Statistical Inference

Lecture 05a

ANU - RSFAS

Last Updated: Sat Mar 18 17:25:25 2017

Evaluating Estimators

- We now have a few approaches to obtain point estimators.
- Now we must decide which one we ought to choose and for what reason.
- We have hinted at the idea of unbiasedness.

Definition 1: The bias of a point estimator T of a parameter θ is the difference between the expected value of T and θ.

$$\mathrm{Bias}_\theta = E[T] - \theta$$

- For $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ where $E[X] = \mu,\ V(X) = \sigma^2$.

$$E[\bar{X}] = \mu, \quad E[S^2] = \sigma^2$$

- $\mathrm{Bias}(\bar{X}) = E[\bar{X}] - \mu = 0$. We say that the estimator is unbiased.

Evaluating Estimators

Eg. Let's consider $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}\,(\mu, \sigma^2)$.

- We know the $E[S^2] = \sigma^2$. So it is unbiased.
- I actually think I like the estimator:

$$\frac{1}{n+1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

What do you think?

Recall: Definition 2: The mean squared error (MSE) of an estimator $T \equiv T(\boldsymbol{X})$ of a parameter θ is the function

$$E_\theta(T - \theta)^2$$

Evaluating Estimators

Eg. Let's reconsider our question for $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{normal}\ (\mu, \sigma^2)$.

- The MoM & ML estimator for σ^2 was $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=i}^{n} (X_i - \bar{X})^2$
- The unbiased estimator is $S^2 = \frac{1}{n-1} \sum_{i=i}^{n} (X_i - \bar{X})^2$

$$\mathrm{MSE}(S^2) = \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

$$\mathrm{MSE}(\hat{\sigma}^2) = \frac{(2n-1)\sigma^4}{n^2}$$

$$\mathrm{MSE}(S^2) > \mathrm{MSE}(\hat{\sigma}^2)$$

Evaluating Estimators

- Let's now consider $\tilde{\sigma}^2 = \frac{1}{n+1}\sum_{i=1}^{n}(x_i - \bar{x})^2$:
- First let's derive the variance of S^2:

$$
\begin{aligned}
(n-1)S^2/\sigma^2 \sim \chi^2_{n-1} \quad &\Rightarrow \quad V\left((n-1)S^2/\sigma^2\right) = 2(n-1) \\
&\qquad \left[(n-1)/\sigma^2\right]^2 V\left(S^2\right) = 2(n-1) \\
&= \quad V\left(S^2\right) = \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{(n-1)}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\sigma}^2 &= \frac{1}{n+1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \frac{n-1}{n+1}S^2
\end{aligned}
$$

Evaluating Estimators

$$E[\tilde{\sigma}^2] = E\left[\frac{n-1}{n+1}S^2\right] = \frac{n-1}{n+1}E\left[S^2\right] = \frac{n-1}{n+1}\sigma^2$$

$$V[\tilde{\sigma}^2] = V\left[\frac{n-1}{n+1}S^2\right] = \left[\frac{n-1}{n+1}\right]^2 V\left[S^2\right] = \left[\frac{n-1}{n+1}\right]^2 \frac{2\sigma^4}{(n-1)}$$

$$MSE[\tilde{\sigma}^2] \quad = \quad \frac{2(n-1)\sigma^4}{(n+1)^2} + \left[\frac{n-1}{n+1}\sigma^2 - \sigma^2\right]^2 = \frac{2\sigma^4}{n+1}$$

Evaluating Estimators

- Now let's compare to the MLE:

$$\begin{aligned} \frac{2\sigma^4}{n+1} &\overset{?}{<} \frac{(2n-1)\sigma^4}{n^2} \\ \frac{2}{n+1} - \frac{(2n-1)\sigma^4}{n^2} &\overset{?}{<} 0 \\ \frac{(1-n)\sigma^4}{n^2(n+1)} &< 0 \quad \text{for n } \geq 1 \end{aligned}$$

$$MSE(\tilde{\sigma}^2) < MSE(\hat{\sigma}^2) < MSE(S^2)$$

Evaluating Estimators

Eg.: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$.

- The MLE is $\hat{p} = \bar{X}$. The bias: $Bias(\hat{p}) = E[\hat{p}] - p = p - p = 0$.
- The variance of $\hat{p} = V(\bar{X}) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$.

$$MSE(\hat{p}) = \frac{p(1-p)}{n} + 0^2 = \frac{p(1-p)}{n}$$

Evaluating Estimators

- Let's consider a Bayesian estimator when $p \sim beta(a, b)$.
- We found $[p|\boldsymbol{x}] \sim beta(y + a, n - y + b)$. Where $Y = \sum_{i=1}^{n} X_i$.
- If we use the mean of the posterior distribution we have the following Bayesian estimator:

$$\hat{p}_B = \frac{y + a}{a + b + n}$$

- Now we will again let Y be random to compare the MSE of the Bayesian and ML estimators.

$$\begin{aligned}
E\left[\frac{Y + a}{a + b + n}\right] &= \frac{E[Y] + a}{a + b + n} = \frac{np + a}{a + b + n} \\
V\left[\frac{Y + a}{a + b + n}\right] &= \left[\frac{1}{a + b + n}\right]^2 V(Y) = \left[\frac{1}{a + b + n}\right]^2 np(1 - p)
\end{aligned}$$

Evaluating Estimators

$$MSE[\hat{p}_B] = MSE\left[\frac{Y+a}{a+b+n}\right] = \left[\frac{np(1-p)}{(a+b+n)^2}\right] + \left[\frac{np+a}{a+b+n} - p\right]^2$$

- To compare with the MLE we need specific values for a and b. Let's pick these to make the $MSE[\hat{p}_B]$ constant for p. You of course can try other values!
- It turns out that if we set $a = b = \sqrt{n/4}$ then we get (which is constant for p):

$$\hat{p}_B = \frac{Y + \sqrt{n/4}}{n + \sqrt{n}}$$

$$MSE[\hat{p}_B] = \frac{1}{(4(1+\sqrt{n})^2)}$$

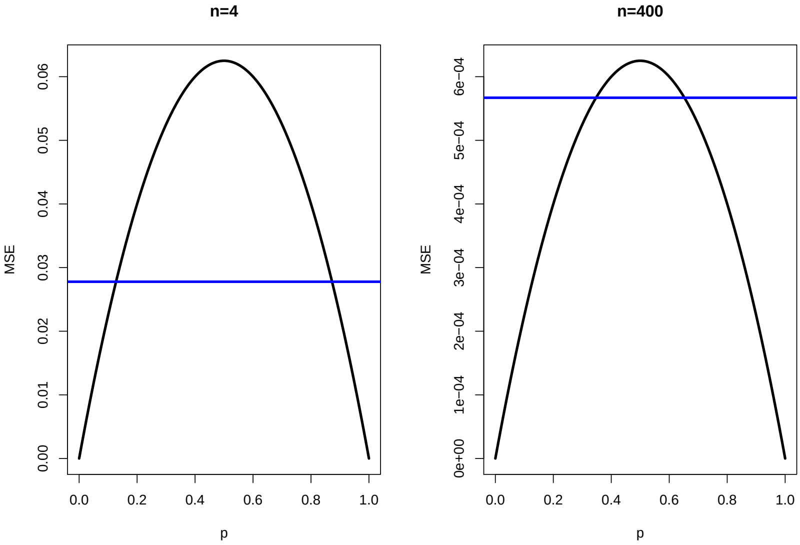
Evaluating Estimators

```r
p <- seq(0,1, by=0.001)

par(mfrow=c(1,2))
n <- 4
MSE.mle <- p*(1-p)/n
plot(p, MSE.mle, type="l", lwd=3, main="n=4", ylab="MSE")
abline(h = 1/(4*(sqrt(n) + 1)^2), lwd=3, col="blue")

n <- 400
MSE.mle <- p*(1-p)/n
plot(p, MSE.mle, type="l", lwd=3, main="n=400", ylab="MSE")
abline(h = 1/(4*(sqrt(n) + 1)^2), lwd=3, col="blue")
```

Evaluating Estimators



Evaluating Estimators

- For small n the Bayesian estimator is the better choice, unless we believe p is near 0 or 1.
- For large n, then the MLE is the better choice, unless we believe p is near $1/2$.

Of course this is based on the way we set our prior.

Evaluating Estimators

Eg.: An estimator T^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E[T^*] = \tau(\theta)$ for all θ and, for any other estimator T with $E[T] = \tau(\theta)$ we have

$$V(T^*) \leq V(T) \;\; \forall \, \theta.$$

T^* is also called a uniform minimum variance unbiased estimator (UMVUE) for $\tau(\theta)$.

- Finding a best unbiased estimator can be difficult.

Evaluating Estimators

Eg.: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$.

- $E[\bar{X}] = \lambda$
- $E[S^2] = \lambda$
- Through a lot of work we can show:

$$V(\bar{X}) < V(S^2)$$

- But we could also consider (infinite number of choices for $a \in [0, 1]$):

$$T(\bar{X}, S^2) = a\bar{X} + (1 - a)S^2$$

- Are there other unbiased estimators?
- We would like to know if we have unbaised estimator whether its variance is the smallest possible!

Evaluating Estimators

Rice 8.7 Theorem A (Cramer-Rao Inequality [lower bound]): Let $X_1, \ldots, X_n$ be a random sample from a distribution family with density function $f_X(x;\theta)$ where θ is a scalar parameter. Also, let $T = t(X_1, \ldots, X_n)$ be an unbiased estimator for $\tau(\theta)$. Then under certain regularity (smoothness) conditions:

$$Var_\theta(T) \geq \frac{\{\tau'(\theta)\}^2}{ni(\theta)} = \{\tau'(\theta)\}^2 I(\theta)^{-1}$$

- $\tau'(\theta) = \frac{d}{d\theta}\tau(\theta)$
- Where $I(\theta) = ni(\theta)$ is called the Expected Fisher Information.
- $I(\theta) = E\left[\left(\frac{\partial l(\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial\theta^2}\right]$

Evaluating Estimators

Theorem A Extended: Let $X_1, \ldots, X_n$ be a sample [note we don't have to have iid] with pdf $f(\boldsymbol{x}|\theta)$ and let $T(\boldsymbol{X})$ be an estimator [doesn't have to be unbiased] then based on regularity conditions we have:

$$V\left[T(\boldsymbol{X})\right] \geq \frac{\left[\frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]\right]^2}{E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right]} = \frac{\left[\frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]\right]^2}{I(\theta)}$$

Cramer-Rao Inequality [lower bound]

- If $E[T(\boldsymbol{X})] = \tau(\theta)$, so $T(\boldsymbol{X})$ is an unbiased estimator for $\tau(\theta)$,

$$V\left[T(\boldsymbol{X})\right] \geq \frac{\left[\tau'(\theta)\right]^2}{I(\theta)}$$

- If we have iid samples:

$$V\left[T(\boldsymbol{X})\right] \geq \frac{\left[\tau'(\theta)\right]^2}{ni(\theta)}$$

Cramer-Rao Inequality [lower bound]

Eg: Poisson distribution: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$.

$$f(x|\lambda) = \frac{\lambda^x exp(-\lambda)}{x!}$$

- $\tau(\lambda) = \lambda \Rightarrow \frac{d}{d\theta}\tau(\lambda) = 1$.

Cramer-Rao Inequality [lower bound]

Eg: Poisson distribution.

$$
\begin{aligned}
i(\lambda) &= E\left[\left(\frac{d}{d\lambda} ln\{f(x|\lambda)\}\right)^2\right] \\
&= E\left[\left(\frac{d}{d\lambda}\{xln(\lambda) - \lambda - x!\ \}\right)^2\right] \\
&= E\left[\left(\frac{X}{\lambda} - 1\right)^2\right] \\
&= \frac{1}{\lambda^2} E\left[(X-\lambda)^2\right] \\
&= \frac{1}{\lambda^2} Var(X) \\
&= \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda}
\end{aligned}
$$

Cramer-Rao Inequality [lower bound]

$$V_{\lambda}(T) \geq \frac{1}{n\lambda^{-1}} = \frac{\lambda}{n}$$

We will show that a sufficient statistic for λ to be $\sum_{i=1}^{n} X_i$ so $\bar{X}$ is also a sufficient statistic.

$$V(\bar{X}) = \frac{\lambda}{n}$$

We see that the lower bound is achieved by $\bar{X}$ thus it is the Uniform Minimum Variance Unbiased Estimator (UMVUE) of λ.

Cramer-Rao Inequaility - Regularity Conditions

- $\frac{\partial}{\partial\theta} ln\{f(x|\theta)\}$ exists for all x and θ;
- interchange of integration and differentiation is permissible;
- The expectation $i(\theta) = E\left[\left[\frac{\partial}{\partial\theta} ln\{f(x|\theta)\}\right]^2\right]$, where X is a generic random variable having distribution with density $f(x|\theta)$, is finite for all $\theta \in \Theta$.

Cramer-Rao Inequaility - Proof

- The proof is based on the Cauchy-Schwarz Inequality: For any two random variables Y and Z:

$$[Cov(Y,Z)]^2 \leq V(Y)V(Z) \;\Rightarrow\; V(Y) \geq \frac{[Cov(Y,Z)]^2}{V(Z)}$$

- We choose Y in this equation to be $T(\boldsymbol{X})$ and Z to be $\frac{\partial}{\partial\theta}$ $log\ f(\boldsymbol{x}|\theta)$.
- Recall:

$$V(Z) = E[Z^2] - (E[Z])^2$$
$$Cov(Y,Z) = E[YZ] - E[Y]\ E[Z]$$

Cramer-Rao Inequaility - Proof

- First note:

$$
\begin{aligned}
\frac{\partial}{\partial\theta} E[T(\boldsymbol{X})] &= \frac{\partial}{\partial\theta} \int_{\mathcal{X}} T(\boldsymbol{x}) f(\boldsymbol{x}|\theta) dx \\
&= \int_{\mathcal{X}} T(\boldsymbol{x}) \left[\frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta) \right] dx \\
&= \int_{\mathcal{X}} T(\boldsymbol{x}) \left[\frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta) \frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} \right] dx \\
&= E\left[T(\boldsymbol{X}) \frac{\frac{\partial}{\partial\theta} f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} \right] \\
&= E\left[T(\boldsymbol{X}) \ \frac{\partial}{\partial\theta} \ log \ f(\boldsymbol{x}|\theta) \right]
\end{aligned}
$$

Cramer-Rao Inequaility - Proof

- Second note:

$$
\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] &= \int_{\mathcal{X}}\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] f(\boldsymbol{x}|\theta)dx \\
&= \int_{\mathcal{X}}\left[\frac{\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)}\right] f(\boldsymbol{x}|\theta)dx \\
&= \int_{\mathcal{X}}\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)dx \\
&= \frac{\partial}{\partial\theta}\int_{\mathcal{X}} f(\boldsymbol{x}|\theta)dx \\
&= \frac{\partial}{\partial\theta}(1) = 0
\end{aligned}
$$

Cramer-Rao Inequaility - Proof

$$\begin{aligned}
Cov\left[T(\boldsymbol{X}),\ \frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] &= E\left[T(\boldsymbol{X})\ \frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] \\
&\quad -E\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]E\left[T(\boldsymbol{X})\right] \\
&= E\left[T(\boldsymbol{X})\ \frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] - [0]E\left[T(\boldsymbol{X})\right] \\
&= E\left[T(\boldsymbol{X})\ \frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] \\
&= \frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]
\end{aligned}$$

Cramer-Rao Inequaility - Proof

$$
\begin{aligned}
V\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] &= E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right] - \left(E\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]\right)^2 \\
&= E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right] - (0)^2 \\
&= E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right]
\end{aligned}
$$

Cramer-Rao Inequaility - Proof

$$
\begin{aligned}
V(Y) &\geq \frac{[Cov(Y, Z)]^2}{V(Z)} \\
V\left[T(\boldsymbol{X})\right] &\geq \frac{\left[Cov\left[T(\boldsymbol{X}),\ \frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]\right]^2}{V\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]} \\
&\geq \frac{\left[\frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]\right]^2}{E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right]}
\end{aligned}
$$

Cramer-Rao Inequaility

- So we have:

$$V\left[T(\boldsymbol{X})\right] \geq \frac{\left[\frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]\right]^2}{E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right]}$$

Corollarly (iid case): If the regularity conditions hold and $T(\boldsymbol{X})$ is an unbiased estimator for $\tau(\theta)$ and we have $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x|\theta)$, then

$$V\left[T(\boldsymbol{X})\right] \geq \frac{\left[\frac{\partial}{\partial\theta}E[T(\boldsymbol{X})]\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x|\theta)\right)^2\right]} = \frac{[\tau'(\theta)]^2}{ni(\theta)} = \{\tau'(\theta)\}^2 I(\theta)^{-1}$$

Cramer-Rao Inequaility - Proof

- We need to show

$$E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x|\theta)\right)^2\right]$$

Cramer-Rao Inequaility - Proof

$$
\begin{aligned}
E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right] &= E\left[\left(\frac{\partial}{\partial\theta}\ log\prod_{i=1}^{n} f(x_i|\theta)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\theta}\ \sum_{i=1}^{n} log\ f(x_i|\theta)\right)^2\right] \\
&= E\left[\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)^2\right] \\
&= \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)^2\right] \\
&\quad + \sum_{i\neq j} E\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\ \frac{\partial}{\partial\theta}\ log\ f(x_j|\theta)\right)
\end{aligned}
$$

Cramer-Rao Inequaility - Proof

$$
\begin{aligned}
&= \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)^2\right] \\
&\quad + \sum_{i\neq j} E\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\ \frac{\partial}{\partial\theta}\ log\ f(x_j|\theta)\right) \\
&= \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)^2\right] \\
&\quad + \sum_{i\neq j} E\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)\ E\left(\frac{\partial}{\partial\theta}\ log\ f(x_j|\theta)\right) \\
&= \sum_{i=1}^{n} E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x_i|\theta)\right)^2\right] + [0][0] \\
&= nE\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x|\theta)\right)^2\right] = ni(\theta)
\end{aligned}
$$

Fisher Information

- The Fisher information, or expected Fisher information, or the information number is:

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right)^2\right] = -E\left[\left(\frac{\partial^2}{\partial\theta^2}\ log\ f(\boldsymbol{x}|\theta)\right)\right]$$

- For one data point we have:

$$i(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ log\ f(x|\theta)\right)^2\right]$$

- For iid data:

$$ni(\theta) = I(\theta)$$

Fisher Information - Proof

- First note:

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\ log\ f(\boldsymbol{x}|\theta) &= \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta) \\
&= \frac{\partial}{\partial\theta}\frac{\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} \\
&= \frac{\left[\frac{\partial^2}{\partial\theta^2}f(\boldsymbol{x}|\theta)\right]f(\boldsymbol{x}|\theta) - \left[\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)\right]\left[\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)\right]}{[f(\boldsymbol{x}|\theta)]^2} \\
&= \frac{\frac{\partial^2}{\partial\theta^2}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} - \left[\frac{\frac{\partial}{\partial\theta}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)}\right]^2 \\
&= \frac{\frac{\partial^2}{\partial\theta^2}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} - \left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2
\end{aligned}
$$

Fisher Information - Proof

- Now let's take expectations:

$$
\begin{aligned}
E\left[\frac{\partial^2}{\partial\theta^2}\ log\ f(\boldsymbol{x}|\theta)\right] &= E\left[\frac{\frac{\partial^2}{\partial\theta^2}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)}\right] - E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] \\
&= -E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] + \int_{\mathcal{X}} \frac{\frac{\partial^2}{\partial\theta^2}f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x}|\theta)} f(\boldsymbol{x}|\theta)dx \\
&= -E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] + \int_{\mathcal{X}} \frac{\partial^2}{\partial\theta^2} f(\boldsymbol{x}|\theta)dx \\
&= -E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] + \frac{\partial^2}{\partial\theta^2}\int_{\mathcal{X}} f(\boldsymbol{x}|\theta)dx \\
&= -E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] + \frac{\partial^2}{\partial\theta^2}[1] \\
&= -E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] + 0
\end{aligned}
$$

Fisher Information - Proof

- So we have:

$$E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] \ = \ -E\left[\frac{\partial^2}{\partial\theta^2}\ log\ f(\boldsymbol{x}|\theta)\right]$$

- Note: We already showed $E\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] = 0$. We also showed the following (but just to make it clear)

$$\begin{aligned} V\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right] &= E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] - \left(E\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]\right)^2 \\ &= E\left[\left[\frac{\partial}{\partial\theta}\ log\ f(\boldsymbol{x}|\theta)\right]^2\right] \end{aligned}$$