Exercises: unsupervised inference with iid observations and the linear Gaussian model

Joaquín Rapela*

August 6, 2023

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% prediction ellipse of these samples.

```
sigma_zx = 1.0
sigma_zy = 1.0
rho_z = -0.8
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

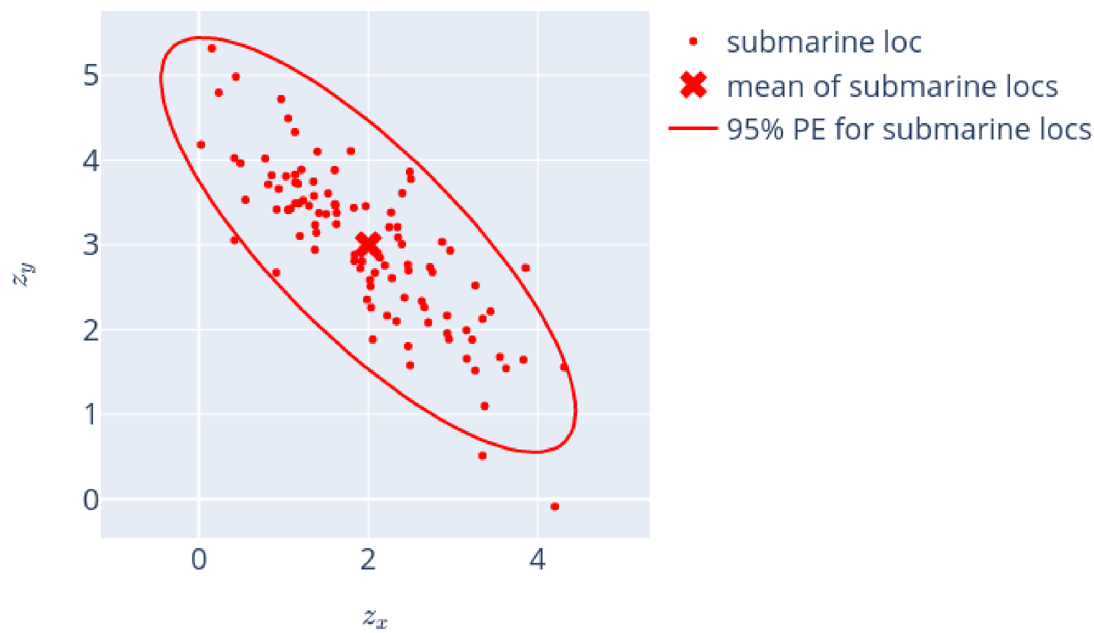

Figure 1: 100 a-priori samples of the submarine location (dots), the mean submarine location (cross) and 95% prediction ellipse (line) for samples of submarine locations. Click on the figure to get its interactive version.

(b)

The modified code lines appear below and Fig. 2 shows the generated measurement samples, and the mean and 95% prediction ellipse of these samples.

```
    sigma_y_x = 1.0
    sigma_y_y = 1.0
    rho_y = 0.8
    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
```

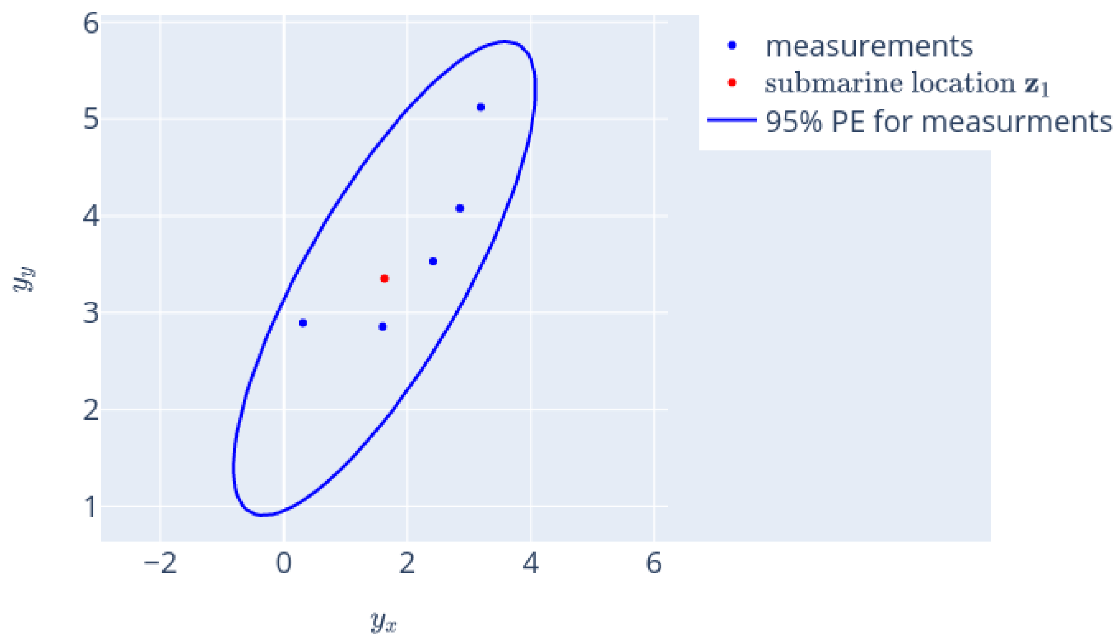

Figure 2: 5 noisy measurements of of the submarine location (dots), and the mean ($\mathbf{z}_1$, cross) and 95% prediction ellipse (line) of these measurements. Click on the figure to get its interactive version.

(c)

$$\begin{aligned}p(\mathbf{z}|\mathbf{y}_1,\ldots,\mathbf{y}_N) &= K_1\ p(\mathbf{z},\mathbf{y}_1,\ldots,\mathbf{y}_N)\\ &= K_1\ p(\mathbf{y}_1,\ldots,\mathbf{y}_N|\mathbf{z})\ p(\mathbf{z})\\ &= K_2\ \mathcal{N}\left(\bar{\mathbf{y}}\left|\mathbf{z},\frac{1}{N}\boldsymbol{\Sigma}_y\right.\right)\mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z\right)\end{aligned} \tag{1}$$

where K_1 and K_2 are constants that do not depend on $\mathbf{z}$, and such that the integral of the corresponding right hand sides are one. The last line follows from the previous one by Claim 1 in the exercise statement.

We know that

$$p(\mathbf{z}|\bar{\mathbf{y}}) = K_3\ p(\bar{\mathbf{y}},\mathbf{z}) = K_3\ p(\bar{\mathbf{y}}|\mathbf{z})p(\mathbf{z}) \tag{2}$$

where K_3 is a constant such that the integral of the right hand side of Eq. 2 is one. Taking

$$p(\mathbf{z}) = \mathcal{N}\left(\mathbf{z}|\boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z\right) \tag{3}$$

$$p(\bar{\mathbf{y}}|\mathbf{z}) = \mathcal{N}\left(\bar{\mathbf{y}}\left|\mathbf{z},\frac{1}{N}\boldsymbol{\Sigma}_y\right.\right) \tag{4}$$

it follow that right hand side of Eq. 1 equals that of Eq. 2, and both K_2 and K_3 are such that the right hand side of these equations integrate to one. Then K_2 equals K_3 and the left hand side of these equations are also equal (i.e., $p(\mathbf{z}|\mathbf{y}_1,\ldots,\mathbf{y}_N) = p(\mathbf{z}|\bar{\mathbf{y}})$).

To find a close-form expression of the $p(\mathbf{z}|\bar{\mathbf{y}})$ we use the following result from the Gaussian linear model, that we discussed in the lecture on linear regression[1]. If

$$\begin{aligned}p(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Lambda}^{-1})\\ p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{L}^{-1})\end{aligned}$$

then

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\left\{\mathbf{A}^\intercal\mathbf{L}(\mathbf{y}-\mathbf{b})+\boldsymbol{\Lambda}\boldsymbol{\mu}\right\},\boldsymbol{\Sigma}) \tag{5}$$

[1] see Eqs. 2.166 and 2.167 in Bishop (2016)

where

$$\mathbf{\Sigma} = (\mathbf{\Lambda} + \mathbf{A}^\intercal \mathbf{L} \mathbf{A})^{-1}$$

Setting $\mathbf{x} = \mathbf{z}$, $\mathbf{y} = \bar{\mathbf{y}}$, $\boldsymbol{\mu} = \boldsymbol{\mu}_z$, $\mathbf{\Lambda}^{-1} = \mathbf{\Sigma}_z$, $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{L}^{-1} = \frac{1}{N}\mathbf{\Sigma}_y$ in Eq. 5, we obtain

$$p(\mathbf{z}|\bar{\mathbf{y}}) = \mathcal{N}(\mathbf{z}|\mathbf{\Sigma}\left\{N\mathbf{\Sigma}_y\bar{\mathbf{y}} + \mathbf{\Sigma}_z^{-1}\boldsymbol{\mu}_z\right\}, \mathbf{\Sigma})$$

where

$$\mathbf{\Sigma} = (\mathbf{\Sigma}_z^{-1} + N\mathbf{\Sigma}_y^{-1})^{-1}$$

Therefore

$$\begin{aligned} p(\mathbf{z}|\mathbf{y}_1, \ldots, \mathbf{y}_N) = p(\mathbf{z}|\bar{\mathbf{y}}) &= \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_{z|\bar{y}}, \mathbf{\Sigma}_{z|\bar{y}}) \\ \mathbf{\Sigma}_{z|\bar{y}}^{-1} &= \mathbf{\Sigma}_z^{-1} + N\mathbf{\Sigma}_y^{-1} \qquad (6) \\ \boldsymbol{\mu}_{z|\bar{y}} &= \mathbf{\Sigma}_{z|\bar{y}}\left[N\mathbf{\Sigma}_y^{-1}\bar{\mathbf{y}} + \mathbf{\Sigma}_z^{-1}\boldsymbol{\mu}_z\right] \qquad (7) \end{aligned}$$

(d)

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

```
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
```

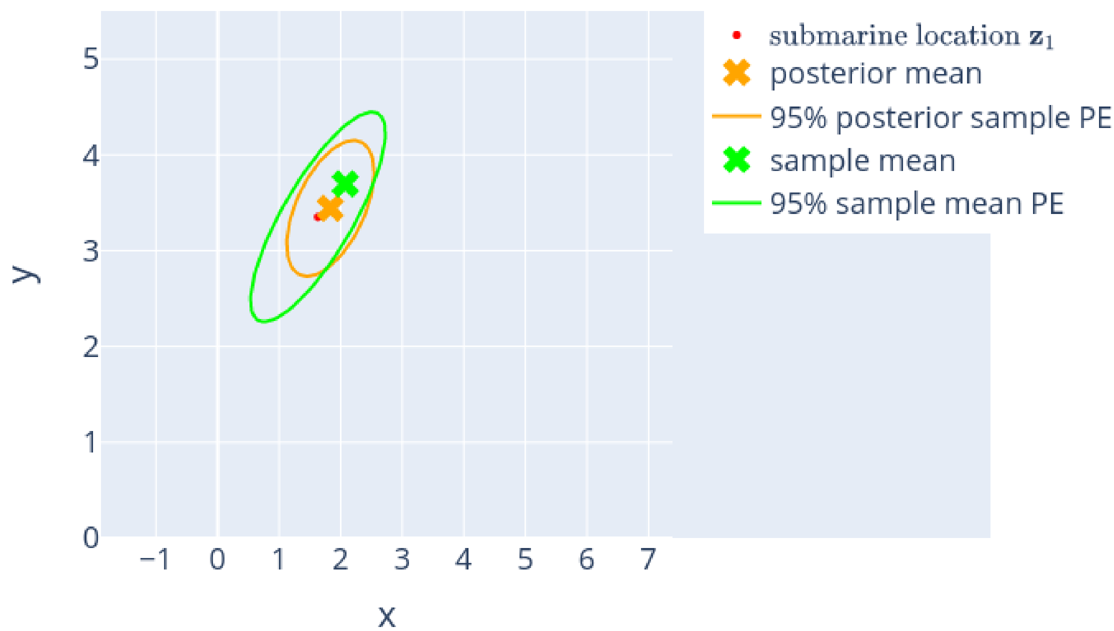

Figure 3: Sample mean of 5 noisy measurements (green cross), 95% prediction ellipse for sample means (green line), mean of the posterior distribution (orange cross), 95% prediction ellipse for samples from the posterior (orange line), and submarine location ($\mathbf{z}_1$, red dot). Click on the figure to get its interactive version.

(e)

Figs. 4-6 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% prediction ellipses become smaller,

3. for one measurement (Fig. 4) the posterior 95% prediction ellipse is the average between the that of the prior (Fig. 1, $\mathbf{\Sigma}_z$ in Eq. 1 of the exercise statement) and that of the likelihood (Fig. 2, $\mathbf{\Sigma}_y$ in Eq. 1 of the exercise statement). As the number of measurements increases, the posterior 95% prediction ellipses become tilted along the 45° orientation, as the 95% prediction ellipse of the measurements likelihood (Fig. 2, $\mathbf{\Sigma}_y$ in Eq. 2 of the exercise statement).

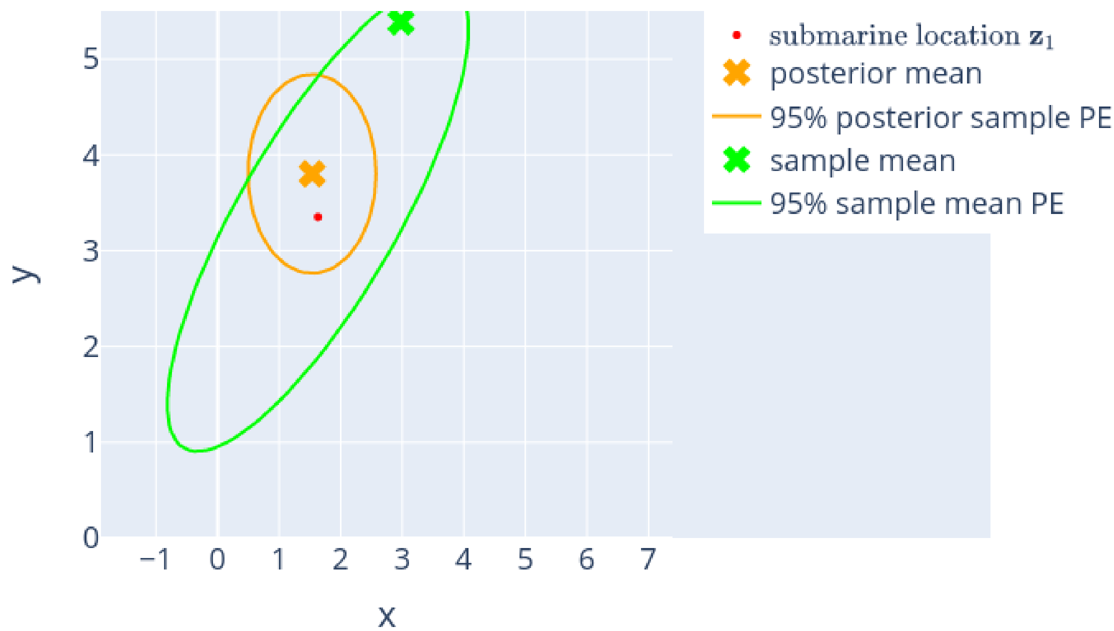

Figure 4: Sample mean of 1 noisy measurements (green cross), 95% prediction ellipse for sample means (green line), mean of the posterior distribution (orange cross), 95% prediction ellipse for posterior samples (orange line), and submarine location ($\mathbf{z}_1$, red dot). Click on the figure to get its interactive version.

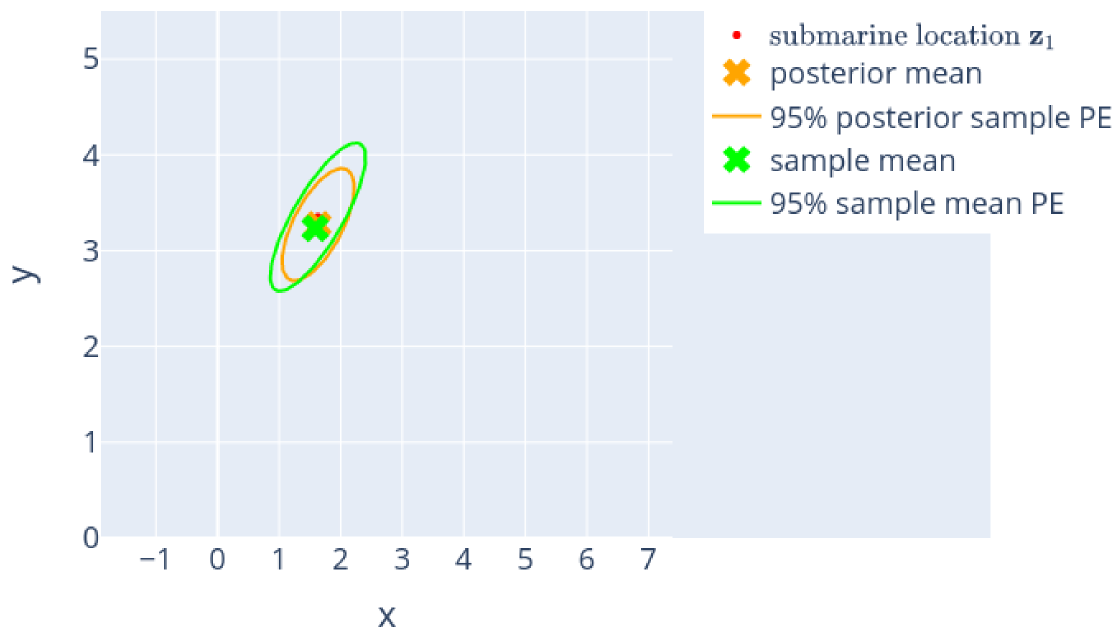

Figure 5: Sample mean of 10 noisy measurements (green cross), 95% prediction ellipse for sample means (green line), mean of the posterior distribution (orange cross), 95% prediction ellipse for posterior samples (orange line), and submarine location ($\mathbf{z}_1$, red dot). Click on the figure to get its interactive version.

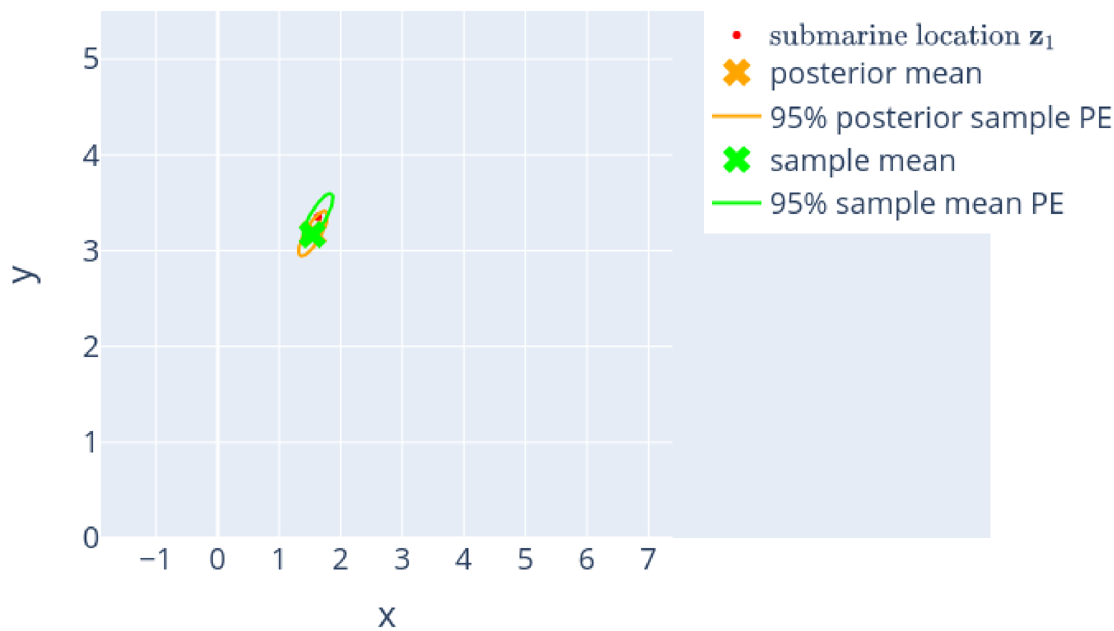

Figure 6: Sample mean of 100 noisy measurements (green cross), 95% prediction ellipse for sample means (green line), mean of the posterior distribution (orange cross), 95% prediction ellipse for posterior samples (orange line), and submarine location ($\mathbf{z}_1$, red dot). Click on the figure to get its interactive version.

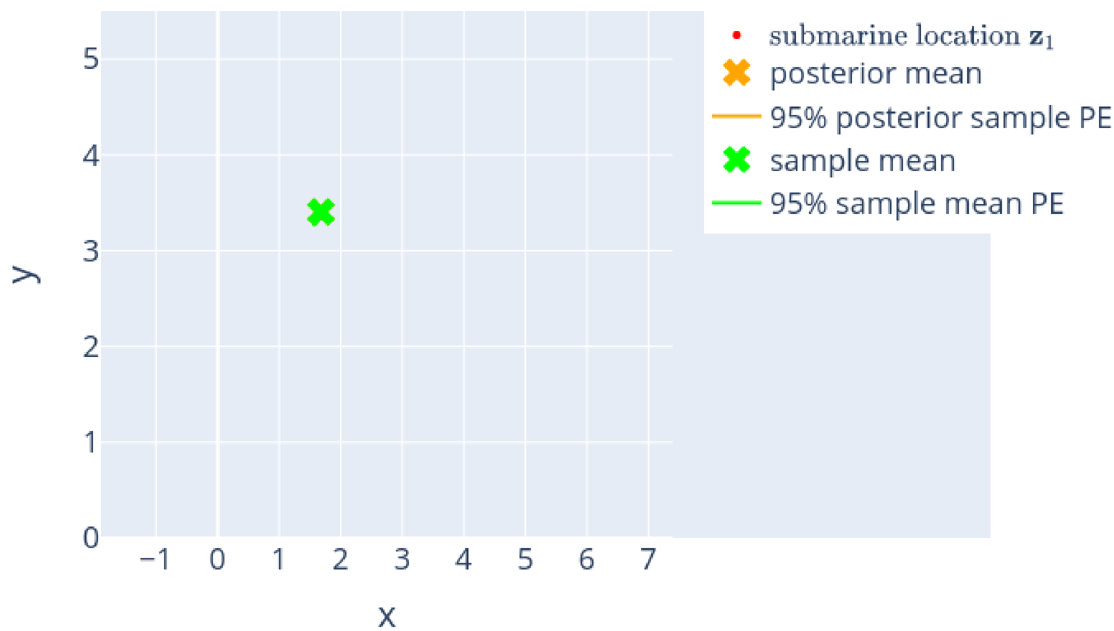

Figure 7: Sample mean of 1000 noisy measurements (green cross), 95% prediction ellipse for sample means (green line), mean of the posterior distribution (orange cross), 95% prediction ellipse for posterior samples (orange line), and submarine location ($\mathbf{z}_1$, red dot). Click on the figure to get its interactive version.

(f)

Eqs. 9 and 8 were obtained by re-arranging Eqs. 7 and 6 to more clearly show the behaviour of the posterior mean and covariance as N increases to infinity.

$$
\begin{aligned}
p(\mathbf{z}|\mathbf{y}_1,\ldots,\mathbf{y}_N) &= \mathcal{N}(\mathbf{z}|\boldsymbol{\mu}_{z|\bar{y}}(N), \boldsymbol{\Sigma}_{z|\bar{y}}(N)) \\
\boldsymbol{\Sigma}_{z|\bar{y}}(N) &= \frac{1}{N}\left(\boldsymbol{\Sigma}_y^{-1} + \frac{1}{N}\boldsymbol{\Sigma}_z^{-1}\right)^{-1} \qquad (8) \\
\boldsymbol{\mu}_{z|\bar{y}}(N) &= \left(\boldsymbol{\Sigma}_y^{-1} + \frac{1}{N}\boldsymbol{\Sigma}_z^{-1}\right)^{-1}\left[\boldsymbol{\Sigma}_y^{-1}\bar{\mathbf{y}}_N + \frac{1}{N}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\mu}_z\right] \qquad (9)
\end{aligned}
$$

From Eq. 8 we observe that as N increases the contributions of the prior covariance, $\boldsymbol{\Sigma}_z$, to the posterior covariance, $\boldsymbol{\Sigma}_{z|\bar{y}}(N)$, becomes smaller, in comparison to the contribution from the likelihood covariance, $\boldsymbol{\Sigma}_y$. When N is very large, the contribution of the prior covariance disappears, the posterior covariance converges to the sample mean covariance, $\frac{1}{N}\boldsymbol{\Sigma}_y$, which becomes zero.

From Eq. 9 we observe

$$
\lim_{N\to\infty}\boldsymbol{\mu}_{z|\bar{y}}(N) = \boldsymbol{\Sigma}_y\left[\boldsymbol{\Sigma}_y^{-1}\bar{\mathbf{y}}_N\right] = \lim_{N\to\infty}\bar{\mathbf{y}}_N
$$

In class we proved that

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

Thus, as N approaches infinity, the covariance 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 and sample the mean, become deterministic and converge to the population mean of the observations; i.e., $\mathbf{z}_1$.

References

Bishop, C. M. (2016). *Pattern recognition and machine learning.* Springer-Verlag New York.