Lecture Notes for STAT-F404, author: Marc Hallin[1]

5 Hypothesis Testing: UMP Tests

5.1 The decision problem

Consider the statistical model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, where $\mathcal{P}$ is partitioned into $\mathcal{P} = H_0 \oplus H_1$, along with the decision space $\mathcal{D} = \{RH_0, \not{R}H_0\} = \{1, 0\}$; here, RH_0 and $\not{R}H_0$ (equivalently, 1 and 0) respectively stand for "reject H_0" and "do not reject H_0". Consider also the loss function defined by

$$L_P(d) = \begin{cases} 1 & \text{if P} \in H_1 \text{ and } d = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The cost of not rejecting H_0 when H_0 is false (the so-called *Type II error*) thus is one, while rejecting H_0 when H_0 is true (*Type I error*) has cost zero.

A randomized decision rule—a collection, indexed by $\mathbf{x} \in \mathcal{X}$, of conditional (on $\mathbf{X}$=$\mathbf{x}$) distributions over the two points "0" ($\not{R}H_0$) and "1" (RH_0)—is entirely described by the $\mathbf{X}$-measurable probability mass it puts on "1" (RH_0), that is, an $\mathbf{X}$-measurable statistic, $\phi(\mathbf{X})$, say, with values in $[0, 1]$. The set of all possible randomized decision rules is thus

$$\mathcal{T} := \{\, \phi : \phi(\mathbf{x}) \text{ a statistic with values in } ([0,1], \mathcal{B}_{[0,1]})\}, \qquad \mathcal{B}_{[0,1]} := \mathcal{B} \cap [0,1],$$

with the interpretation that, in case the randomized decision rule ϕ is adopted, conditional on $\mathbf{X} = \mathbf{x}$, decision "1" (RH_0) will be taken with probability $\phi(\mathbf{x})$. If $\mathbf{x}$ is observed and $\phi(\mathbf{x}) = 1/2$, then the statistician thus can flip a fair coin in order to decide between RH_0 and $\not{R}H_0$; if $\phi(\mathbf{x}) = 1/6$, then she/he can roll a dice, etc. Of course, if $\phi(\mathbf{x}) = 1$ or 0, then she/he will reject or not reject without randomization.

[1]With slight modifications by Davy Paindaveine and Thomas Verdebout.

A specific terminology is associated with this decision problem:

– H_0 is called the *null hypothesis*, H_1 the *alternative hypothesis*; together, they characterize a *testing problem*.

– A decision rule ϕ (a statistic with values in $[0,1]$) is called a *(randomized) test*. If ϕ is such that $\mathrm{P}[\phi(\mathbf{X}) \in \{0,1\}] = 1$ for any $\mathrm{P} \in \mathcal{P}$, it is called a *nonrandomized* or *pure test*.

– The unconditional probability under P that a given test ϕ eventually leads to the rejection of H_0 is

$$\mathrm{E}_{\mathrm{P}}[\phi] = \int_{\mathcal{X}} \phi(x)\, d\mathrm{P}(x);$$

this quantity is called the *size* of ϕ when $\mathrm{P} \in H_0$, the *power* of ϕ when $\mathrm{P} \in H_1$.

– The risk (the expected loss) associated with a test ϕ is

$$\mathrm{R}^{\phi}_{\mathrm{P}} = \begin{cases} 1 - \mathrm{E}_{\mathrm{P}}[\phi] & \text{if } \mathrm{P} \in H_1 \\ 0 & \text{if } \mathrm{P} \in H_0 \end{cases}$$

(under $\mathrm{P} \in H_1$, that risk is the probability of ϕ committing Type II error and is called the *Type II risk*). That risk $\mathrm{R}^{\phi}_{\mathrm{P}}$ is to be minimized uniformly in $\mathrm{P} \in H_1$. Equivalently, the power of ϕ, $\mathrm{E}_{\mathrm{P}}[\phi]$, $\mathrm{P} \in H_1$, is to be maximized uniformly in $\mathrm{P} \in H_1$.

Clearly, if the power is to be maximized with respect to $\phi \in \mathcal{T}$, without placing any restriction on ϕ, then the trivial test $\phi(\mathbf{x}) = 1$ $\mathcal{P}$-almost surely, which rejects H_0 irrespective of the observed value $\mathbf{x}$ of $\mathbf{X}$, qualifies as the uniformly most powerful test, hence the solution of the testing problem. Such a trivial solution is ruled out by the following principle.

The Neyman principle. Fix some $\alpha \in (0,1)$, and restrict to the class $\mathcal{C}_\alpha$ of *α-level tests*, i.e., of the tests ϕ satisfying the *level constraint*

$$\mathrm{E}_{\mathrm{P}}[\phi] \leq \alpha \text{ for all } \mathrm{P} \in H_0. \tag{5.1}$$

A test ϕ^* is said to be *uniformly most powerful* (UMP) within a class $\mathcal{C}$ of tests if

(a) $\phi^* \in \mathcal{C}$, and

(b) for all $\phi \in \mathcal{C}$ and all $\mathrm{P} \in H_1$, $\mathrm{E}_\mathrm{P}[\phi^*] \geq \mathrm{E}_\mathrm{P}[\phi]$.

That principle, often complemented by some further ones, will be considered throughout the chapters on hypothesis testing. A test ϕ^* which is uniformly most powerful within the class $\mathcal{C}_\alpha = \{\phi : \mathrm{E}_\mathrm{P}[\phi] \leq \alpha \text{ for all } \mathrm{P} \in H_0\}$ of α-level tests is called *uniformly most powerful at level* α, or α*-level uniformly most powerful*.

5.2 The Neyman-Pearson Lemma

5.2.1 Testing a simple null against a simple alternative

A hypothesis H (null or alternative) is called *simple* if it contains a single element. Else, it is called *composite*. The simplest of all hypothesis testing problems is that of testing a *simple null* $H_0 = \{\mathrm{P}_0\}$ against a *simple alternative* $H_1 = \{\mathrm{P}_1\}$. The problem then consists in maximizing $\mathrm{E}_1[\phi] := \mathrm{E}_{\mathrm{P}_1}[\phi] = \int \phi(\mathbf{x}) d\mathrm{P}_1(\mathbf{x})$ under the level constraint $\mathrm{E}_0[\phi] := \mathrm{E}_{\mathrm{P}_0}[\phi] = \int \phi(\mathbf{x}) d\mathrm{P}_0(\mathbf{x}) \leq \alpha$. Maximizing such an integral under an integral constraint is a standard variational problem. Its solution, along with some properties, is summarized in the following remark, known as the *Neyman-Pearson Fundamental Lemma*.

Note that P_0 and P_1 are dominated by the sum $\mu := \mathrm{P}_0 + \mathrm{P}_1$; it will be convenient to use the corresponding densities

$$f_0 := \frac{d\mathrm{P}_0}{d\mu} \quad \text{and} \quad f_1 := \frac{d\mathrm{P}_1}{d\mu} \cdot$$

Also, instead of "uniformly most powerful" (UMP), in this context, we simply say "most powerful" (MP); "uniformly" here indeed means "uniformly in $\mathrm{P} \in H_1$", which in the present case is superfluous, as H_1 is simple.

Before stating the Neyman-Pearson Lemma, let us define a *Neyman test with constant* k

(for the simple H_0 against the simple H_1) as a test of the form

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

where $k \in \bar{\mathbb{R}}^+ := \mathbb{R}^+ \cup \{\infty\}$ and $\mathbf{x} \mapsto \gamma(\mathbf{x})$ takes values in $[0, 1]$.

The Neyman-Pearson Lemma generally consists of the following fourfold statement.

Lemma 1 (Neyman-Pearson Lemma)**.** *Consider the statistical model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, with $\mathcal{P} := \{\mathrm{P}_0, \mathrm{P}_1\}$, the null hypothesis $H_0 := \{\mathrm{P}_0\}$, and the alternative $H_1 := \{\mathrm{P}_1\}$. Fix $\alpha \in (0, 1)$. Then, we have the following:*

(i) There exist $k \in \mathbb{R}^+$ and $\gamma \in [0, 1]$ such that the test

$$\phi^*_\alpha(\mathbf{x}) := \begin{cases} 1 & \textit{if } f_1(\mathbf{x}) > k f_0(\mathbf{x}) \\ \gamma & \textit{if } f_1(\mathbf{x}) = k f_0(\mathbf{x}) \\ 0 & \textit{if } f_1(\mathbf{x}) < k f_0(\mathbf{x}) \end{cases}$$

satisfies $\mathrm{E}_0[\phi^*_\alpha] = \alpha$ (size constraint).

(ii) The test ϕ^*_α is most powerful at level α.

(iii) Conversely, if ϕ is such that $\mathrm{E}_0[\phi] \leq \alpha$ and $\mathrm{E}_1[\phi] = \mathrm{E}_1[\phi^*_\alpha]$, then $(\phi^*_\alpha(\mathbf{x}) - \phi'(\mathbf{x}))(f_1(\mathbf{x}) \neq k f_0(\mathbf{x})) = 0$ μ-a.e., or equivalently, $(\phi^*_\alpha(\mathbf{x}) - \phi'(\mathbf{x}))\mathbb{I}[f_1(\mathbf{x}) \neq k f_0(\mathbf{x})] = 0$ μ-a.e. (if an α-level test ϕ' is as powerful as ϕ^*_α, then it is also a Neyman test with constant k).

(iv) $\mathrm{E}_1[\phi^*_\alpha] > \alpha$.

Proof. (i) Let $F_0(z) := \mathrm{P}_0[f_1(\mathbf{X}) \leq z f_0(\mathbf{X})]$ for any z. Noting that $z \mapsto F_0(z)$ is a cumulative distribution function, define

$$k := \inf\{z : F_0(z) > 1 - \alpha\} \quad \text{and} \quad \gamma = \begin{cases} \dfrac{F_0(k) - (1 - \alpha)}{F_0(k) - F_0(k - 0)} & \text{if } F_0(k) > F_0(k - 0) \\ 0 & \text{if } F_0(k) = F_0(k - 0), \end{cases}$$

where $F_0(k-0)$ denotes the limit of $F_0(z)$ when z converges to k from below. Then,

$$\begin{aligned} \mathrm{E}_0[\phi^*_\alpha] &= \mathrm{P}_0[f_1(\mathbf{X}) > kf_0(\mathbf{X})] + \gamma \mathrm{P}_0[f_1(\mathbf{X}) = kf_0(\mathbf{X})] + 0 \times \mathrm{P}_0[f_1(\mathbf{X}) < kf_0(\mathbf{X})] \\ &= 1 - F_0(k) + \frac{F_0(k) - (1-\alpha)}{F_0(k) - F_0(k-0)}(F_0(k) - F_0(k-0)) = \alpha. \end{aligned}$$

<u>Remark 1</u>: Note that if F_0^{-1} is well-defined at $1-\alpha$, then $F_0(k) = 1-\alpha$ and $\gamma = 0$: ϕ^*_α is a *pure test* involving no randomization. If not, $F_0(k-0) \leq 1-\alpha < F_0(k)$, and $0 < \gamma \leq 1$. In case $\gamma < 1$, ϕ^*_α is a randomized test (in case $\gamma = 1$, again, no randomization is involved, but the critical region is of the form $\{\mathbf{x} : f_1(\mathbf{x}) \geq kf_0(\mathbf{x})\}$).

(ii) For any ϕ satisfying $\mathrm{E}_0[\phi] \leq \alpha$, consider the integral (with respect to $\mu = \mathrm{P}_0 + \mathrm{P}_1$)

$$\int_{\mathcal{X}} (\phi^*_\alpha(\mathbf{x}) - \phi(\mathbf{x}))(f_1(\mathbf{x}) - kf_0(\mathbf{x}))d\mu(\mathbf{x}). \tag{5.2}$$

The integrand in (5.2) is nonnegative for all $\mathbf{x}$: indeed,

– either $f_1(\mathbf{x}) - kf_0(\mathbf{x}) < 0$; then $\phi^*_\alpha(\mathbf{x}) - \phi(\mathbf{x}) = -\phi(\mathbf{x}) \leq 0$, and the integrand is nonnegative;

– or $f_1(\mathbf{x}) - kf_0(\mathbf{x}) > 0$; then $\phi^*_\alpha(\mathbf{x}) - \phi(\mathbf{x}) = 1 - \phi(\mathbf{x}) \geq 0$, and the integrand again is nonnegative;

– or $f_1(\mathbf{x}) - kf_0(\mathbf{x}) = 0$, and the integrand is zero, hence in particular nonnegative.

It follows that the integral itself is nonnegative. Developing that integral yields

$$\begin{aligned} 0 &\leq \mathrm{E}_1[\phi^*_\alpha] - \mathrm{E}_1[\phi] - k(\mathrm{E}_0[\phi^*_\alpha] - \mathrm{E}_0[\phi]) \\ &= \mathrm{E}_1[\phi^*_\alpha] - \mathrm{E}_1[\phi] - k(\alpha - \mathrm{E}_0[\phi]), \end{aligned} \tag{5.3}$$

hence (since $k \geq 0$ and $\mathrm{E}_0[\phi] \leq \alpha$)

$$\mathrm{E}_1[\phi^*_\alpha] - \mathrm{E}_1[\phi] \geq k(\alpha - \mathrm{E}_0[\phi]) \geq 0,$$

as was to be shown.

(iii) Assume that ϕ satisfies $\mathrm{E}_0[\phi] \leq \alpha$ and is as powerful as ϕ^*_α. Then, (5.3) yields

$$0 \leq \mathrm{E}_1[\phi^*_\alpha] - \mathrm{E}_1[\phi] - k(\alpha - \mathrm{E}_0[\phi]) = -k(\alpha - \mathrm{E}_0[\phi]) \leq 0, \tag{5.4}$$

so that the integral (5.2) is zero. As an integral with nonnegative integrand, however, (5.2) only can take value 0 if that integrand is μ-almost everywhere zero, which establishes the result (that is, ϕ^*_α and ϕ coincide μ-almost everywhere, except possibly in the possible randomization part where $f_1(\mathbf{x}) = kf_0(\mathbf{x})$).

(iv) Clearly, the trivial test defined by $\phi_0(\mathbf{x}) = \alpha$ for any $\mathbf{x}$ has level α. Since ϕ^*_α is most powerful at level α, we must then have $\mathrm{E}_1[\phi^*_\alpha] \geq \mathrm{E}_1[\phi_0] = \alpha$. Now, assume that $\mathrm{E}_1[\phi^*_\alpha] = \alpha$. Then,

$$\begin{aligned}
&\mu(\{\mathbf{x} : f_1(\mathbf{x}) \neq kf_0(\mathbf{x})\}) \\
&\quad = \mu(\{\mathbf{x} : f_1(\mathbf{x}) \neq kf_0(\mathbf{x}), \phi^*_\alpha(\mathbf{x}) = \phi_0(\mathbf{x})\}) + \mu(\{x : f_1(\mathbf{x}) \neq kf_0(\mathbf{x}), \phi^*_\alpha(\mathbf{x}) \neq \phi_0(\mathbf{x})\}) \\
&\quad =: T_1 + T_2 = 0
\end{aligned}$$

(T_1 is zero because $\phi_0(\mathbf{x}) = \alpha(\in (0,1))$ cannot be equal to $\phi^*_\alpha(\mathbf{x})$ when $f_1(\mathbf{x}) \neq kf_0(\mathbf{x})$, whereas T_2 is zero from Part (iii) of the lemma). Thus, $f_1(\mathbf{x}) = kf_0(\mathbf{x})$ μ-almost everywhere. Since $\int_{\mathcal{X}} f_0(\mathbf{x})d\mu(\mathbf{x}) = \int_{\mathcal{X}} f_1(\mathbf{x})d\mu(\mathbf{x}) = 1$, we then have that $f_1(\mathbf{x}) = f_0(\mathbf{x})$ μ-almost everywhere, which implies that $\mathrm{P}_0 = \mathrm{P}_1$, a contradiction. This completes the proof of the lemma. □

Remark 2: It follows from the proof of the Neyman-Pearson Lemma that

(a) *any test of the form*

$$\phi(\mathbf{x}) = \begin{cases} 1 & \textit{if } f_1(\mathbf{x}) > kf_0(\mathbf{x}) \\ 0 & \textit{if } f_1(\mathbf{x}) < kf_0(\mathbf{x}) \end{cases} \tag{5.5}$$

for some $k \geq 0$ (*no specification in case* $f_1(\mathbf{x}) = kf_0(\mathbf{x})$) *is most powerful, at level* $\mathrm{E}_0[\phi]$, *for* $\{\mathrm{P}_0\}$ *against* $\{\mathrm{P}_1\}$;

(b) *for any test of the form (5.5), there exists a test of the form*

$$\phi'(\mathbf{x}) = \begin{cases} 1 & \text{if } f_1(\mathbf{x}) > kf_0(\mathbf{x}) \\ \gamma & \text{if } f_1(\mathbf{x}) = kf_0(\mathbf{x}) \\ 0 & \text{if } f_1(\mathbf{x}) < kf_0(\mathbf{x}), \end{cases}$$

with $\gamma \in [0,1]$, *such that* $\mathrm{E}_0[\phi'] = \mathrm{E}_0[\phi]$ *and* $\mathrm{E}_1[\phi'] = \mathrm{E}_1[\phi]$;

(c) *unless* $\mathrm{P}_0 = \mathrm{P}_1$, *any test of the form (5.5) with* $\mathrm{E}_0[\phi] < 1$ *is such that* $\mathrm{E}_1[\phi] > \mathrm{E}_0[\phi]$.

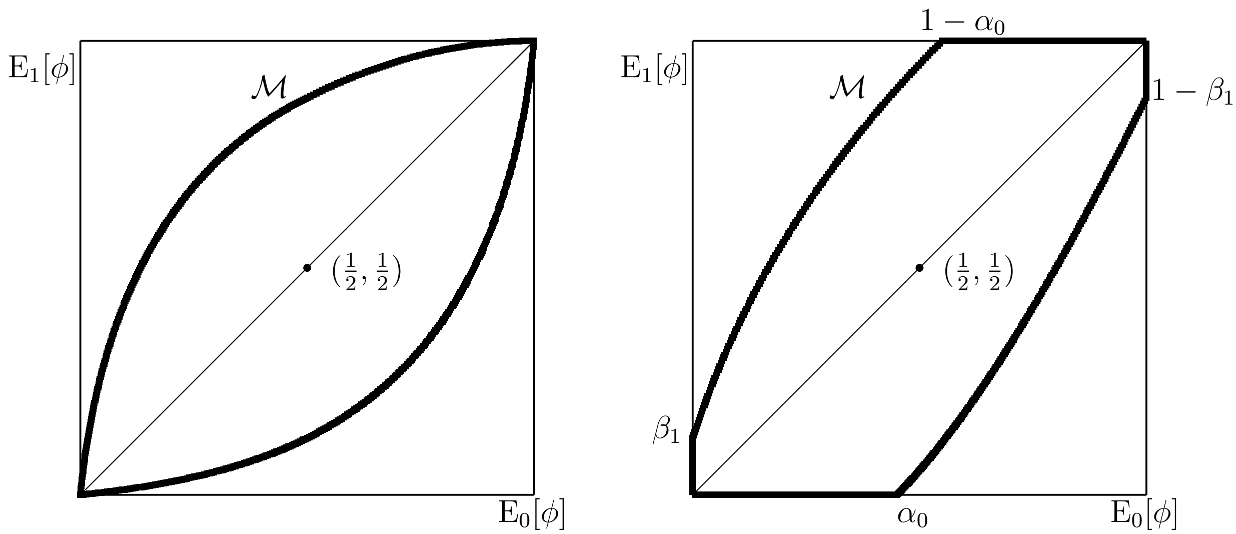
The intuitive interpretation of the optimality property of test of the Neyman type is essentially the following: with P_0-probability one, $f_1(\mathbf{X}) > kf_0(\mathbf{X})$ is equivalent to $f_1(\mathbf{X})/f_0(\mathbf{X}) > k$, where $f_1(\mathbf{x})/f_0(\mathbf{x})$, the *likelihood ratio*, can be interpreted as an "exchange rate" between size and power, between type I risk (the P_0-probability of rejecting) and power (the P_1-probability of rejecting). The optimal test ϕ^*_α in part (ii) of the Lemma thus consists in spending "a total amount α" of type I risk on those points $\mathbf{x}$ where the "exchange rate" is most favorable.

5.2.2 The power diagram

If two tests ϕ' and ϕ'' are such that $\mathrm{E}_0[\phi'] = \mathrm{E}_0[\phi'']$ and $\mathrm{E}_1[\phi'] = \mathrm{E}_1[\phi'']$, they are perfectly equivalent from a decision-theoretic point of view: same size, same power. Therefore, we may identify all tests ϕ having (for a given testing problem, of the form $H_0 = \{\mathrm{P}_0\}$, $H_1 = \{\mathrm{P}_1\}$) the same size $\mathrm{E}_0[\phi]$ and the same power $\mathrm{E}_1[\phi]$ with the point $(\mathrm{E}_0[\phi], \mathrm{E}_1[\phi])$ in the unit square. The set

$$\mathcal{M} := \{(\mathrm{E}_0[\phi], \mathrm{E}_1[\phi]) : \phi \text{ is a test}\}$$

is called the *power diagram* (for P_0 and P_1). It has the following typical form.



The lefthand panel corresponds to the particular case where P_0 and P_1 are absolutely continuous with respect to each other, whereas the righthand panel is the general case. As for the quantities α_0 and β_1,

- $\beta_1 := \mathrm{P}_1[f_0(\mathbf{X}) = 0]$ is the maximal power of a test with size zero (achieved by $\phi(\mathbf{x}) = \mathbb{I}[f_0(\mathbf{x}) = 0]$), whereas
- $(1 - \alpha_0) := \mathrm{P}_0[f_1(\mathbf{X}) > 0]$ is the minimal size of a test with power one (achieved by $\phi(\mathbf{x}) = \mathbb{I}[f_1(\mathbf{x}) > 0]$).

Less importantly, $\alpha_0 = \mathrm{P}_0[f_1(\mathbf{X}) = 0]$ is then the maximal size of a test with power zero (achieved by $\phi(\mathbf{x}) = \mathbb{I}[f_1(\mathbf{x}) = 0]$) and $(1 - \beta_1) = \mathrm{P}_1[f_0(\mathbf{X}) > 0]$ is then the minimal power of a test with size one (achieved by $\phi(\mathbf{x}) = \mathbb{I}[f_0(\mathbf{x}) > 0]$). Whenever P_0 and P_1 are absolutely continuous with respect to each other, $\alpha_0 = \beta_1 = 0$.

The following proposition provides some elementary properties of power diagrams.

Proposition. (i) *The main diagonal of the unit square, representing the tests of the form* $\phi_0 = \alpha$ *μ-almost everywhere* ($\alpha \in [0, 1]$)*, is in* $\mathcal{M}$*;*

(ii) $\mathcal{M}$ *is symmetric with respect to* $(\frac{1}{2}, \frac{1}{2})$*;*

(iii) $\mathcal{M}$ *is convex;*

(iv) *the "upper boundary" of* $\mathcal{M}$ *represents the Neyman-Pearson Lemma tests;*

(v) $\mathcal{M}$ *is closed, hence compact.*

Except for part (v), all statements in this proposition are quite elementary; proofs are left to the reader.

5.3 Families with monotone likelihood ratios

Testing a simple null against a simple alternative is of theoretical rather than practical interest. The simplest problems (for a one-parameter family $\{\mathrm{P}_\theta : \theta \in \Theta\}$, where Θ is an interval of $\mathbb{R}$—possibly, $\mathbb{R}$ itself) that are of practical relevance are of the form

$$H_0 = \{\mathrm{P}_\theta : \theta \leq \theta_0\} \qquad \text{vs} \qquad H_1 = \{\mathrm{P}_\theta : \theta > \theta_0\},$$

which is often simply written as $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Such hypotheses are called *one-sided*. They only make sense, of course, for $\theta_0 \in \mathrm{int}(\Theta)$—an assumption which is tacitly made throughout this section. Of course, the opposite problem, with $H_0 : \theta \geq \theta_0$ and $H_1 : \theta < \theta_0$, is equally interesting, but essentially equivalent.

A family $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ is said to have *monotone likelihood ratio in the (real-valued) statistic* T if (i) it is dominated by some σ-finite measure μ and (ii) there exist versions of the densities $f_\theta := \frac{d\mathrm{P}_\theta}{d\mu}$ such that, for any $\theta < \theta'$, the ratio

$$\frac{f_{\theta'}(\mathbf{x})}{f_\theta(\mathbf{x})}$$

is a nondecreasing function of $T(\mathbf{x})$.

Example 1: Binomial $\mathrm{Bin}(n, p)$ families, with densities (over $\mathbb{R}$, with respect to the counting measure of the set $\{0, 1, \ldots, n\}$)

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

indexed by $\theta = p \in [0,1]$, have monotone likelihood ratio with respect to $T(x) = x$.

Example 2: Poisson families, with densities (over $\mathbb{R}^n$ for a sample of size n, with respect to the counting measure of $\mathbb{N}^n$)

$$f_\lambda(\mathbf{x}) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \qquad \mathbf{x} = (x_1, \ldots, x_n)$$

indexed by $\lambda \in \mathbb{R}_0^+$, have monotone likelihood ratio with respect to $T(\mathbf{x}) = \sum_{i=1}^n x_i$.

Example 3: More generally, one-parameter exponential families, with densities (indexed by $\theta \in \Theta$)

$$f_\theta(\mathbf{x}) = C(\theta)h(\mathbf{x})\exp(\theta T(\mathbf{x}))$$

is a monotone likelihood ratio family with respect to the natural statistic $T(\mathbf{x})$.

As we shall see, the conclusions of the Neyman-Pearson Lemma almost directly extend to one-sided testing problems in families with monotone likelihood ratios—a fact we summarize in the following theorem (Karlin and Rubin, 1956).

Theorem 1. *Let $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ be a family with monotone likelihood ratio with respect to $T(\mathbf{x})$. Fix $\alpha \in (0,1)$ and $\theta_0 \in \mathrm{int}(\Theta)$. Then, (i) There exist $t_\alpha \in \mathbb{R}$ and $\gamma_\alpha \in [0,1]$ such that the test*

$$\phi_\alpha^*(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > t_\alpha \\ \gamma_\alpha & \text{if } T(\mathbf{x}) = t_\alpha \\ 0 & \text{if } T(\mathbf{x}) < t_\alpha \end{cases}$$

has size α under P_{θ_0}, that is, satisfies $\mathrm{E}_{\theta_0}[\phi_\alpha^*] = \alpha$. (ii) The size/power function $\theta \mapsto \mathrm{E}_\theta[\phi_\alpha^*]$ is strictly monotone increasing. (iii) The test ϕ_α^* is uniformly most powerful in the class of α-level tests for the problem of testing $H_0 : \theta \le \theta_0$ against $H_1 : \theta > \theta_0$.

Proof. (i) The proof of this part is very similar to the proof of the first part of the Neyman-Pearson fundamental lemma. Let $t \mapsto F_{\theta_0}^T(t) := \mathrm{P}_0[T(\mathbf{X}) \le t]$ be the cumulative distribution function of T under P_{θ_0}. Then, with

$$t_\alpha := \inf\{z : F^T_{\theta_0}(t) > 1-\alpha\} \quad \text{and} \quad \gamma_\alpha = \begin{cases} \dfrac{F^T_{\theta_0}(t_\alpha) - (1-\alpha)}{F^T_{\theta_0}(t_\alpha) - F^T_{\theta_0}(t_\alpha - 0)} & \text{if } F^T_{\theta_0}(t_\alpha) > F^T_{\theta_0}(t_\alpha - 0) \\ 0 & \text{if } F^T_{\theta_0}(t_\alpha) = F^T_{\theta_0}(t_\alpha - 0), \end{cases}$$

we have

$$\begin{aligned} \mathrm{E}_{\theta_0}[\phi^*_\alpha] &= \mathrm{P}_{\theta_0}[T(\mathbf{X}) > t_\alpha] + \gamma_\alpha \mathrm{P}_{\theta_0}[T(\mathbf{X}) = t_\alpha] + 0 \times \mathrm{P}_{\theta_0}[T(\mathbf{X}) < t_\alpha] \\ &= 1 - F^T_{\theta_0}(t_\alpha) + \frac{F^T_{\theta_0}(t_\alpha) - (1-\alpha)}{F^T_{\theta_0}(t_\alpha) - F^T_{\theta_0}(t_\alpha - 0)} (F^T_{\theta_0}(t_\alpha) - F^T_{\theta_0}(t_\alpha - 0)) = \alpha. \end{aligned}$$

(ii) Fix $\theta' < \theta''$ in Θ. In view of the monotone likelihood property, we have that $T(\mathbf{x})$ is larger than, equal to, or smaller than t_α if and only if $f_{\theta''}(\mathbf{x})/f_{\theta'}(\mathbf{x})$ is larger than, equal to, or smaller than some $k = k(\theta', \theta'', t_\alpha)$. Thus, the test ϕ^*_α rewrites

$$\phi^*_\alpha(\mathbf{x}) := \begin{cases} 1 & \text{if } f_{\theta''}(\mathbf{x}) > k f_{\theta'}(\mathbf{x}) \\ \gamma_\alpha & \text{if } f_{\theta''}(\mathbf{x}) = k f_{\theta'}(\mathbf{x}) \\ 0 & \text{if } f_{\theta''}(\mathbf{x}) < k f_{\theta'}(\mathbf{x}). \end{cases} \tag{5.6}$$

This is the Neyman-Pearson test for $H_0 : \theta = \theta'$ against $H_1 : \theta = \theta''$ at level $\mathrm{E}_{\theta'}[\phi^*]$. From Part (iv) of the Neyman-Pearson lemma, we thus have that $\mathrm{E}_{\theta''}[\phi^*_\alpha] > \mathrm{E}_{\theta'}[\phi^*_\alpha]$.

(iii) It directly follows from (i)–(ii) that ϕ^*_α is an α-level test for the problem of testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. Let then ϕ be an arbitrary α-level test for this problem. Fix $\theta_1 > \theta_0$ arbitrarily. Since ϕ is an α-level test for the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ and since ϕ^*_α is the Neyman-Pearson test for this problem at level α (this is seen by proceeding as in (ii) with $\theta' = \theta_0$ and $\theta'' = \theta_1$), we must have $\mathrm{E}_{\theta_1}[\phi^*_\alpha] \geq \mathrm{E}_{\theta_1}[\phi]$. This establishes the result. □

The result shows in particular that the size/power function $\theta \mapsto \mathrm{E}_\theta[\phi^*_\alpha]$ is strictly monotone increasing. In exponential families, it can actually be shown that

$$\frac{d}{d_\theta} \mathrm{E}_\theta[\phi_\alpha^*] > 0$$

at any θ such that $0 < \mathrm{E}_\theta[\phi_\alpha^*] < 1$ (a proof is available on request), which will actually play an important role in the next chapter.

5.4 A generalized Neyman-Pearson Lemma

Consider next the problem of testing the composite null hypothesis $H_0 = \{\mathrm{P}_1, \ldots, \mathrm{P}_m\}$ against the simple alternative $H_1 = \{\mathrm{P}_{m+1}\}$. Writing $\mu := \mathrm{P}_1 + \ldots + \mathrm{P}_{m+1}$, let

$$f_i := \frac{\mathrm{dP}_i}{\mathrm{d}\mu}, \qquad i = 1, \ldots, m+1,$$

and define the corresponding power diagram as

$$\mathcal{M}_{m+1} := \{(\mathrm{E}_1[\phi], \ldots, \mathrm{E}_{m+1}[\phi]) : \phi \text{ is a test}\}$$

(each point $\mathbf{y}$ in $\mathcal{M}_{m+1}$ represents a class of tests which are all equivalent from the point of view of size and power, therefore essentially the same from a decisional point of view). The power diagram $\mathcal{M}_{m+1}$ enjoys all elementary properties of $\mathcal{M}_2$: convexity, compactness, symmetry, etc. Note that the projection of $\mathcal{M}_{m+1}$ onto the space of its first m components is nothing else but $\mathcal{M}_m$.

A test of the form

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{m+1}(\mathbf{x}) > \sum_{i=1}^m k_i f_i(\mathbf{x}) \\ \gamma(\mathbf{x}) & \text{if } f_{m+1}(\mathbf{x}) = \sum_{i=1}^m k_i f_i(\mathbf{x}) \\ 0 & \text{if } f_{m+1}(\mathbf{x}) < \sum_{i=1}^m k_i f_i(\mathbf{x}), \end{cases}$$

where $k_1, \ldots, k_m$ are real numbers (not necessarily positive) is called a *generalized Neyman test*. The following result extends (in a somewhat weaker form, though) the fundamental Neyman-Pearson Lemma to the present context.

Proposition 1 (The generalized Neyman-Pearson Lemma — 1st version). *(i) For all* $\mathbf{c} = (c_1, \ldots, c_m) \in \mathcal{M}_m$*, there exists a test maximizing* $\mathrm{E}_{m+1}[\phi]$ *under the size constraints* $\mathrm{E}_i[\phi] = c_i$ *for* $i = 1, \ldots, m$*.*

(ii) If ϕ^* *satisfies* $\mathrm{E}_i[\phi] = c_i$*,* $i = 1, \ldots, m$*, with* $\mathbf{c} \in \mathcal{M}_m$ *and is of the generalized Neyman type, then it maximizes* $\mathrm{E}_{m+1}[\phi]$ *under the constraints* $\mathrm{E}_i[\phi] = c_i$ *for* $i = 1, \ldots, m$*.*

(iii) If, moreover, the Neyman test ϕ^* *in (ii) is such that* $k_i \geq 0$ *for* $i = 1, \ldots, m$*, then it also maximizes* $\mathrm{E}_{m+1}[\phi]$ *under the level constraints* $\mathrm{E}_i[\phi] \leq c_i$ *for* $i = 1, \ldots, m$*.*

(iv) if $\mathbf{c} = (c_1, \ldots, c_m)$ *is an interior point of* $\mathcal{M}_m$*, then there exists a Neyman test such that* $\mathrm{E}_i[\phi] = c_i$ *for* $i = 1, \ldots, m$ *(it follows from (ii) that this test automatically maximizes* $\mathrm{E}_{m+1}[\phi]$ *under the constraints* $\mathrm{E}_i[\phi] = c_i$ *for* $i = 1, \ldots, m$*).*

Proof. (i) Denote by D the "vertical" straight line through $\mathbf{c}$. The tests satisfying the constraints $\mathrm{E}_i[\phi] = c_i$, for $i = 1, \ldots, m$, are those represented by $\mathrm{D} \cap \mathcal{M}_{m+1}$. Due to convexity, $\mathrm{D} \cap \mathcal{M}_{m+1}$ is a ("vertical") segment $[\mathrm{B}^-, \mathrm{B}^+]$, with $\mathrm{B}^\pm := (c_1, \ldots, c_m, b^\pm)$ and $b^+ \geq b^-$. Any test represented by B^+ (a nonempty class) achieves the desired maximization, and the maximal value is b^+.

(ii) Let ϕ satisfy $\mathrm{E}_i[\phi] = c_i$. The integrand in

$$\int_{\mathcal{X}} (\phi^*(\mathbf{x}) - \phi(\mathbf{x}))\Big(f_{m+1}(\mathbf{x}) - \sum_{i=1}^{m} k_i f_i(\mathbf{x})\Big)\,\mathrm{d}\mu(\mathbf{x}) \tag{5.7}$$

is nonnegative; hence the integral also is. Thus,

$$\begin{aligned}\int_{\mathcal{X}} (\phi^*(\mathbf{x}) - \phi(\mathbf{x})) f_{m+1}(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x}) &\geq \sum_{i=1}^{m} k_i \int (\phi^*(\mathbf{x}) - \phi(\mathbf{x})) f_i(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x}) \\ &= \sum_{i=1}^{m} k_i (c_i - \mathrm{E}_i[\phi]) = 0,\end{aligned}$$

which reads $\mathrm{E}_{m+1}[\phi^*] \geq \mathrm{E}_{m+1}[\phi]$ (note, however, that this does not tell us anything about the *existence* of such a ϕ^*, nor about the values of k_i, $i = 1, \ldots, m$; on this point, we refer to Part (iv) of the proposition).

(iii) Let ϕ satisfy $\mathrm{E}_i[\phi] \leq c_i$, $i = 1, \ldots, m$. Since $k_i \geq 0$, $i = 1, \ldots, m$, nonnegativity of (5.7) now yields

$$\begin{aligned}\int_{\mathcal{X}} (\phi^*(\mathbf{x}) - \phi(\mathbf{x})) f_{m+1}(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) &\geq \sum_{i=1}^{m} k_i \int (\phi^*(\mathbf{x}) - \phi(\mathbf{x})) f_i(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \\ &= \sum_{i=1}^{m} k_i (c_i - \mathrm{E}_i[\phi]) \geq 0,\end{aligned}$$

which provides again $\mathrm{E}_{m+1}[\phi^*] \geq \mathrm{E}_{m+1}[\phi]$ (that conclusion is invalid as soon as one at least of the k_i's is negative).

(iv) Convexity of $\mathcal{M}_{m+1}$ and the Separating Hyperplane Theorem imply the existence of a hyperplane H such that B^+ (defined in Part (i) of the result) belongs to H and $\mathcal{M}_{m+1}$ lies entirely on one side of H. The point $\mathbf{c}$ belongs to the interior of $\mathcal{M}_m$, so that $\mathrm{B}^- \neq \mathrm{B}^+$. It clearly follows that $\mathrm{B}^- \in \mathcal{M}_{m+1}$ lies "below" B^+, so that $\mathcal{M}_{m+1}$ also entirely lies "below" H. The equation of H is (the general equation of a hyperplane through a point B^+ with coordinates $(c_1, \ldots, c_m, b^+)$)

$$\sum_{i=1}^{m} \tilde{k}_i y_i + \tilde{k}_{m+1} y_{m+1} = \sum_{i=1}^{m} \tilde{k}_i c_i + \tilde{k}_{m+1} b^+, \qquad \mathbf{y} \in \mathbb{R}^{m+1},$$

where the coefficients $\tilde{k}_i$ are defined up to a multiplicative constant. We have $\tilde{k}_{m+1} \neq 0$. Indeed, $\tilde{k}_{m+1} = 0$ would imply that the vertical line through $[\mathrm{B}^-, \mathrm{B}^+]$ belongs to H, which is not compatible with $\mathcal{M}_{m+1}$ being "below" H unless $\mathrm{B}^- = \mathrm{B}^+$; this, however, is ruled out by the assumption that $\mathbf{c}$ is an interior point of $\mathcal{M}_m$. Since $\tilde{k}_{m+1} \neq 0$, we may, without any loss of generality, assume $\tilde{k}_{m+1} = 1$. Putting $k_i = -\tilde{k}_i$, we then have that

$$y_{m+1} - \sum_{i=1}^{m} k_i y_i = b^+ - \sum_{i=1}^{m} k_i c_i \quad \text{if and only if } \mathbf{y} \text{ “on” H}$$
$$y_{m+1} - \sum_{i=1}^{m} k_i y_i < b^+ - \sum_{i=1}^{m} k_i c_i \quad \text{if and only if } \mathbf{y} \text{ “below” H.}$$

Since $\mathcal{M}_{m+1}$ is entirely below H, any test ϕ provides

$$\mathrm{E}_{m+1}[\phi] - \sum_{i=1}^{m} k_i \mathrm{E}_i[\phi] \leq \mathrm{E}_{m+1}[\phi^+] - \sum_{i=1}^{m} k_i \mathrm{E}_{m+1}[\phi^+],$$

where ϕ^+ is a test represented by B^+. This rewrites

$$\int_{\mathcal{X}} \phi(\mathbf{x})\Big(f_{m+1}(\mathbf{x}) - \sum_{i=1}^{m} k_i f_i(\mathbf{x})\Big)\,\mathrm{d}\mu(\mathbf{x}) \leq \int_{\mathcal{X}} \phi^+(\mathbf{x})\Big(f_{m+1}(\mathbf{x}) - \sum_{i=1}^{m} k_i f_i(\mathbf{x})\Big)\,\mathrm{d}\mu(\mathbf{x}),$$

where the k_i's, as the coefficients of the separating hyperplane H, are fixed. Hence, ϕ^+ is a maximizer, over all possible tests, of the integral

$$\int_{\mathcal{X}} \phi(\mathbf{x})\Big(f_{m+1}(\mathbf{x}) - \sum_{i=1}^{m} k_i f_i(\mathbf{x})\Big)\mathrm{d}\mu(\mathbf{x}).$$

That maximum, for ϕ ranging over the set of all possible tests, is clearly

$$\int_{\mathcal{X}} \Big(f_{m+1}(\mathbf{x}) - \sum_{i=1}^{m} k_i f_i(\mathbf{x})\Big)^+ \mathrm{d}\mu,$$

where $(z)^+ = \max(z, 0)$ is the positive part of a number z, and this maximum can only be achieved if ϕ^* is, μ-almost everywhere, of the form

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{m+1}(\mathbf{x}) > \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \\ \gamma(\mathbf{x}) & \text{if } f_{m+1}(\mathbf{x}) = \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \\ 0 & \text{if } f_{m+1}(\mathbf{x}) < \sum_{i=1}^{m} k_i f_i(\mathbf{x}), \end{cases}$$

that is, if it is a Neyman test with constants k_i, $i = 1, \ldots, m$. □

Two important remarks are in order.

Remark 1: The proof of Proposition 1 can actually be extended easily to cover the following slightly more general version of the result, that will be useful in the next chapter.

Proposition 2 (The generalized Neyman-Pearson Lemma — 2nd version)**.** *Let* $g_1, \ldots, g_{m+1} : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ *be measurable functions that are* μ*-integrable, and consider*

$$\mathcal{M}_m = \left\{ \left(\int_{\mathcal{X}} \phi(\mathbf{x}) g_1(\mathbf{x})\, d\mu(\mathbf{x}), \ldots, \int_{\mathcal{X}} \phi(\mathbf{x}) g_m(\mathbf{x})\, d\mu(\mathbf{x}) \right) : \phi \text{ a test} \right\}.$$

In this framework, calling "a generalized Neyman test" a test of the form

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } g_{m+1}(\mathbf{x}) > \sum_{i=1}^m k_i g_i(\mathbf{x}) \\ \gamma(\mathbf{x}) & \text{if } g_{m+1}(\mathbf{x}) = \sum_{i=1}^m k_i g_i(\mathbf{x}) \\ 0 & \text{if } g_{m+1}(\mathbf{x}) < \sum_{i=1}^m k_i g_i(\mathbf{x}), \end{cases}$$

where $k_1, \ldots, k_m$ *are real numbers, we have the following:*

(i) For all $\mathbf{c} = (c_1, \ldots, c_m) \in \mathcal{M}_m$*, there exists a test maximizing* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_{m+1}(\mathbf{x})\, d\mu(\mathbf{x})$ *under the size constraints* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) = c_i$ *for* $i = 1, \ldots, m$*.*

(ii) If ϕ^* *satisfies* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) = c_i$*,* $i = 1, \ldots, m$*, with* $\mathbf{c} \in \mathcal{M}_m$ *and is of the generalized Neyman type, then it maximizes* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_{m+1}(\mathbf{x})\, d\mu(\mathbf{x})$ *under the constraints* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) = c_i$ *for* $i = 1, \ldots, m$*.*

(iii) If, moreover, the Neyman test ϕ^* *in (ii) is such that* $k_i \geq 0$ *for* $i = 1, \ldots, m$*, then it also maximizes* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_{m+1}(\mathbf{x})\, d\mu(\mathbf{x})$ *under the constraints* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) \leq c_i$ *for* $i = 1, \ldots, m$*.*

(iv) if $\mathbf{c} = (c_1, \ldots, c_m)$ *is an interior point of* $\mathcal{M}_m$*, then there exists a Neyman test such that* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) = c_i$ *for* $i = 1, \ldots, m$ *(it follows from (ii) that this test automatically maximizes* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_{m+1}(\mathbf{x})\, d\mu(\mathbf{x})$ *under the constraints* $\int_{\mathcal{X}} \phi(\mathbf{x}) g_i(\mathbf{x})\, d\mu(\mathbf{x}) = c_i$ *for* $i = 1, \ldots, m$*).*

Of course, the first version of the generalized Neyman-Pearson lemma is recovered when g_i is taken as a density function f_i for any $i = 1, \ldots, m$.

<u>Remark 2:</u> In the "favorable" cases described by Proposition 1(iii), the optimal test is of

the form

$$\phi(\mathbf{x}) = 1 \quad \text{if } f_{m+1}(\mathbf{x}) > k\left(\sum_{i=1}^{m} \frac{k_i}{k} f_i(\mathbf{x})\right),$$

where $k := \sum_{i=1}^{m} k_i$ is the sum of the nonnegative coefficients k_i's. Since $k_i/k \geq 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^{m}(k_i/k) = 1$, this test is a Neyman test (in the sense of the fundamental lemma) for a mixture density of the form

$$f_0 := \sum_{i=1}^{m} \frac{k_i}{k} f_i$$

against $\{f_{m+1}\}$. This remark is exploited in the next section.

5.5 Least favorable distributions

5.5.1 Mixtures

The generalized Neyman-Pearson Lemma tells us that, *in the "favorable cases"*, most powerful tests of a composite null hypothesis $H_0 = \{f_1, \ldots, f_m\}$ against $H_1 = \{f_{m+1}\}$, under a level condition $\mathrm{E}_{f_i}[\phi] \leq \alpha$ for $i = 1, \ldots, m$, exist and are of the form

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{m+1}(\mathbf{x}) > \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \\ \gamma & \text{if } f_{m+1}(\mathbf{x}) = \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \\ 0 & \text{if } f_{m+1}(\mathbf{x}) < \sum_{i=1}^{m} k_i f_i(\mathbf{x}), \end{cases}$$

with $k_i \geq 0$ for $i = 1, \ldots, m$ and γ determined by $\mathrm{E}_{f_i}[\phi] \leq \alpha$ for $i = 1, \ldots, m$. By "favorable cases", we mean that such a test exists. Letting $k := \sum_{i=1}^{n} k_i$, this test rewrites as

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } f_{m+1}(\mathbf{x}) > k\left(\sum_{i=1}^{m} \frac{k_i}{k} f_i(\mathbf{x})\right) \\ \gamma & \text{if } f_{m+1}(\mathbf{x}) = k\left(\sum_{i=1}^{m} \frac{k_i}{k} f_i(\mathbf{x})\right) \\ 0 & \text{if } f_{m+1}(\mathbf{x}) < k\left(\sum_{i=1}^{m} \frac{k_i}{k} f_i(\mathbf{x})\right), \end{cases}$$

which is *the* Neyman-Pearson test for the simple hypothesis $\{\sum_{i=1}^{m} \frac{k_i}{k} f_i\}$ against the simple alternative $\{f_{m+1}\}$ under α-level constraint. This density $\sum_{i=1}^{m} \frac{k_i}{k} f_i$ (one easily checks that it is a density) is a *mixture* of $f_1, \ldots, f_m$, with *mixing probabilities* $\frac{k_1}{k}, \ldots, \frac{k_m}{k}$. When testing for a composite null hypothesis H_0, this suggests looking at mixtures of the densities in H_0.

5.5.2 Least Favorable Mixtures

Consider the problem of testing $H_0 = \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ against $H_1 = \{g\}$, where $f_{\boldsymbol{\theta}}$ and g are densities, with respect to some σ-finite measure μ, over $(\mathcal{X}, \mathcal{A})$, and $\Theta \subseteq \mathbb{R}^k$ is equipped with the Borel σ-field $\mathcal{B}^k \cap \Theta$. Let λ denote a probability measure over $(\Theta, \mathcal{B}^k \cap \Theta)$. Then, $h_\lambda : \mathbf{x} \mapsto h_\lambda(\mathbf{x}) := \int_\Theta f_{\boldsymbol{\theta}}(\mathbf{x}) \mathrm{d}\lambda(\mathbf{x})$ is still a probability density with respect to μ over $(\mathcal{X}, \mathcal{A})$—a *mixture* of the densities $f_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$.

For any λ, denote by $H_\lambda := \{h_\lambda\}$ the simple hypothesis under which the observation has density h_λ. Consider the Neyman-Pearson α-level test ϕ_λ of H_λ against H_1, and write $\pi_\lambda := \mathrm{E}_g[\phi_\lambda]$ for its power under H_1. We adopt the following definition.

Definition 1. *The mixing measure λ_{LF}, or the corresponding mixture density $h_{\lambda_{\mathrm{LF}}}$, are called* least favorable *if $\pi_{\lambda_{\mathrm{LF}}} \leq \pi_\lambda$ for any probability measure λ over Θ.*

We then have the following result.

Proposition 3. *Let λ_0 be such that $\mathrm{E}_{\boldsymbol{\theta}}[\phi_{\lambda_0}] \leq \alpha$ for all $\boldsymbol{\theta} \in \Theta$. Then, (i) the test ϕ_{λ_0} is* most powerful*, at level α, for H_0 against H_1; (ii) the density h_{λ_0} is* least favorable.

Proof. (i) By assumption, ϕ_{λ_0} is an α-level for H_0 against H_1. Let then ϕ be an arbitrary α-level test for the same problem, that is, $\mathrm{E}_{\boldsymbol{\theta}}[\phi] \leq \alpha$ for all $\boldsymbol{\theta} \in \Theta$. Then, Fubini's Theorem yields that

$$\mathrm{E}_{h_{\lambda_0}}[\phi] = \int_{\mathcal{X}} \phi(\mathbf{x}) h_{\lambda_0}(\mathbf{x}) d\mu(\mathbf{x}) = \int_{\mathcal{X}} \phi(\mathbf{x}) \Big(\int_\Theta f_{\boldsymbol{\theta}}(\mathbf{x}) d\lambda_0(\boldsymbol{\theta}) \Big) d\mu(\mathbf{x}) \tag{5.8}$$

$$= \int_{\mathcal{X}} \int_\Theta \phi(\mathbf{x}) f_{\boldsymbol{\theta}}(\mathbf{x}) d\lambda_0(\boldsymbol{\theta}) d\mu(\mathbf{x}) = \int_\Theta \Big(\int_{\mathcal{X}} \phi(\mathbf{x}) f_{\boldsymbol{\theta}}(\mathbf{x}) d\mu(\mathbf{x}) \Big) d\lambda_0(\boldsymbol{\theta}) = \int_\Theta \mathrm{E}_{\boldsymbol{\theta}}[\phi] d\lambda_0(\boldsymbol{\theta}) \leq \alpha,$$

so that ϕ is an α-level test for $H_{\lambda_0} = \{h_{\lambda_0}\}$ against $H_1 = \{g\}$. Since ϕ_{λ_0} is the most powerful test at level α for the latter problem, we must then have $\mathrm{E}_g[\phi_{\lambda_0}] \geq \mathrm{E}_g[\phi]$.

(ii) Fix an arbitrary mixture distribution λ. Proceeding as in (5.8), we have

$$\mathrm{E}_{h_\lambda}[\phi_{\lambda_0}] = \int_{\mathcal{X}} \phi_{\lambda_0}(\mathbf{x}) h_\lambda(\mathbf{x}) d\mu(\mathbf{x}) = \int_{\mathcal{X}} \phi_{\lambda_0}(\mathbf{x}) \left(\int_\Theta f_{\boldsymbol{\theta}}(\mathbf{x}) d\lambda(\boldsymbol{\theta}) \right) d\mu(\mathbf{x})$$

$$= \int_{\mathcal{X}} \int_\Theta \phi_{\lambda_0}(\mathbf{x}) f_{\boldsymbol{\theta}}(\mathbf{x}) d\lambda(\boldsymbol{\theta}) d\mu(\mathbf{x}) = \int_\Theta \left(\int_{\mathcal{X}} \phi_{\lambda_0}(\mathbf{x}) f_{\boldsymbol{\theta}}(\mathbf{x}) d\mu(\mathbf{x}) \right) d\lambda(\boldsymbol{\theta}) = \int_\Theta \mathrm{E}_{\boldsymbol{\theta}}[\phi_{\lambda_0}] d\lambda(\boldsymbol{\theta}) \leq \alpha.$$

Thus, ϕ_{λ_0} satisfies the level constraint under H_λ and, therefore, is at most as powerful as ϕ_λ: $\mathrm{E}_g[\phi_{\lambda_0}] \leq \mathrm{E}_g[\phi_\lambda]$. This shows that $\pi_{\lambda_0} \leq \pi_\lambda$ for any λ, so that h_{λ_0} is least favorable. □

5.5.3 Application 1: one-sided tests in one-parameter exponential families

In the one-parameter family of exponential densities

$$f_\theta(\mathbf{x}) = C(\theta) \exp(\theta T(\mathbf{x}))$$

(densities are with respect to some dominating measure μ, and $\theta \in \Theta$, where Θ is an interval of $\mathbb{R}$), consider the one-sided testing problem $H_0 : \theta \leq \theta_0$ versus the alternative $H_1 : \theta > \theta_0$. Uniformly most powerful tests at level α have been obtained for that problem in Section 5.3. As we now show, they also follow from a least-favorable approach.

Proposition 4. *The test*

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > t_\alpha \\ \gamma_\alpha & \text{if } T(\mathbf{x}) = t_\alpha \\ 0 & \text{if } T(\mathbf{x}) < t_\alpha, \end{cases}$$

where γ_α and t_α are determined by the size condition $\mathrm{E}_{\theta_0}[\phi^*] = \alpha$, is uniformly most powerful, at level α, for $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$.

Proof. Fix $\theta_1 > \theta_0$ arbitrarily and consider the problem of testing $\mathcal{H}_0 : \theta \leq \theta_0$ against $H_1 : \theta = \theta_1$. We show that, for this problem, f_{θ_0} (the degenerate mixture associated with $\lambda(\{\theta_0\}) = 1$) is least favorable. The Neyman-Pearson test for $\{f_{\theta_0}\}$ against $\{f_{\theta_1}\}$ has the form

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } C(\theta_1)\exp(\theta_1 T(\mathbf{x})) > k_\alpha C(\theta_0)\exp(\theta_0 T(\mathbf{x})) \\ \gamma_\alpha & \text{if } C(\theta_1)\exp(\theta_1 T(\mathbf{x})) = k_\alpha C(\theta_0)\exp(\theta_0 T(\mathbf{x})) \\ 0 & \text{if } C(\theta_1)\exp(\theta_1 T(\mathbf{x})) < k_\alpha C(\theta_0)\exp(\theta_0 T(\mathbf{x})), \end{cases}$$

with k_α and γ_α determined by $\mathrm{E}_{\theta_0}[\phi] = \alpha$. Equivalently,

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \exp((\theta_1 - \theta_0)T(\mathbf{x})) > k_\alpha C(\theta_0)/C(\theta_1) \\ \gamma_\alpha & \text{if } \exp((\theta_1 - \theta_0)T(\mathbf{x})) = k_\alpha C(\theta_0)/C(\theta_1) \\ 0 & \text{if } \exp((\theta_1 - \theta_0)T(\mathbf{x})) < k_\alpha C(\theta_0)/C(\theta_1), \end{cases}$$

or again

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) > t_\alpha := (\theta_1 - \theta_0)^{-1}\log(k_\alpha C(\theta_0)/C(\theta_1)) \\ \gamma_\alpha & \text{if } T(\mathbf{x}) = t_\alpha := (\theta_1 - \theta_0)^{-1}\log(k_\alpha C(\theta_0)/C(\theta_1)) \\ 0 & \text{if } T(\mathbf{x}) < t_\alpha := (\theta_1 - \theta_0)^{-1}\log(k_\alpha C(\theta_0)/C(\theta_1)), \end{cases}$$

where t_α and γ_α are determined by $\mathrm{E}_{\theta_0}[\phi] = \alpha$. Hence, ϕ coincides with the test ϕ^*_α from Theorem 1(i) (irrespective of θ_1).

Now, for any $\theta' < \theta_0$, note that ϕ^*_α is also the Neyman-Pearson Lemma test for $\{\mathrm{P}_{\theta'}\}$ against $\{\mathrm{P}_{\theta_0}\}$ at level $\mathrm{E}_{\theta'}[\phi^*_\alpha]$, so that the Neyman-Pearson lemma implies that $\mathrm{E}_{\theta'}[\phi^*_\alpha] < \mathrm{E}_{\theta_0}[\phi^*_\alpha] = \alpha$. It follows from Proposition 3(ii) that the degenerate mixture at $\{\theta_0\}$ is indeed least favorable and from Proposition 3(i) that ϕ^*_α is uniformly most powerful, at level α, for $\mathcal{H}_0 : \theta \leq \theta_0$ against $H_1 : \theta = \theta_1$. Since $\theta_1 (> \theta_0)$ was fixed arbitrarily, we conclude that the same test is also uniformly most powerful, at level α, for $\mathcal{H}_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. $\square$

A particular case of an exponential family with one parameter is the Bernoulli family. Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{Bin}(1, p)$ random variables, with $p \in (0, 1)$, and consider the problem of testing

$$H_0 : p \leq p_0 \quad \text{against} \quad H_1 : p > p_0.$$

The Bernoulli family is an exponential family with natural parameter $\theta = \log(\frac{p}{1-p})$ and privileged statistic $\sum_{i=1}^n X_i$. Note that

$$\log\left(\frac{p}{1-p}\right) \leq \log\left(\frac{p_0}{1-p_0}\right) \text{ if and only if } p \leq p_0,$$

so that the testing problem considered is of the form

$$H_0 : \theta \leq \theta_0 \quad \text{against} \quad H_1 : \theta > \theta_0.$$

The test

$$\phi^*_\alpha(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i > t_\alpha \\ \gamma_\alpha & \text{if } \sum_{i=1}^n x_i = t_\alpha \\ 0 & \text{if } \sum_{i=1}^n x_i < t_\alpha, \end{cases}$$

with t_α and γ_α fixed by the condition $\mathrm{E}_{p_0}[\phi^*_\alpha] = \alpha$, is then uniformly most powerful at level α.

5.5.4 Application 2: the sign test

Let $\mathbf{X} = (X_1, \ldots, X_n)$ collect n independently and identically distributed observations of X, with density $f \in \mathcal{F}$, where $\mathcal{F}$ is the family of all probability densities with respect to the Lebesgue measure over $(\mathbb{R}, \mathcal{B})$. Consider, for given $p_0 \in (0, 1)$, the testing problem

$$\begin{cases} H_0 : x_{(p_0)} \geq x_0 \\ H_1 : x_{(p_0)} < x_0, \end{cases}$$

where x_0 is some given real number, and $x_{(p_0)}$ is the quantile of order p_0 of X. Writing $p := \mathrm{P}[X \leq x_0] = \int_{-\infty}^{x_0} f(x)\mathrm{d}\mu(x)$, the same testing problem takes the form

$$\begin{cases} H_0 : p \leq p_0 \\ H_1 : p > p_0. \end{cases}$$

Denote as $M = M(\mathbf{X}) := \#\{i = 1, \ldots, n : X_i \leq x_0\}$ the number of observations that are not larger than x_0: clearly, $M \sim \mathrm{Bin}(n, p)$. The test

$$\phi_{\text{sign}}(\mathbf{x}) := \begin{cases} 1 & \text{if } M(\mathbf{x}) > m_\alpha \\ \gamma_\alpha & \text{if } M(\mathbf{x}) = m_\alpha \\ 0 & \text{if } M(\mathbf{x}) < m_\alpha, \end{cases} \tag{5.9}$$

where m_α and γ_α are determined by $\mathrm{E}_{p_0}[\phi_{\text{sign}}] = \alpha$, is called a *sign test* (here, we write E_{p_0} for the expectation of an M-measurable variable under $M \sim \mathrm{Bin}(n, p_0)$).

Proposition 5. *The sign test (5.9) is* uniformly most powerful *for H_0 against H_1 at level α.*

Proof. Any $f \in \mathcal{F}$ can be characterized as a triple (p, f^+, f^-), where $p := \int_{-\infty}^{x_0} f(x)\mathrm{d}\mu(x) \in (0,1)$, $f^+(x) := f(x)\mathbb{I}[x > x_0]/(1-p)$ is the conditional density of $X \sim f$, conditional on $X \geq x_0$, and $f^-(x) := f(x)\mathbb{I}[x \leq x_0]/p$ is the conditional density of $X \sim f$, conditional on $X < x_0$. The null hypothesis and the alternative then take the forms

$$H_0 = \{(p, f^-, f^+) : p \leq p_0,\ f^- \in \mathcal{F}^-,\ f^+ \in \mathcal{F}^+\}$$

and

$$H_1 = \{(p, f^-, f^+) : p > p_0,\ f^- \in \mathcal{F}^-,\ f^+ \in \mathcal{F}^+\},$$

respectively, where $\mathcal{F}^-$ (resp., $\mathcal{F}^+$) is the set of all possible conditional densities f^+ (resp., f^-). The joint density under (p, f^+, f^-) of $\mathbf{X} = (X_1, \ldots, X_n)$ at $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is

$$p^m(1-p)^{n-m} f^-(x_{i_1}) \ldots f^-(x_{i_m}) f^+(x_{j_1}) \ldots f^+(x_{j_{n-m}}),$$

where $m = M(\mathbf{x}) = \#\{i = 1, \ldots, n : x_i \leq x_0\}$ and $i_1, \ldots, i_m, j_1, \ldots, j_{n-m}$ are such that

$$x_{i_1}, \ldots, x_{i_m} \leq x_0 < x_{j_1}, \ldots, x_{j_{n-m}}.$$

Now, fix a distribution (p_1, f_1^-, f_1^+) in H_1 (hence, $p_1 > p_0$): intuitively, the "closest" distribution in H_0 could well be (p_0, f_1^-, f_1^+). Let us show that indeed (p_0, f_1^-, f_1^+) is the least favorable mixture (against the fixed alternative $\{(p_1, f_1^-, f_1^+)\}$). To this end, we first construct the Neyman-Pearson test for $\{(p_0, f_1^-, f_1^+)\}$ against $\{(p_1, f_1^-, f_1^+)\}$. Recalling that $m = M(\mathbf{x})$, this test is

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } \left(\frac{p_1}{p_0}\right)^m \left(\frac{1-p_1}{1-p_0}\right)^{n-m} > k_\alpha \\ \gamma_\alpha & \text{if } \left(\frac{p_1}{p_0}\right)^m \left(\frac{1-p_1}{1-p_0}\right)^{n-m} = k_\alpha \\ 0 & \text{if } \left(\frac{p_1}{p_0}\right)^m \left(\frac{1-p_1}{1-p_0}\right)^{n-m} < k_\alpha, \end{cases}$$

where k_α and γ_α are determined by $\mathrm{E}_{(p_0,f_1^-,f_1^+)}[\phi^*] = \alpha$ (note that this expectation does not depend on f_1^- nor on f_1^+ since the distribution of M under (p_0, f_1^-, f_1^+) is the $\mathrm{Bin}(n, p_0)$ distribution). Clearly, since $p_1 > p_0$, this test takes the simpler form

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{if } M(\mathbf{x}) > m_\alpha \\ \gamma_\alpha & \text{if } M(\mathbf{x}) = m_\alpha \\ 0 & \text{if } M(\mathbf{x}) < m_\alpha, \end{cases}$$

where m_α and γ_α still are determined by $\mathrm{E}_{(p_0,f_1^-,f_1^+)}[\phi^*] = \alpha$. This test ϕ^* thus coincides with the previously described sign test ϕ_{sign}. Now, in order to use the least favorable argument, it only remains to show that $\mathrm{E}_{(p,f^-,f^+)}[\phi^*] \leq \alpha$ for any $p \leq p_0$, $f^- \in \mathcal{F}^-$ and $f^+ \in \mathcal{F}^+$. But this follows from the fact that $\mathrm{E}_{(p,f^-,f^+)}[\phi^*]$ does not depend on (f^-, f^+) and is increasing in p—one way to show this is to note that ϕ^* is actually the uniformly most powerful test for $\{p \leq p_0\}$ against $\{p > p_0\}$ in the (exponential) Bernoulli model under which $\mathbb{I}[X_1 \leq x_0], \ldots, \mathbb{I}[X_n \leq x_0]$ are independently and identically distributed $\mathrm{Bin}(1, p)$ random variables (see Section 5.5.3). □