Testing Hypotheses II

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

Confidence Intervals and Hypothesis Tests

Example 9.3A

- $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, unknown μ, known σ^2.
- Test hypotheses: $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.
- Use α-level test that rejects H_0 when $|\overline{X} - \mu_0| > t_0$

Critical value: $t_0 = \sigma_{\overline{X}} z(\alpha/2)$
Acceptance Region: $A(\mu_0) = \{\overline{X} : |\overline{X} - \mu_0| < \sigma_{\overline{X}} z(\alpha/2)\}$
which is equivalent to $\overline{X}$ values satisfying:

$$\begin{array}{rcccl} -\sigma_{\overline{X}} z(\alpha/2) & < & \overline{X} - \mu_0 & < & +\sigma_{\overline{X}} z(\alpha/2) \\ \text{or} \quad \overline{X} - \sigma_{\overline{X}} z(\alpha/2) & < & \mu_0 & < & \overline{X} + \sigma_{\overline{X}} z(\alpha/2) \end{array}$$

Confidence Interval for μ:

$$C(\overline{X}) = [\overline{X} - \sigma_{\overline{X}} z(\alpha/2)\ ,\ \overline{X} + \sigma_{\overline{X}} z(\alpha/2)]$$

(Confidence Level $= 100(1-\alpha)\%$)

NOTE: $\overline{X} \in A(\mu_0)$ if and only if $\mu_0 \in C(\overline{X})$ (!!)

Duality of Tests and Confidence Intervals

Theorem 9.3A Suppose

- For every $\theta_0 \in \Theta$ there is a test at level α of the hypothesis $H_0 : \theta = \theta_0$, and
- $A(\theta_0)$ is the acceptance region of the test.

Then the set

$$C(\mathbf{X}) = \{\theta : \mathbf{X} \in A(\theta)\}$$

is a $100(1-\alpha)\%$ confidence region for θ.

Proof: Because A is the acceptance region of a level-α test:

$$P[\mathbf{X} \in A(\theta_0) | \theta = \theta_0] = 1 - \alpha$$

For a given $\mathbf{X} = \mathbf{x}$,

$$\begin{aligned} \theta_0 \in C(\mathbf{x}) &\implies \mathbf{x} \in A(\theta_0) \\ \text{and} \quad \mathbf{x} \in A(\theta_0) &\implies \theta_0 \in C(\mathbf{x}) \, , \end{aligned}$$

so $\{\mathbf{x} \in A(\theta_0)\} \equiv \{\mathbf{x} : C(\mathbf{x}) \ni \theta_0\}$.

$$\implies P[C(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0] = 1 - \alpha.$$

Duality of Tests and Confidence Intervals

Theorem 9.3B Suppose

- $C(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence region for θ, i.e., for every θ_0

$$P[C(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0) = 1 - \alpha.$$

Then, an acceptance region for a test at level α of the hypothesis $H_0 : \theta = \theta_0$ can be constructed as:

$$A(\theta_0) = \{\mathbf{X} : C(\mathbf{X}) \ni \theta_0\}$$

Proof: Because $\{\mathbf{x} : C(\mathbf{x}) \ni \theta_0\} \equiv \{\mathbf{x} \in A(\theta_0)\}$,

$$\Longrightarrow P[\{\mathbf{x} \in A(\theta_0)\}] = P[C(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0] = 1 - \alpha.$$

Outline

Generalized Likelihood Ratio Tests

Likelihood Analysis Framework

- Data observations: $\mathbf{X} = (X_1, \ldots, X_n)$
- Joint distribution of $\mathbf{X}$ given by joint pdf/pmf
  $$f(\mathbf{x} \mid \theta),\ \theta \in \Theta$$
- Null and Alternative Hypotheses
  $$H_0 : \theta \in \Theta_0, \text{ and } H_1 : \theta \notin \Theta_0,$$
  for some proper subset $\Theta_0 \subset \Theta$.
- The MLE of θ solves: $lik(\hat{\theta}) = \max_{\theta \in \Theta} lik(\theta)$
  where $lik(\theta) = f(\mathbf{x} \mid \theta)$ (a function of θ given data $\mathbf{x}$)
- The MLE of θ under H_0 solves $lik(\hat{\theta}_0) = \max_{\theta \in \Theta_0} lik(\theta)$.

Definition: The **generalized likelihood ratio**

$$\Lambda = \frac{lik(\hat{\theta}_0)}{lik(\hat{\theta})} \qquad \text{(for testing } H_0 \text{ vs } H_1\text{)}$$

Generalized Likelihood Ratio Test

- Generalized likelihood ratio for testing H_0 vs H_1:
$$\Lambda = \frac{lik(\hat{\theta}_0)}{lik(\hat{\theta})}$$
- Properties of Λ
$$\Lambda > 0, \qquad \text{since } lik(\theta) > 0$$
$$\Lambda \leq 1, \qquad \text{because } lik(\hat{\theta}) \geq lik(\theta_0)$$
- Higher values of Λ are evidence in favor H_0
- Lower values of Λ are evidence against H_0
- **Rejection Region** of Generalized Likelihood Ratio Test:
$$\{\mathbf{x} : \Lambda < \lambda_0\} \text{ for some } \lambda_0$$
- For level-α test of simple H_0 choose λ_0 :
$$P(\Lambda < \lambda_0 \mid H_0) = \alpha$$
- If H_0 is composite, then choose largest λ_0 :
$$P(\Lambda < \lambda_0 \mid \theta) \leq \alpha, \text{ for all } \theta \in \Theta_0$$

Generalized Likelihood Ratio Test

Define LRStat by Rescaling the Likelihood Ratio

$$LRStat = -2 \times \log(\Lambda) = -2 \times \log[\frac{lik(\hat{\theta}_0)}{lik(\hat{\theta})}]$$

- Since $0 < \Lambda < 1$,
  $$LRStat > 0$$
  Evidence against H_0 given by high values of $LRStat$.
- For simple $H_0 : \theta = \theta_0$,
  $$LRStat = 2[\ell(\hat{\theta}) - \ell(\theta_0)]$$
- From asymptotic theory
  $$\ell(\theta_0) \approx \ell(\hat{\theta}) + (\theta_0 - \hat{\theta})\ell'(\hat{\theta}) + \tfrac{1}{2}(\theta_0 - \hat{\theta})^2 \ell''(\hat{\theta})$$
  so
  $$\begin{aligned} LRStat &\approx [\hat{\theta} - \theta_0]^2 \times [-\ell''(\hat{\theta})] \\ &= [\sqrt{n}(\hat{\theta} - \theta_0)^2] \times [-\ell''(\hat{\theta})/n] \\ &\xrightarrow{\mathcal{D}} [\sqrt{n\mathrm{I}(\theta_0)}(\hat{\theta} - \theta_0)^2] \sim [N(0,1)]^2 \sim \chi^2_1 \end{aligned}$$

Constructing Generalized Likelihood Ratio Tests

Test Statistic for Generalized Likelihood Ratio Test

$$\begin{aligned} LRStat &= -2\log(\Lambda) = -2 \times \log[\frac{lik(\hat{\theta}_0)}{lik(\hat{\theta})}] \\ &= 2 \times [\ell(\hat{\theta}) - \ell(\hat{\theta}_0)] \end{aligned}$$

Example 1: Test for Mean of Normal Distribution

- $X_1, \ldots, X_n$ i.i.d. $N(\theta, \sigma^2)$, (known variance)
- $\log[f(x_i \mid \theta)] = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - \theta)^2$
- $\ell(\theta) = \sum_{i=1}^n \log[f(x_i \mid \theta)] = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2$

For testing $H_0 : \theta = \theta_0$

$$LRStat = 2[\ell(\hat{\theta}) - \ell(\theta_0)] = \frac{1}{\sigma^2}[-\sum_1^n (x_i - \hat{\theta})^2 + \sum_1^n (x_i - \theta_0)^2]$$

- Note that
$$\begin{aligned} \sum_1^n (x_i - \theta_0)^2 &= \sum_1^n (x_i - \overline{x})^2 + n(\overline{x} - \theta_0)^2 \\ &= \sum_1^n (x_i - \hat{\theta})^2 + n(\overline{x} - \theta_0)^2 \end{aligned}$$
- So $LRStat = \dfrac{n(\overline{x} - \theta_0)^2}{\sigma^2} \qquad \sim N(0,1) underH_0$

Constructing Generalized Likelihood Ratio Tests

Test Statistic for Generalized Likelihood Ratio Test

$$\begin{aligned} LRStat &= -2\log(\Lambda) = -2 \times \log[\frac{lik(\hat{\omega}_0)}{lik(\hat{\omega})}] \\ &= 2 \times [\ell(\hat{\omega}) - \ell(\hat{\omega}_0)] \end{aligned}$$

Example 2: Test for Mean of Normal Distribution

- $X_1, \ldots, X_n$ i.i.d. $N(\theta, \sigma^2)$, (unknown variance)
- Parameter $\omega = (\theta, \sigma^2) \in \Omega = (-\infty, +\infty) \times (0, \infty)$
- $\log[f(x_i \mid \theta, \sigma^2)] = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x_i - \theta)^2$
- $\ell(\omega) = \ell(\theta, \sigma^2) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2$

For testing $H_0 : \theta = \theta_0$: use the overall mle and the mle given H_0

- Overall mle's: $\hat{\theta} = \overline{x}$, and $\hat{\sigma}^2 = \sum_1^n (x_i - \overline{x})^2/n$
  $$\ell(\hat{\theta}, \hat{\sigma}^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\hat{\sigma}^2) - \frac{n}{2}$$
- Under $H_0 : \hat{\theta}_0 = \theta_0$, and $\hat{\sigma}_0^2 = \sum_1^n (x_i - \theta_0)^2/n$
  $$\ell(\hat{\theta}_0, \hat{\sigma}_0^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\hat{\sigma}_0^2) - \frac{n}{2}$$

$$LRStat = 2[\ell(\hat{\theta}, \hat{\sigma}^2) - \ell(\theta_0, \hat{\sigma}_0^2)] = n\ln(\hat{\sigma}_0^2/\hat{\sigma}^2)$$

Constructing Generalized Likelihood Ratio Tests

Example 2: Test for Mean of a Normal Distribution

From before,

$$LRStat = 2[\ell(\hat{\theta}, \hat{\sigma}^2) - \ell(\theta_0, \hat{\sigma}_0^2)] = n \ln(\hat{\sigma}_0^2 / \hat{\sigma}^2)$$

- Note that
  $$\begin{aligned} \hat{\sigma}_0^2 &= \tfrac{1}{n}\sum_1^n (x_i - \theta_0)^2 = \tfrac{1}{n}[\sum_1^n (x_i - \overline{x})^2 + n(\overline{x} - \theta_0)^2] \\ &= \hat{\sigma}^2 + (\overline{x} - \theta_0)^2 \end{aligned}$$
- So $LRStat = n \ln \left(1 + \dfrac{(\overline{x} - \theta_0)^2}{\hat{\sigma}^2} \right)$
- $LRStat$ is a monotone function of $|T|$, where
  $$T = \frac{\sqrt{n}(\overline{X} - \theta_0)}{s}$$
  since $s^2 = n\hat{\sigma}^2/(n-1)$
- Under H_0 $T \sim t$-distribution on $(n-1)$ degrees of freedom.

Result: Generalized LR Test $\Longleftrightarrow$ t Test.

Generalized Likelihood Ratio Tests for Multinomial Distributions

Bernoulli Trials

- $B_1, B_2, \ldots, B_n$ i.i.d. $Bernoulli(p)$

$$P(B_i = 1 \mid p) = p = 1 - P(B_i = 0 \mid p)$$

- $X = B_1 + B_2 + \cdots B_n$, count of Bernoulli successes.
- $X \sim Binomial(n, p)$

Multinomial Trials

- $M_1, M_2, \ldots, M_n$ i.i.d. $Multinomial(p_1, p_2, \ldots, p_m)$
- Each M_i has m possible outcomes

  $A_1, A_2, \ldots, A_m$ ("cell outcomes")
  (mutually exclusive and exhaustive)
  $P(M_i = A_j) = p_j,\ j = 1, \ldots, m$ where
  $p_j \geq 0$, for $j = 1, \ldots, m$ and $\sum_1^m p_j = 1$.

- Define counts $X_1, X_2, \ldots, X_m$

$$X_1 = \#(M_i \text{ equal to } A_1), \ldots, X_m = \#(M_i \text{ equal to } A_m),$$

Multinomial Distribution

Multinomial Trials (continued)

- The collection of counts follows a Multinomial Distribution

  $n =$ number of multinomial trials,

  $p = (p_1, p_2, \ldots, p_m)$ (cell probabilities)

  the pmf of $(X_1, X_2, \ldots, X_m)$ is

  $$P(X_1 = x_1, \ldots, X_m = x_m) = \left(\frac{n!}{x_1! \cdots x_m!}\right) p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}$$

- The values of x_i are constrained, $n = \sum_j x_j$.
- The parameter space is $\Omega = \{p : p_j \geq 0, \sum_1^m p_j = 1\}$

  Note: Dimension of Ω is $(m-1)$
- Single counts are binomial random variables

  E.g., $X_1 \sim Binomial(n, p_1)$, and $X_2 \sim Binomial(n, p_2)$, etc.
- Multiple counts are not independent

  E.g., $X_1 \equiv n - (X_2 + X_3 + \cdots X_m)$

Examples Using Multinomial Distributions

- Hardy-Weinberg Equilibrium
  - Data consisting of counts of phenotypes: X_1, X_2, X_3
  - Cell probabilities $(1-\theta)^2$, $2\theta(1-\theta)$ θ^2; $0 < \theta < 1$.

  Hypothesis: the Hardy-Weinberg model is valid for specific data.
- Counts data from various applications
  - Asbestos fiber counts on slides
  - Counts of Bacterial clumps

  Hypothesis: a $Poisson(\lambda)$ model is valid for specific data
- Histogram of sample data
  - The frequency histogram of bin counts follows a multinomial distribution
    (for m fixed bins in a data histogram)

  Hypothesis: the data is a random sample from some fixed distribution or some given family of distributions.

Likelihood Ratio Test for Multinomial Distribution

Null Hypothesis H_0: A model that specifies the cell probabilities

$$p_1(\theta), p_2(\theta), ..., p_m(\theta)$$

which may vary with a parameter θ (taking values in ω_0)

Alternate Hypothesis H_1: General model that assumes

- $p = (p_1, p_2, \ldots, p_m)$ is fixed, but unknown
- Only constraint on p is that $\sum_j p_j = 1$ (and $p_j \geq 0$)

Constructing the Likelihood Ratio Test

- Compute mle under H_0: $\hat{p}_0 = (p_1(\hat{\theta}), \ldots, p_m(\hat{\theta}))$
  $\hat{p}_0$ maximizes $Lik(p)$ for $p \in \Omega_0$
  where $\Omega_0 = \{p = (p_1(\theta), \ldots, p_m(\theta)), \theta \in \omega_0\}$
- Compute overall mle
  $$\hat{p} = (\hat{p}_1, \ldots, \hat{p}_m), \text{ where } \hat{p}_j = x_j/n \text{ for all cells } A_j.$$
- Compute the likelihood ratio
  $$\Lambda = \frac{Lik(\hat{p}_0)}{Lik(\hat{p})} = \prod_{j=1}^{m} \left(\frac{p_j(\hat{\theta})}{\hat{p}_j}\right)^{x_j}$$

Likelihood Ratio Test For Multinomial Distribution

Constructing the Likelihood Ratio Test (continued)

- Compute the likelihood ratio

$$\Lambda = \frac{Lik(\hat{p}_0)}{Lik(\hat{p})} = \prod_{j=1}^{m} \left(\frac{p_j(\hat{\theta})}{\hat{p}_j}\right)^{x_j}$$

- Compute scaled log likelihood ratio:

$$\begin{aligned} LRStat &= -2 \times \log(\Lambda) \\ &= 2\sum_{j=1}^{m} x_j \ln(\hat{p}_j / p_j(\hat{\theta})) \\ &= 2\sum_{j=1}^{m} O_j \ln(O_j / E_j) \end{aligned}$$

  where $O_j = X_j$ and $E_j = np_j(\hat{\theta})$
- Pearson Chi-Square Statistic

$$ChiSqStat = \sum_{j=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

- $LRStat$ and $ChiSqStat$ are almost equivalent
  Use Taylor Series: $f(x) = x\ln(x/x_0) \approx (x - x_0) + \frac{1}{2}\frac{x-x_0{}^2}{x_0}$

Llikelihood Ratio Test for Multinomial Distribution

LRStat and Pearson's ChiSquare Statistic

- $LRStat \quad = 2\sum_{j=1}^{m} O_j \ln(O_j/E_j)$
  where $O_j = X_j$ and $E_j = np_j(\hat{\theta})$
- $ChiSqStat = \sum_{j=1}^{m} \dfrac{(O_i - E_i)^2}{E_i}$

Asymptotic/Approximate Distribution

- Chi-square distribution with q degrees of freedom
- Degrees of freedom q:
  $$q = dim(\Omega) - dim(\omega_0)$$
  Dimension of $\Omega = \{p\}$ (unconstrained)
  minus dimension of $\{p\}$ under H_0 $(\theta \in \omega_0)$
  (Proven in advanced statistics course)
- For *Multinomial* $(X_1, \ldots, X_m)$, with $p = (p_1, \ldots, p_m)$
  $$dim(\Omega) = m - 1$$

Degrees of Freedom for ChiSquare Test Statistic

- For Hardy-Weinberg Model, $m = 3$, $dim(\Omega) = (m-1) = 2$, and $k = dim(\omega_0) = 1$ so
$$q = m - 1 - k = 1.$$
- For distribution of m set of counts and
$$\omega_0 = \{Poisson(\lambda), \lambda > 0\}$$
$dim(\Omega) = m - 1$ and $k = dim(\omega_0) = 1$, so
$$q = m - 1 - 1 = m - 2$$
- For distribution of m set of counts and
$\omega_0 = \{Normal(\theta, \sigma^2)\}$ distributions.
$dim(\Omega) = m - 1$ and $k = dim(\omega_0) = 2$, so
$$q = m - 1 - 2 = m - 3$$