COS 302 Precept 9

Spring 2020

Princeton University

Outline

Outline

Review

- We want to find statistics (like the mean or variance) of our probability distributions for learning and prediction.
- For many common distributions, these are functions of the parameters that we can look up.
- For example, if $X \sim Bin(n, \theta)$ then

$$\mathbb{E}[X] = n\theta$$
$$Var[X] = n\theta(1 - \theta).$$

Motivations

- In many interesting cases, we don't know the distribution or its parameters.
- Sometimes we only have access to samples from a distribution $\pi(x)$, or know it up to a constant, $\pi^*(x) = c\pi(x)$
- Using this information, we can *estimate* the mean and variance, and other quantities of interest, using the Monte Carlo method.

Outline

Definition: Monte Carlo estimation

- Given samples from $\pi(x)$, we want to compute $\mathbb{E}_{\pi(x)}[f(x)]$ for reasonable functions f. Recall that

$$\mathbb{E}_{\pi(x)}[f(x)] = \int f(x)\pi(x)dx.$$

- Any such integral can be estimated from samples:

$$\int f(x)\pi(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)}), \text{ where } x^{(s)} \sim \pi(x).$$

- We call this quantity the Monte Carlo estimator of the mean.

Monte Carlo Approximation: Properties

Monte Carlo estimation has some nice properties:

- It's **unbiased**: the mean of the estimator $\hat{f}$ is the expected value of interest–there's no additional bias in our estimate:

$$E_{\pi(\{x^{(s)}\})}[\hat{f}] = \frac{1}{S}\sum_{s=1}^{S} E_{\pi(x)}[f(x)] = E_{\pi(x)}[f(x)]$$

Monte Carlo Approximation: Properties

Monte Carlo estimation has some nice properties:

- It has a variance that decreases with more samples:

$$\begin{aligned}\mathrm{Var}_{\pi(\{x^{(s)}\})}[\hat{f}] &= \frac{1}{S^2}\sum_{s=1}^{S}\mathrm{Var}_{\pi(x)}[f(x)] \\ &= \frac{1}{S}\mathrm{Var}_{\pi(x)}[f(x)].\end{aligned}$$

  We say that the variance of the estimator *shrinks as* $1/S$.

Monte Carlo Approximation: Analysis

- The properties show us that as the number of samples increases, our estimate $\hat{f}$ approaches the true value $\mathbb{E}[f(X)]$. That's good.
- Note that the *error* of the estimate only shrinks as $1/\sqrt{S}$. This is not great. If we want to reduce our error by $1/m$, we need to increase our number of samples by m^2.

Monte Carlo Approximation: Analysis

- **"Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse."** – Alan Sokal, *Monte Carlo Methods in Statistical Mechanics, 1996*.

Aside: Monte Carlo for Arbitrary Integrals

We can use this method to evaluate an arbitrary integral over a domain Ω. Notice that

$$\begin{aligned}\int_\Omega f(x)dx &= \int_\Omega f(x)\frac{q(x)}{q(x)}dx \\ &= \int_\Omega \frac{f(x)}{q(x)}q(x)dx \\ &= \mathbb{E}_{q(x)}\left[\frac{f(x)}{q(x)}\right],\end{aligned}$$

so long as we choose a probability distribution q so that $q(x) > 0$ for every $x \in \Omega$.

Outline

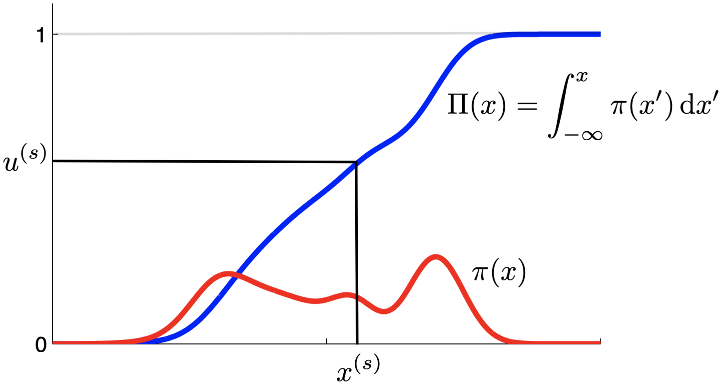
Sampling basics

$$\int f(x)\pi(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)}), \text{ where } x^{(s)} \sim \pi(x)$$

- So, as long as we can sample from $\pi(x)$, we can estimate its statistics
- How can we get samples from a strange or interesting $\pi(x)$?

Review: Inverse Transform Sampling

Recall that if we can compute the CDF of a function, we can also compute its density function:



$u^{(s)} \sim \text{Uniform}(0, 1)$ $\qquad$ $x^{(s)} = \Pi^{-1}(u^{(s)})$

Analysis: Inverse Transform Sampling

- Inverse transform sampling can be a convenient way to map from uniform random variates to a more complex distribution, in simple cases
- **Bad news**: We still had to do an integral $\Pi(x)$. Integrals can be nasty–impossible to compute in closed form, and hard to approximate.

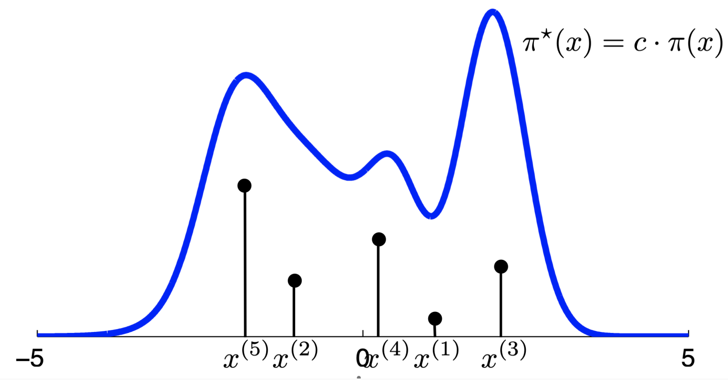
Sampling: Big Picture

- We want to be able to use the Monte Carlo method to figure out the mean and variance of interesting distributions.
- Inverse transform sampling lets us do this, but still requires us to be able integrate $\pi(x)$, and to always know its normalizing constant.
- Rejection sampling is a way to get around both these constraints.

Outline

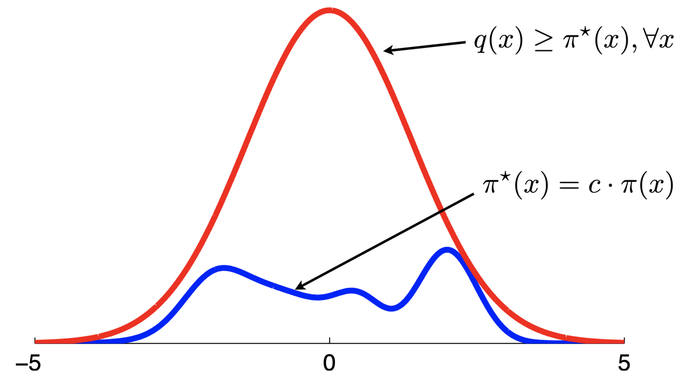
Rejection Sampling Intuition

Notice if I draw samples uniformly from the volume *beneath* a PDF, it must have the correct marginal distribution:



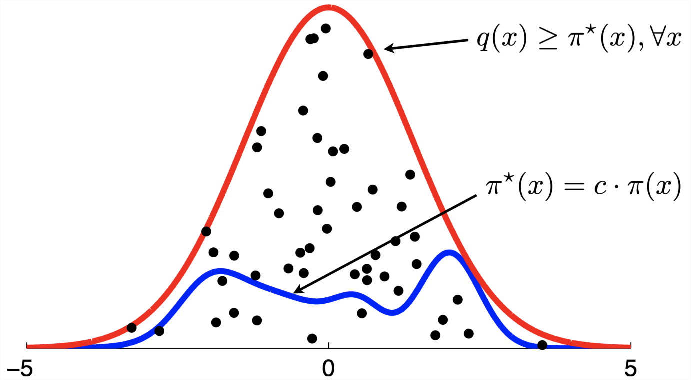
Rejection Sampling Intuition

But how could we sample from this volume? Choose a simple distribution $q(x)$, sample from this, and throw away the ones that are above $\pi^*(x) = c\pi(x)$:



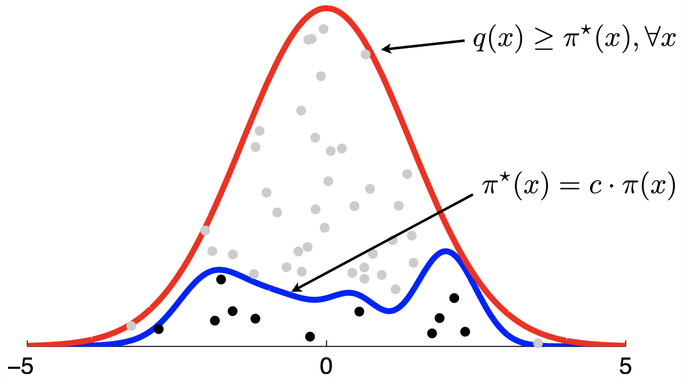
Rejection Sampling Intuition

How do we sample from this volume? Choose a simple distribution $q(x)$, sample from this, and throw away the ones that are above $\pi^*(x) = c\pi(x)$:



Rejection Sampling Intuition

How do we sample from this volume? Choose a simple distribution $q(x)$, sample from this, and throw away the ones that are above $\pi^*(x) = c\pi(x)$:



Rejection Sampling: Algorithm

1. Choose $q(x)$ and c so that $q(x) \geq \pi^*(x) = c\pi(x)$
2. Sample $x^{(s)} \sim q(x)$
3. Sample $u^{(s)} \sim \text{Unif}[0, q(x^{(s)})]$
4. If $u^{(s)} \leq \pi^*(x^{(s)})$ keep $x^{(s)}$, else reject and go to step (2)

If you accept, you get an unbiased sample from $\pi(x)$.