STAB57: An Introduction to Statistics

Shahriar Shams

Week 9 (Likelihood Ratio Test, Goodness of Fit Test)

Winter 2023

Recap of Week 8

- Comparing two independent Normal Populations:
  - equality of two variances
  - equality of two means (variances known)
  - equality of two means (variances unknown)
- Comparing two population means (paired data)

Learning goals for this week

- Likelihood ratio test (LRT)
    - LRT for single population
    - LRT for two populations
    - Confidence Interval using LRT
- Wald and Score test
- Goodness of Fit (GOF) test

Section 1

Likelihood Ratio Test (LRT) [Rice-P339]

Likelihood Ratio Test (LRT)-general definition

- Suppose we are testing $H_0 : \theta \in \Omega_0$ vs. $H_1 : \theta \in \Omega_1$
- Let $L(\theta)$ represent the likelihood function.
- The generalized likelihood ratio is defined as

$$\Lambda^* = \frac{max_{\theta \in \Omega_0}[L(\theta)]}{max_{\theta \in \Omega_1}[L(\theta)]}$$

- A "small" value of Λ^* provides evidence against H_0
- Finding a distribution of Λ^* might be very difficult

LRT: a special case

- A special case of the test statistic

$$\Lambda = \frac{max_{\theta \in \Omega_0}[L(\theta)]}{max_{\theta \in \Omega}[L(\theta)]} = \frac{max_{\theta \in \Omega_0}[L(\theta)]}{L(\hat{\theta})}$$

  where $\hat{\theta}$ is MLE of θ

- If $\hat{\theta} \in \Omega_0$ then $\Lambda = 1 \implies$ We will not reject H_0
- If $\hat{\theta} \notin \Omega_0$, we look for the most likely θ value in Ω_0 and check if it does a "good enough" job as it is done by the MLE.
- Λ value closer to 0 will provide evidence against the H_0

Theorem assigning a distribution to LRT

- Let, p=dim Ω be the number of free parameters in the whole parameter space.
- d= dim Ω_0 be the number of free parameters under the null.
- then we have this following result

  $$-2ln\Lambda \xrightarrow{D} \chi^2_{df=p-d}$$

  when H_0 is true.
- The proof is "out of scope" for the text book and for our course.
- We will only do examples where Ω_0 is a single point (like θ_0)

Subsection 1

LRT for single population

Example of LRT: Normal distribution with known σ

$(X_1, X_2, ..., X_n) \overset{iid}{\sim} N(\mu, \sigma_0^2)$ with σ_0^2 known
Test $H_0 : \mu = \mu_0$ at level of significance, α

1. $L(\mu) = (2\pi\sigma_0^2)^{-n/2} exp(-\frac{1}{2\sigma_0^2}\sum(x_i - \mu)^2)$
2. Under H_0, $L(\mu_0) = (2\pi\sigma_0^2)^{-n/2} exp(-\frac{1}{2\sigma_0^2}\sum(X_i - \mu_0)^2)$
3. The denominator, we have to maximize $L(\mu)$. We know $L(\mu)$ is maximized at $\bar{X}$
4. which gives $L(\hat{\mu}) = (2\pi\sigma_0^2)^{-n/2} exp(-\frac{1}{2\sigma_0^2}\sum(X_i - \bar{X})^2)$
5. Therefore,
$$\Lambda = \frac{L(\mu_0)}{L(\hat{\mu})}$$
6. Ω has 1 parameter and under H_0 it's 0 $\implies p - d = 1$

Example of LRT (cont...)

Recall: $\sum_i (X_i - \mu)^2 = \sum_i (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$ [week-3]

Continuing from previous slide,

$$
\begin{aligned}
\Lambda &= exp(-\frac{1}{2\sigma_0^2}[\sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2]) \\
\implies -2ln\Lambda &= \frac{1}{\sigma_0^2}[\sum (X_i - \mu_0)^2 - \sum (X_i - \bar{X})^2] \\
&= \frac{1}{\sigma_0^2} n(\bar{X} - \mu_0)^2 \\
&= \left(\frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt(n)} \right)^2 \sim \chi^2_{(df=1)}
\end{aligned}
$$

We reject H_0 if $-2ln\Lambda > \chi^2_{1-\alpha,(df=1)}$

LRT for Non-Normal distributions

- LRT allows us to test hypothesis for Non-normal distributions since all we need is the likelihood function evaluated at θ_0 and $\hat{\theta}$
- Here is an example I copied from the lecture slides of Alex Stringer, Assistant Professor, University of Waterloo.

Suppose we have patients arriving at a hospital waiting room, randomly. We can model their wait times X_i according to an exponential distribution,

$$X_i \sim Exp(\theta), E(X) = \theta$$

The hospital claims that the average waiting time is 60 minutes. We go on a randomly selected day and observe that $n = 100$ patients have an average wait time of $\bar{x} = 75$ minutes.
Is the hospitals claim supported by the data?

LRT for $Exp(\theta)$

The hypothesis we wish to test is

$$H_0 : \theta = 60$$
$$H_1 : \theta \neq 60$$

The likelihood is

$$L(\theta|\mathbf{x}) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right)$$

and the MLE is $\bar{X}$.

LRT for $Exp(\theta)$ (cont...)

The likelihood ratio is then

$$\Lambda = \left(\frac{\bar{x}}{\theta_0}\right)^n \exp\left(n\left(1 - \frac{\bar{x}}{\theta_0}\right)\right)$$

and the test statistic is

$$-2\log\Lambda = -2n\left(\log\bar{x} - \log\theta_0 + 1 - \frac{\bar{x}}{\theta_0}\right) \sim \chi^2_1$$

With $\theta_0 = 60$, $n = 100$ and $\bar{x} = 75$, we evaluate

$$-2\log\Lambda = -2(100)\left(\log 75 - \log 60 + 1 - \frac{75}{60}\right) = 5.37$$

which we compare to $\chi^2_{1,0.95} = 3.84$.
Because $5.37 > 3.84$, we reject H_0 at the 5% significance level.

LRT for $Exp(\theta)$ (cont...)

We can also compute the p-value of this test. The p-value is the probability of observing a result with as much or greater evidence against H_0 if H_0 is true. If H_0 is true, then $-2\log\Lambda \sim \chi_1^2$, so

$$p - value = P(\chi_1^2 > 5.37) = 0.02$$

Another example

$(4, 10, 10, 4, 6, 8, 8, 3, 4, 4) \stackrel{iid}{\sim} Pois(\lambda)$. Test $H_0 : \lambda = 5$

1. $-2ln\Lambda = -2ln(0.3230648) = 2.259805$
2. $\chi^2_{0.95,df=1} = 3.841459$
3. Since $-2ln\Lambda < \chi^2_{0.95,df=1}$ we fail to reject H_0
4. Calculate the p-value...

Subsection 2

LRT for two populations

Comparing two Normal populations

- Suppose we have two independent Normal samples

$$X_1, X_2, ...X_n \sim N(\mu_x, \sigma_x^2)$$

  and

$$Y_1, Y_2, ..., Y_m \sim N(\mu_y, \sigma_y^2)$$

  where σ_x^2 and σ_y^2 are known.

- We want to test $H_0 : \mu_x = \mu_y$ using LRT

Likelihood of the full data

- Since σ_x^2 and σ_y^2 are known, we have two unknown parameters: μ_x and μ_y
- Under H_0, these two parameters are equal. So we can write $\mu_x = \mu_y = \mu \implies$ Number of parameters under H_0 is one.
- Unrestricted likelihood,

$$
\begin{aligned}
L(\mu_x, \mu_y) =& (2\pi\sigma_x^2)^{-n/2} exp(-\frac{1}{2\sigma_x^2}\sum(X_i - \mu_x)^2)* \\
& (2\pi\sigma_y^2)^{-m/2} exp(-\frac{1}{2\sigma_y^2}\sum(Y_i - \mu_y)^2)
\end{aligned}
$$

- Finding the MLE of μ_x and μ_y and plugging into this equation gives us $L(\hat{\mu}_x, \hat{\mu}_y)$

Likelihood under $H_0 : \mu_x = \mu_y = \mu$

- Under H_0, μ is the only unknown parameter.
- The likelihood func.

$$
\begin{aligned}
L(\mu) = &(2\pi\sigma_x^2)^{-n/2} exp(-\frac{1}{2\sigma_x^2}\sum(X_i - \mu)^2) * \\
&(2\pi\sigma_y^2)^{-m/2} exp(-\frac{1}{2\sigma_y^2}\sum(Y_i - \mu)^2)
\end{aligned}
$$

- Finding the MLE of μ and plugging into this equation gives us $L(\hat{\mu})$

Test statistic and distribution

- Test statistic,

$$-2ln\Lambda = -2ln\frac{L(\hat{\mu})}{L(\hat{\mu}_x, \hat{\mu}_y)}$$

- Under H_0, $-2ln\Lambda \sim \chi^2_{(df=1)}$

Numerical example

(16.27, 11.66, 14.05, 15.43, 18.74, 13.42, 17.39, 18.71, 11.18, 13.52, 16.74, 5.43, 16.45, 10.75, 19.06) $\sim N(\mu_x, \sigma_x = 3)$;
(10.89, 7.57, 15.39, 8.43, 12.33, 7.43, 5.56, 18.07, 0.35, 7.62) $\sim N(\mu_y, \sigma_y = 4)$

1. $\hat{\mu}_x = \bar{x} = 14.587$ and $\hat{\mu}_y = \bar{y} = 9.364$
2. $\hat{\mu} = 13.162$
3. $L(\hat{\mu}) = 2.098396 * 10^{-34}$
4. $L(\hat{\mu}_x, \hat{\mu}_y) = 1.033094 * 10^{-31}$
5. Test statistic $= -2ln\Lambda = 12.398 \implies p - val = 0.00043$

Subsection 3

Constructing Confidence Interval using LRT

CI using LRT

- Under H_0, $-2ln\Lambda \xrightarrow{D} \chi^2_{df=p-d}$
- We reject H_0 if $-2ln\Lambda > \chi^2_{1-\alpha,(df=p-d)}$
- Conversely, we will fail to reject if $-2ln\Lambda < \chi^2_{1-\alpha,(df=p-d)}$
- Therefore $(1-\alpha)$ level CI for θ is the interval of θ values for which

$$
\begin{aligned}
&-2ln\Lambda < \chi^2_{1-\alpha,(df=p-d)} \\
\implies & L(\theta) > L(\hat{\theta}) * exp\left(-\frac{\chi^2_{1-\alpha,(df=p-d)}}{2}\right)
\end{aligned}
$$

- For the hospital waiting room example, 95% CI for θ is the solution of $-2(100)\left(\log 75 - \log\theta + 1 - \frac{75}{\theta}\right) < 3.84$ which is $(62.037, 91.841)$

Section 2

Wald and Score Test

Large sample property of MLE

We know from week- 4 and 5 that

$$\frac{\hat{\theta} - \theta_0}{\sqrt{1/nI(\theta_0)}} \xrightarrow{D} N(0, 1)$$

and

$$\frac{S(\theta_0)}{\sqrt{nI(\theta_0)}} \xrightarrow{D} N(0, 1)$$

Wald test

A common test statistic proposed by Abraham Wald,

$$\frac{\hat{\theta} - \theta_0}{SE[\hat{\theta}]} \xrightarrow{D} N(0,1)$$

Wald proposed the use of observed-fisher information to estimate $SE[\hat{\theta}]$.

Observed Fisher Information (E&R page 364)

$$= -\frac{\partial^2}{\partial \theta^2} \log f(X_1, X_2, ..., X_n|\theta)\Big|_{\theta=\hat{\theta}}$$

(in the expression of the second-derivative of the negative log-likelihood replace θ by $\hat{\theta}$)

Testing $\theta = \theta_0$ for Bernoulli dist using Wald test

- Suppose $X_1, X_2, ...X_n \overset{iid}{\sim} Bern(\theta)$
- $l(\theta) = \sum X_i log\theta + (n - \sum X_i) log(1-\theta)$
- $l'(\theta) = S(\theta) = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta} \implies \hat{\theta} = \bar{X}$
- $l''(\theta) = -\frac{\sum X_i}{\theta^2} - \frac{n - \sum X_i}{(1-\theta)^2}$
- Obs. Fisher Info $= -l''(\theta)|_{\theta=\bar{X}} = \frac{n}{\bar{X}} + \frac{n}{1-\bar{X}} = \frac{n}{\bar{X}(1-\bar{X})}$
- Wald Test Stat, $\frac{\bar{X} - \theta_0}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \xrightarrow{D} N(0,1)$

Score test

Score test uses the property of

$$\frac{S(\theta_0)}{\sqrt{nI(\theta_0)}} \xrightarrow{D} N(0,1)$$

In the denominator, we calculate the Fisher Information under the null hypothesis.

Testing $\theta = \theta_0$ for Bernoulli dist using Score test

- Suppose $X_1, X_2, ...X_n \overset{iid}{\sim} Bern(\theta)$
- $l(\theta) = \sum X_i log\theta + (n - \sum X_i) log(1-\theta)$
- $l'(\theta) = S(\theta) = \frac{\sum X_i}{\theta} - \frac{n - \sum X_i}{1-\theta}$
- $S(\theta_0) = \frac{\sum X_i}{\theta_0} - \frac{n - \sum X_i}{1-\theta_0} = \frac{n(\bar{X} - \theta_0)}{\theta_0(1-\theta_0)}$
- $l''(\theta) = -\frac{\sum X_i}{\theta^2} - \frac{n - \sum X_i}{(1-\theta)^2}$
- Fisher Info $= -E[l''(\theta)]|_{\theta=\theta_0} = \frac{n}{\theta_0} + \frac{n}{1-\theta_0} = \frac{n}{\theta_0(1-\theta_0)}$

- Score Test Stat, $\frac{\bar{X} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \xrightarrow{D} N(0,1)$

Comments on Wald, Score and LRT

- Computationally, Wald test is easier to conduct.
- For large n, all three of these tests perform the same.
- For small n, LRT is preferred over the others.

- You will find these three tests in almost all the advanced courses

Section 3

Goodness of Fit test

Chi-sq goodness of fit test [E&R-P490]

- This test is used to assess whether or not a categorical random variable W, which takes finite values $\{1, 2, ..., k\}$, has a specified probability measure P.
- When we have discrete random variable which takes infinitely many values, we partition the possible values into k categories.
- When we have a continuous random variable we partition the real line into k sub-intervals.
- Naturally, the counts of these k categories form a multinomial distribution.

Test statistic and distribution

- Let $X_1, X_2, ..., X_k$ be the observed counts of category $1, 2, ..., k$ respectively.
- We can write,

$$(X_1, X_2, ..., X_k) \sim Multinomial(n, p_1, p_2, ..., p_k)$$

- We know, $E[X_i] = np_i$ for $i = 1, 2, ..., k$
- Test statistic,

$$X^2 = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \xrightarrow{D} \chi^2_{(df=k-1)}$$

- It is recommended to ensure that $E[X_i] = np_i \geq 1$ for every i

Proof for a simple case (k=2)

We want to show

$$X^2 = \sum_{i=1}^{2} \frac{(X_i - np_i)^2}{np_i} \xrightarrow{D} \chi^2_{(df=1)}$$

Homework...

Re-wording the test statistic

we can write the same test as following:

$$X^2 = \sum_{i=1}^{k} \frac{(\text{Observed count of i} - \text{expected count of i})^2}{\text{expected count of i}} \xrightarrow{D} \chi^2_{(df=k-1)}$$

Numerical example [Example 9.1.7, E&R-P491]

Suppose we have 10000 random numbers generated from a $Uniform[0, 1]$ distribution. After dividing them into 10 equal length bins we have these following counts.

i	1	2	3	4	5	6	7	8	9	10
x_i	993	1044	1061	1021	1017	973	975	965	996	955

Test if these numbers look uniform or not.

Numerical example (cont...)

- If the numbers are really from a $Uniform[0,1]$ distribution then expected counts for each cell is $10000 * \frac{1}{10} = 1000$
- So we have,

i	1	2	3	4	5	6	7	8	9	10
Observed	993	1044	1061	1021	1017	973	975	965	996	955
Expected	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

- test stat,
  $X^2 = \left(\frac{(993-1000)^2}{1000} + \frac{(1044-1000)^2}{1000} + ... + \frac{(955-1000)^2}{1000} \right) = 11.056$
- p-value = 1-pchisq(11.056,df=9) = 0.27189
- We don't have any evidence to reject the statement that these number are from a $Uniform[0,1]$ distribution.
- In naive words, "they look uniform".

Some comments on the Uniform example

- The choice of 10 bins is completely arbitrary.
- We could have picked 15 or 20 (or some other number) equal bins. The process remains same just the degrees of freedom will be different.
- Not necessarily we need cells with equal probabilities.

- Ques: what if we wanted to check if the numbers are from a $Unif[0, \theta]$ distribution, where θ is unknown...

Theorem 9.1.2 [E&R-P493]

- If $p_1, p_2, ..., p_k$ are **unknown** then we need to estimate them.
- Under H_0 these will be functions of the associated parameter (θ).
- In this case,

$$(X_1, X_2, ..., X_k) \sim Multinomial\,(n, p_1(\theta), p_2(\theta), ..., p_k(\theta))$$

- After estimating θ by $\hat{\theta}$, Test statistic,

$$X^2 = \sum_{i=1}^{k} \frac{(X_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} \xrightarrow{D} \chi^2_{(df=k-1-dim(\Omega))}$$

where $dim(\Omega)$ represents the number of parameters needed to be estimated based on the data in order to calculate the p_i's

Example [Example 9.1.8, E&R-P493]

Testing for exponentiality

Suppose life-lengths of light bulbs (Y_i) follows an $Exponential(\beta)$, where β is unknown. We have the partitions as

$$(0,1],(1,2],(2,3],(3,\infty)$$

Based on sample of size, $n = 30$, the observed counts are 5,16,8,1
H_0 : The true model is $Exponential(\beta)$

- The first goal of this problem is to guess the best β (finding the MLE).

Exponential example continues..

- We can find θ using two different approaches:
    - if the life-lengths of the 30 bulbs were available (lets call them $Y_1, Y_2, ..., Y_{30}$) then

$$L(\beta) = \beta^{30} exp[-\beta \sum y_i] \implies \hat{\beta} = \frac{1}{\bar{y}}$$

    - if all we have is the counts of Y_i's that fall into those four partitions (which is the case in this example), we can define,

$$L(\beta) = (1 - e^{-\beta})^5 (e^{-\beta} - e^{-2\beta})^{16} (e^{-2\beta} - e^{-3\beta})^8 (e^{-3\beta})^1$$

      where, $(1 - e^{-\beta}) = P(Y_i \in (0, 1])$, similarly the other terms. Solving it numerically(using any software), $\hat{\beta} = 0.603535$

Exponential example continues...

- Using $\hat{\beta} = 0.603535$ we can calculate,

$$
\begin{aligned}
p_1 &= 1 - e^{-0.603535} && = 0.453125 \\
p_2 &= e^{-0.603535} - e^{-2*0.603535} && = 0.247803 \\
p_3 &= ... && = 0.135517 \\
p_4 &= ... && = 0.163555
\end{aligned}
$$

- Expected counts: $30 * 0.453125 = 13.59375$, similarly the other three: $7.43409, 4.06551, 4.90665$
- Test stat $= \frac{(5-13.59375)^2}{13.59375} + \frac{(16-7.43409)^2}{7.43409} + ... \approx 22.22$
- p-val=1-pchisq(22.22,df=2) $= 0.000015$
- We reject $\mathrm{H}_0 \implies$ We have strong evidence against $Exp(\beta)$ being the true model for the these data.

Explanation of the $L(\beta)$ from slide 40

- Let's take an example of the first number (Y_1).
- If we knew the actual observed number y_1 (say 0.78) then the contribution of Y_1 in the likelihood function is $f_\beta(y_1) = \beta exp(-\beta * y_1)$
- If we don't observe the actual y_1 rather all we know is $Y_1 \in (0, 1]$ then the contribution of Y_1 in the likelihood function is $P(Y_1 \in (0, 1]) = F_\beta(1) - F_\beta(0) = 1 - e^{-\beta}$
- There are 5 numbers that belong to this range, hence together their contribution is $(1 - e^{-\beta})^5$
- Similarly the other terms...

Homework (Non-credit)

Evans and Rosenthal

Exercise (without the part on residuals) 9.1.5 , 9.1.6, R(9.1.25)

John A. Rice

Exercise 9: 36, 37, 40, 43, 44