Repaso de Probabilidad

Lección 05

Dr. Pablo Alvarado Moya

CE5506 Introducción al reconocimiento de patrones
Área de Ingeniería en Computadores
Tecnológico de Costa Rica

II Semestre, 2019

Contenido

1. Introducción
   - Generalidades
   - Definiciones
2. Variables aleatorias
   - Definiciones
   - Operadores
3. Distribuciones
   - Distribuciones discretas
   - Distribuciones continuas

Teoría de probabilidad

- Marco matemático para representar declaraciones **inciertas**
- Cuantifica la **incertidumbre**
- Provee axiomas para derivar nuevas declaraciones inciertas
- Permite representar
  - Cantidades inciertas
  - Cantidades estocásticas

Fuentes de incertidumbre

Según Pearl, hay tres fuentes de incertidumbre:

1. Sistema modelado es inherentemente estocástico
   (p. ej. mecánica cuántica, juego de dados o cartas, etc.)
2. Observabilidad incompleta
   (p. ej. juegos donde observador no conoce el proceso determinístico)
3. Modelado incompleto
   (p. ej. problemas de cuantificación en representación espacial o temporal)

Conveniencia de incertidumbre

- Con frecuencia es más práctico una regla simple incierta, que una regla compleja exacta:
  1. *La mayoría de los pájaros vuelan*
  2. *Los pájaros vuelan, excepto pichones que no han aprendido, pájaros enfermos o heridos que han perdido la habilidad, especies incapaces como el avestruz, kiwi...*

Frecuentistas vs. Bayesianos

- Hay dos tipos de probabilidades:
  1. Cuantificación de frecuencias (probabilidad frecuentista)
     - Indica en p repeticiones de un experimento, qué fracción de las veces se observa un evento
  2. Grado de creencia (probabilidad Bayesiana)
     - Problemas sin repeticiones, como diagnóstico médico
- Se usan mismas herramientas para tratar ambos tipos de probabilidades
- Podemos interpretar probabilidad como una extensión a la lógica, capaz de manejar incertidumbre:
  - Reglas formales para determinar la posibilidad de veracidad de una proposición dada la posibilidad de otras proposiciones.

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset, \Omega\}$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset, \Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset,\Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset,\Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
  - $P(A) \geq 0, \forall A \in \mathcal{F}$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset,\Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
  - $P(A) \geq 0, \forall A \in \mathcal{F}$
  - $P(\Omega) = 1$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset,\Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
  - $P(A) \geq 0, \forall A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - Si $A_i \cap A_j = \emptyset$, $i \neq j$ entonces $P(\cup_i A_i) = \sum_i P(A_i)$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
    - Ejemplo: Tiro de un dado tiene $\Omega = \{1,2,3,4,5,6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
    - Ejemplo 1: $\mathcal{F} = \{\emptyset, \Omega\}$
    - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
    - $P(A) \geq 0, \forall A \in \mathcal{F}$
    - $P(\Omega) = 1$
    - Si $A_i \cap A_j = \emptyset$, $i \neq j$ entonces $P(\cup_i A_i) = \sum_i P(A_i)$
    - Ejemplo 1: $P(\emptyset) = 0$, $P(\Omega) = 1$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1{,}2{,}3{,}4{,}5{,}6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset, \Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
  - $P(A) \geq 0, \forall A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - Si $A_i \cap A_j = \emptyset$, $i \neq j$ entonces $P(\cup_i A_i) = \sum_i P(A_i)$
  - Ejemplo 1: $P(\emptyset) = 0$, $P(\Omega) = 1$
  - Ejemplo 2: $P(A_i) = |A_i|/6$, es decir, $P(\{1{,}2{,}3{,}4\}) = 4/6$, $P(\{2{,}4{,}6\}) = 3/6$

Definiciones

- Espacio muestral Ω: conjunto de todos los resultados posibles de un experimento.
  - Ejemplo: Tiro de un dado tiene $\Omega = \{1{,}2{,}3{,}4{,}5{,}6\}$
- Espacio de eventos $\mathcal{F}$: conjunto con elementos $A \in \mathcal{F}$ llamados **eventos**, que son $A \subset \Omega$ colecciones de posibles resultados del experimento
  - Ejemplo 1: $\mathcal{F} = \{\emptyset, \Omega\}$
  - Ejemplo 2: $\mathcal{F} = \{A \mid A \subset \Omega\}$
- Medida de probabilidad: Función $P : \mathcal{F} \to \mathbb{R}$ tal que
  - $P(A) \geq 0, \forall A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - Si $A_i \cap A_j = \emptyset$, $i \neq j$ entonces $P(\cup_i A_i) = \sum_i P(A_i)$
  - Ejemplo 1: $P(\emptyset) = 0$, $P(\Omega) = 1$
  - Ejemplo 2: $P(A_i) = |A_i|/6$, es decir, $P(\{1{,}2{,}3{,}4\}) = 4/6$, $P(\{2{,}4{,}6\}) = 3/6$
- Anteriores propiedades son los **axiomas de probabilidad**

Propiedades

- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(A \cap B) \leq \text{mín}(P(A), P(B))$
- $P(A \cup B) \leq P(A) + P(B)$
- $P(\Omega \setminus A) = 1 - P(A)$
- Si $A_1, \ldots, A_k$ son disjuntos y $\bigcup_{i=1}^{k} A_i = \Omega$ entonces $\sum_{i=1}^{k} P(A_k) = 1$

Probabilidades condicionales

- Sea B un evento probable $(P(B) > 0)$
- La probabilidad **condicional** de cualquier evento A **dado** B es

$$P(A|B) \stackrel{!}{=} \frac{P(A \cap B)}{P(B)}$$

Variables aleatorias

- Una **variable aleatoria** es una variable que toma valores aleatoriamente.
- Es una "función" $X : \Omega \to \mathbb{R}$
- Seguimos notación de Dr. Ng, con $X(\omega)$ o símplemente X
- Variable aleatoria puede ser
  - **discreta**: $P(X = k) := P(\{\omega \mid X(\omega) = k\})$
  - **continua**: $P(a \leq X \leq b) := P(\{\omega \mid a \leq X(\omega) \leq b\})$

Tipos de funciones de distribución de probabilidad

Tres tipos de distribuciones se utilizan

- Distribuciones **acumuladas**
- Distribuciones **de masa**
- Distribuciones **de densidad**

Funciones de distribución acumulada

- $F_X(x) \stackrel{!}{=} P(X \leq x)$
- Con $F_X(x)$ puede calcularse cualquier probabilidad de cualquier evento
- Se debe cumplir:
  - $0 \leq F_X(x) \leq 1$
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to +\infty} F_X(x) = 1$
  - $x \leq y \Rightarrow F_X(x) \leq F_X(y)$

Función de masa de probabilidad

- Utilizadas para variables aleatorias **discretas**
- PMF: (*probability mass function*)

$$p_X(x) \stackrel{!}{=} P(X = x)$$

- Propiedades:
  - $0 \leq p_X(x) \leq 1$
  - $\sum_{x \in \mathrm{val}(X)} p_X(x) = 1$
  - $\sum_{x \in A} p_X(x) = P(X \in A)$
- Se utiliza operador $\mathrm{val}(X)$ para obtener conjunto de posibles valores de X

Funciones de densidad de probabilidad (1)

- Útiles para variables aleatorias **continuas**
- Definidas solo si $F_X(x)$ es derivable
- PDF (*probability density function*)
- Se definen como

$$f_X(x) \stackrel{!}{=} \frac{dF_X(x)}{dx}$$

- Se debe cumplir por series de Taylor para $\Delta x \to 0$

$$P(x \leq X \leq x + \Delta x) \approx f_X(x)\Delta x$$

- **¡Cuidado!**: $f_X(x) \neq P(X = x)$
- Nótese que es posible que $f_X(x) > 1$ pero $\int f_X(x)\, dx = 1$

Funciones de densidad de probabilidad (2)

- Propiedades:
  - $f_X(x) \geq 0$
  - $\int_{-\infty}^{\infty} f_X(x) = 1$
  - $\int_{x \in A} f_X(x)\,dx = P(X \in A)$

Esperanza (1)

- Para una variable aleatoria **discreta** X con PMF $p_X(x)$ y $g : \mathbb{R} \to \mathbb{R}$ una función arbitraria
- $g(X)$ se puede considerar como variable aleatoria
- La **esperanza** o **valor esperado** de $g(X)$ se define como

$$E[g(X)] \stackrel{!}{=} \sum_{x \in \mathrm{val}(X)} g(x) p_X(x)$$

- Para una variable aleatoria **continua** X con PDF $f_X(x)$ la **esperanza** o **valor esperado** de $g(X)$ se define como

$$E[g(X)] \stackrel{!}{=} \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Esperanza (2)

- Intuitivamente la esperanza es un promedio **ponderado** de los valores de $g(x)$
- La esperanza de la misma variable aleatoria X (con $g(x) = x$) se conoce como **media** de X
- Propiedades:
  - $\mathrm{E}[a] = a$, con $a \in \mathbb{R}$ constante
  - $\mathrm{E}[af(X) + bg(X)] = a\,\mathrm{E}[f(X)] + b\,\mathrm{E}[g(x)]$ para constantes $a,b \in \mathbb{R}$
  - Para variables aleatorias discretas X, $\mathrm{E}[1\{X = k\}] = P(X = k)$
  - $1\{\cdot\}$ es la **función indicadora** igual a 1 si argumento es verdadero y sino es 0

Varianza (1)

- La **varianza** de una variable aleatoria X mide qué tanto se dispersa X alrededor de su media

$$\mathrm{Var}[X] \stackrel{!}{=} \mathrm{E}[(X - \mathrm{E}[X])^2]$$

- Con las propiedades de la esperanza:

$$\begin{aligned}
\mathrm{E}[(X - \mathrm{E}[X])^2] &= \mathrm{E}[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}^2[X]] \\
&= \mathrm{E}[X^2] - 2E[X\,\mathrm{E}[X]] + \mathrm{E}^2[X] \\
&= \mathrm{E}[X^2] - 2\,\mathrm{E}[X]E[X] + \mathrm{E}^2[X] \\
&= \mathrm{E}[X^2] - \mathrm{E}^2[X]
\end{aligned}$$

Propiedades de la varianza

- $\text{Var}[a] = 0$ para $a \in \mathbb{R}$ constante
- $\text{Var}[af(X)] = a^2 \text{Var}[f(X)]$ para $a \in \mathbb{R}$ constante

Ejemplo

(1)

Ejemplo (Distribución uniforme)

Calcule la esperanza y la varianza de la variable aleatoria X con PDF $f_X(x) = 1, \forall x \in [0,1]$ o 0 en el resto.

Ejemplo (2)

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x\,dx = \frac{1}{2} \\
\mathrm{E}[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_0^1 x^2\,dx = \frac{1}{3} \\
\mathrm{Var}[X] &= E[X^2] - E^2[X] = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}
\end{aligned}
$$

Ejemplo

(1)

Ejemplo (Otro caso)

Suponga que $g(x) = 1\{x \in A\}$ para $A \subseteq \Omega$. Encuentre $\mathrm{E}[g(X)]$

Ejemplo (2)

Caso discreto:

$$E[g(x)] = \sum_{x \in \mathrm{val}(X)} 1\{x \in A\}\, P_X(x) = \sum_{x \in A} P_X(x) = P(x \in A)$$

Caso continuo:

$$E[g(x)] = \int_{-\infty}^{\infty} 1\{x \in A\}\, f_X(x)\, dx = \int_{x \in A} f_X(x)\, dx = P(x \in A)$$

Distribución de Bernoulli

- $\text{val}(X) = \{0,1\}$
- Parámetro $\phi \in [0,1]$ indica probabilidad de que la variable aleatoria sea igual a 1:

$$P(X = 1) = \phi$$
$$P(X = 0) = 1 - \phi$$

- Con eso se puede reexpresar

$$P(X = x) = \phi^x (1 - \phi)^{1-x}$$

- Media y varianza

$$\text{E}[X] = \phi$$
$$\text{Var}[X] = \phi(1 - \phi)$$

Distribución categórica o "multinoulli"

- $\text{val}(X) = \{0,1,\ldots,k\}$
- Parámetro $\underline{\mathbf{p}} \in [0,1]^k$ donde p_i es la probabilidad de i-ésimo estado (hay $k+1$ estados)
- El $(k+1)$-ésimo estado tiene probabilidad $1 - \underline{\mathbf{1}}^T \underline{\mathbf{p}}$.
- Se debe cumplir $\underline{\mathbf{1}}^T \underline{\mathbf{p}} \leq 1$.
- Variable aleatoria no necesariamente asume valores numéricos, sino **categorías**, por lo que no se usa el valor medio o varianza.

Distribución uniforme

- PDF:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{el resto} \end{cases}$$

- Media y varianza

$$\mathrm{E}[X] = \frac{a+b}{2}$$

$$\mathrm{Var}[X] = \frac{(b-a)^2}{12}$$

Distribución gaussiana (o normal)

Caso univariado

- Parámetros $\mu \in \mathbb{R}$, $\sigma \in (0,\infty)$
- PDF:

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

- Media y varianza:

$$\mathsf{E}[X] = \mu$$

$$\mathsf{Var}[X] = \sigma^2$$

- σ se denomina desviación estándar
- Para ahorrar divisiones se usa a menudo (con $\gamma = 1/\sigma^2$):

$$\mathcal{N}(x;\mu,\gamma) = \sqrt{\frac{\gamma}{2\pi}} e^{-\frac{1}{2}\gamma(x-\mu)^2}$$

- γ se denomina **precisión** o varianza inversa

Distribución gaussiana

- Es la distribución más común
- Se utiliza por defecto, si no se tiene más información:
  - Teorema del límite central: suma de muchas variables aleatorias independientes converge a distribución normal, lo que se puede asumir en sistemas complejos
  - Ruido en general se puede asumir distribuido normalmente, incluso si el sistema tiene bloques que no lo son
  - De todas las distribuciones con igual varianza, la distribución normal codifica la mayor incertidumbre sobre los números reales (introduce menos conocimiento previo en un modelo)

Distribución gaussiana (o normal)

Caso multivariado

- En $\mathbb{R}^n$ la distribución normal multivariada es:

$$\mathcal{N}(\underline{\mathbf{x}}; \underline{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})\right)$$

- $\underline{\boldsymbol{\mu}}$ es la media y $\boldsymbol{\Sigma}$ la matriz de covarianza
- Para evitar las inversiones se usa también la matriz de precisión $\boldsymbol{\Gamma}$:

$$\mathcal{N}(\underline{\mathbf{x}}; \underline{\boldsymbol{\mu}}, \boldsymbol{\Gamma}) = \sqrt{\frac{\det(\boldsymbol{\Gamma})}{(2\pi)^n}} \exp\left(-\frac{1}{2}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})^T \boldsymbol{\Gamma}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}})\right)$$

Resumen

1. Introducción
    - Generalidades
    - Definiciones
2. Variables aleatorias
    - Definiciones
    - Operadores
3. Distribuciones
    - Distribuciones discretas
    - Distribuciones continuas

Este documento ha sido elaborado con software libre incluyendo LATEX, Beamer, GNUPlot, GNU/Octave, XFig, Inkscape, GNU-Make y Subversion en GNU/Linux