Probability & Statistics

Dongjiao Ge
djge@cityu.edu.mo
Session 09

4.2 Sample proportion

Suppose now that we have a random sample $X_1, \ldots, X_n$ where the X_i are i.i.d. Bernoulli(p) random variables.

Thus, $\mathrm{P}(X_i = 1) = p$ ('success') and $\mathrm{P}(X_i = 0) = 1 - p$ ('failure'). We know that $E(X_i) = p$ for $i = 1, \ldots, n$ and $Var(X_i) = p(1 - p)$ for $i = 1, \ldots, n$.

The proportion of "successes" is

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

We have

$$\mathrm{E}(\overline{X}_n) = p \text{ and } \mathrm{Var}(\overline{X}_n) = \frac{p(1-p)}{n}$$

4.2 Sample proportion

By the central limit theorem, for large n, $\bar{X}_n$ is approximately distributed as

$$N\left(p, \frac{p(1-p)}{n}\right)$$

We can also say that, for large n, the total number of 'successes' in the sample is approximately normal, i.e., $\sum_{i=1}^{n} X_i \sim N(np, np(1-p))$

Example 1. Suppose that, in a particular country, the unemployment rate is 9.2%. Suppose that a random sample of 400 individuals is obtained. What are the approximate probabilities that:

(i) Forty or fewer were unemployed;

(ii) The proportion of unemployed is greater than 0.125. Solution:

4.2 Sample proportion

Solution:

(i) For $i = 1, \ldots, n$ let the random variable X_i satisfy

$$X_i = \begin{cases} 1 & \text{if the } i\text{th worker is unemployed} \\ 0 & \text{otherwise}. \end{cases}$$

From the question, $\mathrm{P}(X_i = 1) = 0.092$ and $\mathrm{P}(X_i = 0) = 0.908$.

We have $n = 400 \geq \{0.9, 88.8\}$ so that the normal approximation will be valid. Note that $np = 400 \times 0.092 = 36.8$ and $np(1-p) = 400 \times 0.092 \times 0.908 = 33.414$, and $\sum_{i=1}^{n} X_i \sim N(np, np(1-p))$ approximately.

4.2 Sample proportion

$$
\begin{aligned}
\mathrm{P}\left(\sum_{i=1}^{400} X_i \leq 40\right) &= \mathrm{P}\left(\frac{\sum_{i=1}^{400} X_i - 36.8}{\sqrt{33.414}} \leq \frac{40.5 - 36.8}{\sqrt{33.414}}\right) \\
&\approx \mathrm{P}\left(Z \leq 0.640\right), \quad \text{where } Z \sim N(0,1) \text{ approx.} \\
&= \Phi(0.640) \\
&= 0.7390\,.
\end{aligned}
$$

(ii) Here, $\frac{p(1-p)}{n} = \frac{0.092 \times 0.908}{400} = 0.0002088$. Thus,

$$
\begin{aligned}
\mathrm{P}\left(\overline{X}_{400} > 0.125\right) &= \mathrm{P}\left(\frac{\overline{X}_{400} - 0.092}{\sqrt{0.0002088}} > \frac{0.125 - 0.092}{\sqrt{0.0002088}}\right) \\
&\approx 1 - \Phi(2.284) \\
&= 1 - 0.9888 \\
&= 0.0112\,.
\end{aligned}
$$

4.3 Sample variance

The sampling distribution of the sample variance, S^2, is defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

where $X_1, \ldots, X_n$ are a random sample from the distribution with c.d.f. $F_X(\cdot)$ with mean μ and variance σ^2.

可以是连续的也可以是离散的

4.3 Sample variance

If F_X is any discrete or continuous distribution with a finite variance then

$$\begin{aligned}
\mathrm{E}(S^2) &= \frac{1}{(n-1)}\,\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right] \\
&= \frac{1}{(n-1)}\,\mathrm{E}\left[\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2\right] \\
&= \frac{1}{(n-1)}\,\mathrm{E}\left[\sum_{i=1}^{n}[(X_i-\mu)^2-2(X_i-\mu)(\overline{X}-\mu)+(\overline{X}-\mu)^2]\right] \\
&= \frac{1}{(n-1)}\,\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2-2n(\overline{X}-\mu)(\overline{X}-\mu)+n(\overline{X}-\mu)^2\right] \\
&= \frac{1}{(n-1)}\left[\sum_{i=1}^{n}\mathrm{E}\left[(X_i-\mu)^2\right]-2n\,\mathrm{E}[(\overline{X}-\mu)^2]+n\,\mathrm{E}[(\overline{X}-\mu)^2]\right] \\
&= \frac{1}{(n-1)}\left[n\sigma^2-2n\frac{\sigma^2}{n}+n\frac{\sigma^2}{n}\right] \\
&\qquad \text{since } \mathrm{E}[(\overline{X}-\mu)^2]=\mathrm{Var}(\overline{X})=\frac{\sigma^2}{n} \\
&= \frac{1}{(n-1)}\left[(n-1)\sigma^2\right]=\sigma^2\,.
\end{aligned}$$

$E(S^2)=\sigma^2$ means S^2 is the ***unbiased*** estimation of the variance σ^2.

4.3 Sample variance

4.3.1 The chi-squared (χ^2) distribution 卡方分布

We will look more at the case when the X_i are sampled from the $N(\mu, \sigma^2)$.

In order to do so, we first need to introduce a new continuous probability distribution, the **chi-squared** (χ^2) distribution.

The continuous random variable Y is said to have χ^2 distribution with k degrees of freedom (d.f.), written as $\chi^2(k)$, if and only if its p.d.f is given by

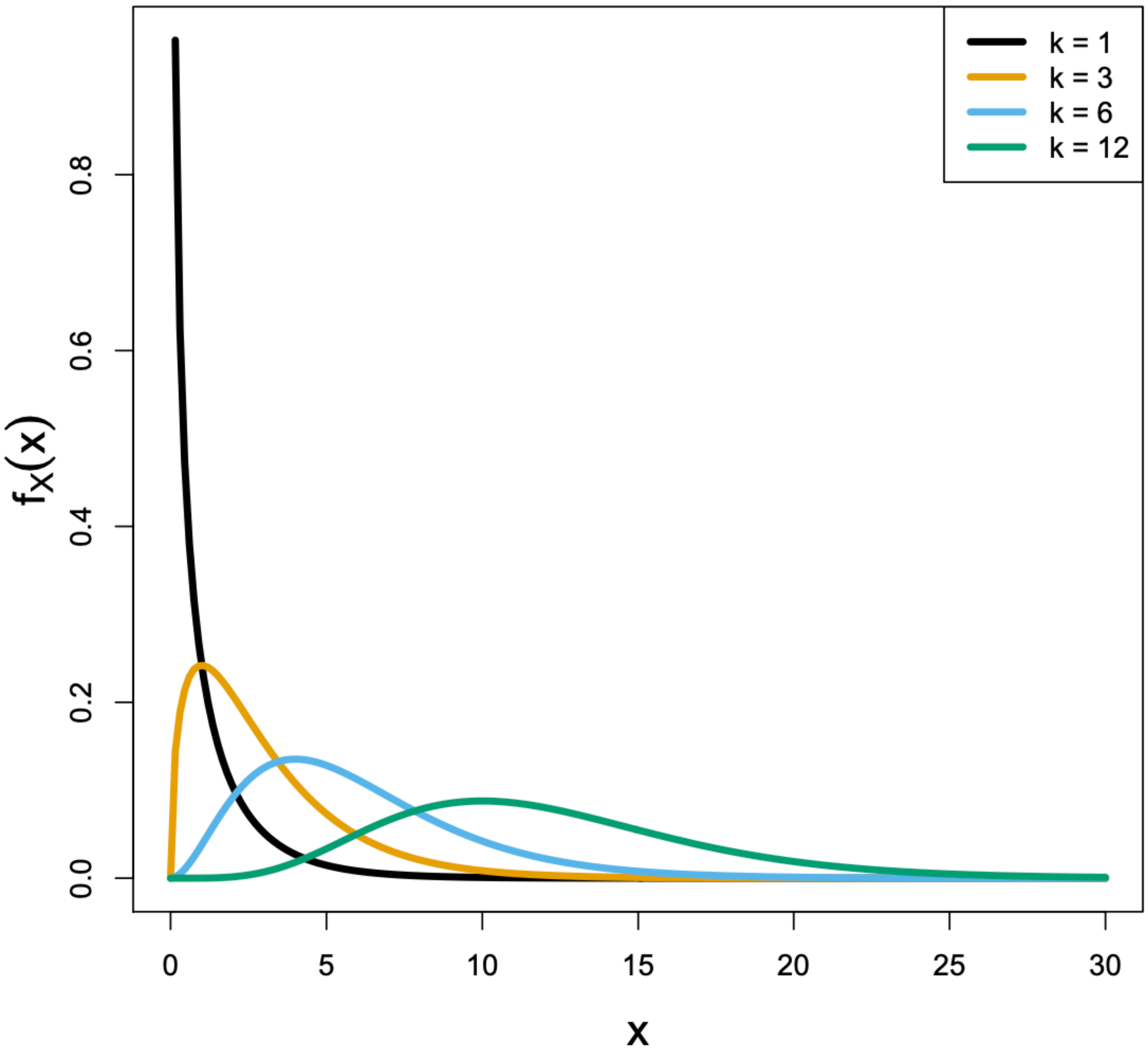
$$f(y) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} y^{(k/2)-1} e^{-y/2}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

4.3 Sample variance

4.3.1 The chi-squared (χ^2) distributio

The mean and variance are given by $E(Y) = k$ and $Var(Y) = 2k$.

- The p.d.f.s of chi-squared random variables with d.f. = 1, 3, 6, and 12 are shown in the figure.
- Note that the p.d.f. becomes **more symmetric** as the number of **degrees of freedom** k becomes **larger**.



4.3 Sample variance

4.3.2 The connection of χ^2 with the normal distribution

Let $Z_1, \dots, Z_k$ be k i.i.d. **standard normal random variables**, i.e. each has a $N(0,1)$. Then, the random variable

$$Y = \sum_{i=1}^{k} Z_i^2$$

has a χ^2 distribution with k degrees of freedom.

Suppose now the random variables $X_1, \ldots, X_n$ are a random sample from the $N(\mu, \sigma^2)$. We have that

$$\frac{X_i - \mu}{\sigma} \sim N(0,1)\,, \quad i = 1, \ldots, n\,,$$

so that

$$\sum_{i=1}^{n} \left[\frac{(X_i - \mu)}{\sigma}\right]^2 \sim \chi^2(n)\,.$$

4.3 Sample variance

4.3.2 The connection of χ^2 with the normal distribution

Cochran's theorem

If $X_1, X_2, \dots, X_n$ are i.i.d standard normal random variables, then $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Example 2. Let $X_1, \dots, X_{40}$ be a random sample of size $n = 40$ from the $N(25, 4^2)$ distribution. Find the probability that the sample variance, S^2, exceeds 20.

Solution.

We need to calculate

$$\begin{aligned} \mathrm{P}\left(S^2 > 20\right) &= \mathrm{P}\left(\frac{39 \times S^2}{16} > \frac{39 \times 20}{16}\right) \\ &= \mathrm{P}(Y > 48.75) \quad \text{where } Y \sim \chi^2(39) \\ &= 1 - \underline{\mathrm{P}(Y < 48.75)} = 1 - 0.8638 = 0.1362\,, \end{aligned}$$

the probability can be computed by using the *pchisq* command in R

5 Point estimation

5.1 Introduction

The objective of a statistical analysis is to make inferences about a population based on a sample.

Usually, we begin by assuming that the data were generated by a probability model for the population. Such a model will typically contain one or more parameters θ whose value is unknown.

The value of θ needs to be estimated using the sample data.

- **Point estimation**
- **Maximum likelihood estimation (MLE)** (will be discussed later on)

5.2 General framework

Let $X_1, \dots, X_n$ be a random sample from a distribution with c.d.f. $F_X(x;\theta)$, where θ is a parameter whose value is unknown. A (point) estimator of θ, denoted by $\hat{\theta}$ is a real, single-valued function of the sample, i.e.

$$\widehat{\theta} = h(X_1, \dots, X_n)\,.$$

- the estimator $\hat{\theta}$ is also a random variable whose probability distribution is called its sampling distribution.
- The value $\hat{\theta} = h(x_1, x_2, \dots, x_n)$ assumed for a particular sample $x_1, x_2, \dots, x_n$ of observed data is called a (point) estimate of θ.

5.3 Properties of estimators

We would like an estimator $\widehat{\theta}$ of θ to be such that:

Unbiased

(i) the sampling distribution of $\widehat{\theta}$ is centered about the target parameter, θ.

(ii) the spread of the sampling distribution of $\widehat{\theta}$ is small.

Small spread

If an estimator has properties (i) and (ii) above then we can expect estimates resulting from statistical experiments to be close to the true value of the population parameter we are trying to estimate.

5.3 Properties of estimators

The bias of a point estimator $\hat{\theta}$ is

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

无偏性

The estimator is said to be **unbiased** if $E(\hat{\theta}) = \theta$ (i.e.,$bias(\hat{\theta})$=0)

Unbiasedness corresponds to property (i) above, and is generally seen as a desirable property for an estimator. Note that sometimes biased estimators can be modified to obtain unbiased estimators.

$$\text{if } \mathrm{E}(\widehat{\theta}) = k\theta, \text{ where } k \neq 1 \text{ a constant, then } \mathrm{bias}(\widehat{\theta}) = (k-1)\theta.$$

$\widehat{\theta}/k$ --- unbiased estimator of $\hat{\theta}$

5.3 Properties of estimators

The **spread** of the sampling distribution can be measured by $Var(\hat{\theta})$

$\sqrt{Var(\hat{\theta})}$ --- **standard error**

Suppose that we have two different unbiased estimators of θ, i.e., $\hat{\theta}_1$ and $\hat{\theta}_2$.
We would prefer to use the estimator with the smallest variance, i.e. choose $\hat{\theta}_1$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, otherwise, choose $\hat{\theta}_2$

5.3 Properties of estimators

Example Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution where σ^2 is assumed **known**. Recall that the $X_i \sim N(\mu, \sigma^2)$ independently in this case. We can estimate μ by the sample mean, i.e.

$$\widehat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \,.$$

$$\mathrm{E}(\overline{X}) = \mu, \text{ thus bias}(\overline{X}) = 0.$$

$$\mathrm{Var}(\overline{X}) = \sigma^2/n$$

6. Likelihood

6.1 The likelihood function 似然函数

One method for deriving an estimator, which works for almost any parameter of interest, is the method of **maximum likelihood**. The estimators derived in this way typically have good properties. The method revolves around the **likelihood function**, which is of great importance throughout Statistics.

Let $X_1, \dots, X_n$ be an i.i.d. random sample from the **discrete/continuous distribution** with **p.m.f./p.d.f** $p(x|\theta)$, where θ is a parameter whose value is unknown. Given observed data values $x_1, \dots, x_n$ from this model, the likelihood function is defined as

$$L(\theta) = \mathrm{P}(X_1 = x_1,\, X_2 = x_2,\, \dots,\, X_n = x_n \,|\, \theta)$$

*In other words, the likelihood is the **joint probability/density** of the observed data considered as a function of the unknown parameter θ.*

By **independence**, we can rewrite the likelihood as follows:

$$L(\theta) = p(x_1 \,|\, \theta) \times \cdots \times p(x_n \,|\, \theta)$$

6.1 The likelihood function

Example Let $x_1, \ldots, x_n$ be a sample obtained from the Poisson(λ) distribution with p.m.f.

$$p(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

The likelihood function for this sample is given by:

$$L(\lambda) = \prod_{i=1}^{n} p(x_i \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}, \qquad \text{for } \lambda > 0\,.$$

6.2 Maximum likelihood estimation

In the discrete case, given sample data $x_1, \ldots, x_n$ the maximum likelihood estimate for θ is the value $\hat{\theta}$ that maximizes the joint probability of the observed data, i.e. that maximizes the value of the likelihood function $L(\theta)$.

Maximization of $L(\theta) = \prod_{i=1}^{n} p(x_i \,|\, \theta) \Longrightarrow$ leads to a numerical value $\hat{\theta}$ for the estimate of θ.

The value of $\hat{\theta}$ depends on the observed sample values $x_1, \ldots, x_n$, i.e. $\hat{\theta}$ is a function of the data,

$$\widehat{\theta} = h(x_1, \ldots, x_n)\,.$$

We can also consider $\hat{\theta}$ as a function of the random sample, $X_1, \ldots, X_n$,

$$\widehat{\theta} = h(X_1, \ldots, X_n)\,,$$

in which case $\hat{\theta}$ is a random variable called the **maximum likelihood estimator**.

6.2 Maximum likelihood estimation

In simple cases, the maximum likelihood estimate can be found by standard calculus techniques, i.e. by solving

$$\frac{dL(\theta)}{d\theta} = 0\,.$$

However, it is usually much easier algebraically to find the maximum of the log-likelihood $l(\theta) = logL(\theta)$ because for i.i.d. data,

$$\log L(\theta) = \log \left[\prod_{i=1}^{n} p(x_i \,|\, \theta) \right] = \sum_{i=1}^{n} \log p(x_i \,|\, \theta)\,.$$

To find the value of θ that maximizes $l(\theta)$ we instead find $\hat{\theta}$ that solves:

$$\frac{dl(\theta)}{d\theta} = \sum_{i=1}^{n} \frac{d \log p(x_i \,|\, \theta)}{d\theta} = 0\,.$$

The solution is a maximum if $\frac{d^2 l(\theta)}{d\theta^2} < 0$ at $\theta = \hat{\theta}$.

6.2 Maximum likelihood estimation

Example Let $X_1, \ldots, X_n$ be a random sample from the Poisson(λ) distribution. Find the maximum likelihood estimator of λ.

We have seen that

$$L(\lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!},$$

so that

$$l(\lambda) = -n\lambda + \left(\sum^n X_i\right)\log\lambda - \log\left(\prod^n X_i!\right).$$

Therefore, $\widehat{\lambda} = \overline{X}$ is indeed the maximum likelihood estimator of λ.

Solving $\frac{dl(\lambda)}{d\lambda} = 0$, we obtain

$$\left.\frac{dl}{d\lambda}\right|_{\lambda=\widehat{\lambda}} = -n + \frac{\sum_{i=1}^n X_i}{\widehat{\lambda}} = 0, \quad \text{which implies that } \widehat{\lambda} = \overline{X}.$$

6.2 Maximum likelihood estimation

Checking the second derivatives, we see that

$$\left.\frac{d^2 l}{d\lambda^2}\right|_{\lambda=\widehat{\lambda}} = \frac{-\sum_{i=1}^{n} X_i}{\widehat{\lambda}^2} = \frac{-n}{\overline{X}} < 0\,.$$

Therefore, $\widehat{\lambda} = \overline{X}$ is indeed the maximum likelihood estimator of λ. If we have a set of data $x_1, \ldots, x_n$ then the maximum likelihood estimate of λ is $\widehat{\lambda} = \overline{x}$, the sample mean. This is an intuitively sensible estimate, as the mean of the Poisson(λ) distribution is equal to λ.

6.2 Maximum likelihood estimation

Example Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with density function $f(x|\sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right)$, please find the maximum likelihood estimate of σ.

Solution: The log-likelihood function is

$$l(\sigma) = \sum_{i=1}^{n} \left[-\log 2 - \log \sigma - \frac{|X_i|}{\sigma} \right]$$

Let the derivative with respect to θ be zero:

$$l'(\sigma) = \sum_{i=1}^{n} \left[-\frac{1}{\sigma} + \frac{|X_i|}{\sigma^2} \right] = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n} |X_i|}{\sigma^2} = 0$$

and this gives us the MLE for σ as

$$\hat{\sigma} = \frac{\sum_{i=1}^{n} |X_i|}{n}$$

Again this is different from the method of moment estimation which is

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}$$

7 Confidence intervals

7.1 Interval estimation 区间估计

- In this module, whenever we have fitted a probability model to a data set, we have done so by calculating point estimates of the values of any unknown parameters θ.
- However, it is very rare for a point estimate to be exactly equal to the true parameter value.
- An alternative approach is to specify an interval, or range, of plausible parameter values. We would then expect the true parameter value to lie within this interval of plausible values. We call such an interval an **interval estimate** of the parameter.

7.1 Interval estimation

Let $X = (X_1, \ldots, X_n)$ be an independent random sample from a distribution $F_X(x; \theta)$ with unknown parameter θ. An interval estimator,

$$I(\mathbf{X}) = [l(\mathbf{X}), u(\mathbf{X})]$$

for θ is defined by two statistics, i.e. functions of the data. The statistic $u(X)$ defines the upper end-point of the interval, and the statistic $l(X)$ defines the lower end-point of the interval.

The key property of an interval estimator for θ is its coverage probability. This defined as the probability that the interval contains, or 'covers', the true value of the parameter, i.e.

$$\mathrm{P}_\theta[\, l(\mathbf{X}) \leq \theta \leq u(\mathbf{X})\,]$$

We use the notation P_θ for probabilities here to emphasize that the probability distributions of $l(X)$ and $u(X)$ depend on θ.

7.1 Interval estimation

Let $\alpha \in (0,1)$, and suppose that we have been able to find statistics l and u such that the coverage probability satisfies

$$\mathrm{P}_\theta[\, l(\mathbf{X}) \leq \theta \leq u(\mathbf{X})\,] = 1 - \alpha\,, \quad \text{for all values of } \theta\,,$$

Then the interval estimator $I(X)$ and, for any particular data set $x = (x_1, \ldots, x_n)$ the resulting interval estimate $I(x)$, is referred to as a $100(1-\alpha)\%$ confidence interval for θ. The proportion $1-\alpha$ is referred to as the **confidence level**, and the interval end points $l(x)$, $u(x)$ are known as the **confidence limits**. 置信区间

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

To illustrate the idea, let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, with μ unknown but σ^2 known. Recall that $X \sim N(\mu, \frac{\sigma^2}{n})$. Thus, if we standardize $\bar{X}$ then we obtain the random variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)\,.$$

A crucial property of Z above is that the distribution of Z does not depend on μ or σ, i.e. the right hand side of the above equation is the same no matter what the value of μ or σ.

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

Let $z_{1-\alpha/2}$ be such that $\mathrm{P}(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha/2$.

By symmetry of the normal distribution, it is also true that

$$\mathrm{P}(Z \leq -z_{1-\frac{\alpha}{2}}) = \alpha/2,$$

and furthermore

$$\mathrm{P}(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Therefore, we have

$$\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha\,.$$

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

Moreover, rearrange the inequality inside the brackets, we have

$$1-\alpha = \mathrm{P}\left(-\frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}} - \overline{X} \le -\mu \le +\frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}} - \overline{X}\right)$$

$$= \mathrm{P}\left(\overline{X} - \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}} \le \mu \le \overline{X} + \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\right).$$

Hence, the interval estimator $I(X)$ for μ defined by

$$I(\mathbf{X}) = \left[\overline{X} - \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\,,\ \overline{X} + \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\right]$$

contains the true value of μ with probability $1-\alpha$.

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

Assume we have a particular set of data values/ observations $x = (x_1, \ldots, x_n)$, the interval estimate

$$I(\mathbf{x}) = \left[\overline{x} - \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\;,\; \overline{x} + \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\right]$$

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

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

We must be careful how to interpret confidence intervals:

- Given a particular realised data set x with corresponding calculated interval $I(x)$, it is not true to say that the parameter θ lies within $I(x)$ with $100(1 - \alpha)\%$ probability. The value of θ is a fixed unknown, and not a random variable.
- Moreover, once we have observed data $x, I(x)$ is also fixed and no longer a random variable. Hence either θ is in $I(x)$ or it is not: there are no random variables remaining about which to make probability statements.

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

The correct presentations:

- ✓ The correct interpretation is that before the experiment the probability that the interval estimator will ultimately contain the true value of θ is $100(1 - \alpha)\%$.
- ✓ Alternatively, if we repeated the experiment a large number of times and calculated a confidence interval for each sample, then approximately $100(1 - \alpha)\%$ of the confidence intervals would contain the true value of θ.

7.2 Single sample procedures

7.2.1 Confidence interval for the mean of a normal distribution with known variance

Example The following $n = 16$ observations are a random sample from a $N(\mu, 2^2)$ distribution, where μ is unknown:

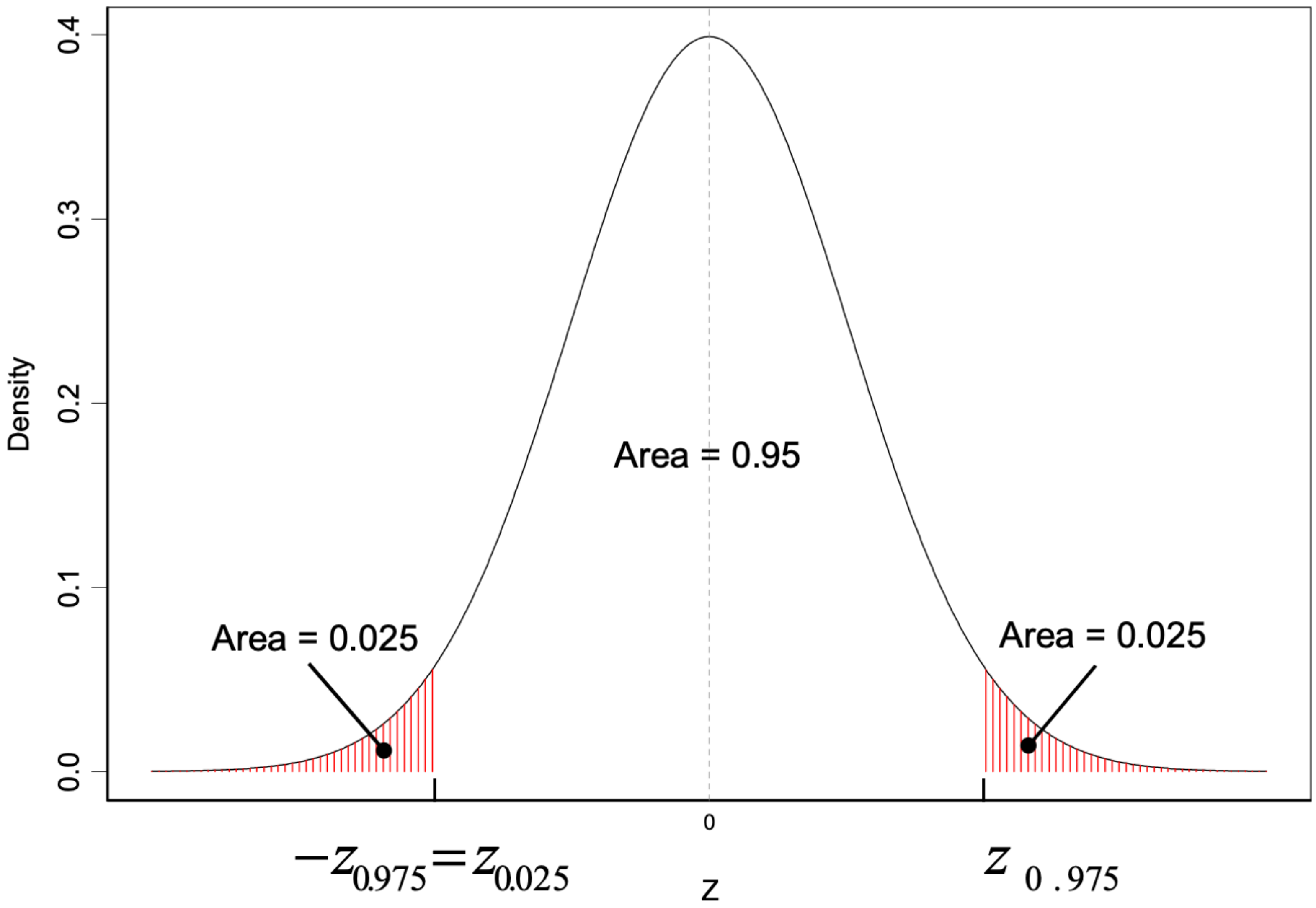
		X2							
X1	10.43	5.42	11.10	12.41	10.14	7.83	8.84	10.42	
	10.44	9.65	10.36	11.48	9.33	6.81	10.55	10.41	X16

We want to use the data to construct a 95% confidence interval for μ, i.e. here $\alpha = 0.05$. The sample mean is $\bar{x} = 9.73$ and $z_{1-\alpha/2} = z_{0.975} = 1.96$ so that the end-points of the 95% CI for μ are given by:

$$9.73 \pm 1.96 \times \sqrt{\frac{4.0}{16}},$$

i.e. the interval is $(8.75, 10.71)$. These data were actually sampled (simulated) from a $N(10, 2^2)$ distribution. Thus the true value $\mu = 10$ is within the CI.

7.2 Single sample procedures



7.2 Single sample procedures
7.2.2 Confidence interval for the mean of a normal distribution, variance unknown

Suppose now that $X_1, \ldots, X_n$ are independent draws from a $N(\mu, \sigma^2)$ distribution where both μ and σ^2 are unknown.

Due to $\sigma$is **unknown**, it is no longer possible to use the confidence interval:

$$\left[\overline{x} - \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}, \overline{x} + \frac{z_{1-\frac{\alpha}{2}}\,\sigma}{\sqrt{n}}\right]$$

Recall, if σ is known, we can baeg a confidence interval on the random variable

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)\,,$$

7.2 Single sample procedures

7.2.2 Confidence interval for the mean of a normal distribution, variance unknown

If σ is unknown, we plug in an estimate of the sample variance in the denominator, namely the sample variance (with divisor n − 1), to obtain

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}\,.$$

- Because both X and S are random variables the distribution of T is **not** $N(0,1)$.
- The fact that S is also random induces extra variability into the distribution of T . Thus, for a given value of n, the distribution of T has a longer tail than that of Z.
- The distribution of T is the **student's t distribution**

7.2 Single sample procedures
7.2.3 Student's t-distribution

The exact distribution of T above is a Student's t-distribution with $(n - 1)$ degrees of freedom, denoted $t(n - 1)$

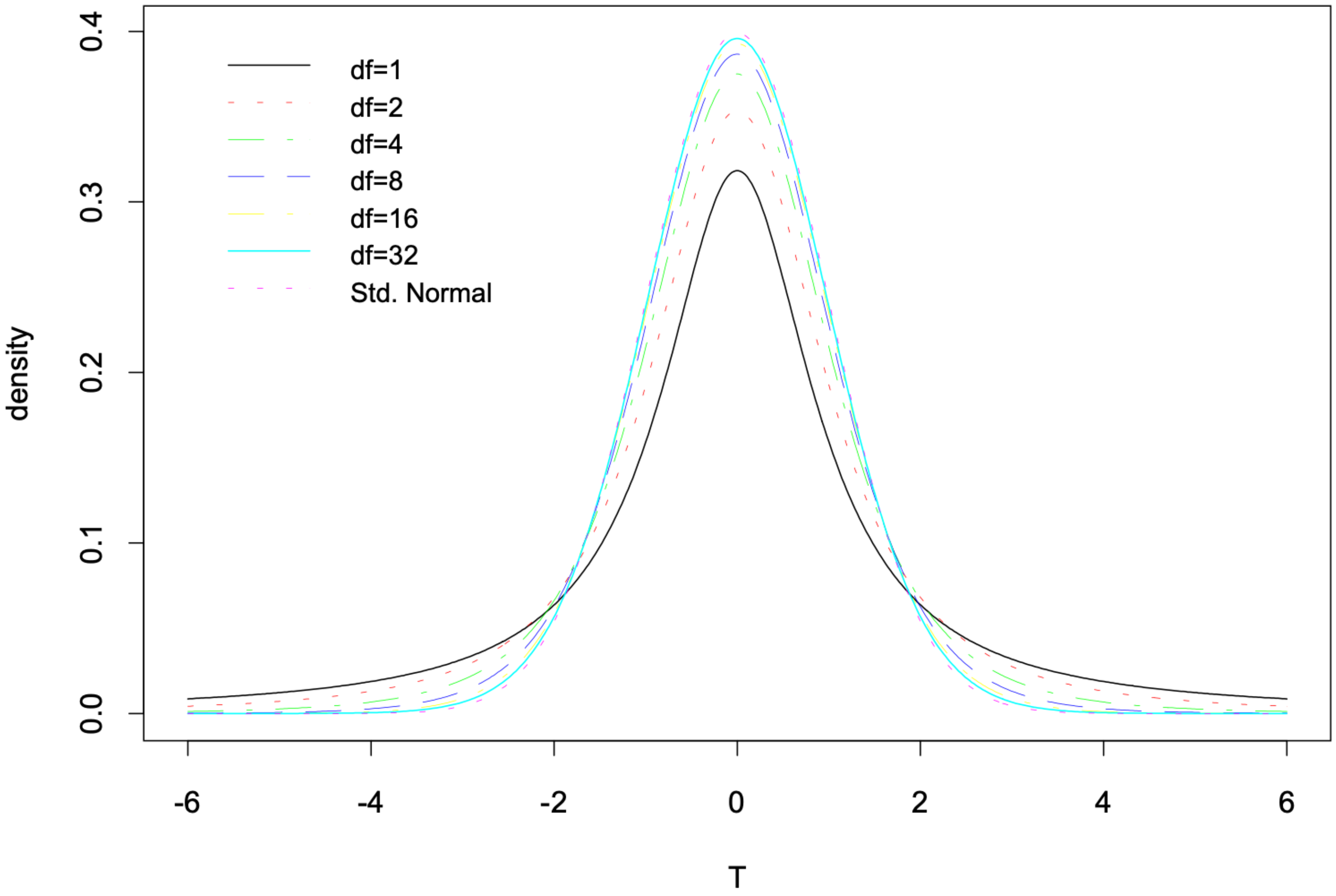
In general, if the random variable T has a t-distribution with ν degrees of freedom then its probability density function is given by:

$$f_T(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{(\nu+1)}{2}},$$

For $\nu > 0$ and $-\infty < x < \infty$. We have that $E(T) = 0$ and $Var(T) = \nu/(\nu - 2)$, for $\nu > 2$. Moreover,the distribution is symmetric about the origin. As the parameter $\nu \rightarrow \infty$, the p.d.f. of T approaches that of the $N(0,1)$ distribution.

7.2 Single sample procedures

7.2.3 Student's t-distribution 尖峰厚尾



7.2 Single sample procedures
7.2.3 Student's t-distribution

Define $t_{1-\frac{\alpha}{2}}$ to be $1-\frac{\alpha}{2}$ point of the $t(n-1)$ distribution, i.e. if $T \sim t(n-1)$, then $P(T \geq t_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Then from the preceding discussion, it follows that the random interval

$$I(\mathbf{X}) = \left[\overline{X} - \frac{t_{1-\frac{\alpha}{2}} S}{\sqrt{n}},\ \overline{X} + \frac{t_{1-\frac{\alpha}{2}} S}{\sqrt{n}}\right]$$

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

7.2 Single sample procedures
7.2.3 Student's t-distribution

Example There are $n = 50$ observations and we found that $\bar{x} = 334.59$ and $s^2 = 15.288$, we do not know the true value of σ^2, we use the critical value $t_{0.975} = 2.0096$ for the $t(49)$ distribution.

The 95% CI for μ has end-points:

$$334.59 \pm 2.0096 \times \sqrt{\frac{15.288}{50}},$$

i.e. $I(x) = (333.48, 335.70)$ which gives a range of plausible values for μ.

Exercises

Q1 Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, Please give the MLE for $\theta = (\mu, \sigma^2)$. Are the MLEs unbiased?