Quantitative Genomics and Genetics
BTRY 4830/6830; PBSB.5201.01

Lecture 8: Hypothesis testing and final general inference topics

Jason Mezey
jgm45@cornell.edu
Feb. 25, 2016 (T) 8:40-9:55
• CLASS TUES, March 1st IS CANCELLED (I will send out an email announcement later today + post an updated schedule)

• Several typos in Homework #3 (!!) I will send out an email and post a correction
  • 2g: nx2 matrix should be Mx2
  • 2i: there are no commas between numbers (!!)
  • 2i: missing negative

• Reminder: Homework #4 will still be available Tues. and due 11:59PM, March 7

• For NYC folks: office hours today will be in a different room (!!) - I will send out an email today (Ithaca will be the same room as always)
Summary of lecture 8

- Last lecture, we completed our general discussion of estimation
- Today we will (briefly) consider the concept of confidence intervals
- We will also begin our discussion of hypothesis testing
Conceptual Overview

System

Question

Sample

Inference

Prob. Models

Statistics

Assumptions
This concept is often introduced to us as the function 

\[ T(x) \quad Pr(T(X)|\theta) \]

Then

\[ \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \]
Review of essential concepts

• **Inference** - the process of reaching a conclusion about the true probability distribution (from an assumed family of probability distributions indexed by parameters) on the basis of a sample

• **System, Experiment, Experimental Trial, Sample Space, Sigma Algebra, Probability Measure, Random Vector, Parameterized Probability Model, Sample, Sampling Distribution, Statistic, Statistic Sampling Distribution, Estimator, Estimator Sampling distribution**
Confidence intervals I

- For the estimation framework we have considered thus far, our goal was to define an estimator that provides a “reasonable guess given the sample” of the true value of the parameter.

- This is called “point” estimation since the true parameter has a single value (i.e. it is a point).

- We could also estimate an interval, where our goal is to say something about the chances that the true parameter (the point) would fall in the interval.

- **confidence interval** (CI) - an estimate of an interval defined such that if it were estimated individually for an infinite number of samples, a specific percentage of the estimated intervals would contain the true parameter value.

- Don’t worry if this concept seems confusing (it is!) let’s first consider an example and then discuss some basics.
Confidence intervals II

- As an example, assume the standard normal r.v. $X \sim N(0,1)$ correctly describes our sampling distribution if we were to produce 50 independent samples, each of size $n=10$ and we were to estimate a CI for each one, we would expect to get the following:
Confidence intervals III

- A CI is therefore calculated from a sample (and reflects uncertainty!)

- A CI is an estimate of an interval, as opposed to an estimate of a parameter, which is a point estimate (more technically, the CI is an estimate of the endpoints of the interval)

- This estimated interval of a CI (generally) includes the estimate of the parameter in the “middle”

- In general, a CI provides a measure of “confidence” in the sense that the smaller the interval, the more “confidence” we have in our estimate (if this seems circular, it is meant to be!)

- In general, we can make the CI smaller with a larger sample size $n$ and by decreasing the probability that the interval contains the true parameter value, i.e. a 95% CI is smaller than a 99% CI

- NOTE THAT A 95% CI estimated from one sample does not contain the true parameter value with a probability of 0.95 (!!!) - the definition of a CI says if we performed an infinite number of samples, and calculated the CI for each, then 95% of these intervals would contain the true parameter value (strange?)
Estimation and Hypothesis Testing

• Thus far we have been considering a “type” of inference, estimation, where we are interested in determining the actual value of a parameter.

• We could ask another question, and consider whether the parameter is NOT a particular value.

• This is another “type” of inference called hypothesis testing.

• We will use hypothesis testing extensively in this course.
Hypothesis testing I

- To build this framework, we need to start with a definition of hypothesis
- **Hypothesis** - an assumption about a parameter
- More specifically, we are going to start our discussion with a *null hypothesis*, which states that a parameter takes a specific value, i.e. a constant

\[ H_0 : \theta = c \]

- For example, for our height experiment / identity random variable, we have \( Pr(X|\theta) \sim N(\mu, \sigma^2) \) and we could consider the following null hypothesis:

\[ H_0 : \mu = 0 \]
Hypothesis testing II

• Our goal in hypothesis testing is to use a sample to reach a conclusion about the null hypothesis.

• To do this, just as in estimation, we will make use of a statistic (a function on the sample), where recall we know the sampling distribution (the probability distribution) of this statistic.

• More specifically, we will consider the probability distribution of this statistic, assuming that the null hypothesis is true:

\[ Pr(T(X = x | \theta = c)) \]

• Note that this means we have a probability distribution of the statistic given the null hypothesis!!

• We will use this distribution to construct a \textit{p-value}.
Hypothesis testing III

- As example, consider our height experiment (reals as sample space) / identity random variable $X$ / normal probability model $\theta = [\mu, \sigma^2]$ / sample $n=1$ (of one height measurement) / identity statistic $T(x) = x$ (takes the height measured height)

- Let’s assume that $\sigma^2 = 1$ and say we are interested in testing the following null hypothesis $H_0 : \mu = 5.5$ such that we have the following probability distribution of the statistic under the null hypothesis:

![Graph of Pr(T(x) | H0)](image_url)
We quantify our intuition as to whether we would have observed the value of our statistics given the null is true with a \textit{p-value}.

\textbf{p-value} - the probability of obtaining a value of a statistic $T(x)$, or more extreme, conditional on $H_0$ being true.

Formally, we can express this as follows:

\begin{equation}
    pval = Pr(|T(x)| \geq t | H_0 : \theta = c)
\end{equation}

Note that a p-value is a function on a statistic (!!) that takes the value of a statistic as input and produces a p-value as output in the range $[0, 1]$: \( pval(T(x)) : T(x) \rightarrow [0, 1] \)
As an intuitive example, let’s consider a continuous sample space experiment / identify r.v. / normal family / \( n=1 \) sample / identity statistic, i.e. \( T(x) = x \)

Assume we know \( \sigma^2 = 1 \) (is this realistic?), let’s say we are interested in testing the null hypothesis \( H_0 : \mu = 0 \) and let’s say that we assume that if we are wrong the value of \( \mu \) will be greater than zero (why?)
**p-value III**

- Same example: let’s consider a continuous sample space experiment / identify r.v. / normal family / $n=1$ sample / identity statistic, i.e. $T(X) = X$ / assume we know $\sigma^2 = 1$ / we test the null hypothesis $H_0 : \mu = 0$ and let’s assume that if we are wrong the value of $\mu$ could be in either direction (again, why?)
p-value IV

- More technically a p-value is determined not just by the probability of the statistic given the null hypothesis is true, but also whether we are considering a “one-sided” or “two-sided” test.

- For a one-sided test (towards positive values), the p-value is:

\[
pval(T(x)) = \int_{T(x)}^{\infty} Pr(T(x)|\theta = c) \,dT(x)
\]

\[
pval(T(x)) = \sum_{T(x)} Pr(T(x)|\theta = c)
\]

- For a two-sided test, the p-value is:

\[
pval(T(x)) = \int_{-\infty}^{-|\text{median}(T(X))|} Pr(T(x)|\theta = c) \,dT(x) + \int_{|\text{median}(T(X))|}^{\infty} Pr(T(x)|\theta = c) \,dT(x)
\]

\[
pval(T(x)) = \sum_{\text{min}(T(X))}^{-|\text{median}(T(X))|} Pr(T(x)|\theta = c) + \sum_{|\text{median}(T(X))|}^{\text{max}(T(X))} Pr(T(x)|\theta = c)
\]
Hypothesis decisions I

- We use the p-value to make a decision about the null hypothesis.
- Specifically, we use the p-value for our sample to decide whether we “accept” (or better stated: “cannot reject”) the null hypothesis or “reject” the null hypothesis.
- To do this, we use a value $\alpha$ such that if the p-value is below this value we “reject”, if it is above we “cannot reject” (note that this corresponds to a critical value of the statistic $C_{\alpha}$).
- For example for a value $\alpha = 0.05$ we have the following for our previous examples:

$$\alpha = \int_{c_{\alpha}}^{\infty} f_X(x) \, dx$$

$$\alpha = \int_{-\infty}^{-c_{\alpha}} f_X(x) \, dx + \int_{c_{\alpha}}^{\infty} f_X(x) \, dx$$
Hypothesis decisions II

- Note that there are two possible outcomes of a hypothesis test: we reject or we cannot reject.
- We never know for sure whether we are right (!!)
- If we cannot reject, this does not mean H0 is true (why? What if our p-value is 0.99?)
- The value $\alpha$ is called the type I error, the probability of incorrectly rejecting H0 when it is true.
- The value $1 - \alpha$ is the probability of making a correct decision not to reject H0.
- Note that we can control the level of type I error because we decide on the value of $\alpha$.
Assume H0 is correct (!): $\mu = 0$

Sample I:
$T(x) = -0.755$
$p = 0.45$

Sample II:
$T(x) = 2.8$
$p = 0.0025$

One-sided test

Two-sided test
Results of hypothesis decisions I: when H0 is correct (!!)

- There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject

<table>
<thead>
<tr>
<th>H0 is true</th>
<th>1-(\alpha), (correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject (H_0)</td>
<td>(1-\alpha), (correct)</td>
</tr>
<tr>
<td>reject (H_0)</td>
<td>(\alpha), type I error</td>
</tr>
</tbody>
</table>

\[
Pr(T(x) \mid H0)
\]

\[
T(x)
\]
Results of hypothesis decisions I: when H0 is correct (!!)

- There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject

<table>
<thead>
<tr>
<th>Decision</th>
<th>H₀ is true</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject H₀</td>
<td>1-α, (correct)</td>
</tr>
<tr>
<td>reject H₀</td>
<td>α, type I error</td>
</tr>
</tbody>
</table>

\( \Pr(T(x) \mid H₀) \)

\( c_α = 1.64 \)
Results of hypothesis decisions I:
when H0 is correct (!!)

- There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject.

<table>
<thead>
<tr>
<th></th>
<th>H0 is true</th>
<th>H0 is not true</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject H0</td>
<td>1-α, (correct)</td>
<td>α, type 1 error</td>
</tr>
<tr>
<td>reject H0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Graph](image)
Assume H0 is wrong (!): \( \mu = 3 \)

**one-sided test**

\[ \alpha = 0.05 \]

\[ c_\alpha = 1.64 \]

\( p = 0.77 \)

**Sample I:**

\( T(\mathbf{x}) = -0.755 \)

\( p = 0.0025 \)

\[ c_\alpha = 1.96 \]

**Sample II:**

\( T(\mathbf{x}) = 2.8 \)

\( p = 0.005 \)

\( -c_\alpha \)
There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject.

<table>
<thead>
<tr>
<th>Decision</th>
<th>$H_0$ is true</th>
<th>$H_0$ is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject $H_0$</td>
<td>$1 - \alpha$, (correct)</td>
<td>$\beta$, type II error</td>
</tr>
<tr>
<td>reject $H_0$</td>
<td>$\alpha$, type I error</td>
<td>$1 - \beta$, power (correct)</td>
</tr>
</tbody>
</table>
Results of hypothesis decisions II: when H0 is wrong (!!)

- There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject.

<table>
<thead>
<tr>
<th></th>
<th>H₀ is true</th>
<th>H₀ is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject H₀</td>
<td>1-α, (correct)</td>
<td>β, type II error</td>
</tr>
<tr>
<td>reject H₀</td>
<td>α, type I error</td>
<td>1 - β, power (correct)</td>
</tr>
</tbody>
</table>

\[ \Pr(T(x) \mid H₀) \] for a given value of \( x \) that could take values from (+∞, −∞).

\[ T(x) \] is how we assess our null hypothesis. However, this is still does not provide us a guideline quite small. Can we interpret this as evidence against H₀?

\[ p \text{-value} \] is a function of our statistic. If our statistic is (equal to) or greater than a particular value, this is an example of a one-sided test. The various critical concepts in hypothesis tests have a close relationship with p-values.

\[ \text{results of hypothesis decisions II} \] (see diagram on board for an example). Also, note in this particular case:

\[ H₀ \] is true

\[ H₀ \] is false

\[ 0 \]

\[ c \]

\[ 1 - \alpha \]

\[ \beta \]

\[ 1 - \beta \]

\[ \alpha \]

\[ \beta \text{, type II error} \]

\[ 1 - \beta \text{, power (correct)} \]

\[ p \text{-value} \]

\[ \text{results of hypothesis decisions II} \] (see diagram on board for an example). Also, note in this particular case:

\[ H₀ \] is true

\[ H₀ \] is false

\[ 0 \]

\[ c \]

\[ 1 - \alpha \]

\[ \beta \]

\[ 1 - \beta \]

\[ \alpha \]

\[ \beta \text{, type II error} \]

\[ 1 - \beta \text{, power (correct)} \]

\[ p \text{-value} \] is a function of our statistic. If our statistic is (equal to) or greater than a particular value, this is an example of a one-sided test. The various critical concepts in hypothesis tests have a close relationship with p-values.
Results of hypothesis decisions II: when H0 is wrong (!!)

- There are only two possible decisions we can make as a result of our hypothesis test: reject or cannot reject

<table>
<thead>
<tr>
<th></th>
<th>$H_0$ is true</th>
<th>$H_0$ is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>cannot reject $H_0$</td>
<td>$1 - \alpha$, (correct)</td>
<td></td>
</tr>
<tr>
<td>reject $H_0$</td>
<td>$\alpha$, type I error</td>
<td>$\beta$, type II error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1 - \beta$, power (correct)</td>
</tr>
</tbody>
</table>

[Diagram showing probability distributions and critical value]
To do this, we define a function $T(x) = \Pr(T(X) | \theta)$. This concept is often introduced to us as a procedure for generating a random sample from a population.

A sample of size $n$ is given by $[X_1 = x_1, \ldots, X_n = x_n]$, and the probability of this sample is $\Pr([X_1 = x_1, \ldots, X_n = x_n])$. This is the sampling distribution of the sample.

A random variable is a function $X = x$, and the probability of $X$ is $\Pr(X)$. The random variable $X$ is a function of the experiment $\mathcal{X}$, which is a mathematical operator that takes an input and produces an output.

The set of all possible outcomes of the experiment is denoted by $\Omega$, and the set of all possible events is denoted by $\mathcal{F}$. The probability of an event $\mathcal{E}$ is denoted by $\Pr(\mathcal{E})$. The parameter space is denoted by $\Theta$.
Hypothesis Tests

\[ H_0 : \theta = c \quad \theta \in \Theta \]

\[ T(x) \quad Pr(T(X)|H_0 : \theta = c) \]

\[ [X_1 = x_1, \ldots, X_n = x_n] \quad Pr([X_1 = x_1, \ldots, X_n = x_n]) \]

Sample of size \( n \)

Sampling Distribution

\[ X = x \quad Pr(X) \]

Random Variable

\[ \mathcal{X} \]

\[ X(\omega), \omega \in \Omega \]

\[ Pr(\mathcal{F}) \]

Experiment

\[ \Omega \]

\[ \mathcal{F} \]
That’s it for today

• Next week: hypothesis testing II (our last general probability and statistic lecture!)