Introduction to Inference

Different types of inference model/methods:

- Nonparametric inference
- Parametric inference: Frequentist inference; Bayesian inference

Different types of inferential problems:

- point estimation
- confidence sets
- hypothesis testing

Parametric and Non-parametric models

A statistical model $\mathcal{F}$ represents a collection of possible distributions.

Parametric models can be represented by a finite number of parameters. Generally we consider a family of distributions indexed by those parameters, e.g.

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), \quad i = 1, \ldots, n$$

Use θ to indicate an arbitrary parameter. Using θ as a subscript, e.g. $P_\theta(Y \in A)$, emphasizes that F_Y depends on θ.

Nonparametric models require an infinite number of parameters. They're sometimes called "distribution free" to indicate that we make few restrictions on the family of distributions.

Frequentist and Bayesian Statistics

Frequentist statistics

- Interprets probability in terms of long-run frequencies of events.
- Treats parameters as unknown, fixed constants.
- Focuses on point estimation, confidence intervals, and hypothesis tests.

Bayesian statistics

- Interprets probability as representing degree of belief.
- Makes probability statements about parameters, reflecting beliefs.
- Bases all inference on the posterior distribution, which we can summarize in various ways.

Inferential Problems (I): Point Estimation

A statistic is any function of the data. A point estimator $\hat{\theta}_n$ is a function of the data intended to provide a single "best guess" of parameter θ.

We call $\hat{\theta}(X_1, \ldots, X_n)$ (the r.v.) an *estimator*, while we call $\hat{\theta}(x_1, \ldots, x_n)$ (the realization) an *estimate*. We use $\hat{\theta}_n$ or $\hat{\theta}$ for both.

Warning! Be careful not to confuse the distribution of X with the distribution of $\hat{\theta}_n$, called the sampling distribution. For example

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

implies a sampling distribution of

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

There are many ways to evaluate and compare estimators, which we'll discuss more formally when we come to decision theory. For now, a few properties to consider are

- Bias: $bias(\hat{\theta}_n) = E_\theta[\hat{\theta}_n] - \theta$
  We say $\hat{\theta}_n$ is unbiased if its bias is zero.

- Standard error: $se(\hat{\theta}_n) = \sqrt{V_\theta(\hat{\theta}_n)}$

- Mean squared error:

$$\begin{aligned} MSE(\hat{\theta}_n) &= E_\theta[(\hat{\theta}_n - \theta)^2] \\ &= bias^2(\hat{\theta}_n) + V_\theta[\hat{\theta}_n] \end{aligned}$$

Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$ and let $\hat{\lambda}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Find the bias, standard error, and MSE of this estimator.

- Consistency: If $\hat{\theta}_n \xrightarrow{P} \theta$, we say $\hat{\theta}_n$ is (weakly) consistent.

  You should have learned that when $X_1, X_2, \ldots$ are iid with $E[X_1] = \mu$ and $V[X_2] = \sigma^2 < \infty$, $\bar{X}_n$ is consistent for μ and S_n^2 is consistent for σ^2. (If not, please read a probability book for the relevant material.)

- Asymptotic normality:

$$\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} N(0,1)$$

  Note that Slutsky's theorem often lets us replace $se(\hat{\theta}_n)$ by some (weakly) consistent estimator $\hat{\sigma}_n$.

Inferential Problems (II): Confidence Sets

A $1-\alpha$ confidence interval for θ is an interval C_n computed from the data such that $P_\theta(\theta \in C_n) \geq 1-\alpha$ for all θ.

$1-\alpha$ is called the coverage of the interval.

Note that the probability statement is about C_n, not θ, which is fixed. To emphasize this, we could write $P(C_n \ni \theta) \geq 1-\alpha$ for all θ.

Suppose $\hat{\theta}_n \approx N(\theta, \hat{\sigma}_n^2)$. Then we can form an approximate $1-\alpha$ confidence interval for θ of

$$C_n = \hat{\theta}_n \pm z_{\alpha/2}\hat{\sigma}_n,$$

where $z_{\alpha/2}$ is chosen such that $P(Z > z_{\alpha/2}) = \alpha/2$ for $Z \sim N(0,1)$.

Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$. Show how we can use this sample to construct an approximate 95% confidence interval for λ.

Inferential Problems (III): Hypothesis Testing

A hypothesis test is a way of evaluating evidence against some default theory, called the null hypothesis. We construct a function of the data called a **test statistic** and consider its sampling distribution, taking an "extreme" value of the test statistic as evidence against the null hypothesis.

In the Neyman-Pearson framework for hypothesis testing, this takes the form of a decision rule:

- If the test statistic exceeds a predetermined threshold, reject the null hypothesis.
- Otherwise, do not reject the null hypothesis.

The tests can be evaluated in terms of the four possible outcomes (null hypothesis true or false; reject or retain null hypothesis) that can occur.