ON THE BIAS IN ESTIMATING GENETIC LENGTH AND OTHER QUANTITIES IN SIMPLEX CONSTRAINED MODELS

BY ARTHUR COHEN[1,2], J. H. B. KEMPERMAN AND HAROLD SACKROWITZ[1,2]

Rutgers University

The genetic distance between two loci on a chromosome is defined as the mean number of crossovers between the loci. The parameters of the crossover distribution are constrained by the parameters of the distribution of chiasmata. Ott (1996) derived the maximum likelihood estimator (MLE) of the parameters of the crossover distribution and the MLE of the mean. We demonstrate that the MLE of the mean is pointwise less than or equal to the empirical mean number of crossovers. It follows that the MLE is negatively biased. For small sample sizes the bias can be nonnegligible. We recommend reduced bias estimators.

Generalizations to many other problems involving linear constraints on parameters are made. Included in the generalizations are a variety of problems involving simplex constraints as studied recently by Liu (2000).

1. Introduction. The probabilistic relationship between chiasmata and crossovers dates back to Mather (1933). This relationship assumes that at each chiasma a crossover occurs with probability $1/2$. Thus, for a given number $C = c$ of chiasmata on a chromosome, the number K of crossover points on a gamete follows a binomial $(c, 1/2)$ distribution. The distribution of K can be obtained from the distribution of C by

$$(1.1) \qquad P(K = k) = \sum_{c=0}^{N} P(k|c) P(C = c),$$

where N is the maximum number of chiasmata that can occur. Thus, if we let p_k denote $P(K = k)$, $k = 0, 1, \ldots, N$, and let q_c denote $P(C = c)$, $c = 0, 1, \ldots, N$, we have that $\mathbf{p} = (p_0, p_1, \ldots, p_N)'$ and $\mathbf{q} = (q_0, \ldots, q_N)'$ are related by

$$(1.2) \qquad \mathbf{p} = A\mathbf{q},$$

where A is an $(N + 1) \times (N + 1)$ triangular matrix with (k, c) element $\binom{c}{k}(1/2)^c$ if $k \leq c$, and 0 if $k > c$. Mather (1938) is credited with defining the genetic distance between two loci on a chromosome as the average (mean) number of

crossover points (per gamete) between them. Ott (1996) used this definition and the relation (1.2) to derive the maximum likelihood estimators (MLE's) of the parameters of the crossover distribution $\mathbf{p}$, the chiasma distribution $\mathbf{q}$ and the mean number of crossovers μ (i.e., the genetic distance). These estimators are based on samples of size n, say, and on frequencies x_k, $k = 0, \ldots, N$, of k crossovers. The relationship (1.2) implies restrictions on the components of $\mathbf{p}$ since $0 \leq q_c \leq 1$ for each $c = 0, 1, \ldots, N$. Thus the empirical frequencies, x_k/n, as potential estimators of p_k, may correspond to an invalid set of estimators of the elements of $\mathbf{q}$. Hence the empirical relative frequencies are not the MLE's of the p_k's. Ott (1996) implicitly and realistically did not assume N is known. He proposed an iterative method, equivalent to the EM algorithm, for obtaining MLE's of crossover distributions via MLE's of $\mathbf{q}$. Using form (1.2), the current setting is seen to belong to a larger class of important models, some of which were recently discussed in Liu (2000). We will return to this issue later.

In this paper we offer the following:

1. Let N^* be the largest index i such that $x_i > 0$. Let $\hat{N}^*$ be the largest index i such that $\hat{p}_i > 0$, where $\hat{p}_i$ is the MLE of p_i. Note that $\hat{N}^* \geq N^*$. We demonstrate that the MLE of $\mu = \sum_{i=0}^{N} i p_i$, denoted by $\hat{\mu} = \sum_{i=0}^{\hat{N}^*} i \hat{p}_i$, is pointwise less than or equal to the empirical mean denoted by $\mu_E = \sum_{i=0}^{N^*} i x_i / n$. This implies that the MLE is a negatively biased estimator of μ. We will also demonstrate numerically that this bias can sometimes be nonnegligible. This result (which we will see to be true in much greater generality) is unexpected. See, for example, Ott's (1996) comment that the MLE of genetic length is sometimes larger and sometimes smaller than the empirical mean.

2. We offer an alternative method of estimating $\mathbf{p}$, μ and $\mathbf{q}$. The alternative method has the following advantages:

   a. The estimator of μ, call it $\tilde{\mu}$, is pointwise greater than or equal to $\hat{\mu}$. Despite the fact that $\tilde{\mu} \leq \mu_E$ pointwise, its bias is considerably less than that of $\hat{\mu}$.

   b. For N known, $\tilde{\mu}$ is easier to compute than $\hat{\mu}$. Note Mather (1933) assumed N is known when studying *Drosophila melanogaster*. He claimed that $N = 4$ (even though $N = 5$ might be possible for some abnormal species).

   c. When N is known, the estimators we propose not only have less bias, but also appear to have slightly better mean square error properties. The latter claim is based on numerical work.

3. We generalize the preceding results for this genetics problem in several directions.

   a. Our results carry over to the case where the (k, c) element $\binom{c}{k} 2^{-c}$ of the matrix A is replaced by $\binom{c}{k} \pi^k (1 - \pi)^{c-k}$. Here $0 < \pi < 1$. Analogous results hold for the columnwise stochastic matrix A whose (k, c) element is $1/(c + 1)$ if $k \leq c$ for $c = 0, 1, \ldots, N$.

   b. Let $\mathbf{u} = (u_0, u_1, \ldots, u_N)'$ be an $(N + 1) \times 1$ vector of constants and let $\mathbf{x} = (x_0, x_1, \ldots, x_N)$ be the vector of observed frequencies. We find sufficient

conditions under which the MLE of $\mathbf{u}'\mathbf{p}$ is pointwise less than or equal to the empirical estimator $\mathbf{u}'\mathbf{x}/n$. Of course, $\mu = \mathbf{u}'\mathbf{p}$ for $\mathbf{u} = (0, 1, 2, \ldots, N)'$ is a special case.

c. Further generalizations involving different matrices A will be indicated.

4. Comparable results are obtainable for a wide variety of models dealing with discrete distributions with simplex constraints. Liu (2000) discussed a variety of problems in which an $m \times 1$ vector of parameters $\boldsymbol{\theta}$, say, is related to another $m \times 1$ vector of parameters $\boldsymbol{\alpha}$, say, by

$$\text{(1.3)} \qquad\qquad \boldsymbol{\theta} = A\boldsymbol{\alpha},$$

where A is an $m \times m$ known matrix of nonnegative elements and the components of $\boldsymbol{\alpha}$ are nonnegative. Here m is known. In particular, suppose $\boldsymbol{\theta}$ in (1.3) represents a vector of means of independent Poisson variables. Or suppose $\boldsymbol{\theta}$ is a vector of multinomial probabilities. Then analogs of our general theorems are developed. For some of the applications in Liu (2000), we give sufficient conditions for a vector $\mathbf{u}$ such that the maximum likelihood estimators of $\mathbf{u}'\boldsymbol{\theta}$ are pointwise greater than or equal to the usual unbiased estimators of $\mathbf{u}'\boldsymbol{\theta}$. For example, in one of Liu's examples the MLE of the average age at death exceeds the unbiased estimator of the average age at death.

Now we define the notion of reversal for an estimator of the mean. Suppose there exist $\mathbf{x}$ and $\mathbf{y}$ such that $\mu_E(\mathbf{x}) \le \mu_E(\mathbf{y})$, where $\mu_E(\cdot)$ is the empirical mean. If $\hat{\mu}(\mathbf{x}) > \hat{\mu}(\mathbf{y})$, then we say the MLE reverses. From a practical point of view, reversing might be regarded as a somewhat negative property. We note, in fact, that the MLE of the mean reverses for the genetics model and for many of the other models in this paper. For the case of known N, the estimator we propose does not reverse.

In Section 2 we pose an abstracted mathematical problem in a fairly general form. The underlying distribution of the data is multinomial. General mathematical results are developed with proofs deferred to the Appendix. The claimed results regarding the bias of the MLE of the mean of the multinomial distribution will fall out as a special case of the general results. This will be given in Section 3 where we also discuss an alternative estimation procedure for $\mathbf{p}$, $\mathbf{q}$ and μ. The alternative estimator $\tilde{\mu}$ of μ satisfies $\mu_E \ge \tilde{\mu} \ge \hat{\mu}$ and so its bias will be less than that of $\hat{\mu}$. The rationale for the alternative estimation procedure is discussed. In Section 4 we discuss generalizations. Section 5 contains numerical work which gives some indication of the magnitude of the bias of $\hat{\mu}$ and $\tilde{\mu}$. Furthermore, mean square errors are presented for both the genetics model and the Poisson model.

2. General mathematical problem. Let C be a random variable taking values in $\{0, 1, \ldots, N\}$ with probability $P(C = j) = q_j$, $j = 0, 1, \ldots, N$. Furthermore, let K be a random variable taking values in $\{0, 1, \ldots, M\}$ and put

$P(K = i) = p_i$, $i = 0, 1, \ldots, M$. Let $\mathbf{p} = (p_0, p_1, \ldots, p_M)'$, $\mathbf{q} = (q_0, \ldots, q_N)'$ and assume

$$(2.1) \qquad \mathbf{p} = A\mathbf{q},$$

where A is a known $(M + 1) \times (N + 1)$ matrix such that each of the columns of A represents a discrete probability distribution (i.e., $\sum_{i=0}^{M} a_{ij} = 1$, $a_{ij} \geq 0$, $j = 0, 1, \ldots, N$). Let S_M denote the M-dimensional simplex consisting of all vectors $\mathbf{w} = (w_0, w_1, \ldots, w_M)'$ with $w_i \geq 0$, $\sum_{i=0}^{M} w_i = 1$. Similarly define S_N. Based on data $\mathbf{x}$, we seek MLE's of $\mathbf{p}$ and $\mathbf{q}$. That is, we seek $\mathbf{p}$ maximizing

$$(2.2) \qquad f(\mathbf{p}) = \sum_{i=0}^{M} \beta_i \log p_i,$$

subject to $\mathbf{p} = A\mathbf{q}$ with $\mathbf{q} \in S_N$, where $\beta_i = x_i/n$. Note that maximizing the underlying multinomial likelihood is the same as maximizing $f(\mathbf{p})$. We recognize that β_i are the empirical estimators of p_i, which would be the MLE's of p_i if there were no restrictions as those imposed by (2.1). Maximizing (2.2) is equivalent to maximizing the concave function

$$(2.3) \qquad g(\mathbf{q}) = \sum_{i=0}^{M} \beta_i \log\left(\sum_{j=0}^{N} a_{ij} q_j \right),$$

subject to $\mathbf{q} \in S_N$. Denoting $\hat{\mathbf{q}}$ as the vector which maximizes $g(\mathbf{q})$, we have from (2.1) that

$$(2.4) \qquad \hat{\mathbf{p}} = A\hat{\mathbf{q}},$$

where $\hat{\mathbf{p}}$ is the vector which maximizes (2.2) subject to $\mathbf{p} = A\mathbf{q}$ for some $\mathbf{q} \in S_N$.

In this section N is considered to be a fixed known quantity. Since N is unknown in the genetics application, some additional argument will be needed in that case. This will be explained more thoroughly later.

Returning to maximizing (2.3), we define $g(\mathbf{q}) = -\infty$ when any index i exists such that $\beta_i > 0$ and $p_i = \sum_{j=0}^{M} a_{ij} q_j = 0$. Hence the MLE of $\hat{\mathbf{p}}$ will be such that $\hat{p}_i = \sum_{j=0}^{M} a_{ij} \hat{q}_j > 0$ whenever $\beta_i > 0$. Thus $g(\hat{q})$ is always a finite number. Furthermore, if $\hat{p}_i = \sum_{s=0}^{N} a_{is} \hat{q}_s = 0$, then $\beta_i = 0$ with probability 1. This allows us to define $y_i = \beta_i/\hat{p}_i$ when $\hat{p}_i > 0$ and for concreteness $y_i = 0$ whenever $\hat{p}_i = 0$. Thus

$$(2.5) \qquad y_i \geq 0 \quad \text{and} \quad \sum_{i=0}^{M} \hat{p}_i y_i = 1.$$

Now let $\mathbf{u} = (u_0, u_1, \ldots, u_M)'$ be a given vector in $\mathbb{R}^{M+1}$. We seek sufficient conditions in order that

$$(2.6) \qquad \hat{\mu} = \sum_{i=0}^{M} u_i \hat{p}_i \leq \sum_{i=0}^{M} \frac{u_i x_i}{n} = \sum_{i=0}^{M} u_i \beta_i = \mu_E$$

for every possible vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_M)'$.

Let

$$(2.7) \qquad \varepsilon_s := 1 - \sum_{i=0}^{M} y_i a_{is}, \qquad s = 0, 1, \ldots, N.$$

LEMMA 2.1. *A necessary and sufficient condition for the concave function* $g(\mathbf{q})$ *to be maximal at* $\hat{\mathbf{q}} \in S_N$ *is that*

$$(2.8) \qquad \varepsilon_s = 0 \quad \textit{if } \hat{q}_s > 0, \qquad \varepsilon_s \geq 0 \quad \textit{if } \hat{q}_s = 0, \ s = 0, 1, \ldots, N.$$

We restate condition (2.6) as

$$(2.9) \qquad \hat{\mu} = \sum_{i=0}^{M} \hat{p}_i u_i \leq \sum_{i=0}^{M} \hat{p}_i u_i y_i = \mu_E$$

each time y_i satisfy (2.5) and (2.8) [with the understanding that the ε_i depend on y_i exactly as in (2.7)].

THEOREM 2.2. *Let* $\hat{\mathbf{q}} \in S_N$ *be given and put* $\hat{\mathbf{p}} = A\hat{\mathbf{q}}$ *so that* $\hat{\mathbf{p}} \in S_M$. *A sufficient condition, in order that* (2.5) *and* (2.8) *together imply* (2.9), *is the existence of real numbers* z_s, $s = 0, 1, \ldots, N$, *satisfying both*

$$(2.10) \qquad z_s \geq 0 \qquad \textit{whenever } \hat{q}_s = 0$$

($z_s \in \mathbb{R}$ *is arbitrary when* $\hat{q}_s > 0$) *and*

$$(2.11) \qquad \sum_{s=0}^{N} a_{is} z_s = \hat{p}_i (\hat{\mu} - u_i) \qquad \textit{for all } i.$$

THEOREM 2.3. *For fixed* $\hat{q}$ (*and* $\hat{\mathbf{p}} = A\hat{\mathbf{q}}$) *satisfying* (2.5) *and* (2.8), *a sufficient condition for* (2.9) *is that there exist real numbers* ξ_s, $s = 0, 1, \ldots, N$, *satisfying*

$$(2.12) \qquad \xi_s \geq 0 \qquad \textit{whenever } \hat{q}_s = 0, \ s = 0, 1, \ldots, N,$$

and

$$(2.13) \qquad \sum a_{is} \xi_s = -\hat{p}_i u_i \qquad \textit{for all } i = 0, 1, \ldots, M.$$

THEOREM 2.4. *A sufficient condition, in order that* (2.9) *holds for each choice of* $\hat{\mathbf{q}} \in S_N$ *satisfying* (2.5) *and* (2.8), *is the existence of a matrix* $H = (h_{st} : s, t = 0, 1, \ldots, N)$ *of size* $(N + 1) \times (N + 1)$ *satisfying*

$$(2.14) \qquad h_{st} \leq 0 \qquad \textit{whenever } s \neq t, \ s, t = 0, 1, \ldots, N$$

and

$$(2.15) \qquad\qquad\qquad AH = DA,$$

where $D = \text{diag}(u_0, u_1, \ldots, u_M)$. *Thus* (2.13) *is equivalent to*

$$(2.16) \qquad \sum_{s=0}^{N} a_{is} h_{st} = u_i a_{it}, \qquad i = 0, 1, \ldots, M; \ t = 0, 1, \ldots, N.$$

The following is an important special case of Theorem 2.4.

THEOREM 2.5. *Suppose* $M = N$ *and further assume* A *is invertible. Thus* $H = (h_{st})$ *is uniquely determined by* (2.15) *as* $H = BDA$, *where* $B = (b_{ij}) = A^{-1}$. *That is,*

$$(2.17) \qquad\qquad h_{st} = \sum_{j=0}^{M} b_{sj} u_j a_{jt}, \qquad s, t = 0, 1, \ldots, M = N.$$

Consequently, a sufficient condition that (2.9) *holds is that numbers defined by* (2.17) *satisfy* $h_{st} \le 0$ *whenever* $s \ne t$.

3. Estimating genetic distance.   In this section we start by demonstrating that the MLE $\hat{\mu}$ of μ is pointwise less than or equal to the empirical mean μ_E. Since μ_E is an unbiased estimator of μ, it follows that $\hat{\mu}$ is negatively biased. In the next section we demonstrate numerically that sometimes this bias can be nonnegligible. In light of this we propose an alternative estimator of μ.

Now consider the model of the previous section with $M = N$.

THEOREM 3.1. *Let the* j*th column of* $A = (a_{ij})$ *correspond to the binomial distribution with parameters* j *and* π, *where* $\pi \in (0, 1)$ *is fixed. Thus*

$$(3.1) \quad a_{ij} = \binom{j}{i} \pi^i (1 - \pi)^{j-i}, \qquad i, j = 0, 1, \ldots, N; \ a_{ij} = 0 \quad \text{if } j < i.$$

Then (2.9) *holds, that is,* $\hat{\mu} \le \mu_E$.

COROLLARY 3.2. *Under the assumptions of Theorem* 3.1, *when* $\pi = 1/2$, $\hat{\mu} \le \mu_E$.

The conclusion of Corollary 3.2 is that for known N the MLE of the genetic distance between two loci on a chromosome is pointwise less than or equal to the empirical estimator. However, the more realistic model is where N is unknown. In this case the MLE's can be found by maximizing (2.2) or (2.3) for each fixed value of N. Since N must be greater than N^*, the value of the empirical estimator of μ is the same regardless of the value of any choice of N. Note further that $\hat{\mu} \le \mu_E$ for any $N \ge N^*$, regardless of which N is true. The conclusion is that $\hat{\mu} = \sum_{i=0}^{\hat{N}^*} i \hat{p}_i \le \mu_E$.

THEOREM 3.3. *Under the assumptions of Theorem* 3.1, $\hat{N}^* \leq [N^*/\pi] - 1$, *where* $[\cdot]$ *denotes the smallest integer greater than or equal to.*

REMARK 3.4. For the case $\pi = 1/2$, Theorem 3.3 is given in Yu and Feingold (2001).

REMARK 3.5. The result of Theorem 3.3 can be useful in the designing of algorithms for obtaining the MLE's. Ott (1996) used 20 dimensions as a "safe" bound to begin his algorithm. However, for his data $N^* = 6$ and $\pi = 1/2$ so that only $N^*/\pi = 12$ dimensions are needed.

In the Introduction we noted that analogous results to those in Section 2 hold for the columnwise stochastic matrix A whose (k, c) element is $1/(c+1)$ if $0 \leq k \leq c$. Namely, it follows easily from Theorem 2.5 that

$$\sum_{i=0}^{N} u_i \left(\frac{x_i}{n} \right) \geq \sum_{i=0}^{N} u_i \hat{p}_i,$$

whenever u_i is increasing in i. Equivalently, the distribution $(\hat{p}_0, \ldots, \hat{p}_N)$ is stochastically smaller than the distribution $(x_0, \ldots, x_N)/n$.

Actually, this result is known. For observe that $\mathbf{p} = A\mathbf{q}$ for some $q \in S_N$ holds iff $\mathbf{p} \in S_N$ in such a way that p_i decreases in i. Thus our MLE distribution $(\hat{p}_0, \ldots, \hat{p}_N)$ is nothing but the (discrete version of the) classical MLE of a decreasing density. And as is well known [see Robertson, Wright and Dykstra (1988), page 328], the graph of the associated d.f. $\hat{F}(x)$ is precisely the smallest concave majorant of the graph of the empirical d.f. $F_n(x)$. In particular, $\hat{F}(x) \geq F_n(x)$, showing that the distribution $(\hat{p}_0, \ldots, \hat{p}_N)$ is stochastically smaller than the distribution $(x_0, \ldots, x_N)/n$. For instance, the former has a smaller mean.

At this point we propose an alternative estimator for μ, $\mathbf{p}$ and $\mathbf{q}$. The rationale for our choice is based on several considerations. We want an estimator of μ, call it $\tilde{\mu}$, that is less biased than the MLE. We want estimators of $\mathbf{p}$ and $\mathbf{q}$ that are consistent with $\tilde{\mu}$. We want the estimator of μ to be pointwise greater than or equal to the MLE. We want estimators of $\mathbf{p}$, $\mathbf{q}$ and μ to be computationally feasible. Furthermore, we want an estimator of μ that does not reverse when used in the N known setting.

An estimator of $\mathbf{p}$ satisfying all these properties when N is unknown is the projection of the sample point $\mathbf{x}$ onto the space $S_{\hat{N}^*} \cap \{\mathbf{p} : \mu = \mu_E\}$.

The resulting estimator of μ is

(3.2) $$\tilde{\mu} = \min((\hat{N}^* - 1)/2, \mu_E).$$

For fixed N, replace $S_{\hat{N}^*}$ by S_N. The resulting estimator has all the desirable properties.

4. Generalizations. Liu (2000) discussed a variety of models in which $\mathbf{X}$ is a random vector of order $m \times 1$ with expectation parameter $\boldsymbol{\theta}$ and, as in (1.3), $\boldsymbol{\theta} = A\boldsymbol{\alpha}$, where A is an $m \times m$ matrix of nonnegative elements and $\boldsymbol{\alpha}$ is an unknown vector of parameters that are assumed nonnegative. Liu sought the maximum likelihood estimator of $\boldsymbol{\theta}$ under several different distributional assumptions on $\mathbf{X}$. In one particular example, $\mathbf{X} = (X_1, \ldots, X_m)'$ and X_i are independent Poisson variables with parameter θ_i. Furthermore, A is a matrix that reflects the fact that the θ_i are increasing and a convex function of age. Liu (2000) specified the A matrix as

$$(4.1) \qquad A = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 2 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \\ 1 & m-1 & (m-2) & \cdots & \cdots & 1 \end{pmatrix}.$$

If one seeks an estimator of $\psi = \mathbf{u}'\boldsymbol{\theta}$ for this model (e.g., $\mathbf{u}'\boldsymbol{\theta}$ could reflect the average age at death in Liu's example), then an analog to our Theorem 2.5 applies and is stated as follows: Let $H = (h_{st}) = A^{-1}DA$, where $D = \text{diag}(u_1, \ldots, u_m)$, and let $\hat{\psi} = \mathbf{u}'\hat{\boldsymbol{\theta}}$ denote the MLE of ψ.

THEOREM 4.1. *Suppose X_i, $i = 1, \ldots, m$, are independent Poisson variables with means θ_i satisfying $\boldsymbol{\theta} = A\boldsymbol{\alpha}$, where A has all nonnegative elements and $\alpha_i \geq 0$, $i = 1, \ldots, m$. If $h_{st} \geq 0$ for all $s \neq t$, then $\hat{\psi} \geq \mathbf{u}'\mathbf{X}$. It follows that $\hat{\psi}$ is a positively biased estimator of ψ.*

REMARK. Theorem 4.1 generalizes to the situation where one has $n_i \geq 1$ independent observations x_{ir} on $X_i \sim P(\theta_i)$, $i = 1, \ldots, m$. In fact, as is easily seen, Theorem 4.1 remains valid provided D is replaced by $D^* = \text{diag}(v_1, \ldots, v_m)$, where $v_i = u_i/n_i$, $i = 1, \ldots, m$. The conclusion would be $\hat{\psi} \geq \sum_{i=1}^m u_i \times (\sum_{r=1}^{n_i} x_{ir}/n_i)$.

Now we apply Theorem 4.1 to the case where A is given in (4.1). Call the inverse matrix of A, B and note

$$(4.2) \qquad B = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & & \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

The matrices A and B essentially appear in Cohen, Kemperman and Sackrowitz (1994), where a hypothesis-testing problem is considered.

Next we calculate H. We have

LEMMA 4.2.   *Given A as in* (4.1), *H is such that*

$$h_{ik} = (i - k)(u_i - 2u_{i-1} + u_{i-2}) + (u_i - u_{i-2}) \quad if \ 2 \le k < i,$$

(4.3) $h_{i1} = u_i - 2u_{i-1} + u_{i-2} \quad if \ i \ge 3, \qquad h_{21} = u_2 - u_1,$

$$h_{ii} = u_i \quad for \ all \ i \ge 1, \ h_{ik} = 0 \quad if \ i < k.$$

THEOREM 4.3.   *A sufficient condition for* $\hat{\psi} \ge \mathbf{u}'\mathbf{X}$ *is that the function* $i \to u_i$ *on* $\{1, 2, \ldots, m\}$ *be increasing* (*nondecreasing*) *and convex.*

As an application of Theorem 4.3, consider estimating the average age at death for Liu's example. If the $\mathbf{u}$'s (ages) are equally spaced and increasing, then clearly the MLE of the average age at death is pointwise greater than or equal to the usual unbiased estimator. One might consider an alternative estimator whose bias is less than that of the MLE. An alternative estimator could be one that is mean preserving and lies in the restricted space.

We note that the phenomenon of the bias of the MLE is typical in many order-restricted parameter problems. Nevertheless, the pointwise stochastic order of the estimators, the direction of the bias and the amount of bias are important issues. As another example, suppose the mean parameters $(\theta_1, \ldots, \theta_m)$ are simply nondecreasing. The distribution of X_i can be any one-parameter exponential family as in Robertson, Wright and Dykstra (1988), page 34. Then, using the PAVA algorithm, one can demonstrate that the MLE of $\psi = \mathbf{u}'\boldsymbol{\theta}$ is positively biased for any vector $\mathbf{u} = (u_1, \ldots, u_m)'$ such that $u_i \le u_{i+1}$, with strict inequality for some i. Similar results are obtainable for cones other than the simple-order cone determined by the nondecreasing θ_i. Included is when the parameters are in the tree order cone, that is, the one in which $\theta_1 \ge \theta_i$, $i = 1, \ldots, m$. One can use these methods to demonstrate that the MLE is a positively biased estimator of θ_1. Lee (1988) demonstrated that this bias can be substantial, especially if m is large. In fact, the bias tends to ∞ as $m \to \infty$.

The methods of this paper may also be suitable for the other models considered in Liu (2000) and for other order-restricted models in Robertson, Wright and Dykstra (1988).

5. Numerical bias and mean square error calculations.   In this section we study the models of Theorem 3.1 with $\pi = 1/2$ and the model of Section 4 with A given in (4.1). We offer numerical calculations, sometimes exact (dimension 3) and sometimes simulated (dimension > 3; 5000 iterations), of bias, relative bias, bias ratio and mean square error. The mean to be estimated is denoted by μ, the empirical mean by μ_E, the MLE by $\hat{\mu}$ and the reduced bias estimator by $\tilde{\mu}$. If T is an estimator, then its bias is $(ET - \mu)$, the relative bias is $|(ET - \mu)|/\mu$ and the

bias ratio is $(E\hat{\mu} - \mu)/(E\tilde{\mu} - \mu)$. The mean square error (MSE) is $E(T - \mu)^2$. The calculations are carried out for selected parameter points and some sample sizes. Parameter points, for the most part, are chosen to be on the boundary of the set to which the parameters are restricted. The order of the bias, as it relates to sample size n, should be higher on boundary points. This is the case in simpler problems when bias can be expressed in a closed form. In the problems of this paper, however, where the estimators do not have closed forms and require algorithms to compute, exact bias formulas are not available. When feasible, a cross section of parameter points is chosen on the boundary. Since, in the genetics model, the MLE is a "shrinker," some parameter points with small means are chosen and some with larger means are chosen. For the Poisson model, a similar selection of parameter points is made.

The numerical study consists of five cases. The first two cases are concerned with the genetic length model assuming N is known with $N = 4$ and $N = 6$. The next two cases are concerned with the Poisson model where there are three populations and five populations, respectively. The fifth case is concerned with the genetics model when N is assumed to be unknown.

The genetic length model is considered in Table 1, where N is assumed to be known and set equal to 4. Sample sizes 8 and 15 are offered. Expected values of $\hat{\mu}$, $\tilde{\mu}$, relative bias (RB) for $\hat{\mu}$, $\tilde{\mu}$, bias ratio (BR) and MSE for $\hat{\mu}$, $\tilde{\mu}$ are given. The entries are obtained by simulation carried out on 5000 iterations. We note that the bias in $\hat{\mu}$ is nonnegligible and is as high as 0.21 with a relative bias of 10%. Bias ratios can be very high. As expected the MSE for $\tilde{\mu}$ is smaller for larger values of μ ($\mu = 2, 1.75, 1.5$) and the MSE for $\hat{\mu}$ is smaller for smaller values of μ ($\mu = 1.25, 1, 0.5$). For the parameter points chosen, the difference in MSE, for $\hat{\mu}$ and $\tilde{\mu}$ is slightly higher when $\tilde{\mu}$ has the smaller MSE, thus favoring $\tilde{\mu}$ somewhat. In summary for this case, unless one had prior information that μ is relatively small, $\tilde{\mu}$ is the preferred estimator because it is much easier to compute, has less bias and has slightly better MSE properties.

The genetic length model is also considered in Table 2, where again N is assumed known and set equal to 6. Fewer parameter points are chosen and only a sample of size 8 is considered. In addition to bias and MSE calculations, we include a column denoting the proportion of times, say P, that $\tilde{\mu}$ differs from μ_E. One expects P to be large for μ large and small for μ smaller. The pattern in bias and MSE is similar to the case where $N = 4$, although we did not choose parameter points for which $\mu < 1.5$. However, for parameter points with $\mu \leq 1.25$, we predict a smaller MSE for $\hat{\mu}$. In Table 2 we note larger differences in the MSE for $\hat{\mu}$ and $\tilde{\mu}$, favoring $\tilde{\mu}$. As predicted P is somewhat large (0.44) when $\mu = 3$ and is down to 0 when $\mu = 1.5$. Our conclusion for this case is to prefer $\tilde{\mu}$ to $\hat{\mu}$.

The Poisson model for three populations is considered in Table 3. The quantity to be estimated is again labeled μ, where $\mu = \mathbf{u}'\boldsymbol{\nu}$, $\boldsymbol{\nu} = \boldsymbol{\theta}/\lambda$, $\boldsymbol{\theta}$ is the mean vector of the independent Poisson variables, $\lambda = \theta_1 + \theta_2 + \theta_3$ and $\mathbf{u}' = (0, 1, 2)$. The choice of $\mathbf{u}$ really reflects equal spacing of a covariate (say, age) because one could add a

TABLE 1
Genetic length model

$\mathbf{q}' = (0, 0, 0, 0, 1)$, $\mathbf{p}' = (0.0625, 0.2500, 0.3750, 0.2500, 0.0625)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	2	1.79	1.86	0.10	0.07	1.47	0.10	0.06
15	2	1.85	1.89	0.08	0.05	1.43	0.06	0.03

$\mathbf{q}' = (0, 0, 0, 0.5, 0.5)$, $\mathbf{p}' = (0.0938, 0.3125, 0.3750, 0.1875, 0.0312)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	1.75	1.62	1.71	0.07	0.03	2.83	0.11	0.08
15	1.75	1.67	1.73	0.05	0.01	4.85	0.06	0.05

$\mathbf{q}' = (0, 0, 0.5, 0, 0.5)$, $\mathbf{p}' = (0.1563, 0.3750, 0.3125, 0.1250, 0.0312)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	1.5	1.40	1.49	0.07	0.01	12.85	0.12	0.11
15	1.5	1.43	1.49	0.05	0.00	12.03	0.07	0.06

$\mathbf{q}' = (0, 0.5, 0, 0, 0.5)$, $\mathbf{p}' = (0.2813, 0.3750, 0.1875, 0.1250, 0.0312)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	1.25	1.15	1.24	0.08	0.01	14.03	0.14	0.14
15	1.25	1.18	1.25	0.06	0.00	24.64	0.08	0.08

$\mathbf{q}' = (0.2, 0.2, 0.2, 0.2, 0.2)$, $\mathbf{p}' = (0.3875, 0.3250, 0.2000, 0.0750, 0.0125)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	1	0.93	1.00	0.07	0.00	87.84	0.12	0.13
15	1	0.96	1.00	0.04	0.00	29.58	0.06	0.07

$\mathbf{q}' = (0.5, 0, 0.5, 0, 0)$, $\mathbf{p}' = (0.625, 0.250, 0.125, 0, 0)$

n	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$
8	0.5	0.47	0.50	0.05	0.00	16.80	0.05	0.06
15	0.5	0.48	0.50	0.04	0.00	23.16	0.03	0.03

Expected values, MSE, relative bias (RB) bias ratio (BR) for $\hat{\mu}$ and $\tilde{\mu}$; $N = 4$ known, sample size n; parameter point $\mathbf{q}' = (q_0, q_1, q_2, q_3, q_4)$, $\mathbf{p}' = (p_0, p_1, p_2, p_3, p_4)$.

TABLE 2
Genetic length model

$\mathbf{q}'$	μ	$E\hat{\mu}$	$E\tilde{\mu}$	RB$\hat{\mu}$	RB$\tilde{\mu}$	BR	MSE$\hat{\mu}$	MSE$\tilde{\mu}$	P
$(0, 0, 0, 0, 0, 0, 1)$	3	2.76	2.83	0.08	0.06	1.40	0.14	0.09	0.44
$(0, 0, 0, 0, 0, 0.5, 0.5)$	2.75	2.59	2.67	0.06	0.03	2.04	0.15	0.11	0.23
$(0, 0, 0, 0.25, 0.25, 0.25, 0.25)$	2.25	2.14	2.23	0.04	0.01	7.03	0.18	0.17	0.03
$(0, 0, 0, 1, 0, 0, 0)$	1.50	1.41	1.50	0.06	0.00	49.25	0.10	0.09	0.00

Expected values, MSE, RB, BR; $P \equiv$ probability $\hat{\mu}$ differs from μ_E; $N = 6$ known; sample size $n = 8$.

TABLE 3
Poisson model

ν'	μ	$E\hat{\mu}$	$E\tilde{\mu}$	$RB\hat{\mu}$	$RB\tilde{\mu}$	BR	$MSE\hat{\mu}$	$MSE\tilde{\mu}$
$(\frac{1}{12}, \frac{1}{12}, \frac{5}{6})$	1.75	1.78	1.75	0.02	0.00	153.38	0.04	0.04
$(0, \frac{1}{3}, \frac{2}{3})$	1.67	1.74	1.67	0.04	0.00	0.00	0.01	0.03
$(\frac{1}{6}, \frac{1}{6}, \frac{2}{3})$	1.50	1.55	1.51	0.03	0.00	9.42	0.06	0.07
$(\frac{1}{12}, \frac{1}{3}, \frac{7}{12})$	1.50	1.56	1.50	0.04	0.00	32.79	0.05	0.05
$(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$	1.33	1.40	1.35	0.05	0.01	4.66	0.06	0.06
$(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$	1.25	1.33	1.28	0.06	0.03	2.44	0.06	0.06
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	1.00	1.15	1.11	0.15	0.11	1.31	0.06	0.04

Expected values, MSE, RB, BR; $\nu' = (\theta'/\lambda) = (\nu_1, \nu_2, \nu_3)$; $S = 8$; $\mu = \mathbf{u}'\mathbf{v}$, $\mathbf{u}' = (0, 1, 2)$.

constant to all components of **u** and also multiply **u** by a constant to generate other relevant linear combinations of the θ_i's. If one is interested in a linear combination of the θ_i's and not the ν_i's, then one multiplies the estimator of μ by S, where $S = X_1 + X_2 + X_3$. (See Section 4.) Thus the bias and MSE entries in Table 3 (and Table 4) would need to be adjusted if one were interested in a linear combination of Poisson means. The bias would be S times the bias in the table. The MSE would be S^2 times the MSE in the table. The entries in Table 3 are based on exact calculations opposed to simulation.

For the Poisson model, the MLE is an "expander" (i.e., overestimates the mean) and so in terms of the MSE it should do better than $\tilde{\mu}$ for large values of μ and $\tilde{\mu}$ should do better for small values of μ. This is precisely the pattern in the MSE entries. In terms of bias, $\hat{\mu}$ has a relative bias as high as 15%.

Table 4 has simulated entries for the Poisson model for five populations and $S = 8$. The MLE has a relative bias as high as 15%. Furthermore, the MSE for $\hat{\mu}$ is quite a bit higher than the MSE for $\tilde{\mu}$ for smaller values of μ. For large values of μ, the MSE for $\hat{\mu}$ is better than that for $\tilde{\mu}$ but the discrepancy is not nearly as large as when $\tilde{\mu}$ is preferred.

TABLE 4
Poisson model

ν'	μ	$E\hat{\mu}$	$E\tilde{\mu}$	$RB\hat{\mu}$	$RB\tilde{\mu}$	BR	$MSE\hat{\mu}$	$MSE\tilde{\mu}$
$(\frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{1}{20}, \frac{16}{20})$	3.5	3.59	3.50	0.03	0.00	81.96	0.11	0.16
$(\frac{2}{20}, \frac{3}{20}, \frac{4}{20}, \frac{5}{20}, \frac{6}{20})$	2.5	2.69	2.53	0.08	0.01	5.56	0.24	0.17
$(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$	2	2.30	2.20	0.15	0.10	1.53	0.22	0.12

Expected values, MSE, RB, BR; $\nu' = (\nu_1, \nu_2, \nu_3, \nu_4, \nu_5)$; $S = 8$; $\mu = \mathbf{u}'\mathbf{v}$; $\mathbf{u} = (0, 1, 2, 3, 4)$.

Based on Tables 3 and 4, unless one had prior information that μ is large, $\tilde{\mu}$ is preferred to $\hat{\mu}$. It is much easier to compute and has superior bias and, in some sense, better MSE properties.

Table 5 is concerned with the genetic length model when N is unknown. This makes the evaluation complicated since parameter points can lie in different dimensional simplexes. For parameter points in small dimensional spaces, the MLE tends to make up for its underestimating μ by somewhat overestimating N. This helps it somewhat in MSE. Yet for parameter points in higher dimensions, it does not seem to overestimate N as much and thus its underestimation is not compensated for. In Table 5 we recognize, as expected, that the bias in the MLE is nonnegligible (relative bias as high as 9%). Furthermore, for μ large, $\tilde{\mu}$ has smaller

TABLE 5
Genetic length model

$n = 8$		$\hat{\mu}$			$\tilde{\mu}$			
$\mathbf{q}'$	μ	$E\{\}$	RB	MSE	$E\{\}$	RB	MSE	BR
$0, 0, 1, \ldots$	1.0	0.91	0.09	0.06	0.97	0.03	0.07	3.9
$0, 0, 0, 1, \ldots$	1.5	1.41	0.06	0.10	1.48	0.02	0.09	4.1
$0, 0, 0, 0, 0, 0, 1, \ldots$	3.0	2.91	0.03	0.19	2.97	0.01	0.18	3.0
$0.5, 0.5, \ldots$	0.25	0.25	0.00	0.02	0.25	0.00	0.02	1.0
$0.5, 0, 0.5, \ldots$	0.50	0.47	0.05	0.06	0.50	0.01	0.06	6.6
$0.5, 0, 0, 0.5, \ldots$	0.75	0.71	0.05	0.11	0.75	0.00	0.12	∞
$0, 0, 0.5, 0.5, \ldots$	1.25	1.16	0.07	0.09	1.23	0.02	0.09	4.6
$0, 0, 0.5, 0, 0.5, \ldots$	1.50	1.41	0.06	0.14	1.48	0.01	0.13	4.8
$0, 0.5, 0, 0, 0.5, \ldots$	1.25	1.19	0.05	0.15	1.24	0.00	0.16	12.0
$0, 0, 0, 0, 0.5, 0.5, \ldots$	2.25	2.16	0.04	0.16	2.22	0.02	0.15	2.9
$0, 0, 0, 0, 0, 0.5, 0.5, \ldots$	2.75	2.65	0.04	0.19	2.72	0.01	0.18	2.9
$0, 0, 0, 0.25, 0.25, 0.25, 0.25, \ldots$	2.25	2.16	0.04	0.19	2.23	0.01	0.18	3.5

$n = 15$		$\hat{\mu}$			$\tilde{\mu}$			
$\mathbf{q}'$	μ	$E\{\}$	RB	MSE	$E\{\}$	RB	MSE	BR
$0, 0, 1, \ldots$	1.0	0.92	0.08	0.03	0.98	0.02	0.03	3.7
$0, 0, 0, 1, \ldots$	1.5	1.43	0.05	0.05	1.49	0.01	0.05	5.0
$0, 0, 0, 0, 0, 0, 1, \ldots$	3.0	2.93	0.02	0.11	2.98	0.01	0.10	3.3
$0.5, 0.5, \ldots$	0.25	0.25	0.00	0.12	0.25	0.00	0.12	1.0
$0.5, 0, 0.5, \ldots$	0.50	0.48	0.04	0.03	0.50	0.00	0.04	∞
$0.5, 0, 0, 0.5, \ldots$	0.75	0.72	0.05	0.06	0.75	0.00	0.06	10.8
$0, 0.5, 0, 0, 0.5, \ldots$	1.25	1.21	0.03	0.08	1.29	0.00	0.08	30.5
$0, 0, 0.5, 0.5, \ldots$	0.75	0.71	0.05	0.03	0.74	0.02	0.03	2.7
$0, 0, 0.5, 0, 0.5, \ldots$	1.50	1.45	0.04	0.07	1.49	0.00	0.07	11.1
$0, 0, 0, 0, 0.5, 0.5, \ldots$	2.25	2.18	0.03	0.08	2.24	0.01	0.08	4.6
$0, 0, 0, 0, 0, 0.5, 0.5, \ldots$	2.75	2.69	0.02	0.10	2.73	0.01	0.10	4.0
$0, 0, 0, 0.25, 0.25, 0.25, 0.25, \ldots$	2.25	2.19	0.03	0.10	2.23	0.01	0.09	3.9

Expected values, MSE, RB, BR; N unknown; $n = 8, 15$.

MSE, while for μ small, $\hat{\mu}$ has smaller MSE. The differences in MSE between $\hat{\mu}$ and $\tilde{\mu}$, regardless of which MSE is bigger, appear to be small. In conclusion, for this case we can only assert that $\tilde{\mu}$ has less bias than $\hat{\mu}$. In terms of computation and MSE, neither estimator has an advantage.

APPENDIX: PROOFS

PROOF OF LEMMA 2.1. Since the function $g(q)$ on the simplex $S_N = \{q : q_s \geq 0 \ (s = 0, \ldots, N); \sum q_s = 1\}$ as defined by (2.3) is concave, it takes its maximal value at $q = \hat{q} \in S_N$ if and only if no strictly larger value $g(q)$ can be obtained by increasing a single $\hat{q}_s < 1$ by a small amount $\delta > 0$ while simultaneously decreasing a single $\hat{q}_t > 0$ by the same amount δ. This in turn is equivalent to the requirement that

$$\frac{\partial}{\partial q_s} g(\hat{q}) - \frac{\partial}{\partial q_t} g(\hat{q}) \leq 0 \qquad \text{if } \hat{q}_s < 1 \text{ and } \hat{q}_t > 0.$$

Equivalently, there exists a value σ such that

$$\frac{\partial}{\partial q_s} g(\hat{q}) = \sigma \quad \text{if } \hat{q}_s > 0, \qquad \frac{\partial}{\partial q_s} g(\hat{q}) \leq \sigma \quad \text{if } \hat{q}_s = 0.$$

Since, from (2.3),

$$\frac{\partial}{\partial q_s} g(\hat{q}) = \sum_{i=0}^{M} \beta_i \frac{a_{is}}{\hat{p}_i} = \sum_{i=0}^{M} y_i a_{is}, \qquad s = 0, 1, \ldots, N,$$

an equivalent condition is that, for some number σ,

$$\sum_{i=0}^{M} y_i a_{is} = \sigma \quad \text{if } \hat{q}_s > 0, \qquad \sum_{i=0}^{M} y_i a_{is} \leq \sigma \quad \text{if } \hat{q}_s = 0.$$

Multiplying by $\hat{q}_s$, summing over s and using $\sum a_{is} \hat{q}_s = \hat{p}_i$ and also (2.5), we see that $\sigma = 1$. Consequently, the above necessary and sufficient condition takes the form (2.8). □

REMARK. The necessary and sufficient condition in Lemma 2.1 is related to the Kuhn–Tucker condition in nonlinear programming. See Zangwill and Mond (1969).

PROOF OF THEOREM 2.2. It suffices to show that (2.5), (2.8), (2.10) and (2.11) imply (2.9). First, observe that (2.11) implies

$$(A.1) \qquad \sum_{s=0}^{N} z_s = 0.$$

Namely, sum (2.11) over i and use that $\sum_i^M a_{is} = 1$, $\sum \hat{p}_i = 1$. Let ε_s be as in (2.7). From (2.8) and (2.10), $z_s \varepsilon_s \geq 0$ for all s. Hence, using (A.1),

$$0 \geq \sum_{s=0}^{N} z_s(-\varepsilon_s) = \sum_{s=0}^{N} z_s \left(-1 + \sum_{i=0}^{M} y_i a_{is} \right) = \sum_{s=0}^{N} z_s \sum_{i=0}^{N} y_i a_{is}$$

$$= \sum_{i=0}^{M} y_i \sum_{s=0}^{N} a_{is} z_s = \sum_{i=0}^{M} y_i \hat{p}_i (\hat{\mu} - u_i) = \hat{\mu} - \sum_{i=0}^{M} \hat{p}_i u_i y_i.$$

Here we used (2.5) and (2.11), as well as $\sum_{i=0}^{M} \hat{p}_i y_i = 1$. Thus $\hat{\mu} \leq \mu_E$. $\quad \square$

PROOF OF THEOREM 2.3.   If ξ_s satisfy (2.12) and (2.13), then numbers z_s satisfying (2.10) and (2.11) are given by

(A.2) $$z_s = \xi_s + \hat{\mu}\hat{q}_s, \qquad s = 0, 1, \ldots, N.$$

Namely, it is obvious that (2.10) and (2.12) are equivalent. Similarly, (2.11) and (2.13) are equivalent since

$$\sum_{s=0}^{N} a_{is} \hat{\mu} \hat{q}_s = \hat{\mu} \sum_{s=0}^{N} a_{is} \hat{q}_s = \hat{\mu} \hat{p}_i. \qquad \square$$

PROOF OF THEOREM 2.4.   Let H be as above. It suffices to show that conditions (2.12) and (2.13) of Theorem 2.3 hold for the numbers ξ_s defined by

(A.3) $$\xi_s = -\sum_{t=0}^{N} h_{st} \hat{q}_t, \qquad s = 0, 1, \ldots, N.$$

That (2.12) is true is obvious from (2.14) and $\hat{q}_t \geq 0$. Now (2.13) follows from

$$\sum_{s=0}^{N} a_{is} \xi_s = -\sum_{s=0}^{N} a_{is} \sum_{t=0}^{N} h_{st} \hat{q}_t = -\sum_{t=0}^{N} \hat{q}_t \sum_{s=0}^{N} a_{is} h_{st}$$

$$= -\sum_{t=0}^{N} \hat{q}_t u_i a_{it} = -\hat{p}_i u_i. \qquad \square$$

PROOF OF THEOREM 2.5.   This is a special case of Theorem 2.4. $\quad \square$

PROOF OF THEOREM 3.1.   We find $B = (b_{ij}) = A^{-1}$ as

(A.4) $$b_{ij} = (-1)^{j-i} \binom{j}{i} \pi^{-j} (1-\pi)^{j-i}, \qquad b_{ij} = 0, \text{ if } j < i.$$

From (2.17),

$$h_{st} = \sum_{j=0}^{N} b_{sj} u_j a_{jt} = \sum_{j=0}^{N} d(s, t, j) u_j,$$

where

$$d(s, t, j) = b_{sj} a_{jt} = (-1)^{j-s} \binom{j}{s} \pi^{-j} (1 - \pi)^{j-s} \binom{t}{j} \pi^j (1 - \pi)^{t-j}$$

$$= \binom{t}{s} (1 - \pi)^{t-s} (-1)^{j-s} \binom{t-s}{j-s}.$$

Hence the condition $h_{st} \leq 0$ whenever $s \neq t$ is equivalent to

$$\sum_{j=s}^{t} (-1)^{j-s} \binom{t-s}{j-s} u_j \leq 0 \qquad \text{whenever } 0 \leq s < t \leq N.$$

Letting $g = s - t$, $k = t - s$, this condition can be written as

$$(\text{A.5}) \quad \sum_{g=0}^{k} (-1)^g \binom{k}{g} u_{s+g} \leq 0 \qquad \text{whenever } 1 \leq k \leq N, \ 0 \leq S \leq N - k.$$

We recognize that the left-hand side of (A.5) is essentially a kth-order difference of the function $i \to u(i) = u_i$, where $i \in \{0, 1, \ldots, N\}$. [See, e.g., Anderson (1971).] In light of this, (A.5) can be expressed as

$$(\text{A.6}) \quad (-1)^k (\Delta^k u)(i) \leq 0 \qquad \text{whenever } 1 \leq k \leq N \text{ and } 0 \leq i \leq N - k,$$

where $\Delta u(i) = u_{i+1} - u_i$ and Δ^k represents a kth-order difference. Thus (A.6) requires that $(\Delta^k u)(i) \geq 0$ if k is odd, $k \geq 1$; $(\Delta^k u)(i) \leq 0$ if k is even, $k \geq 2$. For our problem $u_i = i$, $i = 0, 1, \ldots, N$, so that $(\Delta u)(i) = u_{i+1} - u_i = 1$ for all i; hence $(\Delta^k u)(i) = 0$ for all i and all $k \geq 2$. Thus (A.6) is trivially satisfied and we have $\hat{\mu} \leq \mu_E$. $\square$

PROOF OF THEOREM 3.3. For any fixed $\mathbf{x}$, the MLE's of $\mathbf{q}$ are obtained by finding values of N and $\mathbf{q}$ which maximize $g(\mathbf{q})$ as given in (2.3) with $M = N$. Some simple calculations show that $a_{i,j} \geq a_{i,(j+1)}$ for all $j \geq [i/\pi] - 1$. Thus

$$(\text{A.7}) \qquad a_{i,j} \leq a_{i,[N^*/\pi]-1} \qquad \text{for all } i = 0, 1, \ldots, N^*, \ j \geq [N^*/\pi].$$

Consider any choice of $\mathbf{q}$ and N such that $q_j > 0$ for some $j \geq [N^*/\pi]$. Based on this $\mathbf{q}$, define $\mathbf{r} = (r_1, \ldots, r_{[N^*/\pi]-1})$ by $r_j = q_j$, $j = 0, 1, \ldots, [N^*/\pi] - 2$, and $r_{[N^*/\pi]-1} = \sum_{j=[N^*/\pi]-1}^{N} q_j$. The theorem now follows as (A.7) implies that $g(\mathbf{q}) \leq g(\mathbf{r})$. $\square$

PROOF OF THEOREM 4.1. Let $S = \sum_{i=1}^{m} X_i$, $\lambda = \sum_{i=1}^{m} \theta_i$ and $\boldsymbol{\nu} = \boldsymbol{\theta}/\lambda$. The likelihood may then be expressed as the product of the conditional density of $\mathbf{X}|S$ as a function of $\boldsymbol{\nu}$ and the marginal density of S as a function of λ. The conditional

density of $\mathbf{X}|S$ is multinomial with parameter vector $\boldsymbol{\nu}$ satisfying

$$\text{(A.8)} \qquad\qquad \boldsymbol{\nu} = A\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = \boldsymbol{\alpha}/\lambda$ and $\beta_i \geq 0$, $i = 1, \ldots, m$. The marginal density of S is Poisson with parameter λ, $0 < \lambda < \infty$. As such the MLE of λ is S, since λ is unrestricted. Now A may be taken to be columnwise stochastic without loss of generality. That is, we may write $\boldsymbol{\nu} = G\boldsymbol{\gamma}$, where the elements of G are nonnegative and the sum of the elements of any column of G equals 1. To see this, write

$$\text{(A.9)} \qquad \begin{aligned} \boldsymbol{\nu} &= A\boldsymbol{\beta} \\ &= A\Delta^{-1}\Delta\boldsymbol{\beta} \\ &= G\boldsymbol{\gamma}, \end{aligned}$$

where $G = A\Delta^{-1}$, $\boldsymbol{\gamma} = \Delta\beta$, $\Delta = \operatorname{diag}(a_{.1}, a_{.2}, \ldots, a_{.m})$ and $a_{.j} = \sum_{i=1}^{m} a_{ij}$. Since $1 = \sum_{i=1}^{m} \nu_i$ and $\sum_{i=1}^{m} g_{ij} = 1$, it follows that $\sum_{i=1}^{m} \gamma_i = 1$. Thus we have the same multinomial model as in Section 2. Furthermore, note that if $R = G^{-1}DG$, then $r_{st} \geq 0$ is equivalent to $h_{st} \geq 0$, $s \neq t$. Also recognize that $\hat{\boldsymbol{\theta}} = S\hat{\boldsymbol{\nu}}$. This means that the results of Section 2, in particular, the analog of Theorem 2.4, apply. The analog being that $\hat{\mu} \geq \mu_E$ if $h_{st} \geq 0$, $s \neq t$. This completes the proof of Theorem 4.1. $\quad\square$

PROOF OF LEMMA 4.2.   Since A and B are lower triangular, $h_{ik} = 0$ if $i < k$. Since the diagonal elements of A and B are 1, it follows that $h_{ii} = u_i$. Now let $1 \leq k < i$ for $i \geq 2$. Since $H = BDA$, we have

$$\text{(A.10)} \qquad h_{ik} = b_{ii}u_i a_{ik} + b_{i,k-1}u_{i-1}a_{i-1,k} + b_{i,i-2}u_{i-2}a_{i-2,k}$$

(where the last term vanishes when $i = 2$). In particular, if $i = 2$ and $k = 1$, then $h_{21} = b_{22}u_2 a_{21} + b_{21}u_1 a_{11} = u_2 - u_1$.

For $i \geq 3$, from the nature of B and (A.10) we have

$$\text{(A.11)} \qquad h_{ik} = u_i a_{ik} - 2u_{i-1}a_{i-1,k} + u_{i-2}a_{i-2,k}.$$

First, consider $k = 1$. Since $a_{i1} = 1$ for all i, $h_{i1} = u_i - 2u_{i-1} + u_{i-2}$ if $i \geq 3$. Finally, let $k \geq 2$, $i \geq 3$. Then, from (4.1) and (A.11),

$$\begin{aligned} h_{ik} &= u_i(i - k + 1) - 2u_{i-1}(i - k) + u_{i-2}(i - k - 1) \\ &= (i - k)(u_i - 2u_{i-1} + u_{i-2}) + (u_i - u_{i-2}). \end{aligned}$$

This completes the proof of the lemma. $\quad\square$

PROOF OF THEOREM 4.3.   In light of (4.3), this is an immediate consequence of Theorem 4.1. $\quad\square$

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

COHEN, A., KEMPERMAN, J. H. B. and SACKROWITZ, H. B. (1994). Unbiased testing in exponential family regression. *Ann. Statist.* **22** 1931–1946.

LEE, C. C. (1988). Quadratic loss of order restricted estimators for treatment means with a control. *Ann. Statist.* **16** 751–758.

LIU, C. (2000). Estimation of discrete distributions with a class of simplex constraints. *J. Amer. Statist. Assoc.* **95** 109–120.

MATHER, K. (1933). The relation between chiasmata and crossing-over in diploid and triploid *Drosophila melanogaster*. *J. Genetics* **27** 243–259.

MATHER, K. (1938). Crossing-over. *Biol. Rev. Cambridge Philos. Soc.* **13** 252–292.

OTT, J. (1996). Estimating crossover frequencies and testing for numerical interference with highly polymorphic markers. In *Genetic Mapping and DNA Sequencing* (T. Speed and M. S. Waterman, eds.) 49–63. Springer, New York.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

YU, K. and FEINGOLD, E. (2001). Estimating the frequency distribution of crossovers during meiosis from recombination data. *Biometrics* **57** 427–434.

ZANGWILL, W. I. and MOND, B. (1969). *Nonlinear Programming: A Unified Approach.* Prentice–Hall, Englewood Cliffs, NJ.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY
E-MAIL: artcohen@rci.rutgers.edu