Interval estimation of population means under unknown but bounded probabilities of sample selection

Peter M. Aronow and Donald K. K. Lee*

Preprint, forthcoming at *Biometrika*

November 23, 2012

Abstract

Applying concepts from partial identification to the domain of finite population sampling, we propose a method for interval estimation of a population mean when the probability of sample selection lies within a posited interval. The interval estimate is derived from sharp bounds on the Hajek (1971) estimator of the population mean. We demonstrate the method's utility for sensitivity analysis using a sample of needles collected as part of a syringe tracking and testing program in New Haven, Connecticut.

1 Introduction

Analysts often contend with samples from designs where the probabilities of sample selection are unknown. The question we address here is: what can be learned about the population mean using only limited information about the sampling design? We propose a method for interval estimation of a population mean whenever the probability of sample selection is posited to lie within a closed interval $[\alpha, \beta] \subseteq (0, 1]$.

Our approach serves two primary purposes for applied practitioners. First, when there exists a theoretical basis for bounding the probability of sample selection, the proposed method estimates sharp bounds for the population mean. Second, this estimator provides a means of sensitivity analysis for estimation strategies that presuppose knowledge of the probability of sample selection. For example, when there exists doubt about whether the entire population was covered with equal probability, the method provides a nonparametric way for researchers to assess how robust their conclusions are to deviations from equal probability sampling.

*PM Aronow, Department of Political Science, Yale University, New Haven CT 06520 (peter.aronow@yale.edu); DKK Lee, School of Management, Yale University, New Haven CT 06520 (donald.lee@yale.edu).

2 Setting

Consider a population U indexed by $i = 1, \ldots, N$, with associated responses $y_i \in \mathbb{R}$. If unit i is sampled from the population, let $I_i = 1$; otherwise $I_i = 0$. Under a given sampling design, the probability of i being selected is π_i so that $\mathrm{pr}(I_i = 1) = \pi_i$. Without loss of generality, assume an index ordering such that the first n units are the sampled units and the remaining $N - n$ are the unsampled units. That is, $I_i = 1$ for $i = 1, \ldots, n$, and $I_i = 0$ for $i = n+1, \ldots, N$, so that the set of sampled units is $S = \{1, \ldots, n\}$. Further let the units in S be ordered so that $y_1 \leq \cdots \leq y_n$. The goal is estimation of the population mean, $\mu = \sum_{i=1}^N y_i/N$.

When the probabilities of sample selection are known, and if $\pi_i > 0$ for all i, then the Horvitz–Thompson estimator of μ, $\hat{\mu}_{HT} = \sum_{i=1}^N (I_i y_i)/(N\pi_i)$ is unbiased. Similarly, we may define a random variable that has expectation 1: $\hat{1}_{HT} = \sum_{i=1}^N I_i/(N\pi_i)$. The Hajek (1971) ratio estimator of μ is then

$$\hat{\mu} = \frac{\hat{\mu}_{HT}}{\hat{1}_{HT}} = \frac{\sum_{i \in S} \pi_i^{-1} y_i}{\sum_{i \in S} \pi_i^{-1}} = \frac{w'\mathbf{y}}{w' 1_n}, \tag{1}$$

where $w = (w_1, \ldots, w_n) = \left(\pi_1^{-1}, \ldots, \pi_n^{-1}\right)$, $\mathbf{y} = (y_1, \ldots, y_n)$, and 1_n is an n-vector of ones. Under mild conditions (Lehmann, 1999, Ch. 2), $\hat{\mu}$ is consistent for μ.

3 Unknown but bounded probabilities of sample selection

3.1 Bounds for $\hat{\mu}$

Suppose we no longer have exact knowledge of the sampling design, instead knowing only that, for all i, $0 < \alpha \leq \pi_i \leq \beta \leq 1$. While consistent point estimation is no longer possible without further assumptions, it is possible to obtain an interval estimate for μ. In fact, sharp bounds for the Hajek estimator (1) exist, in the sense that the set of all possible values for $\hat{\mu}$ is precisely the interval $\left[\hat{\mu}^L, \hat{\mu}^H\right]$. If $\pi_i \in [\alpha, \beta]$, an upper bound on $\hat{\mu}$ is

$$\hat{\mu}^H = \max_w \left\{ \frac{w'\mathbf{y}}{w'1_n} : 1/\beta \leq w_1, \ldots, w_n \leq 1/\alpha \right\}. \tag{2}$$

Likewise a lower bound is

$$\hat{\mu}^L = \min_w \left\{ \frac{w'\mathbf{y}}{w'1_n} : 1/\beta \leq w_1, \ldots, w_n \leq 1/\alpha \right\}. \tag{3}$$

Proposition 1. *Let $\gamma = \beta/\alpha \geq 1$ be the relative magnitude of β compared to α. An optimal solution to* (2) *is $w^H = (1/\beta, \ldots, 1/\beta, 1/\alpha, \ldots, 1/\alpha)$ where the first k^H entries equal $1/\beta$, the remaining $n - k^H$ entries equal $1/\alpha$, and k^H is the smallest non-negative integer k*

satisfying $\{\sum_{i=1}^{k} y_i + \gamma \sum_{i=k+1}^{n} y_i\}/\{k + \gamma(n-k)\} \le y_{k+1}$*. The upper bound on* $\hat{\mu}$ *is therefore*

$$\hat{\mu}^H = \frac{\sum_{i=1}^{k^H} y_i + \gamma \sum_{i=k^H+1}^{n} y_i}{k^H + \gamma(n - k^H)}. \tag{4}$$

An optimal solution to (3) *is* $w^L = (1/\alpha, \ldots, 1/\alpha, 1/\beta, \ldots, 1/\beta)$ *where the first* k^L *entries equal* $1/\alpha$*, the remaining* $n - k^L$ *entries equal* $1/\beta$*, and* k^L *is the smallest non-negative integer* k *satisfying* $\{\gamma \sum_{i=1}^{k} y_i + \sum_{i=k+1}^{n} y_i\}/\{\gamma k + (n-k)\} \ge y_{k+1}$*. The lower bound on* $\hat{\mu}$ *is therefore*

$$\hat{\mu}^L = \frac{\gamma \sum_{i=1}^{k^L} y_i + \sum_{i=k^L+1}^{n} y_i}{\gamma k^L + (n - k^L)}.$$

The Hajek estimator, $\hat{\mu}$, can attain either bound $\hat{\mu}^L$ or $\hat{\mu}^H$ since it is possible for $(\pi_1^{-1}, \ldots, \pi_n^{-1})$ to be equal to w^L or w^H, or any convex combination of the two. Thus $[\hat{\mu}^L, \hat{\mu}^H]$ is the set of all possible values of $\hat{\mu}$:

$$\{\hat{\mu} : \alpha \le \pi_1, \ldots, \pi_N \le \beta\} = \left[\frac{\gamma \sum_{i=1}^{k^L} y_i + \sum_{i=k^L+1}^{n} y_i}{\gamma k^L + (n - k^L)}, \frac{\sum_{i=1}^{k^H} y_i + \gamma \sum_{i=k^H+1}^{n} y_i}{k^H + \gamma(n - k^H)}\right]. \tag{5}$$

The bounds do not depend on the actual values of α or β, but only on γ. When $\gamma = 1$, there exists equality on the probability of sample selection. Higher values of γ represent a greater potential for variation in the probability of sample selection; $\gamma = 5$ indicates that some units may be as much as 5 times more likely to be sampled than others. Like methods for sensitivity analysis in other contexts (Robins et al., 1994; Rosenbaum, 2002), observing how the plausible values of $\hat{\mu}$ change with γ allows researchers to assess the sensitivity of their results to deviations from the assumption of equal probability sampling, which necessarily depends on the features of the data at hand.

A property of the optimal weights w^H and w^L is that observations with identical outcomes share the same weight. The following lemma formalizes this and is important for establishing the asymptotic properties of $\hat{\mu}^H$ and $\hat{\mu}^L$.

Lemma 1. *If* $y_k = y_{k+1}$ *then* $w_k^H = w_{k+1}^H$ *and* $w_k^L = w_{k+1}^L$.

A simple closed form solution for the bounds (5) is available when $\mathbf{y}$ is dichotomous. Assuming $y_i \in \{0, 1\}$, define $p = \sum_{i=1}^{n} y_i/n$. Then, by application of Lemma 1,

$$\{\hat{\mu} : \alpha \le \pi_1, \ldots, \pi_N \le \beta, y_i \in \{0,1\}\} = \left[\frac{p}{\gamma(1-p) + p}, \frac{\gamma p}{(1-p) + \gamma p}\right].$$

3.2 Asymptotic sharpness of estimated bounds

The proposed interval estimator yields asymptotically sharp bounds on μ itself under the scaling proposed by Brewer (1979) and articulated in Assumption 1.

Assumption 1. *The sequences of finite populations $\{U_h\}_{h=1}^{\infty}$ and samples $\{S_h\}_{h=1}^{\infty}$ grow as follows: $U_h = \{U_{h-1}, U\} = \{U, \dots, U\}$, that is, h replicates of the original population U so that $|U_h| = hN$. A sample s_h is drawn from U according to the original sampling design, independent of $s_1, \dots, s_{h-1}$. The sample $S_h = \{S_{h-1}, s_h\} = \{s_1, \dots, s_h\}$. The estimators $\hat{\mu}_h^H$ and $\hat{\mu}_h^L$ follow equations* (2) *and* (3) *respectively, but as applied to the population U_h and sample S_h.*

Let Y be the ordered set of distinct outcomes in U so that $y_{(1)} < \cdots < y_{(|Y|)}$, with equivalence classes $\Upsilon_j \equiv \{i = 1, \dots, N : y_i = y_{(j)}\}$. In the limit of the Brewer scaling, the πs can be identified up to units with identical outcomes: as $h \to \infty$, $(\sum_{\{i=1,\dots,Nh: y_i = y_{(j)}\}} I_i)/(\sum_{i=1}^{Nh} I_i) \to \Pi_j/(\sum_{k=1}^{|Y|} \Pi_k)$ where $\Pi_j = \sum_{i \in \Upsilon_j} \pi_i$. Since $\pi_i \geq \alpha > 0$ for all i, Y is identified in the limit. If $W_j \equiv |\Upsilon_j|/\Pi_j$ is known for all $j = 1, \dots, |Y|$, μ is identified:

$$\mu = \frac{\sum_{j=1}^{|Y|} W_j \Pi_j y_{(j)}}{\sum_{j=1}^{|Y|} W_j \Pi_j} = \frac{\sum_{j=1}^{|Y|} W_j \frac{\Pi_j}{\sum_{k=1}^{|Y|} \Pi_k} y_{(j)}}{\sum_{j=1}^{|Y|} W_j \frac{\Pi_j}{\sum_{k=1}^{|Y|} \Pi_k}}.$$

However, W_j is only identified up to the interval $\beta^{-1} \leq W_j \leq \alpha^{-1}$. In the limit, μ can be identified up to the interval bounded by

$$\overline{\mu} = \max_{W} \left\{ \frac{\sum_{j=1}^{|Y|} W_j \Pi_j y_{(j)}}{\sum_{j=1}^{|Y|} W_j \Pi_j} : 1/\beta \leq W_1, \dots, W_{|Y|} \leq 1/\alpha \right\}, \tag{6}$$

$$\underline{\mu} = \min_{W} \left\{ \frac{\sum_{j=1}^{|Y|} W_j \Pi_j y_{(j)}}{\sum_{j=1}^{|Y|} W_j \Pi_j} : 1/\beta \leq W_1, \dots, W_{|Y|} \leq 1/\alpha \right\}. \tag{7}$$

Proposition 2. *In the limit $h \to \infty$, the bound $\underline{\mu} \leq \mu \leq \overline{\mu}$ on μ is sharp. Specifically,*

$$\overline{\mu} = \frac{\sum_{j=1}^{\overline{k}} \Pi_j y_{(j)} + \gamma \sum_{j=\overline{k}+1}^{|Y|} \Pi_j y_{(j)}}{\sum_{j=1}^{\overline{k}} \Pi_j + \gamma \sum_{j=\overline{k}+1}^{|Y|} \Pi_j}, \ \underline{\mu} = \frac{\gamma \sum_{j=1}^{\underline{k}} \Pi_j y_{(j)} + \sum_{j=\underline{k}+1}^{|Y|} \Pi_j y_{(j)}}{\gamma \sum_{j=1}^{\underline{k}} \Pi_j + \sum_{j=\underline{k}+1}^{|Y|} \Pi_j},$$

where $\overline{k}$ is the smallest non-negative integer k satisfying $(\sum_{j=1}^{k} \Pi_j y_{(j)} + \gamma \sum_{j=k+1}^{|Y|} \Pi_j y_{(j)})/(\sum_{j=1}^{k} \Pi_j + \gamma \sum_{j=k+1}^{|Y|} \Pi_j) \leq y_{(k+1)}$, and $\underline{k}$ is the smallest non-negative integer k satisfying $(\gamma \sum_{j=1}^{k} \Pi_j y_{(j)} + \sum_{j=k+1}^{|Y|} \Pi_j y_{(j)})/(\gamma \sum_{j=1}^{k} \Pi_j + \sum_{j=k+1}^{|Y|} \Pi_j) \geq y_{(k+1)}$. The optimal solution to (6) *is $\overline{W} = (1/\beta, \dots, 1/\beta, 1/\alpha, \dots, 1/\alpha)$, where the first $\overline{k}$ entries equal $1/\beta$ and the remaining $|Y| - \overline{k}$ entries equal $1/\alpha$. The optimal solution to* (7) *is $\underline{W} = (1/\alpha, \dots, 1/\alpha, 1/\beta, \dots, 1/\beta)$, where the first $\underline{k}$ entries equal $1/\alpha$ and the remaining $|Y| - \underline{k}$ entries equal $1/\beta$.*

By Proposition 2, if $\hat{\mu}_h^H \to \overline{\mu}$ and $\hat{\mu}_h^L \to \underline{\mu}$ as $h \to \infty$, then $[\hat{\mu}_h^L, \hat{\mu}_h^H]$ is asymptotically sharp. A mild technical condition on $\overline{k}$ and $\underline{k}$ is necessary to guarantee convergence.

Assumption 2. *For any* $k \geq 0$, $y_{(k+1)} \neq (\sum_{j=1}^{k} \Pi_j y_{(j)} + \gamma \sum_{j=k+1}^{|Y|} \Pi_j y_{(j)})/(\sum_{j=1}^{k} \Pi_j + \gamma \sum_{j=k+1}^{|Y|} \Pi_j)$, $y_{(k+1)} \neq (\gamma \sum_{j=1}^{k} \Pi_j y_{(j)} + \sum_{j=k+1}^{|Y|} \Pi_j y_{(j)})/(\gamma \sum_{j=1}^{k} \Pi_j + \sum_{j=k+1}^{|Y|} \Pi_j)$.

Proposition 3. *If Assumptions 1 and 2 hold, then, as* $h \to \infty$, $\mu_h^H \to \overline{\mu}$ *and* $\mu_h^L \to \underline{\mu}$ *almost everywhere.*

4 Application

Kaplan (1991) reports prevalence rates of HIV proviral DNA in needles tested as part of New Haven, Connecticut's syringe tracking and testing program in 1990–1991. Needles were collected from three populations: needles from shooting galleries, locations where drug users inject with rented, often shared, equipment; needles from the street; and needles introduced by a needle exchange program. A sample of tested needles is available for each population. For the shooting gallery population, a sample of 48 needles was collected from a shooting gallery and tested. Of these, 91·7% of needles tested positive for HIV. For the street and needle exchange populations, program clients voluntarily returned needles; 4236 street needles and 5263 exchange needles were returned by program clients. Of those returned, 160 street needles and 579 exchange needles were tested, of which 67·5% and 50·3% tested positive respectively.

Due to uncertainty about the selection processes for collection and testing, we assess the sensitivity of the estimates to deviations from the assumption of equal probability sampling. In Table 1, we report the interval estimates for each population associated with varying levels of γ. When $\gamma = 1$, we reproduce the original point estimates reported by Kaplan (1991), while values of γ greater than one reflect the possibility of unequal probability sampling. In order for the estimates for the shooting gallery and needle exchange populations to overlap, $\gamma > 3$, and for the street needles and needle exchange populations to overlap, $\gamma > 1{\cdot}25$. That is, for the shooting gallery needles to possibly have the same HIV prevalence rate as the exchange needles, infected shooting gallery needles must have been more than 3 times as likely to have been tested than uninfected shooting gallery needles, with the converse holding for exchange needles.

5 Discussion

We note three possible direct extensions of this method. First, when the response variable suffers from nonresponse or corruption, the method could be combined with other methods for partial identification, such as those detailed in Manski (2003). Second, sensitivity analysis could be employed after model-based weights and/or a regression model (e.g., the Cassel et al., 1976, generalized regression estimator) were applied, though the interpretation of γ would change accordingly. Third, additional assumptions, including monotonicity assumptions about the relationship between π_i and y_i, could be used to shrink the bounds further.

Table 1: Estimated bounds for prevalence rates (%) of HIV proviral DNA in syringes collected as part of New Haven syringe tracking and testing program

Shooting Gallery			Street			Needle Exchange		
γ	$\hat{\mu}^L$	$\hat{\mu}^H$	γ	$\hat{\mu}^L$	$\hat{\mu}^H$	γ	$\hat{\mu}^L$	$\hat{\mu}^H$
1	91·7	91·7	1	67·5	67·5	1	50·3	50·3
1·25	89·8	93·2	1·25	62·4	72·2	1·25	44·7	55·9
1·5	88·0	94·3	1·5	58·1	75·7	1·5	40·3	60·3
2	84·7	95·7	2	50·9	80·6	2	33·6	66·9
2·5	81·5	96·5	2·5	45·4	83·9	2·5	28·8	71·7
3	78·6	97·1	3	40·9	86·2	3	25·2	75·2
4	73·4	97·8	4	34·2	89·3	4	20·2	80·2
n	48		n	160		n	579	

Acknowledgements

The authors thank Don Green, Malte Lierl, Chuck Manski, Walter Mebane, Joel Middleton, Cyrus Samii, Allie Sovey, Ed Vytlacil, and Baobao Zhang for helpful comments. Special thanks to Ed Kaplan for valuable suggestions.

Appendix

Proof of Proposition 1

Every $w \in [1/\beta, 1/\alpha]^n$ has a corresponding point $v = w/w'1_n$ on the simplex $\Sigma = \{v \in \mathbb{R}^n_+ : v'1_n = 1\}$ that satisfies $w'y/w'1_n = v'y$. The optimization problem (2) can be transformed into $\max_v v'\mathbf{y}$ such that $v \in \Sigma \cap \Phi$, where Φ is the smallest cone in $\mathbb{R}^n$ containing the hypercube $[1/\beta, 1/\alpha]^n$. The optimal solution set to this linear program must contain an extreme point of $\Sigma \cap \Phi$. Intersecting the ray that passes through this point with $[1/\beta, 1/\alpha]^n$ yields the optimal solution set to (2), which must also contain a vertex of $[1/\beta, 1/\alpha]^n$. The structure of this vertex solution must be of the form $w^H = (1/\beta, \ldots, 1/\beta, 1/\alpha, \ldots, 1/\alpha)$ where the first $k^H \geq 0$ entries are $1/\beta$ and the last $n - k^H$ entries are $1/\alpha$. For otherwise there exists $w_i^H = 1/\alpha$ and $w_{i+1}^H = 1/\beta$. Switching the values of w_i^H and w_{i+1}^H can only increase the value of $w^{H\prime}\mathbf{y}/w^{H\prime}1_n$, proving the assertion.

Given the structure of w^H, it remains to show that the optimal objective value of (2) is (4). The first order condition for k^H to maximize (2) can be characterized as follows. Let $w^k = (1/\beta, \ldots, 1/\beta, 1/\alpha, \ldots, 1/\alpha)$ where the first k entries are $1/\beta$ and the last $n-k$ are $1/\alpha$. Then $(w^{k+1\prime}\mathbf{y})/(w^{k+1\prime}1_n) - (w^{k\prime}\mathbf{y})/(w^{k\prime}1_n) \leq 0$ implies that

$$\frac{\sum_{i=1}^{k} y_i + \gamma \sum_{i=k+1}^{n} y_i}{k + \gamma(n-k)} \leq y_{k+1}.$$

If $(w^{k\prime}\mathbf{y})/(w^{k\prime}1_n) \geq (w^{k+1\prime}\mathbf{y})/(w^{k+1\prime}1_n)$, then $(w^{k+1\prime}\mathbf{y})/(w^{k+1\prime}1_n) \geq (w^{k+2\prime}\mathbf{y})/(w^{k+2\prime}1_n)$ since

$$\begin{aligned}\frac{\sum_{i=1}^{k+1} y_i + \gamma \sum_{i=k+2}^{n} y_i}{k+1+\gamma(n-k-1)} &= \frac{\sum_{i=1}^{k} y_i + \gamma \sum_{i=k+1}^{n} y_i + (1-\gamma)\, y_{k+1}}{k+1+\gamma(n-k-1)} \\ &\leq \frac{[k+\gamma(n-k)]y_{k+1} + (1-\gamma)\, y_{k+1}}{k+1+\gamma(n-k-1)} = y_{k+1} \leq y_{k+2}.\end{aligned}$$

Thus

$$k^H = \min\left\{k \geq 0 : \frac{w^{k\prime}\mathbf{y}}{w^{k\prime}1_n} \geq \frac{w^{k+1\prime}\mathbf{y}}{w^{k+1\prime}1_n}\right\} = \min\left\{k \geq 0 : \frac{\sum_{i=1}^{k} y_i + \gamma \sum_{i=k+1}^{n} y_i}{k+\gamma(n-k)} \leq y_{k+1}\right\}.$$

The solution to (3) can be derived in a similar manner. $\square$

References

BREWER, K.R.W. (1979). A class of robust sampling designs for large-scale surveys. *J. Am. Statist. Assoc.* **74,** 911–915.

CASSEL, C.M., SÄRNDAL, C.-E. & WRETMAN, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika* **63,** 615–620.

HAJEK, J. (1971). Comment on "An Essay on the Logical Foundations of Survey Sampling, Part One." In *The Foundations of Survey Sampling*, Eds. V.P. Godambe, V.P. and D.A. Sprott, p. 246. Toronto: Holt, Rinehart, and Winston.

KAPLAN, E.H. (1991). Evaluating needle-exchange programs via syringe tracking and testing (STT). *AIDS & Public Policy Journal* **6,** 2–8.

LEHMANN, E.L. (1999). *Elements of Large Sample Theory*. New York: Springer-Verlag.

MANSKI, C.F. (2003). *Partial Identification of Probability Distributions*. New York: Springer.

ROBINS, J. M., ROTNITZKY, A. & SCHARFSTEIN, D.O. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Methods in Epidemiology*, Ed. E. Halloran, pp. 1–92. New York: Springer-Verlag.

ROSENBAUM, P.R. (2002). *Observational Studies*. 2nd ed. New York: Springer.

Supplementary Information

Proof of Lemma 1

We establish the result for w^H by considering two separate cases. First, if $y_1 = \ldots = y_n$, then $w^H = (1/\alpha, \ldots, 1/\alpha)$, in which case the result is clearly true. Otherwise at least two outcomes differ, hence by the characterization of k^H in Proposition 1, $k^H > 0$ since

$$\frac{\sum_{i=1}^{0} y_i + \gamma \sum_{i=1}^{n} y_i}{0 + \gamma(n-0)} = \frac{\sum_{i=1}^{n} y_i}{n} > y_1.$$

It then suffices to show that if $k \leq k^H$, then $k+1 \leq k^H$. If $k \leq k^H$,

$$\frac{\sum_{i=1}^{k-1} y_i + \gamma \sum_{i=k}^{n} y_i}{k - 1 + \gamma(n-k+1)} > y_k.$$

It follows that

$$\begin{aligned}\frac{\sum_{i=1}^{k} y_i + \gamma \sum_{i=k+1}^{n} y_i}{k + \gamma(n-k)} &= \frac{(1-\gamma)y_k + \sum_{i=1}^{k-1} y_i + \gamma \sum_{i=k}^{n} y_i}{k + \gamma(n-k)} \\ &> \frac{(1-\gamma)y_k + (k-1+\gamma(n-k+1))y_k}{k+\gamma(n-k)} = y_k = y_{k+1},\end{aligned}$$

so $k+1$ is also less than or equal to k^H. The result for w^L can also be established in the same manner. $\square$

Proof of Proposition 2

We will demonstrate the result for $\overline{\mu}$ only, as the result for $\underline{\mu}$ can be established in the same fashion. Like (2), problem (6) can be transformed into a linear program, hence there exists an optimal solution at a vertex of $[1/\beta, 1/\alpha]^{|Y|}$.

The structure of this vertex solution is of the form $\overline{W} = (1/\beta, \ldots, 1/\beta, 1/\alpha, \ldots, 1/\alpha)$ where the first $\overline{k} \geq 0$ entries are $1/\beta$ and the last $|Y| - \overline{k}$ are $1/\alpha$. To see this, assume that there exists $(\overline{W}_k, \overline{W}_{k+1}) = (1/\alpha, 1/\beta)$. We will show that replacing the value of the pair with either $(1/\beta, 1/\beta)$ or $(1/\alpha, 1/\alpha)$ will lead to an objective value for (6) that is strictly larger. Define $c_1 \equiv \sum_{j \neq \{k,k+1\}} W_j \Pi_j y_{(j)}$, $c_2 \equiv \sum_{j \neq \{k,k+1\}} W_j \Pi_j$, and

$$\mu(W_k, W_{k+1}) \equiv \frac{c_1 + W_k \Pi_k y_{(k)} + W_{k+1}\Pi_{k+1} y_{(k+1)}}{c_2 + W_k \Pi_k + W_{k+1}\Pi_{k+1}}.$$

Then the condition $\mu(1/\beta, 1/\beta) > \mu(1/\alpha, 1/\beta)$ is equivalent to

$$c_2 y_{(k)} - c_1 + \beta^{-1}\Pi_{k+1}(y_{(k)} - y_{(k+1)}) < 0 \qquad \text{(B1)}$$

and the condition $\mu(1/\alpha, 1/\alpha) > \mu(1/\alpha, 1/\beta)$ is equivalent to

$$c_1 - c_2 y_{(k+1)} + \alpha^{-1}\Pi_k(y_{(k)} - y_{(k+1)}) < 0. \tag{B2}$$

The assertion is true if either (B1) or (B2) holds. To show this, observe that the sum of the left hand sides of (B1) and (B2) is negative, so at least one of them has to be negative:

$$\begin{aligned}&\{c_2 y_{(k)} - c_1 + \beta^{-1}\Pi_{k+1}(y_{(k)} - y_{(k+1)})\} + \{c_1 - c_2 y_{(k+1)} + \alpha^{-1}\Pi_k(y_{(k)} - y_{(k+1)})\} \\ &= \big(c_2 + \alpha^{-1}\Pi_k + \beta^{-1}\Pi_{k+1}\big)\{y_{(k)} - y_{(k+1)}\} < 0.\end{aligned}$$

Given the structure of $\overline{W}$, $\overline{k}$ can be characterized using the same approach used to characterize k^H in the proof of Proposition 1. □

Proof of Proposition 3

Fix a sample realization $\omega \in \Omega = \{0,1\}^{\mathbb{N}}$ so that $I_1(\omega), I_2(\omega), \dots$ is a deterministic binary sequence. We drop the explicit dependence on ω for brevity. For a $W \in [\beta^{-1}, \alpha^{-1}]^{|Y|}$, define

$$\mu_h(W) \equiv \frac{\sum_{j=1}^{|Y|} W_j y_{(j)} \sum_{\{i=1,\dots,Nh: y_i = y_{(j)}\}} I_i}{\sum_{j=1}^{|Y|} W_j \sum_{\{i=1,\dots,Nh: y_i = y_{(j)}\}} I_i}.$$

By Lemma 1, the maximization problem (2) yields the same optimum as maximizing $\mu_h(W)$. Denote $P_{h,j} \equiv h^{-1}\sum_{\{i=1,\dots,Nh: y_i = y_{(j)}\}} I_i$. By the strong law of large numbers $P_{h,j} \to \Pi_j$ as $h \to \infty$. Therefore

$$\hat{\mu}_h^H = \max_{W \in [\beta^{-1}, \alpha^{-1}]^{|Y|}} \mu_h(W), \tag{B3}$$

$$\mu_h(W) = \frac{\sum_{j=1}^{|Y|} W_j P_{h,j} y_{(j)}}{\sum_{j=1}^{|Y|} W_j P_{h,j}} \to \frac{\sum_{j=1}^{|Y|} W_j \Pi_j y_{(j)}}{\sum_{j=1}^{|Y|} W_j \Pi_j}. \tag{B4}$$

For a sufficiently large h, $P_{h,j}$ will be close enough to Π_j for the optimal solution to (B3) to be identical to the optimal solution to (6), $\overline{W}$, as characterized in Proposition 2. Formally, Assumption 2 implies that there exists h_0 such that, for all $h \geq h_0$,

$$\frac{\sum_{j=1}^{\overline{k}-1} P_{h,j} y_{(j)} + \gamma \sum_{j=\overline{k}}^{|Y|} P_{h,j} y_{(j)}}{\sum_{j=1}^{\overline{k}-1} P_{h,j} + \gamma \sum_{j=\overline{k}}^{|Y|} P_{h,j}} > y_{(\overline{k})},$$

$$\frac{\sum_{j=1}^{\overline{k}} P_{h,j} y_{(j)} + \gamma \sum_{j=\overline{k}+1}^{|Y|} P_{h,j} y_{(j)}}{\sum_{j=1}^{\overline{k}} P_{h,j} + \gamma \sum_{j=\overline{k}+1}^{|Y|} P_{h,j}} < y_{(\overline{k}+1)}$$

hence $\hat{\mu}_h^H = \mu_h(\overline{W})$ for all $h \geq h_0$ in view of the characterization of $\overline{k}$ in Proposition 2. Then (B4) implies that $\hat{\mu}_h^H \to \overline{\mu}$ along each sample path ω. The convergence of $\hat{\mu}_h^L$ can be established in the same manner. □