Lecture 2. Chp 2.1.

Note Title 4/2/2006

· Critical Issues of Monte Carlo.

(A) Can you sample efficiently from the distribution? If not, how can you get samples?

(B) The error goes as $\frac{1}{\sqrt{m}}\sigma$

where $\sigma^2 = \mathrm{Var}_\pi\, g(x)$.

Want to make σ as small as possible.

How to Sample from any distribution:

(Caveat: if the distribution is represented in a specific form)

Assume you have a uniform pseudo-random number generator (computer programs)

Initial value u_0 (seed)

Output, sequence of values $u_i = \varpi^i(u_0)$ in $[0,1]$

Samples $(u_1 \ldots u_n)$ will be i.i.d. samples from the uniform distribution on $[0,1]$.

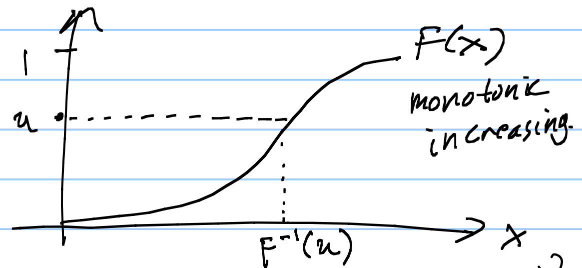
(2) Inversion Method: from uniform to (almost) any other distribution in 1-dimension.

(Liu's book) Lemma 2.1.1 Let $u \sim \text{Uniform}[0,1]$, F is a one-dimensional cummulative distribution function (cdf) — i.e. $F(x) = \int_0^x p(x)dx$ for some density function $p(x)$, — then $X = F^{-1}(u)$ has c.d.f. $F(x)$.

This allows us to sample from $p(x)$ provided we can sample from uniform $[0,1]$ and can compute $F^{-1}(.)$

Note: $F^{-1}(u) = \inf\{x : F(x) \geq u\}$

$F(x)$ is a monotonic funtion $\Rightarrow$ since $p(x) \geq 0$, for all x



If $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ is a Gaussian

then $F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} dy$ is the error function.

Easy to compute $F^{-1}(u)$ — so if we sample from the uniform distribution over $[0,1]$ then we can get samples from a 1-D. Gaussian

To Summarize.

i.i.d samples $u_1, \ldots, u_n$ from Unif $[0,1]$ correspond to i.i.d. samples $F^{-1}(u_1), \ldots, F^{-1}(u_n)$ from the distribution with probability density $p(x)$ and cdf. $F(x) = \int_a^x p(y)\,dy$

Note: 'a' choosen so that $p(y) = 0$ if $y < a$.

(3) Proof of Inversion Lemma

The samples $F^{-1}(u_1), \ldots, F^{-1}(u_n)$ are from the distribution:

$$P_r(X=x) = \int_0^1 \delta(x - F^{-1}(u))\, du.$$

Dirac

$\delta(x-a)$ is the delta function

$\delta(x-a) = 0$, for $x \neq a$

$\int_D \delta(x-a)\, dx = 1$, if $a \in D$

Chain Rule

$$P(x) = \int_0^1 P(x|u) P(u)\, du$$

$$P(x|u) = \delta(x - F^{-1}(u)) \quad P(u) = 1$$

$P_r(X=x)$ has cdf

$$\int_{x=a}^{x_1} P_r(X=x)\,dx = \int_{x=a}^{x_1} dx \int_0^1 du\, \delta(x - F^{-1}(u))$$

$$= \int_0^1 du \left\{ \int_{x=a}^{x_1} dx\, \delta(x - F^{-1}(u)) \right\}_1 = F(x_1).$$

Because

$$\int_{x=a}^{x_1} dx\, \delta(x - F^{-1}(u)) = 1, \text{ if } F^{-1}(u) \in [0, x_1]$$
$$= 0, \text{ otherwise.}$$

i.e. integrand is 1 provided $u \leq F(x_1)$

$F^{-1}(0) = 0 \in [a, x_1]$

$F^{-1}(F(x_1)) = x_1 \in [a, x_1]$

Note: understanding mathematical proofs like this is not essential to the course, but it does give deeper understanding of the methods

(4) Sampling from a Gaussian in higher dimensions.

A Gaussian in m dimension can be expressed as:

$$p(\underline{x} : \underline{\mu}, \underline{\underline{\Sigma}}) = \frac{1}{(2\pi)^{m/2}\sqrt{\det \underline{\underline{\Sigma}}}} \, e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\underline{\Sigma}}^{-1}(\underline{x}-\underline{\mu})}$$

<u>where</u> $\underline{\mu}$ is the mean — an m-dim vector

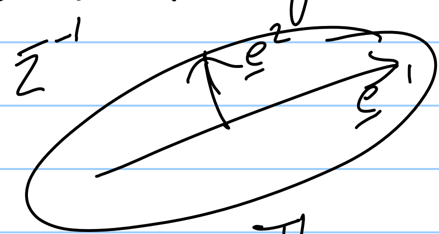
$\underline{\underline{\Sigma}}^{-1}$ is the covariance — an $m \times m$ matrix

<u>We</u> can <u>diagonalize</u> the covariance $\underline{\underline{\Sigma}}^{-1}$ by solving the eigenvector equations.

$$\underline{\underline{\Sigma}}^{-1} \underline{e}^{\mu} = \lambda^{\mu} \underline{e}^{\mu}$$

$\mu = 1$ to m, $\underline{e}^{\mu}$ eigenvector, λ^{μ} eigenvalue.

We can re-express the Gaussian as a product of 1-D Gaussians by changing to the coordinates defined by the eigenvectors



$$P(\underline{y}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{(y_i-\mu_i)^2}{2\sigma_i^2}}$$

Then we can sample from the Gaussian by sampling for each 1-D component separately.

(5) Sampling from a Mixture of Gaussians.

Mixture $p(\underline{x}) = \sum_{i=1}^{n} p_i G(\underline{x}; \underline{\mu}_i, \underline{\underline{\Sigma}}_i)$

$\sum_{i=1}^{n} p_i = 1$, $p_i > 0$, for all i.

$G(\underline{x}; \underline{\mu}_i, \underline{\underline{\Sigma}}_i)$ is a Gaussian in m-dimension with mean $\underline{\mu}_i$ and covariance $\underline{\underline{\Sigma}}_i$.

To sample from $p(x)$

(o) first sample from $(p_1, p_2, \ldots p_n)$
 - i.e. pick component i with probability p_i

(o) second sample from $G(\underline{x}: \underline{\mu}_i, \underline{\underline{\Sigma}}_i)$ (previous page)

This means that we can sample from a mixture of distributions in m-dimensions

Fact → any distribution in m-dimension can be approximated arbitrarily accurately by a mixture of Gaussians.

Conclusion → you now know how to sample from any distribution!
(Inversion Method, Factorized Gausian, Mixture of Gaussians)

But, the distributions you need to sample from will not usually be represented in this form.
 - or sampling would be too easy.

Key Point — some distributions are expressed in a form which makes it very easy to sample from them.

There are methods, like importance sampling, which allow us to exploit this. We can sample from an "easy distribution" and make an adjustment to get a sample from another, more difficult, distribution

(6)

Note Title 4/2/2006

In many applications, we will want to sample for Markov Random Fields (MRF)
(E.G. Genetics, BioInformatics, A.I)

These distributions are of form

$$\pi(\underline{x}) = \frac{1}{Z[T]} e^{-E(\underline{x})/T}$$

where $E(\underline{x})$ is a function that is easy to evaluate

E.G. $\underline{x} = (x_1, \dots, x_N)$ x_i binary-valued $\in \{\pm 1\}$

$$E(\underline{x}) = \sum_i x_i x_{i+1}.$$

But. $Z[T]$ is often very hard to evaluate.
so the distribution is known only up to a normalization constant.

This makes it difficult to sample from.

Two Strategies:

(1) Generate samples from a normalized distribution $p(\underline{x})$ close to $\pi(\underline{x})$ (if you can find one) and adjust.

(2) Markov Chain Monte Carlo. MCMC
(About 2/3 of the course).