq \frac{21,5 - 20}{4}\right\} = 1 - \Phi(0,375)$$

$$P\{X \geq 22\} \cong 0,35383$$

Parte e)

$$P\{18 < X < 21\} = P\{18,5 \leq N \leq 20,5\} = P\left\{\frac{18,5 - 20}{4} \leq Z \leq \frac{20,5 - 20}{4}\right\} = \Phi(0,125) - \Phi(-0,375)$$

$$P\{18 < X < 21\} \cong 0,19591$$

Ю Ю