Parameter Estimation

Mahbub Latif, PhD

December 2023

Plan

- Sampling distribution of sample mean
    - Central limit theorem
- Sampling distribution of sample variance

Introduction

- Let $X_1, \ldots, X_n$ be a random sample from a distribution $F_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ is a vector of unknown parameters
  - The sample could be from a Poisson distribution with $\boldsymbol{\theta} = \lambda$
  - It could be from a normal distribution with $\boldsymbol{\theta} = (\mu, \sigma^2)'$
- In probability theory, the parameters of a distribution is assumed to be known
- In statistics observed data are used to make inferences about unknown parameters

Estimation

- There are two types of statistical methods: descriptive statistics and statistical inference
- Descriptive statistics deals with summarizing observed data using graphical tools or numeric values
- Inferential statistics deals with making inference about population parameters using observed data
    - There are two approaches of statistical inference: estimation and test of hypothesis

Estimation

- Parameters of a distribution are estimated using the methods of estimation
    - There are two types of estimation methods: point estimation and interval estimation
- Test of hypothesis deals with making conclusion on a particular statement about a population (probability distribution)

Point estimates

Maximum likelihood estimators

- Any statistic used to estimate the value of an parameter θ is known as an estimator of θ
  - The observed value of an estimator is called the estimate
- Suppose we have a random sample of three observations $X_1 = 10$, $X_2 = 7$ and $X_3 = 4$ from a population with mean μ
  - The sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimator of μ
  - For the sample, $\bar{X} = (21/3) = 7$ is an estimate

Maximum likelihood estimators

- Let $X_1, \ldots, X_n$ be a random sample from $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown parameter vector

- The likelihood function of $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta} \,|\, \boldsymbol{x}) = f(x_1, \ldots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$$

- The maximum likelihood estimator (mle) $\hat{\boldsymbol{\theta}}$ is a value of $\boldsymbol{\theta}$ that maximizes the likelihood function $L(\boldsymbol{\theta} \,|\, \boldsymbol{x})$ or $\log_e L(\boldsymbol{\theta} \,|\, \boldsymbol{x})$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Omega} L(\boldsymbol{\theta} \,|\, \boldsymbol{x})$$

Exponential distribution

- Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ, i.e. $f(x;\theta) = (1/\theta)\, e^{-x/\theta}$

- The likelihood function of θ

$$L(\theta \mid x) = \prod_{i=1}^{n} (1/\theta)\, e^{-x_i/\theta} = (1/\theta^n)\ \exp\left(-\sum_{i=1}^{n} x_i/\theta\right)$$

- The log-likelihood function

$$\log L(\theta \mid x) = -n \log \theta - \sum_{i=1}^{n} x_i/\theta$$

Exponential distribution

- The maximum likelihood estimator of the exponential parameter θ

$$\hat{\theta} = \arg\max_{\theta \in \Omega} \log L(\theta \mid \boldsymbol{x})$$

- So we can write

$$\frac{d}{d\theta} \log L(\theta \mid \boldsymbol{x}) \Big|_{\theta = \hat{\theta}} = 0$$

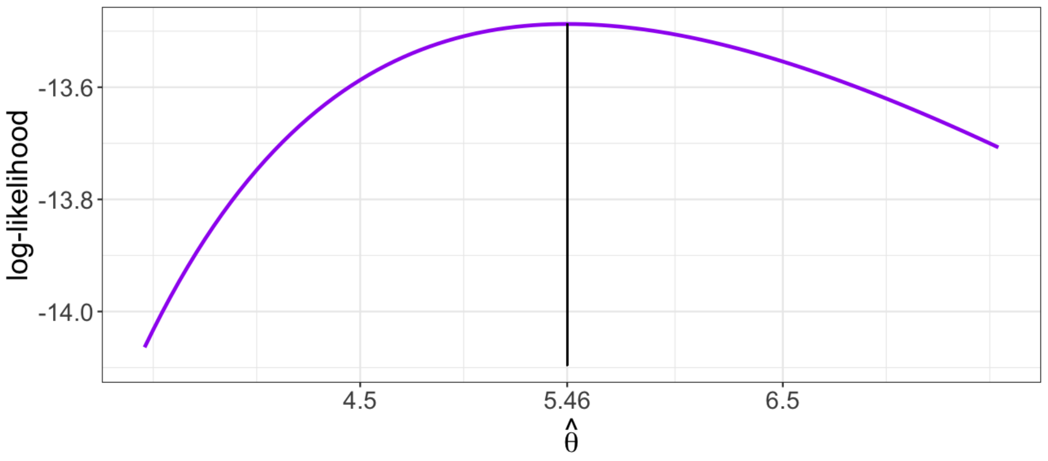
$$-\frac{n}{\hat{\theta}} + \frac{\sum_{i=1}^{n} x_i}{\hat{\theta}^2} = 0 \;\Rightarrow\; \hat{\theta} = \sum_{i=1}^{n} x_i / n = \bar{x}$$

Exponential distribution

- Let $\{4.6, 3.6, 0.5, 15.5, 3.1\}$ be a random sample from an exponential distribution with parameter θ.

- The mle of θ

$$\hat{\theta} = \frac{1}{5}\sum x = \frac{27.3}{5} = 5.46$$



Bernoulli distribution

- Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p, i.e. $f(x;p) = p^x(1-p)^{1-x}$

- The likelihood and log-likelihood function

$$L(p\,|\,\boldsymbol{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\log L(p\,|\,\boldsymbol{x}) = \sum_{i=1}^{n} x_i \log p + (n - \sum_{i=1}^{n} x_i) \log(1-p)$$

  - Show that the mle $\hat{p} = \sum_{i=1}^{n} x_i / n = \bar{x}$

Bernoulli distribution

- Let $\{0, 0, 0, 0, 1, 0, 0, 1, 1, 0\}$ be a random sample from a Bernoulli distribution with parameter p
  - Obtain the mle of p
  - Calculate the value of log-likelihood function at the mle and two other values of p, and compare the results

Poisson distribution

- Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, i.e. $f(x;p) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \ldots$

- The likelihood and log-likelihood function

$$L(\lambda\,|\,\boldsymbol{x}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{\prod x_1!}$$

$$\log L(\lambda\,|\,\boldsymbol{x}) = -n\lambda + \sum_{i=1}^{n} x_i \log\lambda - \log\prod x_1!$$

The normal distribution

- Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, sigma^2)$
- The likelihood and log-likelihood function

$$L(\mu, \sigma^2 \,|\, \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\log L(\mu, \sigma \,|\, \boldsymbol{x}) = -\frac{n}{2}\left(\log 2\pi + \log \sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

The normal distribution

- The log-likelihood function

$$\log L(\mu, \sigma \,|\, \boldsymbol{x}) = -\frac{n}{2}\left(\log 2\pi + \log \sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

- The score functions

$$\frac{d \log L(\mu, \sigma \,|\, \boldsymbol{x})}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)$$

$$\frac{d \log L(\mu, \sigma \,|\, \boldsymbol{x})}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

The normal distribution

$$\left.\frac{d\log L(\mu,\sigma\,|\boldsymbol{x})}{d\mu}\right|_{\mu=\hat{\mu},\sigma=\hat{\sigma}} = \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0$$

$$\Rightarrow\;\hat{\mu} = \bar{x}$$

$$\left.\frac{d\log L(\mu,\sigma\,|\boldsymbol{x})}{d\sigma}\right|_{\mu=\hat{\mu},\sigma=\hat{\sigma}} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0$$

$$\Rightarrow\;\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

Interval estimates

Interval estimates

- Point estimate of a parameter is a single value, e.g. sample mean $\bar{X}$ is a point estimate of the population mean μ

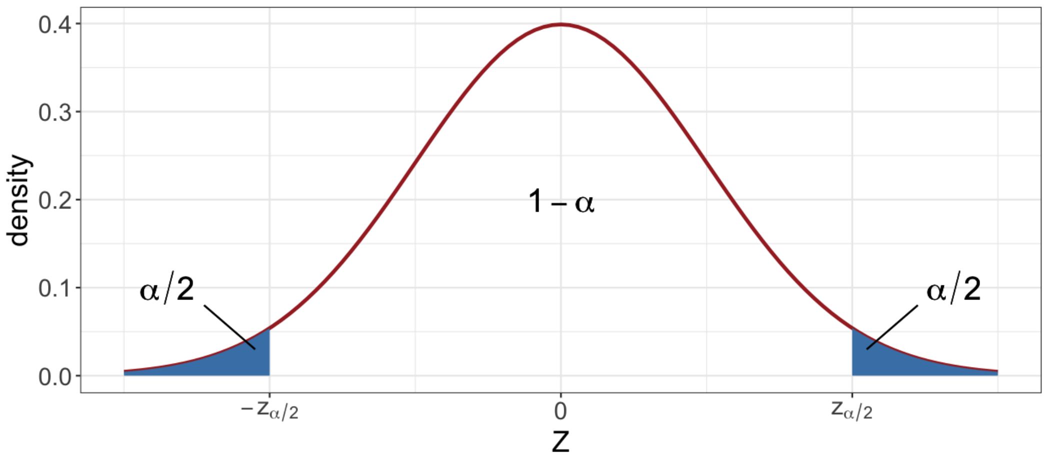
- Interval estimate of a parameter is a range of values that contain the parameter with a certain degree of confidence

Interval estimate of population mean

- Sample mean $\bar{X}$ is an point estimate of μ

- Sampling distribution $\bar{X} \sim N(\mu, \sigma^2/n)$ and

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$



$$P\big(-z_{\alpha/2} < Z < z_{\alpha/2}\big) = 1 - \alpha$$

Interval estimate of population mean

$$P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$P\Big(\bar{X} - z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{X} + z_{\alpha/2}(\sigma/\sqrt{n})\Big) = 1-\alpha$$

- $100(1-\alpha)\%$ two-sided confidence interval for μ

$$\bar{X} \pm z_{\alpha/2}(\sigma/\sqrt{n})$$

Interval estimate of population mean

- $100(1-\alpha)\%$ two-sided confidence interval for μ

$$\bar{X} \pm z_{\alpha/2}(\sigma/\sqrt{n})$$

- 95% confidence interval for μ
  - $100(1-\alpha)\% = 95\% \;\Rightarrow\; \alpha = 0.05$
  - $z_{.05/2} = z_{.025} = 1.96$

$$\bar{X} \pm 1.96(\sigma/\sqrt{n})$$

Interval estimate of population mean

- $100(1-\alpha)\%$ one-sided lower confidence interval for μ

$$P(Z > -z_\alpha) = 1-\alpha \;\Rightarrow\; P\big(\mu < \bar{X} + z_\alpha(\sigma/\sqrt{n})\big) = 1-\alpha$$

$$\Big(-\infty, \bar{X} + z_\alpha(s/\sqrt{n})\Big)$$

- $100(1-\alpha)\%$ one-sided upper confidence interval for μ

$$P(Z < z_\alpha) = 1-\alpha \;\Rightarrow\; P\big(\mu > \bar{X} - z_\alpha(s/\sqrt{n})\big) = 1-\alpha$$

$$\Big(\bar{X} - z_\alpha(s/\sqrt{n}), \infty\Big)$$

Example 7.3a

- Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4.

- If the successive values received are

$$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$$

- Construct a 95 percent two-sided confidence interval for μ.

Example 7.3a

- Signal received $X \sim N(\mu, 4)$
- If the successive values received are

  $$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$$

- Construct a 95 percent two-sided confidence interval for μ.
- Construct a 95 percent lower and upper confidence interval for μ.

- The sample mean $\bar{x} = \frac{81}{9} = 9$
- 95% two-sided confidence interval for μ

  $$\begin{aligned} \bar{x} \pm z_{.025}(\sigma/\sqrt{n}) &= 9 \pm 1.96(2/\sqrt{9}) \\ &= (7.69, 10.31) \end{aligned}$$

- We are 95% confident that the interval (7.69, 10.31) contains the true message value!

Confidence interval for a normal population mean when variance is unknown

- Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2, and both the parameters are unknown

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \Rightarrow \quad t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

- $100(1-\alpha)\%$ confidence interval for μ

$$P\big(-t_{n-1,\alpha/2} < t < t_{n-1,\alpha/2}\big) = 1 - \alpha$$

$$P\big(\bar{x} - t_{n-1,\alpha/2}(s/\sqrt{n}) < \mu < \bar{x} + t_{n-1,\alpha/2}(s/\sqrt{n})\big) = 1 - \alpha$$

Example 7.3a

- Signal received $X \sim N(\mu, \sigma^2)$
- Successive values received are

  $$5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5$$

- 95% two-sided CI for μ?

- The estimates
  - $\bar{x} = 9$
  - $s^2 = 9.5$

Estimating in difference in means of two normal populations

- Let $X_1, \ldots, X_n$ be a sample of size n from a normal population having mean μ_1 and variance σ_1^2
- Let $Y_1, \ldots, Y_m$ be a sample of size m from a different normal population having mean μ_2 and variance σ_2^2
- We are interested to estimate

$$\mu_1 - \mu_2$$

Estimating in difference in means of two normal populations

- Maximum likelihood estimators

$$\hat{\mu}_1 = \bar{X} = (1/n) \sum_i X_i, \quad \hat{\mu}_2 = \bar{Y} = (1/m) \sum_i Y_i, \quad \hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$$

- Sampling distributions

$$\bar{X} \sim N\big(\mu_1, \sigma_1^2/n\big)$$
$$\bar{Y} \sim N\big(\mu_2, \sigma_2^2/m\big)$$
$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Estimating in difference in means of two normal populations

- Since

$$\bar{X} - \bar{Y} \sim N\big(\mu_1 - \mu_2, (\sigma_1^2/n + \sigma_2^2/m)\big)$$

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

- It can be shown

$$P\left\{ - z_{\alpha/2} < \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} < z_{\alpha/2} \right\} = 1 - \alpha$$

Estimating in difference in means of two normal populations

- $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$

$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$$

- If population variances are assumed to be equal, i.e. $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$

$$(\bar{x}-\bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}} = (\bar{x}-\bar{y}) \pm z_{\alpha/2}\,\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}$$

Estimating in difference in means of two normal populations

- If the common population variance σ^2 is unknown, we can estimate it from the data as

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

- The $100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2, n+m-2}\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Approximmate confidence interval for the mean of a Bernoulli random variable

- Let $X \sim B(n, p)$ and the point estimate $\hat{p} = X/n$ and corresponding variance $Var(\hat{p}) = p(1-p)/n$

- For a large n, $\hat{p} \sim N\big(p, p(1-p)/n\big)$, i.e.

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0, 1)$$

Approximate confidence interval for p

- We can show

$$P\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2} \right) = 1 - \alpha$$

- The $100(1-\alpha)\%$ confidence interval for p

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) = 1 - \alpha$$

Problems

- 1, 8, 9, 10, 13, 14, 17, 18, 41, 44, 48, 49,