

# A New Method for Estimation of Resource Selection Probability Function

SUBHASH R. LELE,<sup>1</sup> *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada*

**ABSTRACT** Weighted distributions can be used to fit various forms of resource selection probability functions (RSPF) under the use-versus-available study design (Lele and Keim 2006). Although valid, the numerical maximization procedure used by Lele and Keim (2006) is unstable because of the inherent roughness of the Monte Carlo likelihood function. We used a combination of the methods of partial likelihood and data cloning to obtain maximum likelihood estimators of the RSPF in a numerically stable fashion. We demonstrated the methodology using simulated data sets generated under the log-log RSPF model and a reanalysis of telemetry data presented in Lele and Keim (2006) using the logistic RSPF model. The new method for estimation of RSPF can be used to understand differential selection of resources by animals, an essential component of studies in conservation biology, wildlife management, and applied ecology. (JOURNAL OF WILDLIFE MANAGEMENT 73(1):122–127; 2009)

DOI: 10.2193/2007-535

**KEY WORDS** Bayesian estimation, data cloning, logistic regression, log-log link, maximum likelihood estimation, Monte Carlo, partial likelihood, resource selection function, use-available study design, weighted distributions.

“When resources are used disproportionately to their availability, use is said to be selective” (Manly et al. 2002:15). Understanding differential selection of resources by animals is an essential component of conservation biology, wildlife management, and applied ecology (Boyce and McDonald 1999). Researchers often want to estimate a resource selection probability function (RSPF) and the associated resource selection function (RSF) in pursuit of this understanding (e.g., Boyce and MacDonald 1999, Manly et al. 2002). The most commonly used study design in these studies is the use-versus-available study design, wherein the researcher knows only a random subset of the locations that were visited by the individuals under study and the potentially available set of locations (Manly et al. 2002). The collection of available sites consists of not only the sites that are unused but also sites that might potentially have been used. This problem is sometimes termed as zero contamination (Johnson et al., 2006). Keating and Cherry (2004) noted the problem of assuming no sample contamination and its effect on the estimation of the RSF and RSPF under the use-versus-available study design. Johnson et al. (2006) addressed this problem and showed that if one assumes the exponential form of the RSPF, then standard statistical software to fit logistic regression can be manipulated to obtain estimates of the nonintercept parameters of the exponential RSPF, or equivalently, the exponential RSF.

The exponential RSPF is only one of many possible models to model RSPF. For example, Manly et al. (2002) list logistic, log-log, and probit models among the possible alternatives. However, no practical method was suggested to fit these alternative models to data obtained under the use-versus-available study design. Generalizing the weighted distribution approach described in Johnson et al. (2006), Lele and Keim (2006) showed that these alternative RSPF models can be fit by numerically maximizing the Monte Carlo estimate of the likelihood function. Although valid

computationally, the Monte Carlo maximum-likelihood approach is difficult to implement in practice because the Monte Carlo sampling variation leads to a rough-likelihood surface. Locating the global maximum in the presence of many local maxima is difficult (Press et al. 1986). Furthermore, standard maximization procedures, such as the Newton-Raphson or Nelder-Mead, require specification of a starting value that is close to the true maximum (Press et al. 1986). A starting value that is far from the true maximum leads to difficulties in even starting the optimization procedure (Press et al. 1986). Our goal was to provide a simple, numerically stable method of obtaining maximum-likelihood estimators of the parameters in the general RSPF.

## MATHEMATICAL METHOD

We let  $\underline{X} = (X_1, X_2, \dots, X_p)$  denote the vector of environmental covariates representing resources that may be used by animals. Under the use-versus-available study design, the researcher only knows a sample of the locations that were visited by the individuals under study. These data may be obtained using Global Positioning System collars, radio-telemetry, or other survey methods. We assume that a particular location may potentially be visited repeatedly. The implicit assumption in resource-selection studies is that habitat characteristics of locations that are visited more often are preferred over habitat characteristics of locations that are visited less often. To study the relationship between habitat characteristics  $\underline{X}$  and probability of selection (or, equivalently, visit), we can use a function  $\pi(\underline{X}, \beta)$ , such that  $0 \leq \pi(\underline{X}, \beta) \leq 1$  for all possible values of  $\underline{X}$  and  $\beta$ . This probability function is called the RSPF (Manly et al. 2002).

We assume a sample of  $N$  used locations from the study area. We denote the data by  $\underline{X}_i^U$   $i = 1, 2, \dots, N$  where  $\underline{X}_i^U$  are the set of habitat characteristics associated with used (strictly speaking, visited) locations. The goal of the analysis is to estimate the RSPF,  $\pi(\underline{X}, \beta)$ , or equivalently, to

<sup>1</sup> E-mail: slele@ualberta.ca

estimate the parameters  $\beta$ . To do this, we make an additional assumption, that the habitat characteristics, or equivalently, environmental covariates,  $\underline{X}_i$ , are a random sample from some multivariate distribution  $f^A(\underline{X})$ . This is simply the distribution of covariates at locations that the species can potentially visit. For detailed discussion on the meaning and definition of the available distribution, see Manly et al. (2002). Provided this assumption is reasonable, standard probability argument leads to the result that the used data  $\underline{X}_i^U$ ,  $i = 1, 2, \dots, N$  arise from the distribution:

$$f^U(\underline{X}, \beta) = \frac{\pi(\underline{X}, \beta) f^A(\underline{X})}{P(\beta)}. \quad (1)$$

where  $P(\beta) = \int \pi(\underline{X}, \beta) f^A(\underline{X}) d\underline{X}$ . Notice that  $0 < P(\beta) \leq 1$  (Johnson et al. 2006, Lele and Keim 2006). In non-mathematical terms, the distribution  $f^U(\underline{X}, \beta)$  corresponds to the proportions of various resources represented in the used sample and  $f^A(\underline{X})$  corresponds to the proportions of various resources represented in the available sample. The commonly used selection index, the ratio of the used and available proportions, is applicable if the resource types are categorical variables (Savage 1931). The RSPF,  $\pi(\underline{X}, \beta)$ , is simply an extension of the idea of selection index to the case of continuous covariates.

Given the data  $\underline{X}_i^U$ ,  $i = 1, 2, \dots, N$ , the likelihood function can be written as

$$L(\beta, \underline{X}^U) = \prod_{i=1}^N \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)}.$$

Given a sample from the used distribution, it is not always possible to estimate the parameters  $\beta$  by maximizing the likelihood function. The technical conditions under which it is possible to estimate the parameters  $\beta$  are provided in Lele and Keim (2006). For practitioners, it may be sufficient to know that resource selection models, such as the logistic, log-log, or probit, can be estimated as long as there is  $\geq 1$  numeric (noncategorical) covariate in the model.

We now describe the method to obtain the maximum likelihood estimator (MLE). First, we obtain a random sample with replacement of size  $M$  from the available distribution by randomly selecting with replacement  $M$  locations from the study area and recording their environmental covariate values. We denote this sample from the available distribution by  $\underline{X}_j^A$ ,  $j = 1, 2, \dots, M$ . We denote the partial likelihood by  $PL$  and  $w = N/(M + N)$ , the proportion of used sites ( $N$ ) in the combined sample of used and available sites ( $N + M$ ). Consider the function

$$PL(\beta, \alpha) = \prod_{i=1}^N \left( \frac{w\pi(\underline{X}_i^U, \beta)}{w\pi(\underline{X}_i^U, \beta) + (1-w)\alpha} \right) \times \prod_{j=1}^M \left( \frac{(1-w)\alpha}{w\pi(\underline{X}_j^A, \beta) + (1-w)\alpha} \right). \quad (2)$$

Maximizing  $PL(\beta, \alpha)$  with respect to  $(\alpha, \beta)$ , under the constraint  $\alpha \in (0, 1)$ , leads to estimators that are equivalent to the maximum-likelihood estimators (Gilbert et al. 1999).

See Appendix for derivation of the partial likelihood function. The partial-likelihood function can be maximized using any numerical optimization routine, such as the Newton–Raphson or the Nelder–Mead algorithm. However, we use a different optimization procedure based on data cloning (Lele et al. 2006). One major advantage of the data-cloning algorithm is that it is less sensitive to the specification of a starting value and is more likely to lead to the global maximum than other numerical optimization methods. Furthermore, data cloning facilitates computation of standard errors of the estimates automatically. In our experience, data-cloning-based optimization is computationally slower but numerically more stable than other optimization algorithms. A combination of the 2 methods is also possible where we obtain an initial estimate using data cloning on a subset of the data set. We then use this initial estimate as the starting value for the Nelder–Mead optimization of the Monte Carlo likelihood function. We also note that if we set the number of clones to 1, the data-cloning algorithm provides Bayesian estimates based on flat priors.

The data-cloning algorithm computes the maximum partial-likelihood estimator (MPLE) and its standard error. In practice, we also need the value of the log-likelihood evaluated at the estimated parameters to compute various information criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (Burnham and Anderson 1998). We can easily compute the log-likelihood value at the MPLE  $\hat{\beta}$  using the formula:

$$\log L(\hat{\beta}, \underline{X}^U) = \sum_{i=1}^N \log \pi(\underline{X}_i^U, \hat{\beta}) - N \log \left[ \frac{1}{M} \sum_{j=1}^M \pi(\underline{X}_j^A, \hat{\beta}) \right]. \quad (3)$$

## COMPUTER PROGRAM

A computer program written in R programming language (R Development Core Team 2005) to fit various RSPF models is available from the author. The program can fit the exponential RSPF  $\pi(X, \beta) = \exp(X^T \beta)$ , the logistic RSPF  $\pi(X, \beta) = \exp(X^T \beta) / (1 + \exp(X^T \beta))$ , the log-log RSPF  $\pi(X, \beta) = \exp[-\exp(X^T \beta)]$ , and the probit RSPF  $\pi(X, \beta) = \Phi(X^T \beta)$  where  $\Phi(\cdot)$  denotes the cumulative distribution function for a standard, normal distribution and  $X^T \beta = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ . The parameter  $\beta_0$  represents the intercept, and the other parameters represent coefficients associated with environmental covariates. The program output consists of estimates of these parameters, their standard errors, and the value of the log-likelihood function at the estimated parameters.

The program runs under the Windows operating system. The user needs to install R software (R Development Core Team 2005) and WinBUGS (Spiegelhalter et al. 2004). In addition, the user needs to install R2WinBUGS, BRugs, and MASS packages within the R software.

Data input consists of 2 American Standard Code for Information Interchange (ASCII) text files corresponding to

**Table 1.** Simulation results showing equivalence of maximum partial-likelihood estimator (MPLE) and maximum likelihood estimator (MLE).<sup>a</sup>

True parameter <sup>b</sup>	MPLE				MLE	
	Data cloning		Nelder–Mead		Nelder–Mead	
	$\bar{x}$	SE	$\bar{x}$	SE	$\bar{x}$	SE
$\beta_0 = -1$	-1.003	0.4622	-0.99567	0.4619	-0.9559	0.4914
$\beta_1 = 0.5$	0.4924	0.2194	0.4902	0.2187	0.48298	0.2160
$\beta_2 = 0.8$	0.8186	0.2060	0.8160	0.2065	0.7987	0.2157

<sup>a</sup> MPLE and MLE based on 1,000 used and available points each;  $\bar{x}$  and SE based on 100 simulations.

<sup>b</sup> Log–log resource-selection probability function model  $\pi(X, \beta) = \exp[-\exp(X^T \beta)]$ .

the used points and the random sample of available points. Each row corresponds to the covariates of one location. The first entry in each row should be 1. No row names or column names are allowed. If categorical covariates are used, proper dummy variables should be used in the data input file. Columns should be separated by a space. These files should not be in the data-frame format.

In general, it is advisable to standardize the continuous covariate values by subtracting the mean and dividing by the standard deviation so that they range from approximately  $-3$  to  $3$ . Large positive or negative values of continuous covariates can destabilize the optimization routine. If the number of random samples from the available distribution is small, RSPF estimation can be unstable. As the number of covariates (i.e., no. of columns in the data matrices) increases, the number of samples from the available distribution should also increase to have a fair representation of the available distribution. Furthermore, the weighted distribution method works only if  $\geq 1$  of the covariates is continuous. If all covariates are categorical, only relative probabilities can be estimated (Lele and Keim 2006).

## RESULTS

We tested the validity of the method and the computer program using simulations. We generated 100 data sets under the use-versus-available study design and the log–log

RSPF model with 2 covariates, namely  $\pi(X, \beta) = \exp[-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2)]$ . The first covariate was binary, taking values zero and one (e.g., habitat types), and the second covariate was continuous. For each data set, we computed the MLE and the MPLE using the well-established Nelder–Mead optimization algorithm. For the same data set, we computed the MPLE using the data-cloning optimization algorithm. We used 10 clones and 5 chains, and the number of iterations was 5,000 with a burn-in period of 4,000. For a detailed explanation of the terms clones, burn-in period, number of iterations, number of chains, etc., see Lele et al. (2007) and Spiegelhalter et al. (2004). If the MPLE and MLE estimates are numerically close, these results support the claim that MPLE and MLE are equivalent. Furthermore, if the MPLE estimates using the Nelder–Mead algorithm and the data-cloning optimization algorithm are numerically close for every data set, then it indicates that data cloning is a valid method to optimize the partial-likelihood function. We found that both these claims were valid (Table 1). It is clear that MPLE is approximately unbiased, with mean and variance similar to the MLE.

We then reanalyzed the data set presented in Lele and Keim (2006) using the logistic RSPF model, namely  $\pi(X, \beta) = \exp(X^T \beta) / [1 + \exp(X^T \beta)]$  and the partial-likelihood-based method of estimation. A detailed description of the data set and biological conclusions are available in Lele and Keim (2006). The full data set consisted of 6,338 used points and a random sample of 15,000 points from the available distribution. We used the data-cloning algorithm on the full data set. We also used a 2-step approach described in the previous section. For the 2-step approach, we first considered a random sample of 1,000 used points and a random sample of 1,000 available points to obtain the initial estimates using the data-cloning algorithm. We then used these initial estimates as starting values in maximizing the Monte Carlo likelihood function using the Nelder–Mead algorithm and the full data set (Lele and Keim 2006). To check the effect of the subset size, we repeated the 2-step approach with a random sample of 2,000 used points and a random sample of 2,000 available points to obtain the initial

**Table 2.** Reanalysis of telemetry data on mountain goat (*Oreamnos americanus* de Blaineville 1816) resource use, showing the equivalence of the 2-step and the full-data maximum partial-likelihood estimator (MPLE).

Parameter <sup>a</sup>		2-Step estimation <sup>c</sup>						
		MPLE full-data DC <sup>b</sup>	Subset size $n = 1,000; M = 1,000$			Subset size $n = 2,000; M = 2000$		
			Initial estimate	Final MPLE	Final MLE	Initial estimate	Final MPLE	Final MLE
$\beta_0$	Estimate	-5.3275	-5.8423	-5.3263	-4.9999	-5.2318	-5.3257	-5.0006
	SE	0.1498	0.3703	0.1380	0.086	0.2946	0.1380	0.086
$\beta_1$	Estimate	2.2428	2.5932	2.2482	2.1188	2.3442	2.2426	2.1191
	SE	0.0895	0.3305	0.0939	0.064	0.2289	0.0938	0.064
$\beta_2$	Estimate	-0.0178	-0.0164	-0.01779	-0.01866	-0.0169	-0.0178	-0.01865
	SE	0.0005	0.0013	0.00049	0.00041	0.0010	0.00049	0.00041

<sup>a</sup> Estimated value and its SE.

<sup>b</sup> MPLE based on full data set and data-cloning (DC) maximization procedure.

<sup>c</sup> A subset of the data was used to get an initial estimator using the data-cloning (DC) maximization procedure. This initial estimator is used as a starting value for the Nelder–Mead optimization to obtain the MPLE and max. likelihood estimator (MLE) based on the full data set.

estimates using the data-cloning algorithm. As expected, the Nelder–Mead maximization–based MPLEs were nearly identical to MPLEs based on the full data set and data cloning (Table 2). Different subsets of the original data, along with a 2-step estimation procedure, led to exactly the same MPLEs and MLEs. This indicated that, for large data sets, the 2-step estimation procedure was a reasonable approach. We also point out that the MPLEs are slightly different than the MLEs.

In addition to fitting the logistic RSPF, we also fit log–log, probit, and exponential RSPF to these data. According to the AIC values, the logistic RSPF provided the best fit (AIC = –18,307.66) followed by the probit RSPF (AIC = –18,230.16), the exponential RSPF (–18,133.18), and the log–log RSPF (AIC = –18,102.41). This analysis indicated that it is useful to consider RSPF models more general than the commonly used exponential RSF.

## DISCUSSION

Under use-versus-available study design, it is possible to estimate probability of selection if one uses logistic, log–log, or probit RSPF models (Lele and Keim 2006). On the other hand, using the exponential RSF model leads only to relative probability of selection. Thus these general models of RSPF lead to stronger inferences than were feasible before. Aside from providing estimates of probability of selection, these models also expand the class of models that wildlife managers can use to study resource selection by animals. As shown in the data analysis section, these models sometimes describe the data better than the exponential RSF model. One of the main obstacles in using the general RSPF models, such as the logistic, log–log, and probit, was the lack of availability of a computer program to fit these models. We provided a numerically stable computational algorithm to accomplish this task. The researchers and managers no longer need to be restricted to using the exponential RSF model and can use other, general models of RSPF to study resource selection by animals.

## MANAGEMENT IMPLICATIONS

Understanding which resources are important to species under study is essential for making good management decisions. One of the important goals of resource selection studies is to estimate the probability that a particular environmental attribute or a combination of attributes will be selected by the species. If certain environmental attributes have high probability of selection, they are deemed important when managing the ecosystem. Up until now, researchers could infer the change in the probability of selection but not the probability of selection itself. From management perspective, this information is not very useful. For example, inferring that the construction of the road reduces the probability of selection by a factor of 5 is not sufficient; reducing the probability of selection from 0.9 to 0.18 is not the same as reducing the probability of selection

from 0.005 to 0.001, although in both cases the probability is changing by the same factor of 5. In the first situation, building a road has severe impact on the wildlife whereas in the second case, the impact is negligible, although the relative change is exactly the same. Thus, for effective management, it is important to know the probability of selection and not simply the relative probability of selection. We have now provided a methodology to obtain the probability of selection itself.

## ACKNOWLEDGMENTS

This research was partially supported by a grant from the National Scientific and Engineering Research Council of Canada. Thanks to C. Jerde and the researchers at the Yellowstone Ecological Research Center, Bozeman, Montana, USA, who tested the program and provided feedback on the manuscript. J. Keim of Matrix Solutions, Inc. provided the telemetry data on mountain goats. I am grateful to the Editor, Associate Editor, and 2 referees for many useful suggestions that have improved the manuscript substantially.

## LITERATURE CITED

- Boyce, M., and L. McDonald. 1999. Relating populations to habitats using resource selection function. *Trends in Ecology and Evolution* 14:268–272.
- Burnham, K. P., and D. R. Anderson. 1998. *Model selection and inference: a practical information theoretic approach*. Springer-Verlag, New York, USA.
- Gilbert, P., S. Lele, and Y. Vardi. 1999. Semiparametric inference for selection bias models with application to AIDS vaccine trials. *Biometrika* 86:27–43.
- Johnson, C. J., S. E. Nielsen, E. H. Merrill, T. L. McDonald, and M.S. Boyce. 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation method. *Journal of Wildlife Management* 70:347–357.
- Keating, K. A., and S. Cherry. 2004. Use and interpretation of logistic regression in habitat selection studies. *Journal of Wildlife Management* 68:774–789.
- Lele, S. R., B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters* 10:551–563.
- Lele, S. R., and J. L. Keim. 2006. Weighted distributions and estimation of resource selection probability functions. *Ecology* 87:3021–3028.
- Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. 2002. *Resource selection by animals: statistical analysis and design for field studies*. Second edition. Kluwer Press, Boston, Massachusetts, USA.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. *Numerical recipes: the art of scientific computing*. Cambridge University Press, Cambridge, United Kingdom.
- R Development Core Team. 2005. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Savage, R. E. 1931. The relation between the feeding of the herring off the east coast of England and the plankton of the surrounding waters. *Fishery Investigation, Ministry of Agriculture, Food and Fisheries, Series 2*, 12: 1–88.
- Spiegelhalter, D., A. Thomas, and N. Best. 2004. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Institute of Public Health, London, United Kingdom.

## APPENDIX: DERIVATION OF THE PARTIAL LIKELIHOOD FUNCTION

Let  $\underline{X}_i^U, i = 1, 2, \dots, N$  denote the data set corresponding to the used points and  $\underline{X}_j^A, j = 1, 2, \dots, M$  denote the data set corresponding to a random sample from the available distribution. Let  $w = N/(M + N)$ . The likelihood function can be written as

$$\begin{aligned} \prod_{i=1}^N f^U(\underline{X}_i^U) \prod_{j=1}^M f^A(\underline{X}_j^A) &= \prod_{i=1}^N \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)} \prod_{j=1}^M f^A(\underline{X}_j^A) \\ &= \prod_{i=1}^N \frac{\frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)}}{w \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)} + (1-w) f^A(\underline{X}_i^U)} \times \prod_{j=1}^M \frac{f^A(\underline{X}_j^A)}{w \frac{\pi(\underline{X}_j^A, \beta) f^A(\underline{X}_j^A)}{P(\beta)} + (1-w) f^A(\underline{X}_j^A)} \\ &\times \prod_{i=1}^N \left[ w \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)} + (1-w) f^A(\underline{X}_i^U) \right] \times \prod_{j=1}^M \left[ w \frac{\pi(\underline{X}_j^A, \beta) f^A(\underline{X}_j^A)}{P(\beta)} + (1-w) f^A(\underline{X}_j^A) \right] \end{aligned} \quad (1)$$

The first term in the above product can be written as

$$\prod_{i=1}^N \frac{\frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)}}{w \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)} + (1-w) f^A(\underline{X}_i^U)} = \prod_{i=1}^N \frac{\frac{\pi(\underline{X}_i^U, \beta)}{P(\beta)}}{w \frac{\pi(\underline{X}_i^U, \beta)}{P(\beta)} + (1-w)} = \frac{1}{w^N} \prod_{i=1}^N \frac{w \pi(\underline{X}_i^U, \beta)}{w \pi(\underline{X}_i^U, \beta) + (1-w) P(\beta)} \quad (2)$$

Similar algebraic manipulations lead to representing the second term as

$$\prod_{j=1}^M \frac{f^A(\underline{X}_j^A)}{w \frac{\pi(\underline{X}_j^A, \beta) f^A(\underline{X}_j^A)}{P(\beta)} + (1-w) f^A(\underline{X}_j^A)} = \prod_{j=1}^M \frac{1}{w \frac{\pi(\underline{X}_j^A, \beta)}{P(\beta)} + (1-w)} = \frac{1}{(1-w)^M} \prod_{j=1}^M \frac{(1-w) P(\beta)}{w \pi(\underline{X}_j^A, \beta) + (1-w) P(\beta)} \quad (3)$$

The full-likelihood function can be expressed as a product of the following 2 terms:

$$PL(\beta) = \prod_{i=1}^N \frac{w \pi(\underline{X}_i^U, \beta)}{w \pi(\underline{X}_i^U, \beta) + (1-w) P(\beta)} \prod_{j=1}^M \frac{(1-w) P(\beta)}{w \pi(\underline{X}_j^A, \beta) + (1-w) P(\beta)} \quad (4)$$

and

$$\begin{aligned} \text{Remainder term} &= \frac{1}{w^N} \frac{1}{(1-w)^M} \prod_{i=1}^N \left[ w \frac{\pi(\underline{X}_i^U, \beta) f^A(\underline{X}_i^U)}{P(\beta)} + (1-w) f^A(\underline{X}_i^U) \right] \\ &\times \prod_{j=1}^M \left[ w \frac{\pi(\underline{X}_j^A, \beta) f^A(\underline{X}_j^A)}{P(\beta)} + (1-w) f^A(\underline{X}_j^A) \right] \end{aligned} \quad (5)$$

Let us rewrite the first term, namely,

$$PL(\beta, \alpha) = \prod_{i=1}^N \left( \frac{w \pi(\underline{X}_i^U, \beta)}{w \pi(\underline{X}_i^U, \beta) + (1-w) P(\beta)} \right) \prod_{j=1}^M \left( \frac{(1-w) \alpha}{w \pi(\underline{X}_j^A, \beta) + (1-w) P(\beta)} \right). \quad (6)$$

by suppressing the dependence of  $P(\beta)$  on  $\beta$  and treating it as a free parameter  $\alpha \in (0, 1)$  as follows:

$$PL(\beta, \alpha) = \prod_{i=1}^N \left( \frac{w\pi(\underline{X}_i^U, \beta)}{w\pi(\underline{X}_i^U, \beta) + (1-w)\alpha} \right) \prod_{k=1}^N \left( \frac{(1-w)\alpha}{w\pi(\underline{X}_i^A, \beta) + (1-w)\alpha} \right). \quad (7)$$

The results in Gilbert et al. (1999) show that maximizing this function with respect to the parameters  $(\beta, \alpha)$ , under the constraint  $0 < \alpha \leq 1$ , leads to the same estimators asymptotically as those that would have been obtained by maximizing the full-likelihood function.

*Associate Editor: McCorquodale.*