Appendices to "Sequential Monte Carlo for Sampling Balanced and Compact Redistricting Plans"

Cory McCartan* Kosuke Imai†

March 30, 2023

A Proofs of Propositions

Lemma 4.1. *The probability of splitting a* valid *new district G_i from an existing area $\widetilde{G}_{i-1}$ using Algorithm 1 with parameter $k_i \geq K_i$ is*

$$q(G_i \mid \widetilde{G}_{i-1}, \text{pop}(V_i) \in [P_i^-, P_i^+]) \;=\; \frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i} |\mathcal{C}(G_i, \widetilde{G}_i)|. \tag{1}$$

Proof. Any spanning tree can be decomposed into two other trees and an edge joining them. Let $T \cup e \cup T'$ denote the spanning tree obtained by joining two other spanning trees, T and T', with an edge e. Then Equation 3 can be written as

$$q(G_i \mid \widetilde{G}_{i-1}) = \sum_{\substack{T^{(1)} \in \mathcal{T}(G_i) \\ T^{(2)} \in \mathcal{T}(\widetilde{G}_i)}} \sum_{e \in \mathcal{C}(T^{(1)}, T^{(2)})} q(G_i \mid T^{(1)} \cup e \cup T^{(2)})\, \tau(\widetilde{G}_{i-1})^{-1}.$$

Now, $q(G_i \mid T^{(1)} \cup e \cup T^{(2)})$ is determined by whether whether $e^* = e$, i.e., if e is the edge selected to be cut. If e has d_e in the top k_i (if it induces one of the best k_i balanced splits), then it has a $1/k_i$ probability of being selected in step (c) and cut. If d_e is not in the top k_i, then this probability is zero.

Everything written to this point holds regardless of whether G_i is a valid district (i.e., satisfies $\text{pop}(V_i) \in [P_i^-, P_i^+]$). From here onwards we will restrict our attention to valid districts only. Notice that the forward-looking bounds P_i^- and P_i^+ are stricter than merely ensuring $\text{dev}(G_i) \leq D$. That is, conditional on $\text{pop}(V_i) \in [P_i^-, P_i^+]$, we must also have $\text{dev}(G_i) \leq D$.

Therefore, if a sorted edge e_j *in any spanning tree* induces such a balanced partition, we must have $j \leq K_i$, where as in the main text K_i counts the maximum number of such edges across all possible spanning trees. Thus, so long as we set $k_i \geq K_i$, we will have $d_e \leq D$.

Furthermore, across all spanning trees $T^{(1)} \in \mathcal{T}(G_i)$ and $T^{(2)} \in \mathcal{T}(\widetilde{G}_i)$, and connecting edges $e \in E(T^{(1)}, T^{(2)})$, the value of d_e is constant, since removing e induces the same districting. Combining these two facts, we have, conditional on satisfying the bounds P_i^- and P_i^+,

$$q(e^* = e \mid T^{(1)} \cup e \cup T^{(2)}, \text{pop}(V_i) \in [P_i^-, P_i^+]) \;=\; k_i^{-1},$$

*Ph.D. candidate, Department of Statistics, Harvard University. 1 Oxford Street, Cambridge 02138. Email: cmccartan@g.harvard.edu

†Professor, Department of Government and Department of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge 02138. Email: imai@harvard.edu, URL: https://imai.fas.harvard.edu/

which does not depend on $T^{(1)}$, $T^{(2)}$, or e. We may therefore write the conditional sampling probability as

$$\begin{aligned} q(G_i \mid \widetilde{G}_{i-1}, \text{pop}(V_i) \in [P_i^-, P_i^+]) &= \sum_{\substack{T^{(1)} \in \mathcal{T}(G_i) \\ T^{(2)} \in \mathcal{T}(\widetilde{G}_i)}} \sum_{e \in \mathcal{C}(T^{(1)}, T^{(2)})} \frac{1}{k_i \tau(\widetilde{G}_{i-1})} \\ &= \frac{\tau(G_i)\tau(\widetilde{G}_i)}{\tau(\widetilde{G}_{i-1})k_i} |\mathcal{C}(G_i, \widetilde{G}_i)|, \end{aligned} \tag{2}$$

where as in the main text we let $\mathcal{C}(G, H)$ represent the set of edges joining nodes in a subgraph G to nodes in a subgraph H. □

Proposition 4.2. *Let $\pi_S = \sum_{j=1}^S w^{(j)} \delta_{[\xi^{(j)}]}$ be the weighted particle approximation generated by Algorithm 2. Then for all measurable h on unlabeled plans, as $S \to \infty$,*

$$\sqrt{S}(\mathbb{E}_{\pi_S}[h([\xi])] - \mathbb{E}_{\pi}[h([\xi])]) \xrightarrow{d} \mathcal{N}(0, V_{SMC}(h)),$$

for some asymptotic variance $V_{SMC}(h)$.

The proof proceeds by showing that the weights in Algorithm 2 are of a form derived from an existing SMC algorithm with an established central limit theorem.

Proof. We can associate our target measure $\pi([\xi])$ on unlabeled redistricting plans with a corresponding measure on labeled plans

$$\tilde{\pi}(\xi) \coloneqq \psi([\xi])^{-1} \pi([\xi]),$$

so that the pushfoward measure obtained by mapping $\xi \mapsto [\xi]$ recovers π.

Our SMC algorithm will operate on labeled plans, targeting $\tilde{\pi}$, so that the resulting plans, when considered as representatives of their corresponding unlabeled plans, will be representative of π in the sense given by the theorem statement. Recall that a labeled redistricting plan ξ is just a tuple of graph partitions $(G_1, G_2, \ldots, G_n)$. We begin by extending $\tilde{\pi}(\xi)$ to a series of measures on partial plans,

$$\begin{aligned} \tilde{\pi}_i(G_1, G_2, \ldots, G_i) &:\propto \prod_{j=1}^{i} \frac{\tau(G_j)^\rho \tau(\widetilde{G}_j)}{\tau(\widetilde{G}_{j-1})} \mathbf{1}_{\text{pop}(V_j) \in [P_j^-, P_j^+]} \\ &\propto \tau(\widetilde{G}_j) \prod_{j=1}^{i} \tau(G_j)^\rho \mathbf{1}_{\text{pop}(V_j) \in [P_j^-, P_j^+]}, \end{aligned}$$

for $1 \leq i \leq n-2$, and where we have simplified the telescoping product in the second equality. Recall that the $\widetilde{G}_i$ are determined completely by $G_1, G_2, \ldots, G_i$.

For $i = n-1$, the above definition would yield

$$\tilde{\pi}_{n-1}(G_1, G_2, \ldots, G_i) \propto \tau(\widetilde{G}_{n-1}) \prod_{j=1}^{n-1} \tau(G_j)^\rho \mathbf{1}_{\text{pop}(V_j) \in [P_j^-, P_j^+]}, = \tau(\xi)\tau(G_n)^{1-\rho} \mathbf{1}_{\text{dev}(\xi) \leq D},$$

which is close to but not quite the target measure. So we instead define $\pi_{n-1} \coloneqq \tilde{\pi}$; i.e., we add the additional terms $\exp(-J(\xi))$ and $\psi([\xi])$ and adjust for $\tau(G_n)^{1-\rho}$.

With these partial-plan measures defined, notice that the incremental weight $w_i^{(j)}$ for partial plans with $1 \leq i \leq n-2$ and $\text{pop}(V_j) \in [P_j^-, P_j^+]$ may be written as

$$\begin{aligned} w_i^{(j)} &= \tau(G_i^{(j)})^{\rho-1} \frac{k_i}{|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|} \\ &= \frac{\tau(G_i^{(j)})^{\rho} \tau(\widetilde{G}_i^{(j)})}{\tau(\widetilde{G}_{i-1}^{(j)})} \left(\frac{\tau(G_i^{(j)})\tau(\widetilde{G}_i^{(j)})}{\tau(\widetilde{G}_{i-1}^{(j)})} \frac{|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|}{k_i} \right)^{-1} \\ &= \frac{\tilde{\pi}_i(G_1, \ldots, G_i)}{\tilde{\pi}_{i-1}(G_1, \ldots, G_{i-1}) q(G_i \mid \widetilde{G}_{i-1}, \text{pop}(V_i) \in [P_i^-, P_i^+])}. \end{aligned} \tag{3}$$

For the final weighting at split $i = n-1$, the incremental weight (i.e., not including the residual previous weights $\left(\prod_{i=1}^{n-2} w_i^{(j)}\right)^{1-\alpha}$ given by step (c) of Algorithm 2 is

$$\begin{aligned} &\exp\left(-J(\xi^{(j)})\right) w_{n-1}^{(j)} \psi([\xi])^{-1} \left(\tau(G_{n-1}^{(j)})\right)^{\rho-1} \\ &\qquad = \frac{\tilde{\pi}_{n-1}(G_1, \ldots, G_{n-1})}{\tilde{\pi}_{n-2}(G_1, \ldots, G_{n-2}) q(G_{n-1} \mid \widetilde{G}_{n-2}, \text{pop}(V_{n-1}) \in [P_{n-1}^-, P_{n-1}^+])}, \end{aligned}$$

since this weight includes exactly the same additional terms as $\tilde{\pi}_{n-1}$ mentioned above. So in fact Equation (3) holds for all $1 \leq i \leq n-1$.

These incremental weights are precisely those of the SMC partial rejection control algorithm of Peters et al. (2012) (see also LeGland and Oudjane (2005)), with the weights set to zero for invalid samples and the partial rejection threshold set to the minimum possible nonzero weight. So after pushing forward to unlabeled plans, we gain immediately the theorem proved in that work (its Equation 6), viz., that for all measurable h on unlabeled plans and as $S \to \infty$, we have

$$\sqrt{S}(\mathbb{E}_{\pi_S}[h([\xi])] - \mathbb{E}_{\pi}[h([\xi])]) \xrightarrow{d} \mathcal{N}(0, V_{\text{SMC}}(h)),$$

for asymptotic variance $V_{\text{SMC}}(h)$ given by Equation 7 of the same work. □

B Additional Validation Example

This section reports the results of another validation study applied to a 50-precinct map taken from the state of Florida. As in Section 5, we use the efficient enumeration algorithm of Fifield et al. (2020) to obtain all possible redistricting maps with three and four contiguous districts, and use these plans as a baseline to validate the proposed algorithm. The left plot of Figure 1 below shows the validation map.

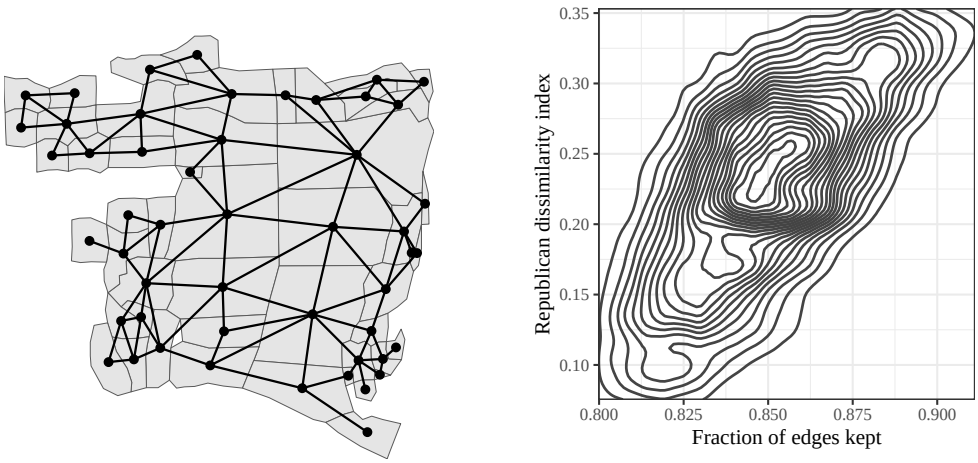

Figure 1: The 50-precinct Florida map used for validation (left) and the joint distribution of Republican dissimilarity and compactness on the map over all partitions into four districts with $\text{dev}(\xi) \leq 0.10$ (right).

There are 112,515,494 partitions of the map into four districts, 33,635 of which have $\text{dev}(\xi) \leq 0.10$. We evaluate the accuracy of the proposed algorithm, and compare it to the same MCMC algorithm used in Section 6, across a range of target distributions, with ρ running from 0.8 to 1.2 The right plot of Figure 1 shows the joint distribution of compactness and the summary statistic we use for the validation, the Republican dissimilarity index (Massey and Denton, 1988). With this validation map, Republican dissimilarity is reasonably sensitive to the compactness of districts. This makes dissimilarity a good test statistic for comparing distributions that differ primarily in their average compactness.

For each target distribution, we sample 1,500 redistricting plans from both the SMC and MCMC algorithms, and repeat the sampling four times in order to produce $\hat{R}$ estimates. The MCMC algorithm is initialized with a random SMC-drawn map and first run for 500 warm-up iterations. To validate and compare the samples, we reweight the enumerated redistricting plans by $\tau(\xi)^\rho$, and then produce quantile-quantile plots of the Republican dissimilarity index, which are shown in Figure 2.

Across the range of ρ values, the agreement between the SMC sample and target distributions is excellent, even in the lower tail. The $\hat{R}$ values for the SMC algorithm were also all less than 1.003. The MCMC algorithm fares well in general but has noticeable bias for $\rho = 0.9$ and $\rho = 1.2$, and has some additional misses for the other target distributions as well, which manifest as small protuberances in the quantile-quantile plot. The $\hat{R}$ values for the MCMC algorithm were all less than 1.004, indicating that the overall location and scale of the Republican dissimilarity were estimated well, but $\hat{R}$ is not designed to capture these smaller-scale deviations from the target distribution.

Here and in Section 5, we validated and compared the SMC and MCMC algorithms with several summary statistics. This reflects the applied use case for redistricting analysis. But it is also informative to study how well the SMC algorithm can target the actual distribution of plans themselves. Of course, we cannot expect the algorithm to perform well in this regard if there are fewer samples then there are plans, which is the case in almost all real-world problems. So we subset the enumerated plans to those with $\text{dev}(\xi) \leq 0.01$, of which there are just 38. We first generate 10,000 samples from the SMC algorithm, targeting a distribution with $\rho = 1$. We measure the discrepancy between the sampled distribution of the 38 plans and the enumerated set weighted to $\tau(\xi)$ with the total variation distance, which is 0.0178 for this sample. Increasing the sample size to 50,000 decreases the total variation distance to 0.0149.

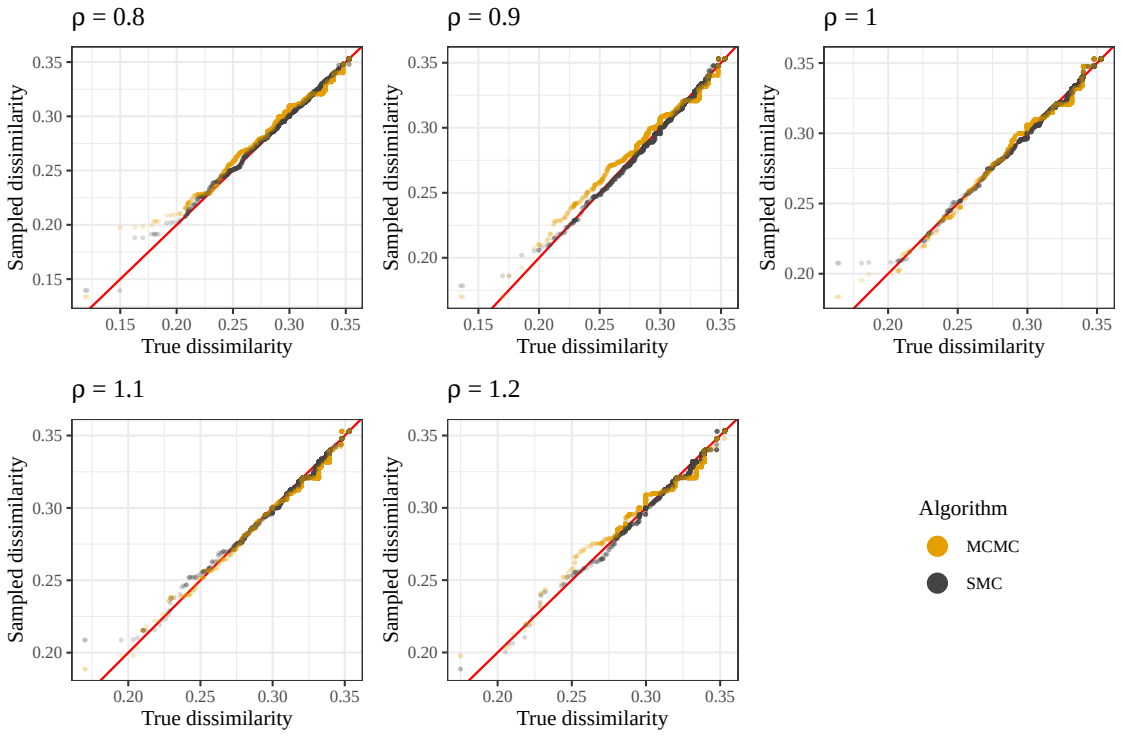

Figure 2: Quantile-quantile plots for 1,500 MCMC and SMC samples of Republican dissimilarity across a range of target distributions with different compactness parameters ρ.

C Algorithm Implementation Details

C.1 Estimating K_i

A natural approach to estimating K_i is to draw a moderate number of spanning trees $\mathcal{T}_i \subseteq \mathcal{T}(\widetilde{G}_i)$ and compute $ok(T)$ for each $T \in \mathcal{T}_i$. The sample maximum, or the sample maximum plus some small buffer amount, would then be an estimate of the true maximum $\widehat{K}_i$ and an appropriate choice of k_i. In practice, we find little noticeable loss in algorithmic accuracy even if $k_i < K_i$. The following proposition theoretically justifies this finding. As above, d_e represents the population deviation of the district induced by removing edge e from a spanning tree.

Proposition C.1. *The probability $q(e = e^* \mid \mathcal{F})$, i.e., the probability that an edge e is selected to be cut at iteration i, given that the tree T containing e has been drawn, and that e would induce a valid district, satisfies*

$$\max\left\{0, q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})\left(1 + \frac{1}{k_i}\right) - 1\right\} \leq q(e = e^* \mid \mathcal{F}) \leq \frac{1}{k_i},$$

where $\mathcal{F}$ is the σ-field generated by $\{T, \text{pop}(V_i) \in [P_i^-, P_i^+]\}$.

Proof. We can write

$$q(e = e^* \mid \mathcal{F}) \;=\; q(e = e^*, d_e \leq d_{e_{k_i}} \mid \mathcal{F}) \;=\; \frac{1}{k_i} q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}),$$

This holds because the edge e will not be cut unless $d_e \leq d_{e_{k_i}}$, i.e., if e is among the top k_i edges. We then have immediately that $q(e = e^* \mid \mathcal{F}) \leq k_i^{-1}$. Additionally, using the lower Fréchet inequality, we find the lower bound

$$\begin{aligned} q(e = e^* \mid \mathcal{F}) &= q(e = e^*, d_e \leq d_{e_{k_i}} \mid \mathcal{F}) \\ &\geq \max\left\{0, q(e = e^* \mid \mathcal{F}) + q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) - 1\right\} \\ &= \max\left\{0, \frac{1}{k_i} q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) + q(d_e \leq d_{e_{k_i}} \mid \mathcal{F}) - 1\right\} \\ &= \max\left\{0, q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})\left(1 + \frac{1}{k_i}\right) - 1\right\}. \end{aligned}$$

□

If $k_i \geq K_i$, then $q(e = e^* \mid \mathcal{F})$ is exactly k_i^{-1}, a fact which is used in the proof of Lemma 4.1. This result, which is proved using a simple Fréchet bound, shows that as long as $q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})$ is close to 1, using k_i^{-1} in Lemma 4.1 is a good approximation to the true sampling probability.

Having sampled $\mathcal{T}_i$, we can compute for each value of k the sample proportion of trees where a randomly selected edge e among the top k of edges of the tree is also among the top k for the other trees—in effect estimating $q(d_e \leq d_{e_{k_i}} \mid \mathcal{F})$. We may then choose k_i to be the smallest k for which this proportion exceeds a pre-set threshold (e.g., 0.99). We have found that this procedure, repeated at the beginning of each sampling stage, efficiently selects k_i without compromising the ability to sample from the target distribution.

C.2 Calculating $\psi([\xi])$

To calculate $\psi([\xi])$, we first observe that sequentially valid labelings of a particular unlabeled plan are in bijective correspondence with certain increasing sequences of connected subgraphs of the quotient graph G/ξ. Specifically, as noted in the text, for every $1 \leq i \leq n-1$, the subgraph $A_i \coloneqq \{i+1, i+2, \ldots, n\}$ of G/ξ is connected if and only if ξ is a sequentially valid labeling. Let $j = n - i$; then the sequence A_j is increasing. Going from A_j to A_{j+1} for any j involves adding a vertex in G/ξ which is adjacent to A_j. This fact provides an easy scheme to generate the number of sequentially valid labelings: for each vertex v in G/ξ, let $A_1 = \{v\}$. Then pick a neighbor of A_1 and add it to the set to form A_2. Continue in this fashion until A_{n-1} contains all but one vertex of G/ξ. Undoing our bijection, this remaining vertex is labeled 1; the vertex added between A_{n-2} and A_{n-1} is labeled 2, and so on.

Figure 3 illustrates this scheme on district-level graph for the Pennsylvania court-imposed plan. We pick an arbitrary vertex, denoted by "A" in Figure 3(a). Then we pick a neighbor, denoted "B," and add it to the subgraph. We continue this way, adding the vertices indicated by the alphabetical ordering, until we have covered the whole district-level graph. Then the last vertex added, "R," is labeled 1 in the district labeling, as shown in Figure 3(b). Vertex "Q" is labeled 2, and so on, until vertex "A" is labeled 18. One can easily check that this labeling is sequentially valid—at every point, the region of the map corresponding to unlabeled districts is contiguous.

As Section 4.4.2 mentioned, we adopt different strategies for calculating $\psi([\xi])$ when $n \leq 13$ and $n > 13$. When there are no more than 13 districts, we simply recurse down the tree of possibilities for generating all sequentially valid plans. In the example of Figure 3, we have 18 options for the first vertex in the subgraph. Once we have picked vertex "A", then we could add "B" or "C". If we add "B", then we could add "C", "D", or "G".

Clearly this tree grows quite large very quickly. Fortunately, there are many duplicate nodes—in our example, adding vertex B and then C in sequence produces the same subgraph as if we had added C and then B. Thus by memoizing our counting function we can significantly reduce the number of tree branches we must explore in full.

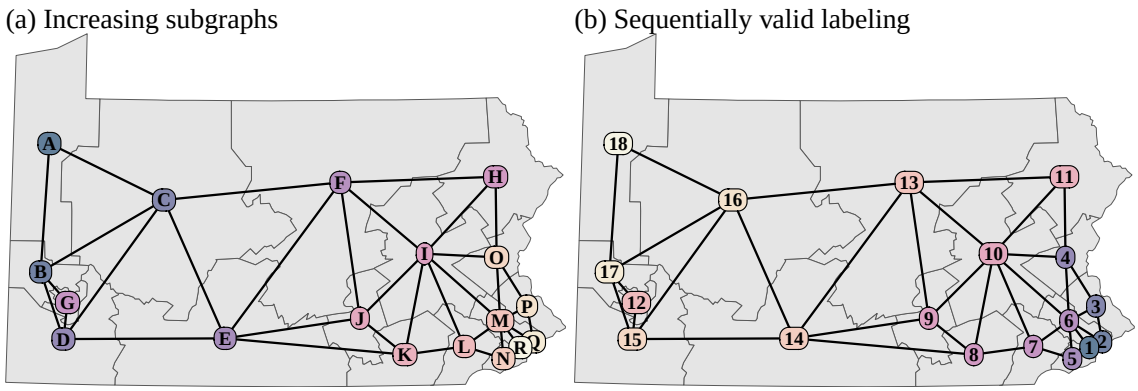

Figure 3: Schematic of the process to generate a sequentially valid labeling, using the district-level graph for the Pennsylvania court-imposed plan. Subgraphs are built in alphabetical order in panel (a): $\{A\}, \{A, B\}, \ldots, \{A, B, \ldots, Q\}$. This corresponds to a sequentially valid labeling shown in panel (b).

When $n > 13$, it is no longer computationally feasible to perform the recursion for each sampled plan in what will often be a large number of sampled plans. In these cases (it should be noted that for the 2020 redistricting cycle, only nine states had more than 13 districts), we can only estimate the number of sequentially valid labelings. To do so, we generate a large number of random sequentially valid labelings by following the scheme outlined above, picking vertices of G/ξ to add to each A_j uniformly at random. Since we observe the number of candidate vertices to add at each stage, and because each set of choices produces a unique sequentially valid labeling, we can compute the probability of drawing each sequentially valid labeling directly.

Denote the probability distribution created by this sampling scheme by p_{sv}, and the (not-probability) measure which assigns mass 1 to all $n!$ relabelings by μ_{all}. Our goal here is to estimate

$$\int \mathbf{1}\{\sigma \text{ is sequentially valid}\} d\mu_{\text{all}}(\sigma)$$

the total number of sequentially valid labelings. This can be done with our sample from p_{sv}, since it is supported on precisely the same set for which the indicator function takes a value of 1. We need only take the mean of the inverse of the probability of each sampled labeling.

This importance sampling estimate is quite accurate, since the proposal support matches the target support exactly, and there is not too much variation in the proposal probabilities. The estimate can be made arbitrarily good by increasing the number of importance samples. This is demonstrated in Figure 4, which shows how the bias in importance sampling estimates varies with the number of importance samples and the overall size of the district-level graph. In our software, we use on the order of 1,500 samples for $n = 14$ by default, and more when n is larger (e.g., around 3,000 for the $n = 28$ Florida congressional districts). These numbers are chosen to ensure that the relative standard error (also known as the *coefficient of variation*) of the $\psi([\xi])$ estimate is controlled to a reasonable amount, something we demonstrate next for the case of Pennsylvania.

For the district-level graph shown above in 3, we have $\log \psi([\xi]) = 29.999$, calculated exactly using the recursive procedure described above. Using the importance sampling procedure with 2,000 samples, we estimate a value of 29.949. This corresponds to a relative error of 0.051. We can also calculate the relative standard error of this example estimate with the delta method, which yields 0.052. In other words, the importance sampling error in the estimate of $\psi([\xi])$ is on the order of 5%. This is also small on the scale of the variation in $\psi([\xi])$ across plans: for the six comparison

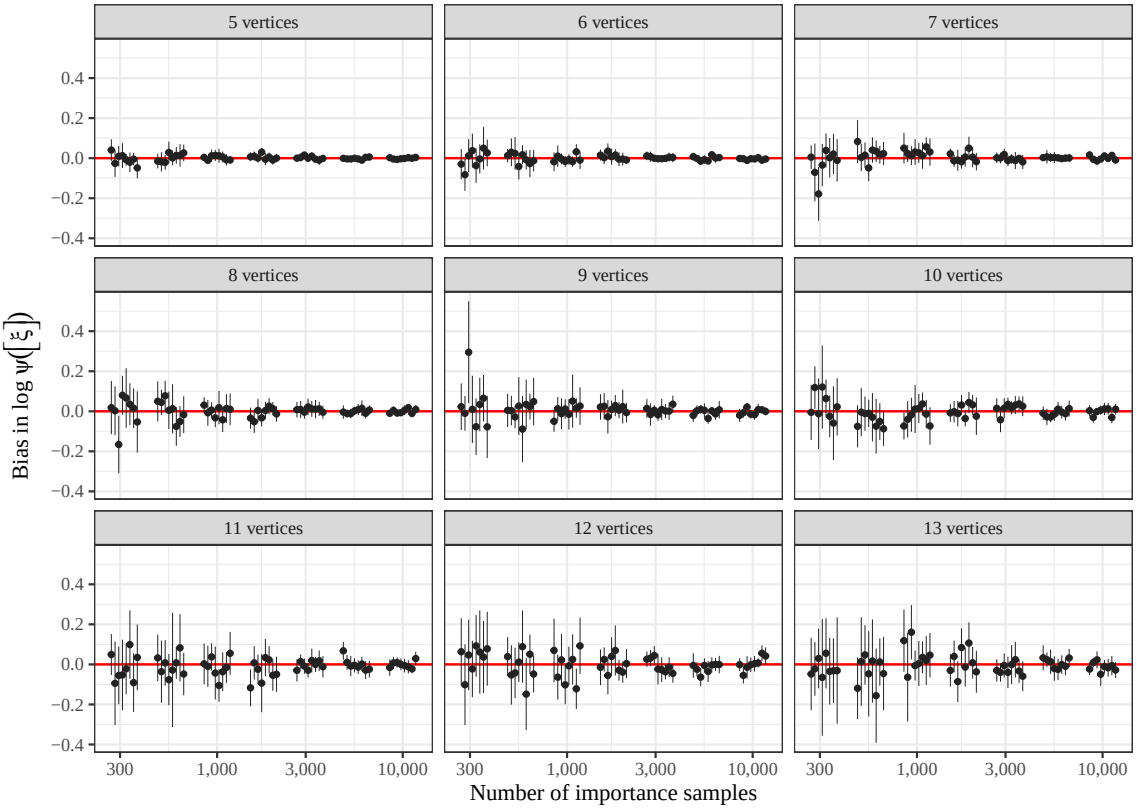

Figure 4: Bias in importance sampling estimates of $\log \psi([\xi])$ by the number of importance samples and the number of vertices in the district-level graph. District-level graphs were selected as random subsets of the graph shown in 3. Eight replicate experiments were performed for each importance sample size and graph size. The vertical bars for each point span two standard errors in either direction. Standard errors were calculated with the delta method.

plans used in the main text, $\log \psi([\xi])$ ranges from 28.456 to 30.317. So the error in the importance sampling estimate ($|29.949 - 29.999| = 0.05$) is just 2.7% of the range of $\log \psi([\xi])$ across these six plans.

Finally, we note that the procedures used here to calculate $\psi([\xi])$ can also be applied to partial plans. Notating this formally is more difficult, but a partial plan can be viewed as a (unbalanced) plan with fewer districts. Calculating ψ for these partial plans and incorporating them into the SMC weights in Algorithm 2 improves the sampling efficiency without changing the target distribution, as long as each partial ψ's contribution to the weights is canceled in the following iteration. This is the approach taken by our software implementation.

C.3 Stabilizing Importance Weights

When $\rho \neq 1$ or when the constraints imposed by J are severe, there can be substantial variance in the importance sampling weights. For large maps with $\rho = 0$, for instance, the weights will generally span hundreds if not thousands of orders of magnitude. This reflects the general computational difficulty in sampling uniformly from constrained graph partitions. As Najt et al. (2019) show, sampling of node-balanced graph partitions is computationally intractable in the worst case. In such cases, the importance sampling estimates will be highly variable, and resampling based on

these weights may lead to degenerate samples with only one unique map.

When the importance weights are variable but not quite so extreme, we find it useful to truncate the normalized final weights (such that their mean is 1) from above at a value $w_{\max}$ at the end of sampling. The theoretical basis for this maneuver is provided by Ionides (2008), who proved that as long as $w_{\max} \to \infty$ and $w_{\max}/S \to 0$ as $S \to \infty$, the resulting estimates are consistent and have bounded variance (since the truncation occurs only after the final SMC step, these conclusions, which were made in the context of importance sampling, carry over.) One such choice we have found to work well for the weights generated by this sampling process is $w_{\max} = S^{0.4}/100$, though for particular maps other choices of exponent and constant multiplier may be superior.

Truncation is no panacea, however. As with any method that relies on importance sampling, it is critical to examine the distribution of importance weights to ensure that they will yield acceptable resamples.

C.4 Computational Complexity

The two asymptotically slowest steps of the SMC algorithm are computing $\tau(G_i)$ for every district G_i and drawing a spanning tree using Wilson's algorithm for each iteration. All other steps, such as computing d_e and $|\mathcal{C}(G_i^{(j)}, \widetilde{G}_i^{(j)})|$, are linear in the number of vertices, and are repeated at most once per iteration.[1] Computing $\tau(G_i)$ requires computing a determinant, which currently has computational complexity $O(|V_i(\xi)|^{2.373})$ though most implementations are $O(|V_i(\xi)|^3)$. Since this must be done for each district of size roughly m/n, the total complexity for sampling one plan is $O(n \cdot (m/n)^{2.373})$. For the spanning trees, the expected runtime of Wilson's algorithm is the mean hitting time of the graph, which is $O(m^2)$ in the worst case. So the total complexity for each sample is roughly $O(nm^2 + m^{2.373}n^{-1.373})$ (ignoring the random rejection procedure). Note that when $\rho = 1$, we need not compute $\tau(G_i)$, and the total complexity is roughly $O(nm^2)$.

References

Fifield, B., Imai, K., Kawahara, J., and Kenny, C. T. (2020). The essential role of empirical validation in legislative redistricting simulation. *Statistics and Public Policy*, 7(1):52–68.

Ionides, E. L. (2008). Truncated importance sampling. *Journal of Computational and Graphical Statistics*, 17(2):295–311.

LeGland, F. and Oudjane, N. (2005). A sequential particle algorithm that keeps the particle system alive. In *2005 13th European Signal Processing Conference*, pages 1–4. IEEE.

Massey, D. S. and Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, 67(2):281–315.

Najt, L., Deford, D., and Solomon, J. (2019). Complexity and geometry of sampling connected graph partitions. *arXiv preprint arXiv:1908.08881*.

Peters, G. W., Fan, Y., and Sisson, S. A. (2012). On sequential monte carlo, partial rejection control and approximate bayesian computation. *Statistics and Computing*, 22(6):1209–1222.

[1] To compute d_e, we walk depth-first over the tree and store, for each node, the total population of that node and the nodes below it. This allows for $O(1)$ computation of d_e for all edges.