Anna Janicka

Mathematical Statistics 2021/2022
Lecture 9

1. Hypothesis Testing – Neyman-Pearson Lemma

In the previous lecture, we looked at the steps of a statistical procedure and discussed how to describe the basic properties of a statistical test. In this lecture, we will tackle the problem of choosing the best test. The basic rule of comparing tests is the following: for a given set of null and alternative hypotheses, for a given significance level, the test which is *more powerful* is better. Formally, we will say that:

Definition 1. *Let $X \sim P_\theta$, where $\{P_\theta : \theta \in \Theta\}$ be a statistical model. Let $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$ describe the null and alternative hypotheses, respectively (we have $\Theta_0 \cap \Theta_1 = \emptyset$). Let C_1 and C_2 be critical regions associated with two tests, both at a significance level α. The test with critical region C_1 is* **more powerful** *than the test with critical region C_2, if*

$$\forall \theta \in \Theta_1 : P_\theta(C_1) \geq P_\theta(C_2) \text{ and } \exists \theta_1 \in \Theta_1 : P_{\theta_1}(C_1) > P_{\theta_1}(C_2)$$

In other words, a test is more powerful than another test, if it is equally as good for all possible values of parameter θ from the alternative hypothesis range, and there is at least one value of the parameter for which it is strictly better. For a given set of null and alternative hypotheses, $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$, we may also define:

Definition 2. *C^* is a* **uniformly most powerful test** *(UMPT) for significance level α, if:*
(1) C^* is a test at significance level α, i.e. for any $\theta \in \Theta_0 : P_\theta(C^*) \leq \alpha$ and
(2) for any test C at significance level α, we have, for any $\theta \in \Theta_1$:

$$P_\theta(C^*) \geq P_\theta(C).$$

In other words, a UMPT is a test which has a power at least as large as any other test of the same hypotheses. If the alternative hypothesis space is simple (Θ_1 only contains of one element), the word 'uniformly' is redundant.

1.1. Likelihood ratio test for testing simple hypotheses.

Let us assume that we wish to test two simple hypotheses: $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$. We can rephrase this to become $H_0 : X \sim f_0$ against $H_1 : X \sim f_1$, where f_0 and f_1 are densities of the distributions defined by θ_0 and θ_1, respectively.

Theorem 1 (Neyman-Pearson Lemma)**.** *Let*

$$C^* = \left\{ x \in \mathcal{X} : \frac{f_1(x)}{f_0(x)} > c \right\},$$

such that $P_0(C^*) = \alpha$ and $P_1(C^*) = 1 - \beta$. Then, for any $C \subseteq \mathcal{X}$, we have:
if $P_0(C) \leq \alpha$, then $P_1(C) \leq 1 - \beta$. In other words, the test with critical region C^* is the most powerful test for testing H_0 against H_1.

The philosophy behind this test is the following: we compare the chances of obtaining the data that we observe under the null and alternative hypotheses. If the likelihood of obtaining the data is much higher for the alternative hypothesis than for the null hypothesis (c times as high, where c is calculated so as to satisfy the condition for the significance level), we reject the null in favor of the alternative.

In many cases – especially when the space of observations is more than one-dimensional – it is easier to write the critical region of the test as $C^* = \{x : \ln f_1(x) - \ln f_0(x) > c_1\}$. Obviously, this expression may (usually) be simplified.

Examples:

(1) Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$ with unknown parameter μ and known σ. Suppose that we want to test the null hypothesis that $H_0 : \mu = 0$ against the alternative that $H_1 : \mu = 1$. We will construct the most powerful test using the Neyman-Pearson Lemma:

We have that $f_0(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum (x_i - 0)^2/2\sigma^2}$, while $f_1(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum (x_i - 1)^2/2\sigma^2}$.

Therefore, the ratio

$$\frac{f_1}{f_0} = \frac{\frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum (x_i - 1)^2/2\sigma^2}}{\frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum (x_i - 0)^2/2\sigma^2}} = e^{\sum (x_i - 0)^2/2\sigma^2 - \sum (x_i - 1)^2/2\sigma^2}.$$

The condition $\frac{f_1}{f_0} > c$ is equivalent to $\ln f_1 - \ln f_0 > \ln c = c_1$, so we must have that

$$\sum (x_i - 0)^2/2\sigma^2 - \sum (x_i - 1)^2/2\sigma^2 > c_1,$$

$$\sum (x_i - 0)^2 - \sum (x_i - 1)^2 > c_1 \cdot 2\sigma^2 = c_2,$$

$$\sum ((x_i - 0)^2 - (x_i - 1)^2) = \sum (x_i - 0 - x_i + 1)(x_i - 0 + x_i - 1) > c_2,$$

$$\sum (2x_i - 1) > c_2,$$

$$\sum (2x_i) > c_2 + n = c_3,$$

$$\sum x_i > c_3/2 = c_4,$$

or, equivalently,

$$\bar{X} > c_4/n = c_5.$$

Therefore, the most powerful test for testing the null hypothesis that $H_0 : \mu = 0$ against the alternative that $H_1 : \mu = 1$ may be expressed in the following form: if the sample average of observations is greater than some constant, we reject the null. An equivalent version of the test would state: if the sample sum of observations is greater than some constant (note that this is a different constant than in the previous sentence), we reject the null.

Now, in order to proceed and provide the constant (critical value) in either version of the test, we need to assume a specific significance level α first. We will then determine a constant for the given significance level, by equating the required significance level to the probability of falling into the critical region if the null hypothesis is true. In the case of our example, since we know that $\bar{X}$ has a normal distribution with mean μ and variance equal to σ^2/n, under the null hypothesis we have that

$$P_0(\bar{X} > c_5) = P_0(\frac{\bar{X} - 0}{\sigma}\sqrt{n} > \frac{c_5 - 0}{\sigma}\sqrt{n}) = 1 - \Phi(\frac{c_5 - 0}{\sigma}\sqrt{n}).$$

If this last value is to be equal to α, then we must have that $\frac{c_5 - 0}{\sigma}\sqrt{n} = u_{1-\alpha}$, where $u_{1-\alpha}$ is the quantile of rank $1 - \alpha$ of the standard normal distribution. Therefore, we should take

$$c_5 = \frac{\sigma u_{1-\alpha}}{\sqrt{n}}.$$

Therefore, the most powerful test for testing the null hypothesis that $H_0 : \mu = 0$ against the alternative that $H_1 : \mu = 1$ for a significance level α has the following form: we reject the null hypothesis in favor of the alternative if $\bar{X} > \frac{\sigma u_{1-\alpha}}{\sqrt{n}}$.

Please note that we absolutely do not need to go back to the initial form of the test (with f_1/f_0 and the constant c) – we have an equivalent, but simpler formula.

(2) Let again $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution $N(\mu, \sigma^2)$ with unknown parameter μ and known σ. Suppose now that we want to test the null hypothesis that $H_0 : \mu = 0$ against the alternative that $H_1 : \mu = -1$. The Neyman-Pearson Lemma procedure will yield almost the same results as in the example above, with one small but significant change. When simplifying the expression f_1/f_0, where this time $f_1(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\sum (x_i+1)^2/2\sigma^2}$, we will get

$$\sum ((x_i - 0)^2 - (x_i+1)^2) = \sum (x_i - 0 - x_i - 1)(x_i - 0 + x_i - 1) > c_2,$$

which leads to a negative left hand side:

$$-\sum (2x_i - 1) > c_2,$$

which in turn translates to

$$\sum (2x_i - 1) < c_2.$$

Therefore, the final form of the test will be $\bar{X} < \frac{\sigma u_\alpha}{\sqrt{n}}$ (we take a quantile of rank α since we take the value of the CDF and not the value of the tail of the CDF). We reject the null hypothesis in favor of the alternative if the sample average is too small.

1.2. **Likelihood ratio test for testing composite hypotheses.** Now, let us assume that we wish to test two more general hypotheses: $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1$. We can rephrase this to become $H_0 : X \sim f_0(\theta_0, \cdot)$ for some $\theta_0 \in \Theta_0$, against the alternative $H_1 : X \sim f_1(\theta_1, \cdot)$, for some $\theta_1 \in \Theta_1$, where f_0 and f_1 are densities of the distributions for θ_0 and θ_1, respectively. Note that this is a generalization of the assumptions of the Neyman-Pearson Lemma, in that we allow statistical models (with unknown parameters) in the hypothesis, rather than probabilistic models (where the value of the parameters is known).

In such a case, we may use the following test statistics:

$$\lambda = \frac{\sup_{\theta_1 \in \Theta_1} f_1(\theta_1, X)}{\sup_{\theta_0 \in \Theta_0} f_0(\theta_0, X)} = \frac{f_1(\hat{\theta}_1, X)}{f_0(\hat{\theta}_0, X)},$$

where $\hat{\theta}_0$ and $\hat{\theta}_1$ are ML estimators of parameter θ for the null and alternative hypotheses, respectively, or

$$\tilde{\lambda} = \frac{\sup_{\theta \in \Theta} f(\theta, X)}{\sup_{\theta_0 \in \Theta_0} f_0(\theta_0, X)} = \frac{f(\hat{\theta}, X)}{f_0(\hat{\theta}_0, X)},$$

where $\hat{\theta}_0$ is a ML estimator of parameter θ for the null hypothesis, while $\hat{\theta}$ is the ML estimator for parameter θ in a model without restrictions (i.e., for $\Theta_0 \cup \Theta_1$).

In both cases, the specification of the LR test will be of the form $\lambda > c$ or $\tilde{\lambda} > c$, where c is a constant corresponding to the adopted significance level. The second version is especially useful if the null hypothesis is simple (for example, we test $\lambda = \frac{1}{2}$ against the alternative $\lambda \neq \frac{1}{2}$ – in such a case we will take the density with parameter $\lambda = \frac{1}{2}$ in the denominator, and the density with the ML estimator of λ in the numerator), or the models are nested (for example, we test the null hypothesis that data come from an exponential distribution, which is a special case of the gamma distribution family, against the alternative that the distribution is gamma, but not exponential).

Please note that in case of composite hypotheses, we can't formulate a theorem which would correspond to the Neyman-Pearson Lemma – i.e., we will not be able to say that the LR test is the most powerful test for a given set of null and alternative hypotheses. This is because in some cases, the UMPT test does not exist (so the LR test will not be UMP because there is no such test)... This might happen, for example, if we wish to test a null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_0 : \theta \neq \theta_0$ if the family of distributions has a monotonic LR property, i.e. if $f_1(x)/f_0(x)$ is an increasing function of a statistic $T(x)$ for any f_0 and f_1 corresponding to parameters $\theta_0 < \theta_1$. In order to have UMPT for $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ we would

need a critical region of the type $T(x) > c$, and to have a UMPT for $H_0 : \theta = \theta_0$ against $H_1 : \theta < \theta_0$ we would need a critical region of the type $T(x) < c$ (please note examples (1)-(2) in the previous subsection!), so it is impossible to find a UMPT for $H_1 : \theta \neq \theta_0$.