Modelação e Física Estatística

Conceitos de probabilidade e teoria da informação[1]

António Luís Ferreira

February 18, 2019

[1] slides baseados no Cap. 1 de Information, Physics and Computation, Oxford University press, 2009

Temas

variáveis aleatórias

- X - variável aleatória discreta
- $\mathscr{X}$ - conjunto de valores tomado pela variável, $p_X(x) = prob(X = x) \geq 0$
- Valor médio - $\overline{X} = \sum_{x \in \mathscr{X}} x\, p_X(x)$.
- Variância - $Var\, X = \overline{(X - \bar{X})^2} = \overline{X^2} - \overline{X}^2 = \sum_{x \in \mathscr{X}} (x - \overline{X})^2 p_X(x)$
- Acontecimento $\mathscr{A} \subseteq \mathscr{X}$ (contido em $\mathscr{X}$) tem probabilidade $prob(\mathscr{A}) = \sum_{x \in \mathscr{A}} p(x)$
- Normalização - $\sum_{x \in \mathscr{X}} p_X(x) = 1$

exemplos

- Variáveis contínuas - $prob(X \in \mathscr{A}) = \int_{x \in \mathscr{A}} p_X(x)\, dx = \int_{\mathscr{X}} I(x \in \mathscr{A}) p_X(x)\, dx$, onde $I(s)$ é a função indicadora igual a 1 se a *afirmação s* é verdadeira e 0 se é falsa; $p_X(x)$ é a densidade de probabilidade.
- Normalização - $\int_{x \in \mathscr{X}} p_X(x)\, dx = 1$
- Bernoulli - $\mathscr{X}$ ={0,1}, $p_X(1) = p$, $p_X(0) = 1 - p$; $\overline{X} = p$; $Var\, X = p(1-p)$
- Distribuição discreta uniforme: $\mathscr{X}$ ={1,2, ..., M}, $p_X(x) = 1/M$
- Variável Gaussiana - $p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$; $\overline{X} = \mu$; $Var\, X = \sigma^2$

entropia

- Entropia de uma variável aleatória com ditribuição de probabilidade $p_X(x)$

$$S_X = S(p_X(x)) = -\sum_{x \in \mathscr{X}} p_X(x) \log_2 p_X(x) = \overline{\log_2 \frac{1}{p_X(X)}}$$

com $0 \log_2 0 = 0$.

- Propriedades:
  - $S_X \geq 0$
  - S_X toma um valor máximo para a distribuição uniforme.
- Bernoulli variable, X: $S_X = p \log_2(p) - (1-p) \log_2(1-p)$ tem máximo para $p = 1/2$,

desigualdade de Jensen

- $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$ com $\alpha \in [0,1]$; equivalente a $\frac{d^2 f}{dx^2} > 0$, para qualquer ponto no intervalo $[x,y]$

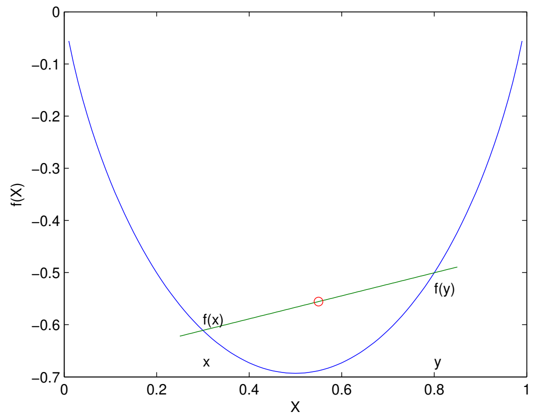

Figure: Função convexa, f(x). A reta secante que passa pelos pontos (x,f(x)) e (y,f(y)) está sempre acima da função entre esses pontos

- Eq. da reta secante

$$\begin{aligned} z(\alpha) &= f(x) + \frac{f(y)-f(x)}{y-x}(\alpha x + (1-\alpha)y - x) \\ &= \alpha f(x) + (1-\alpha)f(y) \end{aligned}$$

desigualdade de Jensen

- Se f(x) é uma função convexa então, $\overline{f(X)} \geq f(\overline{X})$
- Demonstração
  - Se $\mathscr{X} = \{x_1, x_2\}$ então $\overline{f(X)} = p_1 f(x_1) + p_2 f(x_2)$ com $p_1 + p_2 = 1$ e $\overline{f(X)} \geq f(p_1 x_1 + p_2 x_2)$ ou seja $\overline{f(X)} \geq f(\overline{X})$
  - Se $\mathscr{X} = \{x_1, x_2, x_3\}$ então $\overline{f(X)} = p_1 f(x_1) + p_2 f(x_2) + p_3 f(x_3)$ com $p_1 + p_2 + p_3 = 1$
    - $\overline{f(X)} = (p_1 + p_2)\left[\frac{p_1}{p_1+p_2} f(x_1) + \frac{p_2}{p_1+p_2} f(x_2)\right] + p_3 f(x_3)$
    - $\overline{f(X)} \geq (p_1 + p_2) f(\frac{p_1}{p_1+p_2} x_1 + \frac{p_2}{p_1+p_2} x_2) + p_3 f(x_3)$
    - $\overline{f(X)} \geq f(p_1 x_1 + p_2 x_2 + p_3 x_3) = f(\overline{X})$

divergência de Kulback-Leibler

- Divergência de Kulback-Leibler entre duas distribuições de probabilidade
  - $D(q_X||p_X) = \sum_{x \in \mathscr{X}} q_X(x) \log \frac{q_X(x)}{p_X(x)}$
- Propriedades
  - $D(q_X||p_X) = 0$ se $q_X(x) = p_X(x)$;
  - $D(q_X||p_X) \geq 0$.
    - Como $-\log(x)$ é uma função convexa
      $D(q_X||p_X) = \overline{-\log \frac{p_X(x)}{q_X(x)}} \geq -\log \overline{\frac{p_X(x)}{q_X(x)}} = 0$

várias variáveis

- probabilidade condicional (Formula de Bayes)
  - $\text{prob}(\mathscr{X}|\mathscr{Y}) = \frac{\text{prob}(\mathscr{X}\cap\mathscr{Y})}{\text{prob}(\mathscr{Y})}$ ou $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$.
  - $p_Y(y) = \sum_{x\in\mathscr{X}} p_{X,Y}(x,y)$ e portanto $\sum_{x\in\mathscr{X}} p_{X|Y}(x|y) = 1$
- variáveis independentes
  - $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ ou seja $p_{X|Y}(X=x|Y=y) = p_X(x)$;
  - $Var(X+Y) = Var(X) + Var(Y)$
- independentes e idênticamente distribuídas - $p_N(x_1, x_2, \cdots x_N) = \prod_{i=1,N} p(x_i)$;
- Acontecimentos mutuamente exclusivos, $\mathscr{A}_1 \cap \mathscr{A}_2 \equiv \oslash$ temos $prob(A_1 \cup \mathscr{A}_2) = prob(\mathscr{A}_1) + prob(\mathscr{A}_2)$