



JOINT INSTITUTE
交大密西根学院

VE401 Probabilistic Methods in Eng. Solution Manual for RC 6

Chen Xiwen

May 10, 2020

Assignment 5.2

Let X_1, \dots, X_n be i.i.d. exponential random variables with parameter β . Recall that $Y = X_1 + X_2 + \dots + X_n$ follows a Gamma distribution with parameters $\alpha = n$ and β . Transform this expression further to yield a chi-squared random variable.

Let X be an exponential random variable with parameter β . Devise a test statistic for testing $H_0 : \beta = \beta_0$ and $H_0 : \beta \leq \beta_0$ in a Fisher test.

Solution. Since Y follows a Gamma distribution with parameters $\alpha = n$ and β , we have the density function

$$f_Y(y) = \frac{\beta^n}{\Gamma(n)} y^{n-1} e^{-\beta y}, \quad y > 0$$

and $f_Y(y) = 0$ when $y \leq 0$. Let $u = \varphi(y) = 2\beta y$, then

$$y = \varphi^{-1}(u) = \frac{u}{2\beta} \quad \Rightarrow \quad \frac{d}{du} \varphi^{-1}(u) = \frac{1}{2\beta}.$$

Then using transformation of variable, we have

$$\begin{aligned} f_U(u) &= f_Y \circ \varphi^{-1}(u) \cdot \left| \frac{d}{du} \varphi^{-1}(u) \right| \\ &= \frac{\beta^n}{\Gamma(n)} \frac{u^{n-1}}{(2\beta)^{n-1}} e^{-u/2} \cdot \frac{1}{2\beta} \\ &= \frac{1}{2^n \Gamma(n)} u^{n-1} e^{-u/2} \quad u > 0, \end{aligned}$$

and $f_U(u) = 0$ when $u \leq 0$, which is a chi-squared distribution with $2n$ degrees of freedom. Therefore, we have the distribution

$$Y \sim \frac{1}{2\beta} \chi_{2n}^2.$$

Given samples X_1, \dots, X_n with size n , we can use test statistic

$$\chi_{2n}^2 = 2\beta_0 \sum_{i=1}^n X_i = 2n\beta_0 \bar{X}.$$

We see that a maximum likelihood estimator for β is given by

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n f_{X_i}(x_i) = \beta^n e^{\beta \sum_{i=1}^n X_i} \quad \Rightarrow \quad \ell(\beta) = n \ln \beta - \beta \sum_{i=1}^n X_i \\ &\quad \Rightarrow \quad \hat{\beta} = \frac{1}{\bar{X}}. \end{aligned}$$

With larger true parameter β , we would expect a smaller test statistic. Or

$$2n\beta\bar{x} = 2n\beta_0\bar{x} \cdot \frac{\beta}{\beta_0} \leq 2n\beta_0\bar{x}$$

if $\beta \leq \beta_0$.

- For one-tailed test $\beta \leq \beta_0$, “more extreme data” means smaller \bar{X} . Therefore, the p-value is given by

$$P\text{-value} = F_{\chi_{2n}^2}(2n\beta_0\bar{X}),$$

where $F_{\chi_{2n}^2}$ is the cumulative distribution function of chi-squared distribution with $2n$ degrees of freedom.

- For two-tailed test $\beta = \beta_0$, the p-value is given by

$$P\text{-value} = 2 \min \left(F_{\chi_{2n}^2}(2n\beta_0\bar{X}), 1 - F_{\chi_{2n}^2}(2n\beta_0\bar{X}) \right).$$

If we want to have a critical region for the tests, with the test statistic χ_{2n}^2 defined above, we reject H_0 at significance level α

- $H_0 : \beta = \beta_0$ if $\chi_{2n}^2 < \chi_{2n,1-\alpha/2}^2$ or $\chi_{2n}^2 > \chi_{2n,\alpha/2}^2$,
- $H_0 : \beta \leq \beta_0$ if $\chi_{2n}^2 < \chi_{2n,1-\alpha}^2$,
- $H_0 : \beta \geq \beta_0$ if $\chi_{2n}^2 > \chi_{2n,\alpha}^2$.

Critical Region and Confidence Interval

Suppose we have a sample X_1, \dots, X_n of size n from a normal population X with mean μ and known variance σ^2 . We would like to test the hypotheses

$$H_0 : \mu \leq \mu_0, \quad H_1 : \mu > \mu_0.$$

Then for such one-tailed tests, what is the corresponding confidence interval? We know that we reject H_0 at significance level α if

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha,$$

which can be rewritten as

$$\mu_0 < \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Since we know that if the inequality above holds, we will reject H_0 , in which case the null value μ_0 falls outside of the confidence interval. Therefore, the corresponding confidence interval is

given by

$$\text{CI} = \left[\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty \right).$$

Similarly, if the null hypothesis is $H_0 : \mu \geq \mu_0$, the corresponding one-sided confidence interval is given by

$$\text{CI} = \left(-\infty, \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}} \right].$$

Mean and Variance for Rank Sum

Wilcoxon rank-sum test statistic. Let X_1, \dots, X_m and Y_1, \dots, Y_n , where $m \leq n$, be random samples from two continuous populations X and Y and associate the rank $R_i, i = 1, \dots, m+n$, to the R_i th smallest among the $m+n$ total observations. The test statistic is given by

$$W_m = \text{sum of the ranks of } X_1, \dots, X_m$$

with

$$\text{E}[W_m] = \frac{m(m+n+1)}{2}, \quad \text{Var}[W_m] = \frac{mn(m+n+1)}{12}.$$

Proof. The rank of the $m+n$ random variables follows a uniform discrete distribution. Namely,

$$P[R_i = k] = \frac{1}{m+n}, \quad \text{for } k = 1, \dots, m+n.$$

Therefore, the expectation of each rank is given by

$$\text{E}[R_i] = \frac{1}{m+n} \sum_{k=1}^{m+n} k = \frac{m+n+1}{2},$$

giving

$$\text{E}[W_m] = \text{E} \left[\sum_{i=1}^m R_i \right] = \frac{m(m+n+1)}{2}.$$

Denote $N = m+n$. We know that

$$\sum_{i=1}^N i^2 = \frac{N(N+1)(2N+1)}{6}, \quad \sum_{i=1}^N \sum_{j=1}^N ij = \left(\sum_{i=1}^N i \right)^2 = \frac{N^2(N+1)^2}{4}.$$

Therefore,

$$\sum_{i \neq j}^N ij = \frac{N^2(N+1)^2}{4} - \frac{N(N+1)(2N+1)}{6}.$$

By properties of variance, we have

$$\text{Var}[W_m] = \text{Var}\left[\sum_{i=1}^m R_i\right] = \sum_{i=1}^m \text{Var}[R_i] + \sum_{i \neq j}^m \text{Cov}[R_i, R_j],$$

where

$$\begin{aligned} \text{Var}[R_i] &= \text{E}[R_i^2] - \text{E}[R_i]^2 \\ &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{N^2 - 1}{12}, \\ \text{Cov}[R_i, R_j] &= \text{E}[R_i R_j] - \text{E}[R_i]\text{E}[R_j] \\ &= \sum_{i \neq j}^N \frac{ij}{N(N-1)} - \left(\frac{N+1}{2}\right)^2 \\ &= \frac{N(N+1)^2}{4(N-1)} - \frac{(N+1)(2N+1)}{6(N-1)} - \frac{(N+1)^2}{4} \\ &= -\frac{N+1}{12}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}[W_m] &= \sum_{i=1}^m \text{Var}[R_i] + \sum_{i \neq j}^m \text{Cov}[R_i, R_j] \\ &= \frac{m(N^2 - 1)}{12} - m(m-1) \cdot \frac{N+1}{12} \\ &= \frac{m(N-m)(N+1)}{12} \\ &= \frac{mn(m+n+1)}{12}. \end{aligned}$$

Comparison of Means

Exercise 1.

Suppose we have two normally distributed populations $X^{(1)}$ and $X^{(2)}$ with mean μ_1, μ_2 and variances σ_1^2, σ_2^2 , respectively. A sample of size $n = 20$ is gathered for each of these populations

	3.73	2.90	2.58	3.33	3.34
$X^{(1)}$	2.80	3.84	3.01	2.91	0.83
	4.54	3.49	1.12	0.78	0.67
	2.32	2.42	3.21	3.09	1.16

and

	1.30	2.55	3.03	1.71	2.45
$X^{(2)}$	1.58	2.68	1.07	2.45	2.72
	2.54	2.59	2.35	2.42	2.87
	4.13	1.73	2.42	3.03	1.23

We want to test the hypotheses

$$H_0 : \mu_1 = \mu_2, \quad H_1 : |\mu_1 - \mu_2| \geq 0.5,$$

with significance level $\alpha = 0.05$ in the following cases.

1. We know the variances are $\sigma_1^2 = \sigma_2^2 = 1$. Perform the test. What is the required sample size for the power of the test to be at least 80%?

Solution. We calculate

$$\bar{x}^{(1)} = 2.604, \quad \bar{x}^{(2)} = 2.343.$$

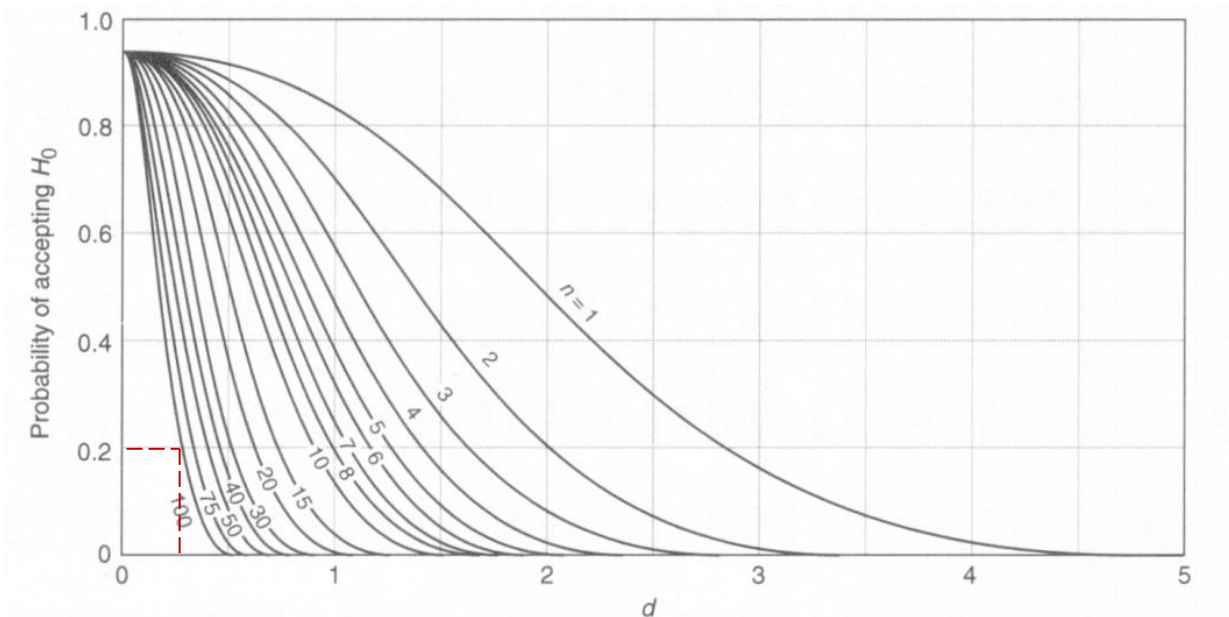
Then the test statistic is given by

$$z = \frac{\bar{x}^{(1)} - \bar{x}^{(2)}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{2.604 - 2.343}{\sqrt{1/20 + 1/20}} = 0.8254.$$

The critical value is given by $z_{\alpha/2} = 1.96 > z$. Therefore, we fail to reject H_0 . We calculate

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{2\sigma^2}} = 0.35,$$

and read from OC curve for normal tests. We would require a sample size of at least 100.



(a) OC curves for different values of n for the two-sided normal test for a level of significance $\alpha = 0.05$.

2. The variances are unknown but equal $\sigma^2 = \sigma_1^2 = \sigma_2^2$. Perform the test. What is the required sample size for the power of the test to be at least 70%?

Solution. We calculate the variances

$$s_1^2 = 1.263, \quad s_2^2 = 0.534, \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{s_1^2 + s_2^2}{2} = 0.899.$$

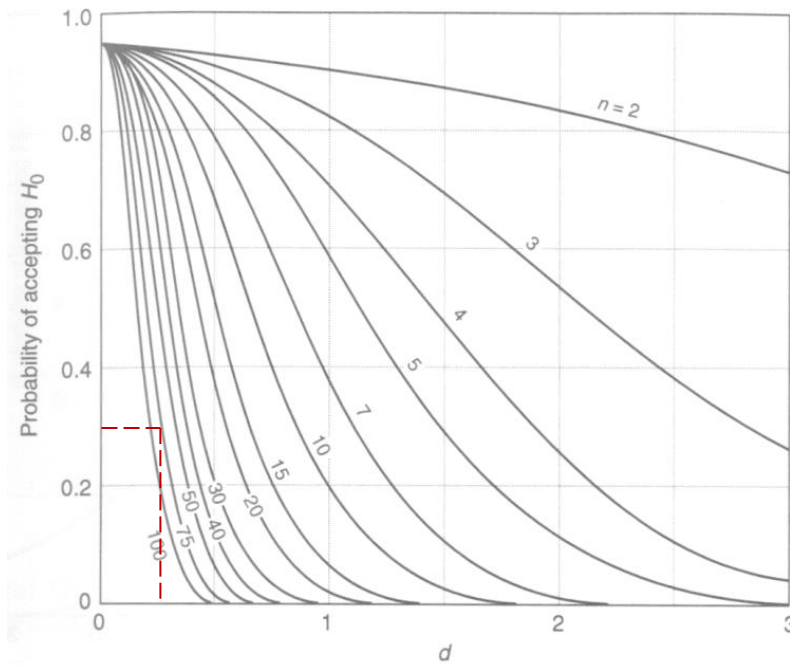
Then the test statistic is given by

$$t_{38} = \frac{\bar{x}^{(1)} - \bar{x}^{(2)}}{\sqrt{s_p^2(1/n_1 + 1/n_2)}} = 0.871.$$

The critical value is given by $t_{\alpha/2, 38} = 2.024 > t_{38}$. Therefore, we fail to reject H_0 . We calculate

$$d = \frac{|\mu_1 - \mu_2|}{2s_p} = \frac{0.5}{2\sqrt{0.899}} = 0.264,$$

where we estimate the variance using pooled variance, and read from OC curve for normal tests. We would require a modified sample size of at least $n^* = 2n - 1 = 75$, giving $n = 38$.



(e) OC curves for different values of n for the two-sided t test for a level of significance $\alpha = 0.05$.

3. The variances are unknown and not necessarily equal. Perform the hypothesis test.

Solution. We calculate

$$\gamma = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 32.64 \approx 32$$

and thus the test statistic

$$t_{32} = \frac{\bar{x}^{(1)} - \bar{x}^{(2)}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = 0.871$$

with critical value $t_{\alpha/2, 32} = 2.04 > t_{32}$. Therefore, we fail to reject H_0 .

Pearson Statistic

Exercise 2.

Let X denote the number of defective components from a pack of 12 such components. 100 such packs are observed, and the corresponding values of X are as follows.

X	0	1	2	3	4
Count	16	42	36	5	1

Perform a test to analyze whether a binomial distribution is an appropriate model for the distribution of X .

Solution. We set up the hypothesis

$H_0 : X$ follows a binomial distribution with parameters $n = 12$ and p ,

where p is unknown and requires an estimate

$$\hat{p} = \frac{1}{12 \times 100}(42 + 2 \times 36 + 3 \times 5 + 4 \times 1) = 0.1108.$$

Then we calculate

$$\begin{aligned} P[X = 0] &= \binom{12}{0}(1 - \hat{p})^{12} = 0.244, \\ P[X = 1] &= \binom{12}{1}\hat{p}(1 - \hat{p})^{11} = 0.365, \\ P[X = 2] &= \binom{12}{2}\hat{p}^2(1 - \hat{p})^{10} = 0.250, \\ P[X = 3] &= \binom{12}{3}\hat{p}^3(1 - \hat{p})^9 = 0.104, \\ P[X \geq 4] &= 1 - P[X < 4] = 0.036. \end{aligned}$$

Therefore, the observations and expectations can be listed as follows.

X	0	1	2	3	4
O	16	42	36	5	1
E	24.423	36.631	25.045	10.406	3.595

We check that Cochran's rule is satisfied. The Pearson statistic is then given by

$$\chi_3^2 = \sum_{i=1}^5 \frac{(O_i - E_i)^2}{E_i} = 13.197,$$

with critical value $\chi_{0.05,3} = 7.81 < \chi_3^2$ and $5 - 1 - 1 = 3$ degrees of freedom. Therefore, we reject H_0 with significance level 0.05. There is evidence that X does not follow a binomial distribution.