Interval Estimation

- What is interval estimation?

Question 6.10

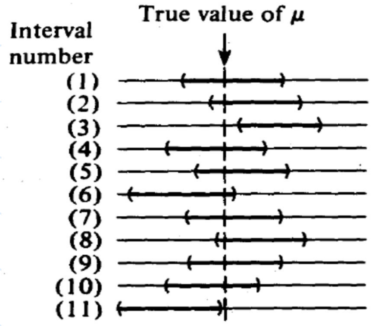
- Is it satisfactory to report only an estimated value of θ?
- Note that
  1. A point estimate, although it will represent our best guess for the true value of the parameter, may be close to that true value but will virtually *never equal* it.
  2. Some measure of how close the point estimate is to the true value is required. One way to do this is to report both the *estimate* and its *estimated standard error*.
- The following questions arise naturally:
  1. A point estimator only gives a value. Wound it be better that we can give customer a range of possible values? This amounts to replacing the point estimate, a single value, by an entire interval of plausible values.
  2. Is there an estimation method that can combine together the two types of information, i.e., estimated value and estimated standard error?

Definition 6.22 (interval estimator, coverage probability, interval estimate, confidence interval, and confidence level, TBp. 217 & 279)

- For a random vector $\mathbf{X} = (X_1, \ldots, X_n)$, an **interval estimator** of a parameter θ with **coverage probability** $1-\alpha$ is a *random interval*

$$(\hat{\theta}_L(\mathbf{X}), \hat{\theta}_U(\mathbf{X})),$$

  where
  1. $\hat{\theta}_L(\mathbf{X})$, $\hat{\theta}_U(\mathbf{X})$ are functions of data,
  2. $\hat{\theta}_L(\mathbf{X}) < \hat{\theta}_U(\mathbf{X})$, and,
  3. $P(\theta \in (\hat{\theta}_L(\mathbf{X}), \hat{\theta}_U(\mathbf{X}))) = 1-\alpha$.
- If $\mathbf{X} = \mathbf{x}$ is observed, the interval $(\hat{\theta}_L(\mathbf{x}), \hat{\theta}_U(\mathbf{x}))$ is called an **interval estimate**.
- The term "$100 \cdot (1-\alpha)\%$ **confidence interval**" is used to denote either an interval estimator with coverage probability $1-\alpha$ or an interval estimate.
- The $100(1-\alpha)\%$ is also referred to as **confidence level**.
- **Note.** The α is usually assigned a small value, e.g. 0.1, 0.05, or 0.01.



Repeated construction of 95% confidence intervals

Question 6.11

How to interpret the $100(1-\alpha)\%$ in a $100(1-\alpha)\%$ interval *estimate* $(\hat{\theta}_L(\mathbf{x}), \hat{\theta}_U(\mathbf{x}))$? For example, for a 95% interval estimate, say (23.5, 47.8), can we say that

$$P(\theta \in (23.5, 47.8)) = 0.95?$$

• methods for constructing interval estimators

Recall. methods for the estimation of standard error

(1) exact distribution (2) asymptotic method (3) bootstrap

The three methods can be applied to construct confidence intervals.

Definition 6.23 (pivotal quantity)

A function of data $X_1, \ldots, X_n$ and parameter θ, denoted by

$$Q(\mathbf{X}, \theta) = Q(X_1, \ldots, X_n, \theta),$$

is called a **pivotal quantity** for θ if the distribution of $Q(\mathbf{X}, \theta)$ is irrelevant to *all* parameters. (cf. ancillary statistics)

Example 6.38 (some pivotal quantities, TBp. 279-281)

1. Let $X_1, \ldots, X_n$ be i.i.d. from Exponential distribution $E(\lambda)$, then
$$\lambda \sum_{i=1}^{n} X_i \sim \Gamma(n, 1).$$
Because the pdf of $\Gamma(n,1)$ is irrelevant to λ, $\lambda \sum_{i=1}^{n} X_i$ is a pivotal quantity.
2. Let $X_1, \ldots, X_n$ be i.i.d. from Uniform distribution $U(0, \theta)$. Then, the pdf of
$$\frac{X_{(n)}}{\theta}$$
is
$$nq^{n-1}, \quad 0 \le q \le 1.$$
Because the pdf is irrelevant to θ, $X_{(n)}/\theta$ is a pivotal quantity.
3. Let $X_1, \ldots, X_n$ be i.i.d. from Normal distribution $N(\mu, \sigma^2)$, where μ is a parameter and the value of σ is known. Then,
$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0, 1).$$
Because the pdf of $N(0,1)$ is irrelevant to μ, $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$ is a pivotal quantity.

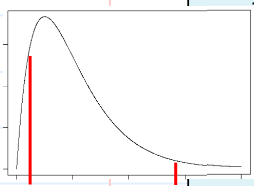
4. Let $X_1, \ldots, X_n$ be i.i.d. from Normal distribution $N(\mu, \sigma^2)$, where both μ and σ are unknown. Then,
$$\frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t_{n-1},$$
where $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$. Because the pdf of t_{n-1} is irrelevant to μ and σ^2, $\frac{\sqrt{n}(\overline{X}-\mu)}{S}$ is a pivotal quantity for μ.
5. Let $X_1, \ldots, X_n$ be i.i.d. from Normal distribution $N(\mu, \sigma^2)$, where both μ and σ are unknown. Then,
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$
Because the pdf of χ^2_{n-1} is irrelevant to μ and σ, $\frac{(n-1)S^2}{\sigma^2}$ is a pivotal quantity for σ^2.

Theorem 6.21 (exact distribution method for constructing confidence intervals of θ)

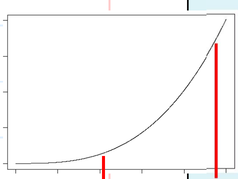
Let $Q(\mathbf{X}, \theta)$ be a pivotal quantity, and $\mathbf{A}$ be a set such that
$$P(Q(\mathbf{X}, \theta) \in \mathbf{A}) = 1 - \alpha,$$
where $\mathbf{A}$ does not depend on all parameters. Then, the set
$$\{\theta : Q(\mathbf{x}, \theta) \in \mathbf{A}\}$$
is a $100(1-\alpha)\%$ confidence interval (or confidence set) of θ.

Example 6.39 (cont. Ex. 6.38, exact confidence intervals, TBp. 279-281)

1. Let a and b satisfy $P(\Gamma(n,1) \le a) = \alpha/2$ and $P(\Gamma(n,1) \ge b) = \alpha/2$, respectively, and $\mathbf{A} = (a, b)$, then

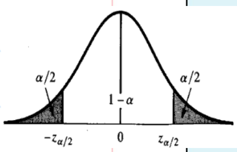


$$1 - \alpha = P\left(a < \lambda \sum_{i=1}^n X_i < b\right) = P\left(\frac{a}{n\overline{X}} < \lambda < \frac{b}{n\overline{X}}\right).$$
Hence, $(a/n\overline{X}, b/n\overline{X})$ is a $100(1-\alpha)\%$ confident interval of λ.
2. Let a and b satisfy $\int_0^a nq^{n-1} dq = \frac{\alpha}{2}$ and $\int_b^1 nq^{n-1} dq = \frac{\alpha}{2}$, respectively (whose solutions are $a = (\frac{\alpha}{2})^{\frac{1}{n}}$ and $b = (1 - \frac{\alpha}{2})^{\frac{1}{n}}$). Let $\mathbf{A} = (a, b)$, then
$$\begin{aligned} 1 - \alpha &= P\left(a < \frac{X_{(n)}}{\theta} < b\right) \\ &= P\left(\frac{a}{X_{(n)}} < \frac{1}{\theta} < \frac{b}{X_{(n)}}\right) = P\left(\frac{X_{(n)}}{b} < \theta < \frac{X_{(n)}}{a}\right). \end{aligned}$$



Hence, $(X_{(n)}/b, X_{(n)}/a)$ is a $100(1-\alpha)\%$ confident interval of θ.

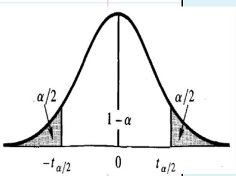
3. Let $z(\alpha)$ denote the point beyond which the $N(0,1)$ distribution has probability α. Note that $z(\alpha) = -z(1-\alpha)$. Then,



$$\begin{aligned} 1-\alpha &= P\left(-z(\alpha/2) < \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} < z(\alpha/2)\right) \\ &= P\left(\overline{X} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

Hence, $\overline{X} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$ forms a $100(1-\alpha)\%$ confident interval of μ.

4. Let $t_n(\alpha)$ denote the point beyond which the t_n distribution has probability α. Note that $t_n(\alpha) = -t_n(1-\alpha)$. Then,

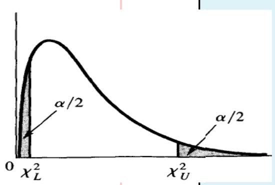


$$\begin{aligned} 1-\alpha &= P\left(-t_{n-1}(\alpha/2) < \frac{\sqrt{n}(\overline{X}-\mu)}{S} < t_{n-1}(\alpha/2)\right) \\ &= P\left(\overline{X} - t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right). \end{aligned}$$

Hence, $\overline{X} \pm t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}$ forms a $100(1-\alpha)\%$ confident interval of μ.

5. Let $\chi^2_n(\alpha)$ denote the point beyond which the χ^2_n distribution has probability α. Then,

$$\begin{aligned} 1-\alpha &= P\left(\chi^2_{n-1}(1-\alpha/2) < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1}(\alpha/2)\right) \\ &= P\left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right). \end{aligned}$$



Hence, $\left(\frac{(n-1)S^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)S^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$ forms a $100(1-\alpha)\%$ confident interval of σ^2.

Question: How to find a $100(1-\alpha)\%$ confident interval of σ?

Note. The construction of confidence intervals based on exact distribution requires detailed knowledge of the sampling distribution as well as some cleverness.

Question: Are there other methods that can offer a more systematic general procedure for the construction of C.I.?

Definition 6.24 (asymptotically pivotal quantity)

For data $\mathbf{X} = (X_1, \ldots, X_n)$, a function of $\mathbf{X}$ and parameter θ, denoted by

$$Q_n(\mathbf{X}, \theta) = Q(X_1, \ldots, X_n, \theta),$$

is called **asymptotically pivotal quantity** if the limiting distribution of $Q_n(\mathbf{X}, \theta)$ as $n \to \infty$, is irrelevant to all parameters.

Question 6.12

What theorems/properties you have learned can help us to identify an asymptotically pivotal quantity?

Theorem 6.22 (asymptotic method for constructing confidence intervals, TBp. 281)

The MLE $\hat{\theta}$ satisfies

$$\sqrt{n\, I(\theta_0)}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1),$$

where θ_0 is the true value of parameter. It can be further argued that

$$\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1).$$

because $$\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) = \sqrt{(nI(\hat{\theta}))/(nI(\theta_0))} \cdot \sqrt{nI(\theta_0)} \cdot (\hat{\theta} - \theta_0)$$

and $(nI(\hat{\theta}))/(nI(\theta_0)) \xrightarrow{\mathcal{P}} 1$.

Therefore,

$$1 - \alpha \approx P\left(-z(\alpha/2) \le \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \le z(\alpha/2)\right).$$

and an asymptotic $100(1-\alpha)\%$ confidence interval for θ_0 is

$$\hat{\theta} \pm z(\alpha/2) \frac{1}{\sqrt{nI(\hat{\theta})}}.$$

$\Rightarrow$ (estimate) $\pm$ (critical value)$\times$(estimated standard error)

Note.

Many confidence intervals have a form like the asymptotic confidence interval of MLE, in which

- the center of confidence interval is the estimated value of θ,
- the length of confidence interval depends on:
  1. confidence level $1 - \alpha$,
  2. sample size n, and
  3. a quantity that is related to the distribution assigned in the statistical modeling step.

Question: We prefer to have a C.I. with longer length or shorter length?

Example 6.40 (cont. Ex. 6.19, asymptotic confidence interval for Poisson mean, TBp. 282)

- Let $X_1, \ldots, X_n$ be i.i.d. from Poisson distribution $P(\lambda)$, then the MLE of λ is $\hat{\lambda} = \overline{X}$ and the Fisher information of $P(\lambda)$ is $I(\theta) = \frac{1}{\lambda}$. Hence an asymptotic $100(1-\alpha)\%$ confidence interval for λ is

$$\overline{X} \pm z(\alpha/2)\sqrt{\overline{X}/n},$$

since the sampling distribution of $\overline{X}$ is approximately Normal.

- Data: Asbestos fibers on filters.

$$\hat{\lambda} = 24.9, \qquad s_{\hat{\lambda}} = \sqrt{\hat{\lambda}/n} = 1.04\,.$$

An asymptotic 95% confidence interval for λ is $(\hat{\lambda} \pm 1.96 s_{\hat{\lambda}}) = (22.9, 26.9)$.

Note. (TBp.283)

- For i.i.d. case, the asymptotic variance of MLE $\hat{\theta}$ is

$$1/(n \cdot I(\theta_0)) = 1/E[\, l'(\theta_0)]^2 = -1/E[\, l''(\theta_0)],$$

where θ_0 is the true value of parameter, and l is the log-likelihood of *all* data.

- For non-i.i.d. case, such as the Multinomial$(n, p_1, \ldots, p_r)$ r.v.'s $(X_1, \ldots, X_r)$,
  - the asymptotic variance of MLE is not of the form $\dfrac{1}{r\, I_{X_1}(\theta_0)}$.
  - However, under some regularity conditions, it can be shown that for non-i.i.d. case,
    - the MLE is asymptotically normal, and
    - its asymptotic variance still equals

$$\frac{1}{E[l'(\theta_0)]^2} \quad \text{or} \quad -\frac{1}{E[l''(\theta_0)]},$$

where l is the log-likelihood of *all* data.

Example 6.41 (cont. Ex. 6.15, asymptotic C.I. for Hardy-Weinberg Equilibrium, TBp. 283)

- Suppose $(X_1, \ldots, X_r) \sim$ Multinomial$(n, p_1, \ldots, p_r)$. Note that the counts $X_1, \ldots, X_r$ are not independent.
- Let us return to the example of Hardy-Weinberg equilibrium, in which the data $(X_1, X_2, X_3) \sim$ Multinomial(n, p_1, p_2, p_3), where

$$p_1 = (1-\theta)^2, \quad p_2 = 2\theta(1-\theta), \quad \text{and} \quad p_3 = \theta^2,$$

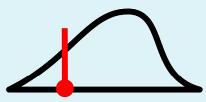
for $0 < \theta < 1$.

- $$l'(\theta) = -\frac{2X_1 + X_2}{1-\theta} + \frac{2X_3 + X_2}{\theta}$$
$$l''(\theta) = -\frac{2X_1 + X_2}{(1-\theta)^2} - \frac{2X_3 + X_2}{\theta^2}$$
$$E(X_1) = n(1-\theta)^2, \quad E(X_2) = 2n\theta(1-\theta), \quad E(X_3) = n\theta^2$$
$$-E[l''(\theta)] = \frac{2n}{\theta(1-\theta)} \qquad \left(\text{Ec: show that } E[l'(\theta)]^2 = \frac{2n}{\theta(1-\theta)}. \right)$$
$$s_{\hat{\theta}} = \sqrt{-\frac{1}{E[l''(\hat{\theta})]}} = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{2n}}$$

- Data: Chinese population of Hong Kong in 1937.
$$\hat{\theta} = 0.4247, \quad \text{and} \quad s_{\hat{\theta}} = 0.011.$$
Hence an approximate 95% confidence interval for θ is
$$\hat{\theta} \pm 1.96\, s_{\hat{\theta}} = (0.403, 0.447).$$
Note that this estimated standard error of $\hat{\theta}$ agree with bootstrap estimate in Ex. 6.15, LNp.26.

Definition 6.25 (confidence intervals constructed by bootstrap, TBp. 284)

- $\hat{\theta}$: an estimator of a parameter θ based on the sample $X_1, \ldots, X_n$
- θ_0: the true, unknown value of θ
- $\underline{\delta}$, $\bar{\delta}$: $\alpha/2$ and $1-\alpha/2$ quantiles of the distribution of $\Delta = \hat{\theta} - \theta_0$, i.e.,
$$\mathrm{P}\left(\hat{\theta} - \theta_0 \le \underline{\delta}\right) = \alpha/2, \quad \text{and}$$
$$\mathrm{P}\left(\hat{\theta} - \theta_0 \le \bar{\delta}\right) = 1 - \alpha/2.$$



Then
$$1-\alpha = \mathrm{P}\left(\underline{\delta} \le \hat{\theta} - \theta_0 \le \bar{\delta}\right) = \mathrm{P}\left(\hat{\theta} - \bar{\delta} \le \theta_0 \le \hat{\theta} - \underline{\delta}\right)$$
and $\left(\hat{\theta} - \bar{\delta}, \hat{\theta} - \underline{\delta}\right)$ is a $100(1-\alpha)\%$ confidence interval for θ.
- Problem: distribution of Δ is unknown. So,
  1. Let us pretend that the *estimate* of θ, $\hat{\theta}_0$, is the true parameter:
$$\theta_0 \text{ (an unknown value)} \longrightarrow \hat{\theta}_0 \text{ (a known value)}$$
$$\hat{\theta} \text{ (an unspecified r.v.)} \longrightarrow \hat{\theta}^* = \hat{\theta}\,|_{\theta=\hat{\theta}_0} \text{ (a specified r.v.)}$$
$$\begin{array}{ccccccc} \Delta & = & \hat{\theta} & - & \theta_0 & \Longrightarrow & \underline{\delta}, \quad \bar{\delta} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \quad \downarrow \\ \Delta^* & = & \hat{\theta}^* & - & \hat{\theta}_0 & \Longrightarrow & \underline{\delta}^*, \quad \bar{\delta}^* \end{array}$$

2. generate many (say, B) bootstrap samples $X_1^*, \ldots, X_n^*$ from a distribution with parameter value being $\hat{\theta}_0$
3. construct for each sample, an estimate of θ, say $\hat{\theta}_j^*$, $j = 1, 2, \ldots, B$
4. $\{\hat{\theta}_1^* - \hat{\theta}_0, \ldots, \hat{\theta}_B^* - \hat{\theta}_0\}$ can be regarded as B samples generated from Δ^*
5. distribution of $\Delta = \hat{\theta} - \theta$ then is approximated by the B samples of $\Delta^* = \hat{\theta}^* - \hat{\theta}_0$.

Example 6.42 (cont. Ex. 6.41, bootstrap C.I. for Hardy-Weinberg Equilibrium, TBp. 284)

- Recall that in Ex. 6.15, LNp.26, $\hat{\theta}_0 = 0.4247$.
- Of 1000 bootstrap estimates $\hat{\theta}_1^*, \ldots, \hat{\theta}_{1000}^*$, the 25^{th} largest is 0.403, which is our estimate of the 0.025 quantile; the 975^{th} largest is 0.446, which is our estimate of the 0.975 quantile.
- Estimates of the 0.025 and 0.975 quantiles of the distribution $\hat{\theta} - \theta$ are

$$\underline{\delta}^* = 0.403 - 0.4247 = -0.0217, \qquad \bar{\delta}^* = 0.446 - 0.4247 = 0.0213.$$

- An approximate 95% C.I. for θ is

$$\left(\hat{\theta}_0 - \bar{\delta}^*, \hat{\theta}_0 - \underline{\delta}^*\right) = (0.404, 0.447),$$

which is about the same as that in Ex. 6.41, LNp.87.

Example 6.43 (cont. Ex. 6.17, bootstrap C.I. for Gamma distribution, TBp. 285)

- Recall that in Ex. 6.17, LNp.29, $\hat{\alpha}_0 = 0.441$, $\hat{\lambda}_0 = 1.96$.
- Of 1000 bootstrap estimates, $\hat{\alpha}_1^*, \ldots, \hat{\alpha}_{1000}^*$, the 50^{th} largest is 0.419, which is our estimate of the 0.05 quantile; the 950^{th} largest is 0.538, which is our estimate of the 0.95 quantile.
- Estimates of the 0.05 and 0.95 quantiles of the distribution $\hat{\alpha} - \alpha$ are

$$\underline{\delta}^* = 0.419 - 0.441 = -0.022, \quad \bar{\delta}^* = 0.538 - 0.441 = 0.097.$$

- An approximate 90% C.I. for α is $\left(\hat{\alpha}_0 - \bar{\delta}^*, \hat{\alpha}_0 - \underline{\delta}^*\right) = (0.344, 0.463)$.
- Similarly, an approximate 90% confidence interval for λ is (1.462, 2.321).

Example 6.44 (sample size determination)

- Question :
  - Suppose that $X_1, \ldots, X_n$ are i.i.d. from $N(\mu, \sigma^2)$, where μ is the average daily yield of a chemical (in tons).
  - Suppose that we would like to estimate μ and we wish the error of estimation to be less than 5 tons with probability 0.95, i.e.

$$P(|\hat{\mu} - \mu| < 5) = 0.95 \iff P(\mu \in (\hat{\mu} - 5, \hat{\mu} + 5)) = 0.95.$$

  - How many measurements should be included in the sample (i.e., n=?) to reach the desired accuracy?

- Suppose that σ is known. The $100(1-\alpha)\%$ exact confidence interval for μ is

$$\left(\overline{X} - z(\alpha/2)\frac{\sigma}{\sqrt{n}},\ \overline{X} + z(\alpha/2)\frac{\sigma}{\sqrt{n}}\right).$$

So,

$$z(\alpha/2)\frac{\sigma}{\sqrt{n}} \le 5 \implies n \ge \frac{z(\alpha/2)^2\sigma^2}{5^2},$$

where $1-\alpha = 0.95$, i.e., $\alpha = 0.05$.

- Suppose that σ is unknown. The $100(1-\alpha)\%$ exact confidence interval for μ is

$$\left(\overline{X} - t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}}\right).$$

So,

$$t_{n-1}(\alpha/2)\frac{S}{\sqrt{n}} \le 5 \implies n \ge \frac{t_{n-1}(\alpha/2)^2 S^2}{5^2},$$

where $1-\alpha = 0.95$ (i.e., $\alpha = 0.05$.), S must be obtained from a previous sample. Note that $t_{n-1} \approx N(0,1)$ when n is large.

❖ **Reading**: textbook, 7.3.3, 8.5.3; **Further reading**: Wackerly et al., 8.5, 8.6, 8.7, 8.8, 8.9