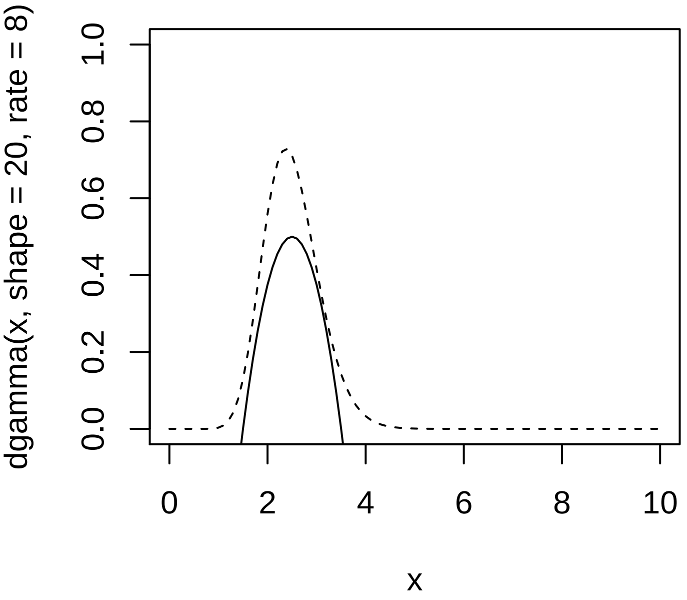
1. *(5 points) Describe the dfferences between accept-reject sampling and importance sampling.*

   Accept-reject and importance sampling are both used to estimate probability distributions and their characteristics and are especially useful when closed form solutions are not apparent. The problem arises from having a space we wish to sample from or compute characteristics of, but the space, say g(x),is not a valid pdf and so sampling from it becomes problematic. Both methods begin with sampling from a named pdf similar to g(x), say k(x).

   Accept-reject sampling only uses the subset of randomly chosen sample points under k(x) which are under g(x) and k(x). The proportion of points "accepted" are an estimate of the proportion of the area under g(x) that is also under k(x), which can be used to estimate a sampling distribution for g(x) along with other quantities such as the area below g(x). k(x) must be greater than g(x) at all sampling points for accept-reject sampling, but not for imporatnce sampling. Importance sampling uses *all* the sampled observations from the chosen randomly from k(x), but weights each based on the ratio g(x)/k(x) so that the closer g(x) is to k(x), the more weight at that point. The notes did not use importance sampling as a method to estimate pdfs, but it seems taking w(x_i) = $\frac{g(x_i)}{k(x_i)}$ a pmf can be found (or possibly estimated with a pdf with smoothing) by taking the density at each x_i to be $\frac{w(x_i)}{\Sigma_i w(x_i)}$.

   While it can be hard to estimate M, which scales a pdf to make it entirely above g(x) in accept-reject sampling, it seems that importance sampling rewards with mass similar g(x) and k(x) rather than putting more mass at peaks of g(x), but I could just not be understanding importance sampling well enough.



Differences:

- Algorithm
- Using all sample points vs. a subset
- Whether or not k(x) can be g(x) at some sample points, which includes finding a constant to make this hold is necessary

2. *(5 points) How are samples from the posterior distribution useful for inference in Bayesian statistics?*

   Samples from a posterior can be used to estimate statistics such as the maximum, median, etc. of a postieror distribution and quantify the uncertainty. Sampling from the posterior using Monte Carlo methods can give estimates of the quantities with estimates of uncertainty/precision.

3. *(5 points) Assume you are interested in obtaining a posterior predictive distribution, complete the following equation.*

   $p(y^*|y_1...y_n) = \int p(\theta|y_1...y_n)p(y^*|\theta)d\theta$

   since

   $p(y^*|y_1...y_n) = \int p(y^*, \theta|y_1...y_n)d\theta$

   $= p(y^*|y_1...y_n) = \int p(y^*|\theta, y_1...y_n)p(\theta|y_1...y_n)d\theta$

   and since all the information contained in $y_1...y_n$ is contained in θ

   $= p(y^*|y_1...y_n) = \int p(y^*|\theta)p(\theta|y_1...y_n)d\theta$

   $= p(y^*|y_1...y_n) = \int p(\theta|y_1...y_n)p(y^*|\theta)d\theta$

4. *(5 points) Define an improper prior and give an example of model where one might be used.*

   Prior distributions do not have to be valid pdfs, as long as the posterior is valid. SAS defines an improper prior to be one where the integral over the parameter space is not finite. `https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect004.htm` Before seeing this definition, I thought an improper prior was simply when the prior did not integrate to 1 over the domain, but I guess that that integral must be for the prior to be improper.

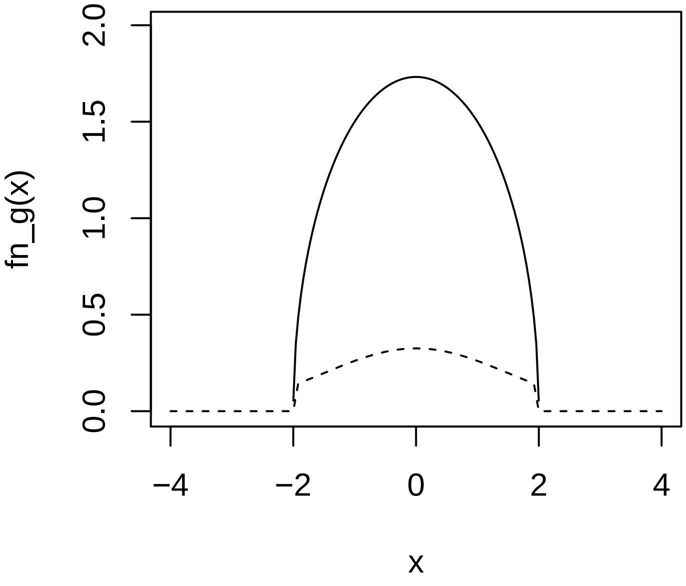
   One way an improper prior can come about is if the input that does not follow exactly the domain of the named pdf at hand.

   Making a prior improper can be used to make it less informative. We know the prior on a Poisson mean (which is also the variance) is strictly positive and that the Gamma distribution is often used as a prior on a Poisson mean, which has a domain of (0,). If we rather used a Normal prior, which has a domain of (-,), we would be using an improper prior because the Poisson mean (variance) is, in reality only positive. Using the Normal may be one way of putting a less informative prior on a

Poisson mean if we do not have many prior observations. An example of a Poisson random variable may be the number of bird species in a confined region, say Benton Lake area.

5. *(20 points) Implement an accept-reject sampler to compute the area under the half-ellipse g(x) =* $\sqrt{}$ *[3(1-$\frac{x^2}{4}$)] Include your R code for full credit.*

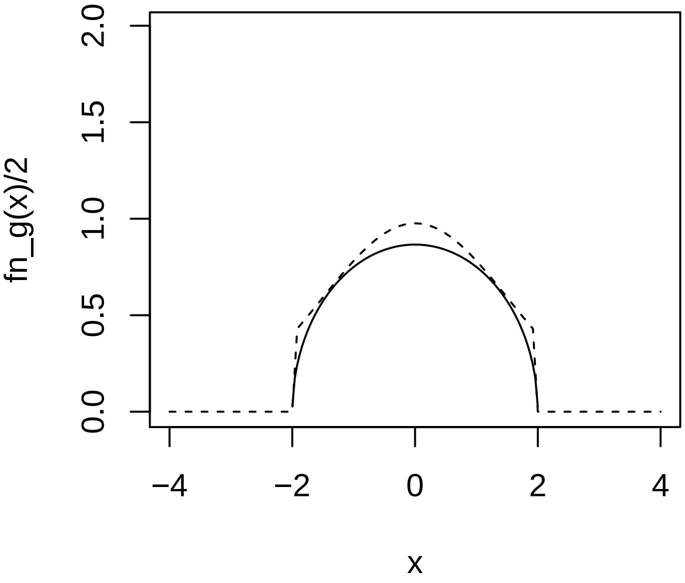
```r
require(truncnorm)

# What is reasonable for fn_f and M
fn_g <- function(x){
  out <- sqrt(3*(1-((x^2/4))))
  return(out)
}

fn_f <- function(x){
  dtruncnorm(x,a=-1.999,b=1.999,0,1.5)
}
curve(fn_g, from = -1.999, to = 1.999,
      xlim = c(-4,4), ylim = c(0,1.99))
curve(fn_f, add=TRUE, lty=2)
```



```r
cat("The normal(0,sd=1.5) trucated at -2 and 2 seems reasonable to
    estimate the area of g. The next step is to find M that normalizes
    the ratio of f and g.")
```

The normal(0,sd=1.5) trucated at -2 and 2 seems reasonable to estimate the area of g. The next step is to find M that normalizes the ratio of f and g.

```
curve(fn_g(x)/2, from = -1.999, to = 1.999,
      xlim = c(-4,4), ylim = c(0,1.99))
curve(fn_f(x)*3, add=TRUE, lty=2)
```



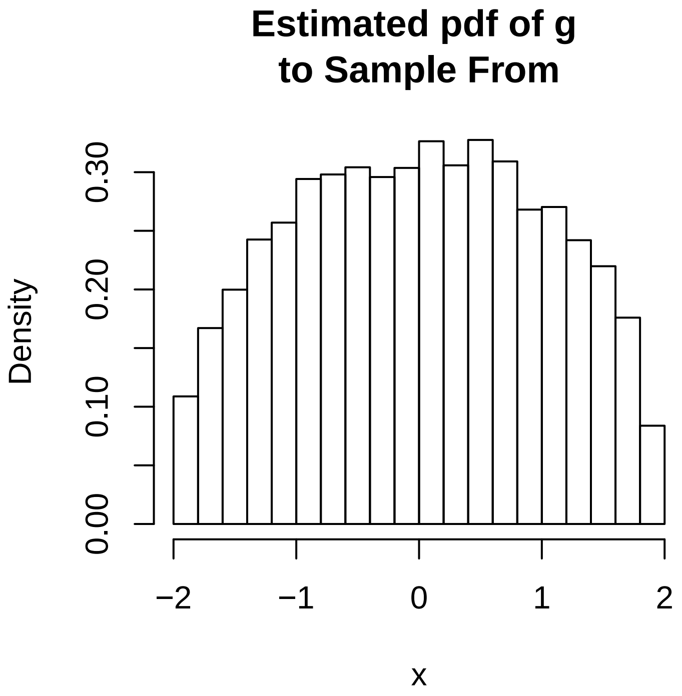
```
cat("Six seems to be a reasonable constant to make f larger than g.
    Below is another method for finding the area.")
```

Six seems to be a reasonable constant to make f larger than g. Below is another method for finding the area.

```
#What is reasonable for M?

# Estimate the area by summing the area of squares (reimman sums), I will use the
# mid point of x to estimate the height, and each bin width = 0.2
s <- c(seq(-2,2, by = 0.2))
mid <- c(s+0.1)[1:length(s)-1]

height <- fn_g(mid)
base <- 0.2

area <- height*base
a <- sum(area)

cat("The estimated area using Reimman Sums is", a, "squared units, so let M=6.")
```

The estimated area using Reimman Sums is 5.460137 squared units, so let M=6.

```
n <- 10000

# I am using the truncated normal distribution
# with mean 0 and sd 1.5, truncated at -2 and 2
# Sample randomly from the trucated distribution
y <- rtruncnorm(n,a=-1.999,b=1.999,0,1.5)
# summary(y)
# Check, all values within -2,2

# Choose M to make the trucated normal distribution
# always above fn_g
# requirement of accept-reject sampling
M <- 6

set.seed(5323)
u.vals <- runif(n,0,1)
# Subset y to only include y such that M*fn_f(y) < fn_g(y)
y.accept <- y[which(u.vals < fn_g(y)/(M*fn_f(y)))]

hist_y.accept <- hist(y.accept, freq = FALSE, breaks = 'FD',
      main = "Estimated pdf of g \n to Sample From ", xlab = "x")
```



```
cat("The area under the curve,", "$M*f$", "is 6 squared units.",
    "The estimated area under g using accept-reject sampling is",
    6*length(y.accept)/length(y), "squared units.")
```

The area under the curve, $M * f$ is 6 squared units. The estimated area under g using accept-reject sampling is 5.4048 squared units.

6. Problem 6 is attached.

7. (a) *(5 points) How would you go about addressing the researcher's questions in a classical framework? Would you be able to compute these probabilities?*

   Under a classical framework, I would have to know the true pdf and its associated parameters to compute these probabilities. The normal pdf may be reasonable, but I would have to know or find a good resource for the parameters associated with the normal distribution.

   If it was reasonable that the mean snowfall at Graf Park did not depend on the mean snow fall at Cherry River, then I would use the bivariate normal distribution for the joint pdf.

   I would compute these probabilities using the probability functions in R.

   (b) *(5 points) How would you address the researcher's questions in a Bayesian framework? Can you compute these probabilities?*

   In a Bayesian framework I would compute each of the three posterior pdfs, and ideally each would be a conjugate pair of a named distribution, so we could again use the probability functions in R.

   Realistically, for example, with $p(\theta|y_1, ..., y_n) = \int_{\sigma^2} p(y_1, ..., y_n|\theta, \sigma^2) * p(\theta|\sigma^2) * p(\sigma^2) d\sigma^2$ with

   $y_1, ..., y_n|\theta, \sigma^2 \sim \text{N}(\theta, \sigma^2)$

   $\theta|\sigma^2 \sim \text{N}(\mu_o, \sigma^2/k_o)$

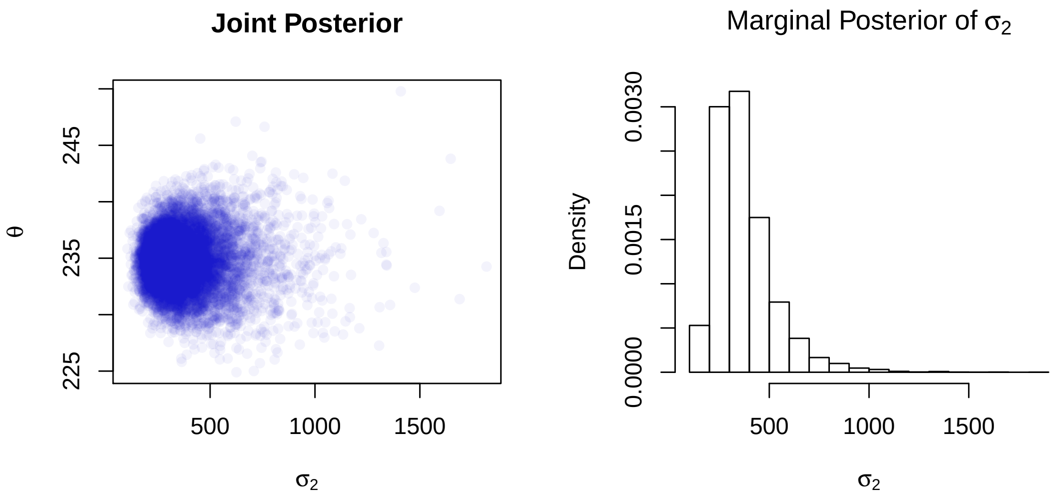
   $\sigma^2 \sim \text{INVGAM}(\frac{\nu_o}{2}, \frac{\nu_o * \sigma_o^2}{2})$

   finding a closed form solution is not feasible and the posteriors have to be estimated using Monte Carlo procedures.

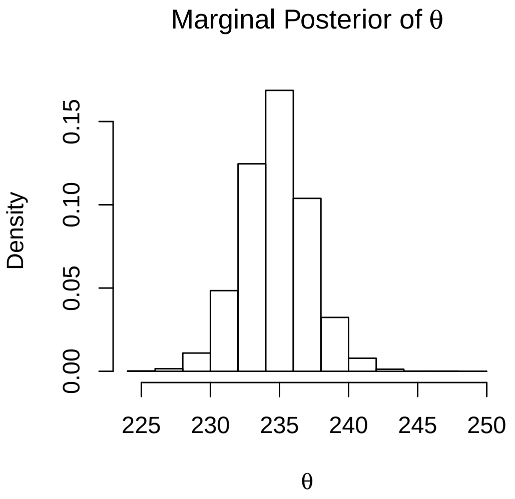
   (c) Kenny says that choosing κ_o larger means we have more information about θ, or a stronger prior distribution on θ whereas choosing a larger ν_o means we have more information about σ^2, meaning a stronger prior on σ^2. In practice, I would plot several priors with different κ_o and ν_o values and ask the researcher which he thought to most resemble the true prior distributions to find estimates of these parameters.

   I estimated θ or the mean, using a website, so I will choose κ_o to be larger, say 50. I have no idea about what σ^2 should be, so I will choose $\nu_o = 1$. I choose $\sigma^2 = 10^2 = 100$ because we expect 95% of data to lie within about two standard deviations of the mean and if I am using about 231.1 cm as the true mean snowfall, I would expect most years the mean snowfall be be between 210 and 250.

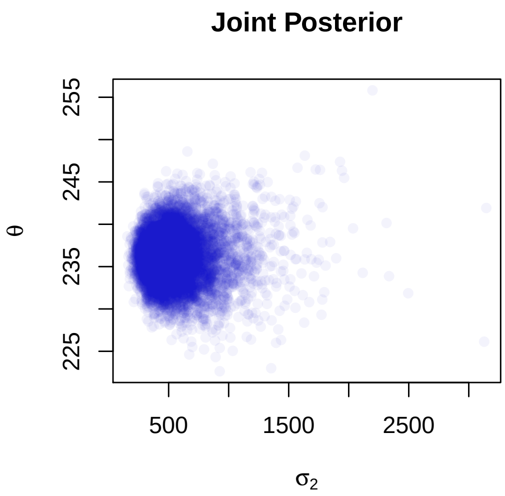
P(θ_{gp})

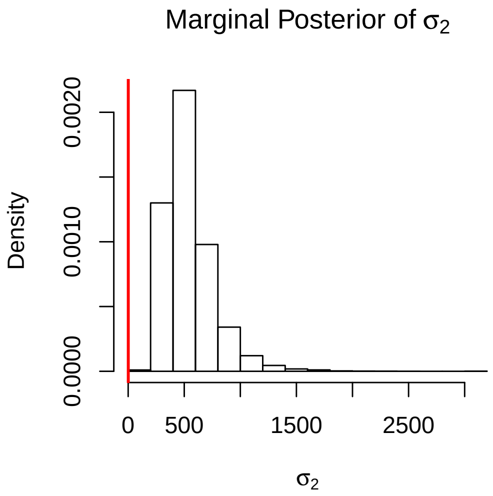


P($\theta_{gp} > 250$) $= 0$

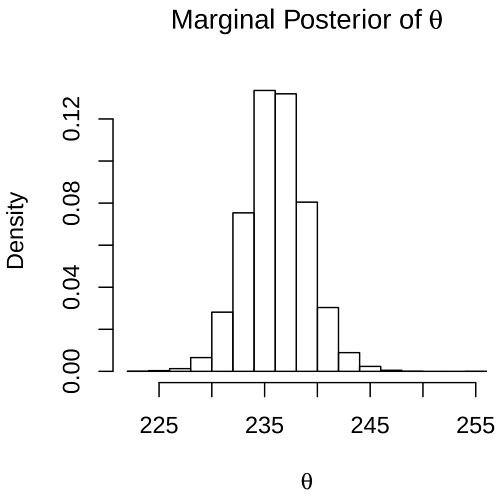


P(θ_{cr})

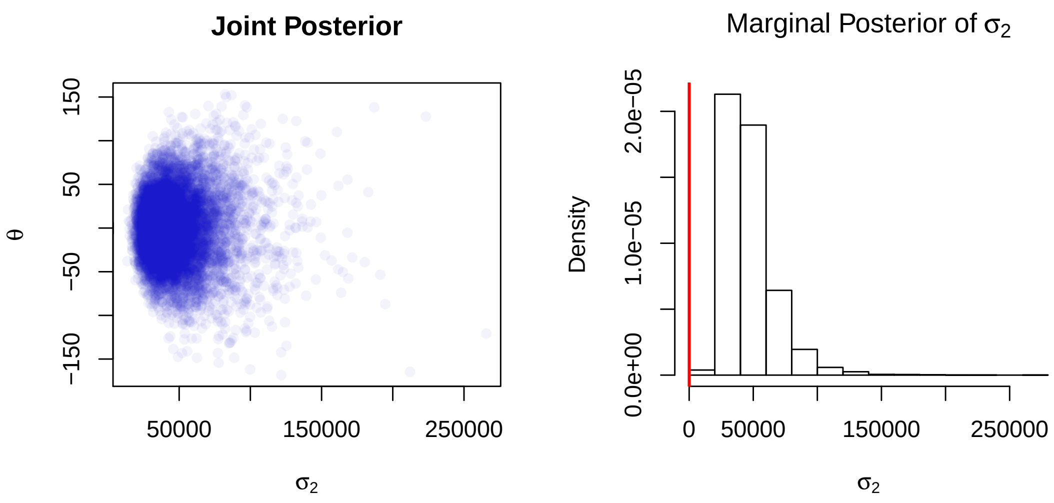




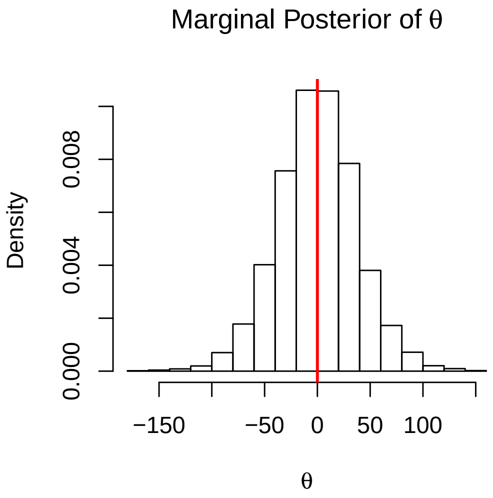
P($\theta_{cr} > 250$) = 1e-04



P(θ_{cr} , θ_{gp})If we can assume θ_{cr} and θ_{gp} are indepdendent, which would mean that knowing the mean snowfall at Cherry River does not tell us anything about the snowfall at Graf Park, then P(θ_{cr} , θ_{gp}) = P(θ_{cr}) × P(θ_{cr}). This does not seem reasonable, but is what I will assume for this problem. Then $\theta_{cr} - \theta_{gp} \sim$ BIVNORM(0 $\frac{2*\sigma_i^2}{k_n}$) where σ_i^2 is generated randomly from an inverse gamma distribution. Note that in addition to assuming independence, I also assumed the two parameters have a common mean and common variance.



P(θ_{cr} - $\theta_{gp} > 0$) = 0.4998



Thoughts: I probably should not have set the prior to having a low chance of being above 250 in the prior, but seeing that I put less strength on the prior using $\nu_o = 1$, I did not think it would have influenced the posterior that much.