Estimation Theory 2024-09-09

Estimator을 찾을 때 Bias가 없고 최소 Variance를 가지는 Estimator을 찾는것이 목표.
그 방법 중 하나 : CRLB 0 -> bias가 없고, CRLB 만족(regularity 만족, var 최소) : efficient

CRAMER-RAO LOWER BOUND (CRLB)

Theorem: Cramer-Rao Lower Bound (CRLB)

Theorem: It is assumed that the probability density function $p(x;\theta)$ satisfies the *regularity* condition

$$\mathbb{E}\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right) = 0 \qquad \text{for all } \theta, \tag{1}$$

where the expectation is taken with respect to $p(x;\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)}. \tag{2}$$

Additional Facts about the CRLB

1. An unbiased estimator achieves the bound if and only if
$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = I(\theta)(g(x)-\theta)$$
for some functions I and g.
2. $\hat{\theta} = g(x)$ is the minimum variance unbiased estimator.
3. The minimum variance is $1/I(\theta)$, where

$$I(\theta) = \mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right).$$

Fisher Information

4) Fisher Information

The Cauchy-Schwarz inequality becomes an equality if $X(z) = cY(z)$ for some constant c, not depending on z.

In our case, this is when
$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = \frac{1}{c}(\hat{\alpha}-\alpha)$$
$$\frac{\partial \ln p(x;\theta)}{\partial \theta} = \frac{1}{c(\theta)}(\hat{\theta}-\theta) \qquad \text{for } \alpha = g(\theta) = \theta$$

To find $c(\theta)$:
$$\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{1}{c(\theta)}\right)(\hat{\theta}-\theta) - \frac{1}{c(\theta)}$$
$$-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right) = \frac{1}{c(\theta)}$$

Hence, $c(\theta) = \dfrac{1}{-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)} = \dfrac{1}{I(\theta)}$

$$\Rightarrow \quad I(\theta) = \mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right).$$

1) regularity proof.

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right) &= \int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)dx \\
&= \int \frac{1}{p(x;\theta)}\frac{\partial p(x;\theta)}{\partial \theta} p(x;\theta)dx \\
&= \int \frac{\partial p(x;\theta)}{\partial \theta} dx \\
&= \frac{\partial}{\partial \theta}\int p(x;\theta)dx \\
&= 0
\end{aligned}$$

2) $\mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)$. proof.

$$\mathbb{E}\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right) = \int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)dx$$

$$\frac{\partial}{\partial \theta}\left(\int \frac{\partial \ln p(x;\theta)}{\partial \theta} p(x;\theta)dx = 0\right)$$

$$\int\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}p(x;\theta) + \frac{\partial \ln p(x;\theta)}{\partial \theta}\frac{\partial p(x;\theta)}{\partial \theta}\right)dx = 0$$

$$\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right) + \int \frac{\partial \ln p(x;\theta)}{\partial \theta}\frac{\partial \ln p(x;\theta)}{\partial \theta}p(x;\theta)dx = 0$$

Hence, $\mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)$.

3) $\mathrm{var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)}$. proof.

Let $\alpha = g(\theta)$. Consider all unbiased estimators $\hat{\alpha}$ such that

$$\mathbb{E}(\hat{\alpha}) = \alpha = g(\theta).$$

$$\frac{\partial}{\partial \theta}\left(\int \hat{\alpha}p(x;\theta)dx = g(\theta)\right) \qquad \int \hat{\alpha}\frac{\partial p(x;\theta)}{\partial \theta}dx = \frac{\partial g(\theta)}{\partial \theta}$$

$$\int \hat{\alpha}\frac{\partial \ln p(x;\theta)}{\partial \theta}p(x;\theta)dx = \frac{\partial g(\theta)}{\partial \theta}$$

$$-\quad \alpha\mathbb{E}\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right) = 0 \quad \text{regularity condition}$$

$$\int(\hat{\alpha}-\alpha)\frac{\partial \ln p(x;\theta)}{\partial \theta}p(x;\theta)dx = \frac{\partial g(\theta)}{\partial \theta}$$

We have $\int(\hat{\alpha}-\alpha)\frac{\partial \ln p(x;\theta)}{\partial \theta}p(x;\theta)dx = \frac{\partial g(\theta)}{\partial \theta}$

Cauchy-Schwarz: $\left|\int X(z)Y(z)p(z)dz\right|^2 \leq \left(\int X^2(z)p(z)dz\right)\cdot\left(\int Y^2(z)p(z)dz\right)$

Let $X(z) \leftarrow \hat{\alpha}-\alpha$, $Y(z) \leftarrow \frac{\partial \ln p(x;\theta)}{\partial \theta}$, and $p(z) \leftarrow p(x;\theta)$.

$$\begin{aligned}
\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 &= \left(\int(\hat{\alpha}-\alpha)\frac{\partial \ln p(x;\theta)}{\partial \theta}p(x;\theta)dx\right)^2 \\
&\leq \underbrace{\left(\int(\hat{\alpha}-\alpha)^2 p(x;\theta)dx\right)}_{\mathrm{var}(\hat{\alpha})}\left(\int\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2 p(x;\theta)dx\right)
\end{aligned}$$

Hence,
$$\mathrm{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{\mathbb{E}\left(\left(\frac{\partial \ln p(x;\theta)}{\partial \theta}\right)^2\right)} = \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)}$$

Let $\alpha = g(\theta) = \theta$, we get the desired result.

Ex 1) $x[n] = A + w[n] \qquad w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2)$

Estimator: $\tilde{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$

$$\mathbb{E}(\tilde{A}) = A \quad \text{unbiased}$$
$$\mathrm{var}(\tilde{A}) = \frac{1}{N}\sigma^2$$

1) regularity

Likelihood of $x := (x[0],\ldots,x[N-1])$

$$\begin{aligned}
p(x;A) &= \prod_{n=0}^{N-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x[n]-A)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right)
\end{aligned}$$

$$\frac{\partial \ln p(x;A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) = \frac{N}{\sigma^2}(\bar{x}-A) \qquad \bar{x} = \frac{1}{N}\sum_{n=0}^{N-1}x[n]$$

Regularity condition: $\mathbb{E}\left(\frac{\partial \ln p(x;A)}{\partial A}\right) = 0$

2) variance

Second derivative: $\frac{\partial^2 \ln p(x;A)}{\partial A^2} = -\frac{N}{\sigma^2}$

CRLB: $\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{N}$ for any unbiased estimator $\hat{A}$

Since $\mathrm{var}(\tilde{A}) = \frac{\sigma^2}{N}$, $\tilde{A}$ is the minumum variance unbiased estimator.

If an unbiased estimator achieves the CRLB, it is said to be *efficient*.

Ex 2)

Consider $x[n] = s[n;\theta] + w[n]$, for $n = 0,1,\ldots,N-1$.

Likelihood of $\mathbf{x} := (x[0],\ldots,x[N-1])$ is

$$\begin{aligned}
p(\mathbf{x};\theta) &= \frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n;\theta])^2\right) \\
\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-s[n;\theta])\frac{\partial s[n;\theta]}{\partial \theta} \\
\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left((x[n]-s[n;\theta])\frac{\partial^2 s[n;\theta]}{\partial \theta^2} - \left(\frac{\partial s[n;\theta]}{\partial \theta}\right)^2\right) \\
\mathbb{E}\left(\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right) &= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial \theta}\right)^2 \\
\mathrm{var}(\hat{\theta}) &\geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial \theta}\right)^2}
\end{aligned}$$

Special case: $s[n;\theta] = \theta$.
Then CRLB $= \frac{\sigma^2}{N}$.

Estimation Theory 2024−09−09

Transformation of Parameters

Let $x[n] = A + w[n]$, where $w[n] \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

1. Nonlinear transformation does not preserve efficiency

$$\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n], \qquad \bar{x} \sim \mathcal{N}\left(A, \frac{\sigma^2}{N}\right)$$

Is $\bar{x}^2$ efficient for A^2?

$$\begin{aligned}\mathbb{E}(\bar{x}^2) &= \mathbb{E}^2(\bar{x}) + \mathbf{var}(\bar{x}) = A^2 + \frac{\sigma^2}{N} \\ &\neq A^2.\end{aligned}$$

Biased => not efficient

2. Linear transform preserves efficiency

Assume: $\hat{\theta}$ efficient for θ, $g(\theta) = a\theta + b$ (affine transform)

$$\widehat{g(\theta)} = g(\hat{\theta}) = a\hat{\theta} + b$$

$$\begin{aligned}E(a\hat{\theta} + b) &= aE(\hat{\theta}) + b = a\theta + b \\ &= g(\theta)\end{aligned}$$

$$\mathbf{var}(\widehat{g(\theta)}) = \mathbf{var}(a\hat{\theta} + b) = a^2\mathbf{var}(\hat{\theta})$$

CRLB for $g(\theta)$:

$$\begin{aligned}\mathbf{var}(\widehat{g(\theta)}) &\geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)} \\ &= \left(\frac{\partial g(\theta)}{\partial \theta}\right)^2 \mathbf{var}(\hat{\theta}) \\ &= a^2\mathbf{var}(\hat{\theta}).\end{aligned}$$

Efficient

3. Efficiency is approximately maintained over a nonlinear transformation for a **large dataset**.

proof 1 : CRLB for $\alpha = g(\theta)$:

$$\mathrm{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{-\mathbb{E}\left(\frac{\partial^2 \ln p(x;\theta)}{\partial \theta^2}\right)}$$

Estimation of A^2 ($\alpha = g(A) = A^2$)

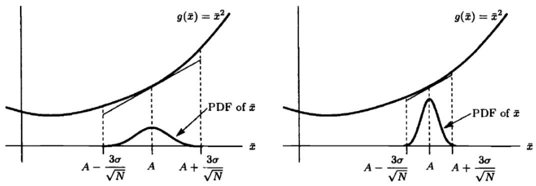
$$\mathrm{var}(\hat{A}^2) \geq \frac{(2A)^2}{N/\sigma^2} = \frac{4A^2\sigma^2}{N}$$

($\frac{\partial^2 \ln p(x;A)}{\partial A^2} = -\frac{N}{\sigma^2}$ at Ex 1)

$$\begin{aligned}\mathbf{var}(\bar{x}^2) &= E(\bar{x}^4) - E^2(\bar{x}^2) \\ &= \frac{4A^2\sigma^2}{N} + \frac{2\sigma^4}{N^2}\end{aligned}$$ as $N \to \infty$, $\mathrm{var}(\bar{x}^2) \to \frac{4A^2\sigma^2}{N}$.

Asymptotically Efficient

proof 2 :



(a) Small N (b) Large N

N이 커지면 PDF가 sharp해지고,
g(x) 대신 Linearized g를 근사적으로 대신 사용 가능

Linearize g about A:

$$g(\bar{x}) \approx g(A) + \frac{dg(A)}{dA}(\bar{x} - A)$$

$$E[g(\bar{x})] = g(A) = A^2$$

$$\begin{aligned}\mathrm{var}[g(\bar{x})] &= \left[\frac{dg(A)}{dA}\right]^2 \mathrm{var}(\bar{x}) \\ &= \frac{(2A)^2\sigma^2}{N} \\ &= \frac{4A^2\sigma^2}{N}\end{aligned}$$

Asymptotically Efficient