UNIVERSITY OF THESSALY – DEPARTMENT OF MECHANICAL ENGINEERING

UNCERTAINTY QUANTIFICATION

BAYESIAN INFERENCE – INFORMATION ENTROPY – MARKOV CHAIN MONTE CARLO ALGORITHM

Table of Contents

Gaussian Distributions

Exercise 1.

The sum $Z = X + Y$ of two independent Gaussian random variables $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ is Gaussian with mean $\mu_Z = \mu_X + \mu_Y$ and variance $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$, i.e. $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.Equation Chapter (Next) Section 1

Since X and Y are independent Gaussian random variables we can write:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\left(\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right)}$$
$$f(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\left(\frac{1}{2\sigma_y^2}(y-\mu_y)^2\right)} \tag{1.1}$$

In addition, the marginal distribution of variable Z is defined as:

$$f(z) = \int_{-\infty}^{\infty} f(z \mid x) f(x) dx \tag{1.2}$$

, where $f(z \mid x)$ is given by:

$$f(z \mid x) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\left(\frac{1}{2\sigma_y^2}(z-(x+\mu_y))^2\right)} \tag{1.3}$$

All we have to do now is to substitute expressions (1.1) and (1.3) into (1.2) and work out some algebraic calculations.

$$f(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\left(\frac{1}{2\sigma_y^2}(z-(x+\mu_y))^2\right)} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\left(\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right)} dx =$$
$$\frac{1}{\sqrt{2\pi}\sigma_x} \frac{1}{\sqrt{2\pi}\sigma_y} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma_y^2}(z-(x+\mu_y))^2 - \frac{1}{2\sigma_x^2}(x-\mu_x)^2\right) dx \tag{1.4}$$

Let us work with the exponent $-\frac{1}{2\sigma_y^2}\left(z-\left(x+\mu_y\right)\right)^2-\frac{1}{2\sigma_x^2}\left(x-\mu_x\right)^2$ just for simplicity. Doing so,

$$A=-\frac{1}{2\sigma_x^2\sigma_y^2}\left\{\sigma_x^2\left[z-\left(x-\mu_y\right)^2\right]+\sigma_y^2\left(x-\mu_x\right)^2\right\}=$$

$$-\frac{1}{2\sigma_x^2\sigma_y^2}\left\{\underbrace{\sigma_x^2z^2}_{3}-\underbrace{2zx\sigma_x^2}_{2}-\underbrace{2z\mu_y\sigma_x^2}_{3}+\underbrace{\sigma_x^2x^2}_{1}+\underbrace{2x\mu_y\sigma_x^2}_{2}+\underbrace{\mu_y^2\sigma_x^2}_{3}+\underbrace{\sigma_y^2x^2}_{1}-2x\mu_x\sigma_y^2+\underbrace{\mu_x^2\sigma_y^2}_{3}\right\}=$$

$$-\frac{1}{2\sigma_x^2\sigma_y^2}\left\{\underbrace{x^2\left(\sigma_x^2+\sigma_y^2\right)}_{1}-\underbrace{2x\sigma_x^2\left(z-\mu_y\right)}_{2}-2x\mu_x\sigma_y^2+\underbrace{\left(\sigma_x^2\left(z-\mu_y\right)^2+\sigma_y^2\mu_x^2\right)}_{3}\right\}=$$

$$-\frac{1}{\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}}\left\{x^2-2x\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}+\frac{\sigma_x^2\left(z-\mu_y\right)^2}{\sigma_x^2+\sigma_y^2}+\frac{\sigma_y^2\mu_x^2}{\sigma_x^2+\sigma_y^2}\right\}=$$

We will now try to complete the square in the brackets so we add and subtract a term.

$$A=-\frac{1}{\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}}\left\{x^2-2x\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}+\left(\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2\right\}+$$

$$-\frac{1}{\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}}\left\{\frac{\sigma_x^2\left(z-\mu_y\right)^2}{\sigma_x^2+\sigma_y^2}+\frac{\sigma_y^2\mu_x^2}{\sigma_x^2+\sigma_y^2}-\left(\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2\right\}\Rightarrow$$

$$A=-\frac{1}{\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}}\left\{\left(x-\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2+\frac{\sigma_x^2\left(z-\mu_y\right)^2+\sigma_y^2\mu_x^2}{\sigma_x^2+\sigma_y^2}-\left(\frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2\right\}$$

If we carefully carry out the algebraic calculations in the second and third term above we end up with:

$$A = -\frac{1}{\frac{2\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}}\left\{\left(x - \frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2 - \frac{\sigma_x^2\sigma_y^2\left(z-\left(\mu_x+\mu_y\right)\right)^2}{\left(\sigma_x^2+\sigma_y^2\right)^2}\right\} \Rightarrow$$

$$A = -\frac{\left(x - \frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2}{2\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}\right)} - \frac{\left(z-\left(\mu_x+\mu_y\right)\right)^2}{\sigma_x^2+\sigma_y^2} \tag{1.5}$$

Now let's substitute expression (1.5) into the marginal distribution expression (1.4) to get:

$$f\left(z\right) = \frac{1}{\sqrt{2\pi}\sigma_x}\frac{1}{\sqrt{2\pi}\sigma_y}\int_{-\infty}^{\infty}\exp\left(-\frac{\left(x - \frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2}{2\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}\right)}\right)\exp\left(-\frac{\left(z-\left(\mu_x+\mu_y\right)\right)^2}{\sigma_x^2+\sigma_y^2}\right)dx$$

$$f\left(z\right) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_y^2}}\exp\left(-\frac{\left(z-\left(\mu_x+\mu_y\right)\right)^2}{\sigma_x^2+\sigma_y^2}\right)\int_{-\infty}^{\infty}B\left(x\right)dx \tag{1.6}$$

, where

$$B\left(x\right) = \frac{1}{\sqrt{2\pi}\frac{\sigma_x\sigma_y}{\sqrt{\sigma_x^2+\sigma_y^2}}}\exp\left(-\frac{\left(x - \frac{\sigma_x^2\left(z-\mu_y\right)+\sigma_y^2\mu_x}{\sigma_x^2+\sigma_y^2}\right)^2}{2\left(\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}\right)}\right)$$

Note that $B\left(x\right)$ is a Gaussian distribution and therefore it must integrate to 1. Hence, equation (1.6) yields:

$$\boxed{f(z)=\frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^{\ 2}+\sigma_y^2}}\exp\left(-\frac{\left(z-\left(\mu_x+\mu_y\right)\right)^2}{\sigma_x^2+\sigma_y^2}\right)} \tag{1.7}$$

, which is nothing but a Gaussian distribution with mean

$$\boxed{\mu_z=\mu_x+\mu_y}$$

, and variance

$$\boxed{\sigma_z=\sigma_x^2+\sigma_y^2}$$

Exercise 2.

The sum $Z=X+Y$ of two Gaussian random variables $X \sim N(\mu_X,\sigma_X^2)$ and $Y \sim N(\mu_Y,\sigma_Y^2)$ is Gaussian with mean $\mu_Z=\mu_X+\mu_Y$ and variance $\sigma_Z^2=\sigma_X^2+\sigma_Y^2+2\rho\sigma_X\sigma_Y$, where ρ is the correlation coefficient given by $\rho=E[XY]/(\sigma_X\sigma_Y)$. Equation Section (Next)

Solution:

In the case of two dependent Gaussian random variables X , Y the sum $Z=X+Y$ is given by:

$$f(z)=\int_{-\infty}^{\infty} f_{xy}(x,z-x)dx \tag{2.1}$$

If the variables were independent then $f_{xy}(x,z-x)=f_x(x)f_y(z-x)$. In this case though this separation is not possible. But:

$$f_{xy}(x,y)=\frac{\exp\left[A(x,y)\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \tag{2.2}$$

, where:

$$A(x,y)=-\frac{1}{2\left(1-\rho^2\right)}\left(\frac{\left(x-\mu_x\right)^2}{\sigma_x^2}+\frac{\left(y-\mu_y\right)^2}{\sigma_y^2}-\frac{2\rho\left(x-\mu_x\right)\left(y-\mu_y\right)}{\sigma_x\sigma_y}\right)$$

To continue we have to substitute the above expression (2.2) into (2.1) and cleverly carry out the integral so that we end up with a distribution in z. This part however, involves a great deal of algebraic calculations so we will skip it and present the result only for the sake of simplicity. The distribution that we end up with is the following:

$$f(z)=\frac{e^{-\frac{\left(z-(\mu_x+\mu_y)\right)^2}{2\left(\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y\right)}}}{\sqrt{2\pi\left(\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y\right)}} \tag{2.3}$$

Therefore, the variable Z is also a Gaussian with mean:

$$\boxed{\mu_z=\mu_x+\mu_y} \tag{2.4}$$

, and variance:

$$\sigma_z=\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y \tag{2.5}$$

Exercise 3.

The mixture distribution is defined by $f(\underset{\sim}{x})=\sum_{i=1}^{n} w_i f_i(\underset{\sim}{x})$, where $f_i(\underset{\sim}{x}), i=1,\dots,n$ are the mixture components and w_i are mixture weights which are non-negative $w_i \geq 0$ and satisfy $\sum_{i=1}^{n} w_i = 1$. The mixture components $f_i(\underset{\sim}{x})$ are probability distributions. Show that $f(\underset{\sim}{x})$ is a probability distribution. Estimate the first and second moment of the mixture distribution in terms of the first and second moment of the mixture components. Estimate the variance of the mixture distribution. Equation Section (Next)

Solution:

a. Integrating $f(\underset{\sim}{x})$ from $-\infty$ to ∞ we get:

$$\begin{gathered}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f(\underset{\sim}{x})\, dx_1 dx_2 \dots dx_m = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\left(\sum_{i=1}^{n} w_i f_i(\underset{\sim}{x})\right) dx_1 dx_2 \dots dx_m = \\ \sum_{i=1}^{n} w_i\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f_i(\underset{\sim}{x})\, dx_1 dx_2 \dots dx_m\right)\end{gathered} \tag{3.1}$$

But since the components $f_i(\underset{\sim}{x})$ are probability distributions from definition we also have:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f_i(\underset{\sim}{x})\, dx_1 dx_2 \dots dx_m = 1 \tag{3.2}$$

Combining (3.1) and (3.2) we get:

$$\boxed{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m = \sum_{i=1}^{n} w_i = 1}$$

Which means that the mixture distribution $f(\underset{\sim}{x})$ is also a probability distribution since it integrates to 1.

b. The first moment of a probability distribution is from definition

$$\underset{\sim}{E}\left[\underset{\sim}{x}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \underset{\sim}{x}\, f\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m \tag{3.3}$$

Let us define the first moments of the mixture components $f_i(\underset{\sim}{x})$ as:

$$\underset{\sim}{\mu}^i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \underset{\sim}{x}\, f_i\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m \tag{3.4}$$

, with the k-th component being

$$\mu_k^i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_k\, f_i\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m$$

Now we can estimate the first moment of the mixture distribution $f(\underset{\sim}{x})$ as:

$$\begin{gathered}\underset{\sim}{E}\left[\underset{\sim}{x}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \underset{\sim}{x}\, f\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \underset{\sim}{x}\, \left(\sum_{i=1}^{n} w_i f_i\left(\underset{\sim}{x}\right)\right) dx_1 dx_2 \ldots dx_m = \\ \sum_{i=1}^{n} w_i \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \underset{\sim}{x}\, f_i\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m\right)\end{gathered} \tag{3.5}$$

And by combining expressions (3.4) and (3.5) we can estimate the first moment of the mixture distribution with respect to the first moments of the mixture components as:

$$\boxed{\underset{\sim}{E}\left[\underset{\sim}{x}\right] = \sum_{i=1}^{n} w_i \underset{\sim}{\mu}^i} \tag{3.6}$$

, with the k-th component being

$$E_k\left[\underset{\sim}{x}\right] = \sum_{i=1}^{n} w_i \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_k\, f_i\left(\underset{\sim}{x}\right)\, dx_1 dx_2 \ldots dx_m\right)$$

c. The second moment of a probability distribution is defined as:

$$E_{ij}\left[\left(x_i-\mu_i\right)\left(x_j-\mu_j\right)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(x_i-\mu_i\right)\left(x_j-\mu_j\right)f\left(\underset{\sim}{x}\right)\,dx_1\ldots dx_i\ldots dx_j\ldots dx_m \tag{3.7}$$

Again we define the second moments of the mixture components $f_i(\underset{\sim}{x})$ as:

$$E^k\left[\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T f_k\left(\underset{\sim}{x}\right)\,d\underset{\sim}{x} \tag{3.8}$$

Hence, the second moment of the mixture distribution can be written as:

$$\begin{aligned}E\left[\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T\right]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T f\left(\underset{\sim}{x}\right)\,d\underset{\sim}{x}=\\ &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T\left(\sum_{k=1}^{n}w_k f_k\left(\underset{\sim}{x}\right)\right)d\underset{\sim}{x}=\\ &\sum_{k=1}^{n}w_k\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T f_k\left(\underset{\sim}{x}\right)\,d\underset{\sim}{x}\right)\end{aligned} \tag{3.9}$$

And by combining equations (3.8) and (3.9) we finally get:

$$\boxed{E\left[\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T\right]=\sum_{k=1}^{n}w_k E^k\left[\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)\left(\underset{\sim}{x}-\underset{\sim}{\mu}\right)^T\right]} \tag{3.10}$$

Exercise 4.

The mixture of Gaussian distributions is defined by $f(\underline{x})=\sum_{i=1}^{n}w_i f_i(\underline{x})$, where the mixture components $f_i(\underline{x}), i=1,\ldots,n$ are Gaussian, i.e. $f_i(\underline{x})=N(\underline{x};\underline{\mu}_i,\Sigma_i)$ and w_i are mixture weights which are non-negative $w_i\geq 0$ and satisfy $\sum_{i=1}^{n}w_i=1$. Estimate the mean and the variance of the mixture distribution. Find the marginal distribution of a parameter x_j in $\underline{x}$.Equation Section (Next)

Solution:

Recalling that the first and second distribution moments are essentially the mean and variance of the distribution respectively, we can use the expressions we derived in the previous Exercise for general distributions to determine the mixture's mean and variance (covariance matrix).

a. Mean:

$$\underset{\sim}{\mu}_i = \underset{\sim}{E}_i\left[\underset{\sim}{x}\right] = \int_{-\infty}^{\infty} \underset{\sim}{x}\, f_i\left(\underset{\sim}{x}\right)\, d\underset{\sim}{x} \tag{4.1}$$

And recalling expression (3.6) the mean of the Gaussian mixture distributions can be written as:

$$\boxed{\underset{\sim}{E}\left[\underset{\sim}{x}\right] = \sum_{i=1}^{n} w_i \underset{\sim}{\mu}^i} \tag{4.2}$$

b. Variance:

$$\Sigma_i = \mathrm{E}_i\left[\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^T\right] = \int_{-\infty}^{\infty} \left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^T f_i\left(\underset{\sim}{x}\right)\, d\underset{\sim}{x} \tag{4.3}$$

Similarly, recalling expression (3.10) we can express the variance of the Gaussian mixture as:

$$\boxed{\Sigma = \sum_{i=1}^{n} w_i \Sigma_i} \tag{4.4}$$

Note that both Σ and Σ_i are matrixes whom dimensions are $m \times m$.

c. Marginal Distribution

Let's define the marginal distribution of x_j in the mixture components as:

$$f_i\left(x_j\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_i\left(x_1, x_2, \ldots, x_m\right) dx_1 dx_2 \ldots dx_{j-1} dx_{j+1} \ldots dx_m$$

Then we can easily write the marginal distribution of x_j of the mixture distribution as:

$$\boxed{f\left(x_j\right) = \sum_{i=1}^{n} f_i\left(x_j\right)} \tag{4.5}$$

Prior System Analysis

Exercise 1.

Consider the mathematical model of a system represented by the equation $Y = aX_1 + bX_2 + e$, where X_1 and X_2 are uncertain parameters of the mathematical model of the system, $Y \in R$ is the output quantity of interest (QoI), and $e \in R$ represents the model error which is quantified by a Gaussian distribution $e \sim N(0, S)$, where $S \in R$. The parameters X_1 and X_2 are assumed to be independent with mean μ_1 and μ_2, respectively. Also the standard deviation of the parameters X_1 and X_2 are σ_1 and σ_2, respectively. The uncertainty in the output QoI is quantified by the simplified measures of uncertainty such as the mean μ_Y and standard deviation σ_Y. The variables a and b are known constants. Given the uncertainty in the parameters X_1 and X_2, find the uncertainty in the output QoI Y, i.e. find μ_Y and standard deviation σ_Y.
Equation Chapter (Next) Section 1

Solution:

The mathematical model can also be expressed by the following equivalent form:

$$Y = \underset{\sim}{A} \cdot \underset{\sim}{X} + e \tag{1.1}$$

, where

$$\underset{\sim}{A} = \begin{bmatrix} a & b \end{bmatrix} \quad , \quad \underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

The mean μ_Y of the QoI can be calculated as:

$$\mu_Y = E[Y] = E[\underset{\sim}{A} \cdot \underset{\sim}{X} + e] = \underset{\sim}{A} \cdot \underset{\sim}{E}[\underset{\sim}{X}] + E[e] = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + 0 \Rightarrow$$

$$\boxed{\mu_Y = a\mu_1 + b\mu_2} \tag{1.2}$$

Similarly, the variance σ_Y of the QoI is:

$$\sigma_Y = E\left[\left(Y-\mu_Y\right)^2\right] = E\left[\left(\underset{\sim}{A}\cdot\underset{\sim}{X}+e-\underset{\sim}{A}\cdot\underset{\sim}{\mu}\right)\left(\underset{\sim}{A}\cdot\underset{\sim}{X}+e-\underset{\sim}{A}\cdot\underset{\sim}{\mu}\right)^T\right] =$$

$$= E\left[\underset{\sim}{A}\cdot\underset{\sim}{X}\cdot\underset{\sim}{X}^T\cdot\underset{\sim}{A}^T+\left(\underset{\sim}{A}\cdot\underset{\sim}{X}\right)e-\underset{\sim}{A}\cdot\underset{\sim}{X}\cdot\underset{\sim}{\mu}^T\cdot\underset{\sim}{A}^T+e\left(\underset{\sim}{X}^T\cdot\underset{\sim}{A}^T\right)+e^2\right]+$$

$$E\left[-e\left(\underset{\sim}{\mu}^T\cdot\underset{\sim}{A}^T\right)-\underset{\sim}{A}\cdot\underset{\sim}{\mu}\cdot\underset{\sim}{X}^T\cdot\underset{\sim}{A}^T-\underset{\sim}{A}\cdot\underset{\sim}{\mu}e+\underset{\sim}{A}\cdot\underset{\sim}{\mu}\cdot\underset{\sim}{\mu}^T\cdot\underset{\sim}{A}^T\right]\Rightarrow$$

$$\sigma_Y = \underset{\sim}{A}\cdot E\left[\underset{\sim}{X}\cdot\underset{\sim}{X}^T\right]\cdot\underset{\sim}{A}^T+E\left[\left(\underset{\sim}{A}\cdot\underset{\sim}{X}\right)e\right]-\underset{\sim}{A}\cdot\underset{\sim}{E}\left[\underset{\sim}{X}\right]\cdot\underset{\sim}{\mu}\cdot\underset{\sim}{A}^T+E\left[e\left(\underset{\sim}{X}^T\cdot\underset{\sim}{A}^T\right)\right]+E\left[e^2\right]+$$

$$-E\left[e\right]\left(\underset{\sim}{\mu}^T\cdot\underset{\sim}{A}^T\right)-\underset{\sim}{A}\cdot\underset{\sim}{\mu}\cdot\underset{\sim}{E}\left[\underset{\sim}{X}^T\right]\cdot\underset{\sim}{A}^T-\underset{\sim}{A}\cdot\underset{\sim}{\mu}E\left[e\right]+\underset{\sim}{A}\cdot E\left[\underset{\sim}{\mu}\cdot\underset{\sim}{\mu}^T\right]\cdot\underset{\sim}{A}^T\Rightarrow$$

$$\sigma_Y = \underset{\sim}{A}\cdot\left[\Sigma+\underset{\sim}{\mu}\cdot\underset{\sim}{\mu}^T\right]\cdot\underset{\sim}{A}^T+0-\mu_Y^2+0+S-0-\mu_Y^2-0+\mu_Y^2\Rightarrow$$

$$\sigma_Y = \underset{\sim}{A}\cdot\left[\Sigma\right]\cdot\underset{\sim}{A}^T+S$$

$$\boxed{\sigma_z^2 = a^2\sigma_1^2+b^2\sigma_2^2+S} \tag{1.3}$$

Exercise 2.

Consider the mathematical model of a physical process represented by the equation $Y = a\cos(X_1-1)+E$ where X_1 is the uncertain parameter of the mathematical model of the system, $Y \in R$ is the output quantity of interest (QoI), and $E \in R$ represents the model error which is quantified by a Gaussian distribution $E \sim N(0,S)$, where $S \in R$. The variable a is known constant. The uncertainty in the output QoI is quantified by the simplified measures of uncertainty such as the mean $\underline{\mu}_Y$ and standard deviation σ_Y. Given the uncertainty in the parameter X_1, find the uncertainty in the output QoI Y in the following cases.

a. The parameter X_1 has mean $\mu_1 = 1$ and standard deviation σ.

b. The parameter X_1 is Gaussian with mean $\mu_1 = 1$ standard deviation σ. Find the result for any other distribution of the uncertain variable X_1 with mean $\mu_1 = 1$ and standard deviation σ.

c. The parameter X_1 is uniform with upper and lower bounds $1-b$ and $1+b$. Use the analytical approximations based on Taylor series expansion up to the quadratic term. Also find the exact estimate of the mean $\underline{\mu}_Y$ and standard deviation σ_Y and investigate the effect of level of the uncertainty in X_1 on the accuracy of the Taylor series expansion estimate by plotting the errors

$$|\mu_Y^{approx}-\mu_Y^{exact}|/\mu_Y^{exact}$$

$$|\sigma_Y^{approx}-\sigma_Y^{exact}|/\sigma_Y^{exact}$$

, as a function of b ranging from 0 to 1. Comment on the results. Equation Section (Next)

Solution:

This is a non-linear prior distribution model. Any non-linear model can be written in the form:

$$Y = g(X) + E$$

, where in this case $g(X) = a\cos(X-1)$. It can be easily proven that if $E \sim N(0,S)$ then:

$$\mu_Y = E[g(X)] \tag{2.1}$$

, and:

$$\sigma_Y^2 = E\left[g(X)^2\right] + S^2 - \mu_Y^2 \tag{2.2}$$

In order to evaluate expressions (2.1) and (2.2) however, we need the explicit expressions for both the non-linear function $g(x)$ but also for the distribution of the model parameter X , $f(X)$

a. If we are interested to quantify the uncertainty in the output quantity Y without any further information about the distribution in X then we have to proceed with a Taylor Series expansion of $g(x)$.

$$g(x) = g(x_0) + \left.\frac{dg(x)}{dx}\right|_{x_0}(x-x_0) + \frac{1}{2}\left.\frac{d^2g(x)}{dx^2}\right|_{x_0}(x-x_0)^2 + \cdots \tag{2.3}$$

, also:

$$g^2(x) = \left(g(x_0) + \left.\frac{dg(x)}{dx}\right|_{x_0}(x-x_0) + \frac{1}{2}\left.\frac{d^2g(x)}{dx^2}\right|_{x_0}(x-x_0)^2 + \cdots\right)^2 \tag{2.4}$$

Using the above Taylor expansions, expressions (2.1) and (2.2) become:

$$\mu_Y = E[g(x)] = g(x_0) + \left.\frac{dg(x)}{dx}\right|_{x_0}(\mu - x_0) + \frac{1}{2}\left.\frac{d^2g(x)}{dx^2}\right|_{x_0}\left(\sigma^2 + (\mu - x_0)^2\right) \tag{2.5}$$

$\sigma_Y^2 = E\left[g^2(x)\right] + S^2 - \mu_Y^2$, with

$$E\left[g^2(x)\right] = g^2(x_0) + 2g(x_0)\left.\frac{dg(x)}{dx}\right|_{x_0}(\mu - x_0) + \left\{\left.\frac{dg(x)}{dx}\right|_{x_0}^2 + g(x_0)\left.\frac{d^2g(x)}{dx^2}\right|_{x_0}\right\}\left(\sigma^2 + (\mu - x_0)^2\right)$$

The central point x_0 in the Taylor series expansion is usually chosen to be the most probable value of parameter X. In this case, we will evaluate the Taylor expansions at the mean $x_0 = \mu$. Doing so, simplifies the above expressions for the mean and variance in Y to:

$$\mu_Y = a\cos(1-1) - \frac{a}{2}\cos(1-1)\sigma^2 \Rightarrow \mu_Y = a\left(1 - \frac{\sigma^2}{2}\right) \tag{2.6}$$

$$\begin{gathered}\sigma_Y^2 = a^2\cos^2(0) + \left\{a^2\sin^2(0) - a^2\cos^2(0)\right\}\sigma^2 + S^2 - \mu_Y^2 \Rightarrow \\ \sigma_Y^2 = S^2 + a^2\left(1 - \sigma^2\right) - \mu_Y^2\end{gathered} \tag{2.7}$$

b. Ambiguous question?

c. In the case of a uniform prior distribution for X the exact uncertainty measures for the output QoI Y are calculated as follows:

$$\mu_Y^{exact} = E\left[g(X)\right] = \int_{-\infty}^{\infty} a\cos(X-1)f(x)dx = \frac{a}{2b}\int_{1-b}^{1+b}\cos(X-1)dx \Rightarrow$$

$$\boxed{\mu_Y^{exact} = \frac{a}{b}\sin(b)} \tag{2.8}$$

$$\left(\sigma_Y^{exact}\right)^2 = E\left[g^2(X)\right] + S^2 - \left(\mu_Y^{exact}\right)^2 = \int_{-\infty}^{\infty} a^2\cos^2(X-1)f(x)dx + S^2 - \left(\mu_Y^{exact}\right)^2 =$$

$$\frac{a^2}{2b}\int_{1-b}^{1+b}\cos^2(X-1)dx + S^2 - \left(\mu_Y^{exact}\right)^2 = \frac{a^2}{2b}\left(b + \frac{\sin(2b)}{2}\right) + S^2 - \left(\mu_Y^{exact}\right)^2 \Rightarrow$$

$$\boxed{\left(\sigma_Y^{exact}\right)^2 = S^2 + \frac{a^2}{2} + \frac{a^2\left(b\cos(b) - 2\sin(b)\right)\sin(b)}{2b^2}} \tag{2.9}$$

Now if we try to approximate these measures using a Taylor series expansion in $g(x)$ like we did before, we will get:

$$\mu_Y^{app} = g(\mu) + \frac{1}{2}\left.\frac{d^2 g(x)}{dx^2}\right|_{\mu} \sigma^2 \tag{2.10}$$

$$\left(\sigma_Y^{app}\right)^2 = g^2(\mu) + \left\{\left.\frac{dg(x)}{dx}\right|_{\mu}^2 + g(\mu)\left.\frac{d^2 g(x)}{dx^2}\right|_{\mu}\right\}\sigma^2 + S^2 - \left(\mu_Y^{app}\right)^2 \tag{2.11}$$

Now recall that for a uniform distribution with bounds (v, w) the mean and variance are known to be:

$$\mu = \frac{1}{2}(v+w) \qquad \sigma^2 = \frac{1}{12}(w-v)^2$$

Thus, the approximate expressions (2.10)-(2.11) for the mean and variance can be written as:

$$\mu_Y^{app} = a\left(1 - \frac{b^2}{6}\right) \tag{2.12}$$

$$\left(\sigma_Y^{app}\right)^2 = S^2 - \frac{1}{36}a^2 b^4 \tag{2.13}$$

Now let us form the percentage errors between the exact and the approximate expressions for the mean and variance as follows:

$$\frac{\left|\mu_Y^{app} - \mu_Y^{ex}\right|}{\mu_Y^{ex}} = \frac{\left|1 - \frac{b^2}{6} - \frac{\sin(b)}{b}\right|}{\frac{\sin(b)}{b}} \tag{2.14}$$

, and:

$$\frac{\left|\sigma_Y^{app} - \sigma_Y^{ex}\right|}{\sigma_Y^{ex}} = \frac{\left|\frac{1}{2} + \frac{\left(b\cos(b) - 2\sin(b)\right)\sin(b)}{2b^2} - \frac{1}{36}b^4\right|}{\frac{1}{2} + \frac{\left(b\cos(b) - 2\sin(b)\right)\sin(b)}{2b^2}} \tag{2.15}$$

We would like to get an idea on how these percentage errors depend on the bound length $2b$ of the uniform distribution of the input model parameter X . This dependence is shown in the two graphs below. The graphs are drown with respect to b .

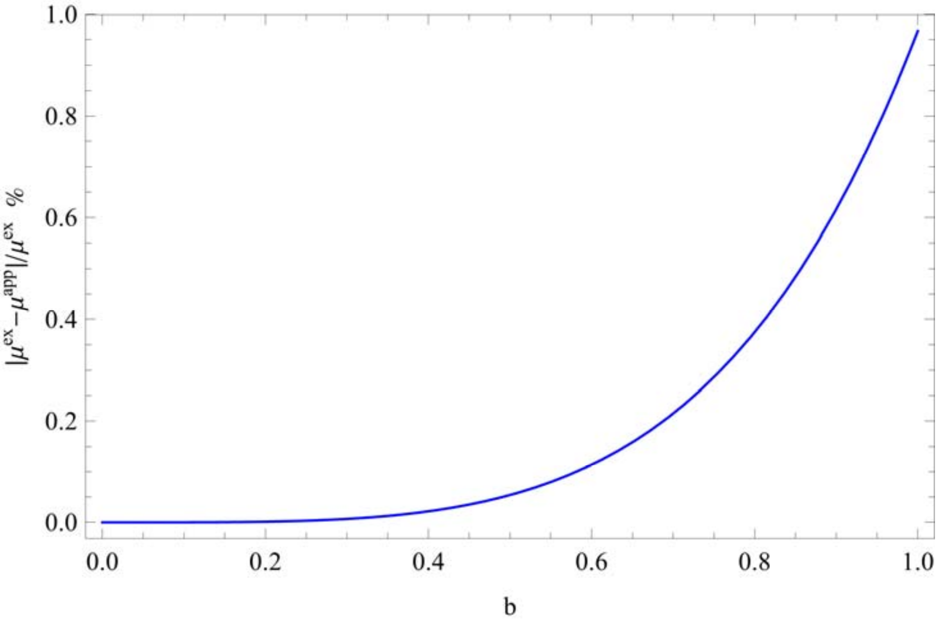


Figure 1. Percentage error between the exact and approximate expressions for the mean

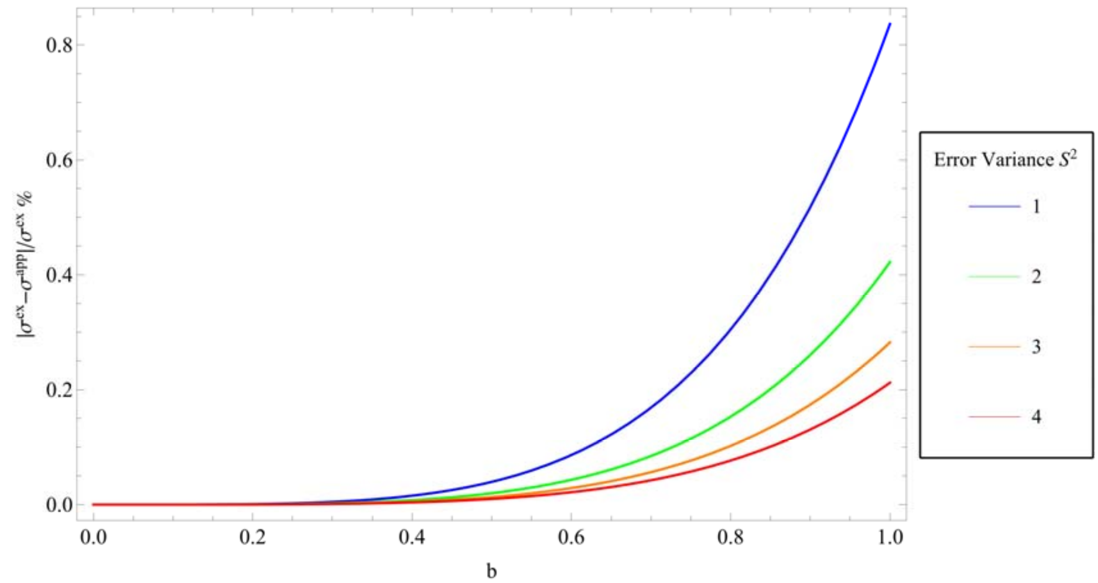


Figure 2. Percentage error between the exact and approximate expressions for the variance

Since the uniform prior distribution has bounds $[1-b, 1+b]$ we can infer that increasing b from 0 to 1 we reduce the length of the prior's support and thus increase uncertainty. We can see that if $b=0$ then the uniform prior's bounds are such to make both percentage errors equal to zero. This means that in the case of $b=0$ the exact and asymptotic expressions are essentially the same. In contrast, increasing b, increases the prior uncertainty and the percentage errors between the exact and asymptotic expressions for the mean and variance increase as well. In fact, in the extreme case of $b=1$, where the prior distribution region degenerates into a single point the percentage error for the mean is 100%. In all cases, larger uniform prior bounds yield a smaller error between exact and asymptotic expressions. In the second graph it is easy to infer that larger error variances S reduce the impact of the prior uncertainty and consequently reduce the percentage error.

Bayesian Inference and Posterior System Analysis

Exercise 1.

The posterior distribution of the parameters of a model is given by

$$p(\theta_1,\theta_2 \mid D,I) \propto \exp\left[-\frac{1}{2}\left(\theta_1^2+\theta_2^2+2\mu\theta_1\theta_2-2\mu\theta_1-2\theta_2+1\right)\right]$$

Find the uncertainty region and plot it in the two-dimensional parameter space (θ_1,θ_2).

<u>Hint</u>: Need to find the most probable point, the Hessian, the covariance matrix and then <u>clearly plot the contour plots</u> of the posterior distribution in the two-dimensional parameter space, indicate the principal direction of the ellipsoid, as well as the length of the uncertainty along the principal axes of the ellipsoid. Equation Chapter (Next) Section 1

Solution:

The posterior distribution can be written in the equivalent form:

$$p(\theta_1,\theta_2 \mid D,I) = A\exp\left[-\frac{1}{2}\left(\theta_1^2+\theta_2^2+2\mu\theta_1\theta_2-2\mu\theta_1-2\theta_2+1\right)\right] \quad (1.1)$$

, where A is a constant that is chosen so that the distribution integrates to 1. We next introduce the function $L(\theta_1,\theta_2)$ that is defined as the negative logarithm of the posterior distribution. Thus,

$$L(\theta_1,\theta_2) = A' + \frac{1}{2}\left(\theta_1^2+\theta_2^2+2\mu\theta_1\theta_2-2\mu\theta_1-2\theta_2+1\right) \quad (1.2)$$

We know that the most probable point $\left(\hat{\theta}_1,\hat{\theta}_2\right)$ is the point that minimizes the above expression. Thus,

$$\left.\begin{aligned}\frac{\partial L}{\partial\theta_1}=0 &\Rightarrow \theta_1+\mu\left(\theta_2-1\right)=0\\ \frac{\partial L}{\partial\theta_2}=0 &\Rightarrow \theta_2+\mu\theta_1-1=0\end{aligned}\right\}\Rightarrow \boxed{\begin{aligned}\hat{\theta}_1&=0\\ \hat{\theta}_2&=1\end{aligned}}$$

For the Hessian matrix we have:

$$H(\theta_1,\theta_2)=\begin{bmatrix}\dfrac{\partial^2 L}{\partial\theta_1^2} & \dfrac{\partial^2 L}{\partial\theta_1\partial\theta_2}\\ \dfrac{\partial^2 L}{\partial\theta_2\partial\theta_1} & \dfrac{\partial^2 L}{\partial\theta_2^2}\end{bmatrix}=\begin{bmatrix}1 & \mu\\ \mu & 1\end{bmatrix}$$

We can now expand the L function around the MPV using a Taylor series and retaining only up to the quadratic terms. Doing so, yields:

$$L(\theta_1,\theta_2)=L(\hat{\theta}_1,\hat{\theta}_2)+\frac{1}{2}Q(\theta_1,\theta_2) \tag{1.3}$$

, where

$$Q(\theta_1,\theta_2)=\left(\underset{\sim}{\theta}-\hat{\underset{\sim}{\theta}}\right)^{\mathrm{T}}\left[H\left(\hat{\underset{\sim}{\theta}}\right)\right]\left(\underset{\sim}{\theta}-\hat{\underset{\sim}{\theta}}\right) \tag{1.4}$$

The spread of uncertainty around the MPV is determined by the contour plots of the posterior distribution. The expansion in (1.3) as well as the quadratic form in (1.4) permit us to write the posterior distribution in the following form:

$$p(\theta_1,\theta_2 \mid D,I)\propto \exp\left[-\frac{1}{2}Q(\theta_1,\theta_2)\right] \tag{1.5}$$

Note however, that the contour plots of (1.5) are essentially the same as the contours of the quadratic form(1.4). We will now try to draw these contours so as to quantify the uncertainty in the posterior distribution. Recall that in order to draw the contours of $Q(\theta_1,\theta_2)$ we need to solve the eigenvalue problem for the Hessian matrix. The principal directions resemble the “directions” of the spread of uncertainty whereas the inverse of the eigenvalues represent the “intensity” of the spread along each direction. Hence,

$$\left[\hat{H}\right]\cdot \underset{\sim}{u}=\lambda \underset{\sim}{u} \tag{1.6}$$

Solving this eigenvalue problem yields the following eigenvalues:

$$\begin{aligned}\lambda_1 &= 1+\mu\\ \lambda_2 &= 1-\mu\end{aligned} \tag{1.7}$$

, and their respective eigenvectors:

$$\begin{aligned}\underset{\sim}{u}^1 &= (1,1)^T\\ \underset{\sim}{u}^2 &= (-1,1)^T\end{aligned} \tag{1.8}$$

Now we can write the quadratic form in (1.4) in the following form:

$$\underset{\sim}{y}^T \cdot [\Lambda] \cdot \underset{\sim}{y} = k^2 \Rightarrow \lambda_1 y_1^2 + \lambda_2 y_2^2 = k^2 \tag{1.9}$$

The new coordinate system (y_1, y_2) originates in the MPV and its directions coincide with the principal directions in(1.8). We are now ready to draw the contours that express the spread of uncertainty in the posterior distribution for different values of the parameter μ .

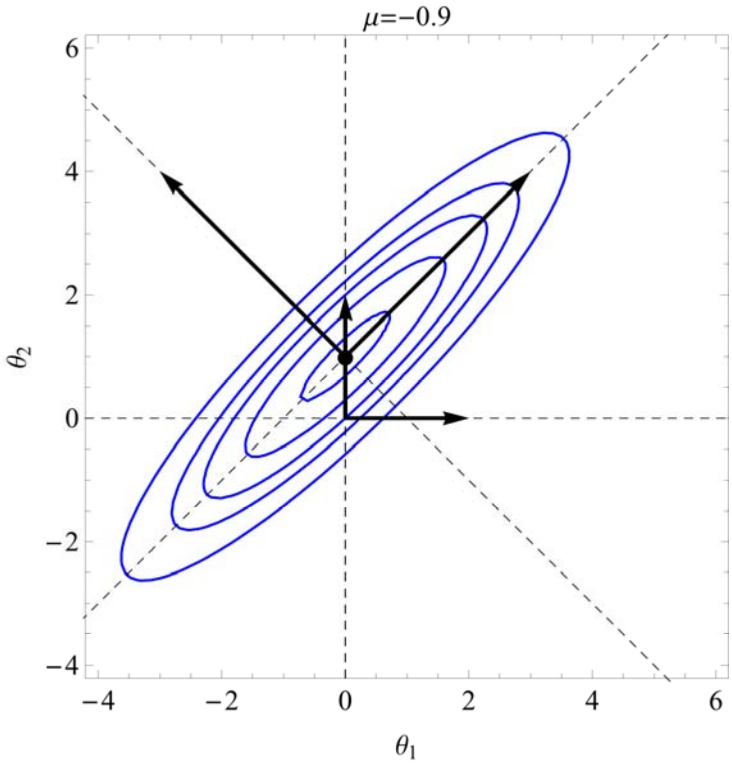


Figure 3. Contours that resemble the spread of uncertainty in the posterior PDF (μ=-0.9)

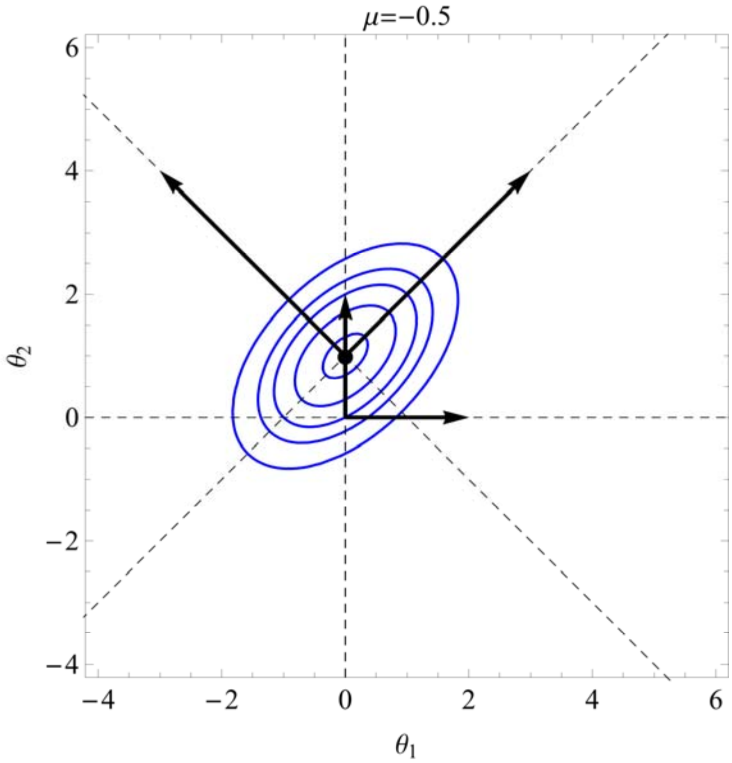


Figure 4. Contours that resemble the spread of uncertainty in the posterior PDF (μ=-0.5)

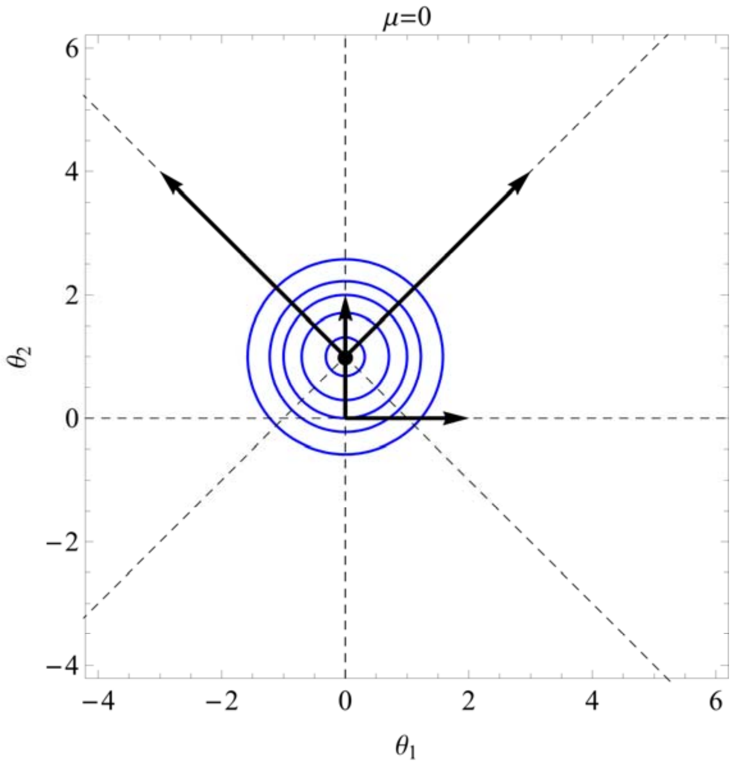


Figure 5. Contours that resemble the spread of uncertainty in the posterior PDF (μ=0)

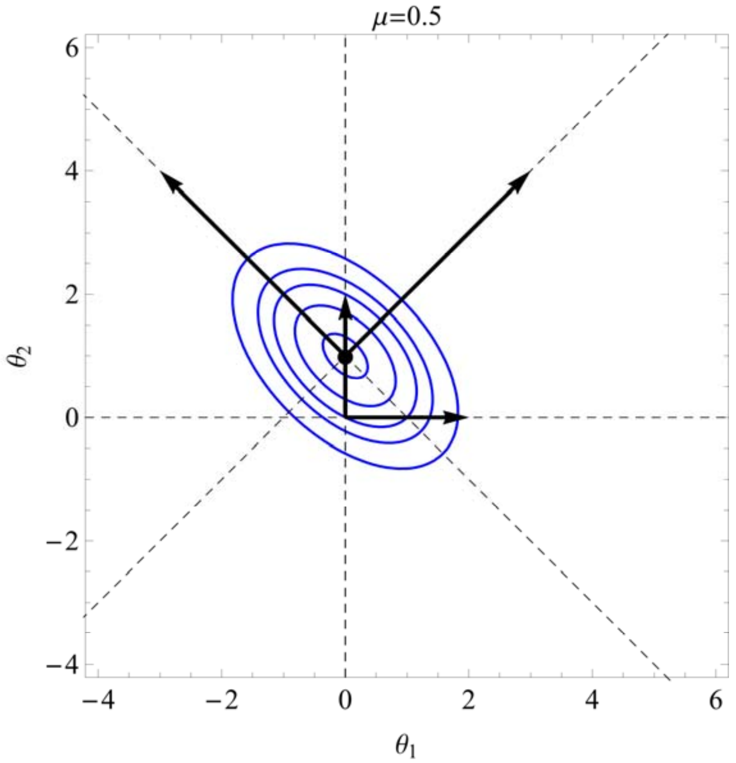


Figure 6. Contours that resemble the spread of uncertainty in the posterior PDF (μ=0.5)

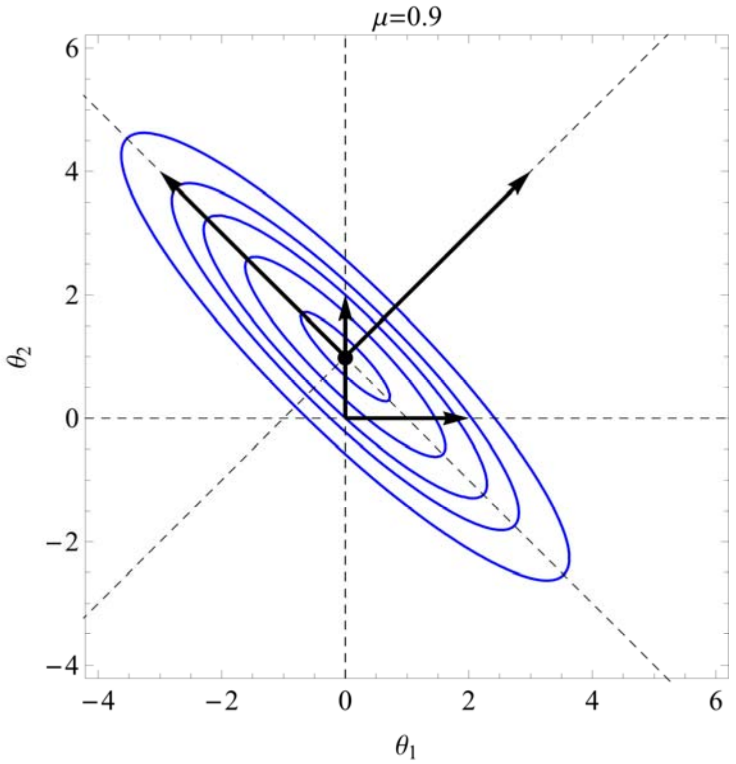


Figure 7. Contours that resemble the spread of uncertainty in the posterior PDF (μ=0.9)

Exercise 2.

Consider the mathematical model of a physical process/system represented by the equation

$$Y = aX_1 + E$$

, where X_1 is the uncertain parameter of the mathematical model of the system, $Y \in R$ is the output quantity of interest (QoI), and $E \in R$ represents the model error which is quantified by a Gaussian distribution $E \sim N(0, S)$, where $S \in R$ is known. Given the single measurement $\hat{Y} = y_0$

a. Find the posterior uncertainty in the model parameter X_1. The prior uncertainty in X_1 is quantified by
   i. a uniform distribution with very large bounds
   ii. a Gaussian distribution with mean μ and standard deviation σ

b. For the case (i), find the uncertainty in the output quantity of interest

$$Z = bY + \eta$$

, where the error term η is a Gaussian distribution with mean zero and variance S_0. Equation Section (Next)

Solution:

a.i. Assuming a uniform prior distribution for X we have:

$$f(X|I) = \begin{cases} \dfrac{1}{X_{\max} - X_{\min}}, & X \in [X_{\min}, X_{\max}] \\ 0, & otherwise \end{cases}$$

Since the bounds $X_{\min}, X_{\max}$ are taken to be very large, the region of uncertainty for the posterior distribution is assumed to be very small and contained within the support of the uniform prior. The posterior distribution according to Bayes theorem, is given by:

$$\underbrace{f(X|\hat{Y}, I)}_{\text{Posterior}} \propto \underbrace{f(\hat{Y}|X, I)}_{\text{Likelihood}} \underbrace{f(X|I)}_{\text{Prior}} \tag{2.1}$$

Hence we only have to determine the likelihood. Based upon our single observation the likelihood is:

$$Y \sim N(aX, S) \rightarrow f(\hat{Y}|X, I) = \frac{1}{\sqrt{2\pi S}} \exp\left[-\frac{1}{2S}(Y - aX)^2\right]$$

Therefore, the posterior PDF is analogous to:

$$f\left(X|\hat{Y},I\right)\propto\left(\frac{1}{\sqrt{2\pi S}}\exp\left[-\frac{1}{2S}\left(\hat{Y}-aX\right)^2\right]\right)\left(cst\right)\propto\frac{1}{\sqrt{2\pi S}}\exp\left[-\frac{1}{2S}\left(\hat{Y}-aX\right)^2\right] \tag{2.2}$$

In this case where the prior distribution is uniform with very large bounds the posterior PDF coincides with the likelihood and:

$$\boxed{f\left(X|\hat{Y},I\right)=\frac{1}{\sqrt{2\pi\frac{S}{\alpha^2}}}\exp\left[-\frac{1}{2\left(\frac{S}{\alpha^2}\right)}\left(X-\frac{\hat{Y}}{\alpha}\right)^2\right]} \tag{2.3}$$

a.ii. In the case of a Gaussian prior we have that:

$$f\left(X|\mu,I\right)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right] \tag{2.4}$$

However the expression for the likelihood remains the same as before and the posterior distribution according to (2.1) is given by:

$$f\left(X|\hat{Y},\mu,I\right)\propto\frac{1}{\sqrt{2\pi S}}\exp\left[-\frac{1}{2S}\left(\hat{Y}-aX\right)^2\right]\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right]\Rightarrow$$
$$\boxed{f\left(X|\hat{Y},\mu,I\right)\propto N\left(\frac{\hat{Y}}{a},\frac{S}{a^2}\right)N\left(\mu,\sigma^2\right)} \tag{2.5}$$

, which means that the posterior PDF is the product of two Gaussians. Recall that the product of two Gaussian distributions with means μ_1,μ_2 and variances σ_1^2,σ_2^2 is another Gaussian with mean and variance given by:

$$\mu=\frac{\frac{\mu_1}{2\sigma_1^2}+\frac{\mu_2}{2\sigma_2^2}}{\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}}\quad,\ \sigma^2=\frac{\left(\sigma_1\sigma_2\right)^2}{\sigma_1^2+\sigma_2^2}$$

b. We already shown that in the case of a uniform prior, the PDF of the linear model $f\left(Y|X,I\right)$ is a Gaussian distribution with mean aX and variance S. Now the model $Z=bY+\eta$ with $\eta\sim N\left(0,S_0\right)$ is essentially the sum of two Gaussian distributions. We have already shown that the sum of two independent Gaussians is another Gaussian with mean the sum of the means and variance the sum of the variances. Thus it is straightforward to write:

$$\boxed{Z\sim N\left(abX,S_0+S\right)}$$

Exercise 3.

Consider the mathematical model of a physical process/system represented by the equation

$$Y = aX^2 + E$$

, where X_1 is the uncertain parameter of the mathematical model of the system, $Y \in R$ is the output quantity of interest (QoI), and $E \in R$ represents the model error which is quantified by a Gaussian distribution $E \sim N(0, S)$, where $S \in R$ is known. Given the single measurement $\hat{Y} = y_0$,

a. Find the posterior uncertainty in the model parameter X_1 using Bayesian central limit theorem. The prior uncertainty in X_1 is quantified by a Gaussian distribution with mean μ and standard deviation σ.
b. Approximate the uncertainty in the output quantity of interest

$$Z = bY + \eta$$

, where the error term η is a Gaussian distribution with mean zero and variance S_0.
Equation Section (Next)

Solution:

a. We just have to repeat the process of determining the posterior distribution presented in the previous exercise but in this case the model is non-linear. Assuming a Gaussian prior we have:

$$f\left(X|\mu,I\right) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right] \tag{3.1}$$

The likelihood, given the single measurement $\hat{Y} = y_0$ is given by:

$$f\left(\hat{Y}\middle|X,I\right) = \frac{1}{\sqrt{2\pi S}}\exp\left[-\frac{1}{2S}\left(y_0 - aX^2\right)^2\right] \tag{3.2}$$

The posterior distribution using Bayes theorem is:

$$f\left(X\middle|\hat{Y},\mu,I\right) \propto f\left(\hat{Y}\middle|X,I\right) f\left(X\middle|\mu,I\right) \Rightarrow$$

$$f\left(X|\hat{Y},\mu,I\right) \propto \frac{1}{\sqrt{2\pi S}}\exp\left[-\frac{1}{2S}\left(y_0 - aX^2\right)^2\right]\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right] \Rightarrow$$

$$f\left(X\middle|\hat{Y},\mu,I\right)\propto\exp\left[-\frac{1}{2S}\left(y_0-aX^2\right)^2-\frac{1}{2\sigma^2}\left(X-\mu\right)^2\right] \tag{3.3}$$

Let us define the function $L(X)$ as:

$$L(X)=\frac{1}{2S}\left(y_0-aX^2\right)^2+\frac{1}{2\sigma^2}\left(X-\mu\right)^2 \tag{3.4}$$

Now using (3.4) the posterior PDF in (3.3) can be written in the following form:

$$f\left(X\middle|\hat{Y},\mu,I\right)\propto\exp\left[-L(X)\right]$$

In order to proceed and approximate the posterior PDF using the CLT we need to determine the MPV. To do so:

$$\hat{X}=\arg\min\left(L(X)\right)$$

$$\frac{\partial L}{\partial X}=0\Rightarrow\frac{X-\mu}{\sigma^2}+\frac{2aX\left(aX^2-y_0\right)}{S}=0$$

Solving the above equation for X however is not only rather difficult but also we will probably end up with a rather complex expression for $\hat{X}$. From now on we will assume that $\hat{X}$ is known and the Hessian matrix in this univariate case evaluated at the MPV is:

$$H\left[\hat{X}\right]=\frac{2a\left(3a\hat{X}^2-y_0\right)}{S}+\frac{1}{\sigma^2}$$

Let's define $\Sigma=\left(H\left[\hat{X}\right]\right)^{-1}$. According to the CLT the posterior PDF can now be approximated by the following Gaussian distribution:

$$f\left(X\middle|\hat{Y},\mu,\sigma,\mathrm{I}\right)=\frac{1}{\sqrt{2\pi\Sigma}}\exp\left[-\frac{1}{2\Sigma}\left(X-\hat{X}\right)^2\right] \tag{3.5}$$

, where recall that $\hat{X}$ stands for the MPV.

b. Now to approximate the posterior uncertainty in the QoI $Z=bY+\eta$. Recall that the distribution in Y is given by the likelihood in (3.2) which is a Gaussian on Y with mean aX^2 . Consequently, the QoI $Z=bY+\eta$ is the sum of two independent Gaussian random variables and its corresponding uncertainty can be quantified as:

$$Z \sim N\left(abX^2, S + S_0\right) \tag{3.6}$$

Exercise 4.

Consider a mathematical model of a system represented by the difference equation

$$Y_k = g(Y_{k-1}, \mu) + E$$

, where E is a Gaussian distribution, i.e. $E \sim N(0,\sigma^2)$. Given the observations $D \equiv (\hat{Y}_0, \hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{0\to N}$ covering all time instances, we are interesting in updating the uncertainty in the variables μ and σ^2. Find the likelihood of the model parameters μ and σ^2 .Equation Section (Next)

Using Bayes theorem, the posterior uncertainty in μ and σ^2 is given by:

$$f\left(\mu,\sigma^2 \middle| \left\{\hat{Y}_k\right\}_{0\to n}, I\right) \propto f\left(\left\{\hat{Y}_k\right\}_{0\to n} \middle| \mu,\sigma^2, I\right) f\left(\mu,\sigma^2 \middle| I\right) \tag{4.1}$$

The likelihood is essentially the distribution:

$$f\left(\left\{\hat{Y}_k\right\}_{0\to n} \middle| \mu,\sigma^2, I\right) \tag{4.2}$$

We can write that:

$$f\left(\left\{\hat{Y}_k\right\}_{1\to N} \middle| \mu,\sigma^2, I\right) = f\left(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N \middle| \mu,\sigma^2, I\right)$$

Due to the nature of the model (difference model) we have that:

$$f\left(\left\{\hat{Y}_k\right\}_{0\to N} \middle| \mu,\sigma^2, I\right) = f\left(\hat{Y}_N \middle| \hat{Y}_{N-1}, \mu,\sigma^2, I\right) f\left(\hat{Y}_{N-1} \middle| \hat{Y}_{N-2}, \mu,\sigma^2, I\right) \cdots f\left(\hat{Y}_1 \middle| \hat{Y}_0, \mu,\sigma^2, I\right)$$

In terms of Gaussian distributions we can write:

$$f\left(\left\{\hat{Y}_k\right\}_{0\to N} \middle| \mu,\sigma^2, I\right) \sim N\left(g\left(Y_{N-1},\mu\right),\sigma^2\right) N\left(g\left(Y_{N-2},\mu\right),\sigma^2\right) \cdots N\left(g\left(Y_0,\mu\right),\sigma^2\right) \Rightarrow$$

$$\boxed{f\left(\left\{\hat{Y}_k\right\}_{0\to N} \middle| \mu,\sigma^2, I\right) \sim \prod_{i=1}^{n} N\left(g\left(Y_{i-1},\mu\right),\sigma^2\right)} \tag{4.3}$$

Exercise 5.

The posterior uncertainty in two parameters x_1 and x_2 is found to be Gaussian with mean $\underline{\hat{x}} = (3,3)^T$ and covariance matrix $C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, where $-1 < \rho < 1$

a. Plot the spread of uncertainty around the best estimate $\underline{\hat{x}}$ (that is, plot a contour plot corresponding to $Q(\underline{x}) = 1$), for values of $\rho = 0, 0.1, 0.5, 0.9$

   Hint: Solve the eigenvalue problem $C\underline{\upsilon} = \mu\underline{\upsilon}$ and use the results in class to draw the contour plots. Note that $H^{-1} = C$ and that $\underline{u}$ and λ obtained from the eigenvalue problem $H\underline{u} = \lambda\underline{u}$ developed in class are related to $\underline{\upsilon}$ and μ as follows:

$$\underline{\upsilon} = \underline{u}$$
$$\mu = \frac{1}{\lambda}$$

b. Also, estimate the uncertainty in the marginal distribution of x_1 or x_2. Can the uncertainty in the marginal distribution of x_1 or x_2 describe the spread of uncertainty in the two dimensional space (x_1, x_2) of the two parameters?Equation Section (Next)

Solution:

a. Recall that the Hessian matrix of the posterior distribution is from definition the inverse of the covariance matrix. Equivalently, the covariance matrix is obtained as the inverse of the Hessian. Thus, in this case where the covariance matrix is known we can derive the Hessian as:

$$H = [C]^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \tag{5.1}$$

With the Hessian matrix known, we can write the quadratic form $Q(\underset{\sim}{x})$ as:

$$Q(\underset{\sim}{x}) = (\underset{\sim}{x} - \underset{\sim}{\hat{x}})^T \cdot [H] \cdot (\underset{\sim}{x} - \underset{\sim}{\hat{x}}) \tag{5.2}$$

, where $\underset{\sim}{\hat{x}}$ stands for the most probable value. Now recall that the contour plots of the posterior distribution are essentially the contours of the quadratic term $Q(\underset{\sim}{x})$ as we showed previously (Equations (1.1)-(1.5) at Exercise 1, Bayesian Inference and Posterior System Analysis Chapter). However, in order to plot the contours of (5.2) we first need to solve the eigenvalue problem for the Hessian to determine the principal directions and the eigenvalues. The latter is important to identify the location and "size" of uncertainty region.

The Eigen problem formulated as $[H]\cdot \underset{\sim}{u} = \lambda \underset{\sim}{u}$ has the following solution:

Eigenvalues: $$\lambda_1 = \frac{1}{1-\rho} \quad , \quad \lambda_2 = \frac{1}{1+\rho} \tag{5.3}$$

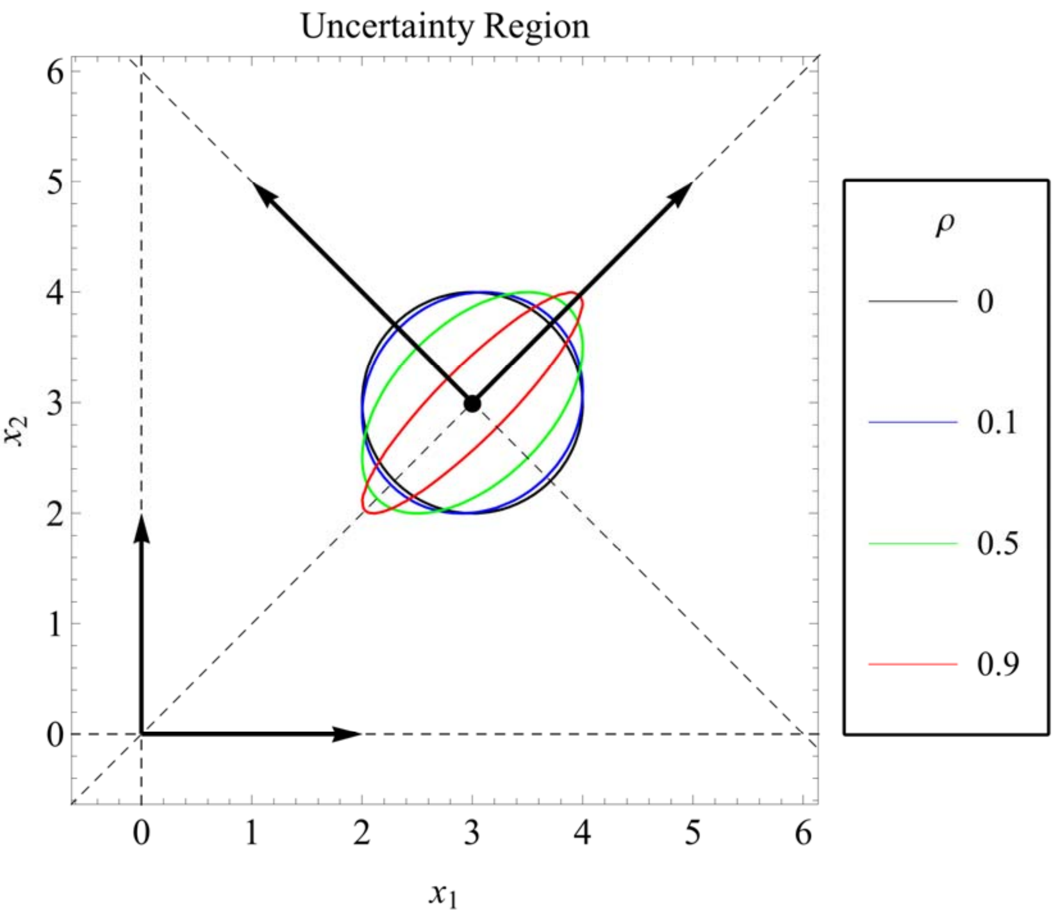
And the corresponding eigenvectors are:

$$\underset{\sim}{u}_1 = (-1,1) \quad , \quad \underset{\sim}{u}_2 = (1,1) \tag{5.4}$$

Recalling that the spread of uncertainty along the principal directions is given by the inverse of the corresponding eigenvalue, we can calculate the spread μ_1, μ_2 along the eigenvectors $\underset{\sim}{u}_1, \underset{\sim}{u}_2$ as:

$$\mu_1 = 1-\rho \quad , \quad \mu_2 = 1+\rho \tag{5.5}$$

Finally the contours are drawn and we summarize the different contours that correspond to different values of ρ for $Q=1$ in the graph below. The graph also indicates the principal directions as black bold vectors that originate at the most probable value (black point).



b. According to the Central Limit theorem (CLT) we can locally approximate our unknown posterior PDF as a multivariate Gaussian distribution centered at the MPV with variance matrix C which are given. Hence:

$$p(\underset{\sim}{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^2 \det C} \exp[-\frac{1}{2}(\underset{\sim}{x} - \hat{\underset{\sim}{x}})^T \cdot [C]^{-1} \cdot (\underset{\sim}{x} - \hat{\underset{\sim}{x}})] \tag{5.6}$$

The marginal distribution in either x_1 or x_2 can be obtained as:

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2)\, dx_2 \tag{5.7}$$

We can easily prove however that the above integration equals to a Gaussian marginal distribution for x_1 $x_1 \sim N(\hat{x}_1, C_{11})$. In complete form, the marginal distribution for x_1 is:

$$p(x_1) = \frac{1}{\sqrt{2\pi C_{11}}} \exp[-\frac{1}{2C_{11}}(x_1 - 3)^2] \tag{5.8}$$

It is reasonable to assume that the marginal distribution for x_2 will also be a Gaussian $x_2 \sim N(\hat{x}_2, C_{22})$. These two marginal distributions however do not suffice to quantify the posterior uncertainty as they completely omit the correlation between the two parameters. Thus, the marginal distributions are unable to describe the spread of uncertainty in the two dimensional space (x_1, x_2).

Exercise 6.

Inference of Acceleration of Gravity and Air Resistance Coefficient for a Falling Object

Consider the mathematical model of a falling object with mass m, acceleration of gravity g and air resistance force $F_{res} = -m\beta\upsilon^2$, where β is the air resistance coefficient. Using Newton's law, the equation of motion of the falling object is

$$m\frac{d\upsilon(t)}{dt} = mg - m\beta\upsilon^2(t)$$

, or equivalently

$$a(t) = g - \beta\upsilon^2(t)$$

Measurements for the acceleration and the velocity of the falling object are obtained at regular time intervals $k\Delta t$. The acceleration measurements are denoted by $(\hat{a}_1,\hat{a}_2,\ldots,\hat{a}_N)\equiv\{\hat{a}_k\}_{1\to N}$ and the corresponding velocity measurements are denoted by $(\hat{\upsilon}_1,\hat{\upsilon}_2,\ldots,\hat{\upsilon}_N)\equiv\{\hat{\upsilon}_k\}_{1\to N}$. Given the observation data $D\equiv(\hat{a}_1,\hat{a}_2,\ldots,\hat{a}_N,\hat{\upsilon}_1,\hat{\upsilon}_2,\ldots,\hat{\upsilon}_N)$ of the acceleration and velocity of the falling object at time instances $t=\Delta t, 2\Delta t,\ldots, N\Delta t$, respectively, we are interesting in estimating the uncertainty of the parameters g and β of the system. Note that the measurements and the model predictions satisfy the model error equation

$$\hat{a}_k = g - \beta\hat{\upsilon}_k^2 + E_k$$

$k=1,\ldots,N$, where the measurement error terms E_k are independent identically distributed (iid) and follow a zero-mean Gaussian distribution $E_k \sim N(0,\sigma^2)$. The value of the variance σ^2 is given.

Assume a uniform prior for the parameter set (g,β) and derive the expressions for the

1. Posterior PDF $p(g,\beta\,|\,D,\sigma,I)$.
2. The function $L(g,\beta) = -\ln p(g,\beta\,|\,D,\sigma,I)$
3. The MPV (or best estimate) $(\hat{g},\hat{\beta})$ of (g,β)
4. The uncertainty in the parameter space (g,β)
5. Derive the Gaussian asymptotic approximation for the posterior PDF of $p(g,\beta\,|\,D,\sigma,I)$. Is the Gaussian representation of the posterior uncertainty exact or approximate for this case?
6. Find the marginal distribution of the parameter β. Specifically,
   a. Give the uncertainty in β in terms of the mean and the standard deviation of the marginal distribution of β.
   b. Find the minimum number of data points required so that the uncertainty in β is less that a given value λ.
7. Find the uncertainty in the resistance force $F_{res} = -m\beta\upsilon^2$ given the uncertainties in the parameters (g,β):
   a. Compute the mean of F_{res}
   b. Compute the standard deviation of F_{res}
   c. Find the probability density function that describes the uncertainty in F_{res} Equation Section (Next)

Solution:

a. Given the set of independent observations $D \equiv (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_N, \hat{\upsilon}_1, \hat{\upsilon}_2, \ldots, \hat{\upsilon}_N)$ of the acceleration and velocity of a falling object we are interested in updating the uncertainty in the model parameters g, β . To begin with we postulate a uniform prior distribution in these parameters so that:

$$p\left(g,\beta \middle| I\right) = \begin{cases} \dfrac{1}{(g_{\max} - g_{\min})(\beta_{\max} - \beta_{\min})} & g \in [g_{\min}, g_{\max}], \beta \in [\mu_{\min}, \mu_{\max}] \\ 0 & otherwise \end{cases} \tag{6.1}$$

Given the model $\hat{a}_k = g - \beta \hat{\upsilon}_k^2 + E_k$ with $E_k \sim N(0, \sigma^2)$ we can infer the likelihood as:

$$\begin{aligned} p\left(D \middle| g,\beta,\sigma,I\right) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}\left(\hat{a}_i - \left(g - \beta\hat{\upsilon}_i^2\right)\right)^2\right] \Rightarrow \\ p\left(D \middle| g,\beta,\sigma,I\right) &= \frac{1}{\sqrt{2\pi}^N \sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\hat{a}_i - \left(g - \beta\hat{\upsilon}_i^2\right)\right)^2\right] \end{aligned} \tag{6.2}$$

Now using Bayes theorem, the posterior PDF for the parameters g, β can be found to be:

$$\boxed{p\left(g,\beta \middle| D,\sigma,I\right) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\hat{a}_i - \left(g - \beta\hat{\upsilon}_i^2\right)\right)^2\right]} \tag{6.3}$$

b. Now the L function is:

$$L(g,\beta) = -\ln\left[p\left(g,\beta \middle| D,\sigma,I\right)\right] \Rightarrow \boxed{L(g,\beta) = \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\hat{a}_k - \left(g - \beta\hat{\upsilon}_k^2\right)\right)^2 + c} \tag{6.4}$$

c. The MPV of the parameters will be found as $\left(\hat{g}, \hat{\beta}\right) = \underset{g,\beta}{\arg\min}\left(L(g,\beta)\right)$. So:

$$\left.\begin{aligned} \frac{\partial L}{\partial g} = 0 &\Rightarrow -\frac{\sum_{i=1}^{N}\left(\hat{a}_i - \left(g - \beta\hat{\upsilon}_i^2\right)\right)}{\sigma^2} = 0 \\ \frac{\partial L}{\partial \beta} = 0 &\Rightarrow \frac{\sum_{i=1}^{N}\hat{\upsilon}_i^2\left(\hat{a}_i - \left(g - \beta\hat{\upsilon}_i^2\right)\right)}{\sigma^2} = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{A} - Ng + \beta\hat{U}^2 = 0 \\ \hat{\Sigma} - g\hat{U}^2 + \beta\hat{V}^2 = 0 \end{aligned}$$

, where:

$$\hat{A}=\sum_{i=1}^{N}\hat{a}_i \quad , \quad \hat{U}^2=\sum_{i=1}^{N}\hat{u}_i^2 \quad , \quad \hat{V}^2=\sum_{i=1}^{N}\hat{u}_i^4 \quad , \quad \hat{\Sigma}=\sum_{i=1}^{N}\hat{a}_i\hat{u}_i^2$$

The MPV for the parameter set is the solution of the above system and it is found to be:

$$\boxed{\begin{aligned}\hat{g}&=\frac{\hat{\Sigma}\hat{U}^2-\hat{A}\hat{V}^2}{\left(\hat{U}^2\right)^2-N\hat{V}^2}\\ \hat{\beta}&=\frac{N\hat{\Sigma}-\hat{A}\hat{U}^2}{\left(\hat{U}^2\right)^2-N\hat{V}^2}\end{aligned}} \tag{6.5}$$

d. To fully define the uncertainty in the parameter space we also need the covariance matrix. Let's first determine the Hessian matrix though and then calculate the covariance matrix as the inverse of the Hessian.

$$H\left[g,\beta\right]=\begin{bmatrix}\dfrac{N}{\sigma^2} & \dfrac{-\sum_{i=1}^{N}\hat{u}_i}{\sigma^2}\\ \dfrac{-\sum_{i=1}^{N}\hat{u}_i}{\sigma^2} & \dfrac{\sum_{i=1}^{N}\hat{u}_i^2}{\sigma^2}\end{bmatrix}=\frac{1}{\sigma^2}\begin{bmatrix}N & -\sum_{i=1}^{N}\hat{u}_i\\ -\sum_{i=1}^{N}\hat{u}_i & \sum_{i=1}^{N}\hat{u}_i^2\end{bmatrix}$$

By defining $\hat{Y}=\sum_{i=1}^{N}\hat{u}_i$ we can write the Hessian in a more "clean" form as:

$$H=\frac{1}{\sigma^2}\begin{bmatrix}N & -\hat{Y}\\ -\hat{Y} & \hat{U}^2\end{bmatrix} \tag{6.6}$$

Now we can inverse the Hessian to find the covariance matrix as:

$$\left[C\right]=\left(\frac{1}{\left(N\hat{U}^2-\hat{Y}^2\right)\sigma^2}\right)\begin{bmatrix}\hat{U}^2 & \hat{Y}\\ \hat{Y} & N\end{bmatrix} \tag{6.7}$$

The most probable value and the covariance matrix completely define the uncertainty in the parameter space $\left(g,\beta\right)$.

e. The asymptotic approximation for the posterior PDF according to the Bayesian CLT is the following Gaussian distribution:

$$p^{app}\left(g,\beta\middle|\sigma^2,D,I\right)=\frac{1}{\sqrt{2\pi\det[C]}}\exp\left[-\frac{1}{2}\left(\underset{\sim}{x}-\hat{\underset{\sim}{x}}\right)^T\cdot[C]^{-1}\cdot\left(\underset{\sim}{x}-\hat{\underset{\sim}{x}}\right)\right] \tag{6.8}$$

, where $\underset{\sim}{x}=\begin{bmatrix} g \\ \beta \end{bmatrix}$. This asymptotic approximation for the posterior uncertainty is approximate. We can refer back to the exact expression for the posterior PDF in (6.3) to see that even if we carry out all algebraic calculations the two expressions are not equivalent. We know however that for a sufficiently large amount of measurements D the two expressions are asymptotically equivalent.

f. The marginal distribution in β can be obtained by the posterior PDF as:

$$p\left(\beta\middle|g,D,\sigma^2,I\right)=\int_{-\infty}^{\infty} p\left(\beta,g\middle|D,\sigma^2,I\right)dg \tag{6.9}$$

If we substitute in the above integral the approximate expression we derived in (6.8) and if we carry out the integration we will find out that the marginal distribution in β is a Gaussian $\beta\sim N\left(\hat{\beta},C_{22}\right)$:

$$p\left(\beta\middle|g,D,\sigma^2,I\right)=\frac{1}{\sqrt{2\pi C_{22}}}\exp\left[-\frac{1}{2C_{22}}\left(\beta-\hat{\beta}\right)^2\right] \tag{6.10}$$

Therefore the uncertainty in β can be quantified by the simple measures of mean and standard deviation as $\beta\sim$ Gaussian with $\mu_\beta=\hat{\beta}$ and $\sigma_\beta=\sqrt{C_{22}}$. Analytical expressions for $\hat{\beta},C_{22}$ can be found in equations (6.5) and (6.7) above.

To determine the minimum data points in order for the uncertainty in β to be less than a specified value λ we write:

$$\sqrt{C_{22}}<\lambda\Rightarrow\sqrt{\frac{N}{\left(N\hat{U}^2-\hat{Y}^2\right)\sigma^2}}<\lambda\Rightarrow\frac{N}{\left(N\hat{U}^2-\hat{Y}^2\right)\sigma^2}<\lambda^2$$

$$N>\frac{\hat{Y}^2}{\hat{U}^2-\dfrac{1}{\lambda^2\sigma^2}} \tag{6.11}$$

g. Since the model for the resistance force is of the form $F_{res}=-m\beta\upsilon^2$ and we know that $\beta\sim N\left(\hat{\beta},C_{22}\right)$ we can rewrite the model as:

$$F_{res}=-m\hat{\beta}\upsilon^2+\eta$$

, where $\eta \sim N\left(0, C_{22}\right)$. Recall that u is known every time interval Δt in terms of measurements $\hat{u}_i$. However, the MPV for β, $\hat{\beta}$ as well as the variance both depend on the measurements $\hat{u}_i, \hat{a}_i$ as we showed in (6.5) and (6.7)

To simplify things we could assume that the model can be equivalently written in the form:

$$\left\{F_{res}\right\}_j = -m\hat{\beta}_j \hat{\upsilon}_j^2 + \eta_j \tag{6.12}$$

, with $\hat{\beta}_j = \dfrac{j \cdot \hat{\Sigma}_j - \hat{A}_j \hat{U}_j^2}{\left(\hat{U}_j^2\right)^2 - j \cdot \hat{V}_j^2}$ and $\hat{A}_j = \sum_{i=1}^{j} \hat{a}_i$, $\hat{U}_j^2 = \sum_{i=1}^{j} \hat{u}_i^2$, $\hat{V}_j^2 = \sum_{i=1}^{N} \hat{u}_i^4$, $\hat{\Sigma}_j = \sum_{i=1}^{j} \hat{a}_i \hat{u}_i^2$.

Also, $\eta_j \sim N\left(0, C_{22}(j)\right)$, where: $C_{22}(j) = \dfrac{j}{\left(j \cdot \hat{U}_j^2 - \hat{Y}_j^2\right)\sigma^2}$.

All the above state that both the mean and variance for the resistance force vary in general when new data is available every time interval Δt .Despite that, we can determine the mean and variance of F_{res} in a closed form as:

$$\text{Mean: } \hat{F}_{res}(j) = -m\hat{\beta}_j \hat{\upsilon}_j^2 \ , \ \hat{\beta}_j = \frac{j \cdot \hat{\Sigma}_j - \hat{A}_j \hat{U}_j^2}{\left(\hat{U}_j^2\right)^2 - j \cdot \hat{V}_j^2} \tag{6.13}$$

$$\text{Variance: } \left[\sigma_{F_{res}}(j)\right]^2 = \frac{j}{\left(j \cdot \hat{U}_j^2 - \hat{Y}_j^2\right)\sigma^2} \tag{6.14}$$

Recall that we denote with j the current time interval. Thus, we use the data from $t = 0$ to $t = j\Delta t$ in order to determine the mean and variance. Alternatively we could calculate the mean and variance of the resistance force at the end of the experiment when all the date would be available and therefore the results would be independent of time.

Each time interval j, the probability density function for the resistance force will be the following Gaussian:

$$p_j\left(F_{res} \middle| \hat{u}_j, \hat{a}_j\right) = \frac{1}{\sqrt{2\pi C_{22}(j)}} \exp\left[-\frac{1}{2}\left(F_{res} - \hat{F}_{res}^j\right)^2\right] \tag{6.15}$$

The MPV and variance for β depend on j in the sense that once new data is available we can use it to update the MPV and variance. Apart from β however, the resistance force is an implicit time function since the velocity varies for different time intervals.

Information Entropy

Exercise 1.

Estimate the information entropy for the exponential distribution $p(x) = \lambda \exp(-\lambda x)$, $x \geq 0$
Equation Chapter (Next) Section 1

Solution:

The Information entropy of a distribution is given by the following integral:

$$I_p = E_{\underset{\sim}{x}}\left[-\ln\left(p\left(\underset{\sim}{x}\right)\right)\right] = -\int_{\underset{\sim}{X}} p\left(\underset{\sim}{x}\right)\ln\left(p\left(\underset{\sim}{x}\right)\right)d\underset{\sim}{x} \tag{1.1}$$

In the case of a univariate exponential distribution the above integral takes the form:

$$I_p = -\lambda\int_0^\infty \exp[-\lambda x](\lambda - \lambda x)\,dx = \mathrm{e}^{-\lambda x}(\lambda - \lambda x - 1)\Big|_0^\infty = 1 - \lambda \tag{1.2}$$

Hence, for the exponential distribution we have that:

$$I_p = 1 - \lambda \tag{1.3}$$

Exercise 2.

Show that the maximum entropy distribution defined within the interval $[a,b]$ is the uniform distribution.
Equation Section (Next)

Solution:

We need to prove that the uniform distribution is the least informative form all other distributions defined within the interval $[a,b]$. The latter translates into the problem of determining the expression $p\left(\underset{\sim}{x}\right)$ that maximizes the information entropy within the desired support. Hence:

$$\begin{aligned} &Max\left(I_p = -\int_X p(x)\ln\left(p(x)\right)dx\right) \\ &s.t. \quad \int_a^b p(x)dx - 1 = 0 \end{aligned} \tag{2.1}$$

Introducing Lagrange multipliers we can express the Lagrange function as:

(2.2)

Let us introduce an arbitrary function $h(x)$ that integrates to 1. That way we can rewrite the above equation as:

$$L\left(p(x)\right)=\int_a^b\left[p(x)\left(\lambda-\ln\left(p(x)\right)\right)-\lambda h(x)\right]dx \tag{2.3}$$

We can alternatively maximize the following functional

$$f\left(p(x)\right)=p(x)\left(\lambda-\ln\left(p(x)\right)\right)-\lambda h(x) \tag{2.4}$$

Hence,

$$\frac{\partial f}{\partial p}=0\Rightarrow\lambda-\ln\left(p(x)\right)-1=0\Rightarrow p(x)=e^{\lambda-1}$$
$$\frac{\partial f}{\partial\lambda}=0\Rightarrow p(x)-h(x)=0$$

In other words, the first equation states that the distribution we are looking for is a constant since λ does not depend on x. Taking this into consideration we can apply to it the constraint equation to find:

$$\left.\begin{array}{l} p(x)=c \\ \int_a^b p(x)dx=1 \end{array}\right\}\Rightarrow c=\frac{1}{b-a}\Rightarrow$$

$$\boxed{p(x)=\frac{1}{b-a}\quad ,x\in[a,b]} \tag{2.5}$$

, which concludes the proof. The distribution that maximizes the information entropy given only that it is defined in the domain $[a,b]$ is the uniform distribution.

Exercise 3.

Show that the maximum entropy distribution given only the mean is the exponential distribution.

Equation Section (Next)

Solution:

Just like we did before, the problem is formulated as:

$$\begin{aligned} & Max\left(I_p = -\int_X p(x)\ln\left(p(x)\right)dx \right) \\ & s.t. \quad \int_X xp(x)dx - \mu = 0 \\ & \qquad \int_X p(x)dx - 1 = 0 \end{aligned} \tag{3.1}$$

Introducing Lagrange multipliers, the Lagrange function can be written as:

$$\begin{aligned} & L\left(p(x)\right) = -\int_X p(x)\ln\left(p(x)\right)dx + \lambda_1\left(\int_X xp(x)dx - \mu\right) + \lambda_2\left(\int_X p(x)dx - 1\right) \Rightarrow \\ & L\left(p(x)\right) = \int_X \left[p(x)\left(\lambda_1 x + \lambda_2 - \ln\left(p(x)\right)\right)\right]dx - \left(\lambda_1\mu + \lambda_2\right) \end{aligned} \tag{3.2}$$

Doing again the same trick, introducing the function $h(x)$ that integrates to 1 we can maximize the following expression:

$$f\left(p(x)\right) = p(x)\left(\lambda_1 x + \lambda_2 - \ln\left(p(x)\right)\right) - \left(\lambda_1\mu + \lambda_2\right)h(x) \tag{3.3}$$

Assigning all derivatives to zero we end up with:

$$\begin{aligned} & \frac{\partial f}{\partial p} = 0 \Rightarrow \ln\left(p(x)\right) = \lambda_1 x + \lambda_2 - 1 \Rightarrow p(x) = e^{\lambda_1 x + \lambda_2 - 1} \\ & \frac{\partial f}{\partial \lambda_1} = 0 \Rightarrow xp(x) = \mu h(x) \\ & \frac{\partial f}{\partial \lambda_2} = 0 \Rightarrow p(x) = h(x) \end{aligned}$$

If we demand that the exponential expression for $p(x)$ satisfies the constraints we have:

$$\left.\begin{aligned}&\int_X xe^{\lambda_1 x+\lambda_2-1}dx=\mu\Rightarrow\frac{\lambda_1 x-1}{\lambda_1^2}e^{\lambda_1 x+\lambda_2-1}\Big|_X=\mu\\&\int_X e^{\lambda_1 x+\lambda_2-1}dx=1\Rightarrow\frac{1}{\lambda_1}e^{\lambda_1 x+\lambda_2-1}\Big|_X=1\end{aligned}\right\}\tag{3.4}$$

Let us support the distribution on the positive x axis just to be able to complete our calculations. Hence:

$$\left.\begin{aligned}&\int_0^\infty xe^{\lambda_1 x+\lambda_2-1}dx=\mu\Rightarrow\frac{\lambda_1 x-1}{\lambda_1^2}e^{\lambda_1 x+\lambda_2-1}\Bigg|_0^\infty=\mu\\&\int_0^\infty e^{\lambda_1 x+\lambda_2-1}dx=1\Rightarrow\frac{e^{\lambda_1 x+\lambda_2-1}}{\lambda_1}\Bigg|_0^\infty=1\end{aligned}\right\}\Rightarrow e^{\lambda_2-1}\frac{e^{\lambda_1 x}}{\lambda_1}\Bigg|_0^\infty=1\overset{\lambda_1<0}{\Rightarrow}\boxed{\lambda_1=-e^{\lambda_2-1}}$$

, and also:

$$\left.\begin{aligned}&\frac{\lambda_1 x-1}{\lambda_1^2}e^{\lambda_1 x+\lambda_2-1}\Bigg|_0^\infty=\mu\\&\lambda_1=-e^{\lambda_2-1}\end{aligned}\right\}\Rightarrow\underbrace{-\lambda_1\frac{x}{\lambda_1}e^{\lambda_1 x}\Big|_0^\infty}_{0}-\int_0^\infty\frac{1}{\lambda_1}e^{\lambda_1 x}dx=\mu\Rightarrow\boxed{\frac{1}{\lambda_1}=-\mu}$$

Finally, if we interpret the first equation as:

$$\lambda_1=-e^{\lambda_2-1}\Rightarrow-\ln\left(\lambda_1\right)=\lambda_2-1\Rightarrow\lambda_2-1=-\ln\left(\frac{1}{\mu}\right)\Rightarrow\boxed{\lambda_2-1=\ln\left(\mu\right)}$$

, then the distribution can be written as:

$$\boxed{p\left(x\right)=\mu e^{-\mu x}}\tag{3.5}$$

, which completes the proof. Therefore we proved that among all distributions with the same mean and defined in the positive axis, the least informative is the exponential distribution.

Markov Chain Monte Carlo

Exercise 1.

The posterior probability density function of a set of two parameters $\underline{\theta} = (\theta_1, \theta_2)^T$ is Gaussian with mean $\underline{0}$ and diagonal covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}$$

Let $\underline{\theta}^{(j)}$ be the current sample in the Markov Chain Monte Carlo algorithm generated using a Metropolis-Hasting algorithm. Following Metropolis-Hasting algorithm, let $\underline{\xi}$ be the candidate sample drawn from a uniform distribution centered at the current sample $\underline{\theta}^{(j)}$. Let $\underline{\theta}^{(j)} = (1,0)^T$. If $\underline{\xi} \sim U([0,6],[0,2])$, is drawn from a uniform distribution with bounds $[0,6]$ for the first component ξ_1 and $[0,2]$ for the second component ξ_2

1. find the probability that the next sample in the chain will be $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,1)^T$
2. find the probability that the next sample in the chain will be $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,3)^T$
3. find the probability that the next sample in the chain will be $\underline{\theta}^{(j+1)} = \underline{\xi} = (3,0)^T$

Equation Chapter (Next) Section 1

Solution:

The proposal PDF for candidate points ξ is a uniform distribution such that $\underline{\xi} \sim U([0,6],[0,2])$. In addition, the posterior PDF is a Gaussian with zero mean and covariance matrix $[\Sigma]$ which is given. From the theory of the Metropolis-Hastings algorithm for Markov Chain – Monte Carlo Integration we know that in order to decide whether to accept or reject the candidate point ξ in the chain we have to evaluate the quotient:

$$Q\left(\underset{\sim}{\xi} \middle| \underset{\sim}{\theta}^{(j)}\right) = \frac{p\left(\underset{\sim}{\xi} \middle| D, I\right)}{p\left(\underset{\sim}{\theta}^{(j)} \middle| D, I\right)} \frac{q\left(\underset{\sim}{\theta}^{(j)} \middle| \underset{\sim}{\xi}\right)}{q\left(\underset{\sim}{\xi} \middle| \underset{\sim}{\theta}^{(j)}\right)} \tag{1.1}$$

After we evaluate the quotient, we accept the candidate point with probability:

$$a\left(\underset{\sim}{\xi} \middle| \underset{\sim}{\theta}^{(j)}\right) = \min\left\{1, Q\left(\underset{\sim}{\xi} \middle| \underset{\sim}{\theta}^{(j)}\right)\right\}$$

, or reject it with probability:

$$1 - a\left(\underset{\sim}{\xi}\middle|\underset{\sim}{\theta}^{(j)}\right) = \max\left\{1, 1 - Q\left(\underset{\sim}{\xi}\middle|\underset{\sim}{\theta}^{(j)}\right)\right\}$$

So, the next point in the chain will be:

$$\underset{\sim}{\theta}^{(j+1)} = \begin{cases} \underset{\sim}{\xi} & \text{with probability} \quad a\left(\underset{\sim}{\xi}\middle|\underset{\sim}{\theta}^{(j)}\right) \\ \underset{\sim}{\theta}^{(j)} & \text{with probability} \quad 1 - a\left(\underset{\sim}{\xi}\middle|\underset{\sim}{\theta}^{(j)}\right) \end{cases} \tag{1.2}$$

We have analytical expressions for the posterior and the proposal distribution which they are:

$$p\left(\underset{\sim}{\theta}^{(j)}\middle|D, I\right) = \frac{1}{6\pi}\exp\left[-\frac{1}{2}\left(\theta_1^2 + \frac{\theta_2^2}{3}\right)\right] \tag{1.3}$$

$$q\left(\underset{\sim}{\xi}\middle|D, I\right) = \begin{cases} 1/12 \ \text{ if } \ \xi_1 \in [0,6] \text{ , and } \xi_2 \in [0,2] \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{1.4}$$

So:

i. If we need the next sample in the chain to be $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,1)^T$ then:

$$Q\left(\underset{\sim}{\xi}^*\middle|\underset{\sim}{\theta}^{(j)}\right) = \frac{p\left(\underset{\sim}{\xi}^*\middle|D, I\right)}{p\left(\underset{\sim}{\theta}^{(j)}\middle|D, I\right)}\underbrace{\frac{q\left(\underset{\sim}{\theta}^{(j)}\middle|\underset{\sim}{\xi}^*\right)}{q\left(\underset{\sim}{\xi}^*\middle|\underset{\sim}{\theta}^{(j)}\right)}}_{1} = \frac{p\left(\{0,1\}^T\middle|D, I\right)}{p\left(\{1,0\}^T\middle|D, I\right)} = 1.55962$$

And therefore the probability that $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,1)^T$ will be the next point in the chain is:

$$a\left(\underset{\sim}{\xi}\middle|\underset{\sim}{\theta}^{(j)}\right) = \min\{1, 1.55962\} = 1$$

ii. . If we need the next sample in the chain to be $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,3)^T$ then:

$$Q\left(\underset{\sim}{\xi}^*\middle|\underset{\sim}{\theta}^{(j)}\right) = \frac{p\left(\underset{\sim}{\xi}^*\middle|D, I\right)}{p\left(\underset{\sim}{\theta}^{(j)}\middle|D, I\right)}\underbrace{\frac{q\left(\underset{\sim}{\theta}^{(j)}\middle|\underset{\sim}{\xi}^*\right)}{q\left(\underset{\sim}{\xi}^*\middle|\underset{\sim}{\theta}^{(j)}\right)}}_{1} = \frac{p\left(\{0,3\}^T\middle|D, I\right)}{p\left(\{1,0\}^T\middle|D, I\right)} = 1.$$

And therefore the probability that $\underline{\theta}^{(j+1)} = \underline{\xi} = (0,3)^T$ will be the next point in the chain is:

$$a\left(\underset{\sim}{\xi}^{*}\middle|\underset{\sim}{\theta}^{(j)}\right)=\min\{1,1\}=1$$

iii. Finally, If we need the next sample in the chain to be $\underline{\theta}^{(j+1)}=\underline{\xi}=(3,0)^{T}$ then:

$$Q\left(\underset{\sim}{\xi}^{*}\middle|\underset{\sim}{\theta}^{(j)}\right)=\frac{p\left(\underset{\sim}{\xi}^{*}\middle|D,I\right)}{p\left(\underset{\sim}{\theta}^{(j)}\middle|D,I\right)}\underbrace{\frac{q\left(\underset{\sim}{\theta}^{(j)}\middle|\underset{\sim}{\xi}^{*}\right)}{q\left(\underset{\sim}{\xi}^{*}\middle|\underset{\sim}{\theta}^{(j)}\right)}}_{1}=\frac{p\left(3,0^{T}\middle|D,I\right)}{p\left(\{1,0\}^{T}\middle|D,I\right)}=0.018$$

And therefore the probability that $\underline{\theta}^{(j+1)}=\underline{\xi}=(3,0)^{T}$ will be the next point in the chain is:

$$a\left(\underset{\sim}{\xi}^{*}\middle|\underset{\sim}{\theta}^{(j)}\right)=\min\{1,0.018\}=0.018$$

While the chain will remain in the previous point with a probability:

$$1-a\left(\underset{\sim}{\xi}^{*}\middle|\underset{\sim}{\theta}^{(j)}\right)=0.982$$