Likelihood Ratio Tests

Statistics · Module 5 — Hypothesis Testing

Node 5.4

Position in Knowledge Graph

System: Statistics

Structural Domain: Hypothesis Testing

Node: 5.4

The likelihood ratio (LR) test is the general-purpose construction for composite hypothesis testing. Where Neyman–Pearson handles simple-vs-simple and UMP handles one-sided MLR families, LR tests apply to arbitrary parametric constraints without requiring special structural conditions. Wilks' theorem provides the asymptotic reference distribution, making LR tests practically computable in large samples. This node requires the testing framework (Node 5.1), density and likelihood theory (Node 2.3), and MLE theory (Node 3.2).

Formal Definition

Generalized Likelihood Ratio Statistic

Let $X_1, \ldots, X_n \sim p_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^k$. Consider

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta \setminus \Theta_0,$$

where $\Theta_0 \subset \Theta$ is defined by r smooth equality constraints, so $\dim(\Theta_0) = k - r$.

The **generalized likelihood ratio statistic** is

$$\Lambda_n = \frac{\sup_{\theta \in \Theta_0} L(\theta; X)}{\sup_{\theta \in \Theta} L(\theta; X)} = \frac{L(\hat{\theta}_0; X)}{L(\hat{\theta}_n; X)},$$

where $\hat{\theta}_n$ is the unrestricted MLE and $\hat{\theta}_0$ is the MLE restricted to Θ_0.

Note $\Lambda_n \in (0, 1]$; the test rejects H_0 for small values of Λ_n.

Log-Likelihood Form

The standard form is

$$-2 \log \Lambda_n = 2 \left[\ell(\hat{\theta}_n) - \ell(\hat{\theta}_0) \right],$$

which is non-negative. The test rejects H_0 for large values of $-2 \log \Lambda_n$.

Intuition

The LR statistic measures how much the data favors the unrestricted model over the restricted model. If the constraint $\theta \in \Theta_0$ costs little in likelihood, Λ_n is near 1 and there is no evidence against H_0. If the constraint forces a substantial loss in fit, Λ_n is small and the data contradicts H_0.

The factor of -2 in the log transform is not arbitrary: it is chosen so that the asymptotic distribution under H_0 is exactly chi-squared, without additional scaling. This normalization arises from the quadratic approximation to the log-likelihood near its maximum, with the Fisher information providing the natural metric.

Structural Role

LR tests provide a universal construction for composite hypotheses. Their advantages:

1. **Generality.** Applicable to any parametric model with smooth constraints.
2. **Invariance.** The LR statistic is invariant under reparameterization of θ.
3. **Asymptotic universality.** Wilks' theorem gives a single reference distribution (χ^2_r) regardless of the specific model.

The LR test, Wald test, and Score (Rao) test form the "holy trinity" of asymptotic tests (Node 4.4). All three converge to the same first-order asymptotic limit, but they differ in finite samples and in computational convenience.

Mathematical Development

Theorem – Wilks (1938)

Regularity conditions. Assume: (i) Θ is an open subset of $\mathbb{R}^k$; (ii) $\Theta_0 = \{\theta \in \Theta : g(\theta) = 0\}$ where $g : \mathbb{R}^k \to \mathbb{R}^r$ is continuously differentiable with rank r Jacobian; (iii) the log-likelihood is three times continuously differentiable with non-singular Fisher information $I(\theta)$; (iv) the MLE $\hat{\theta}_n$ is consistent and asymptotically normal.

Statement. Under H_0 (with true parameter $\theta_0 \in \Theta_0$),

$$-2 \log \Lambda_n \xrightarrow{d} \chi^2_r \quad \text{as } n \to \infty,$$

where $r = \dim(\Theta) - \dim(\Theta_0)$.

Proof Sketch

Step 1: Taylor expansion. Expand $\ell(\theta)$ around $\hat{\theta}_n$ to second order:

$$\ell(\theta) \approx \ell(\hat{\theta}_n) - \frac{1}{2}(\theta - \hat{\theta}_n)^\top \left[-\nabla^2 \ell(\hat{\theta}_n)\right] (\theta - \hat{\theta}_n).$$

Define the observed information $\hat{J}_n = -\nabla^2 \ell(\hat{\theta}_n)/n$. By the law of large numbers, $\hat{J}_n \xrightarrow{p} I(\theta_0)$.

Step 2: Quadratic approximation. Using the expansion at the restricted MLE:

$$-2 \log \Lambda_n = 2[\ell(\hat{\theta}_n) - \ell(\hat{\theta}_0)] \approx n(\hat{\theta}_n - \hat{\theta}_0)^\top \hat{J}_n (\hat{\theta}_n - \hat{\theta}_0).$$

Step 3: Asymptotic normality. Under H_0, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$.

Step 4: Constrained projection. The restricted MLE $\hat{\theta}_0$ is the projection of $\hat{\theta}_n$ onto Θ_0 in the metric defined by $I(\theta_0)$. The difference $\hat{\theta}_n - \hat{\theta}_0$ lies in the r-dimensional normal space to Θ_0 at θ_0.

Step 5: Chi-squared conclusion. The quadratic form $n(\hat{\theta}_n - \hat{\theta}_0)^\top I(\theta_0)(\hat{\theta}_n - \hat{\theta}_0)$ is asymptotically a sum of squares of r independent standard normals, hence χ^2_r. $\square$

Level α LR Test

Reject H_0 when

$$-2\log\Lambda_n > \chi^2_{r,\alpha},$$

where $\chi^2_{r,\alpha}$ is the upper α quantile of χ^2_r. This test has asymptotic level α by Wilks' theorem.

Relationship to Wald and Score Tests

Under the same regularity conditions:

Wald statistic:

$$W_n = n(\hat{\theta}_n - \theta_0)^\top \hat{I}_n(\hat{\theta}_n - \theta_0) \xrightarrow{d} \chi^2_r.$$

Score (Rao) statistic:

$$S_n = \frac{1}{n}\nabla\ell(\hat{\theta}_0)^\top I(\hat{\theta}_0)^{-1}\nabla\ell(\hat{\theta}_0) \xrightarrow{d} \chi^2_r.$$

All three statistics satisfy $W_n = -2\log\Lambda_n + o_p(1) = S_n + o_p(1)$ under H_0, so they are first-order asymptotically equivalent.

Bartlett Correction

The LR statistic admits a Bartlett correction: there exists a constant c (depending on the model) such that

$$(1 - c/n)\cdot(-2\log\Lambda_n)$$

has a χ^2_r distribution accurate to order $O(n^{-2})$, improving finite-sample performance. This correction is specific to the LR statistic and does not apply to the Wald or Score statistics.

Worked Examples

Example 1: Testing a Normal Mean (Known Variance)

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, 1)$. Test $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.

Here $k = 1$, $r = 1$, $\hat{\mu}_n = \bar{X}$, $\hat{\mu}_0 = \mu_0$.

$$\ell(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum(X_i - \mu)^2.$$

$$-2\log\Lambda_n = 2[\ell(\bar{X}) - \ell(\mu_0)] = n(\bar{X} - \mu_0)^2.$$

Under H_0, $\sqrt{n}(\bar{X} - \mu_0) \sim N(0, 1)$, so $-2\log\Lambda_n \sim \chi^2_1$ exactly.

Reject when $n(\bar{X} - \mu_0)^2 > \chi^2_{1,\alpha} = z^2_{\alpha/2}$, which is equivalent to the two-sided z-test $|\bar{X} - \mu_0| > z_{\alpha/2}/\sqrt{n}$.

Example 2: Testing Equality of Two Normal Means

Let $X_1, \ldots, X_m \overset{\text{iid}}{\sim} N(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu_2, \sigma^2)$, σ^2 known. Test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

Parameters: $\theta = (\mu_1, \mu_2) \in \mathbb{R}^2$, constraint $g(\theta) = \mu_1 - \mu_2 = 0$, so $r = 1$.

Unrestricted MLEs: $\hat{\mu}_1 = \bar{X}$, $\hat{\mu}_2 = \bar{Y}$. Restricted MLE: $\hat{\mu}_0 = (m\bar{X} + n\bar{Y})/(m + n)$.

$$-2 \log \Lambda_n = \frac{mn}{(m+n)\sigma^2}(\bar{X} - \bar{Y})^2.$$

Under H_0, $\bar{X} - \bar{Y} \sim N(0, \sigma^2(1/m + 1/n))$, so $-2 \log \Lambda_n \sim \chi^2_1$ exactly.

Example 3: Testing Independence in a Multinomial

Consider a 2×2 contingency table with cell counts $(n_{11}, n_{12}, n_{21}, n_{22})$ from a multinomial with $N = \sum n_{ij}$ and cell probabilities p_{ij}. Test $H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}$ (independence).

Here $k = 3$ (full multinomial has 3 free parameters), $\dim(\Theta_0) = 2$ (row and column marginals determine Θ_0), so $r = 1$.

The LR statistic is

$$-2 \log \Lambda_n = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{e}_{ij}},$$

where $\hat{e}_{ij} = n_{i\cdot} n_{\cdot j}/N$ are the expected counts under independence.

Under H_0, $-2 \log \Lambda_n \xrightarrow{d} \chi^2_1$ as $N \to \infty$.

Cross-Links

- **AI:** Model comparison in machine learning (e.g., nested neural network architectures, feature selection) uses likelihood ratio principles. The deviance in generalized linear models is exactly $-2 \log \Lambda_n$.
- **Economics:** LR tests are standard for testing restrictions in econometric models (e.g., testing coefficient restrictions in regression, testing overidentifying restrictions in GMM).
- **Linear Algebra:** The quadratic form $n(\hat{\theta}_n - \hat{\theta}_0)^\top I(\theta_0)(\hat{\theta}_n - \hat{\theta}_0)$ is a Mahalanobis distance; its chi-squared distribution follows from the spectral decomposition of the information matrix.
- **Quantum:** Quantum likelihood ratio tests for composite hypotheses on quantum states follow the same structure; the quantum Fisher information matrix replaces the classical one, and the SLD (symmetric logarithmic derivative) plays the role of the score.

Structural Compression

Invariant: The LR statistic compresses all information about the constraint $\theta \in \Theta_0$ into a single scalar via the log-likelihood difference.

Mechanism: Second-order Taylor expansion of the log-likelihood under regularity reduces $-2 \log \Lambda_n$ to a chi-squared quadratic form in r asymptotically normal directions, with r equal to the codimension of the constraint.

Cross-System Echo: The chi-squared reference distribution is shared by Wald and Score statistics; all three converge to the same first-order asymptotic limit. The $-2 \log \Lambda$ form reappears in information criteria (AIC $= -2\ell + 2k$ penalizes by model dimension), in deviance analysis for GLMs, and in the quantum Chernoff bound for asymptotic state discrimination.

Exercises

1. Derive the LR statistic for testing $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$ in the model $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with μ known. Verify the asymptotic chi-squared distribution.
2. Show that the LR statistic is invariant under reparameterization: if $\eta = g(\theta)$ is a diffeomorphism, the LR statistic computed in the η-parameterization equals the one in the θ-parameterization.
3. For the multinomial goodness-of-fit test ($H_0 : p = p_0$ with p_0 fully specified), compute $-2 \log \Lambda_n$ and verify that $r = k - 1$ where k is the number of categories.
4. Prove that for a one-parameter model with simple null $H_0 : \theta = \theta_0$, the LR statistic equals the square of the signed likelihood root: $-2 \log \Lambda_n = [\text{sign}(\hat{\theta}_n - \theta_0)\sqrt{-2 \log \Lambda_n}]^2$, and the signed root is asymptotically $N(0, 1)$.
5. Compute the LR test for $H_0 : \lambda_1 = \lambda_2$ vs $H_1 : \lambda_1 \neq \lambda_2$ given independent samples $X_1, \ldots, X_m \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_1)$ and $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda_2)$.
6. State the Bartlett correction for the LR test of $H_0 : \mu = 0$ in a $N(\mu, 1)$ model and verify that the corrected statistic has improved distributional accuracy.
7. Argue, structurally, why the LR test is the natural successor to the NP lemma for composite hypotheses: identify what structural property is lost when moving from simple to composite testing, and explain how the asymptotic chi-squared approximation substitutes for the exact NP optimality.

Statistics · Module 5 — Hypothesis Testing · Node 5.4
Concepts: hypothesis-testing, likelihood, maximum-likelihood-estimation
Prerequisites: 5.1, 2.3, 3.2
Enables: 5.5, 5.6