6.1. $X_i \sim$ uniform on $[0,\theta]$ $\quad i=0,1$

$\hat{\theta} = \max(X_0, X_1)$

$$P[\hat{\theta} \le a] = P[\max(X_0,X_1) \le a] = P[X_0 \le a]\,P[X_1 \le a] = \begin{cases} 0, & a<0 \\ (\frac{a}{\theta})^2, & 0 \le a < \theta \\ 1, & \text{o.w.} \end{cases}$$

$\hat{\theta}$ has density

$$f_{\hat{\theta}}(a) = \begin{cases} 0, & a \notin [0,\theta] \\ \frac{2a}{\theta^2}, & a \in [0,\theta] \end{cases}$$

$$E\hat{\theta} = \int_0^\theta \frac{2a^2}{\theta^2}\,da = \frac{2}{3}\theta \left(\frac{a}{\theta}\right)^3\Big|_0^\theta = \frac{2}{3}\theta$$

$\hat{\theta}$ is biased

$$E\hat{\theta}^2 = \int_0^\theta \frac{2}{\theta^2}a^3\,da = \frac{1}{2}\theta^2$$

$$\operatorname{Var}\hat{\theta} = \frac{1}{2}\theta^2 - \frac{4}{9}\theta^2 = \frac{\theta^2}{18}$$

6.2. $X_i \sim$ uniform on $[0, \theta]$

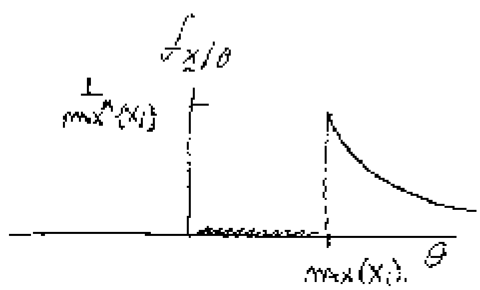
(a) Show $\hat{\theta}_{ML} = \max\{X_i\}_{i=1}^{n}$

$$f(\underline{x}/\theta) = \frac{1}{\theta^n} 1_{[0,\infty]}(\min(x_i)) \; 1_{[-\infty,\theta]}(\max(x_i))$$

$f(\underline{x}/\theta)$ decreases in θ, for $\theta \geq \max(X_i)$.

$$f(\underline{x} / \max\{x_i\}) = \frac{1}{\max^n\{x_i\}}$$

$f(\underline{x}/\theta) = 0$ for $\theta < \max(X_i)$.



(b) $$P[\hat{\theta}_{ML} \leq a] = \prod_{i=1}^{n} P[X_i \leq a] = \begin{cases} 0 & , \; a < 0 \\ \left(\frac{a}{\theta}\right)^n & , \; 0 < a \leq \theta \\ 1 & , \; \theta < a \end{cases}$$

so

$$f_{\hat{\theta}_{ML}}(a) = \begin{cases} \frac{n}{\theta^n} a^{n-1} & , \; 0 \leq a \leq \theta \\ 0 & , \; \text{o.w.} \end{cases}$$

(c) $E\,\hat{\theta}_{ML} = \int_0^{\theta} \frac{M}{\theta^M} a^M \, da = \frac{M}{M+1}\,\theta$

(d) $E\,\hat{\theta}^2_{ML} = \int_0^{\theta} \frac{M}{\theta^M} a^{M+1} \, da = \frac{M}{M+2}\,\theta^2$

$$\begin{aligned} \operatorname{Var} \hat{\theta}_{ML} &= \theta^2 \left[\frac{M}{M+2} - \frac{M^2}{(M+1)^2} \right] \\ &= \theta^2 \; \frac{M(M+1)^2 - M^2(M+2)}{(M+2)(M+1)^2} \\ &= \theta^2 \; \frac{M}{(M+2)(M+1)^2} \end{aligned}$$

6.7 Let $\underline{X}$, denote an observation and suppose that $\underline{y}$ denotes "extra" data.

Let $f_{\underline{X}|\underline{\theta}}(\underline{X}|\underline{\theta})$ denote the likelihood function of $\underline{\theta}$ given $\underline{X}$.

Suppose that $\hat{\underline{\theta}}(\underline{X})$ is unbiased for $\underline{\theta}$.
Consider the estimator $\tilde{\underline{\theta}}(\underline{X},\underline{Y}) = E\left[\hat{\underline{\theta}} \,/\, \underline{X},\underline{Y}\right]$

1) $\tilde{\underline{\theta}}$ is unbiased for $\underline{\theta}$
2) $\operatorname{Var} \tilde{\underline{\theta}} \le \operatorname{Var} \hat{\underline{\theta}}$

These facts follow from the fact that $\underline{X}, \underline{Y}$ is sufficient for $\underline{\theta}$, and the Rao-Blackwell Thm.

6.10 Let $\underline{t} = \underline{t}(\underline{x})$. Show $J_{\underline{x}}(\underline{\theta}) - J_{\underline{t}}(\underline{\theta})$ is PSD.

Let $\underline{x}$ be a discrete or continuous random vector

1.) $p_\theta(\underline{t}, \underline{x}) = p_{\underline{\theta}}(\underline{x}|\underline{t})\, p_{\underline{\theta}}(\underline{t})$.

2.) $p_\theta(\underline{t}, \underline{x}) = p_{\underline{\theta}}(\underline{t}|\underline{x}) \cdot p_\theta(\underline{x}) = p_{\underline{\theta}}(\underline{x})\, \delta(\underline{t} = \underline{t}(\underline{x}))$.

3.) $J_{\underline{x},\underline{t}}(\underline{\theta}) = J_{\underline{x}}(\underline{\theta})$.

proof:

$$J_{\underline{x},\underline{t}}(\underline{\theta}) = -E \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log p_\theta(\underline{x}, \underline{t})\right]^T$$

$$\overset{2}{=} -E \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log p_{\underline{\theta}}(\underline{x}) - \frac{\partial}{\partial \underline{\theta}} \log \delta(\underline{t} = \underline{t}(\underline{x}))\right]^T$$

$$= -E \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log p_{\underline{\theta}}(\underline{x})\right]$$

$$= J_{\underline{x}}(\theta).$$

4.) $J_{\underline{x},\underline{t}}(\underline{\theta}) - J_{\underline{t}}(\underline{\theta})$ is PSD.

proof:

$$J_{\underline{x},\underline{t}}(\underline{\theta}) - J_{\underline{t}}(\underline{\theta}) = -E \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log \frac{p_\theta(\underline{t}, \underline{x})}{p_\theta(\underline{t})}\right]^T$$

$$\overset{1}{=} -E \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log p_\theta(\underline{x}|\underline{t})\right]^T$$

$$= E_{\underline{t}} - E_{\underline{x}|\underline{t}} \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial}{\partial \underline{\theta}} \log p_\theta(\underline{x}|\underline{t})\right]^T$$

$$= E_{\underline{t}} \left[J_{\underline{x}|\underline{t}}(\underline{\theta}) \right]$$

Since the Fisher information matrix $J_{\underline{x}|\underline{t}}(\underline{\theta})$ is always PSD, so is its average.

5.) when $\underline{t}$ is sufficient, $\frac{\partial}{\partial \underline{\theta}} \log p_\theta(\underline{x}|\underline{t}) = \underline{0}$ in 4.)