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.

<table>
<thead>
<tr>
<th>0</th>
<th>u</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_a</td>
<td>p_b</td>
<td>p_c</td>
</tr>
</tbody>
</table>

- **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.**

---

2/20/2013
Continuous Distributions

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

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

Some analytic examples:
- \( \mu(x) = a \ e^{-ax} \quad 0 < x \quad \text{then} \quad x = -\ln(u)/a \)
- \( \mu(x) = (a+1)x^a \quad 0 < x < 1 \quad \text{then} \quad x = u^{1/(a+1)} \)

Problem if inverse mapping is difficult.

Example: Drawing from Poisson Distribution

\[ y = -\ln(x)/a, \quad \mu(y) = a e^{-ay} \quad \text{and} \quad x = F(y) = 1 - e^{-ay} \]

As \( F(y) \) is area under \( \mu(y) \), \( y(x) = F^{-1}(x) \) 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?

\[ p(x)dx = \frac{c(x)q(x)dx}{\text{normalization}} \]

Hence choose \[ c(x) = \frac{a p(x)}{q(x)} \]

Where \( a \) is set so that \[ c(x) \leq 1 \] so \( a \leq \min\frac{q(x)}{p(x)} \)

\( \frac{1}{a} \) 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

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

Gaussian distributed on (-6,6).

- Take maximum of ‘k’ udrns.
  \[ x = \max(u_1, \ldots, u_k) \]
  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 \( y_1 = \sqrt{-2 \ln u_1 \cos 2\pi u_2} \) \( y_2 = \sqrt{-2 \ln u_1 \sin 2\pi u_2} \)

Samples:
- \( p(x,y) \, dx \, dy = (2\pi)^{-1} \exp(-r^2/2) = p(r) \, r \, dr \, d\theta \)
- \( p(v) \, dv = 0.5 \exp(-v/2) \) 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 $T$ : $S S^T = T$

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

$\begin{align*}
    x_a &= S_{ai} y_i \\
    \langle x_a x_b \rangle &= \langle S_{ai} y_i S_{bj} y_j \rangle = S_{ai} S_{bj} \delta_{ij} = SS^T = T
\end{align*}$

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 $Z' = Z + \delta Z$ 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 $<Z>$

$\begin{align*}
    F(Z') &= F(Z) + \frac{dF}{dZ} \delta Z + \frac{1}{2} \frac{d^2F}{dZ^2} (\delta Z)^2 + L \\
    \text{bias}(F) &= < F(Z') - F(Z) > = \frac{1}{2} \left( \frac{d^2F}{dZ^2} \right)_Z (\delta Z)^2 + L = \frac{1}{2} \frac{d^2F}{dZ^2} \text{err}(Z)^2 = O(N^{-1}) \\
    \text{error}(F) &= \left[ (< F(Z') - <F(Z)> )^2 \right]^{1/2} = \left( \frac{dF}{dZ} \right)_Z < \delta Z^2 >^{1/2} + L = \frac{dF}{dZ} \text{err}(Z) = O(N^{-1/2})
\end{align*}$

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