

# Financial Time Series I and Methods of Statistical Prediction

## Homework 2: Review on Basics

1. (a) We know that about 40% of the mathematicians in the United States are women. So the probability that there are 3 or fewer women in a randomly selected group of 11 mathematicians is

$$\sum_{x=0}^3 C_x^{11} \cdot 0.4^x \times 0.6^{11-x} = 0.29628$$

- (b) The command of this simulation is as follows.

```
> sum(rbinom(10000,11,0.4)<=3)/10000
```

2. (a) We need to do the test:

$H_0$ : the distribution of the two year is the same.

$H_a$ : the distribution of the two year is different.

We can use chi-square test here.

$$\sum_{i=1}^5 \frac{(o_i - e_i)^2}{e_i} = 9.7469 > \chi_{4,0.95}^2 = 9.49$$

We have enough evidence to reject  $H_0$  that we can say the two distribution may be different.

- (b) We need to test  $H_0 : \mu_1 = \mu_2$  vs.  $H_a : \mu_1 \neq \mu_2$

$$Z = \frac{2.975 - 2.97}{\sqrt{\frac{1.33^2}{7646} + \frac{1.242^2}{200}}} = 0.0561 < Z_{0.025} = 1.96$$

We don't have enough evidence to reject  $H_0$ . In this sample, the mean of 1998 may be the same as it in 1997.

- (c) In (b), we conclude that the mean of 1997 is equal to the mean of 1998 statistically. But by (a), the grade distribution of 1998 is somehow different from the grade distribution of 1997 in the following way. The performance of students in 1998 is more concentrated than what in 1998. There are less students who get poor grades, and more students get higher grades. At the same time, the proportion of students of getting the best grade decreases. This exercise just wants to remind you that the means of two populations are equal does not mean that their populations distribution are the same.

3. Those studies only tell us that there exists a negative correlation between hours spent watching TV and scored on reading tests. We can't make sure that there really exists a causation between these two. Maybe there exists a factor that it is correlative to hours spent watching TV and it will let people get a poor score on reading test. Actually, the factor affect the score on reading test but we don't know. The factor may be hours spent on reading. In order to find that watching TV would make people less able to read, we need to set up a well-done experiment that can control other effect.
4. If the height of a man's height is denoted by  $X$ , the height of his wife is  $Y = 0.92X$ . The correlation of their hight must be 1 because there exists a linear relationship between  $X$  and  $Y$ .

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X, 0.92X)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(0.92X)}} = \frac{0.92 \times \text{Cov}(X, X)}{0.92 \times \text{Var}(X)} = 1$$

5. (a) In this problem, we only have 10 rates instead of 15 rates. Then we observed all 10 survival times, (14, 17, 27, 18, 12, 8, 22, 13, 19, 12). Suppose the survival times follow exponential distribution  $f(x) = \frac{1}{\theta} \exp(-\frac{x}{\theta})$ ,  $x \geq 0$ .

- i. Maximum likelihood estimate

The likelihood function will be

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{10} f(x_i) = \prod_{i=1}^{10} \left( \frac{1}{\theta} \exp(-\frac{x_i}{\theta}) \right) \\ &= \left( \frac{1}{\theta} \right)^{10} \exp\left(-\frac{\sum_{i=1}^{10} x_i}{\theta}\right) \\ l(\hat{\theta}) &= \frac{d}{d\theta} L(\theta)|_{\theta=\hat{\theta}} = -10 \log \hat{\theta} - \frac{\sum_{i=1}^{10} x_i}{\hat{\theta}} = 0 \\ \hat{\theta} &= \frac{\sum_{i=1}^{10} x_i}{10} = \frac{162}{10} = 16.2 \end{aligned}$$

The MLE of mean survival time is  $\hat{\theta} = 16.2$ .

- ii. Method of moment

We need to calculate  $E(X)$  and sample mean.

$$E(X) = \theta, \frac{\sum_{i=1}^{10} x_i}{10} = 16.2$$

By the method of moment, the MME of mean survival time is  $\tilde{\theta} = 16.2$ .

- (b) We assume that we observe 10 survival times and 5 truncated times, (14, 17, 27, 18, 12, 8, 22, 13, 19, 12, 30<sup>+</sup>, 30<sup>+</sup>, 30<sup>+</sup>, 30<sup>+</sup>, 30<sup>+</sup>). We don't observe when these 5 rats die. Suppose the survival times follow exponential distribution  $f(x) = \frac{1}{\theta} \exp(-\frac{x}{\theta})$ ,  $x \geq 0$ . And define  $c_i$  is the index function of whether a data is truncated or not.

i. Maximum likelihood estimate

The likelihood function will be

$$\begin{aligned}
 L(\theta) &= \prod_{i=1}^{15} f(x_i)^{(1-c_i)} [1 - F(x_i)]^{c_i} \\
 &= \prod_{i=1}^{15} \left( \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) \right)^{(1-c_i)} \left( \exp\left(-\frac{x_i}{\theta}\right) \right)^{c_i} \\
 &= \left( \frac{1}{\theta} \right)^{10} \exp\left(-\frac{\sum_{i=1}^{15} x_i}{\theta}\right) \\
 l(\hat{\theta}) &= \frac{d}{d\theta} L(\theta) \Big|_{\theta=\hat{\theta}} = -10 \log \hat{\theta} - \frac{\sum_{i=1}^{15} x_i}{\hat{\theta}} = 0 \\
 \hat{\theta} &= \frac{\sum_{i=1}^{15} x_i}{10} = \frac{312}{10} = 31.2
 \end{aligned}$$

The MLE of mean survival time is  $\hat{\theta} = 31.2$ .

ii. Method of moment

$$E(X) = \theta = m_1 = m'_1 = \frac{\sum_{i=1}^{15} x_i}{15} = 20.8$$

The MME of mean survival time is  $\tilde{\theta} = 20.8$ .

6. (a) Assume a normal population. Because the sample size only is 20, we need to use t-distribution.

$$\begin{aligned}
 C.I &= \left[ \bar{x} \pm t_{0.025,19} \times \frac{s}{\sqrt{n}} \right] = \left[ 87.395 - 2.093 \times \frac{0.5175}{\sqrt{20}}, 87.395 + 2.093 \times \frac{0.5175}{\sqrt{20}} \right] \\
 &= [87.1528, 87.6372]
 \end{aligned}$$

- (b) The command of normal probability plot is as follows.

```
> qqnorm(data)
```

7. (a) Base on the definition of confidence interval, the distribution of  $Y$  is Binomial distribution with  $n = 100$  and  $p = 0.95$ .

(b)

$$\begin{aligned}
 P\{Y \geq 90\} &= P\left\{ \frac{Y - 95}{\sqrt{100 \times 0.95 \times 0.05}} \geq \frac{90 - 0.5 - 95}{\sqrt{100 \times 0.95 \times 0.05}} \right\} = P\{Z \geq -2.52\} = 0.9941 \\
 P\{Y \geq 95\} &= P\left\{ \frac{Y - 95}{\sqrt{100 \times 0.95 \times 0.05}} \geq \frac{95 - 0.5 - 95}{\sqrt{100 \times 0.95 \times 0.05}} \right\} = P\{Z \geq -0.22\} = 0.5871
 \end{aligned}$$

8.

$$E(X) = \int_0^{\infty} x \cdot \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot x^{\alpha-1} \exp\left[-\frac{x}{\beta}\right] dx = \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot x^{(\alpha+1)-1} \exp\left[-\frac{x}{\beta}\right] dx$$

$$\begin{aligned}
&= \frac{\beta}{\Gamma(\alpha)} \cdot \Gamma(\alpha + 1) \int_0^\infty \frac{1}{\Gamma(\alpha + 1)} \cdot \frac{1}{\beta^{\alpha+1}} \cdot x^{(\alpha+1)-1} \exp\left[-\frac{x}{\beta}\right] dx = \alpha\beta \\
E(X^2) &= \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot x^{\alpha-1} \exp\left[-\frac{x}{\beta}\right] dx = \int_0^\infty \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^\alpha} \cdot x^{(\alpha+2)-1} \exp\left[-\frac{x}{\beta}\right] dx \\
&= \frac{\beta^2}{\Gamma(\alpha)} \cdot \Gamma(\alpha + 2) \int_0^\infty \frac{1}{\Gamma(\alpha + 2)} \cdot \frac{1}{\beta^{\alpha+2}} \cdot x^{(\alpha+2)-1} \exp\left[-\frac{x}{\beta}\right] dx = \alpha(\alpha + 1)\beta \\
Var(X) &= E(X^2) - (E(X))^2 = \alpha(\alpha + 1)\beta - (\alpha\beta)^2 = \alpha\beta^2
\end{aligned}$$

9. We know that  $f_X(x) = (2\pi)^{-\frac{1}{2}} \cdot \exp(-\frac{x^2}{2})$  and  $Y = g(X) = \exp(X)$  which is a monotone increasing function.

$$\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \\
&= (2\pi)^{-\frac{1}{2}} \cdot \exp\left(-\frac{(\log y)^2}{2}\right) \cdot \frac{1}{y}
\end{aligned}$$