Teoría de Probabilidad

Cristian Guarnizo-Lemus
cristianguarnizo@itm.edu.co

Maestria en Automatización y Control Industrial

Experimento aleatorio

- Se define un experimento aleatorio como aquel experimento cuyo resultado no se conoce con anterioridad.
- El enfoque matemático que se emplea para estudiar los resultados de un experimento aleatorio o de un fenónemo aleatorio se conoce como teoría de la probabilidad.

Contenido

Conjuntos

- Un conjunto se define como una colección de elementos.
- **Notación**. Letras mayúsculas denotan conjuntos $A, \ B, \ldots$. Letras minúsculas denotan miembros o elementos de un conjunto $a, \ b, \ldots$.
- El símbolo $\in$ denota pertenencia de un elemento a un conjunto. Por ejemplo, $a \in A$, se lee "el elemento a pertenece al conjunto A".
- El simbolo $\notin$ se lee como "no pertenece a".

Más conjuntos

- Conjunto vacío o nulo. Es un conjunto que no tiene elementos. Se denota como $\emptyset$.
- Conjunto universal. Es un conjunto que contiene a todos los conjuntos bajo consideración.
- Conjunto contable. Si los elementos del conjunto pueden ponerse en correspondencia uno a uno con los enteros.
- Conjunto finito. Conjunto contable con un número finito de elementos.

Relaciones entre conjuntos

- **Subconjunto**. $A \subset B$
  Cada elemento de A también es un elemento de B.

- **Igualdad de conjuntos**. Dos conjuntos son iguales si contienen exactamente los mismos elementos.

Relaciones entre conjuntos

- **Unión.** $A \cup B$.
  Conjunto de todos los elementos que pertenecen a A o pertenecen a B o a ambos.
  Si se tienen N conjuntos

$$A_1 \cup A_2 \cup \ldots A_N = \bigcup_{i=1}^{N} A_i$$

- **Intersección.** $A \cap B$
  Conjunto de todos los elementos que pertenecen a A y a B. Si se tienen N conjuntos

$$A_1 \cap A_2 \cap \ldots A_N = \bigcap_{i=1}^{N} A_i$$

Más relaciones entre conjuntos

- **Mutuamente excluyentes.** Se dice que dos conjuntos son mutuamente excluyentes si no tienen elementos comunes. Los conjuntos A y B son mutuamente excluyentes si

  $$A \cap B = \emptyset$$

  Los conjuntos $A_1,\ A_2, \ldots, A_N$ son mutuamente excluyentes si

  $$A_i \cap A_j = \emptyset,\ \forall i, j \text{ con } i \neq j$$

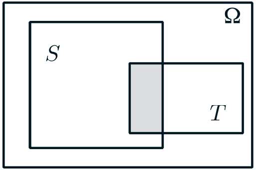
- **Complemento.** El complemento, $\overline{A}$, de un conjunto A relativo a S se define como el conjunto de todos los elementos de S que no están en A.

Álgebra de Conjuntos

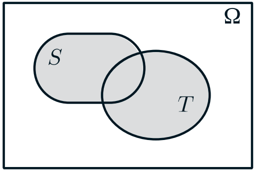
Las operaciones con conjuntos tienes varias propiedades, que son consecuencias elementales de las definiciones. algunos ejemplos son:

$$
\begin{aligned}
A \cup B &= B \cup A, & A \cup (B \cup C) &= (A \cup B) \cup C \\
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), & A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\
\overline{(\overline{A})} &= A, & A \cap \overline{A} &= \emptyset \\
A \cup S &= S, & A \cap S &= A
\end{aligned}
$$

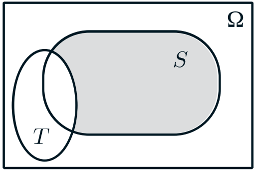
Ejemplos conjuntos



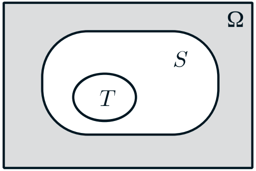
(a)



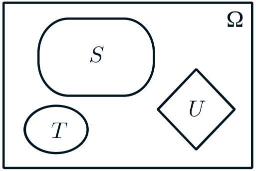
(b)



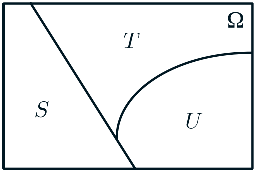
(c)



(d)



(e)



(f)

Contenido

Espacio muestral

- Desde el punto de vista de la teoría de probabilidad, los elementos del **conjunto universal** S son los resultados de un experimento.
- Un **experimento** es una secuencia de acciones que produce unos resultados.
- La totalidad de todos los posibles resultados que produce un experimento se conoce como el **espacio muestral**.
- El espacio muestral es entonces el conjunto universal S.
- Un **evento** se define como una colección de resultados de un experimento.
- Un evento es entonces un subconjunto (por ejemplo, A) del espacio muestral S.

Contenido

Preliminares

- La probabilidad de un evento A ($P(A)$) es un número que se asigna al evento A.
- Existen diferentes formas en las que probabilidades pueden asignarse a los subconjuntos de un espacio muestral.
- La forma coherente de hacerlo se realiza a través de la definición de una *medida de probabilidad*.

Medida de Probabilidad

- **Medida de probabilidad.** Consiste en asignar a los posibles eventos medibles del espacio muestral S un valor en el intervalo $[0, 1]$ que satisface
  1. (Normalización) $P(S) = 1$.
  2. (No-negativo) $P(A) \geq 0$ para todo $A \in S$.
  3. (Adición) Si $A_1, A_2, \ldots$ es una colección de eventos mutuamente excluyentes (ó disyuntos) en S, luego

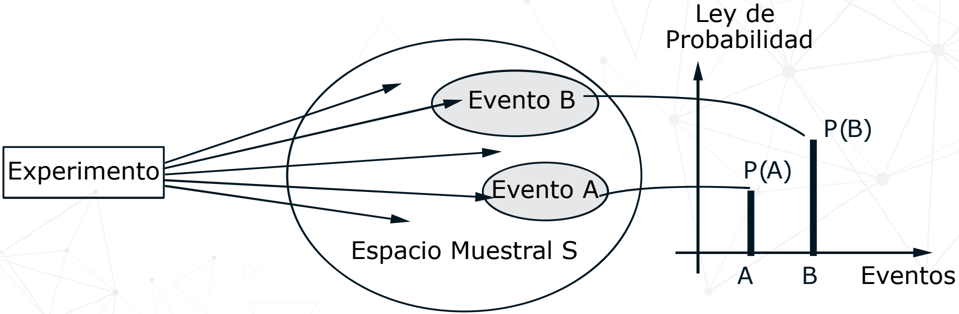
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Espacio de Probabilidad

- Es posible asignarle a los resultados de un experimento aleatorio un espacio de probabilidad (S, P).
- **Ejemplo.** Una moneda, posiblemente parcializada, se tira al aire una vez. Se puede asumir que $S = \{H, T\}$ y que $\mathcal{F} = \{\emptyset,\ H,\ T,\ S\}$. Adicionalmente, una posible medida de probabilidad $P : \mathcal{F} \leftarrow [0, 1]$ está dada como

$$P(\emptyset) = 0, \quad P(H) = \rho, \quad P(T) = 1 - \rho, \quad P(S) = 1.$$

Espacio muestral



Ley de la probabilidad discreta

- Si el espacio muestral consiste en un número finito de posibles resultados, entonces la ley de la probabilidad esta especificada por las probabilidades de los eventos que consisten en un solo elemento. En particular, la probabilidad de cualquier evento $\{s_1, s_2, \ldots, s_n\}$ es la suma de las probabilidades de sus elementos:

$$P(\{s_1, s_2, \ldots, s_n\}) = P(s_1) + P(s_2) + \cdots + P(s_n).$$

Definición de probabilidad

- **Frecuencia relativa.** Un experimento aleatorio se repite n veces. Si el evento A ocurre n_A veces, luego su probabilidad $P(A)$ se define como el límite de la frecuencia relativa nA/n de la ocurrencia de A

$$P(A) \triangleq \lim_{n\to\infty} \frac{n_A}{n}$$

- **Definición clásica.** En esta definición, la probabilidad $P(A)$ de un evento se calcula como la relación entre el número de casos favorables, N_A, sobre el número de posibilidades, N,

$$P(A) \triangleq \frac{N_A}{N}.$$

Asume que todos los eventos son igualmente probables.

Leyes útiles de probabilidad

Dada una definición de probabilidad que satisfaga los tres axiomas de una medida de probabilidad, se pueden establecer las siguientes relaciones

1. $P(\emptyset) = 0$.
2. Para cualquier evento arbitrario A, $P(A) \leq 1$.
3. Si $A \cup \overline{A} = S$ y $A \cap \overline{A} = \emptyset$, $P(\overline{A}) = 1 - P(A)$.
4. Si $A \subset B$, luego $P(A) \leq P(B)$.
5. Si $A \cup B = P(A) + P(B) - P(A \cap B)$.
6. $P(A \cup B) \leq P(A) + P(B)$.
7. $P(A \cup B \cup C) = P(A) + P(\overline{A} \cap B) + P(\overline{A} \cap \overline{B} \cap C)$.

Leyes útiles de probabilidad

Si $A_1,\ A_2, \ldots,\ A_n$ son eventos aleatorios tales que

$$A_i \cap A_j = \emptyset,\ \forall i \neq j$$
$$A_1 \cup A_2 \cup \ldots A_n = S,$$

luego $P(A) = P(A \cap A_1) + P(A \cap A_2) + \ldots + P(A \cap A_n)$. Se dice que los conjuntos son mutuamente excluyentes y exhaustivos, si se cumplen las dos condiciones anteriores.

Contenido

Producto cartesiano

1. Sea E un experimento aleatorio que se compone de dos subexperimentos E_1 y E_2.
2. El espacio muestral S_1 asociado a E_1 se compone de las salidas $a_1, \ a_2, \ \ldots, \ a_{n_1}$.
3. El espacio muestral S_2 asociado a E_2 se compone de las salidas $b_1, \ b_2, \ldots, \ b_{n_2}$.
4. El espacio muestral S del experimento combinado es el producto Cartesiano de S_1 y S_2. Esto es

$$\begin{aligned} S &= S_1 \times S_2 \\ &= \{(a_i, b_j) : i = 1, \ldots, n_1; j = 1, \ldots, n_2\}. \end{aligned}$$

5. Se pueden definir medidas de probabilidad sobre S_1, S_2 y $S = S_1 \times S_2$. Si los eventos $A_1, A_2, \ldots, A_n$ están definidos sobre el subexperimento E_1, y los eventos $B_1, B_2, \ldots, B_m$ están definidos sobre el subexperimento E_2, el evento A_iB_j es un evento sobre el experimento total.

Probabilidad conjunta o marginal

1. **Probabilidad conjunta.** La probabilidad del evento $A_i \cap B_j$ se conoce como *probabilidad conjunta* y se denota $P(A_i \cap B_j)$ ó $P(A_iB_j)$.
2. **Probabilidad marginal.** Si los eventos $A_1,\ A_2, \ldots,\ A_n$ asociados con el subexperimento E_1 son mutuamente excluyentes y exhaustivos, luego

$$
\begin{aligned}
P(B_j) = P(B_j \cap S) &= P\left[B_j \cap (A_1 \cup A_2 \cup \ldots \cup A_n)\right] \\
&= \sum_{i=1}^{n} P(A_iB_j).
\end{aligned}
$$

La probabilidad $P(B_j)$ se conoce como la probabilidad marginal.

Probabilidad condicional

- Con frecuencia, la probabilidad de ocurrencia de un evento B podría depender de la ocurrencia de un evento relacionado A.
- La probabilidad condicional $P(B|A)$ se define como

$$p(B|A) \triangleq \frac{P(AB)}{P(A)}.$$

- Empleando la definición clásica de probabilidad,

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{N_{AB}}{N_A},$$

donde N_{AB} es el numero de casos favorables del evento $A \cap B$, y N_A el numero de casos favorables del evento A.

Relaciones entre prob. conj. marg. y cond.

- $P(AB) = P(A|B)P(B) = P(B|A)P(A)$
- Si $AB = \emptyset$, luego $P(A \cup B|C) = P(A|C) + P(B|C)$.
- $P(ABC) = P(A)P(B|A)P(C|AB)$.
- Si $B_1, B_2, \ldots, B_m$ forman un conjunto de eventos mutuamente excluyentes y exhaustivos, luego

$$P(A) = \sum_{j=1}^{m} P(A|B_j)P(B_j).$$

Teorema de Bayes

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^{m} P(A|B_j)P(B_j)}.$$

Independencia

Dos eventos A_i y B_j son independientes si

$$P(A_i B_j) = P(A_i)P(B_j), \text{ ó } P(A_i|B_j) = P(A_i).$$

Ejercicio

Demostrar que

1. Si $A_1 \subset A_2$ entonces $P(A_1) \leq P(A_2)$ y $P(A_2 - A_1) = P(A_2) - P(A_1)$.
2. $P(\overline{A}) = 1 - P(A)$.
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Contenido

Definición

- Una variable aleatoria es una función cuyo dominio es el conjunto de salidas $\lambda \in S$, y cuyo rango es la recta real $\mathbb{R}$.
- Para cada salida $\lambda \in S$, la variable aleatoria asigna un número, $X(\lambda)$ tal que
  1. El conjunto $\{\lambda : X(\lambda) \leq x\}$ es un evento para todo $x \in \mathbb{R}$.
  2. Las probabilidades de los eventos $\{\lambda : X(\lambda) = -\infty\}$ y $\{\lambda : X(\lambda) = \infty\}$ es igual a cero, esto es
  $$P(X = \infty) = P(X = -\infty) = 0.$$
  3. La variable aleatoria X induce una medida de probabilidad sobre la recta real como sigue
  $$P(X = x) = P\{\lambda : X(\lambda) = x\}, P(X \leq x) = P\{\lambda : X(\lambda) \leq x\},$$
  $$P(x_1 < X \leq x_2) = P\{\lambda : x_1 < X(\lambda) \leq x_2\}.$$
  4. Para todo efecto práctico se puede asumir que una variable aleatoria es una función que mapea el espacio muestral a la recta real, $X : S \to \mathbb{R}$.

Contenido

Función de distribución

- La probabilidad $P(X \leq x)$ también se denota como $F_X(x)$, que se conoce como la función de distribución de la variable aleatoria X.
- La función de distribución tiene las siguientes propiedades
  1. $F_X(x = -\infty) = 0$.
  2. $F_X(x = \infty) = 1$.
  3. $F_X(x_1) \leq F_X(x_2)$ si $x_1 < x_2$.
  4. $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$.

Función de distribución conjunta

- La probabilidad $P(A \cap B)$ se definió anteriormente como la probabilidad de ocurrencia conjunta de los dos eventos A y B.
- Si el evento A es el evento $(X \leq x)$ y el evento B es el evento $(Y \leq y)$, la probabilidad conjunta se conoce como la *función de distribución conjunta* de las variables aleatorias X y Y,

$$F_{X,Y}(x,y) = P\left[(X \leq x) \cap (Y \leq y)\right].$$

- A partir de esta definición se derivan los siguientes resultados

$$F_{X,Y}(-\infty,-\infty) = 0, \quad F_{X,Y}(-\infty,y) = 0, \quad F_{X,Y}(\infty,y) = F_Y(y)$$
$$F_{X,Y}(x,-\infty) = 0, \quad F_{X,Y}(\infty,\infty) = 1, \quad F_{X,Y}(x,\infty) = F_X(x).$$

Contenido

Variables aleatorias discretas y continuas

- Una variable aleatoria puede ser discreta o continua.
- Una *variable aleatoria discreta* toma valores de un número contable de valores diferentes.
- Ejemplos incluyen el número de llamadas telefónicas que recibe una oficina en un intervalo finito de tiempo ó el puntaje de un estudiante en un examen.
- Una *variable aleatoria continua* puede asumir cualquier valor dentro de uno o más intervalos de la recta real.
- El tiempo exacto en que se escribe una llamada telefónica es un ejemplo de una variable aleatoria continua.

Contenido

Función de probabilidad de masa

- Una variable aleatoria discreta está caracterizada por un conjunto permitido de valores $x_1, x_2, \ldots, x_n$ junto con las probabilidades que la variable aleatoria tome esos valores.

- La probabilidad de $X = x_i$ se denota como $P(X = x_i)$ para $i = 1, \ldots, n$, y se conoce como la *función de probabilidad de masa*.

Propiedades de la función de probabilidad de masa

Las siguientes son algunas propiedades de la función de probabilidad de masa

1. $P(X = x_i) \geq 0$, $i = 1, \ldots, n$.
2. $\sum_{i=1}^{n} P(X = x_i) = 1$.
3. $P(X \leq x) = F_X(x) = \sum_{\forall x_j \leq x} P(X = x_i)$.
4. $P(X = x_i) = \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} [F_X(x_i) - F_x(x_i - \epsilon)]$

Dos variables aleatorias

- Considérese dos variables aleatorias X y Y que toman valores $x_1, x_2, \ldots, x_n$ y $y_1, y_2, \ldots, y_m$.

- Estas dos variables aleatorias pueden caracterizarse por una función de masa de probabilidad conjunta $P(X = x_i, Y = y_j)$ que da la probabilidad que $X = x_i$ y $Y = y_j$.

Funciones de masa de prob. conj., marg. y cond.

- **Conjunta**

$$P(X \leq x, Y \leq y) = \sum_{x_i \leq x} \sum_{y_i \leq y} P(X = x_i, Y = y_j).$$

- **Marginal**

$$P(X = x_i) = \sum_{j=1}^{m} P(X = x_i, Y = y_j).$$

- **Condicional**

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}, \text{ con } P(Y = y_j) \neq 0.$$

- **Teorema de Bayes**

$$P(X = x_i | Y = y_j) = \frac{P(Y = y_j | X = x_i) P(X = x_i)}{\sum_{i=1}^{n} P(Y = y_j | X = x_i) P(X = x_i)}.$$

Independencia estadística

Dos variables aleatorias son *independientes estadísticamente* si

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j), \; i = 1, \ldots, n. \; j = 1, \ldots, m.$$

Contenido

Preliminares

- La descripción más detallada posible de una variable aleatoria está dada por su función de masa de probabilidad.

- En algunas ocasiones es mejor describir la variable aleatoria usando un conjunto de números característicos o de medidas, que son representativos de la función de masa de probabilidad.

- Esto números se conocen como *valores esperados* o *promedios estadísticos*.

Valor esperado y primeros momentos estadísticos

- El valor esperado o promedio de una función $g(X)$ de una variable aleatoria discreta X está definida como

$$E\left\{g(X)\right\} = \sum_{i=1}^{n} g(x_i)P(X = x_i).$$

- Dos valores esperados o momentos de la variable aleatoria X, que se usan con frecuencia en la práctica son su media μ_X y su varianza σ_X^2,

$$E\{X\} = \mu_X = \sum_{i=1}^{n} x_i P(X = x_i),$$

$$E\left\{(X - \mu_X)^2\right\} = \sigma_X^2 = \sum_{i=1}^{n} (x_i - \mu_X)^2 P(X = x_i).$$

- La raíz cuadrada de la varianza se conoce como la *desviación estándar*.

Valor esperado de dos variables aleatorias

- El valor esperado de una función de dos variables aleatorias se define como

$$E\{g(X,Y)\} = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i,y_i)P(X=x_i, Y=y_j).$$

- Un valor esperado útil, que entrega una medida de la dependencia entre dos variables aleatorias X y Y es el coeficiente de correlación

$$\rho_{XY} = \frac{E\{(X-\mu_X)(Y-\mu_Y)\}}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

Ortogonalidad de variables aleatorias

Se dice que dos variables aleatorias X y Y son ortogonales si

$$E\{XY\} = 0.$$

Valores esperados condicionales

- Las relaciones entre dos variables aleatorias se describen algunas veces a través de valores esperados condicionales.
- Los valores esperados condicionales se definen como

$$E\{g(X,Y)|Y=y_j\}=\sum_{i=1}^{n} g(x_i,y_i)P(X=x_i|Y=y_j),$$

$$E\{g(X,Y)|X=x_i\}=\sum_{j=1}^{m} g(x_i,y_i)P(Y=y_i|X=x_j).$$

- **Pregunta**. Verifique que $E\{g(X,Y)=E_X\{E_{Y|X}[g(X,Y)|X]\}$.
- Uno de los valores esperados condicionales más importantes es la media condicional definida como

$$E\{X|Y=y_j\}=\mu_{X|Y=y_j}=\sum_{i=1}^{n} x_iP(X=x_i|Y=y_j).$$

Contenido

Uniforme y binomial

- **Función probabilidad de masa uniforme.** Se dice que una variable aleatoria X sigue una función de probabilidad de masa uniforme cuando

$$P(X = x_i) = \frac{1}{n}, \quad i = 1, 2, 3, \ldots, n.$$

- **Función de probabilidad de masa binomial.**
  - Sea ρ la probabilidad de que un evento A ocurra, a partir de un experimento aleatorio E.
  - Si el experimento se repite n veces y las n salidas son independientes, sea X la variable aleatoria que representa el número de veces que A ocurre en las n repeticiones.
  - La probabilidad de que el evento A ocurra k veces está dada por la función de probabilidad de masa

$$P(X = k) = \binom{n}{k} \rho^k (1 - \rho)^{n-k}, \quad k = 0, 1, 2, \ldots, n.$$

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

Poisson

- Sea emplea para modelar eventos como el número de llamadas telefónicas recibidas por una oficina y el número de electrones que emite un cátodo encandilado.
- El número de eventos en un intervalo de tiempo de longitud T sigue una función de probabilidad de masa de Poisson de la forma

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

donde $\lambda = \lambda' T$.

Multinomial

- Suponga un experimento aleatorio que se repite n veces.
- La salida de cada repetición del experimento corresponde a uno de k posibles eventos $A_1, A_2, \ldots, A_k$, mutuamente excluyentes y exhaustivos.
- Sea p_i la probabilidad de que la salida del experimento sea A_i.
- Asumamos que p_i permanece constante a lo largo de las n repeticiones.
- Sea X_i la variable aleatoria que indica el número de veces que el experimento termina en el evento A_i. Luego,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

donde $\sum x_i = n$, $\sum p_i = 1$ y cada x_i puede tomar cualquier valor entero entre 0 y n.

Ejemplo Variable Discreta

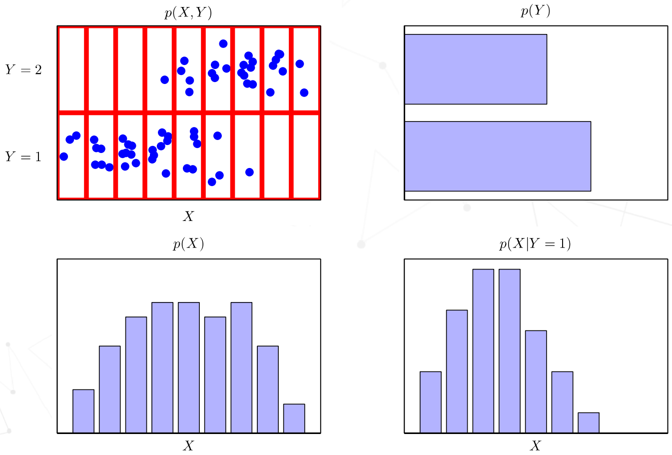


Figure 1

Contenido

Funciones de densidad de probabilidad

- Como se mencó anteriormente, una *variable aleatoria continua* puede tomar valores en un intervalo de la recta real.
- La ley de probabilidad para una variable aleatoria continua X se define en términos de la función de densidad de probabilidad (pdf) $f_X(x)$,

$$f_X(x) = \frac{\mathrm{d}\, F_X(x)}{\mathrm{d}\, x}.$$

Propiedades de una fdp

1. $f_X(x) \geq 0.$
2. $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}\,x = 1.$
3. $P(X \leq a) = F_X(a) = \int_{-\infty}^{a} f_X(x)\,\mathrm{d}\,x.$
4. $P(a \leq x \leq b) = \int_{a}^{b} f_X(x)\,\mathrm{d}\,x.$

Dos variables aleatorias: pdf conjunta

- Si existen dos variables aleatorias X y Y, pueden caracterizarse con la pdf $f_{X,Y}(x,y)$.
- Si la función de distribución conjunta $F_{X,Y}(x,y)$ es continua y tiene derivadas parciales, la pdf conjunta se define como

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

- Igualmente se tien los siguientes resultados
  1. $F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(\alpha,\beta)\, \mathrm{d}\,\alpha\, \mathrm{d}\,\beta.$
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(\alpha,\beta) = 1.$

Fdp marginal y fdp condicional

- Las funciones de densidad de probabilidad marginal están dadas como

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

- Las funciones de densidad de probabilidad condicional se definen como

$$f_{X|Y}(x|y) \triangleq \frac{f_{X,Y}(x,y)}{f_Y(y)},$$
$$f_{Y|X}(y|x) \triangleq \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Teorema de Bayes e Independencia Estadística

- El teorema de Bayes para variables aleatorias continuas está dado como

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(\lambda)\, \mathrm{d}\,\lambda}.$$

- Se dice que dos variables aleatorias X y Y son estadísticamente independientes si

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Valores esperados

Los valores esperados de variables aleatorias continuas se definen como

$$
\begin{aligned}
E\{g(X,Y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,\mathrm{d}\,x\,\mathrm{d}\,y,\\
\mu_x = E\{X\} &= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}\,X,\\
\sigma_X^2 = E\{(X-\mu_X)^2\} &= \int_{-\infty}^{\infty} (x-\mu_x)^2 f_X(x)\,\mathrm{d}\,x,\\
\sigma_{XY} &= E\{(X-\mu_X)(Y-\mu_y)\},\\
\rho_{XY} &= \frac{E\{(X-\mu_X)(Y-\mu_y)\}}{\sigma_X\sigma_Y},\\
E\{g(X,Y)|Y=y\} &= \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\,\mathrm{d}\,x,\\
E\{g(X)h(y)\} &= E\{g(X)\}E\{h(Y)\}.
\end{aligned}
$$

Contenido

Fdp uniforme

Se dice que una variable aleatoria continua X sigue una función de densidad de probabilidad uniforme si

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{de otra manera} \end{cases}$$

Fdp Gaussiana I

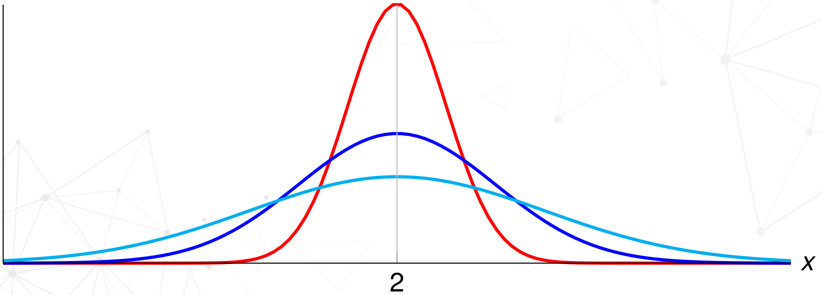
- La fdp Gaussiana está motivada en el *teorema del límite central*: una variable aleatoria determinada como la suma de un gran número de causas independientes tiende a tener una fdp Gaussiana.
- Bajo la presunción del teorema del límite central, la fdp Gaussiana se emplea para modelar el ruido eléctrico.
- La fdp Gaussiana tiene la forma

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}.$$

- La familia de fdp Gaussianas está caracterizada sólo por dos parámetros μ_x y σ_X^2, que son la media y la varianza de la variable aleatoria X.
- En este curso, la fdp Gaussiana se denota como $f_X(x) = \mathcal{N}\left(x|\mu_X, \sigma_X^2\right)$ ó $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.

Fdp Gaussiana II

La media de las tres Gaussianas es $\mu_X = 2$, y las desviaciones estándar son $\sigma_X = 0.5$ (en rojo), $\sigma_X = 1$ (en azul) y $\sigma_X = 1.5$ (en cyan).



Fdp Gaussiana III

- En algunas aplicaciones resulta necesario calcular la probabilidad

$$P(X > a) = \int_a^{\infty} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\} d\,x.$$

- Haciendo un cambio de variable $z = (x - \mu_X)/\sigma_X$, la integral anterior se reduce a

$$P(X > a) = \int_{(a-\mu_X)/\sigma_X}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}.$$

Fdp Gaussiana III

- Esta integral no puede resolverse analíticamente. Se expresa en términos de la función $Q(y)$ definida como

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty \exp\left\{-\frac{z^2}{2}\right\} \mathrm{d}\, z, \quad y > 0.$$

La función $Q(y)$ se puede calcular a través de la relación

$$Q(y) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{y}{\sqrt{2}}\right),$$

donde $\mathrm{erfc}(x)$ es la función de error complementaria.

- **Gamma** (prior conjugado para la varianza de una Gaussiana o la media de una Poisson).
- **Beta** (prior conjugada para los parámetros de una binomial).
- **Dirichlet** (prior conjugada para los parámetros de una multinomial).
- **Exponencial** (se usa para modelar el tiempo de espera de un próximo evento en un proceso de Poisson)
- **Weibull** (se usa para modelar tiempos de fallas en análisis de confiabilidad).

Contenido

Función de distribución de vectores aleatorios

- Un **vector aleatorio** es un vector cuyos elementos individuales son variables aleatorias.
- La ley de probabilidad para un vector de variables aleatorias se especifica en términos de función de distribución conjunta

$$F_{X_1,X_2,\ldots,X_m}(x_1, x_2, \ldots, x_m) = P[(X_1 \leq x_1), (X_2 \leq x_2), \ldots, (X_m \leq x_m)].$$

- También se puede especificar en términos de la función de probabilidad de masa (para variables discretas) o en términos de la función de densidad de probabilidad (para variables continuas).
- En lo que sigue se analiza el caso de vectores de variables aleatorias continuas.

Fdp conjunta y marginal

- La fdp conjunta de un vector aleatorio de m dimensiones es la derivada parcial de la función de distribución

$$f_{X_1,X_2,\ldots,X_m} = \frac{\partial^m F_{X_1,X_2,\ldots,X_m}(x_1, x_2, \ldots, x_m)}{\partial x_1 \partial x_2 \ldots \partial x_m}$$

- La fdp de una de las variables aleatorias X_1 , para $1 \leq i \leq m$ se obtiene integrando con respecto a todas las otras variables aleatorias $X_j \neq X_i$. Por ejemplo, para X_1

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)\, \mathrm{d}\, x_2 \ldots \mathrm{d}\, x_m$$

- Si se quiere la fdp conjunta de (X_i, X_j), se marginaliza integrando con respecto a X_k para $k \neq i$ y $k \neq j$. Por ejemplo, la fdp conjunta (X_1, X_2)

$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)\, \mathrm{d}\, x_3 \ldots \mathrm{d}\, x_m.$$

Fdp condicional

- La fdp condicional de un subconjunto de variables aleatorias del vector aleatorio dado otro subconjunto de variables del mismo vector.
- Ejemplos de fdp condicionales son

$$f_{X_1,X_2,X_3|X_4}(x_1,x_2,x_3|x_4) = \frac{f_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4)}{f_{X_4}(x_4)}$$

$$f_{X_1,X_2|X_3,X_4}(x_1,x_2|x_3,x_4) = \frac{f_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4)}{f_{X_3,X_4}(x_3,x_4)}$$

Valores esperados

- Los valores esperados se calculan usando múltiples integrales.
- Por ejemplo,

$$E\{g(x_1,\ldots,x_m)\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_m) f_{X_1,\ldots,X_m}(x_1,\ldots,x_m)\,\mathrm{d}\,x_1\ldots\mathrm{d}\,x_m.$$

- Igualmente, los valores esperados condicionales se calculan usando funciones de densidad de probabilidad condicionales.

Medias y covarianzas

- Dos momentos estadísticos de importancia en vectores aleatorios son las medias y las covarianzas.
- La media está dada como

$$\mu_{X_i} = E\{X_i\}.$$

- Las covarianzas están dadas como

$$\sigma_{X_i X_j} = E\{X_i X_j\} - \mu_{X_i}\mu_{X_j}.$$

Notación

- La ley de probabilidad para vectores aleatorios puede especificarse de manera concisa usando notación vectorial.
- Supóngase un conjunto de m variables aleatorias $X_1, X_2, \ldots, X_m$.
- Estas variables aleatorias pueden representarse como un vector columna de dimensiones $m \times 1$,

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

- Un valor específico de **X** se denota como $\mathbf{x}^\top = (x_1, x_2, \ldots, x_m)$.

Fdp con notación vectorial

- Con notación vectorial, la probabilidad conjunta está dada como

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\ldots,X_m}(x_1, \ldots, x_m).$$

- El vector de medias se define como

$$\mu_{\mathbf{X}} = E\{\mathbf{X}\} = \begin{bmatrix} E\{X_1\} \\ E\{X_2\} \\ \vdots \\ E\{X_m\} \end{bmatrix}$$

- La matriz de covarianza está definida como

$$\Sigma_{\mathbf{X}} = E\{\mathbf{X}\mathbf{X}^\top\} = \begin{bmatrix} \sigma_{X_1X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_m} \\ \sigma_{X_2X_1} & \sigma_{X_2X_2} & \cdots & \sigma_{X_2X_m} \\ \vdots & \cdots & \cdots & \vdots \\ \sigma_{X_mX_1} & \sigma_{X_mX_2} & \cdots & \sigma_{X_mX_m} \end{bmatrix}$$

Correlación e Independencia

- Se dice que dos componentes del vector aleatorio no están correlacionadas si $\sigma_{X_i X_j} = \sigma_{ij} = 0$.
- Se dice que los componentes del vector aleatorio son independientes si

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{m} f_{X_i}(x_i)$$

- Otra notación para la fdp de un vector aleatorio es $p(\mathbf{x})$.

Contenido

Fdp Gaussiana multivariada

Se dice que un vector aleatorio X es Gaussiano multivariado si su fdp sigue la forma

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{m/2}|\Sigma_{\mathbf{X}}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^{\top}\Sigma_{\mathbf{X}}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})\right]$$

donde $\mu_{\mathbf{X}}$ es el vector de medias, $\Sigma_{\mathbf{X}}$ es la matriz de covarianza, $|\Sigma_{\mathbf{X}}|$ es el determinante de $\Sigma_{\mathbf{X}}$ y $\Sigma_{\mathbf{X}}^{-1}$ es la matriz inversa de $\Sigma_{\mathbf{X}}$.
Una notación alternativa para la fdp Gaussiana multivariada es la siguiente,

$$p(\mathbf{x}) = \mathcal{N}\left(\mathbf{x}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{X}}\right).$$

Fdp Gaussiana multivariada II

Suponga que **X** es un vector aleatorio que sigue una fdp Gaussiana. Si se particiona el vector de variables aleatorias **X** de la siguiente forma

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} X_{k+1} \\ X_{k+2} \\ \vdots \\ X_m \end{bmatrix}$$

y

$$\mu_{\mathbf{X}} = \begin{bmatrix} \mu_{\mathbf{X}_1} \\ \mu_{\mathbf{X}_2} \end{bmatrix}, \Sigma_{\mathbf{X}} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

donde $\mu_{\mathbf{X}_i} = E\{\mathbf{X}_i\}$ y $\Sigma_{ij} = E\{\mathbf{X}_i\mathbf{X}_j^\top\} - \mu_{\mathbf{X}_i}\mu_{\mathbf{X}_j}$, luego $\mathbf{X}_1$ sigue una fdp Gaussiana multivariada de k dimensiones con media $\mu_{\mathbf{X}_1}$ y la matriz de covarianza Σ_{11}.

Fdp Gaussiana multivariada III

- Si $\Sigma_{\mathbf{X}}$ es una matriz diagonal, esto es,

$$\Sigma_{\mathbf{X}} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sigma_m^2 \end{bmatrix},$$

  luego las componentes de $\mathbf{X}$ son independientes (en el caso Gaussiano, la no correlación implica independencia, lo cual no es necesariamente cierto para otras fdps).
- Si $A \in \mathbb{R}^{k \times m}$ es una matriz de rango k, luego $\mathbf{Y} = \mathbf{A}\mathbf{X}$ sigue una fdp Gaussiana de dimensión k con momentos

$$\mu_{\mathbf{Y}} = \mathbf{A}\mu_{\mathbf{X}},$$
$$\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^\top.$$

Fdp Gaussiana multivariada IV

- Empleando la partición del vector $\mathbf{X}$ como se vió anteriormente, la fdp condicional de $\mathbf{X}_1$ dado $\mathbf{X}_2 = \mathbf{x}_2$ es una Gaussiana multivariada con los siguientes momentos

$$\mu_{\mathbf{X}_1|\mathbf{X}_2} = E\{\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2\} = \mu_{\mathbf{X}_1} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_{\mathbf{X}_2}),$$
$$\Sigma_{\mathbf{X}_1|\mathbf{X}_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

- Las propiedades anteriores indican que la fdp condicional, la fdp marginal y las transformaciones lineales derivadas de una fdp Gaussiana multivariada conducen a fdp Gaussianas multivariadas.

Referencias

- Basado en las presentaciones del Prof. Mauricio A. Álvarez, del curso "Procesos Estocásticos".
- S. Shamugan. " Random Signals: Detection, Estimation and Data Analysis", 1988.