1 Chapter 6 Learning as Inference

Problem 6.9 ***In section 6.4.2, we specified how the posterior distribution over the precision parameter J_s is used in prediction.***

(a) Evaluate the integral for the posterior predictive distribution of s_{t+1}, Eq. (6.37).

(b) How does this distribution compare to a situation where J_s is known?

Answer

(a) From 6.4.2 we know that the marginalization equation is:

$$p(s_{t+1} \mid s) = \int p(s_{t+1} \mid J_s)\, p(J_s \mid s)\, dJ_s$$

And we have assumptions that:

- The likelihood $p(s_{t+1} \mid J_s)$ is in Gaussian form:

$$p(s_{t+1} \mid J_s) = \sqrt{\frac{J_s}{2\pi}} \exp\left(-\frac{J_s}{2}(s_{t+1} - s)^2\right)$$

- The posterior over precision $p(J_s \mid s)$ is a Gamma distribution from equation (6.33):

$$p(J_s \mid s) = \frac{\beta^\alpha}{\Gamma(\alpha)} J_s^{\alpha-1} e^{-\beta J_s} \quad \text{where } \alpha = \frac{t}{2} + 1, \quad \beta = \frac{1}{2}\sum_{i=1}^{t}(s_i - \mu)^2$$

Then we can substitute them into the integral:

$$p(s_{t+1} \mid s) = \int_0^\infty \sqrt{\frac{J_s}{2\pi}} \exp\left(-\frac{J_s}{2}(s_{t+1} - \mu)^2\right) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} J_s^{\alpha-1} e^{-\beta J_s}\, dJ_s$$

Combine the powers of J_s and exponential terms:

$$p(s_{t+1} \mid s) = \frac{\beta^\alpha}{\Gamma(\alpha)\sqrt{2\pi}} \int_0^\infty J_s^{\alpha - \frac{1}{2}} \exp\left(-J_s\left(\beta + \frac{(s_{t+1} - \mu)^2}{2}\right)\right) dJ_s$$

This is a Gamma integral with shape $\alpha + \frac{1}{2}$ and rate $A = \beta + \frac{(s_{t+1}-\mu)^2}{2}$, so:

$$p(s_{t+1} \mid s) = \frac{\beta^\alpha}{\Gamma(\alpha)\sqrt{2\pi}} \cdot \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{A^{\alpha + \frac{1}{2}}}$$

where,

$$A = \beta + \frac{(s_{t+1} - \mu)^2}{2} = \frac{1}{2}\sum_{i=1}^{t}(s_i - \mu)^2 + \frac{(s_{t+1} - \mu)^2}{2}$$

(for the substitution part I got completely lost so I asked ChatGPT to form it for me)

Then finally, it can be expressed as:

$$p(s_{t+1} \mid s) = \frac{\beta^\alpha \, \Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha)\sqrt{2\pi}\left(\beta + \frac{(s_{t+1}-\mu)^2}{2}\right)^{\alpha+\frac{1}{2}}}$$

(b) When J_s is known, the posterior predictive distribution over s_{t+1} is simply a Gaussian distribution with known mean and variance:

$$p(s_{t+1} \mid s, J_s) = \mathcal{N}(s_{t+1}; s, J_s^{-1})$$

In this case, the distribution is very narrow, since there is high confidence in the prediction given that the precision is known.

We can thus get the integrate over posterior distribution:

$$p(s_{t+1} \mid s) = \int p(s_{t+1} \mid J_s, s)\, p(J_s \mid s)\, dJ_s$$

The distribution will thus become:

- **More-tailed** than a Gaussian.
- **Wider** than the Gaussian with fixed precision, since it incorporates uncertainty in J_s.