

430.714

Estimation Theory

Prof. Songhwai Oh
ECE, SNU

SAMPLING

Sampling

- An approximation method when a direct computation is not possible.
- Goals
 - (1) Estimate probabilities: likelihood, posterior
 - (2) Estimate expectations

$$\mathbb{E}(g(X)) = \sum_x g(x)p(x)$$

- Note: (1) is a special case of (2), i.e., $g(x) = \delta(x, \bar{x}_E)$
- Sampling

- Collect $x_1, x_2, \dots, x_L \sim p(x)$
- Solve (2) by

$$\mathbb{E}(g(X)) \approx \frac{1}{L} \sum_{l=1}^L g(x_l)$$

Inverse transform sampling

Suppose we want to sample from $p(y)$ with CDF F .

Start with $U \sim \text{Uniform}(0, 1)$ and the target CDF F .

$$F(y) = \int_{-\infty}^y p(t) dt$$

$$Y = F^{-1}(U)$$

A uniform probability level selects a quantile of the target.

For a CDF with jumps, use its generalized inverse.

Why the inverse transform works

Probability mass is preserved

$$p_Y(y) |dy| = p_U(u) |du|$$

$$p_Y(y) = p_U(u) \left| \frac{du}{dy} \right|$$

We have $u = F(y)$ and $p_u(u) = 1$. Also $du/dy = dF/dy = p(y)$.

$$p_Y(y) = p(y)$$

Example: exponential sampling

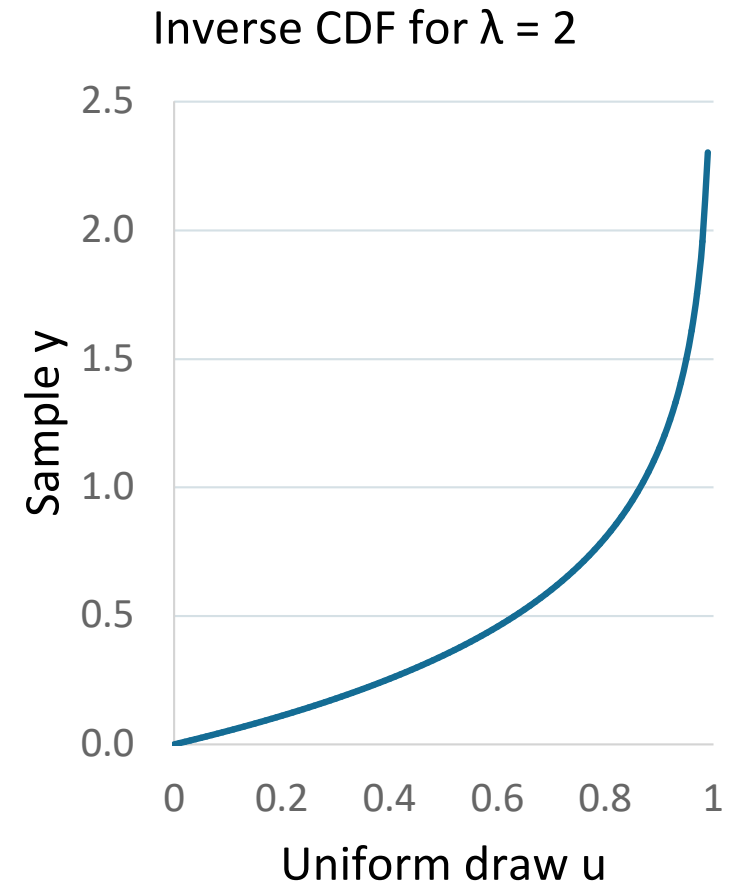
$$p(y) = \lambda e^{-\lambda y}, \quad y \geq 0$$

$$F(y) = 1 - e^{-\lambda y}$$

$$Y = -\frac{\ln(1 - U)}{\lambda}$$

$$\lambda = 2, U = 0.8$$

$$Y \approx 0.805$$



Gaussian sampling by scaling and shifting

$$Z \sim \mathcal{N}(0, 1)$$

$$X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$$

An affine transformation sets the mean and variance

$$E[X] = \mu$$

$$\text{Var}(X) = \sigma^2$$

Example: $\mu = 3$, $\sigma = 2$, $Z = -0.5$ gives $X = 2$.

Multivariate Gaussian sampling

Draw independent standard normals: $Z \sim N(0, I)$.

Cholesky decomposition of the covariance matrix:

$$\Sigma = LL^T$$

$$\mathbf{X} = \boldsymbol{\mu} + L\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Example: correlated components

$$X_1 = Z_1 \quad X_2 = 0.6Z_1 + 0.8Z_2$$

Both variances are 1, and $\text{Cov}(X_1, X_2) = 0.6$.

Rejection sampling

- Goal: generate samples from $p(z)$, where

$$p(z) = \frac{1}{Z_p} \tilde{p}(z).$$

- $\tilde{p}(z)$ can be evaluated.
- Z_p is unknown.

Example: computation of the MMSE estimate $\mathbb{E}(\theta|\mathbf{x})$

- If we can take samples from $P(\theta|\mathbf{x})$, $\mathbb{E}(\theta|\mathbf{x})$ can be estimated using the sample mean.
- But, in general, we cannot take samples from $P(\theta|\mathbf{x})$ since its normalization constant is unknown (i.e., difficult to compute).
- However,

$$P(\theta|\mathbf{x}) = \frac{P(\mathbf{x}|\theta)P(\theta)}{P(\theta)} = \frac{P(\mathbf{x}|\theta)P(\theta)}{\int P(\mathbf{x}|\theta')P(\theta')d\theta'}$$

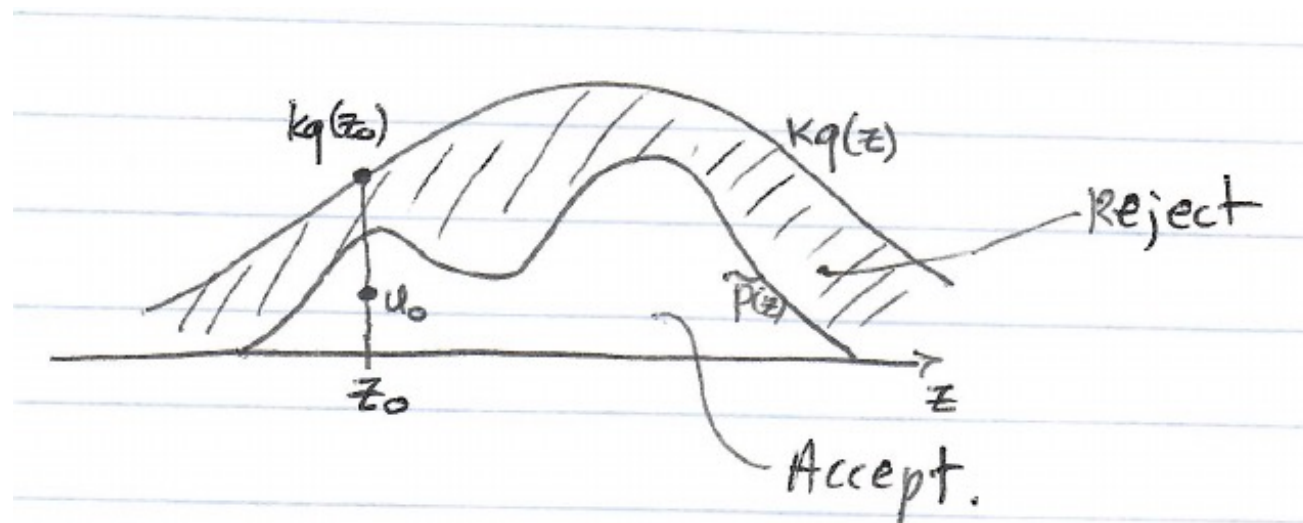
- Here, we can treat $p(z) = P(\theta|\mathbf{x})$, $\tilde{p}(\theta) = P(\mathbf{x}|\theta)P(\theta)$, and $Z_p = \int P(\mathbf{x}|\theta')P(\theta')d\theta'$.

- Suppose we can sample from $q(z)$ such that $kq(z) \geq \tilde{p}(z)$ for all z .

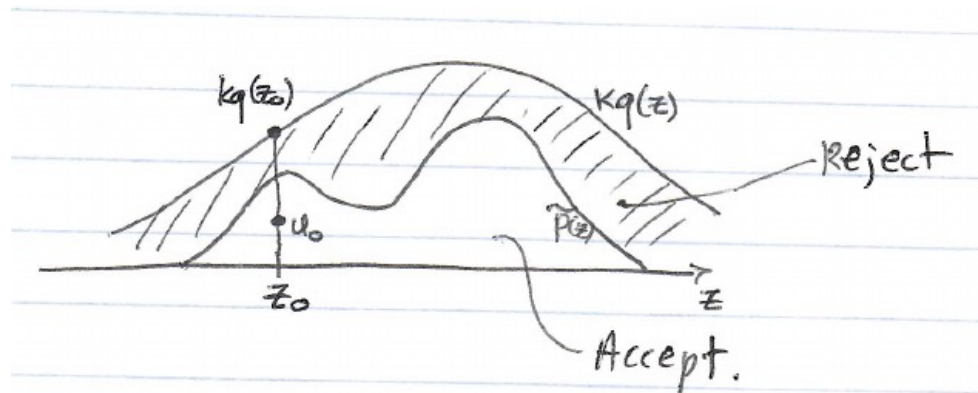
Rejection sampling

Algorithm:

- Sample z_0 from $q(z)$.
- Sample u_0 from $Unif(0, kq(z_0))$.
- Reject u_0 if $u_0 > \tilde{p}(z_0)$, otherwise accept u_0 (only accepted samples are used for estimation).



Rejection sampling



- Acceptance probability:

$$P(\text{accept}) = \int \frac{\tilde{p}(z)}{kq(z)} q(z) dz = \frac{1}{k} \int \tilde{p}(z) dz = \frac{Z_p}{k}$$

- Accepted samples

$$P(z \mid \text{accept}) = \frac{\mathbb{P}(\text{accept} \mid Z = z) q(z)}{\mathbb{P}(\text{accept})} = \frac{\frac{\tilde{p}(z)}{kq(z)} q(z)}{Z_p/k} = \frac{\tilde{p}(z)}{Z_p} = p(z)$$

- A tighter valid envelope yields fewer rejected proposals.

Importance sampling: change of measure

$$\mathbb{E}_p[g(X)] = \int g(x) p(x) dx$$

$$= \int g(x) \frac{p(x)}{q(x)} q(x) dx$$

$$= \mathbb{E}_q \left[g(X) \frac{p(X)}{q(X)} \right] \approx \frac{1}{N} \sum_{i=1}^N \frac{p(x_i)}{q(x_i)} g(x_i)$$

The last expectation uses $X \sim q$.

Required support: $q(x) > 0$ wherever $p(x) > 0$.

The importance sampling estimator

Draw $x_1, \dots, x_n \sim q$. Define an importance weight for each draw.

$$r_i = \frac{p(x_i)}{q(x_i)}$$

$$\hat{\mu}_{\text{IS}} = \frac{1}{N} \sum_{i=1}^N r_i g(x_i)$$

Weights correct each proposal draw's contribution

Unknown normalization constants

$$p(x) = \frac{\tilde{p}(x)}{Z_p}, \quad q(x) = \frac{\tilde{q}(x)}{Z_q} \quad \tilde{r}(x) = \frac{\tilde{p}(x)}{\tilde{q}(x)}$$

$$\mathbb{E}_p f(X) = \int f(x) \frac{p(x)}{q(x)} q(x) dx = \frac{Z_q}{Z_p} \int f(x) \frac{\tilde{p}(x)}{\tilde{q}(x)} q(x) dx \approx \frac{1}{N} \frac{Z_q}{Z_p} \sum \tilde{r}(x_i) f(x_i)$$

$$Z_p = \int \tilde{p}(x) dx = \int \frac{\tilde{p}(x)}{q(x)} q(x) dx = Z_q \int \frac{\tilde{p}(x)}{\tilde{q}(x)} q(x) dx \approx \frac{1}{N} Z_q \sum \tilde{r}(x_i)$$

$$\frac{1}{N} \frac{Z_q}{Z_p} \approx \frac{1}{\sum \tilde{r}(x_i)} \quad w_i = \frac{\tilde{r}(x_i)}{\sum \tilde{r}(x_j)}$$

$$\mathbb{E}_p f(X) \approx \frac{1}{\sum \tilde{r}(x_j)} \sum \tilde{r}(x_i) f(x_i) = \sum \frac{\tilde{r}(x_i)}{\sum \tilde{r}(x_j)} f(x_i) = \sum w_i f(x_i)$$

Self-normalized importance sampling

$$w_i = \frac{\tilde{r}_i}{\sum_{j=1}^N \tilde{r}_j}$$
$$\hat{\mu}_{\text{IS}} = \frac{\sum_{i=1}^N \tilde{r}_i g(x_i)}{\sum_{i=1}^N \tilde{r}_i} = \sum_{i=1}^N w_i g(x_i)$$

The normalized weights sum to 1.

Consistent under suitable conditions, but generally biased at finite N.

Sampling–importance–resampling (SIR)

1. Sample $x_1, \dots, x_n$ from q .
2. Compute and normalize their importance weights.

$$w_i = \frac{\tilde{p}(x_i)/q(x_i)}{\sum_{j=1}^N \tilde{p}(x_j)/q(x_j)}$$

3. Draw an index I with $P(I = i) = w_i$, then return x_i .

Repeat step 3 with replacement to obtain M resampled particles.

Resampling converts weighted particles into equally weighted draws.

Why SIR approaches the target

The CDF of one resampled particle, conditional on the pool, is

$$\begin{aligned} F_N(a) &= \sum_{i=1}^N w_i \mathbf{1}\{x_i \leq a\} \\ &= \frac{\sum_{i=1}^N \tilde{r}_i \mathbf{1}\{x_i \leq a\}}{\sum_{i=1}^N \tilde{r}_i} \end{aligned}$$

As the proposal count N grows

$$F_N(a) \xrightarrow{N \rightarrow \infty} \frac{\int_{-\infty}^a \frac{\tilde{p}(x)}{q(x)} q(x) dx}{\int \frac{\tilde{p}(x)}{q(x)} q(x) dx} = \frac{\int_{-\infty}^a \tilde{p}(x) dx}{\int \tilde{p}(x) dx} = \int_{-\infty}^a p(x) dx = F_p(a)$$

More resamples from a fixed pool do not remove the pool's error.