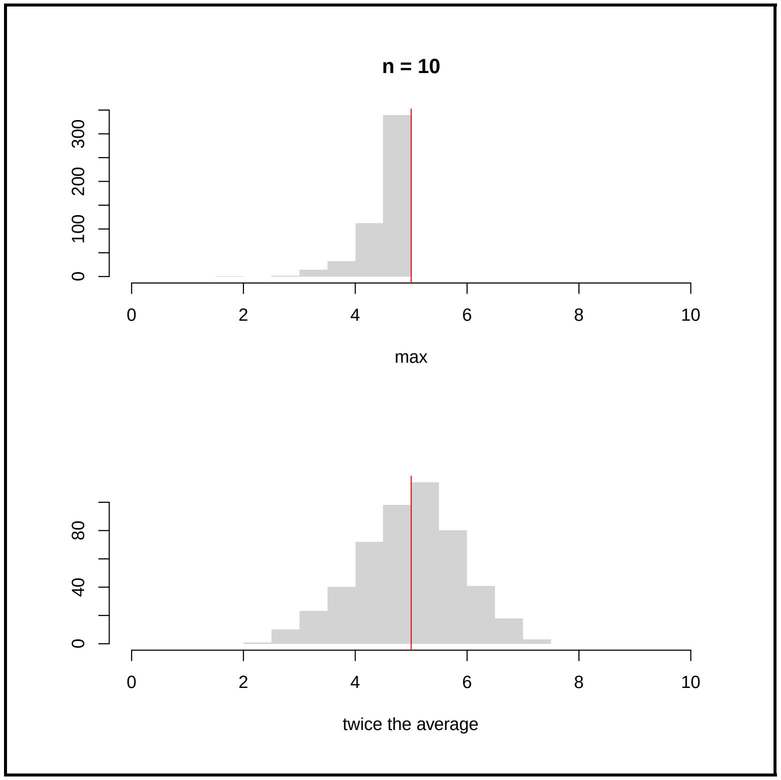
Review

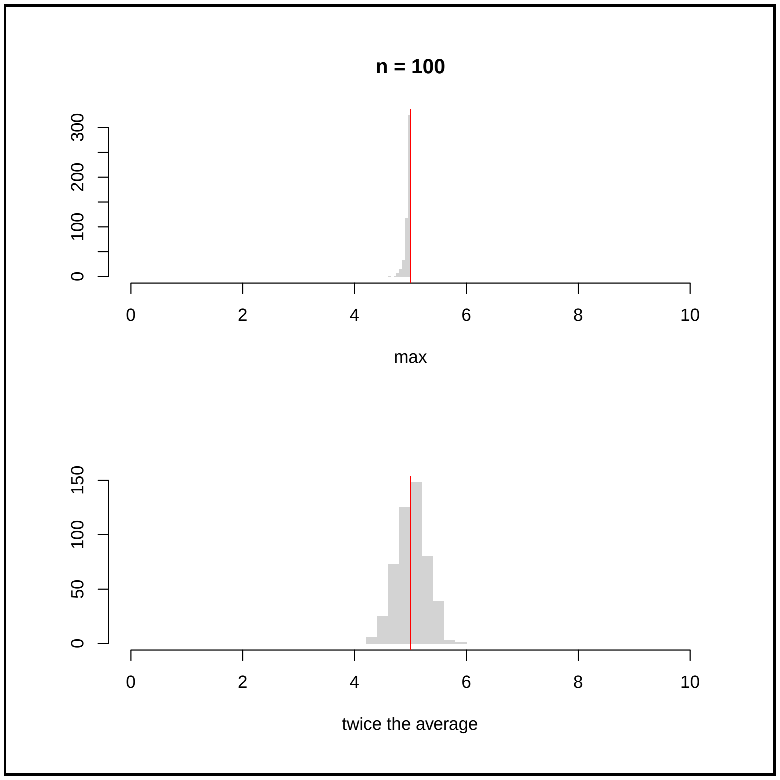
$$f(x_1, ..., x_n; \theta) \rightarrow \left\{ \begin{array}{ccc} \{x_1^{(1)}, \; ... \; , x_n^{(1)}\} & \rightarrow & \widehat{\theta}^{(1)} \\ \{x_1^{(2)}, \; ... \; , x_n^{(2)}\} & \rightarrow & \widehat{\theta}^{(2)} \\ \{x_1^{(3)}, \; ... \; , x_n^{(3)}\} & \rightarrow & \widehat{\theta}^{(3)} \\ \vdots & & \vdots \\ \{x_1^{(B)}, \; ... \; , x_n^{(B)}\} & \rightarrow & \widehat{\theta}^{(B)} \end{array} \right\} \text{distribution of } \widehat{\theta}$$

imagine many data sets from the same data-generating process

Review: Uniform$(0, \theta)$

$$\widehat{\theta}_{mle} = \max\{X_1, \ldots, X_n\} \text{ versus } \widehat{\theta}_{mom} = 2\bar{X}$$






```
> source("https://www.math.uwaterloo.ca/~m3zhu/teaching/stat845/f23dp6lec05code.R")
> do.it(n=10)
> do.it(n=100)
```

Example: Uniform$(0, \theta)$

MOM $\widehat{\theta} = 2\bar{X}$. $\mathbb{E}(X_i) = \theta/2$, $\mathbb{V}\mathrm{ar}(X_i) = \theta^2/12$.

$$\frac{\bar{X} - \theta/2}{\theta/\sqrt{12n}} \sim \mathrm{N}(0,1) \Rightarrow \bar{X} \sim \mathrm{N}\left(\frac{\theta}{2}, \frac{\theta^2}{12n}\right) \text{ approximately}$$

$$\Rightarrow \widehat{\theta} = 2\bar{X} \sim \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right) \text{ approximately}$$

MLE $\widehat{\theta} = \max\{X_1, X_2, ..., X_n\} \equiv Y$. $f_X(x) = 1/\theta, F_X(x) = x/\theta$.

$$\widehat{\theta} \equiv Y \sim f(y) = n[F_X(y)]^{n-1} f_X(y) = n\left[\frac{y}{\theta}\right]^{n-1}\left[\frac{1}{\theta}\right] = \frac{ny^{n-1}}{\theta^n}$$

Bias

Definition An estimator $\widehat{\theta}$ is unbiased if $\mathbb{E}(\widehat{\theta}) = \theta$.

$$\mathbb{B}\text{ias}(\widehat{\theta}) = \mathbb{E}(\widehat{\theta}) - \theta.$$

Exercise 5.7 (p. 78) For $X_1, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2)$, the MLE,

$$\widehat{\sigma}^2_{(mle)} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

is biased, whereas the so-called "sample variance",

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

is unbiased. [*Remark: But the actual distribution of S^2 is much harder to find.*]

Efficiency

Definition If both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are unbiased, then $\widehat{\theta}_1$ is more efficient than $\widehat{\theta}_2$ if $\mathbb{V}\text{ar}(\widehat{\theta}_1) \leq \mathbb{V}\text{ar}(\widehat{\theta}_2)$.

Example Given two independent observations X, Y from the same distribution with mean/expectation θ, which one,

$$\frac{2X + 8Y}{10} \quad \text{or} \quad \frac{3X + 7Y}{10}$$

is a more efficient way of estimating the unknown parameter θ?
[*Think*: *Can you find an even more efficient way?*]

Mean-Squared Error

Definition The mean-squared error of $\widehat{\theta}$ is

$$\mathrm{MSE}(\widehat{\theta}) = \mathbb{E}[(\widehat{\theta} - \theta)^2].$$

Theorem

$$\begin{aligned}
\mathrm{MSE}(\widehat{\theta}) &\equiv \mathbb{E}[(\widehat{\theta} - \theta)^2] = \mathbb{E}\{[\widehat{\theta} - \mathbb{E}(\widehat{\theta}) + \mathbb{E}(\widehat{\theta}) - \theta]^2\} \\
&= \mathbb{E}\{[\widehat{\theta} - \mathbb{E}(\widehat{\theta})]^2\} + [\mathbb{E}(\widehat{\theta}) - \theta]^2 + 2\,\underbrace{\mathbb{E}\{[\widehat{\theta} - \mathbb{E}(\widehat{\theta})][\mathbb{E}(\widehat{\theta}) - \theta]\}}_{=0} \\
&= \mathbb{V}\mathrm{ar}(\widehat{\theta}) + [\mathbb{B}\mathrm{ias}(\widehat{\theta})]^2
\end{aligned}$$

Exercise Show that $\mathbb{E}\{[\widehat{\theta} - \mathbb{E}(\widehat{\theta})][\mathbb{E}(\widehat{\theta}) - \theta]\} = 0$.

Exercise 5.8d (pp. 78–79): Uniform$(0, \theta)$

MOM

$$\widehat{\theta}_{mom} = 2\bar{X} \sim \mathrm{N}\left(\theta, \frac{\theta^2}{3n}\right) \quad \text{approximately}$$

$$\mathbb{E}(\widehat{\theta}_{mom}) = ..., \quad \mathbb{V}\mathrm{ar}(\widehat{\theta}_{mom}) = ..., \quad \mathrm{MSE}(\widehat{\theta}_{mom}) = ...$$

MLE

$$\widehat{\theta}_{mle} = \max\{X_1, X_2, ..., X_n\} \equiv Y \sim f(y) = \frac{ny^{n-1}}{\theta^n}$$

$$\mathbb{E}(\widehat{\theta}_{mle}) = ..., \quad \mathbb{V}\mathrm{ar}(\widehat{\theta}_{mle}) = ..., \quad \mathrm{MSE}(\widehat{\theta}_{mle}) = ...$$

Recall Example 5.2 (pp. 72–73)

$$X_1, X_2, ..., X_n \quad \text{independent}, \quad \text{each} \sim \text{Poisson}(\theta v_i)$$

$$\begin{aligned} \text{e.g.,} \quad X_i &= \text{number of disasters,} \\ v_i &= \text{number of activities} \end{aligned}$$

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta v_i}(\theta v_i)^{x_i}}{x_i!} \quad \Rightarrow \quad \ell(\theta) = \sum_{i=1}^{n} -\theta v_i + x_i \log(\theta v_i) + C$$

$$\ell'(\theta) = 0 \quad \Rightarrow \quad \widehat{\theta}_{mle} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} v_i} = \frac{\#\{\text{disasters over } n \text{ teams}\}}{\#\{\text{activities over } n \text{ teams}\}}$$

Continuation: Example 5.4 (pp. 76–78)

- take slightly different point of view
- n separate "data sets", each of size one
- on each "data set", the same "disaster-to-activity" estimator would be equal to X_i/v_i, for $i = 1, 2, ..., n$
- can then average these n "individual MLEs" to obtain an overall estimator,

$$\widehat{\theta}_{alt} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{v_i}$$

Question Which estimator has smaller MSE, $\widehat{\theta}_{alt}$ or $\widehat{\theta}_{mle}$?