Statistical Inference

Lecture 10a

ANU - RSFAS

Last Updated: Mon May 8 13:59:24 2017

Hypothesis Testing

- So far we have focused on statistical point and interval estimation.
- In many situations, however, the simple estimation of a population characteristic is not the final desired outcome of a statistical analysis.
- We may want to use our estimates to decide whether some previously proposed theory or statement regarding the population of interest is actually true (or at least is plausible given the information provided by the observations at hand).
- This is, of course, the standard framework of statistical hypothesis testing which is familiar from any introductory unit in basic statistics.

Hypothesis Testing

- Consider the following situation:
  - Suppose that we have purchased a light-bulb based on its advertised claim that the mean lifetime of such bulbs is at least 1000 hours.
  - If we then observe the lifetime of the actual bulb we purchased, we have some data with which to assess the advertising claim.
  - This simple scenario is precisely the framework of statistical hypothesis testing.
  - Suppose we believe that the lifetime of the population of bulbs in question is exponentially distributed with mean parameter θ

$$p(x|\theta) = \frac{1}{\theta} exp\left(-\frac{x}{\theta}\right) \qquad \text{for } \theta \in \Theta$$

Hypothesis Testing

- We can formulate a hypothesis test as: $H_0 : \theta \geq 1000$
- **Definitions:** Suppose that $X_1, \ldots, X_n$ represent a simple random sample from a parametric family with density function $f(x|\theta)$ for some parameter $\theta \in \Theta$.
  - A statistical hypothesis is simply a subset of the parameter space, Θ.
  - Any statistical hypothesis of interest, often termed the null hypothesis, is associated with a competing alternative hypothesis.
  - A null hypothesis and its alternative form a partition of the parameter space Θ consisting of the sets Θ_0 and

$$\Theta_1 = \Theta_0^c \cap \Theta$$

Hypothesis Testing

Definition: A hypothesis testing procedure or hypothesis test is a rule that specifies:

1. For which sample values the decision is made to accept H_0.
2. For which sample values H_0 is rejected and H_1 is accepted as true.

The subset of the space for which H_0 will be rejected is called the rejection region or critical region (R). The complement of the rejection region is called the acceptance region.

Hypothesis Testing

- For our light bulb example:
  - We can define a test which rejects H_0 if X is less than 1,000 hours.
  - $R = \{X < 1000\}$
- More generally, we can define a statistical test in terms of a rejection region (R) which is just a set for some statistic $T(X_1, \ldots, X_n)$:

$$R = \{X \in \mathcal{X} : T(X) < k\}$$

Type I and Type II Errors

- Common sense would indicate that the test described in the example of the previous section; namely, rejecting the null hypothesis that the mean lifetime of the bulbs is at least 1000 hours based on a single observation being less than 1000 hours, is not a very good test.
- We will make errors.
- Consider the following possibilities:
  - Type I Error: Reject H_0 given that it is true. Thus the observations fall in the rejection region R when in fact that null hypothesis, H_0, is true.
  - Type II Error: Do not Reject H_0 when it is false: Thus the observed data values fall outside the rejection region when in fact the null hypothesis is false.

Type I and Type II Errors

	Decision	
Truth	Accept H_0	Accept H_1
H_0	Correct Decision	Type I Error
H_1	Type II Error	Correct Decision

Type I and Type II Errors

- Probability of a Type I error (α):

$$\begin{aligned} P(R) &= P_{1000}(X < 1000 | H_0 \text{ is true }) \\ &= \int_0^{1000} \frac{1}{1000} exp\left(-\frac{x}{1000}\right) dx \\ &= 1 - exp(-1000/1000) = 0.632 \end{aligned}$$

- What if $\theta = 1500$:

$$\begin{aligned} P(R) &= P_{1500}(X < 1000 | H_0 \text{ is true }) \\ &= \int_0^{1000} \frac{1}{1500} exp\left(-\frac{x}{1500}\right) dx \\ &= 1 - exp(-1000/1500) = 0.077 \end{aligned}$$

Type I and Type II Errors

- Let's determine the probability of a Type II error.
- Note that we specified $H_0 : \theta \geq 1000$. This means:

$$H_1 : \theta < 1000$$

- Picking a specific value in this region ($\theta = 500$), we have:

$$\begin{aligned} P(R^c) &= P_{500}(X > 1000 | H_0 \text{ is false }) \\ &= \int_{1000}^{\infty} \frac{1}{500} exp\left(-\frac{x}{500}\right) dx \\ &= exp(-1000/500) = 0.135 \end{aligned}$$

- There is a strong relationship between Type I and Type II errors. Note that for a given value of θ, only one type of error can occur (since for any given θ, H_0 either is or is not true).

Type I and Type II Errors

Definition: The power function of a hypothesis test with rejection R is the function of θ defined by:

$$\beta(\theta) = P(\boldsymbol{X} \in R)$$

- **Power = 1 - P(Type II Error)**
  $= 1 - P(\boldsymbol{X} \in R^c | H_1 \text{ is true}) = P(\boldsymbol{X} \in R | H_1 \text{ is true})$
- Given that H_1 is true, what is the probability I reject H_0

Type I and Type II Errors

Definition: For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test if:

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$$

Definition: For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test if:

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

- Many authors don't distinguish the two. I will not try to trip you on an exam on this, so don't worry.

Type I and Type II Errors

- Consider the light bulb example:

$$\beta(\theta) \quad = \quad P(X \in R) = P(X < 1000) = 1 - exp(-1000/\theta)$$

- The size of the test determined by $R = \{X < 1000\}$.
- Again recall: $H_0 : \theta \geq 1000$.
- The power function is a decreasing function of θ in this case. So to maximize it we set $\theta = 1000$.

$$\begin{aligned} sup_{\theta \in \Theta_0} \beta(\theta) &= sup_{\theta \geq 1000} 1 - exp(-1000/\theta) \\ &= 1 - exp(-1000/1000) = 0.632 = \alpha \end{aligned}$$

Type I and Type II Errors

- It is standard to focus on test which have sizes 0.05 or 0.01.
- If we focus on tests with rejection regions of the form $R = \{X < k_\alpha\}$, we can choose k_α such that:

$$\begin{aligned} sup_{\theta \in \Theta_0}\ \beta(\theta) &= sup_{\theta \geq 1000} 1 - exp(-k_\alpha/\theta) \\ &= 1 - exp(-k_\alpha/1000) = \alpha \end{aligned}$$

Based on this $k_\alpha = -1000 ln(1-\alpha)$ so at $\alpha = 0.05$ we have:

$$R = \{X < 51.29\}$$

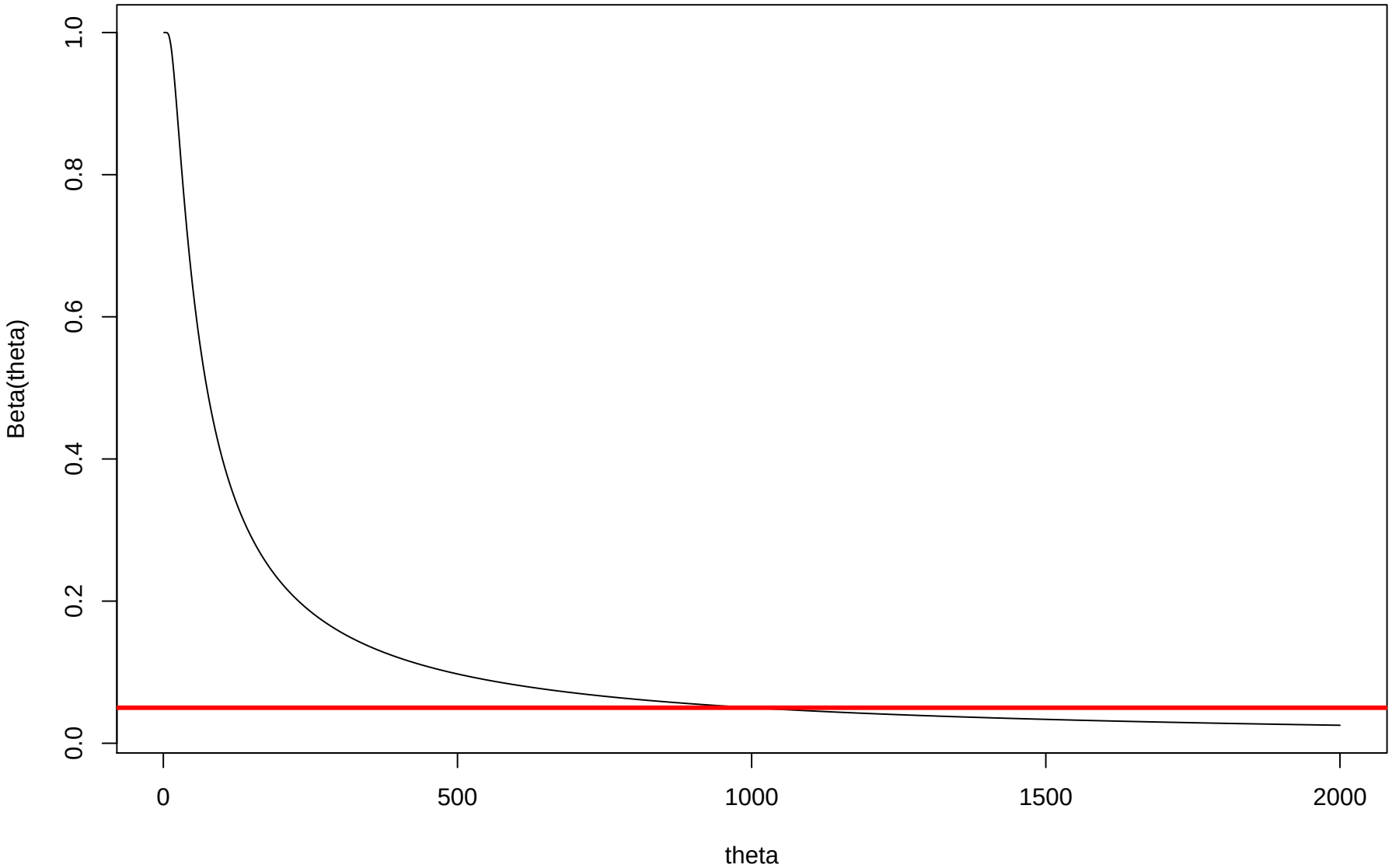
Type I and Type II Errors

- What actually is the power based on this rejection region if truly $\theta = 500 \in \Theta_1$?

$$\begin{aligned}
\beta(500) &= P_{500}(X \in R) \\
&= P_{500}(X < 51.29) \\
&= \int_0^{51.29} \frac{1}{500} exp(-x/500) dx \\
&= 1 - exp(-51.29/500) = 0.0975.
\end{aligned}$$

- So this test has less than a 10% chance of detecting even this drastic departure from the null hypothesis based on $\alpha = 0.05$!!

The Power Function is a Function! $\beta(\theta)$



- Red $= 0.05$

Type I and Type II Errors

- Unfortunately, if our power is not as large as we like, we cannot simply change a rejection region of the form $R = \{X < k\}$ to increase the power without simultaneously affecting the size of our test (α).
- Our task, then, is to find tests (or equivalently rejection regions) of a given size which have the best possible power when $\theta \in \Theta_1$.
- We can increase our sample size (not always possible) or by finding a "good" test statistic.

Essential Nature of a Hypothesis Test (Experimental Design, Hoff 2009)

- Given H_0, H_1 and data $\boldsymbol{x} = \{x_1, \ldots, x_n\}$:

1. From the data, compute a relevant test statistic $T(\boldsymbol{x})$: The test statistic $T(\boldsymbol{x})$ should be chosen so that it can differentiate between H_0 and H_1 in ways that are scientifically relevant. Typically, $T(\boldsymbol{x})$ is chosen so that

$$T(\boldsymbol{x}) \text{ is probably } \begin{cases} \text{small under } H_0 \\ \text{large under } H_1 \end{cases}$$

2. Obtain a null distribution: A probability distribution over the possible outcomes of $T(\boldsymbol{X})$ under H_0. Here, $X = \{X_1, \ldots, X_n\}$ are potential experimental results that could have happened under H_0.

3. Compute the p-value: The probability under H_0 of observing a test statistic $T(\boldsymbol{X})$ as or more extreme than the observed statistic $T(\boldsymbol{x})$.

$$\text{p-value} = P(T(\boldsymbol{X}) \geq T(\boldsymbol{x})|H_0)$$

If the p-value is small $\Rightarrow$ evidence against H_0
If the p-value is large $\Rightarrow$ not evidence against H_0

P-values

- Another way of reporting the results of a hypothesis test is through a certain kind of test statistic called a p-value.

Definition: a p-value $p(\boldsymbol{X})$ is a test statistic satisfying $0 \leq p(\boldsymbol{x}) \leq 1$ for every sample point $\boldsymbol{x}$. Small values of $p(\boldsymbol{X})$ give evidence that H_1 is true.

- A p-value is valid if for every $\theta \in \Theta_0$ and every $0 \leq \alpha \leq 1$

$$P(\text{Reject } H_0 | H_0) = P(p(\boldsymbol{X}) \leq \alpha) \leq \alpha$$

- The p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

Then, $p(\boldsymbol{x})$ is a valid p-value.

P-values

Example: Suppose that $X_1, \ldots, X_n$ are a random sample from a normal distribution with mean μ and unit variance. Consider testing:

$$\begin{aligned} H_0 : &\quad \mu \leq \mu_0 \\ H_1 : &\quad \mu > \mu_0 \end{aligned}$$

- We can show that that $R = \left\{ \frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq c \right\}$. So

$$T(\boldsymbol{X}) = \frac{\bar{X} - \mu_0}{1/\sqrt{n}}$$

$$\text{p-value} = P\left(\frac{\bar{X} - \mu_0}{1/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{1/\sqrt{n}} \right) = P\left(Z \geq \frac{\bar{x} - \mu_0}{1/\sqrt{n}} \right)$$

- The probability, under H_0, of getting the observed test statistic or something more extreme (based on the rejection region).