CPSC 535
Standard Sampling Methods

AD

6th February 2007

- Classical “exact” simulation methods.

- Classical “exact” simulation methods.
- Accept/Reject.

- Classical “exact” simulation methods.
- Accept/Reject.
- Variations over the Accept/Reject algorithm

The Monte Carlo principle

- Let $\pi(x)$ be a probability density on $\mathcal{X}$

The Monte Carlo principle

- Let $\pi(x)$ be a probability density on $\mathcal{X}$
- Monte Carlo approximation is given by

$$\widehat{\pi}_N(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^{(i)}}(x) \text{ where } X^{(i)} \overset{\text{i.i.d.}}{\sim} \pi.$$

The Monte Carlo principle

- Let $\pi(x)$ be a probability density on $\mathcal{X}$
- Monte Carlo approximation is given by

$$\widehat{\pi}_N(x) = \frac{1}{N}\sum_{i=1}^{N}\delta_{X^{(i)}}(x) \text{ where } X^{(i)} \overset{\text{i.i.d.}}{\sim} \pi.$$

- For any $\varphi : \mathcal{X} \to \mathbb{R}$

$$\mathbb{E}_{\widehat{\pi}_N}(\varphi) = \frac{1}{N}\sum_{i=1}^{N}\varphi\left(X^{(i)}\right) \approx \mathbb{E}_{\pi}(\varphi)$$

and more precisely

$$\mathbb{E}_{\{X^{(i)}\}}\left[\mathbb{E}_{\widehat{\pi}_N}(\varphi)\right] = \mathbb{E}_{\pi}(\varphi) \text{ and } var_{\{X^{(i)}\}}\left(\mathbb{E}_{\widehat{\pi}_N}(\varphi)\right) = \frac{var_{\pi}(\varphi)}{N}.$$

- If we could sample from any distribution π easily, then everything would be easy.

- If we could sample from any distribution π easily, then everything would be easy.
- Unfortunately, there is no generic algorithm to sample exactly from any π.

- If we could sample from any distribution π easily, then everything would be easy.
- Unfortunately, there is no generic algorithm to sample exactly from any π.
- Today, we discuss simple methods which are the building blocks of more complex algorithms; i.e. MCMC and SMC.

Pseudo Random Number Generators

- All algorithms discussed here rely on the availability of a generator of independent uniform random variables in $[0, 1]$.

Pseudo Random Number Generators

- All algorithms discussed here rely on the availability of a generator of independent uniform random variables in $[0, 1]$.
- It is impossible to get such numbers and we only get pseudo-random numbers which look like they are i.i.d. $\mathcal{U}[0, 1]$.

Pseudo Random Number Generators

- All algorithms discussed here rely on the availability of a generator of independent uniform random variables in $[0, 1]$.
- It is impossible to get such numbers and we only get pseudo-random numbers which look like they are i.i.d. $\mathcal{U}[0, 1]$.
- There are a few standard very good generators available. We will not give any detail as their constructions are based on techniques very different from the ones we address here.

Inverse CDF Method

- Consider $\mathcal{X} = \{1, 2, 3\}$ and

$$\pi(X = 1) = \frac{1}{6},\ \pi(X = 2) = \frac{2}{6},\ \pi(X = 3) = \frac{1}{2}.$$

Inverse CDF Method

- Consider $\mathcal{X} = \{1, 2, 3\}$ and

$$\pi(X = 1) = \frac{1}{6},\ \pi(X = 2) = \frac{2}{6},\ \pi(X = 3) = \frac{1}{2}.$$

- Define the cdf of X for $x \in [0, 3]$ as

$$F_X(x) = \sum_{i=1}^{3} \pi(X = i)\, \mathbb{I}(i \leq x)$$

and its inverse for $u \in [0, 1]$

$$F_X^{-1}(u) = \inf\{x \in \mathcal{X} : F_X(x) \geq u\}$$

- To sample from this discrete distribution, sample $U \sim \mathcal{U}[0,1]$.

- To sample from this discrete distribution, sample $U \sim \mathcal{U}[0,1]$.
- Find $X = F_X^{-1}(U)$.

- To sample from this discrete distribution, sample $U \sim \mathcal{U}[0,1]$.
- Find $X = F_X^{-1}(U)$.
- The probability of U falling in the vertical interval i is precisely equal to the probability $\pi(X = i)$.

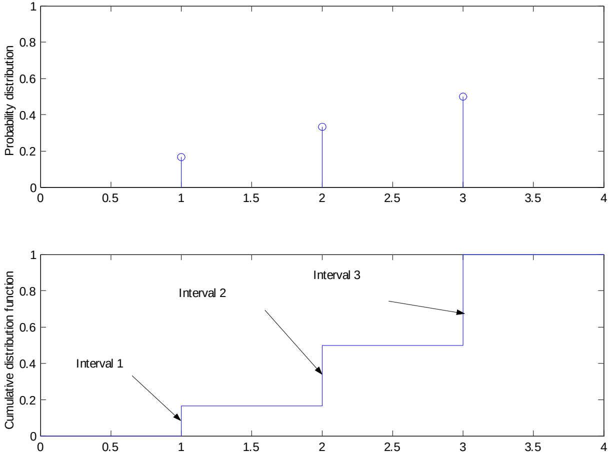


Figure: The distribution and cdf of a discrete random variable

- Assume the distribution has a density, then the cdf takes the form

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{+\infty} \pi(u)\, I(u \leq x) du = \int_{-\infty}^{x} \pi(u)\, du.$$

- Assume the distribution has a density, then the cdf takes the form

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{+\infty} \pi(u)\, I(u \leq x) du = \int_{-\infty}^{x} \pi(u)\, du.$$

- We would like to use the same algorithm; i.e. $U \sim \mathcal{U}[0,1]$ and set $X = F_X^{-1}(U)$.

- Assume the distribution has a density, then the cdf takes the form

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{+\infty} \pi(u)\, I(u \leq x) du = \int_{-\infty}^{x} \pi(u)\, du.$$

- We would like to use the same algorithm; i.e. $U \sim \mathcal{U}[0,1]$ and set $X = F_X^{-1}(U)$.
- **Question**: Do we have $X \sim \pi$?

- Proof of validity:

$$\begin{aligned}
\Pr\left(X \leq x\right) &= \Pr\left(F_X^{-1}\left(U\right) \leq x\right) \\
&= \Pr\left(U \leq F_X\left(x\right)\right) \text{ since } F_X \text{ is non decreasing} \\
&= \int_0^1 \mathbb{I}\left(u \leq F_X\left(x\right)\right) du \text{ since } U \sim \mathcal{U}\left[0,1\right] \\
&= F_X\left(x\right)
\end{aligned}$$

- Proof of validity:

$$
\begin{aligned}
\Pr(X \leq x) &= \Pr\left(F_X^{-1}(U) \leq x\right) \\
&= \Pr(U \leq F_X(x)) \text{ since } F_X \text{ is non decreasing} \\
&= \int_0^1 \mathbb{I}(u \leq F_X(x))\, du \text{ since } U \sim \mathcal{U}[0,1] \\
&= F_X(x)
\end{aligned}
$$

- The cdf of X produced by the algorithm above is precisely the cdf of π!

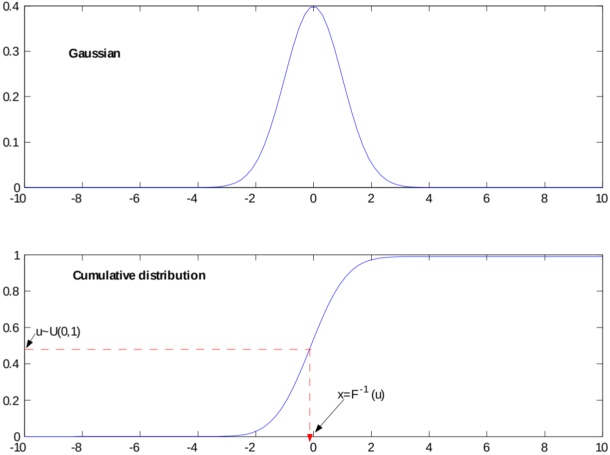


Figure: The density and cdf of a normal distribution

- Consider the exponential of parameter 1 then

$$\pi(x) = \exp(-x)\, \mathbb{I}_{[0,\infty)}$$

thus the cdf of X is

$$F_X(x) = \int_{-\infty}^{x} \pi(u)\, du = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - \exp(-x) & \text{if } x > 0 \end{cases}$$

- Consider the exponential of parameter 1 then

$$\pi(x) = \exp(-x)\, \mathbb{I}_{[0,\infty)}$$

thus the cdf of X is

$$F_X(x) = \int_{-\infty}^{x} \pi(u)\, du = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - \exp(-x) & \text{if } x > 0 \end{cases}$$

- Thus the inverse cdf is

$$1 - \exp(-x) = u \Leftrightarrow x = -\log(1-u) = F_X^{-1}(u).$$

- Consider the exponential of parameter 1 then

$$\pi(x) = \exp(-x)\, \mathbb{I}_{[0,\infty)}$$

thus the cdf of X is

$$F_X(x) = \int_{-\infty}^{x} \pi(u)\, du = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - \exp(-x) & \text{if } x > 0 \end{cases}$$

- Thus the inverse cdf is

$$1 - \exp(-x) = u \Leftrightarrow x = -\log(1-u) = F_X^{-1}(u).$$

- Inverse method: $U \sim \mathcal{U}[0,1]$ then $X = -\log(1-U) \sim \pi$ and $X = -\log(U) \sim \pi$.

- Assume you have $P >> 1$ i.i.d. real-valued rv $X_i \sim f_X$ (cdf F_X) and you are interested in sampling realizations from the distribution of

$$Z = \max(X_1, ..., X_P).$$

- Assume you have $P >> 1$ i.i.d. real-valued rv $X_i \sim f_X$ (cdf F_X) and you are interested in sampling realizations from the distribution of

$$Z = \max(X_1, ..., X_P).$$

- **Brute force direct method**. Sample $X_1, ..., X_P \sim f$ then compute $Z = \max(X_1, ..., X_P)$.

- Assume you have $P >> 1$ i.i.d. real-valued rv $X_i \sim f_X$ (cdf F_X) and you are interested in sampling realizations from the distribution of

$$Z = \max\left(X_1, ..., X_P\right).$$

- **Brute force direct method**. Sample $X_1, ..., X_P \sim f$ then compute $Z = \max\left(X_1, ..., X_P\right)$.
- **Indirect method**. We have

$$\begin{aligned} F_Z(z) &= \Pr\left(X_1 \leq z, ..., X_P \leq z\right) \\ &= \prod_{k=1}^{P} \Pr\left(X_i \leq z\right) = \left[F_X(z)\right]^P \end{aligned}$$

so it follows that for any $U \sim \mathcal{U}[0,1]$

$$Z = F_Z^{-1}(U) = F_X^{-1}\left(U^{1/P}\right)$$

is distributed according to f_Z

- Simple method to sample univariate distributions.

- Simple method to sample univariate distributions.
- This method is only limited to simple cases where the inverse cdf admits a closed form or can be tabulated.

- Simple method to sample univariate distributions.
- This method is only limited to simple cases where the inverse cdf admits a closed form or can be tabulated.
- In practice, it is really very limited.

Change of Variables

- 'Idea': Using the fact that π is related to other distributions easier to sample.

Change of Variables

- 'Idea': Using the fact that π is related to other distributions easier to sample.
- This is very specific!

Change of Variables

- 'Idea': Using the fact that π is related to other distributions easier to sample.
- This is very specific!
- If $X_i \sim \mathcal{E}xp(1)$ then

$$
\begin{aligned}
Y &= 2\sum_{j=1}^{\nu} X_j \sim \chi^2_{2\nu}, \\
Y &= \beta \sum_{j=1}^{\alpha} X_j \sim \mathcal{G}(\alpha, \beta), \\
Y &= \frac{\sum_{j=1}^{\alpha} X_j}{\sum_{j=1}^{\alpha+\beta} X_j} \sim \mathcal{B}e(\alpha, \beta).
\end{aligned}
$$

- Consider $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ then its polar coordinates (R, θ) are independent and distributed according to

$$\begin{aligned} R^2 &= X_1^2 + X_2^2 \sim \mathcal{E}\mathrm{xp}(1/2), \\ \theta &\sim \mathcal{U}[0, 2\pi]. \end{aligned}$$

- Consider $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ then its polar coordinates (R,θ) are independent and distributed according to

$$\begin{aligned} R^2 &= X_1^2 + X_2^2 \sim \mathcal{E}\text{xp}(1/2), \\ \theta &\sim \mathcal{U}[0, 2\pi]. \end{aligned}$$

- It is simple to simulate $R = \sqrt{-2\log(U_1)}$ and $\theta = 2\pi U_2$ where $U_1, U_2 \sim \mathcal{U}[0,1]$ then

$$\begin{aligned} X_1 &= R\cos\theta = \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\ X_2 &= R\sin\theta = \sqrt{-2\log(U_1)}\sin(2\pi U_2). \end{aligned}$$

- Consider $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ then its polar coordinates (R, θ) are independent and distributed according to

$$
\begin{aligned}
R^2 &= X_1^2 + X_2^2 \sim \mathcal{E}\mathrm{xp}(1/2), \\
\theta &\sim \mathcal{U}[0, 2\pi].
\end{aligned}
$$

- It is simple to simulate $R = \sqrt{-2\log(U_1)}$ and $\theta = 2\pi U_2$ where $U_1, U_2 \sim \mathcal{U}[0,1]$ then

$$
\begin{aligned}
X_1 &= R\cos\theta = \sqrt{-2\log(U_1)}\cos(2\pi U_2), \\
X_2 &= R\sin\theta = \sqrt{-2\log(U_1)}\sin(2\pi U_2).
\end{aligned}
$$

- By construction X_1 and X_2 are two independent $\mathcal{N}(0,1)$ rvs.

Sampling via Composition

- Assume we have
$$\pi(x) = \int \overline{\pi}(x, y)\, dy$$
where it is easy to sample from $\overline{\pi}(x, y)$ but difficult/impossible to compute $\pi(x)$.

Sampling via Composition

- Assume we have
$$\pi(x) = \int \overline{\pi}(x, y)\, dy$$
where it is easy to sample from $\overline{\pi}(x, y)$ but difficult/impossible to compute $\pi(x)$.
- In this case, it is sufficient to sample $(X, Y) \sim \overline{\pi} \Rightarrow X \sim \pi$.

Sampling via Composition

- Assume we have
$$\pi(x) = \int \overline{\pi}(x, y)\, dy$$
where it is easy to sample from $\overline{\pi}(x, y)$ but difficult/impossible to compute $\pi(x)$.
- In this case, it is sufficient to sample $(X, Y) \sim \overline{\pi} \Rightarrow X \sim \pi$.
- One can sample from $\overline{\pi}(x, y) = \overline{\pi}(y)\, \overline{\pi}(x|\, y)$ by
$$Y \sim \overline{\pi} \text{ then } X|\, Y \sim \overline{\pi}(\cdot|\, Y).$$

Applications to Scale Mixture of Gaussians

- A very useful application of the composition method is for scale mixture of Gaussians; i.e.

$$\pi(x) = \int \mathcal{N}(x; 0, 1/y)\, \overline{\pi}(y)\, dy.$$

Applications to Scale Mixture of Gaussians

- A very useful application of the composition method is for scale mixture of Gaussians; i.e.

$$\pi(x) = \int \mathcal{N}(x; 0, 1/y)\, \overline{\pi}(y)\, dy.$$

- For various choices of the mixing distributions $\overline{\pi}(y)$, we obtain distributions $\pi(x)$ which are t-student, $\alpha-$stable, Laplace, logistic.

- A very useful application of the composition method is for scale mixture of Gaussians; i.e.

$$\pi(x) = \int \mathcal{N}(x; 0, 1/y)\, \overline{\pi}(y)\, dy.$$

- For various choices of the mixing distributions $\overline{\pi}(y)$, we obtain distributions $\pi(x)$ which are t-student, $\alpha-$stable, Laplace, logistic.
- **Example**: If

$$Y \sim \chi^2_\nu \text{ and } X|\, Y \sim \mathcal{N}(0, \nu/y)$$

then X is marginally distributed according to a t-Student with ν degrees of freedom.

Applications to Scale Mixture of Gaussians

- A very useful application of the composition method is for scale mixture of Gaussians; i.e.

$$\pi(x) = \int \mathcal{N}(x; 0, 1/y)\, \overline{\pi}(y)\, dy.$$

- For various choices of the mixing distributions $\overline{\pi}(y)$, we obtain distributions $\pi(x)$ which are t-student, $\alpha-$stable, Laplace, logistic.
- **Example**: If

$$Y \sim \chi^2_\nu \text{ and } X|\, Y \sim \mathcal{N}(0, \nu/y)$$

then X is marginally distributed according to a t-Student with ν degrees of freedom.

- Conditional upon Y, X is Gaussian: This structure will be used to develop later efficient MCMC algorithms.

Sampling finite mixture of distributions

- Assume one wants to sample from

$$\pi(x) = \sum_{i=1}^{p} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{p} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

Sampling finite mixture of distributions

- Assume one wants to sample from

$$\pi(x) = \sum_{i=1}^{p} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{p} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

- We can introduce $Y \in \{1, ..., p\}$ and introduce

$$\overline{\pi}(x, y) = \pi_y \times \pi_y(x) \Rightarrow \begin{cases} \int \overline{\pi}(x, y)\, dy = \pi(x) \\ \int \overline{\pi}(x, y)\, dx = \overline{\pi}(y) = \pi_y \end{cases}$$

Sampling finite mixture of distributions

- Assume one wants to sample from

$$\pi(x) = \sum_{i=1}^{p} \pi_i . \pi_i(x)$$

  where $\pi_i > 0$, $\sum_{i=1}^{p} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.
- We can introduce $Y \in \{1, ..., p\}$ and introduce

$$\overline{\pi}(x, y) = \pi_y \times \pi_y(x) \Rightarrow \begin{cases} \int \overline{\pi}(x, y)\, dy = \pi(x) \\ \int \overline{\pi}(x, y)\, dx = \overline{\pi}(y) = \pi_y \end{cases}$$

- To sample from $\pi(x)$, then sample $Y \sim \overline{\pi}$ (discrete distribution such that $\Pr(Y = k) = \pi_k$) then

$$X | Y \sim \overline{\pi}(\cdot | Y) = \pi_Y.$$

Sampling infinite mixture of distributions

- Assume you are interested in sampling from the discrete distribution

$$\pi(x) = \sum_{i=1}^{\infty} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{\infty} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

Sampling infinite mixture of distributions

- Assume you are interested in sampling from the discrete distribution

$$\pi(x) = \sum_{i=1}^{\infty} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{\infty} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

- If you try to sample from this distribution by composition, you need to sample from a discrete distribution with infinite support.

Sampling infinite mixture of distributions

- Assume you are interested in sampling from the discrete distribution

$$\pi(x) = \sum_{i=1}^{\infty} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{\infty} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

- If you try to sample from this distribution by composition, you need to sample from a discrete distribution with infinite support.
- Remember that you will set $Y = j$ if

$$\sum_{l=1}^{j-1} \pi_l < U \leq \sum_{l=1}^{j} \pi_l$$

Sampling infinite mixture of distributions

- Assume you are interested in sampling from the discrete distribution

$$\pi(x) = \sum_{i=1}^{\infty} \pi_i . \pi_i(x)$$

where $\pi_i > 0$, $\sum_{i=1}^{\infty} \pi_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

- If you try to sample from this distribution by composition, you need to sample from a discrete distribution with infinite support.
- Remember that you will set $Y = j$ if

$$\sum_{l=1}^{j-1} \pi_l < U \leq \sum_{l=1}^{j} \pi_l$$

- No need to truncate: sample U and then find j such that the above condition is satisfied.

Accept-Reject Method

- The rejection method allows one to sample according to a distribution π defined on $\mathcal{X}$ only known up to a proportionality constant, say $\pi \propto \pi^*$.

Accept-Reject Method

- The rejection method allows one to sample according to a distribution π defined on $\mathcal{X}$ only known up to a proportionality constant, say $\pi \propto \pi^*$.
- It relies on samples generated from a *proposal* distribution q on $\mathcal{X}$. q might as well be known only up to a normalising constant, say $q \propto q^*$.

Accept-Reject Method

- The rejection method allows one to sample according to a distribution π defined on $\mathcal{X}$ only known up to a proportionality constant, say $\pi \propto \pi^*$.
- It relies on samples generated from a *proposal* distribution q on $\mathcal{X}$. q might as well be known only up to a normalising constant, say $q \propto q^*$.
- We need q^* to 'dominate' π^*; i.e.

$$C = \sup_{x \in \mathcal{X}} \frac{\pi^*(x)}{q^*(x)} < +\infty$$

Accept-Reject Method

- The rejection method allows one to sample according to a distribution π defined on $\mathcal{X}$ only known up to a proportionality constant, say $\pi \propto \pi^*$.
- It relies on samples generated from a *proposal* distribution q on $\mathcal{X}$. q might as well be known only up to a normalising constant, say $q \propto q^*$.
- We need q^* to 'dominate' π^*; i.e.

$$C = \sup_{x \in \mathcal{X}} \frac{\pi^*(x)}{q^*(x)} < +\infty$$

- This implies $\pi^*(x) > 0 \Rightarrow q^*(x) > 0$ but also that the tails of $q^*(x)$ must be thicker than the tails of $\pi^*(x)$.

Consider $C' \geq C$. Then the accept/reject procedure proceeds as follows.

1. Sample $Y \sim q$ and $U \sim \mathcal{U}[0, 1]$.

Consider $C' \geq C$. Then the accept/reject procedure proceeds as follows.

1. Sample $Y \sim q$ and $U \sim \mathcal{U}[0,1]$.
2. If $U < \frac{\pi^*(Y)}{C' q^*(Y)}$ then return Y; otherwise return to step 1.

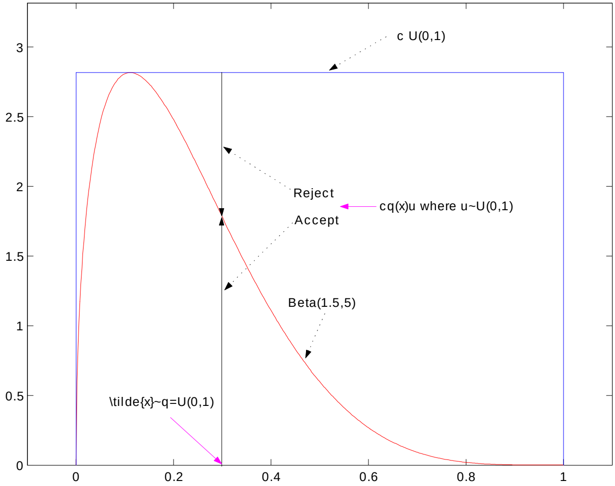


Figure: The idea behind the rejection method for $\pi(x) = \pi^*(x) = \mathcal{B}e(x; 1.5, 5)$, $q(x) = q^*(x) = \mathcal{U}_{[0,1]}(x)$, $C' = C$.

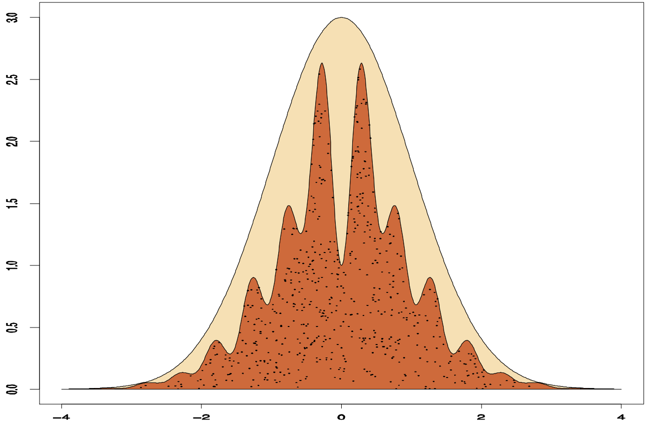


Figure: Sampling from
$$\pi(x) \propto \exp\left(-x^2/2\right)\left(\sin(6x)^2 + 3\cos(x)^2\sin(4x)^2 + 1\right)$$

- We now prove that $\Pr\left(Y \leq x \middle| Y \text{ accepted}\right) = \Pr\left(X \leq x\right)$.

- We now prove that $\Pr\left(Y \leq x \middle| Y \text{ accepted}\right) = \Pr\left(X \leq x\right)$.
- We have for any $x \in \mathcal{X}$

$$
\begin{aligned}
& \Pr\left(Y \leq x \text{ and } Y \text{ accepted}\right) \\
= & \int_0^1 \int_{-\infty}^{x} \mathbb{I}\left(u \leq \frac{\pi^*(y)}{C' q^*(y)}\right) q(y) \times 1 dy du \\
= & \int_{-\infty}^{x} \frac{\pi^*(y)}{C' q^*(y)} q(y)\, dy \\
= & \frac{\int_{-\infty}^{x} \pi^*(y)\, dy}{C' \int_{\mathcal{X}} q^*(y)\, dy}.
\end{aligned}
$$

- We now prove that $\Pr\left(Y \leq x \middle| Y \text{ accepted}\right) = \Pr\left(X \leq x\right)$.
- We have for any $x \in \mathcal{X}$

$$\begin{aligned} & \Pr\left(Y \leq x \text{ and } Y \text{ accepted}\right) \\ = & \int_0^1 \int_{-\infty}^x \mathbb{I}\left(u \leq \frac{\pi^*\left(y\right)}{C' q^*\left(y\right)}\right) q\left(y\right) \times 1 dy du \\ = & \int_{-\infty}^x \frac{\pi^*\left(y\right)}{C' q^*\left(y\right)} q\left(y\right) dy \\ = & \frac{\int_{-\infty}^x \pi^*\left(y\right) dy}{C' \int_{\mathcal{X}} q^*\left(y\right) dy}. \end{aligned}$$

- The probability of being accepted is the marginal of $\Pr\left(Y \leq x \text{ and } Y \text{ accepted}\right)$

$$\Pr\left(Y \text{ accepted}\right) = \frac{\int_{\mathcal{X}} \pi^*\left(y\right) dy}{C' \int_{\mathcal{X}} q^*\left(y\right) dy}.$$

- Thus

$$
\begin{aligned}
\Pr\left(Y \leq x \middle| Y \text{ accepted}\right) &= \frac{\Pr\left(Y \leq x \text{ and } Y \text{ accepted}\right)}{\Pr\left(Y \text{ accepted}\right)} \\
&= \frac{\int_{-\infty}^{x} \pi^{*}\left(y\right) dy}{\int_{\mathcal{X}} \pi^{*}\left(y\right) dy} = \int_{-\infty}^{x} \pi\left(y\right) dy.
\end{aligned}
$$

- Thus

$$\begin{aligned}\Pr\left(Y \leq x \mid Y \text{ accepted}\right) &= \frac{\Pr\left(Y \leq x \text{ and } Y \text{ accepted}\right)}{\Pr\left(Y \text{ accepted}\right)} \\ &= \frac{\int_{-\infty}^{x} \pi^*\left(y\right) dy}{\int_{\mathcal{X}} \pi^*\left(y\right) dy} = \int_{-\infty}^{x} \pi\left(y\right) dy.\end{aligned}$$

- **Example**: We want to sample from $\mathcal{B}e\left(x;\alpha,\beta\right) \propto x^{\alpha-1}\left(1-x\right)^{\beta-1}$ using $\mathcal{U}\left[0,1\right]$. One can find

$$\sup_{x\in[0,1]} \frac{x^{\alpha-1}\left(1-x\right)^{\beta-1}}{1}$$

analytically for $\alpha, \beta > 1$! We do not need the normalizing constant of $\mathcal{B}e$.

- You do not lose anything by not knowing the normalizing constant of q^*.

- You do not lose anything by not knowing the normalizing constant of q^*.
- **Example**: The target π is given by

$$\pi(x) \propto \pi^*(x) = \exp\left(-\frac{x^2}{2}\right) m(x)$$

where $m(x) \leq M$ for any $x \in \mathsf{X}$.

- You do not lose anything by not knowing the normalizing constant of q^*.
- **Example**: The target π is given by

$$\pi(x) \propto \pi^*(x) = \exp\left(-\frac{x^2}{2}\right) m(x)$$

where $m(x) \leq M$ for any $x \in \mathrm{X}$.

- If we use $q(x) = q^*(x) = (2\pi)^{-1/2} \exp\left(-\frac{x^2}{2}\right)$, then we have

$$\frac{\pi^*(x)}{q^*(x)} \leq C_1 = (2\pi)^{1/2} M \text{ and } \Pr(Y \text{ accepted}) = \frac{\int_{\mathrm{X}} \pi^*(y)\, dy}{C_1}.$$

- You do not lose anything by not knowing the normalizing constant of q^*.
- **Example**: The target π is given by

$$\pi(x) \propto \pi^*(x) = \exp\left(-\frac{x^2}{2}\right) m(x)$$

where $m(x) \leq M$ for any $x \in \mathsf{X}$.

- If we use $q(x) = q^*(x) = (2\pi)^{-1/2} \exp\left(-\frac{x^2}{2}\right)$, then we have

$$\frac{\pi^*(x)}{q^*(x)} \leq C_1 = (2\pi)^{1/2} M \text{ and } \Pr(Y \text{ accepted}) = \frac{\int_{\mathsf{X}} \pi^*(y)\, dy}{C_1}.$$

- If we use $q^*(x) = \exp\left(-\frac{x^2}{2}\right)$, then we have $\frac{\pi^*(x)}{q^*(x)} \leq C_2 = M$ and

$$\Pr(Y \text{ accepted}) = \frac{\int_{\mathsf{X}} \pi^*(y)\, dy}{C_2 (2\pi)^{1/2}} = \frac{\int_{\mathsf{X}} \pi^*(y)\, dy}{C_1}$$

- The acceptance probability $\Pr(Y \text{ accepted})$ is a measure of efficiency.

- The acceptance probability $\Pr(Y \text{ accepted})$ is a measure of efficiency.
- The number of trials before accepting a candidate follows a geometric distribution

$$\begin{aligned} \Pr\left(k^{\text{th}} \text{ proposal accepted}\right) &= (1-\rho)^{k-1}\rho \\ \text{where } \rho &= \left(\frac{\int_{\mathcal{X}} \pi^*(y)\,dy}{C' \int_{\mathcal{X}} q^*(y)\,dy}\right) \end{aligned}$$

thus its expected value is

$$\sum_{k=0}^{\infty} k(1-\rho)^{k-1}\rho = \frac{1}{\varrho} = \frac{1}{\Pr(Y \text{ accepted})}.$$

- The acceptance probability $\Pr(Y \text{ accepted})$ is a measure of efficiency.
- The number of trials before accepting a candidate follows a geometric distribution

$$\Pr\left(k^{\text{th}} \text{ proposal accepted}\right) = (1-\rho)^{k-1}\rho$$
$$\text{where } \rho = \left(\frac{\int_{\mathcal{X}} \pi^*(y)\, dy}{C' \int_{\mathcal{X}} q^*(y)\, dy}\right)$$

thus its expected value is

$$\sum_{k=0}^{\infty} k(1-\rho)^{k-1}\rho = \frac{1}{\varrho} = \frac{1}{\Pr(Y \text{ accepted})}.$$

- This is important to better understand the Metropolis-Hastings algorithm.

- Consider a Bayesian model: prior $\pi(\theta)$ and likelihood $f(x|\theta)$.

- Consider a Bayesian model: prior $\pi(\theta)$ and likelihood $f(x|\theta)$.
- The posterior distribution is given by

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int_{\Theta} \pi(\theta) f(x|\theta) d\theta} \propto \pi^{*}(\theta|x)$$

where $\pi^{*}(\theta|x) = \pi(\theta) f(x|\theta)$.

- Consider a Bayesian model: prior $\pi(\theta)$ and likelihood $f(x|\theta)$.
- The posterior distribution is given by

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int_{\Theta} \pi(\theta) f(x|\theta) d\theta} \propto \pi^*(\theta|x)$$
where $\pi^*(\theta|x) = \pi(\theta) f(x|\theta)$.

- We can use the prior distribution as a candidate distribution $q(\theta) = q^*(\theta) = \pi(\theta)$ as long as

$$\sup_{\theta\in\Theta} \frac{\pi^*(\theta|x)}{q^*(\theta)} = \sup_{\theta\in\Theta} f(x|\theta) \leq C.$$

- Consider a Bayesian model: prior $\pi(\theta)$ and likelihood $f(x|\theta)$.
- The posterior distribution is given by

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int_\Theta \pi(\theta) f(x|\theta) d\theta} \propto \pi^*(\theta|x)$$
$$\text{where } \pi^*(\theta|x) = \pi(\theta) f(x|\theta).$$

- We can use the prior distribution as a candidate distribution $q(\theta) = q^*(\theta) = \pi(\theta)$ as long as

$$\sup_{\theta\in\Theta} \frac{\pi^*(\theta|x)}{q^*(\theta)} = \sup_{\theta\in\Theta} f(x|\theta) \leq C.$$

- In many applications, the likelihood is bounded so one can use the rejection procedure and it is accepted with proba $\int_\Theta \pi(\theta) f(x|\theta) d\theta / C$.

- Consider a Bayesian model: prior $\pi(\theta)$ and likelihood $f(x|\theta)$.
- The posterior distribution is given by

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int_\Theta \pi(\theta) f(x|\theta) d\theta} \propto \pi^*(\theta|x)$$
$$\text{where } \pi^*(\theta|x) = \pi(\theta) f(x|\theta).$$

- We can use the prior distribution as a candidate distribution $q(\theta) = q^*(\theta) = \pi(\theta)$ as long as

$$\sup_{\theta\in\Theta} \frac{\pi^*(\theta|x)}{q^*(\theta)} = \sup_{\theta\in\Theta} f(x|\theta) \leq C.$$

- In many applications, the likelihood is bounded so one can use the rejection procedure and it is accepted with proba $\int_\Theta \pi(\theta) f(x|\theta) d\theta / C$.
- Moreover, if we have $q^*(\theta) = \pi(\theta)$ then expected value before acceptance

$$\frac{c}{\int_\Theta \pi(\theta) f(x|\theta) d\theta}.$$

Limitations of Accept-Reject

- Consider the case where $\mathcal{X} = \mathbb{R}^n$

$$\pi(\theta) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\sum_{i=1}^{n} \theta_i^2}{2} \right)$$

and

$$q_\sigma(\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^{n} \theta_i^2}{2\sigma^2} \right)$$

Limitations of Accept-Reject

- Consider the case where $\mathcal{X} = \mathbb{R}^n$

$$\pi(\theta) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \theta_i^2}{2}\right)$$

and

$$q_\sigma(\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \theta_i^2}{2\sigma^2}\right)$$

- We have for any $\sigma > 1$

$$\frac{\pi(\theta)}{q_\sigma(\theta)} = \sigma^n \exp\left(-\sum_{i=1}^n \theta_i^2 \left(1 - \frac{1}{2\sigma^2}\right)\right) \leq \sigma^n \text{ for any } \theta$$

and

$$\Pr(Y \text{ accepted}) = \frac{1}{\sigma^n}$$

Limitations of Accept-Reject

- Consider the case where $\mathcal{X} = \mathbb{R}^n$

$$\pi(\theta) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \theta_i^2}{2}\right)$$

and

$$q_\sigma(\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \theta_i^2}{2\sigma^2}\right)$$

- We have for any $\sigma > 1$

$$\frac{\pi(\theta)}{q_\sigma(\theta)} = \sigma^n \exp\left(-\sum_{i=1}^n \theta_i^2 \left(1 - \frac{1}{2\sigma^2}\right)\right) \leq \sigma^n \text{ for any } \theta$$

and

$$\Pr(Y \text{ accepted}) = \frac{1}{\sigma^n}$$

- Despite having a very good proposal then the acceptance probability decreases exponentially fast with the dimension of the problem.

Advantages.

- Rather universal, and compared to the inverse cdf method requires less algebraic properties.

Drawbacks.

Advantages.

- Rather universal, and compared to the inverse cdf method requires less algebraic properties.
- Neither normalisation constant of π nor that of q are needed.

Drawbacks.

Advantages.

- Rather universal, and compared to the inverse cdf method requires less algebraic properties.
- Neither normalisation constant of π nor that of q are needed.

Drawbacks.

- How to construct the proposal $q(x)$ automatically?

Advantages.

- Rather universal, and compared to the inverse cdf method requires less algebraic properties.
- Neither normalisation constant of π nor that of q are needed.

Drawbacks.

- How to construct the proposal $q\left(x\right)$ automatically?
- Typically the performance of the method decrease exponentially with the dimension of the problem.

Beyond Standard Accept Reject

- In the standard Rejection algorithm, the candidate is sampled before U. This is not necessary.

Beyond Standard Accept Reject

- In the standard Rejection algorithm, the candidate is sampled before U. This is not necessary.
- **Proposition**: Let $(Y_n, I_n)_{n\geq 1}$ be a sequence of i.i.d. rvs taking values in $\mathcal{X} \times \{0,1\}$ such that $Y_1 \sim q$ and

$$\Pr(I_1 = 1 | Y_1 = y) = \frac{\pi^*(y)}{Cq^*(y)}$$

Define $\tau = \min\{i \geq 1 : I_i = 1\}$, then $Y_\tau \sim \pi$.

Beyond Standard Accept Reject

- In the standard Rejection algorithm, the candidate is sampled before U. This is not necessary.
- **Proposition**: Let $(Y_n, I_n)_{n \geq 1}$ be a sequence of i.i.d. rvs taking values in $\mathcal{X} \times \{0, 1\}$ such that $Y_1 \sim q$ and

$$\Pr\left(I_1 = 1 \middle| Y_1 = y \right) = \frac{\pi^* (y)}{Cq^* (y)}$$

  Define $\tau = \min \{ i \geq 1 : I_i = 1 \}$, then $Y_\tau \sim \pi$.
- This result is useful if there are ways of constructing condition for the acceptance or rejection of the current proposed element Y from minimal information about it.

Envelop Accept Reject

- Squeeze principle: Assume we have

$$q_L^*(x) \leq \pi^*(x) \leq Cq^*(x)$$

then we can modify the algorithm as follows.

Envelop Accept Reject

- Squeeze principle: Assume we have

$$q_L^*(x) \leq \pi^*(x) \leq Cq^*(x)$$

  then we can modify the algorithm as follows.

1. Sample $Y \sim q$ and $U \sim \mathcal{U}(0,1)$.

Envelop Accept Reject

- Squeeze principle: Assume we have

$$q_L^*(x) \leq \pi^*(x) \leq Cq^*(x)$$

then we can modify the algorithm as follows.

1. Sample $Y \sim q$ and $U \sim \mathcal{U}(0,1)$.
2. If $U \leq \frac{q_L^*(Y)}{C' q^*(Y)}$ then return Y;

Envelop Accept Reject

- Squeeze principle: Assume we have

$$q_L^*(x) \leq \pi^*(x) \leq Cq^*(x)$$

then we can modify the algorithm as follows.

1. Sample $Y \sim q$ and $U \sim \mathcal{U}(0,1)$.
2. If $U \leq \frac{q_L^*(Y)}{C'q^*(Y)}$ then return Y;
3. Otherwise, accept X if $U < \frac{\pi^*(Y)}{C'q^*(Y)}$, otherwise return to step 1.

Adaptive Rejection Sampling

- Consider the class of univariate log-concave densities; i.e. we have

$$\frac{\partial^2 \log \pi(x)}{\partial x^2} < 0$$

where $\pi(x) = f(x) / \int f(x)\, dx$.

Adaptive Rejection Sampling

- Consider the class of univariate log-concave densities; i.e. we have

$$\frac{\partial^2 \log \pi(x)}{\partial x^2} < 0$$

where $\pi(x) = f(x) / \int f(x)\, dx$.

- The idea is to construct automatically an piecewise linear upper (and lower) bound for the target.

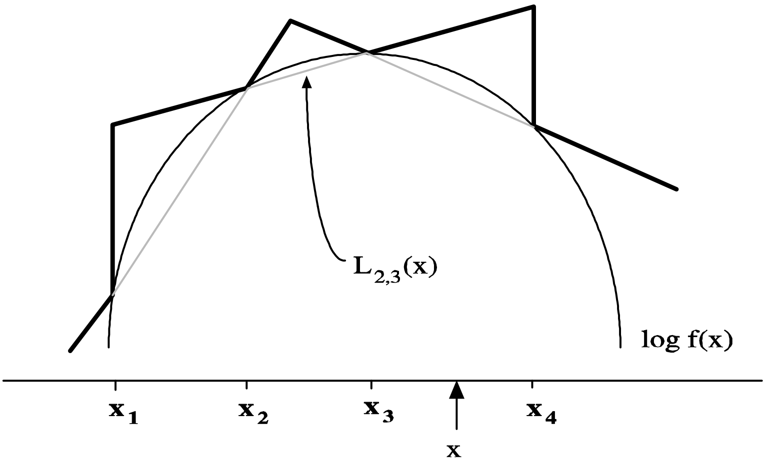
Adaptive Rejection Sampling

- Consider the class of univariate log-concave densities; i.e. we have

$$\frac{\partial^2 \log \pi(x)}{\partial x^2} < 0$$

  where $\pi(x) = f(x) / \int f(x)\, dx$.
- The idea is to construct automatically an piecewise linear upper (and lower) bound for the target.
- Let $\mathcal{S}_n$ be a set of points $\{x_i\}_{i=0}^{n+1}$ in the support of $\pi(x)$ such that $h(x_i) = \log f(x_i)$.

- Because of concavity, the line $L_{i,i+1}$ going through $(x_i, h(x_i))$ and $(x_{i+1}, h(x_{i+1}))$ is below the graph of h in $[x_i, x_{i+1}]$ and is above this graph outside this interval.



- We define $\overline{h}_n(x) = \min\{L_{i-1,i}(x), L_{i+1,i+2}(x)\}$, $\underline{h}_n(x) = L_{i,i+1}(x)$ [where $\overline{h}_n(x) = -\infty$ and $\overline{h}_n(x) = \min\{L_{0,1}(x), L_{n,n+1}(x)\}$ on $[x_0, x_{n+1}]^c$ so that
$$\underline{h}_n(x) \leq h(x) \leq \overline{h}_n(x)$$

- We define $\overline{h}_n(x) = \min\{L_{i-1,i}(x), L_{i+1,i+2}(x)\}$, $\underline{h}_n(x) = L_{i,i+1}(x)$ [where $\overline{h}_n(x) = -\infty$ and $\overline{h}_n(x) = \min\{L_{0,1}(x), L_{n,n+1}(x)\}$ on $[x_0, x_{n+1}]^c$ so that
$$\underline{h}_n(x) \leq h(x) \leq \overline{h}_n(x)$$
- Therefore we have for $\underline{f}_n(x) = \exp \underline{h}_n(x)$, $\overline{f}_n(x) = \exp \overline{h}_n(x)$
$$\underline{f}_n(x) = \exp \underline{h}_n(x) \leq \pi(x) \leq \overline{f}_n(x) = \overline{w}_n g_n(x)$$
where it is easy to compute $\overline{w}_n$ and easy to sample from $g_n(x)$.

- Initialize $n = 0$ and $\mathcal{S}_0$

At iteration $n \geq 1$

- Initialize $n = 0$ and $\mathcal{S}_0$

At iteration $n \geq 1$

1. Generate $Y \sim g_n$.

- Initialize $n = 0$ and $\mathcal{S}_0$

At iteration $n \geq 1$

1. Generate $Y \sim g_n$.
2. If $U \leq \frac{f_n(Y)}{\overline{w}_n \overline{f}_n(Y)}$ then return Y; otherwise set $\mathcal{S}_{n+1} = \mathcal{S}_n \cup \{Y\}$.

- Consider n data (x_i, Y_i)

$$Y_i | \, x_i \sim \mathcal{P}(a + bx_i).$$

and we set the prior

$$\pi(a, b) = \mathcal{N}\left(a; 0, \sigma^2\right) \mathcal{N}\left(b; 0, \tau^2\right)$$

- Consider n data (x_i, Y_i)

$$Y_i | x_i \sim \mathcal{P}(a + bx_i).$$

and we set the prior

$$\pi(a, b) = \mathcal{N}(a; 0, \sigma^2) \mathcal{N}(b; 0, \tau^2)$$

- We have

$$\log \pi(a | x_{1:n}, y_{1:n}, b) = \text{cst} + a \sum y_i - e^a \sum e^{x_i b} - a^2 / 2\sigma^2$$
$$\Rightarrow \frac{\partial^2 \log \pi(a | x_{1:n}, y_{1:n}, b)}{\partial a^2} = -e^a \sum e^{x_i b} - \sigma^{-2} < 0.$$

- Consider n data (x_i, Y_i)

$$Y_i | x_i \sim \mathcal{P}(a + bx_i).$$

and we set the prior

$$\pi(a, b) = \mathcal{N}(a; 0, \sigma^2)\,\mathcal{N}(b; 0, \tau^2)$$

- We have

$$\log \pi(a | x_{1:n}, y_{1:n}, b) = \text{cst} + a \sum y_i - e^a \sum e^{x_i b} - a^2/2\sigma^2$$
$$\Rightarrow \frac{\partial^2 \log \pi(a | x_{1:n}, y_{1:n}, b)}{\partial a^2} = -e^a \sum e^{x_i b} - \sigma^{-2} < 0.$$

- Thus $\pi(a | x_{1:n}, y_{1:n}, b)$ is log-concave, similarly $\pi(b | x_{1:n}, y_{1:n}, a)$ is log-concave.

Monahan's Accept Reject Algorithm

- We want to sample from the cdf
$$F(x) = \frac{H(-G(x))}{H(-1)}$$
where $G(x)$ is a given cdf and
$$H(x) = \sum_{n=1}^{\infty} a_n x^n$$
with $1 = a_1 \geq a_2 \geq \cdots \geq 0$. We only want to use samples from G and $\mathcal{U}[0, 1]$

Monahan's Accept Reject Algorithm

- We want to sample from the cdf
$$F(x) = \frac{H(-G(x))}{H(-1)}$$
where $G(x)$ is a given cdf and
$$H(x) = \sum_{n=1}^{\infty} a_n x^n$$
with $1 = a_1 \geq a_2 \geq \cdots \geq 0$. We only want to use samples from G and $\mathcal{U}[0,1]$
- **Example**: Assume you are interested in sampling from $F(x) = 1 - \cos\left(\frac{\pi x}{2}\right)$ where $0 < x < 1$. You could do it through inversion with $\frac{2}{\pi}\arccos(U)$ but this requires evaluating a complex (transcendental) function. Alternatively we have $G(x) = x^2$ and
$$H(x) = x + \frac{\pi^2}{48}x^2 + \frac{\pi^4}{5760}x^3 + \cdots + \frac{\pi^{2i-2}}{2^{2i-3}(2i)!}x^i + \cdots$$

- Repeat
    - Generate $X \sim G$ and set $K \leftarrow 1$.
    - Repeat
        - Generate $U \sim G$ and $V \sim \mathcal{U}[0,1]$.
        - If $U \leq X$ and $V \leq \frac{a_{K+1}}{a_K}$ then $K \leftarrow K+1$, otherwise stop.

Until K odd, return X.

- We define the event A_n by $X = \max(X, U_1, ..., U_n)$ and $Z_1 = \cdots = Z_n = 1$ where the U_is are the rvs generated in the inner loop and the Z_is are Bernoulli rvs equal to consecutives values $\mathbb{I}_{V \leq \frac{a_{K+1}}{a_K}}$.

- We define the event A_n by $X = \max(X, U_1, ..., U_n)$ and $Z_1 = \cdots = Z_n = 1$ where the U_is are the rvs generated in the inner loop and the Z_is are Bernoulli rvs equal to consecutives values $\mathbb{I}_{V \leq \frac{a_{K+1}}{a_K}}$.
- We have

$$\begin{aligned} P(X \leq x, A_n) &= a_n G(x)^n, \\ P(X \leq x, A_n, A_{n+1}^c) &= P(X \leq x, A_n) - P(X \leq x, A_n, A_{n+1}) \\ &= a_n G(x)^n - a_{n+1} G(x)^{n+1}. \end{aligned}$$

- We define the event A_n by $X = \max(X, U_1, ..., U_n)$ and $Z_1 = \cdots = Z_n = 1$ where the U_is are the rvs generated in the inner loop and the Z_is are Bernoulli rvs equal to consecutives values $\mathbb{I}_{V \leq \frac{a_{K+1}}{a_K}}$.
- We have

$$\begin{aligned} P(X \leq x, A_n) &= a_n G(x)^n, \\ P(X \leq x, A_n, A_{n+1}^c) &= P(X \leq x, A_n) - P(X \leq x, A_n, A_{n+1}) \\ &= a_n G(x)^n - a_{n+1} G(x)^{n+1}. \end{aligned}$$

- The proba that X is accepted is

$$P(K \text{ odd}) = \sum_{n=1}^{\infty} a_n (-1)^{n+1} = H(-1)$$

and the returned X has distribution function

$$F(x) = P(X \leq x) = \frac{\sum_{n=1}^{\infty} a_n G(x)^n (-1)^{n+1}}{H(-1)} = \frac{H(G(-x))}{H(-1)}.$$

- There exists standard techniques to sample from classical distributions.

- There exists standard techniques to sample from classical distributions.
- Rejection is useful for small non-standard distributions but collapses for most "interesting" problems.

- There exists standard techniques to sample from classical distributions.
- Rejection is useful for small non-standard distributions but collapses for most "interesting" problems.
- These algorithms will be building blocks of more complex Monte Carlo algorithms.