Reduction of Variance

As we discussed earlier, the statistical error goes as:
error = sqrt(variance/computer time).

DEFINE: **Efficiency = ζ = 1/vT**

v = *error² of mean* and T = *total CPU time*

How can you make simulation more efficient?
- Write a faster code,
- Get a faster computer
- Work on *reducing the variance*.
- Or all three

We will talk about the third option:
Importance sampling and *correlated sampling*

Importance Sampling

Given the integral: $I = \int dx f(x) = \langle f(x) \rangle$

How should we sample *x* to maximize the efficiency?

Transform the integral: $I = \int dx\; p(x)\left[\frac{f(x)}{p(x)}\right] = \left\langle \frac{f(x)}{p(x)} \right\rangle_p$ Estimator

variance is: $v \approx \sigma^2 = \left\langle \left[\frac{f(x)}{p(x)} - I\right]^2 \right\rangle_p = \int dx p(x) \left[\frac{f(x)}{p(x)}\right]^2 - I^2$

Optimal sampling: $\frac{\delta v}{\delta p(x)} = 0 \quad with \quad \int dx\; p(x) = 1$

Mean value of estimator *I is independent of p(x), but variance v is not!*
Assume CPU-time/sample is independent of p(x), and vary p(x) to minimize v.

Finding Optimal p*(x) for Sampling

Trick to parameterize as a positive definite PDF:

$$p(x) = \frac{[q(x)]^2}{\int dx[q(x)]^2}$$

Solution: $$p^*(x) = \frac{|f(x)|}{\int dx|f(x)|}$$

Estimator: $$\frac{f(x)}{p^*(x)} = \frac{sign(f(x))}{\int dx|f(x)|}$$

1. If f(x) is entirely positive or negative, estimator is constant. "zero variance principle."
2. We can't generally sample $p^*(x)$, because, if we could, then we would have solved problem analytically! But the *form* of *p*(x) is a guide to lowering the variance*.
3. Importance sampling is a general technique: *it works in any dimension*.

Example of importance sampling

Suppose f(x) was given by

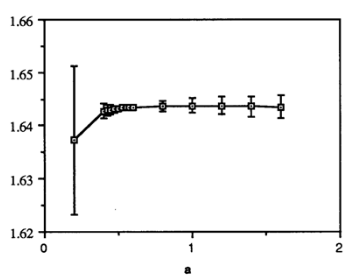
$$f(x) = \frac{e^{-x^2/2}}{1+x^2}$$

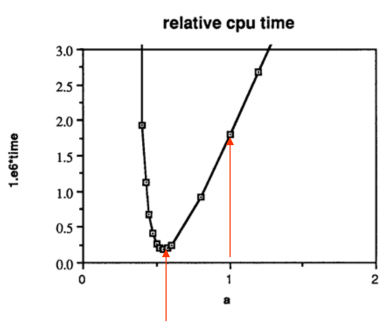
Optimize a Gaussian

$$p(x) = \frac{e^{-x^2/2a}}{(2\pi a)^{1/2}}$$

Value is independent of a.

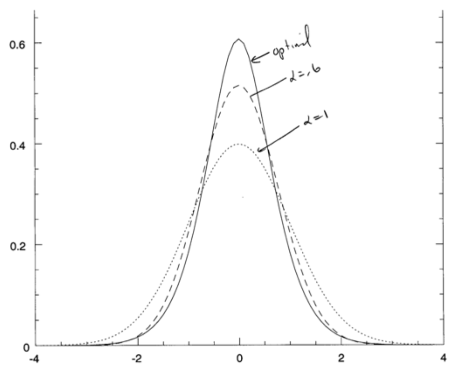
CPU time is not



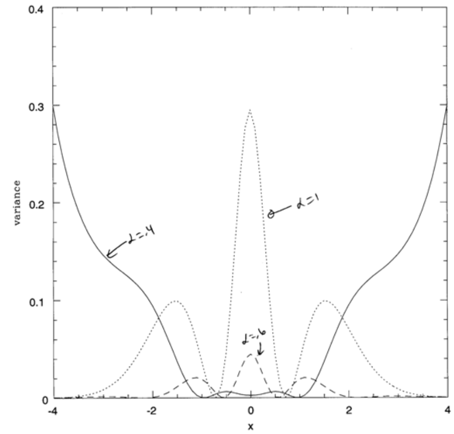


$$v \approx \sigma^2 = \int dx\ p(x)\left[\frac{f(x)}{p(x)} - I\right]^2 = \int dx \left[\frac{e^{-x^2(1-1/2a)}}{(1+x^2)^2}\right] - I^2$$

Importance sampling functions



Variance integrand



What are allowed values of a?

- Clearly for p(x) to exist: 0<a

0 .5 .6 1. a

- For finite estimator 1<a
- For finite variance .5<a
- "Obvious" value a=1
- **Optimal value** a=0.6.

$$v = \int dx\ \frac{f(x)^2}{p(x)} = \int dx \frac{(2\pi a)^{1/2}}{1+x^2}\ e^{-x^2(1-1/2a)}$$

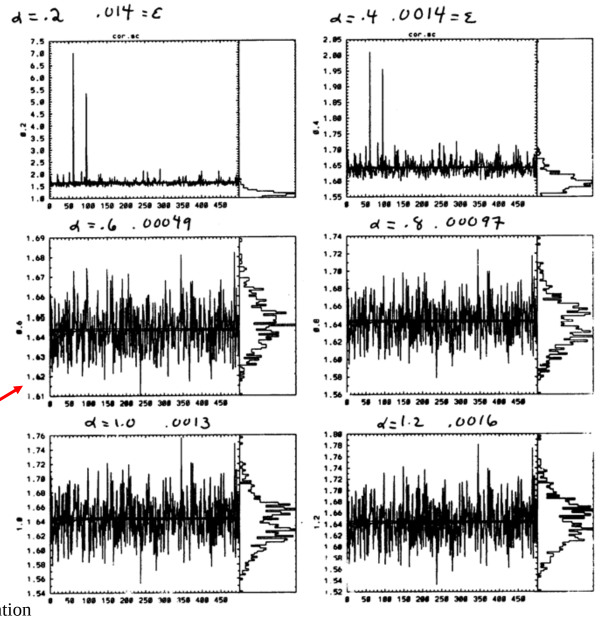
What does infinite variance look like?

Spikes

Long tails on the distributions

Near optimal

Why (visually)?



General Approach to Importance Sampling

- Basic idea of importance sampling is to sample more in regions where function is large.
- Find a convenient approximation to |f(x)|.

- ***Do not under-sample*** -- could cause infinite variance.
- ***Over-sampling*** -- loss in efficiency but not infinite variance.
- *Always derive conditions for finite variance analytically.*
- To debug: *test that estimated value is independent of important sampling*.

- *Sign problem*: zero variance is not possible for oscillatory integral.

"Monte Carlo can add but not subtract."

Correlated Sampling

Suppose we want to compute a function : $G(F_1,F_2)$
where F's are integrals: $F_k = \int dx\, f_k(x)$

Suppose we use same p(x) and same random numbers to do both integrals.

What is optimal p(x)? $$p^*(x) \propto \left| f_1(x)\frac{dG}{dF_1} + f_2(x)\frac{dG}{dF_2} \right|$$

It is a weighted average of the distributions for F_1 and F_2.

Sampling Boltzmann distribution

- Suppose we want to calculate a whole set of averages:

$$< O_k >= \frac{\int dR\, O_k(R) e^{-V(R)/kT}}{\int dR\, e^{-V(R)/kT}}$$

- Optimal sampling is: $$p_k^*(x) \propto \left| e^{-V(R)/kT} (O_k(R) - < O_k >) \right|$$

constant variable

- We need to sample the first factor
  because we want lots of properties
- **Avoid undersampling**.
- The Boltzmann distribution is *very highly peaked*.

Independent Sampling for exp(-V/kT)?

- Try hit or miss MC to get   Z= exp(-V/kT).
- Sample R uniformly in (0,L): $P(R) = \Omega^{-N}=1$

What is the variance of the free energy and how does it depend on the number of particles?

$$Var(Z) = \int dR \left[e^{-\beta V(R)} - Z(\beta) \right]^2 = Z(2\beta) - Z(\beta)^2$$

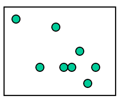
$$\mathrm{var}(\beta F) = \mathrm{var}(Z)/Z^2 = \frac{Z(2\beta)}{Z(\beta)^2} - 1 = e^{2\beta[F(\beta)-F(2\beta)]} - 1$$

$$error(\beta F) = e^{\beta[F(\beta)-F(2\beta)]}$$

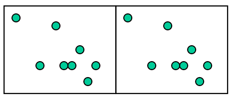
- Blows up exponentially fast at large N since F is extensive!
- The number of sample points must grow exponentially in N, just like a grid based method.

Intuitive explanation

- Throw N points in a box, area A.
- Say *probability of no overlap is q*.



- Throw 2N points in a box, area 2A.
- Probability of no overlap is q^2.



- Throw mN points in a box, area mA
- Probability of no overlap is q^m.

Probability of getting a success is p=exp(m ln(q)). Success defined as a reasonable sample of a configuration.

This is a general argument. We need to sample only near the peak of the distribution: to do so use random walks.