EXERCISE SHEET 02

Ex. 1

a) INDEPENDENCY: $P(x,y) = P(x) \cdot P(y)$

LAW OF TOTAL PROBABILITY: $P(x) = \int_y P(x,y)\,dy$

$$\iint_{xy} P(x,y)\,dx\,dy = 1$$

- For the probability distribution $P(x,y) = l\eta e^{-lx-\eta y}$:

$$P(x) = \int_y l\eta e^{-lx-\eta y}\,dy = le^{-lx}\int_y \eta e^{-\eta y}\,dy$$

$$P(y) = \int_x l\eta e^{-lx-\eta y}\,dx = \eta e^{-\eta y}\int_x le^{-lx}\,dx$$

$$P(x) \cdot P(y) = le^{-lx}\int_y \eta e^{-\eta y}\,dy\ \eta e^{-\eta y}\int_x le^{-lx}\,dx =$$

$$= \underbrace{l\eta e^{-lx-\eta y}}_{=P(x,y)} \underbrace{\iint_{xy} \underbrace{l\eta e^{-lx-\eta y}}_{=P(x,y)}\,dy\,dx}_{=1}$$

$$= P(x,y)$$

AND THEREFORE x AND y ARE INDEPENDENT.

b) Likelihood of the distribution with parameters θ given the data $\mathbf{D}$:

$$P(D|\omega,\theta) = P(D|\theta) = \prod_{i=1}^{N} P(x_i|\theta)$$

Log-Likelihood: $L(\theta) = \log(P(D|\theta))$

- For $P(x,y) = l\eta e^{-lx-\eta y}$:

$$P(D|l) = \prod_{i=1}^{N} P((x,y)|l) = \prod_{i=1}^{N} le^{-lx} \cdot \eta e^{-\eta y}$$

$$L = \sum_{i=1}^{N} \log l - lx_i + \log \eta - \eta y_i$$

$$\frac{\delta L}{\delta l} = \sum_{i=1}^{N} \frac{1}{l} - x_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} \frac{1}{l} = \sum_{i=1}^{N} x_i \Leftrightarrow \frac{N}{l} = \sum_{i=1}^{N} x_i \Leftrightarrow$$

$$\Leftrightarrow \frac{1}{l} = \underbrace{\frac{\sum x_i}{N}}_{\bar{x}} \Leftrightarrow l = \frac{1}{\bar{x}}$$

c) $$P(D|l) = \prod_{i=1}^{N} \cancel{l} e^{-lx} \cdot \cancel{\frac{1}{l}} \cdot e^{-\frac{1}{l}y} = \prod_{i=1}^{N} e^{-lx-\frac{1}{l}y}$$

$$L = \sum_{i=1}^{N} -lx_i - \frac{1}{l} y_i$$

$$\frac{\delta L}{\delta l} = \sum_{i=1}^{N} -x_i + \frac{1}{l^2} y_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \frac{1}{l^2} y_i \Leftrightarrow \bar{x} = \frac{1}{l^2} \sum_{i=1}^{N} y_i \Leftrightarrow$$

$$\Leftrightarrow l = \sqrt{\frac{\sum_{i=1}^{N} y_i}{\bar{x}}} = \sqrt{N \cdot \sum_{i=1}^{N} \frac{y_i}{x_i}}$$

d) $P(D|l) = \prod_{i=1}^{N} l e^{-lx} \cdot (1-l) \cdot e^{-(1-l)Y} =$

$$= \prod_{i=1}^{N} (l - l^2) \cdot e^{-lx - Y + lY}$$

$$L = \sum_{i=1}^{N} \log(l - l^2) - lx - Y + lY$$

$$\frac{\delta L}{\delta l} = \sum_{i=1}^{N} \frac{1-2l}{l-l^2} - x + Y = 0$$

$$\Rightarrow N\left(\frac{1-2l}{l-l^2}\right) - \sum_{i=1}^{N} x_i - Y_i = 0 \Leftrightarrow N\left(\frac{1-2l}{l-l^2}\right) = \sum_{i=1}^{N} x_i - Y_i$$

FOR SIMPLICITY: $A = \sum_{i=1}^{N} -x_i + Y_i$

$$\Rightarrow N\left(\frac{1-2l}{l-l^2}\right) + A = 0 \Leftrightarrow N(1-2l) + (l-l^2)A = 0 \Leftrightarrow$$

$$\Leftrightarrow -Al^2 + (A - 2N)l + N = 0$$

APPLYING THE QUADRATIC EQUATION:

$$l_1 = \frac{2N - A - \sqrt{A^2 + 4N^2}}{-2A}$$

$$l_2 = \frac{2N - A + \sqrt{A^2 + 4N^2}}{-2A}$$

THE PARAMETER l THAT COINCIDES WITH THE PROPER DOMAIN OF $[0, +\infty)$ SHOULD BE CHOSEN (l_1).

$\boxed{\text{Ex. 2}}$

a) $\hat{\beta} = (X^T X)^{-1} X^T y$

$Y = x^T \beta + \varepsilon$ AND $Y = X\beta + \varepsilon$

$Y = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$

$X = (x_1, \dots, x_n)^T \in \mathbb{R}^{n \times d}$

$\varepsilon \sim N(0, \sigma^2 I) \in \mathbb{R}^n$

- Show that $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \cancel{(X^T X)^{-1} X^T X}\, \beta + (X^T X)^{-1} X^T \varepsilon$$

$$= \beta + (X^T X)^{-1} X^T \varepsilon$$

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[\beta + (X^T X)^{-1} X^T \varepsilon] =$$

$$= \mathbb{E}[\beta] + \mathbb{E}[(X^T X)^{-1} X^T \underbrace{\varepsilon}_{\mathbb{E}[\varepsilon] = 0}]$$

$$= \beta$$

$$\mathrm{VAR}(\hat{\beta}) = \mathbb{E}[(\hat{\beta} - \mathbb{E}[\hat{\beta}])^2] =$$

$$= \mathbb{E}[(\cancel{\beta} + (X^T X)^{-1} X^T \varepsilon - \cancel{\beta})(\cancel{\beta} + (X^T X)^{-1} X^T \varepsilon - \cancel{\beta})^T] =$$

$$= \mathbb{E}[(X^T X)^{-1} X^T \varepsilon ((X^T X)^{-1} X^T \varepsilon)^T] =$$

$$= \mathbb{E}[(X^T X)^{-1} X^T \underbrace{\varepsilon \varepsilon^T}_{\mathrm{Var}(\varepsilon) = \mathbb{E}[\varepsilon \varepsilon^T] = \sigma^2 I} X (X^T X)^{-1T}] =$$

$$= \mathbb{E}[(X^T X)^{-1} \cancel{X^T} \sigma^2 I \cancel{X} \underbrace{\cancel{(X^T X)}^{-1T}}_{(X^T X)^T = (X^T X)}]$$

$$= \sigma^2 \cdot \mathbb{E}[(X^T X)^{-1}] =$$

$$= \sigma^2 (X^T X)^{-1}$$

Hence: $\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

b) ———

c) $\hat{Y}_* = x_*^T \hat{\beta}$

$\hat{Y}_* = x_*^T (X^TX)^{-1} X^T Y$

$\hat{Y}_* = x_*^T (X^TX)^{-1} X^T (X\beta + \varepsilon)$

$\hat{Y}_* = x_*^T \cancel{(X^TX)^{-1} X^TX} \beta + x_*^T (X^TX)^{-1} X^T \varepsilon$

$\hat{Y}_* = x_*^T \beta + x_*^T (X^TX)^{-1} X^T \varepsilon$

$\hookrightarrow \mathbb{E}[\varepsilon] = 0$

$\mathbb{E}[\hat{Y}_*] = x_*^T \beta$

$\mathrm{VAR}(\hat{Y}_*) = \mathrm{VAR}[\underbrace{x_*^T \beta}_{=0} + x_*^T (X^TX)^{-1} X^T \varepsilon] =$

$= x_*^T (X^TX)^{-1} \cancel{X^T} \sigma^2 \cancel{I} \cancel{X} \cancel{(X^TX)^{-1}} x_* =$

$= x_*^T (X^TX)^{-1} x_* \cdot \sigma^2$

HENCE: $\hat{Y}_* \sim \mathcal{N}(x_*^T \beta \ , \ \sigma^2 x_*^T (X^TX)^{-1} x_*)$

d) ———