Chapter 4: Estimation and Information (JAW Ch 3)

4.1 CRLB: one-dimensional real parameter (JAW 3.7-10; TPE P. 115-)

(i) (a) $X \sim P_\theta$ **on** $(\mathcal{X}, \mathcal{A})$, $\theta \in \Theta \subset \Re$.
(b) Density $f_\theta \equiv \frac{dP_\theta}{d\mu}$ $\exists$ **where** μ **is** σ**-finite on** $\mathcal{X}$.
(c) $T \equiv T(X)$ **estimates** $q(\theta)$; $\mathrm{E}_\theta|T(X)| < \infty$
(d) $b(\theta) \equiv \mathrm{E}_\theta(T) - q(\theta) \equiv$ **bias of** T
(e) $q'(\theta)$ $\exists$

(ii) Suppose: (a) Θ **is an open subset of** $\Re$
(b) $\exists B,\ \mu(B) = 0$ **s.t. for** $x \notin B$ $\frac{\partial f_\theta(x)}{\partial\theta}$ $\exists$ $\forall\theta$
(c) $A \equiv \{x : f_\theta(x) = 0\}$ **does not depend on** θ
(d) $I(\theta) \equiv \mathrm{E}_\theta((\ell'_\theta(X))^2) > 0$ **where** $\ell'_\theta(x) \equiv \frac{\partial}{\partial\theta}\log f_\theta(x)$ **is the** ***Score function*** **for** θ.
$I(\theta)$ **is the** ***Fisher Information*** **for** θ.
(e) $\int f_\theta(x)d\mu(x)$ **and** $\int T(x)f_\theta(x)d\mu(x)$ **can both be differentiated w.r.t.** θ **under the integral sign.**

(iii) Then, if (ii), $\mathrm{var}_\theta(T(X)) \geq (q'(\theta) + b'(\theta))^2/I(\theta) \quad \forall\theta \in \Theta$ **and equality holds** $\forall\theta$ **iff** $\exists k(\theta)$ **s.t.**
$\ell'_\theta(X) = k(\theta)(T(X) - q(\theta) - b(\theta))$ **a.e.** (μ).

(iv) Proof: $\mathrm{E}_\theta(\ell'_\theta(X)) = 0$ **so** $I(\theta) = \mathrm{E}_\theta((\ell'_\theta(X))^2) = \mathrm{var}(\ell'_\theta(X))$
$(q'(\theta) + b'(\theta)) = \mathrm{Cov}(T(X), \ell'_\theta(X))$ **and result follows from Cauchy-Schwarz, with equality iff** $\ell'_\theta(X) = k(\theta)(T(X)+c(\theta))$; **taking expectations gives** $c(\theta) = -\mathrm{E}_\theta(T) = -q(\theta) - b(\theta)$.

(v) If also $\int f_\theta(x)d\mu(x)$ **can be differentiated twice under the integral**

$$I(\theta) = -\mathrm{E}(\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)) = -\mathrm{E}(\ell''_\theta(X))$$

Prf: Differentiating again, gives $\mathrm{E}_\theta(\ell''_\theta(X)) + \mathrm{E}((\ell'_\theta(X))^2) = 0$

4.2 Assumption verification, and notes

(i) Assumption (ii)(e) can be the hard one to check. **It holds for exponential families: see 2.? and TPE P.27.** **More generally: using** $'$ **to denote** $\frac{\partial}{\partial\theta}$**, if** $X'(\omega,\theta)$ $\exists\forall\theta$ **a.e.(**μ**) and** $|X'(\omega,\theta)| \leq Y(\omega)$ $\forall\theta$**, and** Y **integrable, then DCT will give that** $(\int_\Omega X(\omega,\theta)d\mu)' = \int_\Omega X'(\omega,\theta)d\mu$**.**

(ii) If $b(\theta)=0$ **and** $\mathrm{var}_\theta(T) = (q'(\theta))^2/I(\theta)$**,** T **is MVUE of** $q(\theta)$**, and** $\ell'_\theta(X) = k(\theta)(T(X)-q(\theta))$**. And conversely, if** $\ell'_\theta(X) = \ldots$ **etc.**

(iii) T **is MVUE of** $q(\theta)$ **iff** $aT+b$ **is MVUE of** $aq(\theta)+b$ **but if** π **non-linear,** $\nexists$ **unbiased estimator achieving CRLB for** $\pi(q(\theta))$**.**

(iv) T **is MVUE of** $q(\theta)$ $\Rightarrow$ T **is MLE of** $q(\theta)$**.**

(v) T **a MVUE of** $q(\theta) \Rightarrow \mathrm{var}(T) = |q'(\theta)/k(\theta)|$ **since** $\mathrm{var}(T) = q'(\theta)^2/\mathrm{E}(\ell'_\theta(X))^2 = q'(\theta)^2/(k(\theta)^2(\mathrm{var}_\theta(T))$

(vi) Example: TPE P. 118. JAW 3.8-9
X_i **i.i.d. Poisson mean** $\theta > 0$**.**

(a) Conditions (a)-(d) are trivial. For (e) $\mathrm{E}(T(X^{(n)})) = \Sigma_{x_1,\ldots,x_n} t(x^{(n)})\, e^{-n\theta}\theta^{\Sigma x_i} \;/\; (\Pi\, x_i!)$**, which is absolutely cgt power series in** θ **if** $\mathrm{E}(|T(X^{(n)}|) < \infty$**, so can diffte term-by-term.**

(b) $\ell_\theta(X^{(n)}) = n(\overline{X_n}-\theta)/\theta$**.** $\overline{X_n}$ **attains lower bound for** $q(\theta)=\theta$**, and** $\mathrm{var}(\overline{X_n}) = \theta/n =$ **CRLB.**

(c) For, $q(\theta)=\theta^2$**, CRLB** $= 4\theta^3/n$
$\mathrm{E}(\overline{X_n}^2) = \theta^2+\theta/n$**, so** $T^* = \overline{X_n}^2 - \overline{X_n}/n$ **is unbiased for** θ^2 **and is min variance (by Lehmann-Scheffé & Rao-Blackwell).**
$\mathrm{var}(T^*) = 4\theta^3/n + 2\theta^2/n^2 >$ **CRLB, but** $\to$ **CRLB as** $n\to\infty$**.**
(See TPE P.30 for Poisson moments).

4.3 Examples

(i) Information in an n-sample

If X and Y are independent;

$$\ell'_\theta(X,Y) \equiv \frac{\partial}{\partial\theta}(\log(f_\theta(X,Y))) = \frac{\partial}{\partial\theta}(\log f_\theta(X) + \log f_\theta(Y)) = \ell'_\theta(X) + \ell'_\theta(Y)$$

These two terms are independent, each mean 0, so $\mathrm{E}(\ell'_\theta(X,Y)^2) = \mathrm{E}(\ell'_\theta(X)^2) + \mathrm{E}(\ell'_\theta(Y)^2)$ **or** $I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)$
For $X^{(n)} = (X_1, ..., X_n)$, X_i **i.i.d.;** $I_n(\theta) = nI_1(\theta)$.

(ii) Location parameter: JAW 3.9, TPE P.119

$f_\theta(x) = g(x-\theta)$ **for known** g, $\ell'_\theta(x) = -(g'/g)(x-\theta)$
$I(\theta) = \mathrm{E}((\ell'_\theta(X))^2) = \int \frac{g'(y)^2}{g(y)} dy \equiv I_g$
For n-sample: $I_n(\theta) = nI_g$, $\mathrm{E}_\theta(T(X^{(n)}) = \theta \Rightarrow \mathrm{var}_\theta(T_n) \geq 1/nI_g$; $\mathrm{var}_\theta(n^{\frac{1}{2}}(T_n - \theta)) \geq 1/I_g$.

(iii) Scale parameter: JAW3.9, TPE P.119

$f_\theta(x) = \theta^{-1}g(x/\theta)$, $\ell'_\theta(x) = \theta^{-1}(-1 - (x/\theta)(g'/g)(x/\theta))$
$I(\theta) = \theta^{-2} \int (-1 - y(g'/g)(y))^2 dy \equiv \theta^{-2} I_g^*$.
For n-sample: $I_n(\theta) = nI_g$, $\mathrm{E}_\theta(T(X^{(n)})) = \theta \Rightarrow \mathrm{var}_\theta(T_n) \geq \theta^2/nI_g^*$; $\mathrm{var}_\theta(n^{\frac{1}{2}}(T_n - \theta)/\theta) \geq 1/I_g^*$.

(iv) Reparametrization: $\psi = \psi(\theta)$ a 1-1 transformation.

Then $\ell'_\psi(X) = \frac{\partial}{\partial\psi}(\log f_\theta(X)) = (\psi'((\theta)))^{-1}\ell'_\theta(X)$
$I(\psi) = (\psi'((\theta)))^{-2} I(\theta)$.
But also $\frac{\partial}{\partial\psi}(q(\theta) + b(\theta)) = (q'(\theta) + b'(\theta))/\psi'(\theta)$, **so the CRLB is unchanged – as should be so!!**

4.4 Other lower bounds

(i) Back to Cauchy-Schwarz:
Suppose $\mathrm{E}_\theta(T^2) < \infty$ **and** $\Psi(X;\theta)$ **any function with** $0 < \mathrm{E}_\theta(\Psi(X,\theta)^2) < \infty$, **then** $\mathrm{var}_\theta(T) \geq (\mathrm{Cov}_\theta(T,\psi))^2/\mathrm{var}_\theta(\Psi)$. **In general, this is not useful since the r.h.s. involves** T.

(ii) Blyth's Theorem
(a) $\mathrm{Cov}_\theta(T,\Psi)$ **depends on** T **only through** $\mathrm{E}_\theta(T)$ **iff**
(b) $\mathrm{Cov}_\theta(V,\Psi) = 0 \ \forall \ V$ **s.t.** $\mathrm{E}_\theta(V) = 0 \ \forall\theta$ **and** $\mathrm{E}_\theta(V^2) < \infty$.
Proof: Suppose (b), and let $\mathrm{E}_\theta(T_1) = \mathrm{E}_\theta(T_2)$. **Consider** $V = T_1 - T_2$. **So** $\mathrm{Cov}(T_1,\Psi) = \mathrm{Cov}(T_2,\Psi)$, **Hence (a).**
Suppose (a), and take V **as in (b). So** $\mathrm{E}(T+V) = \mathrm{E}(T)$. **So** $\mathrm{Cov}(T+V,\Psi) = \mathrm{Cov}(T,\Psi)$. **Hence (b).**

(iii) Cor 1: $\Psi(X,\theta) = \ell'_\theta(X)$, **satisfies (b), hence (a), and** $\mathrm{Cov}(T,\Psi) = (\mathrm{E}_\theta(T))'$, **giving the CRLB.**

(iv) Cor 2: Hammersley-Chapman-Robbins Inequality
Assume $f_\theta(x) > 0 \ \forall x \in \mathcal{X}$. **Let** $\Psi(x,\theta) = (f_{\theta+\Delta}(x)/f_\theta(x) \ - 1)$. **So** $\mathrm{E}_\theta(\Psi(X,\theta)) = 0 \ \forall\Delta$ **and for** V **s.t. (b)** $\mathrm{Cov}(V,\Psi) = \mathrm{E}_\theta(\Psi V) = \mathrm{E}_{\theta+\Delta}(V) - \mathrm{E}_\theta(V) = 0$ **and** $\mathrm{Cov}(T,\Psi) = \mathrm{E}_{\theta+\Delta}(T) - \mathrm{E}_\theta(T)$, **so** $\forall\Delta$

$$\mathrm{var}_\theta(T) \ \geq \ (\mathrm{E}_{\theta+\Delta}(T) - \mathrm{E}_\theta(T))^2/\mathrm{E}_\theta\left(\frac{f_{\theta+\Delta}(X)}{f_\theta(X)} \ - 1\right)^2$$

(v) Cor 3: with appropriate differentiability, regularity etc., let $\Delta \to 0$ **in Cor 2, and we get back to CRLB** $((\mathrm{E}_\theta(T))')^2/I(\theta)$.

4.5 Multiparameter CRLB: $\Theta \subset \Re^k$.

(i) Theorem (Vector version of Cauchy-Schwarz/Blyth) For any unbiased estimator T of $q(\theta)$, and any functions $\Psi_i(X, \theta)$ with $\mathrm{E}_\theta(\Psi_i^2(X, \theta)) < \infty$, let $C_{ij} = \mathrm{Cov}_\theta(\Psi_i, \Psi_j)$, and $\gamma_i = \mathrm{Cov}(T, \Psi_i)$. Then (a) $\mathrm{var}(T) \geq \gamma^t C^{-1} \gamma$.

(b) The lower bound depends on T only through $q(\theta)$ provided $\mathrm{Cov}(V, \Psi) = 0 \ \forall V$ s.t. $\mathrm{E}_\theta(V) = 0 \ \forall \theta$ and $\mathrm{E}_\theta(V^2) < \infty$.

(ii) First, let $W = (W_1, ..., W_k)$ and T be r.vs with finite 2nd moments. Let $\rho(a) \equiv \rho(a^t W, T) \leq 1$. We show in (iv),(v) that $\sup_a(\rho^2(a)) \ = \ \gamma^t C^{-1} \gamma / \mathrm{var}(T)$, where $C = \mathrm{var}(W)$ and $\gamma = \mathrm{Cov}(W, T)$.

(iii) $\sup_a(\rho^2(a)) \leq 1$, so putting $W = \Psi$ gives theorem part (a). Then (b) follows exactly as in Blyth's theorem.

(iv) Suppose $C = \mathrm{var}(W)$ is positive definite, so $C = AA^t$, with A non-singular. Then

$$\rho^2(a) \ = \ \frac{(\mathrm{Cov}(a^t W, T))^2}{\mathrm{var}(a^t W)\mathrm{var}(T)} \ = \ \frac{(a^t \gamma)^2}{(a^t C a)\mathrm{var}(T)}$$

and $(a^t \gamma)^2 = (a^t A A^{-1} \gamma)^2 \leq (a^t A A^t a).(\gamma^t (A^{-1})^t A^{-1} 1 \gamma)$ $= (a^t C a)(\gamma^t C^{-1} \gamma)$ gives result.

(v) Note the $\sup$ is attained iff $A^t a \propto A^{-1} \gamma$; that is $a \propto C^{-1} \gamma$.

(vi) Under multidimensional analogues of 4.1 (ii) (a)-(e), $\mathrm{var}(T) \geq \alpha^t (I(\theta))^{-1} \alpha$ where $\alpha = \bigtriangledown \mathrm{E}_\theta(T(X))$, and $I(\theta) = \mathrm{E}_\theta(\bigtriangledown \ell_\theta(X) \bigtriangledown \ell_\theta(X)^t)$.

(vii) Proof: set $\Psi = \bigtriangledown \ell_\theta(X)$, and show $\mathrm{E}_\theta(\Psi) = 0$, $\gamma = \mathrm{Cov}(T, \bigtriangledown \ell_\theta(X)) = \bigtriangledown \mathrm{E}_\theta(T(X)) = \alpha$ and $C = I(\theta)$.

(viii) In case of equality: "$a^t W$" $\propto \ \bigtriangledown \mathrm{E}_\theta(T(X)) I(\theta)^{-1} \bigtriangledown \ell_\theta(X)$: this function is known as the *efficient influence function*

4.6 Nuisance parameters and submodels

(i) If $\mathrm{E}_\theta(T) = c^t\theta$**,** $\alpha = c$**,** $\mathrm{var}(T) \geq c^t(I(\theta))^{-1}c$
If $\mathrm{E}_\theta(T) = \theta_1$**,** $\mathrm{var}(T) \geq ((I(\theta))^{-1})_{11} = (I_{11} - I_{12}I_{22}^{-1}I_{21})^{-1} \geq (I_{11})^{-1}$**,**
and $= (I_{11})^{-1}$ **iff** $I_{12} = 0$**.**

(ii) Exponential families $\ell_\pi(X) = \log(c(\pi)) + \Sigma_{j=1}^k \pi_j T_j(X) + \log h(X)$**.**
$\bigtriangledown(\ell_\pi) = (T - \mathrm{E}_\pi(T)) = (T - \tau(\pi))$
$I(\pi) = \mathrm{E}((\bigtriangledown(\ell_\pi))(\bigtriangledown(\ell_\pi))^t) = \mathrm{var}(T)$
Now $\frac{\partial\tau}{\partial\pi} = \mathrm{Cov}(T, \bigtriangledown(\ell_p i)) = \mathrm{Cov}(T, T - \tau) = \mathrm{var}(T)$**,**
so $\mathrm{var}(T) = I(\pi) = \mathrm{var}(T)I(\tau)\mathrm{var}(T)$**,**
or $I(\tau) = (\mathrm{var}(T))^{-1}$**.**

(iii) Location-scale families: $f_{\theta,\sigma}(x) = \sigma^{-1}g((x-\theta)/\sigma)$**.**
Then $I_{11} = \sigma^{-2}I_g$**,** $I_{22} = \sigma^{-2}I_g^*$ **and** $I_{12} = \sigma^{-2}\int y\frac{g'(y)^2}{g(y)}dy$**.**
Note $I_{12} = 0$ **if** $g()$ **is symmetric about 0.**

(iv) JAW notes etc. on the geometry of this stuff – his class of Nov 7.

4.7 Asymptotic relative efficiency (ARE)

(i) Let $T_{1,n}$ and $T_{2,n}$ be two sequencies of estimators, each consistent for $q(\theta)$ and $T_{i,n}$ being based on an n-sample $X^{(n)}$. Let $n_2(n_1)$ be defined s.t. $\mathrm{var}(T_{2,n_2}) = \mathrm{var}(T_{1,n_1})$. Then the A.R.E. of $(T_{1,n})$ to $(T_{2,n})$ is $\lim_{n_1\to\infty} n_2(n_1)/n_1$.

(ii) Asymptotically Gaussian regular estimators:
If T_n is a consistent estimator of $q(\theta)$ based on i.i.d n-sample $X^{(n)}$, then (T_n) is an asymptotically Gaussian regular estimator if $n^{\frac{1}{2}}(T_n - q(\theta)) \to_d N(0, \tau^2(\theta))$.

(iii) For two Asymptotically Gaussian Regular estimators: $\mathrm{var}(n_1^{\frac{1}{2}} T_{1,n_1}) \to \tau_1^2$, $\mathrm{var}(n_1^{\frac{1}{2}} T_{2,n_2}) = (\sqrt{n_1/n_2})^2 \mathrm{var}(n_2^{\frac{1}{2}} T_{2,n})$. **For equal variance, for large n_1, $(n_1/n_2)\tau_2^2 = \tau_1^2$ or $\lim(n_2/n_1) = \tau_2^2/\tau_1^2$.**

(iv) Example: Suppose $X_1, ..., X_n$ are i.i.d from $F(x-\theta)$ with $F(0) = 1/2$ and $\mathrm{E}(X_i) = \theta$. Then $\overline{X_n}$ and $M = \mathrm{med}(X_i)$ are both consistent estimators of θ.
Suppose F has density f, $f(0) > 0$, and $\mathrm{var}(X_i) = \sigma^2 < \infty$. Then $n^{\frac{1}{2}}(\overline{X_n} - \theta) \to_d N(0, \sigma^2)$ and $n^{\frac{1}{2}}(M - \theta) \to_d N(0, 1/4(f(0))^2)$. So the ARE of $\overline{X_n}$ to M is $1/(4\sigma^2 f(0)^2)$

(v) Exercises: ARE($\overline{X_n} : M$)
$X_i \sim N(\theta, \sigma^2)$: ARE= $\pi/2$. $X_i \sim DE(\theta, \lambda)$: ARE =1. $X_i \sim U(\theta - \psi, \theta + \psi)$: ARE= 3. Note the scale parameter affects $\overline{X_n}$ and M equally.

(vi) Asymptotic efficiency of asymptotically Gaussian regular estimators: (when CRLB conditions apply)
If (T_n) is consistent for $q(\theta)$ and $n^{\frac{1}{2}}(T_n - q(\theta)) \to_d N(0, \tau^2(\theta))$, then $\tau^2(\theta) \geq (q'(\theta))^2/I_1(\theta)$. We define the (absolute) asymptotic efficiency of (T_n) to be $(q'(\theta)^2/(I_1(\theta)\tau^2))$
$= \lim(q'(\theta)^2/(I_n(\theta)\mathrm{var}(T_n)) = \lim(q'(\theta)^2/(I_1(\theta)\mathrm{var}(n^{\frac{1}{2}} T_n)) \leq 1$