Notes on Neyman-style frequentist inference

(1) (large-sample version) The diagram with the 3 datasets (population, sample, repeated sampling) completely describes frequentist inference (Neyman-style) but is cumbersome.

Remarkably, the entire diagram can be reduced to a single line of math:

\[ \mathbb{P}(\hat{\theta} \mid \theta, \Theta) \sim \text{Beta}(\alpha, \beta) \]

\[ (\hat{\theta} = 1, \ldots, n) \]

\[ \text{from which } \hat{\theta} = \frac{\sum \hat{\theta}_i}{n}, \text{ (random variable)} \]

\[ \text{Once } \sum \hat{\theta}_i \geq \frac{n}{2}, \text{ you've decided to use } \hat{\theta} \text{ as your estimate of } \theta, \hat{\theta} \text{ can be reduced even further in this case study:} \]

Under sampling model \( \Theta \), \( (\hat{\theta} \mid \Theta, \beta) \sim \text{Binomial}(n, \theta) \)
Later in the course, $\Theta$ will be a vector of length $k \geq 1$, $\Theta = (\Theta_1, \ldots, \Theta_k)$ parameter but for now, $k = 1$.

**Sampling model $\Theta$ is a special case of the following:**

$$(\mathbf{Y} \sim_\Theta \mathbf{X}) \quad \sim \quad p(\mathbf{Y} \sim_\Theta \mathbf{X})$$

**In which**

$p(\mathbf{Y} \sim_\Theta \mathbf{X})$ is a family of distributions (PMF discrete, PDF continuous) indexed by $\Theta$.

Algorithm for Neyman-style frequentist inference

1. **Specify** $p(\mathbf{Y} \sim_\Theta \mathbf{X})$ in a manner appropriate to the context $C$ of the problem $P$.
   - good
   - Specify a good

2. **Choose a** estimate $\hat{\Theta} = h(\mathbf{Y}, \ldots, \mathbf{Y}_n)$; good = (intuitively captures most or all of the information in $\mathbf{Y} = (\mathbf{Y}, \ldots, \mathbf{Y}_n)$ about $\Theta$; obeys a Central Limit Theorem).
Work out the details of the CLT for \( \hat{\theta} : \) for large \( n, \) \( (\hat{\theta} - \theta) / \text{RS} \sim N(0, \sigma^2 / n) \)

\( \text{RS = repeated sampling} \) (need to be able to figure out \( \text{V}(\hat{\theta}) \))

\( \text{and how to estimate it} ) \)

\( \text{SE}(\hat{\theta}) = \sqrt{\frac{\text{V}(\hat{\theta})}{n}}. \)

Approximate 100(1-\(d\)% CI for \( \hat{\theta} \) is then

\[ \hat{\theta} \pm z_{1-\frac{\alpha}{2}} \cdot \text{SE}(\hat{\theta}). \]