stats6256hyptst 625.5 - Testing statistical hypotheses

Gunnar Stefansson

December 1, 2015

Contents

1 Statistical tests and test functions

1.1 Statistical hypotheses

1.1.1 Handout

A (statistical) hypothesis is a statement concerning the parameter:

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

1.2 Test functions

1.2.1 Handout

A test of the hypothesis is a function

$$\phi : \mathbb{R}^n \to \{0, 1\}$$

where $\mathbf{X} = (X_1, \ldots, X_n)'$ is a random variable (and such that $\phi(\mathbf{X})$ is also a random variable).

Let χ be the domain of $\mathbf{X}$ (sample space) such that $\chi = R \cup A$ and $R \cap A = \emptyset$. Then we define

$$\phi(\mathbf{X}) = \begin{cases} 1 & \text{if } \mathbf{X} \in R \\ 0 & \text{if } \mathbf{X} \in A \end{cases}$$

meaning that we *accept* H_1 if $\mathbf{X} \in R$ and *accept* (or more politically correct *do not reject*) H_0 if $\mathbf{X} \in A$ (see note below on acceptance of hypotheses).

We call R the *rejection region* or *critical region* and A the *acceptance region*. Hypothesis tests produce errors if they *accept* H_0 when in fact H_1 is true and vice versa, see discussion below in 1.4.

1.3 Levels, size and power

1.3.1 Handout

Definition 1 The *power function* is the function

$$\begin{aligned} \beta(\theta) &:= P_\theta[\phi(\mathbf{X}) = 1] \\ &= P_\theta[\text{Reject } H_0]. \end{aligned}$$

The *size* of ϕ is

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$$

(the greatest rejection probability when the null hypothesis is correct).

Remark 1.1. We may also talk in general about a **level** α test if the size is no greater than α.

A typical figure needs to be inserted...power curve

From this we conclude that we never actually use the wording to accept a null hypothesis using e.g. a 5% level test since that leads to a probability of up to 95% of a wrong conclusion. Rather we use the wording that we "can not reject H_0".

A second conclusion from the figure is that we would generally prefer a test to satisfy

$$\beta(\theta_1) \geq \beta(\theta_0) \text{ if } \theta_0 \in \Theta_0 \text{ and } \theta_1 \in \Theta_1$$

Such a test is called "unbiased".

One would typically try to choose the sample size, n, such that in a test for e.g.

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

$\beta(\theta) \geq \beta_0$ if $\theta > \theta_0 + \Delta$ where β_0 and Δ are predetermined. This is a typical criterion when testing drugs (how many patients are needed to be 80% certain that a 5% level test will conclude that the new drug is better than the old one?)

Definition 2 ϕ is an unbiased test if

$$\beta(\theta') \geq \beta(\theta'')$$

for $\theta' \in \Theta_1$ and $\theta'' \in \Theta_0$. If ϕ is of level α, we note in particular that this implies

$$\beta(\theta) \geq \alpha$$

for $\theta \in \Theta_1$.

1.3.2 Examples

Example 1 Assume $X_1, \ldots, X_n \sim n(\mu, \sigma^2),\ iid, \sigma^2$ known. We intend to test

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0.$$

One possible test of size α for this task is

$$\phi(\underline{x}) = \begin{cases} 0 & \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq \zeta_{1-\alpha} \\ 1 & \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > \zeta_{1-\alpha} \end{cases}$$

The power function is

$$\begin{aligned} \beta(\mu) &= P_\mu[\text{Reject } H_0] \\ &= P_\mu[\phi(\mathbf{X}) = 1] \\ &= P_\mu\left[\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \zeta_{1-\alpha}\right] \\ &= P_\mu\left[\bar{X} > \mu_0 + \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right] \\ &= P_\mu\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha}\right] \\ &= P\left[Z > \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha}\right],\ Z \sim n(0,1) \end{aligned}$$

i.e.

$$\beta(\mu) = 1 - \Phi(\frac{\mu_0-\mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha})$$

Note that $\beta(\mu)$ is monotonically increasing in μ and $\beta(\mu) \geq \beta_0$ for $\mu \geq \mu + \Delta$ holds if $\beta(\mu_0 + \Delta) = \beta_0$.

1.4 Error types

1.4.1 Handout

Remark 1.2. Consider the possible errors in hypothesis tests

	Accept	Reject
H_0 true	(correct)	Type I error
H_0 false	Type II error	(correct)

$$\begin{aligned}\theta \in \Theta_0 : P_\theta[\text{Type I error}] &= \beta(\theta) \\ \theta \in \Theta_1 : P_\theta[\text{Type II error}] &= P_\theta[\text{accept } H_0] \\ &= P_\theta[\phi(\mathbf{X}) = 0] \\ &= 1 - \beta(\theta)\end{aligned}$$

In general we want to limit $\beta(\theta)$ on Θ_0 and we also want to limit $1-\beta(\theta)$ on Θ_1. Normally a test is chosen to satisfy

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha \quad (\text{helst } = \alpha)$$

upon which we want to compute

$$\sup_{\theta \in \Theta_1} (1 - \beta(\theta)) = \text{the greates probability of accepting an incorrect hypothesis.}$$

2 Selection of sample sizes

2.1 Determining a sample size based on power

2.1.1 Examples

Example 2 Assume $X_1, \ldots, X_n \sim n(\mu, \sigma^2)$, iid, σ^2 known. We intend to test

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$$

and will select a sample size, n, such that

$$P_\mu[\text{Reject } H_0] \geq \beta_0 \text{ ef } \mu \geq \mu + \Delta.$$

One possible test of size α for this task is

$$\phi(\mathbf{x}) = \begin{cases} 0 & \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq \zeta_{1-\alpha} \\ 1 & \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > \zeta_{1-\alpha} \end{cases}$$

The power function is

$$\begin{aligned}\beta(\mu) &= P_\mu[\text{Reject } H_0] \\ &= P_\mu[\phi(\mathbf{X}) = 1] \\ &= P_\mu\left[\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \zeta_{1-\alpha}\right] \\ &= P_\mu\left[\bar{X} > \mu_0 + \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right] \\ &= P_\mu\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha}\right] \\ &= P\left[Z > \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha}\right], \; Z \sim n(0,1)\end{aligned}$$

i.e.

$$\beta(\mu) = 1 - \Phi(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + \zeta_{1-\alpha})$$

Note that $\beta(\mu)$ is monotonically increasing in μ and $\beta(\mu) \geq \beta_0$ for $\mu \geq \mu + \Delta$ holds if $\beta(\mu_0 + \Delta) = \beta_0$. We thus select n to obtain $\beta(\mu + \Delta) = \beta_0$, i.e.

$$\begin{aligned}
& 1 - \Phi\left(\zeta_{1-\alpha} \frac{\Delta}{\sigma/\sqrt{n}}\right) = \beta_0 \\
\Rightarrow \quad & \Phi\left(\zeta_{1-\alpha} \frac{\Delta}{\sigma/\sqrt{n}}\right) = 1 - \beta_0 \\
\Rightarrow \quad & \zeta_{1-\alpha} - \frac{\Delta}{\sigma/\sqrt{n}} = \Phi^{-1}(1 - \beta_0) \\
\Rightarrow \quad & \zeta_{1-\alpha} - \frac{\Delta}{\sigma/\sqrt{n}} = \zeta_{1-\beta_0} \\
\Rightarrow \quad & \zeta_{1-\alpha} - \zeta_{1-\beta_0} = \frac{\Delta}{\sigma}\sqrt{n} \\
\Rightarrow \quad & n = \left(\sigma \frac{\zeta_{1-\alpha} - \zeta_{1-\beta_0}}{\Delta}\right)^2
\end{aligned}$$

3 The likelihood ratio test

3.1 Likelihood ratio tests

3.1.1 Handout

Likelihood ratio test (LRT)

$\phi(\mathbf{X}) = 1$ if $\lambda(\mathbf{x}) \leq c$

$$\lambda(\mathbf{x}) := \frac{\sup_{\theta \in \Theta_0} L_{\mathbf{x}}(\theta)}{\sup_{\theta \in \Theta} L_{\mathbf{x}}(\theta)}$$

Note that $c = 1 \Rightarrow \phi \equiv 1$ so c is always chosen with $c < 1$. If T is a sufficient statistic for θ, then the LRT can be based on the distribution of T (other components containing θ cancel).

3.1.2 Examples

Example 3 Assume $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim n(\mu, \sigma^2)$, iid, σ^2 known. We intend to test

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0$$

and will select a sample size, n, such that

$$P_\mu[\text{Reject } H_0] \geq \beta_0 \text{ ef } \mu \geq \mu + \Delta.$$

The LRT of size α for this task is

$$\phi(\underline{x}) = \begin{cases} 0 & \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \leq \zeta_{1-\alpha} \\ 1 & \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > \zeta_{1-\alpha} \end{cases}$$

3.2 Monotone likelihood ratio

4 UMP tests

4.1 UMP tests

4.1.1 Handout

Definition 3 Let $\mathbf{C}$ be a collection of tests. ϕ is the *uniformly most powerful* (UMP) test in $\mathbf{C}$ if

$$\beta_\phi(\theta) \geq \beta_{\phi'}(\theta)$$

for all $\theta \in \Theta_1$ and $\phi' \in \mathbf{C}$.

Remark 4.1. $\mathbf{C}$ is never the collection of all tests since we do not want $\phi = 1$ uniformly.

Theorem 4.1 (Neyman-Pearson Lemma) Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random variables with joint density f_θ where $\theta \in \Theta = \{\theta_0, \theta_1\}$.

Consider testing

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

(simple vs. simple) Let

$$\phi(\underline{x}) = \begin{cases} 0 & f_{\theta_1}(\underline{x}) < k f_{\theta_0}(\underline{x}) \\ 1 & f_{\theta_1}(\underline{x}) > k f_{\theta_0}(\underline{x}) \end{cases} \tag{1}$$

with the rejection region

$$R_\phi = \{\underline{x} : \ f_{\theta_1}(\underline{x}) > k f_{\theta_0}(\underline{x})\}$$

and $k > 0$ be such that

$$\alpha = P_{\theta_0}[\underline{\mathbf{X}} \in R_\phi] \tag{2}$$

then the test ϕ is UMP at level α. Furthermore, *if* a test, ϕ, of this type, with $k > 0$, of level α, exists, *then* one must have for any test $\tilde{\phi}$ that if $P[\tilde{\phi}(\underline{\mathbf{X}}) = 1] \leq \alpha$ and $\tilde{\phi}$ is UMP within that category *that* $\tilde{\phi} = \phi$ (outside a set of measure zero), i.e.

$$P[\phi(\underline{\mathbf{X}}) = \tilde{\phi}(\underline{\mathbf{X}})] = 1$$

for $\theta = \theta_0, \theta_1$.

Proof. „⇒âĂİ Note that ϕ as defined with (1) and (2) is automatically a level α test and we need to show that it is ϕ a UMP test among those. Let β be the power function of ϕ and ϕ' be another test also of level α (3), with power function β'. Then $0 \leq \phi'(\underline{x}) \leq 1$ and

$$\phi(\underline{x}) = 1 \quad \text{if} \quad f_{\theta_1}(\underline{x}) > k f_{\theta_0}(\underline{x})$$
$$\phi(\underline{x}) = 0 \quad \text{if} \quad f_{\theta_1}(\underline{x}) < k f_{\theta_0}(\underline{x})$$

and we obtain

$$[\phi(\underline{x}) - \phi'(\underline{x})][f_{\theta_1}(\underline{x}) - k f_{\theta_0}(\underline{x})] = \begin{cases} [1 - \phi'(\underline{x})][f_{\theta_1}(\underline{x}) - k f_{\theta_0}(\underline{x})] \geq 0 & f_{\theta_1}(\underline{x}) > k f_{\theta_0}(\underline{x}) \\ [0 - \phi'(\underline{x})][f_{\theta_1}(\underline{x}) - k f_{\theta_0}(\underline{x})] \geq 0 & f_{\theta_1}(\underline{x}) < k f_{\theta_0}(\underline{x}) \end{cases}$$

so this quantity is always ≥ 0, for all $\underline{x}$ and thus

$$0 \leq \int [\phi(\underline{x} - \phi'(\underline{x})][f_{\theta_1(\underline{x})} - kf_{\theta_0}(\underline{x})]d\underline{x}$$

$$= \beta(\theta_1) - \beta'(\theta_1) - k(\underbrace{\underbrace{\beta(\theta_0)}_{=\alpha,\text{ from}(2)} - \underbrace{\beta'(\theta_0)}_{\leq\alpha,\text{from }(3)}}_{\geq 0})$$

i.e. $\beta(\theta_1) \geq \beta'(\theta_1)$ so ϕ is UMP. Note that in the above we obtain „=âĂIJ because

$$\int \phi(\underline{x}) f_{\theta_1}(\underline{x}) d\underline{x} = E_{\theta_1}[\phi(\underline{x})] = P_{\theta_1}[\phi(\underline{x}) = 1] = \beta(\theta_1)$$

□

Homework: Find and read the rest of the proof.

Note that this test rejects if the value of the density function at θ_1 is sificiently greater than at θ_0 and therefore it is quite related to the LRT but not quite the same.

Also notice how the density function/likelihood keeps cropping up in different situations from estimation to hypothesis testing. In this case by simply knowing the density function we can find the uniformly most powerful test by using the approach of the N-P lemma.

Corrollary 4.1 If T is a sufficient statistic then the UMP test of the N-P lemma can be based on T (and the density of T).

Example 4 Let $X_1, .., X_n \sim n(\mu, \sigma^2)$ be i.i.d. with known σ^2. Suppose we want to test

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu = \mu_1$$

where $\mu_1 < \mu_0$.

In this case we know that $\bar{X} =: T$ is a sufficient statistic for μ and we thus look at the condition

$$g_{\mu_1}(t) > kg_{\mu_0}(t)$$

where g_{μ_i} is the density of T. We know that

$$T \sim n(\mu, \frac{\sigma^2}{n})$$

and

$$g_{\mu_i}(t) = \frac{1}{\sqrt{n\pi}\frac{\sigma}{\sqrt{n}}} e^{-\frac{(t-\mu_i)^2}{2\sigma^2/n}}$$

so that

$$g_{\mu_1}(t) > kg_{\mu_0}(t)$$

if and only if

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

which holds if and only if

$$(t-\mu_1)^2 < (t-\mu_0)^2 + d$$

and this is equivalent to

$$t < e$$

where k, c, d, e are some constants.

For this test to be of level α we need to define which outcome of $T = \bar{X}$ cause rejecton in such a way that the appropriate probability holds, i.e. we reject when

$$\bar{x} < \underbrace{\mu_0 - \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}}}_{e}.$$

Next note that when we want to test

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0$$

then we know that the test

$$\phi(\vec{x}) = \begin{cases} 0 & \bar{x} \geq \mu_0 - \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}} \\ 1 & \bar{x} < \mu_0 - \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}} \end{cases}$$

is UMP for every μ_1 within H_1.

Also note that to test

$$H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0$$

then a test of the form $\phi = 1$ if $\bar{x} < e$ is UMP, but the only choice of e which results in a α-size test is

$$e = \mu_0 - \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}}$$

Also recall that if any other test is as good as this one, then we have

$$\tilde{\phi}(X) = \phi(X) \quad \text{a.s.}$$

Definition 4 A family $\{g_\theta\}_{\theta\in\Theta}$ of univariate densities with $\Theta \subset \mathbb{R}$ has a *monotone likelihood ration* [MLR] if the mapping

$$t \mapsto \frac{g_{\theta_2}(t)}{g_{\theta_1}(t)}$$

is a monotonic function in t when $\theta_2 > \theta_1$, on the set $\{t : g_{\theta_1}(t) > 0 \text{ or } g_{\theta_2}(t) > 0\} = \text{supp}g_{\theta_1} \cup \text{supp}g_{\theta_2}$.

Theorem 4.2 (Karlin-Rubin) Consider testing $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ using a sufficient statistics T with a MLR. For each t_0 the test

$$\phi(t) := \begin{cases} 1 & t > t_0 \\ 0 & t \leq t_0 \end{cases}$$

is UMP test of level $\alpha := P_{\theta_0}[T > t_0]$.

Note that for the Poisson distribution we have

$$\frac{g_{\theta_1}(t)}{g_{\theta_2}(t)} = \frac{e^{-\theta_1}\frac{\theta_1^t}{t!}}{e^{-\theta_2}\frac{\theta_2^t}{t!}} = e^{\theta_2-\theta_1}\left(\frac{\theta_1}{\theta_2}\right)^t$$

which is monotonic in t.

Example 5 Suppose we have a random sample from a normal distribution with a known variance, where we want to test $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Consider the following tests:

$$\begin{aligned} \phi_1(\vec{x}) = 1 \quad \text{iff} \quad |\bar{x} - \mu_0| > \zeta_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \\ \phi_2(\vec{x}) = 1 \quad \text{iff} \quad \bar{x} > \mu_0 + \zeta_{1-\alpha} \frac{\sigma}{\sqrt{n}} \\ \phi_3(\vec{x}) = 1 \quad \text{iff} \quad \bar{x} < \mu_0 - \zeta_{1-\alpha} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

The corresponding power functions are

$$\begin{aligned} \beta_3(\mu) &= P_\mu[\phi_3(X) = 1] = P_\mu\left[X < \mu_0 - \zeta_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right] \\ &= P_\mu\left[\frac{X - \mu}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - \zeta_{1-\alpha}\right] \\ &= \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - \zeta_{1-\alpha}\right) \end{aligned}$$

$$\begin{aligned} \beta_1(\mu) &= P_\mu[\phi_1(X) = 1] = P_\mu\left[|\bar{X} - \mu_0| > \zeta_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \\ &= P_\mu\left[\bar{X} - \mu_0 > \zeta_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] + P_\mu\left[\bar{X} - \mu_0 < \zeta_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \\ &= P_\mu\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + \zeta_{1-\frac{\alpha}{2}}\right] + P_\mu\left[\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - \zeta_{1-\frac{\alpha}{2}}\right] \\ &= 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + \zeta_{1-\frac{\alpha}{2}}\right) + \Phi\left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - \zeta_{1-\frac{\alpha}{2}}\right) \\ &\to \quad 1 \quad \mu \to \infty \\ &\to \quad \alpha \quad \mu \to \mu_0 \quad . \\ &\to \quad 1 \quad \mu \to -\infty \end{aligned}$$

$\beta_1(\mu^{'}) \geq \beta(\mu^{''})$ if $\mu^{''}$ is in H_1 and $\mu^{'}$ is in H_0, so ϕ_1 is unbiased.

Remark 4.2.