# Deep Learning - Foundations and Concepts

## Chapter 14. Sampling

nonlineark@github

March 27, 2025

# Outline

# Expectations

For some applications the goal is to evaluate expectations with respect to the distribution. Suppose we wish to find the expectation of a function $f(z)$ with respect to a probability distribution $p(z)$:

$$E(f) = \int f(z)p(z)\mathrm{d}z$$

The general idea behind sampling methods is to obtain a set of samples $z^{(l)}$ drawn independently from the distribution $p(z)$. This allows the expectation to be approximated by a finite sum:

$$\bar{f} = \frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})$$

# Expectations

Let's calculate the expectation and variance of $\bar{f}$:

$$E(\bar{f}) = E(\frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})) = E(f)$$

$$E(\bar{f}^2) = E(\frac{1}{L^2}\sum_{l,l'} f(z^{(l)})f(z^{(l')})) = (E(f))^2 + \frac{1}{L}\mathrm{var}(f)$$

$$\mathrm{var}(\bar{f}) = E(\bar{f}^2) - (E(\bar{f}))^2 = \frac{1}{L}\mathrm{var}(f)$$

Which shows that:

- $\bar{f}$ is an unbiased estimator of $E(f)$.
- Due to the linear decrease of the variance with increasing $L$, in principle, high accuracy may be achievable with a relatively small number of samples $z^{(l)}$.

# Standard distributions

## Problem

Suppose that $z$ is uniformly distributed over the interval $(0, 1)$. Given a probability density function $p$, find a function $g$ such that the random variable $y = g(z)$ has $p$ as its probability density function.

# Standard distributions

Let $U$ be the probability density function of the uniform distribution over the interval $(0, 1)$, we have:

$$
\begin{aligned}
p(y)\mathrm{d}y &= U(z)\mathrm{d}z \\
f(y_0) &= \int_{-\infty}^{y_0} p(y)\mathrm{d}y = \int_{-\infty}^{z_0} U(z)\mathrm{d}z = z_0 \\
y_0 &= f^{-1}(z_0)
\end{aligned}
$$

So we have to transform the uniformly distributed random numbers using a function that is the inverse of the cumulative distribution function of the desired probability density function.

# Standard distributions

Some examples:

- Exponential distribution $p(y) = \lambda \exp(-\lambda y)$:
  - $z = f(y) = \int_0^y p(t)\mathrm{d}t = 1 - \exp(-\lambda y)$.
  - $y = -\frac{1}{\lambda}\log(1-z)$.
- Cauchy distribution $p(y) = \frac{1}{\pi}\frac{1}{1+y^2}$:
  - $z = f(y) = \int_{-\infty}^y p(t)\mathrm{d}t = \frac{1}{\pi}\arctan y + \frac{1}{2}$.
  - $y = \tan(\pi(z - \frac{1}{2}))$.

# Standard distributions

The generalization to multiple variables involves the Jacobian of the change of variables, so that:

$$p_Y(y_1, \dots, y_M) = p_Z(z_1, \dots, z_M) \left| \frac{\partial(z_1, \dots, z_M)}{\partial(y_1, \dots, y_M)} \right|$$

# Standard distributions

The Box-Muller method for generating samples from a Gaussian distribution. First, suppose we generate pairs of uniformly distributed random numbers $z_1, z_2 \in (-1, 1)$. Next, we discard each pair unless it satisfies $z_1^2 + z_2^2 \leq 1$. This leads to a uniform distribution of points inside the unit circle with $p_Z(z_1, z_2) = \frac{1}{\pi}$. Then, for each pair $z_1, z_2$ we evaluate the quantities:

$$y = z\frac{\sqrt{-4\log ||z||}}{||z||}$$

The joint distribution of $y_1$ and $y_2$ is given by:

$$p_Y(y_1, y_2) = p_Z(z_1, z_2)|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}| = (\frac{1}{\sqrt{2\pi}}\exp(-\frac{y_1^2}{2}))(\frac{1}{\sqrt{2\pi}}\exp(-\frac{y_2^2}{2}))$$

So $y_1$ and $y_2$ are independent and each has a Gaussian distribution with zero mean and unit variance.

# Rejection sampling

Suppose that:

- We wish to sample from a distribution $p(z)$, and sampling directly from $p(z)$ is difficult.
- We are easily able to evaluate $p(z)$ for any given value of $z$, up to some normalizing constant $Z_p$, so that $p(z) = \frac{1}{Z_p}\tilde{p}(z)$, where $\tilde{p}(z)$ can readily be evaluated, but $Z_p$ is unknown.

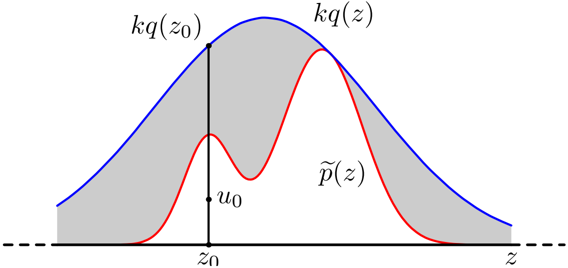
# Rejection sampling

To apply rejection sampling:

- Find a simpler distribution $q(z)$, called a proposal distribution, from which we can readily draw samples.
- Introduce a constant $k$ whose value is chosen such that $kq(z) \geq \tilde{p}(z)$ for all values of $z$.
- Generate a number $z_0$ from the distribution $q(z)$.
- Generate a number $u_0$ from the uniform distribution over $[0, kq(z_0)]$.
- If $u_0 > \tilde{p}(z_0)$ then the sample is rejected, otherwise $u_0$ is retained.
- The corresponding $z$ values in the remaining pairs are distributed according to $p(z)$.

# Rejection sampling

Figure: Illustration of the rejection sampling method



# Rejection sampling

Let's verify the correctness of the rejection sampling method. Suppose that random variable $Z$ is distributed according to $q(z)$, and random variable $U$ is uniformly distributed over $[0, kq(Z)]$. We want to calculate the probability density function of the random variable $Z|0 \leq U \leq \tilde{p}(Z)$:

$$
\begin{aligned}
P(Z \in E|0 \leq U \leq \tilde{p}(Z)) &= \frac{P(Z \in E, 0 \leq U \leq \tilde{p}(Z))}{P(0 \leq U \leq \tilde{p}(Z))} \\
&= \frac{\int_E q(z)\frac{\tilde{p}(z)}{kq(z)}\mathrm{d}z}{\int_{\mathbb{R}} q(z)\frac{\tilde{p}(z)}{kq(z)}\mathrm{d}z} \\
&= \int_E p(z)\mathrm{d}z
\end{aligned}
$$

We see that the random variable $Z|0 \leq U \leq \tilde{p}(Z)$ is indeed distributed according to $p(z)$.

# Rejection sampling

Let's calculate the probability that a sample will be accepted:

$$P_{\text{accept}} = \int_{\mathbb{R}} q(z) \frac{\tilde{p}(z)}{kq(z)} \mathrm{d}z = \frac{Z_p}{k}$$

We see that the constant $k$ should be as small as possible subject to the limitation that $kq(z)$ must be nowhere less than $\tilde{p}(z)$.

# Adaptive rejection sampling

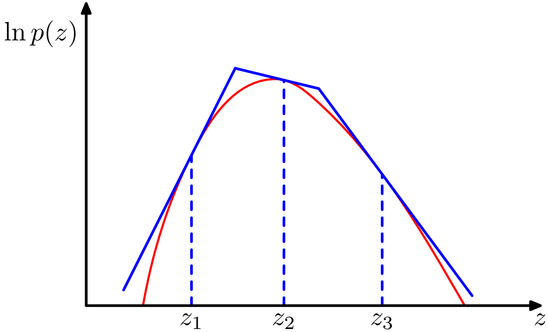
- In many instances, it can be difficult to determine a suitable analytic form for the envelope distribution $q(z)$.
- An alternative approach is to construct the envelope function on the fly based on measured values of the distribution $p(z)$.
- Constructing an envelope function is particularly straightforward when $p(z)$ is log concave.

# Adaptive rejection sampling

- Evaluate the function $\log p(z)$ and its gradient at some initial set of grid points.
- The intersections of the resulting tangent lines are used to construct the envelope function.
- Draw a sample value from the envelope distribution. This is straightforward because the envelope function comprises a piecewise exponential distribution.
- If the sample is accepted, then it will be a draw from the desired distribution.
- If the sample is rejected, then it is incorporated into the set of grid points, a new tangent line is computed, and the envelope function is thereby refined.

# Adaptive rejection sampling

Figure: Illustration of the construction of an envelope function for adaptive rejection sampling



# Importance sampling

The technique of importance sampling provides a framework for approximating expectations directly but does not itself provide a mechanism for drawing samples from a distribution $p(z)$.

Suppose we wish to calculate the expectation of a function $f(z)$ with respect to a distribution $p(z)$:

- The distribution $p(z)$ can be evaluated only up to a normalization constant, so that $p(z) = \frac{\tilde{p}(z)}{Z_p}$, where $Z_p$ is unknown.
- Because it is difficult to draw samples directly from $p(z)$, we rely on a proposal distribution $q(z)$ from which it is easy to draw samples.
- The distribution $q(z)$ also has an unknown normalization constant $Z_q$, so that $q(z) = \frac{\tilde{q}(z)}{Z_q}$.

# Importance sampling

Let's calculate the expectation of $f(z)$ with respect to $p(z)$:

$$
\begin{aligned}
E(f) &= \int f(z)p(z)\mathrm{d}z = \int f(z)\frac{p(z)}{q(z)}q(z)\mathrm{d}z \\
&\approx \frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})\frac{p(z^{(l)})}{q(z^{(l)})} = \frac{Z_q}{Z_p}\frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})\frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})} \\
\frac{Z_p}{Z_q} &= \frac{1}{Z_q}\int \tilde{p}(z)\mathrm{d}z = \frac{1}{Z_q}\int \frac{\tilde{p}(z)}{q(z)}q(z)\mathrm{d}z \\
&\approx \frac{1}{Z_q}\frac{1}{L}\sum_{l=1}^{L}\frac{\tilde{p}(z^{(l)})}{q(z^{(l)})} = \frac{1}{L}\sum_{l=1}^{L}\frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})}
\end{aligned}
$$

where the samples $\{z^{(l)}\}$ are drawn from $q(z)$.

# Importance sampling

Let:

$$\tilde{r}_l = \frac{\tilde{p}(z^{(l)})}{\tilde{q}(z^{(l)})}$$

$$w_l = \frac{\tilde{r}_l}{\sum_{l'} \tilde{r}_{l'}}$$

we see that:

$$E(f) \approx \frac{\sum_{l=1}^{L} f(z^{(l)})\tilde{r}_l}{\sum_{l=1}^{L} \tilde{r}_l} = \sum_{l=1}^{L} w_l f(z^{(l)})$$

# Sampling-importance-resampling

- Draw $L$ samples $z^{(1)}, \dots, z^{(L)}$ from $q(z)$.
- Construct weights $w_1, \dots, w_L$ using $w_l = \frac{\tilde{r}_l}{\sum_{l'} \tilde{r}_{l'}} = \frac{\tilde{p}(z^{(l)})/q(z^{(l)})}{\sum_{l'} \tilde{p}(z^{(l')})/q(z^{(l')})}$.
- Draw $L$ samples from the discrete distribution $(z^{(1)}, \dots, z^{(L)})$ with probabilities given by the weights $(w_1, \dots, w_L)$.

# Sampling-importance-resampling

Let's verify the correctness of the sampling-importance-resampling method.

$$P(z \leq a) = \sum_{l=1}^{L} I(z^{(l)} \leq a) w_l = \frac{\sum_{l=1}^{L} I(z^{(l)} \leq a) \frac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}{\sum_{l=1}^{L} \frac{\tilde{p}(z^{(l)})}{q(z^{(l)})}}$$

where $I$ is the indicator function. Taking the limit $L \rightarrow \infty$:

$$P(z \leq a) = \frac{\int I(z \leq a) \frac{\tilde{p}(z)}{q(z)} q(z) \mathrm{d}z}{\int \frac{\tilde{p}(z)}{q(z)} q(z) \mathrm{d}z} = \frac{\int_{-\infty}^{a} \tilde{p}(z) \mathrm{d}z}{\int_{-\infty}^{\infty} \tilde{p}(z) \mathrm{d}z} = \int_{-\infty}^{a} p(z) \mathrm{d}z$$