b. Examples:

i. Binomial Distribution.

- Rival Unbiased Estimators of $\pi$:
  
  ▶ Suppose $X \sim \text{Bin}(n, \pi)$ and $Y \sim \text{Bin}(m, \pi)$.
  
  ▶ Let $\delta_1(X, Y) = X/n$ and $\delta_2(X, Y) = (X + Y)/(m + n)$.
  
  ▶ By not using some information, $\delta_1$ throws away information. How is this mathematically quantified?

- Calculating Relative Efficiency: Note that
  
  $\text{Var} [\delta_2(X, Y)] = \pi (1 - \pi)/(m + n)$ and
  
  $\text{Var} [\delta_1(X)] = \pi (1 - \pi)/n$. Note that
  
  $\text{Var} [\delta_1(X)] > \text{Var} [\delta_2(X, Y)]$.

ii. Estimating a general mean:

- Consider two ind. measurements $X_1$ and $X_2$, with a common mean $\mu$ and variance $\sigma^2$.

- Then $a_1 X_1 + a_2 X_2$ is unbiased if and only if $a_1 + a_2 = 1$.

- The variance is $(a_1^2 + a_2^2)\sigma^2$, which is minimized when $a_1 = a_2 = \frac{1}{2}$.

- Relative efficiency of the variance minimizing estimator to the general estimator is $2(a_1^2 + a_2^2)$.
iii. Poisson variable.

- Mean and variance of a \( P(\mu) \) random variable are both \( \mu \);
- hence an alternate estimator for \( \mu \) might be the sample variance \( \delta(X) = (n - 1)^{-1}(\sum_{i=1}^{n} X_i^2 - n \bar{X}^2) \).
- To see that this is unbiased, refer to discussion in book about generic variance
- It can be shown that \( \text{Var}[\delta(X)] \approx \mu(1 + 2\mu)/n \).
- sample mean is unbiased and has variance \( \mu/n \).
- relative efficiency of the sample variance to the sample mean is approximately
  \[
  \frac{\mu(1 + 2\mu)/n}{\mu/n} = 1 + 2\mu.
  \]
- Here relative efficiency depends on \( \theta \).
  - This is a relatively simple case, in which one estimator is always better than the other;
  - it need not be the case.

F: 10.4

7. Consistency.

a. Objective: Would like to know the maximum distance our estimator can possibly be from the true parameter value.
i. In many cases, for instance in the case of normal means, the answer is easy: It could be any finite distance away.

b. Realistic Objective: Can we claim that \( \hat{\theta} \) lies within a certain (preferably small) distance from \( \theta \) with a certain probability.

i. As we saw with our efficiency and Cramér-Rao bound calculations, \( \text{Var} [\hat{\theta}] \) usually decreases as \( n \) increases.

ii. Think of \( \hat{\theta} \) as the family of estimators based on various sample sizes,

c. Definition:

i. In words, An estimator \( \hat{\theta} \) is called consistent if
   
   • given
     
     ▶ any high probability of seeing \( \hat{\theta} \) within a certain band, and
     
     ▶ any very small width for this band,
   
   • a large enough \( n \) ensures that the probability that \( \hat{\theta} \) is within the required distance of the true value is as required.

   ii. \( \forall C > 0 \text{ and } \delta > 0 \exists M \text{ possibly depending on } \delta \text{ and } C \text{ such that } P \left[ \left| \hat{\theta} - \theta \right| \leq C \right] > 1 - \delta \text{ for any } n > M \).

d. Example:

i. if \( X_1, \ldots, X_n \sim N(\mu, \sigma^2) \),
• Estimate $\theta$ by $\hat{\theta}_n = \bar{X} \sim N(\mu, \sigma^2/n)$.

• Then $(\hat{\theta}_n - \mu)/\left(\sigma/\sqrt{n}\right) \sim N(0, 1)$.

• Hence $\Pr \left[ \left| \hat{\theta}_n - \mu \right| \leq C \right]$

\[ = \Pr \left[ \left| \frac{\hat{\theta}_n - \mu}{\sigma/\sqrt{n}} \right| \leq \frac{\sqrt{n} C}{C/\sigma} \right] = \Phi \left( \frac{\sqrt{n} C}{\sigma} \right) - \Phi \left( - \frac{\sqrt{n} C}{\sigma} \right), \]

where $\Phi$ is the c.d.f. of a $N(0, 1)$ variable.

• $\lim_{n \to \infty} \Pr \left[ \left| \hat{\theta}_n - \mu \right| \leq C \right] = 1$

• Let $z_{\delta/2}$ satisfy $\Phi(z_{\delta/2}) = 1 - \delta/2$

• For all $n$ such that $\sqrt{n} C/\sigma > z_{\delta/2}$ we have

\[ \Pr \left[ \left| \hat{\theta}_n - \theta \right| \leq C \right] > 1 - \delta. \]

• Hence $n > z_{\delta/2}^2 \sigma^2/C^2 \Rightarrow \Pr \left[ \left| \hat{\theta}_n - \theta \right| \leq C \right] > 1 - \delta$.

ii. An inconsistent Estimator: Suppose $f_X(x; \mu) = \pi^{-1} \left(1 + (x - \mu)^2\right)^{-1}$.

• density of $Z = \frac{1}{2}(X + Y)$ and $W = X$ is

\[ f_{W,Z}(w, z; \mu) = \pi^{-2} \left(1 + (w - \mu)^2\right)^{-1} \left(1 + (2z - w - \mu)^2\right)^{-1} \]

• Integrate re $w$:

\[ f_Z(z; \mu) = \int_{-\infty}^{+\infty} \frac{2 \, dw}{\pi^2 \left(1 + (w - \mu)^2\right)(1 + (2z - w - \mu)^2)}. \]

• Substitute $w - \mu = v + z$ and using partial fractions:
\[
\frac{1}{4z (1 + z^2)} \left[ \frac{2z - v}{(1 + v^2 - 2vz + z^2)} + \frac{v + 2z}{(1 + v^2 + 2vz + z^2)} \right]
\]

- Hence \( Z \) has same distn as \( X \) and \( Y \).

- Hence mean of \( 2^k \) variables has the same distn as \( X \).

- Hence mean is inconsistent.

e. A general rule

i. Often hard: Usually the bounds on \( n \) are not so easily derived explicitly.

ii. Use Chebyshev’s inequality: Relate the probability that a random variable \( T \) is farther than a distance \( C \) from its mean \( \theta \) to its variance.

\[
\text{Var} [T] = \sum_t (t - \theta)^2 p_T(t; \theta)
= \sum_{\{t||t-\theta|<C\}} (t - \theta)^2 p_T(t; \theta) + \sum_{\{t||t-\theta|\geq C\}} (t - \theta)^2 p_T(t; \theta)
\geq 0 + \sum_{\{t||t-\theta|\geq C\}} (C)^2 p_T(t; \theta)
= C^2 \sum_{\{t||t-\theta|\geq C\}} p_T(t; \theta)
= C^2 P [|T - \theta| \geq C].
\]
Densities from Cauchy distribution

Calculated from 200000 random draws
P \left[ \left| \hat{\theta} - \theta \right| \geq C \right] \leq \text{Var} \left[ \hat{\theta} \right] / C^2.

iii. Hence if \( E \left[ \hat{\theta} \right] = \theta \) and \( \lim_{n \to \infty} \text{Var} \left[ \hat{\theta} \right] = 0 \), then \( \hat{\theta} \) is consistent.

iv. Examples

- If \( X \sim \text{Bin}(n, \theta) \), and \( \hat{\theta}_n = X/n \), then
  \[
  \text{Var} \left[ \hat{\theta} \right] = \theta (1 - \theta)/n. \quad \text{Then}
  \]
  \[
  P \left[ \left| \hat{\theta}_n - \theta \right| \geq C \right] \leq \theta (1 - \theta)/(C^2 n) \leq 1/(4C^2 n).
  \]

- If \( X_1, \ldots, X_n \sim P(\mu) \), \( \hat{\mu} = \bar{X} \)

- \( \text{Var} \left[ \hat{\mu} \right] = \mu/n \)

- Applying Chebyshev’s inequality, \( P \left[ \left| \hat{\mu} - \mu \right| \geq C \right] \leq \mu/(C^2 n) \) proves consistency,

- the values of \( n \) making the RHS smaller than some limit \( \delta \) depend on \( \mu \).