STAT 535

'CLASSICAL' MONTE CARLO

A Prelude to MCMC

- Simulating Uniform Random Variables
- Onwards to Other Distributions
- Rejection Sampling
- Importance Sampling
- Sampling and Integration
- Application to Bayesian Inference

Uniform Random Variables

$X_1, \ldots, X_{10}$ *iid* $Unif(0,1)$.

```
x <- runif(10)
```

Q: How does a *purely deterministic* machine generate *truly random* output?

A: It doesn't!

Would you believe (something like) $x_i = z_i/(2^{32})$, where

$$z_{i+1} = 69069\ z_i \bmod 2^{32}$$

Lots of research on variants of this - *pseudo-random number generation* - want algorithm that passes every conceivable GOF test!

Other Univariate Distributions

Can simulate $U \sim Unif(0,1)$ - use this to build up ...

For instance, let $X = -\lambda \log(1 - U)$. Then $X \sim Exp(\lambda)$.

But, few univariate distributions have closed-form quantile functions.

No corresponding general approach for multivariate distributions.

Lots of specialized algorithms. General strategies?

Rejection Sampling

Have algorithm to simulate from the density *proportional to* $f(x)$.

Want algorithm to simulate from the density *proportional to* $g(x)$.

Can find c such that $g(x)/ \leq cf(x)$ for all x.

```
Repeat
  X <- draw from f
  U <- draw from Unif(0,1)
Until U < {g(X)/cf(X)}
Return X
```

Proof?
When will this work well?
Always easy to find such a c?

Importance Sampling

Have algorithm to simulate from the density *proportional to* $f(x)$.

Want to *'learn about'* the distribution with density *proportional to* $g(x)$.

Here 'learn about' really means compute expectations.

$$
\begin{aligned}
E_{g^*}\{h(X)\} &= \frac{\int h(x)g(x)dx}{\int g(x)dx} \\
&= \frac{\int h(x)\{g(x)/f(x)\}f(x)dx}{\int \{g(x)/f(x)\}f(x)dx} \\
&= \frac{E_{f^*}\{h(X)w(X)\}}{E_{f^*}\{w(X)\}},
\end{aligned}
$$

where $w(x) = g(x)/f(x)$.

This suggests an easy estimator of $E_{g^*}\{h(X)\}$ —ratio of two (sample) means.

Note there is no need to bound g!

One Monte Carlo sample or two?

Weighted sample interpretation?

Error assessment?

When will this work well?

Toy example: Have Unif(0,1), Want Beta(5,3)

```
set.seed(13); m <- 4000; samp <- runif(m)
wht <- samp^4*(1-samp)^2; wht <- wht/sum(wht)

print(round(summary(m*wht),3))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.146   0.850   1.000   1.863   2.296

print(c(mean(samp), sqrt(var(samp))), digits=3)
[1] 0.499 0.288

tmp1 <- sum(wht*samp); tmp2 <- sum(wht*samp^2)
print(c(tmp1, sqrt(tmp2-tmp1^2)),digits=3)
[1] 0.627 0.161
```


Integration versus Sampling

Note the connection between integration, learning about distributions, and sampling.

Generally speaking, numerical integration suffers the curse-of-dimensionality

If k gridpoints suffices for 1-D, need k^2 for 2-D, etc.

With m function evaluations can achieve error of $O(m^{-1/d})$ in computing an integral over a subset of $\mathbb{R}^d$.

But

$$
\int_{\mathbb{R}^d} h(x)dx = E_{f^*}\{h(X)/f^*(X)\}
$$

which we can approximate with a sample mean of $h(X)$ from an *iid* sample of size m from f^*.

Think CLT: error is $O_p(m^{-1/2})$ *regardless of* d. Where did the curse go? *Remarkable*!!!

Practical??? Shape of f relative to h?

Error assessment is trivial. Just report a standard error along with the mean.

To cloud the situation, *Quasi Monte Carlo* methods (which are deterministic) can achieve $O(m^{-1}(\log m)^d)$. Practical?

Current wisdom: MC is the only hope for high-dimensional integration.

The Bayesian Application

Have a model for data which gives rise to a likelihood function for the parameter vector θ.

Have a prior distribution which summarizes the pre-experiment belief about θ.

Have Bayes theorem to yield the posterior distribution which summarizes the post-experiment belief about θ.

$$f(\theta|\text{data}) = \frac{f(\text{data}|\theta)f(\theta)}{\int f(\text{data}|\theta^*)f(\theta^*)d\theta^*}$$

Inference about a scalar quantity of interest $\psi = h(\theta)$?

Point estimate: mean, median, or mode of the posterior (marginal) distribution of $h(\theta)$.

Interval estimate: interval have 0.95 probability under this dist.

We can evaluate an unnormalized version of the density of interest, so in principle Monte Carlo methods such as rejection sampling or importance sampling (or others) could be applied to compute these estimates.

In principle we can compute these quantities as precisely as we like (by making the MC sample size m larger).

Potential point of confusion:

Statistical Error: Difference between $E(\psi|\text{data})$ and ψ. In nice models goes to zero as n increases, with rate $n^{-1/2}$. But in most real problems increasing n costs \$\$\$\$\$\$\$\$\$\$.

Monte Carlo or *Simulation* Error: Difference between Monte Carlo estimate (perhaps mean of $h(\theta^{(1)}), \ldots, h(\theta^{(m)})$) and $E(\psi|\text{data})$. Goes to zero as m increases, with rate $m^{-1/2}$. In most real problems increasing m is dirt cheap.