Problem

Mathematical model of a system:

$$Y_k = g(Y_{k-1}, \mu) + E \quad , \quad E = N(0, \sigma^2)$$

Find the likelihood, given the data $D \equiv (\hat{Y}_0, \hat{Y}_1, \dots, \hat{Y}_N)$

$$P(\{\hat{Y}_k\}_{0 \to N} \mid \mu, \sigma^2, I) = P(\hat{Y}_N, \hat{Y}_{N-1}, \dots, \hat{Y}_0 \mid \mu, \sigma^2, I)$$

From the product rule of probability we have that:

$$P(\hat{Y}_N, \hat{Y}_{N-1}, \dots, \hat{Y}_0 \mid \mu, \sigma^2, I) = P(\hat{Y}_{N-1}, \hat{Y}_{N-2}, \dots, \hat{Y}_0 \mid \mu, \sigma^2, I) \times$$
$$P(\hat{Y}_N \mid \hat{Y}_{N-1}, \hat{Y}_{N-2}, \dots, \hat{Y}_0, \mu, \sigma^2, I)$$

Based on the structure of the model, Y_k depends only on Y_{k-1} (given μ, σ^2). Mathematically this means that

$$P(\hat{Y}_k \mid \hat{Y}_{k-1}, \hat{Y}_{k-2}, \dots, \hat{Y}_0, \mu, \sigma^2) = P(\hat{Y}_k \mid \hat{Y}_{k-1}, \mu, \sigma^2)$$

Substituting in the above expression one derives that

$$P(\{\hat{Y}_k\}_{0 \to n} \mid \mu, \sigma^2, I) = \prod_{p=1}^{k} P(\hat{Y}_p \mid \hat{Y}_{p-1}, \mu, \sigma^2, I)$$

$\longrightarrow$

Problem

$Y = X + E$, X and E independent

$f(x) = N(\mu, \sigma^2)$

$E = N(0, 1^2)$

$\hat{Y} = 2$

We have found in class that $f(x) = N(\mu, \sigma^2)$ and

$f(X|\hat{Y}) = N\left(\frac{2+\mu}{2}, \frac{1}{2}\sigma^2\right)$

The prior robust prediction for the QoI Y is

$f(y|I)$. From the equation of the model, $Y = X + E$

we can see that the QoI is the sum of two normally

distributed variables (independent) Thus, it is also Gaussian with mean

and variance the sums of the means and variances

$f(y|I) = N(\mu + 0, \sigma^2 + 1^2) = N(\mu, \sigma^2 + 1)$

For the posterior, $f(y|\hat{Y}) = N\left(\frac{2+\mu}{2}, \frac{1}{2}\sigma^2\right) + N(0, 1^2)$

and for the same reason, $f(y|\hat{Y}) = N\left(\frac{2+\mu}{2}, \frac{1}{2}\sigma^2 + 1\right)$

Proove that $f(x|\hat{Y}_1, I) \sim \exp\left[-\frac{1}{2}\frac{(2-x)^2+(x-\mu)^2}{\sigma^2}\right]$ is Normal

---

$$f(x|\hat{Y}_1, I) \sim \exp\left[-\frac{1}{2}\frac{(2-x)^2+(x-\mu)^2}{\sigma^2}\right]$$

$$\sim \exp\left[-\frac{1}{2}\frac{4-4x+x^2+x^2-2x\mu+\mu^2}{\sigma^2}\right]$$

$$\sim \exp\left[-\frac{1}{2}\frac{2x^2-2x(\mu+2)+\mu^2}{\sigma^2}\right]$$

Completing the square we obtain:

$$2x^2-2x(\mu+2)+\mu^2 = 2\left[x^2 - x(\mu+2) + \frac{\mu^2}{2}\right] =$$

$$2\left[\left(x-\frac{\mu+2}{2}\right)^2 + \underbrace{\frac{\mu^2}{2} - \frac{(\mu+2)^2}{4}}_{\text{independent of } x}\right] \Rightarrow$$

$$f(x|\hat{Y}_1, I) \sim \exp\left[-\frac{1}{2}\frac{2\left(x-\frac{\mu+2}{2}\right)^2}{\sigma^2}\right]$$

$$\sim \exp\left[-\frac{\left(x-\frac{\mu+2}{2}\right)^2}{\sigma^2}\right] \Rightarrow$$

$$f(x|\hat{Y}_1, I) \sim \exp\left[-\frac{1}{2}\frac{\left(x-\frac{\mu+2}{2}\right)^2}{\frac{\sigma^2}{2}}\right] = N\left(\frac{\mu+2}{2}, \frac{\sigma^2}{2}\right)$$

Problem

An output QoI follows the Gaussian distribution $Y = N(X, \sigma^2)$, where the variance σ^2 is known, but the mean is uncertain. We are interested in updating the uncertainty in the mean X of the model, using measurements of the QoI, $D = \{\hat{Y}_1, \dots, \hat{Y}_N\}$.

In the class we have calculated the Posterior PDF which quantifies the uncertainty in the uncertain parameter X, using both a Uniform and a Gaussian prior pdf for X.

Now the task is to plot the posterior pdf as a function of the number of data.

Uniform prior pdf

In the case of the uniform prior we found that the posterior pdf of X is Gaussian with mean $\hat{\mu}$ and Variance $S = \frac{\sigma^2}{N}$ where $\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} \hat{Y}_k$.

We can see that the data affects the mean and variance of the posterior pdf. In figure 1 we see the evolution of the posterior pdf as the data grows larger.

About figure 1: The data have been drawn from a normal distribution with mean = 5 and standard deviation = 2. The length of the data was chosen to be 10, 50, and 100. The means and variances of the respective posterior pdfs were found:

	10 samples	50 samples	100 samples
mean	5,5056	4,8785	4,9988
variance	0,4	0,08	0,04
std	0,6325	0,2828	0,2

We can see that as the number of data grows larger, the most probable value of the posterior pdf (mean of the Gaussian) tends to the actual mean of the normal distribution from which the data were generated, which is the "uncertain" parameter. In this case of simulated data, we ofcourse know the mean of the sample (data) generating distribution since we have to fix it in order to get the data.

Gaussian priors

In this case we assign a Gaussian prior pdf for μ.

$$p(\mu|\sigma^2, I) = N(m, T^2)$$

We have found that in this case the posterior pdf of μ is also Gaussian with mean $\hat{\mu} = Bm + (1-B)\mu_0$

and variance $S = \left(\frac{N}{\sigma^2} + \frac{1}{T^2}\right)^{-1} = BT^2 = (1-B)\frac{\sigma^2}{N}$

where $\mu_0 = \frac{1}{N}\sum_{k=1}^{N} \hat{Y}_k$ is the sample mean and

$B = \left(\frac{N}{\sigma^2} + \frac{1}{T^2}\right)^{-1}\frac{1}{T^2} \leq 1$ with $\frac{\sigma^2}{N}$ being the sample variance

For the Gaussian prior $p(\mu|\sigma^2, I) = N(5, 3^2)$ we have that:

Gaussian posterior pdf:

	10 samples	50 samples	100 samp
mean	5,8150	4,8738	5,1477
variance	0,3830	0,0793	0,0398
std	0,6189	0,2816	0,1996

The corresponding plots can be seen in figure 2.

Gaussian priors

In this case we change the parameters of the prior pdf to $p(\mu|\sigma^2, I) = N(7, 1^2)$. The results are:

Gaussian posterior pdf:

	10 samples	50 samples	100 samples
mean	6,3744	4,6375	5,0875
variance	0,2857	0,0741	0,0385
std	0,5345	0,2722	0,1962

The corresponding plots can be seen in figure 3.

As the data grows larger in size, the posterior pdfs for both priors, $N(5, 3^2)$ and $N(7, 1^2)$ tend to the same Gaussian pdf. So the choice of the prior really matters if we have just a few data in our posession.

The second prior, $N(7, 1^2)$ gives a mpv $= 6,3744$ for 10 samples whereas the first prior $N(5, 3^2)$ gives mpv $= 5,8750$ for 10 samples because the prior mean is closer to the actual mean of the data generating distribution, $(=5)$ and needs less samples to converge. Also, for less samples, B is more important and affects the posterior variance more: $S = B\tau^2$. So for a few samples, the prior variance $\tau = 1$ of the second prior yields a smaller posterior variance than $\tau = 3$.

Problem

In output QoI follows the Gaussian distribution $Y = N(\mu, X)$.
The mean of the Gaussian is known $(=\mu)$, but the variance is uncertain.
We are interested in updating the uncertainty in the variance X of the distribution, given a set of independent data: $D = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N)$
We will use $X = \sigma^2$ to denote the possible values of the uncertain variable X.

We assume a uniform prior for X.

---

1). Posterior pdf: $P(\sigma^2 \mid (\hat{Y}_k)_{k=1 \to N}, \mu, I) =$

$$P(\{\hat{Y}_k\}_{1 \to N} \mid \sigma^2, \mu, I) \cdot p(\sigma^2 \mid \mu, I)$$

likelihood $\qquad\qquad$ prior

$$p(\sigma^2 \mid \mu, I) = \begin{cases} \dfrac{1}{\sigma_{max}^2}, & \text{for } 0 \le \sigma \le \sigma_{max}^2 \\ 0, & \text{otherwise} \end{cases}$$

For the likelihood function, the data are independent so we have from the product rule of probability combined with independence:

$$P(\{\hat{Y}_k\}_{1-N} \mid \sigma^2, \mu, I) = P(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_N \mid \sigma^2, \mu, I) =$$

$$P(\hat{Y}_1 \mid \sigma^2, \mu, I) \cdot P(\hat{Y}_2 \mid \sigma^2, \mu, I) \cdots P(\hat{Y}_N \mid \sigma^2, \mu, I)$$

$\longrightarrow$

2) $L(\sigma^2) = -\log\left[P(\sigma^2 | D, \mu, I)\right] \Rightarrow$

$$L(\sigma^2) = \frac{N}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2 + C$$

3) The mpv of σ^2 is the one that maximizes the posterior pdf, or minimizes the $L(\sigma^2)$.

$$\left.\frac{\partial L}{\partial \sigma^2}\right|_{\sigma^2 = \hat{\sigma}^2} = \frac{N}{2\hat{\sigma}^2} - \frac{1}{2\hat{\sigma}^4}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2 = \frac{1}{2\hat{\sigma}^2}\left[N - \frac{1}{\hat{\sigma}^2}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2\right]:$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{N}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2$$

4) Having found the most probable value of the parameter σ^2, we are now interested in finding the uncertainty in this value. For this we use the second derivative to obtain:

$$\left.\frac{\partial^2 L}{\partial (\sigma^2)^2}\right|_{\sigma^2 = \hat{\sigma}^2} = -\frac{N}{2\hat{\sigma}^4} + \frac{4\hat{\sigma}^2}{4\hat{\sigma}^8}\sum_{k=1}^{N}(\hat{Y}_k - \mu)^2 =$$

$$-\frac{N}{2\hat{\sigma}^4} + \frac{\hat{\sigma}^2}{\hat{\sigma}^8}\cdot\hat{\sigma}^2\cdot N = -\frac{N}{2\hat{\sigma}^4} + \frac{2\cdot N}{2\cdot\hat{\sigma}^4} = \frac{N}{2\hat{\sigma}^4}$$

A measure of the uncertainty based on the second derivative is:

$$\sqrt{s} = \left[\left.\frac{\partial^2 L}{\partial(\sigma^2)^2}\right|_{\sigma^2 = \hat{\sigma}^2}\right]^{-\frac{1}{2}} = \frac{\sqrt{2}\,\hat{\sigma}^2}{\sqrt{N}}$$

5) Taylor series expansion, keeping up to the quadratic terms.

$$L(\sigma^2) \simeq L(\hat{\sigma}^2) + \frac{\partial L}{\partial \sigma^2}\Big|_{\sigma^2=\hat{\sigma}^2} (\sigma^2 - \hat{\sigma}^2) + \frac{1}{2}\frac{\partial^2 L}{\partial (\sigma^2)^2}\Big|_{\sigma^2=\hat{\sigma}^2} (\sigma^2-\hat{\sigma}^2)^2$$

since we expand around the mpv, $\frac{\partial L}{\partial \sigma^2}\Big|_{\sigma^2=\hat{\sigma}^2} = 0$ and the quadratic term remains.

$$L(\sigma^2) \simeq L(\hat{\sigma}^2) + \frac{1}{2}\frac{\partial^2 L}{\partial (\sigma^2)^2}\Big|_{\sigma^2=\hat{\sigma}^2} (\sigma^2 - \hat{\sigma}^2)^2 \Rightarrow$$

$$L(\sigma^2) \simeq L(\hat{\sigma}^2) + \frac{1}{2} \cdot \frac{N}{2\hat{\sigma}^4} \cdot (\sigma^2 - \hat{\sigma}^2)^2$$

Then the posterior pdf is given by:

$$P(\sigma^2 | D, \mu, I) = \exp\left[-L(\sigma^2)\right] \Rightarrow$$

$$P(\sigma^2 | D, \mu, I) = \exp\left[-L(\hat{\sigma}^2) - \frac{1}{2}\frac{N}{2\hat{\sigma}^4}(\sigma^2 - \hat{\sigma}^2)^2\right] \Rightarrow$$

$$P(\sigma^2 | D, \mu, I) \sim \exp\left[-\frac{1}{2}\frac{N}{2\hat{\sigma}^4}(\sigma^2 - \hat{\sigma}^2)^2\right]$$

which is a Gaussian approximation of the posterior pdf with mean $= \hat{\sigma}^2$ and variance $= \frac{2\hat{\sigma}^4}{N}$.

6) The true posterior pdf for σ^2 is the inverse gamma pdf:

$$P(\sigma^2 | D, \mu, I) \sim (\sigma^2)^{-\frac{N}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(Y_k - \mu)^2\right]$$

The Gaussian approximation was found in the previous question.

In the following table we see the evolution of the Gaussian posterior mean and variance as the data grows larger.
The data were drawn from a normal distribution with mean= 10 and var= 4.

# Data	1	2	3	4	10	100	200
Post mean	2,5	1,3	1,1	2,2	1,9	3,9	4,05
Post var	13	1,7	0,9	2,4	0,7	0,3	0,16

We can see that as the data grows larger, the mean of the Gaussian posterior tends to the actual variance of the data-generating distribution

Inverse Gamma exact Posterior mpv

# Data	1	2	3	4	10	100	200
mpv	2,5	1,3	1,1	2,2	1,9	3,9	4,0

By looking at figures we can see that for $N = 100$ and larger the exact IG distribution can be very well approximated by a Gaussian, justifying the Bayesian Central Limit Theorem

For less samples, it is useful to note that the IG pdf is more accurate than the normal because due to its large tails it allows for more possible values for σ^2 than the normal does which is centered around an inaccurate value for a small number of samples.