Sampling Distributions

Allen & Tildesley pgs. 345-351
and "Numerical Recipes" on random numbers

Today I explain how to generate a ***non-uniform probability distributions***.

These are used in *importance sampling* to reduce the variance of a Monte Carlo evaluation or to simulate various physical processes.

We start by assuming that there is software to generate udrn's (uniform RN) in the range (0,1).

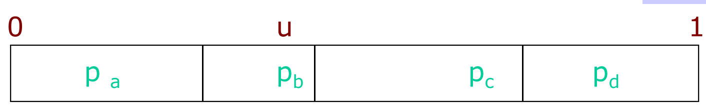
How do we sample an arbitrary p(x)dx ?

There are lots of tricks.

Discrete Distributions

Any discrete distribution p_k can be sampled by constructing the cumulant.

$$c_k = \sum_{i=1}^{k} p_i$$



- **Sample 0<u<1.**
- Find which region it is in, i.e. **find k**: $c_{k-1} < u < c_k$
- **Return label "k".**

- The search can be done by ***bisection*** in $\log_2(N)$ steps.
- For simple distributions, it might be even easier.
- There is a faster more complicated method (see notes).

Make a Table $c_k = c_{k-1} + p_k$ before simulation.

Continuous Distributions

Generalize this to a continuous function.
This is the mapping method.

- Construct the cumulant: $c(y) = \int_{-\infty}^{y} dx p(x)$
- Sample u in (0,1)
- Find $x = c^{-1}(u)$ *An inverse mapping from urnd to x.*
  x is sampled from p(x) dx, which is non-uniform.
  Can always perform this with a table lookup.

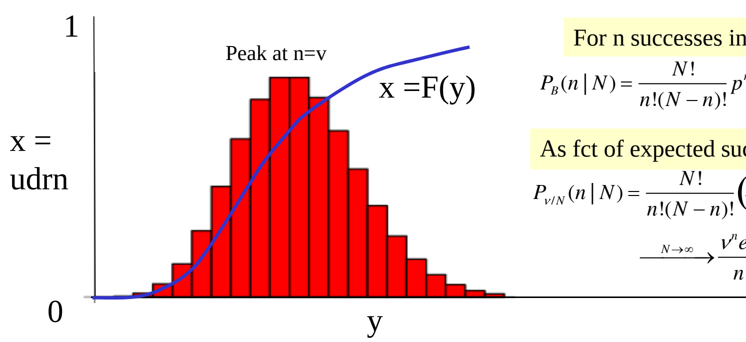
Some analytic examples:

- $P(x) = a\, e^{-ax}$ $0 < x$ then $x = -\ln(u)/a$
- $P(x) = (a+1)x^{a}$ $0 < x < 1$ then $x = u^{1/(a+1)}$

Problem if inverse mapping is difficult.

Example: Drawing from Poisson Distribution

$y = -\ln(x)/a$, $p(y) = ae^{-ay}$ and $x = F(y) = 1 - e^{-ay}$



For n successes in N trials

$$P_B(n\,|\,N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

As fct of expected successes υ=Np

$$P_{\nu/N}(n\,|\,N) = \frac{N!}{n!(N-n)!} \left(\frac{\nu}{N}\right)^n \left(1 - \frac{\nu}{N}\right)^{N-n} \xrightarrow{N\to\infty} \frac{\nu^n e^{-\nu}}{n!} \quad (\textit{indep of } N)$$

As F(y) is area under p(y), $y(x) = F^{-1}(x)$ prescribes that

- **Choose *x=(0,1]*, then find value y that has that *fraction x of area to left of y, or x=F(y)***
- **Return that value of y.**

Rejection technique

- Sample x from q(x)dx
- Accept x with probability c(x), otherwise repeat

What is distribution of accepted x's?

p(x)dx= c(x)q(x)dx/normalization

Hence choose c(x) = a p(x)/q(x)

Where **a** is set so that c(x) ≤ 1 so a ≤ min[q(x)/p(x)]

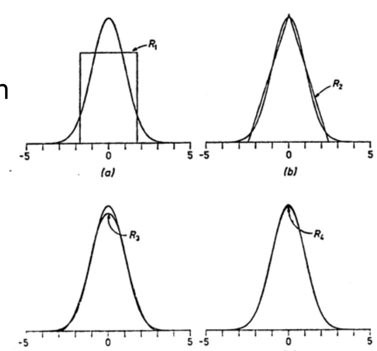
1 / a is the *acceptance probability*, the *efficiency*.

- Do not *under sample*, or else efficiency will be low.
- If inefficient, you will use lots of prns/step.

Composition method

Combine several random numbers

- **Add several udrn's.**
  remember characteristic function
  limit for large k is a Gaussian



Example: add two integers in (1,6)

Example: add 12 udrn, i.e

$$\varsigma = \sum_{i=1}^{12} u_i - 6 \quad u \in [0,1]$$

Gaussian distributed on (-6,6).

- Take maximum of 'k' udrns.
  x= max(u_1, .., u_k) Prove that P(x)= k x^{k-1}

Normal distribution

- Inverse mapping is a little slow, also of infinite range.
- Trick: Generate 2 prns at a time: **r**=(x,y)
  - Generate u_1 and u_2
  - Calculate $y_1 = \sqrt{-2\ln u_1}\cos 2\pi u_2 \qquad y_2 = \sqrt{-2\ln u_1}\sin 2\pi u_2$

Samples:

$p(x,y)\,dxdy = (2\pi)^{-1}\exp(-r^2/2) = p(r)\,r\,dr\,d\theta$

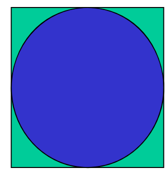
$p(v)\quad dv = 0.5*\exp(-v/2)$ with $v= r^2$

We can sample the angle using a *rejection technique*:

Sample (x,y) in square

Accept if $x^2+y^2 <1$

Normalize to get the correct r.



Example: Box-Mueller Code to sample a Normal Distribution

Normal distribution $\langle x\rangle=x_0$ and $\langle (x-x_0)^2\rangle=\sigma^2$

```
While(r2 > 0.25) then
 {x=sprng()-0.5, y=sprng()-0.5
    r2 = x*x+y*y}
radius= sigma*sqrt(-2*ln(sprng())/r2)
xnormal = x0 + x*radius
ynormal = y0 + y*radius
```

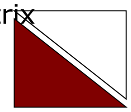
- No trig functions!
- Rejection mixes up regularity of random numbers
- Efficiency of angle generation is $4/\pi$.
- Will get 2 ndrn's each time.

Multivariate normal distributions

How to sample a correlated Gaussian? (say with D components)

- Assume we want $<x_i x_j> = T_{ij}$
- Take the square root of T : $SS^t = T$

  Make "Choleski" decomposition (see Numerical Recipes or notes). We can assume S is a triangular matrix all entries above diagonal are zero.



- Generate D normally distributed numbers y.
- Transform to correlated random distribution

$$x_\alpha = S_{\alpha i} y_i \qquad \langle y_i y_j \rangle = \delta_{ij}$$

$$\langle x_\alpha x_\beta \rangle = \langle S_{ai} y_i S_{\beta j} y_j \rangle = S_{ai} S_{\beta j} \langle y_i y_j \rangle = S_{ai} S_{\beta j} \delta_{ij} = SS^T = T$$

Bias

- Bias is a *systematic error* caused by using a random number in another expression.
- You will get a result systematically too high or low.
- Suppose $Z' = \bar{Z} + \delta Z$ is the result of MC sampling.
- But we want F(Z). *Example:* F=-kT ln(Z).
- What is the statistical error and bias of F(Z')?
- Expand Z in power series about <Z>

$$F(Z') = F(\bar{Z}) + \left.\frac{dF}{dZ}\right|_{\bar{Z}} \delta Z + \frac{1}{2} \left.\frac{d^2F}{dZ^2}\right|_{\bar{Z}} \delta Z^2 + \ldots$$

$$bias(F) = < F(Z) - F(\bar{Z}) > = \frac{1}{2} \left.\frac{d^2F}{dZ^2}\right|_{\bar{Z}} < \delta Z^2 > + \ldots = \frac{1}{2} \left.\frac{d^2F}{dZ^2}\right|_{\bar{Z}} err(Z)^2 \qquad \mathbf{O(N^{-1})}$$

$$error(F) = [< (F(Z') - \langle F(Z) \rangle)^2 >]^{1/2} = \left|\frac{dF}{dZ}\right|_{\bar{Z}} < \delta Z^2 >^{1/2} + \ldots = \left|\frac{dF}{dZ}\right|_{\bar{Z}} err(Z) \qquad \mathbf{O(N^{-1/2})}$$

You may need to correct for the bias unless N is very large.