Tutorial 7

1. Let $X_1, \ldots, X_n$ be a random sample from distribution $p_\theta(0) = 1 - \theta$, $p_\theta(1) = \theta$, where the parameter $\theta \in (0, 1)$. Show that $\overline{X}$ is a UMVU (uniformly minimum variance unbiased) estimator of θ.

2. Let $X \sim \text{Binomial}(n, \theta)$. In other words

$$\mathbb{P}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \qquad x = 0, 1, \ldots, n$$

   Consider the estimators $\widehat{\theta} = \frac{X}{n}$ and $\widetilde{\theta} = \frac{X+1}{n+2}$. Compute the bias of these estimators, their variance and their mean squared errors. Are they consistent?

3. Let $(X_1, \ldots, X_n)$ be a sample from a Poisson distribution with unknown parameter λ. To estimate λ, we estimate $g(\lambda) := \mathbb{P}(X_1 = 0) = e^{-\lambda}$ and we consider two estimators $\widehat{g}_1$, $\widehat{g}_2$ of g, where

$$\widehat{g}_1 = e^{-\overline{X}}, \qquad \widehat{g}_2 = \left(1 - \frac{1}{n}\right)^{n\overline{X}},$$

   where $\overline{X}$ is the sample average. Compute the bias of the estimators $\widehat{g}_1$ and $\widehat{g}_2$.

4. Let $(X_1, \ldots, X_n)$ be a random sample from a Bernoulli(p) distribution (that is, $\mathbb{P}(X_j = 1) = p$, $\mathbb{P}(X_j = 0) = 1 - p$). Show that there do not exist unbiased estimators of the quantities

$$g_1(p) = \frac{p}{1-p}, \qquad g_2(p) = \frac{1}{p}.$$

5. Let $(X_1, X_2, \ldots, X_n)$ be a random sample with unknown expected value μ and known variance σ^2.

   (a) Show that the statistic

$$T(X_1, \ldots, X_n) = \sum_{i=1}^{n} a_i X_i \qquad \sum_{i=1}^{n} a_i = 1$$

   is an unbiased estimator of μ.

   (b) Compute the variance of T and show that, for unbiased estimators of this form, it is minimised for $a_i = \frac{1}{n}$, $i = 1, \ldots, n$.

6. Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p) random variables. Find the Cramer-Rao lower bound for the variance of an unbiased estimator of $g(p) = p(1 - p)$.

7. Let $(X_1, \ldots, X_n)$ be a random sample from an exponential distribution Exp(λ). That is, the density function is:

$$f(x; \lambda) = \lambda e^{-\lambda x} \mathbf{1}_{\{x \geq 0\}}.$$

   Show that the statistic $T(X_1, \ldots, X_n) = nX_{1:n}$ is an unbiased estimator of $\frac{1}{\lambda}$, but that it is not consistent.

8. Suppose θ is an unknown parameter to be estimated, and let $f(X)$ be the estimator. Let $l(\theta, a)$ be a *loss function*, where the *loss* incurred when estimating θ by $f(X)$ is given by $l(\theta, f(X))$. The *risk function* is defined as:

$$R(\theta, f) = \mathbb{E}_\theta \left[l(\theta, f(X)) \right].$$

Suppose that the loss function $l(\theta, a)$ is strictly convex in the variable a. Suppose $g(X)$ is an unbiased estimator of $q(\theta)$ and that $T(X)$ is a sufficient statistic. Let $g^*(X) = \mathbb{E}_\theta \left[g(X) | T(X) \right]$. Show that $R(\theta, g^*) \leq R(\theta, g)$.

Hint Jensen's inequality: if ϕ is a convex function and X a random variable then

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]).$$

9. Let X have density or probability mass function $p(x, \theta)$, where $\theta \in \Theta \subset \mathbb{R}$. Suppose that the following assumptions hold:

- $\{x : p(x, \theta) > 0\}$ is the same for all $\theta \in \Theta$ and
- For any statistic T satisfying $\mathbb{E}[|T(X)|] < +\infty$ for all $\theta \in \Theta$,

$$\int T(x) \frac{\partial}{\partial \theta} p(x, \theta) dx = \frac{\partial}{\partial \theta} \int T(x) p(x, \theta) dx.$$

Suppose that h is monotone increasing and differentiable from Θ to $h(\Theta)$. Let $\eta = h(\theta)$ and $q(x, \eta) = p(x, h^{-1}(\eta))$.

(a) Let $I_p(\theta)$ and $I_q(\eta)$ denote the Fisher information in the two parametrisations. Show that

$$I_q(\eta) = \frac{1}{(h'(h^{-1}(\eta)))^2} I_p(h^{-1}(\eta)).$$

(b) Let $B_p(\theta)$ and $B_q(\eta)$ denote the information inequality lower bound for the two parametrisations. That is,

$$B_p(\theta) = \frac{(\psi_1'(\theta))^2}{I(\theta)}, \qquad B_q(\eta) = \frac{(\psi_2'(\eta))^2}{I(\eta)}$$

where $\psi_1(\theta)$ is the quantity to be estimated and $\psi_2(\eta) = \psi_1(h^{-1}(\eta))$; i.e. the same quantity under the η-parametrisation.

Show that $B_q(\eta) = B_p(h^{-1}(\eta))$. That is, the Fisher information lower bound is the same.

10. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ where μ is known.

(a) Show that

$$\widehat{\sigma}^2 := \frac{1}{n} \sum_{j=1}^{n} (X_j - \mu)^2$$

is a UMVU estimator of σ^2.

(b) For parameter estimation, a *decision rule* d simply means assigning a decision $d(X)$ for the unknown parameter. A decision rule d is *inadmissible* if there is another decision rule d^* such that $R(\theta, d^*) \leq R(\theta, d)$ for all θ and $R(\theta, d^*) < R(\theta, d)$ for some θ. A decision rule is *admissible* if it is not inadmissible. Show that $\widehat{\sigma}^2$ is inadmissible under squared loss error $R(\theta, d) = \mathbb{E}_\theta \left[|\theta - d(X)|^2 \right]$.
Hint Consider bias estimators of the form $a_n \widehat{\sigma}^2$ where $\widehat{\sigma}^2$ is defined above.

11. Let $Y_1, \ldots, Y_n$ be independent Poisson random variables, where

$$\mathbb{E}[Y_j] = \mu_j = \exp\left\{\alpha + \beta z_j\right\}.$$

(For example, z_j could be the level of a drug given to the jth patient with an infectious disease, and Y_j could denote the number of infectious microbes found in a unit of blood taken from patient j 24 hours after the drug was administered.

(a) Write the model for $Y_1, \ldots, Y_n$ in two-parameter canonical exponential form and give the sufficient statistic.

(b) Let $\theta = (\alpha, \beta)$. Compute $I(\theta)$, the information matrix for the model and find the lower bound on the variances of unbiased estimators $\widehat{\alpha}$ and $\widehat{\beta}$ of α and β.

(c) Let $z_i = \log\left(\frac{i}{n+1}\right)$, $i = 1, \ldots, n$. Compute $\lim_{n \to +\infty} \frac{1}{n} I(\theta)$ and give the limit of n times the lower bound on the variances of $\widehat{\alpha}$ and $\widehat{\beta}$.
Hint Use integral approximations for the sums.

Short Answers

1. Firstly,

$$\mathbb{E}_\theta\left[\overline{X}\right] = \mathbb{E}_\theta\left[X_1\right] = 0 \times (1-\theta) + 1 \times \theta = \theta$$

so the estimator is unbiased.

Secondly: there are several ways to show it is UMVU.

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta) = \theta^{\sum_j x_j}(1-\theta)^{n-\sum_j x_j} = \exp\left\{n\overline{x}\log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\}$$

This is an exponential family; take $\overline{X}$ as the sufficient statistic and $\eta := n\log\left(\frac{\theta}{1-\theta}\right)$ as the canonical parameter. Hence, by the result in lectures, $\overline{X}$ is the UMVU estimator of its expected value, which is $\mathbb{E}_\theta[\overline{X}] = \theta$.

Alternatively, we can compute the Cramér-Rao lower bound directly. The estimator has variance

$$\mathbf{V}_\theta\left(\overline{X}\right) = \frac{1}{n}\mathbf{V}_\theta\left(X_1\right) = \frac{\theta(1-\theta)}{n}.$$

Now to show that this achieves the C-R lower bound, compute $I(\theta) = nI_1(\theta)$, the information in the sample.

$$\frac{d}{d\theta}\log p_\theta(0) = -\frac{1}{1-\theta} \qquad \frac{d}{d\theta}\log p_\theta(1) = \frac{1}{\theta}$$

$$I_1(\theta) = \mathbb{E}_\theta\left[\left(\frac{d}{d\theta}\log p_\theta(X)\right)^2\right] = (1-\theta)\frac{1}{(1-\theta)^2} + \theta\frac{1}{\theta^2} = \frac{1}{\theta(1-\theta)}$$

hence the information from a sample size n is:

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$

and the C-R lower bound is

$$\frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{n}.$$

hence estimator is UMVU.

2.

$$\mathbb{E}[\widehat{\theta}] = \frac{1}{n}\mathbb{E}[X] = \frac{1}{n}n\theta = \theta$$

so this estimator is unbiased.

$$\mathbf{V}(\widehat{\theta}) = \frac{1}{n^2}\mathbf{V}(X) = \frac{1}{n^2}n\theta(1-\theta) = \frac{\theta(1-\theta)}{n} < \frac{1}{4n}$$

so

$$\mathbb{P}(|\widehat{\theta}-\theta| > \epsilon) \leq \frac{\theta(1-\theta)}{\epsilon^2 n} \leq \frac{1}{4\epsilon^2 n} \overset{n\to+\infty}{\longrightarrow} 0$$

hence uniformly consistent.

For $\widetilde{\theta}$

$$\mathbb{E}[\widetilde{\theta}] = \frac{n\theta + 1}{n+2} = \frac{n}{n+2}\theta + \frac{1}{n+2}$$

so

$$\text{Bias}(\widetilde{\theta}) = \frac{n}{n+2}\theta + \frac{1}{n+2} - \theta = \frac{1-2\theta}{n+2}$$

$$\mathbf{V}(\widetilde{\theta}) = \frac{n\theta(1-\theta)}{(n+2)^2}$$

so that the mean squared error is:

$$\mathbb{E}\left[|\widetilde{\theta} - \theta|^2\right] = \frac{(1-2\theta)^2}{(n+2)^2} + \frac{n\theta(1-\theta)}{(n+2)^2} = \frac{1+(n-4)\theta + 3\theta^2}{(n+2)^2}$$

Yes - the estimator is uniformly consistent; for $n \geq 4$,

$$\mathbb{P}(|\widetilde{\theta} - \theta|^2 > \epsilon) \leq \frac{1+(n-4)\theta(1-\theta)}{\epsilon^2(n+2)^2} \leq \frac{1+(n-4)/4}{\epsilon^2(n+2)^2} \stackrel{n\to+\infty}{\longrightarrow} 0.$$

3. $Y := n\overline{X} = \sum_{j=1}^n X_j \sim \text{Poiss}(n\lambda)$ so that

$$\mathbb{E}[\widehat{g}_1] = \mathbb{E}\left[e^{-Y/n}\right] = \sum_{x=0}^{\infty} \frac{(\lambda n)^x}{x!} e^{-n\lambda - (x/n)} = \sum_{x=0}^{\infty} \frac{(\lambda n e^{-1/n})^x}{x!} e^{-n\lambda} = e^{-n\lambda(1-e^{-1/n})}$$

so

$$\text{Bias}(\widehat{g}_1) = e^{-\lambda}\left(e^{-\lambda(n(1-e^{-1/n})-1)} - 1\right).$$

$$\mathbb{E}[\widehat{g}_2] = \sum_{x=0}^{\infty} \frac{\left((n\lambda)(1-\frac{1}{n})\right)^x}{x!} e^{-\lambda n} = e^{n\lambda - \lambda - n\lambda} = e^{-\lambda}$$

so that $\widehat{g}_2$ is unbiased.

4. An unbiased estimator of $g_1(p)$ is a function of the n binary variables $T(X_1, \ldots, X_n)$ satisfying

$$\mathbb{E}[T(X_1, \ldots, X_n)] = \sum_{\{0,1\}^n} T(x_1, \ldots, x_n)p^k(1-p)^{n-k} = \frac{p}{1-p}$$

where k denotes the number of 1s in the sequence $(x_1, \ldots, x_n)$.

so that

$$0 = \sum_{\{0,1\}^n} T(x_1, \ldots, x_n)p^{k-1}(1-p)^{n-k+1} - 1.$$

This holds for all p, which is a contradiction, since the equation is a polynomial of degree $n+1$ and hence has at most $n+1$ distinct roots. Similarly for g_2.

5. (a)

$$\mathbb{E}[T] = \mu \sum_{i=1}^{n} a_i = \mu.$$

(b)

$$\mathbf{V}(T) = \sigma^2 \sum_{i=1}^{n} a_i^2$$

$n-1$ free variables; $a_n = 1 - \sum_{j=1}^{n-1} a_j$ so that

$$\frac{\partial}{\partial a_i} \mathbf{V}(T) = 2\sigma^2 (a_i - a_n) = 0$$

so that $a_1 = \ldots = a_n$. With constraint that $\sum_{j=1}^{n} a_j = 1$, it follows that $a_j = \frac{1}{n}$ for each $j = 1, \ldots, n$.

6.

$$\mathbb{P}(x) = p^x (1-p)^{1-x} \Rightarrow \log \mathbb{P}(x) = x \log p + (1-x) \log(1-p) \qquad x \in \{0,1\}$$

$$\frac{d}{dp} \log \mathbb{P}(x) = \frac{d}{dp} \log \mathbb{P}(x) = \left(\frac{x}{p} - \frac{(1-x)}{1-p} \right) = \left(\frac{x}{p(1-p)} - \frac{1}{1-p} \right)$$

$$I(p) = \mathbf{V}_p \left(\frac{d}{dp} \log \mathbb{P}_p(X) \right) = \frac{1}{p^2(1-p)^2} \mathbf{V}_p(X) = \frac{1}{p(1-p)}.$$

For n observations, $I_n(g) = \frac{n}{p(1-p)}$. The Cramér lower bound is therefore:

$$\mathbf{V}(\widehat{g}) \geq \frac{(g'(p))^2}{I_n(g)} = \frac{(1-2p)^2 p(1-p)}{n}.$$

7. $\min_j X_j \sim \text{Exp}(n\lambda)$ hence

$$\mathbb{E}\left[n \min_j X_j \right] = n \frac{1}{n\lambda} = \frac{1}{\lambda}.$$

$$\mathbf{V}\left(n \min_j X_j \right) = n^2 \frac{1}{n^2 \lambda^2} = \frac{1}{\lambda^2},$$

Also, $n \min_j X_j \sim \text{Exp}(\lambda)$, hence $\mathbb{P}\left(|n \min_j X_j - \lambda| > \epsilon \right) \not\to 0$ as $n \to +\infty$.

8. This follows immediately from the definition and Jensen:

$$\begin{aligned} R(\theta, g^*) &= \mathbb{E}_\theta \left[l(\theta, g^*(X)) \right] = \mathbb{E}_\theta \left[l(\theta, \mathbb{E}_\theta[g(X)|T(X)]) \right] \\ &\leq \mathbb{E}_\theta \left[\mathbb{E}_\theta \left[l(\theta, g(X)) | T(X) \right] \right] = \mathbb{E}_\theta \left[l(\theta, g(X)) \right] = R(\theta, g). \end{aligned}$$

Note that we do not need T to be a *sufficient* statistic.

9. (a) This follows simply from the definition: let $\eta = h(\theta)$, then

$$\frac{d}{d\eta}\log q(x,\eta) = \frac{1}{dh(\theta)/d\theta}\frac{d}{d\theta}\log p(x,\theta)$$

$$I_q(\eta) = \mathbb{E}_\eta\left[\left(\frac{d}{d\eta}\log q(x,\eta)\right)^2\right] = \frac{1}{(h'(\theta))^2}\mathbb{E}_\theta\left[\left(\frac{d}{d\theta}\log p(x,\theta)\right)^2\right] = \frac{1}{h'(h^{-1}(\eta))}I_p(h^{-1}(\eta)).$$

(b) Consider a quantity $\psi(\theta) = \mathbb{E}_\theta[T(X)]$, so that $T(X)$ is an unbiased estimator of $\psi(\theta)$. The lower bound using parameter θ is:

$$\frac{(\psi'(\theta))^2}{I_p(\theta)},$$

while with parameter η:

$$\frac{d}{d\eta}\psi(h^{-1}(\eta)) = \psi'(h^{-1}(\eta))\frac{dh^{-1}(\eta)}{d\eta},$$

so that the lower bound is:

$$\frac{\left(\frac{d}{d\eta}\psi(h^{-1}(\eta))\right)^2}{I_q(\eta)} = \frac{\left(\psi'(h^{-1}(\eta))\right)^2}{I_p(h^{-1}(\eta))}.$$

10. (a) Unbiased is clear:

$$\mathbb{E}_{\sigma^2}\left[\widehat{\sigma}^2\right] = \frac{1}{n}\sum_{j=1}^{n}\mathbf{V}_{\sigma^2}(X_j) = \sigma^2$$

The variance is:

$$\mathbf{V}_{\sigma^2}(\widehat{\sigma}^2) = \frac{1}{n^2}\sum_{j=1}^{n}\mathbf{V}((X_j-\mu)^2) = \frac{\sigma^4}{n}\mathbf{V}\left(\left(\frac{X_1-\mu}{\sigma}\right)^2\right) = \frac{2\sigma^4}{n}$$

while the Fisher information is $I(\sigma^2) = nI_1(\sigma^2)$;

$$\frac{d}{d(\sigma^2)}\log p(x,\sigma^2) = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}$$

Let $I_n(\sigma^2)$ denote the information in a sample of size n, then $I_n(\sigma^2) = nI_1(\sigma^2)$ and:

$$I_1(\sigma^2) = \mathbf{V}_{\sigma^2}\left(\left(\frac{d}{d(\sigma^2)}\log p(x,\sigma^2)\right)^2\right) = \frac{1}{4\sigma^4}\mathbf{V}_{\sigma^2}\left(\left(\frac{X-\mu}{\sigma}\right)^2\right)$$

and now use $V = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_1^2$ so that $\mathbf{V}(V) = 2$. Then

$$I_1(\sigma^2) = \frac{1}{2\sigma^4} \Rightarrow I_n(\sigma^2) = \frac{n}{2\sigma^4}.$$

Hence the Cramér-Rao lower bound is $\frac{2\sigma^4}{n}$, which is $\mathbf{V}_\sigma^2(\widehat{\sigma}^2)$.
so that $I(\sigma^2) = \frac{n}{2\sigma^4}$ giving a lower bound of $\frac{2\sigma^4}{n}$, hence $\widehat{\sigma}^2$ is an UMVU estimator.

(b) For an unbiased estimator, this risk function is simply the variance since $\widehat{\sigma}^2$ is UMVU, it follows that any estimator with smaller risk must be biased. Try estimators of the form $a_n\widehat{\sigma}^2$. Then

$$\begin{aligned}\mathbb{E}_{\sigma^2}\left[\left|a_n\widehat{\sigma}^2-\sigma^2\right|^2\right] &= \mathbb{E}_{\sigma^2}\left[\left|a_n(\widehat{\sigma}^2-\sigma^2)+(a_n-1)\sigma^2\right|^2\right]\\ &= a_n^2\mathbf{V}_{\sigma^2}(\widehat{\sigma}^2)+\left((a_n-1)\sigma^2\right)^2\\ &= a_n^2\frac{2\sigma^4}{n}+(a_n-1)^2\sigma^4\end{aligned}$$

Minimising gives: $a_n=\frac{n}{n+2}$. This gives estimator $\widetilde{\sigma}^2=\frac{1}{n+2}\sum_{j=1}^n(X_j-\mu)^2$ and $R(\sigma^2,\widetilde{\sigma}^2)=\frac{2}{n+2}\sigma^4$ which is smaller than the UMVU estimator.

11. (a)

$$p(y_1,\ldots,y_n;\alpha,\beta)=\frac{1}{\prod_{j=1}^n y_j!}\exp\left\{\alpha\sum_{j=1}^n y_j+\beta\sum_{j=1}^n y_jz_j-\sum_{j=1}^n\exp\{\alpha+\beta z_j\}\right\}$$

Sufficient statistic: $T(Y)=(T_1(Y),T_2(Y))$ where

$$T_1(Y)=\sum_{j=1}^n Y_j \qquad \text{and} \qquad T_2(Y)=\sum_{j=1}^n z_jY_j.$$

(b)

$$\frac{\partial}{\partial\alpha}\log p=\sum_{j=1}^n y_j-\sum_{j=1}^n e^{\alpha+\beta z_j}\Rightarrow -\frac{\partial^2}{\partial\alpha^2}\log p=\sum_{j=1}^n e^{\alpha+\beta z_j}$$

$$\frac{\partial}{\partial\beta}\log p=\sum_{j=1}^n y_jz_j-\sum_{j=1}^n z_je^{\alpha+\beta z_j}\Rightarrow -\frac{\partial^2}{\partial\beta^2}\log p=e^{\alpha}\sum_{j=1}^n z_j^2e^{\beta z_j}$$

$$-\frac{\partial^2}{\partial\alpha\partial\beta}\log p=e^{\alpha}\sum_{j=1}^n z_je^{\beta z_j}$$

so

$$I_{\alpha,\alpha}(\theta)=-\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\alpha^2}\log p(X,\alpha,\beta)\right]=\sum_{j=1}^n\mu_j=e^{\alpha}\sum_{j=1}^n e^{\beta z_j}$$

$$I_{\beta,\beta}(\theta)=-\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\beta^2}\log p(X;\alpha,\beta)\right]=e^{\alpha}\sum_{j=1}^n z_j^2e^{\beta z_j}$$

$$I_{\alpha,\beta}(\theta)=-\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\alpha\partial\beta}\log p(X;\alpha,\beta)\right]=e^{\alpha}\sum_{j=1}^n z_je^{\beta z_j}.$$

We have to invert the information matrix:

$$I^{-1}(\theta)=\frac{e^{-\alpha}}{\left(\sum e^{\beta z_j}\right)\left(\sum z_j^2e^{\beta z_j}\right)-\left(\sum z_je^{\beta z_j}\right)^2}\begin{pmatrix}\sum z_j^2e^{\beta z_j} & -\sum z_je^{\beta z_j}\\ -\sum z_je^{\beta z_j} & \sum e^{\beta z_j}\end{pmatrix}$$

$$\mathbf{V}(\widehat{\alpha}) \geq e^{\alpha}\left\{(\sum e^{\beta z_j}) - \frac{(\sum z_j e^{\beta z_j})^2}{(\sum z_j^2 e^{\beta z_j})}\right\}$$

$$\mathbf{V}\left(\widehat{\beta}\right) \geq e^{\alpha}\left\{(\sum z_j^2 e^{\beta z_j}) - \frac{(\sum z_j e^{\beta z_j})^2}{(\sum e^{\beta z_j})}\right\}$$

Note If the information matrix is singular, then the results are correct but useless; they give variances greater than or equal to 0.

(c) Integrals are straightforward :

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

$$\int_0^1 x^{\beta} \log x dx = \int_0^1 e^{\beta \log x} \log x dx = \frac{d}{d\beta}\int_0^1 x^{\beta} dx = -\frac{1}{(1+\beta)^2}$$

$$\int_0^1 x^{\beta} (\log x)^2 dx = \frac{d^2}{d\beta^2}\int_0^1 x^{\beta} (\log x)^2 dx = \frac{2}{(1+\beta)^3}$$

In the limit,

$$\frac{1}{n} I(\theta) \to e^{\alpha}\begin{pmatrix} \frac{1}{1+\beta} & -\frac{1}{(1+\beta)^2} \\ -\frac{1}{(1+\beta)^2} & \frac{2}{(1+\beta)^3} \end{pmatrix}$$

$$n I^{-1}(\theta) \to e^{-\alpha}\begin{pmatrix} 2(1+\beta) & (1+\beta)^2 \\ (1+\beta)^2 & (1+\beta)^3 \end{pmatrix}$$

Lower bounds: $2(1+\beta)e^{-\alpha}$ and $(1+\beta)^3 e^{-\alpha}$ respectively for α and β.