Bayesian Statistics

Nuno Carvalho

May 26, 2025

1 Distribution of a parameter from a noisy estimate

Often times, there is some true parameter, β, we are interested in, but our estimate carries with it some noise, ϵ. If we know (or assume) the prior distributions of the parameter and noise, we can use Bayes' Theorem to derive the expected true parameter value given our estimate. This may be useful if we are producing estimates for a set of parameters that share the same prior distribution, and we want to regularize our estimates down towards zero to avoid overfitting in order to maximize prediction.

Here, we will assume the true parameter value is normally distributed around zero.

1.1 Set-up

Formally, we can write:
$$\hat{\beta} = \beta + \epsilon$$
$$\text{where} \quad \beta \sim \mathcal{N}(0, \sigma_\beta^2) \quad \text{and} \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

meaning that the true parameter value and its measurement noise are independent from each other. Our goal is to obtain the distribution of $\beta|\hat{\beta}$. Using Bayes' Theorem, we know:
$$\Pr(\beta|\hat{\beta}) = \frac{\Pr(\hat{\beta}|\beta)\Pr(\beta)}{\Pr(\hat{\beta})}$$

The denominator, $\Pr(\hat{\beta})$, is just a scalar that ensures that the integral of the entire expression is 1. In other words, it does not actually affect the shape of the distribution, so we can primarily concern ourselves with:

$$\Pr(\beta|\hat{\beta}) \propto \Pr(\hat{\beta}|\beta)\Pr(\beta)$$

We will also be making use of the probability density function for a normally distributed variable with mean μ and standard deviation σ:

$$\Pr(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(\frac{-(x-\mu)^2}{2\sigma^2})$$

1.2 Derivation

1.2.1 $\Pr(\beta)$

We have already defined the distribution of β to be:

$$\Pr(\beta) = (2\pi\sigma_\beta^2)^{-\frac{1}{2}} \exp(-\frac{\beta^2}{2\sigma_\beta^2})$$

1.2.2 $\Pr(\hat{\beta}|\beta)$

As defined, $\hat{\beta}$ is a random variable that is the sum of two independent normally distributed random variables, β and ϵ. By fixing β, $\hat{\beta}$ just becomes the distribution of ϵ plus the constant β. Adding a constant to a normally distributed variable only changes the mean of that distribution, so:

$$\hat{\beta}|\beta \sim \mathcal{N}(\beta + \mathrm{E}[\epsilon], \sigma_\epsilon^2) = \mathcal{N}(\beta, \sigma_\epsilon^2)$$

Therefore:

$$\Pr(\hat{\beta}|\beta) = (2\pi\sigma_\epsilon^2)^{-\frac{1}{2}} \exp(-\frac{(\hat{\beta} - \beta)^2}{2\sigma_\epsilon^2})$$

1.2.3 $\Pr(\hat{\beta}|\beta) \cdot \Pr(\beta)$

We now multiply both of these probabilities together:

$$\Pr(\hat{\beta}|\beta)\Pr(\beta) = (2\pi\sigma_\epsilon^2)^{-\frac{1}{2}} \exp(-\frac{(\hat{\beta} - \beta)^2}{2\sigma_\epsilon^2})(2\pi\sigma_\beta^2)^{-\frac{1}{2}} \exp(-\frac{\beta^2}{2\sigma_\beta^2})$$

$$= ((2\pi\sigma_\epsilon^2)(2\pi\sigma_\beta^2))^{-\frac{1}{2}} \exp(-\frac{1}{2}(\frac{(\hat{\beta} - \beta)^2}{\sigma_\epsilon^2} + \frac{\beta^2}{\sigma_\beta^2}))$$

For simplicity, we'll define $C_1 = ((2\pi\sigma_\epsilon^2)(2\pi\sigma_\beta^2))^{-\frac{1}{2}}$, which is just a scaling factor and does not affect the shape of the distribution. The goal is to rewrite the term inside the exponent, which determines the shape, in a form analogous to a normal distribution. Keep in mind that the sole variable here is β, as we have fixed $\hat{\beta}$ since we are trying to determine $\beta|\hat{\beta}$. We'll start by expanding the square:

$$-\frac{1}{2}(\frac{(\hat{\beta} - \beta)^2}{\sigma_\epsilon^2} + \frac{\beta^2}{\sigma_\beta^2}) = -\frac{1}{2}(\frac{\hat{\beta}^2 - 2\hat{\beta}\beta + \beta^2}{\sigma_\epsilon^2} + \frac{\beta^2}{\sigma_\beta^2})$$

$$= -\frac{1}{2}((\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\beta^2})\beta^2 - \frac{2}{\sigma_\epsilon^2}\hat{\beta}\beta + \frac{1}{\sigma_\epsilon^2}\hat{\beta})$$

$$= -\frac{1}{2}((\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\beta^2})\beta^2 - \frac{2}{\sigma_\epsilon^2}\hat{\beta}\beta) - \frac{\hat{\beta}}{2\sigma_\epsilon^2}$$

2

We now define:

$$\sigma^2_{\beta|\hat\beta} = \frac{\sigma^2_\epsilon \sigma^2_\beta}{\sigma^2_\epsilon + \sigma^2_\beta}$$

such that:

$$\frac{1}{\sigma^2_\epsilon} + \frac{1}{\sigma^2_\beta} = \frac{\sigma^2_\epsilon + \sigma^2_\beta}{\sigma^2_\epsilon \sigma^2_\beta} = \frac{1}{\sigma^2_{\beta|\hat\beta}}$$

We can also pull out the last term as a constant, $C_2 = -\frac{\hat\beta}{2\sigma^2_\epsilon}$, since it doesn't contain β and $\exp(C_2)$ is still a constant. We will also pull out $\sigma^2_{\beta|\hat\beta}$ and complete the square:

$$-\frac{1}{2}\Big(\frac{(\hat\beta - \beta)^2}{\sigma^2_\epsilon} + \frac{\beta^2}{\sigma^2_\beta}\Big) = -\frac{1}{2}\Big(\frac{1}{\sigma^2_{\beta|\hat\beta}}\beta^2 - \frac{2}{\sigma^2_\epsilon}\hat\beta\beta\Big) + C_2$$

$$= -\frac{1}{2\sigma^2_{\beta|\hat\beta}}\Big(\beta^2 - \frac{2\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\beta\Big) + C_2$$

$$= -\frac{1}{2\sigma^2_{\beta|\hat\beta}}\Big(\beta^2 - 2\frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\beta + \big(\frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\big)^2 - \big(\frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\big)^2\Big) + C_2$$

$$= -\frac{1}{2\sigma^2_{\beta|\hat\beta}}\Big(\big(\beta - \frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\big)^2 - \big(\frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\big)^2\Big) + C_2$$

$$= -\frac{1}{2\sigma^2_{\beta|\hat\beta}}\big(\beta - \frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta\big)^2 + \frac{\sigma^2_{\beta|\hat\beta}}{2(\sigma^2_\epsilon)^2}\hat\beta^2 + C_2$$

Similar to before, we'll define $C_3 = \frac{\sigma^2_{\beta|\hat\beta}}{2(\sigma^2_\epsilon)^2}\hat\beta^2$, which is another constant. This means we can now write:

$$\Pr(\hat\beta|\beta)\Pr(\beta) = C_1 \exp(C_2 + C_3)\exp\Big(-\frac{(\beta - \frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta)^2}{2\sigma^2_{\beta|\hat\beta}}\Big)$$

This is equivalent to the probability density function of a normal distribution with a mean of $\frac{\sigma^2_{\beta|\hat\beta}}{\sigma^2_\epsilon}\hat\beta$ and variance of $\sigma^2_{\beta|\hat\beta}$. We can now plug in the definition of $\sigma^2_{\beta|\hat\beta} = \frac{\sigma^2_\epsilon \sigma^2_\beta}{\sigma^2_\epsilon + \sigma^2_\beta}$ to get:

$$\beta|\hat\beta \sim \mathcal{N}\Big(\frac{\sigma^2_\beta}{\sigma^2_\beta + \sigma^2_\epsilon}\hat\beta, \frac{\sigma^2_\epsilon \sigma^2_\beta}{\sigma^2_\epsilon + \sigma^2_\beta}\Big)$$

Therefore:

$$\mathrm{E}[\beta|\hat\beta] = \frac{\sigma^2_\beta}{\sigma^2_\beta + \sigma^2_\epsilon}\hat\beta$$

1.3   Intuition

We can understand this formula as follows: The true parameter value, β, is distributed around zero. Our parameter estimate, $\hat\beta$, can take on a value far from zero not only because the true parameter value is far from zero, but also if it has large measurement

error, ϵ that pushes it farther from zero. Therefore, while given a certain true parameter value, the expected parameter estimate is the same as the true parameter value (that is, $\mathrm{E}[\hat{\beta}|\beta] = \beta$, since the measurement error is centered at zero), the reverse is not true.

Instead, the expected value of the true parameter, given the parameter estimate, is equal to the parameter estimate multiplied by a shrinkage factor, $\frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_\epsilon^2}$. The greater signal ($\sigma_\beta^2$) there is relative to noise (σ_ϵ^2), the less the expected true parameter value is shrunk towards zero. In other words, the less noise there is, the closer the expected true parameter value is to the parameter estimate. If there is no noise, the parameter estimate is always the true parameter value.