

# STAT 238 - Bayesian Statistics

## Lecture Three

Spring 2026, UC Berkeley

Aditya Guntuboyina

26 Jan 2026

In the last lecture, we started discussing this example.

### 1 Example 3: Inference from measurements

Suppose a scientist makes 6 numerical measurements 26.6, 38.5, 34.4, 34, 31, 23.6 on an unknown real-valued physical quantity  $\theta$ . On the basis of these measurements, what can be inferred about  $\theta$ ?

Here is the Bayesian solution to this problem. The first step is modeling where we have to write the likelihood and prior. The likelihood represents the probability of the observed data conditional on parameter values. Here the main parameter is  $\theta$ . In order to write the probability of the observed data, it is helpful to introduce another parameter  $\sigma$  which represents the scale of the noise inherent in the measurement process.

So our parameter vector is  $(\theta, \sigma)$ . We work with the normal likelihood:

$$\text{Likelihood} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right). \quad (1)$$

where  $n = 6$  and  $x_1 = 26.6, x_2 = 38.5, x_3 = 34.4, x_4 = 34, x_5 = 31, x_6 = 23.6$  denote the observed data points. More formally, you can arrive at this likelihood in the following way. Denote potential measurements by  $X_1, \dots, X_n$ . Each actual measurement will have some rounding error so the data point 26.6 should be understood as belonging to the interval  $[26.6 - \delta, 26.6 + \delta]$  for some small rounding error  $\delta$ . So the likelihood is:

$$\begin{aligned} \text{likelihood} &= \mathbb{P}\{\text{observed data} \mid \theta, \sigma\} \\ &= \mathbb{P}\{X_1 \in [x_1 - \delta, x_1 + \delta], \dots, X_n \in [x_n - \delta, x_n + \delta] \mid \theta, \sigma\}. \end{aligned}$$

Assuming  $\delta$  is small, we can use probability-density approximation to write

$$\text{likelihood} \approx \delta^n f_{X_1, \dots, X_n \mid \theta, \sigma}(x_1, \dots, x_n).$$

We are now assuming that:

$$f_{X_1, \dots, X_n \mid \theta, \sigma}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right). \quad (2)$$

This leads to the likelihood (1) (note that  $\delta^n$  is being dropped as it is a constant of proportionality which does not affect any further calculations).

---

Here is a short digression on the likelihood and the assumption (2).

A lot of the time, the likelihood assumption (2) is written as:

$$X_1, \dots, X_n \mid \theta, \sigma \stackrel{\text{i.i.d}}{\sim} N(\theta, \sigma^2). \quad (3)$$

Strictly speaking (2) and (3) are not the same. This is because (3) is equivalent to

$$f_{X_1, \dots, X_n \mid \theta, \sigma}(u_1, \dots, u_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u_i - \theta)^2}{2\sigma^2}\right) \quad \text{for all } u_1, \dots, u_n \in (-\infty, \infty) \quad (4)$$

(4) is much stronger than (2) because  $u_1, \dots, u_n$  are completely arbitrary while in (2) the points  $x_1, \dots, x_n$  are not arbitrary (they simply equal the observed data).

For example, note that if we assume

$$f_{X_1, \dots, X_n \mid \theta, \sigma}(u_1, \dots, u_n) = \begin{cases} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u_i - \theta)^2}{2\sigma^2}\right), & \text{if } u_1, \dots, u_n \in (20, 30), \\ \text{completely arbitrary,} & \text{if } (u_1, \dots, u_n) \notin (20, 30)^n. \end{cases}$$

then again we arrive at the same likelihood. But now (3) is no longer true.

---

To complete the modeling step, we need to describe the prior on  $\theta, \sigma$ . We assume

$$\theta, \log \sigma \stackrel{\text{i.i.d}}{\sim} \text{uniform}(-C, C) \quad (5)$$

for a very large positive constant  $C$ . The idea here is that we are allowing  $\theta$  and  $\log \sigma$  to take values essentially on the entire real line and not expressing preference for any one value over any other. In terms of the densities, (5) is the same as

$$\begin{aligned} \text{prior} &= f_{\theta, \sigma}(\theta, \sigma) = f_{\theta}(\theta) f_{\sigma}(\sigma) \\ &= f_{\theta}(\theta) f_{\log \sigma}(\log \sigma) \frac{1}{\sigma} \\ &= \frac{\mathbf{1}\{-C < \theta < C\}}{2C} \frac{\mathbf{1}\{-C < \log \sigma < C\}}{2C} \frac{1}{\sigma} \\ &\propto \frac{\mathbf{1}\{-C < \theta < C, -C < \log \sigma < C\}}{\sigma}. \end{aligned}$$

To now get the posterior (which is the joint density of  $\theta, \sigma$  given the observed data), we use Bayes rule as

$$\begin{aligned} \text{posterior} &= f_{\theta, \sigma \mid \text{data}}(\theta, \sigma) \\ &\propto f_{\theta, \sigma}(\theta, \sigma) \times \text{likelihood} \\ &\propto \frac{\mathbf{1}\{-C < \theta < C, -C < \log \sigma < C\}}{\sigma} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right) \\ &\propto \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right) \mathbf{1}\{-C < \theta < C, -C < \log \sigma < C\}. \end{aligned}$$

The constant underlying proportionality above is determined by the overall integral being one:

$$\text{posterior} = f_{\theta, \sigma | \text{data}}(\theta, \sigma) = \frac{\sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right) \mathbf{1}\{-C < \theta < C, -C < \log \sigma < C\}}{\int_{-C}^C \int_{e^{-C}}^{e^C} \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right) d\theta d\sigma}.$$

This is the joint posterior density of  $\theta$  and  $\sigma$ . If we only want the posterior density for  $\theta$ , we use the sum rule of probability to integrate out  $\sigma$ :

$$f_{\theta | \text{data}}(\theta) \propto \mathbf{1}\{-C < \theta < C\} \int_{e^{-C}}^{e^C} \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right) d\sigma$$

Because  $C$  is large, the limits of the integral can be taken to be 0 and  $\infty$  leading to

$$\begin{aligned} f_{\theta | \text{data}}(\theta) &\propto \mathbf{1}\{-C < \theta < C\} \int_0^\infty \sigma^{-n-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2\right) d\sigma \\ &= \mathbf{1}\{-C < \theta < C\} 2^{(n/2)-1} \Gamma(n/2) \left[ \sum_{i=1}^n (x_i - \theta)^2 \right]^{-n/2} \end{aligned}$$

The factor  $2^{(n/2)-1} \Gamma(n/2)$  does not depend on  $\theta$  so it can be absorbed in the constant of proportionality as well leading to:

$$f_{\theta | \text{data}}(\theta) \propto \mathbf{1}\{-C < \theta < C\} \left( \frac{1}{S(\theta)} \right)^{n/2}$$

where  $S(\theta)$  is the sum of squares term:

$$S(\theta) = \sum_{i=1}^n (x_i - \theta)^2.$$

If  $C$  is large, then the indicator can be dropped (because it will essentially be always 1) so the posterior becomes:

$$f_{\theta | \text{data}}(\theta) \propto \left( \frac{1}{S(\theta)} \right)^{n/2}.$$

Thus the posterior is inversely proportional to  $S(\theta)^{n/2}$ . This means that the posterior model will be at the least squares estimator which is the mean  $\hat{\theta} = \bar{x} = (x_1 + \dots + x_n)/n$ . It is cleaner to write the above posterior as:

$$f_{\theta | \text{data}}(\theta) \propto \left( \frac{S(\hat{\theta})}{S(\theta)} \right)^{n/2}.$$

Because  $S(\theta) = S(\hat{\theta}) + n(\theta - \hat{\theta})^2$ , we can also rewrite the posterior as:

$$f_{\theta | \text{data}}(\theta) \propto \left( \frac{S(\hat{\theta})}{S(\hat{\theta}) + n(\theta - \hat{\theta})^2} \right)^{n/2} = \left( \frac{1}{1 + \frac{(\theta - \hat{\theta})^2}{S(\hat{\theta})/n}} \right)^{n/2}.$$

It can be shown (left as exercise) that this is related to the  $t$ -density (see [https://en.wikipedia.org/wiki/Student%27s\\_t-distribution](https://en.wikipedia.org/wiki/Student%27s_t-distribution)). Specifically, the above is equivalent to

$$\frac{\sqrt{n}(\theta - \hat{\theta})}{\sqrt{S(\hat{\theta})/(n-1)}} \mid \text{data} \sim t_{n-1} \quad (6)$$

where  $t_{n-1}$  denotes the  $t$ -density with  $n - 1$  degrees of freedom. Note that  $\hat{\theta} = \bar{x}$  and  $S(\hat{\theta}) = \sum_{i=1}^n (x_i - \bar{x})^2$ .

So the Bayesian point estimate is simply  $\hat{\theta} = \bar{x}$  (this is the posterior mean, median and mode!). A  $100(1 - \alpha)\%$  uncertainty interval for  $\theta$  is given by:

$$\left[ \hat{\theta} - \frac{1}{\sqrt{n}} t_{n-1, \alpha/2} \sqrt{\frac{S(\hat{\theta})}{n-1}}, \hat{\theta} + \frac{1}{\sqrt{n}} t_{n-1, \alpha/2} \sqrt{\frac{S(\hat{\theta})}{n-1}} \right] \quad (7)$$

where  $t_{n-1, \alpha/2}$  is the  $(1 - \alpha/2)$  quantile of the Student  $t$ -distribution with  $n - 1$  degrees of freedom. This uncertainty interval is sometimes referred to as the Bayesian Credible Interval.

Check that  $\alpha = 0.05$  leads to the interval  $[25.598, 37.102]$ .

## 1.1 Frequentist Solution

It turns out that the Bayesian credible interval (7) is also the standard frequentist confidence interval in this problem. Specifically, we have (below  $\bar{X} = (X_1 + \dots + X_n)/n$ )

$$\mathbb{P} \left\{ \bar{X} - \frac{t_{n-1, \alpha/2}}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \leq \theta \leq \bar{X} + \frac{t_{n-1, \alpha/2}}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \right\} = 1 - \alpha \quad (8)$$

under the assumption:

$$X_1, \dots, X_n \text{ are i.i.d } N(\theta, \sigma^2).$$

This is because

$$\frac{\sqrt{n}(\bar{X} - \theta)}{S} \sim t_{n-1} \quad \text{where } S := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}. \quad (9)$$

In the above probability statement,  $\theta$  is held fixed, and the probability is taken with respect to  $X_1, \dots, X_n$  which are i.i.d  $N(0, \sigma^2)$ .

Observe the difference between (6) and (9). In (6), the data is held fixed at the observed values and the probability is with respect to  $\theta$ . In (9),  $\theta$  is held fixed and the probability is with respect to the random variables  $X_1, \dots, X_n$  which are supposed to represent data.

In this problem, the standard Bayesian inference and standard frequentist inference exactly coincide.

However, it is very easy to break this coincidence. We shall discuss this in the next lecture.