

Submitted to *INFORMS Journal on Computing*
manuscript (Please, provide the manuscript number!)

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and should not be used to distribute the papers in print or online or to submit the papers to another publication.

Online Supplement for “Sampling Conditionally on a Rare Event via Generalized Splitting”

Zdravko I. Botev

UNSW Sydney, Australia, botev@unsw.edu.au

Pierre L’Ecuyer

Université de Montréal, Canada, and Inria–Rennes, France lecuyer@iro.umontreal.ca

1. Proof of the Theorems

We first recall the working notation. Let \mathcal{A} be a class of measurable sets. For any $A \subseteq \mathcal{A}$ and $i \geq 1$, let M_i and $H_i(A)$ be the cardinalities of \mathcal{Y}_i and of $\mathcal{Y}_i \cap A$. These are the realizations of M and $H(A)$ for replication i of Algorithm 2 (GS sampler 2). Let \bar{M}_n and $\bar{H}_n(A)$ be the respective averages of these n realizations, and let $m := \mathbb{E}_{\text{GS}}[M]$, so that the target distribution is $\mathbb{Q}(A) = \mathbb{E}_{\text{GS}}[H(A)]/m$. For simplicity of notation, unless there is ambiguity, we henceforth drop the GS subscripts from \mathbb{E}_{GS} . When we draw an \mathbf{Y} from $\hat{\mathbb{Q}}_n$, it belongs to A with probability $\bar{H}_n(A)/\bar{M}_n$ (since \mathcal{Y}_{\cup} is not empty, $\bar{M}_n > 0$). Note that $H(A) \leq M$ for all $A \in \mathcal{A}$, and that M_i and H_i take their values in $\{1, \dots, s^{\tau-1}\}$.

In particular, in Algorithm 3 (Sampling an empirical distribution from n iid non-empty GS replications) we obtain the independent sets, $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$, of states \mathbf{Y} . We can (re)label all the states \mathbf{Y} such that:

$$\overbrace{\mathbf{Y}_1, \dots, \mathbf{Y}_{T_1}}^{\mathcal{Y}_1}, \dots, \dots, \overbrace{\mathbf{Y}_{T_{n-1}+1}, \dots, \mathbf{Y}_{T_n}}^{\mathcal{Y}_n}.$$

In this way, $\{\mathbf{Y}_t, t = 1, 2, \dots\}$ is a discrete-time regenerative process with regeneration times $0 = T_0 < \dots < T_n$, and tour lengths $M_i = T_i - T_{i-1}$, $j = 1, 2, \dots, n$ with stationary measure $\mathbb{Q}(A)$. With this notation we have that $N(t) = \min\{n : T_n > t\}$ in Algorithm 4 (Sampling an empirical distribution with more than t retained states). Moreover, if we define the number of renewals in $(0, t]$ as $\tilde{N}(t) := N(t) - 1 = \max\{n : T_n \leq t\}$ with $N(0) = 0$, then $\{\tilde{N}(t), t \geq 0\}$ is a *renewal process* (Asmussen 2008, Chapter 5).

Since $N(t) = \tilde{N}(t) + 1$ is a stopping time with respect to the filtration generated by the sequence of iid random variables $\{M_i, i \geq 1\}$, by the Wald identity we have $\mathbb{E}[T_{N(t)}] = \mathbb{E}[N(t)]\mathbb{E}[M_i]$. We define $\hat{\mathbb{Q}}_n(A) = \bar{H}_n(A)/\bar{M}_n$ and $\hat{\mathbb{Q}}_{N(t)}(A) = \frac{1}{T_{N(t)}} \sum_{i=1}^{N(t)} H_i(A)$. With $Z_i(A) := H_i(A) - M_i\mathbb{Q}(A)$, Wald's identity also gives

$$\mathbb{E} \left[\sum_{i=1}^{N(t)} Z_i(A) \right] = \mathbb{E}[N(t)]\mathbb{E}[Z_i(A)] = 0. \quad (1)$$

REMARK 1 (ELAPSED-TIME PROCESS). Note that the autocorrelation plot of the age (or current lifetime) process, $E(t) := t - T_{\tilde{N}(t)}$, may be used as a graphical tool to diagnose the convergence of $\{\mathbf{Y}_t, t = 1, 2, \dots\}$ to its stationary distribution $\mathbb{Q}(A)$, because (Asmussen 2008, Page 170, Proposition 1.3):

$$\sup_A |\mathbb{P}(\mathbf{Y}_t \in A) - \mathbb{Q}(A)| \leq 2 \sup_A |\mathbb{P}(E(t) \in A) - \mathbb{P}(E(\infty) \in A)|.$$

In other words, ensuring the convergence of the Markov process $\{E(t), t \geq 0\}$ to its stationary measure is sufficient to ensure the convergence of $\{\mathbf{Y}_t, t = 1, 2, \dots\}$ to its stationary measure.

1.1. Proof of Theorem 1 (Sampling via n iid runs of GS)

First, we prove the bound on the TV error. Using the identity, (Meketon and Heidelberger 1982, Page 180)

$$\frac{x}{y} - \frac{\alpha}{\beta} = \frac{x(y-\beta)^2}{y\beta^2} + \frac{x}{\beta} - \frac{\alpha y}{\beta^2} - \frac{(x-\alpha)(y-\beta)}{\beta^2}, \quad (2)$$

with $\alpha = 0$, we have that

$$\begin{aligned} \mathbb{E} \left[\frac{\bar{H}_n(A)}{\bar{M}_n} \right] - \frac{h(A)}{m} &= \mathbb{E} \left[\frac{\bar{Z}_n(A)}{\bar{M}} \right] \\ &= \mathbb{E} \left[\frac{\bar{Z}_n(A)}{\bar{M}_n} \frac{(\bar{M}_n - m)^2}{m^2} \right] - \frac{\text{Cov}(\bar{Z}_n(A), \bar{M}_n)}{m^2} \\ (|Z(A)| \leq M \text{ was used}) &\leq \mathbb{E} \left[\frac{(\bar{M}_n - m)^2}{m^2} \right] - \frac{\text{Cov}(\bar{Z}_n(A), \bar{M}_n)}{m^2} \\ &= \frac{\text{Var}(M) - \text{Cov}(Z_1(A), M_1)}{n m^2}. \end{aligned}$$

Hence, using the fact that $|\text{Cov}(Z_1(A), M_1)|^2 \leq \text{Var}(M_1)\text{Var}(Z_1(A)) \leq \text{Var}(M_1)\mathbb{E}M^2$, we obtain

$$\begin{aligned} \sup_A |\mathbb{Q}_n(A) - \mathbb{Q}(A)| &\leq \frac{\text{Var}(M) + \sup_A |\text{Cov}(Z_1(A), M_1)|}{n m^2} \\ &\leq \frac{\text{Var}(M) + \sqrt{\text{Var}(M)\mathbb{E}M^2}}{n m^2} \end{aligned}$$

We can thus clearly see that the convergence of $\mathbb{Q}_n(A)$ depends on the relative error of M .

Next, we prove the bound for the mean absolute value. First, note that the term $\mathbb{E}(\bar{Z}_n(A)\bar{V}_n)^2$, where $V_k := M_k - m$, can be bounded using the independence of the pairs $(Z_i(A), V_i)$ and $\mathbb{E}Z_i(A) = \mathbb{E}V_i = 0$, as follows:

$$\mathbb{E}\bar{Z}_n^2(A)\bar{V}_n^2 = \frac{\sum_{i,j,k,l} \mathbb{E}[Z_i Z_j V_k V_l]}{n^4} \leq \frac{3n^2 - 2n}{n^4} \max_{i,j,k,l} \mathbb{E}|Z_i Z_j V_k V_l| \leq \frac{3\mathbb{E}M^4}{n^2}.$$

Therefore, using the triangle inequality, we have:

$$\begin{aligned} m\mathbb{E}|\widehat{\mathbb{Q}}_n(A) - \mathbb{Q}(A)| &= \mathbb{E} \left| \bar{Z}_n(A) - \frac{\bar{Z}_n(A)(\bar{M}_n - m)}{\bar{M}_n} \right| \\ (\bar{M}_n \geq 1) &\leq \mathbb{E}|\bar{Z}_n(A)| + \mathbb{E}|\bar{Z}_n(A)(\bar{M}_n - m)| \\ &\leq \sqrt{\mathbb{E}\bar{Z}_n^2(A)} + \sqrt{\mathbb{E}(\bar{Z}_n(A)(\bar{M}_n - m))^2} \\ (|Z(A)| \leq M) &\leq \frac{\sqrt{\mathbb{E}M^2}}{\sqrt{n}} + \frac{\sqrt{3\mathbb{E}M^4}}{n}. \end{aligned}$$

1.2. Proof of Theorem 2 (Sampling until GS returns t states)

Recall that $N(t) = \tilde{N}(t) + 1$ is a stopping time. Let $R(t) := T_{\tilde{N}(t)+1} - t$, so that $r(t) := \mathbb{E}R(t) = m\mathbb{E}[N(t)] - t$. Using Wald's identity (1), we can write:

$$\begin{aligned} \bar{Q}_t(A) - \mathbb{Q}(A) &= \mathbb{E} \frac{\sum_{k=1}^{N(t)} (H_k(A) - M_k \mathbb{Q}(A))}{T_{N(t)}} \\ &= \mathbb{E} \frac{\sum_{k=1}^{N(t)} Z_k(A)}{T_{N(t)}} = \mathbb{E} \frac{\frac{1}{t} \sum_{k=1}^{N(t)} Z_k(A)}{1+R(t)/t} \\ &= \mathbb{E} \left(\frac{1}{1+R(t)/t} - 1 \right) \bar{Z}_t(A), \end{aligned}$$

where $\bar{Z}_t(A) := \frac{1}{t} \sum_{k=1}^{N(t)} Z_k(A)$. Then, using the fact that $\frac{1}{1+R(t)/t} \leq 1$, we obtain the uniform bound

$$\begin{aligned} |\bar{Q}_t(A) - \mathbb{Q}(A)| &= \frac{1}{t} \left| \mathbb{E} \left[\frac{R(t)}{1+R(t)/t} \bar{Z}_t(A) \right] \right| \\ &\leq \frac{1}{t} \mathbb{E} |R(t) \bar{Z}_t(A)| \\ &\leq \frac{\sqrt{\mathbb{E}R^2(t) \mathbb{E}[\bar{Z}_t^2(A)]}}{t} \\ &= \frac{\sqrt{\mathbb{E}R^2(t)}}{t} \sqrt{\mathbb{E}[N(t)] \mathbb{E}[Z^2(A)/t^2]} \\ &= \frac{\sqrt{\mathbb{E}[R^2(t)]}}{t^{3/2}} \sqrt{\mathbb{E}[Z^2(A)] \mathbb{E}[N(t)]/t} \\ &\leq \frac{\sqrt{\mathbb{E}[R^2(t)]}}{t^{3/2}} \sqrt{\frac{\mathbb{E}M^2}{m} (1 + r(t)/t)}, \end{aligned}$$

where in the third last line we used Wald's second-moment identity (see (3) below). To finish the proof we apply Lorden's moment inequalities ($\mathbb{E}[R(t)] \leq \mathbb{E}[M^2]/m$ and $\mathbb{E}[R^2(t)] \leq 4\mathbb{E}[M^3]/(3m)$, see Lorden (1970)) to obtain

$$\sup_A |\bar{Q}_t(A) - \mathbb{Q}(A)| \leq \frac{\sqrt{\frac{4}{3} \mathbb{E}[M^3] \mathbb{E}[M^2] (m + \mathbb{E}[M^2]/t) / m^3}}{(t/m)^{3/2}}.$$

To prove the bound for the mean absolute value, we proceed as follows. Again using $\frac{1}{1+R(t)/t} \leq 1$, we have:

$$\begin{aligned} \mathbb{E} |\bar{Q}_t(A) - \mathbb{Q}(A)| &= \mathbb{E} \left| \frac{\frac{1}{t} \sum_{k=1}^{N(t)} Z_k(A)}{1+R(t)/t} \right| \\ &\leq \sqrt{\mathbb{E} \left(\frac{\frac{1}{t} \sum_{k=1}^{N(t)} Z_k(A)}{1+R(t)/t} \right)^2} = \frac{\sqrt{\mathbb{E}[N(t)] \mathbb{E}[Z^2(A)]}}{t} \\ &\leq \sqrt{\frac{\mathbb{E}[Z^2(A)]}{tm} + \frac{\mathbb{E}[M^2] \mathbb{E}[Z^2(A)]}{t^2 m^2}} \leq \frac{\sqrt{\mathbb{E}M^2}}{\sqrt{tm}} + \frac{\mathbb{E}M^2}{tm}, \end{aligned}$$

where in the second last line we used Cauchy's inequality and Wald's second-moment identity, and in the last line we used Lorden's inequality and the sub-additivity of the square root.

1.3. Proof of Theorem 3 (asymptotic version)

Denote $r(t) := \mathbb{E}R(t)$ and $r := (\mathbb{E}M^2 + m)/(2m)$ and note that under the condition $\mathbb{E}M^{p+5} < \infty$ for some $p \geq 0$, we have (Glynn 2006)

$$r(t) = r + o(1/t^{p+3}).$$

Using $0 \leq \frac{1}{1+x} - 1 + x \leq x^2$ for $x \geq 0$, we have the error bound:

$$\begin{aligned} |\bar{Q}_t(A) - \mathbb{Q}(A)| &= \left| \mathbb{E} \left(\frac{1}{1+R(t)/t} - 1 \right) \bar{Z}_t(A) \right| \\ (\text{triangle ineq.}) &\leq \frac{\mathbb{E}R(t)\bar{Z}_t(A)}{t} + \left| \mathbb{E} \left(\frac{1}{1+R(t)/t} - 1 + \frac{R(t)}{t} \right) \bar{Z}_t(A) \right| \\ &\leq \frac{\mathbb{E}R(t)\bar{Z}_t(A)}{t} + \frac{\mathbb{E}R^2(t)|\bar{Z}_t(A)|}{t^2} \\ &\leq \frac{\mathbb{E}R(t) \sum_{k=1}^{N(t)} Z_k(A)}{t^2} + \frac{\sqrt{\mathbb{E}[R^4(t)]\mathbb{E}[\bar{Z}_t^2(A)]}}{t^2}. \end{aligned}$$

Since $\mathbb{E}M^5 < \infty$, by Lorden's inequality, we have $\mathbb{E}R^4(t) < \infty$ and the second term is $\mathcal{O}(t^{-5/2})$, because by Wald's second-moment identity:

$$\mathbb{E}[\bar{Z}_t^2(A)] = \frac{\mathbb{E}[N(t)]}{t^2} \mathbb{E}[Z_1^2(A)] \leq (1 + r(t)/t) \mathbb{E}[M^2]/t = \mathcal{O}(1/t). \quad (3)$$

For the first term, we verify that $e_A(t) := \mathbb{E}R(t) \sum_{k=1}^{N(t)} Z_k(A) < \infty$ satisfies the renewal equation $e_A(t) = (u * v_A)(t)$ with $v_A(t) := \mathbb{E}[R(t)Z_1(A)] = \mathbb{E}[(R(t) - r)Z_1(A)]$, see (Awad and Glynn 2007, Page 25). The latter is bounded uniformly in A :

$$\begin{aligned} |v_A(t)| &= |\mathbb{E}[(R(t) - r)Z_1(A); M_1 > t] + \mathbb{E}[(R(t) - r)Z_1(A); M_1 \leq t]| \\ &= |\mathbb{E}[(M_1 - r)Z_1(A); M_1 > t] + \mathbb{E}[(r(t - M_1) - r)Z_1(A); M_1 \leq t]| \\ &\leq \mathbb{E}[|M_1 - r|M_1; M_1 > t] + \mathbb{E}[|r(t - M_1) - r|M_1; M_1 \leq t]. \end{aligned}$$

For the first term, we obtain:

$$\mathbb{E}[|M_1 - r|M_1; M_1 > t] = \mathcal{O}(\mathbb{E}[M^{p+5}; M > t]/t^{p+3}) = o(1/t^{p+3}).$$

For the second term,

$$\begin{aligned} \mathbb{E}[|r(t - M) - r|M; M \leq t] &\leq \mathbb{E}[|r(t - M) - r|M; M \leq t/2] + \mathbb{E}[|r(t - M) - r|M; M \geq t/2] \\ &\leq \sup_{s > t/2} |r(s) - r| \mathbb{E}[M] + \sup_{s < t/2} |r(s) - r| \mathbb{E}[M; M > t/2] \\ &= o(1/t^{p+3}) + o(1/t^{p+4}). \end{aligned}$$

Hence, we have the convergence uniformly in A :

$$\begin{aligned} e_A(t) &= \frac{\mathbb{E}(M_1 - 1 - 2r)M_1 Z_1(A)}{2m} + o(1/t^{p+2}) \\ &\leq \frac{\mathbb{E}|M_1 - 1 - 2r|M_1^2}{2m} + o(1/t^{p+2}). \end{aligned}$$

Putting it all together, we obtain

$$\sup_A |\bar{Q}_t(A) - Q(A)| \leq \frac{\mathbb{E}|M - 1 - 2r|M^2}{2mt^2} + \mathcal{O}(t^{-5/2}) + o(1/t^{p+4}).$$

where $r = (\mathbb{E}M^2 + m)/(2m)$. The exponential convergence comes from the fact that $\mathbb{E}M^p < \infty$ for all $p > 0$, because $M \leq s^{\tau-1}$ is always bounded. This completes the proof.

Notational Setup for Proofs of Theorems 4 and 5

We now introduce some working notation that will apply to both the proofs of Theorem 4 and 5. Define

$$\mathcal{F}(\mathcal{Y}_\cup) = \{(b_1, \dots, b_{T_n}) \in \{0, 1\}^{T_n} : \text{there exists an } A \in \mathcal{A} : b_i = \mathbb{I}\{\mathbf{x}_i \in A\}, \mathbf{x}_i \in \mathcal{Y}_\cup\} \quad (4)$$

to be a class of binary functions on $\{0, 1\}^{T_n}$ such that each element of \mathcal{F} corresponds to an intersection of \mathcal{Y}_\cup with a set A in \mathcal{A} . Without any conditions on the class of sets \mathcal{A} , the cardinality of $\mathcal{F}(\mathcal{Y}_\cup)$ grows exponentially in T_n , and we have $|\mathcal{F}(\mathcal{Y}_\cup)| \leq 2^{T_n}$ for any n . Let

$$\mathcal{S}_{\mathcal{A}}(\{\mathbf{y}_i\}_{i=1}^n) := \max_{\mathbf{y}_1, \dots, \mathbf{y}_n} \#|A \cap \{\mathbf{y}_1, \dots, \mathbf{y}_n\}, A \in \mathcal{A}|$$

denote the *Vapnik-Chervonenkis shatter coefficient* (Vapnik 2013). Loosely speaking, the shatter coefficient $\mathcal{S}_{\mathcal{A}}(\mathcal{Y}_\cup)$ is the maximum number of distinct ways in which the point-set \mathcal{Y}_\cup can intersect with elements of \mathcal{A} .

Sauer's Lemma (Sauer 1972) tells us that if \mathcal{A} is a class of sets with Vapnik-Chervonenkis dimension $v < \infty$, then the shatter coefficient eventually grows polynomially in n , instead of exponentially:

$$\mathcal{S}_{\mathcal{A}}(\{\mathbf{y}_i\}_{i=1}^n) \leq (ne/v)^v, \quad n > v. \quad (5)$$

Let ρ_1, \dots, ρ_n be iid random variables with marginal distribution $\mathbb{P}(\rho = \pm 1) = 1/2$. Let \mathcal{Y}'_\cup be a sample independent from \mathcal{Y}_\cup that can, in principle, be obtained from another n independent calls to Algorithm 1 (GS Sampler 1). The \mathcal{Y}'_\cup sample is a “ghost” sample (Giné and Zinn 1984) that does not need to be constructed, but is only used in symmetrization inequalities. We denote quantities computed using \mathcal{Y}'_\cup by $H'_i, \mathbf{Y}'_i, M'_i, T'_i$, etc. For example,

H' is an independent “ghost” copy of H . We will make use of two symmetrization inequalities by Giné and Zinn (1984). The first will be used in Theorem 5 (Bound on Expected TV for Empirical Distribution):

$$\mathbb{E} \sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \mathbb{E}H(A)| \leq \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n \rho_i (H_i(A) - H'_i(A)) \right|. \quad (6)$$

The second will be used in Theorem 4 (Almost-Sure TV Convergence):

$$\mathbb{P}(\sup_{A \subseteq \mathcal{A}} |H(A) - \mathbb{E}H(A)| > \epsilon) \leq 2\mathbb{P}(\sup_{A \subseteq \mathcal{A}} |H(A) - H'(A)| > \epsilon/2) \quad \text{for } \epsilon > \sqrt{8 \sup_{A \subseteq \mathcal{A}} \text{Var}(H(A))}. \quad (7)$$

1.4. Proof of Theorem 4 (Almost-Sure TV Convergence)

If we can show that (with $g_n = o(n) \Leftrightarrow \lim_{n \uparrow \infty} g_n/n = 0$),

$$\mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\hat{Q}_n(A) - \mathbb{Q}(\mathcal{A})| > \epsilon) \leq c_1 \exp(-c_2 n \epsilon^2 + o(n)) \quad (8)$$

for some constants $c_1, c_2 > 0$, then the fact that $\sum_{n \geq 1} \mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\hat{Q}_n(A) - \mathbb{Q}(\mathcal{A})| > \epsilon) < \infty$ for any $\epsilon > 0$ implies the almost sure convergence result of the theorem. To show (8) we will use the symmetrization inequality (7) and the simple union bound:

$$\mathbb{P}(|X \pm Y| > \epsilon) \leq \mathbb{P}(|X| > \alpha \epsilon) + \mathbb{P}(|Y| > (1 - \alpha) \epsilon) \quad \alpha \in (0, 1). \quad (9)$$

Using these two inequalities, we have

$$\begin{aligned} \mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\hat{Q}_n(A) - \mathbb{Q}(\mathcal{A})| > \epsilon) &\stackrel{(9)}{\leq} \mathbb{P}(|\bar{M}_n - m| > \frac{m\epsilon}{2}) + \mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \mathbb{E}H(A)| > \frac{m\epsilon}{2}) \\ (\text{Hoeffding's with } M < s^\tau) &\leq 2 \exp\left(-\frac{nm^2\epsilon^2}{2s^{2\tau}}\right) + \mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \mathbb{E}H(A)| > \frac{m\epsilon}{2}). \end{aligned}$$

Thus, in order to show (8), we only need an exponentially decaying bound on the second term with $\epsilon_1 = m\epsilon/2$:

$$\mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \mathbb{E}H(A)| > \epsilon_1) \stackrel{(7)}{\leq} 2\mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \bar{H}'_n(A)| > \epsilon_1/2) \quad \text{for } \epsilon_1 > \sqrt{8 \frac{\mathbb{E}M^2}{n}} \Leftrightarrow n > \frac{8\mathbb{E}M^2}{\epsilon_1^2}.$$

Recall that ρ_1, \dots, ρ_n is an iid random sample with $\mathbb{P}(\rho = \pm 1) = 1/2$, and that each H'_i is an independent “ghost” copy of H_i . By symmetry, each $H_i(A) - H'_i(A)$ has the same

distribution as $\rho_i(H_i(A) - H'_i(A))$. Using this observation, we obtain (with $\epsilon_2 := \epsilon_1/2$ and for $n > 2\mathbb{E}M^2/\epsilon_2^2$):

$$\begin{aligned} \mathbb{P}(\sup_{A \subseteq \mathcal{A}} |\bar{H}_n(A) - \bar{H}'_n(A)| > \epsilon_2) &= \mathbb{P}\left(\sup_{A \subseteq \mathcal{A}} \frac{1}{n} \left| \sum_{i=1}^n \rho_i(H_i(A) - H'_i(A)) \right| > \epsilon_2\right) \\ &\stackrel{(9)}{\leq} \mathbb{P}\left(\sup_{A \subseteq \mathcal{A}} \frac{1}{n} \left| \sum_{i=1}^n \rho_i H_i(A) \right| > \frac{\epsilon_2}{2}\right) + \mathbb{P}\left(\sup_{A \subseteq \mathcal{A}} \frac{1}{n} \left| \sum_{i=1}^n \rho_i H'_i(A) \right| > \frac{\epsilon_2}{2}\right) \\ &= 2\mathbb{P}\left(\sup_{A \subseteq \mathcal{A}} \frac{1}{n} \left| \sum_{i=1}^n \rho_i H_i(A) \right| > \frac{\epsilon_2}{2}\right). \end{aligned}$$

The proof will be complete if we show that ($\epsilon_3 = \epsilon_2/2$)

$$\mathbb{P}\left(\sup_{A \subseteq \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n \rho_i H_i(A) \right| > \epsilon_3\right) \leq c_1 \exp(-c_2 n \epsilon_3^2 + o(n))$$

for some constants $c_1, c_2 > 0$. Let

$$N_{\mathcal{A}}(\mathcal{Y}_{\cup}) := \#\{A \cap \{\mathbf{Y}_1, \dots, \mathbf{Y}_{T_n}\}, A \in \mathcal{A}\}$$

be the number of different subsets of the points in \mathcal{Y}_{\cup} that can be picked out by the class \mathcal{A} (so that, by definition, the shatter coefficient is $\mathcal{S}_{\mathcal{A}}(T_n) = \max_{\mathcal{Y}_{\cup}} N_{\mathcal{A}}(\mathcal{Y}_{\cup})$). Similarly, let

$$N_{\mathcal{A}}(\bar{\mathcal{Y}}_{\cup}) := \#\{A \cap \bar{\mathcal{Y}}_{\cup}, A \in \mathcal{A}\},$$

where $\bar{\mathcal{Y}}_{\cup} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_{ns^{\tau-1}}\}$ is the collection of all $ns^{\tau-1}$ potential states from n independent runs of splitting (L'Ecuyer et al. 2018)[Section 3.1] (in practice only a small fractions of these trajectories survive till the final level of splitting). Clearly, $N_{\mathcal{A}}(\bar{\mathcal{Y}}_{\cup}) \geq N_{\mathcal{A}}(\mathcal{Y}_{\cup})$.

A well-known result (see (Rao 1962) and (Devroye et al. 2013, Theorem 13.13)) asserts that when the \mathbf{Y} 's have a density and \mathcal{A} is the class of all convex sets, then:

$$\mathbb{E}N_{\mathcal{A}}(\bar{\mathcal{Y}}_{\cup}) = 2^{o(n)}. \tag{10}$$

Thus, by conditioning on \mathcal{Y}_U , we can write:

$$\begin{aligned}
\mathbb{P} \left(\sup_{A \subseteq \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n \rho_i H_i(A) \right| > \epsilon_3 \right) &= \mathbb{E} \left[\mathbb{P} \left(\sup_{A \subseteq \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n \rho_i H_i(A) \right| > \epsilon_3 \middle| \mathcal{Y}_U \right) \right] \\
&\stackrel{\text{(union bound)}}{\leq} \mathbb{E} \left[N_{\mathcal{A}}(\mathcal{Y}_U) \sup_{A \subseteq \mathcal{A}} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \rho_i H_i(A) \right| > \epsilon_3 \middle| \mathcal{Y}_U \right) \right] \\
&\stackrel{\text{(Hoeffding's with } |\rho_i H_i(A)| < s^\tau)}{\leq} \mathbb{E} \left[N_{\mathcal{A}}(\mathcal{Y}_U) \sup_{A \subseteq \mathcal{A}} 2 \exp \left(-\frac{2n\epsilon_3^2}{(2s^\tau)^2} \right) \right] \\
&\stackrel{\text{(using } |\mathcal{Y}_U| \leq |\bar{\mathcal{Y}}_U| = ns^{\tau-1})}{\leq} 2 \exp \left(-\frac{n\epsilon_3^2}{2s^{2\tau}} \right) \mathbb{E} [N_{\mathcal{A}}(\bar{\mathcal{Y}}_U)] \\
&\stackrel{(10)}{=} 2 \exp \left(-\frac{n\epsilon_3^2}{2s^{2\tau}} + o(n) \right).
\end{aligned}$$

This completes the proof.

1.5. Proof of Theorem 5 (Bound on Expected TV for Empirical Distribution)

Our proof follows as closely as possible the proof of the classical VC inequalities, as described in (Devroye and Lugosi 2001, Theorems 3.1 & 3.2).

Applying the triangle inequality and then the symmetrization inequality (6), yields:

$$\begin{aligned}
\mathbb{E} \sup_{A \subseteq \mathcal{A}} |\hat{Q}_n(A) - \mathbb{Q}(A)| &\leq \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \frac{\bar{H}_n(A)}{\bar{M}_n} - \frac{\bar{H}_n(A)}{m} \right| + \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \frac{\bar{H}_n(A)}{m} - \mathbb{Q}(A) \right| \\
&\leq \frac{1}{m} \mathbb{E} |\bar{M}_n - m| + \frac{1}{m n} \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \sum_{i=1}^n \rho_i (H_i(A) - H'_i(A)) \right| \\
&= \frac{\sqrt{\text{Var}(M)}}{m\sqrt{n}} + \frac{1}{m n} \mathbb{E} \Psi(\{\mathbf{Y}_k\}, \{\mathbf{Y}'_k\}),
\end{aligned}$$

where we define the conditional expectation

$$\Psi(\{\mathbf{Y}_k\}, \{\mathbf{Y}'_k\}) := \mathbb{E} \left[\sup_{A \subseteq \mathcal{A}} \left| \sum_{i=1}^n \rho_i (H_i(A) - H'_i(A)) \right| \middle| \mathcal{Y}_U, \mathcal{Y}'_U \right],$$

and the last expectation is with respect to $\boldsymbol{\rho}$. Let $\check{\mathcal{A}} \subset \mathcal{A}$ be the collection of sets such that all intersections with the pointset $\{\mathbf{Y}_1, \dots, \mathbf{Y}_{T_n}, \mathbf{Y}'_1, \dots, \mathbf{Y}'_{T_n}\}$ are represented once, and any two sets in $\check{\mathcal{A}}$ are different. Observe that

$$\Psi(\{\mathbf{Y}_k\}, \{\mathbf{Y}'_k\}) = \mathbb{E} \left[\sup_{A \subseteq \check{\mathcal{A}}} \left| \sum_{i=1}^n \rho_i (H_i(A) - H'_i(A)) \right| \middle| \mathcal{Y}_U, \mathcal{Y}'_U \right]$$

and that $|\check{\mathcal{A}}| \leq \mathcal{S}_{\mathcal{A}}(T_n + T'_n)$.

Let $\|X\|_G < \infty$ denote the sub-Gaussian coefficient of the random variable X . In other words, the moment generating function of X satisfies

$$\mathbb{E} \exp(tX) \leq \exp(t^2 \|X\|_G^2 / 2), \quad \forall t.$$

We shall next use the maximal inequality

$$\mathbb{E} \max_{k \in \mathcal{K}} |X_k| \leq \sqrt{2 \ln(2|\mathcal{K}|)} \max_{k \in \mathcal{K}} \|X_k\|_G \quad (11)$$

for a finite index set \mathcal{K} , which holds even if the X_k 's are dependent. We will also make use of the property that

$$\|\sum_k w_k X_k\|_G^2 = \sum_k w_k^2 \|X_k\|_G^2, \quad (12)$$

whenever X_1, X_2, \dots are independent. Conditioning on all $\{\mathbf{Y}_k\}, \{\mathbf{Y}'_k\}$, and taking expectation over $\boldsymbol{\rho}$, we obtain:

$$\begin{aligned} \Psi(\{\mathbf{Y}_k\}, \{\mathbf{Y}'_k\}) &= \mathbb{E}_{\boldsymbol{\rho}} \sup_{A \subseteq \mathcal{A}} \left| \sum_{k=1}^n \rho_k (H_k(A) - H'_k(A)) \right| \\ \text{(maximal ineq.)} &\stackrel{(11)}{\leq} \sqrt{2 \ln(2\mathcal{S}_{\mathcal{A}}(T_n + T'_n))} \sup_{A \subseteq \mathcal{A}} \left\| \sum_{k=1}^n \rho_k (H_k(A) - H'_k(A)) \right\|_G \\ \text{(Sauer's Lemma)} &\stackrel{(5)+(12)}{\leq} \sqrt{2 \ln(2[(T_n + T'_n)\mathbf{e}/v]^v)} \sup_{A \subseteq \mathcal{A}} \sqrt{\sum_{k=1}^n \|\rho_k (H_k(A) - H'_k(A))\|_G^2} \\ &\leq \sqrt{2 \ln(2[(T_n + T'_n)\mathbf{e}/v]^v)} \sup_{A \subseteq \mathcal{A}} \sqrt{\sum_{k=1}^n (H_k(A) \vee H'_k(A))^2} \\ &\leq \sqrt{2 \ln(2[(T_n + T'_n)\mathbf{e}/v]^v)} \sum_{k=1}^n (M_k \vee M'_k). \end{aligned}$$

Therefore, using the bound $(R_i := (M_i \vee M'_i), r_n^2 := \frac{1}{n} \sum_i R_i^2)$:

$$\begin{aligned} \mathbb{E} r_n^2 \ln(T_n + T'_n) &\leq \ln(2n) \mathbb{E} r_n^2 + \mathbb{E} \left(\frac{1}{n} \sum_i R_i^2 \right) \ln \left(\frac{1}{n} \sum_i R_i \right) \\ \text{(Cauchy-Schwartz)} &\leq \ln(2n) \mathbb{E} r_n^2 + \mathbb{E} \left(\frac{1}{n} \sum_i R_i^2 \right) \frac{1}{2} \ln \left(\frac{1}{n} \sum_i R_i^2 \right) \\ \text{(Jensen's on } x \ln(x)) &\leq \ln(2n) \mathbb{E} r_n^2 + \mathbb{E} \frac{1}{2n} \sum_i R_i^2 \ln(R_i^2) \\ &= \ln(2n) \mathbb{E} R^2 + \mathbb{E} R^2 \ln(R) \\ &\leq 2 \ln(2n) \mathbb{E} M^2 + 2 \mathbb{E} M^2 \ln(M), \end{aligned}$$

we obtain:

$$\begin{aligned} \mathbb{E} \sup_{A \subseteq \mathcal{A}} |\widehat{\mathbb{Q}}_n(A) - \mathbb{Q}(A)| &\leq \frac{\sqrt{\text{Var}(M)}}{m\sqrt{n}} + \frac{\mathbb{E}\sqrt{2(\ln(2)+v\ln(T_n+T'_n)+v-v\ln(v))r_n^2}}{m\sqrt{n}} \\ &\leq \frac{\sqrt{\text{Var}(M)}}{m\sqrt{n}} + \frac{2\sqrt{(\ln(2)+v+v\ln(2n/v))\mathbb{E}M^2+v\mathbb{E}M^2\ln M}}{m\sqrt{n}} \\ &\leq \frac{\sqrt{\text{Var}(M)}}{m\sqrt{n}} + \frac{2\sqrt{v\ln(2n)\mathbb{E}[M^2\ln M]}}{m\sqrt{n}} \psi_1(v, n), \end{aligned}$$

where

$$\psi_1(v, n) = \sqrt{\frac{(\ln(2) + v + v \ln(2n/v))\mathbb{E}M^2}{v \ln(2n)\mathbb{E}M^2 \ln M} + \frac{1}{\ln(2n)}}.$$

This completes the proof of the theorem.

1.6. Proof of Theorem 6 (Second Bound on Expected TV for Empirical Distribution)

We need to introduce more working notation. First, recall a number of standard definitions. Define the weighted $L_p(\mathbb{P})$ metric on the probability space $(\mathbb{R}^d, \mathcal{B}, \mathbb{P})$ via the norm $\|X\|_p := (\int_{\mathbb{R}^d} |X(\omega)|^p d\mathbb{P}(\omega))^{1/p}$, $p \geq 1$. Let \mathcal{F} be a class of functions. An ϵ -cover of \mathcal{F} under the $L_p(\mathbb{P})$ metric is a finite set $\mathcal{C} = \{f_1, \dots, f_{|\mathcal{C}|}\}$ with cardinality $|\mathcal{C}|$ such that for every $f \in \mathcal{F}$ there exists an $f_k \in \mathcal{C}$ that satisfies $\|f - f_k\|_p \leq \epsilon$. Let \mathcal{C}^* be the ϵ -cover with the smallest cardinality. The cardinality of the smallest ϵ -cover of \mathcal{F} under the metric $L_p(\mathbb{P})$ is called the *covering number* and is denoted by $\mathcal{N}(\epsilon, \mathcal{F}, L_p(\mathbb{P}))$. We will write $\mathcal{N}(\epsilon, \mathcal{F}, L_p(\mathbb{P})) = \mathcal{N}(\epsilon, \mathcal{F})$ if the metric is clear from the context.

Recall that \mathcal{Y}_U with $T_n = n\bar{M}_n = |\mathcal{Y}_U|$ is the agglomeration of all the final states from n independent runs of Algorithm 1 (GS Sampler 1). Since the splitting factor is s , we have $M \leq s^\tau$. Denote $\sigma_n^2 := \frac{1}{n} \sum_{j=1}^n M_j^2$. We know that $\sigma_n \leq s^\tau$. For each index $k = 0, 1, \dots, K := \lceil \log_s(s^\tau \sqrt{n}) \rceil$, we define a cover as follows.

Conditional on \mathcal{Y}_U , we let \mathcal{C}_k^* be the smallest $(\sigma_n s^{-k})$ -cover of the set of functions

$$\mathcal{F}(\mathcal{Y}_U) = \{(H_1(A), \dots, H_n(A)); A \in \mathcal{A}\},$$

under the weighted metric with norm $\|\mathbf{h}\|_2 := \sqrt{\frac{1}{n} \sum_{j=1}^n h_j^2}$.

Observe that the zero vector is within $\sigma_n s^{-0}$ radius of all elements of $\mathcal{F}(\mathcal{Y}_U)$, and that $\mathcal{C}_0 = \{\mathbf{0}\}$ is an minimal $(\sigma_n s^{-0})$ -cover, that is, $\mathcal{N}(\sigma_n s^{-0}, \mathcal{F}(\mathcal{Y}_U)) = 1$. Further, the minimal ϵ -cover for $\epsilon \in [0, \frac{1}{\sqrt{n}})$ contains all the elements of $\mathcal{F}(\mathcal{Y}_U)$, that is, $\mathcal{N}(\sigma_n s^{-K}, \mathcal{F}(\mathcal{Y}_U)) = |\mathcal{F}(\mathcal{Y}_U)| = |\mathcal{C}_K^*|$.

Conditional on \mathcal{Y}_U , we let $\mathbf{h} = (h_1, \dots, h_n)$ be the vector with components $h_j(A) = \sum_{k \in \mathcal{Y}_j} \mathbb{I}\{\mathbf{Y}_k \in A\}$ (each h_j is a conditional version of H_j). For a given $\boldsymbol{\rho} = (\rho_1, \dots, \rho_n)^\top$, let \mathbf{h}^* correspond to the vector maximizing

$$\sup_{A \subseteq \mathcal{A}} \left| \sum_{k=1}^n \rho_k h_k(A) \right| = \left| \sum_{k=1}^n \rho_k h_k^* \right| = |\boldsymbol{\rho}^\top \mathbf{h}^*|.$$

Then, for $k = 0, \dots, K$, let \mathbf{h}_k be the vector in the minimal cover \mathcal{C}_k^* , which is closest to \mathbf{h}^* , that is $\|\mathbf{h}_k - \mathbf{h}^*\|_2 = \inf_{\mathbf{h} \in \mathcal{C}_k^*} \|\mathbf{h} - \mathbf{h}^*\|_2 \leq \sigma_n s^{-k}$. It follows that $\mathbf{h}^* = \mathbf{h}_K = \sum_{k=1}^K (\mathbf{h}_k - \mathbf{h}_{k-1})$. By the triangle inequality we have

$$\|\mathbf{h}_k - \mathbf{h}_{k-1}\|_2 \leq \|\mathbf{h}_k - \mathbf{h}^*\|_2 + \|\mathbf{h}_{k-1} - \mathbf{h}^*\|_2 \leq (s+1)\sigma_n s^{-k}.$$

Hence,

$$\begin{aligned} |\boldsymbol{\rho}^\top \mathbf{h}^*| &= \left| \sum_{k=1}^K \boldsymbol{\rho}^\top (\mathbf{h}_k - \mathbf{h}_{k-1}) \right| \leq \sum_{k=1}^K |\boldsymbol{\rho}^\top (\mathbf{h}_k - \mathbf{h}_{k-1})| \\ &\leq \sum_{k=1}^K \max_{\substack{\mathbf{h} \in \mathcal{C}_k^*, \mathbf{h}' \in \mathcal{C}_{k-1}^* \\ \|\mathbf{h} - \mathbf{h}'\|_2 < (s+1)\sigma_n s^{-k}}} |\boldsymbol{\rho}^\top (\mathbf{h} - \mathbf{h}')| \end{aligned} \quad (13)$$

Taking expectation with respect to $\boldsymbol{\rho}$ and using the maximal inequality (11), we thus obtain

$$\mathbb{E}_{\boldsymbol{\rho}} \max_{\substack{\mathbf{h} \in \mathcal{C}_k^*, \mathbf{h}' \in \mathcal{C}_{k-1}^* \\ \|\mathbf{h} - \mathbf{h}'\|_2 < (s+1)\sigma_n s^{-k}}} |\boldsymbol{\rho}^\top (\mathbf{h} - \mathbf{h}')| \leq \sqrt{2 \ln(2|\mathcal{C}_{k-1}^*||\mathcal{C}_k^*|)} (s+1)\sigma_n s^{-k}.$$

Therefore, taking expectation over \mathcal{Y}_U :

$$\mathbb{E} |\boldsymbol{\rho}^\top \mathbf{h}^*| \leq (s+1) \sum_{k=1}^K s^{-k} \mathbb{E} \left[\sigma_n \sqrt{2 \ln(2|\mathcal{C}_k^*|^2)} \right] \leq (s+1) \sum_{k=1}^K s^{-k} \sqrt{2 \mathbb{E} [\ln(2|\mathcal{C}_k^*|^2) \sigma_n^2]}.$$

Finally, from the triangle inequality and symmetrization inequality (6), we have

$$\begin{aligned} \mathbb{E} \sup_{A \subseteq \mathcal{A}} |\widehat{\mathbb{Q}}_n(A) - \mathbb{Q}(A)| &\leq \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \frac{\bar{H}_n(A)}{\bar{M}_n} - \frac{\bar{H}_n(A)}{m} \right| + \mathbb{E} \sup_{A \subseteq \mathcal{A}} \left| \frac{\bar{H}_n(A)}{m} - \mathbb{Q}(A) \right| \\ &\leq \frac{1}{m} \mathbb{E} |\bar{M}_n - m| + \frac{2}{m n} \mathbb{E} \sup_{A \subseteq \mathcal{B}} \left| \sum_{i=1}^n \rho_i H_i(A) \right| \\ &\leq \frac{\sqrt{\text{Var}(M)}}{m \sqrt{n}} + \frac{(s+1)2\sqrt{2}}{m n} \sum_{k=1}^K s^{-k} \sqrt{\mathbb{E} [\sigma_n^2 \ln(2\mathcal{N}^2(\sigma_n s^{-k}, \mathcal{F}(\mathcal{Y}_U)))]}. \end{aligned}$$

It thus remains to bound the metric entropy $\ln \mathcal{N}(\sigma_n \epsilon, \mathcal{F}(\mathcal{Y}_\cup))$. For a fixed \mathcal{Y}_\cup , let $\mathcal{C}'_1, \dots, \mathcal{C}'_n$ be minimal ϵ -covers corresponding to each of the n binary function classes ($j = 1, \dots, n$):

$$\mathcal{F}(\mathcal{Y}_j) = \{(b_1, \dots, b_{M_j}) : A \in \mathcal{A}, b_i = \mathbb{I}\{\mathbf{Y}_i \in A\}, \mathbf{Y}_i \in \mathcal{Y}_j\}.$$

This implies that for any $\mathbf{b}_j \in \mathcal{F}(\mathcal{Y}_j)$, there exists an $\mathbf{s}_j \in \mathcal{C}'_j$ such that:

$$\|\mathbf{b}_j - \mathbf{s}_j\|_2 = \sqrt{\frac{1}{M_j} \sum_{k=1}^{M_j} (b_j^{(k)} - s_j^{(k)})^2} \leq \epsilon.$$

Then, the set $\left\{s_j^{(1)} + \dots + s_j^{(M_j)} : \mathbf{s}_j \in \mathcal{C}'_j, j = 1, \dots, n\right\}$ is an $\sigma_n \epsilon$ -cover of $\mathcal{F}(\mathcal{Y}_\cup)$. To see this, note that for any $\mathbf{h} \in \mathcal{F}(\mathcal{Y}_\cup)$, we have

$$h_j \in \left\{b_j^{(1)} + \dots + b_j^{(M_j)} : \mathbf{b}_j \in \mathcal{F}(\mathcal{Y}_j)\right\}, \quad j = 1, \dots, n$$

and by the Cauchy-Schwartz inequality:

$$\begin{aligned} \left\| \mathbf{h} - \left(\sum_{k=1}^{M_1} s_1^{(k)}, \dots, \sum_{k=1}^{M_n} s_n^{(k)} \right) \right\|_2^2 &= \frac{1}{n} \sum_{j=1}^n \left(h_j - \sum_{k=1}^{M_j} s_j^{(k)} \right)^2 \\ &= \frac{1}{n} \sum_{j=1}^n M_j^2 \left(\frac{1}{M_j} \sum_{k=1}^{M_j} (b_j^{(k)} - s_j^{(k)}) \right)^2 \\ &\leq \frac{1}{n} \sum_{j=1}^n M_j^2 \left(\frac{1}{M_j} \sum_{k=1}^{M_j} |b_j^{(k)} - s_j^{(k)}| \right)^2 \\ &\leq \frac{1}{n} \sum_{j=1}^n M_j^2 \left(\frac{1}{M_j} \sum_{k=1}^{M_j} |b_j^{(k)} - s_j^{(k)}|^2 \right) \leq \sigma_n^2 \epsilon^2. \end{aligned}$$

Using the inequality of Haussler (1995)

$$\ln \mathcal{N}(\epsilon, \mathcal{F}(\mathcal{Y}_j)) \leq \ln(\mathbf{e}[v+1]) + v \ln(2\mathbf{e}/\epsilon^2), \quad \epsilon \in [0, 1] \quad (14)$$

for the cover number of a class of sets \mathcal{A} with VC dimension $v < \infty$, we thus have the bound on the metric entropy of $\mathcal{F}(\mathcal{Y}_\cup)$:

$$\ln \mathcal{N}(\sigma_n \epsilon, \mathcal{F}(\mathcal{Y}_\cup)) \leq \sum_{j=1}^n \ln \mathcal{N}(\epsilon, \mathcal{F}(\mathcal{Y}_j)) \stackrel{(14)}{\leq} n(\ln(\mathbf{e}[v+1]) + v \ln(2\mathbf{e}/\epsilon^2)), \quad \epsilon \in [0, 1].$$

Hence, combining all the results so far we obtain the upper bound for $\mathbb{E} \sup_{A \subseteq \mathcal{A}} |\widehat{\mathbb{Q}}_n(A) - \mathbb{Q}(A)| - \frac{\sqrt{\text{Var}(M)}}{m\sqrt{n}}$:

$$\begin{aligned} \frac{(s+1)2\sqrt{2}}{mn} \sum_{k=1}^K \frac{\sqrt{\mathbb{E}[\sigma_n^2 \ln(2\mathcal{N}^2(\sigma_n s^{-k}, \mathcal{F}(\mathcal{Y}_\cup)))]}}{s^k} &\leq \frac{(s+1)4\sqrt{v\mathbb{E}M^2}}{m\sqrt{n}} \sum_{k=1}^K \frac{\sqrt{\frac{\ln(2)}{2nv} + \frac{\ln(\mathbf{e}[v+1])}{v} + \ln(2\mathbf{e}s^{2k})}}{s^k} \\ &\leq \frac{(s+1)4\sqrt{v\mathbb{E}M^2}}{m\sqrt{n}} \psi_2(\tau, v, n). \end{aligned}$$

Hence, the result of the theorem follows.

References

- Asmussen S (2008) *Applied probability and queues*, volume 51 (Springer Science & Business Media).
- Awad HP, Glynn PW (2007) On the theoretical comparison of low-bias steady-state simulation estimators. *ACM Transactions on Modeling and Computer Simulation* 17(1):4, ISSN 1049-3301, URL <http://dx.doi.org/http://doi.acm.org/10.1145/1189756.1189760>.
- Devroye L, László G, Gábor L (2013) *A probabilistic theory of pattern recognition* (New York: Springer-Verlag).
- Devroye L, Lugosi G (2001) *Combinatorial methods in density estimation* (Springer, New-York).
- Giné E, Zinn J (1984) Some limit theorems for empirical processes. *The Annals of Probability* 12(4):929–989.
- Glynn PW (2006) Simulation algorithms for regenerative processes. Henderson SG, Nelson BL, eds., *Simulation*, 477–500, Handbooks in Operations Research and Management Science (Amsterdam, The Netherlands: Elsevier), chapter 16.
- Haussler D (1995) Sphere packing numbers for subsets of the boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory, Series A* 69(2):217–232.
- L'Ecuyer P, Botev ZI, Kroese DP (2018) On a generalized splitting method for sampling from a conditional distribution. *Proceedings of the 2018 Winter Simulation Conference*, 1694–1705 (IEEE Press).
- Lorden G (1970) On excess over the boundary. *The Annals of Mathematical Statistics* 41(2):520–527.
- Meketon MS, Heidelberger P (1982) A renewal theoretic approach to bias reduction in regenerative simulations. *Management Science* 26:173–181.
- Rao RR (1962) Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics* 33(2):659–680.
- Sauer N (1972) On the density of families of sets. *Journal of Combinatorial Theory, Series A* 13(1):145–147.
- Vapnik V (2013) *The nature of statistical learning theory* (Springer-Verlag).