Probabilistic Modeling and Statistical Computing Fall 2015

October 27, 2015

Hypothesis Testing

Given data (experimental results or observations) and a possible statistical model, are the data **compatible with the model**?

Is the model **plausible**, given the data?

Often, want to examine whether the data show an **effect.**

Often, the statistical model contains a **parameter** θ that describes an effect.

Example: Mice in maze

Null and Alternative

Null Hypothesis H_0

There is no real effect. Observed effects are due to chance. Parameter: $\theta = \theta_0$.

Alternative Hypothesis H_0

There is an effect. Parameter:

- $\theta = \theta_A \neq \theta_0$ (simple alternative)
- $\theta > \theta_0$ (one-sided alternative, composite alternative)
- $\theta \neq \theta_0$ (two-sided, composite)

Mouse Example

Parameters: μ_c = mean time for control population, μ_t = mean time for treatment population. **These are unknown.**

Null hypothesis: No effect. $\mu_c = \mu_t$ or with $\theta = \mu_c - \mu_t$

$$\boxed{\theta = \theta_0 = 0}$$

Alternative Hypothesis: Treated mice are faster. $\mu_c > \mu_t$ or with $\theta = \mu_c - \mu_t$

$$\boxed{\theta > 0}$$

Test statistics

Definition

A test statistic is a function of a sample. *Could be vector valued.*

Possible test statistic: Sample means.
Observed values:

$$\bar{x}_c = 28, \quad \bar{x}_t = 19.9$$

Could also use difference of sample means.

Here: $\bar{x}_c - \bar{x}_t > 0$. Is this evidence for $\theta_c - \theta_t > 0$?

Permutation Model

Explanatory variables are "control" and "treatment". **Response variables** are times in maze.

Null hypothesis becomes: Observed responses have nothing to do with explanatory variables. 10 "control" labels and 14 "treatment" labels have been assigned randomly to 24 mice. **Uniform distribution: All these assignments are equally likely.**

Permutation Model

Alternative hypothesis becomes: Observed responses have something to do with explanatory variables. 10 "control" labels and 14 "treatment" labels could have been assigned randomly to 24 mice with **non-uniform distribution: Faster mice tend to have "treatment" label.**

There are $\boxed{24! = 6.2 \cdot 10^{23}}$ permutations of 24 observed times. *Avogadro's number*

Sampling Distribution

Let's assume the null hypothesis is true.

- Permute the 24 observed times.
- Assign the label "control" to the first 10 times and the label "treatment" to the other 14 times.
- Compute the test statistic $\bar{x}_c - \bar{x}_t$ for this permutation.
- *Do this for all* $24!$ *permutations to obtain a distribution of the test statistics*

Null Distribution and *p* value

Sometimes we can approximate or compute the distribution of the test statistics with a simulation or with a formula. Usually easier if a specific H_0 is assumed.

*How likely is the observed test statistics, **given no effect**?*

The **null distribution** is the distribution of the test statistics if H_0 is true. The *p*-**value** is the probability that the observed test statistic or something more extreme occurs under the null distribution.

Example: Verizon Repair Data

- Verizon repairs phones of its own customers.
- Verizon must also repair phones of customers of competing companies.
- Verizon will be fined if repair times for customers of competing companies are substantially worse.
- Data set with repair times for 1664 of its own customers and 23 customers of others

Verizon Repair Data

- Load the data
- Histograms and boxplots
- Are the means different?
- How likely is this due to chance?

Other test statistics

- Are the median repair times different? Test statistic: median
- How about trimmed means?
- *If in doubt, use several different test statistics and compare the answers.*

One-Way Table

Birth months for players in the 1990 soccer World Cup

Aug-Oct	Nov-Jan	Feb-April	May-July	Total
150	138	140	100	528

Are birth months distributed uniformly over the year?

Null hypothesis: "uniform" distribution.
Technically, this is a multinomial distribution with $n = 528$ and $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$.

Observed and Expected Cell Counts

We have 4 observations in 4 "cells".
Expected count in each cell: $n \cdot p_i = \frac{528}{4} = 132$.
Deviations from expected counts in each cell:

	Aug-Oct	Nov-Jan	Feb-April	May-July
Obs	150	138	140	100
Exp	132	132	132	132
Dev	18	6	8	-32

The deviations add to 0.

Test Statistic

K. Pearson ($\sim$ 1900) proposed

$$X^2 = \sum_{cells} \frac{(observed - expected)^2}{expected}$$

This is a combination of absolute difference and relative difference between observed and expected counts in each cell. Here, $X^2 = 10.97$

	Aug-Oct	Nov-Jan	Feb-April	May-July
Obs	150	138	140	100
Dev	18	6	8	-32
Cell	2.45	0.27	0.48	7.76

Permutation Test

What is the null distribution of X^2? How likely is a value like $X^2 = 10.97$ or larger?

Simulate many observations from a multinomial distribution with $n = 528, p_1 = \cdots = p_4 = \frac{1}{4}$.

Estimate the probability that a value ≥ 8.17 occurs.

P-Value

Under the null hypothesis, values ≥ 10.97 occur with probability ≈ 0.012.

Permutation Test

What is the null distribution of X^2? How likely is a value like $X^2 = 10.97$ or larger?

Simulate many observations from a multinomial distribution with $n = 528, p_1 = \cdots = p_4 = \frac{1}{4}$.

Estimate the probability that a value ≥ 8.17 occurs.

P-Value

Under the null hypothesis, values ≥ 10.97 occur with probability ≈ 0.012.

Conclusion

If the birthday distribution were uniform across quarters, the observed discrepancy would be very unlikely (only in 1 in 80 cases). The observed discrepancy is **statistically significant**.

This tells us nothing about the alternative, only that the null hypothesis is inconsistent with the data.

Look at contributions to X^2 from the cells to see where the biggest discrepancies occur.

Conclusion and Interpretation

	Aug-Oct	Nov-Jan	Feb-April	May-July
Obs	150	138	140	100
Dev	18	6	8	-32
Cell	2.45	0.27	0.48	7.76

Birth date right after August 1 are overrepresented, birth dates just before that are underrepresented.

What would it mean if a very small X^2 had been observed?

Approximation: χ^2 Test

Pearson, Fisher: For sufficiently "full" cells (expected cell count > 5 or so), X^2 has an approximate χ^2 distribution with $k - 1$ degrees of freedom ($k =$ number of cells).
This should be apparent in the sampling distribution of X^2. Check this!

Here, there are 3 degrees of freedom. χ^2 approximation of P value:

$$1 - F_{\chi^2_3}(X^2) = 0.0119$$

Approximation: χ^2 Test

Pearson, Fisher: For sufficiently "full" cells (expected cell count > 5 or so), X^2 has an approximate χ^2 distribution with $k-1$ degrees of freedom ($k =$ number of cells).
This should be apparent in the sampling distribution of X^2. Check this!

Here, there are 3 degrees of freedom. χ^2 approximation of P value:

$$\boxed{1 - F_{\chi^2_3}(X^2) = 0.0119}$$

Two-Way Tables

Example: General Social Survey 2006

- Import the data
- What is recorded here?
- How complete are the data?

We are interested in the relation between gender and happiness. What are the possible levels of each variable?

GSS2006: Gender and Happiness

Need to handle NAs in the table. Where do these occur? Are they equally distributed between males and females?

We could address this by analyzing a suitable one way table. For now, just eliminate those observations from consideration.

Obtain a two-way table with $r = 2$ rows and $c = 3$ columns.

Are levels of happiness the same for males and females?

Null Hypothesis: Same distribution

The null hypothesis is that proportions for the three levels of happiness are the same for both genders.

That is, the same proportions of males and females are "not too happy / pretty happy / happy".

Deviations from these proportions would be due to chance.

Null hypothesis: Homogeneous populations

Expected Cell Counts

- $R_i =$ row sum of row i
- $C_j =$ column sum of column j
- $N = \sum_i R_i = \sum_j C_j =$ total count in the table

Overall fraction of population in column j is $\frac{C_j}{N}$. If If proportions are the same in all rows, then the row total R_i in row i should be distributed according to these proportions.

Expected cell count $\boxed{\frac{R_i C_j}{N}}$.

Statistic

As in the case of one-way tables, the statistic to be used is

$$X^2 = \sum_{rows,cols} \frac{(observed - expected)^2}{expected}$$

Observed value: $X^2_{obs} \approx 0.7969$

Permutation Test

Make random permutations of the original observations:

- Extract observations with complete gender and happiness information.
- Randomly permuted the happiness values.
- Make another two-way table of gender versus happiness.This table will have the same row and column totals.
- Compute the X^2 statistic.
- **Repeat many times.**

Null Distribution

See the histogram.

The p-value is 0.66. What does this mean?

Could we have kept the observations with Happy = NA in the data for permutations?

Is there a simpler way of simulating the null distribution?

χ^2 Approximation

Pearson, Fisher: For sufficiently "full" cells (expected cell count > 5 or so), X^2 has an approximate χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.
This should be apparent in the sampling distribution of X^2. Check this!

Here, there are 2 degrees of freedom. χ^2 approximation of P value:

$$1 - F_{\chi^2_2}(X^2) = 0.6713501$$

χ^2 Approximation

Pearson, Fisher: For sufficiently "full" cells (expected cell count > 5 or so), X^2 has an approximate χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.
This should be apparent in the sampling distribution of X^2. Check this!

Here, there are 2 degrees of freedom. χ^2 approximation of P value:

$$\boxed{1 - F_{\chi^2_2}(X^2) = 0.6713501}$$

Similar: Tests for Independence

Example: Are religious affiliation and views towards the death penalty independent?

Examine this using the GSS data.