Lecture 3. Rejection Sampling Chp 2.2.

Note Title 4/1/2010

First, exploit relationships between distributions

Second, rejection sampling.

Relationships Between Distributions

The Gamma and Beta distribution take the following forms: Gamma distribution

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

Beta distribution

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1].$$

These distribution are closely related:

$$P_{Beta}(z) = \iint \delta\left(z - \frac{x_1}{x_1+x_2}\right) P_{Gamma}(x_1) P_{Gamma}(x_2)\, dx_1 dx_2$$

$\delta(\,)$, Dirac delta function

So if x_1 & x_2 are random samples from $P_{Gamma}(\cdot)$ then $\frac{x_1}{x_1+x_2}$ is a random sample from $P_{Beta}(\cdot)$

If $P_{Gamma}(x_1)$ & $P_{Gamma}(x_2)$ have parameters α and β, then $P_{Beta}(z)$ has parameters α, β.

Can exploit these types of relationships

Rejection Method   von Neumann

Suppose we want to sample from a distribution $\pi(\underline{x})$, but don't know the normalization constant.

i.e. $\pi(\underline{x}) = \frac{l(\underline{x})}{c}$ $l(\underline{x})$ known, c unknown.
(computing $c = \sum_{x} l(\underline{x})$ may be intractible).

or $l(\underline{x}) = \pi(\underline{x})\, c$.

Find a sampling distribution $g(\underline{x})$ (i.e. one we can sample from, and is normalized and a "covering constant" M s.t.

$$M g(\underline{x}) \geq l(\underline{x}), \quad \forall \underline{x}.$$

Procedure:

(a) Draw a sample $\underline{x}$ from $g(\cdot)$ and compute $r = \frac{l(\underline{x})}{M g(\underline{x})}$.   (must be ≤ 1).

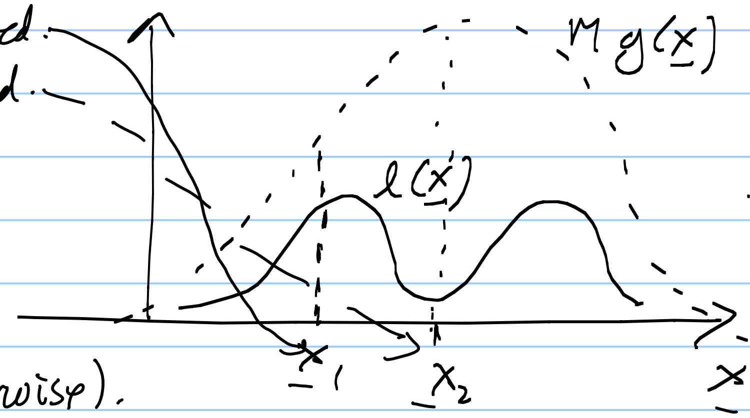
(b) Accept the sample with probability r, (otherwise reject it).

Note: most efficient (fewest rejects) if M is as small as possible.
Can measure the efficiency by counting the fraction of rejected samples.

Page 3

Why does Rejection Sampling work?

Sample at x_1 is probably accepted.
Sample at x_2 is probably rejected.



Let $I(\underline{x})$ be an indicator variable, so that $I(\underline{x})=1$ if sample $\underline{x}$ is accepted ($=0$, otherwise).

Bayes Theorem: $P(x|I) = \dfrac{P(I|x)\,p(x)}{p(I)}$

Proof
$P(x,I) = P(I|x)\,p(x)$
$= P(x|I)\,p(I)$

The distribution over accepted samples is

$$P(\underline{x}|I=1) = \frac{P(I=1|\underline{x})\,P(\underline{x})}{P(I=1)}$$

$$P(I=1|\underline{x}) = \frac{\ell(\underline{x})}{M\,g(\underline{x})}, \quad P(\underline{x}) = g(\underline{x}).$$

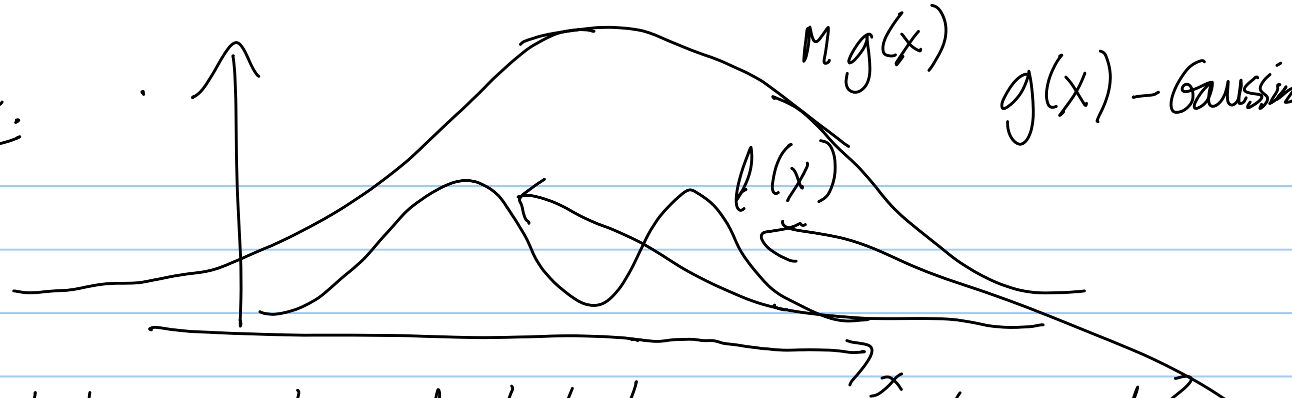
$$P(I=1) = \int P(I=1|\underline{x})\,P(\underline{x})\,d\underline{x} = \int \frac{\ell(\underline{x})}{M\cancel{g(\underline{x})}}\,\cancel{g(\underline{x})}\,d\underline{x} = \frac{c}{M}.$$

Hence, $$P(\underline{x}|I=1) = \frac{\ell(\underline{x})}{\cancel{M}\cancel{g(\underline{x})}}\cdot\cancel{g(\underline{x})}\,\frac{\cancel{M}}{c} = \frac{\ell(\underline{x})}{c} = \pi(\underline{x})$$

what we want.

Problem with this method. If c/M is small, then most samples are rejected, which is inefficient.

Page 4.



What sampling distribution would be good?
A Gaussian is easy to sample from, but might not be efficient if $\ell(x)$ has two peaks.
Most samples from the Gaussian will occur in between the two peaks of $\ell(x)$ – and get rejected.
A better choice is a mixture of Gaussians

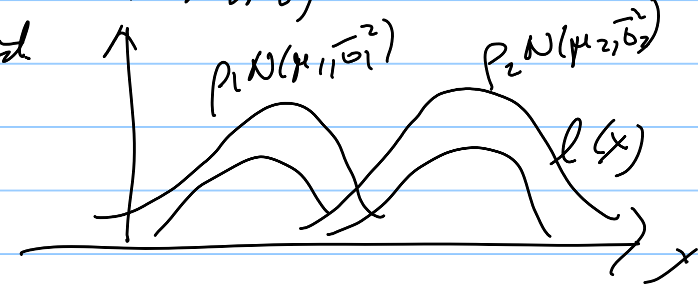
$$g(x) = p_1 N(\mu_1, \sigma_1^2) + p_2 N(\mu_2, \sigma_2^2)$$

$p_1 + p_2 = 1, \quad p_1 > 0, p_2 > 0$

Gaussian $N(\mu, \sigma^2)$ mean, variance

Sample from $g(x)$ in two stages:
stage (i): sample from (p_1, p_2) (biased coin) to select $N(\mu_i, \sigma_i^2)$ $i = 1$ or 2
(ii) sample from $N(\mu_i, \sigma_i^2)$

Better bound – more efficient



Page 5) Example of Rejection Sampling.

Let $\pi(x) \propto N(0,\sigma^2)\, I_{(x \geq c)}$

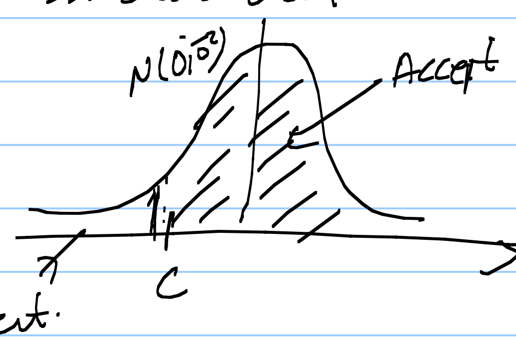
zero mean Gaussian

Set $\ell(x) = N(0,\sigma^2)$

$I_{(x\geq c)} = 1$, if $x \geq c$
$= 0$, if $x < c$.

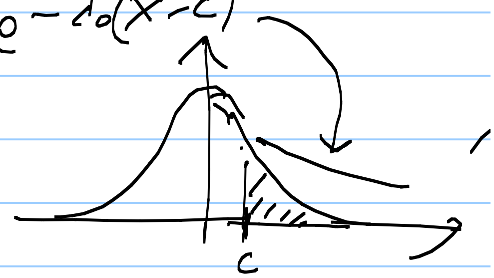
How to sample?

If $c < 0$. Draw samples X from $g(x) = N(0,\sigma^2)$ and $M = 1$, implies reject samples smaller than c and accept samples bigger than c.
(i.e. $r = 0,\ x < c,\ r = 1,\ x > c$)



The efficiency (proportion of samples not rejected) is better than 50%

For $c > 0$, we need a more efficient method. Bound it by $M \lambda_0 e^{-\lambda_0 (x-c)}$



Book: gives a formula for the optimal choice of M (to maximize efficiency).
Is it correct? (Homework Assignment!)