HW2

1-20-15

Using random numbers to generate realizations from non-uniform probability distributions.

Use unif(0,1) realizations to generate realizations from the following distributions:

Discrete
1. Bernoulli   1 to 1   uniform to Bernoulli
2. Geometric
3. Binomial   n to 1   Bernoulli to Binomial
4. Poisson

5. Gamma
6. Normal
7. Cauchy

Let X be Bernoulli with parameter p.
Let $\underline{\underline{u}}$ be a unif(0,1) realization.

To simulate a value of X, assign

$X=1$ if $\underline{\underline{u}} < p$
otherwise
set $X=0$

$p=70\%$
→ prob. of success

Let X be Geometric with parameter p, then X can be thought of as the number of Bernoulli realizations necessary to obtain the first "1".

To simulate a value of X, first simulate a sequence of independent Bernoulli realizations, each with the same "success" probability p. Then count how many Bernoulli realizations were necessary to obtain the first "success", the first 1

Let X be binomial with parameters n and p. Then X can be thought of as the # of "successes" observed in n independent realizations.

To simulate a value of X, first simulate exactly n independent Bernoulli realizations, each with success probability p. The <u>sum</u> of these n Bernoulli realizations is the simulated value of X.

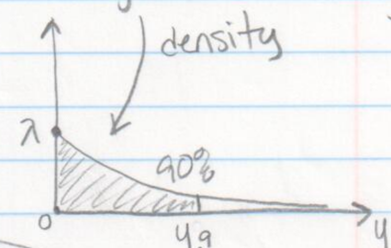
<u>Aside</u> Exponential Distribution

Inverse CDF Sampling Detour

Let Y be a r.v. with density $f(y)$, and CDF $F(y)$.
Let u be a unif(0,1) realization

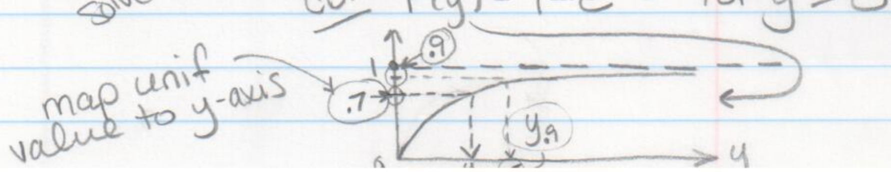
A realization of Y can be simulated as
$$y = F^{-1}(u)$$

<u>Example</u>: Let $Y \sim Exp(\lambda)$ $\quad f(y) = \lambda e^{-\lambda y}$ for $y > 0$



treat as u
solve for y

CDF $F(y) = 1 - e^{-\lambda y}$ for $y > 0$



To simulate a value of $Y \sim \exp(\lambda)$ using inverse CDF sampling:

★ Draw $u \sim \text{unif}(0,1)$

★ Compute inverse CDF: $\dfrac{\ln(1-u)}{-\lambda} = y$

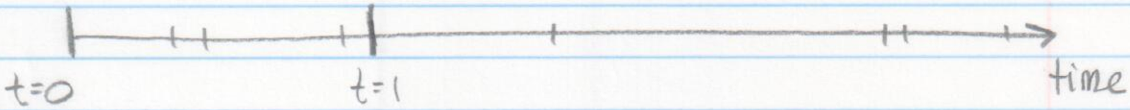
$$\Rightarrow \frac{\ln(u)}{-\lambda} = y$$

← saves a step: $(1-u)$ and u are both random var between 0 and 1.

Detour over

Let X be Poisson with parameter λ.
To simulate a value of X, we use the theory of Poisson process:

The # of observed arrivals in a time window of unit length follows a Poisson distribution. The interarrival times are independent and follow an exponential distribution.



To simulate a value of X: first simulate independent realizations Y from an exponential density with the same λ parameter. Let K be the minimum number of exponential realizations needed to satisfy

$$\sum_{i=1}^{K} Y_i > 1 \qquad \text{Set } X = K-1$$

$$u = \int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\,dt = \frac{1}{\pi}\int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\,dt = \frac{1}{\pi}(\arctan t)\Big|_{-\infty}^{y}$$

$$= \frac{1}{\pi}(\arctan y) - \frac{1}{\pi}(\arctan -\infty)$$

$$\lim_{t \to -\infty} \arctan t = -\frac{\pi}{2}$$

$$u = \frac{1}{\pi}\arctan y - \left[\frac{1}{\pi}\cdot\left(-\frac{\pi}{2}\right)\right] = \frac{1}{\pi}\arctan y + 1/2$$

$$\pi(u - 1/2) = \arctan y$$

$$y = \tan(\pi u - \pi 1/2)$$

Let $X_1, X_2, X_3, \dots, X_n$ be iid exponential (λ) r.v.s
Let $Y = \sum_{i=1}^{n} X_i$

Then $Y \sim \text{Gamma}(n, \lambda)$

use this result to simulate a realization from a $\text{Gamma}(n, \lambda)$ distribution
(sum of expontials equal Gamma)

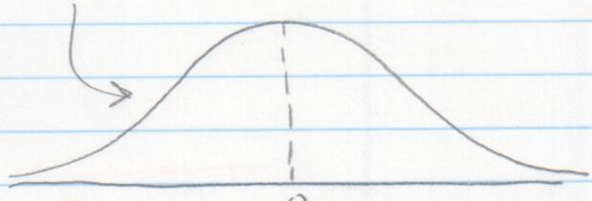
Let Y be a Cauchy r.v.
Then $f(y) = \dfrac{1}{\pi(1+y^2)}$ for $y \in \mathbb{R}$
Use inverse CDF sampling to simulate Y

$$F(y) = \int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\, dt$$

Set $u = \int_{-\infty}^{y} \frac{1}{\pi(1+t^2)}\, dt$ solve for y.

Let Y be normal with mean 0 and variance 1

$f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ for $y \in \mathbb{R}$ Rejection Sampling



Rejection Sampling
To simulate from a target density $f(y)$, first simulate a realization from a proposal density $g(y)$.
Note $g(y)$ must be chosen to satisfy $A \cdot g(y) \geq f(y)$ $\forall y$

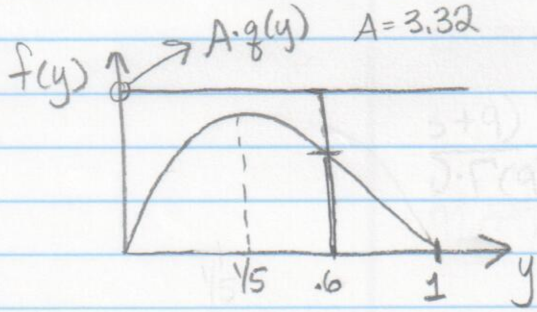
$$A \cdot \pi^{-1}(1+y^2)^{-1} \geq \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

$$A \geq \frac{\pi(1+y^2)}{\sqrt{2\pi}} e^{-y^2/2}$$

$$\frac{d}{dy}\left(\frac{\pi}{\sqrt{2\pi}}(1+y^2)e^{-y^2/2}\right) = \sqrt{\frac{\pi}{2}}\left(-e^{-y^2/2}\right)(y)(y^2-1)$$

Example $f(y) = \dfrac{\Gamma(3+9)}{\Gamma(3)\cdot\Gamma(9)} y^{3-1}(1-y)^{9-1}$ for $y \in (0,1)$

Note: $Y \sim Beta(3,9)$



$A\cdot g(y) \geq f(y)$

$A \geq \dfrac{f(y)}{g(y)}$

$g(y)=1$ so ignored it

Let $g(y)=1 \quad y\in(0,1)$

ignore constant

$\frac{d}{dy} f(y) = -y^2 8(1-y)^7 + 2y(1-y)^8$

$= 0$

Accept $A\cdot g(.6)$ as value from $f(y)$ with prob. equal the height of $\dfrac{f(.6)}{A\cdot g(.6)}$ if value is 0.6

$2y(1-y)^8 = y^2(1-y)^7 8$

$y=0,\ y=1$

$(1-y)^8 = y(1-y)^7 4$

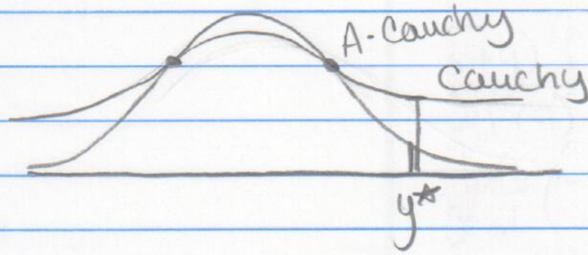
$(1-y) = 4y$

$y = 1/5$

Next simulate a realization from $g(y)$. Call this y^*. Accept y^* as a realization from $f(y)$ with probability $\alpha = \dfrac{f(y^*)}{A\cdot g(y^*)}$ — make A as small as possible to make α as large as possible — avoid too much rejection

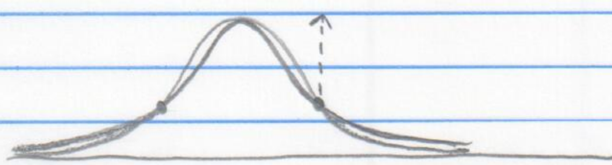
$A = f(1/5)$ plug in above

Pull unif(0,1) and compare to α — reject if unif $> \alpha$
— o/w accept



Find A such that

$$A\cdot\pi^{-1}(1+y^2)^{-1} \ge \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \quad \forall y$$



$$f(1) = \frac{1}{\sqrt{2\pi}} e^{-(1)^2/2}$$

$$\frac{1}{\sqrt{2\pi e}}$$

$$\frac{d}{dy} e^{-y/2} = -y e^{-y^2/2}$$

$$\frac{d^2}{dy^2} =$$

$$\frac{e^{-y^2/2}(y^2-1)}{\sqrt{2\pi}} \qquad y = 1, -1$$

CPMA 573 — Homework #2

Exercise 1: Simulating from a continuous distribution. Write a program in R that uses the **runif** function to complete each of the following:

a. Sample 7500 draws from a gamma density with parameters $\alpha = 10$ and $\beta = 4.0$. Use inverse CDF sampling in conjunction with the fact that the sum of n independent exponential λ random variables is gamma, with $\alpha = n$ and $\beta = \lambda$.

b. Sample 7500 realizations from the standard Cauchy density, using inverse CDF sampling.

c. Sample 7500 realizations from the standard Normal density, using rejection sampling in conjunction with Cauchy reference density $q(x) = \pi^{-1}(1+x^2)^{-1}$ and your program from part **b**. You will need to find the appropriate multiplicative constant A such that $A \cdot q(x)$ is greater than the standard normal density for all x.

Find the mean and variance of each set of 7500 realizations, and compare—except for the Cauchy case—with the corresponding theoretical mean and variance. Finally, plot a histogram of the two sets of realizations—again ignore the Cauchy case—with

```
hist(data,probability=T)
```

where `data` represents the vector in which the 7500 realizations are stored. Then superpose the corresponding theoretical density:

```
lines(seq(0,6,length=250),dgamma(seq(0,6,length=250),10,4))
lines(seq(-4,4,length=250),dnorm(seq(-4,4,length=250)))
```

Exercise 2: Simulating from a discrete distribution. Write a program in R that uses the **runif** function to generate 7500 independent realizations from each of the following distributions:

a. geometric, with parameter $p = 0.83$.

b. binomial, with parameters $n = 200$, $p = 0.11$

c. Poisson, with parameter $\lambda = 2.1$.

Find the mean and variance of each set of 7500 realizations, and compare with the corresponding theoretical mean and variance. Finally, report the theoretical probabilities that the integer k is observed in combination with your simulated probability of observing k, for $k = \{0, 1, 2, 3\}$.

Note that you can save figures generated in R as postscript by *preceding* all plotting commands by

```
postscript(file=''myfile.ps'', horizontal=T)
```

or you can save figures as .pdf files by preceedig all plotting commands by

```
pdf(file=''myfile.pdf'')
```

After all plotting commands have been entered, complete your postscript or .pdf file by entering

```
dev.off()
```