Ann. Inst. Statist. Math.
37 (1985), Part A, 507-517

MAXIMUM LIKELIHOOD PREDICTION

KENNETH S. KAMINSKY* AND LENNART S. RHODIN**

(Received Feb. 2, 1984; revised Jan. 8, 1985)

Summary

The principle of maximum likelihood is applied to the joint prediction and estimation of a future random variable and an unknown parameter. We assume dependence between present and future, and the approach is non-Bayesian. Our principal application is to the prediction of higher order statistics from lower ones in Type II censored random samples. Some simple criteria for existence and uniqueness of the predictor are given for this situation and the methods are illustrated with several examples.

1. Introduction

Prediction of future events (or estimation of events which have occurred but were unobservable) on the basis of past and present knowledge is a fundamental problem of statistics, arising in many contexts, and producing varied solutions. Some recent works on the subject include Faulkenberry [5], Lauritzen [12], Aitchison and Dunsmore [1], and Hinkley [7].

With certain exceptions, point prediction problems have been solved by least squares or more general projection methods and their precision is most often measured by *mean square error* (MSE). Bayesian methods have been used to a considerable extent to obtain prediction intervals by means of predictive distributions and predictive likelihood functions. An important limitation has been that the quantity to be predicted is usually assumed to be independent (conditional on the parameter) of the observed data. Obviously, there often exists a dependence between that which is to be predicted and the data on which the prediction is to be based.

It is our intention in this paper to expound the classical method

Key words and phrases: Prediction, maximum likelihood, order statistics.

* Present address: St. Olaf College, Northfield, Minnesota, U.S.A.
** Present address: International Consultants, Kuwait.

of maximum likelihood as a device for predicting future observations in a framework where present and past are generally assumed to be dependent. Even if they are independent, reasonable predictors may result, but it is possible to construct examples where they do not.

The principal application which we will make of the method outlined in this paper is to prediction of higher order statistics from the lower ones in Type II censored random samples (cf. Kaminsky [8]). Some simple sufficient conditions for the existence of a unique maximum likelihood predictor will be given. A somewhat expanded version [11] of the present paper is available from the Institute of Mathematical Statistics, University of Umeå, 901 87 Umeå, Sweden.

2. Definitions and notation

Let $\boldsymbol{X}^T=(X_1, X_2, \cdots, X_p)$ and $\boldsymbol{Y}^T=(Y_1, Y_2, \cdots, Y_q)$ denote random vectors with joint pdf $f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\beta})$ indexed by the parameter $\boldsymbol{\beta} \in \Omega$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ could both be observed, f would correspond to the usual likelihood function of $\boldsymbol{\beta}$. The problem here will be to predict the unobservable (either future, or past and missing) value of $\boldsymbol{Y}$, having observed $\boldsymbol{X}$. As indicated above, we shall generally assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dependent. Thus, viewed as a function of $\boldsymbol{y}$ and $\boldsymbol{\beta}$, we define

$$L(\boldsymbol{y}, \boldsymbol{\beta}; \boldsymbol{x})=f(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\beta}) ,$$

to be the *predictive likelihood function* (PLF) of $\boldsymbol{y}$ and $\boldsymbol{\beta}$. (Lauritzen [12] and Hinkley [7] independently gave different definitions for related problems. Hinkley's predictive likelihood for example, is based on the pdf of $\boldsymbol{X}$ given the minimal sufficient reduction of $(\boldsymbol{X}, \boldsymbol{Y})$).

Suppose $\boldsymbol{Y}^*=t(\boldsymbol{X})$ and $\boldsymbol{\beta}^*=u(\boldsymbol{X})$ are statistics for which

$$L(\boldsymbol{y}^*, \boldsymbol{\beta}^*; \boldsymbol{x})=\sup_{(\boldsymbol{y}, \boldsymbol{\beta})} L(\boldsymbol{y}, \boldsymbol{\beta}; \boldsymbol{x}) .$$

We call $\boldsymbol{Y}^*$ the *maximum likelihood predictor* (MLP) of $\boldsymbol{Y}$ and $\boldsymbol{\beta}^*$ the *predictive maximum likelihood estimator* (PMLE) of $\boldsymbol{\beta}$. The PLF is not a predictive distribution. That would take the form $p(y \mid x)$ (cf. Aitchison and Dunsmore [1]), the parameter having been eliminated once its prior distribution is specified. Nevertheless, our use of L seems natural to us and as we will see in the examples, the results are reasonable.

Remark 1. Notice that if $\boldsymbol{\beta}$ is known, an MLP for $\boldsymbol{Y}$ is also a mode of the conditional distribution of $\boldsymbol{Y}$ given $\boldsymbol{X}=\boldsymbol{x}$. In Section 3, we will discuss conditions on the underlying pdf itself which guarantee the existence of a unique MLP for $\boldsymbol{Y}$.

3. Some results for order statistics

Let $X_1 \leqq X_2 \leqq \cdots \leqq X_n$ denote the order statistics of a random sample of fixed size n from a population with pdf $f(x;\boldsymbol{\beta})$, $(x,\boldsymbol{\beta}) \in \boldsymbol{D}=[a,b]\times \Omega$; Ω a k-dimensional interval. Assume further that f is positive and continuous on $\boldsymbol{D}$, vanishes outside $\boldsymbol{D}$, has continuous first partial derivatives in x and $\boldsymbol{\beta}$, and that the associated cdf, $F<1$ on $\boldsymbol{D}$. In life-testing and survival analysis, it is common to take a as finite (often zero) and b as infinite. In that case we take the interval to be $[a,\infty)$.

We consider prediction of X_s, having observed $X_1, X_2, \cdots, X_r$ $(1 \leqq r < s \leqq n)$. The PLF of X_s and $\boldsymbol{\beta}$ is

$$L(x_s, \boldsymbol{\beta};\ x_1, \cdots, x_r) = c\cdot \prod_{j=1}^{r} \cdot f_j [F_s - F_r]^{s-r-1} f_s [1-F_s]^{n-s}$$

$(0 \leqq x_1 \leqq \cdots \leqq x_r \leqq x_s)$, where $f_j = f(x_j;\boldsymbol{\beta})$, $F_j = F(x_j;\boldsymbol{\beta})$ and $c = n!/[(s-r-1)!(n-s)!]$.

3.1. *The parameter known*

We will briefly examine conditions under which a unique MLP for X_s exists when $\boldsymbol{\beta}$ is known. Notice that by the continuity of f and F, L converges to zero both as x_s tends toward x_r from the right and as x_s tends toward b, from the left. Also, $L>0$ on $\boldsymbol{D}$. This means that if there exists a unique solution, x_s^*, of the likelihood equation $\partial \log L/\partial x_s = 0$, then X_s^* must be the unique MLP of X_s. Now, this likelihood equation may be written as

$$(1)\qquad \frac{\partial \log L}{\partial x_s} = f_s\left[\frac{f_s'}{f_s^2} + \frac{(s-r-1)}{F_s - F_r} - \frac{n-s}{1-F_s}\right] = 0\ .$$

Notice from (1) (or from the Markov property of order statistics) that when $\boldsymbol{\beta}$ is known, the MLP of X_s, if it exists, is a function of X_r and the known value of $\boldsymbol{\beta}$.

We consider three cases separately:
(i) $r+1<s<n$, (ii) $s=n$ and (iii) $s=r+1$.

Case (i). $r+1<s<n$: In this case, the function

$$(s-r-1)/(F_s - F_r) - (n-s)/(1-F_s)$$

(viewed as a function of x_s on $[x_r, b)$) is continuous, decreasing, converges to $+\infty$ as x_s approaches x_r from the right and converges to $-\infty$ as x_s approaches b from the left. Thus, from the likelihood equation (1), we see that a unique MLP for X_s exists if

$$f'/f^2 = -\partial(1/f)/\partial x$$

is non-increasing on $[x_r, b)$.

Remark 2. The following is a list of some common pdf's for which the above conditions are satisfied: exponential; gamma with shape parameter $\geqq 1$; Weibull with shape parameter $\geqq 1$; logistic; normal; half-normal; Student's t; Cauchy; Pareto; and power function (i.e. $F(x)=x^{\nu}$, $\nu \geqq 1$, $0 \leqq x \leqq 1$; $\nu=1$ gives the uniform distribution). If f is a PF_2 ([3], p. 76) density for which f'' exists, then $\partial(1/f)\partial x$ is non-decreasing. Hence for such densities, a unique MLP for X_s exists. The converse is not true since the Cauchy is not a PF_2 density even though it produces a unique MLP for X_s.

Case (ii). $s=n$: By similar reasoning, we find that a sufficient condition for the existence of a unique MLP for X_n is that f'/f^2 decrease to $-\infty$ in addition to the other conditions already assumed. This added condition is met by all the distributions listed in Remark 2 with the exception of the power-function family. That the condition f'/f^2 decrease to $-\infty$ is not necessary for the existence and uniqueness of an MLP is nicely illustrated by the uniform distribution (cf. Example 4.2).

Case (iii). $s=r+1$: Again by reasoning similar to that in Case (i), we find that a unique MLP for X_{r+1} exists if f'/f^2 is decreasing on $[x_r, b)$ and if $f_r'/f_r \geqq (n-r-1)/[1-F_r]$. That these conditions are not necessary is again illustrated by the uniform distribution.

Another justification for maximum likelihood prediction of order statistics is the fact that in the case were X_r and X_s are *sample quantiles*, with r/n and s/n converging to π_1 and π_2 respectively, with increasing n $(0<\pi_1<\pi_2<1)$, maximum likelihood prediction and best (minimum mean square error) unbiased prediction of X_s are equivalent in large samples. This follows from the joint asymptotic normality of X_r and X_s.

3.2. *The parameter unknown*

If, as is expected in practice, the parameter is not known, the PLF may be maximized by standard means (if a maximum exists) and the MLE of X_s determined along with the PMLE of β.

Remark 3. If f is a location-scale parameter family, so that f takes the form

$$f(x;\mu,\beta)=\frac{1}{\beta}g\left(\frac{x-\mu}{\beta}\right),$$

and if μ and β are assumed known for the moment, then in the sample quantile situation mentioned above, the asymptotically best predictor of X_s takes the simple form

$$(2) \qquad X_s' = X_r + (1-p)\cdot\mu + (u - u^*\cdot p)\cdot\beta$$

(Kaminsky and Nelson [10]), where $p = [(1-\pi_1)/(1-\pi_2)]g(u_1)/g(u_2)$, $\pi_i = G(u_i)$ $(i=1, 2)$ and where $G' = g$. Now, it is often the case that explicit expressions for the PMLE's of unknown μ and β are unattainable. An alternative to numerical solution of the likelihood equations is to substitute *any* reasonable estimates μ' and β' into (2). The MSE of X_s' is then known once the variances and covariances among X_r, μ' and β' are known. For example, substituting the asymptotically best linear unbiased estimates of μ and β (eg. Chapter 4, [14]) in (2) produces the asymptotically best linear unbiased predictor of X_s (Kaminsky and Nelson [10]).

4. Examples involving order statistics

Example 4.1 (*The exponential distribution*). Consider a random sample of size n from a two-parameter exponential population with pdf $f(x; \mu, \beta) = (1/\beta)\exp\{-(x-\mu)/\beta\}$, $x \geqq \mu$, μ real, $\beta > 0$. For the sake of brevity, we will cover only the case of unknown scale parameter β, and without loss of generality, we take $\mu = 0$. The case when both parameters are unknown is handled with little added difficulty. The log PLF of X_s and β is proportional to

$$-(r+1)\log(\beta) - \left\{\sum_{j=1}^{r} x_j + (n-s+1)x_s\right\}$$
$$+ (s-r-1)\log\{\exp(-x_r/\beta) - \exp(-x_s/\beta)\} .$$

As the reader may verify, this function has a unique maximum relative to x_s and β, for any s such that $r+1 \leqq s \leqq n$. Interestingly however, when $s = r+1$, the maximum occurs on the boundary of the region where L is positive. The MLP and PMLE are

$$X_s^* = X_r + \beta^*\cdot q , \qquad \beta^* = T_r/(r+1) ,$$

where $q = \log[(n-r)/(n-s+1)]$, and where T_r, the well-known *total time on test* of all n items up to time x_r, is equal to

$$T_r = \sum_{j=1}^{r} X_j + (n-r)X_r .$$

The mean square error of X_s^* is

$$\text{MSE}(X_s^*) = \beta^2\{d_2(r, s) + d_1^2(r, s) + [r/(r+1)]q[q - 2d_1(r, s)]\} ,$$

where $d_m(r, s) = \sum_{j=r+1}^{s} (n-j+1)^{-m}$ $(m=1, 2)$.

Now, X_s^* is a biased predictor of X_s and β^* is a biased estimator of β. These biases are $\mathrm{E}(X_s - X_s^*) = \beta\{d_1(r, s) - rq/(r+1)\}$ (see Table 1) and $\mathrm{E}(\beta - \beta^*) = \beta/(r+1)$ respectively. An interesting problem would be to determine conditions under which the MLP is unbiased. We will compare X_s^* with the *best unbiased predictor* of X_s (notice that it is also linear),

$$X_s^{**} = X_r + \beta^{**} d_1(r, s) .$$

Its MSE is

$$\mathrm{MSE}\,(X_s^{**}) = \beta^2\{d_2(r, s) + d_1^2(r, s)/r\}$$

(cf. Kaminsky and Nelson [10], and Kaminsky [9]), where $\beta^{**} = T_r/r = (r+1)\beta^*/r$ is the best linear unbiased estimator (BLUE) of β. We have made extensive numerical comparisons between the MSE's of X_s^* and X_s^{**} for assorted values of r, s and n and we have found the following. Evidently, the ratio

$$\mathrm{EFF}\,(X_s^*, X_s^{**}) = \mathrm{MSE}\,(X_s^*)/\mathrm{MSE}\,(X_s^{**})$$

tends to be near unity if s is near $2r$; less than unity if $s<2r$ or if s is near n, and greater than unity if $2r<s<n$. In other words, the MLP is sometimes better, sometimes not as good, and sometimes about the same as the best unbiased predictor. A small table illustrating these observations is given below. In the sequel, we will call the above ratio the *efficiency* of X_s^* relative to X_s^{**}.

Table 1. Efficiency and bias for the MLP for assorted r, s and n.

r	s	n	EFF	BIAS (X_s^*/β)
3	6	15	0.9986	0.1375
15	17	40	0.7111	0.0434
2	9	10	1.2642	0.7937
10	12	100	0.7333	0.0122
10	20	100	0.9998	0.0213
10	50	100	1.0721	0.0670
500	1020	2000	1.0000	0.0017

Example 4.2 (*The uniform distribution*). Consider a random sample of size n from a uniform distribution with pdf $f(x;\beta) = 1/\beta$, $0 \leqq x \leqq \beta$. The log PLF here takes the form

$$\log L = \log\,(\text{constant}) - s\log\,(\beta) + (s-r-1)\log\,(x_s - x_r) + (n-s)\log\,(1 - x_s/\beta) .$$

It is not difficult to show that for any s satisfying $r+1 \leqq s \leqq n$, a unique MLE and PMLE exist for X_s and β respectively (when $s=r+1$ or n, the unique maximum occurs on the boundaries of the region where L is positive). We have

$$X_s^* = sX_r/(r+1) , \qquad \beta^* = nX_r/(r+1)$$

and

$$\mathrm{MSE}\,(X_s^*) = s(s-r+1)\beta^2/[(r+1)(n+1)(n+2)^2] .$$

For the sake of comparison, the *best linear unbiased predictor* (BLUP) of X_s is found to be (cf. Kaminsky and Nelson [10])

$$X_s^{**} = sX_r/r .$$

We omit the derivation. The BLUE of β is

$$\beta^{**} = (n+1)X_r/r$$

(Sarhan and Greenberg [14], p. 389). Straightforward calculations lead to the MSE of X_s^{**}

$$\mathrm{MSE}\,(X_s^{**}) = s(s-r)\beta^2/[r(n+1)(n+2)] ,$$

and so the efficiency of X_s^* relative to X_s^{**} is

$$\mathrm{EFF}\,(X_s^*, X_s^{**}) = (r+1)(s-r)/[r(s-r+1)] .$$

It is clear that

$$\mathrm{EFF}\,(X_s^*, X_s^{**}) \begin{cases} >1 & \text{if } s>2r , \\ =1 & \text{if } s=2r , \\ <1 & \text{if } s<2r . \end{cases}$$

Thus, as with the exponential distribution, the MLP can be better than, as good as, or not as good as the BLUP of X_s.

5. Other examples

Example 5.1 (*The Poisson process*). Let $\{N_t;\ t \geqq 0\}$ be a homogeneous Poisson process with arrival rate $1/\beta$, and let $\{X_j;\ j=0, 1, 2, \cdots\}$ be the associated arrival process (with $X_0=0$). As before, we wish to predict X_s from the first r arrival times X_j $(j=1, 2, \cdots, r)$. By exploiting the independence and identical exponential distributions of the interarrival times $X_j - X_{j-1}$ $(j=1, 2, 3, \cdots)$, it is not difficult to deduce the PLF of X_s and β. It is

$$L(x_s, \beta;\ x_1, \cdots, x_r) = \frac{(x_s - x_r)^{s-r-1} \exp(-x_s/\beta)}{\beta^s \Gamma(s-r)}$$

$0 \leqq x_r \leqq x_s$. If $s \geqq r+1$ and β is unknown, the unique MLP and PMLE are found to be

$$X_s^* = sX_r/(r+1)$$

and

$$\beta^* = X_r/(r+1) ,$$

the maximum occuring along the boundary $x_s = x_r$ in the special case when $s = r+1$.

Both X_s^* and β^* are biased. The biases are

$$\mathrm{E}\,(X_s - X_s^*) = s\beta/(r+1)$$

and

$$\mathrm{E}\,(\beta - \beta^*) = \beta/(r+1) .$$

The BLUP of X_s is easily calculated with the help of Kaminsky and Nelson [10]. It is

$$X_s^{**} = X_r + (s-r)\beta^{**} = sX_r/r ,$$

where β^{**}, the BLUE of β is

$$\beta^{**} = X_r/r .$$

The respective MSE's of X_s^* and X_s^{**} are

$$\mathrm{MSE}\,(X_s^*) = s(s-r+1)\beta^2/(r+1)$$

and

$$\mathrm{MSE}\,(X_s^{**}) = s(s-r)\beta^2/r ,$$

so that the efficiency of X_s^* relative to X_s^{**} is

$$\mathrm{EFF}\,(X_s^*, X_s^{**}) = (s-r)(r+1)/\{r(s-r+1)\} . \tag{3}$$

Notice that this expression is identical to that for the uniform distribution (Example 4.2) and so the same comments apply for the Poisson process.

Example 5.2 (*Occurrence of record values*). Let $X_1, X_2, X_3, \cdots$ denote a sequence of i.i.d. random variables with pdf $f(x;\beta) = (1/\beta)\exp(-x/\beta)$. (The two-parameter case gives very slightly different results, and is omitted for the sake of brevity.) We call X_i an *upper record* *value* of the sequence provided that $X_i > \max\{X_1, X_2, \cdots, X_{i-1}\}$. Let $L(1), L(2), \cdots$ denote the indices at which the records occur. That is, $L(1)=1$ (by convention), $L(n)=\min\{i; X_i > X_{L(n-1)}\}$ $(n \geqq 2)$.

The problem is to predict $X_{L(s)}$ having observed the first r record values $X_{L(j)}$ $(j=1, 2, \cdots, r)$. But, it is well-known (see for example Resnick [13], p. 69; Nagaraja in David [4], pp. 31–32) that the times between successive records $X_{L(j)} - X_{L(j-1)}$ are i.i.d. with pdf $f(x; \beta)$ as above, in the exponential case. In other words, the sequence of record values constitutes the arrival times in a Poisson process with arrival rate $1/\beta$. Thus, the results of Example 5.1 apply giving the MLP of $X_{L(s)}$ and PMLE of β as

$$X^*_{L(s)} = sX_{L(r)}/(r+1)$$

and

$$\beta^* = X_{L(r)}/(r+1) .$$

The efficiency of $X^*_{L(s)}$ relative to $X^{**}_{L(s)}$, the BLUP of $X_{L(s)}$ is again given by (3).

Example 5.3 (*The multivariate normal distribution*). Let $\boldsymbol{Y}^T = (\boldsymbol{Y}_1^T, \boldsymbol{Y}_2^T)$ be a $1 \times (n_1 + n_2)$ $(n = n_1 + n_2)$ random vector with n-variate normal distribution given by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{Y}_i$ is $n_i \times 1$ $(i=1, 2)$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned in the natural way into $\boldsymbol{\mu}^T = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} .$$

Suppose further that $\boldsymbol{\mu}$ can be written as

$$\boldsymbol{\mu} = \boldsymbol{X\beta} = \begin{bmatrix} \boldsymbol{x}_1\boldsymbol{\beta} \\ \boldsymbol{x}_2\boldsymbol{\beta} \end{bmatrix}$$

where $\boldsymbol{X}$ is $n \times p$ $(p < n)$ of rank p, and where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters. That is, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ constitutes a generalized linear regression model of full rank with correlated normal errors. Suppose that the data consist of $\boldsymbol{Y}_1$ and we wish to predict $\boldsymbol{Y}_2$. As mentioned earlier, $\boldsymbol{Y}_2$ may equally well represent a vector of "missing values", so that the experiment which was to produce all of $\boldsymbol{Y}$ may have produced n_1 real observations and n_2 unobserved ones. As is well-known even without the assumption of normality, the BLUP of $\boldsymbol{Y}_2$ is (cf. Goldberger [6], Whittle [15])

$$\boldsymbol{Y}_2^{**} = \boldsymbol{X}_2\boldsymbol{\beta}^{**} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{X}_1\boldsymbol{\beta}^{**}) ,$$

where $\boldsymbol{\beta}^{**}$ is the BLUE of $\boldsymbol{\beta}$ based on $\boldsymbol{Y}_1$ (Aitken [2]),

$$\boldsymbol{\beta}^{**}=(\boldsymbol{X}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{Y}_1 .$$

The MSE of $\boldsymbol{Y}$ is

$$\begin{aligned}\mathrm{MSE}\,(\boldsymbol{Y})=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}+(\boldsymbol{X}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1)\\ \times(\boldsymbol{X}_1^T\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1)^{-1}(\boldsymbol{X}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1)^T .\end{aligned}$$

Of course, if we use the normality of $\boldsymbol{Y}$, it is clear that the MLP and BLUP of $\boldsymbol{Y}$ are identical, as are the MLE and PMLE of $\boldsymbol{\beta}$, since the quadratic form (cf. Whittle [15], p. 53)

$$\begin{aligned}Q=(\boldsymbol{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})=(\boldsymbol{Y}_1-\boldsymbol{X}_1\boldsymbol{\beta})^T\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_1-\boldsymbol{X}_1\boldsymbol{\beta})\\ +\boldsymbol{Z}^T(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{Z} ,\end{aligned}$$

where $\boldsymbol{Z}=\boldsymbol{Y}_2-\boldsymbol{X}_2\boldsymbol{\beta}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_1-\boldsymbol{X}_1\boldsymbol{\beta})$ is to be minimized in both cases.

6. Concluding remarks

We have used the principle of maximum likelihood to derive predictors of random variables based on past observations. The likelihood function is maximized relative to variation in the future random variables as well as to the unknown parameters. We have seen, with the help of several examples, that the method produces reasonable results. Yet, there are some questions which remain unanswered at this time. Under what general conditions can the MLP for $\boldsymbol{Y}$ based upon $\boldsymbol{X}$ be shown to be unbiased, consistent, or efficient (in the sense of attaining the smallest possible MSE)? In general, how does the PMLE for $\boldsymbol{\beta}$ compare with the ordinary MLE based on the usual likelihood function $f(\boldsymbol{x};\boldsymbol{\beta})$? We hope to provide answers to these questions in the near future.

UNIVERSITY OF UMEÅ, SWEDEN

REFERENCES

[1] Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*, Cambridge University Press, Cambridge.

[2] Aitken, A. C. (1933). On least squares and linear combinations of observations, *Proc. Roy. Soc. Edin.*, **55**, 42-48.

[3] Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, Inc.

[4] David, H. A. (1981). *Order Statistics*, 2nd edition, John Wiley and Sons, New York.

[5] Faulkenberry, G. D. (1973). A method of obtaining prediction intervals, *J. Amer. Statist. Ass.*, **68**, 433-435.

[6] Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model, *J. Amer. Statist. Ass.*, **57**, 369-375.

[7] Hinkley, D. (1979). Predictive likelihood, *Ann. Statist.*, **7**, 718-728.

[8] Kaminsky, K. S. (1973). Maximum likelihood prediction (abstract), *Inst. Math. Statist. Bull.*, **2**, 253-254.

[9] Kaminsky, K. S. (1977). Best prediction of exponential failure times when items may be replaced, *Austral. J. Statist.*, **19**, 61-62.

[10] Kaminsky, K. S. and Nelson, P. I. (1975). Best linear unbiased prediction of order statistics in location and scale families, *J. Amer. Statist. Ass.*, **70**, 145-150.

[11] Kaminsky, K. S. and Rhodin, L. S. (1984). Maximum Likelihood Prediction, *Statistical Research Report 1984-1*, University of Umeå.

[12] Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models, *Scand. J. Statist.*, **1**, 128-134.

[13] Resnick, S. I. (1973). Limit laws for record values, *J. Stochastic Processes and their Appl.*, **1**, 67-82.

[14] Sarhan, A. E. and Greenberg, B. G. (1962). *Contributions to Order Statistics*, John Wiley and Sons, New York.

[15] Whittle, P. (1963). *Prediction and Regulation*, D. Van Nostrand Co., Inc., Princeton.