4.13

Intro

In this exercise, we will determine the minimum amount of samples we must drawn from a particular distribution to have a certain posterior credible interval. This situation is quite common in practice. For example, when we take a number of measures in a lab experience we need to report it as a single number plus a degree of uncertainty. One common approach is to report the sample average together with the sample standard deviation. The posterior credible interval is a more Bayesian approach to this situation.

Solution

a)

The r.v. is Gaussian and has the following distribution: $X \sim \mathcal{N}(\mu, \sigma^2 = 4)$. Since μ is unknown, we treat it as a r.v. as well. Its prior is $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2 = 9)$. The posterior is $\mu \sim \mathcal{N}(\mu_n, \sigma_n^2)$, so we know we are dealing with Bayesian inference of Gaussian parameters. This is important because from section 4.6.1 we have an expression that relates the three standard deviations:

$$\begin{aligned} V_N^{-1} &= V_0^{-1} + N\Sigma^{-1} \\ \frac{1}{\sigma_n^2} &= \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \\ N &= \left\lceil \sigma^2 \left(\frac{1}{\sigma_n^2} - \frac{1}{\sigma_0^2} \right) \right\rceil \end{aligned} \tag{1}$$

Now, we need to find the posterior variance σ_n^2 to determine the number of samples. As you may already guess, the posterior variance is deeply related to the credible interval. From the problem statement we know that $|u - l| = 1$. Using the hint of the author about Gaussians we conclude that

$$\begin{aligned} |u - l| = \mu_n + 1.96\sigma_n - (\mu_n - 1.96\sigma_n) &= 3.92\sigma_n \\ 1 &= 3.92\sigma_n \\ \sigma_n &= \frac{1}{3.92} \end{aligned} \tag{2}$$

With the three variances ($\sigma_0^2 = 9$, $\sigma_n^2 = \frac{1}{3.92^2}$, $\sigma^2 = 4$) at hand, we can calculate N.

$$N = \left\lceil \sigma^2 \left(\frac{1}{\sigma_n^2} - \frac{1}{\sigma_0^2} \right) \right\rceil = \lceil 61.02 \rceil = 62 \tag{3}$$

Conclusion

In this exercise we determined the number of samples necessary to have a posterior credible interval of width equal to 1. In order to do that we performed Bayesian inference in the mean parameter of the original Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma^2 = 4)$. We discovered that the minimum number of samples for $p(l \leq \mu \leq u) \geq 0.95$ is N = 62.

It is worth noting that in (1) the number of necessary samples increases if the original standard deviation (σ) increases. The number of samples also increases if we want to get a posterior with better precision ($\frac{1}{\sigma_n^2}$).