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?
   ii. Estimating a general mean:
      • Consider two ind. measurements $X_1$ and $X_2$,
        with a common mean $\mu$ and variance $\sigma^2$.
      • Then $a_1X_1 + a_2X_2$ 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 $\mathcal{P}(\mu)$ random variable are
        both $\mu$;
      • hence an alternate estimator for
        $\mu$ might be the sample variance
        $\hat{\sigma}(X) = (n - 1)^{-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2)$.
      • 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, $\text{Var} \left( \frac{\hat{\theta}}{\bar{X}} \right)$ 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. $\forall C > 0$ and $\delta > 0 \exists M$ possibly depending on $\delta$
      and $C$ such that $\text{P} \left[ \left| \hat{\theta} - \theta \right| \leq C \right] > 1 - \delta$
      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 $(\hat{\theta}_n - \mu)/(\sigma/\sqrt{n}) \sim N(0,1)$.
      • Hence
        $\text{P} \left[ \left| \hat{\theta}_n - \mu \right| \leq C \right] = \Phi \left( \frac{\sqrt{nC}}{\sigma} \right) - \Phi \left( -\frac{\sqrt{nC}}{\sigma} \right)$,
        ▷ where $\Phi$ is the c.d.f. of a $N(0,1)$ variable.
      • $\lim_{n \to \infty} \text{P} \left[ \left| \hat{\theta}_n - \mu \right| \leq C \right] = 1$
      • Let $z_{\delta/2}$ satisfy $\Phi(z_{\delta/2}) = 1 - \delta/2$
      • For all $n$ such that $\sqrt{nC}/\sigma > z_{\delta/2}$ we have
        $\text{P} \left[ \left| \hat{\theta}_n - \theta \right| \leq C \right] > 1 - \delta$.
   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.
Densities from Cauchy distribution

\[
\text{Var}[T] = \sum_{t} (t-\theta)^2 p_T(t; \theta)
\]
\[
= \sum_{\{t|t-\theta<C\}} (t-\theta)^2 p_T(t; \theta) + \sum_{\{t|t-\theta\geq C\}} (t-\theta)^2 p_T(t; \theta)
\]
\[
\geq 0 + \sum_{\{t|t-\theta\geq C\}} (C)^2 p_T(t; \theta)
\]
\[
= C^2 \sum_{\{t|t-\theta\geq C\}} p_T(t; \theta)
\]
\[
= C^2 P[|T-\theta| \geq C].
\]
\[
P[|\hat{\theta} - \theta| \geq C] \leq \text{Var}[\hat{\theta}] / C^2.
\]
iii. Hence if \(E[\hat{\theta}] = \theta\) and \(\lim_{n \to \infty} \text{Var}[\hat{\theta}] = 0\), then \(\hat{\theta}\) is consistent.

iv. Examples
- If \(X \sim \text{Bin}(n, \theta)\), and \(\hat{\theta}_n = X/n\), then
  \[
  \text{Var}[\hat{\theta}] = \theta(1-\theta)/n.
  \]
  Then
  \[
  P[|\hat{\theta}_n - \theta| \geq C] \leq \theta(1-\theta)/(C^2n) \leq 1/(4C^2n).
  \]
- If \(X_1, \ldots, X_n \sim \mathcal{P}(\mu)\), \(\bar{X} \sim \bar{X}\)
  \[
  \text{Var}[\bar{X}] = \mu/n
  \]
- Applying Chebyshev’s inequality,
  \[
  P[|\bar{X} - \mu| \geq C] \leq \mu/(C^2n)\]
  proves consistency,
- the values of \(n\) making the RHS smaller than some limit \(\delta\) depend on \(\mu\).