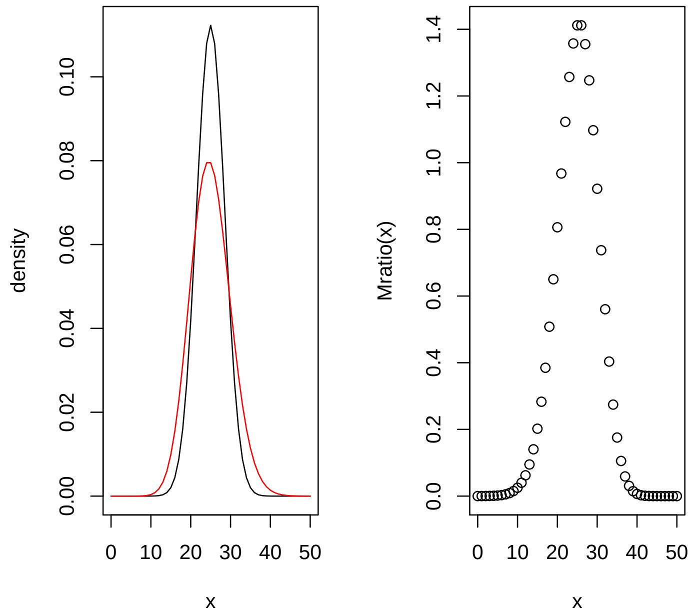
STAT 701 Midterm Practice Problems

1. Assume f is a binomial distribution with parameters n and p and g is the poisson distribution with parameter $\lambda = np$.

   - Write the pseudocode for the accept/reject routine to sample from the distribution f using the density g as a candidate. Be sure to discuss in detail how you would come up with the mulitplier M. You do not need to write out the binomial and poisson densities. You can refer to them in terms of f and g.
     Solution
     - Determine the multiplier M such that $Mg(x) > f(x)$ for $x = 0, \ldots, n$. Do this by finding the maximum of $f(x)/g(x)$. The ratio does not simplify nicely into a function that is easy to find the max. However, one could evaluate the ratio at all values $x = 0, \ldots, n$, or note that since the two distributions have the same mean (np), and the poisson has larger variance (np vs. $np(1-p)$), the highest ratio will happen near the mean. The following code illustrates this.

```
> n=50
> p=.5
> x=c(0:n)
> Mratio=function(x){dbinom(x,n,p)/dpois(x,n*p)}
> c(x[which.max(Mratio(x))], max(Mratio(x)))
[1] 25.000000  1.411859
> par(mfrow=c(1,2))
> plot(c(0:n), dbinom(c(0:n),n,p),type="l", xlab="x", ylab="density")
> lines(c(0:n), dpois(c(0:n), n*p),col=2)
> plot(x, Mratio(x))
```

       Algorithm:
     - Generate y values from the Poisson distribution.
     - Generate u values from the uniform distribution.
     - Accept the $x = y$ values for which $u \leq f(y)/Mg(y)$



- Show that the probability of acceptance of your algorithm is $1/M$.
  Solution
  For $U \sim \mathcal{U}_{[0,1]}$, $Y \sim g(y)$, and $X \sim f(x)$, such that $f/g \leq M$, the acceptance condition in the Accept–Reject algorithm is that $U \leq f(Y)/(Mg(Y))$. The probability of acceptance is thus:

$$\begin{aligned} P(U \leq f(Y)/Mg(Y)) &= \sum_{y=0}^{n} \int_0^{\frac{f(y}{Mg(y)}} \, \mathrm{d}u g(y) \\ &= \sum_{y=0}^{n} \frac{f(y)}{Mg(y)} g(y) \\ &= \frac{1}{M} \sum_{y=0}^{n} f(y) \\ &= \frac{1}{M} \, . \end{aligned}$$

2. Write the pseudocode for an inverse transform routine for sampling from an exponential distribution with pdf $g(x) = \beta e^{-\beta x}, x > 0, \beta > 0$.
Solution Begin by computing the inverse cdf function:

$$\begin{aligned}
f(x) &= \beta e^{-\beta x} \\
F(x) &= \int_0^\infty \beta e^{-\beta x} \\
&= -e^{-\beta x} + c \\
&= 1 - e^{-\beta x} \\
F^-(u) &= \frac{-1}{\beta} ln(1-u)
\end{aligned}$$

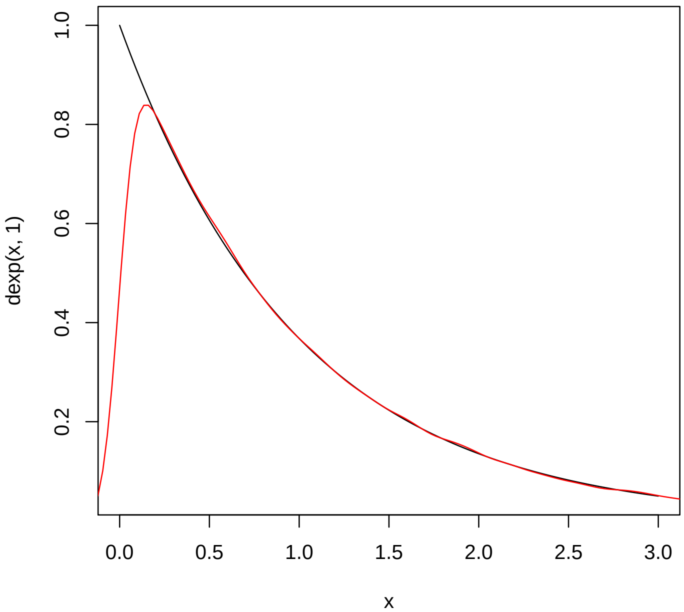
To sample via the inverse transform method:

(a) Generate u values from the uniform(0,1) distribution

(b) Plug these u values into $F^-(u)$ to get draws from the exponential distribution.

```
> Nsim=10^5
> beta=1
> Finv=function(x){-log(1-x)/beta}
> u=runif(Nsim)
> x=Finv(u)
>

> par(mfrow=c(1,1))
> curve(dexp(x,1),from=0, to=3)
> lines(density(x), col=2)
```



3. If you have two densities g_1 and g_2 that you are considering using as importance function to use importance sampling to find the mean of the target density f, discuss at least one means by which you could compare g_1 and g_2 to determine which will lead to more accurate estimation for a fixed sample size.

   Solution
   The better importance function is the one that matches the shape of the target density the best. It gives the most uniform weights. To compare g_1 and $g2$, some samples should be drawn from them and weights $w = f(x)/g(x)$ computed for each. Three measures that could be used to compare are

   - The variance of the weights. The importance function with the lower variance is better since the variance of the weights factors into the multiplier in the variance of the estimate.
   - The effective sample size. The importance function with the ESS closer to n is better. An estimate from an importance sample with effective sample size $n*$ has the same variance as an estimate based on a sample from f of size n.
   - The perplexity. The importance function with perplexity closer to

1 is better. When the perplexity is 1, the shapes of the target and importance functions are a perfect match.

4. You plan to use importance sampling to find the tail probability of a normal distribution $P(Z > 6)$ using importance sampling via a truncated exponential. Identify the following

   - The function $h(x)$ whose expectation is the goal of the analysis.
     h is an indicator function. $h(x) = I[x > 6]$.
   - The target density, f(x).
     The target is the standard normal distribution. $f(x) = \frac{1}{\sqrt{2\pi} exp(x^2/2)}$
   - The importance function, g(x)
     The importance function is the truncated exponential. $g(x) = e^{-(x-6)}, \ \ x > 6$.
   - The weights, w.
     The weights are the ratio of the target to the importance function. $w = f(x)/g(x)$

5. The code below gives a Monte Carlo estimate of the mean of a beta distribution. Two options are provided for estimating the error. Which do you prefer and why?

```
> Nsim=10^3
> x=rbeta(Nsim, 4, 6)
> estint=mean(x)
> ### Variance estimate A
> estvar1=var(x)/Nsim
> ### Variance estimate B
> xm=matrix(nrow=Nsim, ncol=100, data=sample(x, Nsim*100, replace=T))
> estvar2=mean(apply(xm,2,var)/Nsim)
```

Solution
Variance estimate B is preferred. While variance estimate A is asymptotically correct, it is based on only a single sequence of estimators. Variance estimate B takes a bootstrapping approach and will lead to a smoother confidence band