STK4011 and STK9011
Autumn 2015

Hypothesis testing

Covers (most of) the following material from chapters 8 and 10:

- Section 8.3.2
- Section 10.3.1
- Section 10.3.2: to the middle of page 495

Ørnulf Borgan
Department of Mathematics
University of Oslo

Basic concepts

Assume that we have random variables $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pmf or pdf $f(\mathbf{x} \mid \theta) = f(x_1, ..., x_n \mid \theta)$ where $\theta \in \Theta$

We want to test the null hypothesis $H_0 : \theta \in \Theta_0$ versus the alternative hypothesis $H_1 : \theta \in \Theta_0^c$

A hypothesis test is a procedure that specifies:

- for which values of $\mathbf{X}$ we reject H_0 (accept H_1)
- for which values of $\mathbf{X}$ do not reject H_0 (accept H_0)

We may make two types of error:

		Decision	
		Accept H_0	Reject H_0
Truth	H_0	Correct decision	Type I Error
	H_1	Type II Error	Correct decision

Let R be the rejection region of the test, so we reject $H_0 : \theta \in \Theta_0$ if $\mathbf{X} \in R$

Probability of Type I error: $P_\theta\left(\mathbf{X} \in R\right), \quad \theta \in \Theta_0$

Probability of Type II error:

$$P_\theta\left(\mathbf{X} \in R^c\right) = 1 - P_\theta\left(\mathbf{X} \in R\right), \quad \theta \in \Theta_0^c$$

Power function: $\beta(\theta) = P_\theta\left(\mathbf{X} \in R\right)$

We distinguish between the size and level of a test:

- a test with power function $\beta(\theta)$ is a size α test if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$
- a test with power function $\beta(\theta)$ is a level α test if $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$

Let C be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$. A test in the class C, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class C test if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Theta_0^c$ and every $\beta'(\theta)$ that is a power function of a test in class C

Theorem 8.3.12 (Neyman-Pearson Lemma)

Consider testing $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$, where the pdf or pmf corresponding to θ_i is $f(\mathbf{x} \mid \theta_i)$; $i = 0,1$, using a test with rejection region R that satisfies

$$\mathbf{x} \in R \quad \text{if} \quad f(\mathbf{x} \mid \theta_1) > k\, f(\mathbf{x} \mid \theta_0)$$
$$\mathbf{x} \in R^c \quad \text{if} \quad f(\mathbf{x} \mid \theta_1) < k\, f(\mathbf{x} \mid \theta_0) \qquad (8.3.1)$$

for some $k \geq 0$, and

$$\alpha = P_{\theta_0}(\mathbf{X} \in R) \qquad (8.3.2)$$

Then

a) Any test that satisfies (8.3.1) and (8.3.2) is a UMP level α test

b) If there exists a test satisfying (8.3.1) and (8.3.2) with $k > 0$, then every UMP level α test is a size α test and every UMP level α test satisfies (8.3.1) except perhaps on a set A satisfying $P_{\theta_0}(\mathbf{X} \in A) = P_{\theta_1}(\mathbf{X} \in A) = 0$

Corollary 8.3.13

Suppose that $T = T(\mathbf{X})$ is a sufficient statistic for θ and let $g(t \mid \theta_i)$ be the pdf or pmf of T corresponding to θ_i; $i = 0,1$. Then any test based on T with rejection region S is a UMP level α test if it satisfies

$$t \in S \quad \text{if} \quad g(t \mid \theta_1) > k\, g(t \mid \theta_0)$$
$$t \in S^c \quad \text{if} \quad g(t \mid \theta_1) < k\, g(t \mid \theta_0)$$

for some $k \geq 0$, where $\alpha = P_{\theta_0}(T \in S)$

Example 8.3.15 (UMP normal test)

Let $X_1, X_2, \ldots, X_n$ be iid $n(\theta, \sigma^2)$ with σ^2 known

We will find the UMP test for testing test $H_0: \theta = \theta_0$ versus $H_1: \theta = \theta_1$ where $\theta_0 > \theta_1$

The sample mean $\bar{X}$ is sufficient, and has pdf

$$g(\bar{x} \mid \theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} e^{-n(\bar{x}-\theta)^2/(2\sigma^2)}$$

By Corollary 8.3.13, the most powerful test rejects H_0 when

$$\frac{g(\bar{x} \mid \theta_1)}{g(\bar{x} \mid \theta_0)} > k$$

Now we have:

$$\frac{g(\bar{x} \mid \theta_1)}{g(\bar{x} \mid \theta_0)} = \frac{(2\pi\sigma^2/n)^{-n/2} e^{-n(\bar{x}-\theta_1)^2/(2\sigma^2)}}{(2\pi\sigma^2/n)^{-n/2} e^{-n(\bar{x}-\theta_0)^2/(2\sigma^2)}}$$

$$= \exp\left\{\frac{n}{2\sigma^2}\left[2(\theta_1 - \theta_0)\bar{x} + \theta_0^2 - \theta_1^2\right]\right\}$$

So the most powerful test rejects when (since $\theta_0 > \theta_1$)

$$\bar{X} < \frac{(2\sigma^2 \log k)/n - \theta_0^2 + \theta_1^2}{2(\theta_1 - \theta_0)} = c$$

The test has size α if we choose

$$c = \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

Neyman-Pearsons Lemma considers testing of the simple null hypothesis $H_0: \theta = \theta_0$ versus the simple alternative hypothesis $H_1: \theta = \theta_1$

We will use the result to obtain a UMP level α test for the one-sided composite null hypothesis $H_0: \theta \le \theta_0$ versus the one-sided composite alternative hypothesis $H_1: \theta > \theta_0$ (or $H_0: \theta \ge \theta_0$ versus $H_1: \theta < \theta_0$)

We then need to consider a property of the likelihood ratio called the monotone likelihood ratio (MLR) property

Definition 8.3.16

A family of pdfs or pmfs $\{g(t|\theta): \theta \in \Theta\}$ for a univariate random variable T with real-valued parameter θ has a monotone likelihood ratio (MLR) if for every $\theta_2 > \theta_1$ the ratio $g(t|\theta_2)/g(t|\theta_1)$ is a monotone function of *t* (nonincreasing or non-decreasing) on the set $\{t: g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$

(We interpret $c/0$ as ∞ when $c > 0$)

Example (MLR for a normal mean)

Let $X_1, X_2, ..., X_n$ be iid $n(\theta, \sigma^2)$ with σ^2 known

The likelihood ratio for $\bar{X}$ is

$$\frac{g(\bar{x}|\theta_2)}{g(\bar{x}|\theta_1)} = \exp\left\{\frac{n}{2\sigma^2}\left[2(\theta_2 - \theta_1)\bar{x} + \theta_1^2 - \theta_2^2\right]\right\}$$

Theorem 8.3.17

Consider testing $H_0: \theta \le \theta_0$ versus $H_1: \theta > \theta_0$. Suppose that $T = T(\mathbf{X})$ is a sufficient statistic for θ and that the family of pdfs or pmfs $\{g(t|\theta): \theta \in \Theta\}$ of T has a nondecreasing likelihood ratio. Then for any t_0 , the test that rejects H_0 if and only if $T > t_0$ is a UMP level α test, where $\alpha = P_{\theta_0}(T > t_0)$

For testing $H_0: \theta \ge \theta_0$ versus $H_1: \theta < \theta_0$, the test that rejects if and only if $T < t_0$ is a UMP level α test, where $\alpha = P_{\theta_0}(T < t_0)$

If the likelihood ratio is nonincreasing, the inequalities for the rejection regions are reversed

Example 8.3.18 (UMP normal test)

Let $X_1, X_2, ..., X_n$ be iid $n(\theta, \sigma^2)$ with σ^2 known

Consider testing $H_0: \theta \ge \theta_0$ versus $H_1: \theta < \theta_0$

The UMP level α test rejects H_0 if

$$\bar{X} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

The power function is

$$\beta(\theta) = P_\theta\left(\bar{X} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}\right) = P\left(Z < -z_\alpha + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}}\right)$$

where $Z \sim n(0,1)$

Example 8.3.19 (two-sided alternative)

Let $X_1, X_2, ..., X_n$ be iid $n(\theta, \sigma^2)$ with σ^2 known

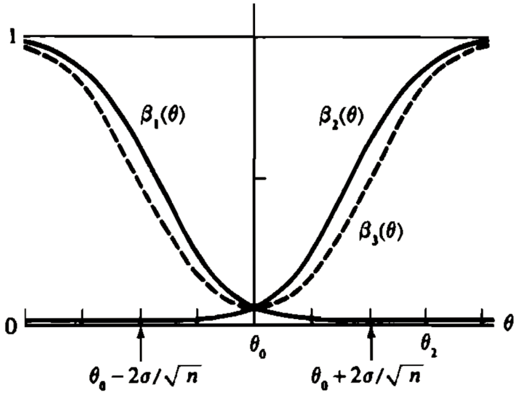
We will test $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$

Consider the three tests:

- Test 1: Reject H_0 if $\bar{X} < \theta_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$
- Test 2: Reject H_0 if $\bar{X} > \theta_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$
- Test 3: Reject H_0 if

$$\bar{X} < \theta_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X} > \theta_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Power function of the three tests:



No UMP level α test exists

But test 3 is UMP unbiased level α test

Likelihood ratio tests

Let $X_1, X_2, ...$ be iid with pdf or pmf $f(x|\theta)$, where θ may be a vector

Then the likelihood based on $\mathbf{X} = (X_1, X_2, ..., X_n)$ is given by (we supress n in the notation)

$$L(\theta|\mathbf{x}) = L(\theta|x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

The likelihood ratio test statistic for testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}$$

Let $\hat{\theta}$ be the unrestricted maximum likelihood estimator of θ, i.e. the value of θ that maximizes the likelihood when $\theta \in \Theta$

Let $\hat{\theta}_0$ be the maximum likelihood estimator of θ under the null hypothesis, i.e. the value of θ that maximizes the likelihood when $\theta \in \Theta_0$

Then the LRT statistics takes the form

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

The likelihood ratio test (LRT) has rejection region of the form $\{\mathbf{x}: \lambda(\mathbf{x}) \leq c\}$

To obtain a level α test, the constant c must be chosen so that

$$\sup_{\theta \in \Theta_0} P_\theta\left(\lambda(\mathbf{X}) \le c\right) \le \alpha$$

We may use asymptotic arguments to obtain an approximation of the distribution of $-2\log\lambda(\mathbf{X})$

Theorem 10.3.1

Consider testing $H_0: \theta = \theta_0$ versus $H_1: \theta \ne \theta_0$, and assume «some regularity conditions» (page 516). Then under H_0

$$-2\log\lambda(\mathbf{X}) \to \chi_1^2$$

in distribution as $n \to \infty$

Example (normal with known variance)

Let $X_1, X_2, \ldots, X_n$ be iid $n(\theta, \sigma^2)$ with σ^2 known

We will test $H_0: \theta = \theta_0$ versus $H_1: \theta \ne \theta_0$

Here we have (cf. example 8.2.2)

$$\lambda(\mathbf{X}) = \exp\left\{-\frac{n}{2\sigma^2}(\bar{X} - \theta_0)^2\right\}$$

Hence

$$-2\log\lambda(\mathbf{X}) = \frac{n}{\sigma^2}(\bar{X} - \theta_0)^2 = \left(\frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

So here the result holds also for finite n

Example 10.3.2 (Poisson)

Let $X_1, X_2, \ldots, X_n$ be iid $\text{Poisson}(\lambda)$

We will test $H_0: \lambda = \lambda_0$ versus $H_1: \lambda \ne \lambda_0$

Here we obtain

$$-2\log\lambda(\mathbf{x}) = 2n\left[(\lambda_0 - \hat{\lambda}) - \hat{\lambda}\log(\hat{\lambda}/\lambda_0)\right]$$

where $\hat{\lambda} = \bar{x}$

We get an approximate size α test if we reject when

$$-2\log\lambda(\mathbf{X}) > \chi^2_{1,\alpha}$$

Theorem 10.3.1

Consider testing $H_0: \theta \in \Theta_0$ versus $H_1: \theta \in \Theta_0^c$, and assume «some regularity conditions» (page 516). Then if $\theta \in \Theta_0$

$$-2\log\lambda(\mathbf{X}) \to \chi^2_{\text{df}}$$

in distribution as $n \to \infty$, where the degrees of freedom (df) is the differences between the number of free parameters specified by $\theta \in \Theta$ and the number of free parameters specified by $\theta \in \Theta_0$

Example (Multinomial)

Let $\mathbf{X}_i = (X_{i1}, X_{i2}, X_{i3});\ i = 1, 2, ..., n$ be iid with pmf

$$f(x_1, x_2, x_3 \mid p_1, p_2, p_3) = p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

where $x_1 + x_2 + x_3 = 1$ and $p_1 + p_2 + p_3 = 1$

$$\left(\sum X_{i1}, \sum X_{i2}, \sum X_{i3}\right) \sim \text{multinomial}(n, p_1, p_2, p_3)$$

We will test (Hardy-Weinberger)

$$H_0:\ p_1 = \theta^2,\ p_2 = 2\theta(1-\theta),\ p_3 = (1-\theta)^2$$

for a $\theta \in (0,1)$ versus the alternative that H_0 does not hold

Here $-2\log\lambda(\mathbf{X}) \to \chi^2_{(3-1)-1}$

Other large-sample tests

Let $X_1, X_2, ...$ be iid with pdf or pmf that depends on a real-valued parameter θ (and possibly on other parameters as well)

Suppose we have an estimator $W_n = W(X_1, X_2, ..., X_n)$ of θ with standard deviation σ_n and assume that

$$(W_n - \theta)/\sigma_n \to n(0,1)$$

Also assume that we (for each n) has an estimator S_n for σ_n such that $\sigma_n / S_n \to 1$ in probability

By Slutsky's theorem we then have that

$$\frac{W_n - \theta}{S_n} = \frac{W_n - \theta}{\sigma_n}\frac{\sigma_n}{S_n} \to n(0,1)$$

We will test $H_0: \theta = \theta_0$ versus $H_1: \theta \neq \theta_0$

An approximate test may be based on the statistic

$$Z_n = \frac{W_n - \theta_0}{S_n}$$

We then reject H_0 if $Z_n < -z_{\alpha/2}$ or $Z_n > z_{\alpha/2}$

If H_0 is true we have that $Z_n \to Z \sim n(0,1)$, so the Type I error probability becomes

$$P_{\theta_0}\left(Z_n < -z_{\alpha/2} \text{ or } Z_n > z_{\alpha/2}\right) \to P\left(Z < -z_{\alpha/2} \text{ or } Z > z_{\alpha/2}\right) = \alpha$$

The test is an asymptotically size α test

For an alternative parameter $\theta > \theta_0$ we may write

$$Z_n = \frac{W_n - \theta_0}{S_n} = \frac{W_n - \theta}{S_n} + \frac{\theta - \theta_0}{S_n}$$

Here we have $(W_n - \theta)/S_n \to n(0,1)$ in distribution and (typically) $(\theta - \theta_0)/S_n \to \infty$ in probability

Then $Z_n \to \infty$ in probability (cf. exercise 5.33), and it follows that

$$P_\theta\left(\text{reject } H_0\right) = P_\theta\left(Z_n < -z_{\alpha/2} \text{ or } Z_n > z_{\alpha/2}\right)$$
$$\geq P_\theta\left(Z_n > z_{\alpha/2}\right) \to 1$$

Wald test

Assume now that $W_n = W(X_1, X_2, ..., X_n)$ is the ML estimator of θ and S_n^2 is its estimated variance given either as (using expected information)

$$S_n^2 = \frac{1}{n I_1(W_n)}$$

or as (using observed information)

$$S_n^2 = \frac{1}{-\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} \log f(X_i \mid \theta)|_{\theta=W_n}}$$

Then $Z_n = \dfrac{W_n - \theta_0}{S_n}$ is the Wald test statistic

Example 10.3.5 (Binomial Wald test)

Let $X_1, X_2, ..., X_n$ be iid $\text{Bernoulli}(p)$

We will test $H_0: p = p_0$ versus $H_1: p \neq p_0$

Here we have the MLE $W_n = \hat{p} = \sum_{i=1}^n X_i / n$

and the variance estimator (both versions)

$$S_n^2 = \frac{\hat{p}(1-\hat{p})}{n}$$

The Wald test statistic takes the form

$$Z_n = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Score test

The score is given by

$$S(\theta) = \frac{\partial}{\partial\theta} \log L(\theta \mid \mathbf{X}) = \sum_{i=1}^n \frac{\partial}{\partial\theta} \log f(X_i \mid \theta)$$

We have that $\mathrm{E}_\theta S(\theta) = 0$ and

$$\mathrm{Var}_\theta S(\theta) = n \mathrm{Var}_\theta \left(\frac{\partial}{\partial\theta} \log f(X \mid \theta) \right) = n I_1(\theta)$$

Under $H_0: \theta = \theta_0$ it follows by the central limit theorem that the score test statistic

$$Z_S = \frac{S(\theta_0)}{\sqrt{n I_1(\theta_0)}}$$

converges in distribution to $Z \sim n(0,1)$

Example 10.3.6 (Binomial score test)

Let $X_1, X_2, ..., X_n$ be iid $\text{Bernoulli}(p)$

We will test $H_0: p = p_0$ versus $H_1: p \neq p_0$

Here we find that

$$S(p) = \frac{\sum X_i}{p} - \frac{n - \sum X_i}{1-p} = \frac{\hat{p} - p}{p(1-p)/n}$$

and

$$I_1(p) = \frac{1}{p(1-p)}$$

Thus the score test statistic takes the form

$$Z_S = \frac{S(p_0)}{\sqrt{n I_1(p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$