The Likelihood Ratio Test

Arthur Ollivier, Yacine Klikel, Simon Poirson, Antoine Chosson

STAT Flipped Classroom

12.12.2025

Outline

Motivations

The goal of the **Likelihood Ratio Test (LRT)** is to construct a statistical procedure to determine if a dataset $x \in E$ corresponds to a null hypothesis or an alternative hypothesis.

Formally, let Θ be the parameter space, partitioned into two disjoint sets Θ_0 and Θ_1:

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

We denote the hypotheses as follows:

$$\mathcal{H}_0 : x \sim P_\theta,\ \theta \in \Theta_0 \quad \text{versus} \quad \mathcal{H}_1 : x \sim P_\theta,\ \theta \in \Theta_1$$

Reminder — Parametric Test

Test

Let $\mathcal{P} = \{P_\theta(\mathrm{d}x),\ \theta \in \Theta\} \subset \mathcal{P}(\mathcal{E})$ be a parametric model. A **parametric test** of the hypothesis $\mathcal{H}_0$ against $\mathcal{H}_1$ is a measurable mapping:

$$\phi : (E, \mathcal{E}) \to (\{0,1\}, 2^{\{0,1\}})$$

Intuition:

- $\phi(x) = 0 \implies$ We accept $\mathcal{H}_0$ (Data supports $\theta \in \Theta_0$).
- $\phi(x) = 1 \implies$ We reject $\mathcal{H}_0$ in favor of $\mathcal{H}_1$ (Data supports $\theta \in \Theta_1$).

Reminder — Likelihood Function

Likelihood Function

Let $\mathcal{P} = \{p(x;\theta)\lambda(\mathrm{d}x),\ \theta \in \Theta\}$ be a parametric model absolutely continuous with respect to a σ-finite measure λ. For any observation $x \in E$, the **likelihood function** is defined as:

$$\theta \in \Theta \mapsto L(x;\theta) := p(x;\theta)$$

Example (i.i.d. case): Consider a model formed by n i.i.d. random variables $\mathbf{X}_{1:n} = (X_1, \ldots, X_n)$. The likelihood of the observation $\mathbf{x}_{1:n} = (x_1, \ldots, x_n) \in E^n$ is:

$$L_n(\mathbf{x}_{1:n};\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$

Likelihood Ratio Test (LRT) Definition (1/2)

Context: Simple Hypotheses

We consider the specific case of **simple hypotheses**, where the parameter space partitions into singletons $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$. The hypotheses are:

$$\mathcal{H}_0 : \theta = \theta_0 \quad \text{and} \quad \mathcal{H}_1 : \theta = \theta_1$$

The Likelihood Ratio Statistic

The test is based on the *likelihood ratio* statistic:

$$\Lambda_n(\mathbf{x}_{1:n}) = \frac{L_n(\mathbf{x}_{1:n}; \theta_1)}{L_n(\mathbf{x}_{1:n}; \theta_0)}$$

Likelihood Ratio Test (LRT) Definition (2/2)

The Rejection Region W_n^{LR}

We define the **rejection region**:

$$W_n^{\mathrm{LR}} = \left\{ \mathbf{x}_{1:n} \in E^n \mid \frac{L_n(\mathbf{x}_{1:n}; \theta_1)}{L_n(\mathbf{x}_{1:n}; \theta_0)} \geq A_n \right\}$$

for a certain A_n to determine in order the test is of level α.

The Test Function

Formally, the Likelihood Ratio Test (LRT) is the binary decision rule associated with this region (taking value 1 for rejection):

$$\phi_{\mathrm{LRT}}(\mathbf{x}_{1:n}) = \mathbb{I}\{\mathbf{x}_{1:n} \in W_n^{\mathrm{LR}}\} = \mathbb{I}\{\Lambda_n(\mathbf{x}_{1:n}) \geq A_n\}$$

Log-Likelihood Ratio Statistics

Normalized Log-Likelihood Ratio

We define the statistic $\lambda_n(\mathbf{X}_{1:n})$ as:

$$\lambda_n(\mathbf{x}_{1:n}) = \frac{1}{n} \log \Lambda_n(\mathbf{x}_{1:n}) = \frac{1}{n} \sum_{i=1}^{n} \left(l_1(x_i; \theta_1) - l_1(x_i; \theta_0) \right)$$

Moments under $\mathcal{H}_0$

We define:

- $h := \mathbb{E}_{\theta_0}[l_1(X_1; \theta_1) - l_1(X_1; \theta_0]$
- $s^2 := \mathrm{Var}_{\theta_0}(l_1(X_1; \theta_1) - l_1(X_1; \theta_0))$

Asymptotics of the LRT

Asymptotics of the LRT

Consider the test with the rejection region defined by the threshold a_n:

$$W_n^{\mathrm{LR,asymp}} = \{\lambda_n(\mathbf{X}_{1:n}) \geq a_n\} \quad \text{with} \quad a_n = h + \frac{s}{\sqrt{n}}\phi_{1-\alpha}$$

where $\phi_{1-\alpha}$ is the quantile of order $1-\alpha$ of $\mathcal{N}(0,1)$.

Then:

1. The test has **asymptotic level** α.
2. If the model is identifiable, the test is **consistent** (power tends to 1).

Remark

Note that $W_n^{\mathrm{LR,asymp}}$ is equivalent to the standard region W_n^{LR} by setting the threshold $A_n = \exp(na_n)$.

Preliminary Lemma (1/2)

Lemma (Expectation of the inverse likelihood ratio)

Under the alternative hypothesis $\mathcal{H}_1$, the expectation of the inverse likelihood ratio for a single observation is equal to 1:

$$\mathbb{E}_{\theta_1}\left[\frac{1}{\Lambda_1(X_1)}\right] = \mathbb{E}_{\theta_1}\left[\frac{p(X_1;\theta_0)}{p(X_1;\theta_1)}\right] = 1$$

Proof

By definition of the expectation:

$$\begin{aligned}
\mathbb{E}_{\theta_1}\left[\frac{p(X_1;\theta_0)}{p(X_1;\theta_1)}\right] &= \int \frac{p(x;\theta_0)}{p(x;\theta_1)} p(x;\theta_1)\mathrm{d}x \\
&= \int p(x;\theta_0)\mathrm{d}x \\
&= 1
\end{aligned}$$

Preliminary Lemma (2/2)

Consequence under $\mathcal{H}_1$ (Jensen's inequality)

Consider the convex function $\psi(t) = -\log(t)$. By Jensen's inequality:

$$\mathbb{E}_{\theta_1}\left[-\log\frac{p(X_1;\theta_0)}{p(X_1;\theta_1)}\right] \geq -\log\underbrace{\left(\mathbb{E}_{\theta_1}\left[\frac{p(X_1;\theta_0)}{p(X_1;\theta_1)}\right]\right)}_{=1 \text{ (by Lemma)}} = 0$$

Rearranging the terms, we define μ_1 and show it is non-negative:

$$\mu_1 := \mathbb{E}_{\theta_1}[l_1(X_1;\theta_1) - l_1(X_1;\theta_0)] \geq 0$$

Comparison with h (under $\mathcal{H}_0$)

By a symmetric argument under $\mathcal{H}_0$,

$$h = \mathbb{E}_{\theta_0}\left[\log\frac{p(X_1;\theta_1)}{p(X_1;\theta_0)}\right] \leq \log(1) = 0$$

Steps of the Asymptotic Proof

Strategy

We recall that $\lambda_n(\mathbf{X}_{1:n})$ is a sample mean of i.i.d. variables $Y_i = l_1(X_i; \theta_1) - l_1(X_i; \theta_0)$.

Steps

1. **Under $\mathcal{H}_0$:** Use the Central Limit Theorem (CLT) to establish the asymptotic level α.
2. **Under $\mathcal{H}_1$:** Use the Law of Large Numbers (LLN) and the **Preliminary Lemma** to lower bound the limit of the statistic.
3. **Consistency:** Conclude by comparing the limits under $\mathcal{H}_1$ (positive) and $\mathcal{H}_0$ (negative).

Asymptotic Proof (1/3)

Step 1: Asymptotic Level under $\mathcal{H}_0$

Under $\mathcal{H}_0$, the variables $Y_i = l_1(X_i;\theta_1) - l_1(X_i;\theta_0)$ are i.i.d. with mean h and variance s^2.

We apply the **Central Limit Theorem** (convergence in distribution):

$$\sqrt{n}\,(\lambda_n(\mathbf{X}_{1:n}) - h) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, s^2)$$

The rejection region is defined by $\lambda_n \geq a_n$ with $a_n = h + \frac{s}{\sqrt{n}}\phi_{1-\alpha}$. The probability of false alarm becomes:

$$\mathbb{P}_{\theta_0}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR,asymp}}) = \mathbb{P}_{\theta_0}\left(\frac{\sqrt{n}(\lambda_n - h)}{s} \geq \phi_{1-\alpha}\right)$$

By definition of the quantile $\phi_{1-\alpha}$, this probability converges to α.

Asymptotic Proof (2/3)

Step 2: Limit behavior under $\mathcal{H}_1$

Under $\mathcal{H}_1$, the variables Y_i are i.i.d. with expectation $\mu_1 = \mathbb{E}_{\theta_1}[Y_1]$.
We apply the **Strong Law of Large Numbers** (almost sure convergence):

$$\lambda_n(\mathbf{X}_{1:n}) \xrightarrow[n\to\infty]{a.s.} \mu_1$$

Application of the Lemma: From the Preliminary Lemma, we established using Jensen's inequality that:

$$\mu_1 = \mathbb{E}_{\theta_1}\left[-\log \frac{p(X_1;\theta_0)}{p(X_1;\theta_1)}\right] \geq 0$$

Asymptotic Proof (3/3)

Step 3: Consistency

- **Statistic:** $\lambda_n(\mathbf{X}_{1:n}) \xrightarrow[n\to\infty]{a.s.} \mu_1 \geq 0$ (under $\mathcal{H}_1$).
- **Threshold:** $a_n = h + \frac{s}{\sqrt{n}}\phi_{1-\alpha} \xrightarrow[n\to\infty]{h}$ (deterministic limit).

From the Lemma, we know that $h \leq 0 \leq \mu_1$. If the model is identifiable ($P_{\theta_0} \neq P_{\theta_1}$), the strict Jensen inequality applies, implying a **strictly positive gap**:

$$\mu_1 > 0 > h \implies \mu_1 - h > 0$$

Therefore, the difference converges almost surely to a positive constant:

$$\lambda_n(\mathbf{X}_{1:n}) - a_n \xrightarrow[n\to\infty]{a.s.} \mu_1 - h > 0$$

This implies $\mathbb{P}_{\theta_1}(\lambda_n(\mathbf{X}_{1:n}) \geq a_n) \to 1$.

Optimality of the LRT or the Neyman-Pearson Lemma

The Neyman-Pearson Lemma

We consider a test with rejection region W_n of level **at most** α, *i.e.* $\mathbb{P}_{\theta_0}(\mathbf{X}_{1:n} \in W_n) \leq \alpha$. Then, **the LRT is more powerful than this test**, that is:

$$\mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n) \leq \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}})$$

Proof's *vademecum*

- Key definition: Let's recall that $\mathbf{x}_{1:n} \in W_n^{\mathrm{LR}} \iff \frac{L_n(\mathbf{x}_{1:n};\theta_1)}{L_n(\mathbf{x}_{1:n};\theta_0)} \geq a_n$
- Main goal: Let's show that $\mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}}) - \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n) \geq 0$

Steps of the Neyman-Pearson Lemma's proof

Proof's *vademecum*

- Key definition: Let's recall that $\mathbf{x}_{1:n} \in W_n^{\mathrm{LR}} \iff \frac{L_n(\mathbf{x}_{1:n};\theta_1)}{L_n(\mathbf{x}_{1:n};\theta_0)} \geq a_n$
- Main goal: Let's show that $\mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}}) - \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n) \geq 0$

Steps of the proof

1. Express $\mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}}) - \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n)$ as the difference of two integrals over W_n^{LR} and W_n, respectively.
2. Modify the regions of integration to apply the definition of W_n^{LR}.
3. Apply the definition to **lower bound** the two terms.
4. Conclude using the levels of the two tests.

Proof of Neyman-Pearson Lemma (1/4)

Step 1: Expressing the difference as integrals

We want to evaluate the difference in power. Let
$\Delta = \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}}) - \mathbb{P}_{\theta_1}(\mathbf{X}_{1:n} \in W_n)$.
Using the integral definition of probability:

$$\Delta = \int_{W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n}; \theta_1)\mathrm{d}\mathbf{x}_{1:n} - \int_{W_n} L_n(\mathbf{x}_{1:n}; \theta_1)\mathrm{d}\mathbf{x}_{1:n}$$

Proof of Neyman-Pearson Lemma (2/4)

Step 2: Decomposition of the regions of integration

We decompose the regions using the intersection $W_n^{\mathrm{LR}} \cap W_n$:

$$\Delta = \left(\int_{W_n^{\mathrm{LR}} \setminus W_n} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n} + \int_{W_n^{\mathrm{LR}} \cap W_n} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n} \right)$$

$$- \left(\int_{W_n \setminus W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n} + \int_{W_n^{\mathrm{LR}} \cap W_n} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n} \right)$$

The intersection $W_n^{\mathrm{LR}} \cap W_n$ cancels out.

$$\Delta = \int_{W_n^{\mathrm{LR}} \setminus W_n} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n} - \int_{W_n \setminus W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n}; \theta_1) \mathrm{d}\mathbf{x}_{1:n}$$

Proof of Neyman-Pearson Lemma (3/4)

Step 3: Using the LRT definition

Recall that $\mathbf{x}_{1:n} \in W_n^{\mathrm{LR}} \iff L_n(\mathbf{x}_{1:n};\theta_1) \geq a_n L_n(\mathbf{x}_{1:n};\theta_0)$.

- On $W_n^{\mathrm{LR}} \setminus W_n$ (inside the LRT region):

$$L_n(\mathbf{x}_{1:n};\theta_1) \geq a_n L_n(\mathbf{x}_{1:n};\theta_0)$$

- On $W_n \setminus W_n^{\mathrm{LR}}$ (outside the LRT region):

$$-L_n(\mathbf{x}_{1:n};\theta_1) > -a_n L_n(\mathbf{x}_{1:n};\theta_0)$$

$$\Delta = \int_{W_n^{\mathrm{LR}} \setminus W_n} L_n(\mathbf{x}_{1:n};\theta_1)\mathrm{d}\mathbf{x}_{1:n} - \int_{W_n \setminus W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n};\theta_1)\mathrm{d}\mathbf{x}_{1:n}$$

becomes:

$$\Delta \geq a_n \int_{W_n^{\mathrm{LR}} \setminus W_n} L_n(\mathbf{x}_{1:n};\theta_0)\mathrm{d}\mathbf{x}_{1:n} - a_n \int_{W_n \setminus W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n};\theta_0)\mathrm{d}\mathbf{x}_{1:n}$$

Proof of Neyman-Pearson Lemma (4/4)

Step 4: Conclusion using the test levels

The intersection $W_n^{\mathrm{LR}} \cap W_n$ cancels out again, but in reverse.

$$
\begin{aligned}
\Delta &\geq a_n \left(\int_{W_n^{\mathrm{LR}} \setminus W_n} L_n(\mathbf{x}_{1:n}; \theta_0) \mathrm{d}\mathbf{x}_{1:n} - \int_{W_n \setminus W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n}; \theta_0) \mathrm{d}\mathbf{x}_{1:n} \right) \\
&\geq a_n \left(\int_{W_n^{\mathrm{LR}}} L_n(\mathbf{x}_{1:n}; \theta_0) \mathrm{d}\mathbf{x}_{1:n} - \int_{W_n} L_n(\mathbf{x}_{1:n}; \theta_0) \mathrm{d}\mathbf{x}_{1:n} \right) \\
&= a_n \left(\mathbb{P}_{\theta_0}(\mathbf{X}_{1:n} \in W_n^{\mathrm{LR}}) - \mathbb{P}_{\theta_0}(\mathbf{X}_{1:n} \in W_n) \right)
\end{aligned}
$$

- $\mathbb{P}_{\theta_0}(W_n^{\mathrm{LR}}) = \alpha$.
- $\mathbb{P}_{\theta_0}(W_n) \leq \alpha$.

Therefore, $\Delta \geq a_n(\alpha - \mathbb{P}_{\theta_0}(W_n)) \geq 0$.

The Generalized Likelihood Ratio (GLR)

Context: Composite Hypotheses

We now consider the general case where the hypotheses are not necessarily simple (i.e., parameters belong to sets):

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in \Theta \setminus \Theta_0$$

where $\Theta_0 \subset \Theta$ is a subset of the full parameter space.

The GLR Statistic

We generalize the likelihood ratio by comparing the best likelihood achievable under the full model versus the null model. We define $\tilde{\Lambda}_n(\mathbf{X}_{1:n})$ as:

$$\tilde{\Lambda}_n(\mathbf{x}_{1:n}) = \frac{\sup_{\theta \in \Theta} L_n(\mathbf{x}_{1:n}; \theta)}{\sup_{\theta \in \Theta_0} L_n(\mathbf{x}_{1:n}; \theta)}$$

Note: By construction, $\Theta_0 \subset \Theta$, so $\tilde{\Lambda}_n(\mathbf{x}_{1:n}) \geq 1$. Larger values indicate evidence against $\mathcal{H}_0$.

Asymptotic Behavior: Wilks' Theorem

To construct a test, we need the distribution of the statistic under $\mathcal{H}_0$.

Theorem 9.2.7 (Wilks' Theorem)

Assume certain regularity conditions on the model. Let p be the dimension of the full parameter space Θ, and q be the dimension of the subspace Θ_0 (with $q < p$).
Under the null hypothesis $\mathcal{H}_0$, as $n \to \infty$:

$$2 \log \tilde{\Lambda}_n(\mathbf{X}_{1:n}) \xrightarrow{d} \chi^2_{p-q}$$

converges in distribution to a Chi-squared distribution with $p - q$ degrees of freedom.

Intuition: The quantity $2 \log \tilde{\Lambda}_n$ measures the "distance" between the restricted model and the full model.

Generalized Likelihood Ratio Test (GLRT)

The Rejection Region

Based on Wilks' Theorem, the test of asymptotic level α is defined by the rejection region:

$$W_n = \left\{ \mathbf{x}_{1:n} \in E^n \mid 2 \log \tilde{\Lambda}_n(\mathbf{x}_{1:n}) \geq \chi^2_{p-q,1-\alpha} \right\}$$

where $\chi^2_{p-q,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_{p-q} distribution.

P-value

For an observed statistic $w_{obs} = 2 \log \tilde{\Lambda}_n(\mathbf{x}_{obs})$, the p-value is:

$$p\text{-value} = \mathbb{P}(Z \geq w_{obs}) \quad \text{where } Z \sim \chi^2_{p-q}$$

Goal of the Application

Detect whether texture (and ultimately all features) feature is relevant in detecting malignant tumors

The Breast Cancer Wisconsin Dataset

Overview

The Breast Cancer Wisconsin (Diagnostic) dataset contains **569 tumor samples** obtained from digitized images of fine needle aspirates. Each sample is labeled:

- **M** = Malignant
- **B** = Benign

Features

Each tumor is described by **30 continuous features** capturing morphology and texture:

- Radius, texture, perimeter, area, concavity
- Smoothness, symmetry, fractal dimension

Exploratory Visualization — Malignant vs Benign: Feature Distribution

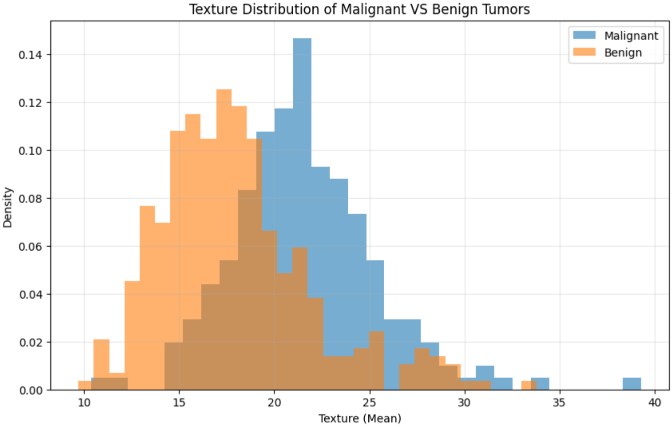

Figure: Texture Distribution of Malignant VS Benign Tumors

Hypothesis Formulation

Goal

Test whether the **texture** feature improves malignancy prediction.

Let δ be the coefficient of texture in the logistic regression.

$$H_0 : \; \delta = 0 \quad \text{(texture does not contribute)},$$
$$H_1 : \; \delta \neq 0 \quad \text{(texture is informative)}.$$

Likelihood in Logistic Regression

We observe independent pairs (Y_i, X_i) for $i = 1, \ldots, n$, where

$$Y_i \in \{0, 1\}, \qquad X_i = (x_{i1}, \ldots, x_{ik}).$$

In a logistic regression model with parameter vector β, we set

$$p_i(\beta) = \mathbb{P}(Y_i = 1 \mid X_i) = \frac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)}.$$

Conditionally on X_i, we assume

$$Y_i \mid X_i \sim \mathrm{B}(p_i(\beta)),$$

and the joint likelihood (under parameter β) is

$$L(\beta) = \prod_{i=1}^{n} p_i(\beta)^{Y_i} \left(1 - p_i(\beta)\right)^{1-Y_i}.$$

Log-Likelihood for Logistic Regression

We compute the **log-likelihood**:

$$\ell(\beta) = \log L(\beta) = \sum_{i=1}^{n} \left[Y_i \log p_i(\beta) + (1 - Y_i) \log\bigl(1 - p_i(\beta)\bigr) \right].$$

Using $p_i(\beta) = \dfrac{\exp(X_i^\top \beta)}{1 + \exp(X_i^\top \beta)}$, we can also write:

$$\ell(\beta) = \sum_{i=1}^{n} \left[Y_i X_i^\top \beta - \log\bigl(1 + \exp(X_i^\top \beta)\bigr) \right].$$

- The **maximum likelihood estimate** (MLE) $\hat{\beta}$ maximizes $\ell(\beta)$.

Full vs Reduced Model: Log-Likelihoods

Reduced model (without texture):

$$\operatorname{logit}\left(p_i^{(0)}\right) = \beta_0 + \gamma_1 z_{i1} + \cdots + \gamma_q z_{iq},$$

with parameter vector $\theta_0 = (\beta_0, \gamma_1, \ldots, \gamma_q)$.

Full model (with texture):

$$\operatorname{logit}\left(p_i^{(1)}\right) = \beta_0 + \gamma_1 z_{i1} + \cdots + \gamma_q z_{iq} + \delta\, c_i,$$

with parameter vector $\theta_1 = (\beta_0, \gamma_1, \ldots, \gamma_q, \delta)$.

We denote:

$$\ell_0 = \ell(\hat{\theta}_0) \quad \text{(maximized log-likelihood under reduced model)},$$

$$\ell_1 = \ell(\hat{\theta}_1) \quad \text{(maximized log-likelihood under full model)}.$$

Likelihood Ratio Test: Definition

We compare:

- H_0: texture has no effect ($\delta = 0$),
- H_1: texture has an effect ($\delta \neq 0$).

The **likelihood ratio** is:

$$\lambda = \frac{L_0}{L_1} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta}_1)} = \exp\big(\ell_0 - \ell_1\big),$$

where L_0 and L_1 are the maximized likelihoods under H_0 and H_1.

The **Likelihood Ratio Test statistic** is defined as:

$$\Lambda = -2 \log \lambda = -2\big(\ell_0 - \ell_1\big) = 2(\ell_1 - \ell_0).$$

- Large Λ means the full model fits **much better** than the reduced model.

Asymptotic Distribution of the LRT Statistic

Under H_0 (here: $\delta = 0$), according to Wilk's Theorem, the statistic has the following asymptotic distribution:

$$\Lambda \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2_{p-q},$$

where

$$p - q = \dim(\theta_1) - \dim(\theta_0)$$

is the difference in the number of parameters.

In our case:

- Full model adds one parameter: δ (concavity).
- Therefore, $p - q = 1$.
- Under H_0:

$$\Lambda \approx \chi^2_1.$$

Practical Computation of the LRT

Step 1. Fit the **reduced model** (without concavity) and record its maximized log-likelihood ℓ_0.

Step 2. Fit the **full model** (with concavity) and record its maximized log-likelihood ℓ_1.

Step 3. Compute the LRT statistic:

$$\Lambda = 2(\ell_1 - \ell_0).$$

Step 4. Under H_0, compare Λ to χ^2_1:

$$p\text{-value} = \mathbb{P}\big(\chi^2_1 \geq \Lambda\big).$$

Decision

- If p-value $< \alpha$ (we take 0.05), reject H_0: concavity significantly improves prediction.
- Otherwise, do not reject H_0.

Likelihood Ratio Test for texture_mean

Model Log-Likelihoods:

$$\ell_0 = -36.5877 \quad \text{(reduced model, without texture\_mean)}$$

$$\ell_1 = -30.1923 \quad \text{(full model, with texture\_mean)}$$

Likelihood Ratio Test:

$$\Lambda = 2(\ell_1 - \ell_0) = 12.7909$$

$$p\text{-value} = 0.000348$$

Conclusion

According to Wilks' theorem, $\Lambda \sim \chi^2(1)$ under H_0. Since $p < 0.05$, we **reject** H_0. `texture_mean` provides significant predictive information and should be included in the model.

Likelihood Ratio Test for All Remaining Features

Variable	Λ	p-value
area_worst	14.513	0.000139
texture_mean	12.791	0.000348
perimeter_se	6.517	0.010687
concave points_mean	4.361	0.036775
symmetry_worst	3.375	0.066205
fractal_dimension_se	3.152	0.075849
concavity_se	3.088	0.078886
fractal_dimension_worst	3.011	0.082680
compactness_se	2.388	0.122271
texture_se	2.040	0.153226
fractal_dimension_mean	1.471	0.225141
smoothness_worst	0.786	0.375446
smoothness_se	0.516	0.472602
symmetry_se	0.495	0.481805
symmetry_mean	0.300	0.583590
concave points_se	0.257	0.612088
smoothness_mean	0.002	0.964313