Creating estimators

- Estimators can be broadly organized into *method of moments* (MoM) estimators and Bayesian estimators
  - The MLE is Bayesian
- Evaluating estimators
  - Bias
  - SE
  - MSE, where $\mathrm{MSE} = \mathrm{Var} + \mathrm{Bias}^2$
  - Probability, e.g., $P(\mid \hat{\theta} - \theta \mid \geq c)$, where c is some error tolerance
- Methods to evaluate
  - Stat 110 or math style **BASHING**
    - Not recommended for the faint of heart or people without the constitution for math
  - Simulation
  - Asymptotics

Empirical CDF

- Estimand is the CDF
- Calculated as follows:

$$\hat{F}(y) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Y_j \leq y)$$

- Fairly intuitive interpretation
  - Percentage of data points less than or equal to a specified point
- Note that this is a *method of moments* estimator, where $F(y) = \mathbb{E}[\mathbb{1}(Y_j \leq y)]$
- Convergence proof follows from LLN
  - $\hat{F}_n(y) \rightarrow F(y)$ as $n \rightarrow \infty$ (with probability 1)

Asymptotics

- 5 tools for asymptotics
  - Law of law numbers, LLN (Stat 110)
  - Central limit theorem, CLT (Stat 110)
  - Continuous mapping theorem, CMT
  - Delta method
  - Slutsky's theorem
    - You can laugh a little

LLN

Ex: Let $Y_1, Y_2, \ldots$ be i.i.d. with mean μ and variance σ^2. MoM for μ is $\bar{Y} = \hat{\theta}$ (this is unbiased). The variance of $\bar{Y}$ is σ^2/n.

- Bias and variance going to zero as $n \to \infty$ is GOOD
- If it doesn't, that's VERY BAD
- The SE of $\bar{Y}$ is $\sigma/\sqrt{n}$
- Note that we can also use the *same data* to estimate the error as $\hat{\text{SE}} = \hat{\sigma}/\sqrt{n}$
- We often need more than the first two moments
- Note $\hat{\theta}$ is asymptotically Normal by CLT
- And $\hat{\theta}$ is consistent by LLN

CLT

Reminder: Expression for CLT. d or D indicates convergence in distribution.

$$\sqrt{n}(\bar{Y} - \mu) \overset{d}{\sim} \mathcal{N}(0, \sigma^2)$$

Equivalently,

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \overset{d}{\sim} \mathcal{N}(0, 1),$$

which may be preferable to use the PDF ϕ or CDF Φ of the standard Normal.

- How large does n have to be?
- $n \geq 30$
  - This may trigger mathematicians
  - Statisticians stay winning, though
  - For typical examples in the real world
    - BAD THINGS may happen for distributions with heavy tails (e.g., Cauchy)
    - Though most cases eventually fall to the might of the CLT

We can also use the statement

$$\bar{Y} \dot{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

which is NOT the same as the above. This approximate distribution is NOT a limit statement; the n appears on the RHS, which would not work at all with a limit.

- This is more useful for approximations than the limit statements
- Interpretation: the distribution of $\bar{Y}$ is asymptotically Normal

Thm: Let $Y_1, Y_2, \ldots$ be i.i.d. continuous r.v.s with PDF f, CDF F, and quantile function $Q = F^{-1}$. We have

$$\hat{Q}_n(p) = Y_{(\lceil np \rceil)}, \quad \sqrt{n}(\hat{Q}_n(p) - Q(p)) \overset{d}{\sim} \mathcal{N}\left(0, \frac{p(1-p)}{f(p)^2}\right)$$

Ex: $Y_1, Y_2, \ldots$ is i.i.d. $\mathcal{N}(\theta, \sigma^2)$ and our estimand is θ. Let M_n be the sample median (an order statistic). Which is better?

From the above theorem and CLT, we have

$$M_n \dot{\sim} \mathcal{N}\left(\theta, \frac{\pi\sigma^2}{2n}\right), \quad \bar{Y} \sim \mathcal{N}(\theta, \sigma^2/n).$$

We can see in the approximate distributions that M_n has higher variance (but the same bias). So by MSE M_n would be worse.

- Efficiency vs. robustness
- What if we do not have the underlying Normal distribution?
  - Median may be more robust
- E.g., Cauchy distribution
  - Heavy-tailed
  - No mean
  - Sample median is more robust against this case
    - Sample mean has indeterminate
  - Exercise: prove that the mean of n Cauchys is distributed Cauchy, not Normal

Forms of convergence

- Convergence... (in descending order of strength)
  - Almost surely (not covered in this lecture)
  - In probability
  - In distribution

Def: X_n converges to X in **probability** if for any $\epsilon > 0$, we have $P(\mid X_n - X \mid \geq \epsilon) \to 0$ as $n \to \infty$.

CMT

Thm (Continuous mapping theorem): If g is continuous and $X_n \to X$, then $g(X_n) \to g(X)$ in the same form of convergence.

- Convergence in probability is stronger than convergence in distribution
- $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$
  - The converse is **NOT TRUE** unless X is a constant (proven in Stat 210)

Slutsky's theorem

Thm (Slutsky's theorem): Suppose $X_n \xrightarrow{d} X, Y_n \xrightarrow{d} Y$.

Does $X_n + Y_n \xrightarrow{d} X + Y$? No, not generally. For example, let X_n, Y_n be i.i.d. $\mathcal{N}(0,1)$ where $X = Y \sim \mathcal{N}(0,1)$. The LHS converges to $\mathcal{N}(0,2)$ and the RHS is $\mathcal{N}(0,4)$.

Now, suppose that $Y_n \overset{d}{\to} c$ where c is a constant. We have

$$\begin{aligned}
X_n + Y_n &\overset{d}{\to} X + c \\
X_n - Y_n &\overset{d}{\to} X - c \\
X_n Y_n &\overset{d}{\to} cX \\
X_n / Y_n &\overset{d}{\to} X/c : c \neq 0, Y_n \neq 0
\end{aligned}$$