Estimation

Contents

1 Probabilistic models

Definition 1.1 (Scientific models)**. Scientific models** are attempts to describe *physical conditions or changes* that occur in the world and universe around us.

Exm

Example 1.1 (Scientific models in epidemiology)**.** Epidemiologists typically study *biological conditions and changes*, such as the spread of infectious diseases through populations, or the effects of environmental factors on individuals.

1.1 All models are wrong, some are useful

Box and Draper (1987), p424 (emphasis added):

> …Essentially, all models are wrong, but some are useful. **However, the approximate nature of the model must always be borne in mind.**

see also Dunn and Smyth (2018), §1.8

1.2 Statistical analysis of scientific models

When we perform statistical analyses, we use data to help us choose between models - specifically, to determine which models best explain that data.

However, physical processes do not produce data on their own. Data is only produced when scientists implement an *observation process* (i.e., a *scientific study*), which is distinct from the underlying *physical process*. In some cases, the observation process and the physical process interact with each other. This phenomenon is called the "observer effect"[1].

In order to learn about the physical processes we are ultimately interested in, we often need to make special considerations for the observation process that produced the data which we are analyzing. In particular, if some of the planned observations in the study design were not completed, we will likely need to account for the incompleteness of the resulting data set in our analysis. If we are not sure why some observations are incomplete, we may need to model the observation process in addition to the physical process we were originally interested in. For example, if some participants in a study dropped out part-way through the study, we may need investigate why those participants dropped out, as opposed to other participants who completed the study.

These kinds of *missing data* issues are outside of the scope of this course; see Van Buuren (2018) for more details.

2 Estimands, estimates, and estimators

2.1 Estimands

Definition 2.1 (Estimand)**.** An **estimand** is an unknown quantity whose value we want to know (Pohl et al. 2021; Lawrance et al. 2020).

Exm

Example 2.1 (Mean height of students)**.** If we are trying to determine the mean height of students at our school, then the *population mean* is our **estimand**.

In statistical contexts, most estimands are parameters of probabilistic models, or functions of model parameters.

Notation for estimands

Model paramaters and other estimands are often symbolized using lower-case Greek letters: $\alpha, \beta, \gamma, \delta$, etc.

2.2 Estimates

Definition 2.2 (Estimate/estimated value)**.** In statistics, an **estimate** or **estimated value** is an informed guess of an estimand's value, based on observed data.

Exm

Example 2.2 (Mean height of students)**.** Suppose we measure the heights of 50 random students from our school, and the sample mean was 175cm. We might use 175cm as an *estimate* of the population mean.

[1] https://en.wikipedia.org/wiki/Observer_effect

2.3 Estimators

Definition 2.3 (Estimator)**.** An **estimator** is a function $\hat{\theta}(x_1, ...x_n)$ that transforms data $x_1, ...x_n$ into an estimate.

i Estimators are random variables

When estimators are applied to random variables, the estimators are also random variables.

i Notation for estimators

Estimators are often symbolized by placing a $\widehat{\ }$ ("hat") symbol on top of the corresponding estimand; for example, $\hat{\theta}$.
Usually, their dependence on the data is implicit:

$$\hat{\theta} \stackrel{\text{def}}{=} \hat{\theta}(x_1, ...x_n)$$

Exm

Example 2.3 (Mean height of students)**.** If we want to estimate the mean height of students at our university, which we will represent as μ, we might measure the heights of $n = 50$ randomly sampled students as random variables $X_1, ..., X_n$. Then we could use the function

$$\hat{\mu}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{\text{def}}{=} \bar{X}$$

as an *estimator* to produce an *estimate* $\hat{\mu} = \bar{x}$ of μ.
Another estimator would be just the height of the first student sampled:

$$\hat{\mu}^{(2)}(X_1, ..., X_n) = X_1$$

A third possible estimator would be the mean of all sampled students' heights, except for the two most extreme; specifically, if we re-order the observations $X_{(1)} = \min_{i \in 1:n} X_i$, $X_{(2)} = \min_{i \in \{1:n\} - \arg X_{(1)}} X_i$, ..., $X_{(n)} = \max_{i \in 1:n} X_i$, then we could define the estimator:

$$\hat{\mu}^{(3)}(X_1, ..., X_n) = \frac{1}{n}\sum_{i=2}^{n-1} X_{(i)}$$

Which of these estimators is best? It depends on how we evaluate them (see Section 3).

2.4 Contrasting estimands, estimates, and estimators

It's helpful to keep in mind the mathematical type of each estimation concept:

- estimands are numbers (or vector of numbers)
- estimates are also numbers (or vectors)
- estimators are functions of random variables, so they are also random variables

3 Accuracy of estimators

3.1 Accuracy

To determine which estimator is best, we need to define *best*. "Accuracy" is usually most important; easy computation is usually secondary.

Definition 3.1 (Accuracy)**.** The **accuracy** of an estimator for a given estimand does not have a consensus formal definition, but all of the usual candidates are related to the distributions of the *estimation errors* made by the resulting estimates.

3.2 Estimation error

Definition 3.2 (Estimation error)**.** The **estimation error** of an estimate $\hat{\theta}$ of a true value θ is the difference between the estimate and the estimand θ:

$$\varepsilon\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \hat{\theta} - \theta$$

3.3 Residuals

See Linear-model residual definitions and terminology[2] for residual definitions and for the relationship between residuals, model deviations, and estimation error.

Some frequently-used measures of accuracy include:

3.4 Mean squared error

Definition 3.3 (Mean squared error)**.** The **mean squared error** of an estimator $\hat{\theta}$, denoted $\operatorname{MSE}\left(\hat{\theta}\right)$, is the expectation of the square of the estimation error[3]:

$$\operatorname{MSE}\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \operatorname{E}\left[\left(\varepsilon\left(\hat{\theta}\right)\right)^2\right]$$

3.5 Mean absolute error

Definition 3.4 (Mean absolute error)**.** The **mean absolute error** of an estimator is the expectation of the absolute value of the estimation error:

$$\operatorname{MAE}\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \operatorname{E}\left[\left|\varepsilon\left(\hat{\theta}\right)\right|\right]$$

3.6 Bias

Definition 3.5 (Bias)**.** The **bias** of an estimator $\hat{\theta}$ for an estimand θ is the expected value of the estimation error:

$$\operatorname{Bias}\left(\hat{\theta}\right) \stackrel{\text{def}}{=} \operatorname{E}\left[\varepsilon\left(\hat{\theta}\right)\right] \tag{1}$$

Theorem 3.1 (Bias equals Expectation minus Truth)**.**

$$\operatorname{Bias}\left(\hat{\theta}\right) = \operatorname{E}\left[\hat{\theta}\right] - \theta$$

[2] Linear-models-overview.qmd#sec-lm-residuals
[3] https://en.wikipedia.org/wiki/Does_exactly_what_it_says_on_the_tin

Proof

Proof.

$$\begin{aligned}\operatorname{Bias}\big(\hat{\theta}\big) &\stackrel{\text{def}}{=} \operatorname{E}\big[\varepsilon\big(\hat{\theta}\big)\big] \\ &= \operatorname{E}\big[\hat{\theta} - \theta\big] \\ &= \operatorname{E}\big[\hat{\theta}\big] - \operatorname{E}[\theta] \\ &= \operatorname{E}\big[\hat{\theta}\big] - \theta\end{aligned}$$

The third equality is by the linearity of expectation. □

Theorem 3.2 (Mean Squared Error equals Bias Squared plus Variance)**.** *For any one-dimensional estimator $\hat{\theta}$:*

$$\operatorname{MSE}\big(\hat{\theta}\big) = \Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 + \operatorname{Var}\big(\hat{\theta}\big) \tag{2}$$

Proof

Proof. Let's start by expanding each term of the right-hand side:

$$\begin{aligned}\Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 &= \Big(\operatorname{E}\big[\hat{\theta}\big] - \theta\Big)^2 \\ &= \Big(\operatorname{E}\big[\hat{\theta}\big]\Big)^2 - 2\operatorname{E}\big[\hat{\theta}\big]\theta + \theta^2\end{aligned}$$

$$\operatorname{Var}\big(\hat{\theta}\big) = \operatorname{E}\big[\hat{\theta}^2\big] - \Big(\operatorname{E}\big[\hat{\theta}\big]\Big)^2$$

Now, add them together and simplify:

$$\begin{aligned}\Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 + \operatorname{Var}\big(\hat{\theta}\big) &= \Big(\operatorname{E}\big[\hat{\theta}\big]\Big)^2 - 2\operatorname{E}\big[\hat{\theta}\big]\theta + \theta^2 + \operatorname{E}\big[\hat{\theta}^2\big] - \Big(\operatorname{E}\big[\hat{\theta}\big]\Big)^2 \\ &= \operatorname{E}\big[\hat{\theta}^2\big] - 2\operatorname{E}\big[\hat{\theta}\big]\theta + \theta^2\end{aligned}$$

Now let's expand the left-hand side to reach the same expression:

$$\begin{aligned}\operatorname{MSE}\big(\hat{\theta}\big) &= \operatorname{E}\Big[\Big(\varepsilon\big(\hat{\theta}\big)\Big)^2\Big] \\ &= \operatorname{E}\big[(\hat{\theta} - \theta)^2\big] \\ &= \operatorname{E}\big[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\big] \\ &= \operatorname{E}\big[\hat{\theta}^2\big] - \operatorname{E}\big[2\hat{\theta}\theta\big] + \operatorname{E}[\theta^2] \\ &= \operatorname{E}\big[\hat{\theta}^2\big] - 2\operatorname{E}\big[\hat{\theta}\big]\theta + \theta^2\end{aligned}$$

$\operatorname{MSE}\big(\hat{\theta}\big)$ and $\Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 + \operatorname{Var}\big(\hat{\theta}\big)$ both equal $\operatorname{E}\big[\hat{\theta}^2\big] - 2\operatorname{E}\big[\hat{\theta}\big]\theta + \theta^2$. Equality is transitive, so $\operatorname{MSE}\big(\hat{\theta}\big)$ and $\Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 + \operatorname{Var}\big(\hat{\theta}\big)$ are equal to each other:

$$\operatorname{MSE}\big(\hat{\theta}\big) = \Big(\operatorname{Bias}\big(\hat{\theta}\big)\Big)^2 + \operatorname{Var}\big(\hat{\theta}\big)$$

□

3.6.1 Unbiased estimators

Definition 3.6 (Unbiased estimator)**.** An estimator $\hat{\theta}$ is **unbiased** if $\text{Bias}\left(\hat{\theta}\right) = 0$.

Theorem 3.3 (Properties of unbiased estimators)**.** *If $\hat{\theta}$ is unbiased, then:*

$$\text{E}\left[\hat{\theta}\right] = \theta \tag{3}$$

$$\text{MSE}\left(\hat{\theta}\right) = \text{Var}\left(\hat{\theta}\right) \tag{4}$$

Proof

Proof. If $\hat{\theta}$ is unbiased, then:
Equation 3:

$$\begin{aligned} \text{Bias}\left(\hat{\theta}\right) &= 0 \\ \text{E}\left[\hat{\theta}\right] - \theta &= 0 \\ \text{E}\left[\hat{\theta}\right] &= \theta \end{aligned}$$

Equation 4:

$$\begin{aligned} \text{MSE}\left(\hat{\theta}\right) &\overset{\text{def}}{=} \text{E}\left[\left(\varepsilon\left(\hat{\theta}\right)\right)^2\right] \\ &= \text{E}\left[\left(\hat{\theta} - \theta\right)^2\right] \\ &= \text{E}\left[\left(\hat{\theta} - \text{E}\left[\hat{\theta}\right]\right)^2\right] \\ &\overset{\text{def}}{=} \text{Var}\left(\hat{\theta}\right) \end{aligned}$$

(*Alternative proof of Equation 4*) We could have started from Theorem 3.2 instead:

$$\begin{aligned} \text{MSE}\left(\hat{\theta}\right) &= \left(\text{Bias}\left(\hat{\theta}\right)\right)^2 + \text{Var}\left(\hat{\theta}\right) \\ &= (0)^2 + \text{Var}\left(\hat{\theta}\right) \\ &= 0 + \text{Var}\left(\hat{\theta}\right) \\ &= \text{Var}\left(\hat{\theta}\right) \end{aligned}$$

□

3.7 Standard error

Definition 3.7 (Standard error)**.** The **standard error** of an estimator $\hat{\theta}$ is just the standard deviation[a] of $\hat{\theta}$; specifically:

$$\text{SE}\left(\hat{\theta}\right) \overset{\text{def}}{=} \text{SD}\left(\hat{\theta}\right)$$

[a] probability.qmd#def-sd

"Standard error" is a confusing concept in a few ways. First of all, it isn't even defined as a characteristic of the estimation error, $\varepsilon\left(\hat{\theta}\right)$! Moreover, it is just a synonym for standard deviation, so it seems like a redundant concept. However, standard errors help us construct p-values and

confidence intervals, so they come up a lot - often enough to give them their own name.

We can relate standard error to estimation error, by rearranging the result from Theorem 3.2:

$$\begin{aligned}\mathrm{Var}\big(\hat{\theta}\big) &= \mathrm{Var}\big(\hat{\theta} - \theta\big) \\ &= \mathrm{Var}\big(\varepsilon\big(\hat{\theta}\big)\big)\end{aligned}$$

So the variance of the estimator is equal to the variance of the estimation error, and the standard error is equal to the standard deviation of the estimation error:

$$\mathrm{SE}\big(\hat{\theta}\big) = \mathrm{SD}\big(\varepsilon\big(\hat{\theta}\big)\big)$$

Corollary 3.1 (Standard error squared equals MSE minus squared bias)**.** *standard error is what is left over of MSE after bias is removed:*

$$\big(\mathrm{SE}\big(\hat{\theta}\big)\big)^2 = \mathrm{MSE}\big(\hat{\theta}\big) - \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2$$

Proof

Proof.

$$\begin{aligned}\mathrm{MSE}\big(\hat{\theta}\big) &= \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 + \mathrm{Var}\big(\hat{\theta}\big) \\ \therefore \mathrm{Var}\big(\hat{\theta}\big) &= \mathrm{MSE}\big(\hat{\theta}\big) - \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2 \\ \therefore \big(\mathrm{SE}\big(\hat{\theta}\big)\big)^2 &= \mathrm{MSE}\big(\hat{\theta}\big) - \big(\mathrm{Bias}\big(\hat{\theta}\big)\big)^2\end{aligned}$$

□

Corollary 3.2 (For unbiased estimators, SE = RMSE)**.** *If* $\mathrm{E}\big[\varepsilon\big(\hat{\theta}\big)\big] = 0$*, then:*

$$\mathrm{SE}\big(\hat{\theta}\big) = \sqrt{\mathrm{MSE}\big(\hat{\theta}\big)}$$

(this result is equivalent to Equation 4)

Exercises

Exercise 3.1 (Binomial likelihood (adapted from Dobson and Barnett (2018), Chapter 3))**.** Let $Y \sim \mathrm{Binomial}(n, \pi)$, so that

$$\mathrm{p}(Y = y \mid \pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad y \in \{0, 1, \ldots, n\}.$$

Assume $0 < y < n$ so the MLE lies in the interior of $(0, 1)$.
(a) Write the log-likelihood $\ell(\pi; y)$ for a single observation y.
(b) Derive the score function $\ell'(\pi; y) \overset{\text{def}}{=} \frac{\partial}{\partial \pi} \ell(\pi; y)$.
(c) Set the score equal to zero and solve for $\hat{\pi}_{ML}$. Confirm that $\hat{\pi}_{ML} = y/n$.
(d) Compute the second derivative $\ell''(\pi; y)$ and verify that it is negative, confirming a maximum.

Solution

Solution. **(a)**

$$\begin{aligned}\ell(\pi; y) &= \log\left\{ \binom{n}{y} \pi^y (1-\pi)^{n-y} \right\} \\ &= \log\left\{ \binom{n}{y} \right\} + y \log\{\pi\} + (n-y)\log\{1-\pi\} \\ &\propto y \log\{\pi\} + (n-y)\log\{1-\pi\}\end{aligned}$$

(b)

$$\begin{aligned}\ell'(\pi; y) &= \frac{\partial}{\partial \pi}[y \log\{\pi\} + (n-y)\log\{1-\pi\}] \\ &= \frac{y}{\pi} - \frac{n-y}{1-\pi} \\ &= \frac{y(1-\pi) - (n-y)\pi}{\pi(1-\pi)} \\ &= \frac{y - n\pi}{\pi(1-\pi)}\end{aligned}$$

(c)
Setting $\ell'(\pi; y) = 0$:

$$\begin{aligned}0 &= \frac{y - n\pi}{\pi(1-\pi)} \\ y &= n\pi \\ \hat{\pi}_{ML} &= \frac{y}{n}\end{aligned}$$

(d)

$$\begin{aligned}\ell''(\pi; y) &= \frac{\partial}{\partial \pi}\left[\frac{y}{\pi} - \frac{n-y}{1-\pi}\right] \\ &= -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2}\end{aligned}$$

Since $y \geq 0$, $n - y \geq 0$, and $\pi \in (0, 1)$, each term is non-positive, and $\ell''(\hat{\pi}_{ML}; y) < 0$ whenever $0 < y < n$, confirming $\hat{\pi}_{ML} = y/n$ is a maximum.

Exercise 3.2 (Gaussian log-likelihood (adapted from Kleinbaum et al. (2014), Chapter 5))**.** Let $X_1, \ldots, X_n \sim_{\text{iid}} \text{N}(\mu, \sigma^2)$.
(a) Write the likelihood $\mathcal{L}(\mu, \sigma^2; \tilde{x})$ for the observed data $\tilde{x} = (x_1, \ldots, x_n)$.
(b) Write the log-likelihood $\ell(\mu, \sigma^2; \tilde{x})$. Show that it can be written as

$$\ell(\mu, \sigma^2; \tilde{x}) = -\frac{n}{2}\log\{2\pi\sigma^2\} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

(c) Derive the MLE $\hat{\mu}_{ML}$ and $\hat{\sigma}^2_{ML}$.

Solution

Solution. **(a)**

$$\mathcal{L}(\mu, \sigma^2; \tilde{x}) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

(b)

$$\begin{aligned} \ell(\mu, \sigma^2; \tilde{x}) &= \log\{\mathcal{L}(\mu, \sigma^2; \tilde{x})\} \\ &= -\frac{n}{2} \log\{2\pi\sigma^2\} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

(c)
Deriving $\hat{\mu}_{ML}$:

$$\begin{aligned} \frac{\partial}{\partial \mu} \ell &= \frac{\partial}{\partial \mu} \left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right] \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \\ &= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} x_i - n\mu \right) \end{aligned}$$

Setting this to zero: $\hat{\mu}_{ML} = \bar{x} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} x_i$.
Deriving $\hat{\sigma}^2_{ML}$:

$$\frac{\partial}{\partial \sigma^2} \ell = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Setting this to zero and solving:

$$\begin{aligned} \frac{n}{2\sigma^2} &= \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \end{aligned}$$

Substituting $\hat{\mu}_{ML} = \bar{x}$:

$$\hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Note: this maximum likelihood estimator is a biased estimator of σ^2; the unbiased sample variance divides by $n - 1$.

Exercise 3.3 (Score at the MLE (adapted from Dobson and Barnett (2018), Chapter 3))**.** Let $X_1, \ldots, X_n \sim_{\text{iid}} \text{Pois}(\lambda)$.
(a) Write the log-likelihood $\ell(\lambda; \tilde{x})$.
(b) Derive the score function $\ell'(\lambda; \tilde{x})$.
(c) Show that $\hat{\lambda}_{ML} = \bar{x}$, and verify that the score equals zero at the MLE.
(d) Provide an intuitive interpretation: why does the score being zero at $\hat{\lambda}_{ML}$ make sense?

Solution

Solution. **(a)**

$$\begin{aligned}\ell(\lambda;\tilde{x}) &= \log\left\{\prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\right\}\\ &= \sum_{i=1}^{n}(x_i\log\{\lambda\}-\lambda-\log\{x_i!\})\\ &= \left(\sum_{i=1}^{n}x_i\right)\log\{\lambda\}-n\lambda-\sum_{i=1}^{n}\log\{x_i!\}\end{aligned}$$

(b)

$$\begin{aligned}\ell'(\lambda;\tilde{x}) &= \frac{\partial}{\partial\lambda}\left[\left(\sum_{i=1}^{n}x_i\right)\log\{\lambda\}-n\lambda\right]\\ &= \frac{\sum_{i=1}^{n}x_i}{\lambda}-n\\ &= \frac{n\bar{x}}{\lambda}-n\end{aligned}$$

(c)
Setting $\ell'(\lambda;\tilde{x})=0$:

$$\begin{aligned}\frac{n\bar{x}}{\lambda} &= n\\ \hat{\lambda}_{ML} &= \bar{x}\end{aligned}$$

If $\bar{x}>0$ (equivalently, at least one $x_i>0$), then

$$\begin{aligned}\ell'(\bar{x};\tilde{x}) &= \frac{n\bar{x}}{\bar{x}}-n\\ &= n-n\\ &= 0.\end{aligned}$$

If instead all $x_i=0$, then $\bar{x}=0$ and $\hat{\lambda}_{ML}=0$ is a boundary value. In that case, the score formula $\ell'(\lambda;\tilde{x})=\frac{n\bar{x}}{\lambda}-n$ is not defined at $\lambda=0$, so the usual interior verification $\ell'(\hat{\lambda}_{ML};\tilde{x})=0$ does not apply.

(d)
The score measures the rate of change of the log-likelihood. When it equals zero, increasing or decreasing λ slightly would not improve the fit; we are at a "flat" point. Intuitively, $\hat{\lambda}_{ML}=\bar{x}$ is the value of λ that makes the expected count per observation (λ) exactly equal to the observed average count ($\bar{x}$), the best possible match between model and data.

Exercise 3.4 (Standard error of an MLE (adapted from Dobson and Barnett (2018), Chapter 5))**.** Let $X_1,\ldots,X_n \sim_{\text{iid}} \text{Pois}(\lambda)$.
(a) Derive the Hessian $\ell''(\lambda;\tilde{x})=\frac{\partial}{\partial}2]\lambda\ell(\lambda;\tilde{x})$.
(b) Derive the observed information $I(\lambda;\tilde{x})=-\ell''(\lambda;\tilde{x})$.
(c) Evaluate $I(\hat{\lambda}_{ML};\tilde{x})$.
(d) Give an approximate 95% confidence interval for λ using the asymptotic normal distribution of the MLE.

Solution

Solution. **(a)**
From Exercise 3.3, $\ell'(\lambda;\tilde{x})=\frac{n\bar{x}}{\lambda}-n$.

$$\ell''(\lambda;\tilde{x}) = \frac{\partial}{\partial\lambda}\left[\frac{n\bar{x}}{\lambda} - n\right]$$
$$= -\frac{n\bar{x}}{\lambda^2}$$

(b)

$$I(\lambda;\tilde{x}) = -\ell''(\lambda;\tilde{x}) = \frac{n\bar{x}}{\lambda^2}$$

(c)
Assuming $\bar{x} > 0$ (i.e., at least one $x_i > 0$), substituting $\hat{\lambda}_{ML} = \bar{x}$:

$$I(\hat{\lambda}_{ML};\tilde{x}) = \frac{n\bar{x}}{\bar{x}^2} = \frac{n}{\bar{x}}$$

(d)
By the asymptotic theory of MLEs:

$$\hat{\lambda}_{ML} \dot{\sim} \mathrm{N}\Big(\lambda,\ \frac{1}{n\mathcal{J}(\lambda)}\Big)$$

We estimate this asymptotic variance using the observed information evaluated at the MLE. Since $I(\hat{\lambda}_{ML};\tilde{x})^{-1} = \bar{x}/n$:

$$\mathrm{SE}\big(\hat{\lambda}_{ML}\big) \approx \sqrt{\frac{\bar{x}}{n}}$$

An approximate 95% CI for λ is:

$$\hat{\lambda}_{ML} \pm 1.96 \times \sqrt{\frac{\bar{x}}{n}} = \bar{x} \pm 1.96\sqrt{\frac{\bar{x}}{n}}$$

Exercise 3.5 (Exponential MLE (adapted from Dobson and Barnett (2018), Chapter 3))**.** Let $X_1, \ldots, X_n \ \sim_{\text{iid}}\ \text{Exponential}(\mu)$, so that

$$\mathrm{p}(X = x \mid \mu) = \frac{1}{\mu}e^{-x/\mu}, \quad x > 0, \quad \mu > 0.$$

(a) Write the log-likelihood $\ell(\mu;\tilde{x})$ for the observed data $\tilde{x} = (x_1, \ldots, x_n)$.
(b) Derive the score function $\ell'(\mu;\tilde{x}) = \frac{\partial}{\partial\mu}\ell(\mu;\tilde{x})$.
(c) Set the score equal to zero and show that $\hat{\mu}_{ML} = \bar{x}$. Compute the second derivative and verify this critical point is a maximum.
(d) Derive the observed information $I(\mu;\tilde{x}) = -\ell''(\mu;\tilde{x})$, evaluate it at $\hat{\mu}_{ML}$, and give an approximate 95% confidence interval for μ.

Solution

Solution. **(a)**

$$\begin{aligned}
\ell(\mu;\tilde{x}) &= \log\left\{\prod_{i=1}^{n}\frac{1}{\mu}e^{-x_i/\mu}\right\} \\
&= \sum_{i=1}^{n}\log\left\{\frac{1}{\mu}e^{-x_i/\mu}\right\} \\
&= \sum_{i=1}^{n}\left(-\log\{\mu\} - \frac{x_i}{\mu}\right) \\
&= -n\log\{\mu\} - \frac{1}{\mu}\sum_{i=1}^{n}x_i \\
&= -n\log\{\mu\} - \frac{n\bar{x}}{\mu}
\end{aligned}$$

(b)

$$\begin{aligned}
\ell'(\mu;\tilde{x}) &= \frac{\partial}{\partial\mu}\left(-n\log\{\mu\} - \frac{n\bar{x}}{\mu}\right) \\
&= -\frac{n}{\mu} + \frac{n\bar{x}}{\mu^2} \\
&= \frac{n}{\mu^2}(\bar{x}-\mu)
\end{aligned}$$

(c)
Setting $\ell'(\mu;\tilde{x}) = 0$:

$$\begin{aligned}
0 &= \frac{n}{\mu^2}(\bar{x}-\mu) \\
\hat{\mu}_{ML} &= \bar{x}
\end{aligned}$$

The second derivative is:

$$\begin{aligned}
\ell''(\mu;\tilde{x}) &= \frac{\partial}{\partial\mu}\left(-\frac{n}{\mu} + \frac{n\bar{x}}{\mu^2}\right) \\
&= \frac{n}{\mu^2} - \frac{2n\bar{x}}{\mu^3}
\end{aligned}$$

Evaluated at $\hat{\mu}_{ML} = \bar{x}$:

$$\ell''(\bar{x};\tilde{x}) = \frac{n}{\bar{x}^2} - \frac{2n\bar{x}}{\bar{x}^3} = \frac{n}{\bar{x}^2} - \frac{2n}{\bar{x}^2} = -\frac{n}{\bar{x}^2} < 0$$

Since the second derivative is negative, $\hat{\mu}_{ML} = \bar{x}$ is a maximum.
(d)
The observed information is:

$$I(\mu;\tilde{x}) = -\ell''(\mu;\tilde{x}) = \frac{2n\bar{x}}{\mu^3} - \frac{n}{\mu^2}$$

Evaluated at $\hat{\mu}_{ML} = \bar{x}$:

$$I(\hat{\mu}_{ML};\tilde{x}) = \frac{2n\bar{x}}{\bar{x}^3} - \frac{n}{\bar{x}^2} = \frac{2n}{\bar{x}^2} - \frac{n}{\bar{x}^2} = \frac{n}{\bar{x}^2}$$

So $\mathrm{SE}(\hat{\mu}_{ML}) \approx \sqrt{I(\hat{\mu}_{ML};\tilde{x})^{-1}} = \frac{\bar{x}}{\sqrt{n}}$.
An approximate 95% CI for μ is:

$$\hat{\mu}_{ML} \pm 1.96 \times \frac{\bar{x}}{\sqrt{n}} = \bar{x} \pm \frac{1.96\bar{x}}{\sqrt{n}}$$

Note: the exponential distribution has $\mathrm{Var}(X) = \mu^2$, so $\mathrm{SE}(\hat{\mu}_{ML}) = \mu/\sqrt{n}$, which is estimated by $\bar{x}/\sqrt{n}$. More generally, the standard error of a sample mean is $\mathrm{SD}(X)/\sqrt{n}$; here that reduces to $\mu/\sqrt{n}$ because $\mathrm{SD}(X) = \mu$ for the exponential distribution.

References

Box, George E. P., and Norman Richard. Draper. 1987. *Empirical Model-Building and Response Surfaces*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models*. 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

Dunn, Peter K, and Gordon K Smyth. 2018. *Generalized Linear Models with Examples in R*. Vol. 53. Springer. https://link.springer.com/book/10.1007/978-1-4419-0118-7.

Kleinbaum, David G, Lawrence L Kupper, Azhar Nizam, K Muller, and ES Rosenberg. 2014. *Applied Regression Analysis and Other Multivariable Methods*. 5th ed. Cengage Learning. https://www.cengage.com/c/applied-regression-analysis-and-other-multivariable-methods-5e-kleinbaum/9781285051086/.

Lawrance, Rachael, Evgeny Degtyarev, Philip Griffiths, et al. 2020. "What Is an Estimand, and How Does It Relate to Quantifying the Effect of Treatment on Patient-Reported Quality of Life Outcomes in Clinical Trials?" *Journal of Patient-Reported Outcomes* 4 (1): 1–8. https://link.springer.com/article/10.1186/s41687-020-00218-5.

Pohl, Moritz, Lukas Baumann, Rouven Behnisch, Marietta Kirchner, Johannes Krisam, and Anja Sander. 2021. "Estimands—A Basic Element for Clinical Trials." *Deutsches Ärzteblatt International* 118 (51-52): 883–88. https://doi.org/10.3238/arztebl.m2021.0373.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. CRC press. https://stefvanbuuren.name/fimd/.