Exercises: inference in the linear Gaussian model

Joaquín Rapela*

November 21, 2022

1 Inferring location of a static submarine from its sonar measurements

(a)

The modified code lines appear below and Fig. 1 shows the generated submarine samples, and the mean and 95% confidence ellipse of the samples probability density function.

```
sigma_zx = 1.0
sigma_zy = 2.0
rho_z = 0.7
cov_z_11 = sigma_zx**2
cov_z_12 = rho_z*sigma_zx*sigma_zy
cov_z_21 = rho_z*sigma_zx*sigma_zy
cov_z_22 = sigma_zy**2
```

*j.rapela@ucl.ac.uk

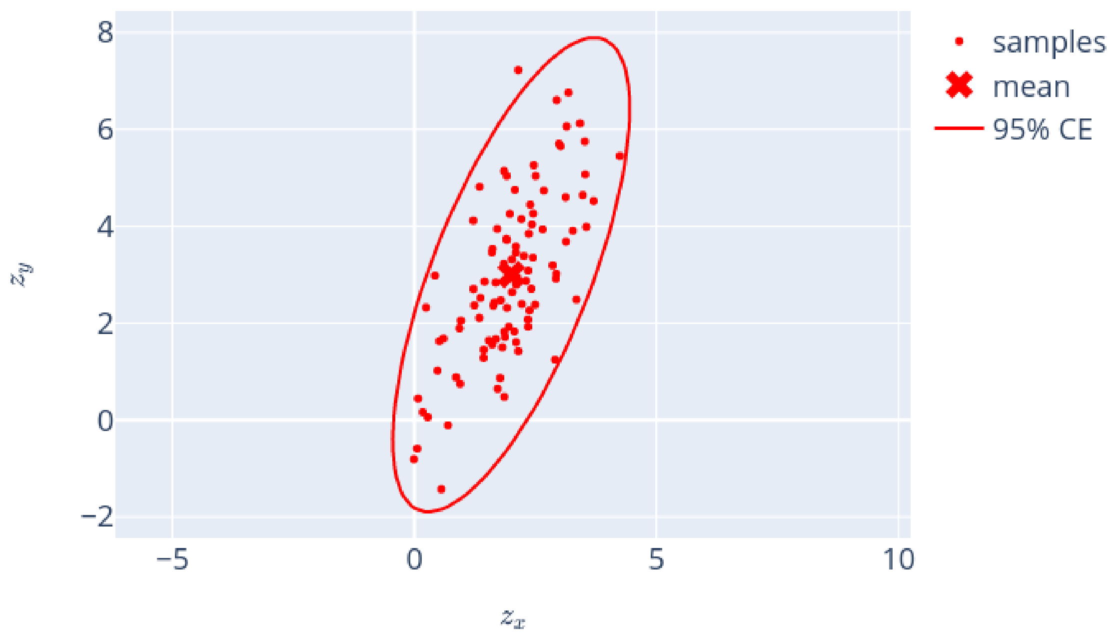

Figure 1: 100 a-priori samples of the submarine location (dots), and the mean (cross) and 95% confidence ellipse (line) of the samples probability density function.

(b)

The modified code lines appear below and Fig. 2 shows the generated measurement samples, and the mean and 95% confidence ellipse of the samples probability density function.

```
sigma_y_x = 1.0
sigma_y_y = 1.0
rho_y = 0.0
cov_y_11 = sigma_y_x**2
```

```
cov_y_12 = rho_y*sigma_y_x*sigma_y_y
cov_y_21 = rho_y*sigma_y_x*sigma_y_y
cov_y_22 = sigma_y_y**2
```

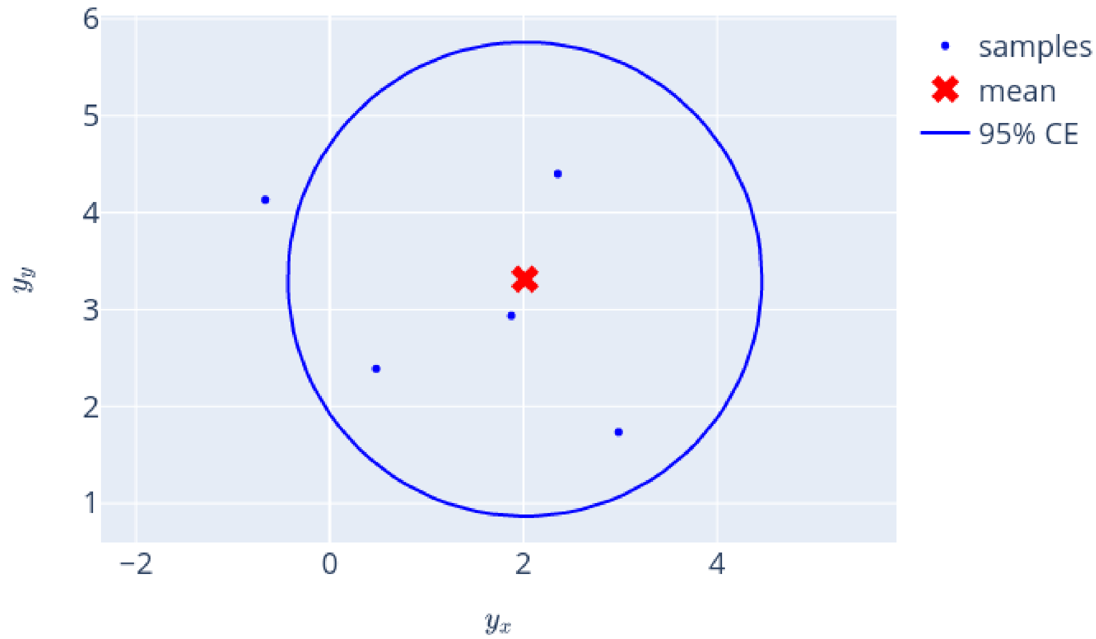

Figure 2: 5 noisy measurements of of the submarine location (dots), and the mean ($\mathbf{z}_1$, cross) and 95% confidence ellipse (line) of the measurements probability density function.

(c)

$$\begin{aligned} p(\mathbf{z}|\mathbf{y}_1, \dots, \mathbf{y}_N) &= K_1 \; p(\mathbf{z}, \mathbf{y}_1, \dots, \mathbf{y}_N) \\ &= K_1 \; p(\mathbf{y}_1, \dots, \mathbf{y}_N|\mathbf{z}) \; p(\mathbf{z}) \\ &= K_2 \; \mathcal{N}\left(\bar{\mathbf{y}} \middle| \mathbf{z}, \frac{1}{N}\Sigma_y\right) \mathcal{N}\left(\mathbf{z}|\mu_z, \Sigma_z\right) \end{aligned} \tag{1}$$

where K_1 is a constant that does not depend on $\mathbf{z}$ and Eq. 1 follows from Claim 1 in the exercise statement. In the right-hand side of Eq. 1 we recognize a linear Gaussian model (i.e., $\bar{\mathbf{y}}$ and $\mathbf{z}$ are Gaussian random variables and the mean of $\bar{\mathbf{y}}$ depends linearly on $\mathbf{z}$).

Defining $p(\bar{\mathbf{y}}|\mathbf{z}) = \mathcal{N}\left(\bar{\mathbf{y}} \middle| \mathbf{z}, \frac{1}{N}\Sigma_y\right)$ and $p(\mathbf{z}) = \mathcal{N}\left(\mathbf{z}|\mu_z, \Sigma_z\right)$, because the right-hand side of Eq. 1 equals a probability density function on $\mathbf{z}$, this right-hand side should be $p(\mathbf{z}|\bar{\mathbf{y}})$. To derive a mathematical expression for $p(\mathbf{z}|\bar{\mathbf{y}})$, we use Eq. 3.37 from Murphy (2022) with $\mathbf{y} = \bar{\mathbf{y}}$, $\mathbf{W} = I$, $\mathbf{b} = \mathbf{0}$, $\Sigma_y = \frac{1}{N}\Sigma_y$, yielding

$$p(\mathbf{z}|\mathbf{y}_1, \dots, \mathbf{y}_N) = p(\mathbf{z}|\bar{\mathbf{y}}) = \mathcal{N}(\mathbf{z}|\mu_{z|\bar{y}}, \Sigma_{z|\bar{y}})$$

$$\Sigma_{z|\bar{y}}^{-1} = \Sigma_z^{-1} + N\Sigma_y^{-1} \tag{2}$$

$$\mu_{z|\bar{y}} = \Sigma_{z|\bar{y}} \left[N\Sigma_y^{-1}\bar{\mathbf{y}} + \Sigma_z^{-1}\mu_z\right] \tag{3}$$

(d)

The modified code lines appear below and Fig. 3 plots the mean of the measurements, the mean of the posterior and its 95% confidence ellipse.

```
cov_y_inv = np.linalg.inv(cov_y)
cov_z_inv = np.linalg.inv(cov_z)
tmp1 = N * cov_y_inv + cov_z_inv
tmp2 = N * np.matmul(cov_y_inv, sample_mean_y) + \
    np.matmul(cov_z_inv, mean_z)
post_mean_z = np.linalg.solve(tmp1, tmp2)
post_cov_z = np.linalg.inv(tmp1)
yBar_mean = z
yBar_cov = 1.0/N*cov_y
```

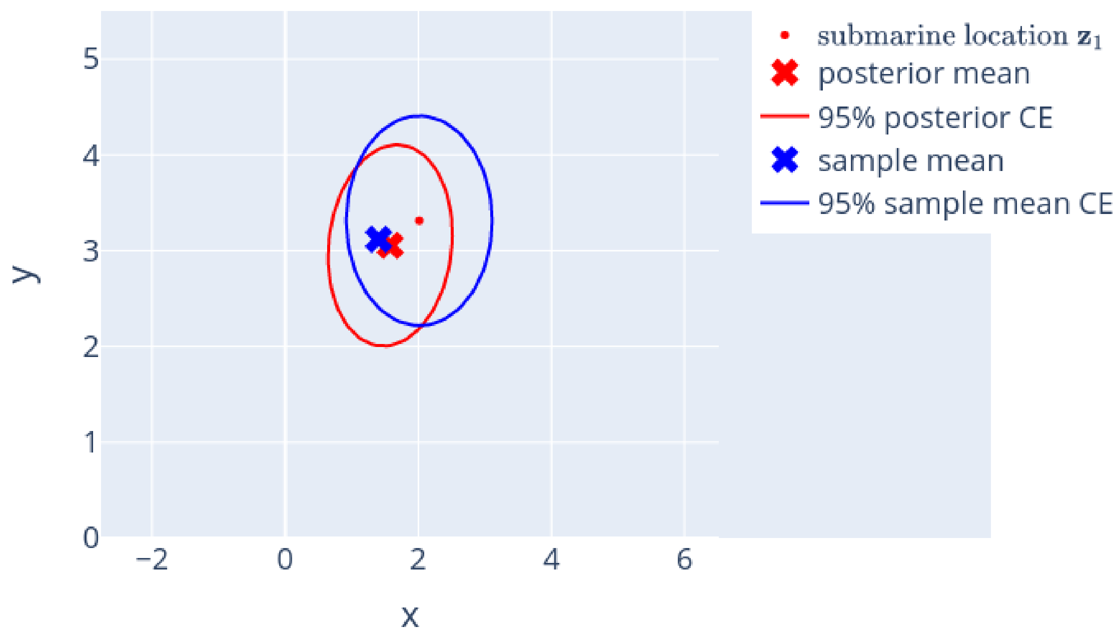

Figure 3: Sample average of 5 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

(e)

Figs. 4-7 plot the posterior estimates computed from an increasing number of measurements.

In these figures we observe that:

1. as the number of measurements increases, the posterior mean approaches the sample mean, and the sample mean approaches the submarine location $\mathbf{z}_1$,

2. as the number of measurements increases, the 95% confidence ellipses become smaller,

3. for three measurements (Fig. 4) the posterior 95% confidence ellipse is tilted, as that of the prior (Fig. 1, Σ_z in Eq. 1 of the exercise statement). As the number of measurments increases, the posterior 95% confidence ellipses become more and more spherical, as the 95% confidence ellipse of the measurements likelihood (Fig. 2, Σ_y in Eq. 2 of the exercise statement).

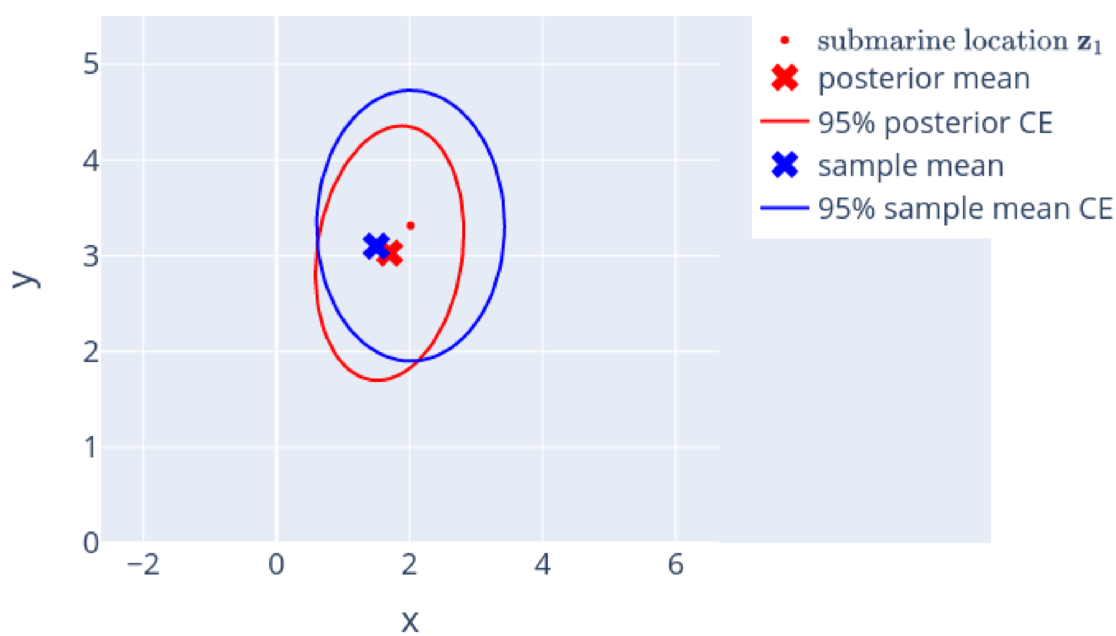

Figure 4: Sample average of 3 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

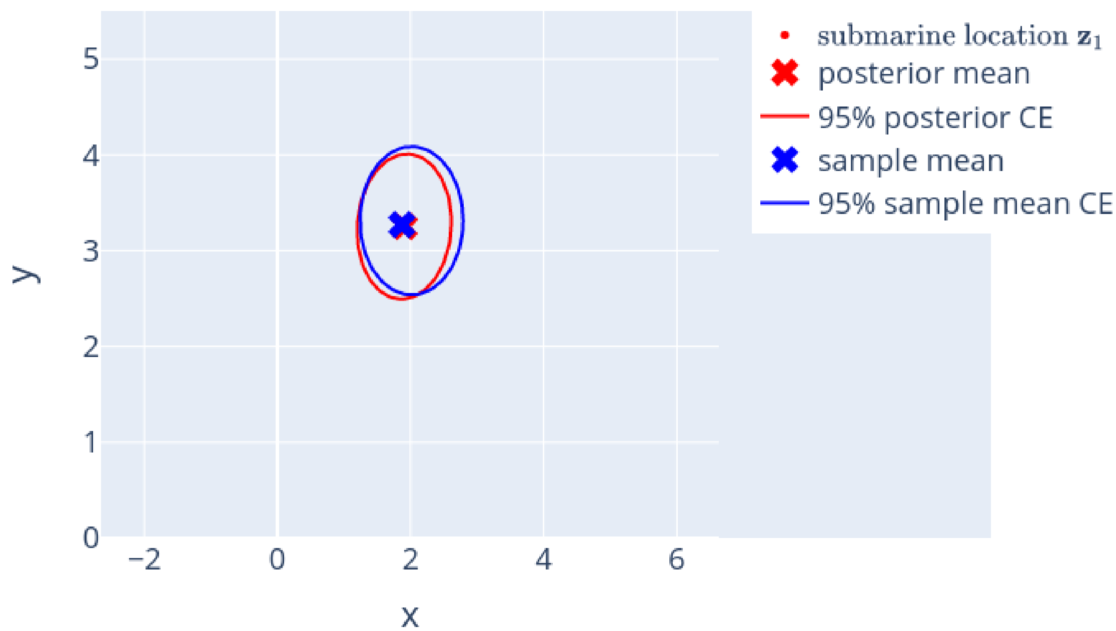

Figure 5: Sample average of 10 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

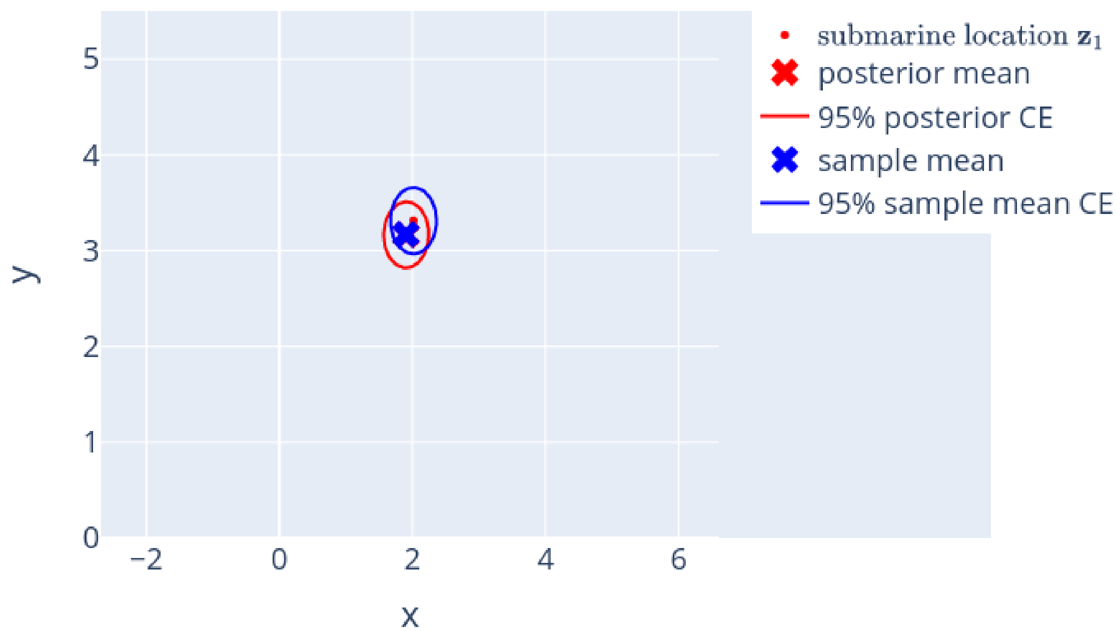

Figure 6: Sample average of 50 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

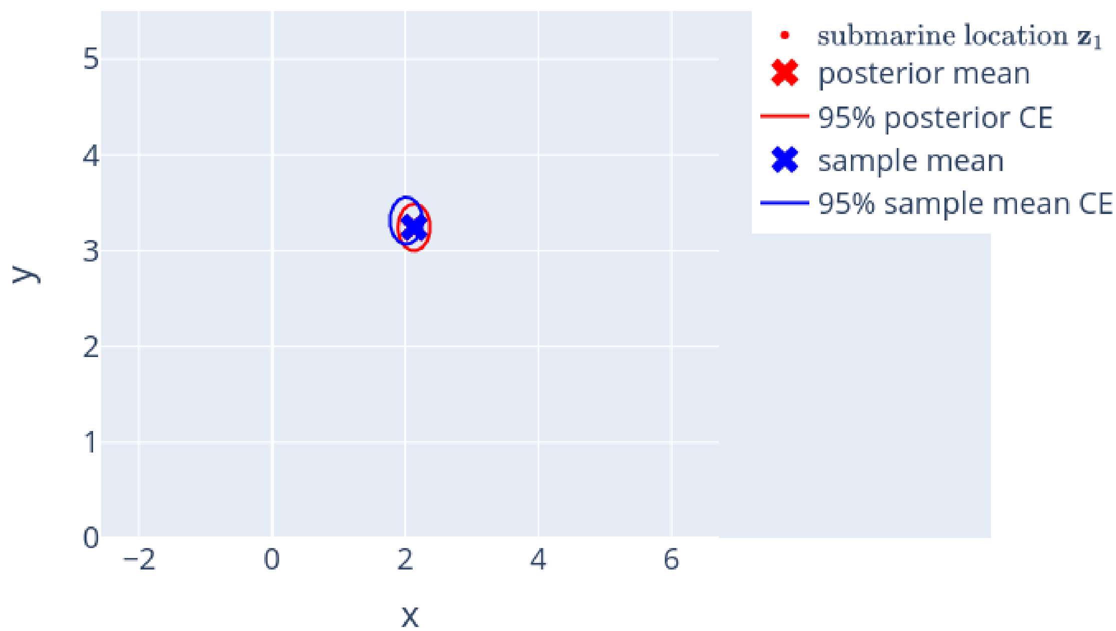

Figure 7: Sample average of 100 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

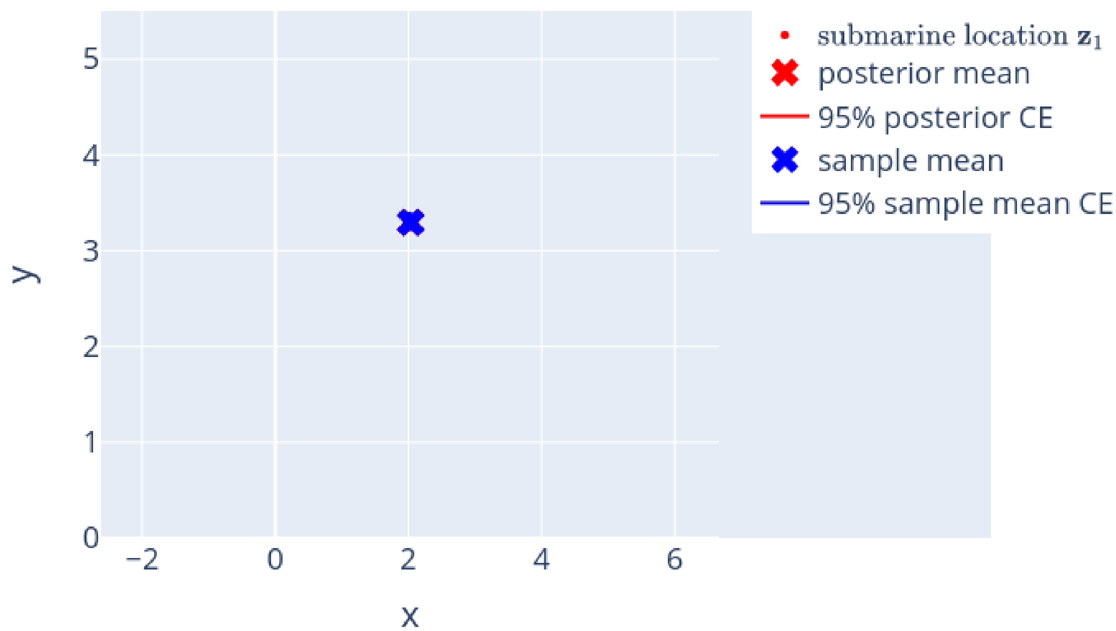

Figure 8: Sample average of 1000 noisy measurements (blue cross), 95% confidence ellipse of this average (blue line), mean of the posterior distribution (red cross), its 95% confidence ellipse (red line), and submarine location ($\mathbf{z}_1$, red dot)

(f)

Eqs 5 and 4 were obtained by re-arrenging Eqs. 3 and 2 to more clearly show the behavior of the posterior mean and covariance as N increases to infinity.

$$
\begin{aligned}
p(\mathbf{z}|\mathbf{y}_1, \ldots, \mathbf{y}_N) &= \mathcal{N}(\mathbf{z}|\mu_{z|\bar{y}}(N), \Sigma_{z|\bar{y}}(N)) \\
\Sigma_{z|\bar{y}}(N) &= \frac{1}{N}\left(\Sigma_y^{-1} + \frac{1}{N}\Sigma_z^{-1}\right)^{-1} \quad (4) \\
\mu_{z|\bar{y}}(N) &= \left(\Sigma_y^{-1} + \frac{1}{N}\Sigma_z^{-1}\right)^{-1}\left[\Sigma_y^{-1}\bar{\mathbf{y}}_N + \frac{1}{N}\Sigma_z^{-1}\mu_z\right] \quad (5)
\end{aligned}
$$

From Eq. 4 we observe that as N increases the contributions of the prior covariance, Σ_z, to the posterior covariance, $\Sigma_{z|\bar{y}}(N)$, becomes smaller and smaller, in comparison to the contribution from the likelihood covariance, $\frac{1}{N}\Sigma_y$. When N is very large, the contribution of the prior covariance dissapears, the posterior covariance converges to the likelihood covariance, which becomes zero.

From Eq. 5 we observe

$$
\lim_{N\to\infty} \mu_{z|\bar{y}}(N) = \Sigma_y\left[\Sigma_y^{-1}\bar{\mathbf{y}}_N\right] = \lim_{N\to\infty} \bar{\mathbf{y}}_N \quad (6)
$$

In class we proved that

$$
\bar{\mathbf{y}}_N \sim \mathcal{N}(\bar{\mathbf{y}}_N|\mathbf{z}_1, \frac{1}{N}\Sigma_y)
$$

Thus, as N approaches infinity, the variance of $\bar{\mathbf{y}}_N$ becomes zero, and $\bar{\mathbf{y}}_N$ collapses to its mean $\mathbf{z}_1$. Therefore, as N approaches infinity, both the posterior mean, Eq. 6, and sample the mean, become deterministic and converge to the population mean of the observatons; i.e., $\mathbf{z}_1$.

References

Murphy, K. P. (2022). *Probabilistic machine learning: an introduction*. MIT press.