Adaptive Rejection Sampling

- Literature
- Derivative based Adaptive Rejection Sampling (Tangent Method)
- Derivative free Adaptive Rejection Sampling
- Efficiency Comparison
- Other Methods
- Adaptive Rejection Sampling in C
- Adaptive Rejection Sampling in Fortran
- BUGS

Literature

1. W. R. Gilks (1992), "Derivative-free Adaptive Rejection Sampling for Gibbs Sampling," Bayesian Statistics 4, (eds. Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M.) Oxford University Press, 641-649.

2. W. R. Gilks, P. Wild (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, Vol. 41, Issue 2, 337-348.

3. P. Wild, W. R. Gilks (1993), "Algorithm AS 287: Adaptive Rejection Sampling from Log-Concave Density," *Applied Statistics*, Vol. 42, Issue 4, 701-709.

Benefits of Adaptive Rejection Sampling

ARS reduces the number of evaluations of $g(x)$ in two ways:

1. Through the assumption of log-concavity of $f(x)$, locating the supremum of $g(x)$ is avoided ($g(x) = cf(x)$).

2. After each rejection step, the probability of needing to evaluate $g(x)$ further is reduced by using the recently acquired information about $g(x)$ to update the envelope and squeezing functions.

ARS is useful in Gibbs Sampling, where full-conditionals can be algebraically messy, but often log-concave. When distributions are not log-concave, ARS can be followed by a single step of the Metropolis-Hastings algorithm (ARMS).

Derivative based ARS

Preliminaries:

- Assume that $g(x)$ is continuous and differentiable everywhere in domain D.
- Let $h(x) = \ln g(x)$, such that $h(x)$ is concave everywhere in D ($h''(x)$ must be negative semi-definite everywhere within its support).

Initialization Step

- Let $T_k = \{x_i; i = 1, \ldots, k\}$ be the k starting points.
- If D is unbounded on the left, choose x_1 such that $h'(x_1) > 0$ and likewise if D is unbounded on the right, choose x_k such that $h'(x_k) < 0$.
- Calculate the following functions with the k starting points:
  1. $u_k(x)$, the piece-wise linear upper bound formed from the tangents to $h(x)$ at each point in T_k
  2. $s_k(x) = \frac{\exp u_k(x)}{\int_D \exp u_h(x')dx'}$
  3. $l_k(x)$, the piece-wise linear lower bound formed from the chords between adjacent points in T_k

The tangent lines at x_j and x_{j+1} intersect at the point

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}$$

for $j = 1, \ldots, k-1$ and where z_0 is the lower bound of D and z_k is the upper bound of D ($-\infty$ or $+\infty$ if D is not bounded).
The piece-wise upper bound is defined as

$$u_k(x) = h(x_j) + (x - x_j)h'(x_j)$$

for $x \in [z_{j-1}, z_j]$ and $j = 1, \ldots, k$. Similarly, the piece-wise linear lower bound is

$$l_k(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}$$

for $x \in [x_j, x_{j+1}]$ and $j = 1, \ldots, k-1$ and if $x < x_1$ or $x > x_k$, $l_k(x)$ is set equal to $-\infty$.

Sampling Step

- Sample a value x^* from $s_k(x)$ and a value $u*$ independently from a $U(0,1)$.
- **Squeezing Test**

  If $u^* \leq \exp\{l_k(x^*) - u_k(x^*)\}$ then accept x^*,
  otherwise evaluate $h(x^*)$ and $h'(x^*)$.
- **Rejection Test**

  If $u^* \leq \exp\{h(x^*) - u_k(x^*)\}$ then accept x^*,
  otherwise reject x^*.

Updating Step

- If $h(x^*)$ and $h'(x^*)$ were evaluated in the sampling step, include x^* in T_k to form T_{k+1}
- Relabel the elements of T_{k+1} in ascending order and reconstruct functions $u_{k+1}(x)$, $s_{k+1}(x)$, and $l_{k+1}(x)$.

Efficiency of ARS

- ARS tends to space the evaluations of $h(x)$ and $h'(x)$ optimally, because new evaluations are most likely to occur at values of x where the rejection envelope and squeezing functions are most discrepant.

  If k evaluations of $h(x)$ and $h'(x)$ have been performed and the current x^* has not been accepted, then the $P(x^* = x) \propto \exp u_k(x) - \exp l_k(x)$.

- Two starting points (T_2) were found to be sufficient for computational efficiency.

- Empirically, the number of evaluations of $h(x)$ and $h'(x)$ required to sample n points from $f(x)$ increases approximately in proportion to $n^{\frac{1}{3}}$, even for very non-normal densities.

Derivative Free ARS

- Similar to the derivative based ARS, except the envelope function is constructed in such a way that evaluation of derivatives is not required but, the density must still be log-concave.
- In addition, it does not require the existence of continuous derivatives.

Checking for Log-Concavity

Let $P = (x, \ln f(x))$ be a point on the logarithm of the density $f(x)$, $f(x)$ is log-concave with respect to x if:

$$\text{gradient of chord } P_2P_3\text{-gradient of chord } P_1P_2 \leq 0$$

for any $x_1 < x_2 < x_3$ within its domain.

Initialization Step

- Evaluate $\ln f(x)$ (up to an additive normalizing constant) at three or more points, such that at least one point lies to each side of the mode of the density.
- Points are on either side of the mode if:
  1. Chord joining two left most points has a positive gradient.
  2. Chord joining two right most points has a negative gradient.
- The lower bound to the log density is constructed from the chords joining the evaluated points of $f(x)$ with the vertical lines at the extreme points.
- The upper bound is constructed by extending the chords to their points of intersection.

Sampling Step

- Exponentiate the piece-wise linear upper bound of $\ln f(x)$ to create $e(x)$, the envelope function.
- Sample a point X from the envelope function, $e(x)$.
- Sample a point U independently from a $U(0, 1)$.

Squeezing Step

- Exponentiate the piece-wise linear lower bound of $\ln f(x)$ to create $s(x)$, the squeezing function.
- If $U \leq \frac{s(X)}{e(X)}$ then X is accepted as a required sample from $f(x)$.
- Otherwise $f(x)$ is evaluated and the rejection step is performed.

Rejection Step

- If $U \leq \frac{f(X)}{e(X)}$ then X is accepted as the required sample from $f(x)$.
- Otherwise X is rejected and the updating step is performed.

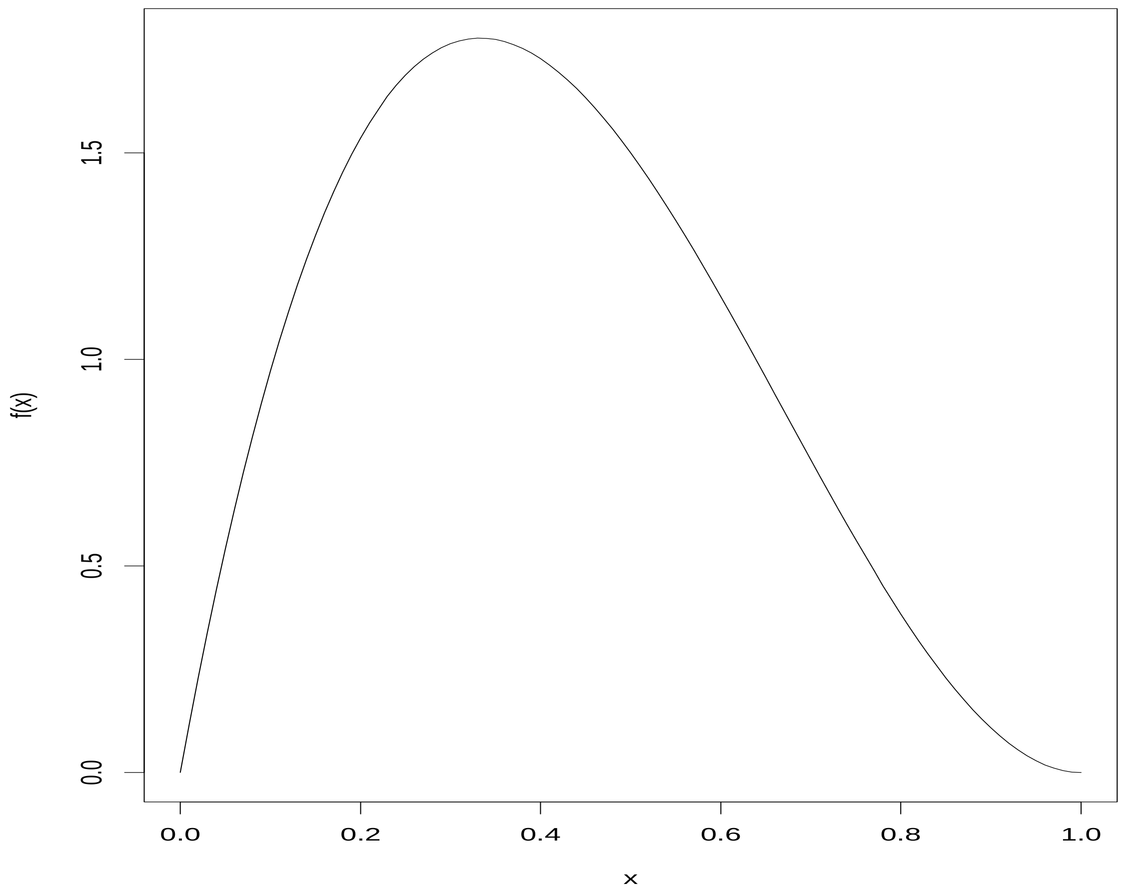
Updating Step

- Upper and lower bounds of $\ln f(x)$ are updated to make use of the information gained in the rejection step.
  1. Include the rejected point X into the set of points P.
  2. Construct new chords which include the new point, P_4.
  3. Reconstruct the piece-wise exponential envelope and squeezing functions, which will be improved due to the addition of the new point.

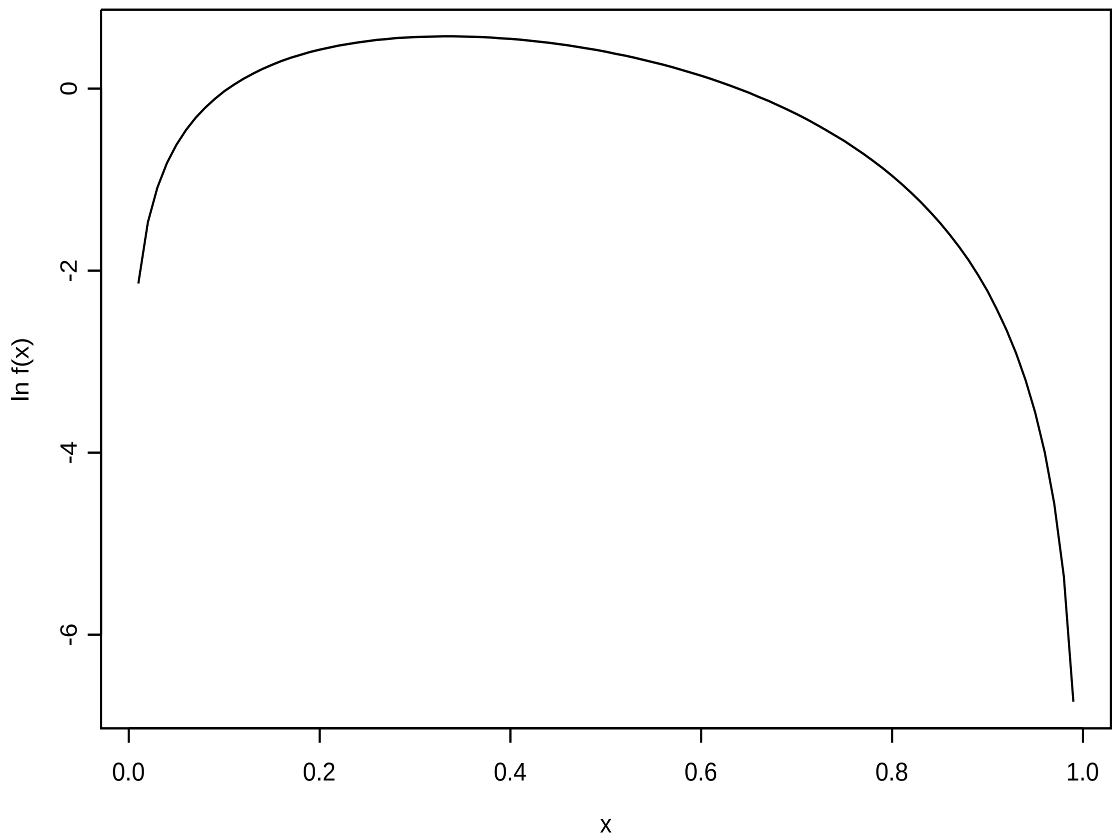
Benefits of ARS

- The probability of acceptance increases at subsequent squeezing and rejection steps.
- The envelope and squeezing functions rapidly adapt to the shape of the density, $f(x)$.

Example: Beta(2,3)



Log Density of a Beta(2,3)



Initialization Step

- Three points were chosen as the starting points and $\ln f(x)$ was evaluated at each point.

  $P_1 = (0.2, 0.4292)$

  $P_2 = (0.4, 0.567)$

  $P_3 = (0.7, -0.28)$

- The log-concavity of $f(x)$ was checked with respect to x.

  $-2.7556 - 0.5889 \leq 0$

Initialization Step

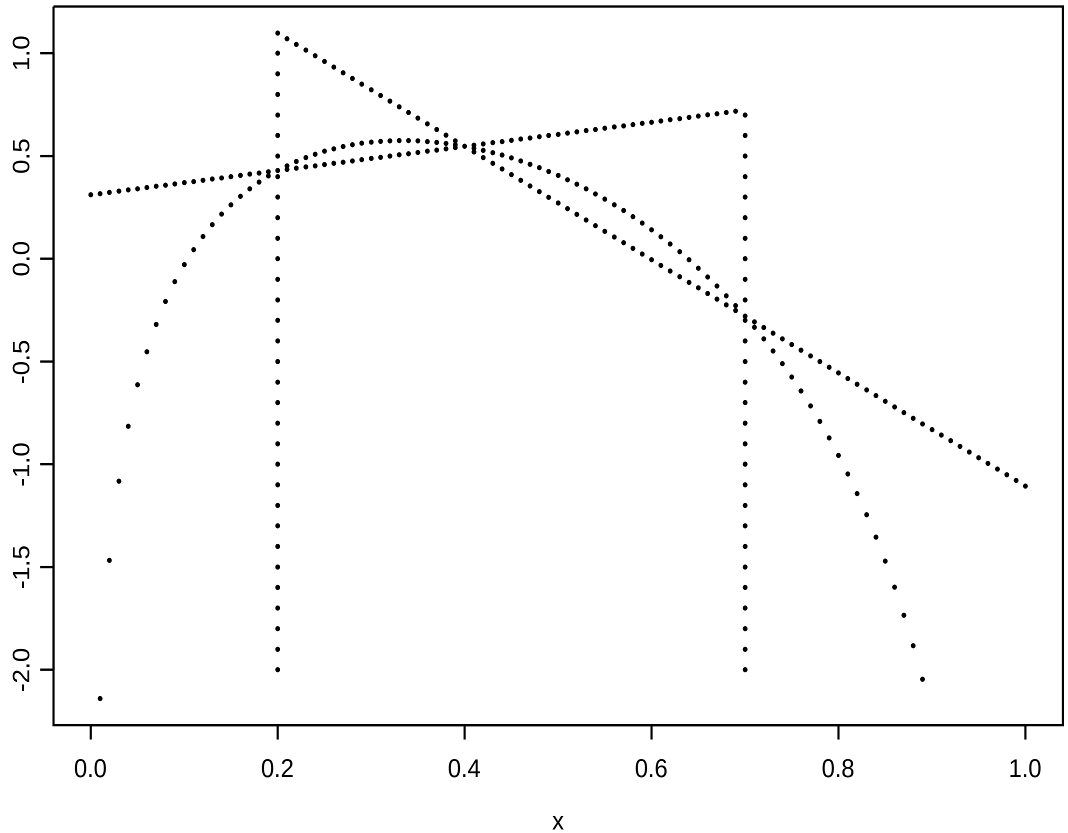
- Linear equations for the chords of P_1P_2 and P_2P_3.

  P_1P_2: $0.5889x + 0.3114$

  P_2P_3: $-2.7556x + 1.6492$

- Vertical lines are constructed at the extreme points, P_1 and P_3.
- The upper bound is created by extending the chords to their points of intersection with the vertical lines at 0.2 and 0.7.

Upper and Lower Bounds



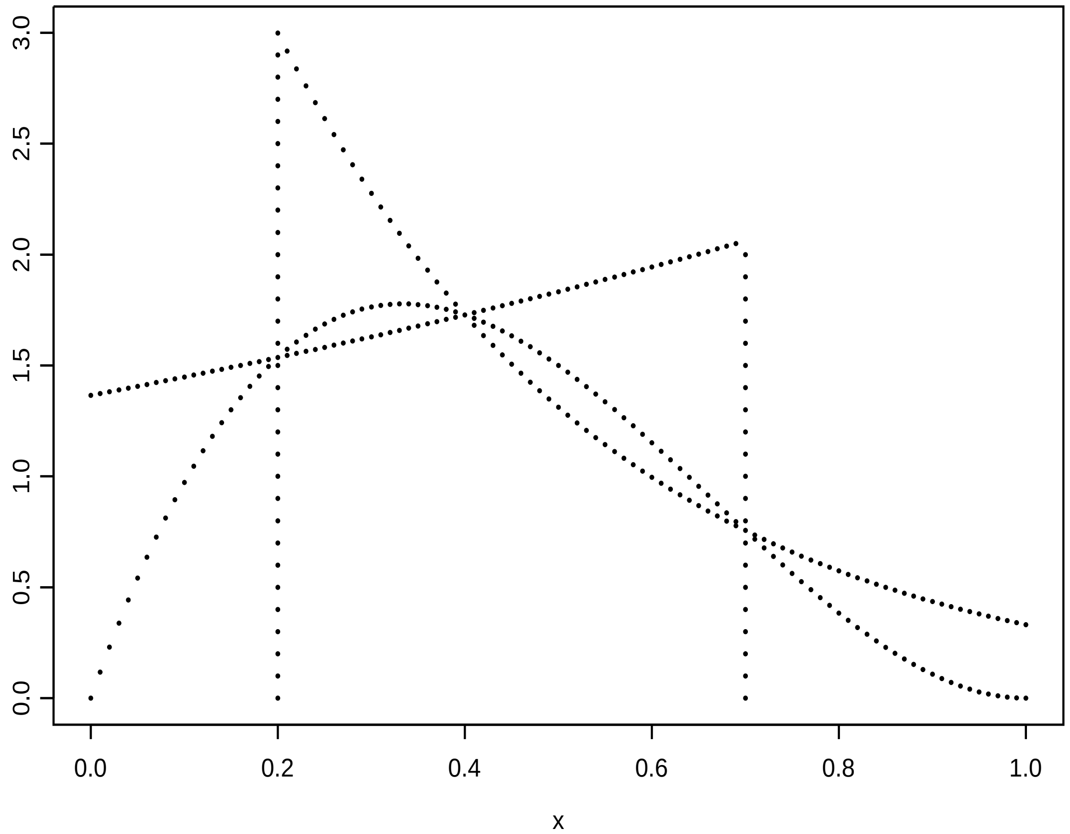
Sampling Step

- The piece-wise linear upper and lower bounds were exponentiated.

$$e(x) = \begin{cases} \exp\{0.5889x + 0.3114\} & 0 < x \leq 0.2,\ 0.4 < x \leq 0.7 \\ \exp\{-2.7556x + 1.6492\} & 0.2 < x \leq 0.4,\ 0.7 < x \leq 1 \end{cases}$$

$$s(x) = \begin{cases} \exp\{0.5889x + 0.3114\} & 0.2 < x \leq 0.4 \\ \exp\{-2.7556x + 1.6492\} & 0.4 < x \leq 0.7 \end{cases}$$

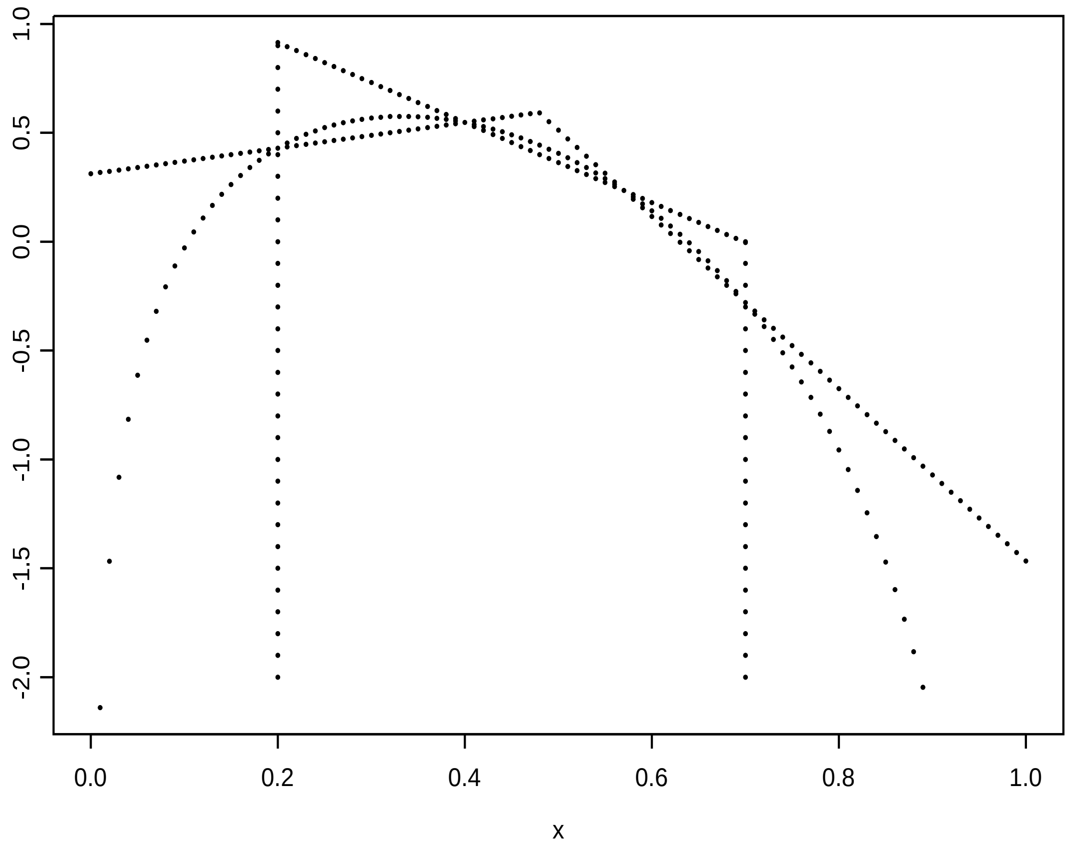
Envelope and Squeezing Functions



Updating Step

- Samples were drawn from $e(x)$ (0.5699) and $U(0, 1)$ (0.7159).
- Both the squeezing and rejection steps failed, so the updating step was performed.
- The rejected point was included in the set P and it was relabeled in ascending order.
- Chords were constructed to include the new point and the piece-wise linear upper and lower bounds were recalculated.

Updating Step



Efficiency Comparison

Tangent Method of ARS

- On average only three evaluations of the log density are required to accept one point, provided good initial points are chosen.

Derivative free Method of ARS

- If four points are used as initial values, then on average about five density evaluations are required.

In general, moderate changes in the scale or location of the initial values do not have a substantial effect on the number of function evaluations required.

Efficiency Comparison

Depending on the application, the additional costs of calculating the derivative of the log density may make the tangent method less efficient than the derivative free methods of ARS.

Other Methods

The derivative based and derivative free approaches of ARS by Gilks and Wild (1992) and Gilks (1992) are only two of the many approaches of ARS. Two other approaches are presented in:

1. Christian P. Robert, George Casella
   Monte Carlo Statistical Methods, Springer, 57-59.

2. Michael Evans, Tim Swartz,
   Approximating Integrals via Monte Carlo and Deterministic Methods, Oxford University Press, 41-47.

C Example

Example comes from the MRC Biostatistics Unit - BUGS began here as statistical research project, and is the correspondence address for W. R. Gilks.

`www.mrc-bsu.cam.ac.uk/pub/methodology/adaptive_rejection/`

ARS on a $N(10, 5)$

```
gcc -o arms.c
gcc -O2 -o example1.exe arms.o example1.c -lm
```

`arms.h, arms.c, example1.c, example2.c`

C Example

Starting values: (0.0, 3.0, 17.0, 20.0)

Number of iterations:	100	1000
Mean number of evaluations per iteration:	5	5
% with more than six evaluations per iteration:	7%	4.1%

Fortran Programs

Algorithm AS 287 provides the fortran subroutine to implement ARS and can be found at `www.statlib.org`.

Sampling within BUGS

ARS is the principle sampling methodology used in the BUGS software, with the preference order of sampling given by:

1. Standard Density
2. Adaptive Rejection Sampling
3. Ratio of Uniforms
4. Inversion

Literature/References

1. W. R. Gilks (1992), "Derivative-free Adaptive Rejection Sampling for Gibbs Sampling," Bayesian Statistics 4, (eds. Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M.) Oxford University Press, 641-649.
2. W. R. Gilks, P. Wild (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, Vol. 41, Issue 2, 337-348.
3. A. Thomas, D. J. Spiegelhalter, W. R. Gilks (1992), "BUGS: a Program to Perform Bayesian Inference using Gibbs Sampling," Bayesian Statistics 4, (eds. Bernardo, J., Berger, J., Dawid, A. P., and Smith, A. F. M.) Oxford University Press, 837-842.
4. P. Wild, W. R. Gilks (1993), "Algorithm AS 287: Adaptive Rejection Sampling from Log-Concave Density," *Applied Statistics*, Vol. 42, Issue 4, 701-709.