A transformed density rejection based algorithm for densities with poles and inflection points

Josef Leydold

joint work with Wolfgang Hörmann

Institute for Statistics and Mathematics, WU Wien, Austria
Boğaziçi University, Istanbul, Turkey

14th MCM, Paris – Sorbonne, June 30, 2023

Non-Uniform Random Number Generation

Given:

- sequence $U_1, U_2, \ldots$ of "truly" IID uniform random numbers.

Task:

- transform into sequence $X_1, X_2, \ldots$ of random variates with **given distribution**

$$U_1, U_2, U_3, U_4, \ldots \quad \longrightarrow \quad X_1, X_2, X_3, \ldots$$

Most popular methods:

- Inversion method
- **Acceptance-Rejection** method

Automatic Algorithm

Idea:

One algorithm works for a large class of distributions.

- Works for non-standard distributions.
- Generators with known structural properties.
- Can be used by “non-experts” for special problems.

Required:

- PDF, CDF, ... of the distribution.
- Location of mode, “main part”, ... of the distribution.
- ...

Automatic Algorithm – Wishlist

- **Fast**.
- **Exact** (at least in $\mathbb{R}$).
- We can control the properties (accuracy, efficiency, ...) of the algorithm.
- Uses as **few** information as possible about distribution.
- User interface for a library implementation should be as **simple** as possible.

  (Complexity of algorithm should be hidden from user.)

Here: We assume that only the (log-)PDF f and its derivative are given.

Inversion Method

If $U \sim \mathcal{U}(0,1)$, then

$$X = F^{-1}(U) = \text{inf}\,\{x\colon F(x) \geq U\} \sim F$$

1. Generate $U \sim U(0,1)$.
2. Compute $X = F^{-1}(U)$. $\Leftarrow$ **Problem (?)**
3. Return X.

Inversion Method

Advantages:

- Most general method for generating non-uniform random variates.
- Get **one** random variate X for each uniform U.
- **Preserves** the structural properties of the underlying uniform PRNG.

Disadvantages:

- CDF and/or its inverse often not given in closed form.
- Numerical methods may be slow and/or require large tables.
- Numerical methods are **not exact**.
- There are issues with poles.

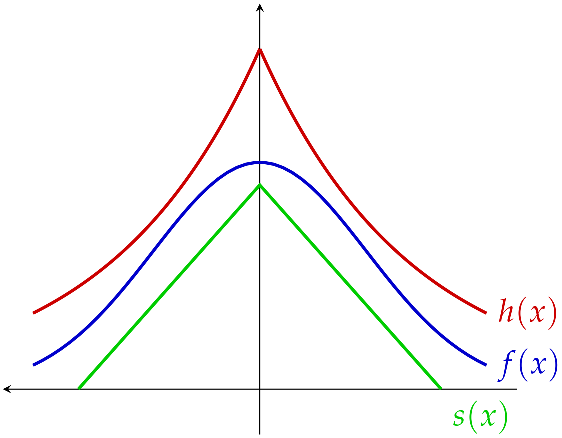
Derflinger et al. (2010) propose an inversion algorithm, when only the PDF is known.

Acceptance-Rejection Method

Need **hat** h and **squeeze** s, s.t.

$$s(x) \leq f(x) \leq h(x)$$

1. Generate $X \sim h$.
2. Generate $U \sim U(0,1)$.
3. If $U \cdot h(X) \leq s(X)$
4. Return X.
5. If $U \cdot h(X) \leq f(X)$,
6. Return X.
7. Else try again.



Acceptance-Rejection

Requirements:

- Hat function h must be a multiple of some PDF.

Properties:

- Works for unnormalized PDFs.
- Performance parameter: rejection constant.

Wishlist:

- $h \approx f$
- Sampling $X \sim h$ should be fast and simple (ideally by inversion).
- Evaluation of squeeze s should be cheap.

Acceptance-Rejection – Rejection Constant

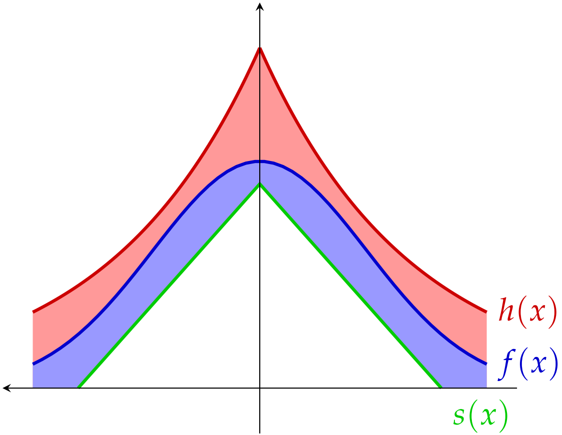
Rejection constant:

$$\alpha = \frac{\int_{\mathbb{R}} h(x)\,dx}{\int_{\mathbb{R}} f(x)\,dx}$$

Ratio hat-squeeze:

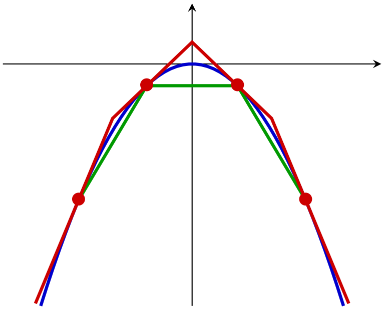
$$\rho = \frac{\int_{\mathbb{R}} h(x)\,dx}{\int_{\mathbb{R}} s(x)\,dx}$$

$$\rho \geq \alpha \geq 1$$

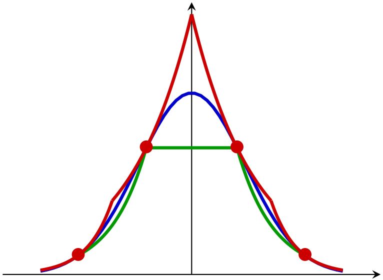


Transformed Density Rejection

Devroye (1984): tangents and secants to construct hat $h(x)$ and squeeze $s(x)$ for **log-concave** PDFs.



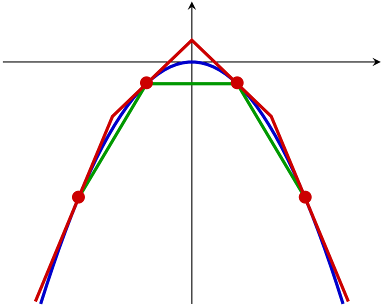
log-scale



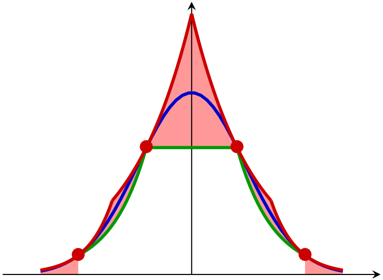
original scale

Transformed Density Rejection

Gilks and Wild (1992): adaptive rejection sampling



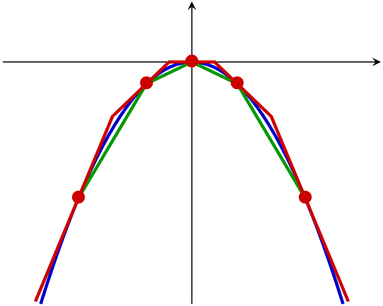
log-scale



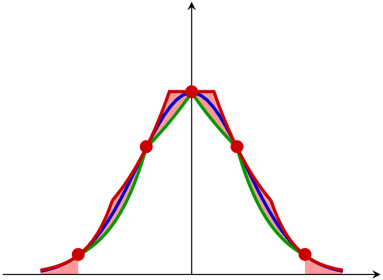
original scale

Transformed Density Rejection

Gilks and Wild (1992): adaptive rejection sampling



log-scale



original scale

$$\rho \to 1 \quad \text{for} \quad N \to \infty$$

Transformed Density Rejection – Algorithm

1. Start with an initial partition.
2. Repeat:
3. Compute $h(x)$ and $s(x)$ for each subinterval.
4. Split every subinterval I where the $\int_I (h(x) - s(x))dx$ is too large.
5. Until ρ is as small as desired.
6. Run acceptance-rejection loop.

Transformed Density Rejection – Properties

- Requires PDF f and derivative f'.
- Performance can be controlled by input parameter ρ.

For $\rho \approx 1$ we find:

- Possibly expensive setup.
- (Very) fast marginal generation time (hardly depends on f.)
- Generation from hat is $\mathcal{O}(1)$
  (by guide table or alias method.)
- Algorithm close to inversion.

Transformed Density Rejection – Properties

- Requires PDF f and derivative f'.
- Performance can be controlled by input parameter ρ.

For $\rho \approx 1$ we find:

- Possibly expensive setup.
- (Very) fast marginal generation time (hardly depends on f.)
- Generation from hat is $\mathcal{O}(1)$
  (by guide table or alias method.)
- Algorithm close to inversion.

Disadvantage: Restricted to **log-concave** distributions!

T-concave Distributions

Hörmann (1995): Generalizes to T_c-concave distribution, i.e., $T_c \circ f$ is concave.

$$T_c(x) = \begin{cases} \log(x) & \text{for } c = 0 \\ \operatorname{sgn}(c)x^c & \text{for } c \neq 0 \end{cases}$$

T-concave Distributions

Hörmann (1995): Generalizes to T_c-concave distribution, i.e., $T_c \circ f$ is concave.

$$T_c(x) = \begin{cases} \log(x) & \text{for } c = 0 \\ \operatorname{sgn}(c) x^c & \text{for } c \neq 0 \end{cases}$$

- If f is T_{c_1}-concave, then it T_{c_2}-concave for all $c_2 \leq c_1$.
- Hat and squeeze are piecewise exponential or power functions.
- In transformed scale and $c \neq 0$: hat and squeeze must not intersect x-axis.
- For unbounded intervals: $c > -1$.
- For unbounded f (pole): $c < -1$.

T-convex

Obviously the idea also works for T_c-convex distributions with the roles of tangents and secants exchanged (Evans and Swartz, 1998).

Idea: Split domain into intervals where f is either T_c-concave or T_c-convex.

Issues:

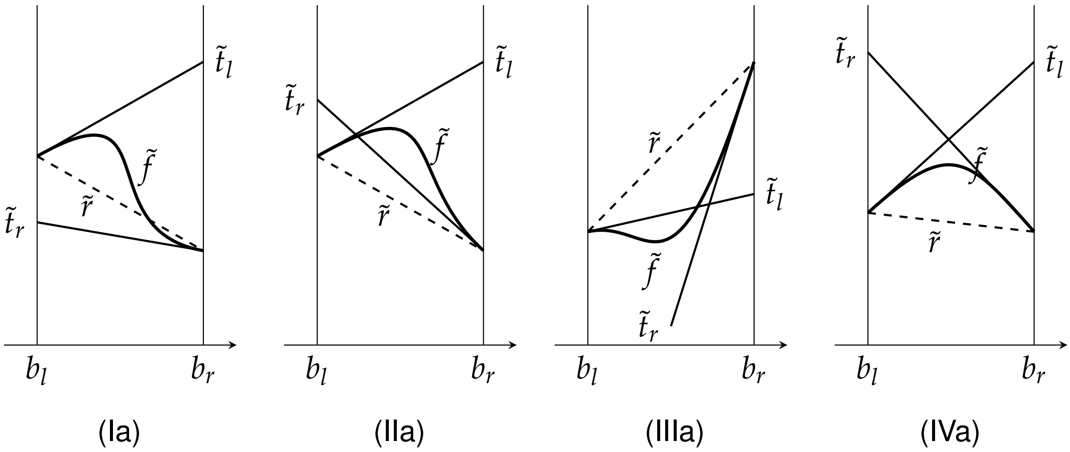
- Need f''.
- Have to compute inflection points of transformed density $T_c \circ f$.
- Is the API still simple?
- Is the algorithm still exact when we have to apply root finding algorithms?
- Can be replace the exact position of inflection points by a **rough** estimate?

A Rough Estimate

Botts, Hörmann, and L (2012):

- Suppose we have a subdivision into intervals.
- Assume that there is **at most one** inflection point of the $T_c \circ f$ in each subinterval $[b_l, b_r]$.
- Then only the four cases below are possible (plus their symmetric analogs).

Possible Cases



Possible Cases

Botts et al. (2012):

- Each interval belongs to (at most) one of these 4+4 types.
- In all cases we can use tangents and secants
  for constructing **hat** and **squeeze**.
- The type can be determined by the inequalities below.

Detect Cases I

Type	$\tilde{f}'$ and R	$\tilde{f}''$	squeeze and hat
Ia	$\tilde{f}'(b_l), \tilde{f}'(b_r) \geq R$	$\tilde{f}''(b_l) \leq 0 \leq \tilde{f}''(b_r)$	$\tilde{t}_r(x) \leq \tilde{f}(x) \leq \tilde{t}_l(x)$
IIa	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \leq 0 \leq \tilde{f}''(b_r)$	$\tilde{r}(x) \leq \tilde{f}(x) \leq \tilde{t}_l(x)$
IIIa	$\tilde{f}'(b_l) \leq R \leq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \leq 0 \leq \tilde{f}''(b_r)$	$\tilde{t}_r(x) \leq \tilde{f}(x) \leq \tilde{r}(x)$
IVa	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$	$\tilde{f}''(b_l), \tilde{f}''(b_r) \leq 0$	$\tilde{r}(x) \leq \tilde{f}(x) \leq \tilde{t}_l(x), \tilde{t}_r(x)$

Observe: We need the slope of the secant:

$$R = \frac{\tilde{f}(b_r) - \tilde{f}(b_l)}{b_r - b_l}$$

Detect Cases II

Type	$\tilde{f}'$ and R	$\tilde{f}''$	squeeze and hat
Ib	$\tilde{f}'(b_l), \tilde{f}'(b_r) \leq R$	$\tilde{f}''(b_l) \geq 0 \geq \tilde{f}''(b_r)$	$\tilde{t}_l(x) \leq \tilde{f}(x) \leq \tilde{t}_r(x)$
IIb	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \geq 0 \geq \tilde{f}''(b_r)$	$\tilde{r}(x) \leq \tilde{f}(x) \leq \tilde{t}_r(x)$
IIIb	$\tilde{f}'(b_l) \leq R \leq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \geq 0 \geq \tilde{f}''(b_r)$	$\tilde{t}_l(x) \leq \tilde{f}(x) \leq \tilde{r}(x)$
IVb	$\tilde{f}'(b_l) \leq R \leq \tilde{f}'(b_r)$	$\tilde{f}''(b_l), \tilde{f}''(b_r) \geq 0$	$\tilde{t}_l(x), \tilde{t}_r(x) \leq \tilde{f}(x) \leq \tilde{r}(x)$

Observe: **We still need f''!** Can we avoid this?

Avoid 2nd Derivative I

Some cases are unique:

Case	$\tilde{f}'$ and R	Type
(1)	$\tilde{f}'(b_l), \tilde{f}'(b_r) \geq R$	(Ia)
(2)	$\tilde{f}'(b_l), \tilde{f}'(b_r) \leq R$	(Ib)

If neither (1) nor (2) holds we need an **additional** test point $p \in (b_l, b_r)$.

Avoid 2nd Derivative II

Case	$\tilde{f}'$ and R	$\tilde{f}(p)$	Type
(3)	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$		—
(3.1)	$\tilde{f}'(p) \leq \tilde{f}'(b_r)$		(IIa)
(3.2)	$\tilde{f}'(p) \geq \tilde{f}'(b_l)$		(IIb)
(3.3)	$\tilde{f}'(b_l) \geq \tilde{f}'(p) \geq \tilde{f}'(b_r)$		—
(3.3.1)		$\tilde{f}(p) > \tilde{t}_l(p)$	(IIb)
(3.3.2)		$\tilde{f}(p) > \tilde{t}_r(p)$	(IIa)
(3.3.3)		$\tilde{f}(p) \leq \tilde{t}_l(p), \tilde{t}_r(p)$	(IIb $\mid$ IVa) + (IIa $\mid$ IVa)

Avoid 2nd Derivative III

Case	$\tilde{f}'$ and R	$\tilde{f}(p)$	Type
(4)	$\tilde{f}'(b_l) \le R \le \tilde{f}'(b_r)$		—
⋮			
(4.3)		$\tilde{f}'(b_l) \le \tilde{f}'(p) \le \tilde{f}'(b_r)$	—
⋮			
(4.3.3)		$\tilde{f}(p) \ge \tilde{t}_l(p), \tilde{t}_r(p)$	(IIIa $\mid$ IVb) + (IIIb $\mid$ IVb)

"(IIIa $\mid$ IVb) + (IIIb $\mid$ IVb)" means that we have to split $[b_l, b_r]$ into two subintervals of the respective types.

Combined Types

We now cannot decide between some case:

Type	$\tilde{f}'$ and R	$\tilde{f}''$	squeeze and hat
IIa | IVa	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \leq 0$	$\tilde{r}(x) \leq \tilde{f}(x) \leq \tilde{t}_l(x)$
IIb | IVa	$\tilde{f}'(b_l) \geq R \geq \tilde{f}'(b_r)$	$\tilde{f}''(b_r) \leq 0$	$\tilde{r}(x) \leq \tilde{f}(x) \leq \tilde{t}_r(x)$
IIIa | IVb	$\tilde{f}'(b_l) \leq R \leq \tilde{f}'(b_r)$	$\tilde{f}''(b_r) \geq 0$	$\tilde{t}_r(x) \leq \tilde{f}(x) \leq \tilde{r}(x)$
IIIb | IVb	$\tilde{f}'(b_l) \leq R \leq \tilde{f}'(b_r)$	$\tilde{f}''(b_l) \geq 0$	$\tilde{t}_l(x) \leq \tilde{f}(x) \leq \tilde{r}(x)$

Nevertheless we still can construct hat and squeeze.

Sign of 2nd Derivative

- Once the type of the interval is known, we can use the above tables to determine the **sign** of $f''(b_l)$ and/or $f''(b_r)$.
- Then for any two points $c, c+\delta$ in (b_l, b_r),
  the sign can be determined for at least one of $f''(c)$ and $f''(c+\delta)$.
- This allows to determine the type when an interval is splitted at c or $c+\delta$ during the setup.
- In particular we thus get rid of the combined types.

Poles

Let p be a pole of f.
Then the secant of $T_c \circ f$ on a small interval $[p, b_r]$ can be used for constructing a hat function whenever

$$-1 > c > \limsup_{x \to p} \mathrm{lc}_f(x)$$

where

$$\mathrm{lc}_f(x) = 1 - \frac{f''(x)\, f(x)}{f'(x)^2}$$

(local concavity).

Work in progress. We test whether the method from Hörmann, L, and Derflinger (2007) can be used to find an appropriate c.

Examples

Generalized Hyperbolic distribution has PDF

$$f(x) = e^{\beta(x-\mu)} \frac{\mathrm{K}_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}},$$

where $\mathrm{K}_\nu(.)$ denotes the modified Bessel function of the third kind.

For $c = -1/2$,
$T_{-1/2} \circ f$ may have a single convex interval on one or both sides of the mode.

Examples

Watson distribution has PDF

$$f(\mathbf{x}) \propto \exp\left(\kappa \boldsymbol{\mu}' \mathbf{x}\right) \qquad \text{on} \quad S^d = \{\mathbf{x} \colon \|\mathbf{x}\|_2 = 1\}$$

It can be decomposed as $\mathbf{X} = \left(\sqrt{1 - W^2}\mathbf{Y}, W\right)$, where $\mathbf{Y}$ is uniformly distributed on the hypersphere orthogonal to $\boldsymbol{\mu}$ and W has log-density

$$g(w) = \kappa w^2 + \frac{d-3}{2} \log\left(1 - w^2\right)$$

on domain $[0, 1]$.
It can be easily shown that $\log \circ g$ has at most one inflection point.

Examples

... and of course it works for all truncated distributions
whenever it works on its entire domain.

Tinflex

The method is implemented in **R** package `Tinflex`:

`https://CRAN.R-project.org/package=Tinflex`

Conclusion

- We have created an automatic algorithm based on the acceptance-rejection method for quite general continuous univariate distributions.
- Requirements: (log-)PDF, its derivative, a partition of the domain s.t. each subinterval contains at most one inflection point of the transformed density.
- We avoid the computation of inflection points and f''.
- Simpler interface for user ...
- ... but higher complexity for the implementation (mostly look-ups in tables of satisfied inequalities).

Thank You

for your attention!

References I

C. Botts, W. Hörmann, and J. L. Transformed density rejection with inflection points. *Statistics and Computing*, 23(2):251–260, 2012. doi: 10.1007/s11222-011-9306-4.

G. Derflinger, W. Hörmann, and J. L. Random variate generation by numerical inversion when only the density is known. *ACM Trans. Model. Comput. Simul.*, 20(4):18:1–18:25, 2010. doi: 10.1145/1842722.1842723.

L. Devroye. A simple algorithm for generating random variates with a log-concave density. *Computing*, 33(3–4):247–257, 1984.

M. Evans and T. Swartz. Random variable generation using concavity properties of transformed densities. *Journal of Computational and Graphical Statistics*, 7(4):514–528, 1998.

W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2): 337–348, 1992.

W. Hörmann. A rejection technique for sampling from T-concave distributions. *ACM Trans. Math. Soft.*, 21(2):182–193, 1995. doi: http://doi.acm.org/10.1145/203082.203089.

References II

W. Hörmann and J. L. A generalized transformed density rejection algorithm. In Z. Botev, A. Keller, C. Lemieux, and B. Tuffin, editors, *Advances in Modeling and Simulation: Festschrift for Pierre L'Ecuyer*, pages 283–300. Springer International Publishing, 2022. doi: 10.1007/978-3-031-10193-9_14. URL `https://doi.org/10.1007/978-3-031-10193-9_14`.

W. Hörmann, J. L, and G. Derflinger. Inverse transformed density rejection for unbounded monotone densities. *ACM Trans. Model. Comput. Simul.*, 17(4):18, 2007. doi: http://doi.acm.org/10.1145/1276927.1276931.