Lecture 4: Rejection Sampling and Sampling from Compositions, Conditionals, Marginals

Deniz Akyildiz

MATH60047/70047 – Stochastic Simulation

October 18, 2022

Imperial College London

Rejection sampling

The algorithm

The rejection sampler:

Rejection sampling

The algorithm

The rejection sampler:

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

Rejection sampling

The algorithm

The rejection sampler:

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

$$a(X') = \frac{p(X')}{Mq(X')}$$

Acceptance rate

We have

$$\begin{aligned}\hat{a} = \mathbb{E}[a(X')] &= \int a(x')q(x')\mathrm{d}x' \\ &= \int \frac{p(x')}{Mq(x')}q(x')\mathrm{d}x' \\ &= \frac{1}{M}\int p(x')\mathrm{d}x' \\ &= \frac{1}{M}.\end{aligned}$$

For the unnormalised case:

$$\begin{aligned}\hat{a} = \mathbb{E}[a(X')] &= \int a(x')q(x')\mathrm{d}x' \\ &= \int \frac{\bar{p}(x')}{Mq(x')}q(x')\mathrm{d}x' \\ &= \int Z\frac{p(x')}{Mq(x')}q(x')\mathrm{d}x' \\ &= \frac{Z}{M}\int p(x')\mathrm{d}x' \\ &= \frac{Z}{M}.\end{aligned}$$

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:

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

Rejection sampling

Example: Optimising rejection sampling

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

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

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,

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

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 $\log$

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

Take the derivative

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

Take the derivative

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

which implies

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

Take the derivative

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

which implies

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

therefore

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

Take the derivative

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

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

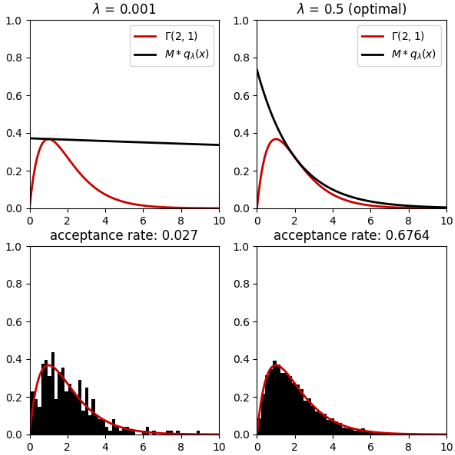
- Sample $X' \sim$ Exp$(1/\alpha)$ and $U \sim$ Unif$(0, 1)$
- If

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

accept X', otherwise start again.

Rejection sampling

Example: Optimising rejection sampling



Example: Rejection sampling

Sample Gaussian using Cauchy

Let

$$\bar{p}(x) = e^{-x^2/2}$$

$$q(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

Compute

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

Example: Rejection sampling

Sample Gaussian using Cauchy

Compute

$$\log \bar{p}(x)/q(x) = -\frac{x^2}{2} + \log(1 + x^2) + \log(1/\pi)$$

Find the roots.

Compute

$$\log \bar{p}(x)/q(x) = -\frac{x^2}{2} + \log(1 + x^2) + \log(1/\pi)$$

Find the roots. Taking the derivative

$$\frac{\mathrm{d}}{\mathrm{d}x} \log \bar{p}(x)/q(x) = -x + \frac{2x}{1 + x^2} = 0$$
$$x = 0, \pm 1.$$

Which one is the maximum?

Compute the second derivative

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} \log \bar{p}(x)/q(x) = -1 + \frac{2(1-x^2)}{(1+x^2)^2} = 0$$

- When $x = 0$, the second derivative is positive - which means $x = 0$ is a minimum.
- When $x = \pm 1$, the second derivative is negative - which means $x = \pm 1$ is a maximum.
- $x^\star = \pm 1$.

So we have

$$M = \frac{\bar{p}(1)}{q(1)} = 2\pi e^{-1/2}.$$

We have seen *independent* sampling methods:

We have seen *independent* sampling methods:

- Given a $p(x)$ (or unnormalised $\bar{p}(x)$, how to draw independent samples
  - Inversion
  - Transformation
  - Rejection

Before we see more of *sampling* methods, we will see how we can use them by

We have seen *independent* sampling methods:

- Given a $p(x)$ (or unnormalised $\bar{p}(x)$, how to draw independent samples
  - Inversion
  - Transformation
  - Rejection

Before we see more of *sampling* methods, we will see how we can use them by

- Composition
- Conditional sampling
- Computing marginals

We have seen *independent* sampling methods:

- Given a $p(x)$ (or unnormalised $\bar{p}(x)$, how to draw independent samples
  - Inversion
  - Transformation
  - Rejection

Before we see more of *sampling* methods, we will see how we can use them by

- Composition
- Conditional sampling
- Computing marginals

Let us look at how to *utilise* sampling methods.

Recall that we denote generic densities $p(x)$ and $p(y)$ etc...

Background

Let us finally settle on some notions

Recall that we denote generic densities $p(x)$ and $p(y)$ etc...

We say that X and Y are independent if $p(x, y)$ satisfies

Recall that we denote generic densities $p(x)$ and $p(y)$ etc...

We say that X and Y are independent if $p(x, y)$ satisfies

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

Similar property naturally follows for expectations.

For dependent X and Y, we have the decomposition

For dependent X and Y, we have the decomposition

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

For dependent X and Y, we have the decomposition

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

The Bayes' rule follows from this formula:

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

We will look at it in detail later.

Let us look into details the structure of joint distributions. Given $p(x, y)$, we have for continuous variables

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

or for discrete variables

$$p(y) = \sum_{x \in \mathsf{X}} p(x, y).$$

Let us look into details the structure of joint distributions. Given $p(x, y)$, we have for continuous variables

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

or for discrete variables

$$p(y) = \sum_{x \in \mathsf{X}} p(x, y).$$

The same is true for

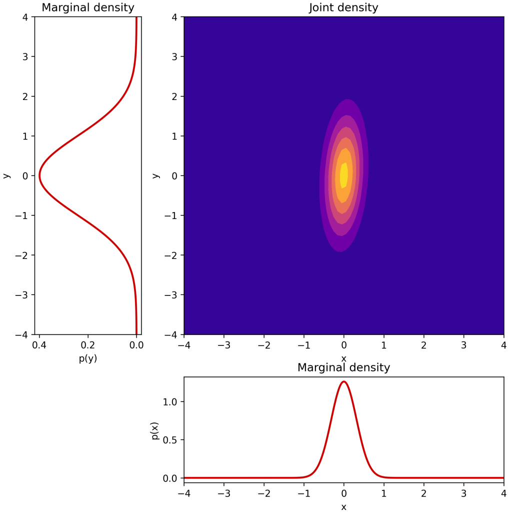
$$p(x) = \int p(x, y) \mathrm{d}y,$$

for continuous variables and

$$p(x) = \sum_{y \in \mathsf{Y}} p(x, y)$$

Background

Joint continuous distribution and marginals



An example table for a joint distribution $p(x, y)$

$p(x, y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	0	0	2/6
$Y = 1$	1/6	0	1/6	0	2/6
$Y = 2$	0	0	1/6	0	1/6
$Y = 3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Let us consider a mixture of two distributions $q_1(x)$ and $q_2(x)$ with weights w_1 and w_2 where $w_1 + w_2 = 1$.

Let us consider a mixture of two distributions $q_1(x)$ and $q_2(x)$ with weights w_1 and w_2 where $w_1 + w_2 = 1$.

We can sample from this mixture by

- Sampling from the discrete distribution $p(k)$ with $p(k = 1) = w_1$ and $p(k = 2) = w_2$, i.e.,

$$k \sim p(k).$$

Let us consider a mixture of two distributions $q_1(x)$ and $q_2(x)$ with weights w_1 and w_2 where $w_1 + w_2 = 1$.

We can sample from this mixture by

- Sampling from the discrete distribution $p(k)$ with $p(k=1) = w_1$ and $p(k=2) = w_2$, i.e.,

$$k \sim p(k).$$

- Sampling from the distribution $q_k(x)$.

The resulting distribution is $p(x) = w_1 q_1(x) + w_2 q_2(x)$.

True for

$$p(x) = \sum_{n=1}^{K} w_n q_n(x).$$

True for

$$p(x) = \sum_{n=1}^{K} w_n q_n(x).$$

We can sample from this mixture by

- Sampling from the discrete distribution with $p(k=1) = w_1, p(k=2) = w_2, \ldots, p(k=n) = w_n$, i.e.,

$$k \sim p(k).$$

- Sampling from the distribution $q_k(x)$.

The resulting distribution is $p(x) = \sum_{n=1}^{K} w_n q_n(x)$.

True for

$$p(x) = \sum_{n=1}^{K} w_n q_n(x).$$

We can sample from this mixture by

- Sampling from the discrete distribution with $p(k=1) = w_1, p(k=2) = w_2, \ldots, p(k=n) = w_n$, i.e.,

$$k \sim p(k).$$

- Sampling from the distribution $q_k(x)$.

The resulting distribution is $p(x) = \sum_{n=1}^{K} w_n q_n(x)$.

Compose the discrete sampling with any other sampling method.

Figure: The density of a mixture of three Gaussians: $p(x) = \sum_{k=1}^{3} w_k \mathcal{N}(x; \mu_k, \sigma_k^2)$ with $\mu_1 = -2, \mu_2 = 0, \mu_3 = 4, \sigma_1 = 0.5, \sigma_2 = 1, \sigma_3 = 0.5, w_1 = 0.2, w_2 = 0.6, w_3 = 0.2.$

Let us consider a conditional distribution $p(y|x)$. This is a density for fixed x, i.e.,

$$\int p(y|x)\mathrm{d}y = 1.$$

Let us consider a conditional distribution $p(y|x)$. This is a density for fixed x, i.e.,

$$\int p(y|x)\mathrm{d}y = 1.$$

Sampling from this distribution is trivial given x with known techniques

- Inversion
- Transformation
- Rejection
- Thousand other methods

Simple example:

$$p(y|x) = \mathcal{N}(y; x, 1).$$

Just sample by fixing x from a Gaussian with mean x.

Recall that we have for any joint distribution $p(x, y)$

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

This decomposition can be used for sampling $(x, y) \sim p(x, y)$. Indeed, we can sample from the joint by

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

Surprisingly, we can use the samples from joint to compute $p(y)$:

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

The above operation is called *marginalisation*[1].

[1] Mathematically same as continuous mixtures, more on this later.

Suppose we have 2D samples $(x_1, y_1), \ldots, (x_n, y_n)$ from $p(x, y)$. How to sample from $p(y)$?

Suppose we have 2D samples $(x_1, y_1), \dots, (x_n, y_n)$ from $p(x, y)$. How to sample from $p(y)$?

Do nothing! Just keep the y's and discard the x's. The resulting samples are from $p(y)$!

Suppose we have 2D samples $(x_1, y_1), \ldots, (x_n, y_n)$ from $p(x, y)$. How to sample from $p(y)$?

Do nothing! Just keep the y's and discard the x's. The resulting samples are from $p(y)$!

This is a neat way to avoid computing an integral.

Example: Consider the following model:

$$p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$$
$$p(y|x) = \mathcal{N}(y; x, \sigma^2).$$

The samples $(x_i, y_i)_{i=1}^n$ can be drawn straightforwardly and y's will be distributed w.r.t. $p(y)$.

Example: Consider the following model:

$$p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$$
$$p(y|x) = \mathcal{N}(y; x, \sigma^2).$$

The samples $(x_i, y_i)_{i=1}^n$ can be drawn straightforwardly and y's will be distributed w.r.t. $p(y)$.

But what is $p(y)$?

Example: Consider the following model:

$$p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$$
$$p(y|x) = \mathcal{N}(y; x, \sigma^2).$$

The samples $(x_i, y_i)_{i=1}^n$ can be drawn straightforwardly and y's will be distributed w.r.t. $p(y)$.

But what is $p(y)$? We can compute it by marginalisation.

$$\begin{aligned} p(y) &= \int p(y|x)p(x)\mathrm{d}x \\ &= \int \mathcal{N}(y; x, \sigma^2)\mathcal{N}(x; \mu, \sigma_0^2)\mathrm{d}x. \end{aligned}$$

This is given as

$$p(y) = \mathcal{N}(y; \mu, \sigma_0^2 + \sigma^2).$$

Two ways to sample $p(y)$

Two ways to sample $p(y)$

Sample from joint and keep y's

- $X \sim p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$
- $Y|X = x \sim p(y|x) = \mathcal{N}(y; x, \sigma^2)$

Sample n times and keep $Y's$ which will be $Y \sim p(y)$.

Two ways to sample $p(y)$

Sample from joint and keep y's

- $X \sim p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$
- $Y|X = x \sim p(y|x) = \mathcal{N}(y; x, \sigma^2)$

Sample n times and keep $Y's$ which will be $Y \sim p(y)$.Or

Two ways to sample $p(y)$

Sample from joint and keep y's

- $X \sim p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$
- $Y|X = x \sim p(y|x) = \mathcal{N}(y; x, \sigma^2)$

Sample n times and keep $Y's$ which will be $Y \sim p(y)$.Or

- Derive $p(y) = \mathcal{N}(y; \mu, \sigma_0^2 + \sigma^2)$
- Sample $Y \sim p(y)$

Marginalisation

Sampling from marginals

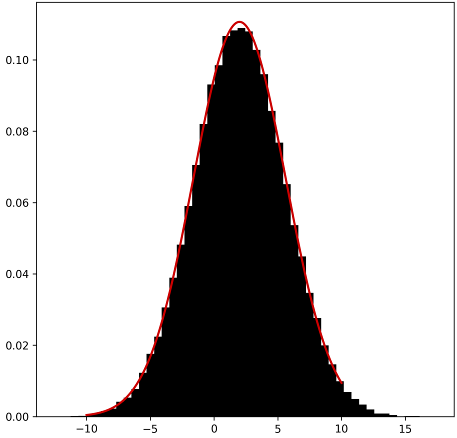


Figure: The sampling from marginal $p(y)$ with $p(x) = \mathcal{N}(x; 2, 2)$ and $p(y|x) = \mathcal{N}(y; x, 3)$. The marginal $p(y)$ is given by $\mathcal{N}(y; 2, 5)$. Samples drawn from $p(x)$ and $p(y|x)$ as described. Red line is true $p(y)$ and the histogram is obtained by drawing (x, y) and keeping only y.

Define a multivariate Gaussian as

$$p(x) = \mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right).$$

Recall the univariate sampling, using $X \sim \mathcal{N}(0,1)$ and $Y = \mu + \sigma X$.

Define a multivariate Gaussian as

$$p(x) = \mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right).$$

Recall the univariate sampling, using $X \sim \mathcal{N}(0, 1)$ and $Y = \mu + \sigma X$.

In the multivariate case, we first compute L such that $\Sigma = LL^T$ and then sample $X \sim \mathcal{N}(0, I)$ (i.e.,)

- Sample $X_1, \ldots, X_d$ from $\mathcal{N}(0, 1)$
- Set $Y = \mu + LX$

Cholesky decomposition can be computed using `np.linalg.cholesky` in `numpy`.

So far, we've seen

So far, we've seen

- Sample from mixtures $p(x) = \sum_{k=1}^{M} w_k q_k(x)$
  - Sample k from a discrete distribution $p(k) = w_k$
  - Sample $x \sim q_k(x)$ from the mixture component

So far, we've seen

- Sample from mixtures $p(x) = \sum_{k=1}^{M} w_k q_k(x)$
  - Sample k from a discrete distribution $p(k) = w_k$
  - Sample $x \sim q_k(x)$ from the mixture component
- Sample from joint $p(x, y)$
  - $X \sim p(x)$
  - $Y|X = x \sim p(y|x)$

So far, we've seen

- Sample from mixtures $p(x) = \sum_{k=1}^{M} w_k q_k(x)$
  - Sample k from a discrete distribution $p(k) = w_k$
  - Sample $x \sim q_k(x)$ from the mixture component
- Sample from joint $p(x, y)$
  - $X \sim p(x)$
  - $Y|X = x \sim p(y|x)$
- Sample from marginal
  - Sample $(X, Y) \sim p(x, y)$ as above
  - Keep only Y samples

So far, we've seen

- Sample from mixtures $p(x) = \sum_{k=1}^{M} w_k q_k(x)$
  - Sample k from a discrete distribution $p(k) = w_k$
  - Sample $x \sim q_k(x)$ from the mixture component
- Sample from joint $p(x, y)$
  - $X \sim p(x)$
  - $Y|X = x \sim p(y|x)$
- Sample from marginal
  - Sample $(X, Y) \sim p(x, y)$ as above
  - Keep only Y samples
- Sample from multivariate Gaussian (using Cholesky decomposition)

So far, we've seen

- Sample from mixtures $p(x) = \sum_{k=1}^{M} w_k q_k(x)$
  - Sample k from a discrete distribution $p(k) = w_k$
  - Sample $x \sim q_k(x)$ from the mixture component
- Sample from joint $p(x, y)$
  - $X \sim p(x)$
  - $Y|X = x \sim p(y|x)$
- Sample from marginal
  - Sample $(X, Y) \sim p(x, y)$ as above
  - Keep only Y samples
- Sample from multivariate Gaussian (using Cholesky decomposition)

Let us see some examples.

Consider

$$p(x) = \mathcal{N}(x; \mu, \sigma_0^2)$$
$$p(y|x) = \mathcal{N}(y; x, \sigma^2).$$

Compute $p(y)$.

The direct computation of the integral

$$p(y) = \int p(y|x)p(x)\mathrm{d}x = \int \mathcal{N}(y; x, \sigma^2)\mathcal{N}(x; \mu, \sigma_0^2)\mathrm{d}x.$$

could be tedious. Note that

$$\begin{aligned} y &= (y - x) + x \\ y - x &\sim \mathcal{N}(y - x; 0, \sigma^2) \\ x &\sim \mathcal{N}(x; \mu, \sigma_0^2). \end{aligned}$$

This is a sum of Gaussians. Therefore, $p(y)$ is also a Gaussian with means and variances summed.

$$p(y) = \mathcal{N}(y; \mu, \sigma_0^2 + \sigma^2).$$

Example 2: Sample from a linear model

Imagine we would like to simulate data from a linear model. The linear relationship we aim to simulate is

$$y = ax + b + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. In order to simulate data, we could choose $p(x)$ (modelling assumption). Assume

$$p(x) = \mathsf{Unif}(x; -10, 10).$$

Next, we need to sample from $p(y|x)$:

$$p(y|x) = \mathcal{N}(y; ax + b, \sigma^2).$$

Example 2: Sample from a linear model

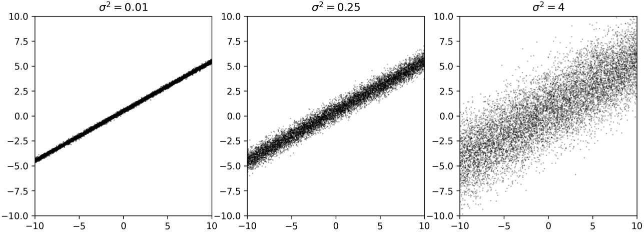

```
import numpy as np
import matplotlib.pyplot as plt

n = 10000
x = np.random.uniform(-10, 10, n)

a = 0.5
b = 0.5
sigma = 2

y = a * x + b + sigma * np.random.normal(0, 1, n)
```

Example 2: Sample from a linear model



Recall our example of a discrete joint distribution $p(x, y)$

	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	0	0	2/6
$Y = 1$	1/6	0	1/6	0	2/6
$Y = 2$	0	0	1/6	0	1/6
$Y = 3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Recall our example of a discrete joint distribution $p(x, y)$

	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

How to sample e.g. $p(y|X=2)$?

Recall our example of a discrete joint distribution $p(x, y)$

$p(x, y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	**0**	0	2/6
$Y = 1$	1/6	0	**1/6**	0	2/6
$Y = 2$	0	0	**1/6**	0	1/6
$Y = 3$	0	0	**0**	1/6	1/6
$p_X(x)$	2/6	1/6	**2/6**	1/6	1

The joint $p(Y = y, X = 2)$ for $y = 0, 1, 2, 3$. The conditional is given by

$$
\begin{aligned}
p(Y = y|X = 2) &= \frac{p(Y = y, X = 2)}{p(X = 2)} \\
&= \frac{p(Y = y, X = 2)}{2/6}, \\
&\rightarrow 3 \times [0, 1/6, 1/6, 0] = [0, 1/2, 1/2, 0].
\end{aligned}
$$

Example 3: Conditionals

Sampling from joint distributions: The discrete case

Recall our example of a discrete joint distribution $p(x,y)$

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	**0**	0	2/6
$Y=1$	1/6	0	**1/6**	0	2/6
$Y=2$	0	0	**1/6**	0	1/6
$Y=3$	0	0	**0**	1/6	1/6
$p_X(x)$	2/6	1/6	**2/6**	1/6	1

The joint $p(Y=y, X=2)$ for $y = 0,1,2,3$. The conditional is given by

$$
\begin{aligned}
p(Y=y|X=2) &= \frac{p(Y=y, X=2)}{p(X=2)} \\
&= \frac{p(Y=y, X=2)}{2/6}, \\
&\rightarrow 3 \times [0, 1/6, 1/6, 0] = [0, 1/2, 1/2, 0].
\end{aligned}
$$

In order to sample, we can use good old inversion.

Example 4: Conditionals

$p(x, y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	0	0	2/6
$Y = 1$	1/6	0	1/6	0	2/6
$Y = 2$	0	0	1/6	0	1/6
$Y = 3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(y\|x)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 0$				
$Y = 1$				
$Y = 2$				
$Y = 3$				

Example 4: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(y\|x)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2			
$Y=1$	1/2			
$Y=2$	0			
$Y=3$	0			

Example 4: Conditionals

Imperial College London

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(y\|x)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2	1		
$Y=1$	1/2	0		
$Y=2$	0	0		
$Y=3$	0	0		

Example 4: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(y\|x)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2	1	0	
$Y=1$	1/2	0	1/2	
$Y=2$	0	0	1/2	
$Y=3$	0	0	0	

Example 4: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(y\|x)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2	1	0	0
$Y=1$	1/2	0	1/2	0
$Y=2$	0	0	1/2	0
$Y=3$	0	0	0	1

Example 5: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(x\|y)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$				
$Y=1$				
$Y=2$				
$Y=3$				

Example 5: Conditionals

$p(x, y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	0	0	2/6
$Y = 1$	1/6	0	1/6	0	2/6
$Y = 2$	0	0	1/6	0	1/6
$Y = 3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(x\|y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 0$	1/2	1/2	0	0
$Y = 1$				
$Y = 2$				
$Y = 3$				

Example 5: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(x\|y)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2	1/2	0	0
$Y=1$	1/2	0	1/2	0
$Y=2$				
$Y=3$				

Example 5: Conditionals

$p(x, y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$	$p_Y(y)$
$Y = 0$	1/6	1/6	0	0	2/6
$Y = 1$	1/6	0	1/6	0	2/6
$Y = 2$	0	0	1/6	0	1/6
$Y = 3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(x\|y)$	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 0$	1/2	1/2	0	0
$Y = 1$	1/2	0	1/2	0
$Y = 2$	0	0	1	0
$Y = 3$				

Example 5: Conditionals

$p(x,y)$	$X=0$	$X=1$	$X=2$	$X=3$	$p_Y(y)$
$Y=0$	1/6	1/6	0	0	2/6
$Y=1$	1/6	0	1/6	0	2/6
$Y=2$	0	0	1/6	0	1/6
$Y=3$	0	0	0	1/6	1/6
$p_X(x)$	2/6	1/6	2/6	1/6	1

Compute and fill in

$p(x\|y)$	$X=0$	$X=1$	$X=2$	$X=3$
$Y=0$	1/2	1/2	0	0
$Y=1$	1/2	0	1/2	0
$Y=2$	0	0	1	0
$Y=3$	0	0	0	1

Example 6: Fundamental Theorem of Simulation

Prove the fundamental theorem of simulation.

Theorem 1 (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

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

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

The proof idea: Start from a uniform distribution $q(x, y)$ on A and show that the marginal in x is $p(x)$.

The proof idea: Start from a uniform distribution $q(x, y)$ on A and show that the marginal in x is $p(x)$.

Proof. Consider the pair (X, Y) uniformly distributed on the region A. We denote their joint density as $q(x, y)$ as

The proof idea: Start from a uniform distribution $q(x, y)$ on A and show that the marginal in x is $p(x)$.

Proof. Consider the pair (X, Y) uniformly distributed on the region A. We denote their joint density as $q(x, y)$ as

$$q(x, y) = \frac{1}{|\mathsf{A}|}, \qquad \text{for } (x, y) \in \mathsf{A}. \tag{2}$$

where |A| is the area of the set A.

The proof idea: Start from a uniform distribution $q(x, y)$ on A and show that the marginal in x is $p(x)$.

Proof. Consider the pair (X, Y) uniformly distributed on the region A. We denote their joint density as $q(x, y)$ as

$$q(x, y) = \frac{1}{|\mathsf{A}|}, \qquad \text{for } (x, y) \in \mathsf{A}. \tag{2}$$

where |A| is the area of the set A. We note that

$$p(x) = \frac{\bar{p}(x)}{|\mathsf{A}|}.$$

We use the standard formula for the joint density $q(x, y) = q(y|x)q(x)$. Note that, since (X, Y) is uniform in A, for fixed x, we have

$$q(y|x) = \frac{1}{\bar{p}(x)} \quad \text{for } (x, y) \in \mathsf{A}.$$

We use the standard formula for the joint density $q(x, y) = q(y|x)q(x)$. Note that, since (X, Y) is uniform in A, for fixed x, we have

$$q(y|x) = \frac{1}{\bar{p}(x)} \quad \text{for } (x, y) \in \mathsf{A}.$$

We therefore write

$$q(x, y) = q(y|x)q(x) = \frac{q(x)}{\bar{p}(x)} \qquad \text{for } (x, y) \in \mathsf{A}. \tag{3}$$

We use the standard formula for the joint density $q(x, y) = q(y|x)q(x)$. Note that, since (X, Y) is uniform in A, for fixed x, we have

$$q(y|x) = \frac{1}{\bar{p}(x)} \quad \text{for } (x, y) \in \mathsf{A}.$$

We therefore write

$$q(x, y) = q(y|x)q(x) = \frac{q(x)}{\bar{p}(x)} \qquad \text{for } (x, y) \in \mathsf{A}. \tag{3}$$

We consider now (2) and (3) which are both valid on $(x, y) \in \mathsf{A}$. Combining them gives

$$q(x) = \frac{\bar{p}(x)}{|\mathsf{A}|},$$

which means $q(x) = p(x)$. ■

See you next Monday!

Martino, Luca, David Luengo, and Joaquín Míguez (2018). *Independent random sampling methods*. Springer.