

Notas Practicas profesionales.

Luis Clemente Michicol Martinez

4 de agosto de 2025

Índice general

I	Repaso Probabilidad y estadística.	1
0.1.	Motivación	3
1.	Descripción probabilista	5
1.1.	Variable Aleatoria	5
1.2.	Descripción probabilística.	5
1.2.1.	Enfoque Frecuentista	6
1.2.2.	Enfoque Axiomático.	6
1.2.3.	Distribuciones de probabilidad.	9
1.2.4.	Distribución acumulativa.	9
1.2.5.	Función característica.	11
1.2.6.	Teorema del limite central	12

Índice de figuras

1.1. Distribución del postulado de iguales probabilidades	7
1.2. Esquema de enfermedades y síntomas, con el modelo de causa consecuencia.	8
1.3. Distribución Gaussiana, la parte con relleno corresponde a $P(X \leq x)$, y la parte que no tiene relleno corresponde $P(X \geq x)$	10
1.4. Función de densidad de probabilidad Gaussiana, con representación de σ	13
1.5. Demostración del teorema del limite central.	14

Índice de cuadros

Parte I

Repaso Probabilidad y estadística.

Introducción

0.1. Motivación

Si consideramos N variables aleatorias x_i , $i = 1, \dots, N$, de las cuales es de nuestro interés saber las correlaciones entre las $x_{i's}$, de esta manera hacemos M observaciones, con lo cual se puede construir una **matriz de correlación**

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1M} \\ x_{21} & x_{22} & x_{23} & \cdot & x_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{NM} \end{bmatrix} \quad (1)$$

Sin embargo, para eso no conocemos a $P(x_1, x_2, \dots, x_N)$ que corresponde a las **distribuciones de probabilidades**. Y algo importante aquí, es que se requiere que para cada $P(x_i)$ sea estacionaria, lo cual implica que $P(x_i)$ es invariante ante traslaciones.

Capítulo 1

Descripción probabilista

1.1. Variable Aleatoria

De manera informal, y para entender el concepto sin la necesidad de ser abstracto, un **variable aleatoria** consiste en el resultado de algún proceso o **evento** del cual no sabemos todos los detalle que influyen en su descripción.

La importancia de este concepto, es que pese a la ignorancia de lo que puede o no influir, se hace uso de la herramienta estadística para tratar de modelar estas variables aleatorias, y así poder hacer predicciones sobre estas. Esto es de una manera muy simplificada y poco formal.

Entonces se puede definir una variable aleatoria de la siguiente manera.

Definición 1.1.1 (Variable aleatoria). *Se define una variable aleatoria, como una función (o un mapeo) que nos lleva del espacio de resultados fundamentales (o espacio muestral), a de los reales*

$$X : \Omega \rightarrow \mathbb{R}$$

Si se quiere profundizar con mayor formalidad se puede buscar en la siguiente referencias Rincón, 2007

1.2. Descripción probabilística.

Se define el **conjunto de resultados fundamentales**¹ como.

$$\Omega = \{r_1, r_2, \dots, r_n\}$$

El objetivo de este conjunto es **agrupar a todos los posibles resultados** de un experimento.

Otro concepto de suma importancia son los **eventos**, los cuales consisten en subconjuntos del espacio muestral. Es decir, consideré un conjunto de varios eventos.

$$\mathcal{E} = \{\{r_i\}, \{r_i, r_j\}, \dots\}$$

Donde cada subconjunto se puede renombrar quedando de la siguiente manera.

$$\mathcal{E} = \{\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_n\}$$

Una vez dado los anteriores conceptos, uno se puede preguntar cómo se puede asignar probabilidades a eventos?. Y aquí es donde entran tres tipos de enfoques, **clásico, geométrico, frecuentista**

¹También es conocido como **espacio muestral**

subjetiva y **axiomático**, de los cuales se hablara únicamente de los enfoques frecuentistas, se puede ver el texto de Rincón, s.f., para tener información acerca de los enfoques que no se abordan.

1.2.1. Enfoque Frecuentista

En este caso se considera hacer N repeticiones de un experimento aleatorio, de los cuales se hace el conteo de las veces ($=m_i$) que se obtuvo cierto resultado r_i , entonces, lo que considera el enfoque frecuentista para asignar probabilidad al resultado r_i es lo siguiente.

$$P(r_i) = \lim_{N \rightarrow \infty} \frac{m_i}{N}$$

Para este enfoque existen varios problemas, el más critico y principal, es el hecho de que requiere que $N \rightarrow \infty$, es decir el experimento se requiere repetir infinidad de veces o al menos una cantidad muy grande de repeticiones, lo cual lo hace imposible hacer, y otro gran problema, es que las repeticiones de los experimentos aleatorios tiene que ser **bajo las mismas condiciones**, lo cual se puede notar que supone igualmente un reto bastante grande.

1.2.2. Enfoque Axiomático.

En este enfoque, en lo que se centra es en establecer reglas que cumplan el cálculo de probabilidades, las cuales están dadas por los siguientes axiomas, dichos axiomas los propuso **Andrey Nikolaevich Kolmogorov**.

Axiomas de la probabilidad

Consideramos un evento A , y P un mapeo que asigna un elemento de \mathbb{R} al evento A .^a

$$P : A \subset \Omega \rightarrow \mathbb{R}$$

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $P(\bigcup_{i=1}^k A_i) = \sum_{i=1}^k P(A_i)$

Además cabe destacar que el conjunto en el que se asigna un elemento de \mathbb{R} esta restringido al intervalo $[0, 1]$

^aEs importante no confundir el mapeo P con el mapeo de una variable aleatoria X , ya que no son lo mismo, lo que principalmente los distingue es que P debe cumplir los 3 postulados vistos.

Como primer hipótesis de trabajo se considera el **postulado de iguales probabilidades a priori**(PEAPP), el dice que todos los resultados del espacio Ω tienen la misma probabilidad de que ocurran, lo cual genera que al graficar su distribución se vea como se muestra en la Figura 1.1

Ahora un concepto importante, son los **eventos mutuamente excluyentes**, lo cual nos dice lo siguiente.

$$A \cap B = \{\emptyset\} \rightarrow P(A \cap B) = 0$$

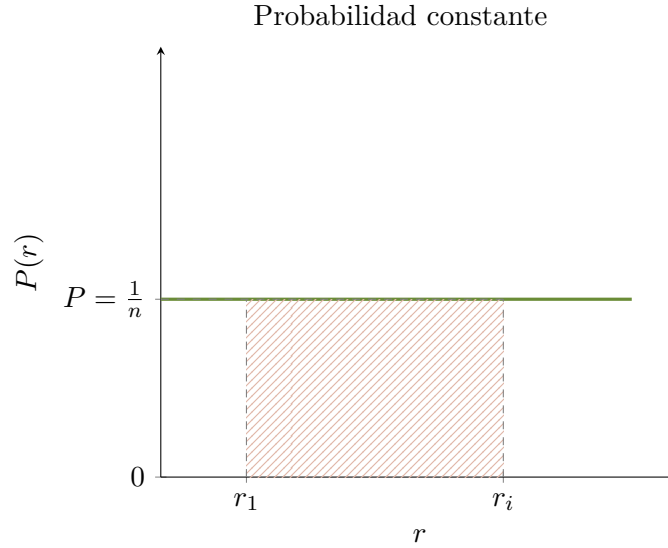


Figura 1.1: Distribución del postulado de iguales probabilidades

Probabilidad condicional

La probabilidad condicional no da una forma de asignar probabilidades a eventos **dependientes** de otros, es decir, se considera un evento A y un evento B la probabilidad de que suceda A debido a que sucede B está dado por lo siguiente.

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (1.1)$$

La ecuación (1.1), se puede reescribir como $P(A \cap B) = P(A | B)P(B)$. Por otro lado, también, con base en (1.1) se puede definir la probabilidad inversa.

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad (1.2)$$

De esta manera se puede notar algo, si la ecuación (1.2) de igual forma la reescribimos como $P(A \cap B) = P(B | A)P(A)$, entonces las ecuaciones (1.1) y (1.2), se pueden igualar, y esto nos lleva a un resultado muy importante llamado el **teorema de Bayes o fórmula de Bayes**.

Teorema de Bayes

$$P(A | B) = \frac{P(A)}{P(B)} P(B | A) \quad (1.3)$$

Esto se puede ver como un sistema de causas y consecuencias

Hay algo importante a notar, y es ¿por qué definir $P(A | B)$ en términos de $P(B | A)$ y no al revés?² y la respuesta es por la facilidad la probabilidad condicional, es más fácil conocer $P(B | A)$ que conocer $P(A | B)$, y con la finalidad de que quede más claro esta afirmación se tomara el ejemplo de enfermedad y síntomas.

Definimos lo siguiente.

²Notar que $P(A | B) \neq P(B | A)$

A :Enfermedad	(X)
B :Síntomas	(S)

Ahora con esta información se puede sustituir en el teorema de Bayes (1.3), y se obtiene lo siguiente.

$$P(X | S) = \frac{P(X)}{P(S)} P(S | X) \quad (1.4)$$

Entonces la ecuación (1.4) no dice que la probabilidad de que tengamos una enfermedad, dado que tenemos ciertos síntomas ($P(X | S)$), es proporcional a la probabilidad de tener síntomas dado que tenemos cierta enfermedad $P(S | X)$. De esta manera se puede notar que es más fácil conocer $P(S | X)$, ya que si conocemos la enfermedad, se pueden conocer los síntomas; sin embargo, en el caso contrario $P(X | S)$ es más difícil, ya que los mismos síntomas que se presentan pueden coincidir con muchas más enfermedades.

Y bueno, finalmente, de (1.3), el término $P(A)/P(B)$ es conocimiento previo que se tiene.

Fórmula de completez de la probabilidad (o probabilidad total)

$$P(B) = \sum_{i=1}^N P(B | A_i) P(A_i) \quad (1.5)$$

Sustituyendo la ecuación de la probabilidad total (1.5), en el teorema de Bayes (1.3), el teorema de Bayes se puede reescribir de la siguiente forma.

$$P(A_i | B_k) = \frac{P(A_i) P(B_k | A_i)}{\sum_{i=1}^N P(B_k | A_i) P(A_i)} \quad (1.6)$$

Esta forma del teorema de Bayes modela sistemas de causa-efecto, anteriormente se dio un ejemplo, sin embargo, en este caso son para k enfermedades, con i síntomas.

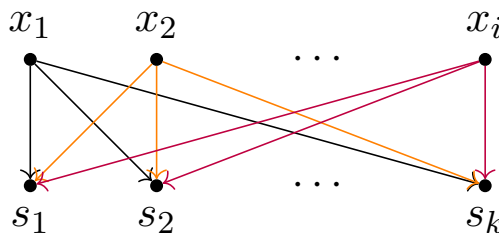


Figura 1.2: Esquema de enfermedades y síntomas, con el modelo de causa consecuencia.

En la Figura 1.3 se esquematiza un sistema con i enfermedades, dadas por x_i y k síntomas, dados por s_k , en el esquema se presentan como si estuvieran cada una de las enfermedades unidas a todos los síntomas; sin embargo, en la realidad puede haber excepciones.³

³Como comentario, este tipo de modelos de causa-efecto, es la base de las redes bayesianas.

1.2.3. Distribuciones de probabilidad.

A toda variable aleatoria se le asocia una función de distribución de probabilidad ⁴, esta función siempre existe, y esta, relaciona una probabilidad de que una variable aleatoria X adquiera un valor x y se representa de la siguiente manera.

$$f(x) = p(x) = P(X = x)$$

Además esta función debe cumplir las siguientes dos condiciones.

1. $p(x) \geq 0$
2. $\sum_x p(x) = 1$

La primera condición, nos dice que la probabilidad de que cierta variable aleatoria tome el valor x , no puede ser menor que cero, y la segunda condición nos dice que la suma de las probabilidades, de todos los valores x que pueda adquirir la variable aleatoria debe ser 1, en caso de que no cumpla esta condición, puede que no este normalizado.

Además otra interpretación importante es que $p(x)dx$ es la probabilidad de que $X \in [x, x + dx]$, de esto se puede obtener lo siguiente.

$$P(a \leq X \leq b) = \int_a^b p(x)dx \quad (1.7)$$

que es la probabilidad de que $X \in [a, b]$

1.2.4. Distribución acumulativa.

Para poder introducir el concepto de distribución acumulativa, se define una función de una variable aleatoria X la cual esta definida en el intervalo de $[0, 1] \in \mathbb{R}$.

$$F(x) = P(X \geq x) = \sum_{x_i \geq x} p(x_i) \quad (1.8)$$

Esta función lo que nos da la probabilidad de que la variable aleatoria X adquiera valores menores o igual a x , donde x esta bien definido, con esto también se nota que esta función es la suma de las probabilidades menores o iguales a x .⁵

De igual forma que en la distribución de probabilidad se puede considerar el caso continuo en lugar de la forma discreta, de la siguiente forma.

$$P(X \geq x) = \int_{-\infty}^x p(x')dx' \quad (1.9)$$

de esta forma tambien se puede definir el caso contrario.

$$P(X \leq x) = 1 - P(X \geq x) \quad (1.10)$$

6

Esta ultima igualdad es posible, ya que si lo notamos, las condiciones de ambas funciones (distribución probabilidad y acumulativa), cumplen los axiomas de probabilidad 1.2.2 propuestos por **Kolmogorov**.

⁴También se le conoce como *función de masa de probabilidad o función de densidad*

⁵Cuando sea necesario especifica de que variable aleatoria se habla, o se esta calculando tanto la función de distribución de probabilidad, como la acumulativa, la notación que se emplea es $f_X(x)$ y $F_X(x)$ respectivamente.

⁶Algo importante a mencionar es que esta igualdad solo es valida para variables aleatorias.

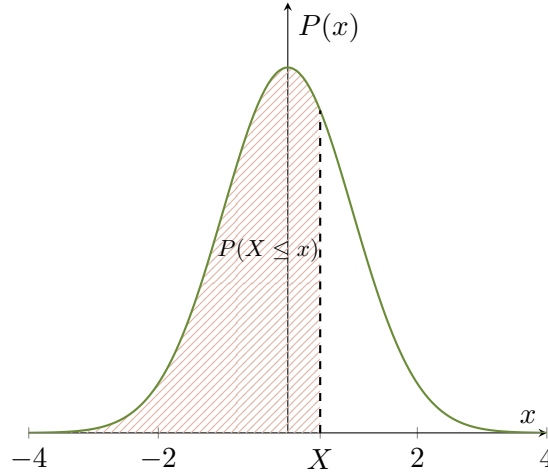


Figura 1.3: Distribución Gaussiana, la parte con relleno corresponde a $P(X \leq x)$, y la parte que no tiene relleno corresponde $P(X \geq x)$.

Todas las variables aleatorias tienen medidas, que nos ayudan a describir algunos comportamientos de estas, de los cuales los siguientes son valores típicos, o características numéricas.

- Valor más probable x^* : corresponde al máximo valor de $P(x)$.
- Mediana x_{med}
- Media: $m = \langle x \rangle = \int x p(x) dx$
- Desviación estándar \rightarrow Volatilidad.

Algo importante a mencionar es acerca del ultimo punto, desviación estándar, es una medida estadística, que nos indica que tan dispersos están los datos del valor medio (o media), sin embargo finanzas esta medida también se entiende como volatilidad, ya que si un activo, tiene una rentabilidad con una desviación estándar alta, su volatilidad es alta, lo cual no lleva a decir que es un activo con riesgo alto.

Además estas medidas también nos sirven para calcular una medida que también es relevante que son **fluctuaciones**, este concepto tiene relación con la variabilidad de los datos, que de igual forma que en el caso de la volatilidad no da una idea de la dispersión de los datos respecto al valor medio, sin embargo no son los mismos conceptos.

Otras medidas importantes son las siguientes.

- Desviación media absoluta (MAD)

$$E_{abs} = \int |x - x_{med}| p(x) dx \quad (1.11)$$

- Varianza

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \int (x - \langle x \rangle)^2 p(x) dx \quad (1.12)$$

- Momentos de orden superior ⁷

⁷Los momentos es la forma más general de medidas estadísticas, que sirven para estudiar las distribuciones de probabilidad.

$$m_n = \langle x^n \rangle = \int x^n p(x) dx \quad (1.13)$$

m_n existe si $p(x)$ decrece lo suficientemente rápido cuando $|x| \rightarrow \infty$

1.2.5. Función característica.

Si consideramos $p(x)$, se define a la función característica, como la **transformada de Fourier** de la distribución de probabilidad $p(x)$, además la función característica existe para cualquier distribución de probabilidad.

Función Característica

$$\phi(t) = \int e^{itx} p(x) dx = E(e^{itx}) \quad (1.14)$$

De igual forma se puede construir la distribución de probabilidad partiendo de la función característica, con la transformada inversa de Fourier.

$$p(x) = \frac{1}{2\pi} \int e^{-itx} \phi(t) dt \quad (1.15)$$

- Momento Ahora hacemos la siguiente consideración, $\phi(t) = \hat{p}(t)$, entonces podemos definir el n - *esimo* momento, mediante la función característica.

$$m_n = (-i)^n \frac{d^n}{dt^n} \hat{p}(t) |_{t=0} \quad (1.16)$$

donde $\hat{p}(0) = 1$

- Momentos acumulados.

$$c_n = (-i)^n \frac{d^n}{dt^n} \log \hat{p}(t) |_{t=0} \quad (1.17)$$

- Momentos acumulados normalizados.

$$\lambda_n = \frac{c_n}{\sigma^n} \quad (1.18)$$

Donde σ corresponde a la desviación.

Los momentos acumulados normalizados son lo que nos van ayudar a describir las distribuciones de probabilidad, dicho de otra forma, estos momentos nos ayudaran a estudiar los datos, con comportamiento aleatorio. Por tal motivo se presenta los cuatro primeros momentos acumulados normalizados.

$$\lambda_1 = m \quad \text{Promedio} \quad (1.19)$$

$$\lambda_2 = \frac{\sigma^2}{\sqrt{m}} \quad (1.20)$$

$$\lambda_3 = \xi = \frac{\langle (x - m)^3 \rangle}{\sigma^3} \quad \text{Skweness} \quad (1.21)$$

$$\lambda_4 = \kappa = \frac{\langle (x - m)^4 \rangle}{\sigma^4} - 3 \quad \text{Kurtosis} \quad (1.22)$$

8

Ahora ¿qué nos dicen estas medidas?. Para el caso del promedio, lo que nos da, es el valor más probable de una distribución de probabilidades. El segundo momento, corresponde a la varianza, este valor nos indica la dispersión de los datos, alrededor del valor medio (promedio), anteriormente mencionamos la desviación estándar que de igual forma nos da la medida, y no es nada extraño, ya que la desviación estándar es la raíz cuadrada de la varianza. El tercer momento, nos da una medida de la asimetría de la distribución de probabilidad $p(x)$.

Finalmente, el cuarto momento, nos da una forma de calcular que tanto se desvían la cola de $p(x)$, respecto una distribución gaussiana⁹

Una condición necesaria, para que exista el momento n –esimo es que $p(x)$ decaiga más rápido que $|x|^{-n-1}$ cuando $|x| \rightarrow \infty$, con lo cual tenemos que

$$p(x) \sim |x|^{-n-1} \rightarrow \text{Comportamiento como ley de potencia } (n+1)$$

Las **transiciones de fase** es concepto muy usado en la física, aun que no exclusivamente para la física, además este mecanismo da lugar a un comportamiento como ley de potencia.

Una de sus propiedades significativas sobre este tipo de distribuciones, es que es invariante sobre transformaciones de escala.

$$f(x) = Ax^\gamma \quad (1.23)$$

Que al aplicar una transformación de escala se obtiene lo siguiente.

$$f(\alpha x) = A\alpha^\gamma x^\gamma = \beta f(x) \quad (1.24)$$

Esto implica un comportamiento fractal o de auto-similitud, además algo importante a mencionar es que para el caso en el que el comportamiento se $f(x) = Ax^\gamma$, las medidas $\langle x \rangle$ y σ^2 son medidas no representativas, con lo cual es necesario explorar con otras medidas.

1.2.6. Teorema del limite central

Distribución Gaussiana (Normal).

La distribución Gaussiana, es de las distribuciones de probabilidad de mayor importancia, ya que esta sale en resultados importantes como lo es **Teorema del limite central**.

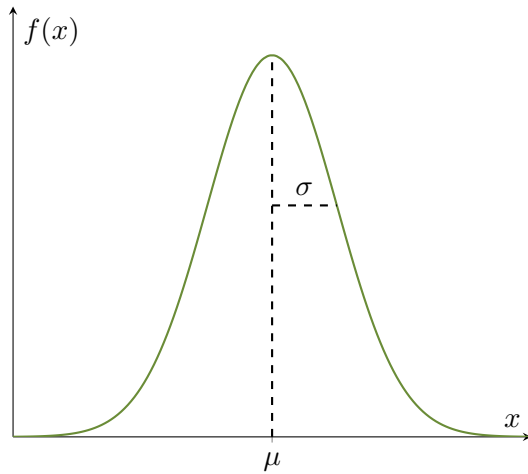
Sea X una variable aleatoria continua, se dice que esta variable aleatoria tiene una distribución normal si su función de densidad es de la siguiente forma.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad (1.25)$$

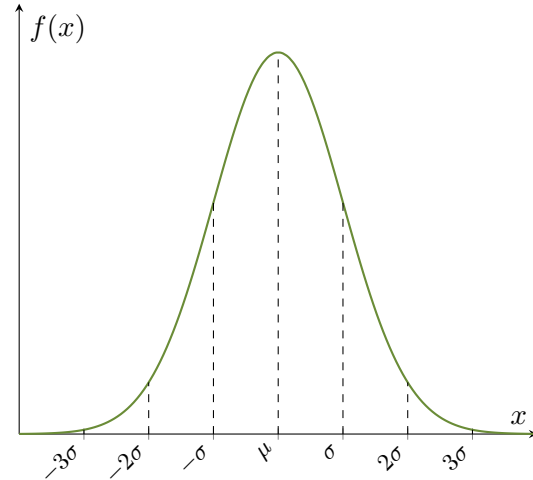
Donde $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, estos dos parámetros nos indica que μ corresponde al valor medio de la distribución de probabilidad, que de igual manera se puede interpretar como que la función de densidad esta centrada en μ y por otro lado, el parámetro σ^2 corresponde a la varianza de la función de densidad, este valor, como ya anteriormente se abordó, nos da una medida de dispersión de los datos, de manera más empírica, que tan ancha es la gráfica de la función de densidad de probabilidad.

⁸En algunos textos, al promedio, también conocido como valor medio se representa por μ , esto es debido a que los momentos en general se representan como μ_n .

⁹Se considera una distribución gaussiana como referencia, ya que esta distribución es simétrica, es decir, $\lambda_3 = 0$, al igual que $\lambda_4 = 0$



(a) Representación de las dos medidas de una distribución Gaussiana.

(b) Los distintos intervalos de σ .Figura 1.4: Función de densidad de probabilidad Gaussiana, con representación de σ .

Como anteriormente se menciona la distribución normal cuenta con dos medidas estadísticas.

$$\mu = \langle x \rangle \quad \text{Promedio} \quad (1.26)$$

$$\sigma^2 = \langle (x - \mu)^2 \rangle \quad \text{Varianza} \quad (1.27)$$

Mientras que momentos de ordenes mayores es igual a cero.

Además algo importante a destacar, es que como se muestra en la Figura 1.4(b), existen distintos intervalos, y dependiendo que intervalo se considere se tendrá cierto porcentaje de los datos. Si consideramos los siguientes intervalos

- $-\sigma \leq x \leq \sigma$ El porcentaje de probabilidad de encontrar datos con valores dentro de ese intervalo, es del 68 %.
- $-2\sigma \leq x \leq 2\sigma$ Del mismo modo, el porcentaje que le corresponde a este intervalo es de 95 %.
- $-3\sigma \leq x \leq 3\sigma$ De igual forma, el porcentaje para este intervalo es de 99,7 %

De esta manera tiene mayor sentido cuando se dice que tanto, la desviación estándar, como la varianza corresponden a la dispersión de los datos.¹⁰

Algo importante, es que la función de densidad $f(x)$ generalmente se representar como $N(\mu, \sigma^2)$.

Ahora bien, la razón por la que se le da particular importancia a la distribución Gaussiana, se relaciona con el **teorema del limite central**.

Teorema del limite central (CLT).

Sea X_1, X_2, \dots, X_n , variables aleatorias, las cuales son independiente, y además con la misma distribución de probabilidad.

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1) \quad (1.28)$$

¹⁰El porcentaje que se obtiene para cada uno de los intervalos, se obtiene mediante distribución acumulativa.

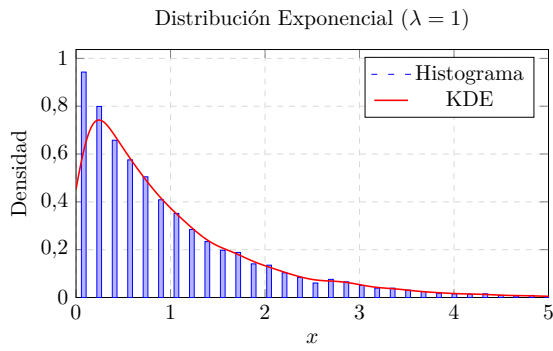
Se puede ver que del CLT 1.28, se puede reescribir de la siguiente manera.

$$\frac{(X_1 + \cdots + X_n)/n - \mu}{\sigma/\sqrt{n}}$$

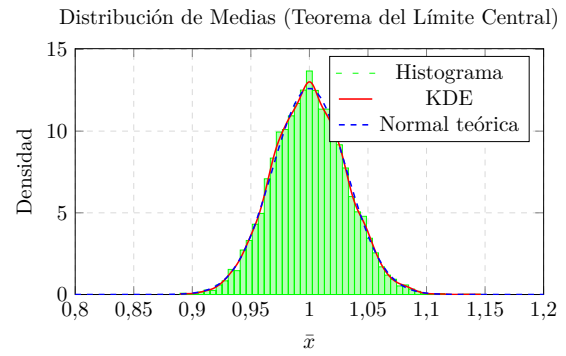
Donde se puede notar que $X_1 + \cdots + X_n/n$ corresponde al promedio, con lo cual se puede reescribir como.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

De esta manera se puede interpretar de mejor manera el CLT, el cual nos dice de manera menos abstracta, si consideramos m muestras, con tamaños n esta debe ser grande, al igual que m , regularmente se recomienda que el tamaño de muestra sea $m \geq 50$, y se calcula el promedio de estas muestra, con los promedios de las muestras son las que siguen una distribución Gaussiana.



(a) Distribución Exponencial con $\lambda = 1$.



(b) Teorema de límite central.

Figura 1.5: Demostración del teorema del límite central.

Referencias Bibliográficas

Rincón, L. (s.f.). *Introducción a la probabilidad*. Universidad Nacional Autónoma de México.

Rincón, L. (2007). *Curso intermedio de probabilidad*. Univesidad Nacional Autónoma de México.