Journal of Heredity 2003:94(2):175–179
DOI: 10.1093/jhered/esg026

Asymptotic Variances of QTL Estimators With Selective DNA Pooling

C. CARLEOS, J. A. BARO, J. CAÑON, AND N. CORRAL

From the Departamento de Estadística, Facultad de Ciencias, C/Calvo Sotelo 33007 Oviedo, Asturias, Spain (Carleos); Departamento de Ciencias Agroforestales, Universidad de Valladolid, Palencia, Spain (Baro); Departamento de Produccion Animal, Universidad Complutense, Madrid, Spain (Cañon); and Departamento de Estadistica, Universidad de Oviedo, Oviedo, Spain (Corral). This work was supported by the European Regional Development Fund project 1FD97-0042.

Address correspondence to Carlos Carleos at the address above, or e-mail: carleos@pinon.ccu.uniovi.es.

Investigation on QTL-marker linkage usually requires a great number of observed recombinations, inferred from combined analysis of phenotypes and genotypes. To avoid costly individual genotyping, inferences on QTL position and effects can instead make use of marker allele frequencies. DNA pooling of selected samples makes allele frequency estimation feasible for studies involving large sample sizes. Linkage studies in outbred populations have traditionally exploited half-sib family designs; within the animal production context, half-sibships provide large families that are highly suitable for DNA pooling. Estimators for QTL position and effect have been proposed that make use of information from flanking markers. We present formulas derived by the delta method for the asymptotic variance of these estimators.

The half-sib design is particularly well suited to animal genetics, for both laboratory and breeding species (Georges et al. 1995; Weller et al. 1990). DNA pooling techniques applied to selected samples allow direct estimation of marker allele frequencies within the best and worst performing animals of a class of half-sib progeny; this allows great savings in terms of genotyping and data collection compared to individual genotype determination. This method does have its drawbacks, such as loss of information about joint marker inheritance—allelic frequencies are known but genotypic frequencies are not. Additionally, imprecise values stemming from technical error constitute an additional source of inaccuracy (Lipkin et al. 1998).

Darvasi and Soller (1994) showed how DNA pooling can be combined with selection to find association between a marker and a QTL using a backcross, an F_2, or a half-sib family. Dekkers (2000) extended this method to consider two-marker (interval) mapping and to allow the estimation of the QTL position.

Methods

We consider a QTL **Q** (with alleles Q and q) flanked by two markers (**M** and **N**), each with two alleles (M, m and N, n, respectively). A half-sib family design is considered, where the common sire has haplotypes MQN/mqn, while not sharing marker alleles with its mates (backcross-like). The family size is n. The recombination fraction between the markers is θ. The recombination rate between **M** and Q is θ_M. Progeny receiving the Q allele (respectively, the q allele) from the sire has a phenotypic distribution following an $N(\mu_Q,\sigma)$ (respectively, an $N[\mu_Q,\sigma]$). The parameter of interest is $\alpha = \mu_Q - \mu_q$. Within the common framework of selective DNA pooling (Darvasi and Soller 1994), α is experimentally determined by mixing DNA samples from either the top (upper tail) or worst (lower tail) performing animals for a given trait. Further, marker frequencies are determined within each tail. Information on the overall distribution is assumed sufficient to suppose the grand mean μ and σ known. Algebraic approximations to the variances of the estimators of θ_M and α will be given.

Let p_{UM} denote the frequency of progeny in the upper tail that received the allele M, that is, $PR[M|U]$. The rest of the subscripts have analogous meanings. The observed proportion corresponding to p_{um} is p_{um}. The estimator of θ_M based on a single tail (Dekkers 2000) is

$$\hat{\theta}_M = \frac{\hat{p}_{UM} - \hat{p}_{UQ}}{1 - 2\hat{p}_{UQ}} \tag{1}$$

where

$$\hat{p}_{UQ} = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{(1 - 2\hat{p}_{UM})(1 - 2\hat{p}_{UN})}{1 - 2\theta}} \tag{2}$$

Let μ_U and μ_L be the means of phenotypes in the upper and the lower tails, respectively. Regarding α, if

$$\hat{\mu}_{ULQ} = \frac{\hat{p}_{UQ}(\hat{\mu}_U - \mu) + \hat{p}_{LQ}(\hat{\mu}_L - \mu)}{\hat{p}_{UQ} + \hat{p}_{LQ}}$$

then an estimator for the other allele, q, is

$$\hat{\alpha} = \frac{\hat{\mu}_{UQ} - \hat{\mu}_{UQ}}{i_s^2} \quad (3)$$

where i_s is the selection intensity associated with a tail of size s (Falconer 1989).

Distribution of the Position Estimator

As seen in the Appendix, the δ-method (Bishop et al. 1975) states that θ_M is approximated by a normal distribution with mean θ_M and variance-covariance matrix

$$\mathrm{Var}(\theta_M) \approx \mathbf{G}' \cdot \mathrm{Var}(\hat{\vec{n}}_M) \cdot \mathbf{G} \quad (4)$$

where

$$\mathbf{G} = \frac{\partial \theta_M}{\partial \vec{n}_M}$$

and n_M is the observed allele frequencies, being its variance $\mathrm{Var}(n_M)$ as shown in equation (9).

Distribution of the Effect Estimator

The distribution of phenotypes did not affect the position estimation, except through the probability of QTL alleles in the tails. For the estimation of α, we assume normal distribution of phenotypes in both groups: offspring inheriting the Q allele, and offspring inheriting the q allele.

Some additional notation is described:

- μ_U and σ_U^2 are the mean and the variance of the phenotypes above the u threshold; it is easily seen that for an overall normal distribution $\mathcal{N}(\mu, \sigma)$,

$$\mu_U = \mu + \sigma \frac{\phi(u')}{1 - \Phi(u')}$$

and (Johnson and Kotz 1970)

$$\sigma_U^2 = \sigma^2 \left[1 + \frac{\phi(u')u'}{1 - \Phi(u')} - \left(\frac{\phi(u')}{1 - \Phi(u')} \right)^2 \right]$$

where $u' = (u - \mu)/\sigma$. In our case, the phenotypic distribution is a mixture of normals, so these expressions should be altered accordingly.

- μ_{UQ} is the mean of offspring above the u threshold inheriting the q allele.

Analogous definitions apply for L and q subscripts.

Let the unobserved sample consist of (X_i, Y_i) values, with $i \in \{1, \ldots, n\}$, where X_i represents the parental QTL allele of individual i:

$$X_i = \begin{cases} 0, & q \\ 1, & Q \end{cases}; \quad X_i \hookrightarrow \mathcal{B}\left(1, \frac{1}{2}\right)$$

and Y_i is the continuous phenotype with conditional Gaussian distribution:

$$Y_i \mid_{X_i} \hookrightarrow \mathcal{N}\left(\mu + \alpha\left[X_i - \frac{1}{2}\right], \sigma\right).$$

Let us define:

- the indicator

$$I_i^U = \chi_{[u,\infty)}(Y_i)$$

signaling whether the individual i is in the upper tail; analogously, I_i^L;

- the tail size

$$n_U = \sum_{i=1}^{n} I_i^U$$

the number of individuals in the upper tail; analogously, n_L;

- the estimator of the tail mean

$$\hat{\mu}_U = \begin{cases} 0, & n_U = 0 \\ \frac{\sum_{i=1}^{n} Y_i I_i}{n_U}, & n_U > 0 \end{cases}$$

analogously for μ_L;

- the unobserved proportion estimators

$$\hat{p}_{UQ} = \begin{cases} 0, & n_U = 0 \\ \frac{\sum_{i=1}^{n} X_i I_i}{n_U}, & n_U > 0 \end{cases}$$

actually estimated by $\hat{p}_{UQ}$ (2); let $\hat{p}_{Uq} = 1 - \hat{p}_{UQ}$ analogously for the lower tail.

The covariance matrix $\mathbf{J} := \mathrm{Var}(\hat{\mu}_U, \hat{\mu}_L, \hat{p}_{UQ}, \hat{p}_{LQ})$ is computed taking into account:

- $\mathrm{Var}(\mu_U)$: it is seen (10):

$$E(\mu_U) \approx \mu_U$$

and similarly for the variance (11):

$$\mathrm{Var}(\mu_U) = \sigma_U^2 \mathrm{E}\left[\frac{1}{\hat{n}_U}\right] \quad (5)$$

where the notation $\mathrm{E}[1/n_U]$ implies substituting zero for the inverse of n_U in the highly unlikely case of an empty upper tail ($n_U = 0$); adequate bounds are computed as shown in the Appendix (p stands for p_{UQ}):

lower bound (Equation 12)

$$\mathrm{E}\left[\frac{1}{\hat{n}_U}\right] \geq \frac{1}{p(n+1)}\left[1 - (n+1)p^n(1-p) - (1-p)^{n+1}\right]$$

upper bound (Equation 13)

$$\mathrm{E}\left[\frac{1}{\hat{n}_U}\right] \leq \frac{1}{p(n+1)} \mathrm{E}\left[\frac{\hat{n}_U + 1}{\hat{n}_U}\right]$$

and according to Lynch and Walsh (1998, p. 818)

$$\mathrm{E}\left(\frac{\hat{n}_U + 1}{\hat{n}_U}\right) \approx \frac{n_U + 1}{n_U}\left(1 + \frac{n - n_U}{n n_U (n_U + 1)}\right)$$

Note that this result is not adequate to directly approximate $\mathrm{E}(1/n_U)$.

Despite unavailability of $E(1/n_U)$, the above bounds show that the approximation achieved with $\hat{\sigma}_U^2/n_U$ is adequate for (5) with n large enough.

The following relations are obtained by the same reasoning:

$$\mathrm{Var}(p_{UQ}) \approx \frac{p_{UQ}(1-p_{UQ})}{n_U}$$

$$\mathrm{Cov}(p_{UQ}, p_{LQ}) = 0$$

$$\mathrm{Cov}(\mu_U, p_{UQ}) \approx \frac{p_{UQ}(\mu_{UQ}-\mu_U)}{n_U}$$

Eventually:

$$\mathbf{J} = \begin{bmatrix} \frac{\sigma_U^2}{\hat{n}_U} & 0 & \frac{p_{UQ}(\mu_{UQ}-\mu_U)}{\hat{n}_U} & 0 \\ 0 & \frac{\sigma_L^2}{\hat{n}_L} & 0 & \frac{p_{LQ}(\mu_{LQ}-\mu_L)}{\hat{n}_L} \\ \frac{p_{UQ}(\mu_{UQ}-\mu_U)}{\hat{n}_U} & 0 & \frac{p_{UQ}(1-p_{UQ})}{\hat{n}_U} & 0 \\ 0 & \frac{p_{LQ}(\mu_{LQ}-\mu_L)}{\hat{n}_L} & 0 & \frac{p_{LQ}(1-p_{LQ})}{\hat{n}_L} \end{bmatrix}$$

The matrix of partial derivatives of the estimator (3) of α with respect to $(\mu_L, \mu_U, p_{LQ}, p_{UQ})$ is:

$$\mathbf{G} := \frac{\partial \alpha}{\partial(\mu_L, \mu_U, p_{LQ}, p_{UQ})} = \frac{1}{i_s^2} \begin{pmatrix} \frac{p_{LQ}}{p_{LQ}+p_{UQ}} - \frac{1-p_{LQ}}{2-p_{LQ}-p_{UQ}} \\ \frac{p_{UQ}}{p_{LQ}+p_{UQ}} - \frac{1-p_{UQ}}{2-p_{LQ}-p_{UQ}} \\ \frac{p_{UQ}(\mu_L-\mu_U)}{(p_{LQ}+p_{UQ})^2} + \frac{(1-p_{UQ})(\mu_L-\mu_U)}{(2-p_{LQ}-p_{UQ})^2} \\ \frac{p_{LQ}(\mu_U-\mu_L)}{(p_{LQ}+p_{UQ})^2} + \frac{(1-p_{LQ})(\mu_U-\mu_L)}{(2-p_{LQ}-p_{UQ})^2} \end{pmatrix}$$

therefore, after application of the δ-method,

$$\mathrm{Var}(\hat{\alpha}) \approx \mathbf{G}' \cdot \mathbf{J} \cdot \mathbf{G}. \quad (6)$$

Results

Simulations were performed to check the adequacy of the proposed approximations. The experimental design and genetic model, both described in the Methods section, led to the results shown in Table 1 for 10,000 iterations. The "Simulation" column displays observed standard deviations of the sampling distribution of the estimator across the 10,000 iterations. The "Predicted" column was obtained from formulas 4 and 6, with parameters replaced with population values, derived from the values chosen for simulation.

The study explored several combinations of sample sizes, interval widths, and substitution effects. Unless specified in Table 1, reference values for those parameters are taken: a family of 5,000 half-sibs, an interval 50 cM wide, and a substitution effect α of 0.5 environmental standard deviations. Upper and lower tails comprise 10% of progeny each.

Discussion

It can be concluded from our results that estimation of the θ_M variance is accurate for marker brackets wider than 20 cM. Shorter intervals lead to distribution of the position estimator not holding within the parameter space (corresponding to the intermarker gap); thus, tail values agglomerate at boundaries. The presented formulas can be used to compute the probability of erroneously locating a QTL exactly at a marker position.

It was also noted that the proposed approximations degrade when the effect α exceeds one standard deviation. A likely explanation is the departure from normality of the overall phenotypic distribution when the Q, q mixture components are too separated.

All of the limitations mentioned so far are related to the single fact that variances of estimators are computed making use of asymptotic theory, which notably relies on regularity conditions.

This study did not address issues such as influence of phase for small QTL effects, and narrow marker intervals, on normality of asymptotic distribution of the effect and position estimators. These deserve further study.

Large samples, above 5,000 half-sibs, are required for the proposed formulas to achieve fair results. Study of small sample distributions must account for lack of normality, and inferences should no longer rely on asymptotic theory. Exploration of the behavior of the estimators under small-sample scenarios requires additional research. A novel approach based on resampling methods is being developed by the authors (Carleos et al. 2002).

Table 1. Variances of position and effect estimators.

	Position estimator		Effect estimator	
Parameters	Predicted	Simulation	Predicted	Simulation
α				
0.25	0.064	0.064	0.032	0.041
0.50	0.032	0.033	0.038	0.039
1.00	0.017	0.018	0.039	0.038
2.00	0.012	0.012	0.003	0.027
cM				
10	0.013	0.027	0.038	0.037
25	0.022	0.022	0.038	0.040
50	0.032	0.033	0.038	0.049
100	0.024	0.026	0.038	0.075
n				
1,000	0.076	0.071	0.086	0.108
5,000	0.032	0.033	0.038	0.039
10,000	0.021	0.020	0.027	0.025
50,000	0.009	0.009	0.012	0.014

The QTL effect α is expressed in environmental standard deviations, and interval width in cM; n is the sample size.

Appendix

Distribution of the Position Estimator

Assuming fixed selection thresholds, the unobserved absolute genotypic frequencies follow a multinomial distribution:

$$\vec{n}_{MN} := (n_{LMN}, n_{LMn}, n_{LmN}, n_{Lmn}, n_C, n_{UMN}, n_{UMn}, n_{UmN}, n_{Umn}) \hookrightarrow \mathcal{B}(n, [p_{LMN}p_L, p_{LMn}p_L, p_{LmN}p_L, p_{Lmn}p_L, p_C, p_{UMN}p_U, p_{UMn}p_U, p_{UmN}p_U, p_{Umn}p_U]) \quad (7)$$

being n_{lmn}, the absolute frequency of genotype *MN* in the lower tail; analogously for the other subscripts. Here, *C* indicates a central class, comprising the individuals not selected.

The "observed" (i.e., estimated by means of DNA pooling) allelic frequencies are:

$$\begin{aligned}\hat{\vec{n}}_M &= (\hat{n}_L, \hat{n}_{LM}, \hat{n}_{LN}, \hat{n}_U, \hat{n}_{UM}, \hat{n}_{UM}) \\ &= (n_{LMN} + n_{LMn} + n_{LmN} + n_{Lmn}, n_{LMN} + n_{LMn}, \\ &\quad n_{LMN} + n_{LmN}, n_{UMN} + n_{UMn} + n_{UmN} + n_{Umn}, \\ &\quad n_{UMN} + n_{UMn}, n_{UMN} + n_{UmN}) = \mathbf{A} \cdot \vec{n}_{MN}\end{aligned} \quad (8)$$

where counts *n* carry subscripts indicating tail and allele, and with *A*, the matrix that relates observed and unobserved absolute frequencies, being

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

so

$$\mathrm{Var}(\hat{\vec{n}}_M) = \mathbf{A} \cdot \mathrm{Var}(\vec{n}_{MN}) \cdot \mathbf{A}' \quad (9)$$

The estimator (1) of θ_M, averaged over the two tails, is rewritten as

$$\hat{\theta}_M = \frac{1}{2} - \frac{1}{4}\sqrt{1-2\theta}\left[\sqrt{\frac{2\hat{n}_{UM} - \hat{n}_U}{2\hat{n}_{UN} - \hat{n}_U}} + \sqrt{\frac{2\hat{n}_{LM} - \hat{n}_L}{2\hat{n}_{LN} - \hat{n}_L}}\right]$$

The multinomial frequencies (7) can be approximated by a normal distribution. The observed frequencies (8) are a linear transformation of those, so they are asymptotically normal. The estimator θ_M is a nonlinear function of the observed frequencies, defined on an open subset and differentiable at their expected values, $\mathrm{E}[n_M] = \mathbf{A}\mathrm{E}[n_{MN}]$. Should these conditions hold, the δ-method states that θ_M is approximated by a normal distribution with mean θ_M and variance-covariance matrix

$$\mathrm{Var}(\hat{\theta}_M) \approx \mathbf{G}' \cdot \mathrm{Var}(\hat{\vec{n}}_M) \cdot \mathbf{G}$$

where

$$\mathbf{G} = \frac{\partial \theta_M}{\partial \vec{n}_M}$$

and $\mathrm{Var}(n_M)$ is (9). **G** can be estimated by:

$$\hat{\mathbf{G}} = \frac{-\sqrt{1-2\theta}}{8}\begin{bmatrix} \frac{1}{\sqrt{\frac{2\hat{n}_{LM}-\hat{n}_L}{2\hat{n}_{LN}-\hat{n}_L}}} \frac{2\hat{n}_{LM}-2\hat{n}_{LN}}{(2\hat{n}_{LN}-\hat{n}_L)^2} \\ \frac{1}{\sqrt{\frac{2\hat{n}_{LM}-\hat{n}_L}{2\hat{n}_{LN}-\hat{n}_L}}} \frac{2(2\hat{n}_{LN}-\hat{n}_L)}{(2\hat{n}_{LN}-\hat{n}_L)^2} \\ \frac{1}{\sqrt{\frac{2\hat{n}_{LM}-\hat{n}_L}{2\hat{n}_{LN}-\hat{n}_L}}} \frac{-2(2\hat{n}_{LM}-\hat{n}_L)}{(2\hat{n}_{LN}-\hat{n}_L)^2} \\ \frac{1}{\sqrt{\frac{2\hat{n}_{UM}-\hat{n}_U}{2\hat{n}_{UN}-\hat{n}_U}}} \frac{2\hat{n}_{UM}-2\hat{n}_{UN}}{(2\hat{n}_{UN}-\hat{n}_U)^2} \\ \frac{1}{\sqrt{\frac{2\hat{n}_{UM}-\hat{n}_U}{2\hat{n}_{UN}-\hat{n}_U}}} \frac{2(2\hat{n}_{UN}-\hat{n}_U)}{(2\hat{n}_{UN}-\hat{n}_U)^2} \\ \frac{1}{\sqrt{\frac{2\hat{n}_{UM}-\hat{n}_U}{2\hat{n}_{UN}-\hat{n}_U}}} \frac{-2(2\hat{n}_{UM}-\hat{n}_U)}{(2\hat{n}_{UN}-\hat{n}_U)^2} \end{bmatrix}.$$

Expectation of μ_U

$$\begin{aligned}\mathrm{E}(\hat{\mu}_U) &= E\left[\frac{\sum Y_i I_i}{\sum I_i}\right] = \mathrm{E}\left[\mathrm{E}\left[\frac{\sum Y_i I_i}{\sum I_i}\middle|\hat{n}_U\right]\right] \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\mathrm{E}\left[\frac{\sum Y_i I_i}{\sum I_i}\middle|\hat{n}_U = n_U\right] \\ &= \left|Y_i' := Y_i\,|_{Y_i>u}\right| = \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\mathrm{E}\left[\frac{\sum_{i=1}^{n_U} Y_i'}{n_U}\right] \\ &= \sum \Pr[n_U]\frac{1}{n_U}\sum_{i=1}^{n_U}\mathrm{E}[Y_i'] = \sum \Pr[n_U]\frac{n_U}{n_U}\mu_U \\ &= \mu_U \Pr[\hat{n}_U \geq 1] \approx \mu_U\end{aligned} \quad (10)$$

Variance of μ_U

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_U) &= \mathrm{E}\left[\left(\frac{\sum Y_i I_i}{\sum I_i} - \mu_U\right)^2\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{\sum Y_i I_i}{\sum I_i} - \mu_U\right)^2\right]\middle|\hat{n}_U\right] \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\mathrm{E}\left[\left[\left(\frac{\sum Y_i I_i}{\sum I_i} - \mu_U\right)^2\right]\middle|\hat{n}_U = n_U\right] \\ &= \left|Y_i' := Y_i\right|_{Y_i>u}\left| \right. \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\mathrm{E}\left[\left(\frac{\sum_{i=1}^{n_U} Y_i'}{n_U} - \mu_U\right)^2\right] \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\mathrm{E}\left[\left(\frac{\sum_{i=1}^{n_U}(Y_i' - \mu_U)}{n_U}\right)^2\right] \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\frac{1}{n_U^2}\mathrm{E}\left[\left(\sum_{i=1}^{n_U} Y_i' - \mu_U)\right)^2\right] \\ &= \sum_{n_u=1}^{n} \Pr[\hat{n}_U = n_U]\frac{n_U}{n_U^2}\sigma_U^2 = \sigma_U^2\mathrm{E}\left[\frac{1}{\hat{n}_U}\right]\end{aligned} \quad (11)$$

Appropriate bounding values for the expectation constituting the last factor, $E[1/n_U]$ in (11), can be determined as follows (p denotes p_{UQ}):

lower bound

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\hat{n}_U}\right] &= \sum_{k=1}^{n} \frac{1}{k} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&\geq \sum_{k=1}^{n} \frac{1}{k+1} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k+1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{1}{p(n+1)} \sum_{k=1}^{n} \frac{(n+1)!}{(k+1)!(n-k)!} p^{k+1} (1-p)^{n-k} \\
&= \frac{1}{p(n+1)} \sum_{k=2}^{n+1} \frac{(n+1)!}{k!(n+1-k)!} p^k (1-p)^{n+1-k} \\
&= \frac{1}{p(n+1)} [1 - (n+1)p^n(1-p) - (1-p)^{n+1}] \qquad (12)
\end{aligned}$$

upper bound

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{\hat{n}_U}\right] &= \sum_{k=1}^{n} \frac{1}{k} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&\leq \sum_{k=1}^{n} \frac{k+1}{k} \frac{n!}{(k+1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \frac{1}{p(n+1)} \sum_{k=1}^{n} \left(\frac{k+1}{k}\right) \frac{(n+1)!}{(k+1)!(n-k)!} \\
&\quad \times p^{k+1} (1-p)^{n-k} \\
&= \frac{1}{p(n+1)} \mathrm{E}\left[\frac{\hat{n}_U + 1}{\hat{n}_U}\right] \qquad (13)
\end{aligned}$$

References

Bishop YMM, Fienberg S, and Holland P, 1975. Discrete multivariate analysis. Cambridge: MIT Press.

Carleos C, Corral N, Baro JA, and Cañon J, 2002. Enhanced precision in QTL mapping with selective DNA pooling. 7th World Congress on Genetics Applied in Livestock Production, Montpellier, France.

Darvasi A, and Soller M, 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373.

Dekkers JCM, 2000. Quantitative trait locus mapping based on selective DNA pooling. J Anim Breed Genet 117:1–16.

Falconer DS, 1989. Introduction to Quantitative Genetics. New York: Longman.

Georges M, Nielsen D, and Mackinnon M, 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907–920.

Johnson NL and Kotz S, 1970. Continuous univariate distributions—1 New York: John Wiley and Sons.

Lipkin E, Mosig MO, Darvasi A, Ezra E, Shalom A, Friedmann A, and Soller M, 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics 149:1557–1567.

Lynch M and Walsh B, 1998. Genetics and analysis of quantitative traits Sunderland, MA: Sinauer Associates.

Weller JI, Kashi Y, and Soller M, 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. J Dairy Sci 73:2525–2537.

Received June 10, 2002
Accepted December 31, 2002

Corresponding Editor: R. C. Woodruff