Simulations and monte carlo methods

The Coupled Rejection Sampler

Thibault Tatou Dekou, Maria Abi Rizk, Patryk Wisniewski

April 30, 2024

ENSAE 2A

Presentation structure

1. Brief recap on coupling
2. Validity of Thorisson's algorithm
3. Alternative to Thorisson and performance comparisons
4. The curse of dimensionality

Coupling 101

Definition: A *coupling* of random variables X and Y is a construction of a joint distribution for (X', Y') such that $X' \overset{d}{=} X$ and $Y' \overset{d}{=} Y$, preserving their individual distributions.

Theorem (Coupling Inequality)
Let μ and ν be probability measures on a measurable space $(S, \mathcal{S})$. For any coupling (X, Y) of μ and ν,

$$\|\mu - \nu\|_{TV} \leq \mathbb{P}[X \neq Y].$$

Useful for proving convergence and limit theorems

Coupling 102

- **Maximal Couplings:** A coupling is maximal if it corresponds to the largest probability of $\{X = Y\}$ or equivalently the smallest probability of $\{X \neq Y\}$.
- **Diagonal Couplings:** Suppose you have two random variables X and Y. In a diagonal coupling, X and Y evolve independently until they happen to take the same value. From that point onward, they remain equal. Often used to bound Markov Chains mixing times.
- **Optimal Transport:** The idea relates to finding the coupling that minimizes some cost function, often represented through a distance metric on the probability measures.

Validity of Thorisson's / Relation to rejection sampling

Conditions: Algorithm 5 is valid for coupling by rejection-sampling if:

1. $X \sim p$, $Y \sim q$
2. $P(X = Y) > 0$
3. A rejection sampling step

Algorithm 5 Modified Thorisson algorithm

function ThorissonCoupling(p, q, C)
    Sample $X \sim p$
    Sample $U \sim U(0, 1)$
    if $U < \min(\frac{q(X)}{p(X)}, C)$ **then**
        Set $Y = X$
    else
        Set $A = 0$
        while $A \neq 1$ **do**
            Sample $U \sim U(0, 1)$
            Sample $Z \sim q$
            if $U > \min(1, C\frac{p(Z)}{q(Z)})$ **then**
                Set $A = 1$
                Set $Y = Z$
            end if
        end while
    end if
    return X, Y
end function

Proof $Y \sim q$

Since in the first step $X \sim p$, let us show that $Y \sim q$. Suppose $A \subseteq \mathcal{X}$ a measurable subset. We have:

$$
\begin{aligned}
P(Y \in A) &= P(Y \in A, \textbf{step1}) + P(Y \in A, \textbf{step2}) \\
P(Y \in A, \text{step1}) &= \mathbb{E}[1(Y \in A, \text{step1})] \\
&= \int_A \int_0^1 \mathbb{1}(u < \min(\frac{q(x)}{p(x)}, C)) \cdot \mathbb{1}_{[0,1]}(u) \cdot p(x) du dx \\
&= \int_A \int_0^1 \mathbb{1}(0 \leq u < \min(\min(\frac{q(x)}{p(x)}, C), 1)) p(x) du dx \\
&= \int_A \left(\int_0^{\min(\min(\frac{q(x)}{p(x)}, C), 1)} du \right) p(x) dx \\
&= \int_A \min(q(x), Cp(x)) dx \\
\implies P(\textbf{step1}) &= \int_{\mathcal{X}} \min(q(x), Cp(x)) dx
\end{aligned}
$$

$$P(Y \in A, \mathbf{step2}) = \int_A q(x) - \min\{q(x), Cp(x)\}\, dx \qquad (1)$$

For (1) to hold it is necessary that:

$$\int_A q(x) - \min\{q(x), Cp(x)\}\, dx = P(Y \in A | \mathbf{step2}) P(\mathbf{step2})$$

We know that,

$$P(\mathbf{step2}) = 1 - P(\mathbf{step1}) = 1 - \int_x \min\{q(x), Cp(x)\}\, dx$$

Thus, given $Y|$step2 has density:

$$\tilde{q}(x) = \frac{q(x) - \min\{q(x), Cp(x)\}}{1 - \int_x \min\{q(s), Cp(s)\}\, ds}$$

It is then sufficient to note that step 2 is a standard case of acceptance-rejection, where we wish to simulate $\tilde{q}(x)$ using the proposal law $q(x)$ to conclude that $Y \sim q$.

Rewriting of step 2 of algorithm 5: Note that,

$$\frac{\tilde{q}(z)}{q(z)} = \frac{1 - \min\{1, \frac{Cp(z)}{q(z)}\}}{1 - \int_x \min\{q(s), Cp(s)\}\, ds} \leq \frac{1}{1 - \int_x \min\{q(s), Cp(s)\}\, ds} = M$$

Thus,

$$1 - \min\{1, Cp(z)\} = \frac{\tilde{q}(z)}{Mq(z)}$$

and instructions 9 to 13 can be rewritten:

- Sample $Z \sim q$
- Sample $V \sim \text{Unif}(0, 1)$ with $V = 1 - U$ (U is the uniform law from the initial version of the algorithm)
- If $V < 1 - \min\{1, \frac{Cp(z)}{q(z)}\} = \frac{\tilde{q}(z)}{Mq(z)}$ then
    - Set $A = 1$
    - Set $Y = Z$

One can thus clearly see the instructions listed in the acceptance-rejection method. Finally, it must be noted that:

$$P(X = Y) = P(\mathbf{step1}) = \int_x \min\{q(x), Cp(x)\}\, dx > 0$$

Variance behaviour - Thorisson's algorithm

To assess execution time, remark that:

- Step 1 contains 3 essential instructions: Draw from p, uniform draw, and conditional instruction on U.
- Similarly at step 2 each "while" iteration has three actions: Draw from q, uniform draw, conditional on V.

These are treated as one block, executed t times.

Remember acceptance probability at step 2:

$$P\left(V < 1 - \min\left\{1, \frac{Cp(Z)}{q(Z)}\right\}\right) = 1 - \int \min\{q(x), Cp(x)\}\, dx$$

Hence, $t \sim \text{geom}(a)$ where: $a = 1 - \int \min\{q(x), Cp(x)\}\, dx$

Therefore, the average number of blocks executed at step 2 is:

$$P(\text{step2}).E[t] = \frac{1 - \int_x \min\{q(x), Cp(x)\}\, dx}{1 - \int_x \min\{q(s), Cp(s)\}\, ds} = 1$$

- On average, two instruction blocks per run, regardless of p and q.
- Variance of Runtime W considers average of conditional variances.
- With step 1 deterministic, step 2's variance dominates:

$$\begin{aligned} W &= P(\textbf{step2}).\mathrm{Var}[t] \\ &= \left(1 - \int_x \min\{q(x), Cp(x)\}\, dx\right)\left(\frac{1-a}{a^2}\right) \\ &= \frac{1}{1-a} = \frac{\int_x \min\{q(x), Cp(x)\}\, dx}{1 - \int_x \min\{q(x), Cp(x)\}\, dx} \end{aligned}$$

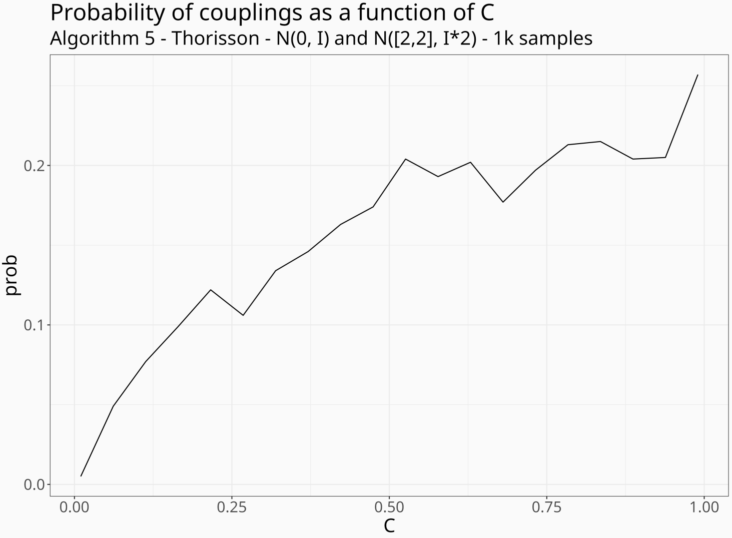
As q approaches p, variance W:

$$W = \frac{C}{1-C}$$

If $C = 1$, algorithm 5's execution time variance explodes.

Empirical behavior Thorisson

The case where C=1 is the original Thorisson algorithm

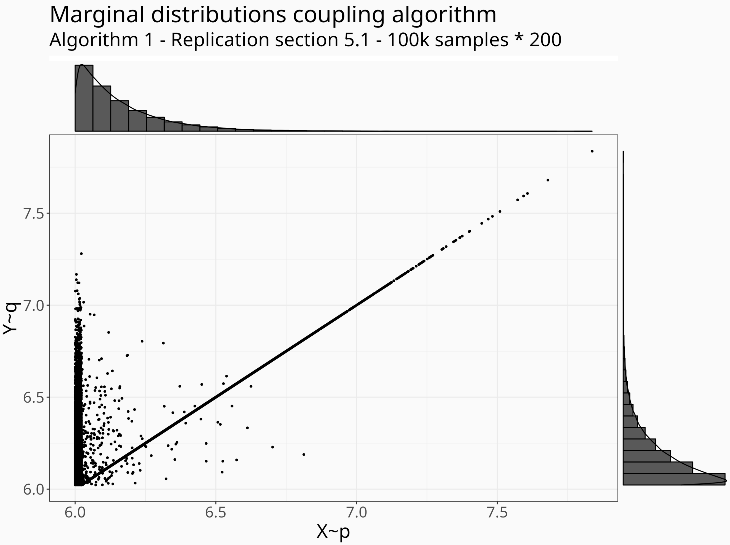


Thorisson alternatives

The authors propose the following alternative to Thorisson

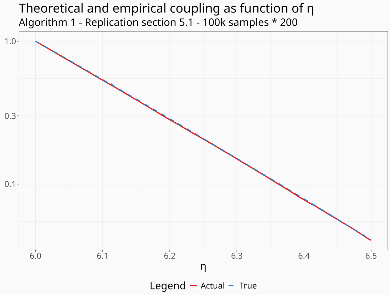
Algorithm 1 Rejection Coupling

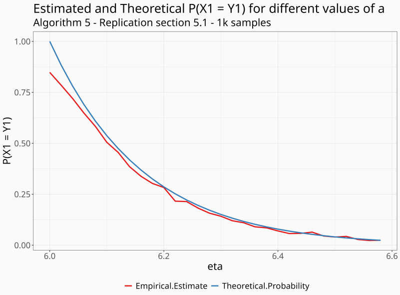
function $\textsc{RejectionCoupling}(\Gamma,\ p,\ q)$ ▷ Supposing $\Gamma \sim p \otimes q$ is a coupling of $\hat{p}$ and $\hat{q}$.
  Set $\Delta_x = 0$ and $\Delta_y = 0$ ▷ Acceptance flags
  while $\Delta_x = 0$ and $\Delta_y = 0$ **do**
    Sample $X_1, Y_1 \sim \Gamma,\ U \sim \mathcal{U}(0,1)$
    if $U < \frac{p(X_1)}{M(p,q)\hat{p}(X_1)}$ **then**
      set $\Delta_x = 1$
    end if
    if $U < \frac{q(Y_1)}{M(p,q)\hat{q}(Y_1)}$ **then**
      set $\Delta_y = 1$
    end if
  end while
  Sample X_2, Y_2 from $p \otimes q$
  return $X = \Delta_x X_1 + (1 - \Delta_x) X_2,\ Y = \Delta_y Y_1 + (1 - \Delta_y) Y_2$
end function

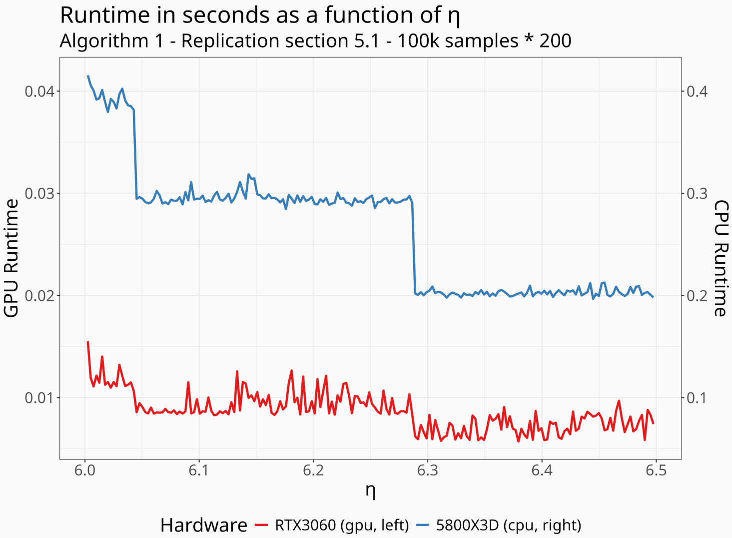

Marginal distributions coupling algorithm
Algorithm 1 - Replication section 5.1 - 100k samples * 200
Y~q
7.5
7.0
6.5
6.0
X~p
6.0
6.5
7.0
7.5

Performance comparison

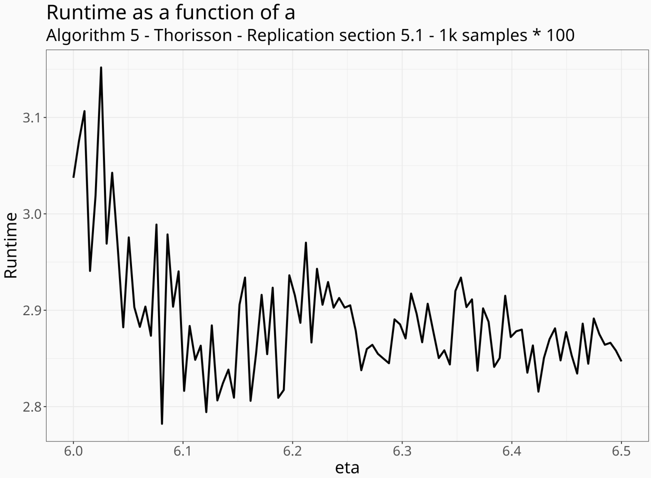
Using the same setup as section 5.1 of the paper we can compare relative performances of Thorisson and the author's approach.





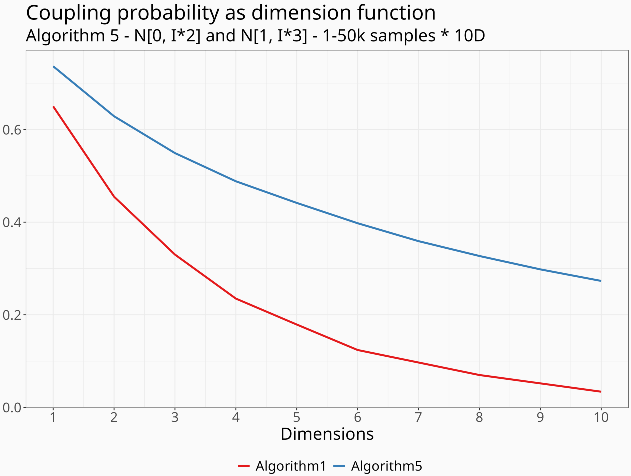

Runtime in seconds as a function of η
Algorithm 1 - Replication section 5.1 - 100k samples * 200
GPU Runtime
CPU Runtime
0.04
0.03
0.02
0.01
0.4
0.3
0.2
0.1
6.0
6.1
6.2
6.3
6.4
6.5
η
Hardware
RTX3060 (gpu, left)
5800X3D (cpu, right)

The runtime evolution follows a similar trajectory to the author's algorithm

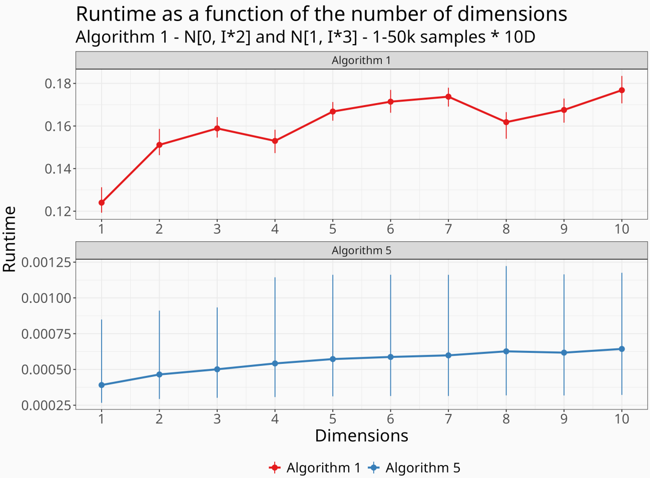


The curse of dimensionality 1/2

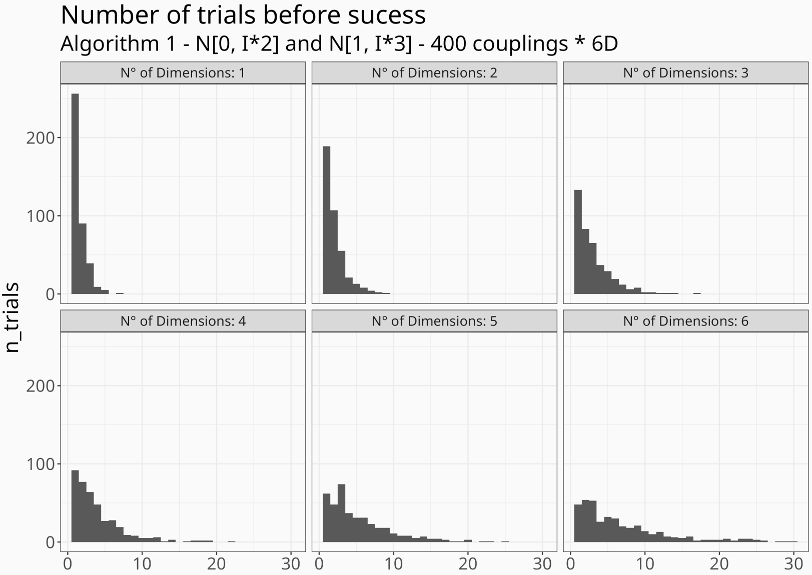
Algorithm 1 suffer much more than Thorisson from the curse of dimensionality



The curse of dimensionality 2/2



A different perspective on coupling probability



- It is possible to view the coupling probability through the lens of successful couplings (X=Y)
- Running the algorithm until a coupling and then counting the number of trials seems to follow a geometric distribution
- This can be tested through a χ^2 test

Table 1: Chi-Test Adequation and Estimated Probabilities of Success

dim	pval	p
1	0.7542079	0.6441224
2	0.8969453	0.4889976
3	0.9405786	0.3325021
4	0.6310333	0.2518892
5	0.2792477	0.1873536
6	0.6644824	0.1403016