# Advanced Simulation - Lecture 3

Patrick Rebeschini

January 22nd, 2018

# From a statistical problem to a sampling problem

- From a statistical model you get a likelihood function and a prior on the parameters.
- Applying Bayes rule, you are interested in

$$\pi(\theta \mid \text{observations}) = \frac{\mathcal{L}(\text{observations}; \theta) p(\theta)}{\int_\Theta \mathcal{L}(\text{observations}; \theta) p(\theta) d\theta}.$$

- Inference $\equiv$ integral w.r.t. posterior distribution.
- Integrals can be approximated by Monte Carlo.
- For Monte Carlo you need samples.
- Today: inversion, transformation, composition, rejection.

# Inversion Method

- Consider a real-valued random variable $X$ and its associated cumulative distribution function (cdf)

$$F(x) = \mathbb{P}(X \leq x) = F(x).$$

- The cdf $F : \mathbb{R} \to [0, 1]$ is
  - increasing; i.e. if $x \leq y$ then $F(x) \leq F(y)$,
  - right continuous; i.e. $F(x + \varepsilon) \to F(x)$ as $\varepsilon \to 0^+$,
  - $F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to +\infty$.
- We define the generalised inverse

$$F^-(u) = \inf \{x \in \mathbb{R}; F(x) \geq u\}$$

also known as the quantile function.

# Inversion Method

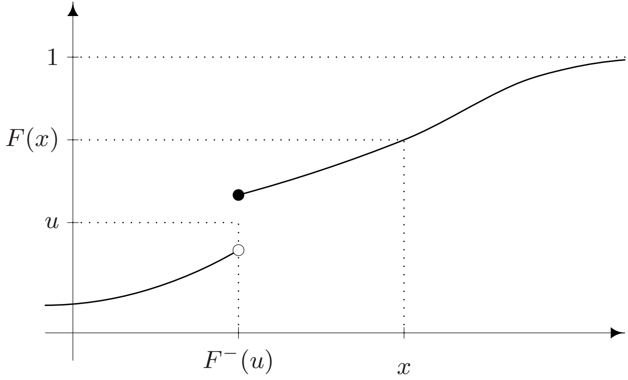


Figure: Cumulative distribution function $F$ and representation of the inverse cumulative distribution function.

# Inversion Method

- **Proposition**. Let $F$ be a cdf and $U \sim \mathcal{U}_{[0,1]}$. Then $X = F^{-}(U)$ has cdf $F$.

- In other words, to sample from a distribution with cdf $F$, we can sample $U \sim \mathcal{U}_{[0,1]}$ and then return $F^{-}(U)$.

- *Proof.* $F^{-}(u) \leq x \Leftrightarrow u \leq F(x)$ so for $U \sim \mathcal{U}_{[0,1]}$, we have

$$\mathbb{P}\left(F^{-}(U) \leq x\right) = \mathbb{P}(U \leq F(x)) = F(x).$$

# Examples

- **Exponential distribution**. If $F(x) = 1 - e^{-\lambda x}$, then $F^{-}(u) = F^{-1}(u) = -\log(1-u)/\lambda$.

  Thus when $U \sim \mathcal{U}_{[0,1]}$, $-\log(1-U)/\lambda \sim \mathcal{E}xp(\lambda)$ and $-\log(U)/\lambda \sim \mathcal{E}xp(\lambda)$.

- **Discrete distribution**. Assume $X$ takes values $x_1 < x_2 < \cdots$ with probability $p_1, p_2, \ldots$ so

$$F(x) = \sum_{x_k \leq x} p_k,$$

$$F^{-}(u) = x_k \text{ for } p_1 + \cdots + p_{k-1} < u \leq p_1 + \cdots + p_k.$$

# Transformation Method

- Let $Y \sim q$ be a $\mathbb{Y}$-valued random variable that we can simulate (e.g., by inversion)
- Let $X \sim \pi$ be $\mathbb{X}$-valued, which we wish to simulate.
- It may be that we can find a function $\varphi : \mathbb{Y} \to \mathbb{X}$ with the property that if we simulate $Y \sim q$ and then set $X = \varphi(Y)$ then we get $X \sim \pi$.
- Inversion is a special case of this idea.

# Transformation Method

- **Gamma distribution**. Let $Y_i$, $i = 1, 2, ..., \alpha$, be i.i.d. with $Y_i \sim \mathcal{E}xp\,(1)$ and $X = \beta^{-1} \sum_{i=1}^{\alpha} Y_i$ then $X \sim \mathcal{G}a\,(\alpha, \beta)$.

  *Proof.* The moment generating function of $X$ is

  $$\mathbb{E}\left(e^{tX}\right) = \prod_{i=1}^{\alpha} \mathbb{E}\left(e^{\beta^{-1} t Y_i}\right) = (1 - t/\beta)^{-\alpha}$$

  which is the MGF of the gamma density $\pi\,(x) \propto x^{\alpha-1} \exp\,(-\beta x)$ of parameters $\alpha, \beta$.

- **Beta distribution**. See Lecture Notes.

# Transformation Method - Box-Muller Algorithm

- **Gaussian distribution.** Let $U_1 \sim \mathcal{U}_{[0,1]}$ and $U_2 \sim \mathcal{U}_{[0,1]}$ be independent and set

$$R = \sqrt{-2\log\left(U_1\right)},\ \vartheta = 2\pi U_2.$$

We have

$$\begin{aligned} X &= R\cos\vartheta \sim \mathcal{N}\left(0,1\right), \\ Y &= R\sin\vartheta \sim \mathcal{N}\left(0,1\right). \end{aligned}$$

- Indeed $R^2 \sim \mathcal{E}xp\left(\frac{1}{2}\right)$ and $\vartheta \sim \mathcal{U}_{[0,2\pi]}$ so

$$q\left(r^2,\theta\right) = \frac{1}{2}\exp\left(-r^2/2\right)\frac{1}{2\pi}.$$

# Transformation Method - Box-Muller Algorithm

- Bijection:

$$\begin{aligned}(x,y) &= \left(\sqrt{r^2}\cos\theta, \sqrt{r^2}\sin\theta\right)\\ \Leftrightarrow \left(r^2,\theta\right) &= \left(x^2+y^2, \arctan\left(y/x\right)\right)\end{aligned}$$

so

$$\pi\left(x,y\right) = q\left(r^2,\theta\right)\left|\det\frac{\partial\left(r^2,\theta\right)}{\partial\left(x,y\right)}\right|$$

where

$$\left|\det\frac{\partial\left(r^2,\theta\right)}{\partial\left(x,y\right)}\right|^{-1} = \left|\det\begin{pmatrix}\frac{\cos\theta}{2r} & -r\sin\theta\\ \frac{\sin\theta}{2r} & r\cos\theta\end{pmatrix}\right| = \frac{1}{2}.$$

- Hence we have

$$\pi\left(x,y\right) = \frac{1}{2\pi}\exp\left(-\left(x^2+y^2\right)/2\right).$$

# Transformation Method - Multivariate Normal

- Let $Z = (Z_1, ..., Z_d) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. Let $L$ be a real invertible $d \times d$ matrix satisfying $L\ L^T = \Sigma$, and $X = LZ + \mu$. Then $X \sim \mathcal{N}(\mu, \Sigma)$.
- We have indeed $q(z) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} z^T z\right)$ and

$$\pi(x) = q(z) \left|\det \partial z / \partial x\right|$$

where $\partial z / \partial x = L^{-1}$ and $\det\left(L^{-1}\right) = \det(\Sigma)^{-1/2}$. Additionally,

$$\begin{aligned} z^T z &= (x - \mu)^T \left(L^{-1}\right)^T L^{-1} (x - \mu) \\ &= (x - \mu)^T \Sigma^{-1} (x - \mu). \end{aligned}$$

- In practice, use a Cholesky factorization $\Sigma = L\ L^T$ where $L$ is a lower triangular matrix.

# Sampling via Composition

- Assume we have a joint pdf $\overline{\pi}$ with marginal $\pi$; i.e.

$$\pi(x) = \int \overline{\pi}_{X,Y}(x,y)\,dy$$

where $\overline{\pi}(x,y)$ can always be decomposed as

$$\overline{\pi}_{X,Y}(x,y) = \overline{\pi}_Y(y)\,\overline{\pi}_{X|Y}(x|y).$$

- It might be easy to sample from $\overline{\pi}(x,y)$ whereas it is difficult/impossible to compute $\pi(x)$.
- In this case, it is sufficient to sample

$$Y \sim \overline{\pi}_Y \text{ then } X|Y \sim \overline{\pi}_{X|Y}(\cdot|Y)$$

so $(X,Y) \sim \overline{\pi}_{X,Y}$ and hence $X \sim \pi$.

# Finite Mixture of Distributions

- Assume one wants to sample from

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

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

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

$$\overline{\pi}_{X,Y}(x, y) = \alpha_y \times \pi_y(x).$$

- To sample from $\pi(x)$, first sample $Y$ from a discrete distribution such that $\mathbb{P}(Y = k) = \alpha_k$ then

$$X | (Y = y) \sim \pi_y.$$

# Rejection Sampling

**Basic idea**: Sample from a proposal $q$ different from the target $\pi$ and correct through rejection step to obtain a sample from $\pi$.

**Algorithm (Rejection Sampling)**. Given two densities $\pi, q$ with $\pi(x) \leq M\ q(x)$ for all $x$, we can generate a sample from $\pi$ by

1. Draw $X \sim q$, draw $U \sim \mathcal{U}_{[0,1]}$.
2. Accept $X = x$ as a sample from $\pi$ if

$$U \leq \frac{\pi(x)}{M\ q(x)},$$

otherwise go to step 1.

# Rejection Sampling

- **Proposition**. The distribution of the samples accepted by rejection sampling is $\pi$.

  *Proof*. We have for any (measurable) set $A$

$$\mathbb{P}\left( X \in A \middle| X \text{ accepted}\right) = \frac{\mathbb{P}\left(X \in A, X \text{ accepted}\right)}{\mathbb{P}\left(X \text{ accepted}\right)}$$

  where

$$\begin{aligned}\mathbb{P}\left(X \in A, X \text{ accepted}\right) &= \int_{\mathbb{X}} \int_0^1 \mathbb{I}_A\left(x\right) \mathbb{I}\left(u \leq \frac{\pi\left(x\right)}{M\ q\left(x\right)}\right) q\left(x\right) du dx \\ &= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M\ q\left(x\right)} q\left(x\right) dx \\ &= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M} dx = \frac{\pi\left(A\right)}{M}.\end{aligned}$$

# Rejection Sampling

So

$$\mathbb{P}\left(X \text{ accepted}\right) = \mathbb{P}\left(X \in \mathbb{X}, X \text{ accepted}\right) = \frac{\pi\left(\mathbb{X}\right)}{M} = \frac{1}{M}$$

and

$$\mathbb{P}\left( X \in A \middle| X \text{ accepted}\right) = \pi\left(A\right).$$

- Rejection sampling produces samples from $\pi$. It requires to be able to evaluate the density of $\pi$ point-wise, and an upper bound $M$ on $\pi(x)/q(x)$.

# Rejection Sampling

- In most practical scenarios, we only know $\pi$ and $q$ up to some normalising constants; i.e.

$$\pi = \widetilde{\pi}/Z_\pi \text{ and } q = \tilde{q}/Z_q$$

where $\widetilde{\pi}, \tilde{q}$ are known but $Z_\pi = \int_{\mathbb{X}} \widetilde{\pi}(x)\, dx$, $Z_q = \int_{\mathbb{X}} \tilde{q}(x)\, dx$ are unknown.

- If $Z_\pi, Z_q$ are unknown but you can upper bound:

$$\widetilde{\pi}(x)/\widetilde{q}(x) \leq \widetilde{M},$$

then using $\widetilde{\pi}$, $\widetilde{q}$ and $\widetilde{M}$ in the algorithm is correct.

- Indeed we have

$$\frac{\widetilde{\pi}(x)}{\widetilde{q}(x)} \leq \widetilde{M} \Leftrightarrow \frac{\pi(x)}{q(x)} \leq \widetilde{M}\frac{Z_q}{Z_\pi} = M.$$

# Rejection Sampling

- Let $T$ denote the number of pairs $(X, U)$ that have to be generated until $X$ is accepted for the first time.
- **Lemma**. $T$ is geometrically distributed with parameter $1/M$ and in particular $\mathbb{E}(T) = M$.
- In the unnormalised case, this yields

$$\mathbb{P}(X \text{ accepted}) = \frac{1}{M} = \frac{Z_\pi}{\widetilde{M} Z_q},$$

$$\mathbb{E}(T) = M = \frac{Z_q \widetilde{M}}{Z_\pi},$$

and it can be used to provide unbiased estimates of $Z_\pi / Z_q$ and $Z_q / Z_\pi$.

# Examples

- **Uniform density on a bounded subset of** $\mathbb{R}^p$**.**
  Consider the problem of sampling uniformly over $B \subset \mathbb{R}^p$, a bounded subset of $\mathbb{R}^p$:

$$\pi(x) \propto \mathbb{I}_B(x).$$

  Let $R$ be a rectangle with $B \subset R$ and

$$q(x) \propto \mathbb{I}_R(x).$$

- Then we can use $\widetilde{M} = 1$ and

$$\widetilde{\pi}(x) / \left(\widetilde{M}' \widetilde{q}(x)\right) = \mathbb{I}_B(x).$$

- The probability of accepting a sample is then $Z_\pi / Z_q$.

# Examples

- **Normal density**. Let $\widetilde{\pi}(x) = \exp\left(-\frac{1}{2}x^2\right)$ and $\widetilde{q}(x) = 1/\left(1+x^2\right)$. We have

$$\frac{\widetilde{\pi}(x)}{\widetilde{q}(x)} = \left(1+x^2\right)\exp\left(-\frac{1}{2}x^2\right) \leq 2/\sqrt{e} = \widetilde{M}$$

which is attained at $\pm 1$. The acceptance probability is

$$\mathbb{P}\left(U \leq \frac{\widetilde{\pi}(X)}{\widetilde{M}\widetilde{q}(X)}\right) = \frac{Z_\pi}{\widetilde{M} Z_q} = \frac{\sqrt{2\pi}}{\frac{2}{\sqrt{e}}\pi} = \sqrt{\frac{e}{2\pi}} \approx 0.66,$$

and the mean number of trials to success is approximately $1/0.66 \approx 1.52$.

# Examples: Genetic linkage model

- We observe

$$(Y_1, Y_2, Y_3, Y_4) \sim \mathcal{M}\left(n; \frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$$

  where $\mathcal{M}$ is the multinomial distribution and $\theta \in (0,1)$.
- The likelihood of the observations is thus

$$\begin{aligned} & p(y_1, ..., y_4; \theta) \\ & = \frac{n!}{y_1! y_2! y_3! y_4!} \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \left(\frac{1}{4}(1-\theta)\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4} \\ & \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}. \end{aligned}$$

- Bayesian approach where we select $p(\theta) = \mathbb{I}_{[0,1]}(\theta)$ and are interested in

$$p(\theta | y_1, ..., y_4) \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4} \mathbb{I}_{[0,1]}(\theta).$$

# Examples: Genetic linkage model

- Rejection sampling using a proposal $q(\theta) = \widetilde{q}(\theta) = p(\theta)$ to sample from $p(\theta | y_1, ..., y_4)$.

- To use accept-reject, we need to upper bound

$$\frac{\widetilde{\pi}(\theta)}{\widetilde{q}(\theta)} = \widetilde{\pi}(\theta) = (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}$$

- Maximum of $\widetilde{\pi}$ can be found using standard optimization procedure to perform rejection sampling. For a realisation of $(Y_1, Y_2, Y_3, Y_4)$ equal to $(69, 9, 11, 11)$ obtained with $n = 100$ and $\theta^\star = 0.6$, results shown in following figure.

# Examples: Genetic linkage model

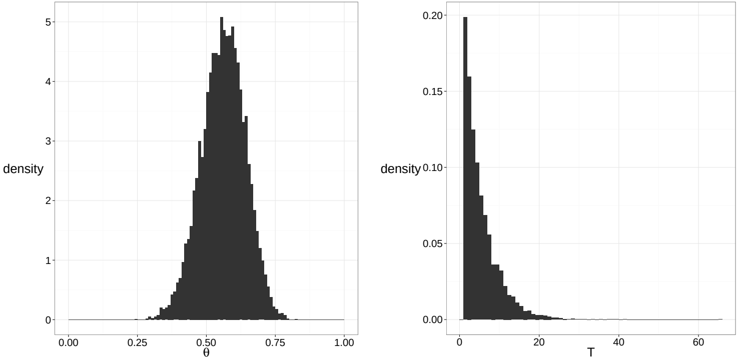


Figure: Histogram of 10,000 samples drawn from posterior obtained by rejection sampling (left); and histogram of waiting time distribution before acceptance (right).