Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4295 Statistical inference

Lecturer Fall 2023: *Gunnar Taraldsen*

Scribe: *Kebene Baissa and Elling Svee*

Lecture 2 in week 44: Distibutions, estimation and uncertainty

Contents

1 Confidence sets

Example 1.1 (Bernoulli (continued from previous lecture)) Let the data be $x = 9$ successes out of $n = 10$ trials, and let the prior be $\pi(\theta) = 1$. What is the success probability θ?

Of course, we cannot find the true θ, but we can estimate it! Let the *estimator* be

$$\hat{\theta} = \frac{x}{n} = 90\% \tag{1}$$

Yet, we are still not finished, as we also need to state the *standard uncertainty*, as SI demands. We can let

$$u_1 = \sqrt{\frac{1}{n}\hat{\theta}(1-\hat{\theta})} \tag{2}$$

Need to do some calculations in order to make sure that (2) is a reasonable uncertainty. We have that $B \sim B(\theta)$. The *expectation* is

$$\begin{aligned} EB &= b_1 P(B = b_1) + b_2 P(B = b_2) \\ &= 0 * P(B = 0) + 1 * P(B = 1) \\ &= \theta \end{aligned}$$

Therefore $\frac{x}{n} = \bar{b}$. Additionally, the *variance* is

$$\begin{aligned} VarB &= E(B - EB)^2 \\ &= EB^2 - (EB)^2 \text{, use that } B^2 = B \\ &= \theta(1-\theta) \end{aligned}$$

Therefore $Var\overline{B} = \frac{1}{n}\theta(1-\theta)$ which implies $Std\overline{B} = \sqrt{\frac{1}{n}\theta(1-\theta)}$. This yields that $u = \sqrt{\frac{1}{n}\hat{\theta}(1-\hat{\theta})}$, meaning that the chosen standard uncertainty in (2) makes sense. As it is a function of the minimal sufficient statistic $\bar{b}$, the estimate is also in harmony with the *sufficiency principle.*

Another alternative of a *standard uncertainty* could the

$$u_2 = \frac{1}{\sqrt{n}} s \tag{3}$$

One find that $E(S^2|T)$ where T is a *complete sufficient statistic* improves S^2 as an estimator of σ^2. This is because of the *Rao-Blackwell* and *Lehmann–Scheffé* theorems. It can also be insightful to look at the definition of *conditional expectation.*

For this example, one can also see that $E(S^2|X) = S^2$. This is because of the fact that S is a function of X.

Want to show that S^2 is a function of $\hat{\theta}$.

$$\begin{aligned} S^2 &= \frac{n}{n-1}\overline{\left(b-\bar{b}\right)^2} \\ &= \frac{n}{n-1}\left(\overline{b^2} - (\bar{b})^2\right) \end{aligned}$$

The *empirical distribution* is $\frac{1}{n}\left(\delta_{x_1} + \ldots + \delta_{x_n}\right)$ where δ_{x_i} is the *point mass* at point x_i. This is a *simple distribution.* In our case the *empirical distribution* will be

$$\frac{1}{n}\left(x\delta_1 + (n-x)\delta_0\right) \tag{4}$$

From numerical calculations, we can find that the numerical value for (2) is

$$u_1 = \frac{1}{10} * \frac{3}{\sqrt{10}} \approx 0.095 = 9.5\% \approx 10\%$$

This means that we can state our estimator as

$$\hat{\theta} = \frac{x}{n} = 90(10)\% \tag{5}$$

Note that from numerical calculations, we can also find that the numerical value for (3) is

$$u_2 = \frac{1}{\sqrt{n-1}}\sqrt{\hat{\theta}(1-\hat{\theta})} = \frac{n}{n-1}\frac{x}{n}\left(1 - \frac{x}{n}\right) \approx 10\%$$

It is not surprising that u_2 is a bit larger than u_1 from the (2) case. Note that it is known that $0 \leq \theta \leq 100\%$.

2 Expanded uncertainty

The *expanded uncertainty* is equivalent with the *confidence interval* (or sometimes the *credibility interval*). It will be given as $\tilde{\theta} \pm k_{\pm} u$ with *coverage factor* $k_{\pm} \approx 2$ for 95% coverage.

Note that any *confidence interval* can be stated on the form $\hat{\theta}_{\text{estimator}} \pm k_{\pm} u$.
An approx fix can be to define $\eta = ln\left(\frac{\theta}{1-\theta}\right) \in \mathbb{R}$, which yields

$$\hat{\eta} = ln\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right) \tag{6}$$

This estimator is not *unbiased*. In fact, no *unbiased estimator* φ of η exists!

An *approximate confidence interval* will be

$$\tilde{\eta} \pm 2u_{\eta} \tag{7}$$

By using the *delta method* stated in *ISO GUM* we can find the *approximated confidence interval*

$$[\hat{\eta}_1, \hat{\eta}_2] \Rightarrow [\tilde{\theta}_1, \tilde{\theta}_2] \tag{8}$$

3 Bayes

We have

$$\pi(\theta|x) \propto L(\theta|x)\pi(\theta) = \theta^x(1-\theta)^{n-x} * \theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1}$$

where $L(\theta)$ is the *likelihood* and $\pi(\theta)$ is the *prior distribution*. Then the posterior

$$\pi(\theta|x) \propto \theta^{(x+\alpha_0)-1}(1-\theta)^{(\beta_0 - y)-1}$$

where $y = n - x$. This yields the *posterior distribution*

$$\theta|x \sim Beta\left(x + \alpha_0, y + \beta_0\right) \tag{9}$$

From (9) we can calculate the *Bayes estimator* by finding the posterior expectation

$$E(\theta|x) = \frac{x + \alpha_0}{x + \alpha_0 + y + \beta_0} = \hat{\theta}_B \tag{10}$$

For the *uniform prior* we have $\pi(\theta) = 1$, which yields $\alpha_0 = \beta_0 = 1$. This is *proper*. For a *Jeffreys prior* we get $\alpha_0 = \beta_0 = \frac{1}{2}$. Lastly, for a *Haldane prior* we get $\alpha_0 = \beta_0 = 0$, which yields that $\hat{\theta}_B = \tilde{\theta}$. The Haldane prior is *improper*.

The *Bayes probabilities* can be computed using a link the Bernoulli. This was done by Bayes in 1761.

4 Exact intervals

Start with $H_0 : \theta \leq \theta_0$. Reject if $X \geq x_0$. This is an *optimal test* because of the *Karlin-Rubin theorem*. (To be continued...)

Definitions

Theorems

Examples

Index