Applied Data Analysis

Exercise Sheet 5 - Solutions

Exercise 18

(a) According to the first hint, applying a Taylor expansion (or the definition of differentiability of vector valued functions) to the function g around μ, we obtain:

$$(1) \qquad g(\boldsymbol{x}) \;=\; g(\boldsymbol{\mu}) + (D_g(\boldsymbol{\mu}))'\,(\boldsymbol{x}-\boldsymbol{\mu}) + r_{\boldsymbol{\mu}}(\boldsymbol{x})\,||\boldsymbol{x}-\boldsymbol{\mu}|| \;,\; \boldsymbol{x}\in\mathbb{R}^p \;,$$

with a function $r_{\boldsymbol{\mu}} : \mathbb{R}^p \longrightarrow \mathbb{R}^q$ fulfilling:

$$(2) \qquad \lim_{\boldsymbol{x}\to\boldsymbol{\mu}} r_{\boldsymbol{\mu}}(\boldsymbol{x}) \;=\; \mathbf{0}$$

Further, by the assumptions of Exercise 18, it holds with $\boldsymbol{Z} \sim \mathcal{N}_p(\mathbf{0},\Sigma)$:

$$(3) \qquad \sqrt{n}\,\big(\boldsymbol{X}_n - \boldsymbol{\mu}\big) \quad \xrightarrow{d} \quad \boldsymbol{Z} \quad \text{for} \quad n \longrightarrow \infty$$

Then, applying the continuous mapping to the continuous norm, it follows by (3):

$$(4) \qquad \big|\big|\sqrt{n}\,\big(\boldsymbol{X}_n - \boldsymbol{\mu}\big)\big|\big| \quad \xrightarrow{d} \quad ||\boldsymbol{Z}|| \quad \text{for} \quad n \longrightarrow \infty$$

Further, we obtain by (3) applying *Slutsky*'s Lemma (hint 2):

$$(5) \qquad \boldsymbol{X}_n - \boldsymbol{\mu} \;=\; \underbrace{\sqrt{n}\,\big(\boldsymbol{X}_n - \boldsymbol{\mu}\big)}_{\xrightarrow{d}\boldsymbol{Z}}\;\underbrace{\frac{1}{\sqrt{n}}}_{\longrightarrow 0} \quad \xrightarrow{d} \quad 0\cdot\boldsymbol{Z} \;=\; \mathbf{0} \quad \text{for} \quad n \longrightarrow \infty$$

Since convergence in distribution to a constant (!) is equivalent to convergence in probability to that constant, (5) yields:

$$(6) \qquad \boldsymbol{X}_n - \boldsymbol{\mu} \quad \xrightarrow{P} \quad \mathbf{0} \quad \text{for} \quad n \longrightarrow \infty$$

(with P denoting the underlying probability measure).

Then, again applying *Slutsky*'s Lemma, we finally obtain for $n \longrightarrow \infty$:

$$(7) \quad \sqrt{n}\,\big(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})\big) \;\overset{(1)}{=}\; \underbrace{(D_g(\boldsymbol{\mu}))'}_{\text{const.}}\;\underbrace{\sqrt{n}\,(\boldsymbol{X}_n - \boldsymbol{\mu})}_{\xrightarrow{d}\boldsymbol{Z}\text{ by (3)}} + \underbrace{r_{\boldsymbol{\mu}}(\boldsymbol{X}_n)}_{\xrightarrow{P}\mathbf{0}\text{ by (2) and (6)}}\;\underbrace{\big|\big|\sqrt{n}\,(\boldsymbol{X}_n - \boldsymbol{\mu})\big|\big|}_{\xrightarrow{d}||\boldsymbol{Z}||\text{ by (4)}}$$

$$\xrightarrow[\text{Slutsky}]{d} \; (D_g(\boldsymbol{\mu}))'\,\boldsymbol{Z} \;\overset{\text{Choice of }\boldsymbol{Z}}{\underset{\text{I.2.11}}{=}}\; \mathcal{N}_q\big(\mathbf{0},(D_g(\boldsymbol{\mu}))'\,\Sigma\,D_g(\boldsymbol{\mu})\big)$$

Remark:

Slutsky's Lemma and the continuous mapping theorem (for the different kinds of convergence) can (for example) be found in R.J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, 1980, Sections 1.5 and 1.6.

(b) (i) A solution to this part of Exercice 18 and/or corresponding remarks to the statements(s) to be proven, will follow.

(ii) First, with $\pi \in (0,1)$, it holds for $n \in \mathbb{N}$:

$$(8) \qquad P(X_n = 0) \overset{\text{Ass.}}{=} (1-\pi)^n \; > \; 0$$

(with P denoting the underlying probability measure).

By (8), we obtain for $n \in \mathbb{N}$:

$$(9) \qquad \mathrm{E}(|Y_n|) \overset{\text{Def.}}{=} \sum_{i=0}^{n} \left| \ln\left(\frac{i}{n}\right) \right| P(X_n = i) \; \geq \; \underbrace{\left| \ln\left(\frac{0}{n}\right) \right|}_{=|\ln(0)|=\infty} \underbrace{P(X_n = 0)}_{>0} \; = \; \infty$$

Thus, for $n \in \mathbb{N}$, even the first moment of Y_n does *not* exist implying that also the second moment and the variance of Y_n does *not* exist.

Note, however, that with

$$(10) \qquad \lim_{n\to\infty} P(X_n = 0) \overset{(8)}{=} \lim_{n\to\infty} (1-\pi)^n \; = \; 0$$

Y_n is well defined in the limit.

Now, let us consider the asymptotic variance of Y_n for $n \to \infty$.

According to the central limit theorem, it holds:

$$(11) \qquad \sqrt{n}\left(\frac{X_n}{n} - \pi\right) \; = \; \frac{X_n - n\,\pi}{\sqrt{n}} \quad \overset{d}{\longrightarrow} \quad Z \; \sim \; \mathcal{N}\big(0, \pi\,(1-\pi)\big)$$

Further, $g : (0,\infty) \longrightarrow \mathbb{R}$, defined by

$$(12) \qquad g(x) \; := \; \ln(x) \;\; , \;\; x \in (0,\infty) \; ,$$

is continuously differentiable on $(0,\infty)$ with

$$(13) \qquad g'(x) \; = \; \frac{1}{x} \;\; , \;\; x \in (0,\infty)$$

Applying the (univariate) delta method to (11), we obtain for $n \longrightarrow \infty$:

$$(14) \qquad \sqrt{n}\,\big(Y_n - \ln(\pi)\big) \underset{(12)}{\overset{\text{Def. } Y_n}{=}} \sqrt{n}\,\left(g\left(\frac{X_n}{n}\right) - g(\pi) \right)$$

$$\overset{d}{\longrightarrow} \; g'(\pi)\,Z \overset{(13)}{=} \frac{1}{\pi}\,Z \overset{(11)}{\sim} \mathcal{N}\left(0, \frac{1-\pi}{\pi}\right)$$

Thus, for sufficiently large $n \in \mathbb{N}$, we obtain:

$$(15) \qquad \mathrm{Var}\big(\sqrt{n}\,(Y_n - \ln(\pi))\big) \; \approx \; \frac{1-\pi}{\pi}$$

where the righthand side of (15) denotes the asymptotic variance of $(Y_n)_{n\in\mathbb{N}}$.

(To be more precise, the the righthand side of (15) denotes asymptotic variance of the sequence $(\sqrt{n}\,(Y_n - \ln(\pi)))_{n\in\mathbb{N}}$.)

Exercise 19

In this Exercise, again, we consider multiple linear regression models according to I.5.3 with stochastically independent and identically distributed error terms $\varepsilon_1, \ldots, \varepsilon_n \sim \mathcal{N}(0, \sigma^2)$, where $\sigma > 0$ is unknown.

For such a model and a realization $\boldsymbol{y} \in \mathbb{R}^n$ of $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$, by I.5.3 and I.2.5, the corresponding likelihood function (i.e. density) $L^{\boldsymbol{Y}}(\bullet \,|\, \boldsymbol{y}) : \mathbb{R}^d \times (0, \infty) \longrightarrow [0, \infty)$ of $\boldsymbol{Y}$ is given by

$$(1)\ \ L^{\boldsymbol{Y}}(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y}) \;=\; \frac{1}{\sqrt{(2\,\pi\,\sigma^2)^n}}\ \exp\left(-\frac{1}{2\,\sigma^2}\,(\boldsymbol{y} - B\,\boldsymbol{\beta})'\,(\boldsymbol{y} - B\,\boldsymbol{\beta})\right)\ ,\ \ (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times (0, \infty)\ ,$$

where d denotes the dimension of the parameter vector $\boldsymbol{\beta}$ in the considered model.

Then, the corresponding log-likelihood function $l^{\boldsymbol{Y}}(\bullet \,|\, \boldsymbol{y}) : \mathbb{R}^d \times (0, \infty) \longrightarrow \mathbb{R}$ of $\boldsymbol{Y}$ is given by

$$\begin{aligned}
(2) \qquad l^{\boldsymbol{Y}}(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y}) \;&:=\; \ln\left(L^{\boldsymbol{Y}}(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y})\right) \\
&\overset{(1)}{=}\; \ln\left((2\,\pi\,\sigma^2)^{-n/2}\right) - \frac{1}{2\,\sigma^2}\,(\boldsymbol{y} - B\,\boldsymbol{\beta})'\,(\boldsymbol{y} - B\,\boldsymbol{\beta}) \\
&=\; -\frac{n}{2}\,\ln(2\,\pi\,\sigma^2) - \frac{1}{2\,\sigma^2}\,\left|\left|\boldsymbol{y} - B\,\boldsymbol{\beta}\right|\right|^2\ ,\ \ (\boldsymbol{\beta}, \sigma^2) \in \mathbb{R}^d \times (0, \infty)
\end{aligned}$$

(For the representations (1) of the likelihood function and (2) of the log-likelihood function, respectively, cf. the solution of Exercise 13, (e) with I_n instead of Σ.)

According to I.4.31, the corresponding maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and $\widehat{\sigma^2}$ of σ^2 are given by

$$(3) \qquad \widehat{\boldsymbol{\beta}} \;=\; (B'\,B)^{-1}\,B'\,y \quad \text{(identical to the LSE of } \boldsymbol{\beta} \text{ given by I.5.4)}$$

and

$$\begin{aligned}
(4) \qquad \widehat{\sigma^2} \;&=\; \frac{1}{n}\,\boldsymbol{y}'\,\underbrace{(I_n - B\,(B'\,B)^{-1}\,B')}_{\text{symmetric and idempotent}}\,\boldsymbol{y} \\
&=\; \frac{1}{n}\,\boldsymbol{y}'\,(I_n - B\,(B'\,B)^{-1}\,B')'\,(I_n - B\,(B'\,B)^{-1}\,B')\,\boldsymbol{y} \\
&\overset{(3)}{=}\; \frac{1}{n}\,\left(\boldsymbol{y} - B\,\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - B\,\widehat{\boldsymbol{\beta}}\right) \;=\; \frac{1}{n}\,\left|\left|\boldsymbol{y} - B\,\widehat{\boldsymbol{\beta}}\right|\right|^2 \overset{\text{I.5.12}}{=} \frac{1}{n}\ \text{SSE}
\end{aligned}$$

Since the likelihood function and the log-likelihood function attain their maxima at the same arguments, we obtain (by definition of maximum likelihood estimates):

$$\begin{aligned}
(5) \qquad &\max\left\{l^{\boldsymbol{Y}}(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y}) \,\middle|\, \boldsymbol{\beta} \in \mathbb{R}^d,\ \sigma > 0\right\} \;=\; l^{\boldsymbol{Y}}\left(\widehat{\boldsymbol{\beta}}, \widehat{\sigma^2} \,|\, \boldsymbol{y}\right) \\
&\overset{(2)}{=}\; -\frac{n}{2}\,\ln\left(2\,\pi\,\widehat{\sigma^2}\right) - \frac{1}{2\,\widehat{\sigma^2}}\,\left|\left|\boldsymbol{y} - B\,\widehat{\boldsymbol{\beta}}\right|\right|^2 \overset{(4)}{=} -\frac{n}{2}\,\ln\left(2\,\pi\,\widehat{\sigma^2}\right) - \frac{1}{2\,\widehat{\sigma^2}}\,n\,\widehat{\sigma^2} \\
&=\; -\frac{n}{2}\left(\ln\left(2\,\pi\,\widehat{\sigma^2}\right) + 1\right)
\end{aligned}$$

Thus, by Definition II.2.39, Akaike's Information Criterion for that model is given by

$$(6)\ \ \text{AIC} \overset{\text{Def.}}{=} -2\,\max\left\{l^{\boldsymbol{Y}}(\boldsymbol{\beta}, \sigma^2 \,|\, \boldsymbol{y}) \,\middle|\, \boldsymbol{\beta} \in \mathbb{R}^d,\ \sigma > 0\right\} + 2\,k \overset{(5)}{=} n\left(\ln\left(2\,\pi\,\widehat{\sigma^2}\right) + 1\right) + 2\,k\ ,$$

where k denotes the total number of parameters in the model.

By (6), for model $\mathscr{M}_1$ with $p+1 = m+2$ parameters, we obtain

$$\text{AIC}(\mathscr{M}_1) \;=\; n\Big(\ln\big(2\,\pi\,\widehat{\sigma_1^2}\big)+1\Big) \;+\; 2\,(p+1) \tag{7}$$

and for model $\mathscr{M}_2$ with $p+1+q$ parameters, we obtain

$$\text{AIC}(\mathscr{M}_2) \;=\; n\Big(\ln\big(2\,\pi\,\widehat{\sigma_2^2}\big)+1\Big) \;+\; 2\,(p+1+q) \tag{8}$$

with $\widehat{\sigma_1^2}$ and $\widehat{\sigma_2^2}$ denoting the maximum likelihood estimates of the unknown variance σ^2 for models $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively.

Thus, finally, we obtain:

$$\begin{aligned}
&\text{AIC}(\mathscr{M}_2) \;<\; \text{AIC}(\mathscr{M}_1) \qquad (9)\\
\overset{(7),(8)}{\iff}\quad & n\Big(\ln\big(2\,\pi\,\widehat{\sigma_2^2}\big)+1\Big) \;+\; 2\,(p+1+q) \;<\; n\Big(\ln\big(2\,\pi\,\widehat{\sigma_1^2}\big)+1\Big) \;+\; 2\,(p+1)\\
\iff\quad & n\Big(\ln\big(2\,\pi\,\widehat{\sigma_2^2}\big)-\ln\big(2\,\pi\,\widehat{\sigma_1^2}\big)\Big) \;<\; -2\,q \quad\iff\quad \ln\left(\frac{\widehat{\sigma_2^2}}{\widehat{\sigma_1^2}}\right) \;<\; -\frac{2\,q}{n}\\
\iff\quad & \frac{\text{SSE}_2}{\text{SSE}_1} \overset{(4)}{=} \frac{n\,\widehat{\sigma_2^2}}{n\,\widehat{\sigma_1^2}} \;=\; \frac{\widehat{\sigma_2^2}}{\widehat{\sigma_1^2}} \;<\; \exp\left(-\,\frac{2\,q}{n}\right)
\end{aligned}$$

Exercise 20

(a) Let a GLM according to section II.2 be given. Then, according to II.2.16, the corresponding likelihood equations are given by

$$(1) \qquad \sum_{i=1}^{n} \left(\frac{y_i - \mathrm{E}(Y_i)}{\mathrm{Var}(Y_i)} \, \frac{\partial g^{-1}}{\partial \eta_i}(\eta_i) \; x_{ik} \right) \;=\; 0 \quad \text{for} \quad k \in \{1, \ldots, p\} \;,$$

where p denotes the dimension of the parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$. Note that these likelihood equations implicitely depend on $\boldsymbol{\beta}$ (as mentioned in the Lecture).

However, in general, the likelihood equations (1) do *not* imply

$$(2) \qquad \sum_{i=1}^{n} \big(y_i - \mu_i \big) \;=\; \sum_{i=1}^{n} \big(y_i - \mathrm{E}(Y_i) \big) \;=\; 0$$

as will be shown by a counter-example at the end of part (a).

Thus, if $\widehat{\mu} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_n)' := g^{-1}(X\,\widehat{\boldsymbol{\beta}})$ for a solution $\widehat{\boldsymbol{\beta}}$ of (1), we will – in general – *not* obtain:

$$(3) \qquad \overline{e} \;=\; \frac{1}{n} \sum_{i=1}^{n} e_i \;\stackrel{\text{Def}}{=}\; \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{\mu}_i) \;=\; 0$$

If, especially, we consider a GLM with a canonical link function, according to II.2.19, the likelihood equations in (1) simplify to

$$(4) \qquad \sum_{i=1}^{n} \big(y_i - E(Y_i) \big)\, x_{ik} \;=\; 0 \quad \text{for} \quad k \in \{1, \ldots, p\}$$

If, additionally, the GLM has an intercept term, i.e. $x_{i1} = 1$ for $i \in \{1, \ldots, n\}$, then (4) evaluated for $k = 1$ yields (2) and, thus, the corresponding estimated residuals have a mean of 0 as derived in (3).

If the GLM has no intercept term, again, we cannot imply the zero mean of the estimated residuals, since then $x_{1k}, \ldots, x_{nk}$ may attain different values for each $k \in \{1, \ldots, p\}$.

As an example, let us consider the observation vector $y := (0, 0, 1, 1, 0)'$ of 5 independent Bernoulli trials with parameter $\pi \in (0, 1)$ and with only one explanatory variable and corresponding data vector $x := (3, 1, 2, 4, 2)'$.

First, we fit the logit model with intercept, i.e. a GLM with canonical link and intercept:

$$(5) \qquad \mathrm{logit}(\pi_i) \;=\; \beta_0 + \beta_1\, x_i \;\;,\;\; i \in \{1, \ldots, 5\}$$

Then, we obtain $\overline{e} = 0$ as provided by the derivations above.

Next, we fit the logit model without intercept:

$$(6) \qquad \mathrm{logit}(\pi_i) \;=\; \beta_1\, x_i \;\;,\;\; i \in \{1, \ldots, 5\}$$

Then, we obtain $\overline{e} = -0.1 \neq 0$.

Finally, we fit the GLM with intercept, but with the identity function instead of the logit function as link function, i.e. a simple linear regression model given by

$$(7) \qquad \pi_i \;=\; \beta_0 + \beta_1\, x_i \;\;,\;\; i \in \{1, \ldots, 5\}$$

Then, we obtain $\overline{e} \approx -0.027 \neq 0$.

(b) According to the given hint, as a counter-example for part (b), we consider the intercept-GLM for a single random variable $Y \sim \mathcal{B}(n, \pi)$ with $n \in \mathbb{N}$ and $\pi \in (0,1)$ and with the identity link function. This is *not* the canonical one, since the logit function is the canonical link for that model.

Thus, the single model equation (according to sample size 1) is just given by:

$$\pi_i = \beta_0 \tag{8}$$

Then, for a realization $y \in \{0, \dots, n\}$ of Y, the corresponding log-likelihood function $l^{\boldsymbol{Y}}(\bullet \,|\, \boldsymbol{y}) : \mathbb{R} \longrightarrow \mathbb{R}$ of $\boldsymbol{Y}$ is given by

$$\begin{aligned} (9) \qquad l^{\boldsymbol{Y}}(\beta_0 \,|\, \boldsymbol{y}) &= \ln\left(\binom{n}{y} \beta_0^y \, (1-\beta_0)^{n-y}\right) \\ &= \ln\left(\binom{n}{y}\right) + y \, \ln(\beta_0) + (n-y) \, \ln(1-\beta_0) \ , \ \beta_0 \in \mathbb{R} \end{aligned}$$

By (9), the log-likelihood function $l^{\boldsymbol{Y}}(\bullet \,|\, \boldsymbol{y})$ is twice differentiable on $\mathbb{R}$, and its first two derivatives are given by

$$\begin{aligned} (10) \qquad \frac{d}{d\beta_0} l^{\boldsymbol{Y}}(\beta_0 \,|\, \boldsymbol{y}) &\overset{(9)}{=} \frac{y}{\beta_0} - \frac{n-y}{1-\beta_0} = \frac{y\,(1-\beta_0) - (n-y)\,\beta_0}{\beta_0\,(1-\beta_0)} \\ &= \frac{y - n\,\beta_0}{\beta_0\,(1-\beta_0)} \ , \ \beta_0 \in \mathbb{R} \\ (11) \qquad \frac{d^2}{d\beta_0^2} l^{\boldsymbol{Y}}(\beta_0 \,|\, \boldsymbol{y}) &\overset{(10)}{=} \frac{-n\,(\beta_0 - \beta_0^2) - (y - n\,\beta_0)\,(1 - 2\,\beta_0)}{\beta_0^2\,(1-\beta_0)^2} \\ &= \frac{-n\,\beta_0 + n\,\beta_0^2 - y + n\,\beta_0 + 2\,y\,\beta_0 - 2\,n\,\beta_0^2}{\beta_0^2\,(1-\beta_0)^2} \\ &= \frac{-y - n\,\beta_0^2 + 2\,y\,\beta_0}{\beta_0^2\,(1-\beta_0)^2} \ , \ \beta_0 \in \mathbb{R} \end{aligned}$$

In the given situation, according to Definition II.2.23, the observed information matrix (which is a real number in this case) is given by

$$\mathcal{J}_F^{\text{obs}}(\beta_0) = -\frac{d^2}{d\beta_0^2} l^{\boldsymbol{Y}}(\beta_0 \,|\, \boldsymbol{y}) \overset{(9)}{=} \frac{y + n\,\beta_0^2 - 2\,y\,\beta_0}{\beta_0^2\,(1-\beta_0)^2} \ , \ \beta_0 \in \mathbb{R} \ , \tag{12}$$

which, obviously, is dependent on the data $y \in \{0, \dots, n\}$ and therefore is different to the expected information (matrix).

Remark:

In the given situation of the considered counter-example, the expected information (matrix) is given by

$$\mathcal{J}_F(\beta_0) = \frac{n}{\beta_0\,(1-\beta_0)} = \frac{1}{\operatorname{Var}_{\beta_0}(\widehat{\beta_0})} = \frac{1}{\operatorname{Var}_{\beta_0}(Y/n)} \ , \ \beta_0 \in \mathbb{R}$$

(c) First, we know from classical introductions to statistics:

$$\text{(13)} \qquad \mathrm{E}_\pi\big(\widehat{\pi}_1\big) \overset{\text{Def.}}{=} \mathrm{E}_\pi\big(\overline{Y}\big) = \pi \;,\; \pi \in (0,1)$$

$$\text{(14)} \qquad \mathrm{Var}_\pi\big(\widehat{\pi}_1\big) \overset{\text{Def.}}{=} \mathrm{Var}_\pi\big(\overline{Y}\big) = \frac{\pi\,(1-\pi)}{100} \;,\; \pi \in (0,1)$$

For the second estimator, we obtain:

$$\text{(15)} \qquad \mathrm{E}_\pi\big(\widehat{\pi}_2\big) \overset{\text{Def.}}{=} \mathrm{E}_\pi\left(\frac{\overline{Y}}{2} + \frac{1}{4}\right) \underset{(13)}{\overset{\text{Lin.}}{=}} \frac{\pi}{2} + \frac{1}{4} \neq \pi \quad \text{for} \quad \pi \neq \frac{1}{2}$$

$$\text{(16)} \qquad \mathrm{Var}_\pi\big(\widehat{\pi}_2\big) \overset{\text{Def.}}{=} \mathrm{Var}_\pi\left(\frac{\overline{Y}}{2} + \frac{1}{4}\right) = \frac{1}{4}\,\mathrm{Var}_\pi\big(\overline{Y}\big) \overset{(14)}{=} \frac{\pi\,(1-\pi)}{400} \;,\; \pi \in (0,1)$$

(i) By (13) and (15), only the first estimator $\widehat{\pi}_1$ of $\pi \in (0,1)$ is unbiased, while the second estimator $\widehat{\pi}_2$ is biased.

(ii) By (14) and (16), it holds:

$$\mathrm{Var}_\pi\big(\widehat{\pi}_2\big) = \frac{\pi\,(1-\pi)}{400} < \frac{\pi\,(1-\pi)}{100} = \mathrm{Var}_\pi\big(\widehat{\pi}_1\big) \;,\; \pi \in (0,1)$$

Thus, the estimator $\widehat{\pi}_2$ has a smaller variance than the estimator $\widehat{\pi}_1$ for each $\pi \in (0,1)$

(iii) For the two estimators, we obtain the following mean squared errors:

$$\begin{aligned}
\text{(17)} \qquad \mathrm{MSE}_\pi\big(\widehat{\pi}_1\big) &\overset{\text{Def.}}{=} E_\pi\Big(\big(\widehat{\pi}_1 - \pi\big)^2\Big) \overset{(13)}{=} E_\pi\Big(\big(\widehat{\pi}_1 - \mathrm{E}_\pi(\widehat{\pi}_1)\big)^2\Big) \\
&\overset{\text{Def.}}{=} \mathrm{Var}_\pi\big(\widehat{\pi}_1\big) \overset{(14)}{=} \frac{\pi\,(1-\pi)}{100} \;,\; \pi \in (0,1)
\end{aligned}$$

$$\begin{aligned}
\text{(18)} \qquad \mathrm{MSE}_\pi\big(\widehat{\pi}_2\big) &\overset{\text{Def.}}{=} E_\pi\Big(\big(\widehat{\pi}_2 - \pi\big)^2\Big) \\
&= \underbrace{E_\pi\Big(\big(\widehat{\pi}_2 - \mathrm{E}_\pi(\widehat{\pi}_2)\big)^2\Big)}_{=\mathrm{Var}_\pi(\widehat{\pi}_2)} + \underbrace{E_\pi\Big(\big(\mathrm{E}_\pi(\widehat{\pi}_2) - \pi\big)^2\Big)}_{=\mathrm{Bias}^2_\pi(\widehat{\pi}_2)} \\
&\overset{(15),(16)}{=} \frac{\pi\,(1-\pi)}{400} + \left(\frac{\pi}{2} + \frac{1}{4} - \pi\right)^2 = \frac{\pi - \pi^2}{400} + \frac{(1-2\,\pi)^2}{16} \\
&= \frac{\pi - \pi^2 + 25 - 100\,\pi + 100\,\pi^2}{400} = \frac{99\,\pi^2 - 99\,\pi + 25}{400} \;, \\
&\qquad \pi \in (0,1)
\end{aligned}$$

Thus, we obtain for $\pi \in (0,1)$:

$$\text{(19)} \qquad \mathrm{MSE}_\pi\big(\widehat{\pi}_2\big) < \mathrm{MSE}_\pi\big(\widehat{\pi}_1\big) \overset{(17),(18)}{\iff}$$

$$\frac{99\,\pi^2 - 99\,\pi + 25}{400} < \frac{\pi - \pi^2}{100} \iff 99\,\pi^2 - 99\,\pi + 25 < 4\,\pi - 4\,\pi^2$$

$$\iff 103\,\pi^2 - 103\,\pi + 25 < 0$$

which is satisfied for all

$$\text{(20)} \qquad \pi \in \left(\frac{1}{2} - \frac{\sqrt{3/103}}{2} \,,\, \frac{1}{2} + \frac{\sqrt{3/103}}{2}\right) \approx (0.415\,,\,0.585)$$

Remark to Exercise 20, (c):

A practical application for the situation considered in part (c) could be given by a polling agency, which wants to estimate the proportion of votes for a candidate A over a second candidate B in an election between this two candidates. If the agency expects the proportion to be close to 50:50, the estimator $\hat{\pi}_2$ is more efficient than $\hat{\pi}_1$.