

STATS300A - Lecture 17

Dominik Rothenhaeusler
Scribed by Michael Howes

11/17/21

Contents

1	Recap	1
2	Optimality of the t-test	2
3	Confidence regions	3
3.1	Definition	3
3.2	Optimality	4
4	P-values	5

1 Recap

Last lecture we spoke about multi-dimensional exponential families and optimal tests in the presence of nuisance parameters. Suppose that $\gamma = (\theta, \lambda) \in \mathbb{R}^{d+1}$ and we have an exponential family $\{P_\gamma\}$ with densities

$$p_\gamma(x) = h(x) \exp \left\{ \theta U(x) + \sum_{j=1}^k \lambda_j T_j(x) - A(\gamma) \right\}.$$

We looked at testing

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

We saw that for fixed θ , the statistic T was sufficient for the model. We said that a test at level α had Neyman structure if:

$$\mathbb{E}_\gamma[\phi|T] = \alpha,$$

almost surely for all $\gamma \in W = \{(\theta, \lambda) : \theta = \theta_0\}$. We proved that if T is complete for $\{P_\gamma : \gamma \in W\}$, then a test is α -similar if and only if the test has Neyman structure. We then concluded that the optimal test with Neyman structure is in fact UMPU. In particular for these hypotheses

$$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0,$$

the optimal test is a function of (U, T) and is given by

$$\phi(u, t) = \begin{cases} 1 & \text{if } u > k_t, \\ \rho_t & \text{if } u = k_t, \\ 0 & \text{if } u < k_t. \end{cases}$$

where ρ_t, k_t are determined by

$$\alpha = \mathbb{P}_{\theta_0}(U(X) > k_t | T(X) = t) + \rho_t \mathbb{P}_{\theta_0}(U(X) = k_t | T(X) = t).$$

Note that since T is sufficient for fixed θ , the above probabilities are well defined.

2 Optimality of the t-test

We will now use this theory to show that the familiar t-test is in fact the UMPU test. Consider data $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. We will consider the one sided test

$$H_0 : \mu \leq 0 \text{ against } H_1 : \mu > 0.$$

The random variables X_1, \dots, X_n have joint density

$$p(x; \mu, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{n\mu}{\sigma^2} \bar{x} \right\},$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. We will thus reparametrize our family with

$$\theta = \frac{n\mu}{\sigma^2} \quad \text{and} \quad \lambda = -\frac{1}{2\sigma^2}.$$

Our null and alternative hypotheses thus become

$$H_0 : \theta \leq 0 \text{ against } H_1 : \theta > 0.$$

Our sufficient statistics are $U(x) = \bar{x}$ and $T(x) = \sum_{i=1}^n x_i^2$. By the above result, the UMPU test if one that rejects when

$$\bar{x} \geq k_{T(x)},$$

and accepts otherwise. The constant $k_{T(x)}$ is chosen such that

$$\mathbb{P}_0(\bar{X} \geq k_t | T = t) = \alpha. \tag{1}$$

We wish to transform our data to get a more recognisable test. Suppose that $h(u, t)$ is any function such that $h(\cdot, t)$ is strictly increasing for each fixed t . Then the UMPU test is equivalent to the test ϕ given by

$$\phi(u, t) = \begin{cases} 1 & \text{if } h(u, t) > k'_t, \\ 0 & \text{if } h(u, t) \leq k'_t. \end{cases}$$

where

$$\mathbb{P}_{\theta_0}(h(U, T) > k'_T | T) = \alpha.$$

We can choose

$$\begin{aligned} h(U, T) &= \frac{\bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \\ &= \frac{\bar{X}}{\sqrt{\sum_{i=1}^n X_i^2 - n\bar{X}^2}} \\ &= \frac{U}{\sqrt{T - nU^2}}. \end{aligned}$$

One can check that

- The function $h(u, t)$ is strictly increasing for fixed t .
- For fixed $\theta = 0$, then statistic $h(U, T)$ is ancillary. This is because when $\theta = 0$ our model becomes a scale model and $h(U, T)$ is scale invariant.

- The statistic T is complete and sufficient on the boundary since T is the sufficient statistic for a full rank exponential family on the boundary.
- Thus by Basu's theorem $h(U, T)$ and T are independent.

Let

$$Y = \frac{\sqrt{n}\bar{X}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}} = \frac{\sqrt{n}}{\sqrt{n-1}} h(U, T).$$

When $\theta = 0$, the statistic Y has Student's t-distribution with $n - 1$ degrees of freedom which we can write as $Y \sim t_{n-1}$. Also since Y is a function of $h(U, T)$, we know that Y is independent of T on the boundary. Thus the constraint

$$\mathbb{P}_0(h(U, T) \leq k'_T | T) = \alpha,$$

becomes

$$\mathbb{P}_0(Y \leq k'') = \alpha,$$

by independence. Thus we can take k'' to be the $1 - \alpha$ quartile of t_{n-1} . Then the test function

$$\phi = \mathbb{I}(Y > t_{1-\alpha}),$$

satisfies condition (1) and is thus UMPU.

3 Confidence regions

3.1 Definition

Definition 1. A *confidence region* with level $1 - \alpha$ is a set valued function $S(X) \subseteq \Omega$ such that for all $\theta \in \Omega$,

$$\mathbb{P}_\theta(\theta \in S(X)) \geq 1 - \alpha.$$

We will see that our optimality theory of testing translates to an optimality theory for confidence intervals. Suppose for all $\theta_0 \in \Omega$ we have a level α test ϕ_{θ_0} that test $H_0 : \theta = \theta_0$ against $H_0 : \theta > \theta_0$. Define $A(\theta_0)$ to be the acceptance region of ϕ_{θ_0} . That is

$$A(\theta_0) = \{x \in \mathcal{X} : \phi_{\theta_0}(x) = 0\}.$$

Then

$$\mathbb{P}_{\theta_0}(X \in A(\theta_0)) = 1 - \mathbb{P}_{\theta_0}(\phi_{\theta_0}(X) > 0) \geq 1 - \mathbb{E}\phi_{\theta_0} \geq 1 - \alpha.$$

This is because $\mathbb{P}_{\theta_0}(X \in A(\theta_0))$ is the probability of accepting the null under the null. Flipping this we can define

$$S(x) = \{\theta \in \Omega : x \in A(\theta)\}.$$

Thus the set $S(X)$ is the set of all null hypotheses that do not get rejected after we see data X . One can think of a subset $C \subseteq \Omega \times \mathcal{X}$ given by

$$C = \{(\theta, x) : \phi_{\theta_0}(x) = 0\}.$$

The set C could be thought of the set of all pairs (θ, x) where the parameter θ is “compatible” with the data x . Where “compatible” means that we do not reject the null corresponding to θ after seeing the data x . Note that we have

$$A(\theta) = \{x \in \mathcal{X} : (\theta, x) \in C\} \text{ and } S(x) = \{\theta \in \Omega : (\theta, x) \in C\}.$$

Thus the sets $A(\theta)$ and $S(x)$ are both sections of the set C . This immediately gives us the duality

$$x \in A(\theta) \iff \theta \in S(x).$$

This duality immediately gives that $S(X)$ is a level $1 - \alpha$ confidence region since

$$\mathbb{P}_\theta(\theta \in S(X)) = \mathbb{P}_\theta(X \in A(\theta)) \geq 1 - \alpha.$$

Remark 1. Note that the $S(X)$ is random but θ is not. This causes subtleties. For example

- The statement “ $S(X)$ has a 95% chance of covering θ ” is a correct statement.
- The statement “95% chance that θ is in $S(x_0)$ ” is not true and doesn’t make sense.

Remark 2. This process of “inverting” a test and going from $A(\theta)$ to $S(x)$ is simple mathematically but is complicated in practice. Suppose we have a model $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ but our data is actually drawn from $\tilde{P} \notin \mathcal{P}$. Then our confidence regions will be artificially small. In this setting, it may appear that our small confidence regions reflect a high degree of certainty about the true value of θ . In reality the small confidence regions are a result of *none* of the models in \mathcal{P} being compatible with our data. See the scribed notes and this [post by Andrew Gelman](#) for more details.

3.2 Optimality

Intuitively we want small confidence regions which correspond to large rejection regions. Thus we wish for optimal confidence regions to correspond to most powerful tests.

Definition 2. A $1 - \alpha$ confidence region $S(X)$ is *uniformly most accurate (UMA)* if for all $1 - \alpha$ confidence regions $S'(X)$ we have

$$\mathbb{P}_{\theta_1}(\theta_0 \in S(X)) \leq \mathbb{P}_{\theta_1}(\theta_0 \in S'(X)),$$

for all $\theta_0 \in \Omega$ and $\theta_1 > \theta_0$.

Thus a UMA test is one which, under the alternative, has the least probability of covering the null θ_0 .

Proposition 1. *If for all θ_0 , the set $A(\theta_0)$ is the acceptance region of a UMP test, then the corresponding confidence region $S(X)$ is UMA.*

Proof. Let $S^*(X)$ be a $1 - \alpha$ confidence region. Define a test $\phi_{\theta_0}^*$ by specifying the acceptance region of $\phi_{\theta_0}^*$ to be

$$A^*(\theta_0) = \{x \in \mathcal{X} : \theta_0 \in S^*(x)\}.$$

Note that

$$\mathbb{P}_{\theta_0}(X \notin A(\theta_0)) = 1 - \mathbb{P}_{\theta_0}(\theta_0 \in S^*(X)) \leq 1 - (1 - \alpha) = \alpha.$$

Thus the test $\phi_{\theta_0}^*$ has level α . Since $A(\theta_0)$ is the acceptance region of a UMP test we have for all $\theta_1 > \theta_0$,

$$\begin{aligned} \mathbb{P}_{\theta_1}(\theta_0 \in S(X)) &= \mathbb{P}_{\theta_1}(X \in A(\theta_0)) \\ &= 1 - \mathbb{P}_{\theta_1}(X \notin A(\theta_0)) \\ &\leq 1 - \mathbb{P}_{\theta_1}(X \notin A^*(\theta_0)) \\ &= \mathbb{P}_{\theta_1}(X \in A^*(\theta_0)) \\ &= \mathbb{P}_{\theta_1}(\theta_0 \in S^*(X)) \end{aligned} \quad \square$$

Remark 3. Two final comments on confidence regions:

- We thus get optimal confidence regions almost for free simply from inverting our hypothesis tests.
- We can also get optimal unbiased confidence regions when doing two sided tests or tests with nuisance parameters.

4 P-values

A informal (and probably familiar) definition of a p-value is as follows. Suppose we have a test that rejects H_0 for large values of $T(X)$ for some statistic T . If we observed some data x , then the p-value for x would be

$$p(x) = \mathbb{P}_{H_0}(T(X) > T(x)).$$

We will now give a more formal definition.

Definition 3. Suppose that for each significance level α we have a level α test ϕ_α . Suppose furthermore that the tests are monotone in the sense that $\phi_\alpha(x) \leq \phi_{\alpha'}(x)$ for all $x \in \mathcal{X}$ and all $\alpha \leq \alpha'$. The *p-value* of observed data $x \in \mathcal{X}$ is defined to be

$$p(x) = \inf\{\alpha : \phi_\alpha(x) = 1\}.$$

The monotone condition means that as the significance level decreases, the rejection regions also decrease. The p-value is thus the smallest significance level at which our family of tests $\{\phi_\alpha\}$ rejects.

Proposition 2. For all $\theta_0 \in \Omega_0$,

$$\mathbb{P}_{\theta_0}(p(X) \leq \alpha) \leq \alpha.$$

Proof. Fix $\alpha_1 > \alpha$. By monotonicity,

$$\begin{aligned} \mathbb{P}_{\theta_0}(p(X) \leq \alpha) &= \mathbb{P}_{\theta_0}(\phi_{\alpha'}(x) = 1, \text{ for all } \alpha' > \alpha) \\ &\leq \mathbb{P}_{\theta_0}(\phi_{\alpha_1}(x) = 1) \\ &\leq \alpha_1. \end{aligned}$$

Thus $\mathbb{P}_{\theta_0}(p(X) \leq \alpha) \leq \alpha_1$ for all $\alpha_1 > \alpha$. Thus $\mathbb{P}_{\theta_0}(p(X) \leq \alpha) \leq \alpha$. \square

The above proposition means that p-values “stochastically dominate” the uniform distribution on $[0, 1]$. See the scribed notes for some comments on the relationship between the informal and formal definitions of p-values.