S721 HW8

John Koo

To save on typing, I will denote $\frac{\partial^k}{\partial x^k} f(x) = \partial_x^k f(x)$.

Part 1

3.28

Part a

If μ is known, it's straightforward:

- $h(x) = 1$
- $c(\sigma^2) = (2\pi\sigma^2)^{-1/2}$
- $w_1(\sigma^2) = -(2\sigma^2)^{-1}$
- $t_i(x) = (x - \mu)^2$

If σ^2 is known, we can first rewrite the density function as $(2\pi\sigma^2)^{-1/2} e^{-x^2/2\sigma^2} e^{\mu x/\sigma^2} e^{-\mu^2/2\sigma^2}$ and we get:

- $h(x) = (2\pi\sigma^2)^{-1/2} \exp(-x^2/2\sigma^2)$
- $c(\mu) = \exp(-\mu^2/2\sigma^2)$
- $w_1(\mu) = \mu$
- $t_1(x) = x/\sigma^2$

Part c

If α is known, we can first write the density function as $\frac{1}{B(\alpha,\beta)} x^{\alpha-1} \exp((\beta - 1)\log(1-x))$ and we get:

- $h(x) = x^{\alpha-1}$
- $c(\beta) = \big(B(\alpha,\beta)\big)^{-1}$
- $w_1(\beta) = \beta - 1$
- $t_1(x) = \log(1 - x)$

If β is known, we can write the density function as $\big(B(\alpha,\beta)\big)^{-1} (1-x)^{\beta-1} \exp((\alpha - 1)\log x)$ and we get:

- $h(x) = (1 - x)^{\beta-1}$
- $c(\alpha) = \big(B(\alpha,\beta)\big)^{-1}$
- $w_1(\alpha) = \alpha - 1$
- $t_1(x) = \log x$

If both α and β are unknown, we can write the density function as $\frac{1}{B(\alpha,\beta)} e^{(\alpha-1)\log x + (\beta-1)\log(1-x)}$ to get:

- $h(x) = 1$
- $c(\alpha, \beta) = \frac{1}{B(\alpha,\beta)}$
- $w_1(\alpha, \beta) = \alpha - 1$
- $t_1(x) = \log x$
- $w_2(\alpha, \beta) = \beta - 1$
- $t_2(x) = \log(1 - x)$

Part d

We can write the mass function as $\frac{1}{x!}e^{-\theta}e^{x\log\theta}$ to get:

- $h(x) = \frac{1}{x!}$
- $c(\theta) = e^{-\theta}$
- $w_1(\theta) = \log\theta$
- $t_1(x) = x$

3.31

Part a

Assuming regular conditions, we have

$0 = \partial_\theta \int h(x)c(\theta)\exp(\sum_i w_i(\theta)t_i(x))dx$
$= \int \partial_\theta(h(x)c(\theta)\exp(\sum_i w_i(\theta)t_i(x)))dx$
$= \int \Big(h(x)c'(\theta)\exp\big(\sum_i w_i(\theta)t_i(x)\big) + h(x)c(\theta)\exp\big(\sum_i w_i(\theta)t_i(x)\big)\big(\sum_i \partial_{\theta_i} w_i(\theta)t_i(x)\big)\Big)dx$

Note that the second part is just $E\big[\sum_i \partial_{\theta_j} w_i(\theta)t_i(x)\big]$ since we are integrating that with the density function. For the first part, we note that $c'(\theta) = c(\theta)\partial_{\theta_j}\log c(\theta)$, so we get:

$0 = \int h(x)c(\theta)\partial_{\theta_j}\big(\log c(\theta)\big)\exp\big(\sum_i w_i(\theta)t_i(x)\big)dx + E\big[\sum_i \partial_{\theta_j} w_i(\theta)t_i(x)\big]$
$= \partial_{\theta_j}\big(\log c(\theta)\big)\int h(x)c(\theta)\exp\big(\sum_i w_i(\theta)t_i(x)\big)dx + E\big[\sum_i \partial_{\theta_j} w_i(\theta)t_i(x)\big]$
$= \partial_{\theta_j}\log c(\theta) + E\big[\sum_i \partial_{\theta_j} w_i(\theta)t_i(x)\big]$
$\implies E\big[\sum_i \partial_{\theta_j} w_i(\theta)t_i(x)\big] = -\partial_{\theta_j}\log c(\theta)$

Part b

Starting with an intermediate step from part (a) and differentiating, we get:

$$\begin{aligned}0 = \int \Big(&h(x)c''(\theta)\exp(\sum_i w_i t_i)+\\ &h(x)c'(\theta)\exp(\sum_i w_i t_i)(\sum_i \partial_{\theta_j} w_i t_i)+\\ &h(x)c'(\theta)\exp(\sum_i w_i t_i)(\sum_i \partial_{\theta_j} w_i t_i)+\\ &h(x)c(\theta)\exp(\sum_i w_i t_i)(\sum_i \partial_{\theta_j} w_i t_i)^2+\\ &h(x)c(\theta)\exp(\sum_i w_i t_i)(\sum_i \partial^2_{\theta_j} w_i t_i)\Big)dx\end{aligned}$$

Substituting the given identities and simplifying the straight-up expected values:

$$0 = \int \Bigg(h(x)\Big(c(\theta)\partial^2_{\theta_j} \log c(\theta) + (\frac{\partial_{\theta_j} c(\theta)}{c(\theta)})^2 c(\theta)\Big) \exp(\sum_i w_i t_i) +$$

$$2h(x)c(\theta)(\partial_{\theta_j} \log c(\theta)) \exp(\sum_i w_i t_i)(\sum_i \partial_{\theta_j} w_i t_i)\Bigg) dx +$$

$$E[(\sum_i \partial_{\theta_j} w_i t_i)^2] +$$

$$E[\sum_i \partial^2_{\theta_j} w_i t_i]$$

$$= \partial^2_{\theta_j} \log c(\theta) + \left(\frac{\partial_{\theta_j} c(\theta)}{c(\theta)}\right)^2 + 2\Big(\partial_{\theta_j} \log c(\theta)\Big) E\big[\sum_i \partial_{\theta_j} w_i t_i\big] + E[(\sum_i \partial_{\theta_j} w_i t_i)^2] + E[\sum_i \partial^2_{\theta_j} w_i t_i]$$

We can substitute:

- $\frac{\partial_{\theta_j} c(\theta)}{c(\theta)} = \partial_{\theta_j} \log c(\theta)$
- $\partial_{\theta_j} \log c(\theta) = -E[\sum_i \partial_{\theta_j} w_i t_i]$
- So $\frac{\partial_{\theta_j} c(\theta)}{c(\theta)} = (E[\sum_i \partial_{\theta_j} w_i t_i])^2$

$$0 = \partial^2_{\theta_j} \log c(\theta) + \big(E[\sum_i \partial_{\theta_j} w_i t_i]\big)^2 - 2\big(E[\sum_i \partial_{\theta_j} w_i t_i]\big) E\big[\sum_i \partial_{\theta_j} w_i t_i\big] + E[(\sum_i \partial_{\theta_j} w_i t_i)^2] + E[\sum_i \partial^2_{\theta_j} w_i t_i]$$

$$= \partial^2_{\theta_j} \log c(\theta) - \big(E[\sum_i \partial_{\theta_j} w_i t_i]\big)^2 + E[(\sum_i \partial_{\theta_j} w_i t_i)^2] + E[\sum_i \partial^2_{\theta_j} w_i t_i]$$

$$= \partial^2_{\theta_j} \log c(\theta) + Var\big(\sum_i \partial_{\theta_j} w_i(\theta) t_i(x)\big) + E\big[\sum_i \partial^2_{\theta_j} w_i(\theta) t_i(x)\big]$$

Then if we rearrange the terms, we get:

$$Var\big(\sum_i \partial_{\theta_j} w_i(\theta) t_i(x)\big) = -\partial^2_{\theta_j} \log c(\theta) - E\big[\sum_i \partial^2_{\theta_j} w_i(\theta) t_i(x)\big]$$

Problem 3.30

Part a

From the example, we have:

- $h(x) = \binom{n}{x}$
- $c(p) = (1-p)^n$
- $w_1(p) = \log \frac{p}{1-p}$
- $t_1(x) = x$

In addition, the example provides:

- $w_1'(p) = \frac{1}{p(1-p)} = (p - p^2)^{-1}$
- $\big(\log c(p)\big)' = -\frac{n}{1-p} = -n(1-p)^{-1}$

Then we can see that:

- $w_1''(p) = -(p - p^2)^{-2}(1 - 2p) = -\frac{1-2p}{p^2(1-p)^2}$
- $\big(\log c(p)\big)'' = n(1 - p)^{-2}(-1) = -\frac{n}{(1-p)^2}$

Putting it all together, we get:

$Var\bigg(\frac{X}{p(1-p)}\bigg) = \frac{n}{(1-p)^2} - E\bigg[-\frac{1-2p}{p^2(1-p)^2}X\bigg]$

$\frac{1}{p^2(1-p)^2}Var(X) = \frac{np^2}{p^2(1-p)^2} + \frac{1-2p}{p^2(1-p)^2}E[X]$

$Var(X) = np^2 + (1 - 2p) np = np^2 + np - 2 np^2 = np - np^2 = n p (1-p)

Part b

From a previous problem:

- $h(x) = \frac{1}{x!}$
- $c(\theta) = e^{-\theta}$
- $w_1(\theta) = \log\theta$
- $t_1(x) = x$

Then:

- $w_1'(\theta) = \theta^{-1}$
- $w_1''(\theta) = -\theta^{-2}$
- $\log c(\theta) = -\theta$
- $\big(\log c(\theta)\big)' = -1$
- $\big(\log c(\theta)\big)'' = 0$

Putting it all together:

$E[X/\theta] = 1 \implies E[X] = \theta$

$Var(X/\theta) = E[X/\theta^2]$
$\implies \theta^{-2}Var(X) = \theta^{-2}E[X]$
$\implies Var(X) = E[X] = \theta$

3.32

Part a

From the text, we are given:

$$c^*(\eta) = \bigg(\int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)dx\bigg)^{-1}$$

Then:

$$\begin{aligned}
-\partial_{\eta_j}\log c^*(\eta) &= -\partial_{\eta_j}\log\bigg(\int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)dx\bigg)^{-1} \\
&= \partial_{\eta_j}\log\bigg(\int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)dx\bigg) \\
&= \frac{\partial_{\eta_j}\int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)dx}{\int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)}
\end{aligned}$$

We can multiply the top and bottom by $c^*(\eta)$ and move that inside the integrals since we are integrating with respect to x. Then the bottom is just $\int f(x)dx = 1$, so we can ignore it, and the top becomes (under regular conditions):

$$\begin{aligned}
&\int h(x)c^*(\eta)\Big(\partial_{\eta_j} \exp\big(\sum_i \eta_i t_i(x)\big)\Big)dx \\
&= \int t_j(x)h(x)c^*(\eta)\exp\big(\sum_i \eta_i t_i(x)\big)dx \\
&= \int t_j(x)f(x)dx \\
&= E[t_j(X)]
\end{aligned}$$

Let $\xi(\eta) = \int h(x)\exp\big(\sum_i \eta_i t_i(x)\big)dx$. Then $-\log c^*(\eta) = \log \xi(\eta)$.

Differentiating once, we get:

$$\partial_{\eta_j} \log \xi(\eta) = \frac{\partial_{\eta_j}\xi(\eta)}{\xi(\eta)}$$

Differentiating twice, we get:

$$\begin{aligned}
\partial^2_{\eta_j} \log \xi(\eta) &= \frac{\partial^2_{\eta_j}\xi - (\partial_{\eta_j}\xi)^2}{\xi^2} \\
&= \frac{\partial^2_{\eta_j}\xi}{\xi} - \left(\frac{\partial_{\eta_j}\xi}{\xi}\right)^2
\end{aligned}$$

From before, we saw that $\frac{\partial_{\eta_j}\xi}{\xi} = E[t_j(X)]$, so all that remains is to show that $\frac{\partial^2_{\eta_j}\xi}{\xi} = E\Big[\big(t_j(X)\big)^2\Big]$

$$\begin{aligned}
\frac{\partial^2_{\eta_j}\xi}{\xi} &= \frac{\partial^2_{\eta_j}\int h(x)\exp(\sum_i \eta_i t_i)dx}{\int h(x)\exp(\sum_i \eta_i t_i)dx} \\
&= \frac{\int t_j^2 h(x)\exp(\sum_i \eta_i t_i)dx}{\int h(x)\exp(\sum_i \eta_i t_i)dx}
\end{aligned}$$

Again, we can multiply the top and bottom by $c^*(\eta)$, which can then be moved inside the integrals:

$$\begin{aligned}
&= \frac{\int t_j^2 h(x)c^*(\eta)\exp(\sum_i \eta_i t_i)dx}{\int h(x)c^*(\eta)\exp(\sum_i \eta_i t_i)dx} \\
&= \frac{\int t_j(x)^2 f(x)dx}{\int f(x)dx}
\end{aligned}$$

The top is just an expected value while the bottom is the integral of a density function, which is just 1. So we have:

$$= E\Big[\big(t_j(X)\big)^2\Big]$$

Therefore,

$$\partial^2_{\eta_j} \log \xi(\eta) = \frac{\partial^2_{\eta_j}\xi}{\xi} - \left(\frac{\partial_{\eta_j}\xi}{\xi}\right)^2 = E\Big[\big(t_j(X)\big)^2\Big] - \Big(E\big[t_j(X)\big]\Big)^2 = Var\big(t_j(X)\big)$$

3.33

Part a

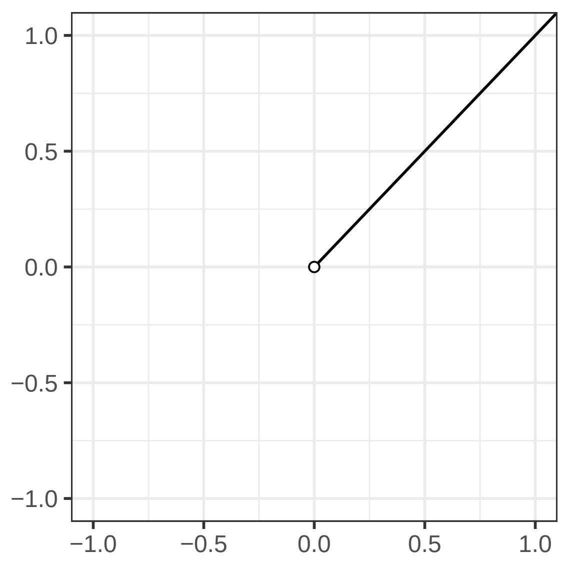
We can write the density function as $f(x \mid \theta) = (2\pi\theta)^{-1/2} \exp(-x^2/2\theta) \exp(x) \exp(-\theta/2)$, so we have:

- $h(x) = \exp(x)$
- $c(\theta) = (2\pi\theta)^{-1/2} \exp(-\theta/2)$
- $w_1(\theta) = (2\theta)^{-1}$
- $t_1(x) = -x^2$

$(\mu, \sigma^2) = (\theta, \theta)$, and $\theta > 0$.

```r
library(ggplot2)

ggplot() +
  geom_line(aes(x = c(0, 5), y = c(0, 5))) +
  coord_cartesian(xlim = c(-1, 1),
                  ylim = c(-1, 1)) +
  geom_point(aes(x = 0, y = 0), shape = 21, fill = 'white') +
  theme_bw() +
  labs(x = NULL, y = NULL)
```



7.39

$\partial_\theta^2 \log f(X \mid \theta)$
$= \partial_\theta(\frac{\partial_\theta f(X|\theta)}{f(X|\theta)})$
$= \frac{(\partial_\theta^2 f)(f) - (\partial_\theta f)^2}{f^2}$
$= \frac{\partial_\theta^2 f(X|\theta)}{f(X|\theta)} - \left(\frac{\partial_\theta f(X|\theta)}{f(X|\theta)}\right)^2$

So $E[\partial_\theta^2 \log f(X \mid \theta)] = E\left[\frac{\partial_\theta^2 f(X|\theta)}{f(X|\theta)}\right] - E\left[\left(\frac{\partial_\theta f(X|\theta)}{f(X|\theta)}\right)^2\right]$

So all we have to show is $E\left[\frac{\partial_\theta^2 f(X|\theta)}{f(X|\theta)}\right] = \int \frac{\partial_\theta^2 f(x|\theta)}{f(x|\theta)} f(x \mid \theta)dx$
$= \int \partial_\theta^2 f(x \mid \theta)dx$
$= \partial_\theta^2 \int f(x \mid \theta)dx$ (under regular condition)
$= \partial_\theta^2 1 = 0.$

Therefore, $E[\partial_\theta^2 \log f(X \mid \theta)] = -E\left[\left(\frac{\partial_\theta f(X|\theta)}{f(X|\theta)}\right)^2\right]$

7.40

$E[\bar{X}] = \frac{1}{n}\sum_i^n E[X_i] = np/n = p$, so $\bar{X}$ is unbiased.

$Var(\bar{X}) = \frac{1}{n^2}\sum_i^n Var(X_i) = \frac{p(1-p)}{n}$.

Here, $g(p) = p$, so $g'(p) = 1$.

The log density is $x \log p + (1-x)\log(1-p)$, and taking the partial w.r.t. p, we get $\frac{x}{p} - \frac{1-x}{1-p}$. Taking another derivative gets us $-\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$

Then $E[\frac{x}{p^2} + \frac{1-x}{(1-p)^2}] = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1-p+p}{p(1-p)} = \frac{1}{p(1-p)}$

Then $I(p) = \frac{n}{p(1-p)}$, so the Cramer-Rao lower bound is $\frac{p(1-p)}{n}$, which is the variance of $\bar{X}$.

Part 2

Problem 1

We know that $\big(Cov(X,Y)\big)^2 \leq Var(X)Var(Y)$. Then $Var(X) \geq \frac{(Cov(X,Y))^2}{Var(Y)}$.

Let "X" in this case be $W(X)$ and "Y" be $\partial_\theta \log f(X \mid \theta)$. Then we need to show:

- $E\Big[\big(\partial_\theta \log f(X \mid \theta)\big)^2\Big] = Var\big(\partial_\theta \log f(X \mid \theta)\big)$
- $Cov\big(W(X), \partial_\theta \log f(X \mid \theta)\big) = \partial_\theta E[W(X)]$

To show the second part, we can use $Cov(X,Y) = E[XY] - E[X]E[Y]$, and the first part of this can be obtained as follows:

$$\begin{aligned}
\partial_\theta E\big[W(X)\big] &= \partial_\theta \int W(x) f(x \mid \theta)dx \\
&= \int W(x)\big(\partial_\theta f(x \mid \theta)\big)dx \\
&= \int W(x)\big(\partial_\theta f(x \mid \theta)\big)\frac{f(x \mid \theta)}{f(x \mid \theta)}dx \\
&= E\Big[W(X)\frac{\partial_\theta f(X \mid \theta)}{f(X \mid \theta)}\Big] \\
&= E\Big[W(X)\partial_\theta \log f(X \mid \theta)\Big]
\end{aligned}$$

We can then note that:

$$E\Big[\partial_\theta \log f(X \mid \theta)\Big] = E\Big[\frac{\partial_\theta f(X \mid \theta)}{f(X \mid \theta)}\Big]$$

$$= \int \frac{\partial_\theta f(x \mid \theta)}{f(x \mid \theta)} f(x \mid \theta) dx$$
$$= \int \partial_\theta f(x \mid \theta) dx$$
$$= \partial_\theta \int f(x \mid \theta) dx$$
$$= \partial_\theta (1) = 0$$

So the second part ($E[X]E[Y]$) is 0. Therefore,

$$\Big(Cov(W(X), \partial_\theta \log f(X \mid \theta))\Big)^2 = \Big(E\Big[W(X)\partial_\theta \log f(X \mid \theta)\Big]\Big)^2 = \Big(\partial_\theta E\big[W(X)\big]\Big)^2$$

Which is precisely the numerator of the Cramer-Rao inequality.

To show the first bullet point, we can note that:

$$Var\big(\partial_\theta \log f(X \mid \theta)\big) = E\Big[\big(\partial_\theta \log f(X \mid \theta)^2\Big] - E\Big[\partial_\theta \log f(X \mid \theta\Big]^2$$

We already saw the second part of this (square of the expectation) is 0, so we can ignore it. Then:

$$Var\big(\partial_\theta \log f(X \mid \theta)\big) = E\Big[\big(\partial_\theta \log f(X \mid \theta)^2\Big]$$

Which is precisely the denominator of the Cramer-Rao inequality.

Problem 2

Part a

W_1

Let $T = \sum_i X_i$. Note that $T \sim Poisson(n\theta)$. Then we can write $E[W_1] = E[e^{-\bar{X}}] = E[e^{-T/n}]$.

$E[e^{-T/n}] = \sum_{t=0} \frac{e^{-t/n} e^{-n\theta} (n\theta)^t}{t!}$
$= e^{-n\theta} \sum_t \frac{(e^{-1/n} n\theta)^t}{t!}$
$= e^{-n\theta} e^{e^{-1/n} n\theta}$
$= \exp\big(-\theta(n - ne^{-1/\theta})\big)$

$E[(e^{-T/n})^2] = E[e^{-2T/n}]$
$= \sum_t \frac{e^{-2t/n} e^{-n\theta} (n\theta)^t}{t!}$
$= e^{-n\theta} e^{e^{-2/n} n\theta} = \exp\big(-\theta(n - ne^{-2/n})\big)$
Then $Var(W_1) = \exp\big(-\theta(n - ne^{-2/n})\big) - \exp\big(-2\theta(n - ne^{-1/\theta})\big)$.

W_2

$E[W_2] = E[(1-1/n)^T]$
$= \sum_t (1-1/n)^t \frac{e^{-n\theta}(n\theta)^t}{t!}$
$= e^{-n\theta} \sum_t \frac{((1-1/n)n\theta)^t}{t!}$
$= e^{-n\theta} e^{(1-1/n)n\theta}$
$= e^{-\theta}$

$E[W_2^2] = E[(1-1/n)^{2T}]$
$= \sum_t \frac{(1-1/n)^{2t} e^{-n\theta}(n\theta)^t}{t!}$
$= e^{-n\theta} \sum_t \frac{(n\theta(1-1/n)^2)^t}{t!}$
$= \exp\Big(-n\theta\big(1-(1-1/n)^2\big)\Big)$
$= e^{-2\theta+\theta/n}$
Then $Var(W_2) = e^{-2\theta+\theta/n} - e^{-2\theta}$
$= e^{-2\theta}(e^{\theta/n} - 1)$.

W_3

$E[\nVdash_{\{X_1=0\}}] = P(X_1 = 0) = e^{-\theta}$.

This is just a Bernoulli trial, so the variance is just $p(1-p)$ where $p = P(X_1 = 0)$. So $Var(W_3) = e^{-\theta}(1-e^{-\theta})$.

W_4

$E[\frac{1}{n}\sum_i \nVdash_{\{X_i=0\}}]$
$= \frac{1}{n}\sum_i E[\nVdash_{\{X_i=0\}}]$
$= \frac{1}{n}\sum_i P(X_i = 0)$
$= \frac{1}{n} n e^{-\theta} = e^{-\theta}$

$Var(\frac{1}{n}\sum_i \nVdash_{\{X_i=0\}})$
$= \frac{1}{n^2}\sum_i Var(\nVdash_{\{X_i=0\}})$
$= \frac{1}{n^2}\sum_i p(1-p) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$
$= \frac{e^{-\theta}(1-e^{-\theta})}{n}$

Part b

$g(\theta) = e^{-\theta} \implies g'(\theta) = -e^{-\theta} \implies \big(g'(\theta)\big)^2 = e^{-2\theta}$

$f(x \mid \theta) = \prod_i \frac{e^{-\theta}\theta_i^x}{x_i!}$
$\implies \log f = -n\theta + \sum_i x_i \log\theta - \sum_i \log x_i!$
$\implies \partial_\theta \log f = -n + \sum_i x_i/\theta$
$\implies (\partial_\theta \log f)^2 = (\sum_i x)^2/\theta^2 - 2n\sum_i x_i/\theta + n^2$
$\implies E[(\partial_\theta \log f)^2] = E[(\sum_i x_i)^2/\theta^2 - 2n\sum_i x_i/\theta + n^2]$
$= \frac{n\theta + (n\theta)^2}{\theta^2} - 2n^2 + n^2$
$= \frac{n}{\theta}$

So the Cramer-Rao lower bound is $\frac{\theta e^{-2\theta}}{n}$.