STAT 243 HW#4

1. We use the Monte Carlo method to evaluate each of the following integrals:

a.

$$\int_0^1 x^2 dx$$

The exact integral is calculated, for comparison, as:

$$\int_0^1 x^2 dx = \left.\frac{x^3}{3}\right|_0^1 = \frac{1}{3} \approx .333\bar{3}$$

Here, we let $f(x) = \frac{1}{1-0}$, a Uniform(0,1) distribution, and we let $h(x) = x^2$. We take 1,000,000 samples from a Uniform(0,1), and using the Monte Carlo method we obtain an estimate of the evaluated integral as .333328, or equivalenty $\frac{1}{3}$. Thus, we see that the Monte Carlo Method estimated the integral accurately.

b.

$$\int_0^1 \int_{-2}^2 x^2 \cos(xy) \; dxdy$$

We solve the exact integral using computational tools for comparison purposes, and obtain

$$\int_0^1 \int_{-2}^2 x^2 \cos(xy) \; dxdy \approx 3.48318$$

Here, since we have two variables, instead of having $f(x)$ and $h(x)$, we will instead have $f(x,y)$ and $h(x,y)$. Here, we let $f(x,y) = \frac{1}{4}$ and we let $h(x,y) = 4x^2\cos(xy)$. Thus, we see that $\int_0^1 \int_{-1}^2 \frac{1}{4} = 1$, which satisfies our condition for $f(x,y)$. We sample from a Uniform(-2,2) for x and from a Uniform(0,1) for y, each 1,000,000 times, and we plug these values into our function $h(x,y)$. Thus, taking 1,000,000 samples from the two Uniform distribution and using the Monte Carlo method, we obtain an estimate of the evaluated integral of 3.480481. Thus, we see that the Monte Carlo method produced an accurate estimate of the integral.

c.

$$\int_0^\infty \frac{3}{4}x^4 e^{-\frac{x^3}{4}} dx$$

We solve the exact integral for comparison. Using computational tools we obtain:

$$\int_0^\infty \frac{3}{4} e^{-\frac{x^3}{4}} \, dx = 2\sqrt[3]{2}\,\Gamma\left(\frac{5}{3}\right) \approx 2.27478$$

Here, we see that the integral is over a very large range (0 to ∞), compared to the other integrals that we have evaluated. As a result, we do not expect to obtain as accurate results for evaluating the integral as we did in parts a and b. Thus, here we estimate the integral 100 times using 1,000,000 samples and the Monte Carlo method, and plot the density to see the value that the estimates center themselves around. For this integral, we let $f(x)$ and $h(x)$ be defined as:

$$f(x) = \frac{\frac{1}{4}^5}{\Gamma(5)} \cdot x^4 \exp\left(-\frac{x}{4}\right) \sim \text{Gamma}(5, 0.25)$$

$$h(x) = \frac{3}{4} \cdot \frac{\Gamma(5)}{\frac{1}{4}^5} \cdot \exp\left(\frac{x}{4} - \frac{x^3}{4}\right)$$

Thus, we take 1,000,000 samples of x from $x \sim \Gamma(5, 0.25)$ and plug these into $h(x)$ to be used in the formula for estimating the integral. Repeating this process 1000 times, we obtain the following density plot of evaluated integral estimates. The vertical red line represents the maximum of the density, which occurs at the value 2.25775. Here we see that the Monte Carlo method did not approximate the integral as accurately as previous integrals that we approximated (see parts a and b). However, this is expected since we have such a large range that we are integrating over (to infinity). In previous problems, we were integrating over a finite interval of numbers, allowing for more accurate approximation.

Monte Carlo Method Integral Estimates

N = 1000   Bandwidth = 0.04357
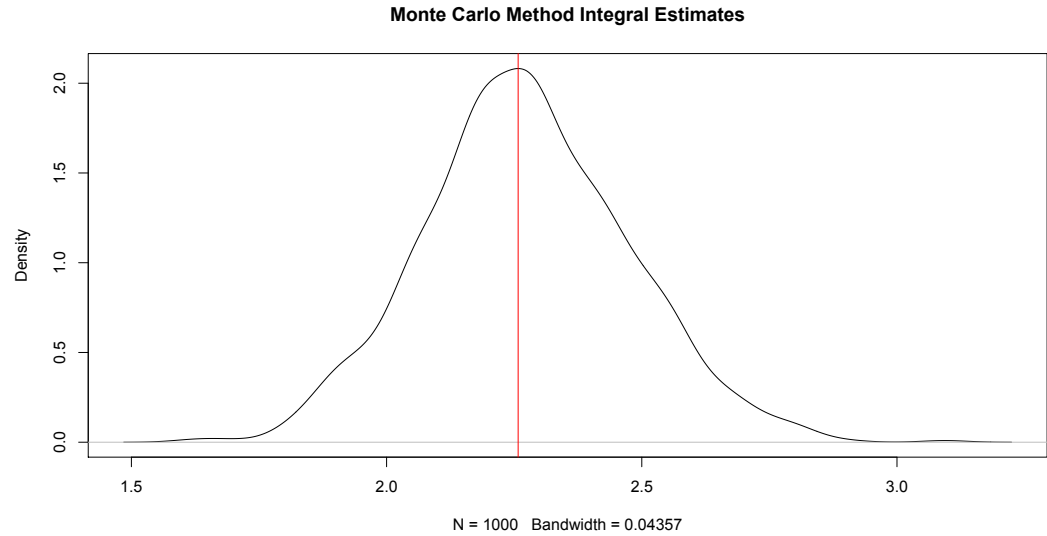
2. We let

$$I = \frac{1}{\sqrt{2\pi}} \int_1^2 e^{-\frac{x^2}{2}}\,dx$$

and estimate I using importance sampling. We take g to be $N(1.5, \nu^2)$ with $\nu = 0.1$ and 10.

Given that $g(x) \sim N(1.5, \nu^2)$, we must find $h(x)$ such that

$$\int_{-\infty}^{\infty} h(x) \cdot \frac{f(x)}{g(x)} \cdot \frac{1}{\sqrt{2\pi\nu^2}} \exp\left\{ -\frac{(x-1.5)^2}{2\nu^2} \right\} = \frac{1}{\sqrt{2\pi}} \int_1^2 e^{-\frac{x^2}{2}}\,dx$$

$$\int_{-\infty}^{\infty} h(x) \cdot \frac{f(x)}{\frac{1}{\sqrt{2\pi\nu^2}} \exp\left\{ -\frac{(x-1.5)^2}{2\nu^2} \right\}} \cdot \frac{1}{\sqrt{2\pi\nu^2}} \exp\left\{ -\frac{(x-1.5)^2}{2\nu^2} \right\} = \frac{1}{\sqrt{2\pi}} \int_1^2 e^{-\frac{x^2}{2}}\,dx$$

Thus, we can easily see that we can let $h(x)$ be the indicator function as follows:

$$h(x) = I_{\{1<x<2\}} = \begin{cases} 1 & \text{if } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

and we can let $f(x) \sim N(0,1)$, the standard normal distribution. With $f(x)$ and $h(x)$ defined as these, we have:

$$\int_{-\infty}^{\infty} I_{\{1<x<2\}} \cdot \frac{\frac{1}{\sqrt{2\pi}\exp\left\{-\frac{x^2}{2}\right\}}}{\frac{1}{\sqrt{2\pi\nu^2}}\exp\left\{-\frac{(x-1.5)^2}{2\nu^2}\right\}} \cdot \frac{1}{\sqrt{2\pi\nu^2}}\exp\left\{-\frac{(x-1.5)^2}{2\nu^2}\right\} = \frac{1}{\sqrt{2\pi}}\int_1^2 e^{-\frac{x^2}{2}}\,dx$$

Now that we have defined $h(x)$ and $f(x)$, we carry out the Importance Sampling algorithm as follows:

1. We generate $\tilde{x}_1, \ldots, \tilde{x}_m$ iid from $g(\tilde{x}) \sim N(1.5, \nu^2)$.

2. We estimate $\mu = \int h(\tilde{x}) \cdot \frac{f(\tilde{x})}{g(\tilde{x})} \cdot g(\tilde{x})dx$ by $\hat{\mu}_g = \frac{1}{m}\sum_{i=1}^m h(\tilde{x}_i | w(\tilde{x}_i))$ where $w(\tilde{x}_i)$ is the importance ration: $w(\tilde{x}_i) = \frac{f(\tilde{x})}{g(\tilde{x})}$

We first use a computational tool to solve the integral directly, for comparison purposes. In doing so, we obtain:

$$\frac{1}{\sqrt{2\pi}}\int_1^2 e^{-\frac{x^2}{2}}\,dx = \frac{1}{2}\left(\text{erf}(\sqrt{2}) - \text{erf}\left(\frac{1}{\sqrt{2}}\right)\right) \approx 0.135905$$

Where **erf** is the errror function. We carry out this process for $\nu = 0.1, 1, 10$, each with m = 1000000. From, these, we obtain the following estimates of the integral:

Integral Estimates

| ν	0.1	1	10
$\hat{\mu}_g$	0.03262997	0.1360167	0.01375336

Looking at these estimates, we see that $\nu = 1$ provided the best estimate for the integral. On the following pages, we plot histograms of the values which we averaged for each value of ν, in order to determine if there are any extreme values.

Values to be Averaged for nu = 0.1

nu = 0.1
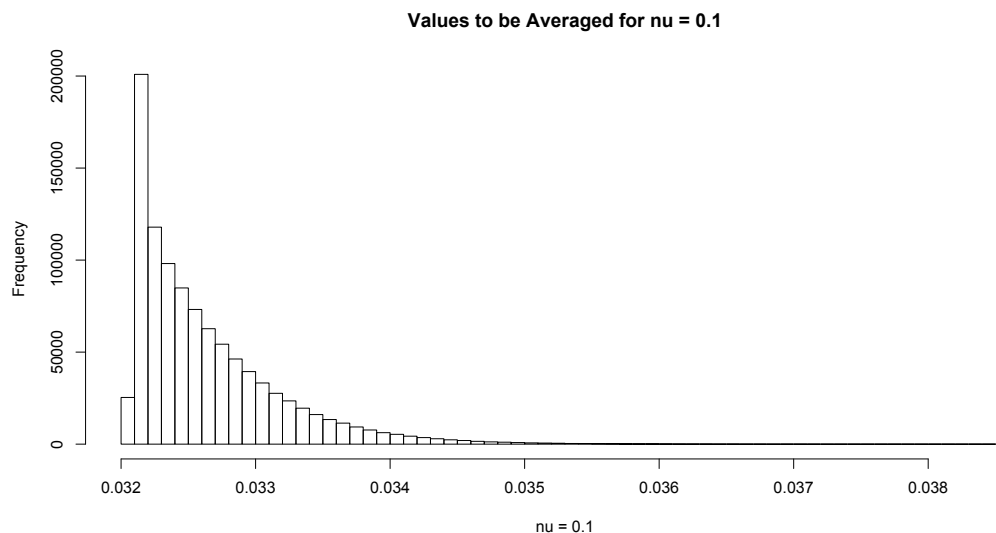
Looking at the histogram for $\nu = 0.1$, we see that it the values of $h(\tilde{x}_i | w(\tilde{x}_i))$ cover a range between 0.03210 and 0.03858. Thus, we see that none of the $\tilde{x}$ values generated from $g(\tilde{x})$ were outside of the range (1, 2), since we have no values that are equal to zero (and thus would have been rejected by the indicator function). However, we also note that the majority of values of $h(\tilde{x}_i | w(\tilde{x}_i))$ appear to be between 0.032 and 0.035, which is a very small range, and is nowhere near what we know the true value of the integral to be. This is because, with such a small ν, we are unable to explore the full space. Here,is also looks as if there are some extreme values on the upper end, as we see that the distribution is heavily skewed.

Values to be Averaged for nu = 1

Frequency

nu = 1
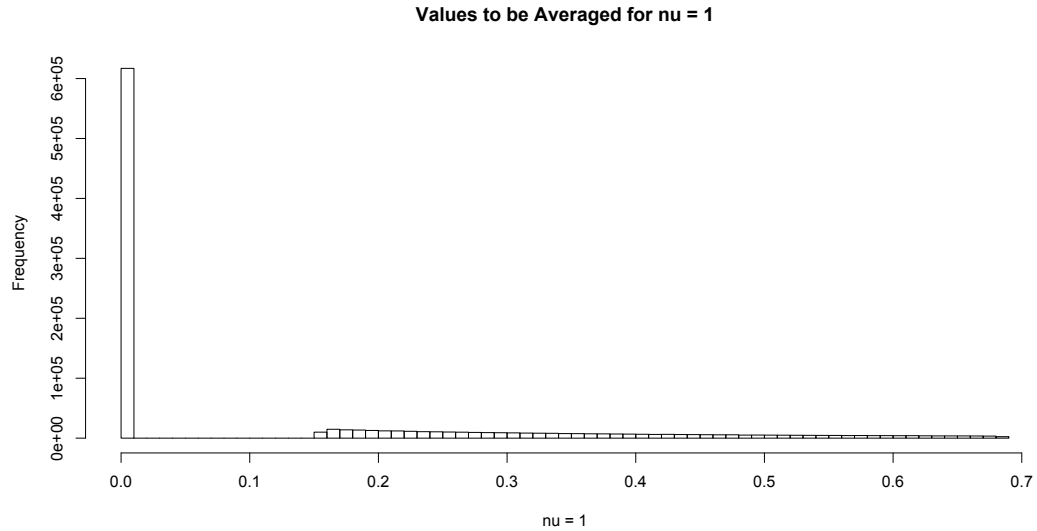
For $\nu = 1$, we see that our $h(\tilde{x}_i|w(\tilde{x}_i))$ values cover a much larger range. Here, we see that our values of $h(\tilde{x}_i|w(\tilde{x}_i))$ range from 0 to 0.6873. Here, we see that we have 617008 zeros, which means that there were 617008 or 61.7008% of $\tilde{x}$'s that were drawn outside of the range from 1 to 2. Here, we see that the ν was large enough to let us explore the entire range, yet not too large, as confirmed by our accurate estimate of the integral. This is to be expected, as in the original integral, we notice that we are simply integrating the standard normal distribution (and thus it has a variance of 1). Here, we let the variance of the underlying distirbution equl to 1, and we see that this leads to a very accurate estimate.

Values to be Averaged for nu = 10
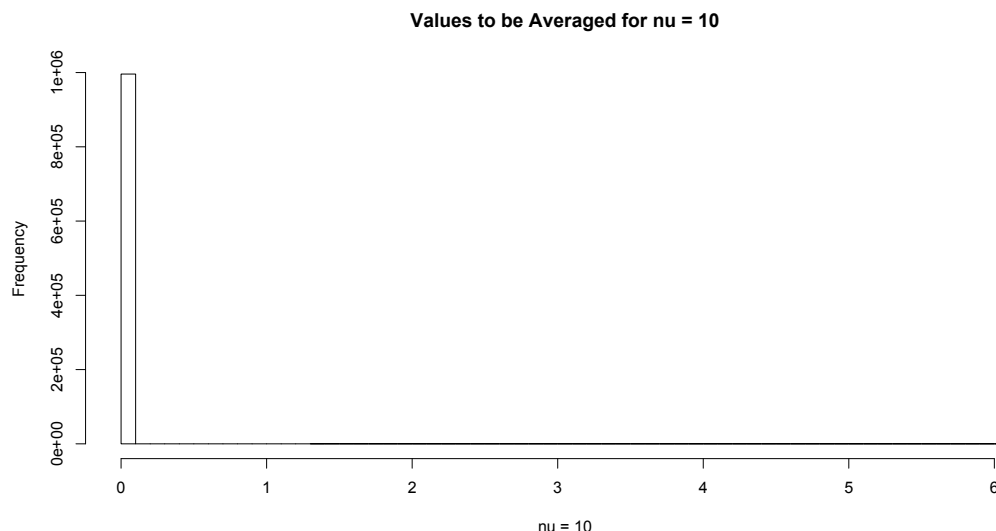
For $\nu = 10$, we see that almost all of the values of $\tilde{x}$ get pushed to zero, as they are almost all out of the range (1,2). Here, we have 995,961, or 99.5961% of $\tilde{x}$'s that are sampled out of the (1,2) range and are thus made to be zero. This is expected, since the $\tilde{x}$'s were sampled from a N(1.5, 10) distibution, and thus with such a large variance, we would expect many of the values to fall outside of (1,2). Thus, here we see that the majority of values of $h(\tilde{x}_i|w(\tilde{x}_i))$ are 0, as expected, with some extreme values on the top end, as we see that the range of the values of $h(\tilde{x}_i|w(\tilde{x}_i))$ extends all the way past 6. This pulls our estimate of the integral for $\nu = 10$ up, yet all of the zero's keep the estimate quite small. Again, this is expected since, with such a large in the underlying distribution, we explore a much larger space than we are actually interested in.

3. We will approximate the following integral using both the simple Monte Carlo integration and the control variate method:

$$I = \int_0^1 \frac{1}{1+x}dx$$

   a. We let $h(x) = \frac{1}{1+x}$ and $U_1, \ldots U_n$ be iid Unif[0,1]. We estimate $\hat{I}_{MC} = \frac{1}{n}\sum_{i=1}^n h(U_i)$ with n = 1500 in R and obtain an estimate of $\hat{I}_{MC} = 0.6941916$. We note that the tru value for I is known to be $I = \ln(2) = 0.6931472$, thus we see that our estimate for I is accurate.

b. Next, we introduct $c(x) = 1 + x$ as a control variate, and we estimate i with:

$$\hat{I}_{CV} = \frac{1}{n} \sum_{i=1}^{n} h(U_i) - b \left[\frac{1}{n} \sum_{i=1}^{n} c(U_i) - E\{c(U)\} \right]$$

In order to carry out this estimation, we must first find the optimal value of b to plug into the expression. From class, we know that the optimal value of b will minimize the variance of the estimator, $\hat{I}_{CV}$. Based on the derivations from class, we have that the optimal value of b which will minimize the variance of $\hat{I}_{CV}$ is equal to:

$$b_{OPT} = \frac{\text{cov}(\hat{I}_{CV}, \hat{\theta}_{MC})}{\text{var}(\hat{\theta}_{MC})} = \frac{\text{cov}(h(\tilde{x}), c(\tilde{U}))}{\text{var}(c(\tilde{U}))}$$

where $\hat{\theta}_{MC} = \frac{1}{n} \sum_{i=1}^{n} c(U_i)$.

Recall from part a that $h(x) = \frac{1}{1+x}$ and $E[h(x)] = \ln 2$. What is left is to find $E[c(U)]$ and $\text{cov}(h(x), c(U))$. Recall that $c(x) = 1 + x$ and since we are on the range of (0,1), we have that the expected value of $c(x)$ is equal to 1.5. Thus, to find our optimal b, we first draw 1500 x's from $x \sim \text{Unif}(0, 1)$, then plug these x's into $h(x) = \frac{1}{1+x}$ and $c(x) = 1 + x$. We then estimate the covariance between these generated $h(x)$ and $c(x)$'s, and the variance of $c(x)$. In doing so, we obtain the following estimate of b: $b_{OPT} = -0.4747049$.

c. We estimate the variances of $\hat{I}_{MC}$ and $\hat{I}_{CV}$ by taking 1,000,000 samples from each in R and obtain:

$$\text{var}(\hat{I}_{MC}) = 1.274619e^{-05}$$

$$\text{var}(\hat{I}_{CV}) = 4.002654e^{-07}$$

Thus, we see that the variance of $\hat{I}_{CV}$ is much smaller than the variance of $\hat{I}_{MC}$, as we would have hoped. We calculate the variance reduction factor derived in class by:

$$\text{variance reduction factor} = \frac{var(\hat{I}_{CV})}{var(\hat{I}_{MC})} \approx 0.03093002$$

d. Although we have already designed an estimator, $\hat{I}_{CV}$, which is much more efficient than our original estimator, $\hat{I}_{MC}$, it is still possible to design a

new estimator for I that would have a smaller variance that $\hat{I}_{CV}$. To do so, we would simply add another control variate, so that instead of just one control variate $(c(x) = 1+x)$, we would have two control variates and thus estimate two b's. Thus, our estimate would be obtained by finding b_1^* and b_2^* that minimize the variance of the following function:

$$\hat{I}_{CV} = \hat{I}_{MC} - b_1(\hat{\theta}_{1MC} - \theta_1) - b_2(\hat{\theta}_{2MC} - \theta_2)$$

where $\hat{\theta}_{1MC} = \frac{1}{n}\sum_{i=1}^{n} c_1(x_i)$ and $\hat{\theta}_{2MC} = \frac{1}{n}\sum_{i=1}^{n} c_2(x_i)$, and $c_1(x_i)$ is the first control variate, and $c_2(x_i)$ is the second control variate. You could theoretically keep adding control variates in this fashion, until the variance reduction factor no longer reduces (adding addition control variates no longer minimizes the variance), or the changes in the variance reduction factor is arbitrary. The variance reduction factor is defined as:

$$\text{variance reduction factor} = \frac{\text{var}(\hat{I}_{CV}^{OPT})}{\text{var}(\hat{I}_{MC}^{OPT})} = 1 - \rho^2$$

where ρ is the correlation coefficient between $h(\tilde{x})$ and $c(\tilde{x})$. However; in general, adding more control variates results equations that are both computationally and analytically difficult to solve, and thus at most only one or two control variates are generally used in practice.

4. We consider a common application in statistics: three different treatments are to be compared by applying them to a randomly selected experimental units. However; instead of the usualy assumption associated with e_{ij} in the typical analysis of variance methods of comparison, we assume that the e_{ij} have independent and identical double exponential distributions centered on zero.

   a. Here we have the model:

   $$y_{ij} = \mu + \alpha_i + e_{ij} \quad 1 \leq i \leq 3, \quad 1 \leq j \leq n_i$$

   $$e_{ij} \sim \quad iid \; \frac{1}{2\sigma}\exp\left(-\frac{|e_{ij}|}{\sigma}\right), \quad -\infty < e_{ij} < \infty$$

   Where α_i corresponds to the treatment i effect, and the treatment means are denoted:

   $$\mu_i = \mu + \alpha_i$$

   We are interested in testing whether or not there are no treatment effects, that is $H_o : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs. $H_a :$ not all $\alpha_i = 0$. To do so, we use the

likelihood-ratio test statistic. In order to derived the likelihood ratio, we first look at the distributions of e_{ij} and y_{ij}. As stated above, we have that the error term is distributed as Laplace: $e_{ij} \sim \text{Laplace}(0, \sigma)$ and thus:

$$e_{ij} \sim \quad iid \ \frac{1}{2\sigma}\exp\left(-\frac{|e_{ij}|}{\sigma}\right), \quad -\infty < e_{ij} < \infty$$

Thus, from properties of the Laplace distribution, it follows that since e_{ij} is distributed Laplace, and y_{ij} is a linear combination of e_{ij}, then y_{ij} is also distributed Laplace as follows:

$$f(y|\mu_1, \mu_2, \mu_3, \sigma) \sim \frac{1}{2\sigma}\exp\left(-\frac{|y_{ij} - \mu_i|}{\sigma}\right)$$

The likelihood-ratio test statistic is defined as:

$$\Lambda =$$

Now, we can implement the Monte Carlo test to test whether or not the null hypothesis holds. We carry out the Monte Carlo test as follows:

1. Sample y

2. Compute the likelihood ratio test statistic Λ

3. Repeat a large number of times.

b. Another computer-intensive test that we could use even if we make not assumptions about the distribution of e_{ij} is the permutation test, also known as the randomization test. The permutation test is a non-parametric test, and thus does not rely on any parameter assumptions to yield accurate results. The main application of the permutation test is to two-sample problems, but we can apply the approch to this three - treatment problem as well. We denote the variables as treatment one being denoted as $x_1, \ldots, x_n \quad iid \quad \sim G$, treatment two is $y_1, \ldots, y_m, \quad iid \quad \sim F$, and treatment three as $z_1, \ldots, z_k \quad iid \quad \sim H$. We would test the hypothesis:

$$\text{H}_o : F = G = H \quad \text{vs.} \quad \text{H}_a : \text{not all distributions are equal}$$

using the test statistic:
$$\hat{\theta} = \bar{z} - \bar{y} - \bar{x}$$

To carry out this test, we would merge $x_1, \ldots x_n$, $y_1, \ldots y_m$ and $z_1, \ldots z_k$ into one large set, and then randomly sample without replacement n observations to serve as group 1, followed by a randomly sampling without

replacement m observations from the remaining to serve as group 2, and then the remaining k observations would serve as group 3. If the null hypothesis is correct - the three groups are equivalent - then we would expect the means from group 1, 2 and 3 all be similar to one another, and thus $\hat{\theta}$ should be zero. Thus, we calculate the value of $\hat{\theta}_1$ for this sample, and then repeat the procedure many, many times (join the treatments again, randomly sample without replacement to form three new groups, calculate $\hat{\theta}_i$ for i = some large number of repetitions). Thus, we will now have our original $\hat{\theta} = \bar{z} - \bar{y} - \bar{x}$ and a large number of $\hat{\theta}_i$'s. We look at the distribution of all of our $\hat{\theta}_i$'s and compare it to the $\hat{\theta}$ that we calculated form the data that was split into their original treatments. If this original $\hat{\theta}$ is outside of the middle 95% of the mean differences, then we would reject the null hypothesis and have evidence to suggest that not all distributions are equal.

5. We consider the zero-inflated Poisson(ZIP) model, in which random data $X_1, \ldots, X_n$ are assumed to be of the form $X_i = R_i Y_i$, where Y_i's have a Poisson(λ) distribution and the R_i's have a Bernoulli(p) distribution, all independent of each other. Given an outcome $x = (x_1, \ldots, x_n)$, the objective is to estimate both λ and p. We consider the following hierarchical Bayes model:

   - $p \sim$ Uniform(0,1) (prior for p)

   - $(\lambda|p) \sim$ Gamma(a, b) (prior for λ)

   - $(r_i|p, \lambda) \sim$ Bernoulli(p) independently (from the model above)

   - $(x_i|r, \lambda, p) \sim$ Poisson(λr_i) independently (from the model above)
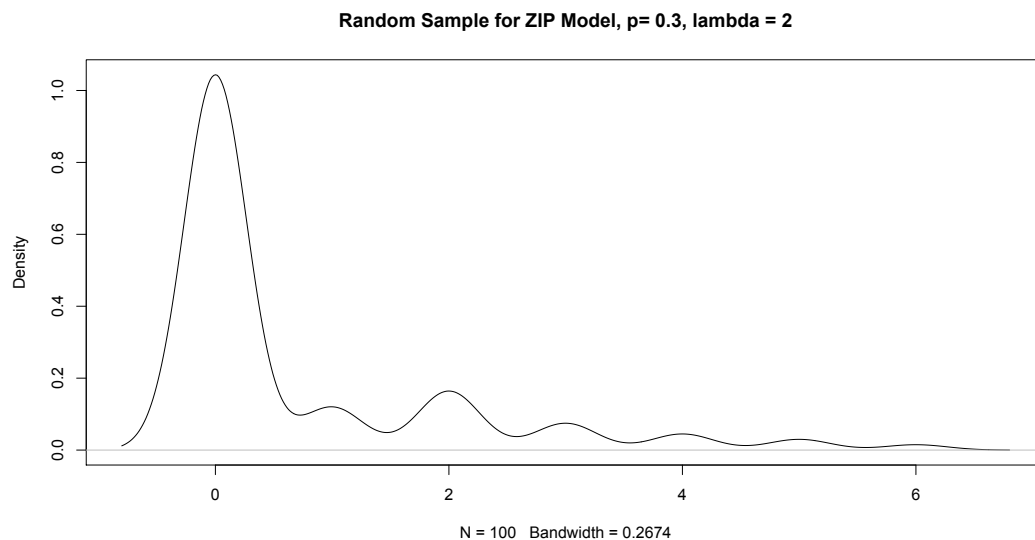
   where a and b are known parameters, and $r = (r_1, \ldots, r_n)$. It follows that:

   $$f(x, r, \lambda, p) = \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \prod_{i=1}^{n} \frac{e^{-\lambda r_i}(\lambda r_i)^{x_i}}{x_i!} p^{r_i}(1-p)^{1-r_i}$$

   We wish to sample from the posterior pdf $f(\lambda, p, r|x)$ using the Gibbs sampler.

   a. We generate a random sample of size n = 100 for the ZIP model using

parameters p $= 0.3$ and $\lambda = 2$. A density plot of this sampe is below.

Random Sample for ZIP Model, p= 0.3, lambda = 2

N = 100   Bandwidth = 0.2674

b. We show that the following conditional distributions hold. In doing so, note that we can prove these up to a constant of proportionality, because the normalizing integrals that would make the following true, get rid of the variables that are "given" in each. Another way of thinking of this is that we can look at $f(\mathbf{x}, \mathbf{r}, \lambda, p)$ for each of (i), (ii), and (iii) given below, and ignore the variables that are on the right hand side of the "given" line. This is because, since these variables are given, we can just treat them as constants.

i. $(\lambda|p, \mathbf{r}, \mathbf{x}) \sim \text{Gamma}(a + \sum_i x_i, b + \sum_i r_i)$

$$(\lambda|p, \mathbf{r}, \mathbf{x}) \propto \lambda^{a-1} e^{-b\lambda} e^{-\lambda \sum_{i=1}^n r_i} \lambda^{\sum_{i=1}^n x_i}$$

$$\Rightarrow (\lambda|p, \mathbf{r}, \mathbf{x}) \propto \lambda^{a + \sum_{i=1}^n x_i - 1} e^{-(b + \sum_{i=1}^n r_i)\lambda}$$

$$\Rightarrow (\lambda|p, \mathbf{r}, \mathbf{x}) \propto\sim \Gamma\left(a + \sum_{i=1}^n, b + \sum_{i=1}^n\right)$$

ii. $(p|\lambda, \mathbf{r}, \mathbf{x}) \sim \text{Beta}(1 + \sum_i r_i, n + 1 - \sum_i r_i)$

$$(p|\lambda, \mathbf{r}, \mathbf{x}) \propto \prod_{i=1}^{n} p^{r_i}(1-p)^{1-r_i}$$

$$\Rightarrow (p|\lambda, \mathbf{r}, \mathbf{x}) \propto p^{\sum_{i=1}^{n} r_i}(1-p)^{n-\sum_{i=1}^{n} r_i}$$

$$\Rightarrow (p|\lambda, \mathbf{r}, \mathbf{x}) \propto \text{Beta}\left(1 + \sum_{i=1}^{n} r_i, n + 1 - \sum_{i=1}^{n} r_i\right)$$

iii. $(r_i|\lambda, p, \mathbf{x}) \sim \text{Bernoulli}\left(\frac{pe^{-\lambda}}{pe^{-\lambda}+(1-p)I_{\{x_i=0\}}}\right)$

$$(r_i|\lambda, p, \mathbf{x}) \propto \prod_{i=1}^{n} \frac{e^{-\lambda r_i}(\lambda r_i)^{x_i}}{x_i!} p^{r_i}(1-p)^{1-r_i}$$

$$(r_i|\lambda, p, \mathbf{x}) \propto e^{-\lambda \sum_{i=1}^{n} r_i} \sum_{i=1}^{n}(r_i)^{x_i} p^{\sum_{i=1}^{n} r_i}(1-p)^{n-\sum_{i=1}^{n} r_i}$$

Here we see that we have a Bernoulli distribution, with probability of success defined as:
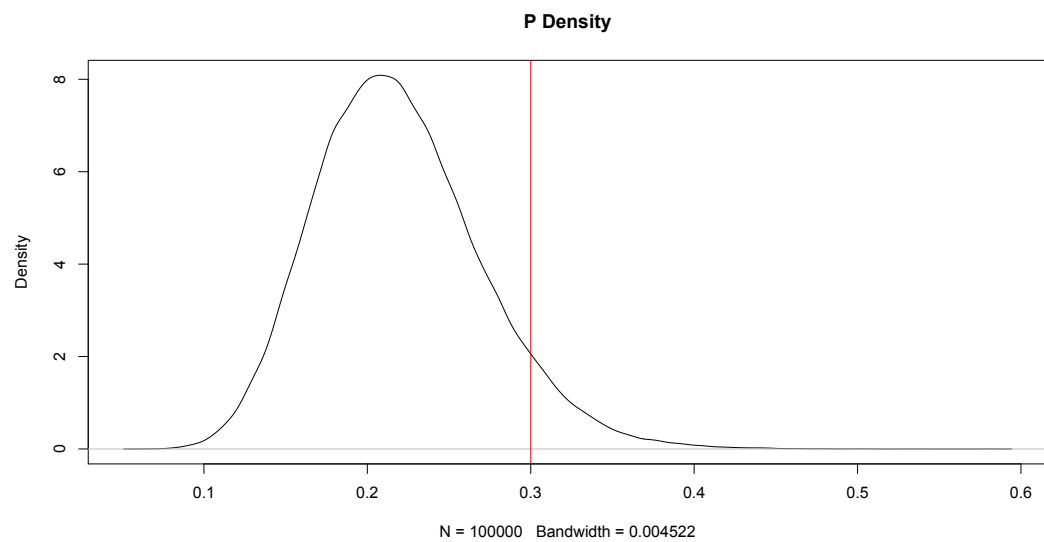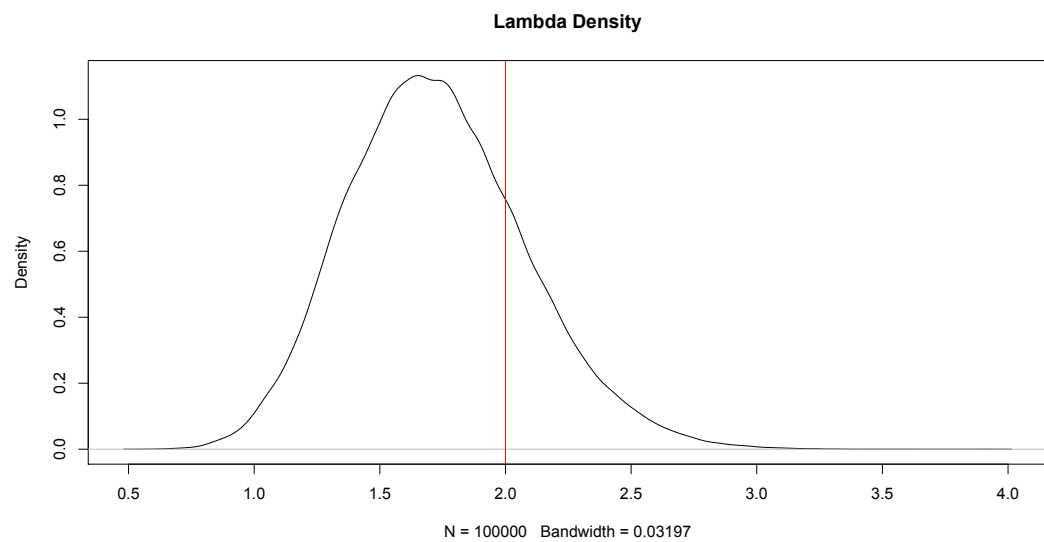$$\frac{pe^{-\lambda}}{pe^{-\lambda} + (1-p)I_{\{x_i=0\}}}$$

This is because r_i is Bernoulli, and thus can either take 0 or 1 values. So thus, we have that $(r_i|\lambda, p, \mathbf{x})$ can either take the following values, for each given i:

$$(r_i|\lambda, p, \mathbf{x}) = \begin{cases} e^{-\lambda}p & \text{if } r_i = 1 \\ (1-p)I_{\{x_i=0\}} & \text{it } r_i = 0 \end{cases}$$

The, the probability of success is the value when $r_i = 1$ over the values of $r_i = 1$ and $r_i = 0$. Thus, we are left with:

$$(r_i|\lambda, p, \mathbf{x}) \sim \text{Bernoulli}\left(\frac{pe^{-\lambda}}{pe^{-\lambda} + (1-p)I_{\{x_i=0\}}}\right)$$

c. We implement the Gibbs sampler to generate a large (dependent) sample from the posterior distribution, and use this to construct 95% Bayesian confidence intervals for p and λ. We begin with a = b = 1 and n = 100,00. Plots of the dependent posterior sample can be seen below, the vertical line in each represents the respective true values ($\lambda = 2$ and $p = 0.30$).

Lambda Density

N = 100000   Bandwidth = 0.03197

P Density

N = 100000   Bandwidth = 0.004522

The 95% Bayesian confidence intervals for λ and p are given by:

95% Bayesian Confidence Intervals

paramater	2.5%	97.5%
λ	1.094145	2.486432
p	0.1318434	0.3300942

Here, we see that both of the 95% Bayesian confidence intervals contain the true values of their parameters. Thus, it appears that the Gibbs sampler method performed well here. We now run the algorithms trying different values of a and b. Here, we let a and b equal all possible pairs of the values 0.5, 1, 1.5, 2, 5, 10. The table on the following page shows thet 95% Bayesian confidence intervals for λ and p for each of these combinations. From these we see that the true λ, $\lambda = 2$, is contained in 80.56% of the intervals, and the true p $p = 0.30$ is contained in 94.44% of the intervals. Below is a table that gives the values of a and b for which the true values of either λ or p were not in the 95% Bayesian confidence interval. We note, however, that there was not a pairing of a and b for which both true parameter values were not in their respective confidence intervals.

a	b	true λ in ci	true p 0.3 in ci
0.50	0.50	Yes	No
0.50	1.50	Yes	No
0.50	5.00	No	Yes
0.50	10.00	No	Yes
1.00	10.00	No	Yes
1.50	5.00	No	Yes
1.50	10.00	No	Yes
2.00	10.00	No	Yes
10.00	10.00	No	Yes

95% Bayesian Confidence Intervals

a	b	lambda 2.5%	lambda 97.5%	p 2.5%	p 97.5%
0.50	0.50	1.18225	2.29708	0.23426	0.46619
0.50	1.00	1.15448	2.44081	0.16008	0.36708
0.50	1.50	1.36638	2.62522	0.17462	0.37379
0.50	2.00	1.25703	2.31445	0.25211	0.48498
0.50	5.00	1.21119	2.35611	0.16121	0.36494
0.50	10.00	1.01142	1.84822	0.27445	0.53661
1.00	0.50	1.10757	2.22492	0.22643	0.46454
1.00	1.00	1.21069	2.17621	0.32107	0.57398
1.00	1.50	1.38823	2.54760	0.21438	0.42483
1.00	2.00	1.79457	3.06473	0.19352	0.38285
1.00	5.00	1.07853	2.15017	0.18716	0.41221
1.00	10.00	1.12645	1.96275	0.29917	0.55593
1.50	0.50	1.30640	2.47520	0.21678	0.43497
1.50	1.00	1.72349	2.97749	0.20466	0.39820
1.50	1.50	1.55256	2.71074	0.22949	0.43750
1.50	2.00	1.28477	2.48313	0.18763	0.39565
1.50	5.00	1.75724	2.91669	0.20459	0.39817
1.50	10.00	0.95236	1.84849	0.20993	0.45693
2.00	0.50	1.61309	2.72191	0.26762	0.48107
2.00	1.00	0.82764	1.88243	0.22352	0.49865
2.00	1.50	1.51056	2.72678	0.20018	0.40144
2.00	2.00	1.59258	2.81992	0.19808	0.39467
2.00	5.00	0.93764	1.82662	0.28969	0.57103
2.00	10.00	0.79592	1.67958	0.18733	0.44780
5.00	0.50	1.67622	2.88744	0.22529	0.42702
5.00	1.00	1.58953	2.83467	0.19800	0.39488
5.00	1.50	1.67541	2.84001	0.23484	0.44006
5.00	2.00	1.81842	2.96621	0.25133	0.45557
5.00	5.00	1.36577	2.40414	0.22801	0.44344
5.00	10.00	1.03043	1.95756	0.19377	0.42256
10.00	0.50	1.36070	2.44332	0.26776	0.49683
10.00	1.00	1.56820	2.73930	0.22811	0.43540
10.00	1.50	1.38723	2.63277	0.18360	0.38616
10.00	2.00	1.69544	2.88688	0.21513	0.41476
10.00	5.00	1.19676	2.16314	0.26816	0.51072
10.00	10.00	1.45066	2.43480	0.21618	0.42394

6. The Independence - Metropolis - Hastings Algorithm is an importance-sampling version of MCMC. We draw the proposal from a fixed distribution g. Generally, g is chosen to be an approximation to f. The acceptance probability becomes:

$$r(x, y) = min \left\{ \frac{f(y)}{f(x)} \cdot \frac{g(x)}{g(y)}, 1 \right\}$$

A random variable Z has an inverge Gaussian distribution it it has density

$$f(z) \propto z^{-3/2} \exp \left\{ -\theta_1 z - \frac{\theta_2}{z} + 2\sqrt{\theta_1 \theta_2} + \log \sqrt{2\theta_2} \right\}, \quad z > 0$$

where $\theta_1 0$ and $\theta_2 > 0$ are parameters. It can be shown that

$$E(Z) = \sqrt{\frac{\theta_2}{\theta_1}} \quad \text{and} \quad E\left(\frac{1}{Z}\right) = \sqrt{\frac{\theta_1}{\theta_2}} + \frac{1}{2\theta_2}$$

We let $\theta_1 = 1.5$ and $\theta_2 = 2$, and draw a sample of size 1,000 using the independence-Metropolis-Hastings algroithm. We use a Gamma(1.5, 2) distribution as the proposal density. Below is a density plot of the sample drawn using the Independence-Metropolis-Hastings algorithm with a Gamma distribution as the proposal density.

Independence - Metropolis-Hastings with Gamma Proposed Density

N = 1000   Bandwidth = 0.1204
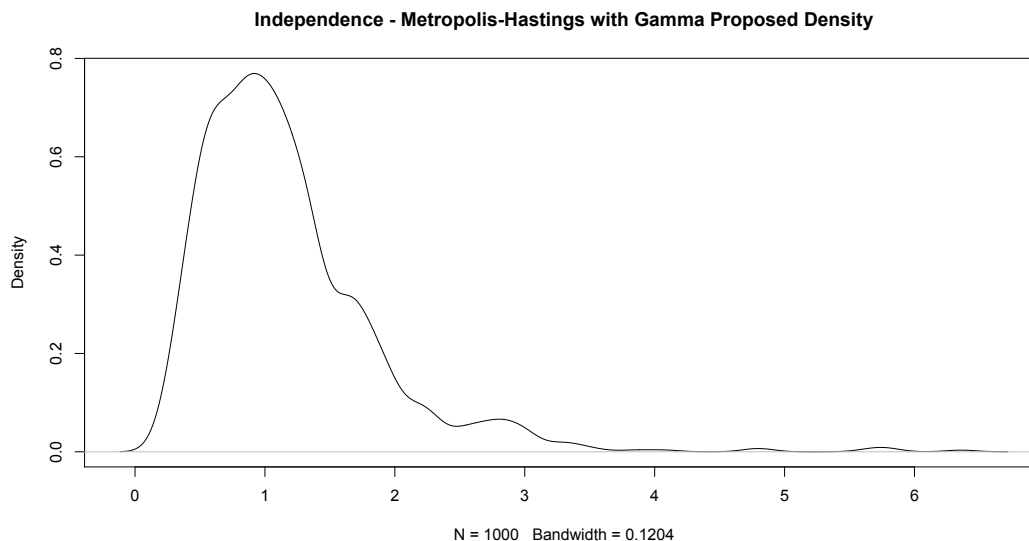
To assess the accuracy of the Independence - Metropolis - Hastings algroithm, we compare the mean of Z and $1/Z$ from the sample to the theoretical means. The theoretical means are derived below:

$$E[Z] = \sqrt{\frac{\theta_2}{\theta_1}} = \sqrt{\frac{2}{1.5}} \approx 1.154701$$

$$E\left[\frac{1}{Z}\right] = \sqrt{\frac{\theta_1}{\theta_2}} + \frac{1}{2\theta_2} = \sqrt{\frac{1.5}{2}} + \frac{1}{2 \cdot 2} \approx 1.116025$$

Our sample of n = 1,000,000 yields the following estimates of the means:

$$\hat{E}[Z] \approx 1.153933, \quad \hat{E}\left[\frac{1}{Z}\right] \approx 1.116973$$

From this, we see that the Independence - Metropolis - Hastings algorithm performed pretty well, as the estimates based on our sample very, very similar to the theoretical values, to several decimal places. The remaining difference is likely due to the sample size, and we believe that the sample values will approach the theoretical values as n $\to \infty$. Thus, it appears that our sample is relatively accurate.

Fixing our $\theta_1 = 1.5$ and $\theta_2 = 2$, we try several different Gamma distributions to see if we can get an accurate sample. For computational ease, we now let n = 1000 as opposed to n = 1000000. Here, we choose distributions Gamma(a, b) where a and b take every possible pair-wise pairing of the values 0.5, 1, 1.5, 2, 2.5, 3, 4 , 5. In carrying out the Independence-Metropolis-Hastings algorithm for all possible pairs, we find that the best pairing that produced the minimum absolute differences were a = 1.5 and b = 2.0 to minimize the absolute differece between E(Z) and $\hat{E}(Z)$, and a = 3, b = 3 to minimize the absolute difference between $E\left(\frac{1}{Z}\right)$ and $\hat{E}\left(\frac{1}{Z}\right)$. However, we note that almost all of the different Gamma distributions used appeard to provide accurate samples, with the largest differences being when Gamma(5, 0.5) was used, with an absolute difference of 1.5401086 for E(Z) and 0.6651700 for E$\left(\frac{1}{Z}\right)$. Below we compare the Gamma distributions with the original and their estimates. Here, we compare

the estimates when n = 1000 for all.

Gamma(a, b) Sample Comparisons

a	b	E.z	target.E.z	abs. diff.	E.inv.z	target.E.inv.z	abs. diff.
1.5	2.0	1.133058	1.154701	0.021643	1.123595	1.116025	0.007569
2.0	1.5	1.154747	1.154701	0.000047	1.120085	1.116025	0.004060
3.0	3.0	1.166071	1.154701	0.011370	1.116180	1.116025	0.000155

Here we see that with a sample size of n = 1000 for each different Gamma(a,b) distribution, the sample from Gamma(2.0, 1.5) was the most accurate, followed by Gamma(3,3). Gamma(1.5, 2) was also quite accurate, but we note that, in order to see as accurate results as mentioned previously, we would need to use a much larger sample size.