What is the value of experimentation & measurement? — Supplementary document on standard Bayesian inference results

Anonymous Author(s)

2019

In this document, intended to be used as a supplement to the paper "What is the value of experimentation & measurement?", we show a number of standard Bayesian inference results used when we derive the value of experimentation & measurement capabilities for prioritization. Section 1 shows the one-dimensional version used in calculating the mean and the variance of certain random variables, and Section 2 shows the multi-dimensional version, used in calculating the covariance of the random variables.

1 Mean & variance of a conditioned normal r.v.

Let $X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $\epsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$, where all X_ns and ϵ_ns are mutually independent. Further let $Y_n = X_n + \epsilon_n$, this means

$$Y_n \sim \mathcal{N}(\mu_X + \mu_\epsilon \, , \, \sigma_X^2 + \sigma_\epsilon^2) \tag{1}$$

and the conditional distribution of Y_n given $X_n = x$ is

$$(Y_n \,|\, X_n = x) \;\sim\; \mathcal{N}(x + \mu_\epsilon, \, \sigma_\epsilon^2). \tag{2}$$

A standard Bayesian inference result states the conditional distribution of X_n given $Y_n = y$ is a normal distribution with

$$\mu_{X_n|Y_n} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2}(y - \mu_\epsilon) + \frac{\sigma_\epsilon^2}{\sigma_X^2 + \sigma_\epsilon^2}\mu_X, \tag{3}$$

$$\sigma^2_{X_n|Y_n} = \frac{\sigma_\epsilon^2 \sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2} \, . \tag{4}$$

We begin by specifying the conditional distribution as the posterior distribution given X_n as the prior and $Y_n \mid X_n$ as the likelihood:

$$f_{X_n|Y_n}(x \mid y) = \frac{f_{Y_n|X_n}(y \mid x)\, f_{X_n}(x)}{f_{Y_n}(y)} \tag{5}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_\epsilon^2}} \exp\left(-\frac{1}{2}\frac{(y-(x+\mu_\epsilon))^2}{\sigma_\epsilon^2}\right) \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\right)}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2+\sigma_\epsilon^2}} \exp\left(-\frac{1}{2}\frac{(y-(\mu_X+\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\right)}. \tag{6}$$

Grouping the fraction terms and exponential terms together and simplifying them, we can rewrite the RHS of Equation (6) as

$$\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\left(-\frac{1}{2}\left[\frac{(y-(x+\mu_\epsilon))^2}{\sigma_\epsilon^2} + \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{(y-(\mu_X+\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\right]\right). \tag{7}$$

We then make x the principal term in the squared expressions. Note we can rewrite $(y-(x+\mu_\epsilon))^2$ as $(x-(y-\mu_\epsilon))^2$ and $(y-(\mu_X+\mu_\epsilon))^2$ as $(\mu_X-(y-\mu_\epsilon))^2$ likewise. Expression (7) can then be written as

$$\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\left(-\frac{1}{2}\left[\frac{(x-(y-\mu_\epsilon))^2}{\sigma_\epsilon^2} + \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{(\mu_X-(y-\mu_\epsilon)))^2}{\sigma_X^2+\sigma_\epsilon^2}\right]\right). \tag{8}$$

We now attempt to complete the square. The expression within the exponent can be grouped to the x^2, x, and constant terms, giving:

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\Bigg(-\frac{1}{2}\Bigg[x^2\left[\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_X^2}\right] - 2x\left[\frac{y-\mu_\epsilon}{\sigma_\epsilon^2}+\frac{\mu_X}{\sigma_X^2}\right] + \\ &\qquad\qquad \frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2} + \frac{\mu_X^2}{\sigma_X^2} - \frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\Bigg]\Bigg). \end{aligned} \tag{9}$$

Note we can rewrite the coefficient of each of the three terms with $\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}$, the intended variance of the posterior distribution, as the denominator. The coefficient for the x^2 and x terms are straightforward:

$$\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_X^2} = \frac{\sigma_X^2+\sigma_\epsilon^2}{\sigma_\epsilon^2\sigma_X^2} = \frac{1}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\ , \tag{10}$$

$$\frac{y-\mu_\epsilon}{\sigma_\epsilon^2} + \frac{\mu_X}{\sigma_X^2} = \frac{\sigma_X^2(y-\mu_\epsilon)+\sigma_\epsilon^2(\mu_X)}{\sigma_\epsilon^2\sigma_X^2} = \frac{\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon) + \frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X)}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\ . \tag{11}$$

To complete the square in Expression (9) properly, we expect to see the numerator for the constant term to be the square of the numerator in the RHS of Expression (11). We first transform the constant term to one that contains the desired denominator:

$$
\begin{aligned}
&\frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2}+\frac{\mu_X^2}{\sigma_X^2}-\frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\\
=&\frac{\sigma_X^2(y-\mu_\epsilon)^2+\sigma_\epsilon^2\mu_X^2-\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X-(y-\mu_\epsilon))^2}{\sigma_\epsilon^2\sigma_X^2}\\
=&\frac{\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)^2+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X^2-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}(\mu_X-(y-\mu_\epsilon))^2}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}.
\end{aligned}
\tag{12}
$$

This is followed by extracting the $(y-\mu_\epsilon)^2$ and μ_X^2 terms from the numerator of Expression (12), giving

$$
\frac{(y-\mu_\epsilon)^2\left[\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}\right]+\mu_X^2\left[\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}\right]+2\mu_X(y-\mu_\epsilon)\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}.
\tag{13}
$$

By observing $\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}=\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2$ and likewise for the σ_ϵ^2 case, as well as expanding the last numerator term of Expression (13), the constant term can be written as

$$
\frac{(y-\mu_\epsilon)^2\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2+\mu_X^2\left(\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2+2(y-\mu_\epsilon)\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\,\mu_X\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}},
\tag{14}
$$

which the numerator is simply the expected result expanded:

$$
\frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2}+\frac{\mu_X^2}{\sigma_X^2}-\frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}=\frac{\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X)\right)^2}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}.
\tag{15}
$$

Substituting Equations (10), (11), and (15) into Expression (9), we arrive at the PDF of the posterior distribution:

$$
\begin{aligned}
f_{X_n|Y_n}(x\,|\,y)&=\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{X_n|Y_n}}}\exp\left(-\frac{1}{2}\left[\frac{x^2-2x(\mu_{X_n|Y_n})+(\mu_{X_n|Y_n})^2}{\sigma^2_{X_n|Y_n}}\right]\right)\\
&=\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{X_n|Y_n}}}\exp\left(-\frac{1}{2}\left[\frac{(x-\mu_{X_n|Y_n})^2}{\sigma^2_{X_n|Y_n}}\right]\right),
\end{aligned}
\tag{16}
$$

where $\mu_{X_n|Y_n}$ and $\sigma^2_{X_n|Y_n}$ are that defined in Equations (3) and (4) respectively. This is clearly the PDF of a normal distribution.

2 Covariance of a conditioned normal r.v.

In this section we state the standard Bayesian inference result in the multi-dimensional case. We then use the generalized result to show that if the X_ns and Y_ns are set up as above, any X_i and X_j $(i \neq j)$ will still be uncorrelated even when the values of the corresponding Y_i and Y_j are known.

2.1 The general result

Eaton [1] has provided the following result regarding the conditional distribution for a multivariate normal distribution. Let $\boldsymbol{x}$ be a n-dimensional multivariate normal random variable. If we partition $\boldsymbol{x}$ into two components of dimension q and $n-q$, and its mean and covariance matrix accordingly such that

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right), \tag{17}$$

where $\boldsymbol{x}_1, \boldsymbol{\mu}_1 \in \mathbb{R}^q$, $\boldsymbol{x}_1, \boldsymbol{\mu}_1 \in \mathbb{R}^{n-q}$, and $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{22}$ have sizes $q \times q$, $q \times (n-q)$, $(n-q) \times q$, and $(n-q) \times (n-q)$ respectively, then the distribution of $\boldsymbol{x}_1$ conditional on $\boldsymbol{x}_2 = \boldsymbol{a}$ is also a multivariate normal with:

$$(\boldsymbol{x}_1 | \boldsymbol{x}_2 = \boldsymbol{a}) \sim \mathcal{N}\big(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_2)\,,\, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big). \tag{18}$$

2.2 Specializing the result to our setup

To show the uncorrelatedness of X_ns conditional on corresponding Y_ns, we revisit the setup in Section 1 and focus on the i^{th}- and j^{th}-indexed random variables. This allows us to construct a four-dimensional vector $\boldsymbol{x} = (X_i, X_j, Y_i, Y_j)^T$ and partition them into two two-dimensional components, so that the general result can be applied.

The setup specifies that $X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$, $\epsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$, and $X_i \perp \epsilon_j \; \forall i, j$. For the i^{th}- and j^{th}-indexed random variables we have

$$X_i \sim \mathcal{N}(\mu_X, \sigma_X^2)\,,\; X_j \sim \mathcal{N}(\mu_X, \sigma_X^2), \tag{19}$$

$$\epsilon_i \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)\,,\; \epsilon_j \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2), \tag{20}$$

where all the four random variables are independent from each other. The setup also specifies that $Y_n = X_n + \epsilon_n$, this yields

$$Y_i \sim \mathcal{N}(\mu_X + \mu_\epsilon,\, \sigma_X^2 + \sigma_\epsilon^2), \text{ and } Y_j \sim \mathcal{N}(\mu_X + \mu_\epsilon,\, \sigma_X^2 + \sigma_\epsilon^2), \tag{21}$$

where $X_i \perp Y_j \; \forall i \neq j$.

We then construct the mean vector and covariance matrix required by the general result. Expressions (19) and (21) provided the information required to complete the entire mean vector and most of the covariance matrix. We only need to obtain $\text{Cov}(X_i, Y_i)$ (and $\text{Cov}(X_j, Y_j)$, which amounts to the same quantity). By definition of covariance:

$$\text{Cov}(X_i, Y_i) = \mathbb{E}(X_i Y_i) - \mathbb{E}(X_i)\mathbb{E}(Y_i). \tag{22}$$

The first term is calculated by observing $Y_i = X_i + \epsilon_i$, $X_i \perp \epsilon_i$, and standard identities between expectations and variances:

$$\begin{aligned}\mathbb{E}(X_iY_i) =& \mathbb{E}(X_i(X_i + \epsilon_i)) = \mathbb{E}(X_i^2) + \mathbb{E}(X_i\epsilon_i) \\ =& \mathbb{E}^2(X_i) + \mathrm{Var}(X_i) + \mathbb{E}(X_i)\mathbb{E}(\epsilon_i).\end{aligned} \tag{23}$$

Using Expressions (19) and (20) when substituting the expectations and variances, we have from the above

$$\begin{aligned}\mathrm{Cov}(X_i, Y_i) =& \mathbb{E}^2(X_i) + \mathrm{Var}(X_i) + \mathbb{E}(X_i)\mathbb{E}(\epsilon_i) - \mathbb{E}(X_i)\mathbb{E}(Y_i) \\ =& \mu_X^2 + \sigma_X^2 + \mu_X\mu_\epsilon - \mu_X(\mu_X + \mu_\epsilon) = \sigma_X^2.\end{aligned} \tag{24}$$

This yields the mean vector and covariance matrix for $\boldsymbol{x} = (X_i, X_j, Y_i, Y_j)^T$ as

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_X \\ \mu_X + \mu_\epsilon \\ \mu_X + \mu_\epsilon \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & 0 & \sigma_X^2 & 0 \\ 0 & \sigma_X^2 & 0 & \sigma_X^2 \\ \sigma_X^2 & 0 & \sigma_X^2 + \sigma_\epsilon^2 & 0 \\ 0 & \sigma_X^2 & 0 & \sigma_X^2 + \sigma_\epsilon^2 \end{pmatrix}. \tag{25}$$

Partitioning $\boldsymbol{x}$ into two components $(X_i, X_j)^T$ and $(Y_i, Y_j)^T$, and using the general result above we then have:

$$\begin{pmatrix} X_i \\ X_j \end{pmatrix} \Bigg| \begin{pmatrix} Y_i \\ Y_j \end{pmatrix} = \begin{pmatrix} y_i \\ y_j \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}') \tag{26}$$

where

$$\begin{aligned}\boldsymbol{\mu}' &= \begin{pmatrix} \mu_X \\ \mu_X \end{pmatrix} + \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix} \begin{pmatrix} \sigma_X^2 + \sigma_\epsilon^2 & 0 \\ 0 & \sigma_X^2 + \sigma_\epsilon^2 \end{pmatrix}^{-1} \left[\begin{pmatrix} y_i \\ y_j \end{pmatrix} - \begin{pmatrix} \mu_X + \mu_\epsilon \\ \mu_X + \mu_\epsilon \end{pmatrix} \right], \\ \boldsymbol{\Sigma}' &= \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix} - \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix} \begin{pmatrix} \sigma_X^2 + \sigma_\epsilon^2 & 0 \\ 0 & \sigma_X^2 + \sigma_\epsilon^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_X^2 \end{pmatrix}\end{aligned} \tag{27}$$

Simplifying the expression by standard matrix operations we arrive at:

$$\boldsymbol{\mu}' = \begin{pmatrix} \frac{\sigma_\epsilon^2}{\sigma_X^2 + \sigma_\epsilon^2}\mu_X + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2}(y_i - \mu_\epsilon) \\ \frac{\sigma_\epsilon^2}{\sigma_X^2 + \sigma_\epsilon^2}\mu_X + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2}(y_j - \mu_\epsilon) \end{pmatrix}, \tag{28}$$

$$\boldsymbol{\Sigma}' = \begin{pmatrix} \frac{\sigma_\epsilon^2 \sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2} & 0 \\ 0 & \frac{\sigma_\epsilon^2 \sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2} \end{pmatrix}. \tag{29}$$

Note the mean and variance of X_i and X_j is that derived in Section 1. Furthermore, the covariance between X_i and X_j given Y_i and Y_j is zero, as claimed at the beginning of this section.

References

[1] Eaton, Morris L. (1983). Multivariate Statistics: a Vector Space Approach. John Wiley and Sons. pp. 116–117. ISBN 978-0-471-02776-8.