C. Want estimate that is *robust*

1. **Definition:** Estimator behaves well when model assumptions aren’t satisfied
2. **Examples:**
   a. In tire example,
      i. estimate $\hat{\theta} = \text{proportion of bad tires in sample}$ is still an OK estimator even if tires on a single car don’t act independently
      ii. Hence estimate is robust to this model failure
   b. If you expect measurements to be normal,
      i. Most efficient estimator is sample mean
      ii. Not robust to model failure: if data are actually Cauchy, performance is very bad.

D. Techniques for generating estimates

F: 10.7

1. **Method of Moments**
   a. **Definition:**
      i. Suppose $X_1, \ldots, X_n \sim f_X(x; \theta)$
      ii. Law of large numbers tells us that $\sum_{j=1}^{n} X_j / n \approx E_\theta [X]$
iii. Method of moments says solve \( \sum_{j=1}^{n} X_j/n = \mathbb{E}_{\hat{\theta}}[X] \) for \( \hat{\theta} \).

iv. Expectations above are functions of \( \theta \).

v. If there are multiple parameters, might solve
\[
\sum_{j=1}^{n} X_j^2/n = \mathbb{E}_{\hat{\theta}}[X^2],
\]
and higher powers.

b. Examples:

i. \( X_1, \ldots, X_k \sim \text{NBin}(\theta, m) \)
   - Number of trials it takes to get \( m \) successes, if each has success probability \( \theta \)
   - \( \mathbb{E}[X_j] = m/\theta \) (Theorem 5.6).
   - Estimate \( \theta \)
   - \( \bar{X} = m/\hat{\theta} \)
   - \( \hat{\theta} = m/\bar{X} \)

ii. The normal dist’n. \( X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2) \).
   - Then \( \sum_{j=1}^{n} X_j^1/n = \hat{\mu}^1, \sum_{j=1}^{n} X_j^2/n = \hat{\mu}^2 + \hat{\sigma}^2 \)
   - Hence \( \hat{\mu} = \bar{X} \) and \( \sum_{j=1}^{n} X_j^2/n = \bar{X}^2 + \hat{\sigma}^2 \), or
   \[
   \hat{\sigma} = \sqrt{\sum_{j=1}^{n} X_j^2/n - \bar{X}^2} = \sqrt{\sum_{j=1}^{n} (X_j - \bar{X})^2/n}.
   \]
   - Recall that this estimate of \( \sigma^2 \) is biased.
   - Contrary to what may seem obvious from their definition,
Lecture 6

these estimators need not be unbiased.

iii. \( X_1, \ldots, X_n \sim \text{Bin}(m, \pi) \)
- \( \mathbb{E}[X_j] = m\pi \)
- \( \hat{\pi} = \bar{X}/m = (\sum_j X_j)/(nm) \).

iv. Same setup as before
- This time estimate \( \psi = \pi/(1 - \pi) \)
  - called odds
    - \( \pi = \psi/(1 + \psi) \)
    - \( \bar{X} = m\hat{\psi}/(1 + \hat{\psi}) \)
  - \( \hat{\psi} = \bar{X}/m/(1 - \bar{X}/m) = \hat{\pi}/(1 - \hat{\pi}) \)

c. Last example demonstrates equivariance: if you change scale of parameter, you change estimate in exactly the same way.

d. Problems with m.o.m.e.s:
   i. No guarantee of near-efficiency.
   - Since \( \text{Var} \left[ \sum_{j=1}^{n} X_j^k/n \right] \) generally large \( k \) large, \( k \) large, then in the estimating equations may add a lot of variability to \( \hat{\theta} \).
   ii. They may not even exist: cf. Cauchy distń.

e. Main advantage:
   i. Intuitive.
Lecture 7

ii. Generally speaking consistent.

2. Extensions

a. Can equate other sample quantities with population quantities
   i. Ex., median
   ii. Works better for some distributions like Cauchy