

# A Proposed QUESO Algebraic Verification Problem

December 19, 2013

## 1 Theoretical Background

Let  $\epsilon_t$  denote a mean-zero Gaussian stochastic process having covariance function  $C(t_1, t_2|\phi) = (1/\lambda)R(t_1, t_2|\phi)$ , where  $\lambda > 0$  and  $R(\cdot, \cdot|\phi)$  is a correlation function for  $\phi \in \Phi$ . We assume the following quantities are specified:

1. Nominal parameter settings  $(\beta_0, \lambda_0, \phi_0)$  with  $\beta_0 \in \mathcal{R}^p$ ,  $\lambda_0 > 0$ , and  $\phi_0 \in \text{int}(\Phi)$
2. Indices  $\{t_1, t_2, \dots, t_n\}$ .

For specified  $n$ , generate a  $n$ -vector of errors

$$(\epsilon_{t_1}, \dots, \epsilon_{t_n}) \sim \mathcal{N}(\mathbf{0}_n, (1/\lambda_0)\mathbf{R}_n(\phi_0)) ,$$

where the  $(i, j)$  element of  $\mathbf{R}_n(\phi_0)$  is given by  $R(t_i, t_j|\phi_0)$ . To complete the data generation process, we sample  $d$ -dimensional covariates  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  independently from the distribution  $\mathcal{N}(\mathbf{0}_d, \mathbf{C})$ , where  $\mathbf{C}$  is a fixed covariance matrix. The  $i$ -th datum is calculated as  $y_{t_i} = \mathbf{f}^T(\mathbf{x}_i)\beta_0 + \epsilon_{t_i}$ , where  $\mathbf{f}(\cdot)$  is a  $p$ -dimensional regression function. In vector-matrix form,

$$\mathbf{y}_n = \mathbf{X}_n\beta_0 + \epsilon_n ,$$

where

$$\begin{aligned} \mathbf{y}_n &= (y_{t_1}, y_{t_2}, \dots, y_{t_n})^T , \\ \mathbf{X}_n &= [\mathbf{f}(\mathbf{x}_1) \ \mathbf{f}(\mathbf{x}_2) \ \dots \ \mathbf{f}(\mathbf{x}_n)]^T , \text{ and} \\ \epsilon_n &= (\epsilon_{t_1}, \epsilon_{t_2}, \dots, \epsilon_{t_n})^T . \end{aligned}$$

To generate  $m$  additional responses, we sample:

1.  $\{\mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+m}\}$  independently from the distribution  $\mathcal{N}(\mathbf{0}_d, \mathbf{C})$
2.  $(\epsilon_{t_{n+1}}, \dots, \epsilon_{t_{n+m}})^T \sim \mathcal{N}(\mathbf{0}_m, (1/\lambda_0)(\mathbf{R}_m(\phi_0) - \mathbf{R}_{m,n}(\phi_0)\mathbf{R}_n^{-1}(\phi_0)\mathbf{R}_{m,n}^T(\phi_0)))$  ,

where the  $(i, j)$  element of  $\mathbf{R}_m(\phi_0)$  is given by  $R(t_{n+i}, t_{n+j}|\phi_0)$ , and the  $(i, j)$  element of  $\mathbf{R}_{m,n}(\phi_0)$  is given by  $R(t_{n+i}, t_j|\phi_0)$ . Then  $y_{t_{n+i}} = \mathbf{f}^T(\mathbf{x}_{n+i})\beta_0 + \epsilon_{t_{n+i}}$  and the matrix-vector form of the augmented data set follows. Note that this process of conditionally sampling errors preserves the correct joint distribution of the augmented data vector.

In the following calculations, we assume the regression matrix  $\mathbf{X}_n$  is fixed, so this will not be explicitly denoted in the notation. Although  $(\beta, \lambda, \phi)$  are fixed at  $(\beta_0, \lambda_0, \phi_0)$  to generate data as above, our statistical analyses will assume some or all of  $(\beta_0, \lambda_0, \phi_0)$  are unknown. The sampling distribution (likelihood function)  $\pi(\mathbf{y}_n|\beta, \lambda, \phi)$  based on a data set of size  $n$  is  $\mathcal{N}(\mathbf{X}_n\beta, (1/\lambda)\mathbf{R}_n(\phi))$ .

(I.) The first analysis assumes that  $\lambda$  and  $\phi$  are fixed at  $\lambda_0$  and  $\phi_0$ , respectively. The Bayesian analysis places a prior distribution on  $\beta$ , which is given by the following:

1.  $\pi(\beta) = \int_{\mathcal{R}^p} \pi(\beta|\mu) \pi(\mu) d\mu$ ,
2.  $\pi(\beta|\mu)$  is  $\mathcal{N}(\mu, \lambda_0^{-1}\Sigma^{-1})$ , and
3.  $\pi(\mu)$  is  $\mathcal{N}(\mu_0, \lambda_0^{-1}\Sigma_0^{-1})$ .

These assumptions imply that the marginal prior  $\pi(\beta)$  is  $\mathcal{N}(\mu_0, \lambda_0^{-1}(\Sigma_0^{-1} + \Sigma^{-1}))$ . The posterior distribution of  $\mu$  is  $\mathcal{N}(\mu_1, \lambda_0^{-1}\Sigma_1)$ , where

$$\Sigma_1^{-1} = \Sigma_0^{-1} + \left( \Sigma^{-1} + (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n)^{-1} \right)^{-1} \text{ and}$$

$$\mu_1 = \Sigma_1 \left[ \left( \Sigma^{-1} + (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n)^{-1} \right)^{-1} \hat{\beta} + \Sigma_0 \mu_0 \right]$$

for  $\hat{\beta} = (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n)^{-1} \mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{y}_n$ . The posterior distribution of  $\beta$  is  $\mathcal{N}(\mu_2, \lambda_0^{-1}\Sigma_2)$ , where

$$\Sigma_2^{-1} = (\Sigma_0^{-1} + \Sigma^{-1})^{-1} + \mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n \text{ and}$$

$$\mu_2 = \Sigma_2 \left[ (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n) \hat{\beta} + (\Sigma_0^{-1} + \Sigma^{-1})^{-1} \mu_0 \right].$$

A noninformative prior for  $\beta$ ,  $\pi(\beta) \propto 1$ , results in an improper posterior distribution of  $\mu$ . Note that assuming a noninformative prior for  $\mu$ ,  $\pi(\mu) \propto 1$ , implies  $\pi(\beta) \propto 1$ . The posterior distribution of  $\beta$  is as above, with

$$\mu_2 = \hat{\beta} \text{ and } \Sigma_2 = (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n)^{-1}. \quad (1)$$

The predictive distribution of  $q$  future responses  $\tilde{\mathbf{y}}_q = (y_{\tilde{t}_1}, y_{\tilde{t}_2}, \dots, y_{\tilde{t}_q})^T$  associated with regression matrix  $\tilde{\mathbf{X}}_q = [\mathbf{f}(\tilde{\mathbf{x}}_1) \mathbf{f}(\tilde{\mathbf{x}}_2) \dots \mathbf{f}(\tilde{\mathbf{x}}_q)]^T$  and the  $q$ -variate error vector  $\tilde{\epsilon}_q = (\epsilon_{\tilde{t}_1}, \epsilon_{\tilde{t}_2}, \dots, \epsilon_{\tilde{t}_q})^T$  is  $\mathcal{N}(\tilde{\mu}_q, \lambda_0^{-1}\tilde{\Sigma}_q)$ , where

$$\tilde{\mu}_q = \tilde{\mathbf{X}}_q \mu_2 + \tilde{\mathbf{R}}_{q,n}(\phi_0) \mathbf{R}_n^{-1}(\phi_0) (\mathbf{y}_n - \mathbf{X}_n \mu_2) \text{ and}$$

$$\tilde{\Sigma}_q = \tilde{\mathbf{R}}_q(\phi_0) - \tilde{\mathbf{R}}_{q,n}(\phi_0) \mathbf{R}_n^{-1}(\phi_0) \tilde{\mathbf{R}}_{q,n}^T(\phi_0) + \tilde{\mathbf{H}}_q \Sigma_2 \tilde{\mathbf{H}}_q^T$$

for  $\tilde{\mathbf{H}}_q = \tilde{\mathbf{X}}_q - \tilde{\mathbf{R}}_{q,n}(\phi_0)\mathbf{R}_n^{-1}(\phi_0)\mathbf{X}_n$ . Here, the  $(i, j)$  element of  $\tilde{\mathbf{R}}_q(\phi_0)$  is given by  $R(\tilde{t}_i, \tilde{t}_j|\phi_0)$ , and the  $(i, j)$  element of  $\tilde{\mathbf{R}}_{q,n}(\phi_0)$  is given by  $R(\tilde{t}_i, t_j|\phi_0)$ .

(II.) The second analysis assumes that  $\phi$  is fixed at  $\phi_0$ . The Bayesian analysis places a prior distribution on  $(\beta, \lambda)$ , which is given by the following:

1.  $\pi(\beta|\lambda) = \int_{\mathcal{R}^p} \pi(\beta|\mu, \lambda) \pi(\mu|\lambda) d\mu$ ,
2.  $\pi(\beta|\mu, \lambda)$  is  $\mathcal{N}(\mu, \lambda^{-1}\mathbf{\Sigma}^{-1})$ ,
3.  $\pi(\mu|\lambda)$  is  $\mathcal{N}(\mu_0, \lambda^{-1}\mathbf{\Sigma}_0^{-1})$ , and
4.  $\pi(\lambda)$  is Gamma( $a, b$ ).

These assumptions imply that the marginal prior  $\pi(\beta|\lambda)$  is  $\mathcal{N}(\mu_0, \lambda^{-1}(\mathbf{\Sigma}_0^{-1} + \mathbf{\Sigma}^{-1}))$ . The posterior distribution of  $\lambda$  is Gamma( $a_1, b_1$ ), where

$$a_1 = (2a + n)/2$$

$$b_1 = \left( 2b + (\mathbf{y}_n - \mathbf{X}_n\hat{\beta})^T \mathbf{R}_n^{-1}(\phi_0)(\mathbf{y}_n - \mathbf{X}_n\hat{\beta}) + (\hat{\beta} - \mu_0)^T \mathbf{\Sigma}_3^{-1}(\hat{\beta} - \mu_0) \right) / 2$$

for  $\mathbf{\Sigma}_3 = \mathbf{\Sigma}_0^{-1} + \mathbf{\Sigma}^{-1} + (\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi_0) \mathbf{X}_n)^{-1}$ .

A noninformative prior for  $\beta$ ,  $\pi(\beta) \propto 1$ , results in the posterior distribution of  $\lambda$  given above, with

$$a_1 = (2a + n - p)/2 \text{ and } b_1 = \left( 2b + (\mathbf{y}_n - \mathbf{X}_n\hat{\beta})^T \mathbf{R}_n^{-1}(\phi_0)(\mathbf{y}_n - \mathbf{X}_n\hat{\beta}) \right) / 2. \quad (2)$$

A noninformative prior for  $\lambda$ ,  $\pi(\lambda) \propto (1/\lambda)$ , results from taking  $a = b = 0$ . Note that  $\pi(\beta, \lambda) \propto (1/\lambda)$  is the Jeffreys noninformative prior.

We denote the  $d$ -variate  $t$  distribution having  $\nu$  degrees of freedom, location vector  $\mu$ , and scale matrix  $\mathbf{\Sigma}$  by  $\mathcal{T}_d(\nu, \mu, \mathbf{\Sigma})$ . The mean of this distribution is  $\mu$  if  $\nu > 1$  and the covariance matrix of this distribution is  $\nu\mathbf{\Sigma}/(\nu - 2)$  if  $\nu > 2$ . The posterior distributions of  $\mu$  and  $\beta$  are given by

$$\pi(\mu|\mathbf{y}_n) \text{ is } \mathcal{T}_p(2a_1, \mu_1, b_1\mathbf{\Sigma}_1/a_1)$$

$$\pi(\beta|\mathbf{y}_n) \text{ is } \mathcal{T}_p(2a_1, \mu_2, b_1\mathbf{\Sigma}_2/a_1) .$$

For the noninformative prior  $\pi(\beta) \propto 1$ , as before the posterior distribution of  $\mu$  is improper. The quantities  $\mu_2, \mathbf{\Sigma}_2$  from (1) and  $a_1, b_1$  from (2) are utilized in the above expression for the posterior distribution of  $\beta$ , where again  $a = b = 0$  if  $\pi(\lambda) \propto (1/\lambda)$ .

The predictive distribution of  $\tilde{\mathbf{y}}_q$  is given by

$$\pi(\tilde{\mathbf{y}}_q|\mathbf{y}_n) \text{ is } \mathcal{T}_q \left( 2a_1, \tilde{\mu}_q, b_1\tilde{\mathbf{\Sigma}}_q/a_1 \right) .$$

(III.) The third analysis allows  $\phi$  to be random. The Bayesian analysis places a prior distribution on  $(\beta, \lambda, \phi)$ , which is given by  $\pi(\beta, \lambda, \phi) = \pi(\beta, \lambda)\pi(\phi)$  where  $\pi(\beta, \lambda)$  is specified in the previous analysis. We will not presently specify  $\pi(\phi)$ . Our goal in this analysis is to numerically approximate the marginal posterior distributions of  $\lambda$ ,  $\mu$ , and  $\beta$ :

$$\begin{aligned}\pi(\lambda|\mathbf{y}_n) &= \int_{\Phi} \pi(\lambda|\mathbf{y}_n, \phi) \pi(\phi|\mathbf{y}_n) d\phi \\ \pi(\mu|\mathbf{y}_n) &= \int_{\Phi} \pi(\mu|\mathbf{y}_n, \phi) \pi(\phi|\mathbf{y}_n) d\phi \\ \pi(\beta|\mathbf{y}_n) &= \int_{\Phi} \pi(\beta|\mathbf{y}_n, \phi) \pi(\phi|\mathbf{y}_n) d\phi.\end{aligned}$$

The distributions  $\pi(\lambda|\mathbf{y}_n, \phi)$ ,  $\pi(\mu|\mathbf{y}_n, \phi)$ , and  $\pi(\beta|\mathbf{y}_n, \phi)$  are given analytically in the previous analysis. That leaves  $\pi(\phi|\mathbf{y}_n)$ , which is given as follows:

$$\pi(\phi|\mathbf{y}_n) \propto \frac{\pi(\phi)}{b_1^{a_1} \det(\mathbf{R}_n(\phi))^{1/2} \det(\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi) \mathbf{X}_n)^{1/2} \det(\boldsymbol{\Sigma}_3)^{1/2}}.$$

Quadrature is used to compute the normalizing constant

$$c(\mathbf{y}_n) = \int_{\Phi} \frac{\pi(\phi) d\phi}{b_1^{a_1} \det(\mathbf{R}_n(\phi))^{1/2} \det(\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi) \mathbf{X}_n)^{1/2} \det(\boldsymbol{\Sigma}_3)^{1/2}}$$

so that

$$\pi(\phi|\mathbf{y}_n) = \frac{c(\mathbf{y}_n)^{-1} \pi(\phi)}{b_1^{a_1} \det(\mathbf{R}_n(\phi))^{1/2} \det(\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi) \mathbf{X}_n)^{1/2} \det(\boldsymbol{\Sigma}_3)^{1/2}}.$$

For the noninformative prior  $\pi(\beta) \propto 1$ ,

$$\pi(\phi|\mathbf{y}_n) = \frac{c(\mathbf{y}_n)^{-1} \pi(\phi)}{b_1^{a_1} \det(\mathbf{R}_n(\phi))^{1/2} \det(\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi) \mathbf{X}_n)^{1/2}}$$

for

$$c(\mathbf{y}_n) = \int_{\Phi} \frac{\pi(\phi) d\phi}{b_1^{a_1} \det(\mathbf{R}_n(\phi))^{1/2} \det(\mathbf{X}_n^T \mathbf{R}_n^{-1}(\phi) \mathbf{X}_n)^{1/2}},$$

where  $a_1, b_1$  are taken from (2) and  $a = b = 0$  if  $\pi(\lambda) \propto (1/\lambda)$ .

The predictive distribution of  $\tilde{\mathbf{y}}_q$  can also be numerically approximated:

$$\pi(\tilde{\mathbf{y}}_q|\mathbf{y}_n) = \int_{\Phi} \pi(\tilde{\mathbf{y}}_q|\mathbf{y}_n, \phi) \pi(\phi|\mathbf{y}_n) d\phi.$$

## 2 Example Preliminaries

A simple linear regression model,  $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2, \dots, x_d)^T$ , is specified for all verification examples. The dimension of  $\beta$  is therefore  $p = d + 1$ . A wide range of covariate dimensions are considered,  $d \in \{2, 10, 50, 100, 500, 1000\}$ .

The verification examples proceed by considering successively more complex specifications of the correlation function  $R(t_1, t_2 | \phi)$ :

1.  $R(t_i, t_j) = \delta_{t_i, t_j}$
2.  $R(t_i, t_j | \phi) = \delta_{t_i, t_j} + \phi(1 - \delta_{t_i, t_j})$ ,  $\phi \in [0, 1)$
3.  $R(t_i, t_j | \phi) = \phi^{|i-j|}$ ,  $-1 < \phi < 1$
4.  $R(t_i, t_j | \phi) = \exp \left[ -(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{D}_\phi (\mathbf{x}_i - \mathbf{x}_j) \right]$ , where  $\mathbf{D}_\phi = \text{diag}(\phi_1, \dots, \phi_d)$  and  $\phi_i \geq 0$  for  $i = 1, \dots, d$ .

The first case describes the standard regression setting in which errors are uncorrelated. The second case describes eqicorrelated errors, for which all pairwise correlations are equal. The third case describes an AR(1) correlation structure for errors associated with observations indexed by time. The pairwise correlation between errors decays as a function of separation in time. The final case specifies a Gaussian correlation for errors associated with observations indexed by covariates. The pairwise correlation between errors decays as a function of weighted distance between the covariate vectors associated with the corresponding observations.

The first three correlation functions admit explicit expressions for the inverse  $\mathbf{R}_n^{-1}(\phi)$  and determinant  $\det(\mathbf{R}_n(\phi))$ :

1.  $\mathbf{I}_n$  and 1
2.  $\frac{1}{1-\phi} \left[ \mathbf{I}_n - \frac{\phi}{1+(n-1)\phi} \mathbf{J}_n \right]$  and  $(1-\phi)^{n-1}(1+(n-1)\phi)$
3.  $\frac{1}{1-\phi^2} \begin{pmatrix} 1 & -\phi & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & \dots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & 0 & \dots & -\phi & 1 \end{pmatrix}$  and  $(1-\phi^2)^{n-1}$

For data sample sizes, we consider  $\{1, 2, 3, 10, 100, 1000, 10^4, 10^5, 10^6\}$  for the first three correlation functions, and  $\{1, 2, 3, 10, 100, 1000\}$  for the fourth correlation function (sample size limited by inverse and determinant calculations). Small sample sizes  $\{1, 2, 3\}$  will allow MCMC sampling from heavy-tailed posterior distributions to be compared with known

analytical forms (specifically, the multivariate  $t$  distributions of the second and third analyses). Large sample sizes allow verification of regression parameter and error variance point estimate convergence ( $\hat{\beta} \rightarrow \beta_0$  and  $b_1/a_1 \rightarrow \lambda_0^{-1}$ ) for the first correlation function. Each case will be structured so that the number of random parameters does not exceed the sample size. For example, if  $n = 1$ , only the first analysis is conducted with a single random regression parameter  $\beta_0$ .

Two prior specifications will be considered: (a)  $\pi(\mu, \beta, \lambda) \propto (1/\lambda)$  (noninformative) and (b)  $\Sigma_0 = s \mathbf{I}_p$  for  $0 < s \ll 1$ , and  $\Sigma = \text{diag}(\{t_i, i = 1, \dots, p\})$  where the  $t_i > 0$  are fixed *a priori* (proper). Both cases take  $a = b = 0$ . For the first correlation function, case (b) allows verification of the convergence of the posterior distribution of  $\mu$  to the Gaussian distribution having mean vector and covariance matrix

$$\mu_1 = \mathbf{D}_\beta \beta_0 + \mathbf{D}_\mu \mu_0 \text{ and } \lambda_0^{-1} \Sigma_1 = \text{diag}(\{\lambda_0^{-1}/(s + t_i), i = 1, \dots, p\})$$

where  $\mathbf{D}_\beta = \text{diag}(\{t_i/(s + t_i), i = 1, \dots, p\})$  and  $\mathbf{D}_\mu = \text{diag}(\{s/(s + t_i), i = 1, \dots, p\})$ .