STAT 801=830
Hypothesis Testing

Richard Lockhart

Simon Fraser University

STAT 801 — Fall 2012

Purposes of These Notes

- Describe hypothesis testing
- Discuss Type I and Type II error.
- Discuss level and power

Hypothesis Testing

- Hypothesis testing: a statistical problem where you must choose, on the basis of data X, between two alternatives.
- Formalized as problem of choosing between two *hypotheses*: $H_o : \theta \in \Theta_0$ or $H_1 : \theta \in \Theta_1$ where Θ_0 and Θ_1 are a partition of the model $P_\theta; \theta \in \Theta$.
- That is $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$.
- A rule for making the required choice can be described in two ways:
  1. In terms of *rejection* or *critical* region of the test.

     $$R = \{X : \text{we choose } \Theta_1 \text{ if we observe } X\}$$

  2. In terms of a function $\phi(x)$ which is equal to 1 for those x for which we choose Θ_1 and 0 for those x for which we choose Θ_0.

Hypothesis Testing

- Each ϕ corresponds to a unique rejection region $R_\phi = \{x : \phi(x) = 1\}$.
- Neyman Pearson approach treats two hypotheses asymmetrically.
- Hypothesis H_o referred to as the *null* hypothesis (traditionally the hypothesis that some treatment has no effect).
  Definition: The power function of a test ϕ (or the corresponding critical region R_ϕ) is

$$\pi(\theta) = P_\theta(X \in R_\phi) = E_\theta(\phi(X))$$

- Interested in **optimality** theory, that is, the problem of finding the best ϕ.
- A good ϕ will evidently have $\pi(\theta)$ small for $\theta \in \Theta_0$ and large for $\theta \in \Theta_1$.
- There is generally a trade off which can be made in many ways, however.

Simple versus Simple testing

- Finding a best test is easiest when the hypotheses are very precise.
- **Definition**: A hypothesis H_i is **simple** if Θ_i contains only a single value θ_i.
- The simple versus simple testing problem arises when we test $\theta = \theta_0$ against $\theta = \theta_1$ so that Θ has only two points in it.
- This problem is of importance as a technical tool, not because it is a realistic situation.
- Suppose that the model specifies that if $\theta = \theta_0$ then the density of X is $f_0(x)$ and if $\theta = \theta_1$ then the density of X is $f_1(x)$.
- How should we choose ϕ?
- To answer the question we begin by studying the problem of minimizing the total error probability.

Error Types

- **Type I error**: the error made when $\theta = \theta_0$ but we choose H_1, that is, $X \in R_\phi$.
- **Type II error**: when $\theta = \theta_1$ but we choose H_0.
- The **level** of a simple versus simple test is

$$\alpha = P_{\theta_0}(\text{We make a Type I error})$$

or

$$\alpha = P_{\theta_0}(X \in R_\phi) = E_{\theta_0}(\phi(X))$$

- Other error probability denoted β is

$$\beta = P_{\theta_1}(X \notin R_\phi) = E_{\theta_1}(1 - \phi(X)).$$

- Minimize $\alpha + \beta$, the total error probability given by

$$\begin{aligned} \alpha + \beta &= \mathrm{E}_{\theta_0}(\phi(X)) + \mathrm{E}_{\theta_1}(1 - \phi(X)) \\ &= \int [\phi(x) f_0(x) + (1 - \phi(x)) f_1(x)] dx \end{aligned}$$

Proof of NP lemma

- Problem: choose, for each x, either the value 0 or the value 1, in such a way as to minimize the integral.
- But for each x the quantity

  $$\phi(x)f_0(x) + (1 - \phi(x))f_1(x)$$

  is between $f_0(x)$ and $f_1(x)$.
- To make it small we take $\phi(x) = 1$ if $f_1(x) > f_0(x)$ and $\phi(x) = 0$ if $f_1(x) < f_0(x)$.
- It makes no difference what we do for those x for which $f_1(x) = f_0(x)$.
- Notice: divide both sides of inequalities to get condition in terms of **likelihood ratio** $f_1(x)/f_0(x)$.

Bayes procedures, in disguise

[redacted]

For each fixed λ the quantity $\beta + \lambda\alpha$ is minimized by any ϕ which has

$$\phi(x) = \begin{cases} 1 & \frac{f_1(x)}{f_0(x)} > \lambda \\ 0 & \frac{f_1(x)}{f_0(x)} < \lambda \end{cases}$$

Neyman-Pearson framework

- Neyman and Pearson suggested that in practice the two kinds of errors might well have unequal consequences.
- Suggestion: pick the more serious kind of error, label it **Type I**.
- Require rule to hold probability α of a Type I error to be no more than some prespecified level α_0.
- α_0 is typically 0.05, chiefly for historical reasons.
- Neyman-Pearson approach: minimize β subject to the constraint $\alpha \leq \alpha_0$.
- Usually this is really equivalent to the constraint $\alpha = \alpha_0$ (because if you use $\alpha < \alpha_0$ you could make R larger and keep $\alpha \leq \alpha_0$ but make β smaller.
- For discrete models, however, this may not be possible.

Binomial example: effects of discreteness

- **Example**: Suppose X is Binomial(n, p) and either $p = p_0 = 1/2$ or $p = p_1 = 3/4$.
- If R is any critical region (so R is a subset of $\{0, 1, \ldots, n\}$) then

$$P_{1/2}(X \in R) = \frac{k}{2^n}$$

  for some integer k.
- Example: to get $\alpha_0 = 0.05$ with $n = 5$: possible values of α are $0, 1/32 = 0.03125, 2/32 = 0.0625$, etc.
- Possible rejection regions for $\alpha_0 = 0.05$:

Region	α	β
$R_1 = \emptyset$	0	1
$R_2 = \{x = 0\}$	0.03125	$1 - (1/4)^5$
$R_3 = \{x = 5\}$	0.03125	$1 - (3/4)^5$

- So R_3 minimizes β subject to $\alpha < 0.05$.
- Raise α_0 slightly to 0.0625: possible rejection regions are R_1, R_2, R_3 and $R_4 = R_2 \cup R_3$.

Discreteness, test functions

- First three have same α and β as before while R_4 has $\alpha = \alpha_0 = 0.0625$ an $\beta = 1 - (3/4)^5 - (1/4)^5$.
- Thus R_4 is optimal!
- Problem: if all trials are failures "optimal" R chooses $p = 3/4$ rather than $p = 1/2$.
- But: $p = 1/2$ makes 5 failures much more likely than $p = 3/4$.
- Problem is discreteness. Solution:
- Expand set of possible values of ϕ to $[0, 1]$.
- Values of $\phi(x)$ between 0 and 1 represent the chance that we choose H_1 given that we observe x; the idea is that we actually toss a (biased) coin to decide!
- This tactic will show us the kinds of rejection regions which are sensible.
- In practice: restrict our attention to levels α_0 for which best ϕ is always either 0 or 1.
- In the binomial example we will insist that the value of α_0 be either 0 or $P_{\theta_0}(X \geq 5)$ or $P_{\theta_0}(X \geq 4)$ or ....

Binomial example: $n = 3$

- 4 possible values of X and 2^4 possible rejection regions.
- Table of levels for each possible rejection region R:

R	α	R	α
$\emptyset$	0	{3}, {0}	1/8
{0,3}	2/8	{1}, {2}	3/8
{0,1}, {0,2}, {1,3}, {2,3}	4/8	{0,1,3}, {0,2,3}	5/8
{1,2}	6/8	{0,1,2}, {1,2,3}	7/8
{0,1,2,3}	1		

- Best level $2/8$ test has rejection region $\{0, 3\}$, $\beta = 1 - [(3/4)^3 + (1/4)^3] = 36/64$.
- Best level $2/8$ test using randomization rejects when $X = 3$ and, when $X = 2$ tosses a coin with $P(H) = 1/3$, then rejects if you get H.
- Level is $1/8 + (1/3)(3/8) = 2/8$; probability of Type II error is $\beta = 1 - [(3/4)^3 + (1/3)(3)(3/4)^2(1/4)] = 28/64$.

Test functions

- **Def'n**: A hypothesis test is a function $\phi(x)$ whose values are always in $[0, 1]$.
- If we observe $X = x$ then we choose H_1 with conditional probability $\phi(X)$.
- In this case we have

$$\pi(\theta) = E_\theta(\phi(X))$$

$$\alpha = E_0(\phi(X))$$

  and

$$\beta = 1 - E_1(\phi(X))$$

- Note that a test using a rejection region C is equivalent to

$$\phi(x) = 1(x \in C)$$

The Neyman Pearson Lemma

[redacted]

When testing f_0 *vs* f_1, β *is minimized, subject to* $\alpha \le \alpha_0$ *by:*

$$\phi(x) = \begin{cases} 1 & f_1(x)/f_0(x) > \lambda \\ \gamma & f_1(x)/f_0(x) = \lambda \\ 0 & f_1(x)/f_0(x) < \lambda \end{cases}$$

where λ *is the largest constant such that*

$$P_0(f_1(X)/f_0(X) \ge \lambda) \ge \alpha_0 \text{ and } P_0(f_1(X)/f_0(X) \le \lambda) \ge 1 - \alpha_0$$

and where γ *is any number chosen so that*

$$E_0(\phi(X)) = P_0(f_1(X)/f_0(X) > \lambda) + \gamma P_0(f_1(X)/f_0(X) = \lambda) = \alpha_0$$

Value γ *is unique if* $P_0(f_1(X)/f_0(X) = \lambda) > 0$.

Binomial example again

- **Example**: Binomial(n, p) with $p_0 = 1/2$ and $p_1 = 3/4$: ratio f_1/f_0 is

$$3^x 2^{-n}$$

- If $n = 5$ this ratio is one of 1, 3, 9, 27, 81, 243 divided by 32.
- Suppose we have $\alpha = 0.05$. λ must be one of the possible values of f_1/f_0.
- If we try $\lambda = 243/32$ then

$$\begin{aligned} P_0(3^X 2^{-5} \geq 243/32) &= P_0(X = 5) \\ &= 1/32 < 0.05 \end{aligned}$$

and

$$\begin{aligned} P_0(3^X 2^{-5} \geq 81/32) &= P_0(X \geq 4) \\ &= 6/32 > 0.05 \end{aligned}$$

- So $\lambda = 81/32$.

Binomial example continued

- Since
$$P_0(3^X 2^{-5} > 81/32) = P_0(X = 5) = 1/32$$
we must solve
$$P_0(X = 5) + \gamma P_0(X = 4) = 0.05$$
for γ and find
$$\gamma = \frac{0.05 - 1/32}{5/32} = 0.12$$
- NOTE: No-one ever uses this procedure.
- Instead the value of α_0 used in discrete problems is chosen to be a possible value of the rejection probability when $\gamma = 0$ (or $\gamma = 1$).
- When the sample size is large you can come very close to any desired α_0 with a non-randomized test.

Binomial again!

- If $\alpha_0 = 6/32$ then we can either take λ to be $243/32$ and $\gamma = 1$ or $\lambda = 81/32$ and $\gamma = 0$.
- However, our definition of λ in the theorem makes $\lambda = 81/32$ and $\gamma = 0$.
- When the theorem is used for continuous distributions it can be the case that the cdf of $f_1(X)/f_0(X)$ has a flat spot where it is equal to $1 - \alpha_0$.
- This is the point of the word "largest" in the theorem.
- **Example**: If $X_1, \ldots, X_n$ are iid $N(\mu, 1)$ and we have $\mu_0 = 0$ and $\mu_1 > 0$ then

$$\frac{f_1(X_1, \ldots, X_n)}{f_0(X_1, \ldots, X_n)} = \exp\{\mu_1 \sum X_i - n\mu_1^2/2 - \mu_0 \sum X_i + n\mu_0^2/2\}$$

which simplifies to

$$\exp\{\mu_1 \sum X_i - n\mu_1^2/2\}$$

Normal, one tailed test for mean

- Now choose λ so that

$$P_0(\exp\{\mu_1 \sum X_i - n\mu_1^2/2\} > \lambda) = \alpha_0$$

- Can make it equal because $f_1(X)/f_0(X)$ has a continuous distribution.
- Rewrite probability as

$$P_0(\sum X_i > [\log(\lambda) + n\mu_1^2/2]/\mu_1) = 1 - \Phi\left(\frac{\log(\lambda) + n\mu_1^2/2}{n^{1/2}\mu_1}\right)$$

- Let z_α be upper α critical point of $N(0,1)$; then

$$z_{\alpha_0} = [\log(\lambda) + n\mu_1^2/2]/[n^{1/2}\mu_1] .$$

- Solve to get a formula for λ in terms of z_{α_0}, n and μ_1.

Simplifying rejection regions

- Rejection region looks complicated: reject if a complicated statistic is larger than λ which has a complicated formula.
- But in calculating λ we re-expressed the rejection region in terms of

$$\frac{\sum X_i}{\sqrt{n}} > z_{\alpha_0}$$

- The key feature is that this rejection region is the same for any $\mu_1 > 0$.
- WARNING: in the algebra above I used $\mu_1 > 0$.
- This is why the Neyman Pearson lemma is a lemma!

Back to basics

- **Def'n**: In the general problem of testing Θ_0 against Θ_1 the level of a test function ϕ is

$$\alpha = \sup_{\theta \in \Theta_0} E_\theta(\phi(X))$$

- The power function is

$$\pi(\theta) = E_\theta(\phi(X))$$

- A test ϕ^* is a Uniformly Most Powerful level α_0 test if
  1. ϕ^* has level $\alpha \le \alpha_o$
  2. If ϕ has level $\alpha \le \alpha_0$ then for every $\theta \in \Theta_1$ we have

$$E_\theta(\phi(X)) \le E_\theta(\phi^*(X))$$

Proof of Neyman Pearson lemma

- Given a test ϕ with level strictly less than α_0 define test

$$\phi^*(x) = \frac{1-\alpha_0}{1-\alpha}\phi(x) + \frac{\alpha_0-\alpha}{1-\alpha}$$

  which has level α_0 and β smaller than that of ϕ.
- Hence we may assume without loss that $\alpha = \alpha_0$ and minimize β subject to $\alpha = \alpha_0$.
- However, the argument which follows doesn't actually need this.

Lagrange Multipliers

- Suppose you want to minimize $f(x)$ subject to $g(x) = 0$.
- Consider first the function

$$h_\lambda(x) = f(x) + \lambda g(x)$$

- If x_λ minimizes h_λ then for any other x

$$f(x_\lambda) \le f(x) + \lambda[g(x) - g(x_\lambda)]$$

- Suppose you find λ such that solution x_λ has $g(x_\lambda) = 0$.
- Then for any x we have

$$f(x_\lambda) \le f(x) + \lambda g(x)$$

  and for any x satisfying the constraint $g(x) = 0$ we have

$$f(x_\lambda) \le f(x)$$

- So for this value of λ quantity x_λ minimizes $f(x)$ subject to $g(x) = 0$.
- To find x_λ set usual partial derivatives to 0; then to find the special x_λ you add in the condition $g(x_\lambda) = 0$.

Return to proof of NP lemma

- For each $\lambda > 0$ we have seen that ϕ_λ minimizes $\lambda\alpha + \beta$ where $\phi_\lambda = 1(f_1(x)/f_0(x) \ge \lambda)$.
- As λ increases the level of ϕ_λ decreases from 1 when $\lambda = 0$ to 0 when $\lambda = \infty$.
- There is thus a value λ_0 where for $\lambda > \lambda_0$ the level is less than α_0 while for $\lambda < \lambda_0$ the level is at least α_0.
- Temporarily let $\delta = P_0(f_1(X)/f_0(X) = \lambda_0)$.
- If $\delta = 0$ define $\phi = \phi_\lambda$.
- If $\delta > 0$ define

$$\phi(x) = \begin{cases} 1 & \frac{f_1(x)}{f_0(x)} > \lambda_0 \\ \gamma & \frac{f_1(x)}{f_0(x)} = \lambda_0 \\ 0 & \frac{f_1(x)}{f_0(x)} < \lambda_0 \end{cases}$$

  where $P_0(f_1(X)/f_0(X) > \lambda_0) + \gamma\delta = \alpha_0$.
- You can check that $\gamma \in [0, 1]$.

End of NP proof

- Now ϕ has level α_0 and according to the theorem above minimizes $\lambda_0\alpha + \beta$.
- Suppose ϕ^* is some other test with level $\alpha^* \le \alpha_0$.
- Then
$$\lambda_0\alpha_\phi + \beta_\phi \le \lambda_0\alpha_{\phi^*} + \beta_{\phi^*}$$
- We can rearrange this as
$$\beta_{\phi^*} \ge \beta_\phi + (\alpha_\phi - \alpha_{\phi^*})\lambda_0$$
- Since
$$\alpha_{\phi^*} \le \alpha_0 = \alpha_\phi$$
the second term is non-negative and
$$\beta_{\phi^*} \ge \beta_\phi$$
which proves the Neyman Pearson Lemma.

NP applied to Binomial(n, p)

- Binomial(n, p) model: test $p = p_0$ versus p_1 for a $p_1 > p_0$
- NP test is of the form

$$\phi(x) = 1(X > k) + \gamma 1(X = k)$$

where we choose k so that

$$P_{p_0}(X > k) \le \alpha_0 < P_{p_0}(X \ge k)$$

and $\gamma \in [0, 1)$ so that

$$\alpha_0 = P_{p_0}(X > k) + \gamma P_{p_0}(X = k)$$

- This rejection region depends only on p_0 and not on p_1 so that this test is UMP for $p = p_0$ against $p > p_0$.
- Since this test has level α_0 even for the larger null hypothesis it is also UMP for $p \le p_0$ against $p > p_0$.

NP lemma applied to $N(\mu, 1)$ model

- In the $N(\mu, 1)$ model consider $\Theta_1 = \{\mu > 0\}$ and $\Theta_0 = \{0\}$ or $\Theta_0 = \{\mu \le 0\}$.
- UMP level α_0 test of $H_0 : \mu \in \Theta_0$ against $H_1 : \mu \in \Theta_1$ is

$$\phi(X_1, \ldots, X_n) = 1(n^{1/2}\bar{X} > z_{\alpha_0})$$

- **Proof**: For either choice of Θ_0 this test has level α_0 because for $\mu \le 0$ we have

$$\begin{aligned} P_\mu(n^{1/2}\bar{X} & > z_{\alpha_0}) \\ &= P_\mu(n^{1/2}(\bar{X} - \mu) > z_{\alpha_0} - n^{1/2}\mu) \\ &= P(N(0,1) > z_{\alpha_0} - n^{1/2}\mu) \\ &\le P(N(0,1) > z_{\alpha_0}) \\ &= \alpha_0 \end{aligned}$$

- Notice the use of $\mu \le 0$.
- Central point: critical point is determined by behaviour on edge of null hypothesis.

Normal example continued

- Now if ϕ is any other level α_0 test then we have

$$E_0(\phi(X_1, \ldots, X_n)) \le \alpha_0$$

- Fix a $\mu > 0$.
- According to the NP lemma

$$E_\mu(\phi(X_1, \ldots, X_n)) \le E_\mu(\phi_\mu(X_1, \ldots, X_n))$$

  where ϕ_μ rejects if

$$f_\mu(X_1, \ldots, X_n)/f_0(X_1, \ldots, X_n) > \lambda$$

  for a suitable λ.
- But we just checked that this test had a rejection region of the form

$$n^{1/2}\bar{X} > z_{\alpha_0}$$

  which is the rejection region of ϕ^*.
- The NP lemma produces the same test for every $\mu > 0$ chosen as an alternative.
- So we have shown that $\phi_\mu = \phi^*$ for any $\mu > 0$.

Monotone likelihood ratio

- Fairly general phenomenon: for any $\mu > \mu_0$ the likelihood ratio f_μ/f_0 is an increasing function of $\sum X_i$.
- So rejection region of NP test always region of form $\sum X_i > k$.
- Value of k determined by requirement that test have level α_0; this depends on μ_0 not on μ_1.
- **Def'n**: The family $f_\theta; \theta \in \Theta \subset R$ has monotone likelihood ratio with respect to a statistic $T(X)$ if for each $\theta_1 > \theta_0$ the likelihood ratio $f_{\theta_1}(X)/f_{\theta_0}(X)$ is a monotone increasing function of $T(X)$.

Monotone likelihood ratio

[redacted]

For a monotone likelihood ratio family the Uniformly Most Powerful level α test of $\theta \le \theta_0$ (or of $\theta = \theta_0$) against the alternative $\theta > \theta_0$ is

$$\phi(x) = \begin{cases} 1 & T(x) > t_\alpha \\ \gamma & T(X) = t_\alpha \\ 0 & T(x) < t_\alpha \end{cases}$$

where

$$P_{\theta_0}(T(X) > t_\alpha) + \gamma P_{\theta_0}(T(X) = t_\alpha) = \alpha_0 \,.$$

Two tailed tests – no UMP possible

- Typical family where this works: one parameter exponential family.
- Usually there is no UMP test.
- Example: test $\mu = \mu_0$ against two sided alternative $\mu \neq \mu_0$.
- There is no UMP level α test.
- If there were its power at $\mu > \mu_0$ would have to be as high as that of the one sided level α test and so its rejection region would have to be the same as that test, rejecting for large positive values of $\bar{X} - \mu_0$.
- But it also has to have power as good as the one sided test for the alternative $\mu < \mu_0$ and so would have to reject for large negative values of $\bar{X} - \mu_0$.
- This would make its level too large.
- Favourite test: usual 2 sided test rejects for large values of $|\bar{X} - \mu_0|$.
- Test maximizes power subject to two constraints: first, level α; second power is minimized at $\mu = \mu_0$.
- Second condition means power on alternative is larger than on the null.