Probability Theory and Mathematical Statistics

Homework 3, Vitaliy Pozdnyakov

Part 1

1. Let $X \sim Pois(a)$, where a is unknown. Apply the maximal likelihood method to estimate a.

$L(x_1, \dots, x_n, a) = \prod_{i=1}^{n} \frac{a^{x_i} e^{-a}}{x_i!}$

$\ln L = \sum_{i=1}^{n} \ln \frac{a^{x_i} e^{-a}}{x_i!} = \sum_{i=1}^{n} \left(\ln a^{x_i} - \ln x_i! + \ln e^{-a}\right) = \sum_{i=1}^{n} (x_i \ln a - \ln x_i! - a)$

$\frac{\partial \ln L}{\partial a} = \sum_{i=1}^{n} \left(\frac{x_i}{a} - 1\right) = \frac{1}{a} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} 1 = \frac{1}{a} \sum_{i=1}^{n} x_i - n = 0$

$a = \frac{\sum x_i}{n} = \overline{X}$

2. Let $X \sim Bi(1, p)$, where p is unknown. Apply the maximal likelihood method to estimate p.

$X \sim Bi(1, p) = \text{Bernoulli}(p)$

$L(x_1, \dots, x_n, p) = p^m (1-p)^{n-m}$, where m is the number of attempts with $x_i = 1$

$\ln L = m \ln p + (n-m) \ln (1-p)$

$\frac{\partial \ln L}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p} = 0$

$p = \frac{m}{n}$

4. Let $X \sim U(a, b)$, where a and b are unknown. Apply the maximal likelihood method to estimate a and b.

Let I be an indicator:

$$I(\text{requirement}) = \begin{cases} 1 \text{ if satisfiable} \\ 0 \text{ otherwise} \end{cases}$$

$L(x_1, \dots, x_n, a, b) = (b-a)^{-n} \prod_{i=1}^{n} I(a \le x_i \le b) = (b-a)^{-n} I(a \le x_{\min}) I(x_{\max} \le b)$

where $x_{\min} = \min(x_1, \dots, x_n)$ and $x_{\max} = \max(x_1, \dots, x_n)$

Thereby if a or $b \in (x_{\min}, x_{\max})$ then $L = 0$

On the other hand, if $(b - a)$ will be increasing, then L will be decreasing.

Consequently $\max L = (x_{\max} - x_{\min})^{-n} \Rightarrow b = x_{\max}$ and $a = x_{\min}$

Part 2

1. Let $X \sim N(\mu, \sigma^2)$, where μ and σ^2 are unknown. Apply the method of moments to estimate μ and σ^2.

$\nu_1 = \int_{-\infty}^{+\infty} x^1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$ — the first theoretical raw moment

$\mu_2 = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$ — the second theoretical central moment

$\tilde{\nu}_1 = \frac{\sum x_i^1}{n} = \overline{X}$ — the first empirical raw moment

$\tilde{\mu}_2 = \frac{\sum (x_i - \overline{x})^2}{n} = s^2$ — the second empirical central moment

Thereby $\mu = \overline{X}$ and $\sigma^2 = s^2$

2. Let $X \sim U(a, b)$, where a and b are unknown. Apply the method of moments to estimate a and b.

$\nu_1 = \int_a^b x^1 \frac{b-a}{2} dx = \frac{1}{2}(b+a)$ — the first theoretical raw moment

$\nu_3 = \int_a^b x^3 \frac{b-a}{2} dx = \frac{1}{3}(a^2 + ab + b^2)$ — the third theoretical raw moment

$\tilde{\nu}_1 = \frac{\sum x_i^1}{n} = \overline{X}$ — the first empirical raw moment

$\tilde{\nu}_3 = \frac{\sum x_i^3}{n} = \overline{X^2} = s^2 + \overline{X}^2$ — the third empirical raw moment

$$\begin{cases} \frac{1}{2}(b+a) = \overline{X} \\ \frac{1}{3}(a^2 + ab + b^2) = s^2 + \overline{X}^2 \end{cases}$$

Let the solution be $a = \overline{X} - \sqrt{3}s$ and $b = \overline{X} + \sqrt{3}s$ and check:

$$\begin{cases} \frac{1}{2}(b+a) = \overline{X} \\ \frac{1}{3}(a^2+ab+b^2) = s^2 + \overline{X}^2 \end{cases} \Rightarrow \begin{cases} \frac{1}{2}(\overline{X} - \sqrt{3}s + \overline{X} + \sqrt{3}s) = \overline{X} \\ \frac{1}{3}((\overline{X} - \sqrt{3}s)^2 + (\overline{X} - \sqrt{3}s)(\overline{X} + \sqrt{3}s) + (\overline{X} + \sqrt{3}s)^2) = s^2 + \end{cases}$$

$$\Rightarrow \begin{cases} \overline{X} = \overline{X} \\ \frac{1}{3}(\overline{X}^2 - 2\sqrt{3}s\overline{X} + 3s^2 + \overline{X}^2 - 3s^2 + \overline{X}^2 + 2\sqrt{3}s\overline{X} + 3s^2 = s^2 + \overline{X}^2 \end{cases}$$

$$\Rightarrow \begin{cases} \overline{X} = \overline{X} \\ \frac{1}{3}(3\overline{X}^2 + 3s^2) = s^2 + \overline{X}^2 \end{cases} \Rightarrow \begin{cases} \overline{X} = \overline{X} \\ s^2 + \overline{X}^2 = s^2 + \overline{X}^2 \end{cases}$$

Thereby $a = \overline{X} - \sqrt{3}s$, $b = \overline{X} + \sqrt{3}s$

Part 3

1. In the city of Perm there are two hospitals. 1000 babies were born in the first maternity hospital and 500 were born in the second hospital during the summer. Let the probability of babies born male be equal to 50%.

- **a. What are the probabilities:**
  - **i. The first hospital has $>45\%$ of male babies;**

$X \sim \text{Bernoulli}(p)$, where $p = 0.5$

n_1 = 1000

Via De Moivre–Laplace theorem:

$P(a \le \frac{m}{n} \le b) \approx \frac{1}{2}[\Phi(z_2) - \Phi(z_1)]$ where a is the left bound, b is the right bound, $z_1 = \frac{a-p}{\sqrt{pq/n_1}}$, $z_2 = \frac{b-p}{\sqrt{pq/n_1}}$, and Φ is the Laplace function $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{\frac{-t^2}{2}}\, dt$

$z_1 = \frac{0.45-0.5}{\sqrt{0.5 \cdot 0.5/1000}} \approx -3.16$; $z_2 = \frac{1-0.5}{\sqrt{0.5 \cdot 0.5/1000}} \approx 31.62$

$P_1(0.45 \le \frac{m}{n} \le 1) \approx \frac{1}{2}[\Phi(31.62) - \Phi(-3.16)] = \frac{1}{2}[\Phi(31.62) + \Phi(3.16)] = \frac{1}{2}[1 + 0.9984] = 0.9992$

- **ii. The second hospital has $>45\%$ of male babies.**

$n_2 = 500$

$z_1 = \frac{0.45-0.5}{\sqrt{0.5 \cdot 0.5/500}} \approx -2.24$; $z_2 = \frac{1-0.5}{\sqrt{0.5 \cdot 0.5/500}} \approx 22.36$

$P_2(0.45 \le \frac{m}{n} \le 1) \approx \frac{1}{2}[\Phi(22.36) - \Phi(-2.24)] = \frac{1}{2}[\Phi(22.36) + \Phi(2.24)] \approx \frac{1}{2}[1 + 0.9749] \approx 0.9875$

- **b. Compare these probabilities and explain.**

We have the RV $X = \frac{X_1 + \cdots + X_n}{n}$ with the mean

$M(X) = M(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n} n M(X_i) = M(X_i) = \mu$ independs of n

But the variance

$D(X) = D(\frac{X_1 + \cdots + X_n}{n}) = \frac{1}{n^2} n D(X_i) = \frac{D(X_i)}{n} = \frac{\sigma^2}{n}$ depends on n

Hence, an increase of n leads to a decrease of $D(X)$ and values concentrate around of the mean and $P_1 > P_2$

- **c. Now imagine the situation that 515 male babies were born in the first maternity hospital. 240 male babies were born in the second maternity hospital. Is the difference statistically significant?**

Let p_1 and p_2 be a share of birth of boys in the first and second maternity hospitals respectively

$H_0 : p_1 = p_2 = p$ where $p = 0.5$ and $q = 1 - p = 0.5$

$H_1 : p_1 \neq p_2$

Significance level $\alpha = 0.05$

We use the z-score for compare of share:

$z = \frac{w_1 - w_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$ where w is emperical share

Critical area $\overline{G} = [-\infty, -z_{cr}] \cup [z_{cr}, +\infty]$

Critical values $[\Phi(z_{cr}) = \Phi(z_{1-\alpha}) = 0.95] \Rightarrow z_{cr} = z_{0.95} = 1.96$

$z = \frac{\frac{515}{1000} - \frac{240}{500}}{\sqrt{0.5 \cdot 0.5\left(\frac{1}{1000} + \frac{1}{500}\right)}} \approx 1.28$

$\text{P-value} = P(p \in [-\infty, -1.28] \cup [1.28, +\infty]) = 1 - \Phi(1.28) = 1 - 0.7984 = 0.2016 > \alpha$

$z \notin \overline{G}$ thereby accept H_0

Answer: the difference is not statistically significant

2. A group of people want to prove that less than half of all voters support the President's policies about managing the financial crisis. Let v be a share of voters who support the president's policies. What is H_1 and H_0 with $\alpha = 0.05$? A recent poll showed that 228 out of 500 voters support the president's policies. Test the hypothesis with $\alpha = 0.05$. Find the P-value and explain the results.

Let $p = v$

$H_0 : p = p_0 = 0.5$ and $q = 1 - p = 0.5$

$H_1 : p < 0.5$

Significance level $\alpha = 0.05$

We use the z-score for a share:

$$z = \frac{(w - p_0)\sqrt{n}}{\sqrt{pq}} \sim N(0, 1)$$

Critical area for the left-tail case $\overline{G} = [-\infty, z_{cr}]$

Critical value for the left-tail case $[\Phi(z_{cr}) = \Phi(z_{1-2\alpha}) = 0.9] \Rightarrow z_{cr} = z_{0.9} = -1.64$

$$z = \frac{(\frac{228}{500} - 0.5)\sqrt{500}}{\sqrt{0.5 \cdot 0.5}} \approx -1.97$$

$$\text{P-value} = P(z \in [-\infty, -1.97]) = \frac{1 - \Phi(-1.97)}{2} = \frac{1 - 0.9512}{2} = 0.0244 < \alpha = 0.05$$

$z \in \overline{G}$ thereby reject H_0

Thereby the fact that less than half of all voters support the President's policies is statistically significant.