Appendix A

Relationship between the Hessian and Covariance Matrix for Gaussian Random Variables

Consider a Gaussian random vector $\boldsymbol{\theta}$ with mean $\boldsymbol{\theta}^\star$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ so its joint probability density function (PDF) is given by:

$$p(\boldsymbol{\theta}) = (2\pi)^{-\frac{N_{\boldsymbol{\theta}}}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)\right] \tag{A.1}$$

The objective function can be defined as its negative logarithm:

$$J(\boldsymbol{\theta}) \equiv -\ln p(\boldsymbol{\theta}) = \frac{N_{\boldsymbol{\theta}}}{2}\ln 2\pi + \frac{1}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}| + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) \tag{A.2}$$

which is a quadratic function of the components in $\boldsymbol{\theta}$. By taking partial differentiations with respect to θ_l and $\theta_{l'}$, the (l, l') component of the Hessian matrix can be obtained:

$$\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) = \left.\frac{\partial^2 J(\boldsymbol{\theta})}{\partial\theta_l \partial\theta_{l'}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} = (\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1})^{(l,l')} \tag{A.3}$$

so the Hessian matrix is equal to the inverse of the covariance matrix:

$$\mathcal{H}(\boldsymbol{\theta}^\star) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \tag{A.4}$$

For Gaussian random variables, the second derivatives of the objective function are constant for all $\boldsymbol{\theta}$ because the objective function is a quadratic function of $\boldsymbol{\theta}$. Therefore, the Hessian matrix can be computed without obtaining the mean vector $\boldsymbol{\theta}^\star$.

The elements in the Hessian matrix carry the conditional information of the random vector because they are obtained by fixing all other parameters. The diagonal elements are the curvature of the objective function in the corresponding direction. The reciprocals of these diagonal

Bayesian Methods for Structural Dynamics and Civil Engineering Ka-Veng Yuen
 ISBN: 978-0-470-82454-2

elements are the conditional variances of the uncertain parameters in $\boldsymbol{\theta}$. However, the diagonal elements in the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ are the marginal variances of the parameters.

In many applications, the objective function is known only implicitly so the components of the Hessian matrix has to be computed numerically, e.g., by the finite difference method, and the diagonal elements are given by:

$$\begin{aligned}
\mathcal{H}^{(l,l)}(\boldsymbol{\theta}^\star) &= \left[\frac{\partial}{\partial\theta_l}\left(\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\
&\approx \frac{1}{\Delta\theta_l}\left[\left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l/2} - \left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l/2}\right] \\
&\approx \frac{1}{\Delta\theta_l}\left[\frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l)-J(\boldsymbol{\theta}^\star)}{\Delta\theta_l} - \frac{J(\boldsymbol{\theta}^\star)-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l)}{\Delta\theta_l}\right] \\
&= \frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l)-2J(\boldsymbol{\theta}^\star)+J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l)}{(\Delta\theta_l)^2}
\end{aligned} \tag{A.5}$$

where $\Delta\boldsymbol{\theta}_l$ is a vector with all elements being zero except the lth element equal to a properly selected step $\Delta\theta_l$ (> 0):

$$\Delta\boldsymbol{\theta}_l = [0, \ldots, 0, \Delta\theta_l, 0, \ldots, 0]^T \tag{A.6}$$

Furthermore, the off-diagonal elements can be computed as follows:

$$\begin{aligned}
\mathcal{H}^{(l,l')}(\boldsymbol{\theta}^\star) &= \left[\frac{\partial}{\partial\theta_{l'}}\left(\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \\
&\approx \frac{1}{2\Delta\theta_{l'}}\left[\left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_{l'}} - \left.\frac{\partial J(\boldsymbol{\theta})}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_{l'}}\right] \\
&\approx \frac{1}{2\Delta\theta_{l'}}\left[\frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})}{2\Delta\theta_l}\right. \\
&\qquad\qquad \left. - \frac{J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})-J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})}{2\Delta\theta_l}\right] \\
&= \frac{1}{4\Delta\theta_l\Delta\theta_{l'}}[J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})-J(\boldsymbol{\theta}^\star+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}) \\
&\qquad\qquad - J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'})+J(\boldsymbol{\theta}^\star-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'})]
\end{aligned} \tag{A.7}$$

where $\Delta\boldsymbol{\theta}_l$ and $\Delta\boldsymbol{\theta}_{l'}$ are vectors with zero elements except the lth and l'th elements equal to $\Delta\theta_l$ and $\Delta\theta_{l'}$, respectively.

Example. Gaussian Random Variable
Assume that θ is a Gaussian random variable with mean μ and variance σ^2 so its PDF is:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right] \tag{A.8}$$

Then, the objective function is given by:

$$J(\theta) \equiv -\ln\ p(\theta) = \frac{1}{2}\ln(2\pi) + \ln\ \sigma + \frac{(\theta-\mu)^2}{2\sigma^2} \tag{A.9}$$

which is a quadratic function of θ.

It is assumed that the objective function is known only implicitly so the finite difference method is used to estimate the Hessian and, hence, the variance of this random variable. By Equation (A.5), the Hessian can be evaluated at an arbitrary point θ':

$$\begin{aligned}
\mathcal{H}(\theta') &= \tfrac{J(\theta'+\Delta\theta)-2J(\theta')+J(\theta'-\Delta\theta)}{(\Delta\theta)^2} \\
&= \tfrac{1}{(\Delta\theta)^2}\left\{\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'+\Delta\theta-\mu)^2}{2\sigma^2}\right]-2\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'-\mu)^2}{2\sigma^2}\right]\right. \\
&\qquad \left.+\left[\tfrac{1}{2}\ln(2\pi)+\ln\ \sigma+\tfrac{(\theta'-\Delta\theta-\mu)^2}{2\sigma^2}\right]\right\} \\
&= \tfrac{1}{(\Delta\theta)^2}\left[\tfrac{(\Delta\theta)^2}{\sigma^2}\right] \\
&= \sigma^{-2}
\end{aligned} \tag{A.10}$$

It does not depend on the point θ' and the step size $\Delta\theta$ and this solution is exact. However, for other distributions, Equation (A.4) is not correct but it provides a good approximation if the uncertainty is small. Nevertheless, the Hessian calculated by the finite difference method depends on the point of evaluation and the step size. The point of evaluation can be fixed at the most probable value of the distribution and selection of the step size has to be carefully handled. This will be discussed in the next example.

Example. Gamma Random Variable
Assume that the positive-valued random variable θ is Gamma distributed with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. The PDF of θ is given by:

$$p(\theta) = \frac{\theta^{\alpha-1}\exp(-\theta/\beta)}{\beta^{\alpha}\Gamma(\alpha)},\ \ \theta > 0 \tag{A.11}$$

where Γ is the Gamma function. The mean and variance of this distribution are given by $\alpha\beta$ and $\alpha\beta^2$, respectively. If $\alpha = 1$, this distribution is deduced to the exponential distribution and the most probable value is $\theta^\star = 0$. If $\alpha > 1$, the most probable parameter is given by:

$$\theta^\star = (\alpha-1)\beta \tag{A.12}$$

Furthermore, the objective function can be defined as the negative logarithm of the PDF:

$$J(\theta) \equiv -\ln p(\theta) = -(\alpha-1)\ln\theta + \frac{\theta}{\beta} + \alpha\ln\beta + \ln\Gamma(\alpha),\ \ \theta > 0 \tag{A.13}$$

Again, it is assumed that the objective function is known only implicitly. By Equation (A.5), the Hessian can be computed:

$$\begin{aligned}
\mathcal{H}(\theta^\star) &= \frac{1}{(\Delta\theta)^2}\left\{\left[-(\alpha-1)\ln(\theta^\star+\Delta\theta)+\frac{\theta^\star+\Delta\theta}{\beta}\right]-2\left[-(\alpha-1)\ln\,\theta^\star+\frac{\theta^\star}{\beta}\right]\right.\\
&\qquad\left.+\left[-(\alpha-1)\ln(\theta^\star-\Delta\theta)+\frac{\theta^\star-\Delta\theta}{\beta}\right]\right\}\\
&= \frac{(\alpha-1)}{(\Delta\theta)^2}\left[-\ln(\theta^\star+\Delta\theta)+2\ln\,\theta^\star-\ln(\theta^\star-\Delta\theta)\right]\\
&= -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\ln\left(1+\frac{\Delta\theta}{\theta^\star}\right)+\ln\left(1-\frac{\Delta\theta}{\theta^\star}\right)\right]
\end{aligned} \tag{A.14}$$

In contrast to the Gaussian random variable, this solution depends on the point for evaluation and also the finite-difference step size.

For a random variable with a small uncertainty, the variance can be approximated by $\mathcal{H}(\theta^\star)^{-1}$. To demonstrate this, let $\alpha = \mu/\beta$ for Equation (A.11) so the mean is μ and the variance is $\mu\beta$. For small β, the Hessian estimated by finite difference method can be approximated by:

$$\begin{aligned}
\mathcal{H}(\theta^\star) &= -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\ln\left(1+\frac{\Delta\theta}{\theta^\star}\right)+\ln\left(1-\frac{\Delta\theta}{\theta^\star}\right)\right]\\
&\approx -\frac{(\alpha-1)}{(\Delta\theta)^2}\left[\frac{\Delta\theta}{\theta^\star}-\frac{1}{2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2+\cdots-\frac{\Delta\theta}{\theta^\star}-\frac{1}{2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2+\cdots\right]\\
&\approx \frac{(\alpha-1)}{(\Delta\theta)^2}\left(\frac{\Delta\theta}{\theta^\star}\right)^2\\
&= \frac{1}{(\mu-\beta)\beta}\\
&\approx \frac{1}{\mu\beta}
\end{aligned} \tag{A.15}$$

Therefore, the Hessian estimated variance provides a good approximation: $\mathcal{H}(\theta^\star)^{-1} \approx \mu\beta$. In the following, two cases are demonstrated and they correspond to a large variance and a small variance.

Case 1: $\alpha = 10$ and $\beta = 0.1$

This distribution corresponds to the more spread PDF in Figure A.1. In this case, the mean and variance are $\alpha\beta = 1$ and $\alpha\beta^2 = 0.1$, respectively. Therefore, the coefficient of variation is $1/\sqrt{10} \approx 32\%$ and it represents a case of large uncertainty. By using Equation (A.14), the Hessian and the estimated variance are shown in Figure A.2 for different finite-difference step sizes up to 0.5. It is clearly seen that the estimation depends on the step size. The correct value of the variance is 0.1 but the estimation is 10% off for this random variable with a large coefficient of variation.

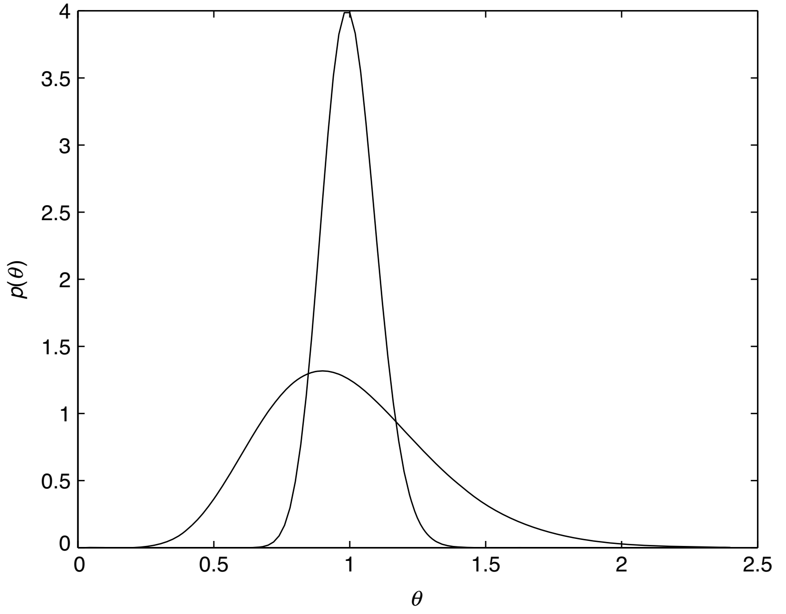

Figure A.1 Gamma distributions

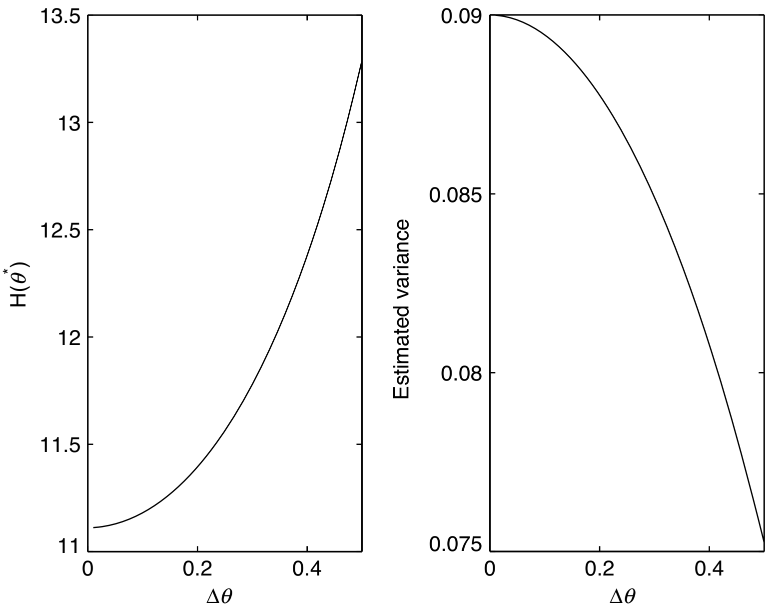

Figure A.2 Finite-difference estimated Hessian and variance ($\alpha = 10$, $\beta = 0.1$)

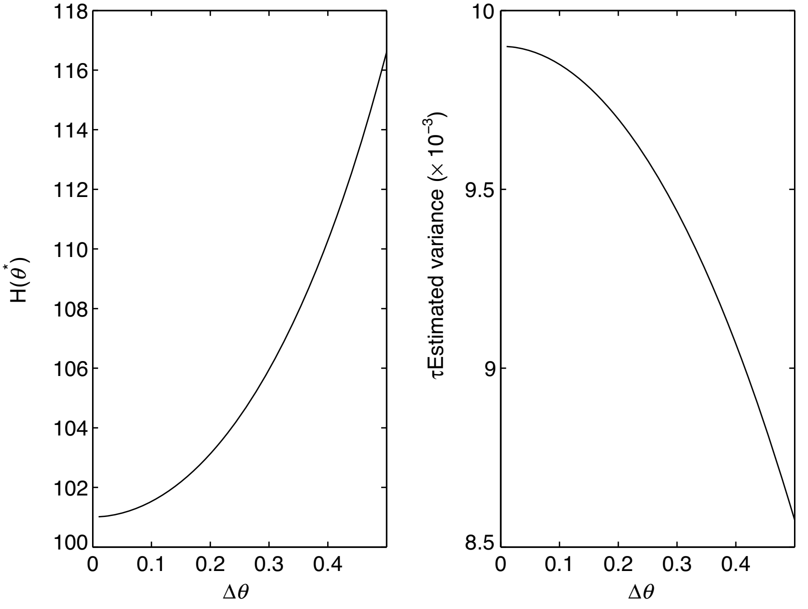

Figure A.3 Finite-difference estimated Hessian and variance ($\alpha = 100$, $\beta = 0.01$)

Case 2: $\alpha = 100$ and $\beta = 0.01$

This distribution corresponds to the more concentrated PDF in Figure A.1. In this case, the mean and variance are $\alpha\beta = 1$ and $\alpha\beta^2 = 0.01$, respectively, so the coefficient of variation is 10%. This random variable has the same mean as Case 1 but a smaller variance. Figure A.3 shows the Hessian and the estimated variance using different finite-difference step sizes. Again, the estimation depends on the step size. The correct value of the variance is 0.01 and the error of the estimation is about 1% when the step size is small. In general, the finite-difference estimated variance is more accurate for a smaller coefficient of variation of the distribution because the local topology of the PDF is more representative for the global distribution in this case. This example shows that this approximation is acceptable up to 10% COV.