Generalized likelihood ratio test

Lecture 16a (STAT 24400 F24)

Simple vs composite hypotheses

Parametric family:

$$\text{Data} \sim f(\cdot \mid \theta)$$

- Recall: For testing a simple H_0 against a simple H_1, e.g.,

$$H_0 : \ \theta = \theta_0 \text{ vs } H_1 : \ \theta = \theta_1 \quad \rightsquigarrow \text{ use LRT: } \frac{\text{Likelihood of } \theta_0}{\text{Likelihood of } \theta_1}$$

- For testing a simple H_0 against a composite H_1, e.g.,

$$H_0 : \ \theta = \theta_0 \text{ vs } H_1 : \ \theta \neq \theta_0$$

  Can we use LRT? What is the "likelihood" for H_1?

- For testing a composite H_0 against a composite H_1, e.g., for $\mathrm{N}(\mu, \sigma^2)$:

$$H_0 : \ \mu = 0 \text{ (and } \sigma^2 \text{ unknown) vs } H_1 : \ \mu \neq 0 \text{ (and } \sigma^2 \text{ unknown)}$$

  What is the "likelihood" for H_0 & for H_1?

Cases of composite hypotheses

General framework: for Data $\sim f(\cdot \mid \theta)$, we test

$$H_0 : \theta \in \Omega_0, \quad H_1 : \theta \in \Omega_1$$

where $\Omega_0, \Omega_1 \subseteq \Theta$ are some sets of possible parameter values.

Examples

- Parametric family $\mathrm{N}(\mu, 1)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\rightsquigarrow \theta = \mu$, $\Theta = \mathbb{R}$, $\Omega_0 = \{0\}$, $\Omega_1 = (-\infty, 0) \cup (0, \infty)$
- Parametric family $\mathrm{N}(\mu, \sigma^2)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\rightsquigarrow \theta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times (0, \infty)$, $\Omega_0 = \{0\} \times (0, \infty)$,
  $\Omega_1 = ((-\infty, 0) \cup (0, \infty)) \times (0, \infty)$
- Parametric family Exponential(λ), testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$
  $\rightsquigarrow \theta = \lambda$, $\Theta = (0, \infty)$, $\Omega_0 = \{1\}$, $\Omega_1 = (0, 1) \cup (1, \infty)$

Generalized likelihood ratio (construction)

To compare H_0 vs H_1 using likelihood, we compare:

Maximum likelihood over all possible $\theta \in \Omega_0$ $\longleftarrow$ best likelihood under H_0

versus

Maximum likelihood over all possible $\theta \in \Omega_1$ $\longleftarrow$ best likelihood under H_1

Due to issues of endpoints / closed sets,
mathematically it works better to compare:

Maximum likelihood over all possible $\theta \in \Omega_0$ $\longleftarrow$ best likelihood under H_0

versus

Max. likelihood over all possible $\theta \in \Omega_0 \cup \Omega_1$ $\longleftarrow$ best likelihood under H_0 or H_1

Generalized likelihood ratio (definition)

We will define a generalized likelihood ratio for the test statistic:

$$\Lambda = \frac{\max_{\theta \in \Omega_0} f(X \mid \theta)}{\max_{\theta \in \Omega_0 \cup \Omega_1} f(X \mid \theta)} \quad \begin{array}{l} \leftarrow \text{ best likelihood under } H_0 \\ \leftarrow \text{ best likelihood under } H_0 \text{ or } H_1 \end{array}$$

Or, for i.i.d. data $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(\cdot \mid \theta)$,

$$\Lambda = \frac{\max_{\theta \in \Omega_0} \prod_{i=1}^n f(X_i \mid \theta)}{\max_{\theta \in \Omega_0 \cup \Omega_1} \prod_{i=1}^n f(X_i \mid \theta)} \quad \begin{array}{l} \leftarrow \text{ best likelihood under } H_0 \\ \leftarrow \text{ best likelihood under } H_0 \text{ or } H_1 \end{array}$$

Generalized likelihood ratio (properties)

- The numerator must be $\leq$ the denominator (deterministically)

$$0 < \Lambda \leq 1$$

- If H_0 is correct, then the denominator might still be slightly larger due to random chance

$$\Lambda \approx 1 \qquad \text{when } H_0 \text{ is true}$$

- If H_0 is not correct, the denominator might be much larger

$$\Lambda << 1 \quad \textit{or} \quad \Lambda \approx 0 \qquad \text{when } H_0 \text{ is far from the truth}$$

Example (normal unknown mean & known variance)

Example 1: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, 1)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$

- The likelihood: $\displaystyle\prod_{i=1}^n f(X_i \mid \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(X_i-\mu)^2/2}$
- The log-likelihood: $\displaystyle -\frac{1}{2}\sum_{i=1}^n (X_i - \mu)^2 + (\text{terms not depending on } \mu)$
- Maximize over $\mu \in \Omega_0 = \{0\} \rightsquigarrow$ the maximizer is $\hat{\mu} = 0$
- Maximize over $\mu \in \Omega_0 \cup \Omega_1 = \mathbb{R} \rightsquigarrow$ the maximizer is $\hat{\mu} = \bar{X}$
- The (generalized) likelihood ratio test statistic

$$\Lambda = \frac{\max_{\theta \in \Omega_0} \prod_{i=1}^n f(X_i \mid \theta)}{\max_{\theta \in \Omega_0 \cup \Omega_1} \prod_{i=1}^n f(X_i \mid \theta)} = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(X_i-0)^2/2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(X_i-\bar{X})^2/2}}$$
$$= e^{-\sum_i X_i^2/2 + \sum_i (X_i - \bar{X})^2/2} = e^{-n\bar{X}^2/2}$$

Discussion: Consider scenarios when H_0 is true vs if H_1 is true.

Example (normal unknown mean & unknown variance)

Example 2: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \sigma^2)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$

- The likelihood:

$$\prod_{i=1}^n f(X_i \mid \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(X_i-\mu)^2/2\sigma^2}$$

- The log-likelihood:

$$-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2 - \frac{n}{2}\log(\sigma^2) + (\text{terms not depending on } \mu \text{ or } \sigma^2)$$

- Max over $(\mu, \sigma^2) \in \Omega_0 = \{0\} \times (0, \infty) \rightsquigarrow$ max at $\hat{\mu} = 0,\ \hat{\sigma}^2 = \frac{1}{n}\sum_i X_i^2$
- Max over $(\mu, \sigma^2) \in \Omega_0 \cup \Omega_1 = \mathbb{R} \times (0, \infty) \rightsquigarrow$ max at $\hat{\mu} = \bar{X},\ \hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$

$$\Lambda = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi \cdot \frac{1}{n}\sum_i X_i^2}} e^{-(X_i-0)^2/2 \cdot \frac{1}{n}\sum_i X_i^2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi \cdot \frac{1}{n}\sum_i (X_i-\bar{X})^2}} e^{-(X_i-\bar{X})^2/2 \cdot \frac{1}{n}\sum_i (X_i-\bar{X})^2}} = \sqrt{1 - \frac{n\bar{X}^2}{\sum_i X_i^2}}$$

Examples (exponential rate)

Example 3: $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$, testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$

- The likelihood:

$$\prod_{i=1}^{n} f(X_i \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$$

- The log-likelihood:

$$n \log(\lambda) - \lambda \sum_{i=1}^{n} X_i$$

- Maximize over $\lambda \in \Omega_0 = \{1\} \rightsquigarrow$ the maximizer is $\hat{\lambda} = 1$
- Maximize over $\lambda \in \Omega_0 \cup \Omega_1 = (0, \infty) \rightsquigarrow$ the maximizer is $\hat{\lambda} = 1/\bar{X}$

$$\Lambda = \frac{\prod_{i=1}^{n} 1 \cdot e^{-1 \cdot X_i}}{\prod_{i=1}^{n} \frac{1}{\bar{X}} \cdot e^{-\frac{1}{\bar{X}} \cdot X_i}} = \frac{\bar{X}^n e^n}{e^{n\bar{X}}}$$

Hypothesis testing with the generalized LR

To use the GLR as a test statistic for testing H_0 vs H_1:

- $\Lambda \leq 1$ deterministically
- $\Lambda \approx 1$ is consistent with H_0
- Λ much lower than 1 is evidence in favor of H_1

How small does Λ need to be, for us to reject H_0?

Our goal:

$$\mathbb{P}_{H_0}(\Lambda < (\text{the threshold we choose})) \approx \alpha$$

for desired Type I error level α.

$\rightsquigarrow$ need to know the (approximate) null distribution of Λ

The null distribution (asymptotic)

Asymptotic result (when sample size n is large):

Under some regularity conditions, $-2\log(\Lambda) \approx \chi^2_{\underbrace{d}_{\text{dimension of } \Omega_0 \cup \Omega_1} - \underbrace{d_0}_{\text{dimension of } \Omega_0}}$

Part of the conditions:
Ω_0 is interior to $\Omega_0 \cup \Omega_1$, not at an endpoint

- Valid — $\Theta = \mathbb{R}$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
- Valid — $\Theta = \mathbb{R} \times (0, \infty)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
- Valid — $\Theta = (0, \infty)$, testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$
- Not valid — $\Theta = \mathbb{R}$, test $H_0 : \mu = 0$ vs $H_1 : \mu > 0$

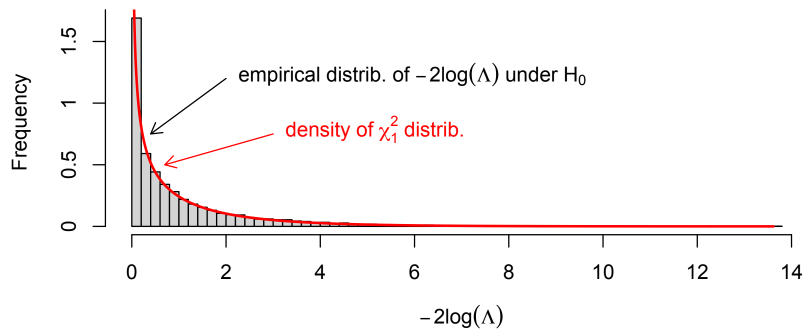
The null distribution (dimension calculation)

How to calculate d & d_0 — examples:

- Parametric family $N(\mu, \sigma^2)$ with μ unknown & σ^2 known,
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 0, \ d = 1$
- Parametric family $N(\mu, \sigma^2)$ with μ & σ^2 unknown,
  test $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$
  $\Rightarrow d_0 = 1, \ d = 2$
- Parametric family $N(\mu, \sigma^2)$ with μ & σ^2 unknown,
  test $H_0 : (\mu, \sigma^2) = (0, 1)$ vs $H_1 : (\mu, \sigma^2) \neq (0, 1)$
  $\Rightarrow d_0 = 0, \ d = 2$
- Parametric family Exponential(λ), test $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$
  $\Rightarrow d_0 = 0, \ d = 1$

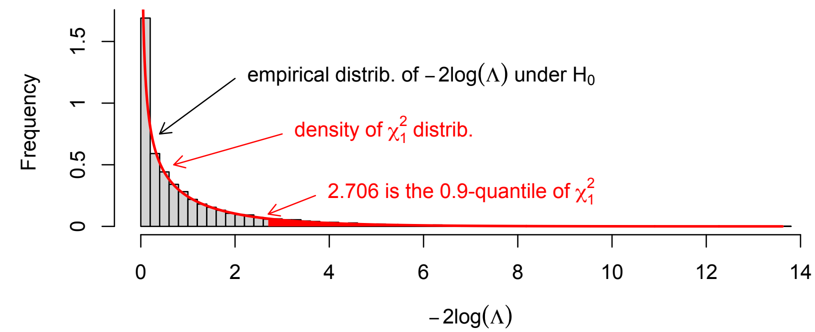
The null distribution (illustration)

A simulation with Exponential(λ) ($n = 10$), testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$:



How to run the test

A simulation with Exponential(λ), testing $H_0 : \lambda = 1$ vs $H_1 : \lambda \neq 1$:



- To run a hypothesis test at level α, compute a threshold $F^{-1}_{\chi^2_{d-d_0}}(1-\alpha)$, and reject H_0 if $-2\log(\Lambda)$ is $>$ the threshold
  (Note — this is always run as a one-tailed test)
- To compute a p-value, compute $1 - F_{\chi^2_{d-d_0}}(-2\log(\Lambda))$

A special case (asymptotic = exact)

For i.i.d. normal data, the asymptotic distribution $-2\log\Lambda \sim \chi^2_{d-d_0}$ is exact.

Back to Example 1: $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathrm{N}(\mu, 1)$, testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$

$$\Lambda = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(X_i - 0)^2/2}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(X_i - \bar{X})^2/2}} = e^{-\sum_i X_i^2/2 + \sum_i (X_i - \bar{X})^2/2} = e^{-n\bar{X}^2/2}$$

$$-2\log(\Lambda) = -2\log\left(e^{-n\bar{X}^2/2}\right) = n\bar{X}^2$$

$$\text{Under } H_0,\ \bar{X} \sim \mathrm{N}(0, 1/n) \ \Rightarrow\ \sqrt{n}\bar{X} \sim \mathrm{N}(0,1) \ \Rightarrow\ n\bar{X}^2 \sim \chi^2_1$$

$\rightsquigarrow$ In this special case, the asymptotic approximation is the exact distribution.

Another example

Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim}$ Geometric(p). Testing $H_0 : p = \frac{1}{2}$ vs $H_1 : p \neq \frac{1}{2}$.

- The likelihood:
$$\prod_{i=1}^n f(X_i \mid p) = \prod_{i=1}^n p(1-p)^{X_i - 1} = p^n (1-p)^{\sum_i X_i - n}$$
- The log-likelihood:
$$n\log(p) + \left(\textstyle\sum_i X_i - n\right)\log(1-p)$$
- Maximize over $p \in \Omega_0 = \{\frac{1}{2}\} \rightsquigarrow$ the maximizer is $\hat{p} = \frac{1}{2}$
- Maximize over $p \in \Omega_0 \cup \Omega_1 = (0,1) \rightsquigarrow$ the maximizer is $\hat{p} = \frac{1}{\bar{X}}$
$$\Lambda = \frac{\prod_{i=1}^n \left(\frac{1}{2}\right) \cdot \left(1 - \frac{1}{2}\right)^{X_i - 1}}{\prod_{i=1}^n \left(\frac{1}{\bar{X}}\right) \cdot \left(1 - \frac{1}{\bar{X}}\right)^{X_i - 1}}$$
- To compute a p-value — $d_0 = 0$ and $d = 1$, so:
$$\text{p-value} = 1 - F_{\chi^2_1}(-2\log(\Lambda))$$