**476 Statistics, Spring 2022.**
**Lecturer: Sergey Nadtochiy.**

**Lecture 8. Hypothesis testing: power of tests, likelihood ratio tests. (Sections 10.10–10.11)**

# 1 Power of tests

Consider a test for $\theta$ with some $H_0$, $H_a$, $U$, and $RR$. Recall that we want

$$\mathbb{P}^{\theta}(U \in RR)$$

to be small for $\theta \in H_0$, and to be large for $\theta \in H_a$.

**Def 1.**

$$\textit{Power}(\theta) := \mathbb{P}^{\theta}(U \in RR)$$

For $H_0 = \{\theta = \theta_0\}$, $H_0 = \{\theta = \theta_a\}$, we have

$$\alpha = \text{Power}(\theta_0), \quad \beta = 1 - \text{Power}(\theta_a).$$

**Typically, we fix $\alpha$ and search for the most powerful $\alpha$-test: i.e., a test whose power function at $\theta_0$ is $\leq \alpha$ and whose power at all $\theta \in H_a$ is "the largest possible" among all $\alpha$-tests.** However, the notion of "the largest possible" may not be well defined unless the alternative hyp. is simple, or if we get lucky.

The following theorem shows how to find the most powerful test for simple $H_0$, $H_a$.

**Thm 1.** *(Neyman-Pearson lemma) For simple $H_0 = \{\theta = \theta_0\}$ and $H_a = \{\theta = \theta_a\}$, the most powerful $\alpha$-test is given by*

$$U = \frac{L(Y_1, \ldots, Y_n; \theta_0)}{L(Y_1, \ldots, Y_n; \theta_a)}, \quad RR = [0, r(\alpha)),$$

*where $L$ is the likelihood function and $r(\alpha)$ is chosen so that*

$$\mathbb{P}^{\theta_0}(U < r(\alpha)) = \alpha. \tag{1}$$

Note that, to design the most powerful test via the above theorem, we need to (i) know ho to compute $L$, and (ii) know the distribution of $U$ under $H_0$.

**Ex 1.** *Consider a sample $Y$ of size $1$ from a distribution with the pdf*

$$f(y; \theta) = \theta y^{\theta - 1}, \quad y \in (0, 1),$$

*and zero for $y \notin (0, 1)$. Consider $H_0 = \{\theta = 2\}$, $H_a = \{\theta = 1\}$.*

**Q 1.** *Find the most powerful test of level $0.05$.*

*We follow the method described in Neyman-Pearson lemma. Notice that*

$$L(y;\theta) = f(y;\theta),$$

$$U = \frac{f(Y;2)}{f(Y;1)} = 2Y, \quad RR = [0,r),$$

$$0.05 = \mathbb{P}^0(Y \leq r/2) = 2\int_0^{r/2} y dy = r^2/4, \quad r = \sqrt{0.2} \approx 0.45,$$

$$RR \approx [0, 0.45).$$

**Q 2.** *Power*$(1) = ?$

$$\textit{Power}(1) = \mathbb{P}^a(U \in RR) = \mathbb{P}^a(Y < 0.45/2) = \int_0^{0.45/2} 1\, dy = 0.225.$$

*Still, the power is not very high (because the sample is small).*

Recall that the notion of a most powerful test is defined for both simple and composite (non-simple) alternatives. However, when the alternative is composite, we emphasize it by saying that the test is **uniformly most powerful**.

There is no general analogue of Neyman-Pearson lemma for constructing the most powerful test, in case $H_a$ is composite. On the other hand, occasionally, we may get lucky and the most powerful test for a carefully chosen simple alternative turns out to be the most powerful test for a given composite alternative.

**Ex 2.** *Consider a sample $Y_1, \dots, Y_n$ of i.i.d. r.v.'s with the distribution $N(\mu, \sigma^2)$, with known $\sigma$ and unknown $\mu$. Consider $H_0 = \{\mu = \mu_0\}$, $H_a = \{\mu > \mu_0\}$.*

**Q 3.** *Find a uniformly most powerful test of level $\alpha$.*

*Let us, for now, pretend that $H_a = \{\mu = \mu_a\}$, for some $\mu_a > \mu_0$, and let us follow Neyman-Pearson:*

$$L(Y_1, \dots, Y_n; \mu) = (2\pi)^{-n/2}\sigma^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2\right),$$

$$U = \frac{(2\pi)^{-n/2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i-\mu_0)^2\right)}{(2\pi)^{-n/2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i-\mu_a)^2\right)} = \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^n \left((Y_i-\mu_a)^2 - (Y_i-\mu_0)^2\right)\right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\mu_a-\mu_0)(2Y_i - \mu_0 - \mu_a)\right) = \exp\left(-\frac{\mu_a-\mu_0}{\sigma^2}\sum_{i=1}^n Y_i + n\frac{(\mu_a-\mu_0)(\mu_0+\mu_a)}{2\sigma^2}\right).$$

*Notice that $U$ is a decreasing function of $\bar{Y}$. Therefore, for any $r$ there exists $r'$ such that: $U \leq r$ if and only if $\bar{Y} \geq r'$. Thus, the test $(U, [0,r])$ is equivalent to $(\bar{Y}, [r', \infty))$. In particular, this means that $(\bar{Y}, [r', \infty))$, with $r'$ such that $\mathbb{P}^0(\bar{Y} \geq r') = \alpha$, is the most powerful $\alpha$-test for $H_0 = \{\mu = \mu_0\}$, $H_a = \{\mu = \mu_a\}$. To find $r'$, we notice that*

$$\alpha = \mathbb{P}^0(\bar{Y} \geq r') = \mathbb{P}^0(\sqrt{n}(\bar{Y} - \mu_0)/\sigma \geq (r' - \mu_0)\sqrt{n}/\sigma)$$

*and recall that $\sqrt{n}(\bar{Y} - \mu_0)/\sigma \sim N(0,1)$ under $H_0$. Thus,*

$$(r' - \mu_0)\sqrt{n}/\sigma = z_\alpha, \quad r' = \mu_0 + z_\alpha \sigma/\sqrt{n}.$$

*We conclude that the most powerful $\alpha$-test for $H_0 = \{\mu = \mu_0\}$, $H_a = \{\mu = \mu_a\}$ is given by the test statistic $\bar{Y}$ and the rejection region*

$$[\mu_0 + z_\alpha \sigma/\sqrt{n}, \infty).$$

*It only remains to notice that, since neither the test statistic nor the rejection region depend on $\mu_a$, the above test is uniformly most powerful for $H_0 = \{\mu = \mu_0\}$, $H_a = \{\mu > \mu_0\}$ at level $\alpha$.*

**Rem 1.** *This would not work for $H_a = \{\mu \neq \mu_0\}$ – there is no uniformly most powerful test for such alternative hypothesis.*

**Rem 2.** *The test in the above example is also uniformly most powerful for $H_0 = \{\mu \leq \mu_0\}$, $H_a = \{\mu > \mu_0\}$, in the sense that it is uniformly most powerful among all tests satisfying*

$$\mathbb{P}^\mu(U \in RR) \leq \alpha, \quad \mu \in H_0.$$

*Indeed, one can check that, with $U = \bar{Y}$, $RR = [\mu_0 + z_\alpha \sigma/\sqrt{n}, \infty)$, we have*

$$\mathbb{P}^\mu(U \in RR) \leq \mathbb{P}^{\mu_0}(U \in RR) = \alpha, \quad \mu \leq \mu_0.$$

*And the fact that the power of this test at any $\mu > \mu_0$ is the largest possible (among all $\alpha$-tests) is shown in the above example.*

# 2 General likelihood ratio tests

**Q 4.** *What is a general method of constructing "good" tests for composite hypotheses $H_0$ and $H_a$? (Recall that $H_0$ and $H_a$ are viewed as subsets of possible values of the unknown parameter $\theta$.)*

Neyman-Pearson lemma tells us that the likelihood ratio test is the best (most powerful) for simple $H_0$ and $H_a$ (i.e. when each of them consists of a single value of $\theta$). Then, heuristically speaking, its modification below is expected to work well for composite hyp.

**Def 2.** *A likelihood ratio test is defined by the test statistic*

$$U = \frac{\max_{\theta \in H_0} L(Y_1, \ldots, Y_n; \theta)}{\max_{\theta \in H_0 \cup H_a} L(Y_1, \ldots, Y_n; \theta)},$$

*and the rejection region*

$$RR = [0, r),$$

*with a constant $r \geq 0$ that is chosen so that the test has a given significance level. (Recall that $L$ is the likelihood function, which is either a product of pdf's or a product of probability functions.)*

The intuition behind the definition is clear. We reject the null hypothesis $H_0$ if the maximum likelihood of the observed sample values under the null hypothesis is significantly smaller than the maximum likelihood under both hypotheses. In other words, we reject $H_0$ if assuming $H_0$ (as opposed to being indifferent between $H_0$ and $H_a$) would reduce the likelihood of the observed data significantly.

Note that, if $H_0$ is composite (i.e. contains more than one element), to ensure that the test is of (significance) level $\alpha$, we need

$$\mathbb{P}^\theta(U < r) \leq \alpha, \quad \forall\, \theta \in H_0.$$

**Ex 3.** *Consider a normal i.i.d. sample $Y_1, \dots, Y_n$ with unknown $\mu$ and $\sigma^2$, and $H_0 = \{\mu = \mu_0\}$, $H_a = \{\mu > \mu_0\}$.*

**Q 5.** *What is the $\alpha$-level likelihood ratio test in this case?*

*Note that we have to include $\sigma$ in the set of unknown parameters even though the hypotheses are not concerned with it. Then, we follow the above definition:*

$$\max_{\mu=\mu_0,\, \sigma^2>0} L(Y_1, \dots, Y_n; \mu, \sigma^2) = \max_{\sigma^2>0} (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \mu_0)^2\right)$$

*The above function converges to zero for $\sigma \to 0, \infty$. Hence, the maximum is attained inside $(0, \infty)$ and must be a zero of the derivative. Taking logarithm of the above, differentiating it w.r.t. $\sigma$, we find*

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \mu_0)^2$$

*(where we recognize the MLE for $\sigma^2$), and conclude*

$$\max_{\mu=\mu_0,\, \sigma^2>0} L(Y_1, \dots, Y_n; \mu, \sigma^2) = (\frac{1}{\sqrt{2\pi}})^n \left(\frac{1}{n}\sum_{i=1}^n (Y_i - \mu_0)^2\right)^{-n/2} e^{-n/2}.$$

*Next, we need to compute*

$$\max_{\mu\geq\mu_0,\, \sigma^2>0} L(Y_1, \dots, Y_n; \mu, \sigma^2) = \max_{\mu\geq\mu_0,\, \sigma^2>0} (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2\right)$$

*First, we maximize over $\mu$ for a fixed $\sigma$. This is easy because after taking log, we obtain a quadratic function (up to additive constant):*

$$\frac{1}{2\sigma^2}(-\sum_{i=1}^n Y_i^2 + 2\mu\sum_{i=1}^n Y_i - n\mu^2)$$

*This function attains its maximum on $[\mu_0, \infty)$ either at the global maximum $\mu = \bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ or at $\mu = \mu_0$ if the latter is above the former. Thus, the maximum is attained at*

$$\hat{\mu} = \max(\bar{Y}, \mu_0).$$

*Next, using the above $\hat{\mu}$ in place of $\mu$, we need to maximize over $\sigma$:*

$$\max_{\sigma^2>0} (\frac{1}{\sqrt{2\pi}\sigma})^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \hat{\mu})^2\right),$$

*which is the same as the problem we already solved (just replace $\mu_0$ by $\hat{\mu}$). Thus, we obtain:*

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\mu})^2,$$

*which yields*

$$\max_{\mu\geq\mu_0,\, \sigma^2>0} L(Y_1, \dots, Y_n; \mu, \sigma^2) = (\frac{1}{\sqrt{2\pi}})^n (\frac{1}{n}\sum_{i=1}^n (Y_i - \max(\bar{Y}, \mu_0))^2)^{-n/2} e^{-n/2}$$

*Thus,*

$$U = \frac{\max_{\mu=\mu_0,\,\sigma^2>0} L(Y_1,\dots,Y_n;\mu,\sigma^2)}{\max_{\mu\geq\mu_0,\,\sigma^2>0} L(Y_1,\dots,Y_n;\mu,\sigma^2)} = \begin{cases} 1, & \text{if } \bar{Y}\leq\mu_0, \\ \left(\frac{\sum_{i=1}^n (Y_i-\bar{Y})^2}{\sum_{i=1}^n (Y_i-\mu_0)^2}\right)^{n/2}, & \text{else.} \end{cases}$$

*Notice that $U \leq 1$, hence it only makes sense to consider $r < 1$. For such $r$, the event*

$$U < r$$

*is equivalent to the combination of $\bar{Y} > \mu_0$ and*

$$\frac{\sum_{i=1}^n (Y_i-\bar{Y})^2}{\sum_{i=1}^n (Y_i-\bar{Y})^2 + 2\sum_{i=1}^n (Y_i-\bar{Y})(\bar{Y}-\mu_0)(=0) + n(\bar{Y}-\mu_0)^2} < r^{2/n}.$$

*The latter is equivalent to*

$$\frac{1}{1+\frac{n(\bar{Y}-\mu_0)^2}{\sum_{i=1}^n (Y_i-\bar{Y})^2}} < r^{2/n}.$$

*Thus, we conclude that $U < r$ is equivalent to $\bar{Y} > \mu_0$ and*

$$V := \frac{\sqrt{n}(\bar{Y}-\mu_0)}{S} > \sqrt{(n-1)(r^{-2/n}-1)} =: r_1,$$

*where we recall the formula for sample standard deviation:*

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (Y_i-\bar{Y})^2}$$

*The above test is natural: we reject the null hypothesis in favor of the alternative (which implies higher mean) if the t-statistic (normalized sample mean) is sufficiently large.*

*Finally, we need to choose $r_1$ so that the probability of rejecting the null, if $\mu=\mu_0$, is at most $\alpha$:*

$$\alpha = \mathbb{P}^0(V > r_1) = \mathbb{P}^0\left(\frac{\sqrt{n}(\bar{Y}-\mu_0)}{S} > r_1\right).$$

*Recall that $\frac{\sqrt{n}(\bar{Y}-\mu_0)}{S}$ has a student distr. with parameter $n-1$, under $H_0$:*

$$\frac{\sqrt{n}(\bar{Y}-\mu_0)}{S} \sim T(n-1).$$

*Then, we conclude that $r_1$ must be the quantile of $T(n-1)$ at the level $1-\alpha$: i.e. in our notation*

$$r_1 = t_\alpha$$

**Exercise 1.** *Show that the likelihood ratio test of level $\alpha$ for $H_0 = \{\mu=\mu_0\}$ vs. $H_a = \{\mu\neq\mu_0\}$ is given by the test statistic*

$$\frac{\sqrt{n}(\bar{Y}-\mu_0)}{S}$$

*and the rejection region is*

$$RR = (-\infty, -t_{\alpha/2}] \cup [t_{\alpha/2}, \infty).$$

The exact distribution of the likelihood ratio test statistic $U$ is often unknown, once we relax the normality assumption (or if the form of sample distribution or hypotheses is non-standard). Nevertheless, we can obtain the asymptotic distribution of $U$.

First, we need an auxiliary definition from linear algebra.

**Def 3.** *The dimension of a set $A \subset \mathbb{R}^n$, denoted $dim(A)$, is the smallest dimension among all linear (affine) subspaces containing this set.*

Examples: a point has dimension zero, an interval on a straight line has dimension one, a circle has dimension two, etc.

**Thm 2.** *(Wilk) Let $r_0 = dim(H_0)$ and $r = dim(H_0 \cup H_a)$, and denote by $U$ the test statistic of the associated likelihood ratio test. Then, under certain technical assumptions (in this course, you can assume that these conditions are satisfied whenever you are asked to use this result),*

$$-2 \log U \to \chi^2(r - r_0),$$

*in distribution, as the sample size goes to infinity.*

The above result allows us to construct likelihood ration tests of asymptotic level $\alpha$.

**Exercise 2.** *Consider two independent samples $\{Y_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$, from Poisson distributions with means $\theta_1$ and $\theta_2$, respectively. Show that the likelihood ratio test for $H_0 = \{\theta_1 = \theta_2\}$ vs. $H_a = \{\theta_1 \neq \theta_2\}$, with asymptotic level $\alpha$, is given by the test statistic*

$$U = \frac{((\bar{Y} + \bar{Z})/2)^{n(\bar{Y}+\bar{Z})}}{\bar{Y}^{n\bar{Y}}\bar{Z}^{n\bar{Z}}}$$

*and by the rejection region*

$$RR = [0, e^{-\chi^2_\alpha/2}].$$