Monte Carlo approximation

Bayesian Methodology in Biostatistics (BST 249)

Jeffrey W. Miller

Department of Biostatistics
Harvard T.H. Chan School of Public Health

Outline

Outline

Introduction

- Sampling-based methods are widely used due to the ease and generality with which they can be applied.

- Basic task: Approximation of expectations such as

$$\mathrm{E}h(X) = \int h(x)p(x)dx$$

in the case of a continuous random variable X with p.d.f. p, or

$$\mathrm{E}h(X) = \sum_x h(x)p(x)$$

in the case of a discrete random variable X with p.m.f. p.

- General principle: Expectations can be approximated by

$$\mathrm{E}h(X) \approx \frac{1}{N}\sum_{i=1}^{N} h(X_i),$$

where $X_1, \ldots, X_N$ are samples from p.

Introduction

- In Bayesian statistics, most inferential tasks require computation of some expectation.

- Examples
  - posterior probabilities
  - posterior densities
  - posterior expected loss
  - posterior predictive distribution
  - marginal likelihood
  - goodness-of-fit statistics

- Samples are also a good way of visualizing where a probability distribution is putting most of its mass.
  - This is especially useful for distributions on complex/high-dim spaces, e.g., the folding of a protein or RNA strand.

Introduction

- Advantages of sampling-based methods:
  - easy to implement
  - general-purpose / widely applicable
  - reliable
  - work in complex and high-dimensional spaces

- Disadvantages of sampling-based methods:
  - slow (require more time to achieve the same level of accuracy)
  - getting "true" samples may be difficult
  - can be difficult to assess accuracy

Outline

Monte Carlo approximation

- Suppose we want to know the expectation of a random variable $X \sim P$.

- *Simple Monte Carlo*: Draw samples $X_1, \dots, X_N \overset{\text{iid}}{\sim} P$ and use

$$\frac{1}{N} \sum_{i=1}^{N} X_i$$

  as an approximation to $\mathrm{E}X$.

- More generally,

$$\mathrm{E}\big(h(Y) \mid Z = z\big) \approx \frac{1}{N} \sum_{i=1}^{N} h(Y_i)$$

  where $Y_1, \dots, Y_N$ are i.i.d. samples from the distribution of $Y \mid Z = z$.
  - This can be viewed as a special case of $\mathrm{E}X$ where X is defined to have the same distribution as $h(Y) \mid Z = z$.

Monte Carlo: Basic properties

- If $\mathrm{E}|X| < \infty$, then $\frac{1}{N}\sum X_i$ is a consistent estimator of $\mathrm{E}X$:

$$\frac{1}{N}\sum_{i=1}^{N} X_i \longrightarrow \mathrm{E}X$$

  as $N \to \infty$, with probability 1, by the law of large numbers.

- $\frac{1}{N}\sum X_i$ is an unbiased estimator of $\mathrm{E}X$, that is,

$$\mathrm{E}\Big(\tfrac{1}{N}\sum X_i\Big) = \mathrm{E}X.$$

- Who can tell me the variance of $\frac{1}{N}\sum X_i$?

Monte Carlo: Basic properties

- The variance is $\mathbb{V}\big(\frac{1}{N}\sum X_i\big) = \frac{1}{N}\mathbb{V}(X)$.

- Due to unbiasedness, the root-mean-squared-error (RMSE) of $\frac{1}{N}\sum X_i$ equals the standard deviation of $\frac{1}{N}\sum X_i$,

$$\begin{aligned}\text{RMSE} &= \Big[\mathrm{E}\big(|\tfrac{1}{N}\sum X_i - \mathrm{E}X|^2\big)\Big]^{1/2} \\ &= \Big[\mathbb{V}\big(\tfrac{1}{N}\sum X_i\big)\Big]^{1/2} \\ &= \frac{1}{\sqrt{N}}\mathbb{V}(X)^{1/2} = \frac{\sigma(X)}{\sqrt{N}}. \end{aligned} \tag{1}$$

- The RMSE tells us how far the approximation will be from the true value, on average. This tells us that the rate of convergence is of order $1/\sqrt{N}$.

Outline

Example: The Pygmalion effect

- To compute the posterior probability that the spurters had a larger mean increase in IQ score, we used the Monte Carlo approximation

$$\mathbb{P}(\boldsymbol{\mu}_S > \boldsymbol{\mu}_C \mid \text{data}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathrm{I}\big(\boldsymbol{\mu}_S^{(i)} > \boldsymbol{\mu}_C^{(i)}\big).$$

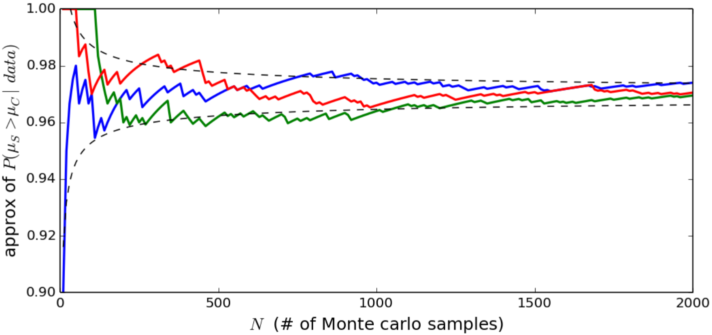
- Using the same approach, we could easily approximate any number of other posterior quantities as well, for example,

$$\mathrm{E}\big(|\boldsymbol{\mu}_S - \boldsymbol{\mu}_C| \,\big|\, \text{data}\big) \approx \frac{1}{N} \sum_{i=1}^{N} |\boldsymbol{\mu}_S^{(i)} - \boldsymbol{\mu}_C^{(i)}|$$

$$\mathrm{E}\big(\boldsymbol{\mu}_S / \boldsymbol{\mu}_C \,\big|\, \text{data}\big) \approx \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\mu}_S^{(i)} / \boldsymbol{\mu}_C^{(i)}.$$

- The posterior density can also be approximated using samples from the posterior.

Example: The Pygmalion effect



- The plot shows Monte Carlo approximations for an increasing number of samples, N.
- Red, blue, and green indicate three repetitions using different sequences of Monte Carlo samples.
- Dotted lines indicate the true value $\pm$ the theoretical RMSE of the Monte Carlo estimator.

Approximating the posterior predictive density

- A Monte Carlo approximation to the posterior predictive p.d.f. or p.m.f. can be made using samples from the posterior:

$$
\begin{aligned}
p(x_{n+1}|x_{1:n}) &= \int p(x_{n+1}|\theta)p(\theta|x_{1:n})d\theta \\
&= \mathrm{E}\big(p(x_{n+1}|\boldsymbol{\theta}) \mid x_{1:n}\big) \\
&\approx \frac{1}{N}\sum_{i=1}^{N} p(x_{n+1}|\boldsymbol{\theta}_i)
\end{aligned}
$$

  where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \overset{\text{iid}}{\sim} p(\theta|x_{1:n})$.

- This is useful when it is difficult or impossible to evaluate the integral analytically.

Outline

Importance sampling approximation

- Importance sampling (IS) is a more powerful type of Monte Carlo approximation.

- The name "importance sampling" is somewhat misleading, since it is not really a method for drawing samples, but rather, a method for approximating expectations—a better name might be importance-weighted Monte Carlo approximation.

- Advantages of importance sampling over simple Monte Carlo:
  - Can significantly improve accuracy, by reducing the variance.
  - Can use samples from a different distribution, say q, to compute expectations with respect to p.
  - Can compute the normalization constant of p.

- Disadvantages
  - Need to be able to evaluate the p.d.f.s/p.m.f.s $p(x)$ and $q(x)$, at least up to proportionality constants.
  - It might be hard to choose a good q.

Importance sampling: Basic idea

- Suppose X is a continuous random variable with p.d.f. $p(x)$.
  - The same thing works in the discrete case — just replace integrals by sums.

- Let q be another p.d.f. on the same space, such that
  1. we can easily sample from q, and
  2. $q(x) > 0$ wherever $p(x) > 0$.

- (Whiteboard:) An *importance sampling approximation* is

$$\begin{aligned} \mathrm{E}h(X) &= \int h(x)p(x)dx \\ &= \int h(x)\frac{p(x)}{q(x)}q(x)dx \\ &\approx \frac{1}{N}\sum_{i=1}^{N} h(Y_i)\frac{p(Y_i)}{q(Y_i)} \end{aligned}$$

where $Y_1, \ldots, Y_N \overset{\text{iid}}{\sim} q$.

Importance sampling: Basic idea

- Importance sampling: $\mathrm{E}h(X) \approx \frac{1}{N}\sum_{i=1}^{N} h(Y_i)\frac{p(Y_i)}{q(Y_i)}$.
- The approximation step here is just a simple Monte Carlo approximation, applied to a different function and distribution.
- q is sometimes called the *proposal distribution*.
- The ratios $w(Y_i) := p(Y_i)/q(Y_i)$ are referred to as the *importance weights*.
- Intuitive interpretation: The importance weights correct for the fact that we are sampling from q rather than p.
  - y's that occur less frequently under q than p have large importance weight $w(y)$, and vice versa.

Importance sampling: Properties

- Evaluating the IS weights
  - For the version of IS above, we need to be able to compute $p(x)$ and $q(x)$ in order to get the importance weights.
  - There is a more general version for which $p(x)$ and $q(x)$ only need to be computable up to constants of proportionality.
- Since the IS approximation is just a simple Monte Carlo approx of $\mathrm{E}\, h(Y)p(Y)/q(Y)$, it has the following properties:
  - consistent, as long as $\mathrm{E}|h(Y)p(Y)/q(Y)| < \infty$, where $Y \sim q$
  - unbiased
  - the variance of the estimator is $\frac{1}{N}\mathbb{V}\big(h(Y)p(Y)/q(Y)\big)$
  - the RMSE is $\sigma\big(h(Y)p(Y)/q(Y)\big)/\sqrt{N}$.

Importance sampling: Choosing the proposal distribution

- The rate of convergence is still of order $1/\sqrt{N}$, however, the constant $\sigma\big(h(Y)p(Y)/q(Y)\big)$ may be smaller or larger.

- How to choose q to minimize RMSE?

- RMSE is minimized when $q(x) \propto h(x)p(x)$, since then $\sigma\big(h(Y)p(Y)/q(Y)\big) = 0$.
  - In other words, the error would be zero after only one sample!
  - In this case, we could trivially compute $\mathrm{E}h(X)$ since $\mathrm{E}h(X) = h(x)p(x)/q(x)$ for almost all x.

- Although this ideal choice of q is unrealistic, it indicates that:
  1. IS can be very accurate if we choose q well, and
  2. we want $q(x)$ to look roughly like $h(x)p(x)$.

Importance sampling: Choosing the proposal distribution

- In practice, we often want to estimate $\mathrm{E}h(X)$ for a variety of different functions h.
- Thus, we often choose $q(x)$ to approximate $p(x)$ rather than $h(x)p(x)$.
  - That way, we can obtain decent estimators for many h's.
  - Plus, we can reuse the same samples Y_i and the same importance weights $w(Y_i) = p(Y_i)/q(Y_i)$.
- When choosing $q(x)$, it is better to err on the side of making it a little more spread out than $p(x)$ (or $h(x)p(x)$).
  - This is done so that $q(x)$ is not too small in areas where $p(x)$ is large.
  - Otherwise, we would occasionally encounter very large IS weights $w(Y_i) = p(Y_i)/q(Y_i)$, which would increase the RMSE.

Outline

GPS example: Approximating the marginal likelihood

- Outside of conjugate priors, computing the marginal likelihood can be very hard and is often not recommended.

- If it must be computed, IS and related techniques are one reasonable approach.

- Suppose we are using a Cauchy model, to handle outliers:

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Cauchy}(\theta, s).$$

- The Cauchy distribution with location θ and scale s has p.d.f.

$$\text{Cauchy}(x \mid \theta, s) = \frac{1}{\pi s \Big(1 + \big(\frac{x-\theta}{s}\big)^2\Big)}.$$

- Unfortunately, there is not a nice conjugate prior for θ.
- Let's put a Cauchy prior on θ:

$$\boldsymbol{\theta} \sim \text{Cauchy}(\theta_0, s_0).$$

GPS example: Data

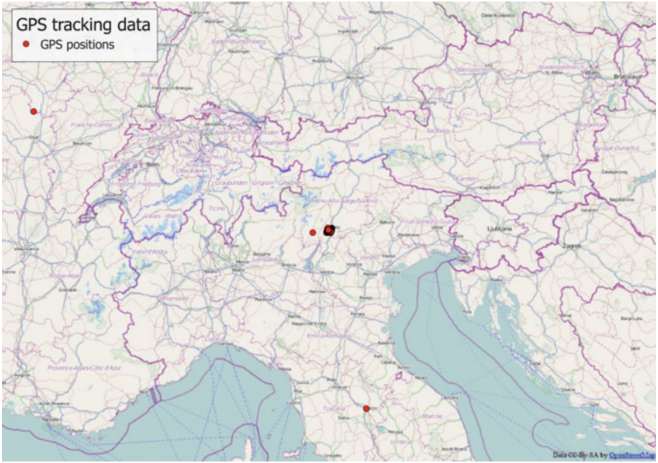
- GPS measurements are usually fairly accurate, but it is not uncommon to get extreme outliers.
- In wildlife management and conservation, animals are tagged with GPS devices in order to track their movements.
- Urbano (2014) provide GPS data on wildlife tracking in northern Italy. Here is a sample of 8 points for illustration:

Latitude	Longitude
36.077916 N	79.009266 W
36.078032 N	79.009180 W
36.078129 N	79.009094 W
36.078048 N	79.008891 W
36.077942 N	79.008962 W
36.089612 N	79.035760 W
36.077789 N	79.008917 W
36.077563 N	79.009281 W

- To keep things simple, let's consider only the latitudes.

GPS example: Data

Animal-tracking GPS measurements with extreme outliers



(Urbano et al., 2014)

GPS example: Approximating the marginal likelihood

- Suppose we need to know the marginal likelihood $p(x_{1:n})$.
- Since we can't compute it analytically (as far as I know), an approximation is needed.
- One approach would be a simple Monte Carlo approximation:

$$p(x_{1:n}) = \int p(x_{1:n}|\theta)p(\theta)d\theta \approx \frac{1}{N}\sum_{i=1}^{N} p(x_{1:n}|\boldsymbol{\theta}_i)$$

  where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \stackrel{\text{iid}}{\sim} p(\theta)$.

- However, this is a poor approximation (the RMSE is large).

GPS example: Approximating the marginal likelihood

- We can do much better with importance sampling if we make a good choice of q:

$$p(x_{1:n}) = \int p(x_{1:n}|\theta)\frac{p(\theta)}{q(\theta)}q(\theta)d\theta \approx \frac{1}{N}\sum_{i=1}^{N} p(x_{1:n}|\boldsymbol{\theta}_i)\frac{p(\boldsymbol{\theta}_i)}{q(\boldsymbol{\theta}_i)}$$

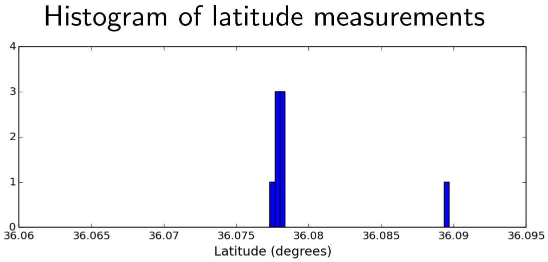
where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \overset{\text{iid}}{\sim} q(\theta)$.

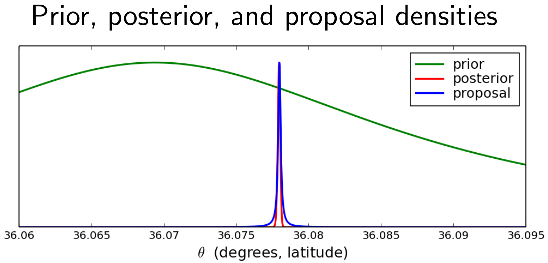
- We want $q(\theta)$ to look as much like $p(x_{1:n}|\theta)p(\theta)$ as possible, and if necessary, to err on the side of being a little more spread out.

- By cheating (a little bit) and looking at a plot of the posterior, let's choose

$$q(\theta) = \text{Cauchy}(\theta \mid \text{median}(x_{1:n}), 10^{-4}).$$

GPS example: Approximating the marginal likelihood

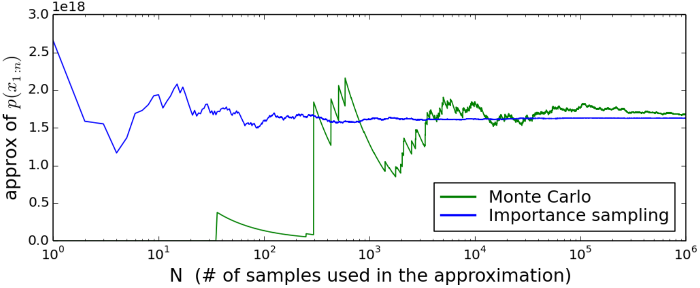




(To make them all visible on one plot, each curve is scaled so that the maximum is 1.)

GPS example: Approximating the marginal likelihood

Convergence of Monte Carlo and IS approximations as N increases



- The IS approximations appear to converge much more quickly, by several orders of magnitude.

GPS example: Approximating the marginal likelihood

- Why does IS do so much better than Monte Carlo here?
- The prior is so spread out, compared to the likelihood, that samples from the prior very rarely land in the small region where the likelihood is large.
- So most of the terms in the Monte Carlo approximation are essentially zero, and a small number of terms are enormous.
- Caution: Approximating the marginal likelihood can be a tricky business, and I would avoid it unless strictly necessary.

Outline

IS with unknown normalization constants

- Often, we cannot compute p and q, but we can compute functions $\tilde{p}$ and $\tilde{q}$ proportional to p and q.

- Fortunately, there is a neat trick that still allows us to make an IS approximation in this situation.

- Suppose

$$p(x) = \tilde{p}(x)/Z_p$$
$$q(x) = \tilde{q}(x)/Z_q$$

  where $\tilde{p}(x)$ and $\tilde{q}(x)$ are easy to compute (but Z_p and Z_q may be intractable).

- Further, assume that $\tilde{q}(x) > 0$ wherever $\tilde{p}(x) > 0$.

IS with unknown normalization constants

- Define
$$\tilde{w}(x) = \begin{cases} \tilde{p}(x)/\tilde{q}(x) & \text{if } \tilde{q}(x) > 0 \\ 0 & \text{if } \tilde{q}(x) = 0. \end{cases}$$
- The general form of the IS approximation is then
$$\begin{aligned} \mathrm{E}h(X) = \int h(x)p(x)dx &\approx \frac{\frac{1}{N}\sum_{i=1}^{N} h(Y_i)\tilde{w}(Y_i)}{\frac{1}{N}\sum_{i=1}^{N} \tilde{w}(Y_i)} \\ &= \sum_{i=1}^{N} h(Y_i)\left(\frac{\tilde{w}(Y_i)}{\sum_{j=1}^{N} \tilde{w}(Y_j)}\right) \end{aligned}$$

  where $Y_1, \ldots, Y_N \overset{\text{iid}}{\sim} q$.
- This can be interpreted as a weighted average of the $h(Y_i)$'s, with weights $\tilde{w}(Y_i)/\sum_j \tilde{w}(Y_j)$.

Example: Approximating posterior expectations

- Consider the animal-tracking GPS example, and now suppose we would like to estimate the posterior mean. Define

$$\pi(\theta) := p(\theta|x_{1:n})$$
$$\tilde{\pi}(\theta) := p(x_{1:n}|\theta)p(\theta)$$
$$Z_\pi := p(x_{1:n}) = \int p(x_{1:n}|\theta)p(\theta)d\theta = \int \tilde{\pi}(\theta)d\theta.$$

- Then $\pi(\theta) = \tilde{\pi}(\theta)/Z_\pi$, and using the general IS formula,

$$\begin{aligned} \mathrm{E}(\boldsymbol{\theta}|x_{1:n}) &= \int \theta p(\theta|x_{1:n})d\theta = \int \theta\pi(\theta)d\theta \\ &\approx \frac{\frac{1}{N}\sum_{i=1}^N \boldsymbol{\theta}_i \tilde{w}(\boldsymbol{\theta}_i)}{\frac{1}{N}\sum_{i=1}^N \tilde{w}(\boldsymbol{\theta}_i)}. \end{aligned}$$

  where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N \overset{\text{iid}}{\sim} q$ and $\tilde{w}(\theta) = \tilde{\pi}(\theta)/q(\theta)$.

- For the GPS data, this yields $\mathrm{E}(\boldsymbol{\theta}|x_{1:n}) \approx 36.0780$.

Derivation of the general IS formula

(Whiteboard activity)

Derivation of the general IS formula (1/2)

- Letting $S = \{x : \tilde{p}(x) > 0\}$, we have

$$
\begin{aligned}
\mathrm{E}h(X) &= \int h(x)p(x)dx \\
&\overset{\text{(a)}}{=} \int_S h(x)\frac{\tilde{p}(x)}{Z_p}\frac{Z_q}{\tilde{q}(x)}q(x)dx \\
&\overset{\text{(b)}}{=} \frac{Z_q}{Z_p}\int h(x)\tilde{w}(x)q(x)dx \\
&\overset{\text{(c)}}{\approx} \frac{Z_q}{Z_p}\frac{1}{N}\sum_{i=1}^{N} h(Y_i)\tilde{w}(Y_i).
\end{aligned}
$$

- In step (a), we use the fact that $\tilde{q}(x) > 0$ whenever $\tilde{p}(x) > 0$.
- In step (b), we use the fact that $\tilde{w}(x) = 0$ for any $x \notin S$.
- Step (c) is a simple Monte Carlo approximation.

Derivation of the general IS formula (2/2)

- Similarly, for the ratio of normalizing constants Z_p/Z_q,

$$
\begin{aligned}
\frac{Z_p}{Z_q} &= \frac{1}{Z_q}\int_S \tilde{p}(x)dx \\
&= \frac{1}{Z_q}\int_S \frac{\tilde{p}(x)}{\tilde{q}(x)}\tilde{q}(x)dx \\
&= \int_S \frac{\tilde{p}(x)}{\tilde{q}(x)}q(x)dx \\
&= \int \tilde{w}(x)q(x)dx \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\tilde{w}(Y_i).
\end{aligned}
$$

- Plugging this into the equation on the previous slide yields the result.

Group activity: Check your understanding

Go to breakout rooms and work together to answer these questions:
https://forms.gle/NKwmWrwNMF6ezoBm6

(Three people per room, randomly assigned. 15 minutes.)

Outline

Inverse c.d.f. method / Smirnov transform

- The *inverse c.d.f. method* is a common way of generating random samples from univariate probability distributions when the inverse of the c.d.f. can be easily computed.

- The basic idea is that if G is the inverse (in a generalized sense) of the c.d.f. F, and $U \sim \mathrm{Uniform}(0,1)$, then $G(U)$ is a random variable with c.d.f. F.

- $\mathrm{Exp}(\theta)$ example:
  - The $\mathrm{Exp}(\theta)$ c.d.f. is $F(x) = (1 - e^{-\theta x})\mathrm{I}(x > 0)$.
  - F is invertible on $(0, \infty)$, with inverse $G(u) = -(1/\theta)\log(1-u)$ for $u \in (0,1)$.
  - Therefore, if $U \sim \mathrm{Uniform}(0,1)$ then $G(U) \sim \mathrm{Exp}(\theta)$.

- Precise statement: Let F be a c.d.f. on $\mathbb{R}$, and define $G(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ for $u \in (0,1)$. If $U \sim \mathrm{Uniform}(0,1)$, then $G(U) \sim F$.

Rejection sampling

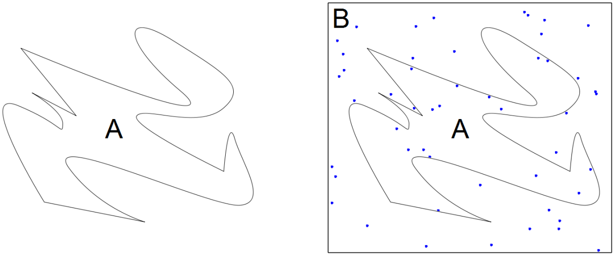
- Rejection sampling is a method for drawing samples when the p.d.f. can be evaluated up to a constant of proportionality.

- Rejection sampling can be applied in multivariate settings, but one has to design a good proposal distribution q.
  - Finding a good q can be difficult, especially in high-dim cases.

- The method relies on two principles:
  1. **The rejection principle:** Rejecting results in conditional samples. That is, if we reject any samples falling outside of a given set, the remaining samples are distributed according to the conditional distribution on that set.

  2. **The projection principle:** If we sample uniformly from the region under the p.d.f. of a distribution (or a function proportional to it), and discard the "height", then we obtain a sample from that distribution.

The rejection principle



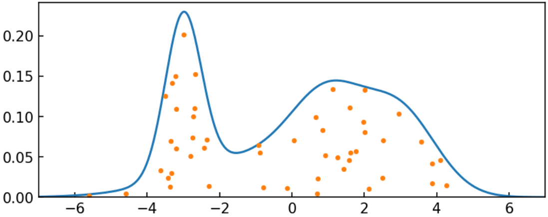
- How to draw samples from the uniform distribution on A?

- Rejection-based approach: Draw samples uniformly from B and keep only those that are in A.

- To minimize rejections, we want the bounding region B to be as small as possible, while still containing A.

- More generally, if we sample X, and reject unless $X \in A$, we get samples from the conditional distribution $X \mid X \in A$.

The projection principle



- Suppose we want to sample from a distribution on $\mathbb{R}^d$ with p.d.f. $\pi(x) = \tilde{\pi}(x)/Z_\pi$.

- Consider the region of $\mathbb{R}^{d+1}$ under $\tilde{\pi}$:

$$A = \Big\{(x, y) : x \in \mathbb{R}^d,\ 0 < y < \tilde{\pi}(x)\Big\}.$$

- It turns out that if $(X, Y) \sim \mathrm{Uniform}(A)$, then $X \sim \pi$.

The projection principle

- To see why, first note that the volume of A is

$$\mathrm{Vol}(A) = \int \tilde{\pi}(x)dx = \int Z_\pi \pi(x)dx = Z_\pi.$$

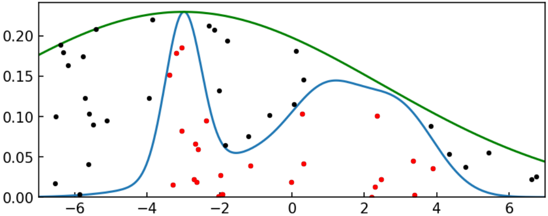
- Since the p.d.f. of the uniform distribution on A is constant, we have

$$p(x,y) = \mathrm{Uniform}(x,y \mid A) = \frac{\mathrm{I}\big(0 < y < \tilde{\pi}(x)\big)}{Z_\pi}.$$

- Therefore,

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} p(x,y)dy = \int_{-\infty}^{\infty} \frac{\mathrm{I}\big(0 < y < \tilde{\pi}(x)\big)}{Z_\pi} dy \\ &= \frac{1}{Z_\pi} \int_0^{\tilde{\pi}(x)} dy = \frac{\tilde{\pi}(x)}{Z_\pi} = \pi(x). \end{aligned}$$

Rejection sampling procedure



- Combining these two principles leads to rejection sampling.
- Basic idea: Generate samples uniformly under a curve that dominates $\tilde{p}(x)$, and only keep samples that land under $\tilde{p}(x)$.
- This generates samples from $p(x) \propto \tilde{p}(x)$.

Rejection sampling procedure

- Objective: Generate samples from a distribution on $\mathbb{R}^d$ with p.d.f. $p(x) \propto \tilde{p}(x)$.

- Choose a proposal distribution q that is easy to sample from, and is as close as possible to being proportional to $\tilde{p}$.

- Choose $c > 0$ such that $c\,q(x) \geq \tilde{p}(x)$ for all x.

- Then, to draw a sample from p:
  1. Sample $X \sim q$.
  2. Sample $Y \sim \mathrm{Uniform}(0, c\,q(X))$.
  3. If $Y \geq \tilde{p}(X)$, then go back to step 1.
  4. Otherwise, output X as a sample.

References and supplements

- Urbano, F., Basille, M., and Cagnacci, F. Data Quality: Detection and Management of Outliers. Spatial Database for GPS Wildlife Tracking Data. Springer International Publishing, 2014. 115-137.
- Radford Neal (2008), The Harmonic Mean of the Likelihood: Worst Monte Carlo Method Ever. `https://radfordneal.wordpress.com/2008/08/17/the-harmonic-mean-of-the-likelihood-worst-monte-carlo-method-ever/`

Individual activity: Exit ticket

Answer these questions individually:
https://forms.gle/mQeZJHMFZdz1oUAh6