

Lecture 6: Composite Hypothesis Testing and Properties of a Random Sample

28 Jan 2016

In the previous lecture, the Bayesian formulation for a composite hypothesis testing problem has been shown. In this lecture, some examples of composite hypothesis testing problems are considered. Further, the notion of a random sample has been defined and some of its properties have been listed and proved.

1 Examples of a composite hypothesis testing problem

Let us consider the following example of a composite hypothesis testing problem.

Example 1.1. Testing on the radius of a point on the plane:

Suppose that $\Gamma = \mathbb{R}^2$, i.e. $Y \in \Gamma$, $Y = (Y_1, Y_2)$.

$$\begin{aligned} Y &= X + n, \\ \text{where, } n &\sim \mathcal{N}(0, \sigma^2 I_2), \\ X &\in \{0, (A \cos \psi, A \sin \psi)\}. \end{aligned}$$

$A > 0$, a constant, $\psi \sim \text{unif}([0, 2\pi))$ and n_1 and n_2 are independent of one another and of ψ . The null hypothesis and the alternate hypothesis are given by,

$$\begin{aligned} \mathcal{H}_0 &: Y \sim \mathcal{N}(0, \sigma^2 I_2). \\ \mathcal{H}_1 &: Y \sim \mathcal{N}((A \cos \psi, A \sin \psi), \sigma^2 I_2). \end{aligned}$$

The observation is a noisy measurement of the coordinates of a point in the plane that is either at the origin or is uniform-randomly distributed on a circle of radius A . So, we have partitions of parametric family of different distributions. For each ψ , we get different distributions.

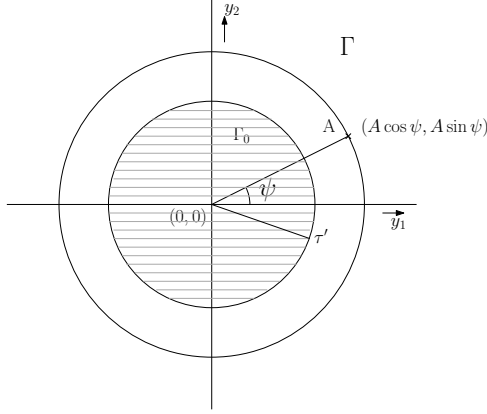


Figure 1: Decision regions for example 1.1($\Gamma_1 = \Gamma_0^c$).

Let the parameter be $\Theta = (\Theta_1, \Theta_2)$, with $\Theta_1 \in \{0, A\}$ and $\Theta_2 \in [0, 2\pi)$. Thus the parameter set is $\Lambda = \{0, A\} \times [0, 2\pi)$, with $\Lambda_0 = \{\theta \in \Lambda : \theta_1 = 0\}$ and $\Lambda_1 = \{\theta \in \Lambda : \theta_1 = A\}$. Let ψ be a random variable that takes the value θ_2 . The conditional distribution of Y given $\Theta = \theta$ is given by,

$$Y | (\Theta = \theta) \sim \mathcal{N}((\theta_1 \cos \theta_2, \theta_1 \sin \theta_2), \sigma^2 I_2). \quad (1)$$

$$P_\theta(y) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{q(u, \theta)}{2\sigma^2}\right] du, \quad y \in \mathbb{R}^2 \quad (2)$$

where,

$$q(y, \theta) \triangleq (y_1 - \theta_1 \cos \theta_2)^2 + (y_2 - \theta_1 \sin \theta_2)^2. \quad (3)$$

The likelihood ratio calculation is as follows,

$$p(y | \Theta \in \Lambda_0) = p_\theta(y) |_{\theta_1=0}, \quad (4)$$

$$= \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(y_1^2 + y_2^2)}{2\sigma^2}\right], \quad (5)$$

and,

$$p(y | \Theta \in \Lambda_1) = \frac{1}{2\pi} \int_0^{2\pi} p_\theta(y) |_{\theta_1=A} d\theta_2, \quad (6)$$

$$= \frac{1}{4\pi^2\sigma^2} \int_0^{2\pi} \exp\left[-\frac{q(y, \theta) |_{\theta_1=A}}{2\sigma^2}\right] d\theta_2. \quad (7)$$

The likelihood ratio is thus given by,

$$L(y) = \frac{dP_1(y | \theta \in \Lambda_1)}{dP_0(y | \theta \in \Lambda_0)} \quad (8)$$

Assume $c(i, \theta)$ is constant over $\theta \in \Lambda_1$, $\theta \in \Lambda_0$.

$$\frac{d}{dy} P_{\Theta}(y|\Theta \in \Lambda_1) = \mathbb{E}[P_{\Theta}(y|\Theta \in \Lambda_1), \Theta_2]. \quad (9)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{-q(y, \theta)}{2\sigma^2}\right] d\theta_2 \quad \forall \theta \in \Lambda. \quad (10)$$

Rewriting the likelihood ratio $L(y)$ as,

$$L(y) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\exp\left[\frac{-1}{2\sigma^2} \left(y_1^2 + A^2 \cos \theta_2 - 2y_1 A \cos \theta_2\right) - \frac{1}{2\sigma^2} \left(y_2^2 + A^2 \sin \theta_2 - 2y_2 A \sin \theta_2\right)\right]}{\exp\left(-\frac{(y_1^2 + y_2^2)}{2\sigma^2}\right)} d\theta_2.$$

$$\begin{aligned} L(y) &= \frac{1}{2\pi} \int_0^{2\pi} \exp\left[-\frac{A^2}{2\sigma^2} + \frac{A}{\sigma^2} (y_1 \cos \theta_2 + y_2 \sin \theta_2)\right] d\theta_2, \\ &= \frac{\exp\left(-\frac{A^2}{2\sigma^2}\right)}{2\pi} \int_0^{2\pi} \exp\left[\frac{A}{\sigma^2} (y_1 \cos \theta_2 + y_2 \sin \theta_2)\right] d\theta_2. \end{aligned} \quad (11)$$

Using the following transformation of variables to make simplifications in the integral,

$$r^2 = y_1^2 + y_2^2, \text{ and } \tan \phi = \frac{y_2}{y_1}, \quad (12)$$

$$\begin{aligned} L(r) &= \frac{\exp\left(-\frac{A^2}{2\sigma^2}\right)}{2\pi} \int_0^{2\pi} \exp\left(\frac{A}{\sigma^2} r \cos(\theta_2 - \phi)\right) d\theta_2, \\ &= \exp\left(-\frac{A^2}{2\sigma^2}\right) I_0\left(\frac{Ar}{\sigma^2}\right). \end{aligned} \quad (13)$$

where $I_0(\cdot)$ is the "zeroth order Bessel function of the first kind".

Note 1. Notice that $L(y)$ is independent of ϕ .

Note 2. I_0 is monotone increasing in its argument.

Let us denote the threshold by τ . Since I_0 is a monotonic function, it can be inverted to find r .

$$\begin{aligned} L(r) &= \tau. \\ r &= L^{-1}(\tau) = \tau'. \\ \tau' &= \frac{\sigma^2}{A} I_0^{-1}\left(\exp\left(\frac{A^2}{2\sigma^2}\right)\right). \end{aligned} \quad (14)$$

The decision rule is,

$$\tilde{\delta}(y) = \begin{cases} 1, & \text{if } r > \tau'. \\ \gamma, & \text{if } r = \tau'. \\ 0, & \text{if } r < \tau'. \end{cases} \quad (15)$$

We have completely characterized decision rule for Bayesian framework. But we know that similar decision rules will hold for minimax. For composite hypothesis testing problems in which the prior distributions of the parameters are unknown, the definition of optimality requires a generalization of the Neyman-Pearson criterion of the simple hypothesis testing problem.

Recall, for a randomized decision rule $\tilde{\delta}$ we want to maximize probability of detection such that probability of false alarm is bounded by "significance level" α .

$$P_F(\tilde{\delta}; \theta) = \mathbb{E}_\theta(\tilde{\delta}(Y)), \quad \theta \in \Lambda_0. \quad (16)$$

$$P_D(\tilde{\delta}; \theta) = \mathbb{E}_\theta(\tilde{\delta}(Y)), \quad \theta \in \Lambda_1. \quad (17)$$

This is well defined if there is a unique $\theta \in \Lambda_0$ and a unique $\theta \in \Lambda_1$. But the problem is that there are many θ 's in Λ_0 and Λ_1 .

Definition 1.2. A Uniformly Most Powerful (UMP) test is a decision rule that maximizes the probability of detection $\forall \theta \in \Lambda_1$, such that the probability of false alarm $\forall \theta \in \Lambda_0$ is below the significance level α .

$$\begin{aligned} & \max_{\tilde{\delta}} P_D(\tilde{\delta}; \theta), \quad \forall \theta \in \Lambda_1. \\ & \text{s.t. } P_F(\tilde{\delta}; \theta) \leq \alpha, \quad \forall \theta \in \Lambda_0. \end{aligned}$$

Although the UMP tests are desirable, they exist only in certain cases. To demonstrate this, let us consider the following example.

Example 1.3. Consider a case where the null hypothesis is simple, i.e., $\Lambda_0 = \{\theta_0\}$, and the alternate hypothesis Λ_1 is composite. For each $\theta \in \Lambda_1$, we can find the rejection region.

$$\Gamma_\theta = \{y \in \Gamma : L(y) > \tau\}. \quad (18)$$

$$\text{where } L(y) = \frac{dP_{\theta_1}(y)}{dP_{\theta_0}}, \quad (19)$$

is the likelihood ratio and τ is the threshold chosen (with possibly randomization) to give significance level α . By the N-P lemma, we know that this test is unique, i.e., for $\theta_1, \theta_2 \in \Lambda_1$, such that $\theta_1 \neq \theta_2$, the test with critical region Γ_{θ_1} will have

smaller power in testing $\mathcal{H}_0 : Y \sim P_{\theta_0}$ versus $\mathcal{H}_1 : Y \sim P_{\theta_2}$ and vice versa unless the two critical regions are identical.

$$P_{\theta_2}(\Gamma_{\theta_1}) \leq P_{\theta_2}(\Gamma_{\theta_2}).$$

with equality iff $\Gamma_{\theta_1} = \Gamma_{\theta_2}$.

Corollary 1.4. *A UMP test exists for simple \mathcal{H}_0 versus composite \mathcal{H}_1 iff Γ_θ is identical $\forall \theta \in \Gamma_1$.*

Example 1.5. UMP testing of location:

Consider the parametric family of distributions $\{P_\theta; \theta \in \Lambda\}$, where,

$$\{P_\theta = \mathcal{N}(\theta, \sigma^2), \theta \in \Lambda\}, \Lambda \subseteq \mathbb{R}, \quad (20)$$

and, suppose that we have the hypotheses,

$$\mathcal{H}_0 : \theta = \mu_0, \quad (21)$$

$$\mathcal{H}_1 : \theta > \mu_0. \quad (22)$$

That is, $\Lambda_0 = \{\mu_0\}$ and $\Lambda_1 = (\mu_0, \infty)$. The most powerful α -level test \mathcal{H}_0 versus $\mathcal{H}_1 : Y \sim \mathcal{N}(\theta, \sigma^2)$ has critical region:

$$\Gamma_\theta = \{y \in \Gamma : y > \sigma\Phi^{-1}(1 - \alpha) + \mu_0\}, \forall \theta \in \Lambda_1.$$

The UMP test, denoted by $\tilde{\delta}$ is:

$$\tilde{\delta} = \mathbb{1}_{\{y \in \Gamma_\theta\}}.$$

Remark 1. Randomization is not required since the continuous random variables are being considered.

$$P_D(\tilde{\delta}; \theta) = 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{(\theta - \mu_0)}{\sigma}\right).$$

Note 3. Γ_θ does not depend on θ .

Note 4. In this case we found a UMP; but this is rather rare.

Example 1.6. Testing of location with different hypotheses pair:

The setup is as in the previous example, except a different \mathcal{H}_1 .

$$\mathcal{H}_0 : \theta = \mu_0 \text{ (as before),}$$

$$\mathcal{H}_1 : \theta \neq \mu_0.$$

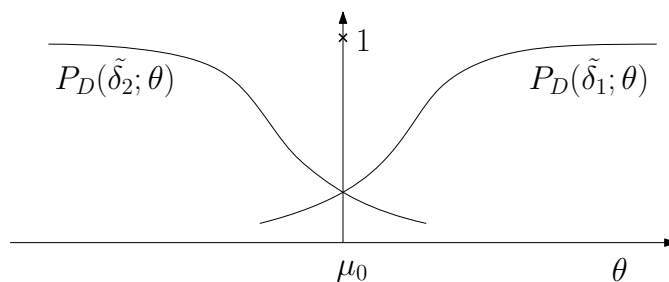


Figure 2: Power curves for test of $\theta = \mu_0$ versus $\theta > \mu_0$ and $\theta = \mu_0$ versus $\theta < \mu_0$, for location testing with Gaussian error.

That is, $\Lambda_0 = \{\mu_0\}$ and $\Lambda_1 = (-\infty, \mu_0) \cup (\mu_0, \infty)$. For $\theta > \mu_0$, the most powerful test is

$$\tilde{\delta}_1(y) = \mathbb{1}_{\{y \in \Gamma_\theta\}}, \quad (23)$$

and the critical region is

$$\Gamma_\theta^1 = \{y \in \Gamma : y > \sigma\Phi^{-1}(1 - \alpha) + \mu_0\}, \quad (24)$$

as in the previous example.

Now, we consider the case when $\theta < \mu_0$. The critical region of the most powerful α -level test becomes:

$$\Gamma_\theta^2 = \{y \in \Gamma : y < \sigma\Phi^{-1}(\alpha) + \mu_0\}. \quad (25)$$

The most powerful test being,

$$\tilde{\delta}_2(y) = \mathbb{1}_{\{y \in \Gamma_\theta^2\}}, \quad (26)$$

resulting in the power of test,

$$P_D(\tilde{\delta}_2; \theta) = \Phi\left(\Phi^{-1}(\alpha) - \frac{(\theta - \mu_0)}{\sigma}\right). \quad (27)$$

Remark 2. Although the critical region is independent of θ , it is different from the critical region for the case of $\theta > \mu_0$. Thus, no UMP exists. So we come up with some other criteria.

Remark 3. Although the UMP is optimal in some sense, it is rare.

Remark 4. The UMP is thus too strong a criterion for many situations.

Other criteria

As observed earlier, the UMP criterion is very strong to be satisfied. Hence by applying some reasonable constraints, we consider a few other criteria.

1. Unbiasedness

Definition 1.7. A test is said to satisfy the condition of unbiasedness if,

$$\begin{aligned} P_D(\tilde{\delta}; \theta) &\geq \alpha, \quad \forall \theta \in \Lambda_1, \\ \text{in addition to} \\ P_F(\tilde{\delta}; \theta) &\leq \alpha, \quad \forall \theta \in \Lambda_0. \end{aligned}$$

This requirement would eliminate both $\tilde{\delta}_1$ and $\tilde{\delta}_2$ in the previous example.

Example 1.8. Consider an example where the parameter set is given by $\Lambda = [\theta_0, \infty)$. Let $\Lambda_0 = \{\theta_0\}$, $\Lambda_1 = (\theta_0, \infty)$. The null hypothesis and the alternate hypothesis are given by:

$$\mathcal{H}_0 : \theta = \theta_0, \tag{28}$$

$$\mathcal{H}_1 : \theta > \theta_0. \tag{29}$$

Such a case arises in signal detection where the amplitude of the signal is represented by θ .

The only case where confusion arises is when the received amplitude is close to θ_0 . Consider a decision rule $\tilde{\delta}$. Under certain regularity conditions, by the application of Taylor's series expansion about θ_0 , we obtain,

$$\begin{aligned} P_D(\tilde{\delta}; \theta) &= P_D(\tilde{\delta}; \theta_0) + (\theta - \theta_0)P'_D(\tilde{\delta}; \theta_0) + O((\theta - \theta_0)^2). \\ &= P_F(\tilde{\delta}) + (\theta - \theta_0)P'_D(\tilde{\delta}; \theta_0) + O((\theta - \theta_0)^2). \\ &\cong \alpha + (\theta - \theta_0)P'_D(\tilde{\delta}; \theta_0). \end{aligned} \tag{30}$$

2. α -level "locally most powerful test" (LMP)

A test that maximizes $P'_D(\tilde{\delta}; \theta_0)$ subject to false alarm constraint $P_F(\tilde{\delta}) \leq \alpha$ is called an α -level locally most powerful test (LMP).

$$\begin{aligned} &\max_{\tilde{\delta}} P'_D(\tilde{\delta}; \theta_0), \\ &\text{subject to } P_F(\tilde{\delta}) \leq \alpha. \end{aligned} \tag{31}$$

Consider,

$$\begin{aligned}
P_D(\tilde{\delta}; \theta) &= \mathbb{E}_\theta\{\tilde{\delta}(Y)\}, \\
&= \int_{y \in \Gamma} \tilde{\delta}(y) dP_\theta(y). \\
P'_D(\tilde{\delta}; \theta_0) &= \int_{y \in \Gamma} \tilde{\delta}(y) d \left[\frac{\partial}{\partial \theta} P_\theta(y) \right]_{\theta=\theta_0}, \text{ and} \\
\tilde{\delta}_{LO}(y) &= \begin{cases} 1, & \text{if } \left[\frac{\partial}{\partial \theta} L(y) \right]_{\theta=\theta_0} > \eta. \\ \gamma, & \text{if } \left[\frac{\partial}{\partial \theta} L(y) \right]_{\theta=\theta_0} = \eta. \\ 0, & \text{if } \left[\frac{\partial}{\partial \theta} L(y) \right]_{\theta=\theta_0} < \eta. \end{cases} \quad (32)
\end{aligned}$$

We are mostly interested in the case when $\Lambda_0 = \{\theta_0\}$ and Λ_1 is composite, where η and γ are chosen such that,

$$P_F(\tilde{\delta}_{LO}) = \alpha.$$

3. Maximum Likelihood Test

A test that is applicable for composite hypothesis testing problems is based on comparing the likelihood ratio $L(y)$ to a threshold τ . The likelihood ratio is given by,

$$L(y) = \frac{\max_{\theta \in \Lambda_1} dP_\theta(y)}{\max_{\theta \in \Lambda_0} dP_\theta(y)}. \quad (33)$$

So, we have three ways to deal with the composite hypothesis test.

2 Properties of a random sample

In this section, we define a random sample and its statistics. Also, some properties of the random sample are listed and proved.

Definition 2.1. The collection of random variables X_1, X_2, \dots, X_n is called a "random sample of size n from the population $p(X)$ ", if they are independent and identically distributed with common distribution p . The joint pdf of X_1, X_2, \dots, X_n is given by

$$p(x_1, x_2, \dots, x_n) = p(x_1)p(x_2)\dots p(x_n) = \prod_{i=1}^n p_X(x_i). \quad (34)$$

Definition 2.2. Let X_1, X_2, \dots, X_n be a random sample of size n from the population $p(X)$ and let T be a real/vector valued function whose domain includes the sample space of X_1, X_2, \dots, X_n . Then the random variable/vector $Y = T(X_1, X_2, \dots, X_n)$ is called a "statistic". The probability distribution of the statistic Y is called the "sampling distribution" of Y .

Example 2.3. $T(X_1, X_2, \dots, X_n) = \min_{i \in [n]} X_i$.

Example 2.4. $T(X_1, X_2, \dots, X_n) = \max_{i \in [n]} X_i$.

Definition 2.5. The "sample mean" is the statistic defined by taking the arithmetic average of values in a random sample, denoted by \bar{X} .

$$\bar{X} = \frac{1}{n} \sum_{i \in [n]} X_i. \quad (35)$$

Definition 2.6. The "sample variance" is the statistic defined by,

$$s^2 = \frac{1}{n-1} \sum_{i \in [n]} (X_i - \bar{X})^2. \quad (36)$$

Definition 2.7. The "sample standard deviation" is the statistic defined by $\sigma = \sqrt{s^2}$.

Theorem 2.8. Let $x_1, x_2, \dots, x_n \in \mathbb{R}$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ then,

$$a) \quad \min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{and} \quad (37)$$

$$b) \quad (n-1)s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2. \quad (38)$$

Proof. a) The proof for the first statement of the above theorem is as follows:

$$\begin{aligned} \sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - a)^2. \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (\bar{x} - a)^2. \\ &\geq \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned} \quad (39)$$

The above inequality holds with equality when $a = \bar{x}$. Therefore,

$$\min_a \sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2. \quad (40)$$

b) Substituting $a = 0$ in the statement in eqn. (39),

$$\begin{aligned} \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n \bar{x}^2. \\ \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n x_i^2 - n\bar{x}^2. \end{aligned} \quad (41)$$

□

Lemma 2.9. *Let X_1, X_2, \dots, X_n be a random sample of size n from a population and let $g(x)$ be a function such that $\mathbb{E}[g(x)]$ and $\text{Var}[g(x)]$ exist. Then,*

$$a). \quad \mathbb{E}\left[\sum_{i=1}^n g(X_i)\right] = n\mathbb{E}[g(X_1)], \text{ and} \quad (42)$$

$$b). \quad \text{Var}\left[\sum_{i=1}^n g(X_i)\right] = n\text{Var}[g(X_1)]. \quad (43)$$

Proof. a) \mathbb{E} is a linear operator and X_1, X_2, \dots, X_n are identically distributed. Thus, the above equation can be written as,

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^n g(X_i)\right] &= \mathbb{E}[g(X_1) + g(X_2) + \dots + g(X_n)]. \\ &= \mathbb{E}[g(X_1)] + \mathbb{E}[g(X_2)] + \dots + \mathbb{E}[g(X_n)]. \\ &= n\mathbb{E}[g(X_1)]. \end{aligned} \quad (44)$$

b) By the definition of variance,

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n g(X_i)\right] &= \mathbb{E}\left[\sum_{i=1}^n g(X_i) - \mathbb{E}\left[\sum_{i=1}^n g(X_i)\right]\right]^2. \\ &= \mathbb{E}\left[\sum_{i=1}^n [g(X_i) - \mathbb{E}[g(X_i)]]\right]^2. \end{aligned} \quad (45)$$

Out of the n^2 terms in the expansion, n terms are of the form,

$$\begin{aligned}\mathbb{E}\left[g(X_i) - \mathbb{E}[g(X_i)]\right]^2 &= \text{Var}(g(X_i)), \\ &= \text{Var}(g(X_1)).\end{aligned}\tag{46}$$

The remaining $n(n-1)$ are the covariance terms which evaluate to zero since X_1, X_2, \dots, X_n are independent. Therefore, for $i \neq j$,

$$\begin{aligned}\mathbb{E}[(g(X_i) - \mathbb{E}[g(X_i)])(g(X_j) - \mathbb{E}[g(X_j)])] &= \text{Cov}(g(X_i), g(X_j)), \\ &= 0.\end{aligned}\tag{47}$$

Therefore,

$$\text{Var}\left[\sum_{i=1}^n g(X_i)\right] = n \text{Var}[g(X_i)].\tag{48}$$

□