FYS-STK3-4155 sept 1

statistics, reminder

$$\mathcal{D} = \{(x_0, y_0), (x_1, y_1) \dots (x_{n-1}, y_{n-1})\}$$

probability distribution

$p(x)$

$$0 \leq p(x) \leq 1$$

$$x \in [a, b]$$

$$\int_a^b p(x)\,dx = 1$$ continuous case

$$\sum_{x_i \in X} p(x_i) = 1$$ discrete case

Expectation values

$$\mathbb{E}[x^n] = \int_a^b p(x)\, x^n\, dx$$

$$\mathbb{E}[x] = \mu = \langle x \rangle = \int_a^b p(x)\, x\, dx$$

Discrete :

$$\mu = \sum_{i=0}^{n-1} x_i \, p(x_i)$$

sample mean

$$\bar{\mu} = \frac{1}{n} \sum_i x_i$$

$\bar{\mu} \neq \mu$ in general

<u>variance</u>

$$\mathrm{var}[x] = \sigma^2$$

$$= \int (x-\mu)^2 p(x)\,dx$$

(Discrete $= \sum_i (x_i - \mu) \, p(x_i)$)

sample variance

$$\bar{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{\mu})^2$$

$$\neq \sigma^2$$

iid = independ and

identically distributed

$$p(x_1, x_2) = p(x_1)\,p(x_2)$$

$$\mathrm{cov}[x_1, x_2] = \int dx_1\, dx_2\, (x_1 - \mu_1)(x_2 - \mu_2)\, p(x_1, x_2)$$

iid: $\mathrm{cov}[x_1 x_2] =$

$$\int dx_1\, dx_2\, (x_1 - \mu)(x_2 - \mu)\, p(x_1)\, p(x_2)$$

$$= \int dx_1 (x_1 - \mu)\, p(x_1) \int dx_2 (x_2 - \mu) \times p(x_2)$$

$$\left[\begin{array}{l} \int dx\, p(x)\, x = \mu \\ \int dx\, p(x) = \underline{1} \end{array} \right.$$

$$= 0.$$

if not

$$\mathrm{cov}[x_1 x_2] = \mathbb{E}[x_1 x_2]$$

$$- \mu_1 \mu_2$$

Discrete case

$$\text{cov}[x_1 x_2] = \sum_{i,j} (x_i^1 - \mu_1)(x_j^2 - \mu_2) \times p(x_i^1 x_j^2)$$

Joint probability (product rule)

$$p(A, B) = p(A \cap B)$$

$$= p(A|B)\, p(B)$$

$$= p(B|A)\, p(A)$$

Marginal probability

$$p(A) = \sum_b p(A \mid B=b)\, p(B=b)$$

conditional prob $(p(B) > 0)$

$$p(A|B) = \frac{p(A,B)}{p(B)}$$

$$= \frac{P(B|A)\,P(A)}{\sum_{a} P(B|A=a)\,P(A=a)}$$

Baye's theorem.

Example

A = output from
a Mammogram
(0, 1)

B = have cancer
(0, 1)

sensitivity

$$P(A=1 \mid B=1) = 0.8$$

$$P(B|A)$$

prior $P(B) = 0.004$

$$P(A=1 \mid B=0) = 0.1$$

Baye's theorem

$$P(B=1 \mid A=1) =$$

$$\frac{P(A=1 \mid B=1)\, P(B=1)}{P(A=1 \mid B=1)\, P(B=1) + P(A=1 \mid B=0)\, P(B=0)}$$

$$= \frac{0.8 \times 0.004}{0.8 \times 0.004 + 0.1 \times 0.996}$$

$$= 0.03 \Rightarrow$$

3%

$$D = \{ (x_0, y_0), \ldots (x_{n-1}, y_{n-1}) \}$$

$$P(y \mid X\beta) = ?$$

y are iid.

$\sim X\beta$

not stochastic

$$y = f(x) + \varepsilon \qquad \backslash ?$$

$$\underbrace{\varepsilon \sim N(0,1)}\ \mathrm{var}_\varepsilon = \sigma_\varepsilon^2$$

$$P(y \mid X\beta) = \prod_{i=0}^{n-1} P_i$$

$$P_i = P(y_i \mid X_{i*}\beta)$$

$$X_{i*}\beta = \sum_{j=0}^{p-1} x_{ij}\beta_j$$

Maximum likelihood estimator (MLE)

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmax}} P(y \mid X\beta) \uparrow P(\beta)$$

$\Rightarrow$ maximization of

$$\frac{\partial}{\partial \beta} \prod_i P_i(y_i \mid x_{i*}\beta)$$

$$\mathbb{E}[y_i] = X_{i*}\beta$$

$[\ \ ]$ 2

$$\operatorname{var}[y_i] = \sigma_\varepsilon$$

$$y_i \sim N(x_{i*}\beta, \sigma_\varepsilon^2)$$

$$C(\beta) = P(y \mid X\beta)$$

$$= \prod_i P_i$$

$$\tilde{C}(\beta) = \sum_{i=0}^{n-1} \log P(y_i \mid x_{i*}\beta)$$

$$C(\beta) = -\tilde{C}(\beta)$$

$$= -\sum_{i=0}^{n-1} \log P(y_i \mid x_{i*}\beta)$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min}\; C(\beta)$$

$$y_i \sim N(x_{i*}\beta, \sigma_\varepsilon^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-(y_i - x_{i*}\beta)^2 / 2\sigma_\varepsilon^2\right)$$

$$= p_i$$

$$\log p_i = -\frac{1}{2}\log(2\pi\sigma_\varepsilon^2) - (y_i - x_{i*}\beta)^2 / 2\sigma_\varepsilon^2$$

$$C(\beta) = \frac{n}{2}\log(2\pi\sigma_\varepsilon^2) + \sum_{i=0}^{n-1} \frac{(y_i - x_{i*}\beta)^2}{2\sigma_\varepsilon^2}$$

$$\frac{\partial C}{\partial \beta} = 0 = X^T(y - X\beta)$$