Monte Carlo

Given rvs $X_1, \ldots, X_n$; distbn specified.

Statistic $T(X_1, \ldots, X_n)$ whose dstbn wanted.

To compute $P(T > t)$:

1. Generate $X_1, \ldots, X_n$ from the density f.
2. Compute $T_1 = T(X_1, \ldots, X_n)$.
3. Repeat N times getting $T_1, \ldots, T_N$.
4. Estimate $p = P(T > t)$ as $\widehat{p} = M/N$ where M is number of repetitions where $T_i > t$.
5. Estimate accuracy of $\widehat{p}$ using $\sqrt{\widehat{p}(1 - \widehat{p})/N}$.

Note: accuracy inversely proportional to $\sqrt{N}$.

Next: tricks to make method more accurate.
Warning: tricks only change constant — SE still inversely proportional to $\sqrt{N}$.

Generating the Sample

Transformation

Basic computing tool: pseudo uniform random numbers — variables U which have (approximately) a Uniform$[0,1]$ distribution.

Other dstbns generated by transformation:

Exponential: $X = -\log U$ has an exponential distribution:

$$\begin{aligned} P(X > x) &= P(-\log(U) > x) \\ &= P(U \le e^{-x}) = e^{-x} \end{aligned}$$

Pitfall: Random uniforms generated on computer sometimes have only 6 or 7 digits.

Consequence: tail of generated dstbn grainy.

Explanation: suppose U multiple of 10^{-6}.

Largest possible value of X is $6\log(10)$.

Improved algorithm:

- Generate U a Uniform[0,1] variable.
- Pick a small ϵ like 10^{-3} say. If $U > \epsilon$ take $Y = -\log(U)$.
- If $U \le \epsilon$: conditional dstbn of $Y - y$ given $Y > y$ is exponential. Generate new U'. Compute $Y' = -\log(U')$. Take $Y = Y' - \log(\epsilon)$.

Resulting Y has exponential distribution.

Exercise: check by computing $P(Y > y)$.

General technique: inverse probability integral transform.

If Y is to have cdf F:

Generate $U \sim Uniform[0,1]$.

Take $Y = F^{-1}(U)$:

$$\begin{aligned} P(Y \le y) &= P(F^{-1}(U) \le y) \\ &= P(U \le F(y)) = F(y) \end{aligned}$$

Example: X exponential. $F(x) = 1 - e^{-x}$ and $F^{-1}(u) = -\log(1-u)$.

Compare to previous method. (Use U instead of $1 - U$.)

Normal: $F = \Phi$ (common notation for standard normal cdf).

No closed form for F^{-1}.

One way: use numerical algorithm to compute F^{-1}.

Alternative: Box Müller

Generate U_1, U_2 two independent Uniform[0,1] variables.

Define

$$Y_1 = \sqrt{-2\log(U_1)}\cos(2\pi U_2)$$

and

$$Y_2 = \sqrt{-2\log(U_1)}\sin(2\pi U_2)\,.$$

Exercise: (use change of variables) Y_1 and Y_2 are independent $N(0,1)$ variables.

Acceptance Rejection

Suppose: can't calculate F^{-1} but know f.

Find density g and constant c such that

1) $f(x) \le cg(x)$ for each x and

2) G^{-1} is computable or can generate observations $W_1, W_2, \ldots$ independently from g.

Algorithm:

1) Generate W_1.

2) Compute $p = f(W_1)/(cg(W_1)) \le 1$.

3) Generate uniform[0,1] random variable U_1 independent of all Ws.

4) Let $Y = W_1$ if $U_1 \le p$.

5) Otherwise get new W, U; repeat until you find $U_i \le f(W_i)/(cg(W_i))$.

6) Make Y be last W generated.

This Y has density f.

Markov Chain Monte Carlo

Recently popular tactic, particularly for generating multivariate observations.

Theorem Suppose $W_1, W_2, \ldots$ is an (ergodic) Markov chain with stationary transitions and the stationary initial distribution of W has density f. Then starting the chain with *any* initial distribution

$$\frac{1}{n}\sum_{i=1}^{n} g(W_i) \to \int g(x)f(x)dx\,.$$

Estimate things like $\int_A f(x)dx$ by computing the fraction of the W_i which land in A.

Many versions of this technique including Gibbs Sampling and Metropolis-Hastings algorithm.

Technique invented in 1950s: Metropolis et al.

One of the authors was Edward Teller "father of the hydrogen bomb".

Importance Sampling

If you want to compute

$$\theta \equiv E(T(X)) = \int T(x)f(x)dx$$

you can generate observations from a different density g and then compute

$$\widehat{\theta} = n^{-1} \sum T(X_i)f(X_i)/g(X_i)$$

Then

$$\begin{aligned} E(\widehat{\theta}) &= n^{-1} \sum E\left\{T(X_i)f(X_i)/g(X_i)\right\} \\ &= \int \{T(x)f(x)/g(x)\}g(x)dx \\ &= \int T(x)f(x)dx \\ &= \theta \end{aligned}$$

Variance reduction

Example: estimate dstbn of sample mean for a Cauchy random variable.

Cauchy density is

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Generate $U_1, \ldots, U_n$ uniforms.

Define $X_i = \tan^{-1}(\pi(U_i - 1/2))$.

Compute $T = \bar{X}$.

To estimate $p = P(T > t)$ use

$$\hat{p} = \sum_{i=1}^{N} 1(T_i > t)/N$$

after generating N samples of size n.

Estimate is unbiased.

Standard error is $\sqrt{p(1-p)/N}$.

Improvement: $-X_i$ also has Cauchy dstbn.

Take $S_i = -T_i$.

Remember that S_i has same dstbn as T_i.

Try (for $t > 0$)

$$\tilde{p} = [\sum_{i=1}^N 1(T_i > t) + \sum_{i=1}^N 1(S_i > t)]/(2N)$$

which is the average of two estimates like $\hat{p}$.

The variance of $\tilde{p}$ is

$$(4N)^{-1}\text{Var}(1(T_i > t) + 1(S_i > t)) = (4N)^{-1}\text{Var}(1(|T| > t))$$

which is

$$\frac{2p(1-2p)}{4N} = \frac{p(1-2p)}{2N}$$

Variance has extra 2 in denominator and numerator is also smaller – particularly for p near 1/2.

So need only half the sample size to get the same accuracy.

Regression estimates

Suppose $Z \sim N(0,1)$. Compute

$$\theta = E(|Z|)\,.$$

Generate N iid $N(0,1)$ variables $Z_1, \ldots, Z_N$.

Compute $\hat\theta = \sum |Z_i|/N$.

But know $E(Z_i^2) = 1$.

Also: $\hat\theta$ is positively correlated with $\sum Z_i^2/N$.

So we try

$$\tilde\theta = \hat\theta - c(\sum Z_i^2/N - 1)$$

Notice that $E(\tilde\theta) = \theta$ and

$$\begin{aligned}\text{Var}(\tilde\theta) = &\\ &\text{Var}(\hat\theta) - 2c\text{Cov}(\hat\theta, \sum Z_i^2/N) \\ &\qquad + c^2\text{Var}(\sum Z_i^2/N)\end{aligned}$$

The value of c which minimizes this is

$$c = \frac{\text{Cov}(\hat\theta, \sum Z_i^2/N)}{\text{Var}(\sum Z_i^2/N)}$$

Estimate c by regressing $|Z_i|$ on Z_i^2!