Lecture 9

17. Hypothesis testing

A special type of decision problem is hypothesis testing. We partition the parameter space into $\Omega_H \cup \Omega_A$ with $\Omega_H \cap \Omega_A = \emptyset$. We write

$$H : \Theta \in \Omega_H$$
$$A : \Theta \in \Omega_A.$$

A decision problem is called *hypothesis testing* if $\aleph = \{0, 1\}$ and

$$L(\theta, 1) > L(\theta, 0),\ \theta \in \Omega_H,$$
$$L(\theta, 1) < L(\theta, 0),\ \theta \in \Omega_A.$$

The action $a = 1$ is called *rejecting the hypothesis* and $a = 0$ is called *not rejecting the hypothesis*. Note that the condition above says that the loss is greater if we reject the hypothesis than if we do not reject when the hypothesis is true, and similarly the loss is greater if we do not reject when the hypothesis is false.

- Type I error: If we reject H when H is true we have made a type I error.
- Tupe II error: If we do not reject H when H is false we have made a type II error.

We can put

$$\begin{aligned} L(\theta, 0) &= 0,\ \theta \in \Omega_H, && \text{not reject when H true,} \\ L(\theta, 1) &= c,\ \theta \in \Omega_H, && \text{reject when H true,} \\ L(\theta, 0) &= 1,\ \theta \in \Omega_A, && \text{not reject when H is false,} \\ L(\theta, 1) &= 0,\ \theta \in \Omega_A, && \text{reject when H false.} \end{aligned}$$

Such a loss function is called a $0 - 1 - c$ loss function. If $c = 1$ it is a $0 - 1$ loss function.

Here are some standard definitions:

- The *test function* of a test is the function $\phi : \mathcal{X} \to [0, 1]$ given by

  $$\phi(x) = \delta(\{1\}; x),$$

  the probability of choosing $a = 1$ (reject) when we observe x.
- The *power function* of a test ϕ is

  $$\beta_\phi(\theta) = E_\theta \phi(X).$$

  It is the probability to reject H given $\Theta = \theta$.
- The *characteristic operating curve* is $\rho_\phi = 1 - \beta_\phi$. It is the probability of not rejecting H given $\Theta = \theta$.
- The *size* of a test is $\sup_{\theta \in \Omega_H} \beta_\phi(\theta)$. It is the maximum probability of rejecting H when H is true.
- The test is called *level* α if its size is at most α.

17.1. **Hypothesis testing in Bayesian case.** In the Bayesian setting the hypothesis is simply the decision problem with $\aleph = \{0,1\}$ and $0-1-c$-loss function. Hence, the posterior risk is

$$\begin{aligned} r(1 \mid x) &= c\mu_{\Theta|X}(\Omega_H \mid x), \\ r(0 \mid x) &= \mu_{\Theta|X}(\Omega_A \mid x). \end{aligned}$$

The optimal decision is to take $a = 1$ "reject the hypothesis" if

$$c\mu_{\Theta|X}(\Omega_H \mid x) < \mu_{\Theta|X}(\Omega_A \mid x).$$

This is equivalent to rejecting the hypothesis if

$$\mu_{\Theta|X}(\Omega_H \mid x) < \frac{1}{1+c},$$

that is, if the posterior odds are too low.

Simple-simple hypothesis.

Definition 27. Let $\Omega = \{\theta_0, \theta_1\}$. The hypothesis $H : \Theta = \theta_0$ versus $A : \Theta = \theta_1$ is called a *simple-simple hypothesis*.

Let us write f_0 for the density when $\Theta = \theta_0$ and f_1 when $\Theta = \theta_1$. Then, if $p_0 = \mu_\Theta(\theta_0)$ and $p_1 = 1 - p_0$, we have

$$\mu_{\Theta|X}(\Omega_H \mid x) = \frac{p_0 f_0(x)}{p_0 f_0(x) + p_1 f_1(x)}.$$

We reject the hypothesis when this ratio is less than $1/(1+c)$.

One-sided tests.

Definition 28. Let $\Omega \subset \mathbb{R}$. A hypothesis of the form $H : \Theta \le \theta_0$ or $H : \Theta \ge \theta_0$ is called a *one-sided hypothesis*.

A test with test function

$$\phi(x) = \begin{cases} 1, & x > x_0, \\ \gamma, & x = x_0, \\ 0, & x < x_0, \end{cases} \quad \text{or} \quad \phi(x) = \begin{cases} 1, & x < x_0, \\ \gamma, & x = x_0, \\ 0, & x > x_0, \end{cases}$$

is called a *one-sided test*.

Bayesian hypothesis testing leads to one-sided tests if the posterior $\mu_{\Theta|X}(\Omega_H \mid x)$ is monotone. Suppose, for instance, $H : \Theta \le \theta_0$ and $A : \Theta > \theta_0$. If $\mu_{\Theta|X}(\Omega_H \mid x)$ is decreasing in x, then rejecting the hypothesis for x_0 implies that one should reject the hypothesis for all $x > x_0$. Thus, the formal Bayes rule is to use a test with test function of the form

$$\phi(x) = \begin{cases} 1, & x > x_0, \\ \gamma, & x = x_0, \\ 0, & x < x_0, \end{cases}$$

for some x_0. Similar remarks apply if $\mu_{\Theta|X}(\Omega_H \mid x)$ is increasing (then the other form of one-sided tests should be used) as well as for one-sided hypothesis of the form $H : \Theta \ge \theta_0$.

Definition 29. If $\Omega \subset \mathbb{R}$, $\mathcal{X} \subset \mathbb{R}$, and $dP_\theta/d\nu = f_{X|\Theta}(x \mid \theta)$, then the parametric family is said to have *monotone likelihood ration (MLR)* if for each $\theta_1 < \theta_2$ the ratio

$$\frac{f_{X|\Theta}(x \mid \theta_2)}{f_{X|\Theta}(x \mid \theta_1)}$$

is a monotone function if x a.e. $P_{\theta_1} + P_{\theta_2}$ in the same direction (inreasing or decreasing) for each $\theta_1 < \theta_2$. If the ratio is increasing the family has *increasing MLR*. If the ratio is decreasing the family has *decreasing MLR*.

Example 29. Let $f_{X|\Theta}$ form a one-parameter exponential family with natural parameter θ and natural statistic $T(X)$. Recall that (Lecture 4) T has a density of the form $c(\theta)\exp\{\theta t\}$ w.r.t. a measure ν_T'. Then

$$\frac{f_{T|\Theta}(t \mid \theta_2)}{f_{T|\Theta}(t \mid \theta_1)} = \frac{c(\theta_1)}{c(\theta_2)}\exp\{t(\theta_2 - \theta_1)\}$$

is increasing for each $\theta_1 < \theta_2$. Hence, it has increasing MLR.

The MLR condition is sufficient to come up with one-sided tests.

Theorem 26. *Suppose the parametric family $f_{X|\ \Theta}$ is MLR and μ_Θ is a prior. Then the posterior probability $\mu_{\Theta|X}([\theta_0, \infty) \mid x)$ and $\mu_{\Theta|X}((-\infty, \theta_0] \mid x)$ are monotone in x for each θ_0.*

Proof. Let us prove the case of increasing MLR and the interval $[\theta_0, \infty)$. We show that $\mu_{\Theta|X}([\theta_0, \infty) \mid x)$ is nondecreasing. Take $x_1 < x_2$. Then

$$\begin{aligned}
&\frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_2)}{\mu_{\Theta|X}((\infty,\theta_0) \mid x_2)} - \frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_1)}{\mu_{\Theta|X}((\infty,\theta_0) \mid x_1)} \\
&= \frac{\int_{[\theta_0,\infty)} f_{X|\Theta}(x_2 \mid \theta)\mu_\Theta(d\theta)}{\int_{(-\infty,\theta_0)} f_{X|\Theta}(x_2 \mid \theta)\mu_\Theta(d\theta)} - \frac{\int_{[\theta_0,\infty)} f_{X|\Theta}(x_1 \mid \theta)\mu_\Theta(d\theta)}{\int_{(-\infty,\theta_0)} f_{X|\Theta}(x_1 \mid \theta)\mu_\Theta(d\theta)} \\
&= \frac{\int_{[\theta_0,\infty)}\int_{(-\infty,\theta_0)}[f_{X|\Theta}(x_2 \mid \theta_2)f_{X|\Theta}(x_1 \mid \theta_1) - f_{X|\Theta}(x_2 \mid \theta_1)f_{X|\Theta}(x_1 \mid \theta_2)]\mu_\Theta(d\theta_1)\mu_\Theta(d\theta_2)}{\int_{(-\infty,\theta_0)} f_{X|\Theta}(x_2 \mid \theta)\mu_\Theta(d\theta)\int_{(-\infty,\theta_0)} f_{X|\Theta}(x_1 \mid \theta)\mu_\Theta(d\theta)}.
\end{aligned}$$

Since the family has increaisgn MLR the integrand in the numerator is nonnegative for each $x_1 < x_2$ and $\theta_1 < \theta_2$. Hence

$$\begin{aligned}
0 &\le \frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_2)}{\mu_{\Theta|X}((\infty,\theta_0) \mid x_2)} - \frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_1)}{\mu_{\Theta|X}((\infty,\theta_0) \mid x_1)} \\
&= \frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_2)}{1-\mu_{\Theta|X}([\theta_0,\infty) \mid x_2)} - \frac{\mu_{\Theta|X}([\theta_0,\infty) \mid x_1)}{1-\mu_{\Theta|X}([\theta_0,\infty) \mid x_1)}.
\end{aligned}$$

The result follows since $x/(1-x)$ is increasing on $[0, 1]$. □

Corollary 2. *Suppose $f_{X|\Theta}$ form a parametric family with MLR and μ_Θ is a prior. Suppose we are testing a one-sided hypothesis against the corresponding one-sided alternative with a $0-1-c$ loss function. Then one-sided tests are formal Bayes rules.*

Proof. We prove the case of increasing MLR and $H : \Theta \ge \theta_0$, $A : \Theta < \theta_0$. Then $\mu_{\Theta|X}([\theta_0, \infty) \mid x)$ is increasing in x and $\mu_{\Theta|X}(-\infty, \theta_0) \mid x)$ is decreasing in x. For a decision rule δ with test function $\phi(x)$ we have

$$r(\delta \mid x) = c\phi(x)\mu_{\Theta|X}([\theta_0, \infty) \mid x) + (1 - \phi(x))\mu_{\Theta|X}(-\infty, \theta_0) \mid x).$$

It is optimal to choose

$$\phi(x) = \begin{cases} 1, & \text{if } \mu_{\Theta|X}([\theta_0, \infty) \mid x) < 1/(1+c), \\ 0, & \text{if } \mu_{\Theta|X}([\theta_0, \infty) \mid x) > 1/(1+c), \\ \gamma, & \text{if } \mu_{\Theta|X}([\theta_0, \infty) \mid x) = 1/(1+c). \end{cases}$$

The one-sided test with

$$\phi(x) = \begin{cases} 1, & x < x_0, \\ \gamma, & x = x_0, \\ 0, & x < x_0, \end{cases} \quad \text{or} \quad \phi(x) = \begin{cases} 0, & x > x_0, \\ \gamma, & x = x_0, \\ 0, & x > x_0, \end{cases}$$

can be written in the form above with x_0 that solves $(1+c)^{-1} = \mu_{\Theta|X}([\theta_0, \infty) \mid x_0)$. Hence, it is a formal Bayes rule with this loss function. $\square$

Point hypothesis. In this section we are concerned with hypothesis of the form $H : \Theta = \theta_0$ vs $A : \Theta \neq \theta_0$. Again it seems reasonable that tests of the form ψ in Theorem 1 are appropriate.

Bayes factors. The Bayesian methodology also has a way of testing point hypothesis. Suppose we want to test $H : \Theta = \theta_0$ against $A : \Theta \neq \theta_0$. If the prior has a continuous distribution then the prior probability and the posterior probability of Ω_H is 0. Either one could replace the hypothesis with a small interval or use what is called Bayes factors. Suppose we assign a probability p_0 to the hypothesis so that the prior is

$$\mu_\Theta(A) = p_0 I_A(\theta_0) + (1 - p_0)\lambda(A \setminus \{\theta_0\})$$

where λ is a probability measure on (Ω, τ). Then the joint density of (X, Θ) is

$$f_{X,\Theta}(x, \theta) = p_0 f_{X|\Theta}(x \mid \theta_0) I_{\{\theta=\theta_0\}} + (1 - p_0) f_{X|\Theta}(x \mid \theta) I_{\{\theta\neq\theta_0\}}.$$

The posterior density is

$$f_{\Theta|X}(\theta \mid x) = p_1 I_{\{\theta=\theta_0\}} + (1 - p_1)\frac{f_{X|\Theta}(x \mid \theta)}{f_X(x)} I_{\{\theta\neq\theta_0\}}$$

where $p_1 = p_0 f_{X|\Theta}(x \mid \theta_0)/f_X(x)$ is the posterior probability of the hypothesis. Note that

$$\frac{p_1}{1 - p_1} = \frac{p_0}{1 - p_0} \frac{f_{X|\Theta}(x \mid \theta_0)}{\int f_{X|\Theta}(x \mid \theta)\lambda(d\theta)}.$$

The second factor on the right is called the *Bayes factor*. Thus, the posterior odds in favor of the hypothesis is the prior odds for the hypothesis times the Bayes factor. It tells you how much the odds has increased or decreased after observing the data. Testing a point hypothesis can be stated as "reject H if the Bayes factor is below a threshold k".

18. Classical hypothesis testing

18.1. **Most powerful tests.** In the classical setting the risk function of a test is closely related to the power function. If the loss function is $0-1-c$ then the risk function is

$$R(\theta,\phi)=\begin{cases} c\beta_\phi(\theta), & \theta\in\Omega_H,\\ 1-\beta_\phi(\theta), & \theta\in\Omega_A.\end{cases}$$

Hence, most attention is on the power function.

Definition 30. Suppose $\Omega=\Omega_H\cup\{\theta_1\}$, where $\theta_1\notin\Omega_H$. A level α test ϕ is called *most powerful (MP) level* α if, for every other level α test ψ, $\beta_\psi(\theta_1)\le\beta_\phi(\theta_1)$.

A level α test ϕ is called *uniformly most powerful (UMP) level* α if, for every other level α test ψ, $\beta_\psi(\theta)\le\beta_\phi(\theta)$ for all $\theta\in\Omega_A$.

Example 30. Suppose that $\Omega=\{\theta_0,\theta_1\}$ and $f_i(x)$ is the density of P_{θ_i} w.r.t. some measure ν for both values of θ (one can take $\nu=P_{\theta_0}+P_{\theta_1}$). Let

$$\begin{aligned} H&:\Theta=\theta_0,\\ A&:\Theta=\theta_1.\end{aligned}$$

Then, the Neyman-Pearson fundamental lemma yields the form of the test functions of all admissible tests. The test corresponding to the test function $\phi_{k,\gamma}$ is

$$\begin{aligned} \text{Reject } H \text{ if } & f_1(x)>kf_0(x),\\ \text{Do not reject } H \text{ if } & f_1(x)<kf_0(x),\\ \text{Reject } H \text{ with probability } \gamma(x) \text{ if } & f_1(x)=kf_0(x).\end{aligned}$$

All these tests are MP of their respective levels. Indeed, since these decision rules form a minimal complete class we have for any other test ψ with the same level that $R(\theta_0,\phi_{k,\gamma})=R(\theta_0,\psi)$, i.e. $\beta_\phi(\theta_0)=\beta_\psi(\theta_0)$ and $R(\theta_1,\phi_{k,\gamma})\le R(\theta_1,\psi)$, i.e. $\beta_\psi(\theta_1)\le\beta_{\phi_{k,\gamma}}(\theta_1)$.

18.2. **Simple-simple hypothesis.**

Definition 31. Let $\Omega=\{\theta_0,\theta_1\}$. The hypothesis $H:\Theta=\theta_0$ versus $A:\Theta=\theta_1$ is called a *simple-simple hypothesis*.

Simple-simple hypothesis are covered by Neyman-Pearson's fundamental lemma. We will now take a closer look at them. Suppose for simplicity that the loss function is $0-1$ so the risk function is

$$R(\theta,\phi)=\begin{cases} \beta_\phi(\theta), & \theta=\theta_0,\\ 1-\beta_\phi(\theta), & \theta=\theta_1.\end{cases}$$

Then the risk function can be represented by a point $(\alpha_0,\alpha_1)\in[0,1]^2$ where $\alpha_0=R(\theta_0,\phi)$ and $\alpha_1=R(\theta_1,\phi)$. The risk set R corresponding to this decision problem is a subset of $[0,1]^2$. Note that the test function $\phi(x)\equiv\alpha_0$ corresponds to the risk function $(\alpha_0,1-\alpha_0)$. As we let α_0 vary in $[0,1]$ we see that R contains the line $y=1-x$, $x\in[0,1]$. Furthermore, R is symmetric around $(1/2,1/2)$. Indeed, if the risk function of a test ϕ is (a,b) then the risk function of the test $1-\phi$ is $(1-a,1-b)$, so this point is also in R. We know from Lecture 9, that R is convex.

Recall the definition of the lower boundary ∂_L of the risk set. By definition ∂_L contains the admissible rules. Hence, the lower boundary is contained in the risk set R, so the risk set is closed from below. By symmetry around $(1/2, 1/2)$ the risk set is closed.

Recall that the admissible rules are given by the minimal complete class $\mathcal{C}$ in Neyman-Pearson's fundamental lemma. Hence, the good tests to choose to test a simple-simple hypothesis are the tests in the class $\mathcal{C}$.

From the Bayesian perspective the tests in $\mathcal{C}$ are Bayes rules with respect to different priors. Indeed, each $\phi_{k,\gamma}$ is a Bayes rule with respect a prior $\lambda = (\lambda_0, \lambda_1)$ with $0 < \lambda_0 < 1$. To see this, note that a Bayes rule w.r.t. λ corresponds to a point (α_0, α_1) that minimizes

$$r(\lambda, \phi) = \lambda_0 \alpha_0 + \lambda_1 \alpha_1.$$

This is the inner product of (λ_0, λ_1) with (α_0, α_1) and graphically it is easy to see that the minimum is on the lower boundary ∂_L of the risk set.

One-sided tests. Recall the definition of one-side hypothesis and one-sided test from Definition 28 (Lecture 15).

In this section we are interested in finding one-sided UMP tests. Recall a test ϕ is UMP level α if for any other level α test ψ, $\beta_\psi(\theta) \leq \beta_\phi(\theta)$ for all $\theta \in \Omega_A$. (Level α is that $\sup_{\theta \in \Omega_H} \beta_\phi(\theta) \leq \alpha$).

In the Bayesian context we saw that the notion of MLR was convenient to determine formal Bayes rules. The situation is similar here.

Theorem 27. *If $f_{X|\Theta}$ forms a parametric family with increasing MLR, then any test of the form*

$$\phi(x) = \begin{cases} 1, & x > x_0, \\ \gamma, & x = x_0, \\ 0, & x < x_0, \end{cases}$$

has nondecreasing power function. Each such test is UMP of its size for testing $H : \Theta \leq \theta_0$ versus $A : \Theta > \theta_0$, for each θ_0. Moreover, for each $\alpha \in [0, 1]$ and each $\theta_0 \in \Omega$ there exists x_0 and $\gamma \in [0, 1]$ such that ϕ is UMP level α for testing H versus A.

Proof. First we show ϕ has nondecreasing power function. Let $\theta_1 < \theta_2$. By Neyman-Pearson's fundamental lemma the MP test of $H_1 : \Theta = \theta_1$ versus $A_1 : \Theta = \theta_2$ is

$$\phi(x) = \begin{cases} 1, & f_{X|\Theta}(x \mid \theta_2) > k f_{X|\Theta}(x \mid \theta_1), \\ \gamma(x), & f_{X|\Theta}(x \mid \theta_2) = k f_{X|\Theta}(x \mid \theta_1), \\ 0, & f_{X|\Theta}(x \mid \theta_2) < k f_{X|\Theta}(x \mid \theta_1). \end{cases}$$

Since the MLR is increasing we can write ϕ as

$$\phi(x) = \begin{cases} 1, & x > t^-, \\ \gamma(x), & t_- \leq x \leq t^-, \\ 0, & x < t_-, \end{cases} \tag{18.1}$$

For ϕ of this form put $\alpha' = \beta_\phi(\theta_1)$. Let $\phi_{\alpha'} \equiv \alpha'$. Then, since ϕ is MP we must have $\beta_\phi(\theta_2) \geq \alpha'$. Hence ϕ has nondecreasing power function.

Next, we show that we can have arbitrary level. Take $\alpha \in [0, 1]$ and put

$$x_0 = \begin{cases} \inf\{x : P_{\theta_0}(-\infty, x] \geq 1 - \alpha, & \alpha < 1, \\ \inf\{x : P_{\theta_0}(-\infty, x] > 0, & \alpha = 1. \end{cases}$$

Then $\alpha^* = P_{\theta_0}(x_0, \infty) \leq \alpha$ and $P_{\theta_0}(\{x_0\}) \geq \alpha - \alpha^*$. Now we take ϕ of the form in (18.1) with $t_- = x_0 = t^-$ and $\gamma(x_0) = \gamma^*$. Then

$$\beta_\phi(\theta_0) = E_{\theta_0}[\phi(X)] = P_{\theta_0}(x_0, \infty) + \gamma^* P_{\theta_0}(\{x_0\}) = \alpha^* + \gamma^* P_{\theta_0}(\{x_0\}).$$

This is equal to α if we take

$$\gamma^* = \begin{cases} 0 & P_{\theta_0}(\{x_0\}) = 0, \\ \frac{\alpha^* - \alpha}{P_{\theta_0}(\{x_0\})} & P_{\theta_0}(\{x_0\}) > 0. \end{cases}$$

This ϕ is MP level α for testing $H_0 = \Theta = \theta_0$ versus $A : \Theta = \theta_1$ for every $\theta_0 < \theta_1$, since it is the same test for all θ_1. Hence ϕ is UMP for testing H_0 versus A. Since $\beta_\phi(\theta)$ is nondecreasing, ϕ has level α for H, so it is UMP level α for testing H versus A. $\square$

Remark 3. There are similar results for testing $H : \Theta \geq \theta_0$ when the family has increasing MLR and for testing either $H : \Theta \leq \theta_0$ or $H : \Theta \geq \theta_0$ when the family has decreasing MLR. The test ϕ has to be modified, interchanging the condition $x > x_0$ to $x < x_0$ accordingly.

Lecture 10

18.3. Two-sided hypothesis.

Definition 32. If $H : \Theta \geq \theta_2$ or $\Theta \leq \theta_1$ and $A : \theta_1 < \Theta < \theta_2$, then the hypothesis is two-sided. If $H : \theta_1 \leq \Theta \leq \theta_2$ and $A : \Theta > \theta_2$ or $\Theta < \theta_1$, then the alternative is two-sided.

Let us consider two-sided hypothesis.

Theorem 28 (c.f. Schervish, Thm 4.82, p. 249)**.** *In a one-parameter exponential family with natural parameter Θ, if $\Omega_H = (-\infty, \theta_1] \cup [\theta_2, \infty)$ and $\Omega_A = (\theta_1, \theta_2)$, with $\theta_1 < \theta_2$ a test of the form*

$$\phi_0(x) = \begin{cases} 1, & c_1 < x < c_2, \\ \gamma_i, & x = c_i, \\ 0, & c_1 > x \text{ or } c_2 < x, \end{cases}$$

with $c_1 \leq c_2$ minimizes $\beta_\phi(\theta)$ for all $\theta < \theta_1$ and for all $\theta > \theta_2$, and it maximizes $\beta_\phi(\theta)$ for all $\theta \in (\theta_1, \theta_2)$ subject to $\beta_\phi(\theta_i) = \alpha_i$ for $i = 1, 2$ where $\alpha_i = \beta_{\phi_0}(\theta_i)$. If $c_1, c_2, \gamma_1, \gamma_2$ are chosen so that $\alpha_1 = \alpha_2 = \alpha$, then ϕ_0 is UMP level α.

Lemma 5. *Let ν be a measure and $p_0, p_1, \dots, p_n$ ν-integrable functions. Put*

$$\phi_0(x) = \begin{cases} 1, & p_0(x) > \sum_{i=1}^n k_i p_i(x), \\ \gamma(x), & p_0(x) = \sum_{i=1}^n k_i p_i(x), \\ 0, & p_0(x) < \sum_{i=1}^n k_i p_i(x), \end{cases}$$

where $0 \leq \gamma(x) \leq 1$ and k_i are constants. Then ϕ_0 minimizes $\int [1 - \phi(x)] p_0(x) \nu(dx)$ subject to the constraints

$$\int \phi(x) p_j(x) \nu(dx) \leq \int \phi_0(x) p_j(x) \nu(dx), \text{ for } j \text{ such that } k_j > 0,$$
$$\int \phi(x) p_j(x) \nu(dx) \geq \int \phi_0(x) p_j(x) \nu(dx), \text{ for } j \text{ such that } k_j < 0$$

Similarly

$$\tilde{\phi}_0(x) = \begin{cases} 0, & p_0(x) > \sum_{i=1}^n k_i p_i(x), \\ \gamma(x), & p_0(x) = \sum_{i=1}^n k_i p_i(x), \\ 1, & p_0(x) < \sum_{i=1}^n k_i p_i(x), \end{cases}$$

Then maximizes $\int [1 - \phi(x)] p_0(x) \nu(dx)$ subject to the constraints

$$\int \phi(x) p_j(x) \nu(dx) \geq \int \tilde{\phi}_0(x) p_j(x) \nu(dx), \text{ for } j \text{ such that } k_j > 0,$$
$$\int \phi(x) p_j(x) \nu(dx) \leq \int \tilde{\phi}_0(x) p_j(x) \nu(dx), \text{ for } j \text{ such that } k_j < 0$$

Proof. Use Lagrange multipliers. See Schervish pp. 246-247. □

Proof of Theorem. A one parameter exponential family has density $f_{X|\Theta}(x \mid \theta) = h(x)c(\theta)e^{\theta x}$ with respect to some measure ν. Suppose we include $h(x)$ in ν (that is, we define a new measure ν' with density $h(x)$ with respect to ν) so that the density is $c(\theta)e^{\theta x}$ with respect to ν'. Then we abuse notation and write ν for ν'.

Let θ_1 and θ_2 be as in the statement of the theorem and let θ_0 be another parameter value. Define $p_i(x) = c(\theta_i)e^{\theta_i x}$ $i = 0, 1, 2$.

Suppose $\theta_0 \in (\theta_1, \theta_2)$. On this region we want to maximize $\beta_\phi(\theta_0)$ subject to $\beta_\phi(\theta_i) = \beta_{\phi_0}(\theta_i)$. Note that $\beta_\phi(\theta_i) = \int \phi(x) p_i(x)\nu(dx)$ and maximizing $\beta_\phi(\theta_0)$ is equivalent to minimizing $\int [1 - \phi(x)] p_0(x)\nu(dx)$. It seems we want to apply the Lemma with $k_1 > 0$ and $k_2 > 0$. Applying the Lemma gives the test maximizing $\beta_\phi(\theta_0)$ as

$$\phi(x) = \begin{cases} 1, & p_0(x) > \sum_{i=1}^2 k_i p_i(x), \\ \gamma(x), & p_0(x) = \sum_{i=1}^2 k_i p_i(x), \\ 0, & p_0(x) < \sum_{i=1}^2 k_i p_i(x), \end{cases}$$

Note that

$$p_0(x) > \sum_{i=1}^2 k_i p_i(x) \iff 1 > k_1 \frac{c(\theta_1)}{c(\theta_0)} e^{(\theta_1 - \theta_0)x} + k_2 \frac{c(\theta_2)}{c(\theta_0)} e^{(\theta_2 - \theta_0)x}.$$

Put $b_i = \theta_i - \theta_0$ and $a_i = k_i c(\theta_i)/c(\theta_0)$, and we get

$$1 > a_1 e^{b_1 x} + a_2 e^{b_2 x}.$$

We want the break points to be c_1 and c_2 so we need to solve two equations

$$a_1 e^{b_1 c_1} + a_2 e^{b_2 c_1} = 1,$$
$$a_1 e^{b_1 c_2} + a_2 e^{b_2 c_2} = 1,$$

for a_1, a_2. The solution exists (check yourself) and has $a_1 > 0$, $a_2 > 0$ as required (recall that we wanted $k_1, k_2 > 0$). So putting $k_i = a_i c(\theta_0)/c(\theta_i)$ gives the right choice of k_i in the minimizing test. Since the minimizing θ does not depend on θ_0 we get the same test for all $\theta_0 \in (\theta_1, \theta_2)$.

For $\theta_0 < \theta_1$ or $\theta_0 > \theta_2$ we want to minimize $\beta_\phi(\theta_0)$. This is done in a similar way using the second part of the Lemma.

Some work also remains to show that one can choose $c_1, c_2, \gamma_1, \gamma_2$ so that the test has level α. We omitt the details. Full details are in the proof of Theorem 4.82, p. 249 in Schervish "Theory of Statistics". □

Interval hypothesis. In this section we consider hypothesis of the form $H : \Theta \in [\theta_1, \theta_2]$ versus $A : \Theta \notin [\theta_1, \theta_2]$, $\theta_1 < \theta_2$. This will be called an interval hypothesis. Unfortunately there is not always UMP tests for testing H vs A. For an example in the case of *point hypothesis* see Example 8.3.19 in Casella & Berger (p. 392). On the other hand, comparing with the situation when the hypothesis and alternative are interchanged, one could guess that the test $\psi = 1 - \phi_0$, with ϕ_0 as in Theorem 28 is a good tests. One can show that this test satisfies a weaker criteria than UMP.

Definition 33. A test ϕ is *unbiased level* α if if has level α and if $\beta_\phi(\theta) \geq \alpha$ for all $\theta \in \Omega_A$. If ϕ is UMP among all unbiased tests it is called UMPU (uniformly most powerful unbiased) level α.

If $\Omega \subset \mathbb{R}^k$, a test ϕ is called *α-similar* if $\beta_\phi(\theta) = \alpha$ for each $\theta \in \overline{\Omega}_H \cap \overline{\Omega}_A$.

Proposition 5. *The following holds:*

(i) *If ϕ is unbiased level α and β_ϕ is continuous, then ϕ is α-similar.*
(ii) *If ϕ is UMP level α, then ϕ is unbiased level α.*
(iii) *If β_ϕ continuous for each ϕ and ϕ_0 is UMP level α and α-similar then ϕ_0 is UMPU.*

Proof. (i) $\beta_\phi \le \alpha$ on Ω_H, $\beta_\phi \ge \alpha$ on Ω_A and β_ϕ continuous implies $\beta_\phi = \alpha$ on $\overline{\Omega}_H \cap \overline{\Omega}_A$.
(ii) Let $\phi^\alpha \equiv \alpha$. Since ϕ is UMP $\beta_\phi \ge \beta_{\psi^\alpha} = \alpha$ on Ω_A. Hence ϕ is unbiased level α.
(iii) Since ϕ^α is α-similar and ϕ_0 is UMP among α-similar tests we have $\beta_{\phi_0} \ge \beta_{\psi^\alpha} = \alpha$ on Ω_A. Hence ϕ_0 is unbiased level α. By continuity of β_ϕ any α-similar level α test ϕ is unbiased level α so $\beta_{\phi_0} \ge \beta_\phi$ on Ω_A. Thus ϕ_0 is UMPU. □

Theorem 29. *Consider a one parameter exponential family with its natural parameter and the hypothesis $H : \Theta \in [\theta_1, \theta_2]$ vs $A : \Theta \notin [\theta_1, \theta_2]$, $\theta_1 < \theta_2$. Let ϕ be any test of H vs A. Then there is a test ψ of the form*

$$\psi(x) = \begin{cases} 1, & x \notin (c_1, c_2), \\ \gamma_i, & x = c_i, \\ 0, & x \in (c_1, c_2), \end{cases}$$

such that $\beta_\psi(\theta_i) = \beta_\phi(\theta_i)$, $\beta_\psi(\theta) \le \beta_\phi(\theta)$ on Ω_H and $\beta_\psi(\theta) \ge \beta_\phi(\theta)$ on Ω_A. Moreover, if $\beta_\psi(\theta_i) = \alpha$, then ψ is UMPU level α.

Proof. Put $\alpha_i = \beta_\phi(\theta_i)$. One can find a test ϕ_0 of the form in Theorem 3, Lecture 15, such that $\beta_{\phi_0}(\theta_i) = 1 - \alpha_i$ (we have not proved this in class, see Lemma 4.81, p. 248) and then this ϕ_0 minimizes the power function on $(\infty, \theta_1) \cup (\theta_2, \infty)$ and maximizes it on (θ_1, θ_2) among all tests ϕ' subject to $\beta_{\phi'}(\theta_i) = 1 - \alpha_i$. But then, $\psi = 1 - \phi_0$ satisfies $\beta_\psi(\theta_i) = \alpha_i$ and maximizes the power function on $(\infty, \theta_1) \cup (\theta_2, \infty)$ and minimizes it on (θ_1, θ_2) among all test subject to the restrictions. This proves the first part.

If $\beta_\psi(\theta_i) = \alpha$, then ψ is α-similar and hence ψ is UMP level α among all α-similar tests. For a one parameter exponential family β_ϕ is continuous for all ϕ so (iii) in the Proposition shows that ψ is UMPU level α. □

Point hypothesis. In this section we are concerned with hypothesis of the form $H : \Theta = \theta_0$ vs $A : \Theta \ne \theta_0$. Again it seems reasonable that tests of the form ψ in Theorem 29 are appropriate.

Theorem 30. *Consider a one parameter exponential family with natural parameter and $\Omega_H = \{\theta_0\}$, $\Omega_A = \Omega \setminus \{\theta_0\}$ where θ_0 is in the interior of Ω. Let ϕ be any test of H vs A. Then there is a test of the form ψ in Theorem 29 such that*

$$\begin{aligned} \beta_\psi(\theta_0) &= \beta_\phi(\theta_0), \\ \partial_\theta \beta_\psi(\theta_0) &= \partial_\theta \beta_\phi(\theta_0) \end{aligned} \tag{18.2}$$

and for $\theta \ne \theta_0$, $\beta_\psi(\theta)$ is maximized among all tests satisfying the two equalities. Moreover, If ψ has size α and $\partial\beta_\psi(\theta_0) = 0$, then it is UMPU level α.

Sketch of proof. First one need to show that there are tests of the form ψ that satisfies the equialities.

Put $\alpha = \beta_\phi(\theta_0)$ and $\gamma = \partial_\theta \beta_\phi(\theta_0)$. Let ϕ_u be the UMP level u test for testing $H : \Theta \ge \theta_0$ vs $A : \Theta < \theta_0$, and for $0 \le u \le \alpha$ put

$$\phi'_u(x) = \phi_u(x) + 1 - \phi_{1-\alpha+u}(x).$$

Note that, for each $0 \le u \le \alpha$,

$$\beta_{\phi'_u}(\theta_0) = \beta_{\phi_u}(\theta_0) + 1 - \beta_{\phi_{1-\alpha+u}}(\theta_0) = u + 1 - (1 - \alpha + u) = \alpha.$$

Then ϕ'_u has the right form, i.e. as in Theorem 29. The test $\phi'_0 = 1 - \phi_{1-\alpha}$ has level α and is by construction UMP level α for testing $H' : \Theta = \theta_0$ vs $A' : \Theta > \theta_0$. Similarly $\phi'_\alpha = \phi_\alpha$ is UMP level α for testing $H' : \Theta = \theta_0$ vs $A'' : \Theta < \theta_0$. We claim that

(i) $\partial_\theta \beta_{\phi'_\alpha}(\theta_0) \le \gamma \le \partial_\theta \beta_{\phi'_0}(\theta_0)$.
(ii) $u \mapsto \partial_\theta \beta_{\phi_u}(\theta_0)$ is continuous.

The first is easy to see intuitively in a picture. A complete argument is in Lemma 4.103, p. 257 in Schervish. The second is a bit involved and we omitt it here. See p. 259 for details. From (i) and (ii) we conclude that there is a test of the form ψ (i.e. ϕ'_{u_0} for some u_0) that satisfies (18.2).

It remains to show that this test maximizes the power function among all level α tests satisfying the restriction on the derivative. For any test η we have

$$\begin{aligned}\partial_\theta \beta_\eta(\theta_0) &= \partial_\theta \int_{\mathcal{X}} \eta(x)c(\theta)e^{\theta x}\nu(dx)|_{\theta=\theta_0} \\ &= \int_{\mathcal{X}} \eta(x)(c(\theta_0)x + c'(\theta_0))e^{\theta_0 x}\nu(dx) \\ &= E_{\theta_0}[X\eta(X)] - \beta_\eta(\theta_0)E_{\theta_0}[X],\end{aligned}$$

where we used integration by parts in the last step. Hence, $\partial_\theta \beta_\eta(\theta_0) = \gamma$ iff

$$E_{\theta_0}[X\eta(X)] = \gamma + \alpha E_{\theta_0}[X].$$

Note that the RHS does not depend on η. For any $\theta_1 \neq \theta_0$ and put

$$\begin{aligned}p_0(x) &= c(\theta_1)e^{\theta_1 x} \\ p_1(x) &= c(\theta_0)e^{\theta_0 x} \\ p_2(x) &= xc(\theta_0)e^{\theta_0 x}.\end{aligned}$$

Then

$$E_{\theta_0}[X\eta(X)] = \int \eta(x)p_2(x)\nu(dx)$$

We know from last time (or Lemma 4.78, p. 247 using Lagrange multipliers) that a test of the form

$$\eta_0(x) = \begin{cases} 1, & p_0(x) > \sum_{i=1}^2 k_i p_i(x), \\ \gamma(x), & p_0(x) = \sum_{i=1}^2 k_i p_i(x), \\ 0, & p_0(x) < \sum_{i=1}^2 k_i p_i(x), \end{cases}$$

where $0 \le \gamma(x) \le 1$ and k_i are constants, maximizes $\int \eta(x)p_0(x)\nu(dx)$ subject to the constraints

$$\begin{aligned}\int \eta(x)p_i(x)\nu(dx) &\le \int \eta_0(x)p_i(x)\nu(dx), \text{ for } i \text{ such that } k_i > 0, \\ \int \eta(x)p_i(x)\nu(dx) &\ge \int \eta_0(x)p_i(x)\nu(dx), \text{ for } i \text{ such that } k_i < 0.\end{aligned}$$

That is, it maximizes $\beta_\eta(\theta_1)$ subject to

$$\begin{aligned}\beta_\eta(\theta_0) &\le (\ge)\beta_{\eta_0}(\theta_0) \\ \mathrm{E}_{\theta_0}[\eta(X)] &\le (\ge)\, \mathrm{E}_{\theta_0}[\eta_0(X)],\end{aligned}$$

where the direction of the inequalities depend on k_i.

The test η_0 corresponds to rejecting the hypothesis if

$$e^{(\theta_1-\theta_0)x} > k_1 + k_2 x.$$

By choosing k_1 and k_2 appropriety we can get a test of the form ψ which is the same for all $\theta_1 \neq \theta_0$.

Finally, we want to show that if the test is level α and $\partial_\theta \beta_\phi(\theta_0) = 0$, the the test is UMPU level α. For this we only need to show that $\partial_\theta \beta_\phi(\theta_0) = 0$ is necessary for the test to be unbiased. But this is obvious because, since the power function is differentiable, if the derivative is either strictly positive or strictly then the power function is less than α in some left- or right-neighborhood of θ_0. □

19. Nuisance parameters

Suppose the parameter Θ is multidimensional $\Theta = (\Theta_1, \dots, \Theta_k)$ and Ω_H is of lower dimension than k, say d dimensional $d < k$, then the remaining parameters are called *nuisance parameters*.

Let $\mathcal{P}_0$ be a parametric family $\mathcal{P}_0 = \{P_\theta : \theta \in \Omega\}$. Let $G \subset \Omega$ be a subparameter space and $\mathcal{Q}_0 = \{P_\theta : \theta \in G\}$ be a subfamily of $\mathcal{P}_0$. Let Ψ be the parameter of the family $\mathcal{Q}_0$.

Definition 34. If T is a sufficient statistic for Ψ in the classical sense, then a test ϕ has *Neyman structure relative to G and T* if $E_\theta[\phi(X) \mid T = t]$ is constant as a function of t P_θ-a.s. for all $\theta \in G$.

Why is Neyman structure a good thing? It is because it sometimes enables a procedure to obtain UMPU tests. Suppose that we can find statistic T such that the distribution of X conditional on T has one-dimensional parameter. Then we can try to find the UMPU test among all tests that have level α conditional on T. Then this test will also be UMPU level α unconditionally.

There is a connection here with α-similar tests.

Lemma 6. *If H is a hypothesis and $\mathcal{Q}_0 = \{P_\theta : \theta \in \overline{\Omega}_H \cap \overline{\Omega}_A\}$ and ϕ has Neyman structure, then ϕ is α-similar.*

Proof. Since

$$\beta_\phi(\theta) = E_\theta[\phi(X)] = E_\theta[E_\theta[\phi(X) \mid T]]$$

and $E_\theta[\phi(X) \mid T]$ is constant for $\theta \in \overline{\Omega}_H \cap \overline{\Omega}_A$ we see that $\beta_\phi(\theta)$ is constant on $\overline{\Omega}_H \cap \overline{\Omega}_A$. □

There is a converse under some slightly stronger assumptions.

Lemma 7. *If T is a boundedly complete sufficient statistic for the subparameter space $G \subset \Omega$, then every α-similar test on G has Neyman structure relative to G and T.*

Proof. By α-similarity $E_\theta[E[\phi(X) \mid T] - \alpha] = 0$ for all $\theta \in G$. Since T is boundedly complete we must have $E[\phi(X) \mid T] = \alpha$ P_θ-a.s. for all $\theta \in G$. □

Now we can use this to find conditions when UMPU tests exists.

Proposition 6. *Let $G = \overline{\Omega}_H \cap \overline{\Omega}_A$. Let I be an index set such that $G = \cup_{i \in I} G_i$ is a partition of G. Suppose there exists a statistic T that is boundedly complete sufficient statistic for each subparameter space G_i. Assume that the power function* *of every test is continuous. If there is a UMPU level α test ϕ among those which have Neyman structure relative to G_i and T for all $i \in I$, then ϕ is UMPU level α.*

Proof. From last time (Proposition 5(i)) we know that continuity of the power function implies that all unbiased level α tests are α-similar. By the previous lemma every α-similar test has Neyman structure. Since ϕ is UMPU level α among all such tests it is UMPU level α. □

In the case of exponential families one can prove the following.

Theorem 31. *Let $X = (X_1, \dots, X_k)$ have a k-parameter exponential family with $\Theta = (\Theta_1, \dots, \Theta_k)$ and let $U = (X_2, \dots, X_k)$.*

(i) *Suppose that the hypothesis is one-sided or two-sided concerning only Θ_1. Then there is a UMP level α test conditional on U, and it is UMPU level α.*
(ii) *If the hypothesis concerns only Θ_1 and the alternative is two-sided, then there is a UMPU level α test conditional on U, and it is also UMPU level α.*

Proof. Suppose that the density is

$$f_{X|\Theta}(x \mid \theta) = c(\theta)h(x)\exp\{\sum_{i=1}^{k} \theta_i x_i\}.$$

Let $G = \overline{\Omega}_H \cap \overline{\Omega}_A$. The conditional density of X_1 given $U = u = (x_1, \dots, x_k)$ is

$$f_{X_1|\Theta,U}(x_1 \mid \theta, u) = \frac{c(\theta)h(x)e^{\sum_{i=1}^{k}\theta_i x_i}}{\int c(\theta)h(x)e^{\sum_{i=1}^{k}\theta_i x_i}dx_1} = \frac{h(x)e^{\theta_1 x_1}}{\int h(x)e^{\theta_1 x_1}dx_1}.$$

This is a one-parameter exponential family with natural parameter Θ_1.

For the hypothesis (one- or two-sided) we have that G is either $G_0 = \{\theta : \theta_1 = \theta_1^0\}$ some θ_1^0 or the union $G_1 \cup G_1$ with $G_1 = \{\theta : \theta_1 = \theta_1^1\}$, $G_2 = \{\theta : \theta_1 = \theta_1^2\}$. The parameter $\Psi = (\Theta_2, \dots, \Theta_k)$ has a complete sufficient statistic $U = (X_2, \dots, X_k)$.

Let η be an unbiased level α test. Then by Proposition 5(i), η is α-similar on G_0, G_1, and G_2. By the previous lemma η has Neyman structure. Moreover, for every test η, $\beta_\eta(\theta) = E_\theta[E_\theta[\eta(X) \mid U]]$ so a test that maximizes the conditional power function uniformly for $\theta \in \Omega_A$ subject to contraints also maximizes the marginal power function subject to the same contstraints.

For part (i) in the conditional problem given $U = u$ there is a level α test that maximizes the conditional power function uniformly on Ω_A subject to having Neyman structure. Since every unbiased level α test has Neyman structure and the power function is the expectation of the conditional power function ϕ is UMPU level α.

For part (ii), if $\Omega_H = \{\theta : c_1 \le \theta_1 \le c_2\}$ with $c_1 < c_2$, then as above the conditional UMPU level α test ϕ is also UMPU level α.

For a point hypothesis $\Omega_H = \{\theta : \theta_1 = \theta_1^0\}$ we must take partial derivative of $\beta_\eta(\theta)$ with respect to θ_1 at every point in G. A little more work... □

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 35. 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 31. 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 32. Suppose $X \sim \text{Bin}(n,p)$ given $P = p$ and let $H : P \le p_0$. 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

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

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 $P(Y \le p_0)$ 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

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

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 33. 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 36. 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 32 (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 34. Let $X_1, \ldots, X_n$ be conditionally IID $\mathrm{N}(\mu, \sigma^2)$ given $(M, \Sigma) = (m, \sigma)$. Let $X = (X_1, \ldots, X_n)$. 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 37. 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 $g(\Theta)$

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

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 33. *Let $g(\theta) = \theta$ for all θ and let $B : \Omega \to \eta$ be as in Definition 37. 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

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

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 $\beta_\psi(\theta) \le \beta_{\phi_y}(\theta)$. That is,

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

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 $P_y(\theta \in R(X)) \le P_y(\theta \in T(X))$. By the calculation above this is equivalent to $\beta_{\phi_\theta}(y) \ge \beta_{\psi_\theta}(y)$ 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 38. 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*.