Maximum Likelihood Estimation and Model Fitting, CMEE MSc. Tin-Yu Hui

Practical 2 (23 Feb 2021)

Question 1

i. If X and Y are independent, then the joint pdf of X and Y is the product of their marginal pdf.

ii. Given data and a statistical model MLE provides estimates to the parameters of interest.

iii. [Circle the correct answer] The likelihood function is a function of parameters/data.

Question 2 [Marginal and conditional distributions]

Given a pair of r.v. X and Y with their joint pdf $f_{XY}(x,y) = y\left(\frac{1}{2} - x\right) + x$, where $0 < x < 1$ and $0 < y < 2$.

i. Show that $f_{XY}(x,y)$ is a valid joint pdf. To show $\iint f_{XY}(x,y)\,dx\,dy = 1$.

$$\int_0^2\left[\int_0^1 \left(y\left(\tfrac{1}{2}-x\right)+x\right)dx\right]dy$$

$$= \int_0^2 \int_0^1 \left(\tfrac{y}{2} - xy + x\right) dx\,dy$$

$$= \int_0^2 \left[\tfrac{yx}{2} - \tfrac{yx^2}{2} + \tfrac{x^2}{2}\right]_0^1 dy$$

$$= \int_0^2 \left[\left(\tfrac{y}{2} - \tfrac{y}{2} + \tfrac{1}{2}\right) - (0-0+0)\right] dy$$

$$= \int_0^2 \tfrac{1}{2}\,dy = 1$$

ii. Find the two marginal distributions, $f_X(x)$ and $f_Y(y)$. What can you say about the mean and variance of X?

$$f_X(x) = \int_0^2 f(x,y)\,dy$$

$$= \int_0^2 \left[y\left(\tfrac{1}{2}-x\right)+x\right] dy$$

$$= \left(\tfrac{1}{2}-x\right)\int_0^2 y\,dy + x\int_0^2 dy$$

$$= \left(\tfrac{1}{2}-x\right) \cdot \tfrac{y^2}{2}\Big|_0^2 + x(2-0)$$

$$= \left(\tfrac{1}{2}-x\right)\left(\tfrac{4}{2} - \tfrac{0}{2}\right) + 2x$$

$$= \left(\tfrac{1}{2}-x\right)(2) + 2x$$

$$= 1 - 2x + 2x = 1$$

$\boxed{E(X) = \tfrac{1}{2} \quad Var(X) = \tfrac{1}{12}}$ marginally

$\boxed{\text{Marginally, } X \text{ is uniformly distributed } X \sim uniform(0,1)}$

$$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 \left(\tfrac{1}{2}y - xy - x\right)dx$$

$$= \tfrac{1}{2}y\int_0^1 dx + (1-y)\int_0^1 x\,dx$$

$$= \tfrac{y}{2}(1-0) + (1-y)\,\tfrac{x^2}{2}\Big|_0^1$$

$$= \tfrac{y}{2} + (1-y)\left(\tfrac{1^2}{2} - 0\right)$$

$$= \tfrac{y}{2} + (1-y)\left(\tfrac{1}{2}\right)$$

$$= \tfrac{y}{2} + \tfrac{1}{2} - \tfrac{y}{2} = \tfrac{1}{2}$$

$\boxed{\text{Marginally, } Y \sim uniform(0,2)}$

iii. Find the conditional distribution of X given $Y = y$.

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = y(1-2x) + 2x$$

Not very informative here, but see below.

iv. Now assume we know $y = 1/2$, what is the conditional distribution of X?

$$f_{X|Y}(x|y=\tfrac{1}{2}) = \tfrac{1}{2}(1-2x) + 2x = \tfrac{1}{2} - x + 2x = \tfrac{1}{2} + x$$

v. Calculate the conditional mean and variance of X, given $y = 1/2$.

$$E(X|Y=\tfrac{1}{2}) = \int_0^1 x \cdot f_{X|Y}(x|y=\tfrac{1}{2})\,dx$$

$$= \int_0^1 x\left(\tfrac{1}{2} + x\right)dx$$

$$= \frac{1}{2}\int_0^1 x\,dx + \int_0^1 x^2\,dx$$

$$= \left(\tfrac{1}{2}\right)\frac{x^2}{2}\Big|_0^1 + \frac{x^3}{3}\Big|_0^1$$

$$= \tfrac{1}{2}\left(\tfrac{1}{2}\right) + \tfrac{1}{3}$$

$$= \tfrac{1}{4} + \tfrac{1}{3}$$

$$= \frac{3+4}{12}$$

$$= \frac{7}{12}$$

If we do not know Y, then $E(X) = \frac{1}{2}$ from the marginal distribution of X. Now, if we know the value of Y, $Y = \frac{1}{2}$, we have extra information of X, and the conditional mean has adjusted accordingly.

Q3 omitted.

```
z2<-rnorm(20000, mean=0, sd=sqrt(1-rho))

# TRANSFORM (z1, z2) INTO (x1, x2), VIA P-TRANSPOSE
x1<-z1/sqrt(2)+z2/sqrt(2)
x2<-z1/sqrt(2)-z2/sqrt(2)

# SCATTER PLOT OF (x1, x2)? LOOKS LIKE AN ELLIPSE?
plot(x1, x2)

# THE SAMPLE VARIANCE AND COVARIANCE OF x1 AND x2?
var(x1)
var(x2)
cov(x1, x2)
```

x. Conversely, we can back transform the correlated $\boldsymbol{X}$ into independent normal r.v., via $\boldsymbol{P}$ (i.e. $\boldsymbol{Z} = \boldsymbol{PX}$)

```
# LET'S REMOVE z1 AND z2 FROM OUR CURRENT R SESSION
rm(r1); rm(r2);

# WE BACK TRANSFORM (x1, x1) INTO (z1, z2), VIA P
z1<-x1/sqrt(2)+x2/sqrt(2)
z2<-x1/sqrt(2)-x2/sqrt(2)

# THE COVARIANCE BETWEEN z1 AND z2 SHOULD BE CLOSE TO ZERO
var(z1)
var(z2)
cov(z1, z2)
cor.test(z1, z2)
```

Question 4

i. $X_1, X_2, \ldots, X_n$ follow i.i.d. $Exponential(\lambda)$. What is the likelihood function $L(\lambda)$?

$L(\lambda) = f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$ ($\because$ independence)

$= (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2})\cdots(\lambda e^{-\lambda x_n})$ ($\because$ iid exponential)

$= \lambda^n e^{-\lambda \sum x_i}$

ii. Please also find the log-likelihood function $l(\lambda)$.

$l(\lambda) = \ln\left[\lambda^n e^{-\lambda \sum x_i}\right] = n\ln\lambda - \lambda \sum x_i$

iii. Find $\lambda = \hat{\lambda}$ such that the log-likelihood function is maximised.

$l'(\lambda) = \frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum x_i$

Find a $\hat{\lambda}$ s.t. $l'(\hat{\lambda}) = 0$

i.e. $\frac{n}{\hat{\lambda}} - \sum x_i = 0$

$$\frac{n}{\lambda} = \sum x_i$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Question 5

i. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $Poisson(\lambda)$. Find the MLE for λ.

$$L(\lambda) = f(x_1, x_2, \cdots, x_n) = f(x_1) f(x_2) \cdots f(x_n) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} \cdots \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

$$= \frac{\lambda^{\sum x_i} e^{-n\lambda}}{x_1! x_2! \cdots x_n!}$$

$$l(\lambda) = \ln\left(\frac{1}{x_1! x_2! \cdots x_n!}\right) + (\sum x_i) \ln \lambda - n\lambda$$

$$l'(\lambda) = 0 + \frac{\sum x_i}{\lambda} - n$$

Find $\hat{\lambda}$ s.t. $l'(\hat{\lambda}) = 0$ i.e. to solve $\frac{\sum x_i}{\hat{\lambda}} - n = 0$

$$\frac{\sum x_i}{\hat{\lambda}} = n \Rightarrow \hat{\lambda} = \frac{\sum x_i}{n}$$

ii. If I observed 5, 3, 2, and 6 events (independently), all within a fixed period of time, what would be the best guess for λ?

$$\hat{\lambda} = \frac{5+3+2+6}{4}$$

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

Question 6

The exercise left to you in class. Let $X_1, X_2, \dots, X_n$ be i.i.d. $N(\mu, \sigma^2)$, both μ and σ^2 are not known. Find the MLE for the two parameters. (Hints: Write down the likelihood and log-likelihood function. Differentiate (partially) the log-likelihood with respect to the two parameters. Set the derivatives to zero, and solve for the unknowns, ...)

$$L(\mu, \sigma^2) = f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \quad (\because \text{iid})$$
$$= \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}$$

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$
$$= -\frac{n}{2}\left[\ln(2\pi) + \ln(\sigma^2)\right] - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$
$$= \text{constant} - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2 \quad \text{---} \quad (*)$$

$$\frac{\partial \ell}{\partial \mu} = 0 + 0 - \frac{1}{2\sigma^2} 2 \sum(x_i-\mu)(-1)$$
$$= \frac{\sum(x_i-\mu)}{\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma^2} = 0 - \frac{n}{2\sigma^2} - \frac{\sum(x_i-\mu)^2}{2\sigma^4}(-1) = \frac{\sum(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2}$$

Find $(\hat{\mu}, \hat{\sigma}^2)$ s.t. $\left.\frac{\partial \ell}{\partial \mu}\right|_{\mu=\hat{\mu}} = \left.\frac{\partial \ell}{\partial \sigma^2}\right|_{\sigma^2=\hat{\sigma}^2} = 0$

i.e.
$$\begin{cases} \frac{\sum(x_i-\hat{\mu})}{\hat{\sigma}^2} = 0 & \text{---} \ (1) \\ \frac{\sum(x_i-\hat{\mu})^2}{2\hat{\sigma}^4} - \frac{n}{2\hat{\sigma}^2} = 0 & \text{---} \ (2) \end{cases}$$

From (1), $\frac{\sum(x_i-\hat{\mu})}{\hat{\sigma}^2} = 0$

$$\Rightarrow \sum(x_i - \hat{\mu}) = 0 \quad (\because \hat{\sigma}^2 \neq 0)$$
$$\Rightarrow \sum x_i - n\hat{\mu} = 0$$
$$\Rightarrow \hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

From (2) $\frac{\sum(x_i-\hat{\mu})^2}{2\hat{\sigma}^4} = \frac{n}{2\hat{\sigma}^2}$

$$\Rightarrow \sum(x_i-\hat{\mu})^2 = n\hat{\sigma}^2 \quad (\because \hat{\sigma}^2 \neq 0)$$
$$\hat{\sigma}^2 = \frac{\sum(x_i-\hat{\mu})^2}{n}$$
$$= \frac{\sum(x_i-\bar{x})^2}{n} \quad (\because \text{sub } \hat{\mu} = \bar{x})$$

Question ~~6~~ 7 [Linear regression exercise]

It is a spin-off exercise from the original marked-recapture experiment for census population size estimation. We measured the difference in their body lengths and how long (in days) they had been hanging around before falling back into our hands. We would like to investigate the relationship between the two variables. There is a short note on the use of `optim()` in today's presentation. You will need the dataset `recapture.csv`