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 26. 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 27. Let $\Omega \subset \mathbb{R}$. A hypothesis of the form $H : \Theta \leq \theta_0$ or $H : \Theta \geq \theta_0$ 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, $H : \Theta \leq \theta_0$ 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 $H : \Theta \geq \theta_0$.

Definition 28. 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 26. 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 24. *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. □

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 29. 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 27. 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

$$H:\Theta=\theta_0,$$
$$A:\Theta=\theta_1.$$

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

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

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 30. 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 27 (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 25. *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 $\alpha^* = P_{\theta_0}(x_0, \infty) \le \alpha$ and $P_{\theta_0}(\{x_0\}) \ge \alpha - \alpha^*$. 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. □

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