STAT 231: Statistics

Charles Shen

Spring 2016, University of Waterloo

Formulas and Notes.

Assume all log **are in base** e **unless specified.**
I've tried to use ln for consistency,
but there may be a few inconsistency.

Feel free to email feedback to me at echen902@gmail.com.

Contents

1 Statistical Models and Maximum Likelihood Estimation

Definition. *The **relative likelihood function** is defined as*

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \quad for\ \theta \in \Omega$$

Note that $0 \leq R(\theta) \leq 1$ *for all* $\theta \in \Omega$.

Definition. *The **log likelihood function** is defined as*

$$l(\theta) = \ln L(\theta) \quad for\ \theta \in \Omega$$

1.1 Likelihood Function for Binomial Distribution

The maximum likelihood estimate of θ is $\bar{\theta} = y/n$.

1.2 Likelihood Function for Poisson Distribution

The value $\theta = \bar{y}$ maximizes $l(\theta)$ and so $\hat{\theta} = \bar{y}$ is the maximum likelihood estimate of θ.

1.3 Likelihood Function for Exponential Distribution

The value $\theta = \bar{y}$ maximizes $l(\theta)$ and so $\hat{\theta} = \bar{y}$ is the maximum likelihood estimate of θ for an Exponential Distribution $\sim Exp(\theta)$.

1.4 Likelihood Function for Gaussian Distribution

The maximum likelihood estimate of θ is $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$, where

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \quad \text{and} \quad \hat{\sigma} = \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}$$

Note that $\hat{\sigma} \neq \sigma$ (sample variance).

1.5 Invariance Property of Maximum Likelihood Estimates

Theorem. *If $\hat{\theta}$ is the maximum likelihood estimate of θ, then $g(\hat{\theta})$ is the maximum likelihood estimate of $g(\theta)$.*

2 Estimation

2.1 Confidence Intervals and Pivotal Quantities

In general, construct a pivot using the estimator, use that to construct coverage interval, estimate it and find the confidence interval.

Definition. A $100p\%$, where $0 \leq p \leq 1$, confidence interval tells $100p\%$ of the intervals constructed from samples will contain the true unknown value of μ (or σ).

Theorem. Central Limit Theorem
If n is large, and if $Y_1, \ldots, Y_n$ are drawn from a distribution with mean μ and variance σ^2, then $\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

For a Binomial Distribution, the confidence interval is

$$\left[\hat{\pi} \pm z^* \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

where $\hat{\pi} = \frac{y}{n}$, y is the observed data.
To determine the sample size

$$n \geq \left(\frac{z^*}{MoE}\right)^2 \hat{\pi}(1-\hat{\pi})$$

where MoE is the margin of error.
To be conservative, we usually pick $\hat{\pi} = 0.5$ as it maximizes $\hat{\pi}(1-\hat{\pi})$.

For a Poisson Distribution, the pivotal quantity is

$$\frac{\bar{Y} - \mu}{\sqrt{\frac{\bar{Y}}{n}}} = Z \sim N(0, 1)$$

and the confidence interval is

$$\left[\bar{y} \pm z^* \sqrt{\frac{\bar{y}}{n}}\right]$$

2.2 Chi-Squared Distribution $\sim X_k^2$

The Gamma function is

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy \quad \text{for } \alpha > 0$$

Properties of the Gamma Function:

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\Gamma(\alpha) = (\alpha - 1)!$
- $\Gamma(1/2) = \sqrt{\pi}$

The X_k^2 distribution is a continuous family of distributions on $(0, \infty)$ with probability density function

$$f(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad \text{for } x > 0$$

where $k \in \{1, 2, \dots\}$ is a parameter of the distribution.
k is referred to as the "degrees of freedom" (d.f) parameter.

For $X \sim X_k^2$

- $E(X) = k$ and $Var(X) = 2k$
- If $k = 1$, $X = Z^2$ and $Z \sim G(0, 1)$
- If $k = 2$, $X \sim Exp(2)$ $(\theta = 2)$
- If k is large, $X \overset{Appr.}{\sim} N(k, 2k)$
- Let X_{k_1}, X_{k_1} be independent random variables with $X_{k_i} \sim X_{k_i}^2$.
  Then $X_{k_1} + X_{k_2} = X_{k_1+k_2}^2$.

Theorem. If $Y_i \sim Exp(\mu)$, then

$$\frac{2Y_i}{\mu} \sim Exp(2) \to X_2^2$$

2.3 Student's t Distribution

Student's t distribution has probability density function

$$f(t;k) = c_k \left(1 + \frac{t^2}{k}\right)^{-(k+1)/2} \qquad \text{for } t \in \Re \text{ and } k = 1, 2, \ldots$$

where the constant c_k is given by

$$c_k = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})} \quad k \text{ is the } degrees\ of\ freedom$$

Properties of T:

i) Range of T: $(-\infty, \infty)$

ii) T is symmetric around 0

iii) As $k \uparrow$, $T \to Z$

Theorem. *Suppose $Z \sim G(0,1)$ and $U \sim X_k^2$ independently. Let*

$$T = \frac{Z}{\sqrt{U/k}}$$

$$\to \frac{\bar{Y} - M}{s/\sqrt{n}} \sim t_{n-1}$$

*Then T has **Student's t distribution with** k* **degrees of freedom**.

2.4 Likelihood-Based Confidence Intervals

Theorem. *A $100p\%$ likelihood interval is an approximate $100q\%$ where $q = 2P(Z \leq \sqrt{-2\ln p}) - 1$ and $Z \sim N(0,1)$.*

Example 2.1. Show that a 1% likelihood interval is an approximate 99.8% confidence interval.
Note that $p = 0.01$

$$\begin{aligned} q &= 2P(Z \leq \sqrt{-2\ln(0.01)}) - 1 \\ &\approx 2P(Z \leq 3.03) - 1 \\ &= 2(0.99878) - 1 \\ &= 0.998 = 99.8\% \end{aligned}$$

Theorem. *If a is a value such that*

$$P = 2P(Z \le a) - 1 \ \ \textit{where } Z \sim N(0,1)$$

then the likelihood interval $\{\theta : R(\theta) \ge e^{-a^2/2}\}$ *is an approximate* $100p\%$ *confidence interval.*

Example 2.2. Since

$$0.95 = 2P(Z \le 1.96) - 1 \text{ where } Z \sim N(0,1)$$

and

$$e^{-(1.96)^2/2} = e^{-1.9208} \approx 0.1465 \approx 0.15$$

therefore a 15% *likelihood interval* for θ is also an approximate 95% *confidence interval* for θ.‘’

2.5 Confidence Intervals for Parameters in the $G(\mu, \sigma)$ Model

If $Y_1, \ldots, Y_n$ are independent $N(\mu, \sigma^2)$, then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$ and

(1) $$\frac{\bar{Y} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

(2) $$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

General Rule:
The Confidence Interval for μ if σ is unknown is

$$\left[\ \bar{y} \pm t^* \frac{s}{\sqrt{n}}\ \right]$$

When σ is unknown, we replace σ by its estimate s, and we use t-pivot.
Confidence interval when σ is known is

$$\left[\ \bar{y} \pm z^* \frac{\sigma}{\sqrt{n}}\ \right]$$

When σ is known, we use z-pivot.

If n is really large, then the t^* value converges to the corresponding z^* value (by Central Limit Theorem).

Confidence Intervals for σ^2 and σ

Theorem. *Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from the $G(\mu, \sigma)$ distribution with sample variance S^2. Then the random variable*

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

has a Chi-squared distribution with $n-1$ degrees of freedom.

Using the theorem, we can construct a $100p\%$ confidence interval for the parameter σ^2 or σ.
Recall this is the same as the equation (2) in this sub-section.
We can find constants a and b such that

$$P(a \leq U \leq b) = p$$

where $U \sim X^2_{n-1}$.
So a $100p\%$ confidence interval for σ^2 is

$$\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$$

and a $100p\%$ confidence interval for σ is

$$\left[\sqrt{\frac{(n-1)s^2}{b}}, \sqrt{\frac{(n-1)s^2}{a}}\right]$$

Unlike confidence interval for μ, the confidence interval for σ^2 is *not symmetric* about s^2. the estimator of σ^2. The X^2_{n-1} distribution is not a symmetric distribution.

Prediction Interval for a Future Observation
Suppose that $Y \sim G(\mu, \sigma)$, then

$$Y - \widetilde{\mu} = Y - \bar{Y} \sim N\left(0, \sigma^2\left(1 + \frac{1}{n}\right)\right)$$

Also

$$\frac{Y - \bar{Y}}{S\sqrt{1 + \frac{1}{n}}} \sim t_{n-1}$$

is a pivotal quantity which can be used to obtain an interval of values for Y.
Let a be a value such that $P(-a \leq T \leq a) = p$ or $P(T \leq a) = (1+p)/2$ which is obtained from tables. Thus

$$\left[\bar{y} \pm as\sqrt{1 + \frac{1}{n}}\right]$$

3 Tests of Hypothesis

Definition. A *hypothesis* in statistic is a claim made about the values of a certain parameter of the population.

There are **two** competing hypotheses:

- *Null* Hypothesis, denoted H_0; current "status quo" assumption.
- *Alternative* Hypothesis, denoted H_1; seeks to challenge H_0.

Definition. A *test statistic* or *discrepancy measure* D is a function of the data $\mathbf{Y}$ that is constructed to measure the degree of "agreement" between the data $\mathbf{Y}$ and the null hypothesis H_0.

For every testing decision, there is a possibility of making two kinds of errors:

Type I H_0 is true; H_0 is rejected.

Type II H_1 is true; H_0 is not rejected.

If Type I error goes down, then Type II error goes up; vice versa holds as well.

3.1 p-value

Suppose there's the test statistic $D = D(\mathbf{Y})$ to test the hypothesis H_0.
Also suppose that $d = D(\mathbf{y})$ is the observed value of D.

Definition. A *p-value* or observed significance level of the test of hypothesis H_0 using test statistic D is

$$\textit{p-value} = P(D \geq d; H_0)$$

Caution: The *p-value* is **not** the probability that H_0 is true.

Table 1: Interpretation of *p-values*

p-value	**Interpretation**
$p\text{-}value > 0.1$	No evidence against H_0 based on the observed data.
$0.05 < p\text{-}value \leq 0.10$	Weak evidence against H_0 based on the observed data.
$0.01 < p\text{-}value \leq 0.05$	Evidence against H_0 based on the observed data.
$0.001 < p\text{-}value \leq 0.01$	Strong evidence against H_0 based on the observed data.
$p\text{-}value \leq 0.001$	Very strong evidence against H_0 based on the observed data.

If the *p-value* is not small, it **cannot be concluded that H_0 is true**. It can only be said that there is **no evidence against the null hypothesis in light of the observed data**.

Confidence Interval vs. Hypothesis Testing
Confidence interval is the range of "reasonable" values for θ, given the level of confidence and sample data.
Hypothesis testing tests whether a particular value of θ is "reasonable" given the *p-value* and sample data.

3.2 Tests of Hypotheses for Parameters in the $G(\mu, \sigma)$ Model

Hypothesis Tests for μ
Using the test statistic

$$D = \frac{|\bar{Y} - \mu_0|}{S/\sqrt{n}}$$

Then using the sample mean $\bar{y}$ and standard deviation s, we get

$$d = \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}$$

The *p-value* can be then obtained via

$$p\text{-}value = P(D \geq d)$$

$$
\begin{aligned}
&= P(|T| \geq d) \\
&= 1 - P(-d \geq T \geq d) \\
&= 2[1 - P(T \leq d)] \quad \text{where } T \sim t_{n-1}
\end{aligned}
$$

One-sided hypothesis tests

Suppose that the null hypothesis is $H_0 : \mu = \mu_0$ and the alternative hypothesis is $H_1 : \mu > \mu_0$.

To test $\mu = \mu_0$, use the same test statistic and observed value. Then *p-value* can be obtained via

$$
\begin{aligned}
p\text{-}value &= P(D \geq d) \\
&= P(T \geq d) \\
&= 1 - P(T \leq d) \quad \text{where } T \sim t_{n-1}
\end{aligned}
$$

Relationship Between Hypothesis Testing and Interval Estimation

Suppose $y_1, y_2, \ldots, y_n$ is an observed random sample from the $G(\mu, \sigma)$ distribution.

Suppose $H_0 : \mu = \mu_0$ is tested, and we have

$$
\begin{aligned}
&p\text{-}value \geq 0.05 \\
&\text{if and only if } P\left(\frac{|\bar{Y} - \mu_0|}{S/\sqrt{n}} \geq \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}; H_0 : \mu = \mu_0 \text{is true}\right) \geq 0.05 \\
&\text{if and only if } P\left(|T| \geq \underbrace{\frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}}_{\text{b}}\right) \geq 0.05 \quad \text{where } T \sim t_{n-1} \\
&\text{if and only if } P\left(|T| \leq \underbrace{\frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}}_{\text{a}}\right) \leq 0.95 \\
&\text{if and only if } \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}} \leq a \quad \text{where } P(|T| \leq a) = 0.95 \\
&\text{if and only if } \mu_0 \in \left[\ \bar{y} - a\frac{s}{\sqrt{n}}, \bar{y} + a\frac{s}{\sqrt{n}}\ \right]
\end{aligned}
$$

which is a 95% confidence interval for μ.

In general, suppose we have data $\mathbf{y}$, a model $f(\mathbf{y}, \theta)$ and we use the same pivotal quantity to construct a confidence interval for θ and a test of the hypothesis $H_0 : \mu = \mu_0$.

Then the parameter value $\theta = \theta_0$ is inside a $100q\%$ confidence interval for θ if and only if the *p-value* for testing $H_0 : \mu = \mu_0$ is greater than $1 - q$.

The disadvantage is that we need to construct the appropriate test statistics D and that may be difficult if the original distribution is complicated.

Hypothesis tests for σ
For testing $H_0 : \sigma = \sigma_0$, use the test statistic

$$U = \frac{(n-1)S^2}{\sigma_0^2}$$

Note that for large values of U and small values of U provide evidence against H_0 due to the asymmetric shape of Chi-squared distributions.
To approximate the *p-value*:

1. Let $u = (n-1)s^2/\sigma_0^2$ denote the observed value of U from the data

2. If u is large (that is, if $P(U \leq u) > 0.5$) compute the *p-value* as

$$p\text{-}value = 2P(U \geq u)$$

   where $U \sim \chi^2_{n-1}$

3. If u is small (that is, if $P(U \leq u) < 0.5$) compute the *p-value* as

$$p\text{-}value = 2P(U \leq u)$$

   where $U \sim \chi^2_{n-1}$

3.3 Likelihood Ratio Tests of Hypotheses - One Parameter

When a pivotal quantity does not exist then a general method for finding a test statistic with good properties can be based on the likelihood function.

Theorem. Suppose

$$\begin{aligned} \theta &= \text{ unknown parameter} \\ n &= \text{ sample size} \\ \hat{\theta} &= \text{ MLE for } \theta \\ \widetilde{\theta} &= \text{ Maximum Likelihood Estimator} \end{aligned}$$

$$
\begin{aligned}
H_0 &: \ \theta = \theta_0 \\
H_1 &: \ \theta \neq \theta_0
\end{aligned}
$$

Then for large n, the Likelihood Ratio Test Statistic is

$$
\begin{aligned}
\Lambda(\theta_0) &= -2 \ln \frac{L(\theta_0)}{L(\widetilde{\theta})} \sim X_1^2 \\
\Lambda(\theta_0) &= 2[L(\widetilde{\theta}) - L(\theta_0)]
\end{aligned}
$$

Using the observed value of $\Lambda(\theta_0)$, denoted by

$$
\lambda(\theta_0) = -2 \ln \left[\frac{L(\theta_0)}{L(\hat{\theta})} \right] = -2 \ln R(\theta_0)
$$

where $R(\theta_0)$ is the relative likelihood function evaluated at $\theta = \theta_0$.
The *p-value* can then be approximated via

$$
\begin{aligned}
p\text{-}value &\approx P[W \geq \lambda(\theta_0)] \quad \text{where } W \sim \chi_1^2 \\
&= P\left(|Z| \geq \sqrt{\lambda(\theta_0)}\right) \quad \text{where } Z \sim G(0,1) \\
&= 2\left[1 - P(Z \leq \sqrt{\lambda(\theta_0)}\right]
\end{aligned}
$$

3.3.1 Likelihood Ratio Test Statistic for Binomial

$$
\lambda(\theta_0) = -2 \ln \left[\left(\frac{\theta_0}{\hat{\theta}} \right)^y \left(\frac{1-\theta_0}{1-\hat{\theta}} \right)^{n-y} \right]
$$

where $\hat{\theta} = y/n$

3.3.2 Likelihood Ratio Test Statistic for Exponential

Suppose $y_1, y_2, \ldots, y_n \sim$ Exponential(θ)

$$
\lambda(\theta_0) = -2 \ln \left[\left(\frac{\hat{\theta}}{\theta_0} \right)^n e^{n(1-\hat{\theta}/\theta_0)} \right]
$$

3.3.3 Likelihood Ratio Test Statistic and $G(\mu, \sigma)$

Suppose $Y \sim G(\mu, \sigma)$ with p.d.f.

$$f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$

Then the likelihood ratio test statistic is

$$\Lambda(\theta_0) = \left(\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}\right)^2$$

Notice that $\Lambda(\theta_0)$ is the square of the standard Normal Distribution random variable

$$\frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$

Therefore, it has exactly a χ_1^2 distribution.