Bernoulli **12**(1), 2006, 157–168

Multivariate prediction

JOSÉ MANUEL CORCUERA[1] and FEDERICA GIUMMOLÈ[2]

[1]*Facultat de Matemàtiques, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain. E-mail: jmcorcuera@ub.edu*
[2]*Dipartimento di Statistica, Università Ca' Foscari di Venezia, San Giobbe, Cannaregio 873, 30121 Venezia, Italy. E-mail: giummole@unive.it*

The problem of prediction is considered in a multidimensional setting. Extending an idea presented by Barndorff-Nielsen and Cox, a predictive density for a multivariate random variable of interest is proposed. This density has the form of an estimative density plus a correction term. It gives simultaneous prediction regions with coverage error of smaller asymptotic order than the estimative density. A simulation study is also presented showing the magnitude of the improvement with respect to the estimative method.

Keywords: ancillary statistic; coverage probability; estimative density; prediction regions; predictive density

1. Introduction

We consider the problem of prediction of an unobservable m-dimensional absolutely continuous random vector $Z = (Z_1, \ldots, Z_m)$, on the basis of an observed sample $y = (y_1, \ldots, y_n)$ from a further n-dimensional random vector $Y = (Y_1, \ldots, Y_n)$. We assume that the density of Y, $p(y; \theta)$, and the conditional density of Z given $Y = y$, $g(z; \theta|y)$, are known except for the d-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^d$. A prediction statement about Z is often given in terms of prediction regions, i.e. regions $R_\alpha(Y) \subset \mathbb{R}^m$ such that

$$P_\theta\{Z \in R_\alpha(Y)\} = \alpha,$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0, 1)$. The above probability is usually called the coverage probability and is calculated with respect to the joint density of Z and Y.

An easy way of making predictions about Z is by means of the so-called estimative predictive density $g(z; \tilde{\theta}|y)$, where $\tilde{\theta}$ is any asymptotically efficient estimator of θ, usually the maximum likelihood estimator (MLE) $\hat{\theta}$. However, prediction regions based on the estimative density are usually imprecise, having coverage error of order $O(n^{-1})$. This is a well-known result in the case of Z being a unidimensional random variable. Indeed, Barndorff-Nielsen and Cox (1996) and Vidoni (1998) suggest a way to correct the quantiles of the estimative density, thus obtaining prediction limits with a coverage error of order $o(n^{-1})$. Unfortunately, their result does not apply to the multidimensional case since it does not take into account the possible dependence between the components of Z nor the interaction among the prediction limits of each component of Z. Thus, prediction regions based on the corrected prediction limits of each component of Z separately still give

1350–7265

coverage error of order $O(n^{-1})$. A resampling approach to the multivariate problem is discussed in Hall *et al.* (1999), using bootstrap calibration of estimative prediction regions. However, no analytical solution is provided by this method.

Here we extend the idea in Barndorff-Nielsen and Cox (1996) to the case when Z is an absolutely continuous m-dimensional random vector.

2. Improved prediction regions

We assume for simplicity that Z is independent of Y, so that its conditional density can be written as

$$g(z;\,\theta) = g^1(z_1;\,\theta)\prod_{i=2}^{m} g^i(z_i;\,\theta|z^{(i-1)}),$$

where $z^{(i)}$ is the value of $Z^{(i)} = (Z_1, \ldots, Z_i)$, the first i components of Z, $g^i(z_i;\,\theta|z^{(i-1)})$ is the conditional density of Z_i given $Z^{(i-1)} = z^{(i-1)}$, $i = 2, \ldots, m$, and $g^1(z_1;\,\theta)$ is the marginal density of Z_1. We denote by $G^i(z_i;\,\theta|z^{(i-1)})$ the conditional cumulative distribution function of Z_i given $Z^{(i-1)} = z^{(i-1)}$, $i = 2, \ldots, m$, and by $G^1(z_1;\,\theta)$ the cumulative distribution function of Z_1. We also assume that g^i and G^i, $i = 1, \ldots, m$, are sufficiently smooth functions of the parameter θ.

Let us suppose that a previous reduction of the data is possible and that $(\hat{\theta}, A)$ is a minimal sufficient statistic, with $\hat{\theta}$ the MLE and A an ancillary statistic. In this context we can apply the conditionality principle and keep the value of A fixed in the evaluation of the long-run properties of the results. Thus, in what follows, any function of the data will be written as a function of the MLE, avoiding writing the dependency on ar A explicitly.

The aim of this work is to find regions $R_\alpha(\hat{\theta}) \subset \mathbb{R}^m$ such that

$$P_\theta\{Z \in R_\alpha(\hat{\theta})\} \doteq \alpha,$$

for every $\theta \in \Theta$ and for any fixed $\alpha \in (0, 1)$, where $\doteq$ means that an equality holds with error terms of order $o(n^{-1})$ or $o_p(n^{-1})$, depending on the context. The above coverage probability is calculated with respect to the joint density of Z and $\hat{\theta}$ conditioned on the observed value of the ancillary statistic A.

We assume that $R_\alpha(\hat{\theta})$ can be expressed through the system of inequalities

$$a^1(\hat{\theta}) \leqslant z_1 \leqslant b^1(\hat{\theta}),$$

$$a^2(z^{(1)};\,\hat{\theta}) \leqslant z_2 \leqslant b^2(z^{(1)};\,\hat{\theta}),$$

$$\ldots$$

$$a^m(z^{(m-1)};\,\hat{\theta}) \leqslant z_m \leqslant b^m(z^{(m-1)};\,\hat{\theta}),$$

so that the m-dimensional integral of $g(z;\,\theta)$ over $R_\alpha(\hat{\theta})$ reduces to m successive integrals in $\mathbb{R}$. Thus, our problem becomes that of finding functions $c^1(\hat{\theta}), c^2(z^{(1)};\,\hat{\theta}), \ldots, c^m(z^{(m-1)};\,\hat{\theta})$ such that

$$P_\theta\{Z_1 \leqslant c^1(\hat\theta),\ Z_2 \leqslant c^2(Z^{(1)};\ \hat\theta),\ \ldots,\ Z_m \leqslant c^m(Z^{(m-1)};\ \hat\theta)\} \doteq \alpha,$$

for every $\theta \in \Theta$. We call a set of such functions a system of approximate simultaneous prediction limits for Z.

In the following we use the Einstein convention so that whenever an index appears twice in an expression, summation on that index is intended.

We have the following result:

Proposition 1. *Let $q^1(\theta)$, $q^i(z^{(i-1)};\ \theta)$, $i = 2, \ldots, m$, be a system of simultaneous prediction limits for Z, such that*

$$P_\theta\{Z_1 \leqslant q^1(\theta),\ Z_2 \leqslant q^2(Z^{(1)};\ \theta),\ \ldots,\ Z_m \leqslant q^m(Z^{(m-1)};\ \theta)\} = \alpha.$$

Then, under the usual asymptotic behaviour of the MLE. $\hat\theta$,

$$\begin{aligned}
&P_\theta\{Z_1 \leqslant q^1(\hat\theta),\ Z_2 \leqslant q^2(Z^{(1)};\ \hat\theta),\ \ldots,\ Z_m \leqslant q^m(Z^{(m-1)};\ \hat\theta)\} \\
&\quad \doteq \alpha - \sum_{i=1}^m \mathbf{b}^r \int^{q^1} \ldots \int^{q^i} g^i_r \,\mathrm{d}z_i \ldots \int^{q^m} \\
&\quad - \frac{1}{2}\sum_{i=1}^m \mathbf{i}^{rs} \left([2] \int^{q^1} \ldots \int^{q^{i-1}} g^i_r(q^i;\ \theta|z^{(i-1)}) q^i_s \int^{q^{i+1}_{(i)}} \ldots \int^{q^m_{(i)}} \right. \\
&\quad \left. + \int^{q^1} \ldots \int^{q^i} g^i_{rs} \mathrm{d}z_i \ldots \int^{q^m} \right) \\
&\quad - \frac{1}{2}\sum_{i=2}^m \sum_{j<i} [2]\mathbf{i}^{rs} \left(\int^{q^1} \ldots \int^{q^{j-1}} g^j(q^j;\ \theta|z^{(j-1)}) q^j_s \int^{q^{j+1}_{(j)}} \ldots \int^{q^i_{(j)}} g^i_{r(j)} \,\mathrm{d}z_i \ldots \int^{q^m_{(j)}} \right. \\
&\quad + \int^{q^1} \ldots \int^{q^j} g^j_r \,\mathrm{d}z_j \ldots \int^{q^{i-1}} g^i(q^i;\ \theta|z^{(i-1)}) q^i_s \int^{q^{i+1}_{(i)}} \ldots \int^{q^m_{(i)}} \\
&\quad \left. + \int^{q^1} \ldots \int^{q^j} g^j_s \,\mathrm{d}z_j \ldots \int^{q^i} g^i_r \,\mathrm{d}z_i \ldots \int^{q^m} \right),
\end{aligned}$$

where $\int^{q^i} \equiv \int^{q^i} g^i \,\mathrm{d}z_i$, $z^{(i-1)}_{(j)} = (z^{(j-1)}, q^j, z_{j+1}, \ldots, z_{i-1})$, $q^i_{(j)} = q^i(z^{(i-1)}_{(j)};\ \theta)$, $\int^{q^i_{(j)}} \equiv \int^{q^i_{(j)}} g^i(z_i;\ \theta|z^{(i-1)}_{(j)})\mathrm{d}z_i$, $g^i_{(j)} = g^i(z_i;\ \theta|z^{(i-1)}_{(j)})$, *subscript r and s denote derivatives with respect to the corresponding components of the parameter*, $[2]$ *indicates the sum of two terms obtained by permutation of the indices r and s, and $\mathbf{b}^r$ and $\mathbf{i}^{rs}$ are the asymptotic bias and variance–covariance matrix of $\hat\theta$:*

$$\mathbf{b}^r(\theta) \doteq \mathrm{E}_\theta\{(\hat\theta - \theta)^r\}$$

and

$$\mathbf{i}^{rs}(\theta) \doteq \mathrm{E}_\theta\{(\hat\theta - \theta)^r(\hat\theta - \theta)^s\}.$$

Proof. Let $R_\alpha(\hat{\theta})$ be the region in $\mathbb{R}^m$ given by $z_1 \leqslant q^1(\hat{\theta})$, $z_2 \leqslant q^2(z^{(1)}; \hat{\theta}), \ldots,$ $z_m \leqslant q^m(z^{(m-1)}; \hat{\theta})$, where $\int_{R_\alpha(\theta)} g(z; \theta)\mathrm{d}z = \alpha$. Then

$$\begin{aligned}
\int_{R_\alpha(\hat{\theta})} g(z; \theta)\mathrm{d}z &= \int_{R_\alpha(\hat{\theta})} g(z; \hat{\theta})\mathrm{d}z + \int_{R_\alpha(\hat{\theta})} (g(z; \theta) - g(z; \hat{\theta}))\mathrm{d}z \\
&\doteq \alpha - (\hat{\theta}-\theta)^r \int_{R_\alpha(\hat{\theta})} g_r(z; \theta)\mathrm{d}z - \frac{1}{2}(\hat{\theta}-\theta)^r(\hat{\theta}-\theta)^s \int_{R_\alpha(\hat{\theta})} g_{rs}(z; \theta)\mathrm{d}z \\
&\doteq \alpha - (\hat{\theta}-\theta)^r \int_{R_\alpha(\theta)} g_r(z; \theta)\mathrm{d}z - \frac{1}{2}(\hat{\theta}-\theta)^r(\hat{\theta}-\theta)^s \int_{R_\alpha(\theta)} g_{rs}(z; \theta)\mathrm{d}z \\
&\quad - (\hat{\theta}-\theta)^r \left(\int_{R_\alpha(\hat{\theta})} g_r(z; \theta)\mathrm{d}z - \int_{R_\alpha(\theta)} g_r(z; \theta)\mathrm{d}z \right).
\end{aligned}$$

Moreover,

$$\int_{R_\alpha(\theta)} g_r(z; \theta)\mathrm{d}z = \sum_{i=1}^m \int^{q^1} \ldots \int^{q^i} g_r^i \,\mathrm{d}z_i \ldots \int^{q^m},$$

$$\begin{aligned}
\int_{R_\alpha(\theta)} g_{rs}(z; \theta)\mathrm{d}z &= \sum_{i=1}^m \int^{q^1} \ldots \int^{q^i} g_r^i \,\mathrm{d}z_i \ldots \int^{q^m} \\
&\quad + \sum_{i=2}^m \sum_{j<i} [2] \int^{q^1} \ldots \int^{q^j} \ldots g_s^j \,\mathrm{d}z_j \ldots \int^{q^i} \ldots g_r^{ai} \,\mathrm{d}z_i \ldots \int^{q^m}
\end{aligned}$$

and

$$\begin{aligned}
&(\hat{\theta}-\theta)^r \left(\int_{R_\alpha(\hat{\theta})} g_r(z; \theta)\mathrm{d}z - \int_{R_\alpha(\theta)} g_r(z; \theta)\mathrm{d}z \right) \\
&= (\hat{\theta}-\theta)^r \sum_{i=1}^m \left(\int^{\hat{q}^1} \ldots \int^{\hat{q}^i} g_r^i \,\mathrm{d}z_i \ldots \int^{\hat{q}^m} - \int^{q^1} \ldots \int^{q^i} g_r^i \,\mathrm{d}z_i \ldots \int^{q^m} \right) \\
&\doteq (\hat{\theta}-\theta)^s(\hat{\theta}-\theta)^r \left(\sum_{i=1}^m \int^{q^1} \ldots \int^{q^{i-1}} g_r^i(q^i; \theta | z^{(i-1)}) q_s^i \int^{q^{i+1}_{(i)}} \ldots \int^{q^m_{(i)}} \right. \\
&\quad + \sum_{i=2}^m \sum_{j<i} \int^{q^1} \ldots \int^{q^{j-1}} g^j(q^j; \theta | z^{(j-1)}) q_s^j \int^{q^{j+1}_{(j)}} \ldots \int^{q^i_{(j)}} g^i_{r(j)} \,\mathrm{d}z_i \ldots \int^{q^m_{(j)}} \\
&\quad \left. + \sum_{i=1}^{m-1} \sum_{j>i} \int^{q^1} \ldots \int^{q^i} g_r^i \,\mathrm{d}z_i \ldots \int^{q^{j-1}} g^j(q^j; \theta | z^{(j-1)}) q_s^j \int^{q^{j+1}_{(j)}} \ldots \int^{q^m_{(j)}} \right).
\end{aligned}$$

Thus, after substituting and taking expectations, we obtain the result. □

Corollary 2. *Let q^i be the α_i-quantile of g^i, $i = 1, \ldots, m$, with $\alpha = \sqcap_{i=1}^m \alpha_i$. Then*

$$P_\theta\{Z_1 \leqslant q^1(\hat\theta),\, Z_2 \leqslant q^2(Z^{(1)}; \hat\theta), \ldots, Z_m \leqslant q^m(Z^{m-1)}; \hat\theta)\}$$

$$\doteq \alpha - \sum_{i=1}^m \mathbf{b}^r \int^{q^1} \ldots \int^{q^{i-1}} G_r^i(q^i; \theta|z^{(i-1)}) \frac{\alpha}{\alpha_1 \ldots \alpha_i}$$

$$+ \frac{1}{2}\sum_{i=1}^m \mathbf{i}^{rs} \int^{q^1} \ldots \int^{q^{i-1}} \left([2]\frac{g_s^i(q^i; \theta|z^{i-1)})G_r^i(q^i; \theta|z^{(i-1)})}{g^i(q^i; \theta|z^{(i-1)}} - G_{rs}^i(q^i; \theta|z^{(i-1)}) \right) \frac{\alpha}{\alpha_1 \ldots \alpha_i}$$

$$+ \frac{1}{2}\sum_{i=2}^m \sum_{j<i} \mathbf{i}^{rs}[2] \int^{q^1} \ldots \int^{q^{j-1}} G_s^j(q^j; \theta|z^{(j-1)}) \int^{q^{j+1}_{(j)}} \ldots \int^{q^{i\,1}_{(j)}} G_r^i(q^i; \theta|z^{(i-1)}_{(j)}) \frac{\alpha}{\alpha_1 \ldots \alpha_i}.$$

Proof. By definition we have that

$$G^i(q^i; \theta|z^{(i-1)}) = \alpha_i.$$

Since q^i is a function of both $z^{(i-1)}$ and θ, by taking the total derivative with respect to the components of θ, we have

$$G_r^i(q^i; \theta|z^{(i-1)}) + g^i(q^i; \theta|z^{(i-1)})q_r^i = 0,$$

where

$$G_r^i(q^i; \theta|z^{(i-1)}) = \partial_{\theta_r} G^i(z_i; \theta|z^{(i-1)})|_{z_i = q^i}.$$

Thus, by substituting $q_r^i = -G_r^i/g^i$ in the expression obtained in Proposition 1, the last two summands cancel and we have the result. □

Corollary 3. *Let q^i be the α_i-quantile of g^i, $i = 1, \ldots, m$, with $\alpha = \sqcap_{i=1}^m \alpha_i$. Then a system of approximate simultaneous prediction limits is given by*

$$c^1(\hat\theta) = q^1(\hat\theta) - \frac{h^1(q^1(\hat\theta); \hat\theta)}{g^1(q^1(\hat\theta); \hat\theta)}$$

and, for $i = 2, \ldots, m$,

$$c^i(z^{(i-1)}; \hat\theta) = q^i(\hat\theta) - \frac{h^i(q^i(\hat\theta), z^{(i-1)}; \hat\theta)}{g^i(q^i(\hat\theta); \hat\theta|z^{(i-1)})} - \sum_{j<i} \frac{h^{ij}(q^i(\hat\theta), z^{(i-1)}; \hat\theta)}{g^i(q^i(\hat\theta); \hat\theta|z^{(i-1)})}, \tag{1}$$

where the terms $h^i(z_i, z^{(i-1)}; \theta)$ and $h^{ij}(z_i, z^{(i-1)}; \theta)$ are of order $O_p(n^{-1})$ and, for $i = 1, \ldots, m$ and $j < i$, are given by

$$h^i(z_i, z^{(i-1)}; \theta)$$

$$= -\mathbf{b}^r(\theta)G_r^i(z_i; \theta|z^{(i-1)}) + \frac{1}{2}\mathbf{i}^{rs}(\theta)\left([2]\frac{g_s^i(z_i; \theta|z^{(i-1)})}{g^i(z_i; \theta|z^{(i-1)})} G_r^i(z_i; \theta|z^{(i-1)}) - G_{rs}^i(z_i; \theta|z^{(i-1)}) \right)$$

and

$$h^{ij}(z_i, z^{(i-1)}; \theta) = \frac{1}{2}\mathbf{i}^{rs}(\theta)[2]\frac{\partial_{z_j}\left(G_s^j(z_j; \theta|z^{(j-1)})g^{j+1}(z_{j+1}; \theta|z^{(j)}) \ldots g^{i-1}(z_{i-1}; \theta|z^{(i-2)})G_r^i(z_i; \theta|z^{(i-1)})\right)}{g^j(z_j; \theta|z^{(j-1)}) \ldots g^{i-1}(z_{i-1}; \theta|z^{(i-2)})}.$$

3. The predictive density

The previous result can be expressed in terms of a predictive density $\hat{g}(z; y)$ for Z, such that

$$\int_{-\infty}^{c^1(\hat{\theta})}\int_{-\infty}^{c^2(z^{(1)};\hat{\theta})}\cdots\int_{-\infty}^{c^m(z^{(m-1)};\hat{\theta})}\hat{g}(z; y)\mathrm{d}z_m\cdots\mathrm{d}z_2\mathrm{d}z_1 = \alpha.$$

In fact we can write

$$\hat{g}(z; y) \doteq \prod_{i=1}^m g^i\left(z_i + \frac{h^i(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})} + \sum_{j<i}\frac{h^{ij}(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})}; \hat{\theta}\right)$$
$$\times\left\{1 + \partial_{z_i}\frac{h^i(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})} + \sum_{j<i}\partial_{z_i}\frac{h^{ij}(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})}\right\}.$$

This is easily shown by a change of variable in

$$\int_{-\infty}^{\hat{q}^1}\int_{-\infty}^{\hat{q}^2}\cdots\int_{-\infty}^{\hat{q}^m} g(w; \hat{\theta})\mathrm{d}w_m\cdots\mathrm{d}w_2\mathrm{d}w_1 = \alpha,$$

putting

$$w_i = z_i + \frac{h^i(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})} + \sum_{j<i}\frac{h^{ij}(z_i, z^{(i-1)}; \hat{\theta})}{g^i(z_i; \hat{\theta}|z^{(i-1)})}.$$

After some calculations, we obtain

$$\hat{g}(z; y) \doteq g(z; \hat{\theta})\left\{1 + \sum_{i=1}^m k^i(z_i, z^{(i-1)}; \hat{\theta}) + \sum_{i=1}^m\sum_{j<i} k^{ij}(z_i, z^{(i-1)}; \hat{\theta})\right\}, \tag{2}$$

where

$$k^i(z_i, z^{(i-1)}; \theta) = \frac{\partial_{z_i} h^i(z_i, z^{(i-1)}; \theta)}{g^i(z_i, z^{(i-1)}; \theta)}$$

and

$$k^{ij}(z_i, z^{(i-1)}; \theta) = \frac{\partial_{z_i} h^{ij}(z_i, z^{(i-1)}; \theta)}{g^i(z_i, z^{(i-1)}; \theta)}.$$

It is important to notice that the terms k^i in (2) are exactly those needed in the unidimensional case, considering the conditional densities. They correct the uncertainty introduced when θ is replaced by $\hat{\theta}$ in the estimative density. The terms k^{ij} correct for the additional dependency introduced among the components of Z, after estimating θ in each conditional density g^i with the same data.

Moreover, we have that

$$\sum_{i=1}^{m} k^i(z_i, z^{(i-1)}; \theta) = -\mathbf{b}^r(\theta)\partial_r \log g(z; \theta) - \frac{1}{2}\mathbf{i}^{rs}(\theta)\partial_{rs} \log g(z; \theta)$$
$$+\frac{1}{2}\mathbf{i}^{rs}(\theta)\sum_{i=1}^{m}\Bigg([2]\partial_s\partial_{z_i} \log g^i(z_i; \theta|z^{(i-1)})\frac{G_r^i(z_i; \theta|z^{(i-1)})}{g^i(z_i; \theta|z^{(i-1)})}$$
$$+\partial_s \log g^i(z_i; \theta|z^{(i-1)})\partial_r \log g^i(z_i; \theta|z^{(i-1)})\Bigg)$$

and

$$k^{ij}(z_i, z^{(i-1)}; \theta) = \frac{1}{2}\mathbf{i}^{rs}(\theta)[2]\Bigg\{\partial_s \log g^j(z_j; \theta|z^{(j-1)})\partial_r \log g^i(z_i; \theta|z^{(i-1)}).$$
$$+\frac{G_s^j(z_j; \theta|z^{(j-1)})}{g^j(z_j; \theta|z^{(j-1)})}\partial_r \log g^i(z_i; \theta|z^{(i-1)})\big(\partial_{z_j} \log g^{j+1}(z_{j+1}; \theta|z^{(j)})$$
$$+\ldots+\partial_{z_j} \log g^{i-1}(z_{i-1}; \theta|z^{(i-2)})\big)$$
$$+\frac{G_s^j(z_j; \theta|z^{(j-1)})}{g^j(z_j; \theta|z^{(j-1)})}\frac{\partial_{z_j} g_r^i(z_i; \theta|z^{(i-1)})}{g^i(z_i; \theta|z^{(i-1)})}\Bigg\}.$$

Thus, we can finally write

$$\hat{g}(z; y) \doteq g(z; \hat{\theta})\Bigg\{1 - \mathbf{b}^r(\hat{\theta})\partial_r \log g + \frac{1}{2}\mathbf{i}^{rs}(\hat{\theta})\Bigg(\partial_r \log g\partial_s \log g - \partial_{rs} \log g.. \tag{3}$$
$$+\sum_{i=1}^{m}[2]\partial_s\partial_{z_i} \log g^i(z_i; \hat{\theta}|z^{(i-1)})\frac{G_r^i(z_i; \hat{\theta}|z^{(i-1)})}{g^i(z_i; \hat{\theta}|z^{(i-1)})}$$
$$+\sum_{i=1}^{m}\sum_{j<i}[2]\frac{G_s^j(z_j; \hat{\theta}|z^{(j-1)})}{g^j(z_j; \hat{\theta}|z^{(j-1)})}\Bigg(\partial_r \log g^i(z_i; \hat{\theta}|z^{(i-1)})\partial_{z_j} log(g^{j+1}\cdots g^{i-1}).$$
$$+\frac{\partial_{z_j} g_r^i(z_i; \theta|z^{(i-1)})}{g^i(z_i; \theta|z^{(i-1)})}\Bigg)\Bigg)\Bigg\}.$$

Remark. Note that $\hat{g}(z; y)$ depends on the factorization we use for the joint density $g(z; \theta) = \sqcap_{i=1}^{m} g^{\sigma(i)}(z_{\sigma(i)}; \theta|z^{(\sigma(i)-1)})$, where σ is a permutation of m elements. If the components of Z are exchangeable, $\hat{g}_\sigma(z; y) = \hat{g}(z_{\sigma(1)}, \ldots, z_{\sigma(m)}; y)$, but in general $\hat{g}_\sigma(z; y) \neq \hat{g}_{\sigma'}(z; y)$ if $\sigma \neq \sigma'$, as we can see in the following example.

Example 1. Suppose that

$$Z = (Z_1, Z_2) \sim N_2\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

If we take

$$g^1(z_1) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{1}{2}z_1^2\right) \quad \text{and} \quad g^2(z_2; \rho|z_1) = \frac{1}{\sqrt{(2\pi(1-\rho^2))}} \exp(-\frac{(z_2 - \rho z_1)^2}{2(1-\rho^2)}),$$

there is a non-symmetric term in expression (3), given by

$$\partial_\rho \partial_{z_2} \log g^2(z_2; \hat{\rho}|z_1) \frac{G_\rho^2(z_2; \hat{\rho}|z_1)}{g^2(z_2; \hat{\rho}|z_1)} = \frac{(\hat{\rho} z_2 - z_1)((1+\hat{\rho}^2)z_1 - 2\hat{\rho} z_2)}{(1-\hat{\rho}^2)^3}.$$

This shows that, in this case, $\hat{g}(z_1, z_2; y) \neq \hat{g}(z_2, z_1; y)$.

Moreover, it is easy to see that the marginal densities of $\hat{g}(z_1, z_2; y)$ are $\hat{g}^1(z_1; y) = g^1(z_1)$ and $\hat{g}^2(z_2; y) \doteq g^2(z_2)(1 + \frac{1}{2}\mathbf{i}^{\rho\rho}(\hat{\rho})(z_2^2 - 1)/(1-\hat{\rho}^2))$. This latter density does not give correct prediction limits for Z_2 since it differs from $g^2(z_2)$. This means that, for some regions in $\mathbb{R}^2$, $\hat{g}(z_1, z_2; y)$ may give a coverage error of order $O(n^{-1})$.

The above situation should not be surprising because the predictive density (3) is constructed to give correct coverage probabilities for prediction limits (1). This limits are the quantiles of $\hat{g}^1(z_1; y)$ and $\hat{g}^i(z_i|z^{(i-1)}; y)$, $i = 2, \ldots, m$, for a prescribed order of the components of Z. Thus, to find a predictive region of approximate confidence $\alpha = \sqcap_{i=1}^m \alpha_i$, we have to find $c^1(\hat{\theta})$, $c^i(z^{(i-1)}; \hat{\theta})$, $i = 2, \ldots, m$, such that

$$\int^{c^1(\hat{\theta})} \hat{g}^1(z_1; y)\mathrm{d}z_1 = \alpha_1, \int^{c^i(z^{i-1}; \hat{\theta})} \hat{g}(z_i|z^{(i-1)}; y)\mathrm{d}z_i = \alpha_i, \qquad i = 2, \ldots, m.$$

Then,

$$\begin{aligned} &z_1 \leqslant c^1(\hat{\theta}), \\ &z_2 \leqslant c^2(z^{(1)}; \hat{\theta}), \\ &\ldots, \\ &z_m \leqslant c^m(z^{(m-1)}; \hat{\theta}) \end{aligned}$$

define a predictive region of approximate confidence α, in the sense that $P_\theta\{Z_1 \leqslant c^1(\theta), Z_2 \leqslant c^2(Z^{(1)}; \theta), \ldots, Z_m \leqslant c^m(Z^{(m-1)}; \theta)\} \doteq \alpha$.

We can look at the joint predictive density (3) as a sequence of conditional predictive densities with increasing uncertainty in the prediction of successive variables. In this sense the factorization chosen should depend on the interest in controlling the different components of Z. For instance, the most convenient ordering will be clear for panel data or for a Markovian process. As an alternative, in the case of no interest, we can take

$$\hat{g}^*(z;\, y) := \frac{1}{m!}\sum_{\sigma} \hat{g}_{\sigma}(z;\, y)$$

as a predictive density, but in general the associated coverage error is of order $O(n^{-1})$. In fact each permutation of the variables gives rise to a predictive density that gives correct coverage probabilities for 'certain' regions (note that, in any case, the prediction limits are always correct up to order n^{-1}). Then, if we take an average of the different predictive densities, we obtain a predictive density that gives good coverage probabilities only if the previous regions are the same in each case. So we cannot, in general, use this averaged predictive density to obtain good prediction limits.

Nevertheless, when a predictive density with correct coverage probability for any region exists, the proposed predictive density captures the first terms of its development in powers of $n^{-1/2}$. This can be appreciated in the example of the next section.

Remark. In the particular case where the components of Z are independent the calculations simplify considerably:

$$h^i(z_i;\, \theta) = -\mathbf{b}^r(\theta)G^i_r(z_i;\, \theta) + \mathbf{i}^{rs}(\theta)B^i_{rs}(z_i;\, \theta),$$

with

$$B^i_{rs}(z;\, \theta) = \frac{1}{2}\left([2]\frac{g^i_r(z;\, \theta)}{g^i(z;\, \theta)}G^i_s(z;\, \theta) - G^i_{rs}(z;\, \theta)\right),$$

$$h^{ij}(z_i,\, z_j;\, \theta) = \mathbf{i}^{rs}(\theta)G^i_r(z_i;\, \theta)\frac{g^j_s(z_j;\, \theta)}{g^j(z_j;\, \theta)}, \tag{4}$$

and therefore

$$k^{ij}(z_i,\, z_j;\, \theta) = \mathbf{i}^{rs}(\theta)\frac{g^i_r(z_i;\, \theta)}{g^i(z_i;\, \theta)}\frac{g^j_s(z_j;\, \theta)}{g^j(z_j;\, \theta)}.$$

It is easily seen that these quantities are invariant with respect to changes in the parametrization. The correction terms involving the h^i are the same as those proposed by Barndorff-Nielsen and Cox (1996) and Vidoni (1998). Moreover, we can write

$$\hat{g}(z;\, y) \doteq g(z;\, \hat{\theta})\Bigg\{1 - \mathbf{b}^r(\hat{\theta})\partial_r \log g + \frac{1}{2}\mathbf{i}^{rs}(\hat{\theta})\Bigg(\partial_r \log g\partial_s \log g - \partial_{rs}\log g..$$
$$+ \sum_{i=1}^{m}[2]\partial_s\partial_{z_i}\log g^i(z_i;\, \hat{\theta})\frac{G^i_r(z_i;\, \hat{\theta})}{g^i(z_i;\, \hat{\theta})}\Bigg)\Bigg\}.$$

Thus, when the components of Z are identically distributed $\hat{g}(z;\, y)$ is a symmetric function of $(z_{1,}z_2,\, \ldots,\, z_m)$.

4. A touchstone example

Consider a sample $y_1, \ldots, y_n$ from a random vector $(Y_1, \ldots, Y_n)^{\mathrm{T}} \sim N(X\beta^{\mathrm{T}}, \sigma^2 I_n)$. We try to predict a further random vector $(Z_1, \ldots, Z_m)^{\mathrm{T}} \sim N(\Lambda\beta', \sigma^2 I_m)$, where X and Λ are known matrices of full rank, β is a p-dimensional unknown parameter, $p \leqslant n-1$, and $\sigma > 0$ is also unknown. We need to calculate the quantities involved in expression (2). The parameter θ is such that $\theta_i = \beta_i$, $i = 1, 2, \ldots, p$, and $\theta_{p+1} = \sigma$, but instead of $p+1$ we use the index σ and reserve the indices $r, s, \ldots$ to indicate the components of β. The asymptotic variance and bias of $\hat{\theta}$ are (see Barndorff-Nielsen and Cox 1994)

$$i = \begin{pmatrix} \dfrac{X^{\mathrm{T}}X}{\sigma^2} & 0 \\ 0 & \dfrac{2n}{\sigma^2} \end{pmatrix}, \qquad b^r = 0, \qquad b^\sigma = -\frac{\sigma}{4n}(1+2p).$$

Let $\mu_i = \lambda_i \beta^{\mathrm{T}}$, where $\lambda_i = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ip})$ is the ith row of Λ, $i = 1, \ldots, m$. Then, for $i = 1, \ldots, m$,

$$g^i(z_i; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\frac{(z_i - \mu_i)^2}{\sigma^2}\right\},$$

$$\frac{g^i_r}{g^i} = \frac{z_i - \mu_i}{\sigma^2}\lambda_{ir}, \qquad \frac{g^i_\sigma}{g^i} = -\frac{1}{\sigma} + \frac{(z_i - \mu_i)^2}{\sigma^3},$$

$$G^i_r = -g^i \lambda_{ir}, \qquad G^i_\sigma = -\frac{z_i - \mu_i}{\sigma} g^i,$$

$$G^i_{rs} = -\frac{z_i - \mu_i}{\sigma^2} g^i \lambda_{ir}\lambda_{is}, \quad G^i_{\sigma\sigma} = g^i \left\{\frac{2(z_i - \mu_i)}{\sigma^2} - \frac{(z_i - \mu_i)^3}{\sigma^4}\right\}.$$

Putting $\Delta_i = (z_i - \mu_i)/\sigma$ and assuming $(X^{\mathrm{T}}X)^{-1} \doteq \Gamma/n$, we have that

$$B^i_{rs} = -\frac{g^i}{2\sigma}\Delta_i \lambda_{ir}\lambda_{is}, \qquad B^i_{\sigma\sigma} = -\frac{g^i}{2\sigma}\Delta_i^3,$$

$$h^i = -\frac{\sigma}{4n}(1+2p) g^i \Delta_i - \frac{\sigma}{2n}\lambda_i \Gamma \lambda_i^{\mathrm{T}} g^i \Delta_i - \frac{\sigma}{4n} g^i \Delta_i^3,$$

$$k^i = \frac{1}{n}\left(\frac{p + \lambda_i \Gamma \lambda_i^{\mathrm{T}} - 1}{2}\Delta_i^2 + \frac{1}{4}\Delta_i^4\right) + C,$$

$$k^{ij} = \frac{1}{n}\left\{\lambda_i \Gamma \lambda_j^{\mathrm{T}} \Delta_i \Delta_j + \frac{1}{2}\Delta_i^2 \Delta_j^2 - \frac{1}{2}(\Delta_i^2 + \Delta_j^2)\right\},$$

where C is a constant term. Hence, up to order n^{-1}, the predictive density is given by

$$\hat{g}(z; y) \propto \exp\left\{\sum_{i=1}^m \left(-\frac{1}{2} + \frac{p + \lambda_i \Gamma \lambda_i^{\mathrm{T}} - m}{2n}\right)\hat{\Delta}_i^2 + \frac{1}{2n}\sum_{i=1}^m \sum_{j \leqslant i}\left(2\lambda_i \Gamma \lambda_j^{\mathrm{T}} \hat{\Delta}_i \hat{\Delta}_j + \hat{\Delta}_i^2 \hat{\Delta}_j^2\right)\right\},$$

Table 1. Comparison of coverage probabilities for the estimative, univariate corrected and multivariate corrected prediction limits. Results are based on 20 000 replications and estimated standard errors are always smaller than 0.004.

	α	Estimative	Univariate	Multivariate
$n = 10$ $m = 5$	0.1	0.123	0.135	0.095
	0.2	0.203	0.235	0.193
	0.3	0.292	0.338	0.298
	0.4	0.373	0.444	0.406
	0.5	0.448	0.533	0.502
	0.6	0.528	0.623	0.606
	0.7	0.609	0.712	0.704
	0.8	0.699	0.805	0.808
	0.9	0.799	0.894	0.902
$n = 20$ $m = 10$	0.1	0.115	0.137	0.096
	0.2	0.209	0.247	0.202
	0.3	0.288	0.340	0.297
	0.4	0.375	0.443	0.406
	0.5	0.450	0.527	0.502
	0.6	0.548	0.633	0.614
	0.7	0.631	0.721	0.714
	0.8	0.724	0.809	0.811
	0.9	0.828	0.898	0.906
$n = 30$ $m = 15$	0.1	0.109	0.130	0.093
	0.2	0.208	0.247	0.201
	0.3	0.294	0.346	0.302
	0.4	0.380	0.443	0.404
	0.5	0.466	0.538	0.510
	0.6	0.551	0.629	0.609
	0.7	0.640	0.718	0.708
	0.8	0.735	0.809	0.808
	0.9	0.845	0.901	0.907
$n = 40$ $m = 20$	0.1	0.111	0.131	0.095
	0.2	0.209	0.245	0.201
	0.3	0.293	0.342	0.299
	0.4	0.386	0.441	0.408
	0.5	0.465	0.531	0.504
	0.6	0.560	0.628	0.613
	0.7	0.655	0.724	0.717
	0.8	0.747	0.809	0.810
	0.9	0.853	0.900	0.908

with $\hat{\Delta}_i = (z_i - \hat{\mu}_i)/\hat{\sigma}$, $i = 1, \ldots, m$. This is equivalent to saying that, up to order n^{-1},

$$\left(\frac{n-p}{n}\right)^{1/2} \hat{\Delta} \sim t_m(0, I_m + \Lambda(X^{\mathrm{T}}X)^{-1}\Lambda^{\mathrm{T}}, n-p),$$

where $\hat{\Delta} = (\hat{\Delta}_1, \ldots, \hat{\Delta}_m)$, as one can prove by expanding this density in powers of n. On the other hand, it is a well-known result (see, for instance, Wang and Chow 1994) that this is the exact distribution of the pivotal quantity $(1 - p/n)^{1/2}\hat{\Delta}$ and it is then recovered by our predictive density.

Unfortunately, closed-form expressions for approximate predictive densities are rarely available. Thus, the performance of different methods should be evaluated through the behaviour of the corresponding approximate simultaneous prediction limits.

Here a simulation study has been performed for the case when $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_m$ are independent and identically distributed random variables having normal distribution $N(\mu, \sigma^2)$ with unknown parameters. This is a particular case of our example with $p = 1$, $\beta = \mu$, $X = 1_n^{\mathrm{T}}$ and $\Lambda = 1_m^{\mathrm{T}}$, where 1_k denotes the k-dimensional row vector with all components equal to 1. In Table 1 we compare the behaviour of three systems of prediction limits: the quantiles of the estimative marginal densities (estimative), the univariate corrected estimative quantiles obtained by disregarding terms ar h^{ij} in (4) (univariate), and those with multivariate correction given in (4) (multivariate). The results show that the proposed multivariate prediction limits improve on both the estimative and the univariate solutions.

Acknowledgements

We would like to thank the referees for helpful comments and suggestions which led to an improved version of the paper.

References

Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*. London: Chapman & Hall.

Barndorff-Nielsen, O.E. and Cox, D.R. (1996) Prediction and asymptotics. *Bernoulli*, **2**, 319–340.

Hall, P., Peng, L. and Tajvidi, N. (1999) On prediction intervals based on predictive likelihood or bootstrap methods. *Biometrika*, **86**, 871–880.

Vidoni, P. (1998) A note on modified estimative prediction limits and distributions. *Biometrika*, **85**, 949–953.

Wang, S.G. and Chow, S.C. (1994) *Advanced Linear Models: Theory and Applications*. New York: Marcel Dekker.

Received April 2004 and revised July 2005