Derivative Based Adaptive Rejection Sampling (a review)

Erwann Rogard

2022-04-06

Abstract

This review was first published on 2007-11-19 as companion to a Mathematica package[1]. The present version is just a revamp[2].

The purpose of the algorithm is to sample from a univariate log–concave (target) density, known up to a constant of proportionality, as commonly arising in Gibbs sampling applications[1][2]. The algorithm is derived from rejection sampling but instead of being pre–specified, the covering (proposal) density is constructed adaptively using only evaluations of the un–normalized density and its derivative. Geometrically, the log of the former is the upper envelope of the convex hull delimited by the tangents to the log un–normalized target at the evaluation points. This modification to the original algorithm requires a new rejection rule which depends on the distances between the true density and the upper hull, evaluated at the candidate draw, and that of the latter to a lower hull, formed by joining the evaluation points. Consider the definitions:

[1] `https://library.wolfram.com/infocenter/ID/7071/`

[2] Revisions:`https://github.com/erwannr/statistics/commits/main/deriv_ars`

Definitions		
D		domain (connected)
$D_\pm$		lower/upper bound
$f(.)$		log–concave density
$g(.)$	$f(.) \propto \mathrm{g}(.)$	un–normalized density
$h(.)$	$\log g(.)$	
$h'(.)$	$h'(x) = \partial h(.)/\partial x$	
x_k	$x_k \in D$	kth abscicae
$x_{(k)}$	$x_{(k)} \leq x_{(k+1)}$	kth ordered abscicae
$t^{\mathrm{u}}_k(.)$		tangent to $h(.)$ at $x_{(k)}$, $k = 1, ...K$
$t^{\mathrm{l}}_k(.)$		line through $\{(x_{(j)}, h(x_{(j)})) : j = k, k+1\}$, $k = 1, ..., K-1$
X	$\{x_k : k = 1, ..., K\}$	set of $K \geq 2$ points such that $h'(x_{(1)}) > 0$ and $h'(x_{(K)}) < 0$, if $D_- = -\infty$ and , $D_+ = +\infty$, respectively
$z_{(k)}$	$t^{\mathrm{u}}_k(z_{(k)}) = t^{\mathrm{u}}_{k+1}(z_{(k)})$	tangent intersection abscicae, $k = 1, ..., K-1$
$z_{(0)}$	D_-	
$z_{(K)}$	D_+	
$u(.)$		upper hull envelope of $h(.)$
$s(.)$	$s(x) \propto \exp(u(x))$	proposal density
c	$\int_D \exp(u(x))dx$	normalizing constant
$S(.)$	$\int_{z_{(0)}}^{(.)} s(x)dx$	cumulative density
$l(.)$		lower hull envelope

In particular,

$$t_k^{\mathrm{u}}(x) = h(x_{(k)}) + (x - x_{(k)})h'(x_{(k)}) \tag{1}$$

$$t_k^{\mathrm{l}}(x) = h(x_{(k)}) + (x - x_{(k)})\frac{h(x_{(k+1)}) - h(x_{(k)})}{x_{(k+1)} - x_{(k)}} \tag{2}$$

$$z_{(k)} = -(t_{k+1}^{\mathrm{u}}(0) - t_k^{\mathrm{u}}(0))/(h'(x_{(k+1)}) - h'(x_{(k)})) \tag{3}$$

$$u(x) = \sum_{k=1}^{K} 1\{z_{(k-1)} \leq x < z_{(k)}\}t_k^{\mathrm{u}}(x) \tag{4}$$

$$S_{\mathrm{un}}^u(x) = \sum_{k=1}^{K} \int_{x\wedge z_{(k-1)}}^{x\wedge z_{(k)}} \exp(t_k^{\mathrm{u}}(y))dy \tag{5}$$

$$= \sum_{k=1}^{K} \frac{1}{h'(x_{(k)})} \left(\exp(t_k^{\mathrm{u}}(x \wedge z_{(k)})) - \exp(t_k^{\mathrm{u}}(x \wedge z_{(k-1)}))\right) \tag{6}$$

$$S_u^{-1}(u) = \inf\{x : S_u^{\mathrm{un}}(x) \geq uc\} \tag{7}$$

$$= x_{(k_*)} + \frac{1}{h'(x_{(k_*)})} \left(\log\left(e^{t_{k_*}^{\mathrm{u}}(z_{(k_*-1)})} + h'(x_{(k_*)})(cu - S_u^{\mathrm{un}}(z_{(k_*-1)}))\right) - h(x_{(k_*)})\right) \tag{8}$$

$$l(x) = \sum_{k=1}^{K-1} 1\{x_{(k)} \leq x < x_{(k+1)}\}t_k^{\mathrm{l}}(x) - \infty(1 - 1\{x_{(1)} \leq x < x_{(K)}\}) \tag{9}$$

where $k_* = \min\{k : S^u(z_{(k)}) \geq u\}$.

The exponential terms in the un–normalized cumulative density are suceptible to overflows. To avoid this we do

$$h(.) \leftarrow h(.) - \max_k h(z_{(k)}) \tag{10}$$

which is permissible because $h(.)$ is the log of an un–normalized version of $f(.)$.

The algorithm below, initialized with some suitable X, generates one draw from f:

Algorithm 1 (ARS).

1. $\omega_l \overset{\mathrm{iid}}{\sim} \mathrm{Unif}(0,1), l \in \{0,1,2\}$; $x^* \leftarrow S_u^{-1}(\omega_0)$
2. if$(\omega_1 \leq \exp(l(x^*) - u(x^*)))\{$ return x^* $\}$else$\{$
   if$(\omega_2 \leq \exp(h(x^*) - u(x^*))$ $\{$
   (a) $X \leftarrow (X, x^*)$, $K \leftarrow K + 1$,

(b) update derived quantities;

(c) return x^*

} else{ goto 1.}
}

Bibliography

[1] Walter R. Gilks and Pascal Wild. "Adaptive Rejection Sampling for Gibbs Sampling". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 41.2 (1992), pp. 337–348. URL: `http://www.jstor.org/stable/2347565`.

[2] Wild. Pascal and Walter R. Gilks. "Algorithm AS 287: Adaptive Rejection Sampling from Log-Concave Density Functions". In: *Applied Statistics* 42.4 (1993), pp. 701–709.