Probabilidad y Estadística para Inteligencia Artificial

Dr. Ing. Pablo Briff

Laboratorio de Sistemas Embebidos - FIUBA

pbriff@fi.uba.ar

27 de Junio de 2020

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Tabla de Contenidos I

Tabla de Contenidos II

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Acerca del docente

- Pablo Briff

Acerca del docente

- Pablo Briff
- Ingeniero electrónico, FIUBA, 2006

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Acerca del docente

- Pablo Briff
- Ingeniero electrónico, FIUBA, 2006
- Doctor en ingeniería electrónica, FIUBA, 2015

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Acerca del docente

- Pablo Briff
- Ingeniero electrónico, FIUBA, 2006
- Doctor en ingeniería electrónica, FIUBA, 2015
- Email: pbriff@fi.uba.ar

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Programa de la Materia

- Clase 1: Repaso de probabilidad y variables aleatorias. Probabilidad conjunta, marginal, condicional. Verosimilitud. Distribuciones comunes. Teorema Central del Limite. Ley de números grandes. Ejercicios teóricos.

Programa de la Materia

- Clase 1: Repaso de probabilidad y variables aleatorias. Probabilidad conjunta, marginal, condicional. Verosimilitud. Distribuciones comunes. Teorema Central del Limite. Ley de números grandes. Ejercicios teóricos.
- Clase 2: Repaso de MATLAB/Octave para probabilidad y estadística. Producto de convolución. Suma de variables aleatorias. Funciones de variables aleatorias. Transformación de pdfs. Modelos multivariable. Ejercicios de correlación. Histograma. Generación de variables aleatorias con método inverso generalizado. Ejemplos.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Programa de la Materia

- Clase 1: Repaso de probabilidad y variables aleatorias. Probabilidad conjunta, marginal, condicional. Verosimilitud. Distribuciones comunes. Teorema Central del Limite. Ley de números grandes. Ejercicios teóricos.
- Clase 2: Repaso de MATLAB/Octave para probabilidad y estadística. Producto de convolución. Suma de variables aleatorias. Funciones de variables aleatorias. Transformación de pdfs. Modelos multivariable. Ejercicios de correlación. Histograma. Generación de variables aleatorias con método inverso generalizado. Ejemplos.
- Clase 3: Condicionamiento de variables aleatorias. Probabilidad condicional como variable aleatoria. Procesos estocásticos. Ejercicios prácticos.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Programa de la Materia

- Clase 4: Estimadores puntuales. Esperanza, varianza, sesgo y MSE del estimador. Media y varianza muestral. Máxima verosimilitud. Estimación de densidad de kernel.

Programa de la Materia

- Clase 4: Estimadores puntuales. Esperanza, varianza, sesgo y MSE del estimador. Media y varianza muestral. Máxima verosimilitud. Estimación de densidad de kernel.
- Clase 5: Estimadores por intervalo. Intervalo de confianza. Reglas de decisión. Errores tipo I y II.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Programa de la Materia

- Clase 4: Estimadores puntuales. Esperanza, varianza, sesgo y MSE del estimador. Media y varianza muestral. Máxima verosimilitud. Estimación de densidad de kernel.
- Clase 5: Estimadores por intervalo. Intervalo de confianza. Reglas de decisión. Errores tipo I y II.
- Clase 6: Enfoque Bayesiano. Maximo a posteriori.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Clase 4: Estimadores puntuales. Esperanza, varianza, sesgo y MSE del estimador. Media y varianza muestral. Máxima verosimilitud. Estimación de densidad de kernel.
- Clase 5: Estimadores por intervalo. Intervalo de confianza. Reglas de decisión. Errores tipo I y II.
- Clase 6: Enfoque Bayesiano. Maximo a posteriori.
- Clase 7: Clase práctica y repaso.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Programa de la Materia

- Clase 4: Estimadores puntuales. Esperanza, varianza, sesgo y MSE del estimador. Media y varianza muestral. Máxima verosimilitud. Estimación de densidad de kernel.
- Clase 5: Estimadores por intervalo. Intervalo de confianza. Reglas de decisión. Errores tipo I y II.
- Clase 6: Enfoque Bayesiano. Maximo a posteriori.
- Clase 7: Clase práctica y repaso.
- Clase 8: Evaluación final.

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$
- Probabilidad: noción de frecuencia o confianza en un evento, $P(E) \in [0,1]$, $P(S) = 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$
- Probabilidad: noción de frecuencia o confianza en un evento, $P(E) \in [0, 1]$, $P(S) = 1$
- En términos simples, $P(E) = \dfrac{\text{casos favorables}}{\text{casos posibles}}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$
- Probabilidad: noción de frecuencia o confianza en un evento, $P(E) \in [0, 1]$, $P(S) = 1$
- En términos simples, $P(E) = \frac{\text{casos favorables}}{\text{casos posibles}}$
- Se aplican la teoría de conjuntos: si E^c es el complemento de E, $P(E \cup E^c) = P(E) + P(E^c) = 1$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$
- Probabilidad: noción de frecuencia o confianza en un evento, $P(E) \in [0, 1]$, $P(S) = 1$
- En términos simples, $P(E) = \frac{\text{casos favorables}}{\text{casos posibles}}$
- Se aplican la teoría de conjuntos: si E^c es el complemento de E, $P(E \cup E^c) = P(E) + P(E^c) = 1$
- En general, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Axiomas de Probabilidad

- Experimento aleatorio: salida desconocida
- Una *variable aleatoria* (v.a.) es una función que asigna un número a cada salida en el espacio muestral de un experimento aleatorio
- Espacio muestral S con todas las salidas posibles, S puede ser discreto o continuo
- Eventos E, i.e. $E \subset S$
- Probabilidad: noción de frecuencia o confianza en un evento, $P(E) \in [0, 1]$, $P(S) = 1$
- En términos simples, $P(E) = \frac{\text{casos favorables}}{\text{casos posibles}}$
- Se aplican la teoría de conjuntos: si E^c es el complemento de E, $P(E \cup E^c) = P(E) + P(E^c) = 1$
- En general, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- ¿Ejemplo?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?
- El término $P(E \cap F)$ limita las posibilidades a la ocurrencia de F

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?
- El término $P(E \cap F)$ limita las posibilidades a la ocurrencia de F
- Propiedad conmutativa: $P(E \cap F) = P(E|F)P(F) = P(F|E)P(E)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?
- El término $P(E \cap F)$ limita las posibilidades a la ocurrencia de F
- Propiedad conmutativa: $P(E \cap F) = P(E|F)P(F) = P(F|E)P(E)$
- Fórmula de Bayes: $P(F|E) = \frac{P(E|F)P(F)}{P(E)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?
- El término $P(E \cap F)$ limita las posibilidades a la ocurrencia de F
- Propiedad conmutativa: $P(E \cap F) = P(E|F)P(F) = P(F|E)P(E)$
- Fórmula de Bayes: $P(F|E) = \frac{P(E|F)P(F)}{P(E)}$
- Si hay subconjuntos mutualmente excluyentes F_i tal que $U_{i=1}^{n} F_i = S$ entonces $P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_j P(E|F_j)P(F_j)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Probabilidad Condicional

- La probabilidad condicional $P(E|F)$ denota la probabilidad de que un evento E suceda dado que otro evento F ha sucedido
- $P(E|F) = \frac{P(E \cap F)}{P(F)}$
- La probabilidad condicional se puede graficar en un diagrama de Venn
- ¿Ejemplo? Que pasa si los eventos son *independientes*?
- El término $P(E \cap F)$ limita las posibilidades a la ocurrencia de F
- Propiedad conmutativa: $P(E \cap F) = P(E|F)P(F) = P(F|E)P(E)$
- Fórmula de Bayes: $P(F|E) = \frac{P(E|F)P(F)}{P(E)}$
- Si hay subconjuntos mutualmente excluyentes F_i tal que $U_{i=1}^{n} F_i = S$ entonces $P(F_i|E) = \frac{P(E|F_i)P(F_i)}{\sum_j P(E|F_j)P(F_j)}$
- Es decir, podemos calcular $P(F_i|E)$ a partir de $P(E|F_i)$ y $P(F_j)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)
- $F(\cdot)$ también es conocida como cdf (cumulative distribution function)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)
- $F(\cdot)$ también es conocida como cdf (cumulative distribution function)
- La pdf es la derivada de la cdf, $p(x) \triangleq dF(x)/dx$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)
- $F(\cdot)$ también es conocida como cdf (cumulative distribution function)
- La pdf es la derivada de la cdf, $p(x) \triangleq dF(x)/dx$
- $\int_{-\infty}^{+\infty} p(x)dx = 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)
- $F(\cdot)$ también es conocida como cdf (cumulative distribution function)
- La pdf es la derivada de la cdf, $p(x) \triangleq dF(x)/dx$
- $\int_{-\infty}^{+\infty} p(x)dx = 1$
- La cdf y la pdf son no-negativas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución de Probabilidad

- La función de distribución de probabilidad $F(\cdot)$ de una variable aleatoria X para un número x es $F(x) = P(X \leq x)$
- Para v.a. discretas, $F(x)$ es no-negativa, y monótonamente creciente de a saltos
- Para v.a. continuas, $F(x)$ es no-negativa, monótonamente creciente y continua
- Para un intervalo $[a, b]$ tenemos que $P(a < X \leq b) = F(b) - F(a)$
- Para variables discretas, $F(a) = \sum_{\forall x \leq a} P(X = x)$, donde $P(X = x)$ es la función de masa de probabilidad
- Para variables continuas, $F(x) = \int_{-\infty}^{x} p(u)du$, donde $p(\cdot)$ es la función de densidad de probabilidad (pdf, en inglés)
- $F(\cdot)$ también es conocida como cdf (cumulative distribution function)
- La pdf es la derivada de la cdf, $p(x) \triangleq dF(x)/dx$
- $\int_{-\infty}^{+\infty} p(x)dx = 1$
- La cdf y la pdf son no-negativas
- Si X es una v.a. *continua* y $x \in \mathcal{R}$, ¿cuánto vale $P(X = x)$?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Es de interés conocer la relación entre dos o más variables

Función de Distribución Conjunta y Marginal

- Es de interés conocer la relación entre dos o más variables
- Se usa la distribución conjunta para tal fin

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución Conjunta y Marginal

- Es de interés conocer la relación entre dos o más variables
- Se usa la distribución conjunta para tal fin
- La probabilidad conjunta de X, Y es $F(x, y) = P\{X \leq x, Y \leq y\}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución Conjunta y Marginal

- Es de interés conocer la relación entre dos o más variables
- Se usa la distribución conjunta para tal fin
- La probabilidad conjunta de X, Y es $F(x, y) = P\{X \leq x, Y \leq y\}$
- La distribución marginal es
  $F_X(x) = P(X \leq x) = P\{X \leq x, Y \leq \infty\} = F(x, \infty)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución Conjunta y Marginal

- Es de interés conocer la relación entre dos o más variables
- Se usa la distribución conjunta para tal fin
- La probabilidad conjunta de X, Y es $F(x, y) = P\{X \leq x, Y \leq y\}$
- La distribución marginal es $F_X(x) = P(X \leq x) = P\{X \leq x, Y \leq \infty\} = F(x, \infty)$
- Caso discreto: $P(X \leq x) = \sum_j P(x, y_j)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Distribución Conjunta y Marginal

- Es de interés conocer la relación entre dos o más variables
- Se usa la distribución conjunta para tal fin
- La probabilidad conjunta de X, Y es $F(x, y) = P\{X \leq x, Y \leq y\}$
- La distribución marginal es $F_X(x) = P(X \leq x) = P\{X \leq x, Y \leq \infty\} = F(x, \infty)$
- Caso discreto: $P(X \leq x) = \sum_j P(x, y_j)$
- Caso continuo: $p_X(x) = \int_{-\infty}^{+\infty} p(x, y)dy$

Distribuciones Condicionales

- La probabilidad condicional denota la probabilidad de ocurrencia de un evento X dado que Y ha ocurrido

Distribuciones Condicionales

- La probabilidad condicional denota la probabilidad de ocurrencia de un evento X dado que Y ha ocurrido
- Caso discreto:
  $P_{X|Y} = P\{X = x|Y = y\} = \frac{P\{X=x,Y=y\}}{P(Y=y)} = \frac{P\{X=x,Y=y\}}{P_Y(y)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- La probabilidad condicional denota la probabilidad de ocurrencia de un evento X dado que Y ha ocurrido
- Caso discreto: $P_{X|Y} = P\{X = x|Y = y\} = \frac{P\{X=x, Y=y\}}{P(Y=y)} = \frac{P\{X=x, Y=y\}}{P_Y(y)}$
- Caso continuo: $p_{X|Y} = \frac{p\{X=x, Y=y\}}{p_Y(y)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones Condicionales

- La probabilidad condicional denota la probabilidad de ocurrencia de un evento X dado que Y ha ocurrido
- Caso discreto:
  $P_{X|Y} = P\{X = x|Y = y\} = \frac{P\{X=x,Y=y\}}{P(Y=y)} = \frac{P\{X=x,Y=y\}}{P_Y(y)}$
- Caso continuo: $p_{X|Y} = \frac{p\{X=x,Y=y\}}{p_Y(y)}$
- ¿Qué sucede si X, Y son independientes?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones Condicionales

- La probabilidad condicional denota la probabilidad de ocurrencia de un evento X dado que Y ha ocurrido
- Caso discreto:
  $P_{X|Y} = P\{X = x|Y = y\} = \frac{P\{X=x,Y=y\}}{P(Y=y)} = \frac{P\{X=x,Y=y\}}{P_Y(y)}$
- Caso continuo: $p_{X|Y} = \frac{p\{X=x,Y=y\}}{p_Y(y)}$
- ¿Qué sucede si X, Y son independientes?
- ¿Qué sucede si $X = Y$?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Teorema de Probabilidad Total

- Sean $A_1, A_2, \ldots, A_n$ eventos disjuntos que forman una partición del espacio muestral

Teorema de Probabilidad Total

- Sean $A_1, A_2, \ldots, A_n$ eventos disjuntos que forman una partición del espacio muestral
- Cada posible salida está incluida una y sólo una vez en alguno de los eventos $A_1, A_2, \ldots, A_n$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Teorema de Probabilidad Total

- Sean $A_1, A_2, \ldots, A_n$ eventos disjuntos que forman una partición del espacio muestral
- Cada posible salida está incluida una y sólo una vez en alguno de los eventos $A_1, A_2, \ldots, A_n$
- Asumamos $P(A_i) > 0$. Entonces para cada evento B tenemos:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Teorema de Probabilidad Total

- Sean $A_1, A_2, \ldots, A_n$ eventos disjuntos que forman una partición del espacio muestral
- Cada posible salida está incluida una y sólo una vez en alguno de los eventos $A_1, A_2, \ldots, A_n$
- Asumamos $P(A_i) > 0$. Entonces para cada evento B tenemos:
- $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots P(A_n \cap B)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Teorema de Probabilidad Total

- Sean $A_1, A_2, \ldots, A_n$ eventos disjuntos que forman una partición del espacio muestral
- Cada posible salida está incluida una y sólo una vez en alguno de los eventos $A_1, A_2, \ldots, A_n$
- Asumamos $P(A_i) > 0$. Entonces para cada evento B tenemos:
- $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots P(A_n \cap B)$
- $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots + P(A_n)P(B|A_n)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia
- Regla de Bayes: $P(y|x) = \frac{P(x|y)P_Y(y)}{P_X(x)} = \frac{P(x|y)P_Y(y)}{\sum_y P(x|y)P_Y(y)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia
- Regla de Bayes: $P(y|x) = \frac{P(x|y)P_Y(y)}{P_X(x)} = \frac{P(x|y)P_Y(y)}{\sum_y P(x|y)P_Y(y)}$
- La forma de la p. posterior depende del numerador

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia
- Regla de Bayes: $P(y|x) = \frac{P(x|y)P_Y(y)}{P_X(x)} = \frac{P(x|y)P_Y(y)}{\sum_y P(x|y)P_Y(y)}$
- La forma de la p. posterior depende del numerador
- El denominador normaliza la p. posterior

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia
- Regla de Bayes: $P(y|x) = \frac{P(x|y)P_Y(y)}{P_X(x)} = \frac{P(x|y)P_Y(y)}{\sum_y P(x|y)P_Y(y)}$
- La forma de la p. posterior depende del numerador
- El denominador normaliza la p. posterior
- En resumen, Bayes permite refinar la p. a priori en una p. a posteriori con los datos X

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Regla de Bayes

- Supongamos que queremos calcular la probabilidad (abreviatura p.) de las v.a. discretas Y a partir de los datos de X
- $P_{Y|X}(y|x)$ es la p. posterior
- $P_{X|Y}(x|y)$ es la verosimilitud
- $P_Y(y)$ es la p. a priori
- $P_X(x)$ es la p. total de los datos, o evidencia
- Regla de Bayes: $P(y|x) = \frac{P(x|y)P_Y(y)}{P_X(x)} = \frac{P(x|y)P_Y(y)}{\sum_y P(x|y)P_Y(y)}$
- La forma de la p. posterior depende del numerador
- El denominador normaliza la p. posterior
- En resumen, Bayes permite refinar la p. a priori en una p. a posteriori con los datos X
- Se puede extender Bayes al caso continuo usando las pdf

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos
- Caso discreto: $\mu \triangleq E[X] = \sum_i x_i P(X = x_i)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos
- Caso discreto: $\mu \triangleq E[X] = \sum_i x_i P(X = x_i)$
- Caso continuo: $\mu \triangleq E[X] = \int_{-\infty}^{+\infty} xp(x)dx$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos
- Caso discreto: $\mu \triangleq E[X] = \sum_i x_i P(X = x_i)$
- Caso continuo: $\mu \triangleq E[X] = \int_{-\infty}^{+\infty} x p(x) dx$
- La esperanza es un operador lineal:
  $E[aX + bY + c] = aE[X] + bE[Y] + c$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos
- Caso discreto: $\mu \triangleq E[X] = \sum_i x_i P(X = x_i)$
- Caso continuo: $\mu \triangleq E[X] = \int_{-\infty}^{+\infty} xp(x)dx$
- La esperanza es un operador lineal: $E[aX + bY + c] = aE[X] + bE[Y] + c$
- Funciones reales: discreta $E[g(X)] = \sum_i g(x_i)P(X = x_i)$, continua $E[g(x)] = \int g(x)p(x)dx$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Esperanza

- La esperanza, valor esperado o media de una v.a. es el promedio en varios experimentos
- Caso discreto: $\mu \triangleq E[X] = \sum_i x_i P(X = x_i)$
- Caso continuo: $\mu \triangleq E[X] = \int_{-\infty}^{+\infty} x p(x) dx$
- La esperanza es un operador lineal: $E[aX + bY + c] = aE[X] + bE[Y] + c$
- Funciones reales: discreta $E[g(X)] = \sum_i g(x_i) P(X = x_i)$, continua $E[g(x)] = \int g(x) p(x) dx$
- La esperanza es el centro de masa de la distribución, y no siempre es el valor más probable (moda)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Varianza

- Sea $\mu_X = E[X]$ la media de X, la varianza de X es $\text{var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$

Varianza

- Sea $\mu_X = E[X]$ la media de X, la varianza de X es $\text{var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$
- Normalmente designaremos $\sigma_X^2 \triangleq \text{var}[X]$

Varianza

- Sea $\mu_X = E[X]$ la media de X, la varianza de X es $\text{var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$
- Normalmente designaremos $\sigma_X^2 \triangleq \text{var}[X]$
- $\text{var}[aX + b] = a^2\sigma_X^2$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Varianza

- Sea $\mu_X = E[X]$ la media de X, la varianza de X es $\text{var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$
- Normalmente designaremos $\sigma_X^2 \triangleq \text{var}[X]$
- $\text{var}[aX + b] = a^2\sigma_X^2$
- Por qué b desapareció de la expresión anterior?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Varianza

- Sea $\mu_X = E[X]$ la media de X, la varianza de X es $\text{var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2$
- Normalmente designaremos $\sigma_X^2 \triangleq \text{var}[X]$
- $\text{var}[aX + b] = a^2\sigma_X^2$
- Por qué b desapareció de la expresión anterior?
- El desvío estándar es $\sigma_X = \sqrt{\sigma_X^2}$, medido en unidades de X

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Covarianza

- La covarianza indica la relación entre dos v.a.

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
- $\text{cov}[X, X] = \text{var}[X]$

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
- $\text{cov}[X, X] = \text{var}[X]$
- $\text{cov}[X + Y, Z] = \text{cov}[X, Z] + \text{cov}[Y, Z]$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
- $\text{cov}[X, X] = \text{var}[X]$
- $\text{cov}[X + Y, Z] = \text{cov}[X, Z] + \text{cov}[Y, Z]$
- $\text{var}[X + Y] = \text{var}[X] + \text{var}[Y] + 2\text{cov}[X, Y]$

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
- $\text{cov}[X, X] = \text{var}[X]$
- $\text{cov}[X + Y, Z] = \text{cov}[X, Z] + \text{cov}[Y, Z]$
- $\text{var}[X + Y] = \text{var}[X] + \text{var}[Y] + 2\text{cov}[X, Y]$
- La correlación, o índice de correlación, se define como $\text{corr}[X, Y] = \frac{\text{cov}[X,Y]}{\sqrt{\text{var}[X]\text{var}[Y]}} \in [-1, +1]$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Covarianza

- La covarianza indica la relación entre dos v.a.
- $\text{cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
- $\text{cov}[X, X] = \text{var}[X]$
- $\text{cov}[X + Y, Z] = \text{cov}[X, Z] + \text{cov}[Y, Z]$
- $\text{var}[X + Y] = \text{var}[X] + \text{var}[Y] + 2\text{cov}[X, Y]$
- La correlación, o índice de correlación, se define como $\text{corr}[X, Y] = \frac{\text{cov}[X,Y]}{\sqrt{\text{var}[X]\text{var}[Y]}} \in [-1, +1]$
- Normalmente usamos $\rho_{XY} \triangleq \text{corr}[X, Y]$

- Si $\rho_{XY} > 0$, un cambio positivo en X influencia un cambio positivo en Y

Covarianza

- Si $\rho_{XY} > 0$, un cambio positivo en X influencia un cambio positivo en Y
- Si $\rho_{XY} < 0$, un cambio positivo en X influencia un cambio negativo en Y

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Covarianza

- Si $\rho_{XY} > 0$, un cambio positivo en X influencia un cambio positivo en Y
- Si $\rho_{XY} < 0$, un cambio positivo en X influencia un cambio negativo en Y
- Si $\rho_{XY} = 0$, un cambio en X no influencia un cambio en Y, se dice que las variables están descorrelacionadas

Covarianza

- Si $\rho_{XY} > 0$, un cambio positivo en X influencia un cambio positivo en Y
- Si $\rho_{XY} < 0$, un cambio positivo en X influencia un cambio negativo en Y
- Si $\rho_{XY} = 0$, un cambio en X no influencia un cambio en Y, se dice que las variables están descorrelacionadas
- $\rho_{XY} = 0$ no implica que X, Y sean v.a. independientes (excepto si X, Y están normalmente distribuidas)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Covarianza

- Si $\rho_{XY} > 0$, un cambio positivo en X influencia un cambio positivo en Y
- Si $\rho_{XY} < 0$, un cambio positivo en X influencia un cambio negativo en Y
- Si $\rho_{XY} = 0$, un cambio en X no influencia un cambio en Y, se dice que las variables están descorrelacionadas
- $\rho_{XY} = 0$ no implica que X, Y sean v.a. independentes (excepto si X, Y están normalmente distribuidas)
- Sin embargo, si X, Y son independientes, $\rho_{XY} = 0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A.

- Definimos el n-ésimo momento de X como $E[X^n]$

Momento de una V.A.

- Definimos el n-ésimo momento de X como $E[X^n]$
- Definimos la función generadora de momentos de X como:

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A.

- Definimos el n-ésimo momento de X como $E[X^n]$
- Definimos la función generadora de momentos de X como:
- $M_X(t) = E(e^{tX})$

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A.

- Definimos el n-ésimo momento de X como $E[X^n]$
- Definimos la función generadora de momentos de X como:
- $M_X(t) = E(e^{tX})$
- Asumimos que la esperanza anterior $E(e^{tX})$ existe (i.e. es finita) para algun intervalo de $t \in (-\delta, +\delta)$ que contiene $t = 0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Definimos el n-ésimo momento de X como $E[X^n]$
- Definimos la función generadora de momentos de X como:
- $M_X(t) = E(e^{tX})$
- Asumimos que la esperanza anterior $E(e^{tX})$ existe (i.e. es finita) para algun intervalo de $t \in (-\delta, +\delta)$ que contiene $t = 0$
- Para calcular $E[X^n]$, tomamos la derivada de $M_X(t)$ respecto de t y la evaluamos en $t = 0$, es decir:

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Momento de una V.A.

- Definimos el n-ésimo momento de X como $E[X^n]$
- Definimos la función generadora de momentos de X como:
- $M_X(t) = E(e^{tX})$
- Asumimos que la esperanza anterior $E(e^{tX})$ existe (i.e. es finita) para algun intervalo de $t \in (-\delta, +\delta)$ que contiene $t = 0$
- Para calcular $E[X^n]$, tomamos la derivada de $M_X(t)$ respecto de t y la evaluamos en $t = 0$, es decir:
- $E[X^n] = \frac{d^n M_X(t)}{dt^n}(0) \triangleq M_X^{(n)}(0)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ
- Usando las propiedades de esperanza de una función $g(X)$ de una v.a. X con distribución conocida, tenemos:

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ
- Usando las propiedades de esperanza de una función $g(X)$ de una v.a. X con distribución conocida, tenemos:
- $M_X(t) = E[e^{tX}] = \int_0^{\infty} e^{tx}\,\lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ
- Usando las propiedades de esperanza de una función $g(X)$ de una v.a. X con distribución conocida, tenemos:
- $M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}\, \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$
- $E[X] = M_X^{(1)}(0) = \frac{1}{\lambda}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ
- Usando las propiedades de esperanza de una función $g(X)$ de una v.a. X con distribución conocida, tenemos:
- $M_X(t) = E[e^{tX}] = \int_0^{\infty} e^{tx}\, \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$
- $E[X] = M_X^{(1)}(0) = \frac{1}{\lambda}$
- $E[X^2] = M_X^{(2)}(0) = \frac{2}{\lambda^2}$

Momento de una V.A. (Cont.)

- Encontrar $M_X(t)$ de una v.a. X con distribución exponencial y parámetro λ
- Usando las propiedades de esperanza de una función $g(X)$ de una v.a. X con distribución conocida, tenemos:
- $M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx}\, \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$
- $E[X] = M_X^{(1)}(0) = \frac{1}{\lambda}$
- $E[X^2] = M_X^{(2)}(0) = \frac{2}{\lambda^2}$
- $\mathrm{var}[X] = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ley Débil de los Grandes Números

- Sea $X = \{X_t\}_{t=1}^{N}$ un conjunto de v.a. independedientes e idénticamente distribuidas (i.i.d.) cada una con media μ y varianza σ^2, para $\epsilon > 0$ se tiene que

Ley Débil de los Grandes Números

- Sea $X = \{X_t\}_{t=1}^{N}$ un conjunto de v.a. independedientes e idénticamente distribuidas (i.i.d.) cada una con media μ y varianza σ^2, para $\epsilon > 0$ se tiene que
- $P\{|\frac{\Sigma_t X_t}{N} - \mu| > \epsilon\} \to 0$ cuando $N \to \infty$

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Ley Débil de los Grandes Números

- Sea $X = \{X_t\}_{t=1}^{N}$ un conjunto de v.a. independedientes e idénticamente distribuidas (i.i.d.) cada una con media μ y varianza σ^2, para $\epsilon > 0$ se tiene que
- $P\{|\frac{\Sigma_t X_t}{N} - \mu| > \epsilon\} \to 0$ cuando $N \to \infty$
- Es decir, N ensayos convergen a la media con N creciente

Teorema Central del Límite

- Sean $X_1, X_2, \ldots, X_N$ un conjunto de v.a. i.i.d. cada una con media μ y varianza σ^2; para N grande se tiene que

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Teorema Central del Límite

- Sean $X_1, X_2, \ldots, X_N$ un conjunto de v.a. i.i.d. cada una con media μ y varianza σ^2; para N grande se tiene que
- $X_1 + X_2 + \ldots + X_N \sim \mathcal{N}(N\mu, N\sigma^2)$

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Teorema Central del Límite

- Sean $X_1, X_2, \ldots, X_N$ un conjunto de v.a. i.i.d. cada una con media μ y varianza σ^2; para N grande se tiene que
- $X_1 + X_2 + \ldots + X_N \sim \mathcal{N}(N\mu, N\sigma^2)$
- Es decir, la suma de N v.a. i.i.d. converge a una distribución normal con media y varianza igual a la suma de cada una

Teorema Central del Límite

- Sean $X_1, X_2, \ldots, X_N$ un conjunto de v.a. i.i.d. cada una con media μ y varianza σ^2; para N grande se tiene que
- $X_1 + X_2 + \ldots + X_N \sim \mathcal{N}(N\mu, N\sigma^2)$
- Es decir, la suma de N v.a. i.i.d. converge a una distribución normal con media y varianza igual a la suma de cada una
- ¿Cuál será, aproximadamente, la distribución de $\frac{X_1+X_2+\ldots+X_N}{N}$?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso
- $P\{X = i\} = p^i(1 - p)^{1-i}$, $i = \{0, 1\}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso
- $P\{X = i\} = p^i(1 - p)^{1-i}$, $i = \{0, 1\}$
- $E[X] = p$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso
- $P\{X = i\} = p^i(1 - p)^{1-i}$, $i = \{0, 1\}$
- $E[X] = p$
- $\text{var}[X] = p(1 - p)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso
- $P\{X = i\} = p^i(1-p)^{1-i}$, $i = \{0, 1\}$
- $E[X] = p$
- $\text{var}[X] = p(1-p)$
- Binomial: N i.i.d. v.a. Bernoulli, X es la p. de i éxitos en N tiradas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Bernoulli y Binomial

- Bernoulli: Ensayo de éxito (1) o fracaso (0)
- La v.a. del proceso toma los valores $X = \{0, 1\}$
- $P\{X = 1\} = p$ es la p. de éxito, $P\{X = 0\} = 1 - p$ la p. de fracaso
- $P\{X = i\} = p^i(1 - p)^{1-i}$, $i = \{0, 1\}$
- $E[X] = p$
- $\text{var}[X] = p(1 - p)$
- Binomial: N i.i.d. v.a. Bernoulli, X es la p. de i éxitos en N tiradas
- $P\{X = i\} = \binom{N}{i} p^i(1 - p)^{N-i}$, $i = \{0, 1, \dots, N\}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Uniforme

- Una v.a. uniforme X en el intervalo cerrado $[a, b]$ tiene pdf:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Uniforme

- Una v.a. uniforme X en el intervalo cerrado $[a, b]$ tiene pdf:
- $p(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{otro } x \end{cases}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Uniforme

- Una v.a. uniforme X en el intervalo cerrado $[a, b]$ tiene pdf:
- $p(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{otro } x \end{cases}$
- $E[X] = \frac{a+b}{2}$

Distribución Uniforme

- Una v.a. uniforme X en el intervalo cerrado $[a, b]$ tiene pdf:
- $p(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{otro } x \end{cases}$
- $E[X] = \frac{a+b}{2}$
- $\text{var}[X] = \frac{(b-a)^2}{12}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \ -\infty < x < +\infty$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \; -\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:
- $p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:
- $p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$
- La cdf es $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} p_Z(z)$

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:
- $p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$
- La cdf es $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} p_Z(z)$
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:
- $p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$
- La cdf es $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} p_Z(z)$
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
- Normalización: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Normal o Gaussiana

- Una v.a. normal o Gaussiana con media μ y varianza σ^2, denotada $\mathcal{N}(\mu, \sigma^2)$ tiene una pdf:
- $p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$, $-\infty < x < +\infty$
- La normal cero-uno $\mathcal{Z} = \mathcal{N}(0, 1)$ es:
- $p_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$
- La cdf es $\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} p_Z(z)$
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$
- Normalización: Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$
- El teorema central del límite se puede aplicar aproximar la suma de v.a. i.i.d. como una variable normal

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Si Z_i son v.a. i.i.d $\mathcal{N}(0,1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \ldots + Z_n^2$ es:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Si Z_i son v.a. i.i.d $\mathcal{N}(0,1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \ldots + Z_n^2$ es:
- $X \sim \mathcal{X}_n^2$, donde $\mathcal{X}_n^2$ es una chi-cuadrado de orden n

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Distribución Chi-Cuadrado

- Si Z_i son v.a. i.i.d $\mathcal{N}(0,1)$, la distribución de la suma de los cuadrados $X = Z_1^2 + Z_2^2 + \ldots + Z_n^2$ es:
- $X \sim \mathcal{X}_n^2$, donde $\mathcal{X}_n^2$ es una chi-cuadrado de orden n
- Usaremos la distribución chi-cuadrado para ensayar tests de hipótesis de varianza, y tests de bondad de ajuste (en inglés, goodness of fit) usando el teorema de Pearson

Distribución t de Student

- Si $Z \sim \mathcal{Z}$ y $X \sim \mathcal{X}_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad

Distribución t de Student

- Si $Z \sim \mathcal{Z}$ y $X \sim \mathcal{X}_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad
- $E[T_n] = 0,\ n > 1$

Distribución t de Student

- Si $Z \sim \mathcal{Z}$ y $X \sim \mathcal{X}_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad
- $E[T_n] = 0,\ n > 1$
- $\text{var}[T_n] = \frac{n}{n-2}$ para$n > 2$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{Z}$ y $X \sim \mathcal{X}_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad
- $E[T_n] = 0,\ n > 1$
- $\text{var}[T_n] = \frac{n}{n-2}$ para$n > 2$
- T_n se parece a $\mathcal{Z}$ cuando $n \to \infty$ pero con colas más notorias (mayor variabilidad que la normal cero-uno)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución t de Student

- Si $Z \sim \mathcal{Z}$ y $X \sim \mathcal{X}_n^2$ son v.a. independientes entonces el cociente $T_n = \frac{Z}{\sqrt{X/n}}$ tiene distribución t con n grados de libertad
- $E[T_n] = 0,\ n > 1$
- $\text{var}[T_n] = \frac{n}{n-2}$ para$n > 2$
- T_n se parece a $\mathcal{Z}$ cuando $n \to \infty$ pero con colas más notorias (mayor variabilidad que la normal cero-uno)
- Usaremos la distribución t de Student para ensayar un test de hipótesis con media y varianza desconocidas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución F

- Si $X_1 \sim \mathcal{X}_n^2$ y $X_2 \sim \mathcal{X}_m^2$ son v.a. independientes con distribución chi-cuadrado con grados de libertad n y m, entonces el cociente:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución F

- Si $X_1 \sim \mathcal{X}_n^2$ y $X_2 \sim \mathcal{X}_m^2$ son v.a. independientes con distribución chi-cuadrado con grados de libertad n y m, entonces el cociente:
- $F_{n,m} = \frac{X_1/n}{X_2/m}$ tiene distribución F con grados de libertad n, m

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución F

- Si $X_1 \sim \mathcal{X}_n^2$ y $X_2 \sim \mathcal{X}_m^2$ son v.a. independientes con distribución chi-cuadrado con grados de libertad n y m, entonces el cociente:
- $F_{n,m} = \frac{X_1/n}{X_2/m}$ tiene distribución F con grados de libertad n, m
- Usaremos la distribución F de Snedecor para ensayar un análisis de varianza

Ejercicio 1

a. ¿Cuál es la probabilidad de obtener exactamente 3 cecas en 10 tiradas de una moneda balanceada?
b. ¿Cuál es la probabilidad de obtener al menos 3 cecas en 10 tiradas de una moneda cargada donde la probabilidad de ceca es 0.4?
c. Una moneda cargada con probabilidad de ceca 0.4 es arrojada al aire. El resultado es cara. ¿Cuál es la probabilidad de obtener al menos 3 cecas en las próximas 10 tiradas?
d. Simular los items anteriores en Octave usando: i) una distribución uniforme con la función `rand` para simular el proceso Bernoulli; ii) las funciones `binopdf`,`binocdf` para calcular la probabilidad binomial (Octave statistics package).

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 2

(a) En una competencia se tienen 3 puertas de las cuales se debe elegir una. Dos puertas tienen la foto de un chancho y la tercera tiene la foto de un automóvil. Si el participante acierta la puerta del automóvil, lo gana. En caso contrario, no gana nada. Una vez que el participante elige una puerta, aún con todas las puerta cerradas, el organizador de la competencia - que sabe en qué puerta se encuentra la foto del automóvil - abre una de las tres puertas que tiene la foto de un chancho. ¿Qué le conviene hacer al participante, cambiar su elección o no? Justificar usando probabilidad a priori y probabilidad condicional.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 3

Sean X, Y dos v.a. i.i.d. $U[0, 1]$. Encontrar la expresión de la función de densidad de probabilidad conjunta.

a) ¿Cuál es la probabilidad de que $X > 0.7$ y $Y < 0.4$ simultáneamente?
b) ¿Cuál es el percentil 40 de X, i.e. x_{40} ?
c) Simular en Octave escribiendo la pdf conjunta o usando funciones de Octave `rand,unifpdf,unifcdf`
d) Graficar la cdf en función de x, y

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 4

Sean X, Y dos v.a. i.i.d. $\mathcal{N}(0,1)$. Encontrar la expresión de la función de densidad de probabilidad conjunta.

a. ¿Cuál es la probabilidad de que $X > 0.7$ y $Y < 0.4$ simultáneamente?
b. Simular en Octave escribiendo la pdf conjunta o usando funciones de Octave `mvnrnd`,`mvnpdf`,`mvncdf`
c. Graficar la cdf en función de x, y

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Bibliografía I

E. Alpaydin, *Introduction to machine learning*.
MIT press, 2020.

"MIT - Lecture Notes Course 6.041-6.431, Fall 2000."
https://vfu.bg/en/e-Learning/Math--Bertsekas_Tsitsiklis_Introduction_to_probability.pdf.
Accessed: 2020-05-15.

"MIT Mathematics Statistics for applications."
https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/.
Accessed: 2020-05-05.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires