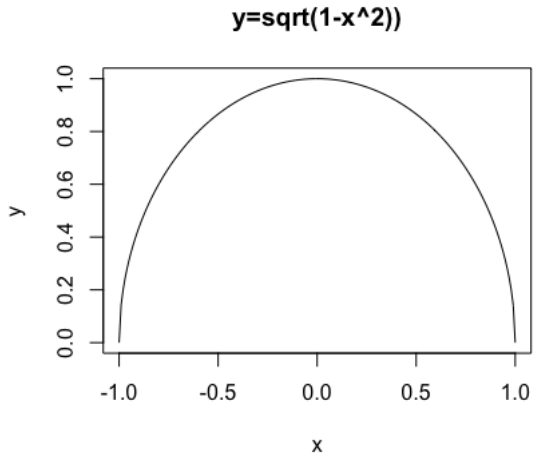
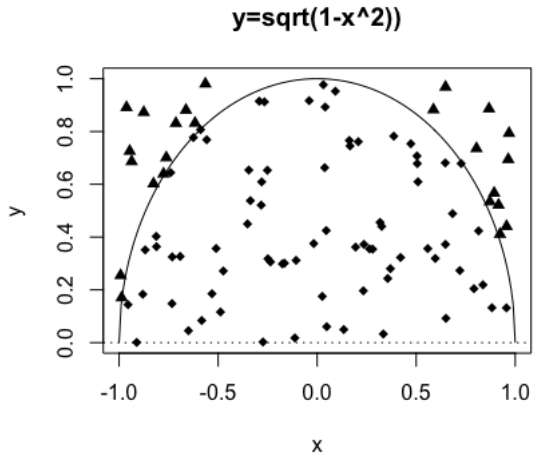
Week 4: Sept 19 - Sept 23

Exercise 1. Consider the function $g(x) = \sqrt{1-x^2}$, where $x \in [0,1]$. The goal is to estimate $I = \int g(x)dx$.



One way to do this is to simulate points from a uniform distribution with a known area. Then we compute the proportion of points that fall under the curve for $g(x)$. Specifically, we create samples from a uniform function on $[-1,1]$. The area under this function is 2. To compute I we estimated the proportion of responses in I that are also under $g(x)$.




```
> x <- seq(-1,1, by=.01)
> y <- sqrt(1-x^2)
> plot(x,y,type='l', main= "y=sqrt(1-x^2))")
> abline(h=0,lty=3)
```

```
> num.sims <- 100000
> f.x <- runif(num.sims,-1,1)
> f.y <- runif(num.sims)
> accept.samples <-
> points(f.x[accept.samples],f.y[accept.samples],pch=18)
> reject.samples <-
> points(f.x[reject.samples],f.y[reject.samples],pch=17)
> accept.proportion <-
> area <-
```

This matches with the analytical solution of

Monte Carlo Procedures

Monte Carlo procedures use random sampling to estimate mathematical or statistical quantities. These computational algorithms are defined by running for a fixed time (number of samples/iterations) and result in a random estimate. There are three main uses for Monte Carlo procedures: 1.)
, 2.) and 3.)

Monte Carlo methods were introduced by John von Neumann and Stanislaw Ulam at Los Alamos. The name Monte Carlo was a code name referring the Monte Carlo casino in Monaco. Monte Carlo methods were central to the Manhattan project and continued development of physics research related to the hydrogen bomb.

An essential part of many scientific problems is the computation of the integral, $I = \int_{\mathcal{D}} g(x)dx$ where $\mathcal{D}$ is often a region in a high-dimensional space and $g(x)$ is the target function of interest. If we can draw independent and identically distributed random samples $x_1, x_2, \ldots, x_n$ uniformly from $\mathcal{D}$ (by a computer) an approximation to I can be obtained as

The law of large numbers states that the average of many independent random variables with common mean and finite variances tends to stabilize at their common mean; that is

A related procedure that you may be familiar with is the *Riemann approximation*. Consider a case where

$\mathcal{D} = [0, 1]$ and $I = \int_0^1 g(x)dx$ then

where $b_i = i/n$. Essentially this method is a grid based evaluation. This works will for a smooth function in low dimension, but quickly runs into the "curse of dimensionality".

Accept - Reject Sampling

In many scenarios, sampling from a function $g(x)$ can be challenging (in that we cannot use `rg.function()` to sample from it. The general idea of accept reject sampling is to simulation observations from another distribution $f(x)$ and accept the response if it falls under the distribution $g(x)$.

Formally the algorithm for the Accept-Reject Method follows as:

1.

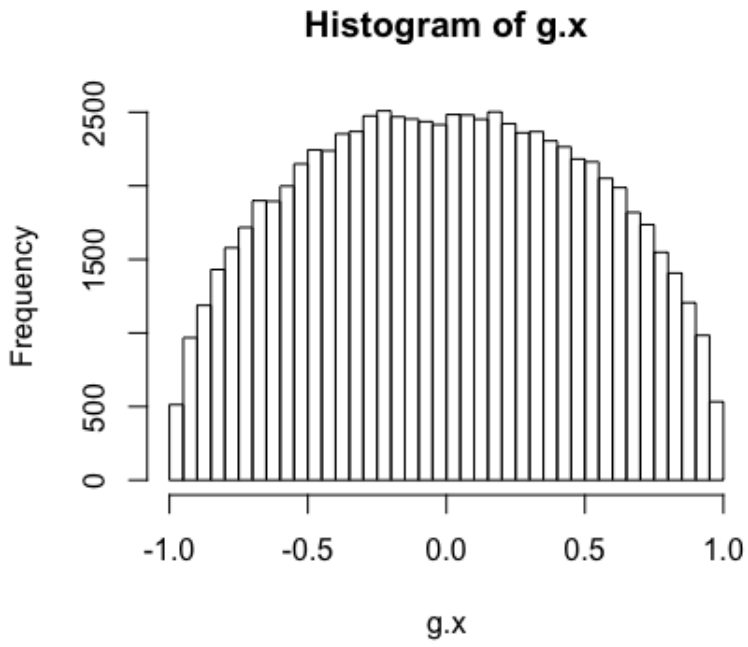
2.

where $f(x)$ and $g(x)$ are normalized probability distributions and M is a constant ≥ 1.

Similar to the context of the problem above, suppose we want to draw samples from a normalized form of $g(x) = \frac{\sqrt{1-x^2}}{\pi/2}$. Then $f(x) = \frac{1}{2}$ for $x \in [-1, 1]$ and $M =$

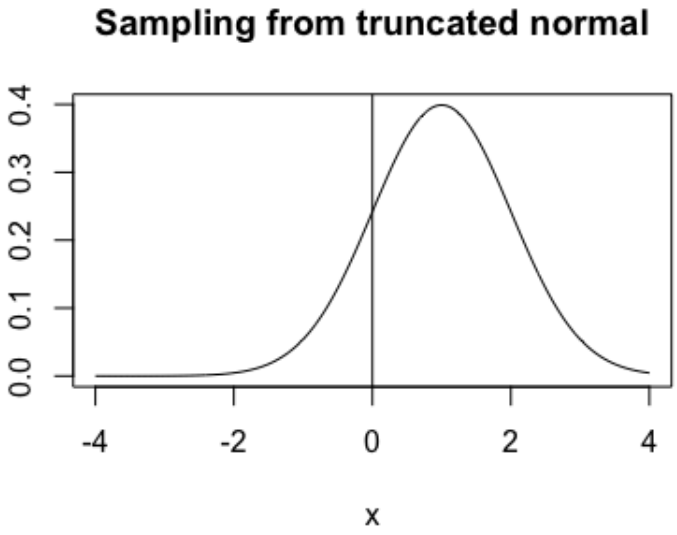
```
x <- runif(num.sims,-1,1) # simulate samples
u.vals <- runif(num.sims)
M <- 4/pi
f.x <- 1/2

accept.ratio <-                                                    # g/fM
accept.index <- (1:num.sims)[u.vals < accept.ratio]
g.x = x[accept.index]
hist(g.x,breaks='FD')
```



Now we have samples from $g(x)$ and could, for example, compute $I = \int xg(x)dx$

Exercise 2. Without the use of packages such as `rtnorm()` how would you draw a sample from a truncated normal distribution?



Again we will use accept-reject sampling. Simulate points from a normal distribution with the same mean and variance. Let M = $\frac{1}{\Phi(c;\mu,\sigma^2)}$, where $\Phi(.;\mu,\sigma^2)$ is the cdf function a normal random variable and c is the truncation point. Then all values of x greater than c are accepted and all values less than c are rejected. We will use this formulation for binary Bayesian regression models using a probit link.

Importance Sampling

Importance sampling is related to the idea of accept-reject sampling, but emphasizes focusing on the "important" parts of the distribution and uses all of the simulations. In accept-reject sampling, computing the value M can be challenging in its own right. We can alleviate that problem with importance sampling.

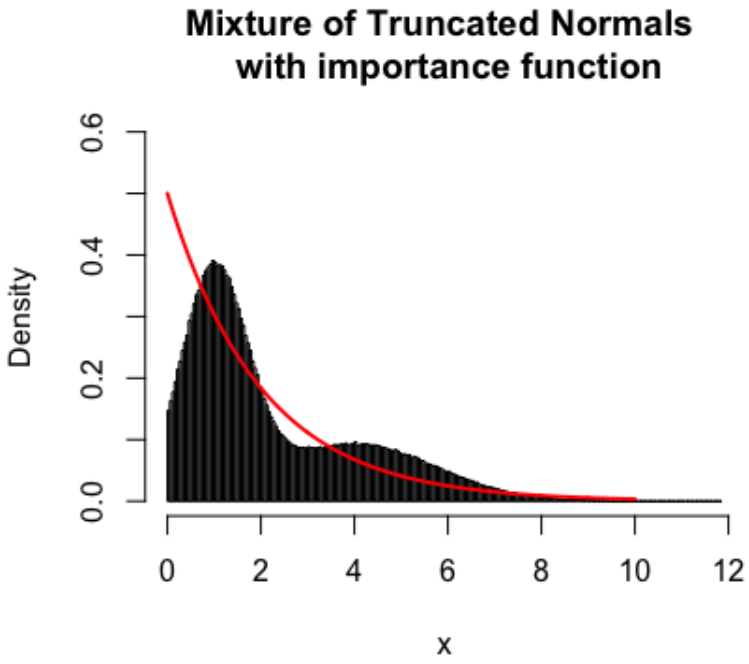
Again the goal is to compute some integral, say $I = \int h(x)g(x)dx = E[h(x)]$. We cannot sample directly from $g(x)$ but we can evaluate the function. So the idea is to find a distribution that we can simulate observations from, $f(x)$, that ideally looks similar to $g(x)$. The importance sampling procedure follows as:

1.

2.

3.

Example. Compute the mean of a mixture of truncated normal distributions. Let $X \sim (.4)N(4,3)_{+(0)} + (.6)N(1,.5)_{+(0)}$.



Let the trial distribution be an Exponential(.5) distribution. Then the importance sampling follows as:

```
> f <- rexp(1e8,.5)
> w <-
+
+
> sum(w*f) / sum(w)
```

Note the `rtnorm()` function has this implementation using importance sampling with an exponential distribution.

Later in class we will spend a little time on Bayesian time series models using dynamic linear models. One way to fit these models is using a specific type of importance sampling in a sequential Monte Carlo framework, known as particle filters.

Monte Carlo Variation

To assess the variation and, more specifically, the convergence of Monte Carlo methods, the Central Limit Theorem is used. The nice thing about this setting is that samples are relatively cheap to come by, only requiring a little more computational run time.

A general prescription is to run the Monte Carlo procedure a few times and assess the variability between the outcomes. Increase the sample size, if needed.

Posterior Inference

Monte Carlo procedures fit naturally in the Bayesian paradigm for creating draws from the posterior distribution. A nice feature of Monte Carlo procedures and posterior inference is for assessing functions of the posterior.

Example. Consider a binomial model where we have a posterior distribution for the probability term, θ. To make inferences on the log-odds $\gamma = \log \frac{\theta}{1-\theta}$, the following procedure is used:

1.

2.

3.

Example. Consider making comparisons between two properties of a distribution. For example a simple contrast, $\gamma = \theta_1 - \theta_2$, or a more complicated function of θ_1, and θ_2 such as $\gamma = \log(\frac{\theta_1}{\theta_2})$. In a classical setting computing distributions, and associated asymptotic properties can be challenging and require large sample approximations. However, this is simple in a Bayesian framework using the same prescription as above.

Posterior Predictive Distribution

Recall the posterior predictive distribution $p(y^*|y_1, ..., y_n)$ is the predictive distribution for an upcoming data point given the observed data.

How does θ factor into this equation?

(1)

Often the predictive distribution is hard to sample from, so a two-step procedure is completed instead.

1.

2.

Similar ideas extend to posterior predictive model checking, which we will return to after studying Bayesian regression.

The Normal Model

A random variable Y is said to be normally distributed with mean θ and variance σ^2 if the density of Y is:

$$p(y|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2\right]$$

Key points about the normal distribution:

-
-
-
-

Inference for θ, conditional on σ^2

When sigma is known, we seek the posterior distribution of $p(\theta|y_1, ..., y_n, \sigma^2)$. A conjugate prior, $p(\theta|\sigma^2)$ is of the form:

$$\begin{aligned} p(\theta|y_1, ..., y_n, \sigma^2) &\propto p(\theta|\sigma^2) \times \exp\left[-\frac{1}{2\sigma^2}\sum (y_i - \theta)^2\right] \\ &\propto \exp\left[c_1 (\theta - c_2)^2\right] \end{aligned}$$

thus a conjugate prior for $p(\theta|y_1, ..., y_n, \sigma^2)$ is from the normal family of distributions. **Note: verifying this will most likely be a homework problem**

Now consider a prior distribution $p(\theta|\sigma^2) \sim N(\mu_0, \tau_0^2)$ and compute the posterior distribution.

$$\begin{aligned} p(\theta|y_1, \ldots, y_n, \sigma^2) &\propto p(y_1, \ldots, y_n|\theta, \sigma^2)p(\theta|\sigma^2) \\ &\propto \\ &\quad \text{now combine terms with the powers of } \theta \\ &\propto \\ &\quad \text{Note from here we could complete the square} \end{aligned}$$

However, there is a shortcut here that you will probably do *approximately* 50 times over the course of this class. Note if $\theta \sim N(E, V)$ then

$$p(\theta) \propto \exp\left[-\frac{1}{2V}(\theta - E)^2\right] \tag{2}$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{\theta^2}{V} - \frac{2\theta E}{V} + c(E, V)\right)\right] \tag{3}$$

Hence from above, the variance of the distribution is the reciprocal of the term with θ^2. That is:

$$V[\theta] =$$

Similarly the term associated with 2θ is E/V, so the expectation is this term times the variance. So the expectation is calculated as:

$$E[\theta] =$$

Notes about the posterior and predictive distributions

- It is common to reparameterize the variance using the inverse, which is known as the precision. Then:
  - $\tilde{\sigma}^2 = 1/\sigma^2$ = sampling precision

- $\tilde{\tau}_0^2 = 1/\tau_0^2$ = prior precision
- $\tilde{\tau}_n^2 = 1/\tau_n^2$ = posterior precision, where τ_n^2 is the posterior variance

Now the posterior precision (i.e. how close the data are to θ) is a function of the prior precision and information from the data: $\tilde{\tau}_n^2 =$

- The posterior mean can be reparameterized as a weighted average of the prior mean and the sample mean.

$$\mu_n =$$

where μ_n is the posterior mean and $\bar{y}$ is the sample mean.

- The predictive distribution of $p(y* | \sigma^2, y_1, \ldots, y_n) \sim N($ $)$.
This will be a homework problem.