Homework 4

Jared Yu

May 22, 2018

Problem 1.

Use Monte Carlo to evaluate each of the following integrals:
(a) If we allow $X \sim U(0,1)$.
then $E(x^2) = \int_0^1 x^2 dx$.
Since $f(x) = 1$
The approximation to $\int_0^1 x^2 dx$ is the same as approximation to $E(x^2)$.
So, we can simulate $X_i \overset{\text{iid}}{\sim} U(0,1)$, $i = 1, \cdots, n$
$\int_0^1 x^2 dx \approx \frac{\Sigma x_i^2}{n}$
After doing this simulation in RStudio, sampling from runif() and taking the mean of x^2, the result is 0.3335771. Which is quite close to the result from just taking the integral:
$\int_0^1 x^2 dx = \frac{1}{3} x^3 |_0^1 = \frac{1}{3}$
Which evaluate to something quite similar.

(b) If we allow X, Y to be independent, and $X \sim U(-2,2)$, and $Y \sim U(0,1)$.
Then we have $f(x) = \frac{1}{4}$ and $f(y) = 1$, so $f(x,y) = f(x)f(y) = \frac{1}{4}$
The original equation, $\int_0^1 \int_{-2}^2 x^2 cos(xy) dx dy = 4 \int_0^1 \int_{-2}^2 x^2 cos(xy) f(x,y) dx dy$
$= 4E[x^2 cos(xy)]$
After sampling both X and Y from their Uniform distributions in RStudio, and taking the mean of what's inside the expecation and multiplying by 4 the result is 3.550365.

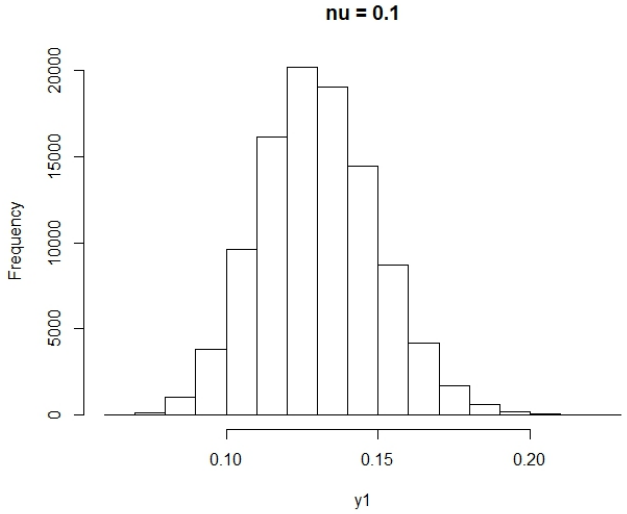
(c) Using $k = 3$ and $\lambda = \sqrt[3]{4}$, it is possible to determine the original distribution is the Weibull distribution.

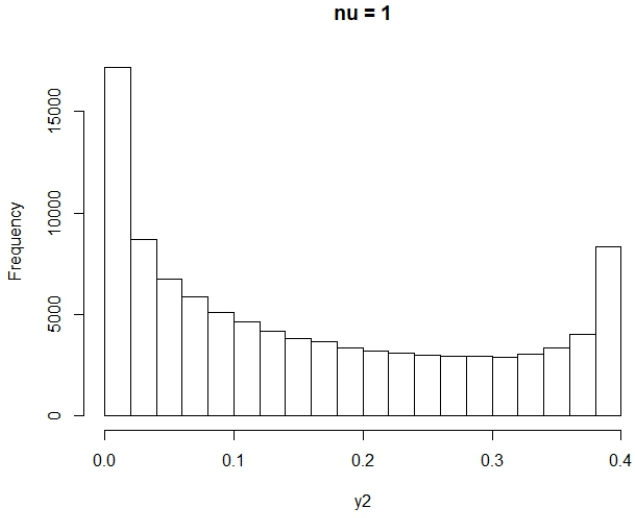
$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k}$, $x \geq 0$
$= \frac{3}{\sqrt[3]{4}}(\frac{x}{\sqrt[3]{4}})^{3-1}e^{-(x/\sqrt[3]{4})^3}$
$= \frac{3}{(4)^{(1/3)}}\frac{x^2}{(4)^{(2/3)}}e^{-x^3/4}$
$= \frac{3x^2}{4}e^{-x^3/4}$
Plugging this equation back into the given question from the problem, it's possible to see what it equates to.
$\int_0^\infty \frac{3x^4}{4}e^{-x^3/4}dx$
$\int_0^\infty x^2\frac{3x^2}{4}e^{-x^3/4}dx$
$\int_0^\infty x^2 f(x)dx$
$= E(x^2)$
After using RStudio to sample from the rweibull() distribution using the corresponding parameters, the result of the mean of x^2 is 2.289859.

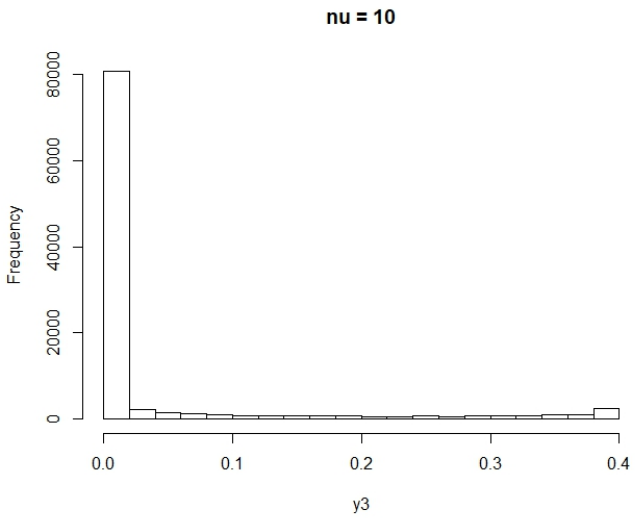
Problem 2.

If we let $\mu = \frac{1}{\sqrt{2\pi}}\int_1^2 e^{-x^2/2}dx$
$= \int_1^2 \frac{\frac{1}{\sqrt{2\pi}}e^{-x^2/2}}{g(x)}g(x)dx$
$= \int_1^2 \frac{\underbrace{\frac{1}{\sqrt{2\pi}}exp-x^2/2}_{f(x)}}{\underbrace{\frac{1}{\sqrt{2\pi\nu^2}}exp\frac{-(x-1.5)^2}{2\nu^2}}_{w^*(x)}} * \underbrace{\frac{1}{\sqrt{2\pi\nu^2}}exp\frac{-(x-1.5)^2}{2\nu^2}dx}_{h(\vec{x})}$

(1) $X_i \overset{\text{iid}}{\sim} g(x)$, $N(1.5, \nu^2)$, $i = 1, \cdots, m$
(2) $\hat{\mu}_g = \frac{1}{m}\Sigma_{i=1}^n h(x_i)w^*(x_i)$


nu = 0.1
Frequency
0
5000
10000
15000
20000
0.10
0.15
0.20
y1


nu = 1
Frequency
0
5000
10000
15000
0.0
0.1
0.2
0.3
0.4
y2



For the 3rd plot, at $\nu = 10$, there begins to have some extreme values that are apparent in the histogram.

Problem 3.

(a) The functions were written in RStudio, and after testing the functions with $n = 1500$, the result is 0.693318, which is quite close to $ln(2)$, 0.6931472.

(b) Let $U \sim U(0,1)$
$c(U) = 1 + U$
$E(c(U)) = 1 + E(U) = 1.5$
The optimal value for b: $\hat{b}_n = \dfrac{\Sigma(h(U_i) - \hat{\mu}_{MC})(c(U_i) - \hat{\theta}_{MC})}{\Sigma(c(U_i) - \hat{\theta}_{MC})^2}$
where $\hat{\theta}_{MC} = \dfrac{1}{n}\Sigma_{i=1}^{n} c(U_i)$
$\hat{I}_{CV} = \dfrac{1}{n}\Sigma_{i=1}^{n} h(U_i) - b_n[\dfrac{1}{n}\Sigma_{i=1}^{n}(1 + U_i) - 1.5]$
After inputting the functions into RStudio, the result for $\hat{I}_{CV} = 0.6933005$.

(c) After determining the variances for $\hat{I}_{CV}$ and $\hat{I}_{MC}$, it was found that the values are $4.165724e-07$ for the former, and $1.304905e-05$ for the latter. Therefore, the variance of CV is smaller.

(d) Instead of using $c(X) = 1 + X$ as the control variate, can we find a new control variate such that we get a smaller variance for $\hat{I}_{CV}$? Consider the variance of the optimal control covariate estimator $\hat{I}_{CV}^{opt}$,

$$Cov(\hat{I}_{CV}^{opt}) = Cov(\hat{I}_{MC})(1 - \rho^2),$$

where ρ is the correlation coefficient between $\hat{I}_{MC}$ and $\hat{\theta}_{MC}$. To minimize the variance, we want to find a control variate $c(X)$ with a large correlation with $h(X)$. The best choice is to use $h(X)$ itself. That means we want to use $c(X) = \frac{1}{1+X}$. Then, the variance becomes 0. Although it doesn't make sense to estimate $I = h(X) = c(X)$ because we are required to know $E(c(X))$, it is "possible" to reduce the variance. For particular, we can calculate $E(c(X)) = E(h(X))$ in this example.

Problem 4.

(a) As usual ANOVA test, the null H_0 and alternative H_a hypothesis can be stated as below:

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = 0$;

H_a: at least one of the α's is not 0.

Step 1: Using our data, under null hypothesis (reduced model) we can obtain the maximum likelihood estimates $\hat{\mu}_r$ of μ and $\hat{\lambda}_r$ of λ where λ is the scale parameter of the double exponential distribution. Also, we can obtain the maximum likelihood estimates in the full model: $\hat{\mu}$ of μ and $\hat{\alpha}_i$ for $i = 1, 2, 3$ and $\hat{\lambda}$ of λ. Then we compute the difference of the log likelihood under the two models $l(\hat{\mu}, \hat{\alpha}, \hat{\lambda}|y) - l(\hat{\mu}_r, \hat{\lambda}_r|y)$ as our test statistic value.

Step 2: Simulate $\{y_{ij}\}_k$ $k = 1, \cdots, M$ (M should be large like 1000)times according to the reduced model estimated under null hypothesis in step 1. And randomly divide each sample into three groups

with the same number of observations in each group as the original data. For each k, calculate the difference of the log likelihood under the two models specified by the null and alternative hypothesis. The 0.05 significance level critical value would be the 0.95 quantile of the M differences computed.

Step 3: If the test statistic value obtained in step 1 is larger than the critical value obtained in step 2, then we conclude that we can reject the null hypothesis. Otherwise, we will accept the null hypothesis.

(b) Yes, we can use the random permutation test. The null and alternative hypothesis are the same as in part (a).

Step 1: We can estimate α_i by the least square approach to get $\hat{\alpha}_i = \bar{Y}_{\cdot\cdot} - \bar{Y}_{\cdot j}$ for $i = 1, 2, 3$. And our test statistic value is the sum of the absolute values of these estimates.

Step 2: We randomly permute the data set M times and divide the permuted data set naturally into three groups. And for each permuted data set, we compute the least square estimates $\hat{\alpha}_i^k = \bar{Y}_{\cdot\cdot}^k - \bar{Y}_{\cdot j}^k$ for $i = 1, 2, 3$ for $k = 1, \cdots, M$. For each k we also compute the sum of the absolute values of these estimates. Then the critical value is the 0.95 quantile of the M sum of absolute estimates.

Step 3: If the test statistic value is bigger than the critical value, then we reject the null hypothesis. Otherwise we accept the null hypothesis.

Problem 5.

(a) Generate 100 points from $R \sim B(1, 0.3)$, $r_1, \ldots, r_{100}$, and $Y \sim Poi(2)$, $y_1, \ldots, y_{100}$, respectively. Then set $X = RY$, we get the 100 points, $x_i = r_i y_i$ for $i = 1, \ldots 100$, generated from the ZIP model.

(b)(i) $f(\lambda|p, \vec{r}, \vec{x}) = \dfrac{f(\vec{x}, \vec{r}, \lambda, p)}{f(\vec{x}, \vec{r}, p)}$

$= \dfrac{f(\vec{x}, \vec{r}, \lambda, p)}{\int f(\vec{x}, \vec{r}, p, \lambda) d\lambda}$

$= \dfrac{f(\vec{x}|\vec{r}, \lambda, p) f(\vec{r}|\lambda, p) f(\lambda|p) f(p)}{\int f(\vec{x}, \vec{r}, p, \lambda) d\lambda}$

Consider $\int_0^\infty f(\vec{x}, \vec{r}, p, \lambda) d\lambda$

$$= \int_0^\infty \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \frac{e^{-\lambda \Sigma r_i} \lambda^{\Sigma x_i} (\Pi_{i=1}^n {r_i}^{x_i})}{(\Pi_{i=1}^n x_i!)} p^{\Sigma r_i} (1-p)^{n-\Sigma r_i} d\lambda$$

$$= \frac{b^a (\Pi r_i)^{x_i} p^{\Sigma r_i} (1-p)^{n-\Sigma r_i}}{\Gamma(a)(\Pi x_i!)} \int_0^\infty \lambda^{a+\Sigma x_i - 1} e^{-\lambda(b+\Sigma r_i)} d\lambda$$

$$= \frac{b^a (\Pi r_i)^{x_i} p^{\Sigma r_i} (1-p)^{n-\Sigma r_i}}{\Gamma(a)(\Pi x_i!)} * \frac{\Gamma(a+\Sigma x_i)}{(b+\Sigma r_i)^{a+\Sigma x_i}} * \underbrace{\int_0^\infty \frac{(b+\Sigma r_i)^{a+\Sigma x_i}}{\Gamma(a+\Sigma x_i)} \lambda^{a+\Sigma x_i - 1} e^{-\lambda(b+\Sigma r_i)} d\lambda}_{=1}$$

$$= \frac{\Gamma(a+\Sigma x_i)}{\Gamma(a)} \frac{b^a}{(b+\Sigma r_i)^{a+\Sigma r_i}} (\frac{\Pi r_i^{x_i}}{\Pi x_i!}) p^{\Sigma r_i} (1-p)^{n-\Sigma r_i}$$

So, $f(\lambda | p, \vec{r}, \vec{x})$

$$= \frac{f(\vec{x}, \vec{r}, \lambda, p)}{f(\vec{x}, \vec{r}, p)}$$

$$= \frac{\frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \frac{e^{-\lambda \Sigma r_i} \lambda^{\Sigma x_i} (\Pi_{i=1}^n {r_i}^{x_i})}{(\Pi_{i=1}^n x_i!)} p^{\Sigma r_i} (1-p)^{n-\Sigma r_i}}{\frac{\Gamma(a+\Sigma x_i)}{\Gamma(a)} \frac{b^a}{(b+\Sigma r_i)^{a+\Sigma r_i}} (\frac{\Pi r_i^{x_i}}{\Pi x_i!}) p^{\Sigma r_i} (1-p)^{n-\Sigma r_i}}$$

$$= \frac{(b+\Sigma r_i)^{a+\Sigma r_i}}{\Gamma(a+\Sigma x_i)} \lambda^{a+\Sigma x_i - 1} e^{-\lambda(b+\Sigma r_i)}$$

So, $\lambda | p, \vec{r}, \vec{x} \sim Gamma(a+\Sigma x_i, b+\Sigma r_i)$

(b)(ii) $f(p | \lambda, \vec{r}, \vec{x})$

$$= \frac{f(\vec{x}, \vec{r}, \lambda, p)}{f(\vec{x}, \vec{r}, \lambda)}$$

$$= \frac{f(\vec{x}, \vec{r}, \lambda, p)}{\int_0^1 f(\vec{x}, \vec{r}, \lambda, p) dp}$$

Consider:

$\int_0^1 f(\vec{x}, \vec{r}, \lambda, p) dp$

$$= \int_0^1 \frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \frac{e^{-\lambda \Sigma r_i} \lambda^{\Sigma x_i} (\Pi {r_i}^{x_i})}{\Pi x_i!} p^{\Sigma r_i} (1-p)^{n-\Sigma r_i} dp$$

$$= \frac{b^a \lambda^{a+\Sigma x_i - 1} e^{-\lambda(b+\Sigma r_i)} (\Pi {r_i}^{x_i})}{\Gamma(a)(\Pi x_i!)} \int_0^1 p^{\Sigma r_i} (1-p)^{n-\Sigma r_i} dp$$

$$= \frac{b^a \lambda^{a+\Sigma x_i - 1} e^{-\lambda(b+\Sigma r_i)} (\Pi {r_i}^{x_i})}{\Gamma(a)(\Pi x_i!)} * \frac{\Gamma(\Sigma r_i + 1)\Gamma(n - \Sigma r_i + 1)}{\Gamma((\Sigma r_i + 1) + (n - \Sigma r_i + 1))} *$$

$$\underbrace{\int_0^1 \frac{\Gamma((\Sigma r_i + 1) + (n - \Sigma r_i + 1))}{\Gamma(\Sigma r_i + 1)\Gamma(n - \Sigma r_i + 1)} * p^{(\Sigma r_i + 1) - 1} (1-p)^{(n - \Sigma r_i + 1) - 1}}_{=1}$$

(b)(iii) $f(r_i|\lambda, p, x_i) = \dfrac{f(x_i|r_i, \lambda, p)f(r_i|\lambda, p)}{\Sigma_{r=0}^{1} f(x_i|r_i, \lambda, p)f(r_i|\lambda, p)}$

$$= \frac{\dfrac{e^{-\lambda r_i}(\lambda r_i)^{x_i}}{x_i!} p^{r_i}(1-p)^{1-r_i}}{\Sigma_{r=0}^{1}[\dfrac{e^{-\lambda r}(\lambda r)^{x}}{x!} p^{r}(1-p)^{1-r}]}$$

$$= \frac{e^{-\lambda r} p^{r}(1-p)^{1-r}}{e^{-\lambda}p + (1-p)I_{(x=0)}}$$

$$= \frac{(pe^{-\lambda})^{r}(1-p)^{1-r}}{pe^{-\lambda} + (1-p)I_{(x=0)}}$$

$$= (\frac{pe^{-\lambda}}{pe^{-\lambda} + (1-p)I_{(x=0)}})^{r} (\frac{1-p}{pe^{-\lambda} + (1-p)I_{(x=0)}})^{1-r}$$

Which follows the distribution similar to $p^r(1-p)^{1-r}$

(c) Here we utilize the conditional densities derived in part (b) and the data generated in part(a) to generate a sample of $(\lambda, p, \boldsymbol{r_i})$ of size 20000 using the Gibbs-Sampling method. And the first 10000 runs are used as burn-in for the approximately convergence to be achieved. Then from the generated data, we can compute the sample mean and construct Empirical 95% Bayesian' confidence intervals for λ and p. The result using different a, b are given in Table 1.

From Table 1, we see that when $a = 1, b = 1$ or $a = 2, b = 2$, our estimates for λ are pretty close to the true value 2. And this can be verified when we look at the plot of density functions of $Gamma(a, b)$ distribution for different a and b. We see that when the density function is concentrated around 2, then the estimates for λ is good. However the estimate for p depends largely on the data X we generated in part (a). Since our data is only of size 100 in part (a), the estimate for p is kind of arbitrary.

Problem 6.

Using the Metropolis-Hastings Algorithm described in the class, we generate a sample of size 1000 using $Gamma(k, \theta)$ as our proposal for different k and θ. Notice that the proposal density is not depending on the last draw. Only the acceptance probability is depending on the last draw. Therefore this is referred to as the Independence-

Metropolis-Hastings Algorithm.

To check how good are we doing, we calculate the sample mean of Z and $\frac{1}{Z}$ and compare them with the true values: $E(Z) = \sqrt{\theta_2/\theta_1} =$ **1.154701** and $E(1/Z) = \sqrt{\theta_1/\theta_2} + \frac{1}{2\theta_2} =$ **1.116025**.

From Table 2, we can see that the estimated expected values are closest to the true value when $k = 2$ and $\theta = 1$. This is because in this case, the proposal density is very similar to the density we want to sample from.

When $k = 2$ and $\theta = 1$, the procedure has the highest acceptance rate and also walks at a reasonable size of steps. The second best one is when $k = 2$ and $\theta = 2$.

Appendix: Tables and Figures

Table 1: Estimates of λ and p using different a and b.

		a=1,b=1	a=2,b=2	a=10,b=1	a=0.5,b=10
λ	estimate	2.083765	2.061086	2.407426	1.570616
	95% lower limit	1.592924	1.586919	1.865521	1.195691
	95% upper limit	2.639425	2.598692	2.99697	1.990016
p	estimate	0.3823204	0.3810588	0.3672957	0.4164825
	95% lower limit	0.2804221	0.281569	0.2707288	0.3082988
	95% upper limit	0.4902653	0.4926738	0.4721491	0.5349219

Table 2: Estimates of $E(Z)$ and $E(1/Z)$ using different k and θ.

	E(Z)	E(1/Z)
True value	1.154701	1.116025
$k = 2, \theta = 1$	1.173821	1.115827
$k = 1, \theta = 10$	1.089192	1.16363
$k = 10, \theta = 0.5$	1.837936	0.5617463
$k = 2, \theta = 2$	1.218463	1.038873