Lecture 7: Probabilistic Inference and Conditional Independence

Deniz Akyildiz

MATH60047/70047 – Stochastic Simulation

October 31, 2022

Imperial College London

Turbulent flow split into two parts from a common source, white background, mix of dark red and black smoke

Announcements

Announcements

Imperial College London

- If you are submitting using Jupyter notebook:
  - Be sure to submit PDF (not .ipynb) – export PDF from Jupyter

Announcements

- If you are submitting using Jupyter notebook:
  - Be sure to submit PDF (not .ipynb) – export PDF from Jupyter
- If you are submitting otherwise, the page limit is 8 without text/code separation (put code to Appendix or inline)

Announcements

- If you are submitting using Jupyter notebook:
  - Be sure to submit PDF (not .ipynb) – export PDF from Jupyter
- If you are submitting otherwise, the page limit is 8 without text/code separation (put code to Appendix or inline)
- Submit anonymized (no names)

Announcements

- If you are submitting using Jupyter notebook:
  - Be sure to submit PDF (not .ipynb) – export PDF from Jupyter
- If you are submitting otherwise, the page limit is 8 without text/code separation (put code to Appendix or inline)
- Submit anonymized (no names)
- Exam contents

A brief recap of what happened:

Imperial College London

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)
- Rejection sampling

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)
- Rejection sampling
- Probabilistic Inference (introduction)

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)
- Rejection sampling
- Probabilistic Inference (introduction)
  - The notions of prior, likelihood, posterior

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)
- Rejection sampling
- Probabilistic Inference (introduction)
  - The notions of prior, likelihood, posterior
  - Bayes update (Bayesian Inference)

A brief recap of what happened:

- Direct sampling methods (Inversion, transformation)
- Rejection sampling
- Probabilistic Inference (introduction)
  - The notions of prior, likelihood, posterior
  - Bayes update (Bayesian Inference)

Today, we will talk about conditional independence.

Recall that, we consider the following model:

$$X \sim p(x)$$
$$Y|X = x \sim p(y|x),$$

where

- $p(x)$ is the prior distribution of X, Recall that, we consider the following model:

$$X \sim p(x)$$
$$Y|X = x \sim p(y|x),$$

where

- $p(x)$ is the prior distribution of X,
- $p(y|x)$ is the likelihood of Y given $X = x$,

Recap: Bayes Rule

Recall that, we consider the following model:

$$X \sim p(x)$$
$$Y|X = x \sim p(y|x),$$

where

- $p(x)$ is the prior distribution of X,
- $p(y|x)$ is the likelihood of Y given $X = x$,

In this context, our interest was to compute the posterior distribution for fixed (given) y:

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

We have also seen that, the posterior distribution is typically intractable (cannot be computed in closed form)

Recap: Bayes Rule

We have also seen that, the posterior distribution is typically intractable (cannot be computed in closed form)

However, we always know the unnormalised posterior

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

or equivalently, we can evaluate

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

Recap: Bayes Rule

We have also seen that, the posterior distribution is typically intractable (cannot be computed in closed form)

However, we always know the unnormalised posterior

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

or equivalently, we can evaluate

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

We can then use this to implement sampling algorithms.

Recap: Bayes Rule

We have also seen that, the posterior distribution is typically intractable (cannot be computed in closed form)

However, we always know the unnormalised posterior

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

or equivalently, we can evaluate

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

We can then use this to implement sampling algorithms.

- Rejection sampler

A tractable example is the Gamma-Poisson model.

A tractable example is the Gamma-Poisson model.

We have the prior defined as

$$p(x) = \mathsf{Gamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

A tractable example is the Gamma-Poisson model.

We have the prior defined as

$$p(x) = \mathsf{Gamma}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x).$$

and the likelihood is given as

$$p(y|x) = \mathsf{Poisson}(y; x) = \frac{x^y}{y!} \exp(-x).$$

This can be seen as an observation model of a count.

Derive the posterior.

Derive the posterior.

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

Derive the posterior.

$$
\begin{aligned}
p(x|y) &\propto p(y|x)p(x) \\
&= \frac{x^y}{y!}\exp(-x)\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x).
\end{aligned}
$$

Derive the posterior.

$$
\begin{aligned}
p(x|y) &\propto p(y|x)p(x) \\
&= \frac{x^y}{y!}\exp(-x)\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x).
\end{aligned}
$$

This is proportional to

$$p(x|y) \propto x^{y+\alpha-1}\exp(-x(\beta+1)).$$

Derive the posterior.

$$\begin{aligned} p(x|y) &\propto p(y|x)p(x) \\ &= \frac{x^y}{y!}\exp(-x)\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x). \end{aligned}$$

This is proportional to

$$p(x|y) \propto x^{y+\alpha-1}\exp(-x(\beta+1)).$$

This is another Gamma density, i.e.:

$$p(x|y) = \mathsf{Gamma}(x; \alpha + y, \beta + 1).$$

Let us see how we can use rejection sampling to sample from the posterior.

Let us see how we can use rejection sampling to sample from the posterior.

We will use the same model and compare the samples to the exact posterior we computed.

Let us see how we can use rejection sampling to sample from the posterior.

We will use the same model and compare the samples to the exact posterior we computed.

We choose a prior

$$p(x) = \mathsf{Gamma}(x; \alpha, 1) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x),$$

and a likelihood

$$p(y|x) = \mathsf{Poisson}(y; x) = \frac{x^y}{y!} \exp(-x).$$

Recall that our unnormalised posterior is written as

$$\bar{p}(x|y) = x^{y+\alpha-1} \exp(-2x).$$

Recall that our unnormalised posterior is written as

$$\bar{p}(x|y) = x^{y+\alpha-1} \exp(-2x).$$

Let us choose an exponential proposal:

$$q_\lambda(x) = \lambda \exp(-\lambda x).$$

Let us compute

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

Let us compute

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

The ratio is given by

$$\begin{aligned}\frac{\bar{p}(x|y)}{q_\lambda(x)} &= \frac{x^{\alpha-1+y}\, e^{-2x}}{\lambda e^{-\lambda x}}, \\ &= \frac{x^{\alpha-1+y}\, e^{-(2-\lambda)x}}{\lambda}.\end{aligned}$$

We aim at optimising this w.r.t. x, so first compute $\log$:

$$\log \frac{\bar{p}(x|y)}{q_\lambda(x)} = \log x^{\alpha-1+y} + \log e^{-(2-\lambda)x} - \log \lambda$$
$$= (\alpha - 1 + y) \log x - (2-\lambda)x - \log \lambda.$$

We aim at optimising this w.r.t. x, so first compute $\log$:

$$\begin{aligned}\log \frac{\bar{p}(x|y)}{q_\lambda(x)} &= \log x^{\alpha-1+y} + \log e^{-(2-\lambda)x} - \log \lambda \\ &= (\alpha - 1 + y)\log x - (2-\lambda)x - \log \lambda.\end{aligned}$$

We now take the derivative of this w.r.t. x:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[(\alpha - 1 + y)\log x - (2-\lambda)x - \log \lambda\right] = \frac{\alpha - 1 + y}{x} - (2-\lambda),$$

and set it to zero:

$$\frac{\alpha - 1 + y}{x} - (2 - \lambda) = 0.$$

This gives us the maximiser

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

We can now compute $M_\lambda = \bar{p}(x^\star|y)/q_\lambda(x^\star)$::

$$\begin{aligned}
M_\lambda &= \frac{\bar{p}(x^\star|y)}{q_\lambda(x^\star)} \\
&= \frac{x^{\star\alpha-1+y} e^{-(2-\lambda)x^\star}}{\lambda} \\
&= \frac{1}{\lambda}\left(\frac{\alpha-1+y}{2-\lambda}\right)^{\alpha-1+y} e^{-(2-\lambda)\left(\frac{\alpha-1+y}{2-\lambda}\right)} \\
&= \frac{1}{\lambda}\left(\frac{\alpha-1+y}{2-\lambda}\right)^{\alpha-1+y} e^{-(\alpha-1+y)}.
\end{aligned}$$

We can now optimise this further to choose our optimal proposal.

We will first compute the $\log$ of M_λ:

$$\begin{aligned}\log M_\lambda &= \log \frac{1}{\lambda} + (\alpha - 1 + y) \log \left(\frac{\alpha - 1 + y}{2 - \lambda} \right) - (\alpha - 1 + y) \\ &= -\log \lambda + (\alpha - 1 + y) \log \left(\frac{\alpha - 1 + y}{2 - \lambda} \right) - (\alpha - 1 + y).\end{aligned}$$

Taking the derivative of this w.r.t. λ, we obtain

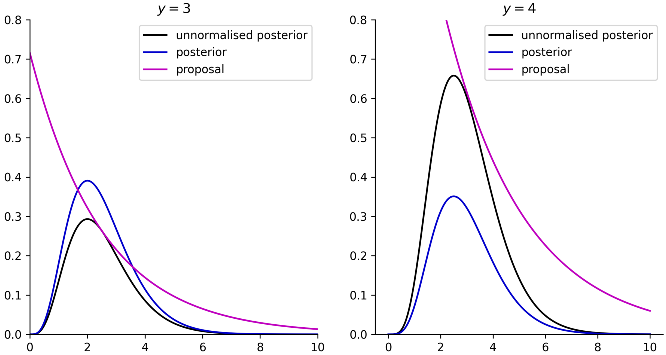
$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \log M_\lambda = -\frac{1}{\lambda} + \frac{(\alpha - 1 + y)}{2 - \lambda}$$

Setting this to zero, we obtain

$$\frac{1}{\lambda} = \frac{(\alpha - 1 + y)}{2 - \lambda},$$

which implies that

$$\lambda^\star = \frac{2}{\alpha + y}.$$


$y = 3$
unnormalised posterior
posterior
proposal
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
2
4
6
8
10
$y = 4$
unnormalised posterior
posterior
proposal

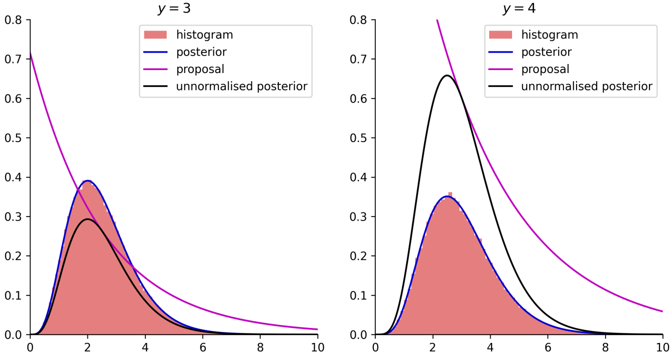


Figure: Histogram of the samples drawn using rejection sampling.

So far, we have only considered a single observation y and Bayes rule

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

So far, we have only considered a single observation y and Bayes rule

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

But what if we have multiple observations $y_1, \ldots, y_n$?

So far, we have only considered a single observation y and Bayes rule

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

But what if we have multiple observations $y_1, \ldots, y_n$?

In order to talk about multiple observations, we need to introduce conditional independence.

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

So far, we only considered two variable case: X and Y.

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

So far, we only considered two variable case: X and Y. However, the real world is more complex:

- Multiple variables (sometimes millions)

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

So far, we only considered two variable case: X and Y. However, the real world is more complex:

- Multiple variables (sometimes millions)
- Interactions between variables

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

So far, we only considered two variable case: X and Y. However, the real world is more complex:

- Multiple variables (sometimes millions)
- Interactions between variables
- Conditional dependence/independence

Conditional Independence

Before going into other sampling methods, we will also discuss the notion of conditional independence.

So far, we only considered two variable case: X and Y. However, the real world is more complex:

- Multiple variables (sometimes millions)
- Interactions between variables
- Conditional dependence/independence

We will learn how to deal with conditional independence.

Imperial College London

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

The definition is really simple: Usual independence *given* Z.

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

The definition is really simple: Usual independence *given* Z.

Take age, foot size, and literacy.

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

The definition is really simple: Usual independence *given* Z.

Take age, foot size, and literacy.

- X: Foot size, Y: Literacy, Z: Age

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

The definition is really simple: Usual independence *given* Z.

Take age, foot size, and literacy.

- X: Foot size, Y: Literacy, Z: Age
- X and Y are dependent: *bigger* foot size *tends* to be associated with *higher* literacy

Imperial College London

Definition 1 (Conditional Independence)

Let X and Y be random variables. We say that X and Y are conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z).$$

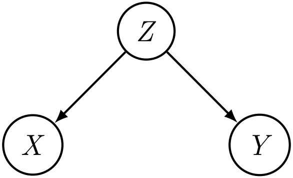
The definition is really simple: Usual independence *given* Z.

Take age, foot size, and literacy.

- X: Foot size, Y: Literacy, Z: Age
- X and Y are dependent: *bigger* foot size *tends* to be associated with *higher* literacy
- X and Y are independent given Z: Both *tend* to be driven by age.

Visualise Conditional Independence

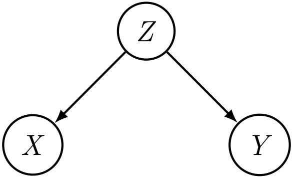
Example



- Z is sometimes called a *confounding* variable in statistics

Visualise Conditional Independence

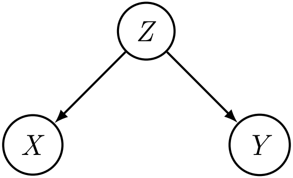
Example



- Z is sometimes called a *confounding* variable in statistics
- You may observe a spurious correlation between (X, Y), when in fact a *latent* variable Z is driving the correlation

Visualise Conditional Independence

Example

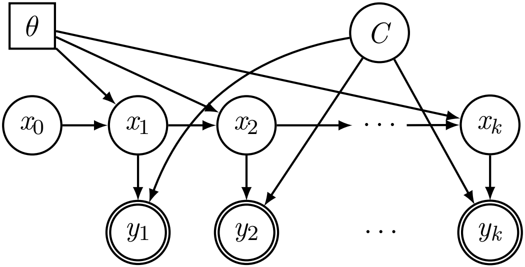


- Z is sometimes called a *confounding* variable in statistics
- You may observe a spurious correlation between (X, Y), when in fact a *latent* variable Z is driving the correlation

Small things to watch out while modelling.

Visualise Conditional Independence

Example



Conditional Independence

Conditionally independent observations

Our main interest is the case where we observe multiple random variables $Y_1, Y_2, \ldots, Y_n$ which are conditionally independent given a single latent variable.

Our main interest is the case where we observe multiple random variables $Y_1, Y_2, \ldots, Y_n$ which are conditionally independent given a single latent variable.

Assume that we have the following model

$$X \sim p(x)$$
$$Y_i | X = x \sim p(y_i | x), \quad i = 1, \ldots, n.$$

Note the difference:

- Previously, we simulated (X_i, Y_i) from $p(x, y)$

Conditional Independence

Conditionally independent observations

Our main interest is the case where we observe multiple random variables $Y_1, Y_2, \dots, Y_n$ which are conditionally independent given a single latent variable.

Assume that we have the following model

$$X \sim p(x)$$
$$Y_i | X = x \sim p(y_i | x), \quad i = 1, \dots, n.$$

Note the difference:

- Previously, we simulated (X_i, Y_i) from $p(x, y)$
- Here, the model assumes multiple y_i's are generated for *single* x.

Our main interest is the case where we observe multiple random variables $Y_1, Y_2, \ldots, Y_n$ which are conditionally independent given a single latent variable.

Assume that we have the following model

$$\begin{aligned} X &\sim p(x) \\ Y_i | X = x &\sim p(y_i | x), \quad i = 1, \ldots, n. \end{aligned}$$

Note the difference:

- Previously, we simulated (X_i, Y_i) from $p(x, y)$
- Here, the model assumes multiple y_i's are generated for *single* x.
- Furthermore, we will assume we *observe* y_i's.

Conditional Independence

Conditionally independent observations

Assume that $y_1, \ldots, y_n$ are observed and conditionally independent given x.

Conditional Independence

Conditionally independent observations

Assume that $y_1, \ldots, y_n$ are observed and conditionally independent given x.

How do we factorise the joint distribution?

$$\begin{aligned} p(y_1, \ldots, y_n | x) &= p(y_1|x) \cdots p(y_n|x) \\ &= \prod_{i=1}^{n} p(y_i|x). \end{aligned}$$

Bayes rule can be written equivalently:

$$
\begin{aligned}
p(x|y_1,\ldots,y_n) &= \frac{p(y_1,\ldots,y_n|x)p(x)}{p(y_1,\ldots,y_n)} \\
&= \frac{p(x)\prod_{i=1}^n p(y_i|x)}{p(y_1,\ldots,y_n)}.
\end{aligned}
$$

Bayes rule can be written equivalently:

$$p(x|y_1, \ldots, y_n) = \frac{p(y_1, \ldots, y_n|x)p(x)}{p(y_1, \ldots, y_n)}$$
$$= \frac{p(x)\prod_{i=1}^n p(y_i|x)}{p(y_1, \ldots, y_n)}.$$

where

$$p(y_1, \ldots, y_n) = \int p(y_1, \ldots, y_n|x)p(x)\mathrm{d}x$$

This quantity is called the *marginal likelihood* (we will see why later).

Let us define a notation for multiple observations:

$$y_{1:n} = (y_1, \ldots, y_n).$$

Let us define a notation for multiple observations:

$$y_{1:n} = (y_1, \ldots, y_n).$$

Similarly Bayes rule can be rewritten in this new notation

$$\begin{aligned} p(x|y_{1:n}) &= \frac{p(x)p(y_{1:n}|x)}{p(y_{1:n})}, \\ &= \frac{p(x)\prod_{i=1}^{n} p(y_i|x)}{p(y_{1:n})}, \end{aligned}$$

where

$$p(y_{1:n}) = \int p(y_{1:n}|x)p(x)\mathrm{d}x.$$

Consider the following model:

$$
\begin{aligned}
X &\sim \mathcal{N}(x; \mu_0, \sigma_0^2) \\
Y_i | X = x &\sim \mathcal{N}(y_i; x, \sigma^2), \quad i = 1, \ldots, n.
\end{aligned}
$$

Consider the following model:

$$X \sim \mathcal{N}(x; \mu_0, \sigma_0^2)$$
$$Y_i | X = x \sim \mathcal{N}(y_i; x, \sigma^2), \quad i = 1, \ldots, n.$$

We are now interested in the posterior distribution of X given $y_{1:n}$.

$$p(x|y_{1:n}) = \frac{p(x) \prod_{i=1}^{n} p(y_i|x)}{p(y_{1:n})},$$

Consider the following model:

$$X \sim \mathcal{N}(x; \mu_0, \sigma_0^2)$$
$$Y_i | X = x \sim \mathcal{N}(y_i; x, \sigma^2), \quad i = 1, \ldots, n.$$

We are now interested in the posterior distribution of X given $y_{1:n}$.

$$p(x|y_{1:n}) = \frac{p(x) \prod_{i=1}^{n} p(y_i|x)}{p(y_{1:n})},$$

This can be done via matching the terms as we did before (Exercise: Do it using the previous derivation)

Given the model and *fixed* observations $y_1, \ldots, y_n$:

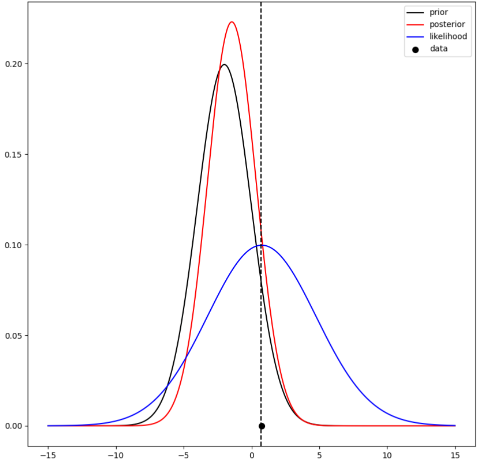
$$p(x|y_{1:n}) = \mathcal{N}(x; \mu_p, \sigma_p^2),$$

where

$$\begin{aligned}
\mu_p &= \frac{\sigma_0^2}{\sigma_0^2 + n\sigma^2}\mu_0 + \frac{\sigma^2 \sum_{i=1}^n y_i}{\sigma_0^2 + n\sigma^2}, \\
\sigma_p^2 &= \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + n\sigma^2}.
\end{aligned}$$

Conditional Independence

Example: Conditionally independent observations



Rejection Sampling

How easy is it to perform rejection sampling in this setting?

Imperial College London

Let us assume that we have $y_1, \ldots, y_n$ observed and our unnormalised posterior is given by

$$\bar{p}(x|y_{1:n}) = p(x) \prod_{i=1}^{n} p(y_i|x).$$

Let us assume that we have a proposal distribution $q(x)$ and assume that we have been lucky to identify some M such that

$$\bar{p}(x|y_{1:n}) \leq Mq(x).$$

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can now perform rejection sampling as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0,1)$
- If $U \leq \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \frac{p(X')\prod_{i=1}^{n} p(y_i|X')}{Mq(X')}$ then accept X'
- Otherwise reject X' and go back to step 1.

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can now perform rejection sampling as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \frac{p(X') \prod_{i=1}^{n} p(y_i|X')}{Mq(X')}$ then accept X'
- Otherwise reject X' and go back to step 1.

What could be an immediate problem as n grows?

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can now perform rejection sampling as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \frac{p(X')\prod_{i=1}^{n} p(y_i|X')}{Mq(X')}$ then accept X'
- Otherwise reject X' and go back to step 1.

What could be an immediate problem as n grows?

- The multiplication $\prod_{i=1}^{n} p(y_i|X')$ would not be numerically stable. Numerical underflow.

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can now perform rejection sampling as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \frac{p(X')\prod_{i=1}^{n} p(y_i|X')}{Mq(X')}$ then accept X'
- Otherwise reject X' and go back to step 1.

What could be an immediate problem as n grows?

- The multiplication $\prod_{i=1}^{n} p(y_i|X')$ would not be numerically stable. Numerical underflow.
- What can you do?

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can now perform rejection sampling as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0, 1)$
- If $U \leq \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \frac{p(X') \prod_{i=1}^{n} p(y_i|X')}{Mq(X')}$ then accept X'
- Otherwise reject X' and go back to step 1.

What could be an immediate problem as n grows?

- The multiplication $\prod_{i=1}^{n} p(y_i|X')$ would not be numerically stable. Numerical underflow.
- What can you do? Work with log-probabilities.

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

We can still perform rejection sampling (provided that $\bar{p}(x|y) \leq Mq(x)$) as follows:

- Sample $X' \sim q(x)$
- Sample $U \sim \text{Unif}(0, 1)$
- Compute log-acceptance probability

$$\log a(X') = \log \frac{\bar{p}(X'|y_{1:n})}{Mq(X')} = \log \frac{p(X') \prod_{i=1}^{n} p(y_i|X')}{Mq(X')},$$

$$= \log p(X') + \sum_{i=1}^{n} \log p(y_i|X') - \log M - \log q(X').$$

- If $\log U \leq \log a(X')$ then accept X'

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

Would this solve our issues though?

Rejection Sampling

How easy is it to perform rejection sampling in this setting?

Would this solve our issues though?

- It is often impossible to find M and $q(x)$ such that $\bar{p}(x|y) \leq Mq(x)$.
- It is not easy to plot the unnormalised posterior $\bar{p}(x|y)$ (without log)
- Bounds found to log-unnormalised posterior can be very loose
  - Super low acceptance probability

Rejection Sampling

Another failure mode

Consider the rejection sampling in 2D for sampling the circle within a square. (See Lecture 1)

Rejection Sampling

Another failure mode

Consider the rejection sampling in 2D for sampling the circle within a square. (See Lecture 1)

The acceptance probability for this case:

$$a = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4} \approx 0.78.$$

Consider the same sampler for the sphere and the cube (3D).

Rejection Sampling

Another failure mode

Imperial College London

Consider the rejection sampling in 2D for sampling the circle within a square. (See Lecture 1)

The acceptance probability for this case:

$$a = \frac{\text{area of the circle}}{\text{area of the square}} = \frac{\pi}{4} \approx 0.78.$$

Consider the same sampler for the sphere and the cube (3D). The acceptance probability for this case:

$$a = \frac{\text{volume of the sphere}}{\text{volume of the cube}} = \frac{\pi}{6} \approx 0.52.$$

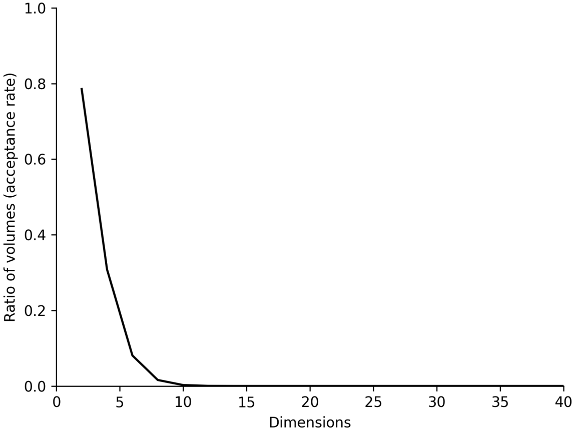
If we were doing this in d dimensions, the acceptance rate would be

$$a = \frac{\text{volume of the unit ball}}{\text{volume of the unit cube}}$$

What do you think happens to it?

Rejection Sampling

Another failure mode



The curse of dimensionality!

Almost there...

Imperial College London

Almost there...

Almost there...

We have done today

Almost there...

We have done today

- Rejection sampler for Bayesian inference

Almost there...

We have done today

- Rejection sampler for Bayesian inference
- Conditional independence

Almost there...

We have done today

- Rejection sampler for Bayesian inference
- Conditional independence
- Bayes update for conditionally i.i.d observations

Almost there...

We have done today

- Rejection sampler for Bayesian inference
- Conditional independence
- Bayes update for conditionally i.i.d observations
- Rejection sampling for complex posteriors

See you tomorrow!

halloween pumpkin, with an aggressive but sneaky look, smoky, against a white background

References I