Math 152 - Fall 2006
Hardin
Efficiency & Rao-Cramér Lower Bound

Below are the important ideas from section 6.2 (Rao-Cramér Lower Bound and Efficiency) in your book. I've annotated so that you can see how the idea pertains to what we have been doing.

- Regularity conditions (pages 313, 319, 325)

(R0): The pdfs are distinct (identifiable); i.e., $\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x_i;\theta')$.

(R1): The pdfs have common support for all θ.

(R2): The point θ_o is an interior point in Ω.

(R3): The pdf $f(x;\theta)$ is twice differentiable as a function of θ.

(R4): The integral $\int f(x;\theta)dx$ can be differentiated twice under the integral sign as a function of θ.

(R5): The pdf $f(x;\theta)$ is three times differentiable as a function of θ. Further, for all $\theta \in \Omega$, there exists a constant c and a function $M(x)$ such that

$$\left| \frac{\partial^3}{\partial \theta^3} \ln f(x;\theta) \right| \leq M(x)$$

with $E_{\theta_o}[M(X)] < \infty$ for all $\theta_o - c < \theta < \theta_o + c$ and all x in the support of X.

- **Fisher's Information** (pages 319-320) The information was derived in class. It is used in the lower bound of the variance. Note that we can find Fisher's information in three different ways:

$$I(\theta) = E\left[\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^2\right]$$

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

$$I(\theta) = Var\left(\frac{\partial \ln f(X;\theta)}{\partial \theta}\right)$$

- **Theorem 6.2.1 Rao-Cramér Lower Bound** (page 322) Note that this theorem gives us a lower bound on the *variance* of any statistic. Let $X_1, X_2, \ldots, X_n$ be iid with common pdf $f(x;\theta)$ for $\theta \in \Omega$. Assume that regularity conditions (R0) - (R4) hold. Let $Y = u(X_1, X_2, \ldots, X_n)$ be a statistic with mean $E[Y] = E[u(X_1, X_2, \ldots, X_n)] = k(\theta)$. Then:

$$Var(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

  Wow!! Proof done in class as well as pages 322-323.

- **Corollary 6.2.1** (page 323) The previous theorem simplifies if we know that Y is unbiased for θ. Under the assumptions of Theorem 6.2.1, if $Y = u(X_1, X_2, \ldots, X_n)$ is an unbiased estimator of θ, then the Rao-Cramér inequality becomes:

$$Var(Y) \geq \frac{1}{nI(\theta)}$$

- **Definition 6.2.1** (page 323) Let Y be an unbiased estimator of a parameter θ in the case of point estimation. The statistic Y is called an **efficient estimator** of θ if and only if the variance of Y attains the Rao-Cramér lower bound. That is iff:

$$Var(Y) = \frac{1}{nI(\theta)}$$

- **Definition 6.2.2** (page 324) In cases in which we can differentiate with respect to a parameter under an integral (or summation for discrete random variable) symbol, the ratio of the Rao-Cramér lower bound to the actual variance of any unbiased estimator of a parameter is called the **efficiency** of that estimator. That is, if Y is unbiased,

$$\text{efficiency} = \frac{1/nI(\theta)}{\text{var}(Y)}$$

- Why all the unbiasedness? It's so that we can compare apples to apples. It doesn't make sense to compare the efficiencies of two estimators that aren't estimating the same parameter. What if we have consistency? Well, we need slightly more than consistency, we need that the estimator converges in distribution to a Normal random variable with mean zero.

- **Definition 6.2.3** (pages 326-327) Here are some of the similar ideas with estimators that are not only consistent but also asymptotically normal. Let $X_1, X_2, \ldots X_n$ be independent and identically distributed with pdf $f(x;\theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, X_2, \ldots, X_n)$ is an estimator of θ_o such that $\sqrt{n}(\hat{\theta}_{1n} - \theta_o) \xrightarrow{D} N(0, \sigma^2_{\hat{\theta}_{1n}})$. Then

  (a) The **asymptotic efficiency** of $\hat{\theta}_{1n}$ is defined to be:

  $$e(\hat{\theta}_{1n}) = \frac{1/I(\theta_o)}{\sigma^2_{\hat{\theta}_{1n}}}$$

  (b) The estimator $\hat{\theta}_{1n}$ is said to be **asymptotically efficient** if the ratio in (a) is 1.

  (c) Let $\hat{\theta}_{2n}$ be another estimator such that $\sqrt{n}(\hat{\theta}_{2n} - \theta_o) \xrightarrow{D} N(0, \sigma^2_{\hat{\theta}_{2n}})$. Then the **asymptotic relative efficiency (ARE)** of $\hat{\theta}_{1n}$ to $\hat{\theta}_{2n}$ is the reciprocal of the ratio of their respective asymptotic variances:

  $$e(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \frac{\sigma^2_{\hat{\theta}_{2n}}}{\sigma^2_{\hat{\theta}_{1n}}}$$

- **Theorem 6.2.2** (page 325) MLEs are asymptotically normal and asymptotically efficient!! Assume $X_1, X_2, \ldots, X_n$ are iid with pdf $f(x;\theta_o)$ for $\theta_o \in \Omega$ such that the regularity conditions (R0) - (R5) hold. Suppose further that Fisher's information satisfies: $0 < I(\theta_o) < \infty$. Then any consistent sequence of solutions of the mle equations satisfies

  $$\sqrt{n}(\hat{\theta} - \theta_o) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_o)}\right)$$

  proof in the text pages 325-326.

- **Corollary 6.2.2** (page 328) Functions of MLEs have the same nice asymptotic properties as MLEs. Under the assumptions of Theorem 6.2.2, suppose $g(x)$ is a continuous function of x which is differentiable at θ_o such that $g'(\theta_o) \neq 0$. Then

  $$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_o)) \xrightarrow{D} N\left(0, \frac{g'(\theta_o)^2}{I(\theta_o)}\right)$$

  proof is straightforward using the delta method, but we aren't going to review the delta method for this course.