MTH211A: Theory of Statistics

Module 6: Testing of Hypothesis

1 Some Concepts and Definitions

Hypothesis testing deals with evaluating the feasibility of two competing statements about the underlying population based on a random sample drawn from the same.

Definition 1. *(Hypothesis, Null and Alternative) A statement or conjecture about the population or population parameters is called a hypothesis.*

The two contradictory statements (hypotheses) in a hypothesis testing problem are called the null hypothesis, denoted by H_0, and the alternative hypothesis, denoted by H_1.

Testing of the hypothesis mainly deals with accepting or rejecting the null hypothesis H_0 given the data, given H_1 as the alternative.

Example 1. Suppose $X_1, \cdots, X_n$ is a random sample from $N(\mu, 1)$ distribution, and we are interested in testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$. While testing the hypothesis H_0 we will collect evidence from the data in support of H_0, given H_1 as the alternative. Thus if the sample indicates that the population mean is much larger than zero, we reject H_0, otherwise, we accept H_0. Note that, if the sample indicates that the population mean is much smaller than zero then also we accept H_0 (in view of H_1).

Definition 2. *(One-sided or Two-sided Alternatives) Suppose we are testing $H_0 : \theta = \theta_0$. The possible alternatives can be $H_{1,1} : \theta = \theta_1$ where $\theta_1 > \theta_0$, $H_{1,2} : \theta = \theta_1$ where $\theta_1 < \theta_0$, $H_{1,3} : \theta > \theta_0$, $H_{1,4} : \theta < \theta_0$, $H_{1,5} : \theta \neq \theta_0$, etc. The first four alternatives are one-sided, whereas $H_{1,5}$ is a two-sided alternative.*

Definition 3. *(Simple and Composite Hypotheses) Under a hypothesis H, if the population distribution is completely specified, then H is called a simple hypothesis, otherwise, it is called a composite hypothesis.*

Suppose under H, it is specified that the underlying parameter vector $\theta \in \Theta_0$, then H is simple if Θ_0 is singleton, and it is composite otherwise.

Example 1. (continue) Suppose $X_1, \cdots, X_n$ is a random sample from $N(\mu, 1)$ distribution, and we are interested in testing $H_0 : \mu = 0$ against $H_1 : \mu > 0$. Here H_0 is simple, but H_1 is composite.

Definition 4. *(Hypothesis Test) A hypothesis test is a set of rules that indicates which sample values lead to acceptance of H_0, and which sample values lead to rejection of H_0. A testing procedure partitions the sample space into two regions: one, called acceptance region, leads to acceptance of H_0, and the other, called critical region, leads to rejection of H_0. In other words, if the observed sample falls in the critical region then H_0 is rejected, otherwise, H_0 is accepted.*

Definition 5. *(Critical and Acceptance Regions) Let the support of $\mathbf{x}$ be $S_X \subseteq \mathbb{R}^n$. A subset C of S_X (or, $\mathbb{R}^n$) such that if the data $\mathbf{x} \in C$ then H_0 is rejected, is called the critical region. A subset A of S_X (or, $\mathbb{R}^n$) such that if the data $\mathbf{x} \in A$ then H_0 is accepted, is called the acceptance region. Note that $S_X \subseteq C \cup A$.*

Sometimes it is convenient to define the test in terms of a function from $\phi : S_X \to [0, 1]$, such that

$$\phi(\mathbf{x}) = 1 \quad \text{if} \quad \mathbf{x} \in C, \quad \text{and} \quad \phi(\mathbf{x}) = 0 \quad \text{if} \quad \mathbf{x} \in A.$$

Such a function is called a test function.

Definition 6. *(Test Function) Any function ϕ from $S_X \to [0,1]$ is known a test function.*

Remark 1. *Given a test of hypothesis (i.e., a rule indicating the samples to be rejected or accepted) T_1, one can define a test function, say, $\phi_1(\mathbf{x})$, as $\phi_1(\mathbf{x}) = P\left(\text{Reject } H_0 \mid \mathbf{X} = \mathbf{x}\right)$. Conversely, given a test function $\phi_1(\mathbf{x})$ one can construct a test, say, T_1, where the testing rule states that given the realization $\mathbf{x}$, H_0 is rejected with probability $\phi_1(\mathbf{x})$. Thus, setting a hypothesis test is equivalent to setting a test function.*

Definition 7. *(Type I and Type II Errors) In following a test procedure two types of errors can occur. One may reject the null hypothesis when it is indeed true, or one may accept the null hypothesis when it is indeed false. The first type of error is called Type I error, and the second type of error is called Type II error.*

True State $\to$ *Decision* $\downarrow$	H_0 *is true*	H_0 *is false*
Accept H_0	*Correct decision*	*Type II error*
Reject H_0	*Type I error*	*Correct decision*

Definition 8. *(Probabilities of Type I and Type II Errors, Power) The probability of type I error is $P(\mathbf{X} \in C \mid H_0)$, where C is the critical region. The probability of type II error is $P(\mathbf{X} \in \bar{C} \mid H_1)$. The probability of the complement of type II error is called power. Thus* power *is the probability of rejecting H_0 when it is indeed false.*

Example 1. (continue) In the above example, suppose we construct the following test procedure based on n samples: If the sample mean is greater than or equal to 6 then we reject H_0. Then the probability of type I error is $P\left[\bar{X}_n \geq 6 \mid \bar{X}_n \sim N(0, 1/n)\right]$.

Definition 9. *(Power Function) Suppose we want to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\Theta_0 \cup \Theta_1 = \Theta$, based on a random sample $\mathbf{X}$. Suppose further that a test function ϕ is proposed for testing H_0 against H_1. Then the power function of the test ϕ is*

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

When $\theta \in \Theta_0$, then $\beta_\phi(\theta)$ provides the probability of type I error at θ, and when $\theta \in \Theta_1$, $\beta_\phi(\theta)$ provides the complement of probability of type II error (power) at θ.

Remark 2. *Ideally, one would like to set a test procedure for minimizing the probabilities of both types of errors. However, in general, if one tends to minimize the probability of one error then the probability of the other error increases. Thus, in practice one bounds the maximum probability of type I error to a pre-assigned level α (which is small enough), and then minimizes the probability of type II error. The pre-assigned threshold of maximum probability of type I error α is called the level of significance, or just the level of the test.*

Definition 10. *(Level of Significance) Let ϕ be a test function for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. Then ϕ is called a level-α test, or a test with level of significance α if*

$$E_\theta\left[\phi(\mathbf{X})\right] \leq \alpha, \quad \textit{for all} \quad \theta \in \Theta_0, \quad \textit{or, equivalently,} \quad \sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leq \alpha. \tag{1}$$

Definition 11. *(Size of a Test) Let ϕ be a test function for testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. Then the size of ϕ is $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$.*

Example 2. Suppose $X_1, \ldots, X_{10}$ be a random sample of size 10 from $\texttt{Bernoulli}(p)$, and consider the testing problem $H_0 : p = 0.5$ against $H_1 : p = 0.75$. One would reject H_0 if more number of heads appear. What would be a good test procedure? Consider the following test procedures and the corresponding probabilities of type I errors and powers.

Test	Procedure	P(Type I error)	Power = 1 - P(Type II error)
ϕ_1	Reject if 8 or more heads appear	$\sum_{x=8}^{10} \binom{10}{x}(0.5)^{10}$ ≈ 0.0547	$\sum_{x=8}^{10} \binom{10}{x}(0.75)^{x}(0.25)^{10-x}$ ≈ 0.5256
ϕ_2	Reject if 9 or more heads appear	$\sum_{x=9}^{10} \binom{10}{x}(0.5)^{10}$ ≈ 0.0107	$\sum_{x=9}^{10} \binom{10}{x}(0.75)^{x}(0.25)^{10-x}$ ≈ 0.2440
ϕ_3	Reject if 10 heads heads appear	$\binom{10}{10}(0.5)^{10}$ ≈ 0.001	$\binom{10}{10}(0.75)^{10}$ ≈ 0.0563

Remark 3. *Given the above remark, a usually recommended test procedure is as follows: Suppose we want to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ at level α. Then we will only consider tests satisfying the level-α conditions (1). Then among the tests satisfying (1), we will consider the test having the highest power.*

Definition 12. *(Most Powerful (MP) Test) Suppose we are interested in testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ at level α, and Φ_α is the class of tests satisfying the level-α condition (i.e., for any test $\phi \in \Phi_\alpha$, (1) is satisfied). A test $\phi_0 \in \Phi_\alpha$ is called most powerful test against an alternative $\theta_1 \in \Theta_1$ if*

$$\beta_{\phi_0}(\theta_1) \geq \beta_\phi(\theta_1) \quad \textit{for all } \phi \in \Phi_\alpha.$$

Definition 13. *(Uniformly Most Powerful (UMP) Test) Suppose we are interested in testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ at level α, and Φ_α is the class of tests satisfying the level-α condition. A test $\phi_0 \in \Phi_\alpha$ is called uniformly most powerful test if*

$$\beta_{\phi_0}(\theta_1) \geq \beta_\phi(\theta_1) \quad \textit{for all } \phi \in \Phi_\alpha, \quad \textit{unifromly in } \theta \in \Theta_1 \ (\textit{i.e., for all } \theta \in \Theta_1).$$

Example 2. (continue) Suppose we want to find the MP test for testing H_0 against H_1 in the last example at level 0.05. Of course the test ϕ_1 does not satisfy the level condition. Both ϕ_2 and ϕ_3 satisfy the level condition. However, ϕ_2 has higher power than ϕ_3. So, ϕ_2 should be preferred over ϕ_3. Can we construct a better test than ϕ_2?

Suppose I consider a test as follows:

- If the number of heads is 9 or more, then H_0 is rejected.
- If the number of heads is 8, then select a random number U from Uniform$(0, 1)$. If the realized value of U, say u, satisfies $u < 0.85$, then reject H_0, otherwise accept H_0.
- If the number of heads is 7 or less, then accept H_0.

Does this test satisfy the level condition? What is the power of this test?

We may write the test in terms of a test function as follows:

$$\phi_4(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i = 9, 10, \\ 0.85 & \text{if } \sum_{i=1}^n x_i = 8 \\ 0 & \text{otherwise.} \end{cases}$$

Checking the level condition of ϕ_4: The probability of type I error is

$$P(\text{Reject } H_0 \mid H_0) = P\left(\sum_{i=1}^{10} X_i \in \{9, 10\} \mid p = 0.5\right) + P\left(\text{Reject } H_0, \sum_{i=1}^{10} X_i = 8 \mid p = 0.5\right)$$

The last probability can be calculated as follows:

$$\begin{aligned} &P\left(\text{Reject } H_0, \sum_{i=1}^{10} X_i = 8 \mid p = 0.5, U < 0.85\right) \times P(U < 0.85) \\ &\quad + P\left(\text{Reject } H_0, \sum_{i=1}^{10} X_i = 8 \mid p = 0.5, U \geq 0.85\right) \times P(U \geq 0.85) \\ &= 0.85 \times P\left(X_i = 8 \mid p = 0.5\right). \end{aligned}$$

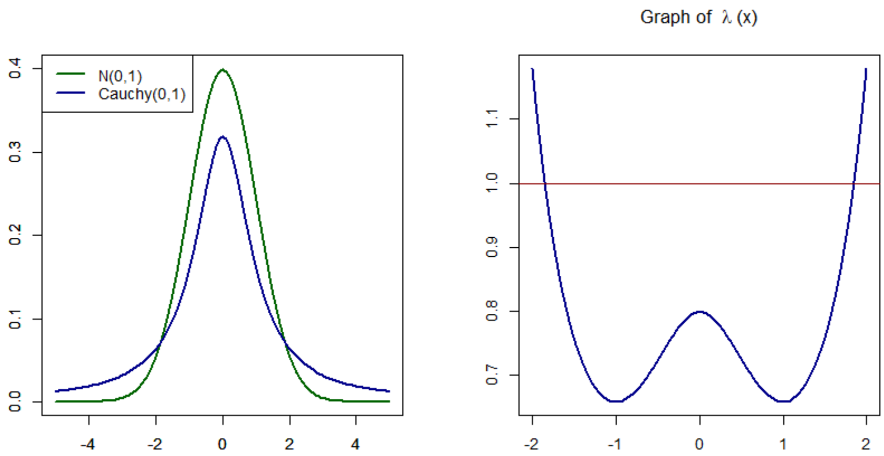

Figure 1: Density plot and $\lambda(x) = f_1(x)/f_0(x)$ for $\texttt{normal}(0, 1)$ against $\texttt{Cauchy}(0, 1)$ problem.

Thus, $P(\text{Type I error}) = 0.0481$, which also satisfies the level condition.

Next, consider the power of the test. A similar calculation would lead to

$$\text{Power} = 1 - P\left(\text{Type II error}\right) = P\left(\text{Reject } H_0 \mid H_1\right) = 0.4834,$$

which is much higher than the power of ϕ_2. Is ϕ_4 the MP test? No, because one may further adjust the test function for the case $\sum x_i = 8$ to get better power, at the cost of reduced probability of type I error. One may continue to make adjustments as long as the level condition is valid.

Definition 14. *(Randomized and Non-randomized Tests) A test function of the form $I_C(\mathbf{x})$ is called a non-randomized test. Any other test function (a function from the sample space S_x to $[0, 1]$) corresponds to a randomized test.*

2 Neyman Pearson Lemma

The following theorem prescribes a method of obtaining the most powerful test for testing a simple versus simple hypothesis.

Theorem 1 (Neyman-Pearson Lemma)**.** *Consider the problem of testing simple vs simple hypotheses, $H_0 : \mathbf{X} \sim f_0(\mathbf{x})$ against $H_1 : \mathbf{X} \sim f_1(\mathbf{x})$, using a test ϕ that satisfies*

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \ f_1(\mathbf{x}) > k f_0(\mathbf{x}), \\ 0 & \text{if } \ f_1(\mathbf{x}) < k f_0(\mathbf{x}), \end{cases} \tag{2}$$

for some $k \geq 0$, and

$$\alpha = E_{H_0}\left[\phi(\mathbf{X})\right]. \tag{3}$$

i. Any test that satisfies (2) and (3) is a MP level α test.

ii. If there exists a test satisfying (2) and (3) with $k > 0$, then every MP level α test is a size α test (i.e., satisfies (3)), and every MP level α test satisfies (2) except perhaps on a set A where $P(\mathbf{X} \in A \mid H_0) = P(\mathbf{X} \in A \mid H_1) = 0$.

[Proof]

Example 3. Let X be a random variable with p.m.f. under H_0 and H_1 are as follows. Find an MP level $\alpha = 0.025$ test using Neyman-Pearson Lemma.

x	1	2	3	4	5	6
$f_0(x)$	0.01	0.01	0.01	0.01	0.01	0.95
$f_1(x)$	0.05	0.04	0.03	0.02	0.01	0.85
$\lambda(x)$	5	4	3	2	1	17/19

Example 4. Let $X \sim N(0,1)$ under H_0 and $X \sim \texttt{Cauchy}(0,1)$ under H_1. Find a UMP level $\alpha = 0.05$ test using Neyman-Pearson Lemma.

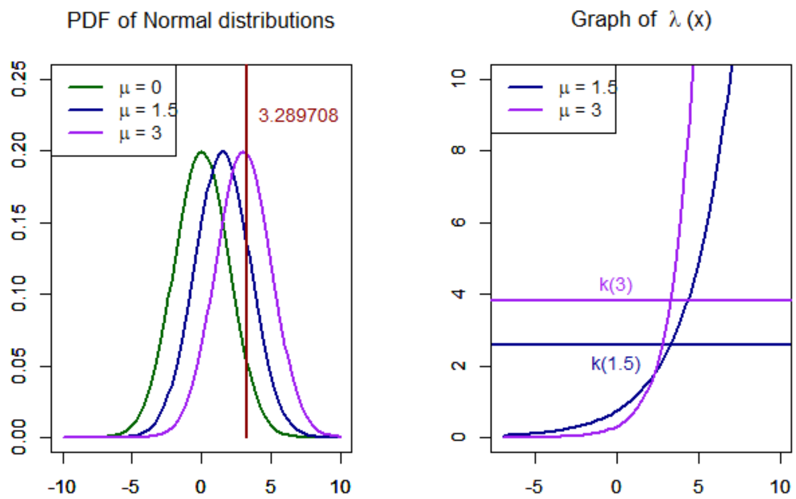

Figure 2: Density plot (panel 1) and $\lambda(x) = f_1(x)/f_0(x)$ (panel 2) for normal mean testing problem. The region $[3.289708, \infty)$ is the rejection region of the MP level 0.05 test for both the alternatives $\mu = 1.5$ and $\mu = 3$.

Corollary 1. *Consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, based on a random sample of size n, $X_1, \cdots, X_n$, from the parametric family $f_\theta; \theta \in \Theta$. Let $T(\mathbf{X})$ be a sufficient statistic for θ with pdf/pmf $f_T(t \mid \theta)$. Next, consider a test $\phi^\star$ of the following form:*

$$\phi^\star(\mathbf{x}) = \begin{cases} 1 & \text{if } \; f_T(t \mid \theta_1) > k f_T(t \mid \theta_0), \; t = T(\mathbf{x}) \\ 0 & \text{if } \; f_T(t \mid \theta_1) < k f_T(t \mid \theta_0), \; t = T(\mathbf{x}), \end{cases}$$

and $E_{\theta_0}\big(\phi^\star(\mathbf{x})\big) = \alpha$. The $\phi^\star$ is a MP level α test.

Example 5. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Test $H_0 : \mu = \mu_0, \sigma = \sigma_0$ against $H_1 : \mu = \mu_1, \sigma = \sigma_0$.

Example 6. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Test $H_0 : \mu = \mu_0, \sigma = \sigma_0$ against $H_1 : \mu = \mu_0, \sigma = \sigma_1$.

Example 7. Let $X_1, \ldots, X_n$ be a random sample from Poisson(λ). Test $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$.

Definition 15. *(Test Statistics) As the above corollary suggests, typically, a hypothesis test is specified in terms of the values of a statistic $T(\mathbf{X})$. This statistic is called* test statistic.

Definition 16. *(p-value) The p-value is the probability of seeing data at least as extreme as the experimental data, assuming the null hypothesis to be true. What 'extreme' means depends on the experimental design and alternative hypothesis. The smaller the p-value, the more extreme the outcome and the stronger the evidence against H_0. It is defined as*

$$p = p(\mathbf{x}) := \inf\{\alpha : \; \textit{given } \mathbf{x}, \; H_0 \textit{ is rejected against } H_1 \textit{ at level } \alpha\}.$$

Revisit Examples 5-7, and calculate the p-values.

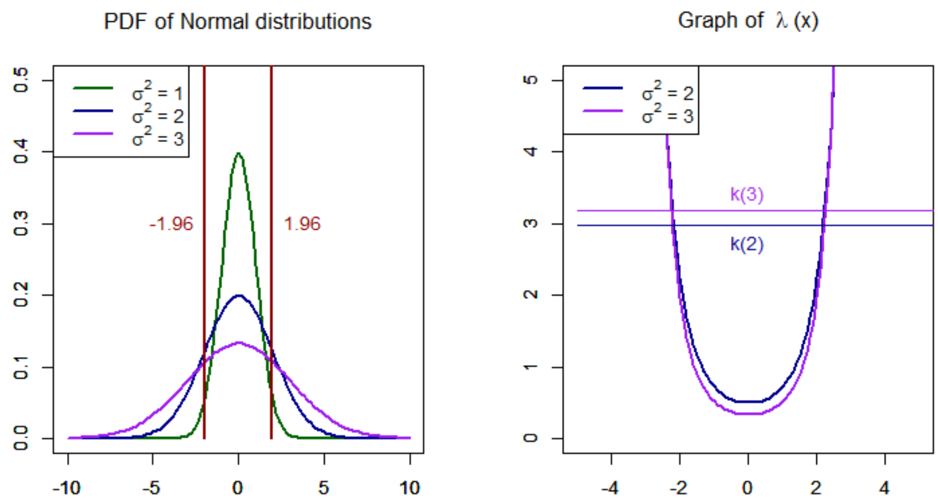

Figure 3: Density plot (panel 1) and $\lambda(x) = f_1(x)/f_0(x)$ (panel 2) for normal variance testing problem. The region $(-\infty, -1.96] \cup [1.96, \infty)$ is the rejection region of the MP level 0.05 test for both the alternatives $\sigma^2 = 2$ and $\sigma^2 = 3$.

3 Families with Monotone Likelihood Ratio

Neyman-Pearson's lemma prescribes a method of finding UMP test for simple vs simple hypothesis. The next step would be to find a UMP test for composite hypotheses. But, a UMP test for composite H_0 and/or composite H_1 may not generally exist. Consider, for example, the testing problem $H_0 : X \sim \text{Cauchy}(0, 1)$ against $H_1 : X \sim \text{Cauchy}(\theta, 1)$ for $\theta \in \{2, 4\}$. The graphs of $\lambda(x, \theta) = f_0(x)/f_\theta(x)$ is given in Figure 4. Observe that, by the NP lemma, the MP test for testing H_0 against $\theta = 2$ would reject H_0 given the point $x = 2.5$. However, the MP test for testing H_0 against $\theta = 4$ may not reject H_0 given the point $x = 2.5$, as the value of $\lambda(2.5)$ is quite small under the alternative $\theta = 4$. Thus, a separate MP test (depending on θ_1) is required, and a UMP test does not exist.

In what follows, we will consider a special class of distributions for which a UMP test of one-sided hypotheses exists.

Definition 17. *(Monotone Likelihood Ratio, MLR) A family of distributions with pdf/pmf $\{f_{\mathbf{X}}(;\theta), \theta \in \Theta \subseteq \mathbb{R}\}$ is said to have a monotone likelihood ratio (MLR) in a statistic $T(\mathbf{x})$ if for any $\theta_1 < \theta_2$ the ratio $\lambda(\mathbf{x}, \theta_1, \theta_2) = f_{\mathbf{X}}(\mathbf{x}; \theta_1)/f_{\mathbf{X}}(\mathbf{x}; \theta_2)$ is a monotone (non-increasing or non-decreasing) function of $T(\mathbf{x})$ for the set of values $\mathbf{x}$ for which at least one of $f_{\mathbf{X}}(\mathbf{x}; \theta_1)$ and $f_{\mathbf{X}}(\mathbf{x}; \theta_2)$ is positive.*

Example 8. `Uniform`$(0, \theta),\ \theta > 0$ distribution is MLR in $T(\mathbf{X}) = X_{(n)}$.

The MLR family is large enough to include the exponential class of distributions. The following theorem states that result.

Theorem 2. *The one-parameter exponential family $f_{\mathbf{X}}(\mathbf{x}; \theta) = \exp\{Q(\theta)T(\mathbf{x}) + S(\mathbf{x}) + D(\theta)\}$ where $Q(\theta)$ is non-decreasing has MLR in $T(\mathbf{x})$.* [Homework]

Theorem 3. *Let $X_1, \cdots, X_n$ be a random sample with some distribution with pdf/pmf $\{f_X(\cdot; \theta) : \theta \in \Theta\}$, which has MLR in $T(\mathbf{X})$ (such that for $\theta_1 < \theta_2$, $\lambda(\mathbf{x}, \theta_1, \theta_2)$ is non-increasing in $T(\mathbf{x})$). For testing $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$, $\theta_0 \in \Theta$, and test of the form*

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > t_0 \\ \gamma & \text{if } T(\mathbf{x}) = t_0 \\ 0 & \text{if } T(\mathbf{x}) < t_0 \end{cases}, \tag{4}$$

has a non-decreasing power function and is UMP of its size $E_{\theta_0}[\phi(\mathbf{X})] = \alpha$. [Without proof]

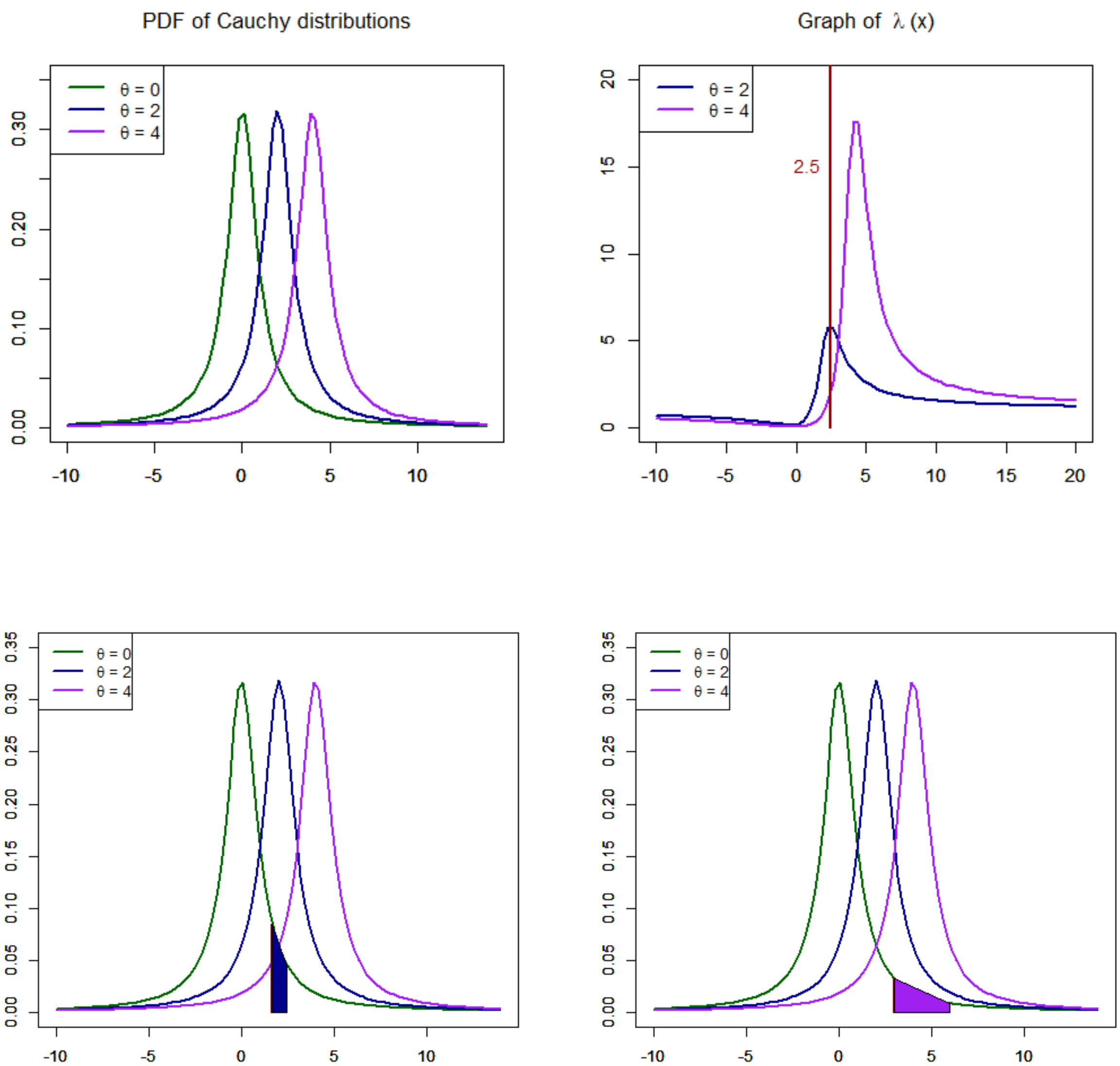

Figure 4: Density plots (in 1st, 3rd and 4th panel) and $\lambda(x) = f_1(x)/f_0(x)$ (2nd panel) for $\texttt{Cauchy}(0,1)$ against $\texttt{Cauchy}(\theta,1)$ problem. The shaded regions in 3rd and 4th panels indicate the rejection regions of the MP level 0.05 test for $\theta = 2$ and $\theta = 4$, respectively.

	$\lambda(\mathbf{x}, \theta_1, \theta_2)$ is non-decreasing in T	$\lambda(\mathbf{x}, \theta_1, \theta_2)$ is non-increasing in T
$H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$	ϕ' as in (5) is an UMP	ϕ as in (4) is an UMP
$H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$	ϕ as in (4) is an UMP	ϕ' as in (5) is an UMP

Table 1: UMP test for different types of MLR families. Here $\theta_1 < \theta_1$ and $\lambda(\mathbf{x}, \theta_1, \theta_2)$ is as define in the definition of MLR.

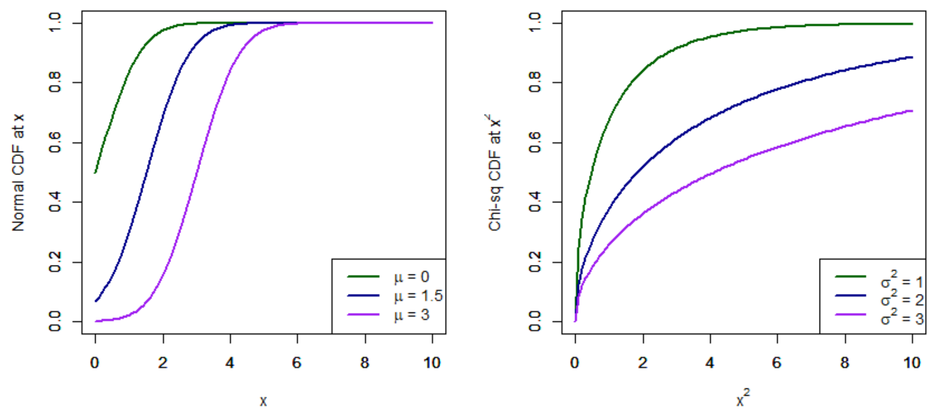

Figure 5: Some examples of MLR families

Remark 4. *If a family of pdfs/pmfs $\{f_X(x;\theta), \theta \in \Theta\}$ has an MLR in T, then the corresponding family of CDFs of T is stochastically increasing in θ.*

Remark 5. *Remark 4 implies that the power function of the test in Theorem 3 is monotone in θ. This property plays a key role in the proof of Theorem 3 (see Casella Berger).*

Remark 6. *If for $\theta_1 < \theta_2$, the ratio $f_{\mathbf{X}}(\mathbf{x};\theta_1)/f_{\mathbf{X}}(\mathbf{x};\theta_2)$ is monotonically non-decreasing in T, then a test of the form*

$$\phi'(\mathbf{x}) = \begin{cases} 1 & \text{if } \; T(\mathbf{x}) < t_0 \\ \gamma & \text{if } \; T(\mathbf{x}) = t_0 \\ 0 & \text{if } \; T(\mathbf{x}) > t_0 \end{cases}, \tag{5}$$

is a UMP for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ with size $E_{\theta_0}[\phi(\mathbf{X})] = \alpha$.

In general, four possible situations may arise, w.r.t. the relation of T in $f_X(\cdot;\theta)$ and the testing problem. Table 1 indicates a UMP test in all these situations.

Example 8. (continue) Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Uniform}(0, \theta)$. Find a UMP test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

Example 9. Let $X_1, \ldots, X_n$ be a random sample from Location-scale exponential(μ, σ_0). Find a UMP test for testing $H_0 : \mu \geq \mu_0,\ \sigma = \sigma_0$ against $H_1 : \mu < \mu_0,\ \sigma = \sigma_0$.