Chapter 2

Efficiency and the Cramér-Rao Inequality

Clearly we would like an unbiased estimator \( \hat{\phi}(X) \) of \( \phi(\theta) \) to produce, in the long run, estimates which are fairly concentrated i.e. have high precision. An estimator \( \hat{\phi} \) is unlikely to impress anybody if it produces widely varying estimates on different occasions. This suggests that \( \text{Var}(\hat{\phi}) \) should be as small as possible. But how small can \( \text{Var}(\hat{\phi}) \) be? The answer is given by the following result.

**The Cramér-Rao Inequality** Let \( X = (X_1, X_2, \ldots, X_n) \) be a random sample from a distribution with p.m/d.f. \( f(x|\theta) \), where \( \theta \) is a scalar parameter. Under certain regularity conditions on \( f(x|\theta) \), for any unbiased estimator \( \hat{\phi}(X) \) of \( \phi(\theta) \)

\[
\text{Var}(\hat{\phi}) \geq \frac{[\frac{d}{d\theta} \phi(\theta)]^2}{\mathbb{E}(S^2(X))} \tag{2.1}
\]

where

\[
S(X) = \frac{\partial}{\partial \theta} \log f_X(X|\theta) = \frac{\partial}{\partial \theta} \log \left( \prod_{i=1}^{n} f(X_i|\theta) \right) \tag{2.2}
\]

**Remark**

(1) The function \( S(X) \) is called the Score statistic and will be seen to play an important role in statistical inference.
(2) The bound \( \left[ \frac{d^2\phi(\theta)}{d\theta^2} \right]^2 \) on the r.h.s. of the Cramér-Rao inequality is known as the Cramér-Rao Lower Bound (CRLB) for the variance of an unbiased estimator of \( \phi(\theta) \).

(3) An estimator whose variance is equal to the CRLB is called a most efficient estimator.

The efficiency \( eff(\hat{\phi}) \) of an unbiased estimator \( \hat{\phi} \) of \( \phi(\theta) \) is defined to be the ratio

\[
\frac{CRLB}{Var(\hat{\phi})}
\]

Note that \( 0 \leq eff(\hat{\phi}) \leq 1 \) for all \( \theta \in \Theta \).

(4) The quantity

\[
I(\theta) = E(S^2(X))
\]

is called the Fisher information on the parameter \( \theta \) contained in the sample \( X \). Note that

(i) the Fisher information is, in general, a function of the value of \( \theta \),

(ii) the higher the Fisher information i.e. the more information there is in the sample, the smaller the CRLB and consequently the smaller the variance of the most efficient estimator (when it exists).

**Proof of the Cramér-Rao inequality** (for case when observations are continuous. For discrete observations replace integrals by sums.)

Since \( \hat{\phi} \) is an unbiased estimator,

\[
E(\hat{\phi}) = \phi(\theta) \quad \text{for all } \theta
\]

i.e.

\[
\int \ldots \int \hat{\phi}(x_1, x_2, \ldots, x_n) f_X(x_1, x_2, \ldots, x_n|\theta) dx_1 dx_2 \ldots dx_n = \phi(\theta)
\]

and for convenience we write

\[
\int \hat{\phi}(x) f_X(x|\theta) dx = \phi(\theta)
\]
Differentiating both sides with respect to $\theta$ and interchanging the order of differentiation and integration (assuming this is possible i.e. regularity conditions hold) we get

$$\int \hat{\phi}(x) \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx = \frac{d}{d\theta} \phi(\theta)$$

(2.5)

We now note that from the definition of the score statistic we have

$$S(x) = \frac{\partial}{\partial \theta} \log f_X(x|\theta) = \frac{1}{f_X(x|\theta)} \frac{\partial}{\partial \theta} f_X(x|\theta)$$

and hence

$$\frac{\partial}{\partial \theta} f_X(x|\theta) = S(x) f_X(x|\theta)$$

(2.6)

Substituting in 2.4 we have

$$\int \hat{\phi}(x) S(x) f_X(x|\theta) \, dx = \frac{d}{d\theta} \phi(\theta)$$

i.e.

$$E \left( \hat{\phi}(X) S(X) \right) = \frac{d}{d\theta} \phi(\theta)$$

(2.7)

Next we note that the score statistic has zero expectation since

$$E(S(X)) = \int S(x) f_X(x|\theta) \, dx = \int \frac{\partial}{\partial \theta} f_X(x|\theta) \, dx$$

because of 2.6. Again interchanging differentiation with integration (i.e. assuming necessary regularity conditions hold) we get

$$E(S(X)) = \frac{\partial}{\partial \theta} \int f_X(x|\theta) \, dx = \frac{\partial}{\partial \theta} 1 = 0$$

(2.8)

Hence

1. $Cov(\hat{\phi}, S) = E(\hat{\phi}, S) - E(\hat{\phi}) E(S) = E(\hat{\phi}, S) = \frac{d}{d\theta} \phi(\theta)$

(2.9)

the last equality following from 2.7, and

2. $Var(S) = E(S^2) - E(S)^2 = E(S^2) = I(\theta)$

(2.10)
Finally, since for any two random variables the correlation coefficient between them is always less than 1 in absolute value, we have
\[ \rho^2(\hat{\phi}, S) \leq 1 \]
i.e.
\[ \frac{[Cov(\hat{\phi}, S)]^2}{Var(\hat{\phi})Var(S)} \leq 1 \]
or
\[ Var(\hat{\phi}) \geq \frac{[Cov(\hat{\phi}, S)]^2}{Var(S)} = \frac{\left[ \frac{d}{d\theta} \phi(\theta) \right]^2}{I(\theta)} \]
which proves the Cramér-Rao inequality.

This proof has also provided us with the following result.

**Corollary** Under certain regularity conditions the score statistic \( S(X) \) has mean zero and variance equal to the Fisher information i.e.
\[
\begin{align*}
\mathbf{E}(S(X)) &= 0 \quad (2.11) \\
Var(S(X)) &= I(\theta) \quad (2.12)
\end{align*}
\]

**Efficiency:** Since we now know that the CRLB is the smallest value that the variance of any estimator of \( \phi(\theta) \) can be, the efficiency of an estimator of \( \theta \) can be assessed by comparing its variance against the CRLB. In fact, as we have seen, we have the following definition for the efficiency \( eff(\hat{\phi}) \) of an estimator \( \hat{\phi} \) of \( \phi(\theta) \).

**Definition** The efficiency of an estimator \( \hat{\phi} \) of \( \phi(\theta) \) is defined to be
\[
eff(\hat{\phi}) = \frac{CRLB}{Var(\hat{\phi})} = \frac{\left[ \frac{d}{d\theta} \phi(\theta) \right]^2}{I(\theta)Var(\hat{\phi})} \quad (2.13)
\]
Remark Note that
\[ 0 \leq \text{eff}(\hat{\phi}) \leq 1 \]
and that the closer to 1 the value of \( \text{eff}(\hat{\phi}) \) is, the more efficient the estimator \( \hat{\phi} \) is; indeed when \( \hat{\phi} \) is most efficient then \( \text{eff}(\hat{\phi}) = 1 \). Note that a most efficient estimator of \( \phi(\theta) \) need not always exist. When dealing with unbiased estimators, if a most efficient unbiased estimator does not exist (i.e. if there is no unbiased estimator whose variance is as small as the CRLB), we seek out the estimator whose variance is the closest to the CRLB i.e. the unbiased estimator with the highest efficiency.

When does a most efficient estimator exist? The following result not only gives the necessary and sufficient conditions for a most efficient estimator to exist, it also provides this estimator.

Theorem Under the same regularity conditions as in the Cramèr-Rao Inequality, there exists an unbiased estimator \( \hat{\phi} \) of \( \phi(\theta) \) whose variance attains the CRLB (i.e. there exists a most efficient unbiased estimator \( \hat{\phi} \) of \( \phi(\theta) \)) if and only if the score statistic \( S(x) \) can be expressed in the form
\[
S(X) = \alpha(\theta) \left[ \hat{\phi}(X) - \phi(\theta) \right]
\]
in which case
\[
\alpha(\theta) = \frac{I(\theta)}{d\phi(\theta)}
\]
or equivalently if and only if the function
\[
S(X). \frac{d\phi(\theta)}{I(\theta)} + \phi(\theta)
\]
is independent of \( \theta \) and is only dependent on \( X \), in which this statistic is the unbiased most efficient estimator of \( \phi(\theta) \).

Proof. Recall that the Cramèr-Rao inequality was a result of the fact that for any unbiased estimator \( \hat{\phi} \) of \( \phi(\theta) \), \( \rho^2(\hat{\phi}, S) \leq 1 \). Hence the CRLB is attained if and only if \( \rho^2(\hat{\phi}, S) = 1 \). But this can only happen if and only if \( \hat{\phi} \) and \( S \) are linearly related random variables i.e. if and only if there exist functions \( \alpha(\theta) \) and \( \beta(\theta) \) which are independent of \( X \) such that
\[
S(X) = \alpha(\theta) \hat{\phi}(X) + \beta(\theta)
\]
Recall, however, (from corollary 4) that \( S(X) \) has zero expectation and that \( \hat{\phi} \) is an unbiased estimator of \( \phi(\theta) \). Hence taking expectations of both sides of 2.15 we get
\[
0 = \alpha(\theta) \phi(\theta) + \beta(\theta)
\]
or \( \beta(\theta) = -\alpha(\theta) \phi(\theta) \) which, on substituting in 2.15 we get
\[
S(X) = \alpha(\theta) \hat{\phi}(X) - \alpha(\theta) \phi(\theta) = \alpha(\theta) \left[ \hat{\phi}(X) - \phi(\theta) \right]
\]
confirming (2.14). Multiplying both sides of (2.16) by \( S(X) \) and taking expectations throughout we get
\[
E \left( \left( S(X) \right)^2 \right) = \alpha(\theta) \left[ E \left( \hat{\phi}(X) S(X) \right) - \phi(\theta) E( S(X)) \right]
\]
and in view of corollary 4 we have
\[
E \left( \left( S(X) \right)^2 \right) = \alpha(\theta) \text{Cov} \left( \hat{\theta}(X), S(X) \right)
\]
which by (2.9) and (2.12) reduces to
\[
I(\theta) = \alpha(\theta) \cdot \frac{d}{d\theta} \phi(\theta)
\]
completing the proof.

**2.0.3 A computationally useful result**

The expression
\[
I(\theta) = E \left( S^2(X) \right) = E \left( \left( \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right)^2 \right)
\]
for the Fisher information does not lend itself to easy computation. The alternative form given below is computationally more attractive and you are strongly recommended to use it when finding the Fisher information in examples, exercises or exam questions.

**Proposition** Under the same regularity conditions as in the Cramér-Rao Inequality
\[ I(\theta) = -E \left( \frac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) \right) \]  

(2.17)

**Proof**

\[
\frac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) = \frac{\partial}{\partial \theta} \left\{ \frac{1}{f_X(X|\theta)} \frac{\partial}{\partial \theta} f_X(X|\theta) \right\} \\
= -\frac{1}{[f_X(X|\theta)]^2} \left[ \frac{\partial}{\partial \theta} f_X(X|\theta) \right]^2 + \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta) \\
= -\left[ \frac{\partial}{\partial \theta} \log f_X(X|\theta) \right]^2 + \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta) \\
= -S^2(X) + \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta)
\]

Taking expectations throughout we get

\[
E \left( \frac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) \right) = -E \left( S^2(X) \right) + E \left( \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta) \right) \\
= -I(\theta) + E \left( \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta) \right) \quad (2.18)
\]

But (assuming continuous observations – replace integrals by summations if observations are discrete.)

\[
E \left( \frac{1}{f_X(X|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(X|\theta) \right) = \int \left( \frac{1}{f_X(x|\theta)} \frac{\partial^2}{\partial \theta^2} f_X(x|\theta) \right) f_X(x|\theta) \, dx \\
= \int \frac{\partial^2}{\partial \theta^2} f_X(x|\theta) \, dx \\
= \frac{\partial^2}{\partial \theta^2} \int f_X(x|\theta) \, dx \\
= \frac{\partial^2}{\partial \theta^2} \cdot 1 = 0
\]

i.e. the second term in 2.18 is zero and the result follows.

**Example** Let \( X = (X_1, X_2, \ldots, X_n) \) be a random sample from the exponential distribution with parameter \( \theta \). Find the Cramér-Rao lower bound for
unbiased estimators of \( \theta \). Does there exist an unbiased estimator of \( \theta \) whose variance is equal to the CRLB?

**Solution:** The Cramér-Rao inequality is

\[
Var(\hat{\theta}) \geq \frac{1}{I(\theta)}
\]

with

\[
I(\theta) = \mathbb{E}\left( S^2(X) \right) = -\mathbb{E}\left( \frac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) \right)
\]

But

\[
f_X(x|\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}
\]

\[
\Rightarrow \log f_X(x|\theta) = n \log \theta - \theta \sum_{i=1}^{n} x_i
\]

\[
\Rightarrow S(x) = \frac{\partial}{\partial \theta} \log f_X(x|\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i
\]

\[
\Rightarrow \frac{\partial^2}{\partial \theta^2} \log f_X(x|\theta) = -\frac{n}{\theta^2}
\]

Hence

\[
I(\theta) = -\mathbb{E}\left( \frac{\partial^2}{\partial \theta^2} \log f_X(X|\theta) \right) = -\mathbb{E}\left( -\frac{n}{\theta^2} \right) = \frac{n}{\theta^2}
\]

and

\[
Var(\hat{\theta}) \geq \frac{1}{I(\theta)} = \frac{\theta^2}{n}.
\]

Now

\[
S(X) = \frac{n}{\theta} - \sum_{i=1}^{n} X_i
\]

and this cannot be expressed in the form \( \alpha(\theta) \left[ \hat{\theta}(X) - \theta \right] \) i.e. there does not exist an unbiased estimator of \( \theta \) whose variance is equal to the CRLB. This can be confirmed by the fact that with \( \phi(\theta) = \theta \) and \( \phi'(\theta) = \frac{d\phi(\theta)}{d\theta} \),

\[
S(X) \frac{\phi'(\theta)}{I(\theta)} + \phi(\theta) = \left( \frac{n}{\theta} - \sum_{i=1}^{n} X_i \right) \frac{\theta^2}{n} + \theta
\]

which is not independent of \( \theta \).
**Example** In the last example, what is the CRLB for the variances of unbiased estimators of $1/\theta$? Does there exist an unbiased estimator of $1/\theta$ whose variance is equal to the CRLB?

**Solution:** For any unbiased estimator $\hat{\psi}$ of $1/\theta$

$$
\text{Var}(\hat{\psi}) \geq \left[ \frac{\partial}{\partial \theta} \left( \frac{1}{\theta} \right) \right]^2 = \left[ \frac{-1}{\theta^2} \right]^2 = \frac{1}{n\theta^2}
$$

Further from the last example

$$
S(X) = \frac{n}{\theta} - \sum_{i=1}^{n} X_i = -n \left[ \frac{\sum_{i=1}^{n} X_i}{n} - \frac{1}{\theta} \right] = -n \left[ \bar{X} - \frac{1}{\theta} \right]
$$

which implies that $\bar{X}$ is an unbiased estimator of $1/\theta$ whose variance is equal to the CRLB $= \frac{1}{n\theta^2}$. Further, by (??)

$$
-n = \frac{I(\theta)}{\frac{\partial}{\partial \theta} \left( \frac{1}{\theta} \right)} = \frac{I(\theta)}{-\frac{1}{\theta^2}}
$$

$$
\Rightarrow I(\theta) = -n \left( -\frac{1}{\theta^2} \right) = \frac{n}{\theta^2}
$$

which agrees with the direct computation of $I(\theta)$.

**A Remark on the regularity conditions:** We have seen that the regularity conditions we require are those which allows us to interchange differentiation with integration on two occasions. These required regularity conditions will normally fail to hold if

1. the density/mass function $f(x|\theta)$ does not tail off rapidly enough (i.e. high values of $x$ can be produced with relatively high probability) so that the integrals converge slowly. In practice most of the distributions we deal with satisfy this regularity condition.
2. the effective range of integration depends on $\theta$. This will be the case if the range of possible values $X$ of $X$ depends on $\theta$. One such case is the case when $X$ has the Uniform $U(0, \theta)$ distribution over the interval $(0, \theta)$ so that

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} & \text{for } 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

2.0.4 Multivariate version of the Cramér-Rao Inequality.

Theorem: Let $X$ be a sample of observations with joint p.d.m.f. $f_X(x|\theta)$ dependent on a vector parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$. Let $\phi(\theta)$ be a real valued function of $\theta$. Then under some regularity conditions, for any unbiased estimator $\hat{\phi}(X)$ of $\phi(\theta)$

$$\text{Var} \left( \hat{\phi} \right) \geq \delta^T [I(\theta)]^{-1} \delta$$

where $\delta$ is the vector of derivatives of $\phi(\theta)$ i.e.

$$\delta = \left( \frac{\partial}{\partial \theta_1} \phi(\theta), \frac{\partial}{\partial \theta_2} \phi(\theta), \ldots, \frac{\partial}{\partial \theta_k} \phi(\theta) \right) \quad (2.19)$$

and the matrix $I(\theta)$ is the Fisher Information matrix with $(i, j)$th element

$$I_{ij}(\theta) = \mathbf{E} \left( S_i(X) S_j(X) \right)$$

$$= - \mathbf{E} \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f_X(X|\theta) \right) \quad (2.20)$$

$i, j = 1, 2, \ldots, k$. Here

$$S_i(X) = \frac{\partial}{\partial \theta_i} \log f_X(X|\theta) \quad (2.21)$$

is the $i$th score statistic $i = 1, 2, \ldots, k$. In particular, if $\hat{\theta}_r(X)$ is an unbiased estimator of the $r$th parameter $\theta_r$ then under the same regularity conditions

$$\text{Var} \left( \hat{\theta}_r \right) \geq J_{rr}(\theta)$$
where $J_{rr}(\theta)$ is the $r$th diagonal element of the inverse matrix $[I(\theta)]^{-1}$, $r = 1, 2, \ldots, k$.

**Proof:** The proof is very similar to that of the scalar case and its details are not provided. Below, however, is a sketch of the proof. The details can be filled in by you.

1. Following exactly the same approach as in the scalar case show that
   \[ \frac{\partial}{\partial \theta_j} \phi(\theta) = Cov(\hat{\phi}(X), S_j(X)) \]

2. Make use of the fact that for any coefficients $c_1, c_2, \ldots, c_k$,
   \[ Var(\hat{\phi}(X) - \sum_{i=1}^{k} c_i S_i(X)) \geq 0 \quad (2.22) \]
   In particular, this is true for the values $c_1 = c_1^*, c_2 = c_2^*, \ldots, c_k = c_k^*$ which minimize $Var(\hat{\phi}(X) - \sum_{i=1}^{k} c_i S_i(X))$.

3. Using 1. show that
   \[ Var(\hat{\phi}(X) - \sum_{i=1}^{k} c_i S_i(X)) = Var(\hat{\phi}(X)) + c^T I(\theta)c - 2c^T \delta \]
   with $c = (c_1, c_2, \ldots, c_k)^T$ and then using calculus show that
   \[ c^* = (c_1^*, c_2^*, \ldots, c_k^*)^T = [I(\theta)]^{-1} \delta \]

4. Thus show that
   \[ \min(Var(\hat{\phi}(X) - \sum_{i=1}^{k} c_i S_i(X))) = Var(\hat{\phi}(X)) - \delta^T [I(\theta)]^{-1} \delta \]
   and because of 2.22 that the result follows.

**Remark:** As in the univariate case the proof also shows that

1. \[ E(S_i(X)) = 0 \quad i = 1, 2, \ldots, k. \] (2.23)
28 CHAPTER 2. EFFICIENCY AND THE CRAMÉR-RAO INEQUALITY

2.

\[ I_{ij}(\theta) = \text{Cov}(S_i(X), S_j(X)) \quad i, j = 1, 2, \ldots, k \]  \quad (2.24)

Example Let \( X_1, X_2, \ldots, X_n \) be a random sample from the \( N(\mu, \sigma^2) \) distribution. Find the CRLB and, in cases 1. and 2. check whether it is equalled, for the variance of an unbiased estimator of

1. \( \mu \) when \( \sigma^2 \) is known,
2. \( \sigma^2 \) when \( \mu \) is known
3. \( \mu \) when \( \sigma^2 \) is unknown
4. \( \sigma^2 \) when \( \mu \) is unknown
5. the coefficient of variation \( \frac{\sigma}{\mu} \) when both \( \mu \) and \( \sigma \) are unknown.

Solution: The sample joint p.d.f. is

\[ f_X(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma}\right)^2\right) \]

and

\[ \log f_X(x|\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \]

1. When \( \sigma^2 \) is known \( \theta = \mu \) and

\[ \log f_X(x|\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{x_i - \theta}{\sigma}\right)^2 \]

\[ S(x) = \frac{\partial}{\partial \theta} \log f_X(x|\theta) = \sum_{i=1}^{n} \frac{(x_i - \theta)}{\sigma^2} = \frac{n}{\sigma^2} [\bar{x} - \theta] \]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \). By the theorem on page 21 this implies that \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) is an unbiased estimator of \( \theta = \mu \) whose variance equals the CRLB and that \( \frac{n}{\sigma^2} = I(\theta) \) i.e. CRLB = \( \frac{\sigma^2}{n} \). Thus \( \bar{X} \) is a most efficient estimator.
2. When \( \mu \) is known but \( \sigma^2 \) is unknown, \( \theta = \sigma^2 \) and

\[
\log f_X(x|\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\theta) - \frac{1}{2\theta} \sum_{i=1}^{n} (x_i - \mu)^2
\]

Hence

\[
S(x) = \frac{\partial}{\partial \theta} \log f_X(x|\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - \mu)^2
\]

\[
= \frac{n}{2\theta^2} \left[ \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 - \theta \right]
\]

By the theorem on page 21, \( \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \) is an unbiased estimator of \( \theta = \sigma^2 \) and \( \frac{n}{2\theta^2} = I(\theta) \) i.e. the CRLB = \( \frac{2\sigma^4}{n} \)

3. and 4. Case both \( \mu \) and \( \sigma^2 \) unknown Here \( \theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \) i.e. \( \theta_1 = \mu \) and \( \theta_2 = \sigma^2 \)

\[
f_X(x|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} \exp(-\frac{1}{2} (x_i - \mu)^2 / \sigma^2)
\]

\[
\propto \theta_2^{n/2} \exp \left( -\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2 \right)
\]

and

\[
\log f_X(x|\theta) = -\frac{n}{2} \log \theta_2 - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2
\]

Thus

\[
\frac{\partial}{\partial \theta_1} \log f_X(x|\theta) = \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)
\]

\[
\frac{\partial^2}{\partial \theta_1^2} \log f_X(x|\theta) = -\frac{n}{\theta_2^2}
\]

\[
\frac{\partial^2}{\partial \theta_2 \partial \theta_1} \log f_X(x|\theta) = -\frac{1}{\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)
\]

\[
\frac{\partial}{\partial \theta_2} \log f_X(x|\theta) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2
\]
\[
\frac{\partial^2}{\partial \theta^2} \log f_X(x|\theta) = n \frac{1}{\theta^2} - \frac{1}{\theta^2} \sum_{i=1}^{n} (x_i - \theta_1)^2
\]

Consequently
\[
I_{11}(\theta) = -E \left( n \frac{1}{\theta^2} \right) = n \frac{\theta_2}{\theta_1}
\]
\[
I_{12}(\theta) = -E \left( \frac{1}{\theta^2} \sum_{i=1}^{n} (X_i - \theta_1) \right) = 0
\]
\[
I_{22}(\theta) = -E \left( \frac{n}{2\theta^2} - \frac{1}{\theta^2} \sum_{i=1}^{n} (X_i - \theta_1)^2 \right) = \frac{n}{2\theta_2^2}
\]
i.e.
\[
I(\theta) = \begin{bmatrix}
\frac{n}{\theta_2} & 0 \\
0 & \frac{n}{2\theta_2^2}
\end{bmatrix}
\]
and
\[
[I(\theta)]^{-1} = J(\theta) = \begin{bmatrix}
\theta_2 & 0 \\
n \frac{2\theta_2^2}{n}
\end{bmatrix}
\begin{bmatrix}
\sigma^2 & 0 \\
0 & 2\sigma^4
\end{bmatrix}
\]

Consequently, for unbiased estimators \( \hat{\mu}, \hat{\sigma}^2 \) of \( \mu \) and \( \sigma^2 \) respectively

\[
Var(\hat{\mu}) \geq \frac{\sigma^2}{n}
\]
and
\[
Var(\hat{\sigma}^2) \geq \frac{2\sigma^4}{n}
\]

5. We were now interested in estimating the ratio \( \frac{\sigma}{\mu} = \frac{\sqrt{\theta_2}}{\theta_1} \) when both \( \mu = \theta_1 \) and \( \sigma^2 = \theta_2 \) are unknown. Since
\[
\left( \frac{\partial}{\partial \theta_1} \left( \frac{\sqrt{\theta_2}}{\theta_1} \right), \frac{\partial}{\partial \theta_2} \left( \frac{\sqrt{\theta_2}}{\theta_1} \right) \right) = \left( -\frac{\sqrt{\theta_2}}{\theta_1^2}, \frac{1}{2\theta_1\sqrt{\theta_2}} \right) = \left( -\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma} \right)
\]
it follows that if \( \hat{\phi} \) is an unbiased estimator of \( \frac{\sigma}{\mu} \) then
\[
Var(\hat{\phi}) \geq \left( -\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma} \right) I(\theta)^{-1} \left( -\frac{\sigma}{\mu^2}, \frac{1}{2\mu\sigma} \right)
\]
\[
\begin{bmatrix}
-\frac{\sigma}{\mu^2} & \frac{1}{2\mu \sigma} \\
\frac{\sigma^2}{n} & 0 \\
0 & \frac{2\sigma^4}{n}
\end{bmatrix}
\begin{bmatrix}
-\frac{\sigma}{\mu^2} \\
\frac{1}{2\mu \sigma}
\end{bmatrix}
\]

\[
\frac{\sigma^4}{n \mu^4} \left[ 1 + \frac{\mu^2}{2\sigma^2} \right]
\]

**Example** Let \(X_1, X_2, \ldots, X_n\) be a random sample from the trinomial distribution with p.m.f.

\[
f(x|\theta) = \frac{m!}{x_1!x_2!(m-x_1-x_2)!} \theta_1^{x_1} \theta_2^{x_2} (1-\theta_1-\theta_2)^{m-x_1-x_2}
\]

where \(\theta = (\theta_1, \theta_2)\) are unknown values. Find lower bounds for the variance of unbiased estimators of \(\theta_1\) and \(\theta_2\).

**Solution:** The sample joint p.m.f. is

\[
f_X(x|\theta) = \prod_{i=1}^{n} \frac{m!}{r_{i1}!r_{i2}!r_{i3}!} \theta_1^{r_{i1}} \theta_2^{r_{i2}} (1-\theta_1-\theta_2)^{m-r_{i1}-r_{i2}}
\]

and

\[
\log f_X(x|\theta) \propto \sum_{i=1}^{n} x_{i1} \log \theta_1 + \sum_{i=1}^{n} x_{i2} \log \theta_2 + \\
+ \sum_{i=1}^{n} (m-x_{i1}-x_{i2}) \log (1-\theta_1-\theta_2)
\]

Hence

\[
\frac{\partial}{\partial \theta_1} \log f_X(x|\theta) = \frac{\sum_{i=1}^{n} x_{i1}}{\theta_1} - \frac{\sum_{i=1}^{n} (m-x_{i1}-x_{i2})}{(1-\theta_1-\theta_2)}
\]

\[
\frac{\partial^2}{\partial \theta_1^2} \log f_X(x|\theta) = -\frac{\sum_{i=1}^{n} x_{i1}^2}{\theta_1^2} - \frac{\sum_{i=1}^{n} (m-x_{i1}-x_{i2})^2}{(1-\theta_1-\theta_2)^2}
\]

and since \(E(x_{i1}) = m\theta_1,\ E(x_{i2}) = m\theta_2\)

\[
I_{11}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_1^2} \log f_X(X|\theta) \right) = \frac{nm}{\theta_1} + \frac{nm}{(1-\theta_1-\theta_2)} \quad (2.25)
\]

By symmetry

\[
I_{22}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_2^2} \log f_X(X|\theta) \right) = \frac{nm}{\theta_2} + \frac{nm}{(1-\theta_1-\theta_2)} \quad (2.26)
\]
Finally
\[
\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log f_X(x|\theta) = -\sum_{i=1}^n \frac{(m - x_{i1} - x_{i2})}{(1 - \theta_1 - \theta_2)^2}
\]
and hence
\[
I_{12}(\theta) = -E \left( \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \log f_X(X|\theta) \right) = \frac{nm}{(1 - \theta_1 - \theta_2)} \quad (2.27)
\]
Hence the Fisher Information matrix is
\[
I(\theta) = \begin{bmatrix} \frac{1}{\theta_1} + \frac{1}{(1 - \theta_1 - \theta_2)} & \frac{1}{1} \\ \frac{1}{1} & \frac{1}{\theta_2} + \frac{1}{(1 - \theta_1 - \theta_2)} \end{bmatrix}
\]
and
\[
J(\theta) = [I(\theta)]^{-1} = \frac{\theta_1 \theta_2 (1 - \theta_1 - \theta_2)}{nm} \begin{bmatrix} \frac{1}{\theta_2} + \frac{1}{(1 - \theta_1 - \theta_2)} & -\frac{1}{(1 - \theta_1 - \theta_2)} \\ -\frac{1}{(1 - \theta_1 - \theta_2)} & \frac{1}{\theta_1} + \frac{1}{(1 - \theta_1 - \theta_2)} \end{bmatrix}
\]
Thus if \( \hat{\theta}_1 \) is an unbiased estimator of \( \theta_1 \)
\[
Var(\hat{\theta}_1) \geq J_{11}(\theta) = \frac{\theta_1 \theta_2 (1 - \theta_1 - \theta_2)}{nm} \left( \frac{1}{\theta_2} + \frac{1}{(1 - \theta_1 - \theta_2)} \right) = \frac{\theta_1 (1 - \theta_1)}{nm}
\]
Similarly if \( \hat{\theta}_2 \) is an unbiased estimator of \( \theta_2 \)
\[
Var(\hat{\theta}_2) \geq J_{22}(\theta) = \frac{\theta_2 (1 - \theta_2)}{nm}
\]
Notice that if \( \theta_2 \) was assumed to be known then the CRLB for the variance of \( \hat{\theta}_1 \) would have been
\[
Var(\hat{\theta}_1) \geq \frac{1}{I_{11}(\theta)}
\]
i.e.
\[
Var(\hat{\theta}_1) \geq \frac{1}{nm \left( \frac{1}{\theta_1} + \frac{1}{(1 - \theta_1 - \theta_2)} \right)} = \frac{\theta_1 (1 - \theta_1 - \theta_2)}{nm (1 - \theta_2)}
\]
Since \((1 - \theta_1) (1 - \theta_2) > (1 - \theta_1 - \theta_2)\) for \(0 < \theta_1, \theta_2 < 1\) it follows that

\[
J_{11}(\theta) > \frac{1}{I_{11}(\theta)}
\]

which indicates that if \(\theta_2\) is unknown it would be wrong to use the inequality

\[
\text{Var} \left( \hat{\theta}_1 \right) \geq \frac{1}{I_{11}(\theta)}
\]

since this inequality is not as tight as the correct inequality

\[
\text{Var} \left( \hat{\theta}_1 \right) \geq J_{11}(\theta)
\]

In general it can be shown that

\[
J_{ii}(\theta) \geq \frac{1}{I_{ii}(\theta)}
\]

for all \(i\), with strict inequality (and hence improved lower bound for the variance of an unbiased estimator of \(\theta_i\)) unless the Fisher information matrix \(I(\theta)\) is diagonal (see C.R. Rao (1973) Linear Statistical Inference and its applications).