Neyman–Pearson Theory

Statistics · Module 5 — Hypothesis Testing

Node 5.2

Position in Knowledge Graph

System: Statistics

Structural Domain: Hypothesis Testing

Node: 5.2

The Neyman–Pearson Lemma is the foundational optimality result in hypothesis testing. It characterizes the most powerful test for simple-vs-simple problems and establishes the likelihood ratio as the fundamental statistic for hypothesis testing. All optimal testing theory (UMP tests, LR tests) descends from this result. It requires the testing framework of Node 5.1 and the distributional theory (densities, likelihood) of Nodes 2.3–2.4.

Formal Definition

Setting

Let P_0 and P_1 be two distributions with densities p_0 and p_1 with respect to a common dominating measure μ. Consider the simple-vs-simple testing problem:

$$H_0 : P = P_0 \quad \text{vs} \quad H_1 : P = P_1.$$

Most Powerful Test

Among all level α tests, a test ϕ^* is **most powerful (MP)** if

$$\mathbb{E}_1[\phi^*(X)] \geq \mathbb{E}_1[\phi(X)]$$

for every test ϕ satisfying $\mathbb{E}_0[\phi(X)] \leq \alpha$.

Intuition

The Neyman–Pearson Lemma answers the question: given a budget α on false alarm probability, how should one allocate rejection probability across the sample space to maximize detection?

The answer is greedy allocation. Rank every point x by how much more likely it is under P_1 than under P_0, i.e., by the likelihood ratio $p_1(x)/p_0(x)$. Reject starting from the points with the highest ratio, spending the budget α on the points that contribute the most power per unit of size.

This is an isoperimetric principle: among all sets of prescribed P_0-measure α, the one with the largest P_1-measure is the upper level set of the likelihood ratio.

Structural Role

The Neyman–Pearson Lemma characterizes optimal tests for simple hypotheses. The likelihood ratio p_1/p_0 is the sufficient statistic for the simple-vs-simple problem: all relevant information for this test is contained in this ratio.

Extensions to composite hypotheses require additional structure: - Monotone likelihood ratio yields UMP tests for one-sided problems (Node 5.3). - Asymptotic approximations yield LR tests for general composite problems (Node 5.4). - The p-value (Node 5.5) calibrates the NP test statistic on the probability scale.

Mathematical Development

Theorem – Neyman–Pearson Lemma

There exists a most powerful level α test of $H_0 : P_0$ vs $H_1 : P_1$, and it has the form

$$\phi^*(x) = \begin{cases} 1 & \text{if } p_1(x) > k\, p_0(x), \\ \gamma & \text{if } p_1(x) = k\, p_0(x), \\ 0 & \text{if } p_1(x) < k\, p_0(x), \end{cases}$$

where $k \geq 0$ and $\gamma \in [0, 1]$ are chosen to satisfy $\mathbb{E}_0[\phi^*(X)] = \alpha$.

Moreover, ϕ^* is unique almost everywhere with respect to μ, except possibly on the set $\{p_1 = k\, p_0\}$.

Proof

Let ϕ be any level α test, i.e., $\mathbb{E}_0[\phi(X)] \leq \alpha$. Define $\Delta(x) = \phi^*(x) - \phi(x)$. We must show $\mathbb{E}_1[\Delta(X)] \geq 0$.

Consider the decomposition of the sample space into three regions:

- On $S_+ = \{x : p_1(x) > k\, p_0(x)\}$: $\phi^*(x) = 1 \geq \phi(x)$, so $\Delta(x) \geq 0$, and $p_1(x) - k\, p_0(x) > 0$.
- On $S_0 = \{x : p_1(x) = k\, p_0(x)\}$: $p_1(x) - k\, p_0(x) = 0$.
- On $S_- = \{x : p_1(x) < k\, p_0(x)\}$: $\phi^*(x) = 0 \leq \phi(x)$, so $\Delta(x) \leq 0$, and $p_1(x) - k\, p_0(x) < 0$.

Therefore $\Delta(x)(p_1(x) - k\, p_0(x)) \geq 0$ everywhere, giving

$$\int \Delta(x)\, p_1(x)\, d\mu(x) \geq k \int \Delta(x)\, p_0(x)\, d\mu(x).$$

The right-hand side is $k(\mathbb{E}_0[\phi^*] - \mathbb{E}_0[\phi]) = k(\alpha - \mathbb{E}_0[\phi]) \geq 0$, since $\mathbb{E}_0[\phi] \leq \alpha$.

Hence $\mathbb{E}_1[\phi^*] \geq \mathbb{E}_1[\phi]$. $\square$

Existence of k and γ

Define $F_0(t) = P_0(p_1(X)/p_0(X) \leq t)$, the CDF of the likelihood ratio under P_0. Set k to be the $(1-\alpha)$-quantile of $p_1(X)/p_0(X)$ under P_0:

$$k = \inf\{t : F_0(t) \geq 1 - \alpha\}.$$

If $P_0(p_1(X)/p_0(X) = k) = 0$, then γ is irrelevant and the test is non-randomized. Otherwise, choose

$$\gamma = \frac{\alpha - P_0(p_1(X)/p_0(X) > k)}{P_0(p_1(X)/p_0(X) = k)}.$$

Corollary: Power Is at Least α

If $P_0 \neq P_1$, then the power of the NP test satisfies $\mathbb{E}_1[\phi^*] \geq \alpha$, with equality only when $p_1 = k\,p_0$ a.e. In particular, the NP test is unbiased.

Proof. From the proof above, $\mathbb{E}_1[\phi^*] \geq k \cdot 0 + \mathbb{E}_1[\phi]$ for the trivial test $\phi \equiv \alpha$, which has $\mathbb{E}_1[\phi] = \alpha$. Since $\int \Delta(x)(p_1(x) - kp_0(x))\,d\mu \geq 0$ with the choice $\phi \equiv \alpha$, and equality requires $p_1 = kp_0$ a.e. on sets where $\Delta \neq 0$. $\square$

Sufficiency of the Likelihood Ratio

The NP test depends on the data only through the likelihood ratio $\Lambda(x) = p_1(x)/p_0(x)$. By the factorization theorem (Node 2.4), $\Lambda(X)$ is a sufficient statistic for the family $\{P_0, P_1\}$. Thus the NP test achieves optimality by reducing the data to a one-dimensional sufficient statistic.

Worked Examples

Example 1: Normal Mean Shift

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, 1)$. Test $H_0 : \mu = 0$ vs $H_1 : \mu = 1$ at level α.

The likelihood ratio is

$$\frac{p_1(x)}{p_0(x)} = \prod_{i=1}^{n} \frac{e^{-(x_i-1)^2/2}}{e^{-x_i^2/2}} = \exp\left(\sum_{i=1}^{n} x_i - \frac{n}{2}\right).$$

This is monotone increasing in $\bar{x} = n^{-1} \sum x_i$. The NP test rejects when $\bar{X} > c$, where c satisfies $P_0(\bar{X} > c) = \alpha$.

Under H_0, $\bar{X} \sim N(0, 1/n)$, so $c = z_\alpha/\sqrt{n}$.

The power is

$$\beta(\mu = 1) = P_1(\bar{X} > z_\alpha/\sqrt{n}) = 1 - \Phi(z_\alpha - \sqrt{n}).$$

For $n = 25$, $\alpha = 0.05$: $c = 1.645/5 = 0.329$, and power $= 1 - \Phi(1.645 - 5) = 1 - \Phi(-3.355) \approx 0.9996$.

Example 2: Exponential Rate

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$ with density $\lambda e^{-\lambda x}$. Test $H_0 : \lambda = 1$ vs $H_1 : \lambda = 2$ at level $\alpha = 0.05$.

The likelihood ratio is

$$\frac{p_1(x)}{p_0(x)} = \frac{2^n e^{-2\sum x_i}}{e^{-\sum x_i}} = 2^n \exp\left(-\sum_{i=1}^{n} x_i\right).$$

This is decreasing in $T = \sum x_i$, so the NP test rejects for small T.

Under H_0, $2T = 2\sum X_i \sim \chi^2_{2n}$. For $n = 5$, reject when $2T < \chi^2_{10,0.95}$. From tables, $\chi^2_{10,0.95} = 3.940$, so reject when $T < 1.970$.

Power: Under H_1, $4T = 4\sum X_i \sim \chi^2_{10}$, so $P_1(T < 1.970) = P(\chi^2_{10} < 7.880) \approx 0.36$.

Cross-Links

- **AI:** The likelihood ratio is the Bayes-optimal classifier under equal priors and 0-1 loss. The NP lemma formalizes the optimality of log-odds thresholding in binary classification.
- **Economics:** In mechanism design, the NP structure appears in optimal auction theory (Myerson): allocate the good to the bidder with the highest virtual value, analogous to rejecting for the highest likelihood ratio.
- **Linear Algebra:** In Gaussian models, the NP critical region is a half-space in $\mathbb{R}^n$, determined by the projection of the data onto the direction $\mu_1 - \mu_0$.
- **Quantum:** The Helstrom bound on quantum state discrimination is the quantum analogue of the NP bound: the optimal POVM element projects onto the positive part of $\rho_1 - k\rho_0$.

Structural Compression

Invariant: Among level α tests of P_0 vs P_1, the likelihood ratio test is most powerful.

Mechanism: Reject when $p_1(x)/p_0(x) > k$: the decision boundary is a level set of the likelihood ratio, allocating rejection probability greedily to the highest-evidence points.

Cross-System Echo: The NP lemma is an instance of the general isoperimetric principle: among all sets of fixed measure under one distribution, the one with maximal measure under another is the upper level set of the Radon–Nikodym derivative. This principle recurs in information theory (channel coding), optimal transport (Monge sets), and quantum information (Helstrom measurement).

Exercises

1. Prove that the Neyman–Pearson test is unique almost everywhere when the distribution of $p_1(X)/p_0(X)$ under P_0 is continuous.

2. Show that the power $\mathbb{E}_1[\phi^*(X)]$ is a non-decreasing function of α, and strictly increasing when P_0 and P_1 are mutually absolutely continuous.

3. Let $X \sim \text{Poisson}(\lambda)$. Derive the NP test of $H_0 : \lambda = 1$ vs $H_1 : \lambda = 3$ at level $\alpha = 0.05$. Compute the threshold k, the randomization probability γ, and the exact power.

4. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with σ^2 known. Show that the NP test of $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$ (with $\mu_1 > \mu_0$) rejects for large $\bar{X}$, and that the critical value does not depend on μ_1.

5. Prove that if ϕ^* is the NP test at level α and ϕ^* has power β^*, then the NP test of $H_0 : P_1$ vs $H_1 : P_0$ at level $1 - \beta^*$ has power $1 - \alpha$.

6. Let $P_0 = N(0, 1)$ and $P_1 = N(0, \sigma^2)$ with $\sigma^2 > 1$. Show that the NP test rejects for large $|X|$, and compute the threshold.

7. Argue, structurally, why the Neyman–Pearson Lemma is an instance of Lagrangian optimization: identify the objective, the constraint, and the role of the multiplier k, and explain why the complementary slackness condition corresponds to the randomization on the boundary $\{p_1 = kp_0\}$.

Statistics · Module 5 — Hypothesis Testing · Node 5.2
Concepts: hypothesis-testing, likelihood
Prerequisites: 5.1, 2.3, 2.4
Enables: 5.3, 5.5

hbar.university — own.your.cognition