Lecture 3: Rejection Sampling

Deniz Akyildiz

MATH60047/70047 – Stochastic Simulation

October 17, 2022

Imperial College London

Announcements

- Next week Tuesday, 25th of October, we are at
  - HXLY 414 - Mathematics Learning Centre

Announcements

Imperial College London

- Next week Tuesday, 25th of October, we are at
  - HXLY 414 - Mathematics Learning Centre
- Exercises 1 posted. Solutions and the second set after tomorrow.

Announcements

- Next week Tuesday, 25th of October, we are at
  - HXLY 414 - Mathematics Learning Centre
- Exercises 1 posted. Solutions and the second set after tomorrow.
- Python
  - Tell us if you cannot run Python (email me)

Announcements

- Next week Tuesday, 25th of October, we are at
  - HXLY 414 - Mathematics Learning Centre
- Exercises 1 posted. Solutions and the second set after tomorrow.
- Python
  - Tell us if you cannot run Python (email me)
- Assignment
  - We will require a report written in Latex with figures (page limit to be announced with the assignment)

Lecture summary

Recap

The Fundamental Theorem of Simulation

Rejection Sampling

Recap: We have seen

Recap: We have seen

- Uniform random variate generation

Recap: We have seen

- Uniform random variate generation
- Direct sampling from variety of distributions

Recap: We have seen

- Uniform random variate generation
- Direct sampling from variety of distributions
  - Inversion method
    - Draw $U \sim \text{Unif}(0, 1)$
    - Compute $X = F^{-}(U) = \min\{x : F_X(x) \geq u\}$

Recap: We have seen

- Uniform random variate generation
- Direct sampling from variety of distributions
  - Inversion method
    - Draw $U \sim \text{Unif}(0, 1)$
    - Compute $X = F^{-}(U) = \min\{x : F_X(x) \geq u\}$
  - Transformation method.
    - Draw $U \sim \text{Unif}(0, 1)$,
    - Obtain $X = g(U)$ for some general transformation g.

Recap: We have seen

- Uniform random variate generation
- Direct sampling from variety of distributions
  - Inversion method
    - Draw $U \sim \text{Unif}(0, 1)$
    - Compute $X = F^{-}(U) = \min\{x : F_X(x) \geq u\}$
  - Transformation method.
    - Draw $U \sim \text{Unif}(0, 1)$,
    - Obtain $X = g(U)$ for some general transformation g.

However, those methods required a quite specific structure for us to be able to sample.

What if inverse of CDF or a nice transformation is not available?

What if inverse of CDF or a nice transformation is not available?

What if we cannot evaluate $p(x)$ – only evaluate an unnormalised density $\bar{p}(x)$

What if inverse of CDF or a nice transformation is not available?

What if we cannot evaluate $p(x)$ – only evaluate an unnormalised density $\bar{p}(x)$

Can we still do *exact* sampling?

The Fundamental Theorem of Simulation

Is there a more general structure?

Theorem 1 (Theorem 2.2, Martino et al., 2018)

Drawing samples from one dimensional random variable X with a density $p(x)$ is equivalent to sampling uniformly on the two dimensional region defined by

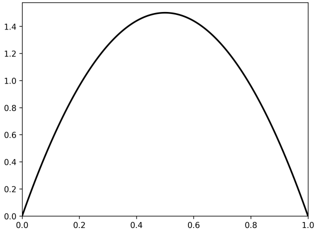
$$\mathsf{A} = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq p(x)\}. \tag{1}$$

In other words, if (x', y') is uniformly distributed on A*, then x' is a sample from $p(x)$.*

Let

$$p(x) = \mathsf{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha - 1}(1 - x)^{\beta - 1},$$

where $\Gamma(n) = (n-1)!$ for integers. For Beta$(2, 2)$:



Its maximum is 1.5 in this specific case. Can we sample uniformly?

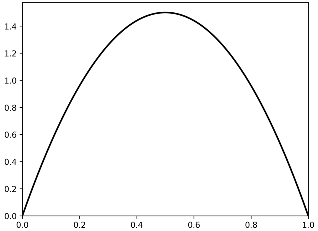
Sampling Beta density

Testing the theorem

Let

$$p(x) = \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

where $\Gamma(n) = (n-1)!$ for integers. For $\text{Beta}(2,2)$:



Its maximum is 1.5 in this specific case. Can we sample uniformly?

- Sample from the box $[0,1] \times [0,1.5]$ and keep the ones inside.

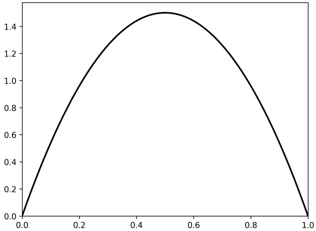
Sampling Beta density

Testing the theorem

Let

$$p(x) = \text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha - 1} (1 - x)^{\beta - 1},$$

where $\Gamma(n) = (n-1)!$ for integers. For $\text{Beta}(2, 2)$:

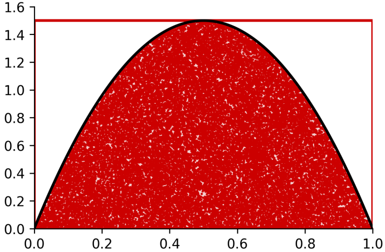


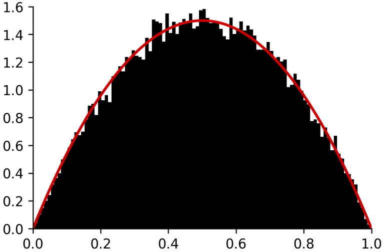
Its maximum is 1.5 in this specific case. Can we sample uniformly?

- Sample from the box $[0, 1] \times [0, 1.5]$ and keep the ones inside.
- Note though our aim is to 'test the x-marginal'

Sampling Beta density

Testing the theorem





Note the key aspects:

Note the key aspects:

- We draw from the *uniform* but we do not have to

Note the key aspects:

- We draw from the *uniform* but we do not have to
- We have not used any inverse CDF or transformation but

Note the key aspects:

- We draw from the *uniform* but we do not have to
- We have not used any inverse CDF or transformation but
  - We have used the expression of $p(x)$

Note the key aspects:

- We draw from the *uniform* but we do not have to
- We have not used any inverse CDF or transformation but
    - We have used the expression of $p(x)$

We will see that we can get away with an unnormalised expression of $p(x)$ (later)

Rejection sampling

More than a box

Using a box wrapping the density is very inefficient:

Rejection sampling

More than a box

Using a box wrapping the density is very inefficient:

- You need the maximum of the density

$$p^\star = \max p(x)$$

which could be as hard as the sampling problem!

Using a box wrapping the density is very inefficient:

- You need the maximum of the density

$$p^\star = \max p(x)$$

  which could be as hard as the sampling problem!
- For densities that are peaky, this could be wildly inefficient

Rejection sampling

More than a box

Using a box wrapping the density is very inefficient:

- You need the maximum of the density

$$p^\star = \max p(x)$$

  which could be as hard as the sampling problem!
- For densities that are peaky, this could be wildly inefficient

Idea: Design a *proposal* density that tightly wraps the target density

Rejection sampling

Choice of the proposal

Consider a (target) density $p(x)$ and a *proposal* density $q(x)$.

Rejection sampling

Choice of the proposal

Consider a (target) density $p(x)$ and a *proposal* density $q(x)$.

For rejection sampling, we always choose a proposal such that

$$p(x) \leq Mq(x),$$

for $M \geq 1$.

Consider a (target) density $p(x)$ and a *proposal* density $q(x)$.

For rejection sampling, we always choose a proposal such that

$$p(x) \leq Mq(x),$$

for $M \geq 1$. Intuitively, the $Mq(x)$ curve should be **above** $p(x)$.

Imperial College London

The rejection sampler:

The rejection sampler:

- $x' \sim q(x)$,

The rejection sampler:

- $x' \sim q(x)$,
- Accept the sample x' with probability

$$a(x') = \frac{p(x')}{Mq(x')} \leq 1.$$

The rejection sampler:

- $x' \sim q(x)$,
- Accept the sample x' with probability

$$a(x') = \frac{p(x')}{Mq(x')} \leq 1.$$

How does this relate to the Fundamental Theorem of Simulation?

Rejection sampling

The algorithm: A closer look

To implement the method:

- Sample $x' \sim q(x)$,

Rejection sampling

The algorithm: A closer look

To implement the method:

- Sample $x' \sim q(x)$,
- Sample $u \sim \text{Unif}(u; 0, 1)$

To implement the method:

- Sample $x' \sim q(x)$,
- Sample $u \sim \mathsf{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

Rejection sampling

The algorithm: A closer look

To implement the method:

- Sample $x' \sim q(x)$,
- Sample $u \sim \text{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

This is how it is generally implemented!

Rejection sampling

The algorithm: A closer look

But how to show that we obtain a sample under the curve?

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

Rejection sampling

The algorithm: A closer look

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

- Sample $u \sim \text{Unif}(u; 0, 1)$

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

- Sample $u \sim \mathrm{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

- Sample $u \sim \text{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

An equivalent way to implement it is

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

- Sample $u \sim \text{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

An equivalent way to implement it is

- Sample $u' \sim \text{Unif}(u'; 0, Mq(x'))$

But how to show that we obtain a sample under the curve?

Note that after $x' \sim q(x)$, the acceptance step goes as

- Sample $u \sim \mathsf{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{p(x')}{Mq(x')}.$$

An equivalent way to implement it is

- Sample $u' \sim \mathsf{Unif}(u'; 0, Mq(x'))$
- Accept

$$u' \leq p(x'),$$

This would give us (x', u') uniformly under the curve! (hence x' samples would be distributed w.r.t. $p(x)$)

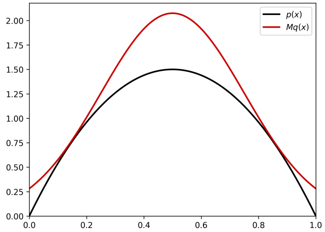
Rejection sampling

Examples: Same Beta$(2, 2)$, better proposal

Choose

$$q(x) = \mathcal{N}(0.5, 0.25),$$

with $M = 1.3$.



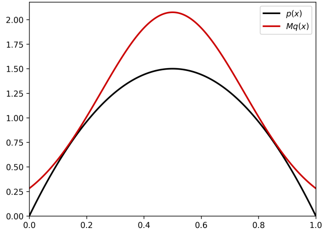
Rejection sampling

Examples: Same Beta$(2, 2)$, better proposal

Choose

$$q(x) = \mathcal{N}(0.5, 0.25),$$

with $M = 1.3$.



Simulation.

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

Luckily, we can evaluate $p(x)$ up to a normalising constant:

$$p(x) = \frac{\bar{p}(x)}{Z}.$$

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

Luckily, we can evaluate $p(x)$ up to a normalising constant:

$$p(x) = \frac{\bar{p}(x)}{Z}.$$

Note the terminology and convention:

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

Luckily, we can evaluate $p(x)$ up to a normalising constant:

$$p(x) = \frac{\bar{p}(x)}{Z}.$$

Note the terminology and convention:

- $\bar{p}(x)$ is called the *unnormalised* density

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

Luckily, we can evaluate $p(x)$ up to a normalising constant:

$$p(x) = \frac{\bar{p}(x)}{Z}.$$

Note the terminology and convention:

- $\bar{p}(x)$ is called the *unnormalised* density
- Z is called the normalising constant
    - It is a super important quantity for many other purposes

We have seen that rejection sampling works for general densities but requires the evaluation of $p(x)$.

In many (many) cases, we cannot evaluate $p(x)$!

Luckily, we can evaluate $p(x)$ up to a normalising constant:

$$p(x) = \frac{\bar{p}(x)}{Z}.$$

Note the terminology and convention:

- $\bar{p}(x)$ is called the *unnormalised* density
- Z is called the normalising constant
  - It is a super important quantity for many other purposes
- We write $p(x) \propto \bar{p}(x)$ to say p is proportional to $\bar{p}(x)$ but normalised to integrate (or sum) to one.

The fundamental theorem of simulation is better than it looks. It works with unnormalised densities (recap, unnormalised version)

Theorem 2 (Theorem 2.2, Martino et al., 2018)

Drawing samples from one dimensional random variable X with a density $p(x) \propto \bar{p}(x)$ is equivalent to sampling uniformly on the two dimensional region defined by

$$\mathsf{A} = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq \bar{p}(x)\}. \tag{2}$$

In other words, if (x', y') is uniformly distributed on A*, then x' is a sample from $p(x)$.*

To implement, choose M and q such that $\bar{p}(x) \leq Mq(x)$

- Sample $x' \sim q(x)$,

Rejection sampling

The algorithm: A closer look

To implement, choose M and q such that $\bar{p}(x) \leq Mq(x)$

- Sample $x' \sim q(x)$,
- Sample $u \sim \text{Unif}(u; 0, 1)$

Rejection sampling

The algorithm: A closer look

To implement, choose M and q such that $\bar{p}(x) \leq Mq(x)$

- Sample $x' \sim q(x)$,
- Sample $u \sim \text{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{\bar{p}(x')}{Mq(x')}.$$

Rejection sampling

The algorithm: A closer look

To implement, choose M and q such that $\bar{p}(x) \leq Mq(x)$

- Sample $x' \sim q(x)$,
- Sample $u \sim \text{Unif}(u; 0, 1)$
- Accept if

$$u \leq \frac{\bar{p}(x')}{Mq(x')}.$$

Exactly same – $\bar{p}$ used instead of p provided that $\bar{p}(x) \leq Mq(x)$

Unnormalised densities

Why would you have them?

Easy to imagine the discrete case: Imagine you want to obtain the probability of observing something discrete

- People with black jumpers: 530
- People with red jumpers: 403
- People with yellow jumpers: 304

In order to talk about 'probability of seeing a black jumper', you'd normalise this numbers (normalisation is easy to compute here).

Unnormalised densities

Why would you have them?

In physics, engineering, and even optimisation, we do not start from densities, instead one defines:

Unnormalised densities

Why would you have them?

In physics, engineering, and even optimisation, we do not start from densities, instead one defines:

$$p(x) \propto e^{-f(x)},$$

for some function f (which is generally called a *potential*).

Unnormalised densities

Why would you have them?

In physics, engineering, and even optimisation, we do not start from densities, instead one defines:

$$p(x) \propto e^{-f(x)},$$

for some function f (which is generally called a *potential*).

f usually comes from a rule which determines how probability mass should be spread . Ising model:

$$p_\beta(\sigma) = \frac{e^{-\beta H(\sigma)}}{Z_\beta},$$

where $H(\sigma)$ defines the energy of a spin configuration of a magnetic material. The normalising constant is

$$Z_\beta(\sigma) = \sum_\sigma e^{-\beta H(\sigma)}.$$

All configurations for atomic spins!

Unnormalised densities

Why would you have them?

In Bayesian statistics (what we care about), we are given

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

Unnormalised densities

Why would you have them?

In Bayesian statistics (what we care about), we are given

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

- $p(y|x)$ is the likelihood of data point y (we know)

Unnormalised densities

Why would you have them?

In Bayesian statistics (what we care about), we are given

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

- $p(y|x)$ is the likelihood of data point y (we know)
- $p(x)$ is the prior of x (we know)

Unnormalised densities

Why would you have them?

In Bayesian statistics (what we care about), we are given

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

- $p(y|x)$ is the likelihood of data point y (we know)
- $p(x)$ is the prior of x (we know)
- $p(y)$ is the normalising constant

$$p(y) = \int p(y|x)p(x)\mathrm{d}x,$$

we do not know.

Unnormalised densities

Why would you have them?

In Bayesian statistics (what we care about), we are given

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}.$$

- $p(y|x)$ is the likelihood of data point y (we know)
- $p(x)$ is the prior of x (we know)
- $p(y)$ is the normalising constant

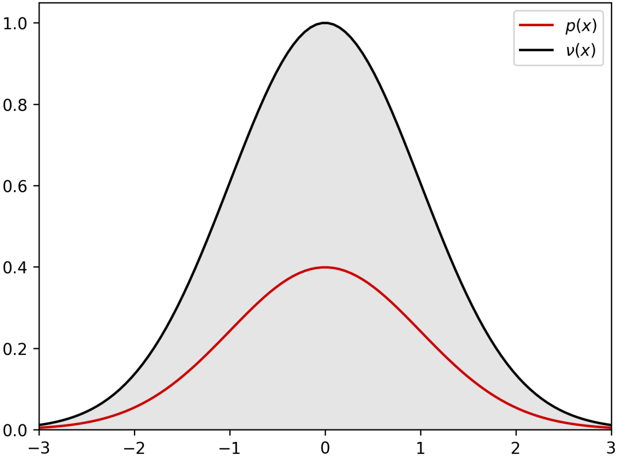
  $$p(y) = \int p(y|x)p(x)\mathrm{d}x,$$

  we do not know.

Therefore, we can evaluate the posterior density only in unnormalised way:

$$p(x|y) \propto p(y|x)p(x).$$

An example of an unnormalised density.

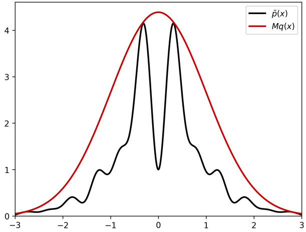


Consider (Robert and Casella, 2004)

$$p(x) \propto \exp(-x^2/2)(\sin^2(6x) + 3\cos^2(x)\sin^2 4x + 1)$$

and

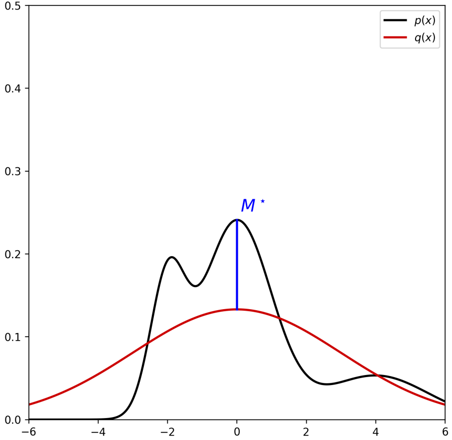
$$q(x) = \mathcal{N}(0,1) = \exp(-x^2/2)/\sqrt{2\pi}.$$



$M = 11.$

A standard choice for M is

$$M^\star = \sup_x \frac{p(x)}{q(x)}.$$

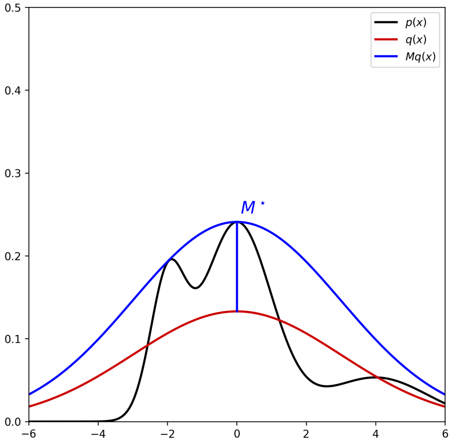


Rejection sampling

Choice of M

A standard choice for M is

$$M^\star = \sup_x \frac{p(x)}{q(x)}.$$



An important quantity for a rejection sampler is *the acceptance rate*. For the normalised case, it can be analytically derived:

$$\hat{a} = \frac{1}{M}.$$

For the unnormalised case, similarly,

$$\hat{a} = \frac{Z}{M}.$$

Imperial College London

An important quantity for a rejection sampler is *the acceptance rate*. For the normalised case, it can be analytically derived:

$$\hat{a} = \frac{1}{M}.$$

For the unnormalised case, similarly,

$$\hat{a} = \frac{Z}{M}.$$

We often would like to optimise our proposal q, such that the acceptance rate is maximised. $\implies$ Minimize M!

Assume that we would like to sample from

$$X \sim \Gamma(\alpha, 1),$$

for $\alpha > 1$. The density is given by

$$p(x) = \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, \qquad \text{for } x > 0,$$

where $\Gamma(\alpha)$ is the Gamma function. $\Gamma(n) = (n-1)!$

Choose as a *proposal*:

$$q_\lambda(x) = \mathsf{Exp}(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0,$$

with $0 < \lambda < 1$.

Choose as a *proposal*:

$$q_\lambda(x) = \mathsf{Exp}(x;\lambda) = \lambda e^{-\lambda x}, \quad x > 0,$$

with $0 < \lambda < 1$.How do we ensure that $p(x) \leq Mq(x)$?

Choose as a *proposal*:

$$q_\lambda(x) = \mathsf{Exp}(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0,$$

with $0 < \lambda < 1$.How do we ensure that $p(x) \leq Mq(x)$? Choose

$$M_\lambda = \sup_x \frac{p(x)}{q_\lambda(x)}.$$

Find M_λ for fixed λ first:

$$\frac{p(x)}{q_\lambda(x)} = \frac{x^{\alpha-1} e^{(\lambda-1)x}}{\lambda \Gamma(\alpha)}.$$

Choose as a *proposal*:

$$q_\lambda(x) = \mathsf{Exp}(x; \lambda) = \lambda e^{-\lambda x}, \quad x > 0,$$

with $0 < \lambda < 1$.How do we ensure that $p(x) \leq Mq(x)$? Choose

$$M_\lambda = \sup_x \frac{p(x)}{q_\lambda(x)}.$$

Find M_λ for fixed λ first:

$$\frac{p(x)}{q_\lambda(x)} = \frac{x^{\alpha-1} e^{(\lambda-1)x}}{\lambda \Gamma(\alpha)}.$$

Maximise this w.r.t. x?

How to compute a maximum? It is useful to take $\log$:

$$\arg\max_x f(x) = \arg\max_x \log f(x).$$

Take the $\log$ of

$$\frac{p(x)}{q_\lambda(x)} = \frac{x^{\alpha-1}e^{(\lambda-1)x}}{\lambda\Gamma(\alpha)}.$$

So we want to optimise

$$G(x) = \log\frac{p(x)}{q_\lambda(x)} = (\alpha-1)\log x + (\lambda-1)x - \log\lambda\Gamma(\alpha).$$

Set $\frac{\mathrm{d}G(x)}{\mathrm{d}x} = 0$.

The derivative

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + (\lambda - 1) = 0$$

which implies

$$x^\star = \frac{\alpha - 1}{1 - \lambda}.$$

Imperial College London

The derivative

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + (\lambda - 1) = 0$$

which implies

$$x^\star = \frac{\alpha - 1}{1 - \lambda}.$$

How do we understand if this is a maximum?

The derivative

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + (\lambda - 1) = 0$$

which implies

$$x^{\star} = \frac{\alpha - 1}{1 - \lambda}.$$

How do we understand if this is a maximum? Compute

$$\frac{\mathrm{d}G^2(x)}{\mathrm{d}^2 x} = -\frac{\alpha - 1}{x^2},$$

plug $x^{\star}$ into this

$$\frac{\mathrm{d}^2 G(x^{\star})}{\mathrm{d}x^2} = -\frac{(\alpha - 1)(1 - \lambda)^2}{(\alpha - 1)^2} < 0,$$

as $\alpha > 1$ and $0 < \lambda < 1$.

Therefore,

$$
\begin{aligned}
M_\lambda &= \frac{p(x^\star)}{q_\lambda(x^\star)}, \\
&= \frac{{x^\star}^{\alpha-1} e^{(\lambda-1)x^\star}}{\lambda\Gamma(\alpha)}, \\
&= \frac{\left(\frac{\alpha-1}{1-\lambda}\right)^{\alpha-1} e^{(\lambda-1)\frac{\alpha-1}{1-\lambda}}}{\lambda\Gamma(\alpha)} \\
&= \frac{\left(\frac{\alpha-1}{1-\lambda}\right)^{\alpha-1} e^{-(\alpha-1)}}{\lambda\Gamma(\alpha)}.
\end{aligned}
$$

Imperial College London

Recall that, we are interested in the acceptance probability (or maximising it)

$$\frac{p(x)}{M_\lambda q_\lambda(x)} = \left(\frac{x(1-\lambda)}{\alpha - 1} \right)^{\alpha - 1} e^{(\lambda - 1)x + \alpha - 1}.$$

Now, the task is to minimise M_λ w.r.t. λ so we get the *optimal* proposal ($\hat{a} = 1/M_\lambda$ would be maximised).

Recall

$$M_\lambda = \frac{\left(\frac{\alpha-1}{1-\lambda}\right)^{\alpha-1} e^{-(\alpha-1)}}{\lambda\Gamma(\alpha)}.$$

Compute

$$\begin{aligned}\log M_\lambda = {} & (\alpha-1)\log(\alpha-1) - (\alpha-1)\log(1-\lambda) \\ & - (\alpha-1) - \log\lambda - \log\Gamma(\alpha).\end{aligned}$$

$$\frac{\mathrm{d}\log M_\lambda}{\mathrm{d}\lambda} = \frac{\alpha - 1}{1-\lambda} - \frac{1}{\lambda} = 0,$$

which implies

$$\lambda\alpha - \lambda = 1 - \lambda,$$

therefore

$$\lambda^\star = \frac{1}{\alpha}.$$

Finally we get the optimal M by computing

$$M_{\lambda^\star} = \frac{\alpha^\alpha e^{-(\alpha-1)}}{\Gamma(\alpha)}.$$

In order to sample from $\Gamma(\alpha, 1)$, we perform

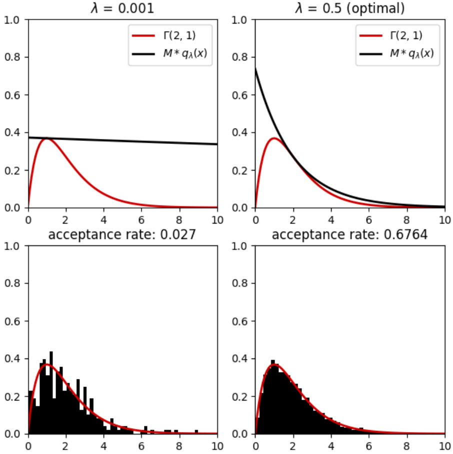
- Sample $X' \sim \mathsf{Exp}(1/\alpha)$ and $U \sim \mathsf{Unif}(0, 1)$
- If

$$U \leq (x/\alpha)^{\alpha-1} e^{(1/\alpha-1)x+\alpha-1},$$

accept X', otherwise start again.

Rejection sampling

Example: Optimising rejection sampling



Rejection Sampling

Example: Sampling truncated distributions

Given $\mathcal{N}(x; 0, 1)$, suppose we are interested in sampling this density between $[-a, a]$. We can write this truncated normal density as

$$p(x) = \frac{\bar{p}(x)}{Z} = \frac{\mathcal{N}(x; 0, 1)\mathbf{1}_{\{x:|x|\leq a\}}(x)}{\int_{-a}^{a} \mathcal{N}(y; 0, 1)\mathrm{d}y}.$$

We can choose $q(x) = \mathcal{N}(x; 0, 1)$ anyway, and we have $\bar{p}(x) \leq q(x)$ (i.e. we can take $M = 1$). The resulting algorithm is extremely intuitive: All you need is to sample from $q(x) = \mathcal{N}(x; 0, 1)$ and reject if this sample is out of bounds $[-a, a]$.

Rejection Sampling

Example: Beta$(2, 2)$

$$p(x) = \mathsf{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

Choose

$$\bar{p}(x) = x^{\alpha-1}(1-x)^{\beta-1},$$

Choose

$$q(x) = \mathsf{Unif}(0, 1)$$

What is

$$M = \sup_x \frac{\bar{p}(x)}{q(x)}$$

Rejection Sampling

Example: Beta$(2, 2)$

Compute

$$\log \bar{p}(x)/q(x) = (\alpha - 1)\log x + (\beta - 1)\log(1 - x)$$

Compute

$$\log \bar{p}(x)/q(x) = (\alpha - 1)\log x + (\beta - 1)\log(1 - x)$$

The derivative

$$\frac{\mathrm{d}\log \bar{p}(x)/q(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + \frac{1 - \beta}{1 - x}$$

Compute

$$\log \bar{p}(x)/q(x) = (\alpha - 1)\log x + (\beta - 1)\log(1 - x)$$

The derivative

$$\frac{\mathrm{d}\log \bar{p}(x)/q(x)}{\mathrm{d}x} = \frac{\alpha - 1}{x} + \frac{1 - \beta}{1 - x}$$

The maximum is

$$x^\star = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

Since

$$M = \frac{\bar{p}(x^\star)}{q(x^\star)}$$

We obtain

$$M = \frac{(\alpha-1)^{\alpha-1}(\beta-1)^{\beta-1}}{(\alpha+\beta-2)^{\alpha+\beta-2}}.$$

Rejection Sampling

Example: Beta$(2, 2)$

The algorithm:

- Sample $X' \sim q(x) = \text{Unif}(0, 1)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \bar{p}(X')/Mq(X')$,
  - Accept X'

See you tomorrow!

References I

- Martino, Luca, David Luengo, and Joaquín Míguez (2018). *Independent random sampling methods*. Springer.
- Robert, Christian P and George Casella (2004). *Monte Carlo statistical methods*. Springer.