Tutorial – Random variables and sampling —— July 6, 2011

1. How to "view" the random numbers you generate – histograms
   (a) Play with 'binscript.m'
      - modify the number of bins, number of samples
      - change uniform on $(0, 1)$ to uniform on (a, b)

   (b) Work with random variables from a Gaussian distribution, in MatLab 'randn' gives $N(0, 1)$ (N for Normal, mean=0, variance=1).

   (c) Repeat ideas above, but change $N(0, 1)$ to $N(\mu, \sigma^2)$ (you can play with values of μ and σ). Note, an $N(0, 1)$ random variable is called a "standard Normal", and usually referred to as Z. The transformation from Z to $X \sim N(\mu, \sigma^2)$ is given by $X = \mu + \sigma Z$. (Note, you'll like need to change the domain you are binning random variables over.)

2. Random variables from "integrate and invert" method.
   (a) Let's do this for the Exponential distribution. Here we have

$$q(y) = \lambda e^{-\lambda y} H(y) \qquad \text{where} \qquad H(y) = \begin{cases} 1 & \text{for } y > 0 \\ 0 & \text{for } y < 0. \end{cases}$$

   Let's integrate (since we already know the CDF, $F(y)$, we could just write this down)

$$\begin{aligned} \int_{-\infty}^{y} q(t)dt &= \int_{-\infty}^{y} \lambda e^{-\lambda y} H(y) dy \\ &= \int_{0}^{y} \lambda e^{-\lambda y} dy = 1 - e^{-\lambda y} = g^{-1}(y) = F(y) = x \end{aligned}$$

   and now let's invert this to find $g(x)$

$$\begin{aligned} x = g^{-1}(y) &= 1 - e^{-\lambda y} \\ e^{-\lambda y} &= 1 - x \\ -\lambda y &= \ln(1 - x) \\ y &= \frac{-1}{\lambda} \ln(1 - x) = g(x). \end{aligned}$$

   <u>Questions</u>: *There is a negative sign in g, should y be negative? Is it negative?* Note, using the transformation g can be rewritten as $g(x) = -\ln(x)/\lambda$. *Why is this true?*

   (b) Generate Exponential random variables using this transformation, use your script to plot the resulting histogram and approximate pdf. Again, play with different numbers of samples and different numbers of bins.

(c) Go through the same integrate and invert process to find the transformation $g(x)$ for the Cauchy distribution

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

for some constant a of your choosing (note, if you don't remember how to do this integral, you can refer to the CDF from yesterday's lecture.) Use your script to bin your samples. Recall, sampling random variables in the tail is quite likely for this distribution. You'll need to be careful when you're binning in this case. The easiest way to do so is to put all samples that fall beneath your "bottom" bin into the bottom bin and all samples that fall above the top bin into the top bin. Effectively this makes the width of you top and bottom bins infinite.

3. Rejection sampling. Our goal is to generate random variables from Gamma distribution

$$q_n(x) = \frac{x^{n-1} e^{-x}}{\Gamma(n)} \qquad \text{for} \qquad x > 0.$$

Consider the case where $n = 5$. For an $f(x) \geq q_n(x)$ we will use a modification of the Cauchy distribution from 2c. That is, we'll consider

$$f(x) = \frac{c_o}{1 + (x - x_o)^2 / a_o} H(x).$$

The key to rejection sampling is to be able to draw random variables from $f(x)/A$.

(a) Plot $q_n(x)$ and $f(x)$ where you choose the parameters a_o, c_o, and x_o to ensure that $f(x) \geq q_n(x)$ for all $x > 0$. To do this, pick some values and plot $f(x)$ and adjust them visually until the condition $f(x) \geq q_n(x)$.

(b) With the parameter values you found in 3a (actually, better to do this calculation for generic parameters), find the number $A > 1$ where

$$A = \int_0^\infty \frac{c_o}{1 + (x - x_o)^2 / a_o} dx.$$

(c) Find the inverse of the integral of your $f(x)$ and invert it to find the transformation $g(x)$. Note, it will be quite similar to the one you found in 3c (note, the big difference is the bottom limit of integration changed from $-\infty$ to 0, *why was that again?*).

(d) Use rejection sampling to generate N random variables from $q_n(x)$ using these steps:

i. Draw a Uniform random variable from $U(0, A)$, call it Z.
ii. Calculate $X = g(Z)$ with the $g(x)$ you found in part 3c.
iii. Calculate $B = f(X)$.
iv. Draw a Uniform random variable from $U(0, B)$, call it Y
v. If $Y > q_n(X)$ throw away X, if $Y \leq q_n(X)$ keep X
vi. Repeat this process until you've kept N $X's$ while keeping track of how many you keep and how many you throw away.

Why do we want to know the ratio of the number we're keeping?