Models

Statistical Model

A set of probability distribuions (typically a strict subset of all probability distributions)

This can be specified as either a pdf, cdf, or a group of known of models (i.e. normal distributions)

- We will refer to a general model as P_θ where $\theta \in \Theta$
  - Θ is the parameter space

Parametric vs Nonparametric Models

A **parametric** model has a parameter space with finite dimension while a **nonparametric** model does not

Examples of parametric:

- Normal distributions
- Polynomials of degree at most d

Examples of nonparametric:

- Any polynomial

Point Estimation

Point Estimate

A **point estimate** $\hat{\theta}$ is a single best guess for a parameter θ

The notation $\theta \rightsquigarrow \hat{\theta}$ is used to denote that θ is estimated by $\hat{\theta}$

This can be written with as $\hat{\theta}_n$ to indicate the dependence on n datapoints

Properties of Estimators

Bias

$$\mathrm{bias}(\hat{\theta}) = \mathbb{E}\left[\hat{\theta}\right] - \theta$$

- The expectation is taken over all possible θ

- An estimator is **unbiased** if the bias is 0
- It is **asymptotically unbiased** if the bias approaches 0 as $n \to \infty$

Standard Error

$$\mathrm{se}(\hat{\theta}) = \sqrt{\mathrm{Var}\left[\hat{\theta}\right]}$$

Mean Squared Error

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \mathrm{bias}(\hat{\theta})^2 + \mathrm{se}(\hat{\theta})^2$$

- The second half can be derived by subtracting and adding $\mathbb{E}\left[\hat{\theta}\right]$ inside the parentheses, rewriting in terms of bias, and then expanding
- Crucially, $\mathbb{E}\left[\hat{\theta} - \mathbb{E}\left[\hat{\theta}\right]\right] = 0$

Consistency

An estimator $\hat{\theta}$ is **consistent** if $\hat{\theta} \xrightarrow{\mathbb{P}} \theta$ as $n \to \infty$

- The LLN is one way of showing consistency
  - I.e. that the sample mean is a consistent estimator of μ
- Continuous mapping theorem also helps
  - Shows that $\overline{X_n}^2$ is a consistent estimator of μ^2

Showing Consistency with MSE

If MSE approaches 0 as $n \to 0$, then the estimator is consistent

Asymptotically Normal

Asymptotically Normal

An estimator $\hat{\theta}$ is **asymptotically normal** if there is an **asymptotic variance** σ^2 such that:

$$\sqrt{n}\left(\hat{\theta} - \mu\right) \rightsquigarrow \mathcal{N}(0, \sigma^2)$$

- Can then think of $\hat{\theta}$ as:
  - $\hat{\theta} = \theta + \frac{\sigma}{\sqrt{n}} Z$
  - Z is standard normal
- CLT can be used to prove sample means are asymptotically normal with variance equal to the variance of the original distribution
- Delta method can be used to show functions of the sample mean are also asymptotically normal
  - Tells us that the asymptotic variance is equal to the variance of the underlying distribution multiplied by $g'(\mu)^2$

Alternative Definition

Alt Def for Asymptotically Normal

An estimator is **asymptotically normal** if

$$\hat{\theta} - \theta \rightsquigarrow \mathcal{N}(0, \mathrm{se}(\hat{\theta}))$$

It is often difficult to compute the standard error of our estimator exactly than to derive the asymptotic variance

- As a result, we prefer our previous definition

Confidence Intervals

Confidence Interval

A $1 - \alpha$ confidence interval for θ is a random interval of the form $C_n = (A_n, B_n)$ such that:

$$P(\theta \in (A_n, B_n)) \geq 1 - \alpha$$

$1 - \alpha$ is the coverage of

- This means that with $\alpha = 0.05$, we expect that 95% of the time we construct our confidence interval, it contains the true value of θ
- Confidence intervals typically look like $\hat{\theta} \pm c_\alpha$

Constructing

- If $\hat{\theta}$ is asymptotically normal, then we can construct an interval using the normal distribution
- We set $c_\alpha = \frac{\sigma}{\sqrt{n}} = z_{\alpha/2}$
- $P(Z \geq z_{\alpha/2}) = \alpha/2$

- However, σ typically depends on θ, which we don't know!
- We can use the asymptotic variance of the estimator instead
- This means our confidence interval only works asymptotically (as n increases)
- Formally, this means that the confidence interval we construct only has an **asymptotic coverage**