CPSC 486/586: Probabilistic Machine Learning	February 22, 2023
Lecture 12	
Lecturer: Andre Wibisono	*Scribe: Jake Tae*

1 Introduction

In the last lecture, we discussed sampling from discrete distributions. In this lecture, we discuss sampling from continuous space. More concretely, assume we can sample from the uniform distribution on $[0, 1] \in \mathbb{R}$. Let $\nu(x)$ be the target distribution on $\mathbb{R}$. How can we sample $x \sim \nu$?

We will explore two different algorithms that can achieve this objective: inverse CDF transform and rejection sampling.

2 Inverse CDF Transform

Let $\nu : \mathbb{R} \to \mathbb{R}$ be the PDF of the target distribution. Then the CDF (cumulative density function) $F_\nu : \mathbb{R} \to [0, 1]$ of ν is given by

$$\begin{aligned} F_\nu(t) &= \mathbb{P}_\nu(X \leq t) \\ &= \int_{-\infty}^{t} \nu(x)\, dx. \end{aligned}$$

Then the inverse CDF transform sampling algorithm is given as follows.

Algorithm 1 Inverse CDF Transform

repeat
    $u \sim \text{Uniform}[0, 1]$
    return $x = F_\nu^{-1}(u)$
until convergence

Lemma 1. *Output x has distribution ν.*

Proof. WTS x has the same CDF F_ν as ν. It suffices to show $\forall t \in \mathbb{R}, \mathbb{P}(X \leq t) = F_\nu(t)$. By the algorithm,

$$
\begin{aligned}
\mathbb{P}(X \leq t) &= \mathbb{P}(F_\nu^{-1}(u) \leq t) \\
&= \mathbb{P}(u \leq F_\nu(t)) \\
&= F_\nu(t).
\end{aligned}
$$

Therefore, the inverse CDF transform works if we can evaluate F_ν^{-1}. $\square$

Example 1. $\nu = \text{Uniform}[a, b]$.

The CDF of ν is given by

$$
\begin{aligned}
F_\nu(t) &= \int_{-\infty}^{t} \nu(x)\, dx \\
&= \frac{t-a}{b-a}.
\end{aligned}
$$

$F_\nu^{-1} = x \iff u = F_\nu(x) = (x-a)/(b-a)$. So we can sample from this distribution by applying the transform $x = a + u(b-a)$.

Example 2. $\nu = \text{Exp}(\lambda)$ with density $\nu(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0}$.

The CDF of ν is given by

$$
\begin{aligned}
F_\nu(t) &= \int_{0}^{t} \nu(x)\, dx \\
&= \int_{0}^{t} \lambda e^{-\lambda x}\, dx \\
&= \left(-e^{-\lambda x}\right)\Big|_0^t \\
&= 1 - e^{-\lambda t}.
\end{aligned}
$$

Applying the inverse CDF transform, we have

$$
\begin{aligned}
x = F_\nu^{-1}(u) &\iff u = F_\nu(x) = 1 - e^{-\lambda x} \\
&\iff e^{-\lambda x} = 1 - u \\
&\iff x = -\frac{1}{\lambda} \log(1-u) = \frac{1}{\lambda} \log \frac{1}{1-u} \sim \text{Exp}(\lambda).
\end{aligned}
$$

3 Rejection Sampling

Suppose we have a target probability distribution $\nu(x)$ on X and its unnormalized version $\tilde{\nu(x)} \propto \nu(x)$, i.e.,

$$\nu(x) = \frac{\tilde{\nu}(x)}{Z_\nu}, Z_\nu = \int_X \tilde{\nu}(x)\, dx,$$

where Z_ν is unknown an intractable.

Suppose we can sample from a proposal distribution $\mu(x)$ and we have the unnormalized version $\tilde{\mu}(x)$, i.e.,

$$\mu(x) = \frac{\tilde{\mu}(x)}{Z_\mu}, Z_\mu = \int_X \tilde{\mu}(x)\, dx.$$

Assume $\forall x \in X, \tilde{\mu}(x) \geq \tilde{\nu}(x)$. Then the rejection sampling algorithm is given as follows.

Algorithm 2 Rejection Sampling

repeat
- $x \sim \mu$
- $u \sim \text{Uniform}[0, 1]$
- $\alpha(x) \leftarrow \frac{\tilde{\nu}(x)}{\tilde{\mu}(x)}$
- **if** $u \leq \alpha$ **then**
  - **return** x
- **else**
  - continue
- **end if**

until convergence

Example 3. Suppose ν is uniform on $\mathbb{B}_2^d = \{x \in \mathbb{R}^d : \|x\|_2^2 = x_1^2 + \cdots + x_d^2 \leq 1\}$, and μ is uniform on $\mathbb{B}_\infty^d = \{x \in \mathbb{R}^d : \|x\|_\infty = \max_i |x_i| \leq 1\}$. We can view ν and μ as a unit sphere and cube respectively, with $\mathbb{B}_2^d \subset \mathbb{B}_\infty^d$.

$$\nu(x) = \frac{1}{\text{Vol}(\mathbb{B}_2^d)} \mathbb{1}_{\{x \in \mathbb{B}_2^d\}},$$
$$\mu(x) = \frac{1}{\text{Vol}(\mathbb{B}_\infty^d)} \mathbb{1}_{\{x \in \mathbb{B}_\infty^d\}}.$$

$$\tilde{\nu}(x) = \mathbb{1}_{\{x \in \mathbb{B}_2^d\}},$$
$$\tilde{\mu}(x) = \mathbb{1}_{\{x \in \mathbb{B}_\infty^d\}}.$$

Note that $\tilde{\mu}(x) \geq \tilde{\nu}(x)$ as required, since the L_∞ norm ball contains the L_2 norm ball.

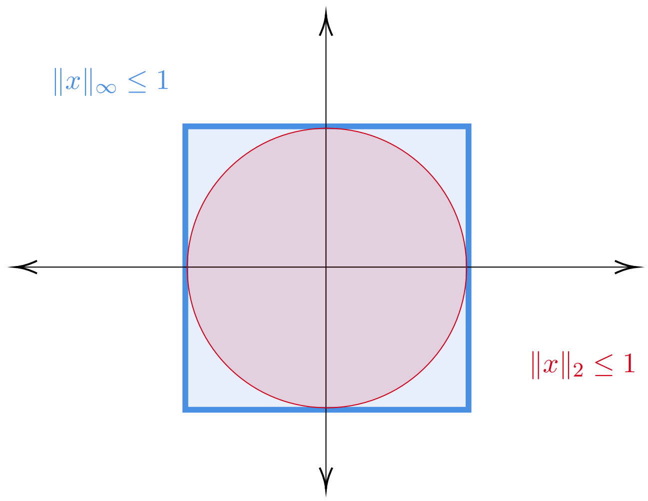

Figure 1: In $d = 2$, recall that $\mathbb{B}_\infty$ is the area inside the square $[-1, 1] \times [-1, 1]$ and $\mathbb{B}_2$ is that of the unit circle. This generalizes to higher dimensions d.

Then the rejection sampling algorithm can be written as follows.

Algorithm 3 Rejection Sampling on d-hypersphere

```
repeat
    x ~ Uniform(B^d_∞)
    α(x) ← ν̃(x)/μ̃(x) = ν̃(x)
    if x ∈ 𝔹^d_2 then
        return x
    else
        continue
    end if
until convergence
```

Lemma 2. *Properties of rejection sampling.*

1. *The output of rejection sampling x follows distribution ν.*

2. *Let N be the number of samples drawn from ν until acceptance. Then N has a geometric distribution with mean $Z_\mu/Z_\nu \geq 1$.*

Proof. See Chewi, Theorem 7.1.1. $\square$

Example 4. From $\nu = \text{Uniform}(B_2^d)$ and $\mu = \text{Uniform}(B_\infty^d)$. Then

$$\begin{aligned} Z_\nu &= \int_X \tilde{\nu}(x)\,dx \\ &= \text{Vol}(B_2^d) \\ &= \frac{\pi^{d/2}}{(d/2)!} \\ &\overset{d\to\infty}{=} 0. \end{aligned}$$

$$\begin{aligned} Z_\mu &= \int_X \tilde{\mu}(x)\,dx \\ &= \text{Vol}(B_\infty^d) \\ &= 2^d \\ &\overset{d\to\infty}{=} \infty. \end{aligned}$$

Then the average number of samples from μ is $Z_\mu/Z_\nu = 2^d(d/2)!/\pi^{d/2}$. When $d = 2$, this quantity becomes $4/\pi$, and it is one way one might approximate π.

Example 5. Let $\nu(x) \propto e^{-f(x)}$ on $\mathbb{R}^d$. WLOG, assume $f(0) = 0, \nabla f(0) = 0$. Assume f is α-strongly convex and L-smooth, i.e., $0 \preceq \alpha I \preceq \nabla^2 f(x) \preceq LI$.

$$\begin{aligned} \frac{\alpha}{2} \leq f(x) \leq \frac{L}{2}\|x\|^2 &\implies \exp\Big(-\frac{\alpha}{2}\Big) \geq \exp(-f(x)) \geq \exp\bigg(-\frac{L}{2}\|x\|^2\bigg) \\ &\implies \tilde{\mu}(x) \geq \tilde{\nu}(x) \\ &\implies \mu = \mathcal{N}(0, \alpha^{-1}I) \text{ as proposal distribution.} \end{aligned}$$

The rejection sampling algorithm for $x \sim \nu$ can be written as follows.

Next, we consider the notion of the consider number, which closely relates to the second property of Lemma 2.

Algorithm 4 Rejection Sampling

repeat
  $x \sim \mu = \mathcal{N}(0, \alpha^{-1} I)$
  $u \sim \text{Uniform}[0,1]$
  $\alpha(x) \leftarrow \frac{\tilde{\nu}(x)}{\tilde{\mu}(x)} = \exp\Big(\frac{\alpha}{2}\|x\|^2 - f(x)\Big)$
  if $u \leq \alpha$ **then**
    return x
  else
    continue
  end if
until convergence

$$Z_\mu = \int_{\mathbb{R}^d} \exp\Big(-\frac{\alpha}{2}\|x\|^2\Big)\, dx = \left(\frac{2\pi}{\alpha}\right)^{d/2}.$$

$$\begin{aligned} Z_\nu &= \int_{\mathbb{R}^d} \exp(-f(x))\, dx \\ &\geq \int_{\mathbb{R}^d} \exp\left(-\frac{L}{2}\|x\|^2\right) dx \\ &= \left(\frac{2\pi}{L}\right)^{d/2}. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbb{E}[\#\text{ samples drawn from } \mu] &= \frac{Z_\mu}{Z_\nu} \\ &\leq \frac{\left(\frac{2\pi}{\alpha}\right)^{d/2}}{\left(\frac{2\pi}{L}\right)^{d/2}} \\ &= \left(\frac{L}{\alpha}\right)^{d/2} \\ &= \kappa^{d/2}, \end{aligned}$$

where $\kappa = L/\alpha \geq 1$ is the condition number of ν. If $\kappa \approx 1$, the problem is "nice"; otherwise, if $\kappa \gg 1$, it is "ill-conditioned."