Simulation of Exam- 2024

Maura Mezzetti

Exercise 1

Let $(X_1, \ldots, X_n)$ be a random sample of i.i.d. random variables. Let $f_\theta(x)$ and $F_\theta(x)$ be the density function and cumulative distribution function respectively. Let $\hat{\theta}$ be the MLE of θ, θ_0 be the true parameter, $L(\theta)$ be the likelihood function, $logL(\theta)$ be the loglikelihood function, and $I(\theta)$ be the Fisher information matrix

Exercise 1

Suppose $f_\theta(x)$ is the probability distribution function of a Possion with parameter λ. The Fisher information is

1. $I(\theta) = \frac{1}{\lambda}$
2. $I(\theta) = \lambda$
3. $I(\theta) = \frac{1}{\lambda^2}$
4. $I(\theta) = \lambda^2$

Exercise 1

Suppose $f_\theta(x)$ is the probability distribution function of a Possion with parameter λ. The Fisher information is

1. $I(\theta) = \frac{1}{\lambda}$
2. $I(\theta) = \lambda$
3. $I(\theta) = \frac{1}{\lambda^2}$
4. $I(\theta) = \lambda^2$

Exercise 1

Suppose $\theta = (\theta_1, \theta_2)$, given the system of hypotheses: $H_0 : \theta_1 = 0$ and $\theta_2 = 2$ versus $H_1 : \theta_1 \neq 0$ and $\theta_2 \neq 2$, the LR test statistics $[-2 log \Lambda]$ has an asymptotic distribution that is

1. $\chi(1)$, chi-squared with one degree of freedom,
2. $\chi(2)$ with two degrees of freedom
3. $F(2, n-2)$
4. $N(0, 1)$

Exercise 1

Suppose $\theta = (\theta_1, \theta_2)$, given the system of hypotheses: $H_0 : \theta_1 = 0$ and $\theta_2 = 2$ versus $H_1 : \theta_1 \neq 0$ and $\theta_2 \neq 2$, the LR test statistics $[-2 log \Lambda]$ has an asymptotic distribution that is

1. $\chi(1)$, chi-squared with one degree of freedom,
2. $\chi(2)$ with two degrees of freedom
3. $F(2, n-2)$
4. $N(0, 1)$

Exercise 1

Which statement is true about confidence intervals?

1. If we construct two 95% confidence intervals for a population mean μ (population with a Gaussian distribution with unknown variance) based on two different random samples with different sample size, the interval from a random sample with bigger sample size always gives a narrower confidence interval.
2. A confidence interval is an interval of values computed from sample data that is likely to include the true population parameter value.
3. A confidence interval between 20% and 40% means that the population proportion definitely lies between 20% and 40%.
4. A 99% confidence interval procedure has a lower probability of producing intervals that will include the population parameter than a 95% confidence interval procedure.

Exercise 1

Which statement is true about confidence intervals?

1. If we construct two 95% confidence intervals for a population mean μ (population with a Gaussian distribution with unknown variance) based on two different random samples with different sample size, the interval from a random sample with bigger sample size always gives a narrower confidence interval.
2. A confidence interval is an interval of values computed from sample data that is likely to include the true population parameter value.
3. A confidence interval between 20% and 40% means that the population proportion definitely lies between 20% and 40%.
4. A 99% confidence interval procedure has a lower probability of producing intervals that will include the population parameter than a 95% confidence interval procedure.

Exercise 1

Assume the regularity conditions hold, which of the following term is zero?

1. $E(\hat{\theta} - \theta_0)$
2. $E(logL'(\theta_0))$
3. $Var(\hat{\theta})$
4. $E(L'(\theta_0))$

Exercise 1

Assume the regularity conditions hold, which of the following term is zero?

1. $E(\hat{\theta} - \theta_0)$
2. $E(logL'(\theta_0))$
3. $Var(\hat{\theta})$
4. $E(L'(\theta_0))$

Exercise 2

Let $(X_1, \ldots, X_n)$ be independent identically distributed random variables with p.d.f.

$$f(x; \beta, \alpha) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0$$

1. Find a sufficient statistics for (β, α)

Consider parameter α known and equal to 4:

2. Find $\hat{\beta}_{MLE}$ maximum likelihood estimator (MLE) for β
3. Compute the score function and the Fisher information.
4. Specify asymptotic distribution of $\hat{\beta}_{MLE}$.

Exercise 2

$$
\begin{aligned}
f(x_i;\beta,\alpha) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x_i} \qquad x_i > 0 \\
f(x_1.x_2,\ldots,x_n;\beta,\alpha) &= \prod_i \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \\
f(x_1.x_2,\ldots,x_n;\beta,\alpha) &= \frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \left(\prod_i x_i\right)^{\alpha-1} e^{-\beta\sum_i x_i} \\
f(x_1.x_2,\ldots,x_n;\beta,\alpha) &= \frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} \left(\prod_i x_i\right)^{\alpha-1} e^{-\beta\sum_i x_i}
\end{aligned}
$$

- $\prod_i x_i$ sufficient statistics for α
- $\sum_i x_i$ sufficient statistics for β

Exercise 2

$$
\begin{aligned}
f(x_i;\beta) &= \frac{\beta^4}{6} x^3 e^{-\beta x_i} \qquad x_i > 0 \\
f(x_1,\ldots,x_n;\beta) &= \frac{\beta^{4n}}{6}\left(\prod_i x_i\right)^3 e^{-\beta \sum_i x_i} \\
L(\beta) &= \frac{\beta^{4n}}{6}\left(\prod_i x_i\right)^3 e^{-\beta \sum_i x_i} \\
log(L(\beta)) &= 4nlog(\beta) - log(6) + 3\sum_i log(x_i) - \beta \sum_i x_i \\
\frac{dlog(L(\beta))}{d\beta} &= \frac{4n}{\beta} - \sum_i x_i \\
\frac{d^2 log(L(\beta))}{d^2\beta} &= -\frac{4n}{\beta^2}
\end{aligned}
$$

Exercise 2

$$\frac{dlog(L(\beta))}{d\beta} = \frac{4n}{\beta} - \sum_i x_i$$

$$\frac{d^2 log(L(\beta))}{d^2\beta} = -\frac{4n}{\beta^2} < 0$$

$$\frac{dlog(L(\beta))}{d\beta} = 0 \quad \Rightarrow \quad \hat{\beta}_{MLE} = \frac{4n}{\sum_i x_i}$$

Exercise 2

Score Function

$$\frac{4n}{\beta} - \sum_i x_i$$

Fisher Information

$$\frac{4n}{\beta^2}$$

Exercise 2

Score Function

$$\frac{4n}{\beta} - \sum_i x_i$$

Fisher Information

$$\frac{4n}{\beta^2}$$

Asymptotic distribution of $\hat{\beta}_{MLE}$

$$\hat{\beta}_{MLE} \sim N\left(\beta_0, \frac{\beta_0^2}{4n}\right)$$

Exercise 2

Score Function

$$\frac{4n}{\beta} - \sum_i x_i$$

Fisher Information

$$\frac{4n}{\beta^2}$$

Asymptotic distribution of $\hat{\beta}_{MLE}$

$$\hat{\beta}_{MLE} \sim N\left(\beta_0, \frac{\beta_0^2}{4n}\right)$$

Exercise 2

1. Find the Likelihood Ratio test statistic for testing $H_0 : \beta = 1.2 \quad versus \quad H_1 : \beta \neq 1.2$, specify the asymptotic distribution and verify the null hypothesis, with $\alpha = 0.05$.
2. Find the Wald test statistic for testing $H_0 : \beta = 1.2 \quad versus \quad H_1 : \beta \neq 1.2$, specify the distribution and verify the null hypothesis, with $\alpha = 0.05$.
3. Find the Score test statistic for testing $H_0 : \beta = 1.2 \quad versus \quad H_1 : \beta \neq 1.2$, specify the distribution and verify the null hypothesis, with $\alpha = 0.05$.

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0 : \beta = 1.2 \quad \textit{versus} \quad H_1 : \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0 : \beta = 1.2 \quad \textit{versus} \quad H_1 : \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Likelihood Ratio test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Likelihood Ratio test statistic for testing
$H_0: \beta = 1.2 \quad \textit{versus} \quad H_1: \beta \neq 1.2$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Gamma- Wald Test

$$\hat{\beta} \sim N\left(\beta_0, \frac{1}{nI(\beta_0)}\right)$$

$$\left(\hat{\beta} - \beta_0\right)^2 \times nI(\beta_0) \sim \chi_1$$

Wald test Statistics

$$\left(\hat{\beta} - \beta_0\right)^2 nI(\hat{\beta})$$

Wald test Statistics

$$\left(\hat{\beta} - \beta_0\right) \times \sqrt{nI(\hat{\beta})}$$

Gamma- Wald Test

$$\hat{\beta} \sim N\left(\beta_0, \frac{1}{nI(\beta_0)}\right)$$

$$\left(\hat{\beta} - \beta_0\right)^2 \times nI(\beta_0) \sim \chi_1$$

Wald test Statistics

$$\left(\hat{\beta} - \beta_0\right)^2 nI(\hat{\beta})$$

Wald test Statistics

$$\left(\hat{\beta} - \beta_0\right) \times \sqrt{nI(\hat{\beta})}$$

Gamma- Wald Test

Wald test Statistics

$$\frac{\frac{4n}{\sum_{i=1}^{n} x_i} - \beta_0}{\frac{\frac{4n}{\sum_{i=1}^{n} x_i}}{\sqrt{n}}}$$

Distribution of test statistics

$$N(0,1)$$

Gamma- Wald Test

Wald test Statistics

$$\frac{\frac{4n}{\sum_{i=1}^{n} x_i} - \beta_0}{\frac{\frac{4n}{\sum_{i=1}^{n} x_i}}{\sqrt{n}}}$$

Distribution of test statistics

$$N(0,1)$$

Observed value of test statistics

$$(1-1.2) * \sqrt{800} = -5.66 \qquad p - value < 0.01$$

Gamma- Wald Test

Wald test Statistics

$$\frac{\frac{4n}{\sum_{i=1}^{n} x_i} - \beta_0}{\frac{\frac{4n}{\sum_{i=1}^{n} x_i}}{\sqrt{n}}}$$

Distribution of test statistics

$$N(0,1)$$

Observed value of test statistics

$$(1-1.2) * \sqrt{800} = -5.66 \qquad p - value < 0.01$$

Decision: REJECT

Gamma- Wald Test

Wald test Statistics

$$\frac{\frac{4n}{\sum_{i=1}^{n} x_i} - \beta_0}{\frac{\frac{4n}{\sum_{i=1}^{n} x_i}}{\sqrt{n}}}$$

Distribution of test statistics

$$N(0,1)$$

Observed value of test statistics

$$(1-1.2) * \sqrt{800} = -5.66 \qquad p - value < 0.01$$

Decision: REJECT

Score Test

$$\frac{\left(\frac{\partial logL(\beta_0)}{\partial \beta}\right)^2}{nI(\beta_0)} \sim \chi^2_1$$

Score Test Statistics

$$\frac{\left(\frac{4n}{\beta_0} - \sum_i x_i\right)^2}{\frac{4n}{\beta_0^2}}$$

Score Test

$$\frac{\left(\frac{\partial logL(\beta_0)}{\partial \beta}\right)^2}{nI(\beta_0)} \sim \chi_1^2$$

Score Test Statistics

$$\frac{\left(\frac{4n}{\beta_0} - \sum_i x_i\right)^2}{\frac{4n}{\beta_0^2}}$$

Distribution of Score Test Statistics

$$\chi_1^2$$

Score Test

$$\frac{\left(\frac{\partial logL(\beta_0)}{\partial\beta}\right)^2}{nI(\beta_0)} \sim \chi^2_1$$

Score Test Statistics

$$\frac{\left(\frac{4n}{\beta_0} - \sum_i x_i\right)^2}{\frac{4n}{\beta_0^2}}$$

Distribution of Score Test Statistics

$$\chi^2_1$$

Score Test

Score Test Statistics

$$\frac{\frac{4n}{\beta_0} - \sum_i x_i}{\sqrt{\frac{4n}{\beta_0^2}}}$$

Distribution of the Score Test Statistics

$$N(0,1)$$

Score Test

Score Test Statistics

$$\frac{\frac{4n}{\beta_0} - \sum_i x_i}{\sqrt{\frac{4n}{\beta_0^2}}}$$

Distribution of the Score Test Statistics

$$N(0,1)$$

Observed value of test statistics

$$score = (800/1.2 - 800)/\sqrt{800/(1.2)^2} = -5.66, \quad p-value < 0.01$$

Score Test

Score Test Statistics

$$\frac{\frac{4n}{\beta_0} - \sum_i x_i}{\sqrt{\frac{4n}{\beta_0^2}}}$$

Distribution of the Score Test Statistics

$$N(0,1)$$

Observed value of test statistics

$$score = (800/1.2 - 800)/\sqrt{800/(1.2)^2} = -5.66, \quad p-value < 0.01$$

Decision: REJECT

Score Test

Score Test Statistics

$$\frac{\frac{4n}{\beta_0} - \sum_i x_i}{\sqrt{\frac{4n}{\beta_0^2}}}$$

Distribution of the Score Test Statistics

$$N(0,1)$$

Observed value of test statistics

$$score = (800/1.2 - 800)/\sqrt{800/(1.2)^2} = -5.66, \quad p-value < 0.01$$

Decision: REJECT

Likelihood ratio Test

Log Likelihood

$$log(L(\beta)) = 4nlog(\beta) - log(6) + 3\sum_i log(x_i) - \beta\sum_i x_i$$

Likelihood ratio Test

$$2 \times (logL(\hat{\beta}) - logL(\beta_0)) \quad \sim \quad \chi^2_1$$

Likelihood ratio Test

Log Likelihood

$$log(L(\beta)) = 4nlog(\beta) - log(6) + 3\sum_i log(x_i) - \beta \sum_i x_i$$

Likelihood ratio Test

$$2 \times (logL(\hat{\beta}) - logL(\beta_0)) \quad \sim \quad \chi^2_1$$

Likelihood ratio Test Statistics

$$2 \times (logL(\hat{\beta}) - logL(\beta_0))$$
$$2 \times \left(4nlog\left(\frac{\hat{\beta}}{\beta_0}\right) - (\hat{\beta} - \beta_0)\sum_i x_i\right)$$

Likelihood ratio Test

Log Likelihood

$$log(L(\beta)) = 4nlog(\beta) - log(6) + 3\sum_i log(x_i) - \beta \sum_i x_i$$

Likelihood ratio Test

$$2 \times (logL(\hat{\beta}) - logL(\beta_0)) \quad \sim \quad \chi_1^2$$

Likelihood ratio Test Statistics

$$2 \times (logL(\hat{\beta}) - logL(\beta_0))$$
$$2 \times \left(4nlog\left(\frac{\hat{\beta}}{\beta_0}\right) - (\hat{\beta} - \beta_0)\sum_i x_i\right)$$

Likelihood ratio Test

Likelihood ratio Test Statistics

$$2 \times \left(4nlog \left(\frac{\hat{\beta}}{\beta_0} \right) - (\hat{\beta} - \beta_0) \sum_i x_i \right)$$

Distribution of the Test Statistics

$$\chi_1^2$$

Likelihood ratio Test

Likelihood ratio Test Statistics

$$2 \times \left(4nlog \left(\frac{\hat{\beta}}{\beta_0} \right) - (\hat{\beta} - \beta_0) \sum_i x_i \right)$$

Distribution of the Test Statistics

$$\chi_1^2$$

Observed Test Statistics

$$2 \times (logL(\hat{\beta}) - logL(\beta_0)) = 28.28 \qquad p - value < 0.01$$

Likelihood ratio Test

Likelihood ratio Test Statistics

$$2 \times \left(4nlog \left(\frac{\hat{\beta}}{\beta_0} \right) - (\hat{\beta} - \beta_0) \sum_i x_i \right)$$

Distribution of the Test Statistics

$$\chi^2_1$$

Observed Test Statistics

$$2 \times (logL(\hat{\beta}) - logL(\beta_0)) \quad = \quad 28.28 \qquad p - value < 0.01$$

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Likelihood Ratio test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Gamma: Large Sample Test

Find the Wald test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Score test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Find the Likelihood Ratio test statistic for testing
$H_0 : \beta = 1.5 \quad \textit{versus} \quad H_1 : \beta \neq 1.5$

Specify the distribution and Verify the null hypothesis, with $\alpha = 0.05$

Exercise 3

Eggs are thought to be infected with a bacterium salmonella enteriditis so that the number of organisms, Y , in each has a Poisson distribution with mean μ. The value of Y cannot be observed directly, but after a period it becomes certain whether the egg is infected $(Y > 0)$ or not $(Y = 0)$. Out of n such eggs, r are found to be infected. Find the maximum likelihood estimator of μ and its asymptotic variance.

Exercise 3: Solution

$$
\begin{aligned}
P(Y=y) &= \frac{\mu^y exp(-\mu)}{y!} \\
\theta = P(Y=0) &= exp(-\mu) \\
\hat{\theta}_{MLE} &= \frac{n-r}{n} \\
\mu &= -log(\theta) \\
\hat{\mu}_{MLE} &= -log\left(\frac{n-r}{n}\right)
\end{aligned}
$$

Exercise 3: Solution

Asymptotic variance can be found by delta method, REMIND:
Let $Y_n, \quad n = 1, 2, \ldots,$ be a sequence of random variables such that

$$\sqrt{n}\,(Y_n - \theta) \longrightarrow_d N(0, \sigma^2)$$

Furthermore let $g()$ be a twice differentiable real function defined on R such that $g'(\theta) \neq 0$. Then

$$\sqrt{n}\,(g(Y_n) - g(\theta)) \longrightarrow_d N(0, \left(g'(\theta)\right)^2 \sigma^2)$$

Exercise 3: Solution

- $Y_n = \hat{p}$
- $\theta = E(\hat{p}) = P(Y = 0)$
- $\sigma^2 = Var(\hat{p}) = \frac{\theta(1-\theta)}{n}$

$$
\begin{aligned}
g(t) &= -log(t) \\
g\prime(t) &= -\frac{1}{t} \\
(g\prime(t))^2 &= \frac{1}{t^2} \\
(g\prime(\theta))^2 &= \frac{1}{\theta^2} \\
(g\prime(\theta))^2\sigma^2 &= \frac{(1-\theta)}{\theta n} \\
Var(\hat{\mu}) &= \frac{1}{n}\frac{(1-exp(-\mu))}{exp(-\mu)}
\end{aligned}
$$

Exercise 4

The Economist collects data each year on the price of a Big Mac in various countries around the world. A sample of McDonald's restaurants in Europe in July 2018 resulted in the following Big Mac prices (after conversion to U.S. dollars).

4.44, 3.94, 2.40, 3.97, 4.36, 4.49, 4.19, 3.71, 4.61, 3.89

Assuming that the price of a Big Mac, X, is well modeled by a normal distribution

- Compute an estimate of $P(X < 4.2)$.
- Compute the MLE estimate of $P(X < 4.2)$.

Solution: Exercise 4

- Compute an estimate of $\theta = P(X < 4.2)$.
  - $T = \frac{1}{n}\sum_i I(X_i < 4.2)$
  - $\hat{\theta} = 0.6$

Solution: Exercise 4

- Compute the MLE estimate of $\theta = P(X < 4.2)$.
  - $\theta = \Phi\left(\frac{4.2-\mu}{\sigma}\right)$
  - $\hat{\mu}_{MLE} = \bar{x} = 4$
  - $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_i (x_i - \bar{x})^2 = 0.36$
  - $\hat{\theta}_{MLE} = \Phi\left(\frac{4.2-\hat{\mu}}{\sqrt{\hat{\sigma}^2}}\right) = \Phi\left(\frac{4.2-4}{\sqrt{0.36}}\right) = \Phi(0.33) = 0.63$

Exercise 5

1. Provide correct statement for Neyman Pearson Lemma
2. Provide correct statement for Factorization Theorem
3. Provide correct statement for Cramér-Rao inequality
4. Provide correct statement for Likelihood Principle and Describe method to reach Maximum Likelihood Estimation
5. Describe method to moments estimation
6. Provide correct statement for Likelihood ratio test
7. Define pivot quantities and explain their use for confidence intervals
8. Define and compare the three large sample tests (Wald, Score and Likelihood Ratio Test)