Chapter 9: Hypothesis Testing

Sections

- 9.1 Problems of Testing Hypotheses
- Skip: 9.2 Testing Simple Hypotheses
- Skip: 9.3 Uniformly Most Powerful Tests
- Skip: 9.4 Two-Sided Alternatives
- 9.5 The t Test
- 9.6 Comparing the Means of Two Normal Distributions
- 9.7 The F Distributions
- 9.8 Bayes Test Procedures
- 9.9 Foundational Issues

Bayesian test procedures

All inference about a parameter is based on the posterior distribution, including hypothesis testing. Let

$$H_0 : \theta \in \Omega_0 \quad \text{vs.} \quad H_1 : \theta \in \Omega_1$$

Then we can obtain:

- $P(\theta \in \Omega_0 | \mathbf{x}) =$ probability that H_0 is true
- $P(\theta \in \Omega_1 | \mathbf{x}) =$ probability that H_1 is true

A straightforward test procedure:

- Reject H_0 if $P(\theta \in \Omega_0 | \mathbf{x}) < P(\theta \in \Omega_1 | \mathbf{x})$
- Critical region: $S_1 = \left\{ \mathbf{x} : P(\theta \in \Omega_1 | \mathbf{x}) > \frac{1}{2} \right\}$

However, since hypothesis testing is a decision problem we should also consider a loss function

Bayesian test procedures

- *Loss function*: $L(\theta, a) =$ the loss that occurs when θ is the true value of the parameter and action a is taken
- *Bayes test procedure*: Minimize posterior expected loss

We will first consider simple hypotheses:

- Let $X_1, \ldots, X_n$ be a random sample from $f(x|\theta)$ where the parameter space contains only two values: $\Omega = \{\theta_0, \theta_1\}$.
- We want to test

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$$

- See chapter 9.2 for a frequentist take on this situation

Bayesian test procedures

Possible decisions (actions):

- d_0: do not reject H_0 ("accept H_0")
- d_1: reject H_0 ("accept H_1")

We specify the losses from making a wrong decision. For example:

- $L(\theta_0, d_1) = w_0$: Loss for d_1 when H_0 is true (type I error)
- $L(\theta_1, d_0) = w_1$: Loss for d_0 when H_1 is true (type II error)

No loss if we make the correct decision:

- $L(\theta_1, d_1) = 0$: The loss when d_1 is chosen and H_1 is true
- $L(\theta_0, d_0) = 0$: The loss when d_0 is chosen and H_0 is true

This loss function $L(\theta, d)$ can be summarized as

$L(\theta_i, d_j)$	d_0	d_1
θ_0	0	w_0
θ_1	w_1	0

This is the *generalized 0-1 loss*

Bayesian test procedures

$L(\theta_i, d_j)$	d_0	d_1
θ_0	0	w_0
θ_1	w_1	0

- Let $p(\theta)$ be the prior pf of θ
- Let $p(\theta_0) = p_0$ and $p(\theta_1) = p_1$
- Expected loss for a test procedure δ:

$$r(\delta) = p_0 w_0 \alpha(\delta) + p_1 w_1 \beta(\delta)$$

where

$$\alpha(\delta) = P(\text{chose } d_1 | \theta = \theta_0) = \text{ Prob. of type I error}$$
$$\beta(\delta) = P(\text{chose } d_0 | \theta = \theta_1) = \text{ Prob. of type II error}$$

We want to minimize $r(\delta)$

Bayes test procedure for simple hypotheses

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$$

Let $X_1, \ldots, X_n$ be a random sample from $f(\mathbf{x}|\theta)$ and let $f_0(\mathbf{x}) = f(\mathbf{x}|\theta_0)$ and $f_1(\mathbf{x}) = f(\mathbf{x}|\theta_1)$

Theorem 9.2.1

Let δ^* be the test that rejects H_0 if $af_0(\mathbf{x}) < bf_1(\mathbf{x})$. Then for every other test procedure δ

$$a\alpha(\delta^*) + b\beta(\delta^*) \leq a\alpha(\delta) + b\beta(\delta)$$

δ^* can either reject or not for $af_0(\mathbf{x}) = bf_1(\mathbf{x})$.

- Set $a = p_0 w_0$ and $b = p_1 w_1$ and it follows that δ^* minimizes the expected loss and is therefore a Bayes test procedure

Bayes test procedure for simple hypotheses

In summary: The Bayes test procedure for testing the simple hypotheses

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1$$

is to reject H_0 if

$$p_0 w_0 f_0(\mathbf{x}) < p_1 w_1 f_1(\mathbf{x})$$

This is the same as the test that rejects H_0 if

$$p(\theta_0|\mathbf{x}) \leq \frac{w_1}{w_0 + w_1}$$

or equivalently if $p(\theta_1|\mathbf{x}) > \frac{w_0}{w_0+w_1}$

Bayes test procedure in general

- Now lets come back to the general hypotheses

$$H_0 : \theta \in \Omega_0 \quad \text{vs.} \quad H_1 : \theta \in \Omega_1$$

and consider the generalized 0-1 loss:

$L(\theta, d_j)$	d_0	d_1
$\theta \in \Omega_0$	0	w_0
$\theta \in \Omega_1$	w_1	0

- More generally: w_0 and w_1 could be functions of θ

Bayes test procedure under generalized 0-1 loss

The Bayes test procedure is to reject H_0 if

$$P(H_0 \text{ is true } |\mathbf{x}) = P(\theta \in \Omega_0|\mathbf{x}) \leq \frac{w_1}{w_0 + w_1}$$

So the test we saw on slide 2 is a special case of using a 0-1 loss function with $w_0 = w_1$

Example: Test about the mean of the normal

- Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1/\tau)$ and assume that the prior distribution of (θ, τ) is the Normal-Gamma distribution.
- Suppose we want to test the hypotheses

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

- Suppose also we assume the generalized 0-1 loss from the previous slide
- The Bayes test procedure rejects H_0 if

$$\left(\frac{\lambda_1 \alpha_1}{\beta_1}\right)^{1/2} (\mu_1 - \theta_0) \geq T^{-1}_{2\alpha_1}\left(1 - \frac{w_1}{w_1 + w_0}\right)$$

  where μ_1, λ_1, α_1 and β_1 are the parameters of the posterior Normal-Gamma distribution
- What is the Bayes test procedure for the improper prior $p(\theta, \tau) = 1/\tau$?

Note: Typo Example 9.8.5: $U \leq T^{-1}_{n-1}(1 - \alpha_0)$ should be $U \leq -T^{-1}_{n-1}(1 - \alpha_0) = T^{-1}_{n-1}(\alpha_0)$

Two-sided alternatives

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

- If the posterior distribution of θ is continuous then

$$P(\theta \in \Omega_0|\mathbf{x}) = P(\theta = \theta_0|\mathbf{x}) = 0$$

In stead we can consider the hypotheses

$$H_0 : |\theta - \theta_0| \leq d \quad \text{vs.} \quad H_1 : |\theta - \theta_0| > d$$

where d represents what is a meaningful difference between θ and θ_0

- This forces us to think about what is a meaningful difference
- The Bayes test procedure (under the generalize 0-1 loss) is then simply to reject H_0 if

$$P(|\theta - \theta_0| \leq d|\mathbf{x})) \leq \frac{w_1}{w_0 + w_1}$$

Significance level and sample size

Standard practice:

- Specify a level of significance α_0 and then find a test that has a large power function on Ω_1 (small type II probability)
- Traditional α_0: 0.10, 0.05, 0.01 (0.05 is the most commonly used)
- Significance level α_0 is chosen in accordance to how serious the consequences of type I error are.
  Worse consequences $\Rightarrow$ smaller α_0
- A cautious experimenter: Choose $\alpha_0 = 0.01$ (Doesn't want to reject H_0 unless there is strong evidence that H_0 is not true)

Problem: For a large sample size, using $\alpha_0 = 0.01$ can actually lead to a test procedure that will reject H_0 for certain samples that, in fact, provide stronger evidence for H_0 than H_1

Significance level and sample size

Example

- Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ and we want to test

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta = 1$$

- We set $\alpha_0 = 0.01$, so the probability of type I error is $\alpha(\delta^*) = 0.01$
- The test procedure δ^* that then minimizes the probability of type II error rejects H_0 if $\overline{X}_n \geq 2.326/\sqrt{n}$
- The probability of type II error is

$$\beta(\delta^*) = \Phi(2.326 - \sqrt{n})$$

- This test is equivalent to rejecting H_0 if (shown in Chapter 9.2)

$$\frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} \geq k = \exp(2.326\sqrt{n} - 0.5n)$$

  That is, we reject if the data is at least k times as likely under H_1 as they are under H_0

Significance level and sample size

Example

The probabilities of type I and type II errors ($\alpha(\delta^*)$ and $\beta(\delta^*)$) and k for $n = 1$, $n = 25$ and $n = 100$:

n	$\alpha(\delta^*)$	$\beta(\delta^*)$	k
1	0.01	0.91	6.21
25	0.01	0.0038	0.42
100	0.01	8×10^{-15}	2.5×10^{-12}

- For larger n we get much more cautious about they type II error than the type I error!
- For $n = 1$: H_0 will be rejected if the observed data are at least 6.21 times as likely under H_1 as they are under H_0
- For $n = 100$: H_0 will be rejected for observed data that are millions of times more likely under H_0 than they are under H_1

Significance level and sample size

The problem:

- The α_0 level is fixed and then we minimize prob. of type II error, $\beta(\delta)$
- For large sample sizes we can get extremely low $\beta(\delta)$, relative to $\alpha(\delta)$

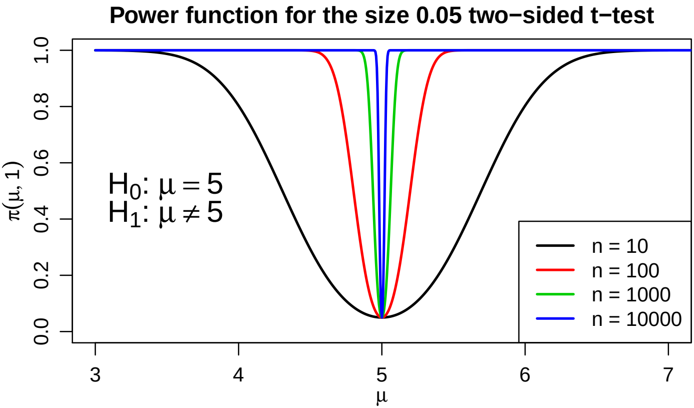
Possible solutions:

- Pick smaller α_0 for larger sample sizes
- Rather than fixing α_0, take both $\alpha(\delta)$ and $\beta(\delta)$ into account e.g.

$$100\alpha(\delta) + \beta(\delta)$$

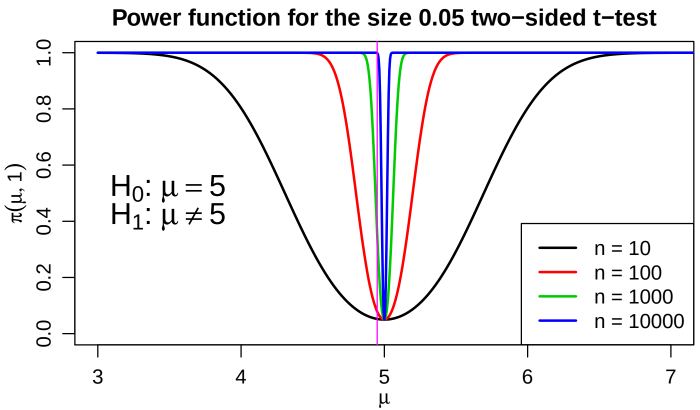
  here the type I error is deemed 100 times more serious that type II error
- Bayesian methods achieve this
- There are also frequentist methods that do this (Lehman 1958)

Power and sample size



- Power tends to increase as sample size increases
- Suppose the true value of μ is 4.95. Would we want to reject $H_0 : \mu = 5$ in that case?

Power and sample size



- Power tends to increase as sample size increases
- Suppose the true value of μ is 4.95. Would we want to reject $H_0 : \mu = 5$ in that case?

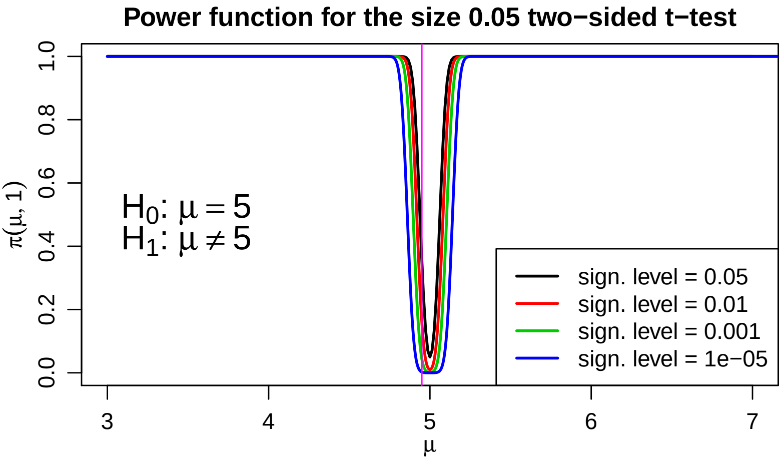
Statistically significant

- Suppose the true value of μ is 4.95.
- For a large n it is very likely that we reject $H_0 : \mu = 5$.
- The results will be called "*statistically significant*"
- That does not necessarily mean that μ is "significantly" different from 5 in a practical way

Possible solutions

- Use a much smaller significance level (see figure on next slide)
- Use an interval in the null hypothesis, e.g. $H_0 : a_1 \leq \mu \leq a_2$
- Consider doing estimation instead of hypothesis testing - use confidence intervals

Sample size and power



Power function for $n = 1000$ for different α_0 levels.

END OF CHAPTER 9