Chapter 4: Estimation and Information (JAW Ch 3)

4.1 CRLB: one-dimensional real parameter (JAW 3.7-10; Severini 3.6)

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

(ii) Suppose: (a) $\Theta$ is an open subset of $\mathbb{R}$
(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 $\theta$
(d) $I(\theta) \equiv \mathbb{E}_\theta((\ell'_\theta(X))^2) > 0$ where $\ell'_\theta(x) \equiv \frac{\partial}{\partial \theta} \log f_\theta(x)$ is the Score function for $\theta$.
$I(\theta)$ is the Fisher Information for $\theta$.
(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. $\theta$ under the integral sign.

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

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

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

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

Prf: Differentiating again, gives $\mathbb{E}_\theta(\ell''_\theta(X)) + \mathbb{E}((\ell'_\theta(X))^2) = 0$
4.2 Assumption verification, and notes

(i) Assumption (ii)(e) can be the hard one to check. It holds for exponential families: see 2.1 and Sev. P. 81-82. More generally: using ' to denote $\frac{\partial}{\partial \theta}$, if $X'(\omega, \theta) \exists \forall \theta$ a.e.(\mu) 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 $\text{var}_\theta(T) = (q'(\theta))^2 / I(\theta)$, $T$ is MVUE of $q(\theta)$, and $\ell'_\theta(X) = k(\theta)(T(X) - q(\theta))$. 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 $\pi$ non-linear, $\nexists$ unbiased estimator achieving CRLB for $\pi(q(\theta))$.

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

(v) $T$ a MVUE of $q(\theta) \Rightarrow$
\[\text{var}(T) = q'(\theta)^2 / E(\ell'_\theta(X))^2 = q'(\theta)^2 / (k(\theta)^2 \text{var}_\theta(T))\]
so $\text{var}(T) = |q'(\theta)/k(\theta)|$ and $I(\theta) = q'(\theta)^2 / \text{var}(T) = |q'(\theta)k(\theta)|$.

(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) $E(T(X^{(n)})) = \sum_{x_1, \ldots, x_n} t(x^{(n)}) e^{-n \theta} \theta^{\sum x_i} / (\Pi x_i!)$, which is absolutely cgt power series in $\theta$ if $E(|T(X^{(n)})|) < \infty$, 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 $\text{var}(\overline{X}_n) = \theta/n = \text{CRLB}$. $I(\theta) = n/\theta$.

(c) For, $q(\theta) = \theta^2$, CRLB $= 4\theta^3/n$
$E(\overline{X}_n^2) = \theta^2 + \theta/n$, so $T^* = \overline{X}_n^2 - \overline{X}_n/n$ is unbiased for $\theta^2$ and is min variance (by Lehmann-Scheffé & Rao-Blackwell).
$\text{var}(T^*) = 4\theta^3/n + 2\theta^2/n^2 > \text{CRLB}$, but $\rightarrow \text{CRLB}$ as $n \rightarrow \infty$.
(Note $\overline{X}_n \sim n^{-1} P(n\theta)$: 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

$$
E(\ell'_\theta(X, Y)^2) = E(\ell'_\theta(X)^2) + E(\ell'_\theta(Y)^2) \text{ or } I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)
$$

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

$$
f_\theta(x) = g(x - \theta) \text{ for known } g, \ell'_\theta(x) = -(g'(x - \theta)/g(x - \theta))
$$

$$
I(\theta) = E((\ell'_\theta(X))^2) = \int \frac{g'(y)^2}{g(y)}dy \equiv I_g
$$

For n-sample: $I_n(\theta) = nI_g$, $E_\theta(T(X^{(n)}) = \theta \Rightarrow \text{var}_\theta(T_n) \geq 1/nI_g$;

$$
\text{var}_\theta(n^{1/2}(T_n - \theta)) \geq 1/I_g.
$$

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

$$
f_\theta(x) = \theta^{-1}g(x/\theta), \ell'_\theta(x) = \theta^{-1}(-1 - (x/\theta)(g'(x/\theta)/g(x/\theta)))
$$

$$
I(\theta) = \theta^{-2} \int(-1 - yg'(y)/g(y))^2g(y)dy \equiv \theta^{-2}I_g^*.
$$

For n-sample: $I_n(\theta) = nI_g$, $E_\theta(T(X^{(n)})) = \theta \Rightarrow \text{var}_\theta(T_n) \geq \theta^2/nI_g^*$;

$$
\text{var}_\theta(n^{1/2}(T_n - \theta)/\theta) \geq 1/I_g^*.
$$

(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 \( E_\theta(T^2) < \infty \) and \( \Psi(X; \theta) \) any function with \( 0 < E_\theta(\Psi(X, \theta)^2) < \infty \), then \( \text{var}_\theta(T) \geq (\text{Cov}_\theta(T, \psi))^2 / \text{var}_\theta(\Psi) \). In general, this is not useful since the r.h.s. involves \( T \).

(ii) Blyth’s Theorem
(a) \( \text{Cov}_\theta(T, \Psi) \) depends on \( T \) only through \( E_\theta(T) \) iff
(b) \( \text{Cov}_\theta(V, \Psi) = 0 \) \( \forall V \) s.t. \( E_\theta(V) = 0 \) \( \forall \theta \) and \( E_\theta(V^2) < \infty \).

Proof: Suppose (b), and let \( E_\theta(T_1) = E_\theta(T_2) \). Consider \( V = T_1 - T_2 \). So \( \text{Cov}(T_1, \Psi) = \text{Cov}(T_2, \Psi) \), Hence (a).
Suppose (a), and take \( V \) with \( E_\theta(V) = 0 \) \( \forall \theta \). So \( E_\theta(T + V) = E_\theta(T) \). So \( \text{Cov}_\theta(T + V, \Psi) = \text{Cov}_\theta(T, \Psi) \). Hence (b).

(iii) Cor 1: \( \Psi(X, \theta) = \ell'_\theta(X) \), satisfies \( \text{Cov}_\theta(V, \Psi) = (E_\theta(V))' \) for any \( V \), hence (b), hence also (a), and \( \text{Cov}_\theta(T, \Psi) = (E_\theta(T))' \), 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 \( E_\theta(\Psi(X, \theta)) = 0 \) \( \forall \Delta \) and for \( V \) s.t. (b) \( \text{Cov}(V, \Psi) = E_\theta(\Psi V) = E_{\theta+\Delta}(V) - E_\theta(V) = 0 \) and \( \text{Cov}(T, \Psi) = E_{\theta+\Delta}(T) - E_\theta(T) \), so \( \forall \Delta \)

\[
\text{var}_\theta(T) \geq (E_{\theta+\Delta}(T) - E_\theta(T))^2 / E_\theta \left( \frac{f_{\theta+\Delta}(X)}{f_\theta(X)} - 1 \right)^2
\]

(v) Cor 3: with appropriate differentiability, regularity etc., let \( \Delta \to 0 \) in Cor 2, and we get back to CRLB \( ((E_\theta(T))')^2 / I(\theta) \).
4.5 Multiparameter CRLB: $\Theta \subset \mathbb{R}^k$. (Sev. P90-91).

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

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

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

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

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

and $(a^t \gamma)^2 = (a^t A A^{-1} \gamma)^2 = ((A^t a)^t (A^{-1} \gamma))^2 \leq (a^t A A^{-1} a)(\gamma^t (A^{-1})^t A^{-1} \gamma) = (a^t C a)(\gamma^t C^{-1} \gamma)$ 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), $\text{var}(T) \geq \alpha^t (I(\theta))^{-1} \alpha$ where $\alpha = \nabla E_\theta(T(X))$, and $I(\theta) = E_\theta(\nabla \ell_\theta(X) \nabla \ell_\theta(X)^t)$.

(vii) Proof: set $\Psi = \nabla \ell_\theta(X)$, and show $E_\theta(\Psi) = 0$, $\gamma = \text{Cov}(T, \nabla \ell_\theta(X)) = \nabla E_\theta(T(X)) = \alpha$ and $C = I(\theta)$.

(viii) For $a \propto C^{-1} \gamma$, $a^t \Psi \propto (\nabla E_\theta(T))^{-1} \nabla \ell_\theta(X) = \tilde{\ell}_\theta(X)$. This function is known as the efficient influence function.
4.6 Nuisance parameters and submodels (Sev P.93-95)

(i) If \( E_\theta(T) = c^T \theta, \quad \alpha = c \), \( \text{var}(T) \geq c^T (I(\theta))^{-1} c \)

If \( E_\theta(T) = \theta_1 \), \( \text{var}(T) \geq ((I(\theta))^{-1})_{11} = (I_{11} - I_{12} I_{22}^{-1} I_{21})^{-1} \geq (I_{11})^{-1} \), and \( \text{var}(T) = (I_{11})^{-1} \) iff \( I_{12} = 0 \).

(ii) Attaining the bound: For \( \text{var}(T) = \gamma^T C^{-1} \gamma, \)

\( (T - E_\theta(T)) \propto \hat{\ell}_\theta(X) = \nabla (q(\theta)) I(\theta)^{-1} \nabla \ell_\theta(X) \).

(iii) Exponential families

\( \ell_\pi(X) = \log(c(\pi)) + \sum_{j=1}^k \pi_j T_j(X) + \log h(X) \),

\( \nabla(\ell_\pi) = (T - E_\pi(T)) = (T - \tau(\pi)) \)

\( I(\pi) = E((\nabla(\ell_\pi))(\nabla(\ell_\pi))^t) = \text{var}(T) \)

Now \( \frac{\partial \tau}{\partial \pi} = \text{Cov}(T, \nabla(\ell_\pi)) = \text{Cov}(T, T - \tau) = \text{var}(T) \),

so \( \text{var}(T) = I(\pi) = \text{var}(T) I(\tau) \text{var}(T) \),

or \( I(\tau) = (\text{var}(T))^{-1} \).

(iv) 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.

(v) Orthogonal parameters (Sev. 3.6.4): \( \theta = (\psi, \phi) \)

If \( \ell(\psi, \phi) = \ell^{(1)}(\psi) + \ell^{(2)}(\phi) \), \( \psi \) and \( \phi \) are orthogonal.

Inferences about \( \psi \) and \( \phi \) can be made separately.

If \( I_{\psi\phi}(\theta) \equiv E(\ell_\psi \ell_\phi^t) = -E_\theta(\ell_\psi \ell_\phi(\theta)) = 0 \), then \( \psi \) and \( \phi \) are (approx/asymptotically) orthogonal.

(Here and on following page: subscripts on \( \ell \) denote derivatives.)
4.7 Finding orthogonal parametrization (Severini P.95)

ψ is parameter of interest, λ a nuisance parameter; log-likelihood ℓ(ψ, λ)

Reparametrize as ℓ*(ψ, φ), where λ = λ(ψ, φ), φ = φ(ψ, λ).

\[ ℓ^*_\phi(ψ, φ)^t = ℓ_\lambda(ψ, λ)^t \frac{∂λ}{∂φ} \]

\[ ℓ^*_ψ(ψ, φ) = ℓ_ψ(ψ, λ) + \left( \frac{∂λ}{∂ψ} \right)^t ℓ_\lambda(ψ, λ) \]

So \( I^*_ψ(ψ, φ) = I_ψ(ψ, λ) \frac{∂λ}{∂φ} + \left( \frac{∂λ}{∂ψ} \right)^t I_\lambda(ψ, λ) \frac{∂λ}{∂φ} \)

For \( I^*_ψ(ψ, φ) = 0 \), (\( ∂λ/∂ψ \)) = \(- (I_\lambda(ψ, λ))^{-1} I_ψ(ψ, λ)\).

Can, in principle, be solved to find (many) φ(ψ, λ).

Example: Weibull dsn: with ψ the parameter of interest

\[ f(y; ψ, λ) = ψ λ^ψ y^{(ψ-1)} \exp(-(λy)^ψ) I_{(0,∞)}(y) \]

\[ I(ψ, λ) = \begin{pmatrix} \ldots & (1 - γ)/λ \\ (1 - γ)/λ & (ψ/λ)^2 \end{pmatrix} \]

so ψ and λ are not orthogonal. (γ = Euler’s const)

To find an orthogonal reparametrization

(\( ∂λ/∂ψ \)) = \(- (1 - γ)λ/ψ^2 \) or \( λ(ψ) = C \exp((1 - γ)/ψ) \).

That is \( C = \exp(\log(λ) + (γ - 1)/ψ) \), so \( φ = g(\log(λ) + (γ - 1)/ψ) \)

will work, where g is any smooth function.
4.8 Asymptotic relative efficiency (ARE)

(i) Let $T_{1,n}$ and $T_{2,n}$ be two sequences 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. $\text{var}(T_{2,n_2}) = \text{var}(T_{1,n_1})$. 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 $n_2^n(T_n - q(\theta)) \rightarrow_d N(0, \tau^2(\theta))$.

(iii) For two Asymptotically Gaussian Regular estimators:
$\text{var}(n_1^n T_{1,n_1}) \to \tau_1^2$, $\text{var}(n_2^n T_{2,n_2}) = (\sqrt{n_1/n_2})^2 \text{var}(n_2^n T_{2,n})$. For equal variance, for large $n_1$, $(n_1/n_2)\tau_2^2 \approx \tau_1^2$ 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 $E(X_i) = \theta$. Then $\overline{X}$ and $M = \text{med}(X_i)$ are both consistent estimators of $\theta$.
Suppose $F$ has density $f$, $f(0) > 0$, and $\text{var}(X_i) = \sigma^2 < \infty$.
Then $n_1^n(\overline{X} - \theta) \rightarrow_d N(0, \sigma^2)$ and $n_2^n(M - \theta) \rightarrow_d N(0, 1/4(f(0))^2)$.
So the ARE of $\overline{X}$ to $M$ is $1/(4\sigma^2 f(0)^2)$

(v) Examples (Note scale parameter cancels out)
$X_i \sim N(\theta, \sigma^2)$: $\overline{\theta} = \overline{X}$, $\text{var}(X_i) = \sigma^2$, ARE$= \pi/2$.
$X_i \sim DE(\theta, \lambda)$: $\overline{\theta} = M$, $\text{var}(X_i) = 2\lambda^2$, ARE$= 1/2$.
$X_i \sim U(\theta - \psi, \theta + \psi)$: $\text{var}(X_i) = \psi^2/3$, $f(0) = (2\psi)^{-1}$, ARE$= 3$.

(vi) Asymmetric efficiency of asymptotically Gaussian regular estimators: (when CRLB conditions apply)
If $(T_n)$ is consistent for $q(\theta)$ and $n_2^n(T_n - q(\theta)) \rightarrow_d N(0, \tau^2(\theta))$, then $\tau^2(\theta) \geq (q'(\theta))^2/I_1(\theta)$. We define the (absolute) asymptotic efficiency of $(T_n)$ to be $(q'(\theta)/I_1(\theta)\tau^2))$
$= \lim(q'(\theta)^2/(I_n(\theta)\text{var}(T_n))) = \lim(q'(\theta)^2/(I_1(\theta)\text{var}(n_1^n T_n)) \leq 1$