Testing Statistical Hypotheses II

Statistical Theory

Guillaume Dehaene
Ecole Polytechnique Fédérale de Lausanne

1. Uniformly Most Powerful Tests
2. Situations When UMP Exist
3. Locally Most Powerful Tests
4. Likelihood Ratio Tests

Uniformly Most Powerful Tests

Neyman-Pearson Framework for Testing Hypotheses

The Problem of Hypothesis Testing

- $\boldsymbol{X} = (X_1, ..., X_n)$ random variables with joint density/frequency $f(\boldsymbol{x}; \theta)$
- $\theta \in \Theta$ where $\Theta = \Theta_0 \cup \Theta_1$ and $\Theta_0 \cap \Theta_1 = \emptyset$
- Observe realization $\boldsymbol{x} = (x_1, ..., x_n)$ of $\boldsymbol{X} \sim f_\theta$
- Decide on the basis of $\boldsymbol{x}$ whether $\theta \in \Theta_0$ (H_0) or $\theta \in \Theta_1$ (H_1)

Neyman-Pearson Framework:

1. Fix a significance level α for the test
2. Among all rules respecting the significance level, pick the one that uniformly maximizes power

When H_0/H_1 both simple$\rightarrow$ Neyman-Pearson lemma settles the problem.

$\hookrightarrow$ What about more general structure of Θ_0, Θ_1?

Uniformly Most Powerful Tests

A *uniformly most powerful test* of $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ at level α:

1. Respects the level for all $\theta \in \Theta_0$, i.e.

$$\delta \in \mathscr{D}(\Theta_0, \alpha) = \{\delta : \mathcal{X} \to \{0,1\} : \mathbb{E}_\theta[\delta] \leq \alpha, \ \forall\, \theta \in \Theta_0\}$$

2. Is most powerful for all $\theta \in \Theta_1$ (i.e. for all possible simple alternatives),

$$\mathbb{E}_\theta[\delta] \geq \mathbb{E}_\theta[\delta'] \qquad \forall \theta \in \Theta_1 \quad \& \quad \delta' \in \mathscr{D}(\Theta_0, \alpha)$$

Unfortunately UMP tests rarely exist. Why?

$\hookrightarrow$ Consider $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

- A UMP test must be MP test for any $\theta \neq \theta_1$.
- But the form of the MP test typically differs for $\theta_1 > \theta_0$ and $\theta_1 < \theta_0$!
  - $\hookrightarrow$ e.g. recall exponential mean example

Example (No UMP test exists)

Let $X \sim$Binom(n, θ) and suppose we want to test:

$$H_0 : \theta = \theta_0 \qquad vs \qquad H_1 : \theta \neq \theta_0$$

at some level α. To this aim, consider first

$$H_0' : \theta = \theta_0 \qquad vs \qquad H_1' : \theta = \theta_1$$

Neyman-Pearson lemma gives test statistics

$$T = \frac{f(X;\theta_1)}{f(X;\theta_0)} = \left(\frac{1-\theta_0}{1-\theta_1}\right)^n \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^X$$

- If $\theta_1 > \theta_0$ then T increasing in X
  $\hookrightarrow$ MP test would reject for large values of X
- If $\theta_1 < \theta_0$ then T decreasing in X
  $\hookrightarrow$ MP test would reject for small values of X

Example (A UMP test exists)

Let $X_1, ..., X_n \stackrel{iid}{\sim} \text{Exp}(\lambda)$ and suppose we wish to test

$$H_0 : \lambda \leq \lambda_0 \qquad vs \qquad H_1 : \lambda > \lambda_0$$

at some level α. To this aim, consider first the pair

$$H_0' : \lambda = \lambda_0 \qquad vs \qquad H_1' : \lambda = \lambda_1$$

with $\lambda_1 > \lambda_0$ which we saw last time to admit a MP test $\forall\ \lambda_1 > \lambda_0$:

Reject H_0 for $\displaystyle\sum_{i=1}^n X_i \leq k$, with k such that $\mathbb{P}_{\lambda_0}\left[\displaystyle\sum_{i=1}^n X_i \leq k\right] = \alpha$

But for $\lambda < \lambda_0$, $\mathbb{P}_{\lambda_0}\left[\sum_{i=1}^n X_i \leq k\right] = \alpha \implies \mathbb{P}_{\lambda}\left[\sum_{i=1}^n X_i \leq k\right] < \alpha$.

So the same test respects level α for all singletons under the the null.
$\hookrightarrow$ The test is UMP of H_0 vs H_1

Situations When UMP Exist

When do UMP tests exist?

Examples: insight on which composite pairs typically admit UMP tests:

1. Hypothesis pair concerns a single real-valued parameter
2. Hypothesis pair is “one-sided”

But existence of UMP test does not only depend on hypothesis structure...
$\hookrightarrow$ Also depends on specific model. Sufficient condition?

Definition (Monotone Likelihood Ratio Property)

A family of density (frequency) functions $\{f(\boldsymbol{x};\theta) : \theta \in \Theta\}$ with $\Theta \subseteq \mathbb{R}$ is said to have monotone likelihood ratio if there exists a real-valued function $T(\boldsymbol{x})$ such that, for any $\theta_0 < \theta_1$, the function

$$\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)}$$

is a non-decreasing function of $T(\boldsymbol{x})$ for x such that $\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} \in (0,\infty)$.

Such a statistic T will necessarily be sufficient for θ. (Fisher-Neyman)

MLR example

Example

Let $X_i \overset{IID}{\sim} \mathcal{B}(\theta)$. $S = \sum X_i \sim \mathcal{B}(n, \theta)$ has mass function:

$$\begin{aligned} f_S(s, \theta) &= \binom{n}{s} \theta^s (1-\theta)^{n-s} \\ &= \binom{n}{s} (1-\theta)^n \left(\frac{\theta}{1-\theta} \right)^s \end{aligned}$$

The likelihood ratio is:

$$\frac{f_S(s, \theta_1)}{f_S(s, \theta_0)} = \left(\frac{1-\theta_1}{1-\theta_0} \right)^n \left(\frac{\frac{\theta_1}{1-\theta_1}}{\frac{\theta_0}{1-\theta_0}} \right)^s$$

Intuition: increasing T shifts the likelihood to the right.

When do UMP tests exist?

Theorem (MLR and UMP)

Let $\boldsymbol{X} = (X_1, ..., X_n)$ have joint distribution of monotone likelihood ratio with respect to a statistic T, depending on $\theta \in \mathbb{R}$. Further assume that T is a continuous random variable. Then, the test function given by

$$\delta(\boldsymbol{X}) = \begin{cases} 1 & \text{if } T(\boldsymbol{X}) \geq k \\ 0 & \text{if } T(\boldsymbol{X}) < k \end{cases} \qquad k \text{ such that } \mathbb{E}_{\theta_0}[\delta(\boldsymbol{X})] = \alpha$$

is UMP among all tests at level α for the hypothesis pair

$$\begin{cases} H_0 : & \theta \leq \theta_0 \\ H_1 : & \theta > \theta_1 \end{cases}$$

[The assumption of continuity of the random variable T can be removed, by considering randomized tests as well, similarly as before]

Proof.

We need to show that:

1. This procedure defines a rule with level $\alpha : \delta \in \mathscr{D}(\Theta_0, \alpha)$, i.e. $\mathbb{E}_\theta[\delta] \leq \alpha \, (= \mathbb{E}_{\theta_0}[\delta])$ for all $\theta \in \Theta_0 = (-\infty, \theta_0]$.
2. It is a most powerful rule: for any $\delta' \in \mathscr{D}(\Theta_0, \alpha)$ and all $\theta_1 \in \Theta_1$, $\mathbb{E}_{\theta_1}[\delta'] \leq \mathbb{E}_{\theta_1}[\delta]$.

To show (1) it suffices to show that $\mathbb{E}_{\theta_0}[\delta] - \mathbb{E}_\theta[\delta] \geq 0$ for $\theta \leq \theta_0$. It is easiest to write down these expected values as functions of t.

$$
\begin{aligned}
\mathbb{E}_\theta(\delta) &= \int \mathbf{1}(t \geq k) f(t;\theta) dt \\
&= \int \mathbf{1}(t \geq k) \frac{f(t;\theta)}{f(t;\theta_0)} f(t;\theta_0) dt \\
&= \int \mathbf{1}(t \geq k) \frac{d(t)}{Z} f(t;\theta_0) dt \\
&= \frac{\int \mathbf{1}(t \geq k) d(t) f(t;\theta_0) dt}{\int d(t) f(t;\theta_0) dt}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_\theta(\delta) &= \frac{\int \mathbf{1}(t \geq k) d(t) f(t;\theta_0) dt}{\int d(t) f(t;\theta_0) dt} \\
&= \frac{1}{r(\theta)+1} \\
r(\theta) &= \frac{\int_k^\infty d(t) f(t;\theta_0) dt}{\int_{-\infty}^k d(t) f(t;\theta_0) dt} \\
&= \frac{\int_k^\infty \frac{d(t)}{d(k)} f(t;\theta_0) dt}{\int_{-\infty}^k \frac{d(t)}{d(k)} f(t;\theta_0) dt} \\
&\geq \frac{\int_k^\infty f(t;\theta_0) dt}{\int_{-\infty}^k f(t;\theta_0) dt} \\
r(\theta) &\geq r(\theta_0)
\end{aligned}
$$

which finally implies:

$$\mathbb{E}_\theta(\delta) \leq \mathbb{E}_{\theta_0}(\delta) = \alpha$$

For (2), note that $\delta \in \mathscr{D}(\{\theta_0\}, \alpha)$. Thus, we can prove that δ is most-powerful by proving that it is most powerful for every pair θ_0 and $\theta_1 \in \Theta_1$.

Since $\theta_0 < \theta_1$, we have $f(X;\theta_1)/f(X;\theta_0) = h(T)$ for some non-decreasing h, by the MLR property of T. Let $K = h(k)$ and let

$$I_k = [k-a, k+b], \qquad a, b > 0,$$

the interval on which $h(t) = K$ (this set is an interval, since h is non-increasing; it could also be half open, or open).
Now consider the MP test of θ_0 vs θ_1.

$$\psi(\boldsymbol{X}) = \begin{cases} 1, & \text{if } f(X;\theta_1) > Kf(X;\theta_0), \\ \mathbb{P}[k \leq T < k+b]/\mathbb{P}[T \in I_k], & \text{if } f(X;\theta_1) = Kf(X;\theta_0) \\ 0, & \text{if } f(X;\theta_1) < Kf(X;\theta_0) \end{cases}$$

Now we note that (recall that T is cts RV, so strict inequalities irrelevant)

$$\begin{aligned} \mathbb{E}_\theta[\psi] &= 0 \times \mathbb{P}_\theta[T < k - a] \\ &\quad + \frac{\mathbb{P}_\theta[k \le T < k + b]}{\mathbb{P}_\theta[T \in I_k]} \mathbb{P}_\theta[T \in I_k] \\ &\quad + 1 \times \mathbb{P}_\theta[T \ge k + b] \\ &= \mathbb{P}_\theta[T \ge k] \\ &= \mathbb{E}_\theta[\delta]. \end{aligned}$$

We have proved that δ is equivalent to the most-powerful ψ. Thus:

- For $\theta = \theta_0$, $\mathbb{E}_{\theta_0}[\psi] = \mathbb{E}_{\theta_0}[\delta]$. Therefore, it follows from the generalised NP-lemma that ψ is most powerful at level $\mathbb{E}_{\theta_0}[\delta]$. In other words, $\mathbb{E}_{\theta_1}[\delta'] \le \mathbb{E}_{\theta_1}[\psi]$ for all $\delta' \in \mathscr{D}(\{\theta_0\}, \alpha)$.
- On the other hand, for $\theta = \theta_1$, we have $\mathbb{E}_{\theta_1}[\psi] = \mathbb{E}_{\theta_1}[\delta]$.

We conclude that $\mathbb{E}_{\theta_1}[\delta'] \le \mathbb{E}_{\theta_1}[\delta]$ for all $\delta' \in \mathscr{D}(\{\theta_0\}, \alpha)$ and the proof is complete.

- T yielding monotone likelihood ratio necessarily a sufficient statistic

Example (One-Parameter Exponential Family)

Let $\boldsymbol{X} = (X_1, ..., X_n)$ have a joint density (frequency)

$$f(\boldsymbol{x};\theta) = \exp[c(\theta)T(\boldsymbol{x}) - b(\theta) + S(\boldsymbol{x})]$$

and assume WLOG that $c(\theta)$ is strictly increasing. For $\theta_0 < \theta_1$,

$$\frac{f(\boldsymbol{x};\theta_1)}{f(\boldsymbol{x};\theta_0)} = \exp\{[c(\theta_1) - c(\theta_0)]T(\boldsymbol{x}) + b(\theta_0) - b(\theta_1)\}$$

is increasing in T by monotonicity of $c(\cdot)$.

Hence the UMP test of $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ would reject iff $T(\boldsymbol{x}) \geq k$, with $\alpha = \mathbb{P}_{\theta_0}[T \geq k]$.

Locally Most Powerful Tests

Locally Most Powerful Tests

$\hookrightarrow$ What if MLR property fails to be satisfied? Can optimality be "saved"?

- Consider $\theta \in \mathbb{R}$ and wish to test: $H_1 : \theta \leq \theta_0$ vs $H_0 : \theta > \theta_0$
- Intuition: if true θ far from θ_0 any reasonable test powerful
- ⋆ So focus on maximizing power in small neighbourhood of θ_0

$\rightarrow$ Consider power function $\beta(\theta) = \mathbb{E}_\theta[\delta(\boldsymbol{X})]$ of some δ.
$\rightarrow$ Require $\beta(\theta_0) = \alpha$ (notice that $\theta_0 \in \Theta_0$ so $\beta(\theta_0)$ is type I at θ_0)
$\rightarrow$ Assume that $\beta(\theta)$ is differentiable, so for θ close to θ_0

$$\beta(\theta) \approx \beta(\theta_0) + \beta'(\theta_0)(\theta - \theta_0) = \alpha + \beta'(\theta_0)\underbrace{(\theta - \theta_0)}_{>0}$$

Since $\Theta_1 = (\theta_0, \infty)$, this suggests approach for locally most powerful test

Choose δ to Maximize $\beta'(\theta_0)$ Subject to $\beta(\theta_0) = \alpha$

$\boxed{\text{How do we solve this constrained optimization problem?}}$

Supposing that $\boldsymbol{X} = (X_1, \ldots X_n)$ has density $f(\boldsymbol{x}; \theta)$, then

$$
\begin{aligned}
& \beta(\theta) = \int_{\mathbb{R}^n} \delta(\boldsymbol{x}) f(\boldsymbol{x}; \theta) d\boldsymbol{x} \\
\Longrightarrow \quad & \frac{\partial}{\partial \theta} \beta(\theta) = \int_{\mathbb{R}^n} \delta(\boldsymbol{x}) \frac{\partial}{\partial \theta} f(\boldsymbol{x}; \theta) d\boldsymbol{x} \quad \text{[provided interchange possible]} \\
= \quad & \int_{\mathbb{R}^n} \delta(\boldsymbol{x}) \frac{f(\boldsymbol{x}; \theta)}{f(\boldsymbol{x}; \theta)} \frac{\partial}{\partial \theta} f(\boldsymbol{x}; \theta) d\boldsymbol{x} \\
= \quad & \int_{\mathbb{R}^n} \delta(\boldsymbol{x}) \left[\frac{\partial}{\partial \theta} \log f(\boldsymbol{x}; \theta) \right] f(\boldsymbol{x}; \theta) d\boldsymbol{x} \\
= \quad & \mathbb{E}_\theta \left[\delta(\boldsymbol{X}) \underbrace{\frac{\partial}{\partial \theta} \log f(\boldsymbol{x}; \theta)}_{S(\boldsymbol{X}; \theta)} \right] = \mathrm{Cov}(\delta, S(X, \theta))
\end{aligned}
$$

The last equality follows if we can differentiate under the integral, in which case $\mathbb{E}[S(X; \theta)] = 0$. So δ must be a "linear functional" of $S(X; \theta)$!

Locally Most Powerful Tests

Theorem (Score Tests are Locally Most Powerful)

Let $\boldsymbol{X} = (X_1, ..., X_n)$ have joint density (frequency) $f(\boldsymbol{x}; \theta)$ and define the test function

$$\delta(\boldsymbol{X}) = \begin{cases} 1 & \text{if } S(\boldsymbol{X}; \theta_0) \geq k, \\ 0 & \text{otherwise} \end{cases}$$

where k is such that $\mathbb{E}_{\theta_0}[\delta(\boldsymbol{X})] = \alpha$. Then δ maximizes

$$\mathbb{E}_{\theta_0}\left[\psi(\boldsymbol{X}) S(\boldsymbol{X}; \theta_0)\right]$$

over all test functions ψ satisfying the constraint $\mathbb{E}_{\theta_0}[\psi(\boldsymbol{X})] = \alpha$.

- Gives recipe for constructing LMP test
- We were concerned about power *only locally around* θ_0
- **BEWARE !** May not even give a level α test for some $\theta < \theta_0$

Proof.

Consider ψ with $\psi(\boldsymbol{x}) \in \{0,1\}\ \forall\ \boldsymbol{x}$ and $\mathbb{E}_{\theta_0}[\psi(\boldsymbol{X})] = \alpha$. Then,

$$\delta(\boldsymbol{x}) - \psi(\boldsymbol{x}) = \begin{cases} \geq 0 & \text{if } S(\boldsymbol{x};\theta_0) \geq k, \\ \leq 0 & \text{if } S(\boldsymbol{x};\theta_0) \leq k \end{cases}$$

Therefore

$$\mathbb{E}_{\theta_0}[(\delta(\boldsymbol{X}) - \psi(\boldsymbol{X}))(S(\boldsymbol{X};\theta_0) - k)] \geq 0$$

expanding the product and since $\mathbb{E}_{\theta_0}[\delta(\boldsymbol{X}) - \psi(\boldsymbol{X})] = 0$ it must be that

$$\mathbb{E}_{\theta_0}\left[\delta(\boldsymbol{X})S(\boldsymbol{X};\theta_0)\right] \geq \mathbb{E}_{\theta_0}\left[\psi(\boldsymbol{X})S(\boldsymbol{X};\theta_0)\right]$$

$\square$

How is the critical value k evaluated in practice? (obviously to give level α)

- When $\{X_i\}$ are iid then $S(\boldsymbol{X};\theta) = \sum_{i=1}^{n} \ell'(X_i;\theta)$
- Under regularity conditions: sum of iid rv's mean zero variance $I(\theta)$
- So, for $\theta = \theta_0$ and large n, $S(\boldsymbol{X};\theta) \stackrel{d}{\approx} \mathcal{N}(0, nI(\theta))$

Example (Cauchy distribution)

Let $X_1, ..., X_n \overset{iid}{\sim}$Cauchy($\theta$), with density,

$$f(x;\theta) = \frac{1}{\pi(1+(x-\theta)^2)}$$

and consider the hypothesis pair $\begin{cases} H_0: & \theta \geq 0 \\ H_1: & \theta \leq 0 \end{cases}$

We have

$$S(\boldsymbol{X};0) = \sum_{i=1}^{n} \frac{2X_i}{1+X_i^2}$$

so that the LMP test at level α rejects the null if $S(\boldsymbol{X};0) \leq k$, where

$$\mathbb{P}_0[S(\boldsymbol{X};0) \leq k] = \alpha$$

While the exact distribution is difficult to obtain, for large n, $S(\boldsymbol{X};0) \overset{d}{\approx} \mathcal{N}(0, n/2)$.

Likelihood Ratio Tests

Likelihood Ratio Tests

We have seen tests for simple vs simple and one sided vs one sided
$\hookrightarrow$ Extension to multiparameter case $\boldsymbol{\theta} \in \mathbb{R}^p$? General Θ_0, Θ_1?

- Unfortunately, optimality theory breaks down in higher dimensions
- General method for constructive *reasonable* tests?

$\rightarrow$ The idea: Combine Neyman-Pearson paradigm with Max Likelihood

Definition (Likelihood Ratio)

The *likelihood ratio statistic* corresponding to the pair of hypotheses $H_0 : \boldsymbol{\theta} \in \Theta_0$ vs $H_1 : \boldsymbol{\theta} \in \Theta_1$ is defined to be

$$\Lambda(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_1} f(\boldsymbol{X}; \theta)}{\sup_{\boldsymbol{\theta} \in \Theta_0} f(\boldsymbol{X}; \theta)} = \frac{\sup_{\boldsymbol{\theta} \in \Theta_1} L(\theta)}{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\theta)}$$

- "*Neyman-Pearson*"-esque approach: reject H_0 for large Λ.
- Intuition: choose the "most favourable" $\theta \in \Theta_0$ (in favour of H_0) and compare it against the "most favourable" $\theta \in \Theta_1$ (in favour of H_1) in a simple vs simple setting (applying NP-lemma)

Example

Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Consider:

$$H_0 : \mu = \mu_0 \qquad \text{vs} \qquad H_1 : \mu \neq \mu_0$$

$$\Lambda(\boldsymbol{X}) = \frac{\sup_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}^+} f(\boldsymbol{X};\mu,\sigma^2)}{\sup_{(\mu,\sigma^2)\in\{\mu_0\}\times\mathbb{R}^+} f(\boldsymbol{X};\mu,\sigma^2)} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{\sum_{i=1}^n (X_i-\mu_0)^2}{\sum_{i=1}^n (X_i-\bar{X})^2}\right)^{\frac{n}{2}}$$

So reject when $\Lambda \geq k$, where k is s.t. $\mathbb{P}_0[\Lambda \geq k] = \alpha$. Distribution of Λ?
By monotonicity look only at

$$\begin{aligned}
\frac{\sum_{i=1}^n (X_i-\mu_0)^2}{\sum_{i=1}^n (X_i-\bar{X})^2} &= 1 + \frac{n(\bar{X}-\mu_0)^2}{\sum_{i=1}^n (X_i-\bar{X})^2} = 1 + \frac{1}{n-1}\left(\frac{n(\bar{X}-\mu_0)^2}{S^2}\right) \\
&= 1 + \frac{T^2}{n-1}
\end{aligned}$$

With $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X})^2$ and $T = \sqrt{n}(\bar{X}-\mu_0)/S \stackrel{H_0}{\sim} t_{n-1}$.
So $T^2 \stackrel{H_0}{\sim} F_{1,n-1}$ and k may be chosen appropriately.

Example

Let $X_1, ..., X_m \stackrel{iid}{\sim} \text{Exp}(\lambda)$ and $Y_1, ..., Y_n \stackrel{iid}{\sim} \text{Exp}(\theta)$ and $\boldsymbol{X}$ indep $\boldsymbol{Y}$.

$$\text{Consider:} \quad H_0 : \theta = \lambda \qquad \text{vs} \qquad H_1 : \theta \neq \lambda$$

$$\underset{\sup_{(\lambda,\theta)\in\mathbb{R}_+^2} f(\boldsymbol{X},\boldsymbol{Y};\lambda,\theta)}{\text{Unrestricted MLEs:}} \quad \hat{\lambda} = 1/\bar{X} \qquad \& \qquad \hat{\theta} = 1/\bar{Y}$$

$$\underset{\sup_{(\lambda,\theta)\in\{(x,y)\in\mathbb{R}_+^2 : x=y\}} f(\boldsymbol{X},\boldsymbol{Y};\lambda,\theta)}{\text{Restricted MLEs:}} \quad \hat{\lambda}_0 = \hat{\theta}_0 = \left[\frac{m\bar{X} + n\bar{Y}}{m+n}\right]^{-1}$$

$$\Longrightarrow \Lambda = \left(\frac{m}{m+n} + \frac{n}{n+m}\frac{\bar{Y}}{\bar{X}}\right)^m \left(\frac{n}{n+m} + \frac{m}{m+n}\frac{\bar{X}}{\bar{Y}}\right)^n$$

Depends on $T = \bar{X}/\bar{Y}$ and can make Λ large/small by varying T.

$\hookrightarrow$ But $T \stackrel{H_0}{\sim} F_{2m,2n}$ so given α we may find the critical value k.

Distribution of Likelihood Ratio?

More often than not, dist(Λ) intractable
$\hookrightarrow$(and no simple dependence on T with tractable distribution either)

Consider asymptotic approximations?

Setup

- Θ open subset of $\mathbb{R}^p$
- either $\Theta_0 = \{\boldsymbol{\theta}_0\}$ or Θ_0 open subset of $\mathbb{R}^s$, where $s < p$
- Concentrate on $\boldsymbol{X} = (X_1, ..., X_n)$ has iid components.
- Initially restrict attention to $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. LR becomes:

$$\Lambda_n(\boldsymbol{X}) = \prod_{i=1}^{n} \frac{f(X_i; \hat{\boldsymbol{\theta}}_n)}{f(X_i; \boldsymbol{\theta}_0)}$$

  where $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta}$.
- Impose regularity conditions from MLE asymptotics

Asymptotic Distribution of the Likelihood Ratio

Theorem (Wilks' Theorem, case $p = 1$)

Let $X_1, \ldots, X_n$ be iid random variables with density (frequency) depending on $\theta \in \mathbb{R}$ and satisfying conditions (A1)-(A6), with $I(\theta) = J(\theta)$. If the MLE sequence $\hat{\theta}_n$ is consistent for θ, then the likelihood ratio statistic Λ_n for $H_0 : \theta = \theta_0$ satisfies

$$2 \log \Lambda_n \xrightarrow{d} V \sim \chi_1^2$$

when H_0 is true.

- Obviously, knowing approximate distribution of $2 \log \Lambda_n$ is as good as knowing approximate distribution of Λ_n for the purposes of testing (by monotonicity and rejection method).
- Theorem extends immediately and trivially to the case of general p and for a hypothesis pair $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.
  (i.e. when null hypothesis is simple)

Proof.

Under the conditions of the theorem and when H_0 is true,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\theta))$$

Now take logarithms and expand in a Taylor series around $\hat{\theta}_n$,

$$\begin{aligned}
\log \Lambda_n &= \sum_{i=1}^{n} [\ell(X_i; \hat{\theta}_n) - \ell(X_i; \theta_0)] = \sum_{i=1}^{n} [\ell(X_i; \hat{\theta}_n) - \ell(X_i; \hat{\theta}_n)] + \\
&\quad + (\theta_0 - \hat{\theta}_n) \sum_{i=1}^{n} \ell'(X_i; \hat{\theta}_n) - \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 \sum_{i=1}^{n} \ell''(X_i; \theta_n^*) \\
&= -\frac{1}{2} n(\hat{\theta}_n - \theta_0)^2 \frac{1}{n} \sum_{i=1}^{n} \ell''(X_i; \theta_n^*)
\end{aligned}$$

where θ_n^* lies between $\hat{\theta}_n$ and θ_0.

Asymptotic Distribution of the Likelihood Ratio

If H_0 is true, and since $\hat{\theta}_n$ is a consistent sequence, θ_n^* is sandwiched so

$$\theta_n^* \xrightarrow{p} \theta_0.$$

Hence under assumptions (A1)-(A6), and when H_0 is true, a first order Taylor expansion about θ_0, the continuous mapping theorem and the LLN give

$$\frac{1}{n}\sum_{i=1}^{n} \ell''(X_i; \theta_n^*) \xrightarrow{p} -\mathbb{E}_{\theta_0}[\ell''(X_i; \theta_0)] = I(\theta_0)$$

On the other hand, by the continuous mapping theorem,

$$n(\hat{\theta}_n - \theta_0)^2 \xrightarrow{d} \frac{\chi_1^2}{I(\theta_0)}$$

Applying Slutsky's theorem now yields the result. $\square$

Asymptotic Distribution of the Likelihood Ratio

Theorem (Wilk's theorem, general p, general $s \leq p$)

Let $X_1, \ldots, X_n$ be iid random variables with density (frequency) depending on $\boldsymbol{\theta} \in \mathbb{R}^p$ and satisfying conditions (B1)-(B6), with $I(\boldsymbol{\theta}) = J(\boldsymbol{\theta})$. If the MLE sequence $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}$, then the likelihood ratio statistic Λ_n for $H_0 : \{\theta_j = \theta_{j,0}\}_{j=1}^{s}$ satisfies $2 \log \Lambda_n \xrightarrow{d} V \sim \chi^2_s$ when H_0 is true.

Exercise

Prove Wilks' theorem. Note that it may potentially be that $s < p$: some of the components of $\boldsymbol{\theta}$ might be adjustable under H_0 !

Hypotheses of the form $H_0 : \{g_j(\boldsymbol{\theta}) = a_j\}_{j=1}^{s}$, for g_j differentiable real functions, can also be handled by Wilks' theorem:

- Define $(\phi_1, \ldots, \phi_p) = g(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), \ldots, g_p(\boldsymbol{\theta}))$
- $g_{s+1}, \ldots, g_p$ defined so that $\boldsymbol{\theta} \mapsto g(\boldsymbol{\theta})$ is 1-1
- Apply theorem with parameter ϕ

Other Tests?

Many other tests possible once we "liberate" ourselves from strict optimality criteria. For example:

- Wald's test
  - $\hookrightarrow$ For a simple null, may compare the unrestricted MLE with the MLE under the null. Large deviations indicate evidence against null hypothesis. Distributions are approximated for large n via the asymptotic normality of MLEs.
- Score Test
  - $\hookrightarrow$ For a simple null, if the null hypothesis is false, then the loglikelihood gradient at the null should not be close to zero, at least when n reasonably large: so measure its deviations form zero. Use asymptotics for distributions (under conditions we end up with a χ^2)
- ...

Summary

In general, MP tests do not exist, because they would need to be MP for all pairs: $\theta_0 \in \Theta_0, \theta_1 \in \Theta_1$. However:

- If there is a monotone LR, one-sided vs one-sided has a MP.
- We can consider locally MP tests like the score test.
  However, these can be silly.

Thus, in high-dimensions and for testing: $\theta = \theta_0$ vs $\theta \neq \theta_0$, we need to give up on optimality.

We can extend the likelihood-ratio test to these two situations.
Wilks' theorem gives us the asymptotic sampling distribution of the likelihood-ratio under the null hypothesis.

Other tests can also be used.