Rejection Sampling

Goal: Generate samples uniformly from some complicated distrib

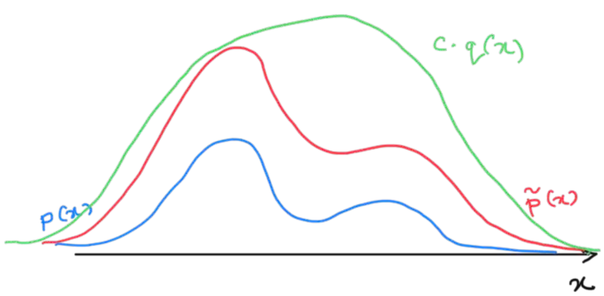
We assume that we can evaluate : $\mathbb{1}_{\{x \in A\}} = \begin{cases} 1, & \text{if} \\ 0, & \text{if } x \end{cases}$

Basic Idea: Draw samples from a simpler "proposal" distribution. Then evaluate some acceptance / rejection criterion to choose whether the sample should be kept or not.

In a more general setting our goal is to :

Sample $x_i \in \mathbb{R}^d$ from some pdf $p(x)$.

Assume we are given $\tilde{p}(x)$, $p(x) = \dfrac{\tilde{p}(x)}{Z_p}$, $Z_p := \int \tilde{p}(x)\,dx >$



Rejection sampling:

1. Choose a proposal dist. $q(x)$,

i.) $\exists c > 0 : c \cdot q(x) \geq \tilde{p}(x) \ \forall x$

ii.) $q(x)$ is easy to sample from.

2. Sample $x \sim q(x)$, sample $u \sim U[0, c \cdot q(x)]$.

3.) If $u \leq \tilde{p}(x)$ then accept this sample x. Otherwise, reject

4.) Go back to step #2 and repea

Output: A collection of accepted samples :

$x_1, x_2, \dots, x_m \sim p(x)$

Questions : 1. How to choose the constant c ? ... $c = \max \left(\dfrac{\tilde{p}(x)}{q(x)} \right.$

Remark : Our intuition on choosing an appropriate proposal $q(x$ breaks down in high-dimension !

Gibbs Sampling

Setup: Given some ML model with parameters $\theta := (\theta_1, \dots, \theta_d)$ and some data $\mathcal{D}$.

Goal: Generate samples from the posterior distribution $p(\theta|\mathcal{D})$.

Gibbs sampling:

1. Pick some initial $\theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_d^{(i)})$.

2. Sample:

$$\theta_1^{(i+1)} \sim p(\theta_1 | \theta_2^{(i)}, \theta_3^{(i)}, \dots, \theta_d^{(i)}, \mathcal{D})$$

conditional posterior distributi

$$\theta_2^{(i+1)} \sim p(\theta_2 | \theta_1^{(i+1)}, \theta_3^{(i)}, \dots, \theta_d^{(i)}, \mathcal{D})$$

$$\vdots$$

$$\theta_d^{(i+1)} \sim p(\theta_d | \theta_1^{(i+1)}, \theta_2^{(i+1)}, \dots, \theta_{d-1}^{(i+1)}, \mathcal{D})$$

3. Increment $i = i+1$, and repeat m times to generate m samples.

Pros: Does not need tuning any free parameters or choosing a propo

Cons: Assumes knowledge of the conditional densities which may be hard to derive in practice.

Example: Bayesian Linear regression. $\mathcal{D} := \{X, y\}$

$$y_i \overset{i.i.d}{\sim} \mathcal{N}(y_i | \alpha x + \beta, \gamma^{-1}) \iff y_i = \alpha x + \beta + \varepsilon, \ \varepsilon \sim \mathcal{N}(0,$$

Likelihood: $p(y | \alpha, \beta, X) = \prod^{n} \mathcal{N}(y_i | \alpha x + \beta, \gamma^{-1})$

Likelihood: $p(y|X,\alpha,\beta,\gamma) = \prod_{i=1} \mathcal{N}(y_i | \alpha x + \beta, \gamma)$

Priors: $\alpha \sim \mathcal{N}(0, I)$, $\beta \sim \mathcal{N}(0, I)$, $\gamma \sim \mathrm{Gam}(2, 1)$

Goal: Draw samples $\boxed{\theta \sim p(\theta|D) \propto p(D|\theta)\,p(\theta)}$

$$\theta := \{\alpha, \beta, \gamma\}$$

Gibbs sampling: Need to derive: $\begin{cases} p(\alpha|\beta,\gamma,x,y) \\ p(\beta|\alpha,\gamma,x,y) \\ p(\gamma|\alpha,\beta,x,y) \end{cases}$

General approach:

1. Write down the posterior conditional density in log form.
2. Throw away all terms that do not depend on the current sampling variable.
3. Pretend that what we are left with is the density of the current sampling variable, while all other variables are kept fixed.

Gibbs update for α:

$$\underbrace{p(\alpha|\beta,\gamma,x,y)}_{\text{conditional posterior}} \overset{\text{Bayes}}{\propto} \underbrace{p(y|x,\alpha,\beta,\gamma)}_{\text{likelihood}}\,\underbrace{p(\alpha)}_{\mathcal{N}(0,I)}$$

$$\log p(\alpha|\beta,\gamma,x,y) \propto -\frac{\gamma}{2}\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2 - \frac{1}{2}\alpha^2 \qquad \text{step \#1}$$

$\propto$ expand and throw away all terms that do not depend on α. step #2

$\propto$ re-arrange terms and try to derive a simple density for α, : e.g. step #3 known density

$$p(\alpha \mid \beta, \gamma, x, y) = \underbrace{\mathcal{N}(\alpha \mid ?, ?)}$$