stats 135 final review

Haoyu Chen

- ★ Simple random sampling
- ★ Distribution derived from the norm
- ★ Estimation of parameters
- ★ Hypotheses testing and assessing goodness of fit
- ★ Summarizing data
- ★ Two samples
- ★ Linear regression

The most important parts

Hypotheses testing

The decision of whether to reject H_0 is determined by whether the sample $X = (X_1, \ldots, X_n)$ falls into a predefined rejection region R.

Usually, the rejection region R has the form

$$R = \{x_1, \ldots, x_n : T(x_1, \ldots, x_n) > c\}$$

where T is called a test statistic and c is called a critical value.

The idea is to construct R so that the probability of the data falling into it when H_0 is true is small.

Hypotheses testing

H_0 null hypothesis; H_0 alternative hypothesis

	H_0 is True	H_1 is True
Reject H_0	type I error	Case1
Accept H_0		type II error

P(type I error) = significance level of the test (alpha)
P(type II error) = beta
power of test = P(Case 1) = 1 - beta = P(X in Reject region)

p-value

Suppose that for every $\alpha \in (0,1)$ we have a size α test with rejection region R_α. When $R_\alpha = \{x : T(x) \geq c_\alpha\}$,

$$\text{p-value} = \sup_{\theta \in \Theta_0} P_\theta(T(X) \geq T(x))$$

where x is the observed data.

Therefore, the p-value is the probability under H_0 of observing a value $T(X)$ the same as or more extreme than what was actually observed.

Warning! A large p-value is not strong evidence in favor of H_0. A large p-value can occur for two reasons: (i) H_0 is true or (ii) H_0 is false but the test has low power.

But do not confuse the p-value with $\mathbb{P}(H_0|\text{Data})$. **The p-value is not the probability that the null hypothesis is true.**

Hypotheses testing

Simple likelihood ratio test

$$H_0\colon \mu = \mu_0$$
$$H_A\colon \mu = \mu_1$$

$$\Lambda = \frac{lik(\theta_0)}{lik(\theta_1)}$$

$$\{\Lambda < c\}$$

Hypotheses testing

General likelihood ratio test

$$H_0 : \theta \in \omega_0 \text{ vs } H_1 : \theta \in \omega_1$$

$$\Lambda = \frac{\max\limits_{\theta \in \omega_0}[\text{lik}(\theta)]}{\max\limits_{\theta \in \Omega}[\text{lik}(\theta)]}$$

reject region: $R = \{X : \Lambda(X) < c\}$

$$\text{power} \;= P(X \in R | H_1) = P(\frac{\max_{\theta \in \omega_0} lik(\theta)}{\max_{\theta \in \Omega} lik(\theta)} < c | \theta \in \omega_1)$$

$$\alpha = P(X \in R | H_0) = P(\frac{\max_{\theta \in \omega_0} lik(\theta)}{\max_{\theta \in \Omega} lik(\theta)} < c | \theta \in \omega_0)$$

For equvalent reject region, if $\Lambda(X)$ can be written as $g(f(X))$ and function g is monotanical increase/decrease, R is equvalent to $\{X : f(X) < / > c\}$

Asymptotic distribution

$$-2 \log \Lambda \sim \chi^2(dim(\Omega) - dim(\omega_0))$$

Hypotheses testing

The multinomial distribution

$$H_0 : p = p(\theta) \text{ vs } H_1 : \theta \in \omega_1$$

$$\text{under } \Omega \quad \hat{p}_i = \frac{x_i}{n} \qquad \dim \Omega = m - 1$$

likelihood ratio

$$\Lambda = \prod_{i=1}^{m} \left(\frac{p_i(\hat{\theta})}{\hat{p}_i} \right)^{x_i}$$

$$\begin{aligned} -2 \log \Lambda &= -2n \sum_{i=1}^{m} \hat{p}_i \log \left(\frac{p_i(\hat{\theta})}{\hat{p}_i} \right) \\ &= 2 \sum_{i=1}^{m} O_i \log \left(\frac{O_i}{E_i} \right) \end{aligned}$$

where $O_i = n\hat{p}_i$ and $E_i = np_i(\hat{\theta})$

Pearson's chi-square test

$$X^2 = \sum_{i=1}^{m} \frac{[x_i - np_i(\hat{\theta})]^2}{np_i(\hat{\theta})}$$

Hypotheses testing

duality of confidence interval and hypothesis test

THEOREM **A**

Suppose that for every value θ_0 in Θ there is a test at level α of the hypothesis H_0: $\theta = \theta_0$. Denote the acceptance region of the test by $A(\theta_0)$. Then the set

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

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

THEOREM **B**

Suppose that $C(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence region for θ; that is, for every θ_0,

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

Then an acceptance region for a test at level α of the hypothesis H_0: $\theta = \theta_0$ is

$$A(\theta_0) = \{\mathbf{X} | \theta_0 \in C(\mathbf{X})\}$$

Hypotheses testing

duality of confidence interval and hypothesis test

THEOREM **A**

Suppose that for every value θ_0 in Θ there is a test at level α of the hypothesis H_0: $\theta = \theta_0$. Denote the acceptance region of the test by $A(\theta_0)$. Then the set

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

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

THEOREM **B**

Suppose that $C(\mathbf{X})$ is a $100(1-\alpha)\%$ confidence region for θ; that is, for every θ_0,

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

Then an acceptance region for a test at level α of the hypothesis H_0: $\theta = \theta_0$ is

$$A(\theta_0) = \{\mathbf{X} | \theta_0 \in C(\mathbf{X})\}$$

Condition	Test	Statistics	Reject rejion
data draw from one norm sample, var known	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$\lvert z \rvert \geq u_{1-\alpha/2}$
data draw from one norm sample, var unknown	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$ $s = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$	$\lvert t \rvert \geq t_{1-\alpha/2}$
data draw from two paired norm samples	$H_0: \mu_D = \mu_x - \mu_y = 0$ $H_1: \mu_D \neq 0$	$t = \bar{D}/s_{\bar{D}} \sim t(n-1)$ $s_{\bar{D}}^2 = \frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$	$\lvert t \rvert \geq t_{1-\alpha/2}$
two sample draw from two norm, but var is known	$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$\lvert z \rvert \geq u_{1-\alpha/2}$

Condition	Test	Statistics	Reject rejion
two sample draw from two norm, var is unknown, but two var are the same	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} \sim t(n_1 + n_2 - 2)$ $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$\lvert t \rvert \geq t_{1-\alpha/2}$
two sample draw from two norm, var is unknown, but two var are not the same	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(n_{fd})$ $n_{fd} = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 / \left(\frac{s_1^2}{n_1^2(n_1 - 1)} + \frac{s_2^2}{n_2^2(n_2 - 1)}\right)$	$\lvert t \rvert \geq t_{1-\alpha/2}$
two sample draw from two norm, test on var	$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	$F = \frac{s_1^2}{s_2^2} \sim F(n_1 - 1, n_2 - 1)$ $s_1^2 \geq s_2^2$	$F \geq F_{1-\alpha/2}$

Q-Q plot

- always non-decreasing
- two distributions being compared are identical, the Q–Q plot follows the 45° line $y = x$
- follows some line, if two distribution are linearly transforming
- If the general trend of the Q–Q plot is flatter than the line y = x, the distribution plotted on the horizontal axis is more dispersed than the distribution plotted on the vertical axis.
- Q–Q plots are often arced, or "S" shaped, indicating that one of the distributions is more skewed than the other, or that one of the distributions has heavier tails than the other.
- common use of Q–Q plots is to compare the distribution of a sample to a theoretical distribution, such as the standard normal distribution N(0,1)

Parametric and nonparametric bootstrap

Parametric

estimate $\hat{\theta}$ from $\{x_1, ..., x_n\}$

For i in range(B):

$$\text{draw}\{x_1, ..., x_m\} \text{ from } F_{\hat{\theta}}$$

$$\text{compute } t_i = T(x_1, ..., x_m)$$

use $\{t_i\}$ to do task

Non-parametric

For i in range(B):

$$\text{draw } \{x_1, ..., x_m\} \text{ from } \{x_1, ..., x_n\} \text{with replacement}$$

$$\text{compute } t_i = T(x_1, ..., x_m)$$

use $\{t_i\}$ to do task

$$V_F[T_n] \underbrace{\approx}_{\substack{\text{ECDF;} \\ \text{depends on } n}} V_{\hat{F}_n}(T_n) \underbrace{\approx}_{\substack{\text{MC integration;} \\ \text{depends on } B}} \widehat{V}_{\hat{F}_n}(T_n)$$

Non-parametric method

$$X_1, X_2, \dots, X_{n_1}, Y_1, Y_2, \dots, Y_{n_2}$$

$$Z_1, Z_2, \dots, Z_{n_1+n_2}.$$

$$Z_{(1)} \le Z_{(2)} \le \dots \le Z_{(n_1+n_2)}$$

$$H_0 \colon F = G.$$

Rank sum

$$R_1 := \sum_{i=1}^{n_1} \mathsf{rank}(X_i) \quad \text{respectively} \quad R_2 := \sum_{i=1}^{n_2} \mathsf{rank}(Y_i).$$

$$\mathbb{E}R_1 = \frac{n_1(n_1+n_2+1)}{2}$$

$$\mathrm{Var}(R_1) = \frac{n_1 n_2(n_1+n_2+1)}{2}.$$

Mann-Whitney test

Under H_0 we have $\pi = \frac{1}{2}$. $\quad \pi := \mathbb{P}\{X < Y\}$

$$U_Y := n_1 n_2 \hat{\pi} = \#\{(X_i, Y_j) \colon X_i < Y_j\} = R_2 - \frac{n_2(n_2+1)}{2}.$$

$$\mathbb{E}U_Y = \frac{n_1 n_2}{2}, \qquad \mathrm{Var}(U_Y) = \mathrm{Var}(R_2) = \frac{n_1 n_2(n_1+n_2+1)}{2}.$$

$$\frac{U_Y - \mathbb{E}U_Y}{\sqrt{\mathrm{Var}(U_Y)}} \to N,$$

Simple linear regression

$$S(\beta_0,\ \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Some basic properties of β_0 and β_1:

1. They are unbiased: $E[\hat{\beta}_0] = \beta_0$ and $E[\hat{\beta}_1] = \beta_1$.
2. They are consistent: $\hat{\beta}_0 \xrightarrow{P} \beta_0$ and $\hat{\beta}_1 \xrightarrow{P} \beta_1$.
3. They are asymptotically normal:

$$\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \sim t_{n-2}$$

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{n\sigma^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{e}_i^2$$

Examples1

Suppose that $X_1, \ldots, X_n$ form a random sample from a uniform distribution on the interval $(0, \theta)$, and that the following hypotheses are to be tested:

$$H_0: \quad \theta \geq 2$$
$$H_1: \quad \theta < 2$$

Let $Y_n = max\{X_1, \ldots, X_n\}$, and consider a test whose rejection region contains all the outcomes for which $Y_n \leq 1.5$.

(a) Determine the power function of the test.

(b) Determine the size of the test.

Examples2

Let $X_1, \ldots, X_n$ be *iid* with density $f(x;\beta) = \beta e^{-\beta x}$ for $x > 0$ and $\beta > 0$. Find the asymptotic (large sample) likelihood ratio test of size α for $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$.

Examples3

Suppose $X_1, \dots, X_n \overset{iid}{\sim} Poisson(\lambda)$. Let $\lambda_0 > 0$. Consider testing $H_0 : \lambda \leq \lambda_0$ versus $H_1 : \lambda > \lambda_0$.

(a) Show that the likelihood ratio test rejects H_0 when $S \equiv \sum_{i=1}^{n} X_i > c$ for some constant c. (You don't need to show what c is yet; this will depend on the size of the test.) You may use the fact that the MLE for λ is $\bar{X}_n = S/n$ and the restricted MLE under $\lambda \leq \lambda_0$ is $\min\{\lambda_0, \bar{X}_n\}$.

(b) What is the power function of the likelihood ratio test with rejection region $\{x : S = \sum_{i=1}^{n} x_i > c\}$? You may use the fact that if $X_1, \dots, X_n \overset{iid}{\sim} Poisson(\lambda)$, then $S \equiv \sum_{i=1}^{n} X_i \sim Poisson(n\lambda)$.

Examples4

Suppose $X_1, \ldots, X_n$ are iid with PDF

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Find the likelihood ratio test statistic $T(X) = \frac{\sup_{\theta \in \Theta} \mathcal{L}_n(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}_n(\theta)}$. Hint: Consider separately the cases $\min\{X_1, \ldots, X_n\} \leq \theta_0$ and $\min\{X_1, \ldots, X_n\} > \theta_0$.

Exa
mpl
e 5

(20 pts) Suppose we take a random sample of size n from a bag of colored balls (red, blue and yellow balls). Let X_1 denote the number of red balls, X_2 denote the number of blue balls, and X_3 denote the number of yellow balls in the sample. Assuming we know that the total number of yellow balls is twice more than the total number of red balls in the bag. Or in other words, the red, blue and yellow balls occur with probability p_1, p_2 and $p_3 = 3p_1$, respectively in the bag.

(a) Find the method of moments estimate of p_2.

(b) Find the maximum likelihood estimator for p_1, p_2 and p_3 under the above model.

(c) Find the aymptotic distribution (after appropriate normalization) for the MLEs in (b).

(d) Construct the likelihood ratio test statistic for the null hypothesis that $p_1 = p_2 = p_3/3$. What is the asymptotic distribution of your test statistic under null?

Example 6

(Multinomial tests of homogeneity) When Jane Austen died, her novel Sandition was incomplete. Someone else finished the novel and it was published. Morton (1978) examined word frequencies to see if the new author was distinguishable from Austen. The data are as follows:

Word	Sense and Sensibility	Emma	Sandition I (Austen)	Sandition II (New author)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Totals	375	440	202	196

Treat each column as an independent Multinomial sample.

(a) Construct the likelihood ratio statistic and calculate the p-value for the null hypothesis that the first three columns (by Austen) have the same set of probabilities for each column.

(b) Now sum the first three columns to give a single column of counts for Austen and another for the new author. Construct the likelihood ratio statistic and calculate the p-value for the null hypothesis that probabilities are the same across authors.

(a) Define the variables corresponding to the first three columns as follows:

Word	Sense and Sensibility	Emma	Sandition I (Austen)
a	X_{11}	X_{12}	X_{13}
an	X_{21}	X_{22}	X_{23}
this	X_{31}	X_{32}	X_{33}
that	X_{41}	X_{42}	X_{43}
with	X_{51}	X_{52}	X_{53}
without	X_{61}	X_{62}	X_{63}
Totals	n_1	n_2	n_3

The corresponding probabilities are in the following table.

Word	Sense and Sensibility	Emma	Sandition I (Austen)
a	p_{11}	p_{12}	p_{13}
an	p_{21}	p_{22}	p_{23}
this	p_{31}	p_{32}	p_{33}
that	p_{41}	p_{42}	p_{43}
with	p_{51}	p_{52}	p_{53}
without	p_{61}	p_{62}	p_{63}
Totals	1	1	1

The null hypothesis is

$$H_0 : p_{1i} = p_{2i} = p_{3i}, \quad i = 1, \ldots, 6.$$

$$H_1 : p_{1i}, p_{2i} \text{ and } p_{3i} \text{ are not all equal for some } i.$$

The unconstrained MLEs are

$$\hat{p}_{ij} = X_{ij}/n_i, \quad i = 1, \ldots, 6; j = 1, 2, 3.$$

Under H_0, the constrained MLEs are

$$\hat{p}_{0i1} = \hat{p}_{0i2} = \hat{p}_{0i3} = \frac{X_{i1} + X_{i2} + X_{i3}}{n_1 + n_2 + n_3}, \quad i = 1, \ldots, 6.$$

The likelihood ratio statistic is

$$\lambda = 2 \log \frac{L(\hat{p})}{L(\hat{p}_0)} = 2 \sum_{i=1}^{6} \sum_{j=1}^{3} X_{ij} \log \frac{\hat{p}_{ij}}{\hat{p}_{0ij}} = 12.587,$$

which has an asymptotic χ^2_{10} under H_0. The degrees of freedom is $(6 - 1) \times 3 - (6 - 1) = 10$.

The p-value is $P(\lambda > 12.587) \approx 0.248$.

(b) Now we want to compare the works by Austen to works of the new author.
The variables we are looking at are

Word	Austen	New Author
a	$X_1 = X_{11} + X_{12} + X_{13}$	X_1^*
an	$X_2 = X_{21} + X_{22} + X_{23}$	X_2^*
this	$X_3 = X_{31} + X_{32} + X_{33}$	X_3^*
that	$X_4 = X_{41} + X_{42} + X_{43}$	X_4^*
with	$X_5 = X_{51} + X_{52} + X_{53}$	X_5^*
without	$X_6 = X_{61} + X_{62} + X_{63}$	X_6^*
Totals	$n = n_1 + n_2 + n_3$	n^*

The corresponding probabilities are

Word	Austen	New Author
a	p_1	p_1^*
an	p_2	p_2^*
this	p_3	p_3^*
that	p_4	p_4^*
with	p_5	p_5^*
without	p_6	p_6^*
Totals	1	1

The null hypothesis is

$$H_0 : p_i = p_i^*, \quad i = 1, \ldots, 6.$$

$$H_1 : p_i \neq p_i^* \quad \text{for some } i.$$

The unconstrained MLEs are

$$\hat{p}_i = X_i/n, \quad \hat{p}_i^* = X_i^*/n^* \quad i = 1, \ldots, 6.$$

Under H_0, the constrained MLEs are

$$\hat{p}_{0i} = \hat{p}_{0i}^* = \frac{X_i + X_i^*}{n + n^*}, \quad i = 1, \ldots, 6.$$

The likelihood ratio statistic is

$$\lambda = 2 \log \frac{L(\hat{p})}{L(\hat{p}_0)} = 2 \left(\sum_{i=1}^{6} X_i \log \frac{\hat{p}_i}{\hat{p}_{0i}} + \sum_{i=1}^{6} X_i^* \log \frac{\hat{p}_i^*}{\hat{p}_{0i}^*} \right) = 31.737,$$

which has an asymptotic χ_5^2 under H_0. The degrees of freedom is $(6 - 1) \times 2 - (6 - 1) = 5$.
The p-value is $P(\lambda > 31.737) \approx 6.699 \times 10^{-6}$.

Example 7

Consider the **regression through the origin** model:

$$Y_i = \beta X_i + \epsilon_i$$

(a) Find the least squares estimate for β.

(b) Find the standard error of the estimate.

(c) Find conditions that guarantee that the estimator is consistent.

Example 8

Let's consider adding some penalty on the parameter of simple linear regression. Thus, we consider the loss function: $l(\beta_0, \beta_1) = \sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2 + \lambda(\beta_0^2 + \beta_1^2)$, where β_0, β_1 are parameter we want to estimate, and λ is a known parameter.

a) Derive the optimal straight line under this new loss.

b) Assume β_i has a prior distribution $N(0, \tau^2)$, and $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, show that the mle estimator of β_i for posterior distribution is the same as the results you got from a), where $\lambda = \sigma^2/\tau^2$.

soln

6. (a) $\beta(\theta) = P_\theta(\{X_1, \ldots, X_n\} \in R)$

$= P_\theta(\max\{X_1, \ldots, X_n\} \leq 1.5)$

$= P_\theta(X_i \leq 1.5,\ i = 1, \ldots, n)$

$= P_\theta(X \leq 1.5)^n$, where $X \sim Unif(0, \theta)$

$$= \begin{cases} 1 & \theta \leq 1.5 \\ \left(\frac{1.5}{\theta}\right)^n & \theta > 1.5 \end{cases}$$

(b) $\alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$

$= \sup_{\theta \geq 2} \left(\frac{1.5}{\theta}\right)^n$

$= \left(\frac{1.5}{2}\right)^n = \left(\frac{3}{4}\right)^n$, since $\left(\frac{1.5}{\theta}\right)^n$ is strictly decreasing w/ θ

soln

Let $X_1, \ldots, X_n$ be *iid* with density $f(x;\beta) = \beta e^{-\beta x}$ for $x > 0$ and $\beta > 0$. Find the asymptotic (large sample) likelihood ratio test of size α for $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$.

Solutions:

The MLE of β can be calculated as follows.

$$L_n(\beta) = \prod_{i=1}^{n} f(X_i;\beta) = \beta^n e^{-\beta \sum_{i=1}^n X_i}.$$

$$\Rightarrow \ell_n(\beta) = \log L_n(\beta) = n \log \beta - \beta \sum_{i=1}^{n} X_i.$$

$$\Rightarrow \frac{d\ell_n}{d\beta}(\beta) = \frac{n}{\beta} - \sum_{i=1}^{n} X_i.$$

$$\frac{d\ell_n}{d\beta}(\beta) = 0 \Rightarrow \hat{\beta} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}.$$

$$\frac{d^2\ell_n}{d\beta^2}(\beta) = -\frac{n}{\beta^2} < 0.$$

So the MLE is $\hat{\beta} = \frac{n}{\sum_{i=1}^n X_i}$.

The likelihood ratio statistic is

$$\lambda = 2 \log \frac{L_n(\hat{\beta})}{L_n(\beta_0)} = 2(\ell_n(\hat{\beta}) - \ell_n(\beta_0)) = 2\left[-n \log \bar{X}_n - n - \left(n \log \beta_0 - n\beta_0 \bar{X}_n\right)\right]$$
$$= 2n\left[\beta_0 \bar{X}_n - \log(\beta_0 \bar{X}_n) - 1\right].$$

The asymptotic distribution of λ is χ_1^2. So the rejection region is

$$\lambda > \chi^2_{1,\alpha},$$

where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$th quantile of χ^2_1 distribution.

(a) The likelihood ratio test statistic is

$$
\begin{aligned}
T(x) &= \frac{sup_\lambda \mathcal{L}_n(\lambda)}{\sup_{\lambda \le \lambda_0} \mathcal{L}_n(\lambda)} \\
&= \frac{\mathcal{L}_n(S/n)}{\mathcal{L}_n(\min\{\lambda_0, S/n\})} \\
&= \begin{cases} \mathcal{L}_n(S/n)/\mathcal{L}_n(\lambda_0), & S > n\lambda_0 \\ 1, & S \le n\lambda_0 \end{cases}
\end{aligned}
$$

The form of the rejection region for the LRT is $\{x : T(x) > b\}$ for some constant b. (By definition, $T(x) \ge 1$, so we only need to consider tests with $b > 1$. Otherwise the test will always reject.) The question asks us to show that an equivalent form of the rejection region is $\{x : S = \sum_{i=1}^n x_i > c\}$ for some constant c. To do this, we just need to show that $T(x)$ is an increasing function of S when $S > n\lambda_0$.

$$
\begin{aligned}
\frac{\mathcal{L}_n(S/n)}{\mathcal{L}_n(\lambda_0)} &= \frac{e^{-n\frac{S}{n}}(S/n)^S}{e^{-n\lambda_0}\lambda_0^S} \\
&= e^{n\lambda_0 - S}(\frac{S}{n\lambda_0})^S \\
\log \frac{\mathcal{L}_n(S/n)}{\mathcal{L}_n(\lambda_0)} &= n\lambda_0 - S + S\log(\frac{S}{n\lambda_0}) \\
\frac{\partial}{\partial S}\log \frac{\mathcal{L}_n(S/n)}{\mathcal{L}_n(\lambda_0)} &= -1 + \log(\frac{S}{n\lambda_0}) + S\frac{n\lambda_0}{S}\frac{1}{n\lambda_0} \\
&= \log(\frac{S}{n\lambda_0})
\end{aligned}
$$

which is positive when $S > n\lambda_0$. Therefore, since $\log T(x)$ is increasing with S when $S > n\lambda_0$, so is $T(x)$, and so an equivalent way to express the rejection region is $\{x : S = \sum_{i=1}^n x_i > c\}$ for some constant c.

(b)

$$
\begin{aligned}
\beta_c(\lambda) &= P_\lambda(S > c) \\
&= 1 - P_\lambda(S \le c) \\
&= 1 - \sum_{k=0}^{\lfloor c \rfloor} \frac{e^{-n\lambda}(n\lambda)^k}{k!}
\end{aligned}
$$

where $\lfloor c \rfloor$ is the largest integer less than or equal to c.

soln

Solution:

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \prod_{i=1}^{n} e^{-(X_i-\theta)} I(X_i \geq \theta) = e^{-\sum_{i=1}^{n} X_i + n\theta} I(\min\{X_1,\ldots,X_n\} \geq \theta)$$

Since $\mathcal{L}_n(\theta)$ is an increasing function in θ on $-\infty < \theta \leq \min\{X_1,\ldots,X_n\}$ and is zero for $\theta > \min\{X_1,\ldots,X_n\}$, $\sup_{\theta\in\Theta} \mathcal{L}_n(\theta) = \mathcal{L}_n(\min\{X_1,\ldots,X_n\})$.

When $\min\{X_1,\ldots,X_n\} \leq \theta_0$, $\sup_{\theta\in\Theta_0} \mathcal{L}_n(\theta) = \mathcal{L}_n(\min\{X_1,\ldots,X_n\})$.

When $\min\{X_1,\ldots,X_n\} > \theta_0$, $\sup_{\theta\in\Theta_0} \mathcal{L}_n(\theta) = \mathcal{L}_n(\theta_0)$.

Therefore, the likelihood ratio test statistic

$$T(X) = \begin{cases} 1 & \text{if } \min\{X_1,\ldots,X_n\} \leq \theta_0 \\ \frac{\mathcal{L}_n(\min\{X_1,\ldots,X_n\})}{\mathcal{L}_n(\theta_0)} = \exp\{n(\min\{X_1,\ldots,X_n\} - \theta_0)\} & \text{if } \min\{X_1,\ldots,X_n\} > \theta_0 \end{cases}$$

soln

(a) $\vec{X} := (X_1, X_2, X_3) \sim \text{Multinomial}(n, \vec{p})$ where $\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$.

$$f(\vec{x}; \vec{p}) = \binom{n}{x_1\ x_2\ x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$= \frac{n!}{x_1!\, x_2!\, x_3!} p_1^{x_1} (1-4p_1)^{x_2} (3p_1)^{x_3}$$

(b) Note that $X_2 \sim \text{Bin}(n, p_2)$ since each time a ball is drawn from the bag, there is p_2 chance that the ball is yellow. We repeat the draws n times and the draws are indept as the draws are done w/ replacement.

Thus, $EX_2 = np_2 \Rightarrow \hat{p}_{2,MOM} = \frac{X_2}{n}$.

(c) We take log of $f(\vec{x}; \vec{p})$ in part (a).

$$\ell(\vec{x}; \vec{p}) = \text{constant} + x_1 \log p_1 + x_2 \log(1-4p_1) + x_3(\log 3 + \log p_1)$$

$$\frac{\partial L}{\partial p_1} = \frac{x_1}{p_1} - \frac{4x_2}{1-4p_1} + \frac{x_3}{p_1} = \frac{x_1 + x_3 - 4p_1(x_1+x_3) - 4x_2 p_1}{p_1(1-4p_1)}$$

$$= \frac{x_1 + x_3 - 4np_1}{p_1(1-4p_1)}$$

$\frac{\partial L}{\partial p_1} = 0 \Rightarrow p_1 = \frac{x_1+x_3}{4n}$. $\frac{\partial L}{\partial p_1}$ exists $\forall p_1 \in (0, \frac{1}{4})$, so the only critical pt is $\frac{x_1+x_3}{4n}$. (Note p_1 cannot be greater than $\frac{1}{4}$; otherwise, $p_2 = 1 - 4p_1 < 0$.)

$\frac{\partial L}{\partial p_1}$ is positive for $p_1 \in (0, \frac{x_1+x_3}{4n})$, negative for $p_1 \in (\frac{x_1+x_3}{4n}, \frac{1}{4})$.

Thus, $\ell(\vec{x}; \vec{p})$ is increasing for $p_1 \in (0, \frac{x_1+x_3}{4n})$ and decreasing for $p_1 \in (\frac{x_1+x_3}{4n}, \frac{1}{4})$. Thus, the MLE for p_1 is

$$\hat{p}_1 = \frac{X_1+X_3}{4n}.$$

By the equivariance of MLE's the MLE's for p_2 & p_3 are

$$\hat{p}_2 = 1 - 4\frac{X_1+X_3}{4n} = \frac{X_2}{n}$$

$$\hat{p}_3 = 3\hat{p}_1 = \frac{3(X_1+X_3)}{4n}.$$

(d) $$\frac{\hat{p}_2 - p_2}{\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n}}} \xrightarrow{d} N(0,1)$$

#4. (e) In the set up of this problem, we have $p_3 = 3p_1$.

Then $H_0: \ p_1 = p_2 = \frac{p_3}{3}$

$H_1: \ p_1 \neq p_2, \quad p_1 = \frac{p_3}{3}$

Thus under H_0: $\begin{cases} p_1 + p_2 + p_3 = 1 \\ p_1 = p_2 = \frac{p_3}{3} \end{cases} \Rightarrow 5p_1 = 1 \Rightarrow p_1 = \frac{1}{5}$

$\Rightarrow p_2 = \frac{1}{5}, \quad p_3 = \frac{3}{5} \quad (df = 0)$

Under H_1, $p_1 \neq p_2$, $p_1 = \frac{p_3}{3}$, the MLEs are obtained in part (c).

$(df = 1)$

Hence,

$$T(x) = \frac{\sup_{\theta \in \Theta} L(\theta)}{\sup_{\theta \in \Theta_0} L(\theta)} = \frac{L(\hat{p}_1, \hat{p}_2, \hat{p}_3)}{L(\frac{1}{5}, \frac{1}{5}, \frac{3}{5})}$$

$$= \frac{\binom{n}{x_1, x_2, x_3} \left(\frac{x_1+x_3}{4n}\right)^{x_1} \left(\frac{x_2}{n}\right)^{x_2} \left(\frac{3(x_1+x_3)}{4n}\right)^{x_3}}{\binom{n}{x_1, x_2, x_3} \left(\frac{1}{5}\right)^{x_1} \left(\frac{1}{5}\right)^{x_2} \left(\frac{3}{5}\right)^{x_3}}$$

$$= \left(\frac{5(x_1+x_3)}{4n}\right)^{x_1} \left(\frac{5x_2}{n}\right)^{x_2} \left(\frac{5(x_1+x_3)}{4n}\right)^{x_3}$$

$$\lambda(x) = 2\log T(x) = 2\cdot\left(x_1 \cdot \log\left(\frac{5(x_1+x_3)}{4n}\right) + x_2 \log\left(\frac{5x_2}{n}\right) + x_3 \log\left(\frac{5(x_1+x_3)}{4n}\right)\right)$$

$$= 2\left[(x_1+x_3)\log\left(\frac{5(x_1+x_3)}{4n}\right) + x_2 \log\left(\frac{5x_2}{n}\right)\right].$$

The asymptotic distribution of $\lambda(x)$ is χ^2_1, since

$df = 1 - 0 = 1.$