

STAT 238 - Bayesian Statistics

Lecture Eleven

Spring 2026, UC Berkeley

Aditya Guntuboyina

13 Feb 2026

1 Bayesian Inference with Normal Likelihoods

Data is y and parameter is θ , the likelihood is given by:

$$y | \theta \sim N(\theta, \sigma^2)$$

for a fixed $\sigma > 0$. We will look at the case of unknown σ later. The frequentist estimator of θ is the MLE y . For Bayesian inference, we use the normal $N(\mu, \tau^2)$ prior on θ . The basic fact is:

$$\theta \sim N(\mu, \tau^2) \text{ and } y | \theta \sim N(\theta, \sigma^2) \implies \theta | y \sim N\left(\frac{y/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2}, \frac{1}{1/\sigma^2 + 1/\tau^2}\right)$$
$$\text{and } y \sim N(\mu, \sigma^2 + \tau^2).$$

The mean of the posterior distribution is thus

$$\frac{y/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2} = \frac{1/\sigma^2}{1/\sigma^2 + 1/\tau^2}y + \frac{1/\tau^2}{1/\sigma^2 + 1/\tau^2}\mu$$

which is a weighted linear combination of the prior mean μ and the data y . The weights are inversely proportion to the variances of the corresponding normal distributions. For a normal distribution, we use the term “precision” to denote the inverse of variance. Thus the weights of the linear combination above are proportional to the precisions of the prior and the likelihood.

Also note that the precision of the posterior equals the sum of the prior and likelihood precisions.

Note that when $\tau^2 \rightarrow \infty$ and μ is a fixed constant, we have $\theta | y \sim N(y, \sigma^2)$ (in this case, the posterior mean coincides with the MLE y). Operationally, the $N(\mu, \tau^2)$ prior with fixed μ and $\tau^2 = +\infty$ has the same behavior as the uniform $(-\infty, \infty)$ prior. These are uninformative priors in this problem.

Often the data will consist of multiple numbers y_1, \dots, y_n with the likelihood being:

$$y_1, \dots, y_n \stackrel{\text{i.i.d}}{\sim} N(\theta, \sigma^2).$$

In this case (for the same prior $\theta \sim N(\mu, \tau^2)$), the posterior is

$$\theta | y_1, \dots, y_n \sim N\left(\frac{n\bar{y}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2}, \frac{1}{n/\sigma^2 + 1/\tau^2}\right).$$

where $\bar{y} := (y_1 + \dots + y_n)/n$. This can be proved directly or use the fact that \bar{y} is sufficient for θ and the data can be reduced to the single number \bar{y} with likelihood $N(\theta, \sigma^2/n)$.

2 Multiple Instances of the Same Problem

Consider the problem of estimating $\theta_1, \dots, \theta_N$ from data y_1, \dots, y_N under the likelihood:

$$y_i \stackrel{\text{i.i.d.}}{\sim} N(\theta_i, \sigma^2)$$

for a fixed σ^2 . We further assume:

$$\theta_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \tau^2).$$

If we fix μ to be some constant value (say $\mu = 0$) and τ^2 to be very large, then the estimate of θ_i would be equal to y_i .

Instead of fixing these values, we can treat μ and τ^2 also as unknown parameters and attempt to estimate them from the observed data. Marginalizing θ_i , it is easy to see that

$$y_i \mid \mu, \tau \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2 + \tau^2).$$

One can estimate μ and τ from this model by some estimates $\hat{\mu}$ and $\hat{\tau}$. θ_i is then estimated by $\mathbb{E}(\theta_i \mid y_i, \mu = \hat{\mu}, \tau = \hat{\tau})$. How to obtain $\hat{\mu}$ and $\hat{\tau}$? It is natural to take

$$\hat{\mu} = \frac{y_1 + \dots + y_N}{N}.$$

The estimate of τ should be based on the sample variance:

$$\sum_{i=1}^N (y_i - \bar{y})^2 \sim (\sigma^2 + \tau^2) \chi_{N-1}^2.$$

The formula for $\mathbb{E}(\theta_i \mid y_i, \mu, \tau)$ is:

$$\mathbb{E}(\theta_i \mid y_i, \mu, \tau) = \frac{y_i/\sigma^2 + \mu/\tau^2}{1/\sigma^2 + 1/\tau^2} = \mu + \left(1 - \frac{\sigma^2}{\sigma^2 + \tau^2}\right) (y_i - \mu).$$

So we need to estimate $1/(\sigma^2 + \tau^2)$ as opposed to τ^2 directly. The following fact:

$$\mathbb{E}\left(\frac{1}{\chi_v^2}\right) = \frac{1}{v-2} \quad \text{for } v > 2$$

shows that

$$\mathbb{E}\left(\frac{1}{\sum_{i=1}^n (y_i - \bar{y})^2}\right) = \frac{1}{\sigma^2 + \tau^2} \mathbb{E}\left(\frac{1}{\chi_{N-1}^2}\right) = \frac{1}{(N-3)(\sigma^2 + \tau^2)}.$$

So we use

$$\text{estimate of } \frac{1}{\sigma^2 + \tau^2} = \frac{N-3}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (1)$$

The estimate of τ^2 implied by (1) can be nonpositive (depending on the data and σ^2). We shall ignore this however.

The estimates of θ_i are therefore given by:

$$\hat{\theta}_i^{\text{JS}} := \bar{y} + \left(1 - \frac{(N-3)\sigma^2}{\sum_{i=1}^n (y_i - \bar{y})^2}\right) (y_i - \bar{y}).$$

This is the James-Stein estimator.

The James-Stein estimator can thus be seen as an empirical Bayes procedure. The James-Stein estimator has the following remarkable frequentist property: its risk is uniformly better than the naive estimator $\hat{\theta}_{i,\text{naive}} = y_i$ in mean squared error for all values of $\theta_1, \dots, \theta_N$ (this is true for $N \geq 3$; although we only defined it for $N > 3$, one can create a more naive version of James-Stein replacing \bar{y} by any fixed constant like 0 and this would work also for $N = 3$). This property is beyond the scope of this class (and also irrelevant to us as it is a frequentist property).

Let us now present the full Bayes estimate. This requires specification of (uninformative) priors of μ and τ . For μ , we use the uniform $(-\infty, \infty)$ prior. For τ , we use the fact that the marginal distribution of the data given μ, τ is $y_i \mid \mu, \tau \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2 + \tau^2)$, which is in terms of the parameter $\gamma^2 := \sigma^2 + \tau^2$. For the $N(\mu, \gamma^2)$ model, the standard uninformative prior for γ is

$$\log \gamma \sim \text{uniform}(-\infty, \infty) \quad \text{or} \quad f_\gamma(\gamma) \propto \frac{I\{\gamma > 0\}}{\gamma}.$$

Now we have the additional information that $\gamma > \sigma$ (because $\gamma^2 = \sigma^2 + \tau^2$), so it is natural to modify the above prior as:

$$f_\gamma(\gamma) \propto \frac{I\{\gamma > \sigma\}}{\gamma}.$$

It is easy to check that this prior on γ leads to the following prior on $\tau = \sqrt{\gamma^2 - \sigma^2}$:

$$f_\tau(\tau) \propto \frac{\tau}{\sigma^2 + \tau^2} I\{\tau > 0\}.$$

To recap, we are using the following prior for μ and τ :

$$\mu \sim \text{uniform}(-\infty, \infty) \quad \text{and} \quad f_\tau(\tau) \propto \frac{\tau}{\sigma^2 + \tau^2} I\{\tau > 0\}.$$

We will combine this with the likelihood $y_i \mid \mu, \tau \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2 + \tau^2)$ to obtain the posterior of μ, τ . It can be checked that:

$$\mu \mid \gamma, y_1, \dots, y_N \sim N\left(\bar{y}, \gamma^2/N\right) \quad \text{and} \quad f_{\gamma \mid y_1, \dots, y_N}(\gamma) \propto \gamma^{-N} \exp\left(-\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{2\gamma^2}\right) I\{\gamma > \sigma\}.$$

You will see how to compute these posteriors given data in the next lab.

References

- [1] Efron, B. (2010) *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Cambridge University Press, 2010.