Lecture 4

Statistical Inference on the Sample Mean

Oscar Volpe

10/6/2021

1. Hypothesis Testing
   - Frequentism
   - Test Statistics
   - Distributional Properties

2. Confidence Regions
   - Finite Sample Coverage
   - Asymptotic Confidence Intervals

Frequentist versus Bayesian Inference

Frequentist Approach

To test hypotheses, one must model what has not occurred.

- Verify/reject hypotheses about a model that is assumed to be *true*.
- Must design your *experiment* and *stopping rule* before you test

"No isolated experiment, however significant in itself, can suffice for the experimental demonstration of any natural phenomenon; for the 'one chance in a million' will undoubtedly occur, with no less and no more than its appropriate frequency, however surprised we may be that it should occur to us."

-Ronald Fisher, 1935

Bayesian Approach

Use prior information in conjunction with new data in your sample.

- All model parameters are assumed to be random variables.
- Start with a *prior*, and then update your beliefs based on likelihoods.

Null Hypotheses & Alternative Hypotheses

Let θ be an unknown parameter in the population. Suppose you wish to test whether θ equals or lies above/below some value using data $\{X_i\}_{i=1}^{n}$. You write down a *null hypothesis* (H_0) and an *alternative hypothesis* (H_1).

- *Example 1 (One-Sided Test).* $H_0 : \theta \leq (\geq) c$ and $H_1 : \theta > (<) c$.
- *Example 2 (Two-Sided Test).* $H_0 : \theta = c$ and $H_1 : \theta \neq c$.

Idea

Given data $\{X_i\}_{i=1}^{n}$, is there sufficient evidence to reject H_0 in favor of H_1?

- If so, then we can "reject" the null hypothesis.
- If not, then we "fail to reject" the null hypothesis.
  - Importantly, we do not "accept" H_0. There is just not enough evidence to rule out the possibility of H_0. Think of H_0 as signifying "no effect".

Significance Level

Given this setup, we are susceptible to two types of errors:

- *Type-I Error:* reject the null hypothesis when it is true
- *Type-II Error:* fail to reject the null hypothesis when it is false

	Reject H_0	Fail to Reject H_0
H_0 True	Type-I Error	Correct
H_0 False	Correct	Type-II Error

The consensus is generally that false positives are worse than false negatives, i.e. that *Type-I Error* is typically worse than *Type-II Error*.

Definition (Significance Level)

The *significance level* of a test of H_0 against H_1 is the probability of incorrectly rejecting H_0, and it is denoted by $\alpha = P(\text{reject } H_0 | H_0 \text{ true})$.

Constructing a Test

Let β denote the probability of failing to reject H_0 when H_1 is true. We say that a test has "high power" if there is a small probability of *Type-II Error*.

Definition (Power of a Test)

The *power* of a test of H_0 against H_1 is the probability of rejecting a false H_0 under a specific alternative H_1, i.e. $\pi = 1 - \beta = P(\text{reject } H_0 | H_1 \text{ true})$

We restrict our attention to tests of the form $\phi_n = \mathbb{I}\{T_n > c_n\}$.

- T_n is our *test statistic* (constructed from data)
- c_n is the *critical value* (our notion of "large")

Intuition: if the test statistic T_n is larger than some critical value c_n, then we reject the null hypothesis in favor of H_1; otherwise, we fail to reject H_0.

Rejection Rules

Definition (Test Statistic)

Let θ be an unknown parameter, and let $\{X_i\}_{i=1}^n$ be a sample. A *test statistic* T_n is a function of $\{X_i\}_{i=1}^n$ used to test an hypothesis about θ.

We "reject" H_0 when $T_n > c_n(\alpha)$. Otherwise, we "fail to reject" H_0.

Definition (p-value)

The *p-value* is the smallest significance level α at which the null hypothesis would be rejected, i.e. $\hat{p}_n = \inf\{\alpha \in (0,1) : T_n > c_n(\alpha)\}$.

- In other words, the p-value is the probability, under H_0, that a future experiment would produce a test statistic value that is *at least at extreme* as that which is observed in the current experiment.
- Small $\hat{p}_n$ implies that such an extreme outcome is unlikely under H_0.

Using Asymptotic Theory for Testing

Consider an i.i.d. sample $\{X_i\}_{i=1}^n$ with mean μ, variance σ^2, and n "large". By the Central Limit Theorem and Slutsky's theorem, we write:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \quad \Rightarrow T_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \xrightarrow{d} N(0, 1)$$

We let $\Phi(\cdot)$ denote the cumulative distribution function of $Z \sim N(0, 1)$. Since the limiting distribution of T_n is $N(0, 1)$, we can write:

$$F_{T_n}(z_{1-\alpha}) = P\Big(\frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \leq z_{1-\alpha}\Big) \to P(Z \leq z_{1-\alpha}) = \Phi(z_{1-\alpha}) = 1 - \alpha,$$

where $z_{1-\alpha} = \Phi^{-1}(1 - \alpha)$ is the $(1 - \alpha)$th quantile of a standard normal.

- For "large" samples, we can use the $N(0, 1)$ distribution as an approximation for the distribution of T_n (see next two slides).

Example 1: One-Sided Test for $E(X) = \mu$

Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} P$ on $\mathbb{R}$ and $\sigma^2 = Var(X_i) < \infty$. We seek to test:

$$H_0 : \mu \leq c \text{ versus } H_A : \mu > c$$

By the Central Limit Theorem and Slutsky's theorem, we write:

$$\sqrt{n}(\bar{X}_n - c) \xrightarrow{d} N(0, \sigma^2) \quad \Rightarrow \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} \xrightarrow{d} N(0, 1),$$

under the null hypothesis H_0. Therefore, we have:

$$T_n = \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} \xrightarrow{d} N(0, 1) \quad \text{and} \quad c_n = \Phi^{-1}(1 - \alpha) = z_{1-\alpha}$$

Thus, we define our test to be: $\phi_n = \mathbb{I}\{T_n > c_n\}$.

Example 2: Two-Sided Test for $E(X) = \mu$

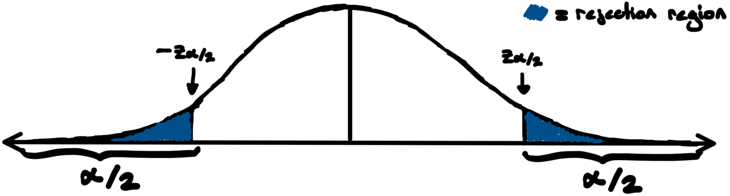
Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$ on $\mathbb{R}$ and $\sigma^2 = Var(X_i) < \infty$. We seek to test:

$$H_0 : \mu = c \text{ versus } H_A : \mu \neq c$$

Following the same steps as before, under the null hypothesis H_0, we have:

$$T_n = \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} \xrightarrow{d} N(0,1) \quad \text{and} \quad c_n = \Phi^{-1}(1 - \frac{\alpha}{2}) = z_{1-\frac{\alpha}{2}}$$

Thus, we define our test to be: $\phi_n = \mathbb{I}\{|T_n| > c_n\}$.



Computing p-values

The p-value for a one-sided test is:

$$\begin{aligned}\hat{p}_n &= \inf\{\alpha \in (0,1) : \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} > z_{1-\alpha}\} \\ &= \inf\{\alpha \in (0,1) : \alpha > 1 - \Phi(\frac{\sqrt{n}(\bar{X}_n - c)}{s_n})\} \\ &= 1 - \Phi(\frac{\sqrt{n}(\bar{X}_n - c)}{s_n})\end{aligned}$$

The p-value for a two-sided test is:

$$\begin{aligned}\hat{p}_n &= \inf\{\alpha \in (0,1) : \frac{\sqrt{n}(|\bar{X}_n - c|)}{s_n} > z_{1-\frac{\alpha}{2}}\} \\ &= 2\Big[1 - \Phi(\frac{\sqrt{n}(|\bar{X}_n - c|)}{s_n})\Big]\end{aligned}$$

A Multidimensional Hypothesis

Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$ on $\mathbb{R}^k$ and let $\Sigma < \infty$ be the $k \times k$ (invertible) variance-covariance matrix. We want to test the following hypothesis:

$$H_0 : \mu = \mathbf{0} \text{ versus } H_A : \mu \neq \mathbf{0}$$

By the CLT and the CMT, we write:

$$\begin{aligned}\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} Z \sim N(0, \Sigma), \quad &\text{where } Z'\Sigma^{-1}Z \sim \chi^2_k \\ &\Rightarrow n(\bar{X}_n - \mu)'\Sigma^{-1}(\bar{X}_n - \mu) \sim \chi^2_k \\ &\Rightarrow n(\bar{X}_n - \mu)'\hat{\Sigma}^{-1}(\bar{X}_n - \mu) \sim \chi^2_k,\end{aligned}$$

where $\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^n [X_i - \bar{X}_n][X_i - \bar{X}_n]'$ and $\hat{\Sigma} \xrightarrow{P} \Sigma$. Then, under H_0:

$$T_n = n(\bar{X}_n)'\hat{\Sigma}^{-1}(\bar{X}_n) \xrightarrow{d} \chi^2_k$$

Let $\phi_n = \mathbb{I}\{T_n > c_n\}$, where $c_n = c_{k,1-\alpha}$ is the $(1-\alpha)$th quantile of χ^2_k.

Student t-Distributions

For smaller samples, these normal approximations are often inadequate. When $\{X_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ and n is small, then the test statistic T_n follows a student t-distribution with $n-1$ degrees of freedom:

$$T_n = \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} \sim t_{n-1}$$

We set the critical value to $c_n = t_{n-1,1-\alpha}$, and we reject when $T_n > c_n$.

- We can compute p-values similarly, using t_{n-1} instead of $N(0,1)$.
- The difference between t_{n-1} and $N(0,1)$ vanishes as $n \to \infty$.

1. Hypothesis Testing
   - Frequentism
   - Test Statistics
   - Distributional Properties

2. Confidence Regions
   - Finite Sample Coverage
   - Asymptotic Confidence Intervals

Coverage Probabilities

Definition (Confidence Interval)

A $(1-\alpha)$-level *confidence interval* for θ is the set of values θ^* for which the null hypothesis $H_0 : \theta = \theta^*$ is not rejected at significance level α.

For finite samples, choose the confidence interval C_n so that:

$$P(\theta \in C_n) \geq 1 - \alpha$$

We call $P(\theta \in C_n)$ the *coverage probability* of C_n.

- We want C_n as small as possible so that this inequality holds, given α.
  - Exact coverage occurs when $P(\theta \in C_n) = 1 - \alpha$.
- Same idea generalizes for multidimensional θ (i.e. *confidence regions*).
- A $(1-\alpha)$-level *asymptotic confidence interval*: $P(\mu \in C_n) \to 1 - \alpha$.

Example: Exact Finite Sample Coverage

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$. We want to compute an interval C_n with coverage probability $P(\mu \in C_n) = 1 - \alpha$. We know that:

$$T_n = \frac{\sqrt{n}(\bar{X}_n - c)}{s_n} \sim t_{n-1}$$

Let us define the confidence interval to be:

$$C_n = \left[\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}\right]$$

We can now show that:

$$P(\mu \in C_n) = P\left(\frac{\sqrt{n}|\bar{X}_n - \mu|}{s_n} \leq t_{n-1,1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Example: An Asymptotic Confidence Interval

Suppose $X_1, \ldots, X_n \in \mathbb{R}$ is an i.i.d. sample. We want to compute C_n so that $P(\mu \in C_n) \to 1 - \alpha$ as $n \to \infty$. By the CLT and Slutsky's theorem:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \quad \Rightarrow \frac{\sqrt{n}(\bar{X}_n - \mu)}{s_n} \xrightarrow{d} N(0,1)$$

Let us define the following terms:

$$C_n = [\bar{X}_n - z_n, \bar{X}_n + z_n], \quad \text{where } z_n = z_{1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}$$

We can now show that:

$$P(\mu \in C_n) = P(\bar{X}_n - z_n \leq \mu \leq \bar{X}_n + z_n) = P\left(|\bar{X}_n - \mu| \leq z_{1-\frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right)$$

$$= P\left(\frac{\sqrt{n}|\bar{X}_n - \mu|}{s_n} \leq z_{1-\frac{\alpha}{2}} \right) \to P(|Z| \leq z_{1-\frac{\alpha}{2}}) = 1 - P(|Z| > z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$