STAT 457 - FINAL

Martha Eichlersmith

2019-12-12

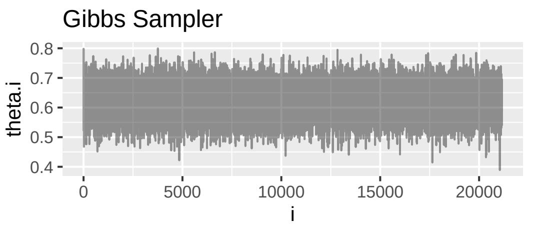
Problem 1

Problem 1a

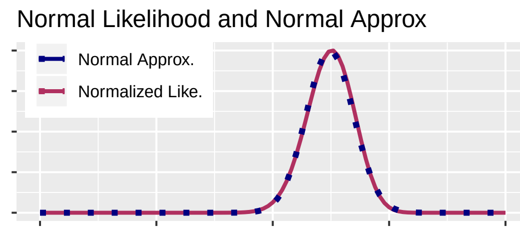
For the data $Y = (125, 18, 20, 34)$, implement the Gibbs sampler algorithm. Use a flat prior on θ.
Plot θ^i versus iteration i. $Y = (y_1, y_2, y_3, y_4) \propto (2 + \theta, 1 - \theta, 1 - \theta, \theta)$

1. Draw a starting value, $t \sim$Uniform(0,1)
2. Draw a latent value, $Z \sim \text{Binomial}\left(y_1,\ \frac{\theta}{2+\theta}\right)$
3. Draw a parameter, $\theta \sim \text{Beta}(Z + y_4 + 1, y_2 + y_3 + 1)$

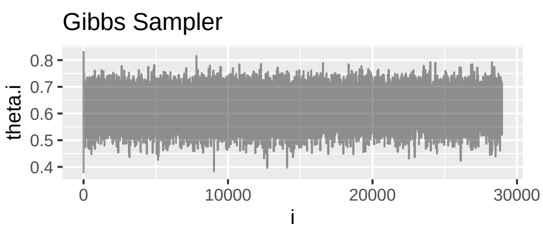
Chain 1 for data for Y=(125,18,20,34)



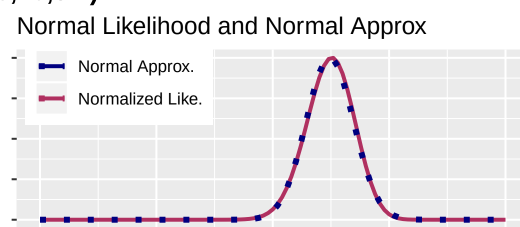
	Name	Value
1	Mean	0.62270
2	SD	0.05086
3	it	21156
4	Start	0.59298



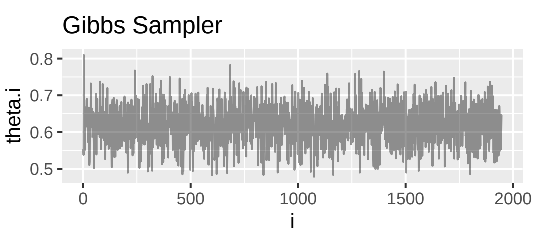
Chain 2 for data for Y=(125,18,20,34)



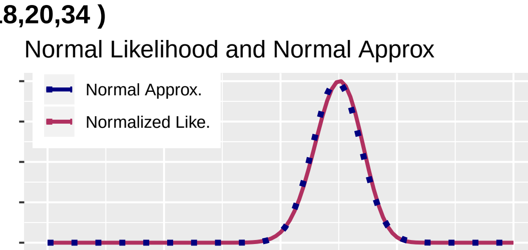
	Name	Value
1	Mean	0.62266
2	SD	0.05097
3	it	28962
4	Start	0.37667



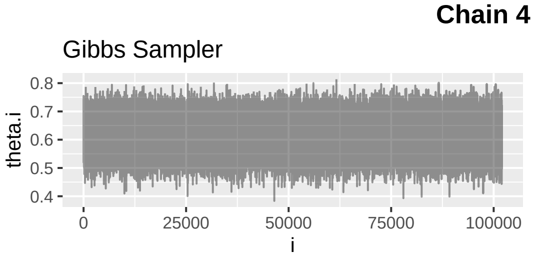
Chain 3 for data for Y=(125,18,20,34)



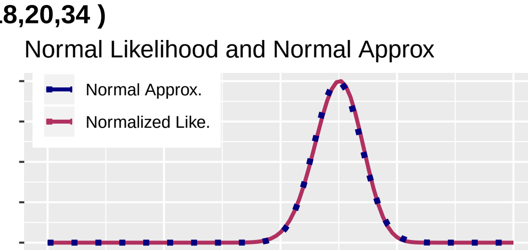
	Name	Value
1	Mean	0.62083
2	SD	0.05121
3	it	1947
4	Start	0.55304



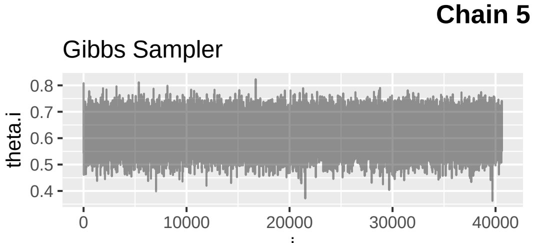
Chain 4 for data for Y=(125,18,20,34)



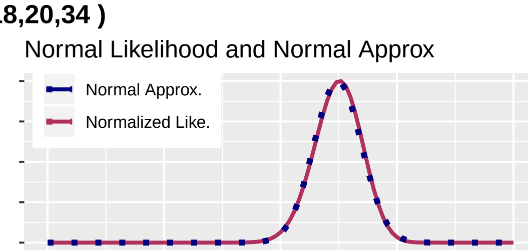
	Name	Value
1	Mean	0.62306
2	SD	0.05108
3	it	102018
4	Start	0.56521



Chain 5 for data for Y=(125,18,20,34)



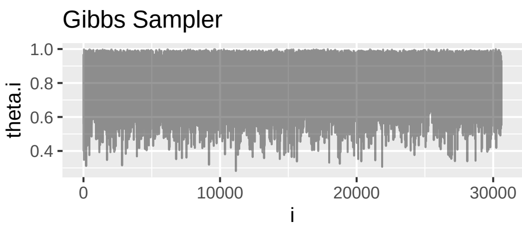
	Name	Value
1	Mean	0.62276
2	SD	0.05112
3	it	40623
4	Start	0.73643



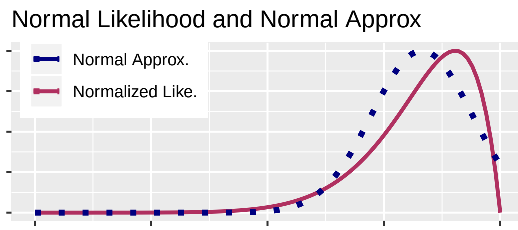
Problem 1b

Repeat 1a for $Y = (14, 0, 1, 5)$.

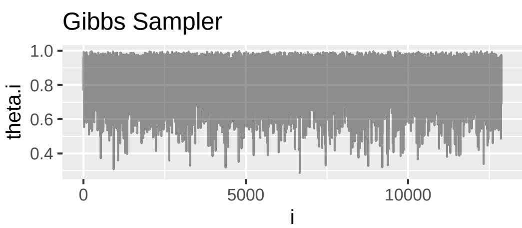
Chain 1 for data for Y=(14,0,1,5)



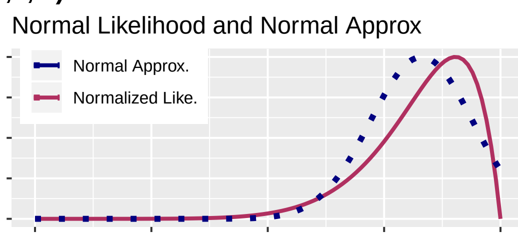
	Name	Value
1	Mean	0.83158
2	SD	0.10790
3	it	30607
4	Start	0.39922



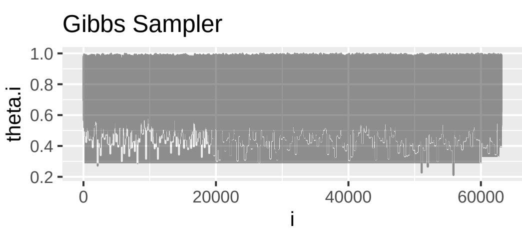
Chain 2 for data for Y=(14,0,1,5)



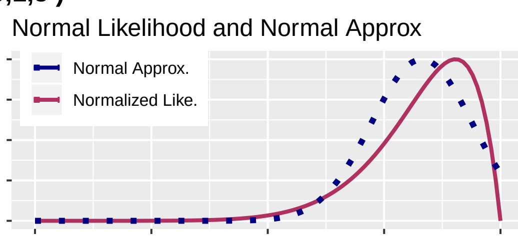
	Name	Value
1	Mean	0.83034
2	SD	0.10933
3	it	12875
4	Start	0.90502



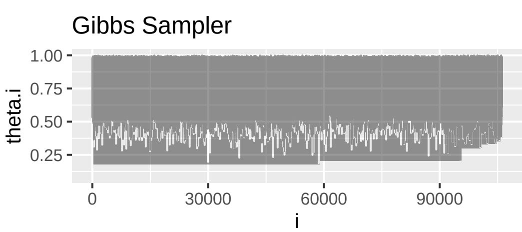
Chain 3 for data for Y=(14,0,1,5)



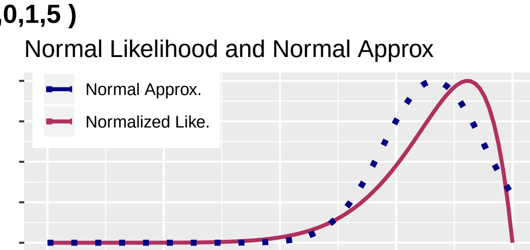
	Name	Value
1	Mean	0.83093
2	SD	0.10811
3	it	63107
4	Start	0.28758



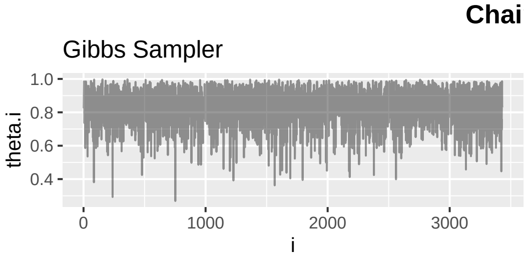
Chain 4 for data for Y=(14,0,1,5)



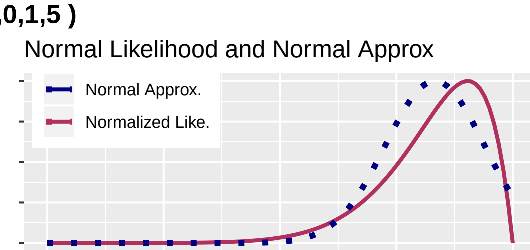
	Name	Value
1	Mean	0.83093
2	SD	0.10790
3	it	106106
4	Start	0.08302



Chain 5 for data for Y=(14,0,1,5)



	Name	Value
1	Mean	0.83272
2	SD	0.10589
3	it	3432
4	Start	0.82467



There is a lack of fit for the data in 1b, where the fit appears to be better for data in 1a. Convergence was assessed when values were had a difference less than 10^{-7}.

Problem 1c

20 Chains for Y=(125,18,20,34)

Mean	Standard Deviation	Standard Error	Iterations
0.62265	0.05147	0.00042	15243
0.62279	0.05103	0.00013	161378
0.62191	0.05203	0.00066	6229
0.62261	0.05126	0.00024	44139
0.62286	0.05113	0.00020	64679
0.62250	0.05089	0.00026	36893
0.62237	0.05087	0.00039	16865
0.62240	0.05085	0.00054	8893
0.62274	0.05114	0.00015	111670
0.62294	0.05100	0.00018	80166
0.62302	0.05108	0.00025	42245
0.62287	0.05109	0.00030	28917
0.62262	0.05099	0.00016	95966
0.62376	0.05110	0.00030	29344
0.62261	0.05111	0.00021	57875
0.62272	0.05060	0.00054	8760
0.62249	0.05108	0.00029	31756
0.62281	0.05068	0.00026	37926
0.62325	0.05090	0.00021	57245
0.62295	0.05102	0.00022	54595

```
## [1] "0.000371 is the standard deviation of the 20 averages of theta"
```

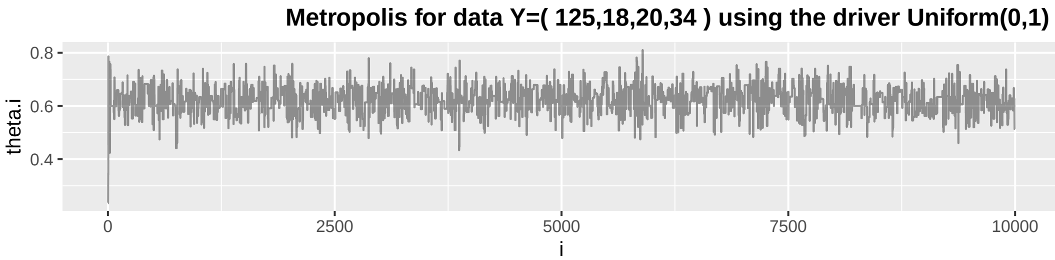
```
## [1] "0.000296 is the average of the standard errors of the 20 chains of theta"
```

You would expected these values to be similar but it appears that our standard error average is under-estimating the variation in the chain means of θ in this case.

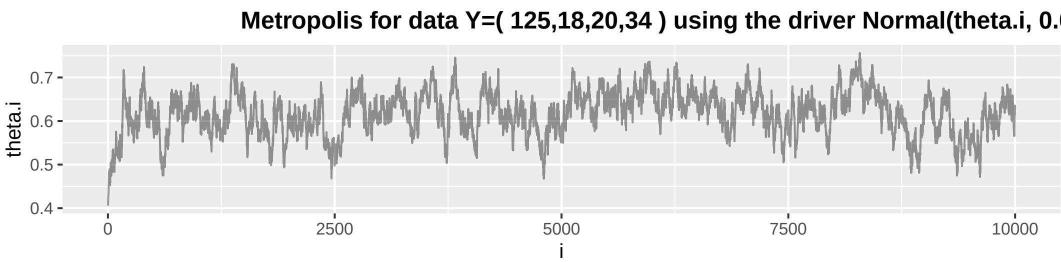
Problem 2

Problem 2a

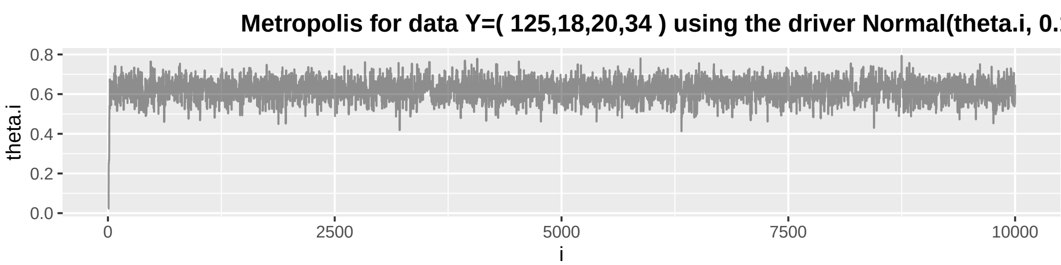
For the genetic linkage model applied to $Y = (125, 18, 20, 34)$, implement the Metropolis algorithm. (use a flat prior on θ). Use one long chain and plot θ^i versus i.
Try several driver functions:



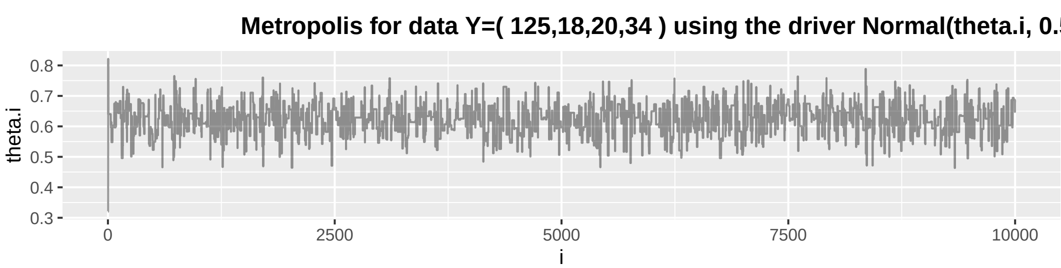
	Name	Value
1	Mean	0.62281
2	SD	0.05181
3	it	10000
4	Start	0.23480



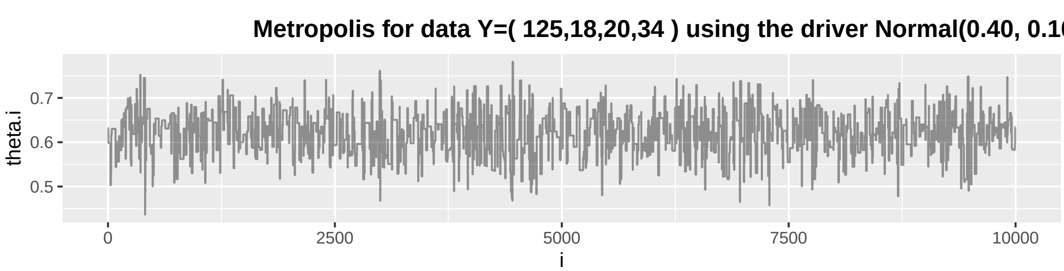
	Name	Value
1	Mean	0.61705
2	SD	0.04995
3	it	10000
4	Start	0.40549



	Name	Value
1	Mean	0.62133
2	SD	0.05491
3	it	10000
4	Start	0.02234



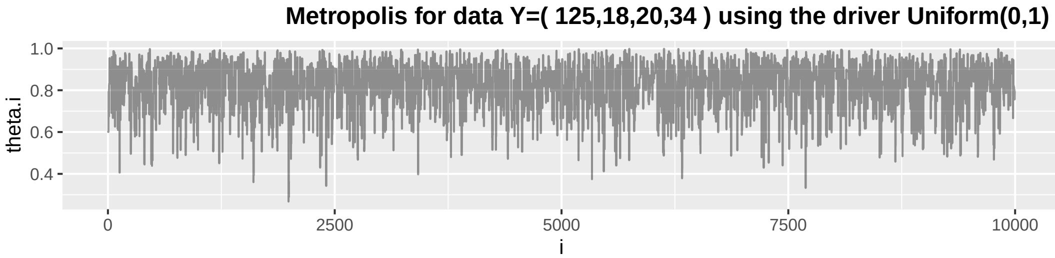
	Name	Value
1	Mean	0.62279
2	SD	0.05158
3	it	10000
4	Start	0.31998



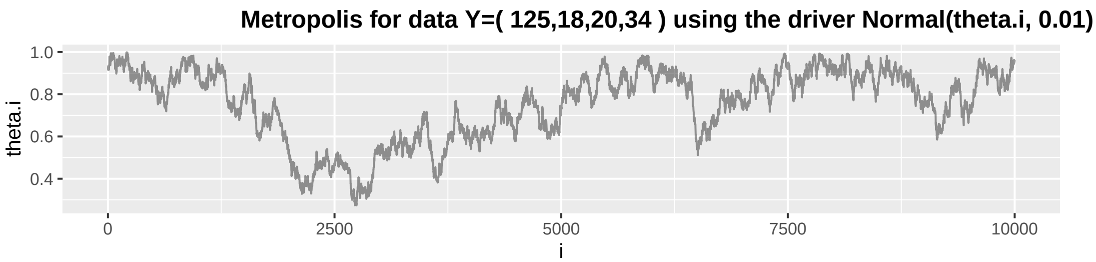
	Name	Value
1	Mean	0.62122
2	SD	0.04950
3	it	10000
4	Start	0.63208

Problem 2b

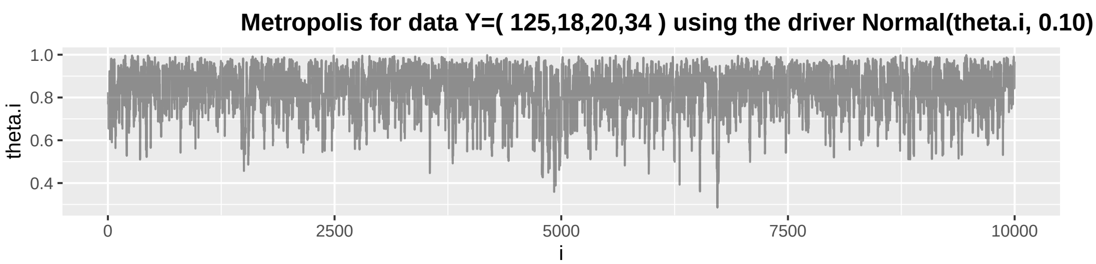
Repeat 2a for $Y = (14, 0, 1, 5)$



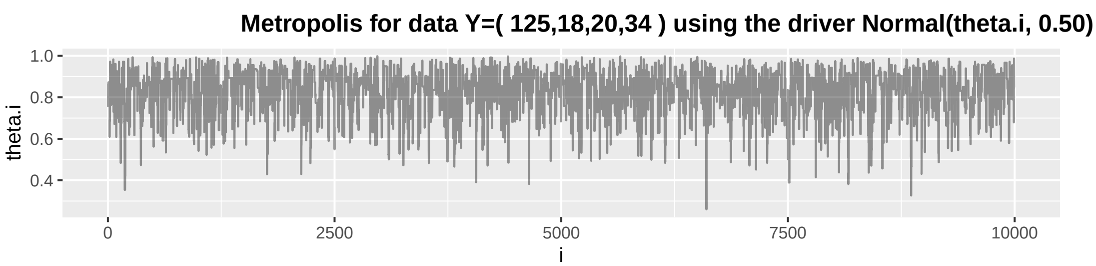
	Name	Value
1	Mean	0.82968
2	SD	0.10753
3	it	10000
4	Start	0.60549



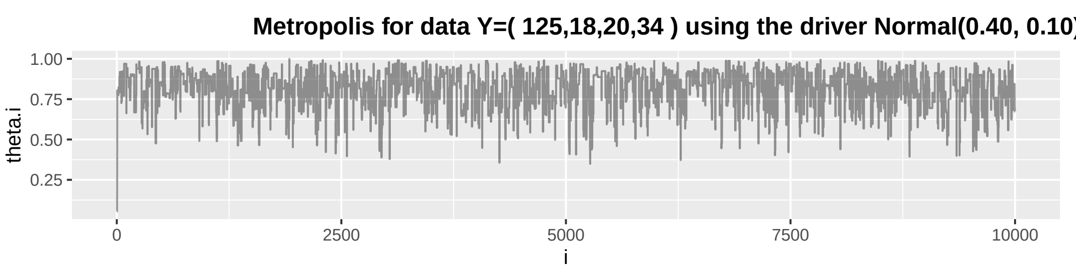
	Name	Value
1	Mean	0.75321
2	SD	0.16762
3	it	10000
4	Start	0.93159



	Name	Value
1	Mean	0.83338
2	SD	0.10637
3	it	10000
4	Start	0.82435



	Name	Value
1	Mean	0.82739
2	SD	0.10997
3	it	10000
4	Start	0.75380



	Name	Value
1	Mean	0.81881
2	SD	0.11488
3	it	10000
4	Start	0.05359

Problem 2c

Compute both the posterior mean and standard deviation for both data sets.
Compare to results from the previous problem.

In problem 1 the means were similar, but in problem 2 the means vary depending on the driver. Some of the means in problem 2 are close to the means in problem 1.

Problem 2d

20 Chains for Y=(125,18,20,34) with Different Drivers

Uniform(0, 1)			Normal(theta.i, 0.01)			Normal(theta.i, 0.10)			Normal(theta.i, 0.50			Normal(0.40, 0.10)		
Mean	SD	SE	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
0.6230	0.0525	5e-04	0.6194	0.0579	6e-04	0.6222	0.0511	5e-04	0.6223	0.0496	5e-04	0.6193	0.0513	5e-04
0.6232	0.0510	5e-04	0.6117	0.0534	5e-04	0.6218	0.0514	5e-04	0.6225	0.0502	5e-04	0.6200	0.0492	5e-04
0.6193	0.0513	5e-04	0.6171	0.0655	7e-04	0.6214	0.0530	5e-04	0.6233	0.0501	5e-04	0.6198	0.0516	5e-04
0.6232	0.0533	5e-04	0.6196	0.0456	5e-04	0.6224	0.0501	5e-04	0.6239	0.0517	5e-04	0.6195	0.0530	5e-04
0.6232	0.0495	5e-04	0.6016	0.0610	6e-04	0.6220	0.0504	5e-04	0.6206	0.0522	5e-04	0.6222	0.0504	5e-04
0.6225	0.0492	5e-04	0.6268	0.0510	5e-04	0.6214	0.0530	5e-04	0.6239	0.0497	5e-04	0.6195	0.0488	5e-04
0.6247	0.0525	5e-04	0.6264	0.0527	5e-04	0.6223	0.0528	5e-04	0.6217	0.0544	5e-04	0.6165	0.0562	6e-04
0.6243	0.0510	5e-04	0.6270	0.0483	5e-04	0.6247	0.0520	5e-04	0.6211	0.0524	5e-04	0.6187	0.0518	5e-04
0.6225	0.0517	5e-04	0.6162	0.0541	5e-04	0.6225	0.0524	5e-04	0.6233	0.0523	5e-04	0.6180	0.0524	5e-04
0.6229	0.0506	5e-04	0.6267	0.0539	5e-04	0.6230	0.0507	5e-04	0.6227	0.0500	5e-04	0.6214	0.0513	5e-04
0.6236	0.0503	5e-04	0.6215	0.0545	5e-04	0.6230	0.0520	5e-04	0.6212	0.0524	5e-04	0.6196	0.0536	5e-04
0.6255	0.0529	5e-04	0.6134	0.0672	7e-04	0.6224	0.0548	5e-04	0.6219	0.0534	5e-04	0.6203	0.0507	5e-04
0.6230	0.0504	5e-04	0.6321	0.0502	5e-04	0.6231	0.0513	5e-04	0.6229	0.0517	5e-04	0.6191	0.0508	5e-04
0.6218	0.0486	5e-04	0.6135	0.0536	5e-04	0.6207	0.0515	5e-04	0.6262	0.0520	5e-04	0.6203	0.0527	5e-04
0.6233	0.0536	5e-04	0.6216	0.0528	5e-04	0.6239	0.0506	5e-04	0.6215	0.0528	5e-04	0.6185	0.0519	5e-04
0.6227	0.0516	5e-04	0.6378	0.0449	4e-04	0.6241	0.0510	5e-04	0.6207	0.0516	5e-04	0.6212	0.0496	5e-04
0.6246	0.0530	5e-04	0.6280	0.0610	6e-04	0.6232	0.0512	5e-04	0.6231	0.0535	5e-04	0.6187	0.0508	5e-04
0.6209	0.0480	5e-04	0.6175	0.0509	5e-04	0.6223	0.0505	5e-04	0.6282	0.0504	5e-04	0.6187	0.0535	5e-04
0.6211	0.0518	5e-04	0.6085	0.0579	6e-04	0.6213	0.0524	5e-04	0.6231	0.0521	5e-04	0.6246	0.0469	5e-04
0.6219	0.0489	5e-04	0.6242	0.0502	5e-04	0.6215	0.0514	5e-04	0.6215	0.0534	5e-04	0.6193	0.0488	5e-04

SD of Average theta's versus Average SE

	sd.of.avgs	avg.of.se
Uniform(0,1)	0.0014301	0.0005108
Normal(theta.i, 0.01)	0.0085221	0.0005433
Normal(theta.i, 0.10)	0.0010238	0.0005169
Normal(theta.i, 0.50)	0.0018357	0.0005179
Normal(0.40, 0.10)	0.0016973	0.0005127

Similar to problem 1, it appears that our estimation of the variation in the mean of theta is lower than the actual variation between the means.

Problem 3

Problem 3a

Consider the 1-way variance components model

$$Y_{ij} = \theta_i + \epsilon_{ij}$$

where Y_{ij} is the jth observation from the ith group, θ_i is the effect, ϵ_{ij}=error, $i = 1, ..., K$ and $j = 1, ..., J$. It is assume that $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and $\theta_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_\theta^2)$. Under the prior specification $p(\sigma_\epsilon^2, \sigma_\theta^2, \mu) = p(\sigma_\epsilon^2)p(\sigma_\theta^2)p(\mu)$, with $p(\sigma_\theta^2) = \text{InverseGamma}(a_1, b_1)$, $p(\sigma_\epsilon^2) = \text{InverseGamma}(a_2, b_2)$, and $p(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$.
Let $\overline{Y}_i = \frac{1}{J}\sum_{j=1}^J Y_{ij}$ and $\theta = (theta_1, \cdots, \theta_k)$.
Show the following:

$$\begin{aligned}
p(\mu, \sigma_\epsilon^2, \sigma_\theta^2) &= p(\mu)p(\sigma_\theta^2)p(\sigma_\epsilon^2) \\
&= \mathcal{N}(\mu_0, \sigma_0^2)\text{IG}(a_1, b_1)\text{IG}(a_2, b_2) \\
&= \left[\frac{1}{\sqrt{2\pi}\ \sigma_0} \exp\left\{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right\}\right] \cdot \left[(\sigma_\theta^2)^{a_1-1}\exp\left\{-b_1\sigma_\theta^2\right\}\right]^{-1} \cdot \left[(\sigma_\epsilon^2)^{a_2-1}\exp\left\{-b_2\sigma_\epsilon^2\right\}\right]^{-1} \\
&\propto {\sigma_\theta}^{-2(a_1-1)} \cdot {\sigma_\epsilon}^{-2(a_2-1)} \cdot \exp\left\{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2} + b_1\sigma_\theta^2 + b_2\sigma_\epsilon^2\right\}
\end{aligned}$$

$$\begin{aligned}
p(\mu, \theta, \sigma_\theta^2, \sigma_\epsilon^2 \mid Y) &\propto p(Y \mid \mu, \theta, \sigma_\theta^2, \sigma_\epsilon^2) \cdot p(\theta \mid \mu, \sigma_\theta^2, \sigma_\epsilon^2) \cdot p(\mu, \sigma_\theta^2, \sigma_\epsilon^2) \\
&= \left[\prod_{i=1}^K\prod_{j=1}^J \left(p(y_{ij}) \sim \mathcal{N}(\theta_i, \sigma_\epsilon^2)\right)\right] \cdot \left[\prod_{i=1}^K \left(p(\theta_i) \sim \mathcal{N}(\mu, \sigma_\theta^2)\right)\right] \cdot \prod_{i=1}^K p(\mu, \sigma_\theta^2, \sigma_\epsilon^2) \\
&= \left[\sigma_\epsilon^{-(KJ)} \cdot \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K\sum_{j=1}^J(y_{ij}-\theta_i)^2}{\sigma_\epsilon^2}\right\}\right] \\
&\quad \cdot \left[\sigma_\theta^{-K}\exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K(\theta_i-\mu)^2}{\sigma_\theta^2}\right\}\right] \\
&\quad \cdot \left[{\sigma_\theta}^{-2K(a_1-1)} \cdot \sigma_\epsilon^{-2K(a_2-1)} \cdot \exp\left\{K\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2} + b_1\sigma_\theta^2 + b_2\sigma_\epsilon^2\right)\right\}\right]
\end{aligned}$$

Problem 3a(1)

$$p(\mu \mid \theta, \sigma_\epsilon^2, \sigma_\theta^2, Y) \propto = \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K (\theta_i - \mu)^2}{\sigma_\theta^2}\right\} \cdot \exp\left\{K\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right)\right\}$$
$$= \mathcal{N}\left(\frac{\sigma_\theta^2 \mu_0 + \sigma_0^2 \sum \theta_i}{\sigma_\theta^2 + K\sigma_0^2},\ \frac{\sigma_\theta^2 \sigma_0^2}{\sigma_\theta^2 + K\sigma_0^2}\right)$$

Problem 3a(2)

$$p(\theta_i \mid \mu, \sigma_\epsilon^2, \sigma_\theta^2, Y) \propto \cdot \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K \sum_{j=1}^J (y_{ij} - \theta_i)^2}{\sigma_\epsilon^2}\right\} \cdot \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K (\theta_i - \mu)^2}{\sigma_\theta^2}\right\}$$
$$= \mathcal{N}\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \overline{Y}_i + \frac{\sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \mu, \quad \frac{\sigma_\theta^2 \sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2}\right)$$

Problem 3a(3)

$$p(\sigma_\epsilon^2 \mid \mu, \theta, \sigma_\theta^2, Y) \propto = \sigma_\epsilon^{-(KJ)} \cdot \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K \sum_{j=1}^J (y_{ij} - \theta_i)^2}{\sigma_\epsilon^2}\right\} \cdot \sigma_\epsilon^{-2K(a_2-1)} \cdot \exp\left\{K\left(b_2 \sigma_\epsilon^2\right)\right\}$$
$$= \mathrm{IG}\left(a_2 + \frac{KJ}{2},\ b_2 + \frac{1}{2}\sum_{i=1}^K \sum_{j=1}^J (Y_{ij} - \theta_i)^2\right)$$

Problem 3a(4)

$$p(\sigma_\theta^2 \mid \mu, \theta, \sigma_\epsilon^2, Y) \propto \sigma_\theta^{-K} \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^K (\theta_i - \mu)^2}{\sigma_\theta^2}\right\} \cdot \sigma_\theta^{-2K(a_1-1)} \cdot \exp\left\{K\left(b_1 \sigma_\theta^2\right)\right\}$$
$$= \mathrm{IG}\left(a_1 + \frac{K}{2},\ b_1 + \frac{1}{2}\sum_{i=1}^K (\theta_i - \mu)^2\right)$$

Problem 3b

Run the Gibbs sampler for the data below.

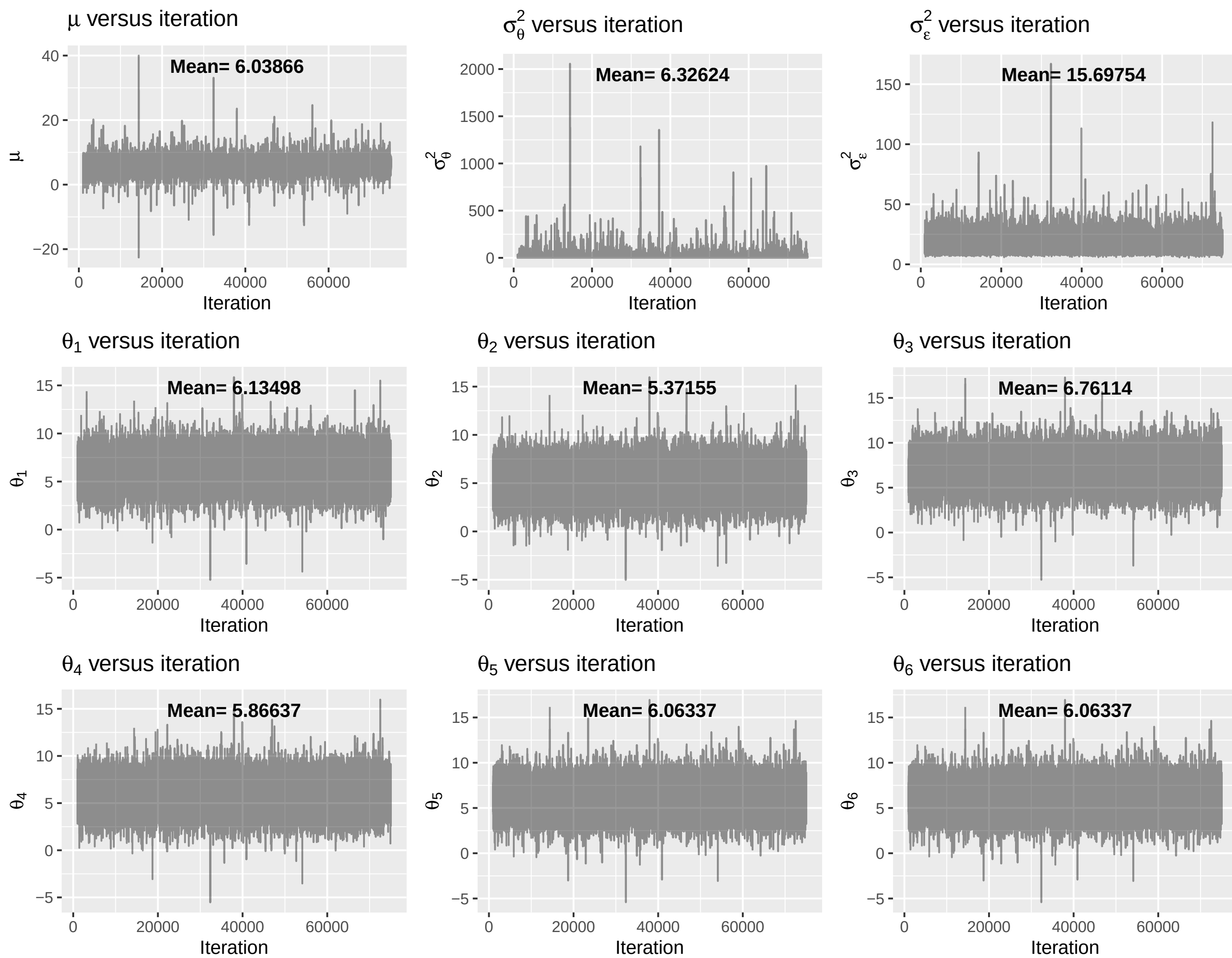
```
##      j1       j2       j3       j4       j5       j6
##      "7.298"  "5.22"   "0.11"   "2.212"  "0.282"  "1.722"
##      "3.846"  "6.556"  "10.386" "4.852"  "9.014"  "4.782"
##      "2.434"  "0.608"  "13.434" "7.092"  "4.458"  "8.106"
##      "9.566"  "11.788" "5.51"   "9.288"  "9.446"  "0.758"
##      "7.99"   "-0.892" "8.166"  "4.98"   "7.198"  "3.758"
##      " "      " "      " "      " "      " "      " "
## y_j. "6.2268" "4.656"  "7.5212" "5.6848" "6.0796" "3.8252"
```

```
## [1] "y_..= 5.6656"
```

Use one chain of length 75,000. Take $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0,0)$, and $p(\sigma_\theta^2) = IG(1,1)$. For each θ_i, for σ_ϵ, and for θ_θ, plot the simulated value at iteration j versus j. Summarize each posterior marginal.

Cannot use $a_1, b_1 = 0$, otherwise $\sigma_\theta^2 = \infty$, resulting in N/A so will use 0.01 instead.

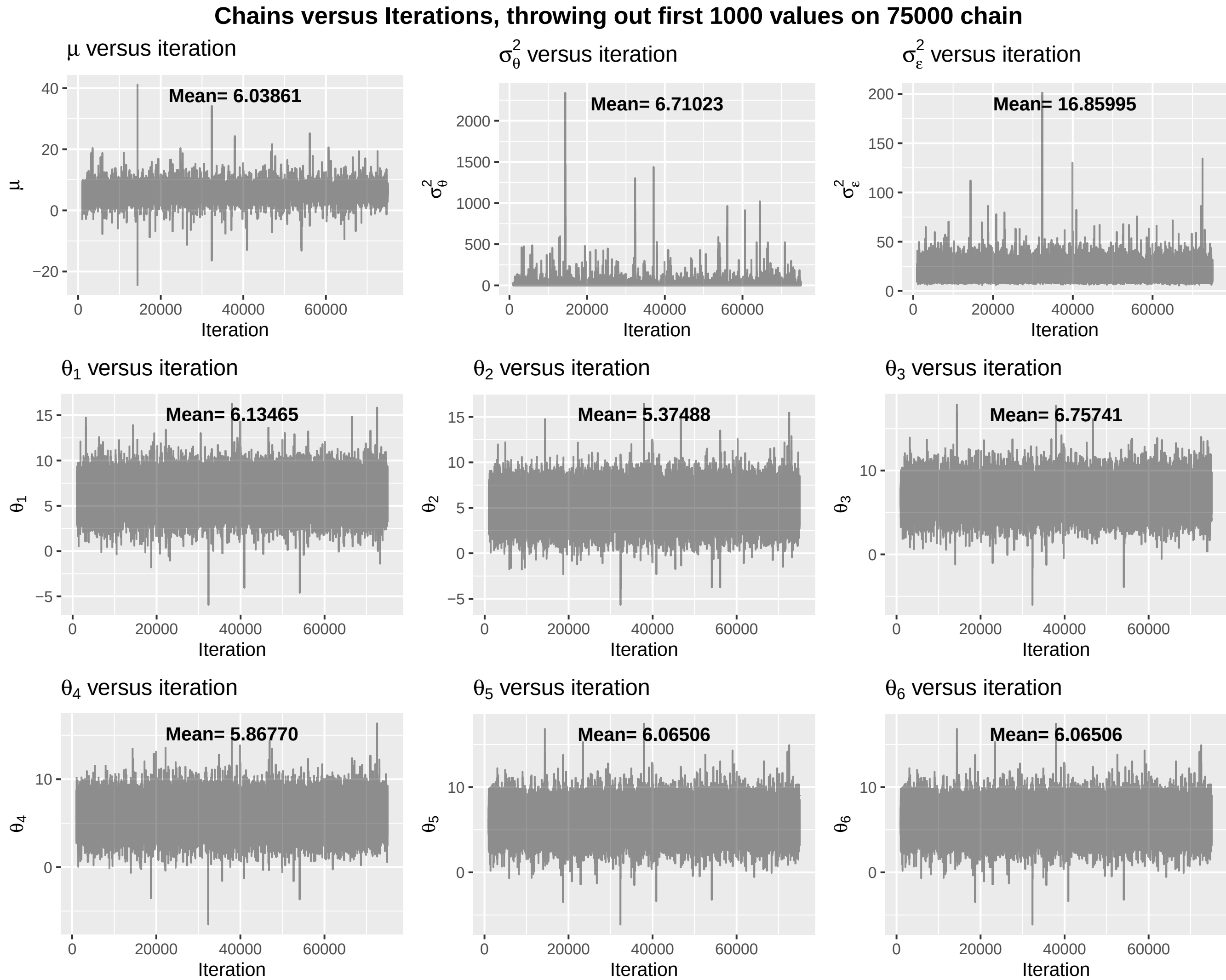
Chains versus Iterations, throwing out first 1000 values on 75000 chain



Problem 3c

Repeat 3b using the prior specification $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0,0)$, and $p(\sigma_\theta^2) = IG(0,0)$. Does this specification violate the Hobart-Casella conditions? Describe what happens to the Gibbs sampler chain in this case.

This violates the Hobart-Casella conditions. If left at zero, the result will "blow up" (i.e., the chains are full of NA). Cannot use $a_1, b_1, a_2, b_2 = 0$, otherwise $\sigma_\theta^2 = \sigma_\epsilon^2 = \infty$, resulting in N/A so will use 0.01 instead.

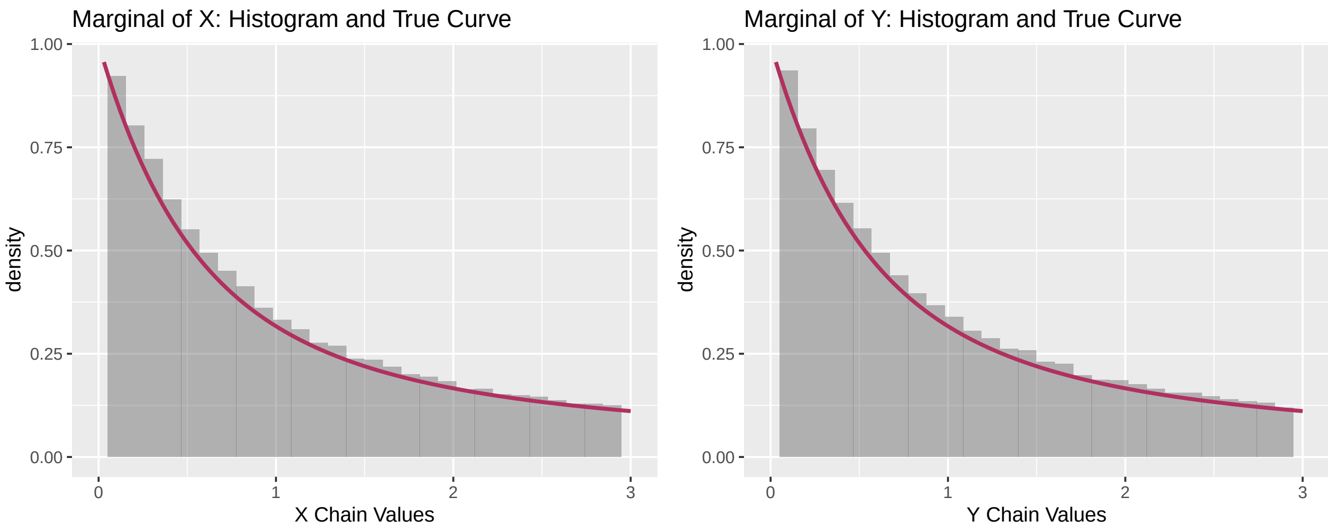


Problem 5

Suppose that X and Y have exponential conditional distributions restricted over the interval $(0, B)$, i.e. $p(x \mid y) \propto y \exp\{-yx\}$ for $0 < x < B < \infty$ and $p(y \mid x) \propto x \exp\{-xy\}$ for $0 < y < B < \infty$, where B is known constant.

Problem 5a

Take $m = 1$ and $B = 3$. Run the data augmentation algorithm using these conditionals. (Hist: Reject the exponential deviates that lie outside $(0, B)$). Obtain the marginal for x using the mixture of conditionals $p(x \mid y)$, mixed over the simulated y deviates in your chain.



Problem 5b

Show that the marginal for x is proportional to $(1 - \exp\{-Bx\})/x$. Compare your results in 5a to this curve.

$$
\begin{aligned}
p_{X|Y}(x \mid y) &= y e^{-yx} \\
p_X(x) &\propto \int_0^B e^{-yx} dy \\
&= \frac{1}{x} \int_0^B x e^{-xy} dy \\
&= \frac{1}{x} \Big[-e^{-xy} \Big]_{y=0}^{y=B} \\
&= \frac{1}{x} \Big(\left[-e^{-Bx} \right] - \left[-e^{-0 \cdot x} \right] \Big) \\
&= \frac{1}{x} \Big(-e^{-Bx} + 1 \Big) \\
&= \frac{1 - e^{-Bx}}{x} \\
p_X(x) &\propto \frac{1 - e^{-Bx}}{x}
\end{aligned}
$$

See above for the histogram of the marginals with the curves. Note the curves are normalized by dividing by B so that both the histogram density and the curve are on scale of 0 to 1. The curve matches the histogram very well.

Problem 5c

Repeat 5a and 5b using $B = \infty$. Describe what happens. Is the marginal for x a proper density in this case?

$$\lim_{B\to\infty} \frac{1 - e^{-Bx}}{x} = \frac{1 - \lim_{B\to\infty} e^{-Bx}}{x} = \frac{1}{x}$$

This is not a proper marginal.

$$\int_0^B \frac{1}{x} dx = \ln(x)\Big|_0^B = \ln(B) - \ln(0) = \ln(B) + \infty = \infty$$