Exam 1: Solution
Stat 428
March 3, 2015

1. Consider the probability density function $f(x) = \lambda e^{-\lambda(x-\eta)}$, for $x \geq \eta$, where λ and η are positive constants.
a. Find the cumulative distribution function $F(x)$.
b. Given two draws from a uniform distribution on (0,1), $u_1 = .343$ and $u_2 = .879$, use the inverse cdf method to obtain two variates x_1 and x_2 from $f(x)$.
c. Let $\lambda = 2$ and $\eta = 1$, and the probability density function $g(x) = e^{-x}$ for $x > 0$. Find a constant c that satisfies $f(x)/g(x) \leq c$ for $x > 1$, and write R code for generating a sample of size $m = 10000$ from $f(x)$ using acceptance rejection sampling with proposals drawn from $g(x)$ (you may use `rexp()`).
d. Using `rexp()`, discuss how you could use a simple transformation method by drawing observations from a well chosen exponential distribution and transforming them into a sample from $f(x) = 2e^{-2(x-1)}$ for $x > 1$.

Solution:
a.Find out cdf. For $x \geq \eta$,

$$F(x) = \int_{\eta}^{x} f(t)dt = \int_{\eta}^{x} \lambda e^{-\lambda(t-\eta)}dt = \lambda \times \frac{1}{(-\lambda)} e^{-\lambda(t-\eta)}|_{\eta}^{x} = -e^{-\lambda(t-\eta)}|_{\eta}^{x} = 1 - e^{-\lambda(x-\eta)} \tag{1}$$

b. Find out inverse cdf.

$$\begin{aligned} u = F(x) = 1 - e^{-\lambda(x-\eta)} &\Rightarrow e^{-\lambda(x-\eta)} = 1 - u \\ &\Rightarrow -\lambda(x-\eta) = ln(1-u) \\ &\Rightarrow x = -\frac{1}{\lambda} ln(1-u) + \eta \end{aligned} \tag{2}$$

Plug in u_1 and u_2, then we get:
$x_1 = -\frac{1}{\lambda} ln(1 - u_1) + \eta = -\frac{1}{\lambda} ln(0.657) + \eta$
$x_2 = -\frac{1}{\lambda} ln(1 - u_2) = -\frac{1}{\lambda} ln(0.121) + \eta$
It is ok if you plug in λ, η given in part c. Then the answer would be $x_1 = 1.210$ and $x_2 = 2.056$.

c. Since $f(x) = 2e^{-2(x-1)}$ and $g(x) = e^{-x}$, for $x \geq 1$,

$$\frac{f(x)}{g(x)} = \frac{2e^{-2(x-1)}}{e^{-x}} = 2e^{-x+2} \leq 2e \tag{3}$$

Therefore, we can set $c = 2e \approx 5.437$. Then, $\frac{f(x)}{cg(x)} = e^{-x+1}$

```
m=0 #our sample size
n=0 #how many times we draw from g(x)
x=NULL
while (m<=10000)
{
  u=runif(1)
  xx=rexp(1)
  if (xx>=1) #notice that the supports of g(x) and f(x) are not the same
  {
    if (u<exp(1-xx)) #simplify f(x)/cg(x)
    {
     x=c(x,xx)
     m=m+1
   }
  }
  n=n+1
}
```

`Note: You can also use the method on classnotes. Noticing that` $c = 2e \approx 5.437$`, you may want to draw roughly 54370` samples.

d.$f(x) = 2e^{-2(x-1)}$ for $x > 1$ is a shifted exponential distribution for exp(2). That is, it just shifts the density curve 1 unit to the left. Therefore, what we can do is, simply draw samples X_i from exp(2), or $rexp(2)$. Transform samples to be $Y_i = X_i + 1$. Then Y_i is what we want from target distribution.

2. Recall that if X has a chi-squared distribution with w degrees of freedom and Y is independent of X and has a chi-squared distribution with v degrees of freedom, then $F = \frac{X/w}{Y/v}$ has an F-distribution with w and v degrees of freedom. Write an R function for drawing from an F distribution with user supplied w and v that makes use of the `rnorm()` function for drawing from a normal distribution.

Solution:
Notice that the chi-squared distribution with freedom w is the sum of w's square of indepentdent standard normal variables. That is, if $X \sim \chi^2(w)$, then $X = S_1^2 + S_2^2 + \cdots + S_w^2$, and $S_i \sim N(0, 1), i = 1, 2, \ldots, w$. It's the same thing for $Y \sim \chi^2(v)$.
Then transform X,Y, two independent Chi-square samples, to get a sample from F distribution, using $F = \frac{X/w}{Y/v}$.

```
rfdist=function(w,v){
  xchisq=0
  ychisq=0
  for (i in 1:w)
  xchisq=xchisq+(rnorm(1))^2
  for (i in 1:v)
  ychisq=ychisq+(rnorm(1))^2
  sample=xchisq/w/(ychisq/v)
  return(sample)
}
```

3. Let $\theta = \int_0^{\pi/2} e^{sinx} dx$.
a. Write R code to construct an estimate $\hat{\theta}$ of θ using Monte Carlo integration with respect to a well chosen uniform distribution.
b. Derive an upper bound for the variance $\hat{\theta}$. It doesn't have to be the precise variance, but a reasonably close upper bound.

Solution:
a.
1. Draw m (say, m=10000) samples $X_1, X_2 \ldots, X_m$ from $f(x) \sim Unif[0, \frac{\pi}{2}]$
2. Transform them into: $g(X_i) = e^{\sin X_i}, i = 1, 2, \ldots, m.$
3. Estimate the expectation value of $g(X_i)$ using:

$$\hat{\theta} = \int_0^{\frac{\pi}{2}} e^{\sin t} dt = E_f(g(X)) \approx \frac{1}{m} \sum_{i=1}^{m} \frac{g(X_i)}{f(X_i)} = \frac{\frac{\pi}{2} - 0}{m} \sum_{i=1}^{m} g(X_i) \tag{4}$$

R code:

```
xx=runif(10000,0,pi/2)
theta_est=(pi/2)*mean(exp(sin(xx)))
theta_est
```

b. Since X_i's are i.i.d from $Unif[0, \frac{\pi}{2}]$. After some derivation, we try to bound $Var(e^{\sin X})$. Notice that since $X \in [0, \frac{\pi}{2}]$, $e^{sinX} \in [0, 1]$.

$$Var(\hat{\theta}) = Var(\frac{\frac{\pi}{2} - 0}{m} \sum_{i=1}^{m} e^{\sin X_i}) = \frac{\pi^2}{4m^2} Var(\sum_{i=1}^{m} e^{\sin X_i})) = \frac{\pi^2}{4m} Var(e^{\sin X}) \tag{5}$$

$$\begin{aligned} Var(e^{\sin X}) &= E(e^{2sinX}) - E^2(e^{sinX}) \\ &\leq E(e^2) - E^2(1) = e^2 - 1 \end{aligned} \tag{6}$$

Note: Other reasonable answers are $e^2, (e-1)^2$ etc.

4. Let $\phi(x) = ce^x$ for $0 < x < \pi/2$, and $\theta = \int_0^{\pi/2} e^{sinx}dx$.
a. Find c so that $\phi(x)$ is a probability density function.
b. Using the inverse cdf sampling technique to draw observations from $\phi(x)$, write R code to construct an estimator $\hat{\theta}$ of θ using importance sampling with importance function $\phi(x)$.
c. Assuming the estimator of (4b) uses the same number of draws m as the estimator of (3a), discuss which you would expect to be more efficient, and why.

Solution:
a.

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \phi(t)dt &= \int_0^{\frac{\pi}{2}} ce^t dt = ce^t|_0^{\frac{\pi}{2}} = c(e^{\frac{\pi}{2}} - 1) = 1 \\ \Rightarrow c &= \frac{1}{e^{\frac{\pi}{2}} - 1} \approx 0.2624\end{aligned} \tag{7}$$

b. First of all, find out cdf of $\phi(x)$. For $x \in (0, \frac{\pi}{2})$,

$$\begin{aligned}F(x) &= \int_0^x \phi(t)dt = \int_0^x ce^t dt = ce^t|_0^x = c(e^x - 1) \\ &= \frac{e^x - 1}{e^{\frac{\pi}{2}} - 1}\end{aligned} \tag{8}$$

Then, find out inverse cdf,

$$\begin{aligned}u = F(X) = \frac{e^X - 1}{e^{\frac{\pi}{2}} - 1} &\Rightarrow e^X - 1 = u(e^{\frac{\pi}{2}} - 1) \\ &\Rightarrow e^X = u(e^{\frac{\pi}{2}} - 1) + 1 \\ &\Rightarrow X = ln(u(e^{\frac{\pi}{2}} - 1) + 1)\end{aligned} \tag{9}$$

Let $g(x) = e^{sinx}$. We first got samples from $\phi(x)$ using inverse cdf sampling technique. Then, we can estimate θ using importance sampling with importance function $\phi(x)$.

$$\theta = \int_0^{\pi/2} e^{sinx}dx = \int_0^{\pi/2} \frac{g(x)}{\phi(x)} \times \phi(x)dx \tag{10}$$

R code:

```
#sample size=10000
#first get 10000 samples from phi(x)
c=1/(exp(pi/2)-1)
```

```
u=runif(10000)
x=log(u/c+1)
xx=exp(sin(x))/(c*exp(x))    #importance sampling samples
theta_hat=mean(xx)           #estimate
```

c. The method in (4b) is more efficient. Compared with uniform distribution in (3a), $\phi(x)$ and $g(x)$ are more similar in shape (they are both exponential-like density funcitions). Therefore, we can expect that the importance sampling in (4b) is more efficient.