Lecture 11

20. Likelihood ratio tests

When no UMP or UMPU tests exists one sometimes consider likelihood ratio tests (LR). You consider the *likelihood ratio*

$$LR = \frac{\sup_{\theta \in \Omega_H} f_{X|\Theta}(X \mid \theta)}{\sup_{\theta \in \Omega} f_{X|\Theta}(X \mid \theta)}.$$

To test a hypothesis you reject H if $LR < c$ for some number c. One chooses c so that the test has a certain level α. The difficulty is often that to find the appropriate c we need to know the distribution of LR. This can be difficult.

21. P-values

In the Bayesian framework $\mu_{\Theta|X}(\Omega_H \mid x)$ gives the posterior probability that the hypothesis is true given the observed data. This is quite useful information when one is interested to know more than just if the hypothesis should be rejected or not. For instance, if the hypothesis is rejected one could ask if the hypothesis was close to being not rejected and the other way around. In the Bayesian setting we get quite explicit information of this kind. In the classical framework there is no such simple way to quantify how well the data supports the hypothesis. However, in many situations the set of α-values such that the level α test would reject H will be an interval starting at some lower value p and extending to 1. In that case this p will be called the P-value.

Definition 34. Let H be a hypothesis. Let Γ be a set indexing non-randomized tests of H. That is, $\{\phi_\gamma : \gamma \in \Gamma\}$ are non-randomized tests of H. For each γ let $\varphi(\gamma)$ be the size of the test ϕ_γ. Then

$$p_H(x) = \inf\{\varphi(\gamma) : \phi_\gamma(x) = 1\},$$

is called the *P-value* of x for the hypothesis H.

Example 28. Suppose $X \sim N(\theta, 1)$ given $\Theta = \theta$ and $H : \Theta \in [-1/2, 1/2]$. The UMPU level α test of H is $\phi_\alpha(x) = 1$ if $|x| > c_\alpha$ for some number c_α. Suppose we observe $X = x = 2.18$. The test ϕ_α will reject H iff $2.18 > c_\alpha$. Since c_α increases as α decreases, the P-value is that α such that $c_\alpha = 2.18$. That is,

$$\begin{aligned}
p_H(2.18) &= \inf\{\varphi(\gamma) : \phi_\gamma(2.18) = 1\} \\
&= \inf\{\sup_{\theta \in [-1/2,1/2]} \beta_{\phi_\gamma}(\theta) : c_\gamma < 2.18\} \\
&= \sup_{\theta \in [-1/2,1/2]} \beta_{\phi_\gamma}(\theta) \text{ s.t. } c_\gamma = 2.18 \\
&= \sup_{\theta \in [-1/2,1/2]} 1 - \Phi(2.18 - \theta) + \Phi(-2.18 - \theta) \\
&= 1 - \Phi(1.68) + \Phi(-2.68) = 0.0502.
\end{aligned}$$

It is tempting to think of P-values as if it were the probability that the hypothesis is true. This interpretation can sometimes be motivated. One example is the following.

Example 29. Suppose $X \sim \text{Bin}(n,p)$ given $P = p$ and let $H : P \leq p_0$. The UMP level α test rejects H when $X > c_\alpha$ where c_α increases as α decreases. The P-value of an observed x is the value of α such that $c_\alpha = x - 1$ unless $x = 0$ in which case the P-value is equal to 1. In mathematical terms the P-value is

$$\begin{aligned} p_H(x) &= \inf\{\varphi(\gamma) : \phi_\gamma(x) = 1\} \\ &= \inf\{\sup_{p \leq p_0} \beta_{\phi_\gamma}(p) : c_\gamma < x\} \\ &= \sup_{p \leq p_0} \sum_{i=x}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=x}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}. \end{aligned}$$

Note that $p_H(0) = 1$. To see how this can correspond to the probability that the hypothesis is true, consider an improper prior of the form $\text{Beta}(0,1)$. Then the posterior distribution of P would be $\text{Beta}(x, n+1-x)$. If $x > 0$ the posterior probability that H is true is $P(Y \leq p_0)$ where $Y \sim \text{Beta}(x, n+1-x)$. Note that Y is the distribution of the xth order statistic from n IID uniform $(0,1)$ variables and hence

$$\begin{aligned} P(Y \leq p_0) &= P(x \text{ out of } n \text{ IID } U(0,1) \text{ variables less than } p_0) \\ &= \sum_{i=x}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} = p_H(x). \end{aligned}$$

Hence, $p_H(x)$ is the posterior probability that the hypothesis is true for this choice of prior.

The usual interpretation of P-values is that the P-value measures the "degree of support" for the hypothesis based on the observed data x. However, one should be aware of that P-values does not always behave in a nice way.

Example 30. Consider Example 1 but with the hypothesis $H' : \Theta \in [-0.82, 0.52]$. Note that $\Omega_{H'} \supset \Omega_H$. The UMPU level α test is $\psi_\alpha(x) = 1$ if $|x + 0.15| > d_\alpha$. If $X = x = 2.18$ then $d_\alpha = 2.33$ and

$$p_{H'}(2.18) = \Phi(-3) + 1 - \Phi(1.66) = 0.0498.$$

This is smaller than $p_H(2.18)$!!! Hence, if we interpret the P-value as the "degree of support" for the hypothesis then the degree of support for H' is less than the degree of support for H. But this is rediculus because $\Omega_{H'} \supset \Omega_H$. This shows that it is not always easy to interpret P-values.

22. Set estimation

We start with the classical notion of set estimation. Suppose we are interested in a function $g(\Theta)$. The idea of set estimation is, given an observation $X = x$, to find a set $R(x)$ that contains the true value $g(\theta)$. Typically, we want the probability $\Pr(g(\theta) \in R(X) \mid \Theta = \theta)$ to be high.

Definition 35. Let $g : \Omega \to G$ be a function, η the collection of all subsets of G and $R : \mathcal{X} \to \eta$ a function. The function R is a *coeffecient* γ *confidence set for* $g(\Theta)$

if for every $\theta \in \Omega$,

$$\{x : g(\theta) \in R(x)\} \text{ is measurable, and } \Pr(g(\theta) \in R(X) \mid \Theta = \theta) \geq \gamma.$$

The confidence set R is exact if $\Pr(g(\theta) \in R(X) \mid \Theta = \theta) = \gamma$. If $\inf_{\theta\in\Omega} \Pr(g(\theta) \in R(X) \mid \Theta = \theta) > \gamma$ the confidence set is *conservative*.

The interpretation of a level γ confidence set R is the following.

- *For any value of θ, if the experiment of generating X from $f_{X|\Theta}(\cdot \mid \theta)$ is repeated many times, the confidence set $R(X)$ will contain the true parameter $g(\theta)$ a fraction γ of the time.*

The relation between hypothesis testing and confidence sets is seen from the following theorem.

Theorem 30 (c.f. Casella & Berger Thm 9.2.2 p. 421)**.** *Let $g : \Omega \to G$ be a function.*

- *For each $y \in G$, let ϕ_y be a level α nonrandomized test of $H : g(\Theta) = y$. Let $R(x) = \{y : \phi_y(x) = 0\}$. Then R is a coefficient $1 - \alpha$ confidence set for $g(\Theta)$. The confidence set R is exact if and only if ϕ_y is α-similar for all y.*
- *Let R be a coefficient $1 - \alpha$ confidence level set for $g(\Theta)$. For each $y \in G$, let*

$$\phi_y(x) = I\{y \notin R(x)\}.$$

Then, for each y, ϕ_y has level α as a test of $H : g(\Theta) = y$. The test ϕ_y is α-similar for all y if and only if R is exact.

Proof. Let ϕ_y be a nonrandomized level α test. Then $\phi_y : \mathcal{X} \to \{0, 1\}$ is measurable for each y, because the corresponding decision rule is measurable. Hence the set

$$\{x : g(\theta) \in R(x)\} = \{x : \phi_{g(\theta)}(x) = 0\} = \phi^{-1}_{g(\theta)}(\{0\})$$

is measurable. Moreover,

$$\begin{aligned}\Pr(g(\theta) \in R(X) \mid \Theta = \theta) &= \Pr(\phi_{g(\theta)}(X) = 0 \mid \Theta = \theta) \\ &= 1 - \Pr(\phi_{g(\theta)}(X) = 1 \mid \Theta = \theta) \\ &= 1 - \beta_{\phi_{g(\theta)}}(\theta) \geq 1 - \alpha\end{aligned}$$

with equality iff $\beta_{\phi_{g(\theta)}}(\theta) = \alpha$. That is, there is equality iff $\phi_{g(\theta)}$ is α-similar. This proves the first part.

Let R be a coefficient $1 - \alpha$ confidence set and $\phi_y(x) = I\{y \notin R(x)\}$. Then

$$\phi^{-1}_{g(\theta)}(\{0\}) = \{x : \phi_{g(\theta)}(x) = 0\} = \{x : g(\theta) \in R(x)\}$$

which is measurable. Hence $\phi_{g(\theta)}$ is measurable and then the corresponding decision rule is measurable. Moreover,

$$\begin{aligned}\beta_{\phi_{g(\theta)}}(\theta) &= \Pr(\phi_{g(\theta)}(X) = 1 \mid \Theta = \theta) \\ &= 1 - \Pr(\phi_{g(\theta)}(X) = 0 \mid \Theta = \theta) \\ &= 1 - \Pr(g(\theta) \in R(X) \mid \Theta = \theta) \leq \alpha.\end{aligned}$$

We have equality in the last step iff R is exact, and this is the same as $\phi_{g(\theta)}$ being α-similar. □

Example 31. Let $X_1, \dots, X_n$ be conditionally IID $\mathrm{N}(\mu, \sigma^2)$ given $(M, \Sigma) = (m, \sigma)$. Let $X = (X_1, \dots, X_n)$. The UMPU level α test of $H : M = y$ is $\phi_y = 1$ if $\sqrt{n}(\overline{x} - y)/s > T_{n-1}^{-1}(1-\alpha/2)$ where T_{n-1} is the cdf of a student-t distribution with $n-1$ degrees of freedom. This translates into the confidence interval $[\overline{x} - T_{n-1}^{-1}(1-\alpha/2)s/\sqrt{n}, \overline{x} + T_{n-1}^{-1}(1-\alpha/2)s/\sqrt{n}]$.

One can suspect that there is an analog of UMP tests for confidence sets. The corresponding concept is called UMA (uniformly most accurate) confidence set.

Definition 36. Let $g : \Omega \to G$ be a function and R a coefficient γ confidence set for $g(\Theta)$. Let $B : G \to \eta$ be a function such that $y \notin B(y)$. Then R is *uniformly most accurate (UMA) coefficient γ against B* if for each $\theta \in \Theta$ and each $y \in B(g(\theta))$ and each coefficient γ confidence set T for $g(\Theta)$

$$\Pr(y \in R(X) \mid \Theta = \theta) \leq \Pr(y \in T(X) \mid \Theta = \theta).$$

For $y \in G$, the set $B(y)$ can be thought of a set of points that you don't want to include in the confidence set. The condition above says that for $y \in B(g(\theta))$ (we don't want y in the confidence set) the probability that the confidence set contains y is smaller if we use R than with any other level α confidence set T.

Note also that the condition $y \notin B(y)$ implies that $g(\theta) \notin B(g(\theta))$. We would like the true value $g(\theta)$ to be in the confidence set so it should not be in $B(g(\theta))$.

Now we can see how UMP tests are related to UMA confidence sets.

Theorem 31. *Let $g(\theta) = \theta$ for all θ and let $B : \Omega \to \eta$ be as in Definition 36. Put*

$$B^{-1}(\theta) = \{y : \theta \in B(y)\}.$$

Suppose $B^{-1}(\theta)$ is nonempty for each θ. For each θ, let ϕ_θ be a test and $R(x) = \{y : \phi_y(x) = 0\}$. Then ϕ_θ is UMP level α for testing $H : \Theta = \theta$ vs $A : \Theta \in B^{-1}(\theta)$ for all θ if and only if R is UMA coefficient $1-\alpha$ randomized against B.

Proof. Suppose first that for each θ, ϕ_θ is UMP level α for testing H vs A. Let T be another coefficient $1-\alpha$ randomized confidence set. Let $\theta \in \Omega$ and $y \in B(\theta)$. We need to show that

$$P_\theta(y \in R(X)) \leq P_\theta(y \in T(X)).$$

First we can observe that $\theta \in B^{-1}(y)$. Define $\psi(x) = I(y \notin T(x))$. This test has level α for testing $H' : \Theta = y$ vs $A' : \Theta \in B^{-1}(y)$. Since ϕ_y is UMP for H' vs A' it follows that $\beta_\psi(\theta) \leq \beta_{\phi_y}(\theta)$. That is,

$$\begin{aligned} P_\theta(y \in R(X)) &= 1 - P(y \notin R(X)) = 1 - E_\theta \phi_y(X) = 1 - \beta_{\phi_y}(\theta) \\ &\leq 1 - \beta_\psi(\theta) = 1 - E_\theta \psi(X) = 1 - P_\theta(y \notin T(X)) = P_\theta(y \in T(X)). \end{aligned}$$

This shows the desired inequality.

For the other direction suppose R is UMA coefficient $1-\alpha$ randomized confidence set against B. For $\theta \in \Omega$ let ψ_θ be a level α test of H and put $T(x) = \{y : \psi_y(x) = 0\}$. Then T is a coefficient $1-\alpha$ confidence set. Put

$$\Omega' = \{(y, \theta) : y \in \Omega, \theta \in B(y)\} = \{(y, \theta) : \theta \in \Omega, y \in B^{-1}(\theta)\}.$$

For each $(\theta, y) \in \Omega'$ we know $P_y(\theta \in R(X)) \leq P_y(\theta \in T(X))$. By the calculation above this is equivalent to $\beta_{\phi_\theta}(y) \geq \beta_{\psi_\theta}(y)$ for all $\theta \in \Omega$ and all $y \in B^{-1}(\theta)$. That is, ϕ_θ is UMP level α for H vs A. $\square$

The theorem shows how to get a UMA confidence set from a UMP test. Nevertheless, one has to be careful when constructing confidence sets. **See Example 5.57, p. 319 in Schervish.** This example shows that in some situations a naive computation of the UMP level α test and the corresponding UMA confidence set can sometimes be inadequate.

22.1. **Prediction sets.** One attempt to do predictive inference in the classical setting is the following.

Definition 37. Let $V : S \to \mathcal{V}_0$ be a random quantity. Let η be all subsets of $\mathcal{V}_0$ and $R : \mathcal{X} \to \eta$ a function. If

$$\{(x, v) : v \in R(x)\} \text{ is measurable, and } \Pr(V \in R(X) \mid \Theta = \theta) \geq \gamma, \text{ for each } \theta \in \Omega,$$

then R is called a *coefficient γ prediction set for V*.

22.2. **Bayesian set estimation.** In the Bayesian setting we can, given a set $R(x) \subset G$ compute the posterior probability $\Pr(g(\Theta) \in R(x) \mid X = x)$. However, to construct confidence sets we should go the other way and specify a coefficient γ and then construct R to have this probability. There can be many such sets. To choose between them one usually argues according to one of the following:

- Determine a number t such that $T = \{\theta : f_{\Theta|X}(\theta \mid x) \geq t\}$ satisfies $\Pr(\Theta \in T \mid X = x) = \gamma$. This is called the *highest posterior density region (HDP)*.
- If $\Omega \subset \mathbb{R}$ and a bounded interval is desired, choose the endpoints to be the $(1-\gamma)/2$ and $(1+\gamma)/2$ quantiles of the posterior distribution of Θ.

Sometimes (for instance in Casella & Berger) such sets are called *credibility sets*.