ORF 245 Fundamentals of Statistics
Chapter 9
Hypothesis Testing

Robert Vanderbei

Fall 2014

Slides last edited on November 24, 2014

Coin Tossing Example

Consider two coins. Coin 0 is fair ($p = 0.5$) but coin 1 is biased in favor of heads: $p = 0.7$. Imagine tossing one of these coins n times. The number of heads is a binomial random variable with the corresponding p values. The probability mass function is therefore

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

For $n = 8$, the probability mass functions are

x	0	1	2	3	4	5	6	7	8
$p(x\|H_0)$	0.0039	0.0313	0.1094	0.2188	0.2734	0.2188	0.1094	0.0313	0.0039
$p(x\|H_1)$	0.0001	0.0012	0.0100	0.0467	0.1361	0.2541	0.2965	0.1977	0.0576

Suppose that we know that our coin is either the fair coin or the biased coin. We flip it $n = 8$ times and it comes up heads 3 times. The probability of getting 3 heads with the fair coin is $p(3|H_0) = 0.2188$ whereas the probability of getting 3 heads with the biased coin is $p(3|H_1) = 0.0467$. Hence, it is $0.2188/0.0467 = 4.69$ times more likely that our coin is the fair coin than it is the biased coin. The ratio $p(x|H_0)/p(x|H_1)$ is called the *likelihood ratio*:

x	0	1	2	3	4	5	6	7	8
$\frac{p(x\|H_0)}{p(x\|H_1)}$	59.5374	25.5160	10.9354	4.6866	2.0086	0.8608	0.3689	0.1581	0.0678

Bayesian Approach

We consider two hypotheses:

$$H_0 = \text{"Coin 0 is being tossed"} \qquad \text{and} \qquad H_1 = \text{"Coin 1 is being tossed"}$$

Hypothesis H_0 is called the *null hypothesis* whereas H_1 is called the *alternative hypothesis*.

Suppose we have a prior believe as to the probability $P(H_0)$ that coin 0 is being tossed. Let $P(H_1) = 1 - P(H_0)$ denote our prior belief regarding the probability that coin 1 is being tossed. A natural initial choice is $P(H_0) = P(H_1) = 1/2$.

If we now toss the coin n times and observe x heads, then the *posterior* probability estimate that coin 0 is being tossed is

$$P(H_0|x) = \frac{p(x|H_0)P(H_0)}{p(x|H_0)P(H_0) + p(x|H_1)P(H_1)}$$

Hence, we get a simple intuitive formula for the ratio

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)} \frac{p(x|H_0)}{p(x|H_1)}$$

Making a Decision

If we must make a choice, we'd probably go with H_0 if

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)} \frac{p(x|H_0)}{p(x|H_1)} > 1$$

and H_1 otherwise. This inequality can be expressed in terms of the likelihood ratio:

$$\frac{p(x|H_0)}{p(x|H_1)} > \frac{P(H_1)}{P(H_0)} =: c$$

where c denotes the ratio of our priors.

If we use an unbiased prior, then $c = 1$ and we will accept H_0 if $X \leq 4$ and we will reject H_0 is $X > 4$.

Type I and Type II Errors

With $c = 1$, we will accept H_0 if $X \leq 4$ and we will reject H_0 if $X > 4$.

Type I Error: Reject H_0 when it is correct:

$$P(\text{reject } H_0 \mid H_0) = P(X > 4 | H_0) = 0.3634$$

Type II Error: Accept H_0 when it is incorrect:

$$P(\text{accept } H_0 \mid H_1) = P(X \leq 4 | H_1) = 0.1941$$

With $c = 5$, we will accept H_0 if $X \leq 2$ and we will reject H_0 is $X > 2$. In this case, we get

Type I Error: $P(\text{reject } H_0 \mid H_0) = P(X > 2 | H_0) = 0.8556$

Type II Error: $P(\text{accept } H_0 \mid H_1) = P(X \leq 2 | H_1) = 0.0113$

The parameter c controls the trade-off between Type-I and Type-II errors.

Type-I and Type-II Errors

Reality / Decision	H_0 true	H_1 true
Accept H_0	Yes!	Type-II error
Reject H_0	Type-I error	Yes!

A Type-I error is also called a *false discovery* or a *false positive*.

A Type-II error is also called a *missed discovery* or a *false negative*.

Neyman-Pearson Paradigm

Abandon the Bayesian approach.

Instead, focus on probability of Type-I and Type-II errors associated with a threshold c.

Things to Consider

It is conventional to choose the simpler of two hypotheses as the null.

The consequences of incorrectly rejecting one hypothesis may be graver than those of incorrectly rejecting the other. In such a case, the former should be chosen as the null hypothesis, because the probability of falsely rejecting it could be controlled by choosing α. Examples of this kind arise in screening new drugs; frequently, it must be documented rather conclusively that a new drug has a positive effect before it is accepted for general use.

In scientific investigations, the null hypothesis is often a simple explanation that must be discredited in order to demonstrate the presence of some physical phenomenon or effect.

In criminal matters: *Innocent until proven guilty!*

Terminology

Null Hypothesis:	H_0
Alternative Hypothesis:	H_1
Type I Error:	Rejecting H_0 when it is true.
Significance Level:	The probability of a Type-I error. Usually denoted α.
Type II Error:	Accepting H_0 when it is false. Probability usually denoted β.
Power:	$1 - \beta$.
Test Statistic:	The measured quantity on which a decision is based.
Rejection region:	The set of values of the test statistic that lead to rejection of the null hypothesis.
Acceptance region:	The set of values of the test statistic that lead to acceptance of the null hypothesis.
Null Distribution:	$p(x\|H_0)$

Example 9.2.A (Unrealistic)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution having *known variance* σ^2 but *unknown mean* μ. We consider two hypotheses:

$$H_0 : \ \mu = \mu_0$$
$$H_1 : \ \mu = \mu_1$$

where μ_0 and μ_1 are two specific possible values (suppose that $\mu_0 < \mu_1$).
The likelihood ratio is the ratio of the joint density functions:

$$\frac{f_0(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)} = \frac{\exp\left(-\sum_{i=1}^{n}(x_i - \mu_0)^2/2\sigma^2\right)}{\exp\left(-\sum_{i=1}^{n}(x_i - \mu_1)^2/2\sigma^2\right)}$$

We will accept the null hypothesis if this ratio is larger than some positive threshold.
Let's call the threshold e^c and take logs of both sides:

$$-\sum_{i=1}^{n}(x_i - \mu_0)^2/2\sigma^2 + \sum_{i=1}^{n}(x_i - \mu_1)^2/2\sigma^2 > c$$

Example 9.2.A – Continued

Expanding the squares, combining the two sums, and simplifying, we get

$$\sum_{i=1}^{n} \left(2x_i(\mu_0 - \mu_1) - \mu_0^2 + \mu_1^2\right) > 2\sigma^2 c$$

Dividing both sides by $2n$ and by $\mu_1 - \mu_0$, we get

$$-\bar{x} + (\mu_0 + \mu_1)/2 > \sigma^2 c/n(\mu_1 - \mu_0)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Finally, isolating $\bar{x}$ from the other terms, we get

$$\bar{x} < (\mu_0 + \mu_1)/2 - \sigma^2 c/n(\mu_1 - \mu_0)$$

Choosing $c = 0$ corresponds to equally probable priors (since $e^0 = 1$). In this case, we get the intuitive result that we should accept the null hypothesis if

$$\bar{x} < (\mu_0 + \mu_1)/2$$

From these calculations, we see that the *test statistic* is the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and that we should reject the null hypothesis when the test statistic is larger than some threshold value, let's call it x_0.

Example 9.2.A – Significance Level

In hypothesis testing, one designs the test according to a set value for the significance level. In this problem, the significance level is given by

$$\alpha = P(\bar{X} > x_0 \mid H_0)$$

If we subtract μ_0 from both sides and then divide by $\sigma/\sqrt{n}$, we get a random variable with a standard normal distribution:

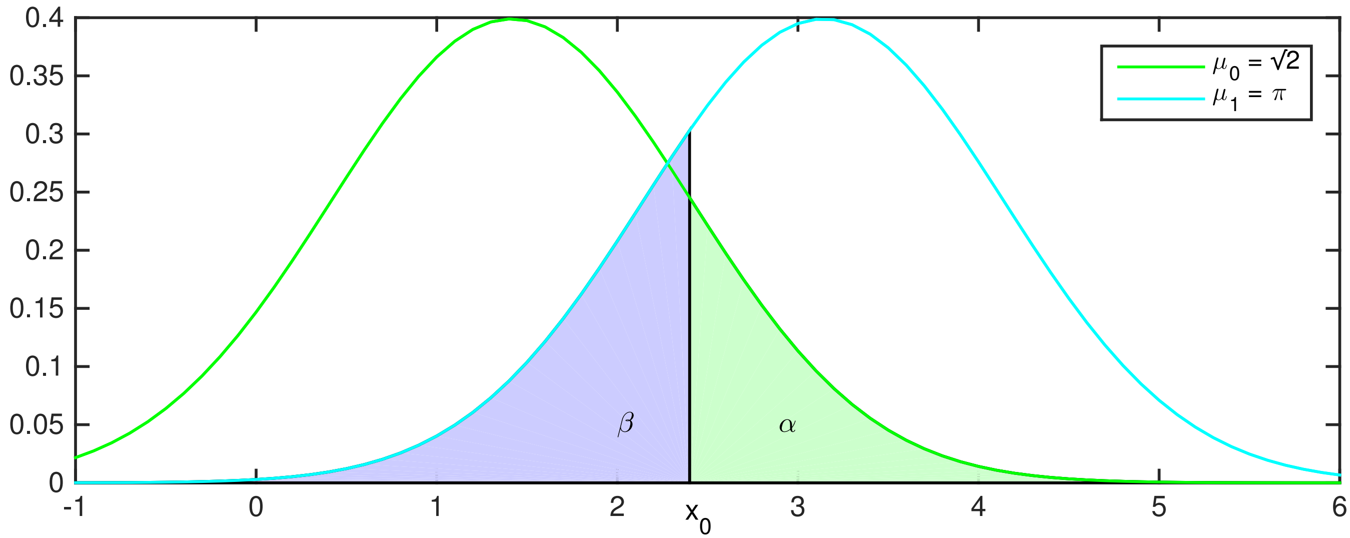
$$\alpha = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_0 - \mu_0}{\sigma/\sqrt{n}} \;\middle|\; H_0\right) = 1 - F_N\left(\frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right)$$

where F_N denotes the cumulative distribution function for a standard Normal random variable.

In a similar manner, we can compute the probability of a Type-II error:

$$\beta = P(\bar{X} \leq x_0 \mid H_1) = P\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \leq \frac{x_0 - \mu_1}{\sigma/\sqrt{n}} \;\middle|\; H_1\right) = F_N\left(\frac{x_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$

Example 9.2.A – α vs. β Trade-Off



As x_0 slides from left to right, α goes down whereas β goes up.

The sum $\alpha + \beta$ is minimized at $x_0 = (\mu_0 + \mu_1)/2$.

To make both α and β smaller, need to make $\sigma/\sqrt{n}$ smaller; i.e., increase n.

Generalized Likelihood Ratios—Example 9.4.A

Let $X_1, \ldots, X_n$ be iid Normal rv's with unknown mean μ and known variance σ^2.
We consider two hypotheses:

$$\begin{aligned} H_0 &: \ \mu = \mu_0 \\ H_1 &: \ \mu \neq \mu_0 \end{aligned}$$

The null hypothesis H_0 is *simple*. The alternative H_1 is *composite*.
For simple hypotheses, the *likelihood* that the hypothesis is true given specific observed values $x_1, x_2, \ldots, x_n$ is just the joint pdf. Hence, the likelihood that H_0 is true is given by:

$$f_0(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2}$$

For composite hypotheses, the *generalized likelihood* is taken to be the largest possible likelihood that could be obtained over all choices of the distribution. So, the likelihood that H_1 is true is given by:

$$f_1(x_1, \ldots, x_n) \;=\; \max_\theta \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2} \;\overset{\text{MLE}}{=}\; \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2}$$

NOTE: The maximum is achieved by the *maximum likelihood estimator*, which we have already seen is just $\bar{x}$.

Example 9.4.A Continued

As before, we define our acceptance region based on the *log-likelihoood-ratio*:

$$\Lambda(x_1, \ldots, x_n) \;=\; \log \frac{f_0(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)} \;=\; -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Note that $f_0 \leq f_1$ and therefore $\Lambda \leq 0$. We accept the null hypothesis when Λ is not too negative:

$$\Lambda(x_1, \ldots, x_n) > -c$$

for some negative constant: $-c$.
Again as before, we expand the quadratics and simplify the inequality to get

$$\sum_{i=1}^{n}\left(2x_i(\mu_0 - \bar{x}) - \mu_0^2 + \bar{x}^2\right) \;>\; -2\sigma^2 c$$

Now, we use the fact that $\sum_i x_i = n\bar{x}$ to help us further simplify:

$$2\bar{x}(\mu_0 - \bar{x}) - \mu_0^2 + \bar{x}^2 \;>\; -2\frac{\sigma^2}{n}c$$

Some final algebraic manipulations and we get:

$$(\bar{x} - \mu_0)^2 \;<\; 2\frac{\sigma^2}{n}c$$

Example 9.4.A Continued

And, switching to capital-letter random-variable notation, we get that we should accept H_0 if

$$|\bar{X} - \mu_0| \ < \ \sqrt{2c}\frac{\sigma}{\sqrt{n}}$$

or, in other words,

$$\mu_0 - \sqrt{2c}\frac{\sigma}{\sqrt{n}} \ < \ \bar{X} \ < \ \mu_0 + \sqrt{2c}\frac{\sigma}{\sqrt{n}}$$

Example 9.3.A (More Realistic)

As before, let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution having known variance σ^2 but unknown mean μ.
This time, let's consider these two hypotheses:

$$\begin{aligned} H_0 &: \ \mu = \mu_0 \\ H_1 &: \ \mu \neq \mu_0 \end{aligned}$$

where μ_0 is a specific/given value.
Reject null hypothesis if

$$\left| \bar{X} - \mu_0 \right| > x_0$$

where x_0 is chosen so that

$$P_{H_0}\left(\left| \bar{X} - \mu_0 \right| > x_0 \right) = \alpha$$

(note the probability is computed assuming H_0 is true). Since we know that the distribution is normal, we see that

$$x_0 = z(\alpha/2)\, \sigma_{\bar{X}}$$

Confidence Interval vs Hypothesis Test

It's easy to check that

$$P_{H_0}\left(\left|\bar{X} - \mu_0\right| > z(\alpha/2)\,\sigma_{\bar{X}}\right) = \alpha$$

is equivalent to

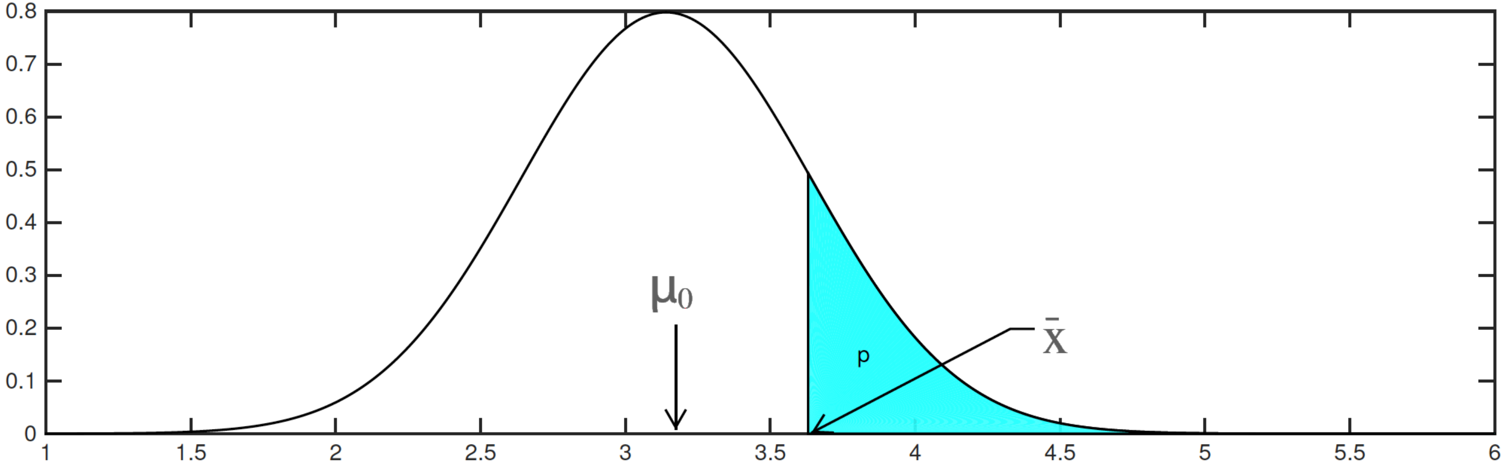
$$P_{H_0}\left(\bar{X} - z(\alpha/2)\,\sigma_{\bar{X}} \le \mu_0 \le \bar{X} + z(\alpha/2)\,\sigma_{\bar{X}}\right) = 1 - \alpha$$

The $100(1-\alpha)\%$ confidence interval for μ is

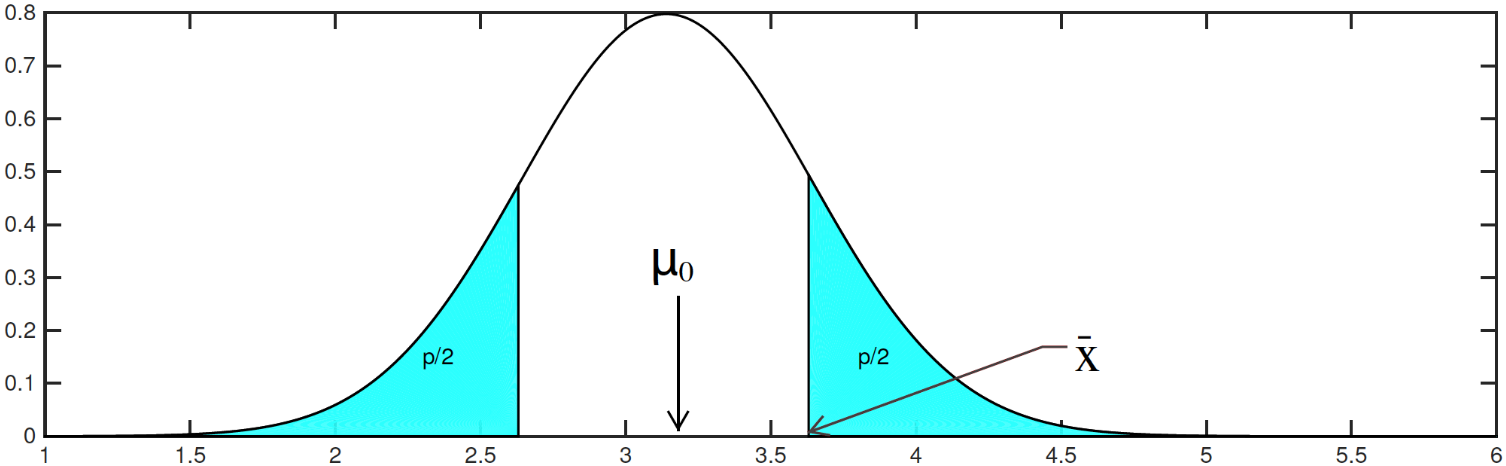
$$P_{\mu}\left(\bar{X} - z(\alpha/2)\,\sigma_{\bar{X}} \le \mu \le \bar{X} + z(\alpha/2)\,\sigma_{\bar{X}}\right) = 1 - \alpha$$

Comparing the acceptance region for the test to the confidence interval, we see that μ_0 lies in the confidence interval for μ if and only if the hypothesis test accepts the null hypothesis. In other words, the confidence interval consists precisely of all those values of μ_0 for which the null hypothesis $H_0 : \mu = \mu_0$ is accepted.

p-Values



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
μ0
p
x̄



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6
μ0
p/2
p/2
x̄

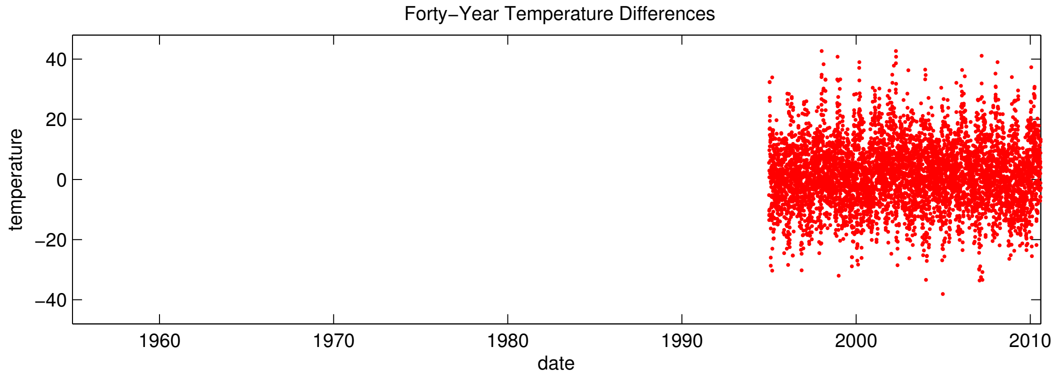
Local Climate Data – Forty Year Differences

On a per century basis...

$$\bar{X} = 4.25°\text{F/century},$$

$$S = 26.5°\text{F/century},$$

$$S/\sqrt{n} = 0.35°\text{F/century}$$



Hypotheses:

$$H_0 : \mu = 0, \qquad H_1 : \mu \neq 0$$

Number of standard deviations out:

$$z = \bar{X}/S_{\bar{X}} = 4.25/0.35 = 12.1$$

Probability that such an outlier happened by chance:

$$p = P(N > 12.1) = 5 \times 10^{-34}$$

Now I'm *way convinced!*

Caveat: The random variables are not fully independent. If it's warmer than average today, then it is likely to be warmer than average tomorrow. Had we used daily averages from just one day a week, the variables would be closer to independent. We'd still reject the null hypothesis but z would be smaller (by a factor of $\sqrt{7}$) and so the p value would be larger.