

On a Generalized Splitting Method for Sampling From a Conditional Distribution

Pierre L'Ecuyer

Université de Montréal, Canada, and Inria–Rennes, France

Zdravko I. Botev

The University of New South Wales, Sydney, Australia

Dirk P. Kroese

The University of Queensland, Brisbane, Australia

Winter Simulation Conference, Göteborg, December 2018

Sampling Conditional on a Rare Event

Random vector \mathbf{Y} in \mathbb{R}^d , with density f . Suppose we know how to sample \mathbf{Y} from f .

We want to sample it from f conditional on $\mathbf{Y} \in B$, for some $B \subset \mathbb{R}^d$ for which $p = \mathbb{P}[\mathbf{Y} \in B]$ is very small.

Sampling Conditional on a Rare Event

Random vector \mathbf{Y} in \mathbb{R}^d , with density f . Suppose we know how to sample \mathbf{Y} from f .

We want to sample it from f conditional on $\mathbf{Y} \in B$, for some $B \subset \mathbb{R}^d$ for which $p = \mathbb{P}[\mathbf{Y} \in B]$ is very small.

What for?

It could be to estimate the conditional expectation $\mathbb{E}[h(\mathbf{Y}) \mid \mathbf{Y} \in B]$ for some real-valued cost function h , or to estimate $\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in B)]$ or to estimate the conditional density, for example.

There are many applications (CVaR, approximate zero-variance importance sampling, etc.).

In Bayesian statistics, \mathbf{Y} may represent a vector of parameters with given prior distribution and we may want to sample it from the posterior distribution given the data.

Rejection sampling

Algorithm 1: Standard rejection

```
while true do  
  Sample  $\mathbf{Y}$  from its unconditional density  $f$   
  if  $\mathbf{Y} \in B$  then  
    return  $\mathbf{Y}$ 
```

To get an independent sample of size n , we can repeat n times independently.

But if $p = \mathbb{P}[\mathbf{Y} \in B]$ is very small, i.e., $\{\mathbf{Y} \in B\}$ is a **rare event**, this is **too inefficient**.

Markov chain Monte Carlo (MCMC)

Suppose we can construct an artificial Markov chain whose **stationary distribution** is the target one, i.e., the distribution of \mathbf{Y} conditional on $\mathbf{Y} \in B$.

We can start this Markov chain at some arbitrary state $\mathbf{y}_0 \in B$, run it for $n_0 + n$ steps for some large enough n_0 , and retain the last n visited states as our sample.

Markov chain Monte Carlo (MCMC)

Suppose we can construct an artificial Markov chain whose **stationary distribution** is the target one, i.e., the distribution of \mathbf{Y} conditional on $\mathbf{Y} \in B$.

We can start this Markov chain at some arbitrary state $\mathbf{y}_0 \in B$, run it for $n_0 + n$ steps for some large enough n_0 , and retain the last n visited states as our sample.

But many issues arise. How large should be n_0 ? Convergence to the stationary distribution may be very slow and the n retained states are typically highly dependent.

Sometimes, we may also not know how to pick a valid state \mathbf{y}_0 from B .

Generalized splitting (GS)

Botev and Kroese (2012) proposed a **generalized splitting** approach as an alternative, to sample **approximately** from the conditional density. To apply GS, we need to choose:

Generalized splitting (GS)

Botev and Kroese (2012) proposed a **generalized splitting** approach as an alternative, to sample **approximately** from the conditional density. To apply GS, we need to choose:

1. an **importance function** S for which $\{\mathbf{y} : S(\mathbf{y}) > \gamma^*\} = B$ for some $\gamma^* > 0$,
2. an integer **splitting factor** $s \geq 2$, and
3. a number $\tau > 0$ of **levels** $0 = \gamma_0 < \gamma_1 < \dots < \gamma_\tau = \gamma^*$ for which

$$\mathbb{P}[S(\mathbf{Y}) > \gamma_t \mid S(\mathbf{Y}) > \gamma_{t-1}] \approx 1/s, \quad \text{for } t = 1, \dots, \tau - 1.$$

Generalized splitting (GS)

Botev and Kroese (2012) proposed a **generalized splitting** approach as an alternative, to sample **approximately** from the conditional density. To apply GS, we need to choose:

1. an **importance function** S for which $\{\mathbf{y} : S(\mathbf{y}) > \gamma^*\} = B$ for some $\gamma^* > 0$,
2. an integer **splitting factor** $s \geq 2$, and
3. a number $\tau > 0$ of **levels** $0 = \gamma_0 < \gamma_1 < \dots < \gamma_\tau = \gamma^*$ for which

$$\mathbb{P}[S(\mathbf{Y}) > \gamma_t \mid S(\mathbf{Y}) > \gamma_{t-1}] \approx 1/s, \quad \text{for } t = 1, \dots, \tau - 1.$$

4. For each level t , an artificial **Markov chain** with transition density $\kappa_t(\mathbf{y} \mid \mathbf{x})$ and whose **stationary density** f_t is the density of \mathbf{Y} conditional on $S(\mathbf{Y}) > \gamma_t$:

$$f_t(\mathbf{y}) := f(\mathbf{y}) \frac{\mathbb{I}(S(\mathbf{y}) > \gamma_t)}{\mathbb{P}[S(\mathbf{Y}) > \gamma_t]}.$$

There are many ways of constructing these chains.

Algorithm 2: Generalized splitting

Require: $s, \tau, \gamma_1, \dots, \gamma_\tau$

Generate \mathbf{Y} from its unconditional density f

if $S(\mathbf{Y}) \leq \gamma_1$ **then**

return $\mathcal{Y}_\tau = \emptyset$ and $M = 0$ // state \mathbf{Y} does not reach first level; return empty list

else

$\mathcal{Y}_1 \leftarrow \{\mathbf{Y}\}$ // state \mathbf{Y} has reached at least the first level

for $t = 2$ **to** τ **do**

$\mathcal{Y}_t \leftarrow \emptyset$ // list of states that have reached level γ_t

for all $\mathbf{Y} \in \mathcal{Y}_{t-1}$ **do**

 set $\mathbf{Y}_0 = \mathbf{Y}$ // we will simulate this chain for s steps

for $j = 1$ **to** s **do**

 sample \mathbf{Y}_j from the density $\kappa_{t-1}(\cdot \mid \mathbf{Y}_{j-1})$

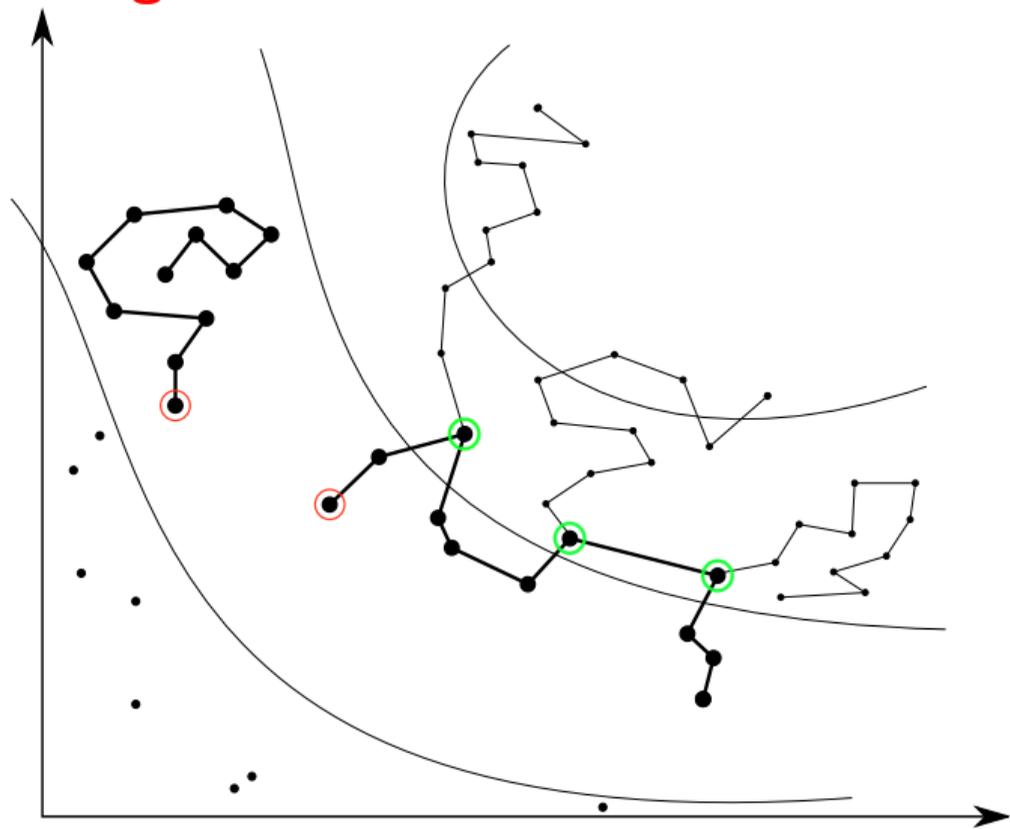
if $S(\mathbf{Y}_j) > \gamma_t$ **then**

 add \mathbf{Y}_j to \mathcal{Y}_t // this state has reached the next level

return the list \mathcal{Y}_τ and its cardinality $M = |\mathcal{Y}_\tau|$. // list of states that have reached B

At each step, \mathcal{Y}_t is the set of states that have reached level t .

Generalized Splitting



Choice of parameters

In many applications there is a natural choice for the importance function S .

Good values for s , τ , and the levels $\{\gamma_t\}$ can typically be found adaptively via an (independent) pilot experiment.

Based on our experience, taking $s = 2$ is usually best.

Some questions left partially open in the 2012 paper

- ▶ Are the final states of the set of trajectories **exactly or approximately** distributed according according to the density f conditional on B ?
In empirical experiments with rare events, we observed that the empirical was very close to the conditional, and it **was unclear if there was bias or not**.

Some questions left partially open in the 2012 paper

- ▶ Are the final states of the set of trajectories **exactly or approximately** distributed according to the density f conditional on B ?
In empirical experiments with rare events, we observed that the empirical was very close to the conditional, and it **was unclear if there was bias or not**.
- ▶ Does GS provide an unbiased estimator of the **conditional expectation** of a function of Y , given $\mathbf{Y} \in B$?

Some questions left partially open in the 2012 paper

- ▶ Are the final states of the set of trajectories **exactly or approximately** distributed according to the density f conditional on B ?
In empirical experiments with rare events, we observed that the empirical was very close to the conditional, and it **was unclear if there was bias or not**.
- ▶ Does GS provide an unbiased estimator of the **conditional expectation** of a function of Y , given $\mathbf{Y} \in B$?
- ▶ Let $M = |\mathcal{Y}_\tau|$ be the number of particles that end up in B . If we pick at random one of those M terminal particles from a given run of GS, assuming that $M > 1$, **is this particle distributed according to $f_\tau(\cdot) = f(\cdot | \mathbf{Y} \in B)$?**

Some questions left partially open in the 2012 paper

- ▶ Are the final states of the set of trajectories **exactly or approximately** distributed according to the density f conditional on B ?
In empirical experiments with rare events, we observed that the empirical was very close to the conditional, and it **was unclear if there was bias or not**.
- ▶ Does GS provide an unbiased estimator of the **conditional expectation** of a function of Y , given $\mathbf{Y} \in B$?
- ▶ Let $M = |\mathcal{Y}_\tau|$ be the number of particles that end up in B . If we pick at random one of those M terminal particles from a given run of GS, assuming that $M > 1$, **is this particle distributed according to $f_\tau(\cdot) = f(\cdot | \mathbf{Y} \in B)$?**
- ▶ If we run GS r times, independently, and collect the terminal states of all the trajectories that have reached the rare event over the r runs, does their empirical distribution **converge to the conditional distribution given B** , and how fast?

Unbiasedness for each fixed potential branch

To study the previous questions, we will consider an imaginary version of the GS algorithm in which all $s^{\tau-1}$ potential trajectories are considered. For those that do not reach the next level in the GS algorithm, we assume that there are phantom trajectories that are continued at all levels.

For $t = 1, 2, \dots, \tau$, denote by $\bar{\mathcal{Y}}_t$ the corresponding set of s^{t-1} states at step t .

Let $\mathbf{Y}(1, j_2, \dots, j_t) \in \bar{\mathcal{Y}}_t$ denote the state coming from the branch going through the j_2 th state of step 2, j_3 th state of step 3, \dots , and currently the j_t th state at step t , where each $j_i \in \{1, \dots, s\}$.

Unbiasedness for each fixed potential branch

The trajectories that are kept alive up to level t in the original algorithm are those for which the following event occurs:

$$\mathcal{E}_t(1, j_2, \dots, j_t) := \{\mathbf{Y}(1) > \gamma_1, \dots, \mathbf{Y}(1, j_2, \dots, j_t) > \gamma_t\}.$$

Proposition 1. For any fixed level t and index $(1, j_2, \dots, j_t)$, conditional on $\mathcal{E}_t(1, j_2, \dots, j_t)$, the state $\mathbf{Y}(1, j_2, \dots, j_t)$ has density f_t (exactly).
For $t = \tau$, this is the density of \mathbf{Y} conditional on $\{\mathbf{Y} \in B\}$.

This can be proved by induction on t .

Unbiasedness for expectation

GS provides an unbiased estimator of $\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A)]$:

Proposition 2. For any (measurable) function h and subset $A \subseteq B$, we have

$$\mathbb{E} \left[s^{1-\tau} \sum_{\mathbf{Y} \in \mathcal{Y}_\tau} h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A) \right] = \mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A)],$$

where the left expectation is with respect to \mathcal{Y}_τ and the one on the right is with respect to the original density f of \mathbf{Y} .

Unbiasedness for expectation

GS provides an unbiased estimator of $\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A)]$:

Proposition 2. For any (measurable) function h and subset $A \subseteq B$, we have

$$\mathbb{E} \left[s^{1-\tau} \sum_{\mathbf{Y} \in \mathcal{Y}_\tau} h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A) \right] = \mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in A)],$$

where the left expectation is with respect to \mathcal{Y}_τ and the one on the right is with respect to the original density f of \mathbf{Y} .

By taking $A = B$ and $h = 1$ we find that $\mathbb{E}[M] = \mathbb{P}(\mathbf{Y} \in B)s^{\tau-1}$.

Sampling from the Conditional Density and Estimating the Conditional Expectation

Each run of the GS algorithm returns a sample $\mathbf{Y}_1, \dots, \mathbf{Y}_M$ of random size M , taking values in B . In view of the previous unbiasedness results, one may expect that if we pick \mathbf{Y}^* at random uniformly from $\mathbf{Y}_1, \dots, \mathbf{Y}_M$, conditional on $M \geq 1$ (if $M = 0$ we just retry), then this \mathbf{Y}^* will have the **conditional density** $f_{\tau}(\cdot) = f(\cdot \mid \mathbf{Y} \in B)$ and that $h(\mathbf{Y}^*)$ will be an unbiased estimator of the **conditional expectation** $\mathbb{E}[h(\mathbf{Y}) \mid \mathbf{Y} \in B]$.

Sampling from the Conditional Density and Estimating the Conditional Expectation

Each run of the GS algorithm returns a sample $\mathbf{Y}_1, \dots, \mathbf{Y}_M$ of random size M , taking values in B . In view of the previous unbiasedness results, one may expect that if we pick \mathbf{Y}^* at random uniformly from $\mathbf{Y}_1, \dots, \mathbf{Y}_M$, conditional on $M \geq 1$ (if $M = 0$ we just retry), then this \mathbf{Y}^* will have the conditional density $f_{\tau}(\cdot) = f(\cdot | \mathbf{Y} \in B)$ and that $h(\mathbf{Y}^*)$ will be an unbiased estimator of the conditional expectation $\mathbb{E}[h(\mathbf{Y}) | \mathbf{Y} \in B]$.

Unfortunately, this is not true.

The problem is that \mathbf{Y}^* and M are not independent!

Formulation as a ratio of expectations

To see what goes on, let $H = \sum_{\mathbf{Y} \in \mathcal{Y}_\tau} h(\mathbf{Y})$. We can write

$$\mathbb{E}[h(\mathbf{Y}) \mid \mathbf{Y} \in B] = \frac{\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in B)]}{\mathbb{P}[\mathbf{Y} \in B]} = \frac{\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in B)]s^{\tau-1}}{\mathbb{E}[M]} = \frac{\mathbb{E}[H]}{\mathbb{E}[M]} := \nu,$$

which is a **ratio of expectations**.

For \mathbf{Y}^* picked uniformly from $\{\mathbf{Y}_1, \dots, \mathbf{Y}_M\}$, we have

$$\mathbb{E}[h(\mathbf{Y}^*)] = \mathbb{E} \left[\frac{1}{M} \sum_{\mathbf{Y}^* \in \mathcal{Y}_\tau} h(\mathbf{Y}^*) \right] = \mathbb{E}[H/M],$$

which is the **expectation of a ratio**.

Formulation as a ratio of expectations

To see what goes on, let $H = \sum_{\mathbf{Y} \in \mathcal{Y}_\tau} h(\mathbf{Y})$. We can write

$$\mathbb{E}[h(\mathbf{Y}) \mid \mathbf{Y} \in B] = \frac{\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in B)]}{\mathbb{P}[\mathbf{Y} \in B]} = \frac{\mathbb{E}[h(\mathbf{Y})\mathbb{I}(\mathbf{Y} \in B)]s^{\tau-1}}{\mathbb{E}[M]} = \frac{\mathbb{E}[H]}{\mathbb{E}[M]} := \nu,$$

which is a **ratio of expectations**.

For \mathbf{Y}^* picked uniformly from $\{\mathbf{Y}_1, \dots, \mathbf{Y}_M\}$, we have

$$\mathbb{E}[h(\mathbf{Y}^*)] = \mathbb{E}\left[\frac{1}{M} \sum_{\mathbf{Y}^* \in \mathcal{Y}_\tau} h(\mathbf{Y}^*)\right] = \mathbb{E}[H/M],$$

which is the **expectation of a ratio**.

Special case: $h(\mathbf{Y}) = \mathbb{I}(\mathbf{Y} \in A)$ to estimate $\mathbb{P}(A)$, for any $A \subseteq B$.

How to estimate ν consistently?

We can estimate $\nu = \mathbb{E}[h(\mathbf{Y}) \mid \mathbf{Y} \in B]$ via standard techniques for ratios of expectations.

Classical approach:

1. Simulate n independent replicates of the GS estimator, let $\mathcal{Y}_{\tau,1}, \dots, \mathcal{Y}_{\tau,n}$ be the n sets \mathcal{Y}_{τ} obtained from these realizations, and let $M_i = |\mathcal{Y}_{\tau,i}|$ and $H_i = \sum_{\mathbf{Y} \in \mathcal{Y}_{\tau,i}} h(\mathbf{Y})$ be the realizations of M and H for replicate i , for $i = 1, \dots, n$.
2. Return the **ratio estimator**

$$\hat{\nu}_n = \frac{H_1 + \dots + H_n}{M_1 + \dots + M_n}.$$

This estimator is asymptotically normal and standard formulas are available to compute a confidence interval on ν . Can also use bootstrap.

Sampling from the Conditional Distribution

To generate independent realizations of \mathbf{Y} **approximately** from its conditional distribution given B , we can pick the realizations at random with replacement from $\mathcal{Y}_{\cup} = \mathcal{Y}_{\tau,1} \cup \dots \cup \mathcal{Y}_{\tau,n}$. Let

$$\hat{Q}_n[A] = |\mathcal{Y}_{\cup} \cap A|/|\mathcal{Y}_{\cup}| \text{ (empirical dist.);} \quad Q_n[A] = \mathbb{E}[\hat{Q}_n[A]]; \quad Q[\cdot] = \mathbb{P}[\cdot | B] \text{ (true cond. dist.).}$$

Sampling from the Conditional Distribution

To generate independent realizations of \mathbf{Y} **approximately** from its conditional distribution given B , we can pick the realizations at random with replacement from $\mathcal{Y}_U = \mathcal{Y}_{\tau,1} \cup \dots \cup \mathcal{Y}_{\tau,n}$. Let

$$\hat{Q}_n[A] = |\mathcal{Y}_U \cap A|/|\mathcal{Y}_U| \text{ (empirical dist.);} \quad Q_n[A] = \mathbb{E}[\hat{Q}_n[A]]; \quad Q[\cdot] = \mathbb{P}[\cdot | B] \text{ (true cond. dist.).}$$

The empirical distribution \hat{Q}_n **converges to** Q in the following **large-deviation sense**:

Proposition 4. For any $\epsilon \in (0, p/2]$, we have $\sup_{A \subseteq B} \mathbb{P} \left[\left| \hat{Q}_n[A] - Q[A] \right| > 2\epsilon/p \right] \leq 4e^{-2n\epsilon^2}$.

The proof uses Hoeffding's inequality for numerator and denominator of $\hat{Q}_n[A] = \bar{H}_n(A)/\bar{M}_n$.

Sampling from the Conditional Distribution

To generate independent realizations of \mathbf{Y} **approximately** from its conditional distribution given B , we can pick the realizations at random with replacement from $\mathcal{Y}_{\mathcal{U}} = \mathcal{Y}_{\tau,1} \cup \dots \cup \mathcal{Y}_{\tau,n}$. Let

$$\widehat{\mathbb{Q}}_n[A] = |\mathcal{Y}_{\mathcal{U}} \cap A|/|\mathcal{Y}_{\mathcal{U}}| \text{ (empirical dist.);} \quad \mathbb{Q}_n[A] = \mathbb{E}[\widehat{\mathbb{Q}}_n[A]]; \quad \mathbb{Q}[\cdot] = \mathbb{P}[\cdot \mid B] \text{ (true cond. dist.).}$$

The empirical distribution $\widehat{\mathbb{Q}}_n$ **converges to \mathbb{Q}** in the following **large-deviation sense**:

$$\textbf{Proposition 4.} \text{ For any } \epsilon \in (0, p/2], \text{ we have } \sup_{A \subseteq B} \mathbb{P} \left[\left| \widehat{\mathbb{Q}}_n[A] - \mathbb{Q}[A] \right| > 2\epsilon/p \right] \leq 4e^{-2n\epsilon^2}.$$

The proof uses Hoeffding's inequality for numerator and denominator of $\widehat{\mathbb{Q}}_n[A] = \bar{H}_n(A)/\bar{M}_n$.

We also have **convergence of \mathbb{Q}_n** (but not of $\widehat{\mathbb{Q}}_n$) to \mathbb{Q} **in total variation** (proof in the paper).

$$\textbf{Corollary 5.} \quad \lim_{n \rightarrow \infty} \sup_{A \subseteq B} |\mathbb{Q}_n[A] - \mathbb{Q}[A]| = \frac{2}{p} \int_0^\infty 4e^{-2n\epsilon^2} d\epsilon = \frac{2\sqrt{2\pi}}{p\sqrt{n}} \rightarrow 0.$$

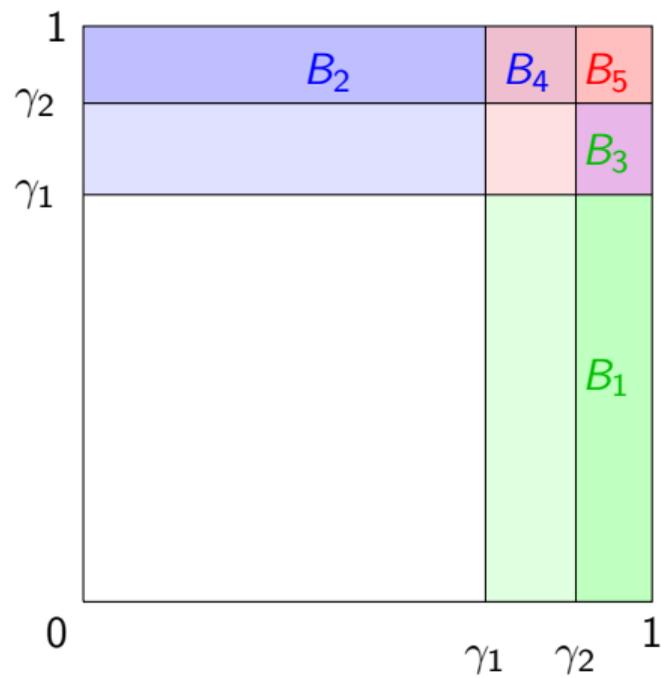
A simple bivariate uniform (counter)example

We illustrate the convergence behavior on a baby example, initially designed as a **counter-example** to show that the sampling is not exact. We have:

- ▶ $\mathbf{Y} = (Y_1, Y_2)$ uniform over $[0, 1]^2$, so f is the uniform density.
- ▶ $B = \{\mathbf{y} \in \mathcal{Y} : S(\mathbf{y}) > \gamma_2\}$.
- ▶ $S(\mathbf{y}) = S(y_1, y_2) = \max(y_1, y_2)$
- ▶ Number of levels $\tau = 2$.
- ▶ Two splitting level cases: $s = 2$ and $s = 10$.
- ▶ Level parameters $\gamma_1 = \sqrt{1 - s^{-1}}$ and $\gamma_2 = \sqrt{1 - s^{-2}}$.
- ▶ The true conditional density on B is constant, equal to s^2 , for $s = 2, 10$.

We want to compare the true conditional density with the density obtained from GS. We consider two choices of **Markov chain kernel** κ_t : Two-way and one-way resampling.

Bivariate Uniform Example



Two-way resampling

Here we use a **symmetric Gibbs sampling**: always resample the two coordinates one after the other, in random order, **conditional on $S(\mathbf{Y}) > \gamma_{t-1}$** .

Specifically, whenever Y_1 is resampled, if $Y_2 > \gamma_{t-1}$ we resample Y_1 uniformly over $(0, 1)$, otherwise we resample Y_1 uniformly over $(\gamma_{t-1}, 1)$. And similarly for Y_2 .

This produces a Markov chain trajectory over s steps, in the colored region. We retain the particles that fall in the set $B = S(\mathbf{Y}) > \gamma_2$ and they form the multiset \mathcal{Y}_2 .

This GS procedure is repeated n times independently and the n realizations of \mathcal{Y}_2 (in case $\tau = 2$) are merged in a single multiset \mathcal{Y}_U .

We wish to compare the density of \mathbb{Q}_n with that of the true conditional density in each region $B_i \subset B$, $i = 1, \dots, 5$.

Two-way resampling

We replicate the following experiment $r = 10^7$ times, independently:

- ▶ Perform n independent runs of GS and construct the random multiset \mathcal{Y}_U .
- ▶ If \mathcal{Y}_U is empty, this replicate has no contribution and we move to the next replicate.
- ▶ Otherwise, we compute the proportion of states in \mathcal{Y}_U that fall in each of the five regions B_1, \dots, B_5 , and divide each proportion by the area of the corresponding region, to obtain a conditional density given \mathcal{Y}_U .

To estimate the exact densities $d(B_1), \dots, d(B_5)$ over the five regions for the distribution of the retained state under in this setting, we simulated this process r times and averaged the conditional densities over the R_0 replications for which \mathcal{Y}_U was nonempty. We did this with $n = 1, 10, 100, 1000$.

Two-way resampling

Density estimates in each region with $r = 10^7$ independent replicates, with n independent runs of GS per replicate, for $s = 2$ and 10.

s	n	R_0	B_1	B_2	B_3	B_4	B_5
2	1	3756298	3.98	3.98	4.06	4.06	4.05
2	10	9909982	3.994	3.995	4.016	4.019	4.017
2	100	10000000	4.000	3.999	4.001	4.002	4.001
2	1000	10000000	4.000	4.000	4.000	4.000	4.000
10	1	651966	99.8	100.1	101.2	100.5	99.2
10	10	4902162	100.0	100.0	99.8	100.3	100.0
10	100	9987952	100.0	100.0	100.0	100.0	100.0
10	1000	10000000	100.0	100.0	100.0	100.0	99.9

\mathbb{Q}_n converges to \mathbb{Q} very quickly with n and is already quite close even with $n = 1$.

One-way resampling

Now we **resample only the first coordinate** Y_1 conditional on $S(\mathbf{Y}) > \gamma_1$.

We do not resample Y_2 .

In this case, all points $\mathbf{Y} \in \mathcal{Y}_\tau$ returned by GS on a given run have the same value of Y_2 .

We repeated the same simulation experiment with this poor resampling scheme, still with $r = 10^7$.

Although the bias is now much larger, we still observe convergence to the uniform density when n increases and this convergence is slower when s is larger.

Density estimates in each region, with $r = 10^7$ independent replicates and n independent runs of GS per replicate, for the one-way resampling, for $s = 2$ and 10.

s	n	R_0	B_1	B_2	B_3	B_4	B_5
2	1	3199391	4.82	3.12	5.84	3.13	3.13
2	10	9788291	4.30	3.68	4.67	3.68	3.68
2	100	10000000	4.022	3.977	4.049	3.978	3.978
2	1000	10000000	4.002	3.998	4.005	3.998	3.998
10	1	385940	170.5	25.9	254	25.9	26.4
10	10	3251227	167.8	28.9	244	28.9	29.2
10	100	9803488	140.8	58.1	166.7	58.1	57.9
10	1000	10000000	104.3	95.7	105.2	95.7	95.6

The bias in \mathbb{Q}_n is now much larger than for the two-way case, and is larger for $s = 10$ than for $s = 2$. A larger s amplifies the bias because it creates more dependence.

The density is higher in B_1 and B_3 .

Conclusions

- ▶ GS returns a random-sized sample of points such that **unconditionally** on the sample size, each point is distributed **exactly** according to the original distribution conditional on the rare event.
- ▶ For any measurable cost function which is nonzero only when the rare event occurs, the method provides an **unbiased** estimator of the expected cost.
- ▶ However, if we select at random one of the returned points, its distribution **differs** in general from the exact conditional distribution given the rare event.
- ▶ But if we **repeat** the algorithm n times and select one of the returned points at random, the distribution of the selected point **converges** to the exact one in total variation.
- ▶ The **empirical distribution** of the set of all points returned over all n replicates also converges to the conditional distribution given the rare event.

Some references

-  Botev, Z. I., and D. P. Kroese. 2012.
“Efficient Monte Carlo Simulation via the Generalized Splitting Method”.
Statistics and Computing 22(1):1–16.
-  Botev, Z. I., P. L'Ecuyer, G. Rubino, R. Simard, and B. Tuffin. 2013a.
“Static Network Reliability Estimation Via Generalized Splitting”.
INFORMS Journal on Computing 25(1):56–71.
-  Botev, Z. I., P. L'Ecuyer, and B. Tuffin. 2013b.
“Markov Chain Importance Sampling with Application to Rare Event Probability Estimation”.
Statistics and Computing 25(1):56–71.
-  Botev, Z. I., P. L'Ecuyer, and B. Tuffin. 2016.
“Static Network Reliability Estimation under the Marshall-Olkin Copula”.
ACM Transactions on Modeling and Computer Simulation 26(2):Article 14, 28 pages.
-  Botev, Z. I., P. L'Ecuyer, and B. Tuffin. 2018.
“Reliability Estimation for Networks with Minimal Flow Demand and Random Link Capacities”.
Submitted, available at <https://hal.inria.fr/hal-01745187>.
-  Choquet, D., P. L'Ecuyer, and C. Léger. 1999.
“Bootstrap Confidence Intervals for Ratios of Expectations”.
ACM Transactions on Modeling and Computer Simulation 9(4):326–348.

 L'Ecuyer, P., F. LeGland, P. Lezaud, and B. Tuffin. 2009.

“Splitting Techniques”.

In *Rare Event Simulation Using Monte Carlo Methods*, edited by G. Rubino and B. Tuffin, 39–62.

Chichester, U.K.: Wiley.