Statistics HW04 Solution

Chapter 8

Problem 7

(a)

$When\ q := 1-p\ \in (0,1),\ we\ have$

$$\mu_1 = E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty}\left(\frac{d}{dq}q^x\right) = p\frac{d}{dq}\left(\frac{1}{1-q}\right) = \boxed{\frac{1}{p}}$$

$Check\ LN,\ chapter 1 \sim 6,\ page\ 62.$

$To\ obtain\ an\ moment\ estimator\ of\ p,\ denoted\ by\ \hat{p}_{MME}:$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n = \hat{\mu}_1 \overset{SET}{=} \frac{1}{\hat{p}_{MME}}$$

$Thus,\ we\ find\ an\ MME\ of\ p:$

$$\boxed{\hat{p}_{MME} = \frac{1}{\overline{X}_n}}$$

$$It\ is\ easy\ to\ see\ that \begin{cases} \hat{p}_{MME}\ is\ not\ defined\ (or\ meaningless)\ because\ \overline{X}_n\ is\ always\ 0 & ,\ if\ p=0 \\ \hat{p}_{MME} = 1\ because\ \overline{X}_n\ is\ always\ 1 & ,\ if\ p=1 \end{cases}.$$

$MME\ is\ not\ unique\ and\ different\ answers\ may\ be\ calculated\ using\ other\ moments\ (such\ as\ the\ 2^{nd}\ moment).$

$\left(INTENTIONALLY\ LEFT\ BLANK\right)$

(b)

The likelihood function of our sample $\boxed{\underset{\sim}{x} = \left(x_1, \dots, x_n\right)}$ *is*

$$\mathscr{L}(p|\underset{\sim}{x}) = f(\underset{\sim}{x}|p) = \prod_{i=1}^{n} f(x_i|p) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

So the log − likelihood $\ell(p)$ *is*

$$\ell(p) = log\left(\mathscr{L}(p|\underset{\sim}{x})\right) = n \times log(p) + \left(\sum_{i=1}^{n} x_i - n\right) \times log(1-p)$$

To find the MLE of p, *denoted by* $\hat{p}_{MLE}$:

$$\frac{d}{dp}\ell(p)\Big|_{p=p^*} = \left(\frac{n}{p} - \frac{\sum_{i=1}^n x_i - n}{1-p}\right)\Bigg|_{p=p^*} \overset{SET}{=} 0 \implies p^* = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{X}_n}$$

Check that p^* *is indeed a maximizer of* $\ell(p)$:

$$\frac{d^2}{dp^2}\ell(p) = \left(-\frac{n}{p^2} - \frac{\sum_{i=1}^n x_i - n}{\left(1-p\right)^2}\right)\Bigg|_{p=p^*} = -\boxed{\left(\frac{n}{{p^*}^2} + \frac{\sum_{i=1}^n x_i - n}{\left(1-p^*\right)^2}\right)} < 0 \qquad \boxed{\textit{Note that } \sum_{i=1}^{n} x_i - n \geq 0}$$

$$\boxed{\textit{Or notice } \frac{n}{{p^*}^2} + \frac{\sum_{i=1}^n x_i - n}{\left(1-p^*\right)^2} = \frac{\left(n\left(\left(1-p^*\right)^2\right) + \left(\sum_{i=1}^n x_i - n\right){p^*}^2\right)}{{p^*}^2\left(1-p^*\right)^2} = \frac{n - (n^2/\sum_{i=1}^n x_i)}{{p^*}^2\left(1-p^*\right)^2} \geq 0.}$$

Thus, we find

$$\hat{p}_{MLE} = \frac{1}{\overline{X}_n}$$

<u>*FYI :*</u>

We usually put a tilde under a $\begin{cases} \textit{lowercase} \\ \textit{uppercase} \end{cases}$ *letter to emphasize that it actually represents a* $\begin{cases} \textit{realization of a random vector} \\ \textit{random vector} \end{cases}$.

You may see bold letter like $\mathbf{x}$ *in textbooks, which means the same thing.*
But it is more convenient to add a tilde underneath when writing.

Problem 9

Let's start with a more general situation and return to §8.4 Example A later.

When the sample is from a population (unknown), which is described by the density $f(\cdot|\underset{\sim}{\theta})$, *the knowledge of* $\underset{\sim}{\theta}$ *(unknown) yields knowledge of the entire population.*

We would like to find a good estimator for $\underset{\sim}{\theta}$ *so that we can roughly guess the population distribution.*

An estimator is a function of the sample, while an estimate is the realization of an estimator which is obtained once the sample is actually drawn.

The estimator is random because samples are drawn randomly.

$\Longrightarrow$ *An estimator is a random variable, while an estimate is actually a number.*

In §8.4 Example A, $\overline{X}$ *is the estimator and its estimate,* $\overline{x}$, *is observed to be 24.9, the author is not wrong.*

(If we do the same experiment again and again, we may observe $\overline{x}$ *to be 25.1, 24.5, 25.3, ... and so on.)*

Problem 16

(a)

Note that $f(\cdot|\sigma)$ *is symmetric about 0, so* $E(X)=0$, *hence we cannot use* $\overline{X}_n$ *to find any MME.*

So we use the integration by parts (**IBP**) *technique* ($\int_a^b u\,dv = uv\big|_a^b - \int_a^b v\,du$) *to calculate* $E(X^2)$:

$$
\begin{aligned}
\mu_2 = E(X^2) &= \int_{-\infty}^{\infty} \frac{x^2}{2\sigma} exp\left(\frac{-|x|}{\sigma}\right) dx \\
&= \underbrace{\int_{-\infty}^{0} \frac{x^2}{2\sigma} exp\left(\frac{x}{\sigma}\right) dx}_{(\star)} + \underbrace{\int_{0}^{\infty} \frac{x^2}{2\sigma} exp\left(\frac{-x}{\sigma}\right) dx}_{(\star\star)} \\
&\overset{(1)}{=} \left(\left(\frac{x^2}{2\sigma}\sigma exp\left(\frac{x}{\sigma}\right)\right)\Bigg|_{-\infty}^{0} - \underbrace{\int_{-\infty}^{0} x exp\left(\frac{x}{\sigma}\right) dx}_{(\diamond)}\right) + \left(\left(\frac{-x^2}{2\sigma}\sigma exp\left(-\frac{x}{\sigma}\right)\right)\Bigg|_{0}^{\infty} + \underbrace{\int_{0}^{\infty} x exp\left(-\frac{x}{\sigma}\right) dx}_{(\diamond\diamond)}\right) \\
&\overset{(2)}{=} \left(\left(-x\sigma exp\left(\frac{x}{\sigma}\right)\right)\Bigg|_{-\infty}^{0} + \int_{-\infty}^{0} \sigma exp\left(\frac{x}{\sigma}\right) dx\right) + \left(\left(-x\,\sigma exp\left(-\frac{x}{\sigma}\right)\right)\Bigg|_{0}^{\infty} - \int_{0}^{\infty} \sigma exp\left(-\frac{x}{\sigma}\right) dx\right) \\
&= \left(\sigma^2 exp\left(\frac{x}{\sigma}\right)\right)\Bigg|_{-\infty}^{0} + \left(-\sigma^2 exp\left(-\frac{x}{\sigma}\right)\right)\Bigg|_{0}^{\infty} \\
&= 2\sigma^2
\end{aligned}
$$

$$(1)\begin{cases} (\star): do\ \textbf{IBP}\ with \begin{pmatrix} u=\frac{x^2}{2\sigma} & du=\frac{x}{\sigma}dx \\ v=\sigma exp\left(\frac{x}{\sigma}\right) & dv=exp\left(\frac{x}{\sigma}\right)dx \end{pmatrix} & over\ \left\{x\in\mathscr{R}\ \middle|\ -\infty<x<0\right\} \\ (\star\star): do\ \textbf{IBP}\ with \begin{pmatrix} u=\frac{x^2}{2\sigma} & du=\frac{x}{\sigma}dx \\ v=-\sigma exp\left(\frac{-x}{\sigma}\right) & dv=-exp\left(\frac{-x}{\sigma}\right)dx \end{pmatrix} & over\ \left\{x\in\mathscr{R}\ \middle|\ 0\le x<\infty\right\} \end{cases}$$

$$(2)\begin{cases} (\diamond): do\ \textbf{IBP}\ with \begin{pmatrix} u=x & du=dx \\ v=\sigma exp\left(\frac{x}{\sigma}\right) & dv=exp\left(\frac{x}{\sigma}\right)dx \end{pmatrix} & over\ \left\{x\in\mathscr{R}\ \middle|\ -\infty<x<0\right\} \\ (\diamond\diamond): do\ \textbf{IBP}\ with \begin{pmatrix} u=x & du=dx \\ v=-\sigma exp\left(\frac{-x}{\sigma}\right) & dv=exp\left(\frac{-x}{\sigma}\right)dx \end{pmatrix} & over\ \left\{x\in\mathscr{R}\ \middle|\ 0\le x<\infty\right\} \end{cases}$$

To find an MME of σ, denoted by $\hat{\sigma}_{MME}$:

$$\frac{1}{n}\sum_{i=1}^{n}X_i^2=\hat{\mu}_2\overset{SET}{=}2\hat{\sigma}^2_{MME}\implies \boxed{\hat{\sigma}_{MME}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}}$$

MME is not unique and different answers may be calculated using other moments (such as the 4^{th} *moment).*

(b)

The likelihood function of our sample $\underset{\sim}{x}=\left(x_1,\dots,x_n\right)$ *is*

$$\mathscr{L}(\sigma|\underset{\sim}{x})=f(\underset{\sim}{x}|\sigma)=\prod_{i=1}^{n}f(x_i|\sigma)=\prod_{i=1}^{n}\frac{1}{2\sigma}exp\left(-\frac{|x_i|}{\sigma}\right)$$

So the log – likelihood, say $\ell(\sigma)$, *is*

$$\ell(\sigma)=log\Big(\mathscr{L}(\sigma|\underset{\sim}{x})\Big)=\sum_{i=1}^{n}\left(-\left(log(2)+log(\sigma)-\frac{|x_i|}{\sigma}\right)\right)$$

To find the MLE of σ, say $\hat{\sigma}_{MLE}$:

$$\frac{d}{d\sigma}\ell(\sigma)\Big|_{\sigma=\sigma^*}=\left(-\frac{n}{\sigma}+\frac{\sum_{i=1}^{n}|x_i|}{\sigma^2}\right)\Bigg|_{\sigma=\sigma^*}\overset{SET}{=}0\implies\sigma^*=\frac{\sum_{i=1}^{n}|x_i|}{n}$$

Check that σ^* *is indeed a maximizer of* $\ell(\sigma)$:

$$\frac{d^2}{d\sigma^2}\ell(\sigma)=\left(\frac{n}{\sigma^2}-\frac{2\sum_{i=1}^{n}|x_i|}{\sigma^3}\right)\Bigg|_{\sigma=\sigma^*}=\frac{1}{\sigma^{*3}}\left(n\sigma^*-2\sum_{i=1}^{n}|x_i|\right)=-\frac{\sum_{i=1}^{n}|x_i|}{\sigma^{*3}}<0$$

Thus, we find

$$\hat{\sigma}_{MLE}=\frac{\sum_{i=1}^{n}|X_i|}{n}$$

Problem 18

(a)

Since $E(X) = \frac{1}{3}$ *has nothing to do with* α*, we cannot use* $\frac{1}{n}\sum_{i=1}^{n} X_i$ *to estimate* α.

Thus, we try to estimate α *by the* 2^{nd} *sample moment,* $\frac{1}{n}\sum_{i=1}^{n} X_i^2$*. Note that*

$$\mu_2 = E(X^2) = Var(X) + \Big(E(X)\Big)^2 = \frac{2}{9(3\alpha+1)} + \frac{1}{9}$$

Using the method of moments to estimate α*, denoted by* $\hat{\alpha}_{MME}$*, we require*

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\mu}_2 \overset{SET}{=} \frac{2}{9(3\hat{\alpha}_{MME}+1)} + \frac{1}{9}$$

Thus, we find an MME of α :

$$\hat{\alpha}_{MME} = \frac{2}{3}\left(\frac{9}{n}\sum_{i=1}^{n} X_i^2 - 1\right)^{-1} - \frac{1}{3}$$

(b)

The likelihood function of our sample $\underset{\sim}{x} = \Big(x_1, \dots, x_n\Big)$ *is*

$$\mathscr{L}(\alpha|\underset{\sim}{x}) = f(\underset{\sim}{x}|\alpha) = \prod_{i=1}^{n} f(x_i|\alpha) = \prod_{i=1}^{n} \frac{\Gamma(3\alpha)}{\Gamma(\alpha)\Gamma(2\alpha)} x^{\alpha-1}(1-x)^{2\alpha-1}$$

So the log − likelihood $\ell(\alpha)$ *is*

$$\ell(\alpha) = log\Big(\mathscr{L}(\alpha|\underset{\sim}{x})\Big) = \sum_{i=1}^{n}\Big(log(\Gamma(3\alpha)) - log(\Gamma(\alpha)) - log(\Gamma(2\alpha)) + (\alpha-1)log(x_i) + (2\alpha-1)log(1-x_i)\Big)$$

To find the MLE of α*, denoted by* $\hat{\alpha}_{MLE}$:

$$\frac{d}{d\alpha}\ell(\alpha)\Big|_{\alpha=\alpha^*} = \left(n\frac{3\Gamma'(3\alpha)}{\Gamma(3\alpha)} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\frac{2\Gamma'(2\alpha)}{\Gamma(2\alpha)} + \sum_{i=1}^{n}\Big(log(x_i) + 2log(1-x_i)\Big)\right)\Bigg|_{\alpha=\alpha^*} \overset{SET}{=} 0$$

So the MLE of α *must satisfy*

$$\boxed{\left(\frac{3\Gamma'(3\alpha)}{\Gamma(3\alpha)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{2\Gamma'(2\alpha)}{\Gamma(2\alpha)}\right)\Bigg|_{\alpha=\hat{\alpha}_{MLE}} = -\frac{1}{n}\sum_{i=1}^{n}\Big(log(x_i) + 2log(1-x_i)\Big)}$$

where $\hat{\alpha}_{MLE}$ *is the ML estimate of* α*, which is not a random variable when* $\underset{\sim}{x}$ *is observed.*

Note that an MLE must satisfy this equation does NOT imply that every value satisfying this equation will be an MLE, the concavity of $\ell(\alpha)$ *should be further calculated!*

Problem 21

(a)

$Let\ Y = X - \theta,\ then\ f_Y(y) = e^{-y}\mathbf{I}_{[0,\infty)}(y),\ i.e.,\ Y \sim Exp(1).$

$So\ E(Y) = E(X-\theta) = E(X) - \theta = 1,\ hence\ \mu_1 = E(X) = 1 + \theta.$

$To\ obtain\ an\ moment\ estimator\ of\ theta,\ denoted\ by\ \hat{\theta}_{MME}:$

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n = \hat{\mu}_1 \overset{SET}{=} 1 + \hat{\theta}_{MME}$$

$Thus,\ we\ find\ an\ MME\ of\ \theta:$

$$\hat{\theta}_{MME} = \overline{X}_n - 1$$

(b)

$Note\ that\ the\ support\ of\ X\ is\ not\ independent\ of\ \theta,\ so\ we\ cannot\ find\ the\ MLE\ by\ direct\ differentiation.$

$The\ likelihood\ function\ of\ our\ sample\ \underset{\sim}{x} = \Big(x_1, \dots, x_n\Big)\ is$

$$\mathscr{L}(\theta|\underset{\sim}{x}) = f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n}\Big(exp\Big(-(x_i-\theta)\Big)\Big)\mathbf{I}_{[\theta,\infty)}(x_i) = \Big(exp\Big(n\theta - \sum_{i=1}^{n} x_i\Big)\Big)\mathbf{I}_{[\theta,\infty)}(x_{(1)}),$$

$where\ \mathbf{I}\ is\ the\ indicator\ function,\ x_{(1)} = min\Big(x_1, \dots, x_n\Big),\ and\ the\ last\ equality\ holds\ because$

$$\Big\{\underset{\sim}{x}\ \Big|\ \theta \le x_i\ \forall i = 1, \dots, n\Big\} = \Big\{\underset{\sim}{x}\ \Big|\ \theta \le x_{(1)}\Big\}.$$

$So\ the\ log-likelihood\ \ell(\theta)\ is$

$$\ell(\theta) = log\Big(\mathscr{L}(\theta|\underset{\sim}{x})\Big) = \Big(n\theta - \sum_{i=1}^{n} x_i\Big)\mathbf{I}_{[\theta,\infty)}(x_{(1)})$$

$Since\ \frac{d}{d\theta}\ell(\theta) = n > 0,\ when\ \theta \le x_{(1)},\ i.e.,\ \ell(\theta)\ is\ an\ increasing\ function\ when\ \theta \le x_{(1)},\ we\ conclude\ that$

$$\hat{\theta}_{MLE} = X_{(1)}$$

$\Big(INTENTIONALLY\ LEFT\ BLANK\Big)$

Problem 26

$$\text{Let} \begin{cases} N & \text{be the size of the whole population (unknown but fixed)} \\ r & \text{be the size of the recaptured population (known)} \\ n & \text{be the size of the tagged population (known)} \\ N-n & \text{be the size of the untagged population (unknown but fixed)} \end{cases}.$$

Let X be the number of tagged animals among the r recaptured animals.
We may view this problem as observing (doing an experiment) $X \sim Hypergeometric\Big(r, n, N-n\Big)$, where $r = 50,\ n = 100$ and the pmf of X is

$$p_X(x|r,n) = \begin{cases} \dfrac{\binom{n}{x}\binom{N-n}{r-x}}{\binom{N}{r}} & , \text{ if } x \in \Big\{0, \dots, min(r,n)\Big\} \text{ and } r - x \le N - n \\ 0 & , \text{ otherwise} \end{cases}$$

and then we observed that $x = 20$ (after the experiment).

Based on this result (observed that $x = 20$), we try to estimate N by its MLE, $\widehat{N}$:

$$\widehat{N} = \underset{N \in \mathbb{Z}^+}{argmax}\, \mathscr{L}(N \mid x),\ \text{where } \mathscr{L}(N \mid x) = p_X(x|N) = \frac{\binom{n}{x}\binom{N-n}{r-x}}{\binom{N}{r}}$$

Since $\widehat{N}$ maximizes $\mathscr{L}(\cdot \mid x)$, we have

$$\begin{cases} \dfrac{\mathscr{L}(\widehat{N}+1 \mid x)}{\mathscr{L}(\widehat{N} \mid x)} = \left[\dfrac{\binom{n}{x}\binom{\widehat{N}+1-n}{r-x}}{\binom{\widehat{N}+1}{r}}\right] \Big/ \left[\dfrac{\binom{n}{x}\binom{\widehat{N}-n}{r-x}}{\binom{\widehat{N}}{r}}\right] = \dfrac{\Big(\widehat{N}+1-r\Big)\Big(\widehat{N}+1-n\Big)}{\Big(\widehat{N}+1\Big)\Big(\widehat{N}+1-n-r+x\Big)} \le 1 & (1) \\ \dfrac{\mathscr{L}(\widehat{N}-1 \mid x)}{\mathscr{L}(\widehat{N} \mid x)} = \left[\dfrac{\binom{n}{x}\binom{\widehat{N}-1-n}{r-x}}{\binom{\widehat{N}-1}{r}}\right] \Big/ \left[\dfrac{\binom{n}{x}\binom{\widehat{N}-n}{r-x}}{\binom{\widehat{N}}{r}}\right] = \dfrac{\Big(\widehat{N}-1-r\Big)\Big(\widehat{N}-1-n\Big)}{\Big(\widehat{N}-1\Big)\Big(\widehat{N}-1-n-r+x\Big)} \le 1 & (2) \end{cases}$$

$$\left.\begin{matrix} \overset{(1)}{\Longrightarrow} \widehat{N}_{MLE} \ge \dfrac{nr}{x} - 1 \\ \overset{(2)}{\Longrightarrow} \widehat{N}_{MLE} \le \dfrac{nr}{x} \end{matrix}\right\} \Longrightarrow \begin{cases} \dfrac{nr}{x} - 1 \text{ or } \dfrac{nr}{x} & , \text{ if } \dfrac{nr}{x} \in \mathbb{Z}^+ \cup \{0\} \\ \left\lfloor \dfrac{nr}{x} \right\rfloor & , \text{ if } \dfrac{nr}{x} \notin \mathbb{Z}^+ \cup \{0\} \end{cases},$$

$$\text{where } \lfloor k \rfloor := \begin{cases} k & , \text{ if } k \in \mathbb{Z} \\ max\{s \in \mathbb{Z} : s < k\} & , \text{ if } k \notin \mathbb{Z} \end{cases} \text{ is the floor function.}$$

Plugging all the known parameters into these inequalities and setting $x = 20$, we find

$$\frac{nr}{x} = \frac{100 \times 50}{20} = 250\ \in \mathbb{Z}^+,\ \text{so } \widehat{N}_{MLE} = 250 \text{ or } 249.$$

made by 林宸緯，馬翌翔 助教

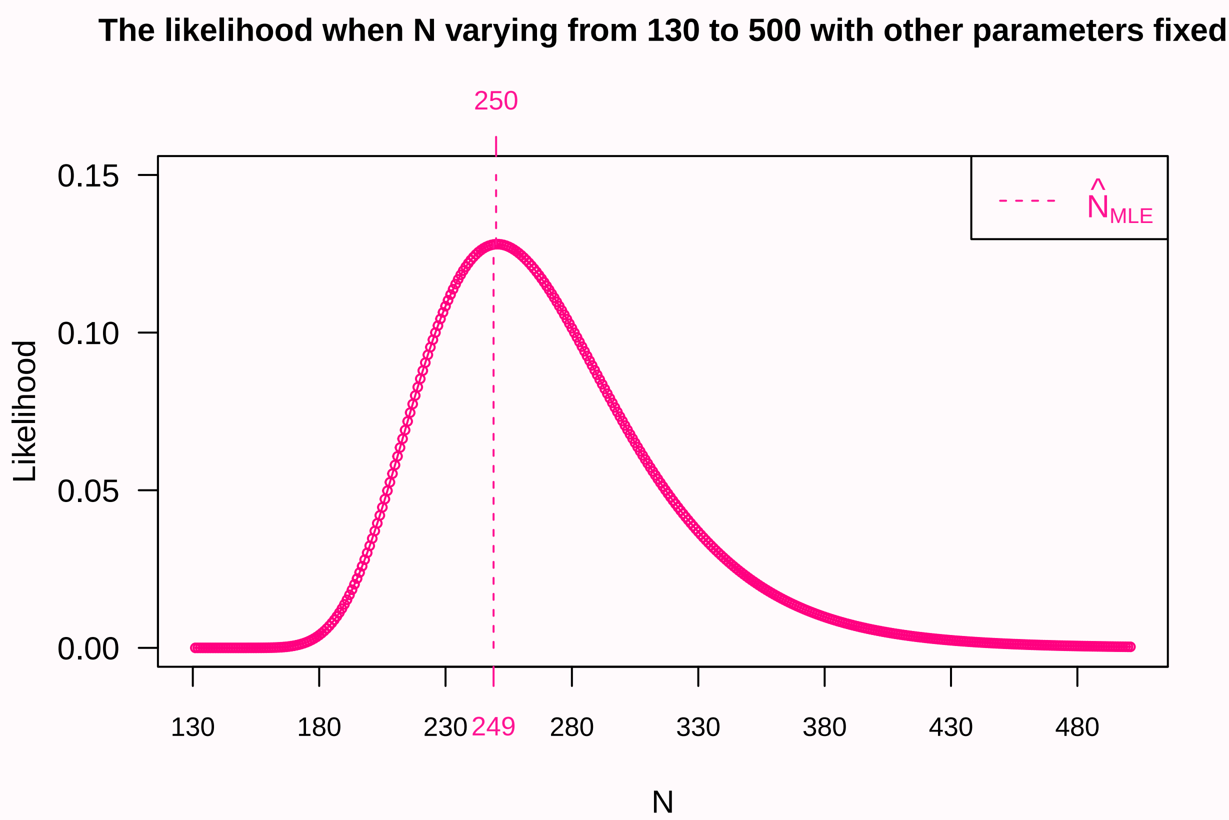
We can verify our answer by **R** :

```
# Define the likelihood function
likelihood <- function(N){
  m <- 100 # size of tagged population
  n <- N   # size of the whole population
  k <- 50  # size of the captured population
  x <- 20  # number of the tagged population among the captured ones
  return(dhyper(x, m, n-m, k))
}
N <- 130 # N should fall in a suitable range, which is determined by the support of the pmf
# Use the above inequalities to find the MLE of N
while(likelihood(N+1)/likelihood(N) > 1 | likelihood(N-1)/likelihood(N) > 1){
  N <- N+1
}
if(likelihood(N+1) == likelihood(N)) print(paste("The MLE we found are:", N, "and", N+1))
```

```
## [1] "The MLE we found are: 249 and 250"
```

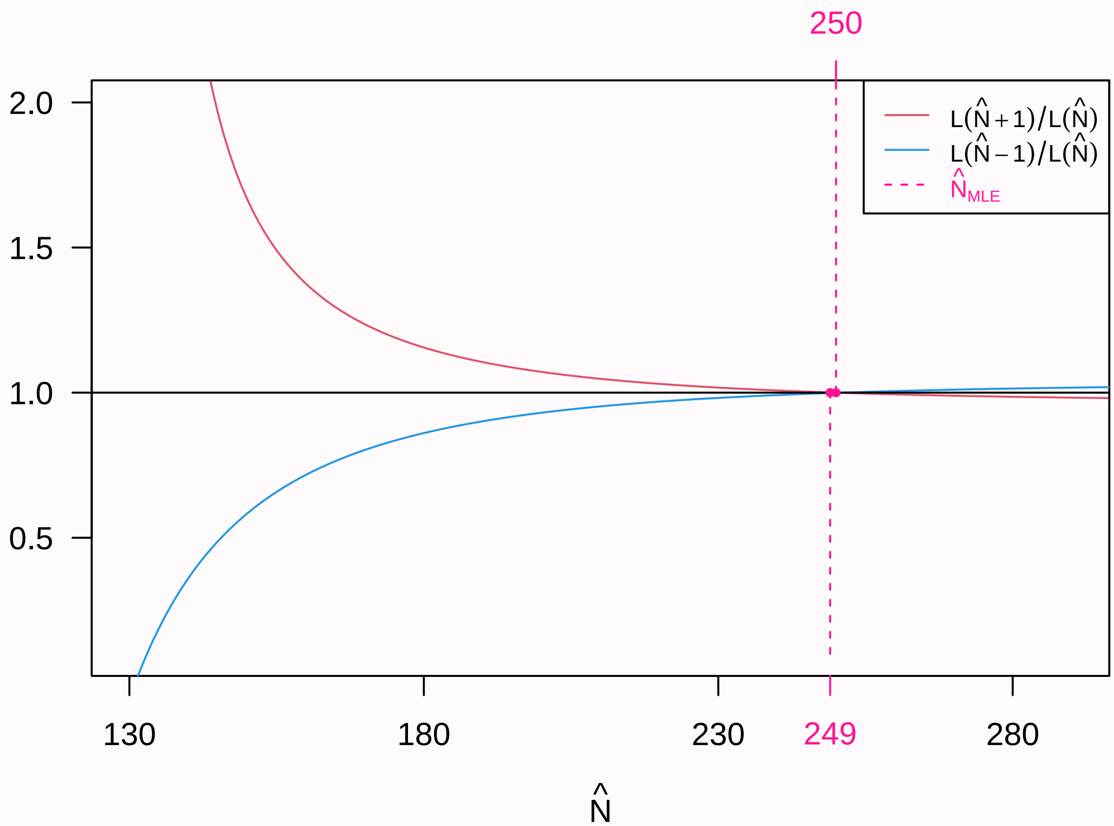
```
if(likelihood(N+1) != likelihood(N)) print(paste("The MLE we found is:", N))
```

Let's look at the likelihood function, note the different shades of color represents different speeds of change. (the code is omitted)

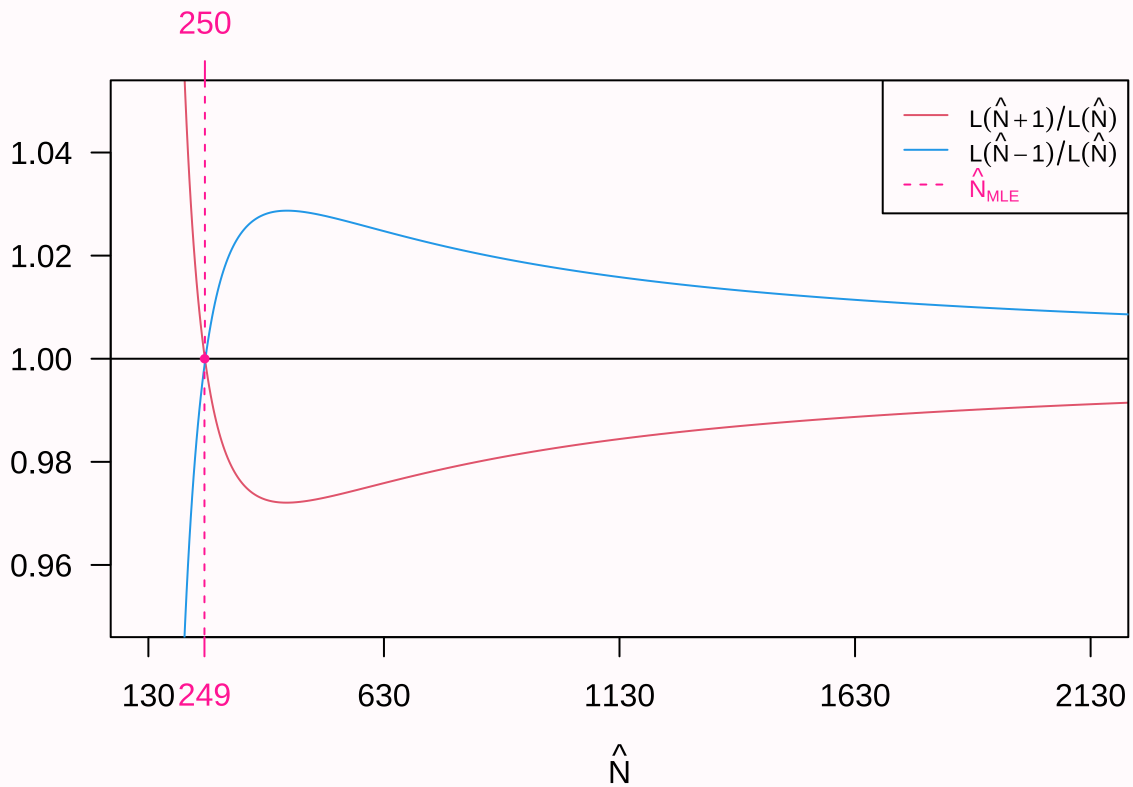


Comparing the behavior of the 2 ratios, they cross near 1 when $\widehat{N}_{MLE}$ *is plugged into these ratios :*

```
par(mar = c(4, 3, 3, 3))

plot(likelihood((130:380)+1)/likelihood(130:380), type = "l", xlim = c(0,160), ylim = c(0.1,2)
     , las = 1, col = 2, main = "", xlab = expression(hat(N)), ylab = "", xaxt = "n")

xticks <- c(seq(0, 150, by = 50))
axis(1, at = xticks, col.axis = 1, labels = xticks + 130)
axis(1, at = 119, labels = expression(bold(249)), col.ticks = "deeppink", col.axis = "deeppink")
axis(3, at = 120, labels = expression(bold(250)), col.ticks = "deeppink", col.axis = "deeppink")

lines(likelihood((130:380)-1)/likelihood(130:380), col = 4)
abline(h = 1, lwd = 1, lty = 1, col = 1)

points(c(119, 120), c(1, 1), col = "deeppink", pch = 19, cex = 0.5)

segments(119, 0.1, 119, 1, col = "deeppink", lwd = 1, lty = 2)
segments(120, 1, 120, 2.2, col = "deeppink", lwd = 1, lty = 2)

legend("topright",
       legend = c(expression(L(hat(N)+1)/L(hat(N))), expression(L(hat(N)-1)/L(hat(N)))
                  , expression(hat(N)[MLE])), lty = c(1, 1, 2), lwd = c(1, 1, 1)
       , col = c(2, 4, "deeppink"), cex = 0.75, text.col = c(1, 1, "deeppink"))
```



View the above plot on a smaller scale : (*the code is omitted*)



$$\textit{In fact, we can find that}\begin{cases}\textit{the red line}\left(\dfrac{\mathscr{L}(\widehat{N}+1\mid x)}{\mathscr{L}(\widehat{N}\mid x)}\right)=1 \textit{ at } \widehat{N}=249.\\[2ex] \textit{the blue line}\left(\dfrac{\mathscr{L}(\widehat{N}-1\mid x)}{\mathscr{L}(\widehat{N}\mid x)}\right)=1 \textit{ at } \widehat{N}=250.\end{cases}$$

(*INTENTIONALLY LEFT BLANK*)

made by 林宸緯，馬翊翔 助教

Problem 51

The likelihood function of our sample $\underset{\sim}{x} = \left(x_1, \ldots, x_n\right)$ *is*

$$\mathscr{L}(\theta|\underset{\sim}{x}) = f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{2^n} exp\Big(-\sum_{i=1}^{n}|x_i-\theta|\Big)$$

So maximizing $\mathscr{L}(\theta|\underset{\sim}{x})$ *is equivalent to minimizing* $h(\theta) := \sum_{i=1}^{n}|x_i-\theta|$, *hence* $\hat{\theta}_{MLE} = argmin_\theta\, h(\theta)$.

We can find (treating $\underset{\sim}{x}$ *as fixed in* $h(\cdot)$*) that* $h(\theta)$ *is a continuous function of* θ *and*

$$h(\theta) = \sum_{i=1}^{n}|x_i-\theta| = \begin{cases} -(2m+1)(\theta) + {}^*C_1 & , if\ \theta \in (-\infty, x_{(1)}) \\ -(2m)(\theta) + {}^*C_2 & , if\ \theta \in [x_{(1)}, x_{(2)}) \\ \vdots & \\ -(1)(\theta) + {}^*C_m & , if\ \theta \in [x_{(m)}, x_{(m+1)}) \\ C_{m+1} & , if\ \theta = x_{m+1} \\ (1)(\theta) + {}_*C_{m+2} & , if\ \theta \in (x_{(m+1)}, x_{(m+2)}] \\ \vdots & \\ (2m)(\theta) + {}_*C_{2m} & , if\ \theta \in (x_{(2m)}, x_{(2m+1)}] \\ (2m+1)(\theta) + {}_*C_{2m+1} & , if\ \theta \in (x_{2m+1}, \infty) \end{cases},$$

$$where \begin{cases} {}^*C_i = \sum_{j=i}^{2m+1} x_{(j)} - \sum_{j=1}^{i-1} x_{(j)} & , for\ i = 1, \ldots, m \\ C_{m+1} = \sum_{j=m+2}^{2m+1} x_{(j)} - \sum_{j=1}^{m} x_{(j)} & \\ {}_*C_i = \sum_{j=i+1}^{2m+1} x_{(j)} - \sum_{j=1}^{i} x_{(j)} & , for\ i = m+1, \ldots, 2m+1 \end{cases} \quad are\ some\ constants\ free\ of\ \theta.$$

We see that $\frac{d}{d\theta}h(\theta) \begin{cases} < 0 & , if\ \theta < x_{(m+1)} \\ = 0 & , if\ \theta = x_{(m+1)} \\ > 0 & , if\ \theta > x_{(m+1)} \end{cases}$, *so the minimum of* h *is attained when* $\theta = x_{(m+1)}$.

We should avoid those points that are not differentiable (h *is not differentiable at* $x_1, \ldots, x_{2m+1}$*).*

That is, $x_{(m+1)} = argmin_\theta\, h(\theta)$, *so*

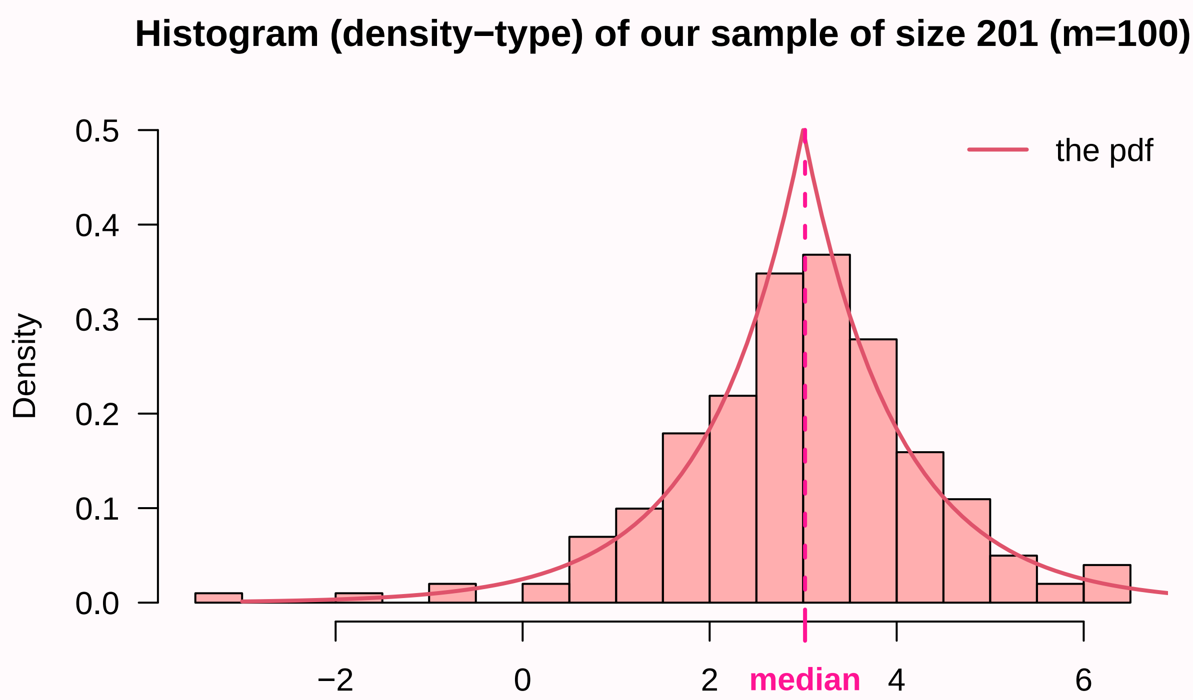
$$\hat{\theta}_{MLE} = X_{(m+1)}$$

made by 林宸緯，馬翌翔 助教

We can verify our answer by **R**, *consider the case* $m = 100,\ \theta = 3,\ \sigma = 1.$

Generate a sample of size 201 *and observe the sample :*

```
# Generate samples from the given distribution
rdouble_exp <- function(n, theta, sigma){
  U <- runif(n, -1, 1)
  X <- rexp(n, rate = 1/sigma)
  return(theta + sigma * sign(U) * X)
}
set.seed(0); n <- 100*2 + 1 # m = 100
x <- rdouble_exp(n, 3, 1) # theta = 3, sigma = 1
hist(x, breaks = 31, col = rgb(1, 0, 0, 0.3), prob = T, ylim = c(0, 0.5), las = 1,
     main = "Histogram (density-type) of our sample of size 201 (m=100)", xlab = "")
curve((exp(-abs(x-3)/1))/2, add = T, col = 2, lwd = 2, from = -3, to = 7, xlim = median(x)+c(-5,5))
axis(1, at = median(x), labels = expression(bold(median))
     , col.ticks = "deeppink", col.axis = "deeppink", lwd.ticks = 2)
abline(v = median(x), col = "deeppink", lwd = 2, lty = 2)
legend("topright", legend = "the pdf", lty = 1, lwd = 2, col = 2, bty = "n")
```



Some descriptive statistics of our sample : (*Note that the median is* 3.02)

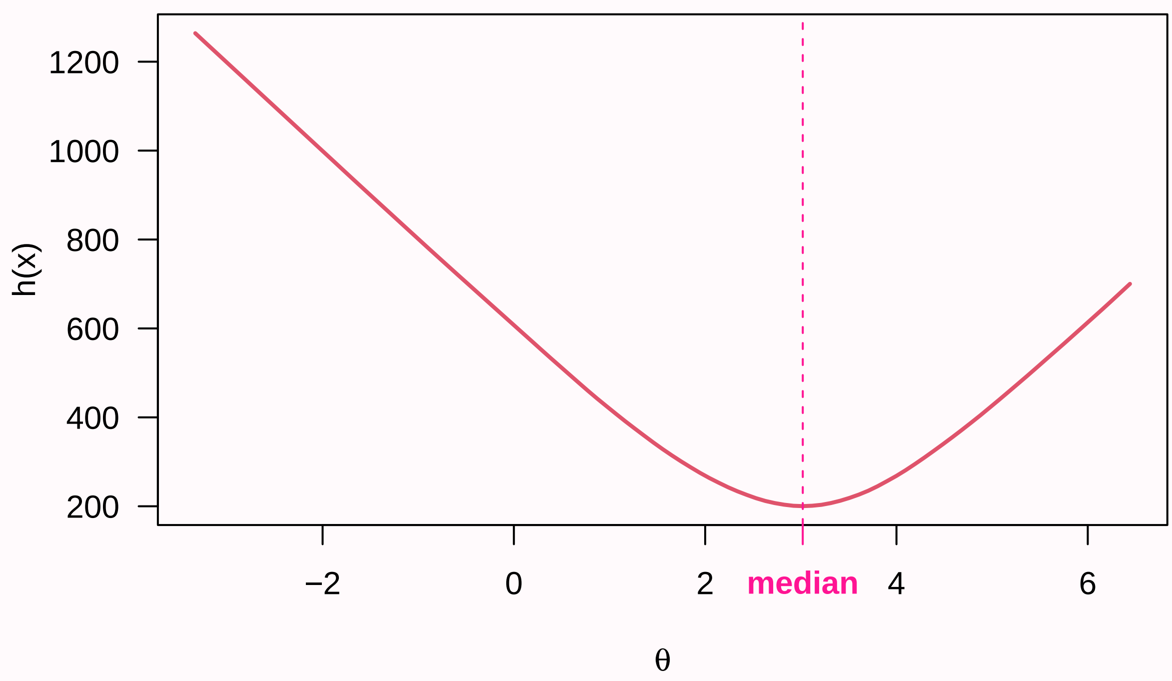
```
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -3.331   2.249   3.020   2.958   3.722   6.441
```

made by 林宸緯，馬翌翔 助教

We can also check $x_{(m+1)} = argmin_\theta h(\theta)$:

```
h <- function(theta){
  return(sapply(theta, function(t) sum(abs(x - t))))
}
curve(h, from = min(x), to = max(x), col = 2, las = 1, lwd = 2, xlab = expression(theta))
axis(1, at = 3.020, labels = expression(bold(median)), col.ticks = "deeppink", col.axis = "deeppink")
abline(v = median(x), col = "deeppink", lwd = 1, lty = 2)
```

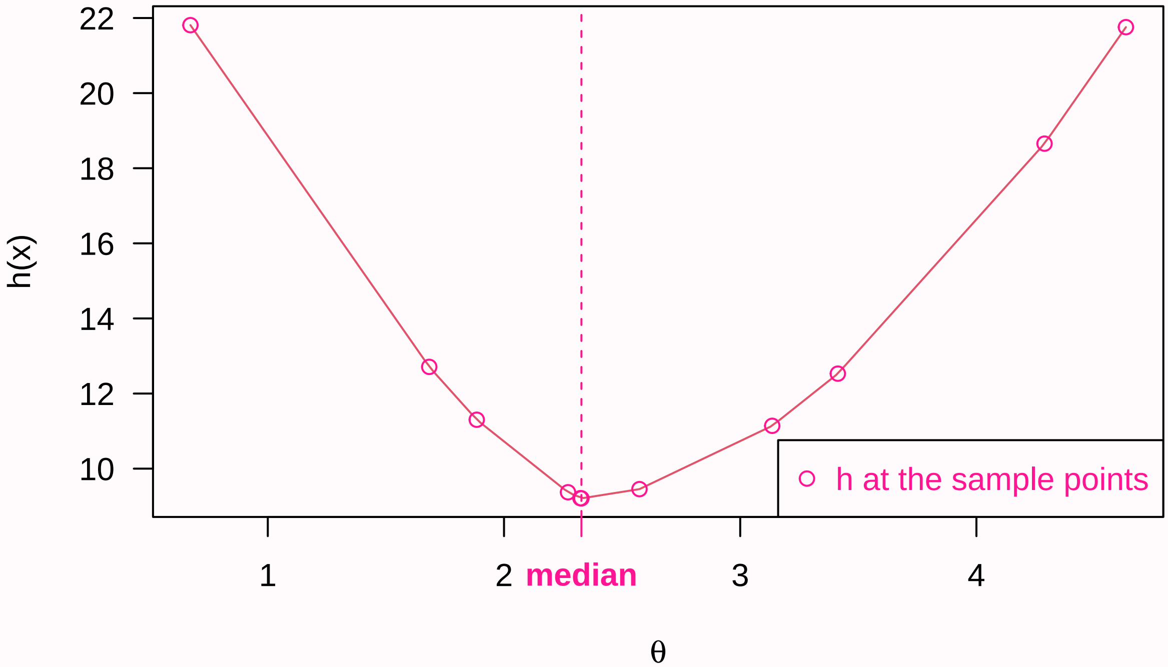


$\left(\textit{INTENTIONALLY LEFT BLANK} \right)$

made by 林宸緯，馬翌翔 助教

Note that h is not differentiable at x_i $\forall i = 1, \ldots, n$.

This can be seen in a small sample size case, we do another simulation with $m = 5$:



We see that h is uneven (hence not differentiable) at each sample point.

$\Big($*INTENTIONALLY LEFT BLANK*$\Big)$

made by 林宸緯，馬翌翔 助教

Problem 60

(a)

The likelihood function of our sample $\underset{\sim}{x} = \left(x_1, \dots, x_n\right)$ *is*

$$\mathscr{L}(\tau|\underset{\sim}{x}) = f(\underset{\sim}{x}|\tau) = \prod_{i=1}^{n} f(x_i|\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-x_i/\tau} = \frac{1}{\tau^n} exp\Big(-\sum_{i=1}^{n} x_i/\tau\Big)$$

So the log − likelihood $\ell(\tau)$ *is*

$$\ell(\tau) = log\Big(\mathscr{L}(\tau|\underset{\sim}{x})\Big) = -nlog(\tau) - \frac{1}{\tau}\sum_{i=1}^{n} x_i$$

To find the MLE of τ*, denoted by* $\hat{\tau}_{MLE}$:

$$\frac{d}{d\tau}\ell(\tau)\Big|_{\tau=\tau^*} = \left(-\frac{n}{\tau} + \frac{\sum_{i=1}^{n} x_i}{\tau}\right)\Bigg|_{\tau=\tau^*} \overset{SET}{=} 0 \implies \tau^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n$$

Check that τ^* *is indeed a maximizer of* $\ell(\tau)$:

$$\frac{d^2}{d\tau^2}\ell(\tau) = \left(\frac{n}{\tau^2} - \frac{2\sum_{i=1}^{n} x_i}{\tau^3}\right)\Bigg|_{\tau=\tau^*} = \frac{1}{\tau^{*2}}\left(n - \frac{2\sum_{i=1}^{n} x_i}{\tau^*}\right) = \frac{1}{\tau^{*2}}(n-2n) = -\frac{n}{\tau^{*2}} < 0$$

Thus, we find

$$\hat{\tau}_{MLE} = \overline{X}_n$$

(b)

We may use MGF to solve this problem.

Since $X_1, \dots, X_n \overset{i.i.d.}{\sim} Exp(\frac{1}{\tau})$*, the mgf of* X_i *is*

$$\boxed{M_{X_i}(t) = \frac{1}{1-\tau t},\ where\ t < \frac{1}{\tau},\ \forall i \in \Big\{1, \dots, n\Big\}} \qquad (\star)$$

So the mgf of $\hat{\tau}_{MLE}$ *is*

$$M_{\hat{\tau}_{MLE}}(s) = E\Big(e^{\overline{X}_n\ s}\Big) = E\left(exp\Big(\frac{s}{n}\sum_{i=1}^{n} X_i\Big)\right) = \prod_{i=1}^{n}\left(E\left(exp\Big(X_i\frac{s}{n}\Big)\right)\right) = \prod_{i=1}^{n} M_{X_i}(\frac{s}{n})$$

$$\overset{(\star)}{=} \left(\frac{1}{1-(\tau \times \frac{s}{n})}\right)^n = \left(\frac{\frac{n}{\tau}}{\frac{n}{\tau} - s}\right)^n, \ \boxed{where\ s < \frac{n}{\tau}}$$

$$\boxed{Since\ the\ last\ equality\ holds\ when\ \frac{s}{n} < \frac{1}{\tau},\ as\ required\ in\ (\star).}$$

made by 林宸緯，馬翌翔 助教

By the uniqueness of the MGF, we conclude that

$$\boxed{\hat{\tau}_{MLE} \sim Gamma\Big(n, \frac{n}{\tau}\Big) \text{ with } E(\hat{\tau}_{MLE}) = \tau}$$

$$\boxed{\text{Recall that if } W \sim Gamma(\alpha, \beta) \text{ with } E(W) = \frac{\alpha}{\beta}, \text{ then its mgf is } M_W(s) = \Big(\frac{\beta}{\beta - s}\Big)^{\alpha}, \text{ where } s < \beta.}$$

We can reach the same conclusion using propositions 4 and 6 at Lecture Notes, Chapter1 ∼ 6, page 74.

(c)

We have

$$(\star\star)\begin{cases} E(\hat{\tau}_{MLE}) = E(\overline{X}_n) & = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} E(X_i) & = \dfrac{1}{n} \times n \times \tau & = \tau \\ Var(\hat{\tau}_{MLE}) = Var(\overline{X}_n) & = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} Var(X_i) & = \dfrac{1}{n^2} \times n \times \tau^2 & = \dfrac{\tau^2}{n} \end{cases}$$

So the Central Limit Theorem implies that

$$\frac{\overline{X}_n - \tau}{\tau/\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1)$$

Thus, we find a normal approximation to the sampling distribution:

$$\overline{X}_n \xrightarrow{d} Z^* \sim N\Big(\tau, \frac{\tau^2}{n}\Big)$$

FYI:

Although it would not be confusing to write $\dfrac{\overline{X}_n - \tau}{\tau/\sqrt{n}} \xrightarrow{d} N(0, 1)$, *it is more correct and accurate to write* $\dfrac{\overline{X}_n - \tau}{\tau/\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1)$.

(d)

By $(\star\star)$ *in part (c), we have*

$$\begin{cases} E(\hat{\tau}_{MLE}) = E(\overline{X}_n) & = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} E(X_i) & = \dfrac{1}{n} \times n \times \tau & = \tau \\ Var(\hat{\tau}_{MLE}) = Var(\overline{X}_n) & = \dfrac{1}{n^2}\displaystyle\sum_{i=1}^{n} Var(X_i) & = \dfrac{1}{n^2} \times n \times \tau^2 & = \dfrac{\tau^2}{n} \end{cases},$$

which shows that $\hat{\tau}_{MLE}$ *is unbiased.*