1.010 Uncertainty in Engineering
Fall 2008

Example Application 15

(Conditional second-moment analysis)

UNCERTAINTY UPDATING USING NOISY OBSERVATIONS

One of the uses of conditional distributions is in updating uncertainty on a variable of interest X based on observation of one or more other variables. For example, one may want to update uncertainty on rainfall tomorrow based on observation of rainfall today, the strength of beam 1 based on observation of the strength of beam 2, or soil compressibility at location A given soil compressibility at some other location B.

In certain cases, the observed variable is itself a measurement of X. For example, one may measure the strength of a concrete column by some nondestructive test, measure topographic elevation at a point using a satellite instrument with limited accuracy, or sample the water of a stream with an imprecise device to determine its degree of contamination. In all these cases, the measurement is not exact. We want to see how, based on such "noisy data", one can update uncertainty on the quantity of interest X.

The method described below is exact if the random variables involved are normally distributed, but is often used as an approximation for variables with any distribution.

Conditional Distributions of Variables with Joint Normal Distribution

Let X_1 and X_2 be jointly normal variables with mean values m_1 and m_2, variances σ_1^2 and σ_2^2, and correlation coefficient ρ. 0ne can show that the conditional distribution of $(X_1|X_2=x_2)$ is also normal, with mean value $m_{1|2}$ and variance $\sigma_{1|2}^2$ given by

$$\begin{aligned} m_{1|2} &= m_1 + \rho\frac{\sigma_1}{\sigma_2}\left(x_2 - m_2\right) \\ \sigma_{1|2}^2 &= \sigma_1^2(1-\rho^2) \end{aligned} \qquad (1)$$

Notice that the conditional mean depends on the observed value x_2 of X_2, whereas the conditional variance does not. Moreover, the conditional variance differs from the unconditional variance by the factor $(1 - \rho^2)$, which is smaller than 1 whenever X_1 and X_2 are dependent.

Application to Noisy Observations

Next we show how Eq. 1 can be used to update uncertainty on a quantity of interest X (e.g., X = load bearing capacity of the soil or concentration of a pollutant at a given location) after making a measurement of it.

The quantity of interest, X, is initially uncertain with mean value m and variance σ^2. To reduce this uncertainty (and for example determine whether X is below a critical level x* with probability at least P*), a measurement Z of X is made. If the measurement had no error, then X could be recovered exactly from Z, but in practice measurements are affected by errors (they are "noisy"). A simple model with noise is the so-called linear model, according to which Z is related to X as

$$Z = a + bX + \varepsilon \qquad (2)$$

where a and b are given deterministic constants and ε is an error term independent of X, with mean value zero and variance σ_ε^2. The problem is to update uncertainty on X based on the observed value of Z, say z.

To use the conditional moment results in Eq. 1, we need to find the mean value and variance of Z and the correlation coefficient between X and Z. After this is done, we may rename $X \rightarrow X_1$ and $Z \rightarrow X_2$ and use that equation. Second-moment propagation of uncertainty through linear functions gives

$$\begin{aligned} m_Z &= a + bm \\ \sigma_Z^2 &= b^2\sigma^2 + \sigma_\varepsilon^2 \qquad (3) \\ \mathrm{Cov}[X,Z] &= b\sigma^2 \end{aligned}$$

Using these results and the relationships

$$\rho\frac{\sigma_1}{\sigma_2} = \frac{\rho\sigma_1\sigma_2}{\sigma_2^2} = \frac{Cov[X_1,X_2]}{Var[X_2]}$$
$$\sigma_1^2(1-\rho^2) = Var[X_1] - \frac{\{Cov[X_1,X_2]\}^2}{Var[X_2]} \tag{4}$$

one obtains from Eq. 1:

$$E[X|Z = z] = m + h\left[\frac{z-a}{b} - m\right]$$
$$Var[X|Z = z] = \sigma^2(1-h) \tag{5}$$

where $h = \left(1 + \frac{\sigma_\varepsilon^2}{b^2\sigma^2}\right)^{-1}$. Like Eq. 1, Eq. 5 holds exactly if both X and ε have normal distribution and in approximation for other distributions.

A key role in Eq. 5 is played by the quantity h, for which some special cases may be noted:

1. suppose that $\sigma_\varepsilon^2 = 0$, or more in general that $\sigma_\varepsilon^2 << b^2\sigma^2$. This means that observations are without error or the contribution from X to the variance of Z far exceeds the contribution from ε (high "signal-to-noise ratio"). In this case h = 1 and Eq. 5 gives $E[X|Z = z] = (z - a)/b$ and $Var[X|Z = z] = 0$. This is of course the solution to the deterministic problem;

2. At the other extreme is the case of very noisy measurements, when $\sigma_\varepsilon^2 >> b^2\sigma^2$. In this case h is close to zero and Eq. 5 gives $E[X|Z = z] = m$ and $Var[X|Z = z] = \sigma^2$, i.e. no change in the state of uncertainty on X as a result of observing Z.

Problem 15.1

(a) Cases of practical interest are intermediate between the above two limiting cases. To understand the role of different factors in the informativeness of a linear experiment, set b = 1 and plot the posterior-to-prior variance ratio $\gamma = Var[X|Z = z]/\sigma^2$ *against* $\sigma_\varepsilon^2/b^2\sigma^2$*. Notice that* γ *is a measure of the information value of the experiment and that the ratio* $\sigma_\varepsilon^2/b^2\sigma^2$ *can be decreased by either reducing the variance of the measurement error* σ_ε^2 *or increasing the "gain" b.*

(b) Think of an application of the observation model presented above to a context of interest to you. Postulate a plausible prior uncertainty state and realistic observation model parameters. Derive the uncertainty updating equation and the posterior variance using Eq. 5.

Problem 15.2

(a) Extend the previous analysis to the vector case, i.e. consider $\underline{X}$ *to be a vector with n components and* $\underline{Z}$ *to be a vector with r components. Assume a linear relation between* $\underline{X}$ *and* $\underline{Z}$ *of the type* $\underline{Z} = \underline{a} + \underline{BX} + \underline{\varepsilon}$*, where* $\underline{a}$ *is a given vector,* $\underline{B}$ *is a given matrix, and* $\underline{\varepsilon}$ *is a random measurement error vector. Assume that* $\underline{X}$ *has joint normal distribution,* $\underline{\varepsilon}$ *has joint normal distribution, and* $\underline{X}$ *and* $\underline{\varepsilon}$ *are independent.*

(b) Extend the results for Part (a) to include dependence between $\underline{X}$ *and* $\underline{\varepsilon}$*.*

Best Linear Unbiased Estimation (BLUE) Theory

Equation 1 is often used also when X_1 and X_2 do not have joint normal distribution. In that case Eq. 1 may be regarded as an approximation or may be used with a different interpretation. Specifically, we show that, irrespective of the type of distribution, the expression for the conditional mean in Eq. 1 has the meaning of best linear unbiased estimator of X_1 from X_2 and the conditional variance in Eq. 1 has the meaning of associated estimation error variance.

Suppose that X_1 and X_2 have mean values, variances, and correlation coefficient as above, but are not necessarily normally distributed. Based on the observation of X_2, we form a linear estimator of X_1, $\hat{X}_1 = a + bX_2$, and look for coefficients a and b such that

1. The estimator is (unconditionally) unbiased, i.e. $E[\hat{X}_1] = E[X_1] = m_1$. This gives $a + bE[X_2] = a + bm_2 = m_1$. Therefore, $a = m_1 - bm_2$.

2. Among all linear unbiased estimators, $\hat{X}_1$ has minimum error variance. The error is $e = \hat{X}_1 - X_1$ and its variance is $\sigma_e^2 = \mathrm{var}[bX_2 - X_1] = \sigma_1^2 + b^2\sigma_2^2 - 2b\rho\sigma_1\sigma_2$. Taking the derivative with respect to b and setting it to zero gives $2b\sigma_2^2 - 2\rho\sigma_1\sigma_2 = 0$. Hence $b = \rho\, \sigma_1/\sigma_2$ and $a = m_1 - \rho\, m_2\, (\sigma_1/\sigma_2)$.

We conclude that the BLUE estimator of X_1 from X_2 is

$$\hat{X}_1 = m_1 + \rho(\sigma_1/\sigma_2)(x_2 - m_2) \tag{6}$$

The associated error variance is obtained by substituting $b = \rho\, \sigma_1/\sigma_2$ into the expression for σ_e^2. This gives

$$\sigma_e^2 = \sigma_1^2(1 - \rho^2) \tag{7}$$

Comparison of Eqs. 6 and 7 with Eq. 1 shows that the BLUE estimator for any joint distribution of X_1 and X_2 is identical to the conditional mean $m_{1|2}$ for jointly normal variables and that the conditional variance in Eq. 1 is also the error variance of the BLUE estimator. This correspondence significantly broadens the applicability of the normal distribution results.