

ESTIMATION: EXERCISES SOLUTIONS

- (1) 100 students each perform an experiment to estimate a parameter μ , and each one independently computes a 99% confidence interval for μ . What is the probability that there will be at least 3 students whose confidence intervals do not include μ ? (Hint: Use the binomial distribution or the Poisson distribution.)

The probability of a confidence interval not including μ is 0.01 and there are 100 of them, so the expected number X of “bad confidence intervals” is 1. Assuming they are independent, we may model the number X as Poisson with parameter 1. The probability mass function is

$$\mathbb{P}\{X = k\} = e^{-1} \cdot \frac{1^k}{k!}.$$

Thus

$$\begin{aligned}\mathbb{P}\{X \geq 3\} &= 1 - \mathbb{P}\{X \leq 2\} \\ &= 1 - e^{-1} \left(1 + 1 + \frac{1}{2}\right) \\ &= 0.080.\end{aligned}$$

We could also compute this with the R command `1-ppois(2,1)`. The more exact solution would use the binomial distribution with parameters (100,0.01). In R this would be `1-pbinom(2,100,.01)`, yielding 0.079.

- (2) A study was carried out to test the prevalence of side effects from the pertussis vaccine.¹ Of 339 infants who received their first injection of vaccine, 69 showed adverse reactions.

- (a) Compute 95% and 99% confidence intervals for the probability of an adverse reaction to the vaccine. We estimate the probability by $\hat{p} = 69/339 = 0.204$, the observed proportion.

The distribution of the number of adverse reactions is Binomial with parameters (339, p), where p is unknown. The SD of this random number is $\sqrt{339p(1-p)}$. The SD of the proportion of adverse reactions is $\sqrt{p(1-p)/339}$. We then use this to estimate the standard error (which is the same as the SD of \hat{p}) by substituting \hat{p} for the unknown p , to obtain

$$SE \approx \sqrt{0.204 \cdot 0.796/339} = 0.022.$$

The 95% confidence interval for p will be $\hat{p} \pm 1.96 SE = 0.204 \pm 0.043$. The 99% confidence interval for p will be $\hat{p} \pm 2.58 SE = 0.204 \pm 0.056$.

- (b) What do these confidence intervals mean? What assumptions go into the interpretation? We are 95% certain that the true value of p is between 0.161 and 0.247. If we performed similar studies many times, and used this procedure each time to compute confidence intervals, we would expect the first interval to include the true p 95% of the time, and the second 99% of the time.

- (3) Let X_1, \dots, X_n be a sample from a normal distribution with unknown mean μ and unknown variance σ^2 .

¹Published as “Whooping-cough vaccination: An assessment,” Miller et al., *The Lancet*, 1974. Described in Samuels and Witmer, *Statistics for Life Sciences*, p.212.

(a) Write down the likelihood for the parameters (μ, σ) .

$$\begin{aligned} L(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(X_i - \mu)^2} \\ &= (2\pi)^{-n/2} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right\}. \end{aligned}$$

(b) Show that the likelihood may be written in terms of the sample mean and the sample variance. (We say that these are **sufficient statistics**.)

We may write

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n ((X_i - \bar{X}) + (\bar{X} - \mu))^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 + 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \bar{X}) \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 \text{ since } \sum_{i=1}^n (X_i - \bar{X}) = 0 \\ &= (n-1)S^2 + n(\bar{X} - \mu)^2, \end{aligned}$$

where

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

is the sample variance. So

$$L(\mu, \sigma) = (2\pi)^{-n/2} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} ((n-1)S^2 + n(\bar{X} - \mu)^2)\right\}.$$

(c) Compute the MLEs for these parameters.

The log likelihood is

$$\ell(\mu, \sigma) = -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} ((n-1)S^2 + n(\bar{X} - \mu)^2).$$

We maximise by setting the partial derivatives equal to 0. We have

$$\frac{\partial \ell}{\partial \mu} = -\frac{1}{\sigma^2} \cdot n(\mu - \bar{X}),$$

which is 0 only at $\hat{\mu} = \bar{X}$. Also

$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot ((n-1)S^2 + n(\bar{X} - \mu)^2).$$

Setting $\mu = \hat{\mu} = \bar{X}$, we get

$$\frac{\partial \ell}{\partial \sigma}(\bar{X}, \sigma) = -\frac{n}{\sigma} + \frac{(n-1)S^2}{\sigma^3}.$$

This is 0 precisely when $\hat{\sigma} = \sqrt{\frac{n}{n-1}}S$. The MLE for $\hat{\mu}$ is just what we would have expected. The MLE for S is not the unbiased sample variance, but rather the variance of the sample, which is slightly biased, but asymptotically unbiased.

- (d) Compute the Fisher information, and use it to estimate the standard error of $\hat{\mu}$ for large n . How does this compare with the true standard error?

The second partial derivatives are

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \mu^2}(\mu, \sigma) &= -\frac{n}{\hat{\sigma}^2}, \\ \frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \cdot ((n-1)S^2 + n(\bar{X} - \mu)^2) \\ \frac{\partial^2 \ell}{\partial \mu \partial \sigma} &= 0.\end{aligned}$$

The fact that the mixed partials are 0 tells us that the estimators are asymptotically independent. (In fact, they are independent, but that requires a bit more theory.) The second partial wrt μ is always negative. The second partial wrt σ at the critical point is

$$\frac{n-1}{S^2} - \frac{3(n-1)^3}{n^2 S^2},$$

which is also negative. Since both are negative, this tells us that the critical point really is a maximum.

The asymptotic standard error for $\hat{\mu}$ from the Fisher information is thus

$$\sqrt{\frac{\hat{\sigma}^2}{n}},$$

which is exactly the true SD of the estimator $\hat{\mu} = \bar{X}$.

- (4) **[Copied from N. Laws, 2014 Part A Statistics]** In 1931 J. W. MacArthur reported the results of a test of genetic linkage theory in tomato plants. The theory predicts that four phenotypes determined by two bi-allelic genes, where one of each pair of alleles is dominant, should appear in the ratios $\frac{9}{16} + \theta : \frac{3}{16} - \theta : \frac{3}{16} - \theta : \frac{1}{16} + \theta$. He found the following frequencies:

Phenotype	Frequency
Tall, cut-leaf	926
Tall, potato-leaf	288
Dwarf, cut-leaf	293
Dwarf, potato-leaf	104
Total	1611

- (a) Write down the likelihood of θ given these observations.

Write $n_1 = 926$, $n_2 = 288$, $n_3 = 293$, $n_4 = 104$, $n = \sum n_i = 1611$. The likelihood is

$$L(\theta) = \frac{n!}{n_1!n_2!n_3!n_4!} \left(\frac{9}{16} + \theta\right)^{n_1} \left(\frac{3}{16} - \theta\right)^{n_2+n_3} \left(\frac{1}{16} + \theta\right)^{n_4}$$

- (b) Find the MLE of θ .

The log likelihood $\ell(\theta) = \log L(\theta)$ has derivative

$$\begin{aligned}\ell'(\theta) &= 16 \left(\frac{n_1}{9 + 16\theta} - \frac{n_2 + n_3}{3 - 16\theta} + \frac{n_4}{1 + 16\theta} \right) \\ &= \frac{16(-412416\theta^2 - 73312\theta + 357)}{(9 + 16\theta)(3 - 16\theta)(1 + 16\theta)}.\end{aligned}$$

This is 0 if the numerator is 0, which has two solutions: $\theta = \theta_1 = 0.0047$ and $\theta = \theta_2 = -0.1825$. Since θ needs to be between $-1/16$ and $+3/16$, only θ_1 is feasible. So we estimate $\hat{\theta} = 0.0047$.

- (c) Use it to calculate expected frequencies for the four phenotypes, and compare them to the observed frequencies. Does the theory look plausible?

Phenotype	Observed Frequency	Expected Frequency
Tall, cut-leaf	926	$1611(\frac{9}{16} + \hat{\theta}) = 913.8$
Tall, potato-leaf	288	$1611(\frac{3}{16} - \hat{\theta}) = 294.4$
Dwarf, cut-leaf	293	$1611(\frac{3}{16} - \hat{\theta}) = 294.4$
Dwarf, potato-leaf	104	$1611(\frac{1}{16} + \hat{\theta}) = 108.3$
Total	1611	1611

The observed and expected numbers are pretty close, so the theory seems plausible.