Abstract 0941

Derivation of Bayes and Minimax decision rules for allelic frequencies estimation in biallelic loci

C.A. Martínez[1,2], K. Khare[2], and M. A. Elzo[1]

[1]Department of Animal Sciences, University of Florida, Gainesville, FL , USA.
[2]Department of Statistics, University of Florida, Gainesville, FL , USA.

SUMMARY

The aim of this study was to derive Bayes and Minimax estimators (ME) of allele frequencies for biallelic loci using decision theory. Because an optimal decision rule with uniformly smallest risk rarely exists, an approach is to establish principles that allow ordering of decision rules according to their risk function. To this end, two general methods were used: The Bayes and the Minimax principles. For an arbitrary locus, the sampling model was a trinomial distribution for numbers of individuals for each genotype and the prior was a Beta distribution, chosen because of mathematical convenience, flexibility and genetic interpretation of its parameters. Three types of loss functions were considered: squared error (SEL), Kullback-Leibler (KLL) and a quadratic error loss (QEL). The SEL and KLL yielded the same estimator, which was a convex combination of the prior mean and the MLE. Using the Bayes estimator from QEL, an ME was derived by applying a theorem which states that a Bayes estimator with constant risk is also Minimax. The constant risk was obtained by finding appropriate hyperparameter values. This estimator was shown to be equivalent to MLE. The prior associated with this ME was uniform [0,1]. Extension to several loci under linkage equilibrium and independent priors was discussed. The estimators derived here have the appealing property of allowing variation in allelic frequencies, which is more congruent with the reality of finite populations exposed to evolutionary forces. In addition, from a Bayesian perspective they permit modelling uncertainty and incorporation of previous genotypic information from the population.

INTRODUCTION

In population genetics, allelic frequencies are typically estimated via maximum likelihood (MLE). Under this setting, allele frequencies are treated as unknown fixed parameters. However, according to population genetics theory, allele frequencies can have random variation, thus they should be treated as random variables (Wright, 1930; 1937; Crow and Kimura, 1970).

Under the decision theory framework, given a parameter space Θ, a decision space D and a loss function $L(\theta, \delta(X))$, the average loss for a decision rule δ when the true state of nature is $\theta \in \Theta$, is defined as $R(\theta,\delta) = E_\theta[L(\theta,\delta(X))]$. The ideal decision rule is one having uniformly smaller risk, that is, it minimizes the risk for every $\theta \in \Theta$ (Lehmann and Casella, 1998). However, such a decision rule rarely exists unless restrictions like unbiasedness and invariance are posed over the estimators. Another approach is to allow all kinds of estimators and to use an optimality criterion weaker than a uniform minimum risk. Such a criterion looks for minimization of $R(\theta,\delta)$ in some general sense and there are two principles to achieve that goal: the Bayes principle and the Minimax principle (Lehman and Casella, 1998; Ghosh, personal communication).

Given a loss function and a prior distribution, the Bayes principle looks for an estimator minimizing the Bayesian risk $r(\Lambda,\delta)$, that is, a decision rule δ is defined to be a Bayes decision rule with respect to a prior distribution Λ if it satisfies:

$$r(\Lambda,\delta) = \int_\Theta R(\theta,\delta)d\Lambda(\theta) = \inf_{\delta\in D} R(\theta,\delta)$$

This kind of estimators can be interpreted as those minimizing the posterior risk. On the other hand, the Minimax principle consists of finding decision rules that minimize the supremum (over the parameter space) of the risk function (the worst scenario). Thus δ^* is said to be a Minimax decision rule (or ME) if:

$$\sup_{\theta\in\Theta} R(\theta,\delta^*) = \inf_{\delta\in D}\sup_{\theta\in\Theta} R(\theta,\delta)$$

The aim of this study was to derive Bayes and Minimax estimators of allele frequencies for biallelic loci under a decision theory framework.

DERIVATION OF BAYES RULES

Here we assume Hardy-Weinberg equilibrium for an arbitrary locus. Let X_1, X_2 and X_3 be random variables indicating the number of animals having genotypes AA, AB and BB following a trinomial distribution conditional on θ (the frequency of the "reference" allele B) with corresponding frequencies: $(1-\theta)^2, 2\theta(1-\theta)$ and θ^2 and let $\boldsymbol{x} = (X_1, X_2, X_3)$. Thus, we are interested in estimating $\theta \in [0,1]$. The sampling model was a trinomial distribution and the prior was a Beta(α,β) distribution. This family of priors was chosen because of mathematical convenience, flexibility and because the hyperparameters α and β have a genetic interpretation (Wright, 1937). Under this setting, three loss functions were used to derive Bayes decision rules: squared error loss (SEL), Kullback-Leibler loss (KLL) and a quadratic error loss (QEL).

Squared error loss: Under SEL, the Bayes estimator is the posterior mean. Thus we need to derive the posterior distribution of the parameter:

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\boldsymbol{x}|\theta)\pi(\theta) \propto (\theta)^{x_2+2x_3+\alpha-1}(1-\theta)^{2x_1+x_2+\beta-1}$$

The posterior is a Beta$(x_2+2x_3+\alpha, 2x_1+x_2+\beta)$ distribution. Therefore, the Bayes estimator under the given prior and SEL is:

$$\hat{\theta}^{SEL} = \frac{x_2+2x_3+\alpha}{2n+\alpha+\beta} = \frac{x_2+2x_3}{2n}\left(\frac{2n}{2n+\alpha+\beta}\right) + \frac{\alpha}{\alpha+\beta}\left(\frac{\alpha+\beta}{2n+\alpha+\beta}\right)$$

which is a convex combination of the MLE and the prior mean.
The frequentist risk for this estimator is:

$$R(\theta,\hat{\theta}^{SEL}) = E_\theta\left[\left(\frac{x_2+2x_3+\alpha}{2n+\alpha+\beta}-\theta\right)^2\right] = Var_\theta\left[\frac{x_2+2x_3+\alpha}{2n+\alpha+\beta}\right] + \left(E_\theta\left[\frac{x_2+2x_3+\alpha}{2n+\alpha+\beta}-\theta\right]\right)^2$$

After some algebra we obtain:

$$R(\theta,\hat{\theta}^{SEL}) = \frac{2n\theta(1-\theta)+[(1-\theta)\alpha-\theta\beta]^2}{(2n+\alpha+\beta)^2}$$

Kullback-Leibler loss: Under this loss, the Bayes decision rule is the one minimizing (with respect to δ):

$$\int_0^1 L_{KL}(\theta,\delta)\pi(\theta|\boldsymbol{x})d\theta$$

where:

$$L_{KL}(\theta,\delta) = E_\theta\left[log\left(\frac{\pi(\boldsymbol{x}|\theta)}{\pi(\boldsymbol{x}|\delta)}\right)\right] = 2n\left[(1-\theta)log\left(\frac{1-\theta}{1-\delta}\right)+\theta log\left(\frac{\theta}{\delta}\right)\right]$$

After differentiating $\int_0^1 L_{KL}(\theta,\delta)\pi(\theta|\boldsymbol{x})d\theta$ with respect to δ, doing some algebra and checking the second order condition we obtain the following Bayes estimator:

$$\hat{\theta}^{KLL} = \frac{x_2+2x_3+\alpha}{2n+\alpha+\beta} = \hat{\theta}^{SEL}$$

Thus, the Bayes decision rules are the same under SEL and KLL.
Quadratic error loss: The general form of this loss is: $w(\theta)(a-\theta)^2, w(\theta)>0, \forall\,\theta\in\Theta$ Let $w(\theta) = [\theta(1-\theta)]^{-1}$. This form of $w(\theta)$ was chosen for mathematical convenience. Under the given loss and prior, the Bayes estimator is:

$$\hat{\theta}^{QEL} = \begin{cases} \dfrac{x_2+2x_3+\alpha-1}{2n+\alpha+\beta-2}, & if\ x_2+2x_3+\alpha-1>0\ and\ 2x_1+x_2+\beta-1>0 \\ 0, & if\ x_2+2x_3+\alpha-1\le 0 \\ 1, & if\ 2x_1+x_2+\beta-1\le 0 \end{cases}$$

In most of real life situations $x_2+2x_3+\alpha-1>0\ and\ 2x_1+x_2+\beta-1>0$, and in this case:

$$R(\theta,\hat{\theta}^{QEL}) = E_\theta[w(\theta)(\theta-a)^2] = \frac{2n}{(2n+\alpha+\beta-2)^2} + \frac{-\theta(\alpha+\beta-2)+\alpha-1}{\theta(1-\theta)(2n+\alpha+\beta-2)^2}$$

DERIVATION OF MINIMAX RULES

To derive Minimax rules the following theorem is used (Lehman and Casella, 1998):
Theorem 1: Let Λ be a prior and δ_Λ a Bayes rule with respect to Λ with Bayes risk satisfying $r(\Lambda,\delta_\Lambda) = \sup_{\theta\in\Theta} R(\theta,\delta_\Lambda)$. Then: $i)$ δ_Λ is Minimax and $ii)$ Λ is least favorable.

A corollary that follows from this theorem is that a if δ is a Bayes decision rule with respect to a prior Λ and it has constant (not depending on θ) frequentist risk $R(\theta,\delta)$ then it is also an ME and Λ is least favorable, that is, it causes the greatest average loss. This is the result that was used to find ME. Once a Bayes estimator has been derived, if it is possible to find appropriate values for the hyperparameters such that $R(\theta,\delta)$ is constant, then under the prior with these particular values for the hyperparameters, the resulting estimator is also an ME. Among the previously derived Bayes estimators, it is easy to notice that provided $x_2+2x_3+\alpha-1>0\ and\ 2x_1+x_2+\beta-1>0$, $\hat{\theta}^{QEL}$ will have a constant risk for $\alpha=\beta=1$, that is, under a uniform [0,1] prior. In that case:

$$\hat{\theta}^{Minimax} = \frac{x_2+2x_3}{2n} = \hat{\theta}^{MLE}\ and\ R\left(\theta,\hat{\theta}^{Minimax}\right) = \frac{1}{2n}\ \forall\,\theta\in\Theta$$

This shows that this ME is equivalent to the MLE and that the uniform [0,1] prior is least favorable. The same idea does not work for $\hat{\theta}^{SEL}$.

EXTENSION TO K LOCI

Let $\boldsymbol{\theta} = (\theta_1,\theta_2,\dots,\theta_k)$ be the vector containing the frequencies of the "reference" alleles for k loci, $\boldsymbol{X} = (\boldsymbol{x_1},\boldsymbol{x_2},\dots,\boldsymbol{x_k})$; $\boldsymbol{x_i} = (X_{1i},X_{2i},X_{3i}), i = 1,2,\dots,k$ the vector containing the number of individuals for every genotype at every locus and $\boldsymbol{\delta} = (\delta_1,\delta_2,\dots,\delta_k)$ a vector valued estimator of $\boldsymbol{\theta}$. Consider a general additive loss function of the form: $L(\boldsymbol{\theta},\boldsymbol{\delta}(\boldsymbol{X})) = \sum_{i=1}^k L(\theta_i,\delta_i(\boldsymbol{X}))$. Assuming linkage equilibrium (LE) we have that $\pi(\boldsymbol{X}|\boldsymbol{\theta}) = \prod_{i=1}^k \pi(\boldsymbol{x_i}|\theta_i)$, and using independent priors it follows that $\pi(\boldsymbol{\theta}|\boldsymbol{X}) = \prod_{i=1}^k \pi(\theta_i|\boldsymbol{x_i})$. To obtain a Bayes estimator, we have to minimize the following expression with respect to $\delta_i, \forall\, i = 1,2,\dots,k$:

$$\int_{\Theta_1}\dots\int_{\Theta_k} L(\boldsymbol{\theta},\boldsymbol{\delta}(\boldsymbol{X}))\,\pi(\boldsymbol{\theta}|\boldsymbol{X})d\theta_1\cdots d\theta_k = \int_{\Theta_1}\dots\int_{\Theta_k}\left(\sum_{i=1}^k(\theta_i,\delta_i(\boldsymbol{X}))\right)\pi(\boldsymbol{\theta}|\boldsymbol{X})d\theta_1\cdots d\theta_k$$
$$= \sum_{i=1}^k\int_{\Theta_1}\dots\int_{\Theta_k} L(\theta_i,\delta_i(\boldsymbol{X}))\prod_{\boldsymbol{j=1}}^{\boldsymbol{k}}\pi(\theta_j|\boldsymbol{x_j})\,d\theta_1\cdots d\theta_k$$

The h^{th} integral in the summation $(h = 1,2,\dots,k)$ can be written as:

$$\int_{\Theta_h} L(\theta_h,\delta_h(\boldsymbol{X}))\pi(\theta_h|\boldsymbol{x_h})d\theta_h \int_{\boldsymbol{\Theta_{-h}}}\pi(\theta_i|\boldsymbol{x_i})\,d\boldsymbol{\theta_{-h}} = \int_{\Theta_h} L(\theta_h,\delta_h)\pi(\theta_h|\boldsymbol{x_h})d\theta_h$$

where the subscript " $-h$" indicates that we consider all $i \neq h$. From the result above, it follows that Bayes estimation of $\boldsymbol{\theta}$ reduces to that of its components. Therefore, under LE, independent priors and an additive loss it follows that $\hat{\boldsymbol{\theta}}^{Bayes} = \left(\hat{\theta}_1^{Bayes},\hat{\theta}_2^{Bayes},\dots,\hat{\theta}_k^{Bayes}\right)$. Applying the results derived previously, an ME of $\boldsymbol{\theta}$ is obtained by posing k independent uniform [0,1] priors and the i^{th} element of $\hat{\boldsymbol{\theta}}^{Minimax}$ has the form $\frac{x_{2i}+2x_{3i}}{2n}$, which is equivalent to the MLE.

FINAL REMARKS

The estimators derived here have the appealing property of allowing variation in allelic frequencies, which is more congruent with the reality of finite populations exposed to evolutionary forces. From a Bayesian perspective they permit modelling uncertainty and incorporation of previous genotypic information from the population.

REFERENCES

Crow J. and Kimura M. 1970. Introduction to population genetics theory. Harper & Row Publishers Inc. New York, NY.
Lehmann E.L. and Casella G. 1998. Theory of point estimation. Springer-Verlag, New York, Inc. New York, NY.
Wright S. 1930. Genetics. 16: 98-159.
Wright S. 1937. Proc. Natl. Acad. Sci. USA 23: 307–320.