Abstract: We consider the problem of comparing predictive intervals for a future observation via their expected lengths at a given confidence level. A higher order asymptotic theory is developed. This yields an explicit formula for expected length comparison and associated admissibility results. Illustrative examples are given.

Key words: Admissibility, confidence level, estimative density, higher order asymptotics, invariance, parametric model, predictive density.

1 Introduction

The problem of predicting a future observation from a parametric model given past observations has a long history and a substantial current interest is continuing in this area; see Barndor-Nielsen and Cox (1996, Section 2) for a brief but illuminating review. While attractive methods have been advocated in the literature from various considerations, it appears that an important and intuitively appealing criterion, namely that of expected length, has not been investigated so far. The present article aims at addressing this issue by developing a higher order asymptotic theory for comparison of predictive intervals directly via their expected lengths at a given confidence level.

We consider a general class of predictive intervals that includes the ones arising from the well-known estimative and predictive density approaches (Harris, 1989).
The setup is described in the next section. In section 3, we work out the conditions that ensure attainment of the stipulated confidence level up to the desired order of approximation. These conditions, in turn, yield an explicit and readily applicable formula for the comparison of expected length at a given confidence level. It is noted that this key formula is invariant with respect to the parametrization chosen. Admissibility results arising from the formula for expected length are presented in Section 4 where examples, including a numerical one, are also given.

In order to give a flavor of the basic ideas without making the notation too involved, we focus on the case of scalar $\mu$, where $\mu$ is the parameter indexing the model. However, in Section 5, we indicate the form that the main formula assumes in the case of vector $\mu$. Throughout, our emphasis has been on unimodal models that are bell-shaped. The treatment of other unimodal models is simpler and this point has also been noted in Section 5.

2 Setup

Let $X_1$; $\cdots$; $X_n$ and $X_0$ be scalar-valued observations that are independent and identically distributed with common density $f(x; \mu)$, where $\mu$ is an unknown scalar parameter (the case of vector $\mu$ will be indicated in Section 5). Here $X_1$; $\cdots$; $X_n$ are available observations and, on the basis of these, it is intended to predict $X_0$, a future observation. Standard regularity conditions are assumed. In particular, the parameter space for $\mu$ is supposed to be the real line or an open interval therein, and we assume the existence of a valid Edgeworth expansion for the distribution of $n^{1/2}(\hat{\mu} - \mu)$, where $\hat{\mu}$ is the maximum likelihood estimator of $\mu$ based on $X_1$; $\cdots$; $X_n$.

We also assume that $f(x; \mu)$ is unimodal for each $\mu$. Furthermore, in this and the next two sections we consider the case of bell-shaped $f(x; \mu)$ such that given any $\alpha(0 < \alpha < 1)$, there exist unique $a(\mu)$ and $b(\mu)$ ($a(\mu) < b(\mu)$) satisfying

$$f(a(\mu); \mu) = f(b(\mu); \mu); \quad F(b(\mu); \mu) - F(a(\mu); \mu) = 1 - \alpha. \quad (2.1)$$
where \( F(x; \mu) \) is the distribution function corresponding to the density \( f(x; \mu) \). The functions \( a(\mu) \) and \( b(\mu) \), that depend on \( \sigma \), are supposed to be smooth.

For predicting the future observation \( X_0 \), we consider predictive intervals of the form

\[
S = [a_0(\mu) + n^{-1}a_1(\mu) + n^{-2}a_2(\mu); b_0(\mu) + n^{-1}b_1(\mu) + n^{-2}b_2(\mu)]
\]

where the \( a_i(\phi) \) and \( b_i(\phi) \) are smooth functions, with functional forms free from \( n \), to be so chosen that \( S \) has confidence coefficient \( 1 - \sigma + o(n^{-2}) \). We propose to examine the influence that the choice of these functions exercises over the expected length of \( S \). Predictive intervals of the form (2.2) can arise quite commonly. In the estimative density approach, predictive intervals are of this form. The same happens in the predictive density approach (Harris, 1989), where in the integrated average of \( f(x; \hat{\mu}) \), with respect to the density of \( \hat{\mu} \), one replaces \( \mu \) by \( \hat{\mu} \) or a perturbed version thereof of the form \( \hat{\mu} + n^{-1}g(\hat{\mu}) \).

As hinted above, here we are considering asymptotics with margin of error \( o(n^{-2}) \). The reason for this will be clear in the next section. Incidentally, asymptotics of the same order are needed in comparing different approaches to prediction via the Kullback-Leibler divergence from the true density (Aitchison, 1975; Harris, 1989; Vidoni, 1995).

Before concluding this section, we note that the optimal determination of \( a_0(\phi) \) and \( b_0(\phi) \) in (2.2) is quite easy. By (2.2), writing \( l(S) \) for the length of \( S \),

\[
\begin{align*}
P_\mu(X_0 \notin S) &= E_\mu[Z f(x; \mu)dx] = F(b_0(\mu); \mu) - F(a_0(\mu); \mu) + o(1); \\
E_\mu[l(S)] &= b_0(\mu) - a_0(\mu) + o(1);
\end{align*}
\]

On the basis of the leading terms in the above, we need to minimize \( b_0(\mu) - a_0(\mu) \) subject to \( F(b_0(\mu); \mu) - F(a_0(\mu); \mu) = 1 \). By (2.1), the obvious solution is

\[
a_0(\mu) = a(\mu); \quad b_0(\mu) = b(\mu);
\]

Hereafter, we work with \( a_0(\phi) \) and \( b_0(\phi) \) as in (2.3) and consider optimal selection of \( a_i(\phi) \) and \( b_i(\phi) \) (\( i = 1, 2 \)).
3 Main result

3.1 Some notation

Let \( \hat{\mu} \) be the maximum likelihood estimator of \( \mu \). Under standard regularity conditions,

\[
E_\mu(\hat{\mu}) = n^{-2}R_{11} + n^{-3}R_{12} + o(n^{-2}); \quad E_\mu(\hat{\mu}^2) = 1^{-1} + n^{-1}R_2 + o(n^{-1}); \\
E_\mu(\hat{\mu}^3) = n^{-2}R_3 + o(n^{-1}); \quad E_\mu(\hat{\mu}^4) = 3n^{-2} + O(n^{-1});
\]

where \( I(\mu) \) is the per observation Fisher information, and \( R_{11}; R_{12}; R_2; R_3 \) are smooth functions of \( \mu \) that depend only on the model and have functional forms free from \( n \). The detailed expressions for these are not hard to obtain, for example, from Chandra and Joshi (1983) though for some models a first principle derivation can be even simpler; see the numerical example in Section 4. In particular, \( R_{11} = I_1^2E_\mu(V_1V_2 + \frac{1}{2}V_3) \) where \( V_j \) is the \( j \)-th derivative of \( \log f(X; \mu) \) with respect to \( \mu \) for \( j = 1; 2; \ldots \).

For \( j = 1; 2; \ldots \) let

\[
G_j = \frac{d}{d\mu} F(b(\mu); \mu)j_u=\mu; \quad H_j = \frac{d}{d\mu} f(b(\mu); \mu)j_u=\mu;
\]

Similarly, let \( K_j \) and \( L_j \) be defined replacing \( b(\mu) \) by \( a(\mu) \) in the expressions for \( G_j \) and \( H_j \) respectively. For notational simplicity, we shall often write \( b \sim b(\mu), b_1 \sim b_1(\mu), a \sim a(\mu), a_1 \sim a_1(\mu), a_0 \sim a_0(\mu), b_0 \sim b_0(\mu), b_0^0 \sim b_0^0(\mu) \) etc, where primes denote differentiation with respect to \( \mu \). Let \( f_1(x; \mu) = @f(x; \mu)@x \).

3.2 Attainment of confidence level

By (2.2) and (2.3),

\[
P_\mu(X_0 2 S) = E_\mu(U_1 \cup U_2);
\]

where

\[
U_1 = F (b(\hat{\mu}) + n^{-1}b_2(\hat{\mu}) + n^{-2}b_2(\hat{\mu}); \mu); \quad U_2 = F (a(\hat{\mu}) + n^{-1}a_1(\hat{\mu}) + n^{-2}a_2(\hat{\mu}); \mu);
\]

Note that

\[
U_1 = F (b(\hat{\mu}); \mu) + n^{-1}b_2(\hat{\mu}) f(b(\hat{\mu}); \mu)
\]
\[ n^2 b_2(\mu) f(b(\mu); \mu) + \frac{1}{2} f b_1(\mu) g^2 f_1(b(\mu); \mu) + o(n^2) \quad (3.4) \]

Since \( \hat{\mu} = \mu + n^2 \), a stochastic expansion about \( \mu \) and use of (3.1) and (3.2) yield

\[ E_{\mu} F(b(\hat{\mu}); \mu) g = F(b; \mu) + n^2 B_1 + n^2 B_2 + o(n^2) \]

where

\[ B_1 = G_1 R_{11} + \frac{1}{2} G_2 I^1; \quad B_2 = G_1 R_{12} + \frac{1}{2} G_2 R_2 + \frac{1}{6} G_3 R_3 + \frac{1}{8} G_4 I^2 \quad (3.5) \]

By a similar computation for the other terms in (3.4),

\[ E_{\mu}(U_1) = F(b; \mu) + n^2 [B_1 + b_1 f(b; \mu)] + n^2 [B_2 + R_{11} f b_1^0 (b; \mu) + b_1 g + \frac{1}{2} I^1 f b_1^0 (b; \mu) + 2 b_1 g + b_1 f_1 (b; \mu)] + o(n^2) \]

\[ (3.6) \]

In view of (2.1), let \( !^i(\mu) > 0 \) be the common value of \( f(a; \mu) \) and \( f(b; \mu) \).

Then by (2.1), (3.3), (3.6) and a similar expression for \( E_{\mu}(U_2) \),

\[ P_{\mu}(X_0 = 2 S) = n^2 [! (b_1 a_1) A_1 + B_1] + n^2 [! (b_2 a_2) D] + o(n^2) \]

\[ (3.7) \]

where

\[ D = A_2 i \quad B_2 + R_{11} ! (a_1^0 b_1^0) + a_1 L_1 i \quad b_1 H_1) + \frac{1}{2} a_1^2 f_1 (a; \mu) i \quad \frac{1}{2} b_1^2 f_1 (b; \mu) \]

\[ + \frac{1}{2} I^1 ! (a_1^0 b_1^0) + 2 (a_1^0 L_1 i \quad b_1^2 H_1) + a_1 L_1 i \quad b_1 H_2 \]  

\[ (3.8) \]

and, analogously to (3.5),

\[ A_1 = K_1 R_{11} + \frac{1}{2} K_2 I^1; \quad A_2 = K_1 R_{12} + \frac{1}{2} K_2 R_2 + \frac{1}{6} K_3 R_3 + \frac{1}{8} K_4 I^2 \]

\[ (3.9) \]

By (3.7), \( S \) has confidence level \( 1 \) \( \oplus \) \( o(n^i^2) \) if and only if

\[ b_1 i \quad a_1 = ! i^1 Q \quad \text{and} \quad b_2 i \quad a_2 = ! i^1 D \]

\[ (3.10) \]
where \( Q = A_1 \cap B_1 \). Note that by (3.5) and (3.9),

\[
Q = R_{11}(K_1 \cap G_1) + \frac{1}{2} l \cap \frac{1}{2}(K_2 \cap G_2) : \tag{3.11}
\]

Also observe that \( G_j \cap H_j \cap K_j \) and \( L_j \) and hence \( A_1 \cap A_2 \cap B_1 \cap B_2 \) and \( Q \) do not depend on the choice of \( a_1 ; a_2 ; b_1 \) or \( b_2 \).

### 3.3 Formula for expected length

By (2.2) and (2.3), the predictive interval \( S \) has length

\[
l(S) = b(\hat{\mu})_i a(\hat{\mu}) + n^i 1[b_1(\hat{\mu}) \cap a_1(\hat{\mu})] + n^i 2[b_2(\hat{\mu}) \cap a_2(\hat{\mu})] : \tag{3.12}
\]

Recalling that \( \hat{\mu} = \mu + n^i 1 \hat{\mu} \), a stochastic expansion about \( \mu \) and use of (3.1) yields

\[
E_\mu[b(\hat{\mu})_i a(\hat{\mu})] = C_0 + n^i 1 C_1 + n^i 2 C_2 + o(n^i 2) ;
\]

where the \( C_i \) \( C_i(\mu) \) neither involve \( n \) nor depend on the choice of \( a_1 ; a_2 ; b_1 \) or \( b_2 \).

A similar computation for the other terms in (3.12) shows that

\[
E_\mu[l(S)] = C_0 + n^i 1 (C_1 + b_1 \cap a_1) + n^i 2 (C_2 + R_{11}(b_0 \cap a_1))
+ \frac{1}{2} l \cap \frac{1}{2}(b_1 \cap a_1) + b_2 \cap a_2] + o(n^i 2) \tag{3.13}
\]

By the \( \cap \)rst condition in (3.10), \( C_1 + b_1 \cap a_1 = C_1 + \frac{1}{2} Q \) which does not depend on the choice of \( a_1 ; a_2 ; b_1 \) or \( b_2 \). This is why we need to consider the term of order \( O(n^i 2) \) in the expression (3.13) for expected length. The second condition in (3.10) eliminates \( b_2 \) and \( a_2 \) from this term and then one can again employ the \( \cap \)rst condition in (3.10) to express it in terms of \( a_1 \) alone. By (3.8), (3.10) and (3.13), these steps yield

\[
E_\mu[l(S)] = C_0 + n^i 1 - 1 + n^i 2 (- 2 + \frac{1}{2} 4 ) + o(n^i 2) \tag{3.14}
\]

after some simplification, where

\[
4 = M_1 a_1 + \frac{1}{2} M_2 a_1^2 + l \cap \frac{1}{2} M_3 a_1^0 \tag{3.15}
\]
with
\[
\begin{align*}
M_1 &= R_{11}(L_1 H_1) + \frac{1}{2} I_{11}^{(-1)}(L_2 H_2) i_{11}^{(-1)} Q_{11}(b; \mu); \\
M_2 &= f_1(a; \mu) i f_1(b; \mu); \\
M_3 &= L_1 H_1.
\end{align*}
\]

Like \( C_0 \), the quantities \( \eta_1, \eta_2 \) in (3.14) are functions of \( \mu \) that neither involve \( n \) nor depend on \( a_2; b_2 \).

Thus, up to the order of approximation considered in (3.14), the choice of \( a_2 \) and \( b_2 \) does not influence \( E_{\mu}(l(S)) \). They only need to satisfy the second condition in (3.10) arising from the target confidence level. This is in the spirit of what happens in the higher order asymptotic theory of testing of hypothesis; compare, for example, Mukerjee (1992). The choice of \( a_1 \), however, does have an impact, via \( 4 \), on \( E_{\mu}(l(S)) \) as in (3.14). One needs to choose \( a_1 \) such that \( 4 \) is small. After \( a_1 \) is chosen appropriately, \( b_1 \) can be obtained from the first condition in (3.10).

The expression (3.15) for \( 4 \) is the key formula for the comparison in which we are interested here. It is satisfying to note that \( 4 \) is invariant under one-to-one transformation of \( \mu \). This can be checked from (3.15) proceeding somewhat along the line of Datta and Ghosh (1996) in a different context. The details are omitted.

\section{Implications and examples}

Since we are considering bell-shaped models, by (2.1) and (3.16), \( M_2 > 0 \) for all \( \mu \).

If, in addition, \( M_3 = 0 \) identically in \( \mu \) (this happens, for example, in scale models that are symmetric about zero), then by (3.15), the optimal solution for \( a_1 \) that minimizes \( 4 \) uniformly in \( \mu \) is \( a_1 = M_1 M_2 \).

We now consider models for which \( M_3 \) is either positive for all \( \mu \) or negative for all \( \mu \). This class of models is quite general. In particular, every location or scale model is of this type unless it has \( M_3 = 0 \) identically in \( \mu \). For models of this kind, the formula (3.15) yields second-order admissibility results akin to those in the context of point-estimation (Ghosh and Sinha, 1981). The term "second-order"
is being used since we are working with margin of error $o(n^{1/2})$.

Without loss of generality, let $M_3 > 0$ for all $\mu$ (if $M_3 < 0$ for all $\mu$, one needs to consider $a_1$ rather than $a_1$). A choice of $a_1$ such that $a_1(\mu) = c(\mu)$ will be called second-order admissible if there exists no other choice, say $d(\mu)$, such that

$$4j_{a_1(\mu)} = d(\mu) \cdot 4j_{a_1(\mu)} = c(\mu)$$

for all $\mu$, with strict inequality for some $\mu$. By (3.15), writing $g(\mu) = d(\mu) \cdot c(\mu)$,

$$4j_{a_1(\mu)} = d(\mu) \cdot 4j_{a_1(\mu)} = c(\mu)$$

after some simplification, where

$$q(\mu) = c(\mu) + \frac{1}{2} M_2 \{ g^2(\mu) + 2q(\mu)g(\mu) + 2fT(\mu)g^2q(\mu) \} \quad (4.1)$$

Let the interval $(\mu_1; \mu_2)$ be the parameter space for $\mu$, where $\mu_1 = -1$ or $\mu_2 = +1$ are possible. For $\mu; \mu_0 2 (\mu_1; \mu_2)$, let

$$\tilde{A}(\mu; \mu_0) = T(\mu) \exp \int_{\mu_0}^{\mu} q(u)T(u)du:$$

As in Ghosh and Sinha (1981), it now follows from (4.1) that a choice $c(\mu)$ of $a_1(\mu)$ is second-order admissible if and only if for some $\mu_0 2 (\mu_1; \mu_2)$,

$$Z_{\mu_0}^{\mu} \tilde{A}(\mu; \mu_0)d\mu = 1 \quad \text{and} \quad Z_{\mu_0}^{\mu} \tilde{A}(\mu; \mu_0)d\mu = 1: \quad (4.4)$$

Example 1: Consider a symmetric bell-shaped location model $f(x; \mu) = h(x \mid \mu)$, where $-x < \mu < 1$ and $h(\phi)$ is a density on the real line satisfying $h(u) = h(i \mid u)$ for all $u$. Writing $h_i(u) = d^i h(u) = du^i$, then

$$h_1(u) = i h_1(i \mid u); \quad h_2(u) = h_2(i \mid u); \quad (4.5)$$

for all $u$. By (2.1), here $a = \mu \mid c$ and $b = \mu + c$, where $c(> 0)$ is a constant that depends on $\phi$ but not $\mu$. Hence $h = h(c)$ and, using (4.5), $f_1(b; \mu) = i f_1(a; \mu) = h_1(c)$. Also, $I = v$, where $v(> 0)$ is a constant free from $\mu$, and because of symmetry, $R_{11} = 0$. Furthermore, by (3.2) and (4.5),

$$G_1 = K_1 = h(c); \quad G_2 = i K_2 = H_1 = i L_1 = h_1(c); \quad H_2 = L_2 = h_2(c);$$
so that by (3.11), \( Q = i v h_1(c) \). Hence (3.16) yields
\[
M_1 = [vh(c)] h_1^2(c); \quad M_2 = M_3 = i 2h_1(c);
\]
(4.6)

Note that \( h_1(c) < 0 \) as a bell-shaped model is being considered.

Consider now natural choices of \( a_1 \) of the form \( a_1 = r \), where \( r \) is any constant free form \( \mu \). Using (4.6) in (3.15), the optimal \( r \) that minimizes \( \Phi \) is then \( r_0 = h_1(c) = 2vh(c) \). Thus, as long as \( a_1 \) is a constant, its best choice is \( a_1 = r_0 \). By (3.10), the corresponding \( b_1 \) is \( b_1 = i r_0 \). This is in agreement with the intuitive expectation that \( a_1 \) and \( b_1 \) should be equidistant from zero.

We next show that the choice \( a_1 = r_0 \) is second-order admissible among all choices of \( a_1 \). With \( a_1 = r_0 \), by (4.2), (4.3) and (4.6), \( q(\mu) = 0 \); \( T(\mu) = v \), and \( \tilde{A}(\mu; \mu_0) = v \). Here, \( \mu_1 = i 1 \), \( \mu_2 = +1 \), and by (4.4), the claimed second-order admissibility of \( a_1 = r_0 \) is now evident.

Example 2: As a specific example of a scale model, consider the univariate normal model with mean \( \mu \) and variance \( \mu^2 \), where \( \mu > 0 \). Then
\[
I = 3\mu^2; \quad R_{11} = i \frac{2\mu}{\varphi};
\]
(4.7)
and by (2.1) and (3.2),
\[
a = \mu(1 + z); \quad b = \mu(1 + z); \quad ! = \tilde{A}(z) = \mu;
\]
\[
f_1(a; \mu) = i f_1(b; \mu) = z\tilde{A}(z) = \mu^2;
\]
\[
G_1 = (1 + z)\tilde{A}(z) = \mu; \quad G_2 = i z(1 + z)^2\tilde{A}(z) = \mu^2;
\]
\[
H_1 = i zG_1 = \mu; \quad H_2 = i (z^2 + 1)G_2 = (\mu z);
\]
(4.8)
where \( \tilde{A}(\phi) \) is the standard univariate normal density and \( z \) is the corresponding \( (1 i \frac{1}{2} \&) \)th quantile. Furthermore, \( K_1; K_2; L_1 \) and \( L_2 \) can be obtained replacing \( z \) by \( i z \) in \( G_1; G_2; H_1 \) and \( H_2 \) respectively. Hence by (3.11), \( Q = \frac{1}{9}z(3z^2 + 7)\tilde{A}(z) \). Now (3.16) yields
\[
M_1 = \frac{1}{9}z(3z^3 i 6z^2 + 7z + 2)\tilde{A}(z) = \mu; \quad M_2 = M_3 = 2z\tilde{A}(z) = \mu^2;
\]
(4.9)
Consider natural choices of $a_1$ of the form $a_1 = r\mu$, where $r$ is any constant free from $\mu$. Using (4.9) in (3.15), the optimal $r$ that minimizes $4$ is then $r_0 = i \frac{1}{18}(3z^3 + 6z^2 + 7z + 8)$. By (3.10), the corresponding $b_1$ is $b_1 = r^n\mu$, where $r^n = i \frac{1}{18}(3z^3 + 6z^2 + 7z + 8)$. As in the previous example, one can check that $a_1 = r_0\mu$ is also second-order admissible among all choices of $a_1$.

Example 2(continued): Continuing with the last example, we now numerically examine how close the coverage probability of an optimal interval suggested there is to the target in small samples. In this example,

$$\hat{\mu} = \frac{1}{2}[i \hat{X} + (\hat{X}^2 + \frac{4}{n} \sum_{i=1}^{n} X_i^2)]^{1/2};$$

(4.10)

where $\hat{X} = (1-n)^{\frac{1}{2}} \sum_{i=1}^{n} X_i$. Hence from first principles one can check that

$$R_{11} = i \frac{2}{243} \mu; \quad R_2 = \frac{2}{81} \mu^2; \quad R_3 = i \frac{4}{27} \mu^3.$$  

(4.11)

Furthermore,

$$G_3 = (1 + z)^3(z^2 \cdot 1)\hat{A}(z) = \mu^3; \quad G_4 = i (1 + z)^4(z^3 \cdot 3z)\hat{A}(z) = \mu^4;$$  

(4.12)

and $K_3$ and $K_4$ can be obtained replacing $z$ by $z$ in $G_3$ and $G_4$ respectively.

We take $a_1 = r_0\mu$ and $b_1 = r^n\mu$, where $r_0$ and $r^n$ are as described above. Also, let $a_2 = s_0\mu; b_2 = s^n\mu$, where $s_0$ and $s^n$ are constants free from $\mu$. Suppose $1 \frac{1}{2} \mu = 0:95$. Then $z = 1:96, r_0 = i 1:8111, r^n = 2:8532$, and by (4.8), $a = i 0:96\mu; b = 2:96\mu$. Also, by (3.10), together with (3.5), (3.8), (3.9), (4.7), (4.8), (4.11) and (4.12), $s^n \mu; s_0 = 3:8993$. Thus from (2.2) and (2.3), the optimal predictive interval is of the form $S_0 = [\frac{1}{4}_n \mu; \frac{1}{2}_n \mu]$, where

$$\frac{1}{4}_n = i 0:96 \frac{1}{n} \frac{1:8111}{n} + \frac{s_0}{n^2}; \quad \frac{1}{2}_n = 2:96 + \frac{2:8532}{n} + \frac{s_0 + 3:8993}{n^2};$$

In view of (4.10), the coverage probability of $S_0$ equals $P(\frac{1}{4}_n \cdot Y_0 = \hat{Y} \cdot \frac{1}{2}_n)$, where $\hat{Y}$ is as in the right-hand side of (4.10) with $X_1; \ldots; X_n$ there replaced by $Y_1; \ldots; Y_n$, the latter random variables as well as $Y_0$ being independently normal each with mean unity and variance unity.
Table 1: Simulated coverage probability of the optimal predictive interval in Normal $(\mu; \mu^2)$ model

<table>
<thead>
<tr>
<th>$s_0$</th>
<th>n=1</th>
<th>n=3</th>
<th>n=5</th>
<th>n=7</th>
<th>n=10</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.9054</td>
<td>0.9452</td>
<td>0.9477</td>
<td>0.9495</td>
<td>0.9505</td>
</tr>
<tr>
<td>0</td>
<td>0.9045</td>
<td>0.9463</td>
<td>0.9477</td>
<td>0.9495</td>
<td>0.9504</td>
</tr>
<tr>
<td>1</td>
<td>0.8877</td>
<td>0.9448</td>
<td>0.9475</td>
<td>0.9488</td>
<td>0.9503</td>
</tr>
</tbody>
</table>

The above coverage probability was studied by simulation for $s_0 = 1; 0; 1$, and various small values of $n$. The results, each based on 20,000 simulations, are summarized in Table 1. They reveal that the convergence towards the target coverage 0.95 is quite fast for each $s_0$ considered. Furthermore, as anticipated from asymptotic considerations, the specific choice of $s_0$ is not of much consequence from $n = 3$ onwards.

5 Further developments

5.1. Another kind of model

The development so far excludes models like simple exponential that are unimodal but not bell-shaped. To cover models of this kind, we now suppose that $f(x; \mu)$ has support $[\gamma; 1)$ for some constant $\gamma$ (free from $\mu$) and is strictly decreasing in $x$ over $[\gamma; 1)$. The dual case where $f(x; \mu)$ has support $(1; \gamma]$ and is strictly increasing in $x$ over $(1; \gamma]$ can be treated similarly. In contrast with what happened in Section 3, it is seen below that for models of this kind the choice of $a_1$ and $b_1$ has an impact on the expected length even when asymptotics are carried out with margin of error $o(n^{-1})$.

Let $b(\mu)$ be the $(1 - \beta)$th quantile of the model $f(x; \mu)$. Since $f(x; \mu)$ is strictly decreasing in $x$ over $[\gamma; 1)$, a simple analysis as in (2.3) shows we need to consider only predictive intervals of the form

$$S = [\gamma + n^{-1}a_1(\hat{\mu}); b(\hat{\mu}) + n^{-1}b_1(\hat{\mu})];$$
where \( a_1(\phi) \) and \( b_1(\phi) \) are smooth functions, with functional forms free from \( n \), to be so chosen that \( S \) has con- dence coe± cient \( 1 \pm o(n^{-1}) \). We also require \( a_1(\mu) \geq 0 \) for all \( \mu \). An algebra similar to but simpler than that in Section 3 shows that

\[
P_{\mu}(X_0 \in S) = 1 + n^{-1} \{ b_1 f (b; \mu) + a_1 f (a; \mu) + B_1 \} + o(n^{-1});
\]

\[
E_{\mu}[\ell(S)] = E_{\mu}[b(\mu)] + n^{-1} \{ b_1 \} + o(n^{-1});
\]

where \( B_1 \) is as in (3.5) and as before, \( a_1 \neq a_1(\mu) \), \( b_1 \neq b_1(\mu) \). Hence we need to choose \( a_1 \) and \( b_1 \) so that \( b_1 - a_1 \) is minimized subject to \( b_1 = a_1 f (a; \mu) - B_1 \). Hence the optimal solution to the above constrained optimization problem turns out to be \( a_1 = 0; b_1 = b(\mu) \). Thus, discrimination among competing choices of \( a_1 \) and \( b_1 \) is possible even working under asymptotics with margin of error \( o(n^{-1}) \).

5.2. Case of Vector \( \mu \)

We now return to unimodal bell-shaped models and consider the case of vector \( \mu \), with a view to indicating the form that the key formula (3.15) assumes in this situation. Let \( \mu = (\mu_1; \ldots; \mu_p)^T \), the parameter space for \( \mu \) being the \( p \)-dimensional Euclidean space or some open subset thereof. Let \( a(\mu) \) and \( b(\mu) \) be de±ned by (2.1). As in Section 2, we consider predictive intervals of the form (2.2) where \( a_0(\mu) \) and \( b_0(\mu) \) satisfy (2.3).

Let \( \gamma = (\gamma_1; \ldots; \gamma_p)^T = n^{1/2}(\mu - \mu_0) \). For \( 1 \leq i; j \leq p \), under standard regularity conditions,

\[
E_{\mu}(\gamma_i) = n^{1/2} \gamma_i + o(n^{-1}); \quad E_{\mu}(\gamma_i \gamma_j) = I_{ij} + O(n^{-1})
\]

where \( I = (I_{ij}) \) is the per observation Fisher information matrix, \( I_{ij} = (I_{ij}) \), and the \( \gamma_i \), which generalize \( R_{11} \) of (3.1), are smooth functions of \( \mu \) that depend only on the model and have functional forms free from \( n \). For \( u = (u_1; \ldots; u_p)^T \) and
\[ G^{(i)} = \frac{\partial}{\partial u_i} F(b(u); \mu) j_{u=\mu}; \quad G^{(ij)} = \frac{\partial^2}{\partial u_i \partial u_j} F(b(u); \mu) j_{u=\mu}; \quad (5.1) \]

Similarly, define \( H^{(i)} \) and \( H^{(ij)} \), \( K^{(i)} \) and \( K^{(ij)} \), and \( L^{(i)} \) and \( L^{(ij)} \) replacing \( F(b(u); \mu) \) in (5.1) by \( f(b(u); \mu) \), \( F(a(u); \mu) \) and \( f(a(u); \mu) \) respectively. Let

\[ Q = -i[K^{(i)} i G^{(i)}] + \frac{1}{2} l^{ij} [K^{(ij)} i G^{(ij)}]; \]

where summation convention is followed.

As before, we write \( a \sim a(\mu), b \sim b(\mu), a_1 \sim a_1(\mu), b_1 \sim b_1(\mu) \) etc, and denote the common value of \( f(a; \mu) \) and \( f(b; \mu) \) by \( ! \). For \( 1 \cdot i; j \cdot p \), let \( a^{(i)}_1 = \circ a_1 = \circ a \).

An algebra similar to but heavier than that in Section 3 shows that if \( a_1; a_2; b_1 \) and \( b_2 \) are chosen so as to ensure the attainment of a confidence level \( 1 \circ \circ + o(n^{-2}) \), then the expression (3.14) for expected length remains valid. As before, \( C_0; C_1 \) and \( C_2 \) in (3.14) continue to be functions of \( \mu \) that neither involve \( n \) nor depend on \( a_1, a_2, b_1 \) or \( b_2 \). The expression (3.15) for \( 4 \) appearing in (3.14) now gets modified to

\[ 4 = M_1 a_1 + \frac{1}{2} M_2 a_1^2 + M_3 a^{(i)}_1; \]

where

\[ M_1 = -i[L^{(i)} i H^{(i)}] + \frac{1}{2} l^{ij} f L^{(ij)} i H^{(ij)} g_j ! i Q f_1(b; \mu); \]
\[ M_2 = f_1(a; \mu) i f_1(b; \mu); \quad M_3 = l^{ij} [L^{(i)} i H^{(j)}]; \]

using summation convention again. It is not hard to see that the above is in agreement with (3.15) and (3.16) when \( p=1 \).

Acknowledgement  This work was supported by a grant from the Center for Management and Development Studies, Indian Institute of Management Calcutta and a grant from the OAP program of the National University of Singapore.
REFERENCES


