Another look at sufficiency
Testing binomial data
Type I and Type II errors
Summary

Testing Binomial Data, Type I and Type II Errors

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

1 Another look at sufficiency

2 Testing binomial data

3 Type I and Type II errors

4 Summary

The idea behind sufficiency

- Estimators, by their very nature, discard data, $\hat{\theta}(\vec{X})$
- In doing so, they accomplish a kind of *data reduction*.
- For example, if you can reduce 10^6 normally distributed numbers to a mean and a variance, you have accomplished substantial data reduction.
- You need *all* 10^6 numbers to estimate the mean and variance, since you want to estimate those as accurately as possible.
- Once you have $\hat{\mu}(\vec{X})$ and $\hat{\sigma^2}(\vec{X})$, however, you don't need $\vec{X}$ any longer
- A sufficient estimator is one that does not *needlessly* discard data.
- If estimator $\hat{\theta}$ is sufficient, everything that can be known about the parameter θ has been extracted from the data, and nothing has been left behind.

Relation between sufficiency and other properties

- Given n pieces of data $\vec{X} = \langle X_1, \ldots, X_n \rangle$
- An estimator can be unbiased, but not sufficient
- An estimator can be consistent, but not sufficient
- $\hat{\mu}_n(\vec{X}) = X_1$ is unbiased and consistent, but not sufficient
  - Unbiased because $E(\hat{\mu}_n) = E(X_1) = \mu$
  - Consistent because $\text{Prob}\left(|\hat{\mu}_n - \mu| < \epsilon\right) > 1 - \frac{\sigma^2}{n\epsilon^2}$
  - Not sufficient because it wastes $n-1$ of the numbers in the sample $\vec{X}$
- An estimator can be sufficient, but not unbiased
  - If $\hat{\theta}$ is sufficient for θ, then any invertible function of $\hat{\theta}$ is likewise.
  - e.g., $\hat{\theta}_2 = \hat{\theta}^3$ has the same information content as $\hat{\theta}$.
  - One would not expect $E\left(\hat{\theta}^3\right)$ to equal $E\left[\left(\hat{\theta}\right)\right]^3$ so not unbiased

Testing binomial data – $H_0: \ p = p_0$

Bruce M. Boghosian

Another look at sufficiency
Testing binomial data
Type I and Type II errors
Summary

- Outcome of n Bernoulli trials: $k_1, k_2, \ldots, k_n$ where $k_j \in \{0, 1\}$
- Bernoulli distribution for one trial $p_Y(k; p) = p^k(1-p)^{1-k}$
- $E(Y) = p$ and $\text{Var}(Y) = p(1-p)$, but parameter p is unknown
- We wish to test a null hypothesis $H_0: \ p = p_0$
- We conduct n trials and let $X = Y_1 + Y_2 + \cdots + Y_n$
- Then X is distributed according to binomial distribution

$$p_X(k; p) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $k = 0, \ldots, n$.

Large-sample versus small-sample testing

- We have $0 \leq X \leq n$ and $E(X) = np$ and $\sigma_X = \sqrt{np(1-p)}$
- If n is sufficiently large, $[pn - 3\sigma_X, pn + 3\sigma_X] \subset [0, n]$
- Of course, p is unknown, so we use null hypothesis p_0 to make this judgement
- We do a *large-sample test*, relying on the Central Limit Theorem, if

$$\boxed{0 < np_0 - 3\sqrt{np_0(1-p_0)} < np_0 + 3\sqrt{np_0(1-p_0)} < n.}$$

- If the above is true, the normal distribution obtained from the CLT will comfortably fit in $[0, n]$.
- If the above is not true, we must conduct a *small-sample test*.

Tufts UNIVERSITY

How does large-sample testing work?

Bruce M. Boghosian

Another look at sufficiency

Testing binomial data

Type I and Type II errors

Summary

- We suppose that $\frac{X-np_0}{\sqrt{np_0(1-p_0)}}$ is distributed as a standard normal.
- Find thresholds for $100(1-\alpha)\%$ confidence of not rejecting H_0 if it is true.
- Take n samples, $k_1,\ldots,k_n$, let $k=\sum_j^n k_j$, and define $z := \frac{k-np_0}{\sqrt{np_0(1-p_0)}}$
- Three tests similar to our earlier work on hypothesis testing
  - To test $H_0 : p = p_0$ versus $H_1 : p > p_0$ at the α level of significance, reject H_0 if $z \geq +z_\alpha$
  - To test $H_0 : p = p_0$ versus $H_1 : p < p_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$
  - To test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ at the α level of significance, reject H_0 if z is either $\leq -z_{\alpha/2}$ or $\geq +z_{\alpha/2}$.

Example: Do people postpone their deaths?

- Study of $n = 747$ obituaries
- Only 60, or 8%, died in the three months prior to their birthday.
- If deaths were distributed uniformly, one would expect this to be 25%.
- Is the decrease from 25% to 8% statistically significant?
- Define k_j, for $j = 1, \ldots, 747$ to be
  - $= 1$ if the jth person died in the three months prior to their birthday
  - $= 0$ if the jth person died at any other time of the year
- Then $p_e = \frac{1}{n} \sum_j^n k_j$ is fraction of deaths three months prior to a birthday
- Take $H_0 : p = 0.25$ since the contrary seems perverse
- Take $H_1 : p < 0.25$ and demand confidence with $\alpha = 0.05$

Example: Do people postpone their deaths?

- First note that $np_0 = 747(0.25) = 186.75$ and $\sigma = \sqrt{747(0.25)(1-0.25)} = 11.83$
- Note np_0 is more than 3σ greater than zero and less than $n = 747$.
- Large-sample testing is warranted.
- Our null hypothesis is that $p = p_0 = 0.25$, and $n = 747$
- Calculate

$$z = \frac{k - np_0}{\sqrt{np_0(1-p_0)}} = \frac{60 - 747(0.25)}{\sqrt{747(0.25)(1-0.25)}} = -10.7.$$

- The above is far, far less than $-z_{0.05} = -1.64$.
- There is very strong evidence (much $>$ 95% confidence) that effect is real.

How does small-sample testing work?

- Example: Experimental drug test with $n = 19$ patients
- Old treatment is known to be 85% effective
- We wish to test $H_0 : p = 0.85$ versus $H_1 : p \neq 0.85$
- For $n = 19$ and $p_0 = 0.85$,
  - $np_0 = 16.15$
  - $\sigma = \sqrt{19(0.85)(0.15)} = 1.556$
- Note that $np_0 + 3\sigma = 16.15 + 3(1.556) = 20.819$
- Indicates that small-sample testing is necessary

How does small-sample testing work?

- List 19 possibilities, note $P(X \leq 13) = 0.053\ldots$, and $P(X = 19) = 0.045\ldots$

k	$P(X = k) = \binom{19}{k}(0.85)^k(0.15)^{19-k}$
6	1.99151×10^{-7}
7	2.09582×10^{-6}
8	0.0000178145
9	0.000123382
10	0.000699164
11	0.00324158
12	0.012246
13	0.0373659
14	0.0907457
15	0.171409
16	0.242829
17	0.242829
18	0.152892
19	0.0455994

- Hence we reject H_0 if $k \leq 13$ or $k = 19$.
- Note that confidence interval is asymmetric.

Type I and Type II errors

	True state of nature	
Our decision	H_0 is true	H_1 is true
Fail to reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

- Analysis of Type I error:

$$
\begin{aligned}
P(\text{Type I error}) =& P\left(\text{Reject } H_0 \mid H_0 \text{ is true}\right) \\
=& P\left(Z \geq z_\alpha \mid \mu = \mu_0\right) \\
=& P\left(\frac{X - \mu_0}{\sigma / \sqrt{n}} \geq z_\alpha \mid \mu = \mu_0\right) \\
=& \alpha.
\end{aligned}
$$

Type II errors

Bruce M. Boghosian

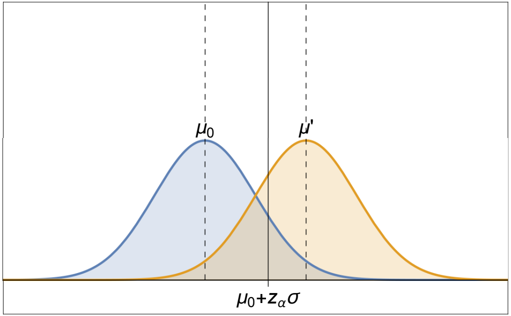
Another look at sufficiency
Testing binomial data
Type I and Type II errors
Summary

- Analysis of Type II errors

$$
\begin{aligned}
P(\text{Type II error}) =& P\left(\text{Fail to reject } H_0 \mid H_1 \text{ is true}\right) \\
=& P\left(Z \leq z_\alpha \mid \mu = \mu' > \mu_0\right) \\
=& P\left(\frac{X - \mu'}{\sigma/\sqrt{n}} \leq z_\alpha \mid \mu = \mu'\right) \\
=& \beta.
\end{aligned}
$$

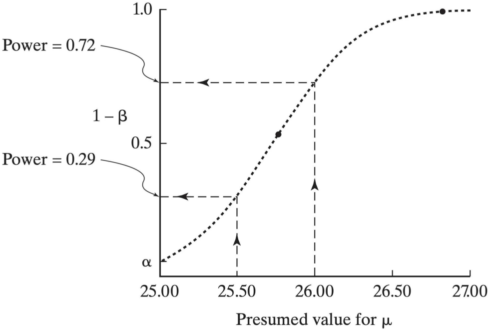
- Note that β depends on the assumed mean $\mu' > \mu_0$.
- A plot of $1 - \beta$ versus $\mu' > \mu_0$ is called a *power curve*.

Graphical depiction of Type I and Type II errors



- For fixed μ', there is a tradeoff between α and β.
- $1-\beta$ is the probability that we reject H_0 when H_1 is true – *power of the test*
- Plot of $1-\beta$ versus μ' is called a *power curve*

Power curves



From Larsen & Marx, Fig. 6.4.4, p. 362

- $1 - \beta$ is the probability that we reject H_0 when H_1 is true – *power of the test*
- Plot of $1 - \beta$ versus μ' is called a *power curve*
  - If $\mu' = 26$, easy to distinguish μ' from μ, so power is 0.72
  - If $\mu' = 25.5$, difficult to distinguish μ' from μ, so power is 0.29

Summary

Another look at sufficiency
Testing binomial data
Type I and Type II errors
Summary

- We have taken another look at sufficiency
- We have studied the testing of binomial data
- We have studied Type I and Type II errors