# Stat 535 C - Statistical Computing & Monte Carlo Methods

Arnaud Doucet

Email: `arnaud@cs.ubc.ca`

## 1.1– Outline

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

## 2.1– Summary of Last Lecture

- Let $\pi(x)$ be a probability density.

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

$$E_{\widehat{\pi}_N}(\varphi) = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(X^{(i)}\right) \simeq E_\pi(\varphi)$$

and more precisely

$$E_X\left[E_{\widehat{\pi}_N}(\varphi)\right] = E_\pi(\varphi) \text{ and } var_X\left(E_{\widehat{\pi}_N}(\varphi)\right) = \frac{var_\pi(\varphi)}{N}.$$

## 2.1– Summary of Last Lecture

• If we could sample from any distribution $\pi$ easily, then everything would be easy.

• Unfortunately, there is no generic algorithm to sample exactly from any $\pi$.

• Today, we discuss simple methods which are the building blocks of more complex algorithms; i.e. MCMC and SMC.

## 3.1– Sampling from Uniform Random Variables

• All algorithms discussed here rely on the availability of a generator of 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.

## 3.2– Sampling from A Discrete Distribution

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

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

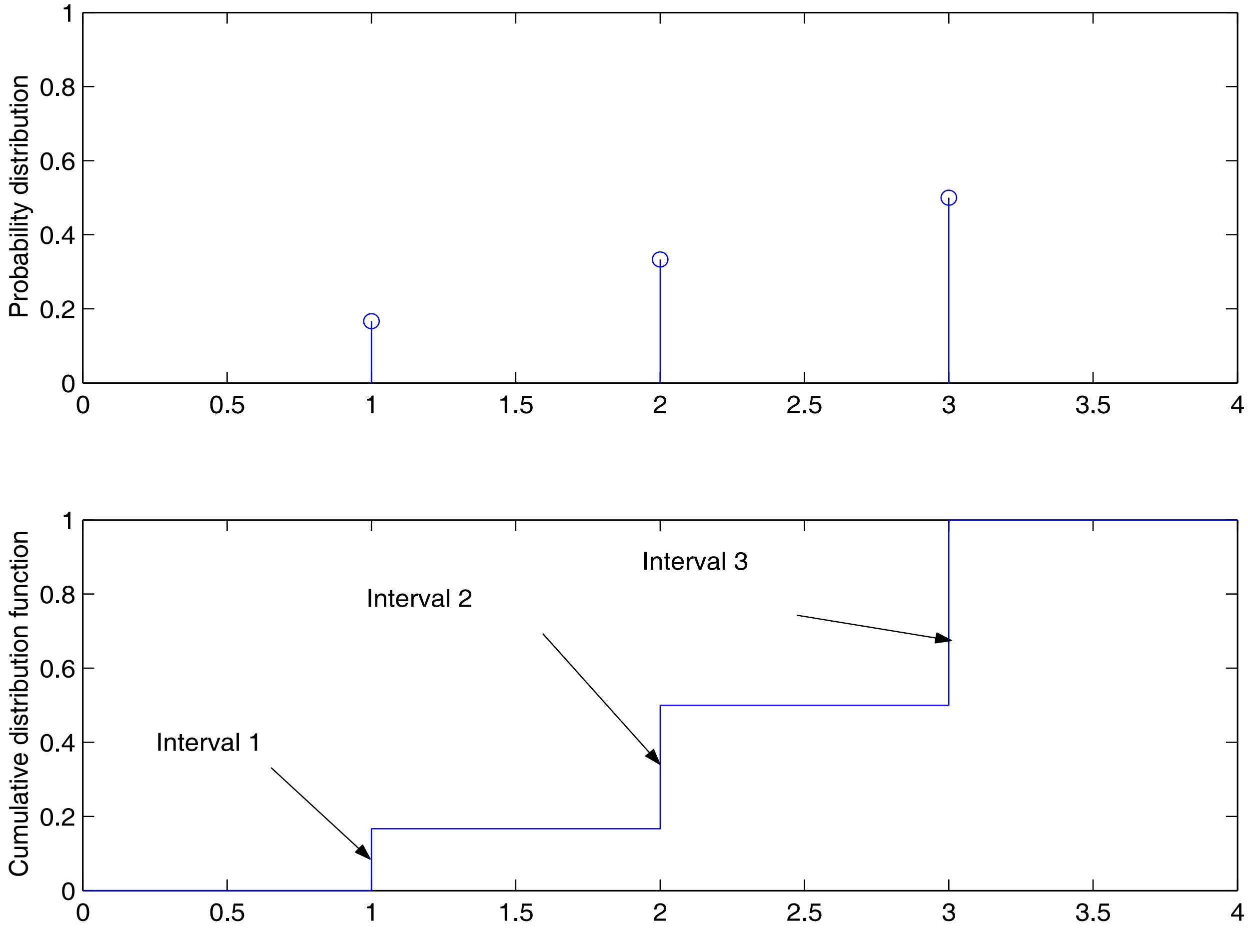
- Define the cdf of $X$ for $x \in [0, 3]$ as

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

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

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

The distribution and cdf of a discrete random variable



# 3.2– Sampling from A Discrete Distribution

- 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 precidely equal to the probability $\pi(X = i)$.

## 3.3– Sampling from a continuous distribution: Inverse Method

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

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

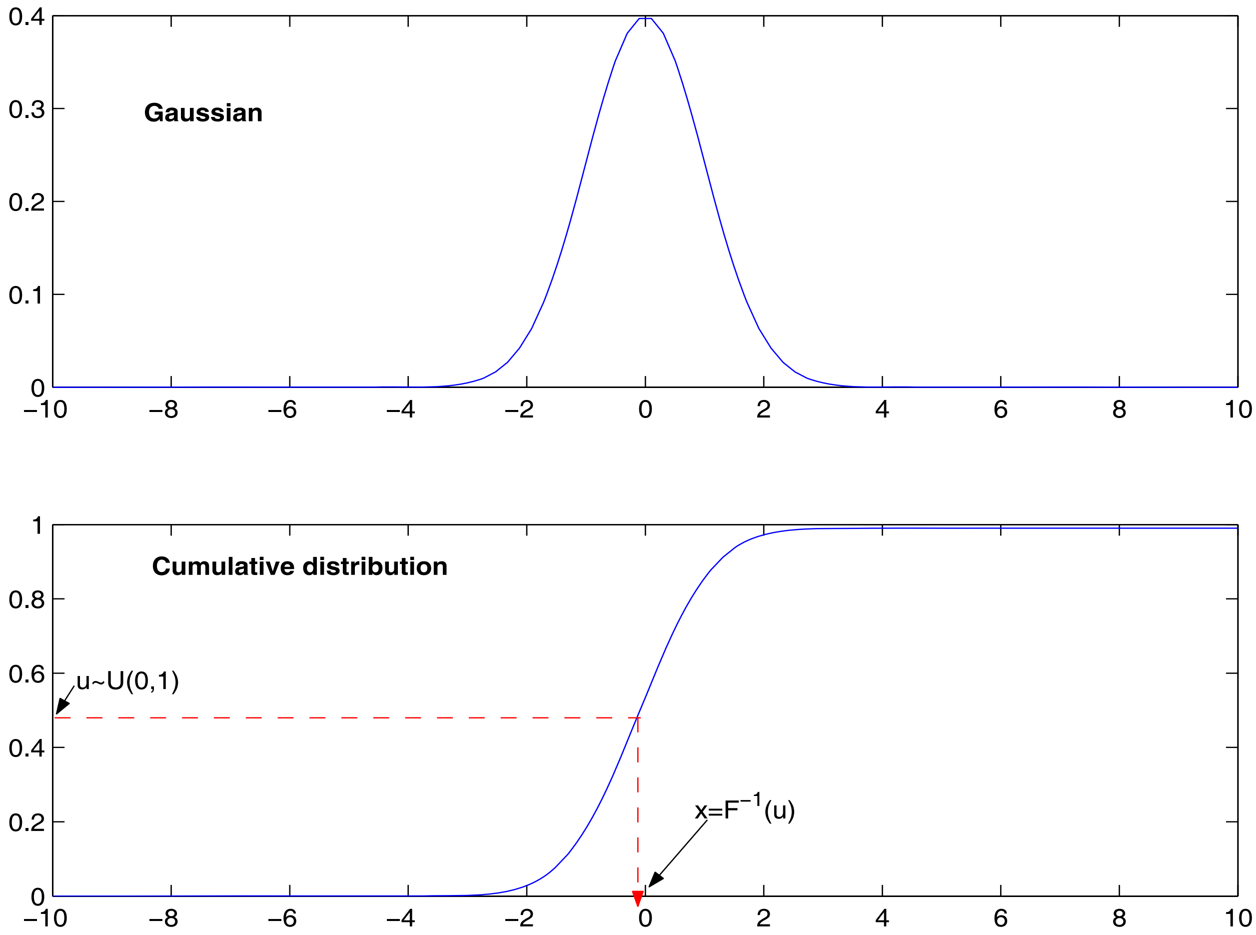
- We would like to use the same algorithm; i.e.

$$U \sim \mathcal{U}[0,1] \text{ and set } X = F_X^{-1}(U).$$

- Question: Do we have $X \sim \pi$?

## 3.3– Sampling from a continuous distribution: Inverse Method

The distribution and cdf of a normal distribution



## 3.3– Sampling from a continuous distribution: Inverse Method

- 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 $\pi$!

## 3.4– Inverse Method: Example

- 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 $X = -\log(U) \sim \pi$.

## 3.4– Inverse Method: Example

- 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.

## 3.5– General Transformation Methods

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

- This is very specific!

- If $X_i \sim \mathcal{E}xp\,(1)$ then

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

## 3.6– Box Muller Algorithm to sample Gaussian random variables

- Consider $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ then its polar coordinates $(R,\theta)$ 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.

# 3.7– Simulation by Composition

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

$$Y \sim \overline{\pi} \text{ then } X|\, Y \sim \overline{\pi}(\cdot|\, Y).$$

## 3.8– Simulation by Composition for Mixture

- Assume one wants to sample from

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

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

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

## 3.9– Simulation by Composition: 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 $\nu$ degrees of freedom.

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

## 3.10– Accept Reject - The first generic method

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

• It relies on samples generated from a *proposal* distribution $q$ on $\mathsf{X}$. $q$ might as well be known only up to a normalising constant, say $q \propto q^*$.

• We need $q$ to 'dominate' $\pi$; i.e.

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

• 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:

**Accept/Reject procedure**

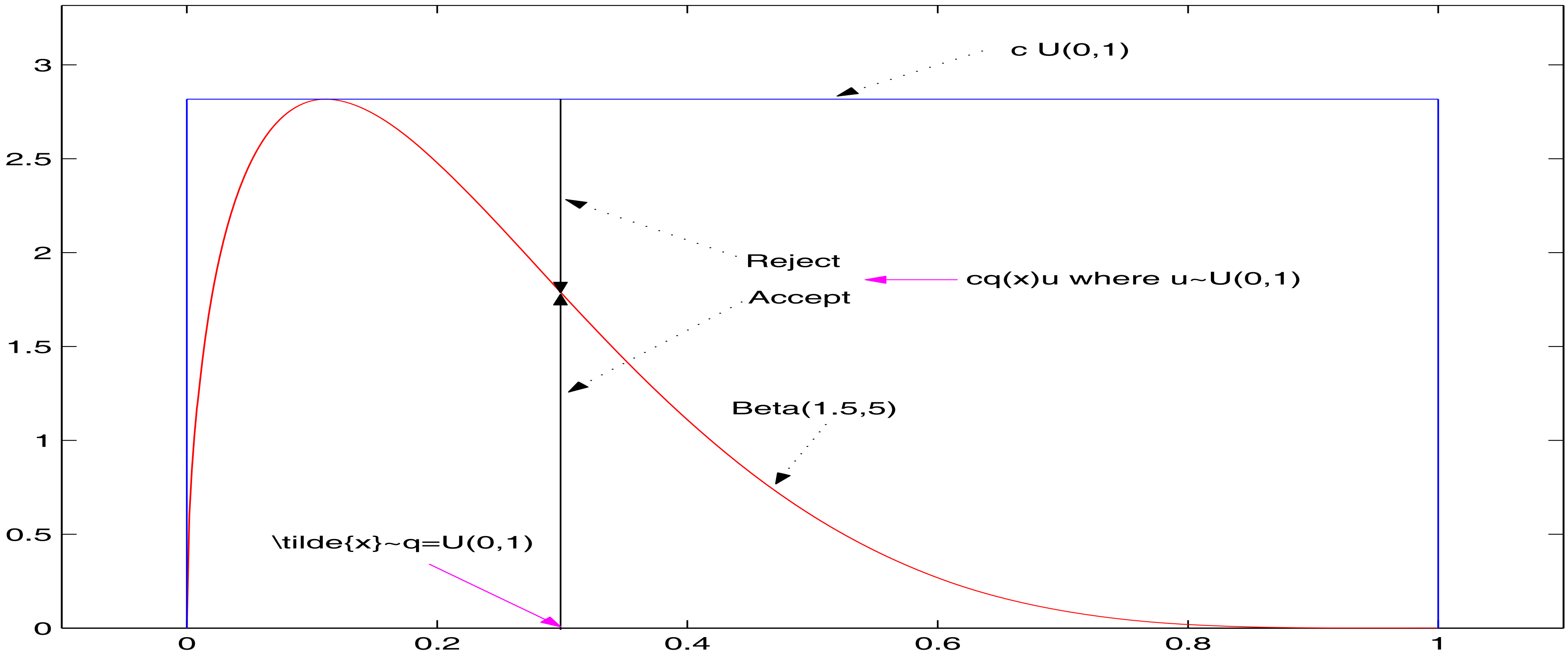
1. Sample $Y \sim q$ and $U \sim \mathcal{U}(0,1)$.

2. If $U < \frac{\pi^*(Y)}{C' q^*(Y)}$ then return $Y$; otherwise return to step 1.

## 3.11– Accept Reject - Illustration

The idea behind the rejection method for $\pi(x) = \pi^*(x) = \mathcal{B}e(x; 1.5, 5)$,

$$q(x) = q^*(x) = \mathcal{U}_{[0,1]}(x),\ C' = C.$$



## 3.12– Accept Reject - Proof of Validity

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

- We have for any $x \in \mathsf{X}$

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

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

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

## 3.12– Accept Reject - Proof of Validity

- Thus

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

- *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 $\mathcal{U}_{[0,1]}$.

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 $\mathcal{B}e$.

## 3.13– Accept Reject - Efficiency

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

$$\frac{1}{\Pr(Y \text{ accepted})}.$$

- The number of trials before success is thus an unbiased estimate of

$$\frac{1}{\Pr(Y \text{ accepted})}.$$

- This is important to better understand Metropolis-Hastings.

## 3.14– Accept Reject as a Sampling through composition

- Almost unknown result (Peterson & Kronmal, 1982): One can rewrite

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

where $p_i = p(1-p)^{i-1}$ and

$$p = \Pr\left(U \leq \frac{\pi^*(X)}{Cq^*(X)}\right).$$

- Instead of simulating from $\mathcal{G}\text{eo}(p)$ directly which is impossible, one simulate an element which admits this probability distribution.

## 3.15– Alternative Formulation of Accept Reject

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

• **Proposition** (Beskos et al., 2005): Let $(Y_n, I_n)_{n\geq 1}$ be a sequence of i.i.d. rvs taking values in $\mathsf{X}\times\{0,1\}$ 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.

## 3.16– Accept Reject - Example

- The target $\pi$ 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

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

- You don't lose anything by not knowing the normalizing constant of $q^*$.

## 3.17– Accept Reject - Application to Bayesian Estimation

- 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 $\int_\Theta \pi(\theta) f(x|\theta) d\theta / C$. End of the course???

## 3.18– The Rejection method does not work well in high-dimension

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

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

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

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

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

## 3.19– Advantages and Drawbacks of the Rejection method

**Advantages.**

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

- Neither normalisation constant of $\pi$ nor that of $q$ are needed.

**Drawbacks.**

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

- Typically the performance of the method decrease exponentially with the dimension of the problem.

## 3.20– Envelope Rejection method

- Squeeze principle: Assume we have

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

then we can modify the algorithm as follows.

**Envelope Accept/Reject procedure**

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.

# 3.21– Adaptive Rejection Sampling

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

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

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

## 3.21– Adaptive Rejection Sampling

Here a nice graph should appear but it does not for whatever reason.

See Fig. 2.5, page 57 in Monte Carlo Statistical Methods.

# 3.21– Adaptive Rejection Sampling

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

At iteration $n \geq 1$

1. Generate $Y \sim q_n$.

2. If $U \leq \frac{\pi(Y)}{C' q_n(Y)}$ then return $Y$; otherwise set $\mathcal{S}_{n+1} = \mathcal{S}_n \cup \{Y\}$.

## 3.22– Adaptive Rejection Sampling: Example

- Consider $n$ data $(x_i, Y_i)$

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

and we set the prior

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

- We have

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

- 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.

## 3.23– Summary

- 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.