Stat 544 Exam 2

May 5, 2008

I have neither given nor received unauthorized assistance on this examination.

____
signature

____
name printed

There are 10parts on this exam. I will score every part out of 10 points.

1. Suppose that $\mathbf{Y}=(Y_{11},Y_{12},Y_{21},Y_{22})$ can be modeled as $\text{Multinomial}_4\left(10,\left(p_{11},p_{12},p_{21},p_{22}\right)\right)$. That is, think of $\mathbf{Y}$ as generated by making $n=10$ independent trials placing an outcome into a 2×2 table (using probabilities as indicated below) and counting up the numbers of outcomes landing in the 4 different cells to make the entries of $\mathbf{Y}$.

p_{11}	p_{12}
p_{21}	p_{22}

Suppose further that one adopts a $\text{Dirichlet}_4(1,1,1,1)$ prior for the vector $\mathbf{p}=\left(p_{11},p_{12},p_{21},p_{22}\right)$ and subsequently observes $\mathbf{Y}=(2,2,2,4)$.

a) What is the posterior distribution of $\mathbf{p}$? What is the posterior conditional of $p_{12}\,|\left(p_{21}=.3 \text{ and } p_{22}=.4\right)$? (In this latter case, an appropriate multiple of p_{12} has some standard distribution. What multiple and what standard distribution?)

b) The quantity $\lambda=\ln\left(p_{21}\right)-\ln\left(p_{11}\right)-\ln\left(p_{22}\right)+\ln\left(p_{12}\right)$ is a measure of dependence or "interaction" in the two-way table of probabilities. Carefully describe how you could simulate a value from the posterior distribution of λ using only iid $\text{Exp}(1)$ random variables.

2. Suppose that $Y_1 \sim \mathrm{N}(0,1)$, that for $i \geq 2$ $Y_i \mid Y_{i-1} \sim \mathrm{N}(Y_{i-1},1)$, and that the DAG below represents the joint distribution of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$.

$$Y_1 \longrightarrow Y_2 \longrightarrow Y_3 \longrightarrow Y_4$$

What is the conditional distribution of $Y_2 \mid (Y_1 = 3, Y_3 = 4, Y_4 = 0)$? (Carefully say why your answer is correct.)

3. Let $\gamma(x) = \exp(-x^2)$. Random variables $Y_1, Y_2, Y_3,$ and Y_4 have joint pdf f such that

$$f(\mathbf{y}) \propto \gamma(y_1)\gamma(y_2)\gamma(y_3)\gamma(y_4)\gamma(y_1 - y_2)\gamma(y_2 - y_3)\gamma(y_3 - y_4)\gamma(y_4 - y_1)$$

This distribution can be represented by some undirected graph, $\mathcal{G}$. Draw this graph. Then identify two Y_i's that are conditionally independent given the other two. (Argue carefully that you have this identification right.)

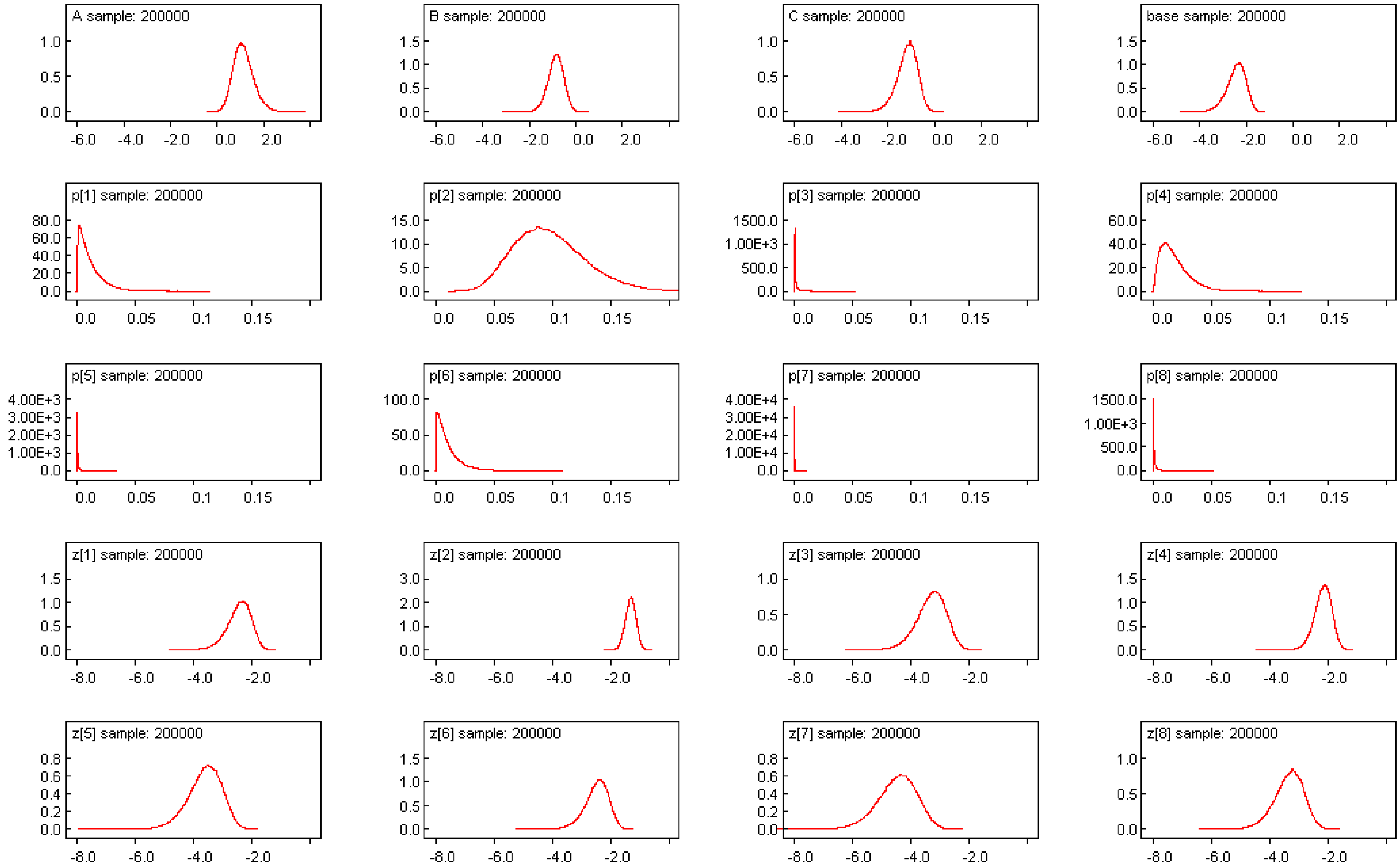
4. A Ni-Cad battery manufacturer was interested in finding what factors affect the faction of cells produced having "shorts" (and the set of process conditions produces the smallest fraction of cells with shorts). For $2\times2\times2=8$ different process set-ups, counts were made of batteries with shorts. The sample sizes varied set-up to set-up (from 88 to 100). Factors and their levels in the study were

A- Rolling Order	1-negative first vs 2-positive first
B- Rolling Direction	1-lower edge first vs 2-upper edge first
C- Nylon Sleeve	1-no vs 2-yes

Below is a summary of a `WinBUGS` session for a Bayes analysis of the manufacturer's experimental results.

```
model {

base ~ dflat()
A ~ dflat()
B ~ dflat()
C ~ dflat()

for (i in 1:8) {
 z[i] <- base + A*IndA2[i] + B*IndB2[i] + C*IndC2[i]
 p[i]<- phi(z[i])
 Y[i] ~ dbin(p[i],n[i])
 }
}

list(Y=c(1,8,0,2,0,1,0,0), n=c(80,88,90,100,90,90,90,90), IndA2=c(0,1,0,1,0,1,0,1),
IndB2=c(0,0,1,1,0,0,1,1), IndC2=c(0,0,0,0,1,1,1,1))

list(base=-1,A=0,B=0,C=0)
```

node	**mean**	**sd**	**MC error**	**2.5%**	**median**	**97.5%**	**start**	**sample**
A	1.133	0.4337	0.003924	0.3843	1.097	2.086	10001	200000
B	-0.8596	0.3367	9.838E-4	-1.565	-0.8438	-0.2446	10001	200000
C	-1.174	0.4356	0.00104	-2.14	-1.135	-0.4291	10001	200000
base	-2.448	0.4089	0.003755	-3.357	-2.409	-1.762	10001	200000
p[1]	0.01112	0.0105	8.067E-5	3.938E-4	0.008005	0.03907	10001	200000
p[2]	0.09777	0.0308	8.382E-5	0.04647	0.09483	0.1661	10001	200000
p[3]	0.001321	0.002125	1.098E-5	5.433E-6	5.43E-4	0.007322	10001	200000
p[4]	0.01831	0.01245	3.148E-5	0.002429	0.01561	0.04932	10001	200000
p[5]	6.427E-4	0.001306	5.663E-6	5.208E-7	1.767E-4	0.004184	10001	200000
p[6]	0.01	0.009467	2.15E-5	3.187E-4	0.007222	0.0352	10001	200000
p[7]	5.617E-5	2.034E-4	6.393E-7	1.568E-9	4.655E-6	4.605E-4	10001	200000
p[8]	0.001178	0.001932	4.973E-6	4.209E-6	4.742E-4	0.006622	10001	200000
z[1]	-2.448	0.4089	0.003755	-3.357	-2.409	-1.762	10001	200000
z[2]	-1.315	0.1813	5.004E-4	-1.68	-1.312	-0.9699	10001	200000
z[3]	-3.308	0.5007	0.003737	-4.399	-3.267	-2.441	10001	200000
z[4]	-2.175	0.2975	7.372E-4	-2.816	-2.154	-1.651	10001	200000
z[5]	-3.621	0.576	0.003806	-4.884	-3.573	-2.637	10001	200000
z[6]	-2.489	0.4095	9.171E-4	-3.415	-2.446	-1.809	10001	200000
z[7]	-4.481	0.6689	0.003856	-5.924	-4.433	-3.314	10001	200000
z[8]	-3.348	0.5054	0.00129	-4.454	-3.305	-2.477	10001	200000



`IndA2=1` for those Y's from the 2nd level of Factor A, `IndB2=1` for those Y's corresponding to the 2nd level of Factor B, and `IndC2=1` for those Y's from the 2nd level of Factor C.

a) Based on the output above, is there clear indication that one of the factors A,B, or C is the "most important determiner" of the rate of shorts produced? If so, which is it and why do you think so. If not, what about the analysis leaves the issue unresolved?

b) Both Y_7 and Y_8 are based on samples of size 90 and have value 0. Based on the Bayes analysis, is any preference indicated between the set-ups represented by these two observed outcomes? If so, is the indication definitive? Explain.

c) The prior used in the `WinBUGS` analysis is improper. Consider two simpler models for these data involving proper but hopefully "nearly flat" priors for z-values. That is, carefully describe a "flat" proper (joint) prior for z-values (for p's) in a model having different rates of shorts (only) for Y's from the two different levels of A. Next specify a corresponding prior for a single z-value (for a single p) in a model that says none of these factors A,B, or C really affects short rates. Then set up the ratio of integrals that is a Bayes factor for comparing these two models for rates of shorts production.

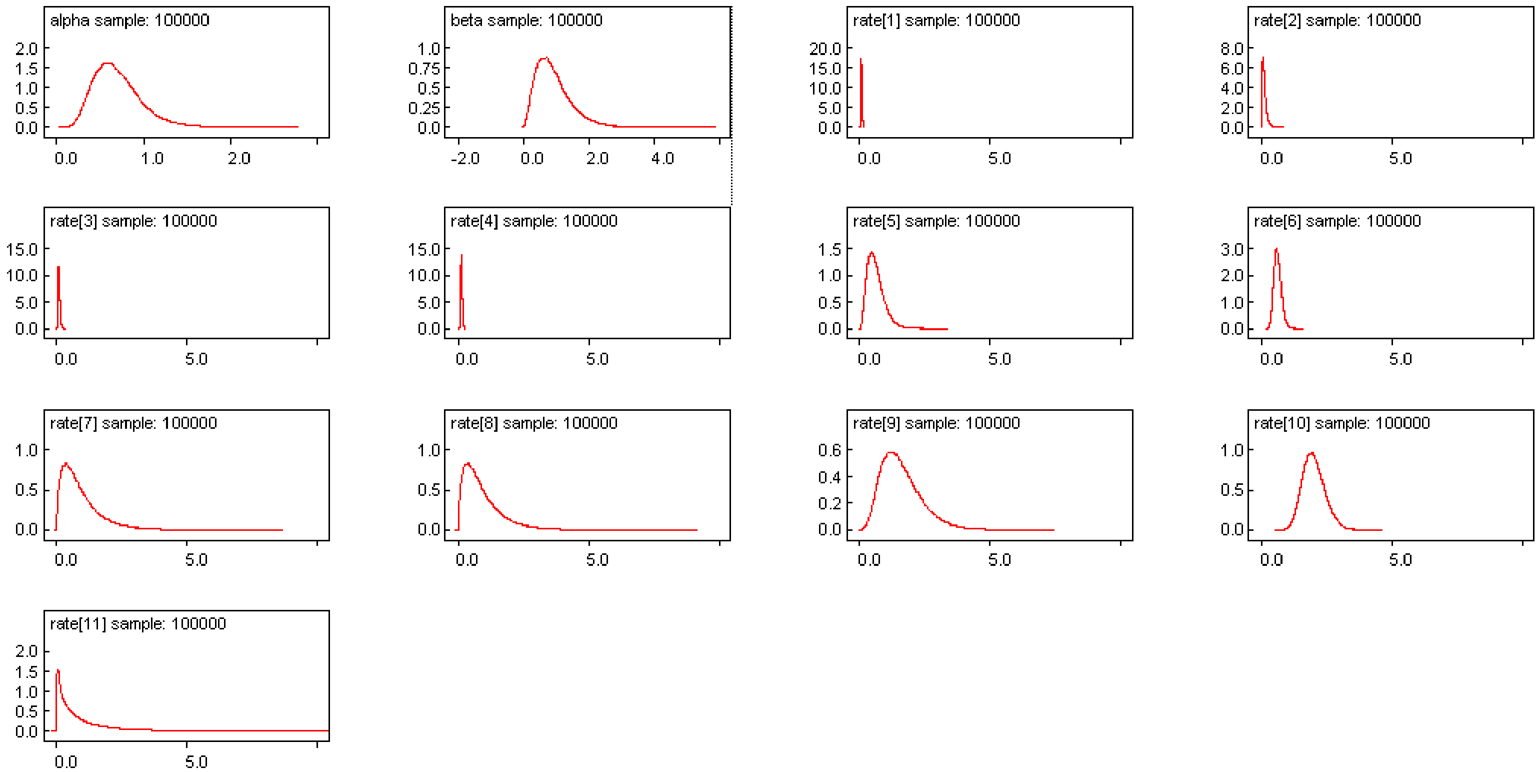
5. The "Pumps Example" in the `WinBUGS` User Manual as an illustration of the gamma distribution concerns failures for 10 different pumps at a rower plant. Failure counts Y_i for $i = 1,\ldots,10$ are modeled as independent $\text{Poisson}(\mu_i)$ variables where

$$\mu_i = rate_i \times time_i$$

for $rate_i$ the pump i failure rate and $time_i$ the recorded running time for pump i. The analysis offered in `WinBUGS` User Manual is based on a hierarchical model due originally to George et al. Below is a slightly modified version of the example code and corresponding summaries of the `WinBUGS` output.

```
model  {

    for (i in 1 : n) {
      rate[i] ~ dgamma(alpha, beta)
      mu[i] <- rate[i] * time[i]
      Y[i] ~ dpois(mu[i])
    }
      rate[n+1] ~ dgamma(alpha, beta)
    alpha ~ dexp(1)
    beta ~ dgamma(0.1, 1.0)
}

list(time = c(94.3, 15.7, 62.9, 126, 5.24, 31.4, 1.05, 1.05, 2.1, 10.5),
  Y = c( 5, 1, 5, 14, 3, 19, 1, 1, 4, 22), n = 10)
```

node	**mean**	**sd**	**MC error**	**2.5%**	**median**	**97.5%**	**start**	**sample**
alpha	0.6971	0.2708	0.001757	0.2869	0.6581	1.336	2001	100000
beta	0.9275	0.5425	0.00357	0.1879	0.8234	2.259	2001	100000
mu[1]	5.636	2.378	0.007946	1.97	5.307	11.14	2001	100000
mu[2]	1.596	1.241	0.004302	0.1265	1.296	4.777	2001	100000
mu[3]	5.615	2.366	0.007355	1.984	5.281	11.11	2001	100000
mu[4]	14.59	3.814	0.01268	8.082	14.27	23.0	2001	100000
mu[5]	3.153	1.652	0.005133	0.7853	2.867	7.091	2001	100000
mu[6]	19.14	4.312	0.01331	11.68	18.82	28.47	2001	100000
mu[7]	0.939	0.763	0.00261	0.07637	0.7417	2.931	2001	100000
mu[8]	0.9391	0.7662	0.002534	0.07682	0.7418	2.932	2001	100000
mu[9]	3.327	1.617	0.006121	0.9874	3.055	7.183	2001	100000
mu[10]	20.87	4.46	0.01538	13.08	20.54	30.5	2001	100000
rate[1]	0.05977	0.02522	8.426E-5	0.02089	0.05627	0.1181	2001	100000
rate[2]	0.1017	0.07903	2.74E-4	0.008057	0.08258	0.3043	2001	100000
rate[3]	0.08926	0.03761	1.169E-4	0.03154	0.08395	0.1766	2001	100000
rate[4]	0.1158	0.03027	1.007E-4	0.06414	0.1132	0.1825	2001	100000
rate[5]	0.6017	0.3152	9.795E-4	0.1499	0.5471	1.353	2001	100000
rate[6]	0.6095	0.1373	4.239E-4	0.372	0.5994	0.9067	2001	100000
rate[7]	0.8943	0.7266	0.002485	0.07274	0.7064	2.791	2001	100000
rate[8]	0.8944	0.7297	0.002413	0.07316	0.7065	2.792	2001	100000
rate[9]	1.584	0.7699	0.002915	0.4702	1.455	3.42	2001	100000
rate[10]	1.987	0.4248	0.001465	1.245	1.956	2.904	2001	100000
rate[11]	0.9326	1.541	0.005263	0.001235	0.4476	4.702	2001	100000



a) A natural non-Bayes way of estimating the rates is as $\widehat{rate_i} = Y_i / time_i$. Values of these estimates are collected in the table below.

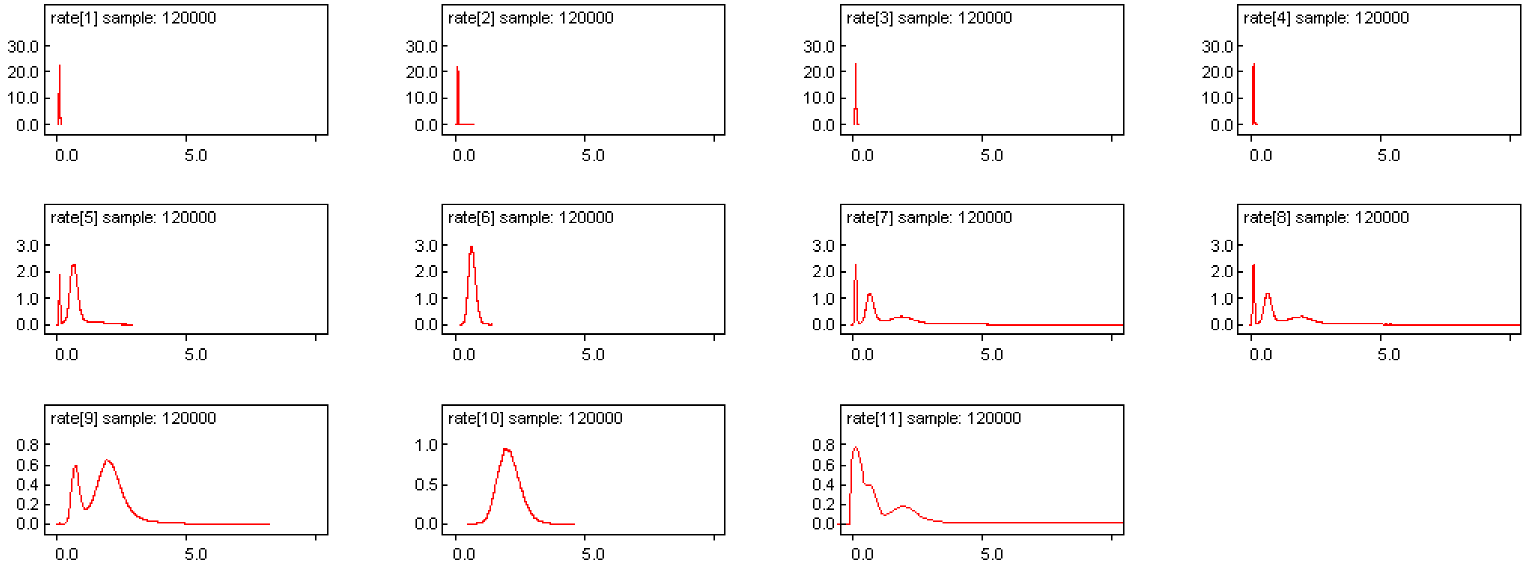
i	1	2	3	4	5	6	7	8	9	10
$Y_i / time_i$	.053	.064	.080	.111	.573	.605	.952	.952	1.905	2.095

How do Bayes inferences for the rates based on the George et al. model compare to these simple estimates? Why (at least in retrospect) do you find these comparisons unsurprising? Explain.

A second Bayes analysis of the Pumps data (using a different prior for the rates) is provided below.

```
model  {
      for (i in 1 : N) {
         X[i] ~ dgamma(alpha, beta)
         theta[i] ~ dbeta(1,priormass)
      }
sum[1] <- theta[1]
      for (i in 2 : (N -1)) {
         sum[i] <- sum[i-1] + theta[i]*(1-sum[i-1])
      }
p[1] <- theta[1]
     for (i in 2 : (N-1)) {
         p[i] <- sum[i] - sum[i-1]
      }
p[N] <- 1 - sum[(N -1)]
      for (i in 1 : n) {
         ind[i] ~ dcat(p[])
         rate[i] <- X[ind[i]]
         mu[i] <- rate[i] * time[i]
         Y[i] ~ dpois(mu[i])
      }
ind[11] ~ dcat(p[])
rate[11] <- X[ind[11]]
}

list(alpha=1, beta=.1, N = 8, priormass=8,
time = c(94.3, 15.7, 62.9, 126, 5.24, 31.4, 1.05, 1.05, 2.1, 10.5),
Y = c( 5, 1, 5, 14, 3, 19, 1, 1, 4, 22), n = 10)
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mu[1]	8.311	1.737	0.007171	5.124	8.234	11.93	10001	120000
mu[2]	1.429	0.4985	0.001567	0.8755	1.382	2.095	10001	120000
mu[3]	5.602	1.164	0.003951	3.557	5.526	8.083	10001	120000
mu[4]	11.35	2.375	0.007775	7.327	11.15	16.54	10001	120000
mu[5]	3.453	1.773	0.008779	0.4428	3.331	8.535	10001	120000
mu[6]	20.37	4.325	0.01529	12.77	20.08	29.66	10001	120000
mu[7]	1.076	0.906	0.004627	0.07514	0.7576	3.015	10001	120000
mu[8]	1.069	0.894	0.004743	0.07519	0.7562	2.989	10001	120000
mu[9]	3.773	1.668	0.008016	1.144	3.889	7.294	10001	120000
mu[10]	21.86	4.548	0.0142	13.82	21.57	31.6	10001	120000
rate[1]	0.08814	0.01842	7.604E-5	0.05434	0.08731	0.1265	10001	120000
rate[2]	0.09101	0.03175	9.981E-5	0.05576	0.08804	0.1334	10001	120000
rate[3]	0.08906	0.0185	6.281E-5	0.05655	0.08786	0.1285	10001	120000
rate[4]	0.09009	0.01885	6.171E-5	0.05815	0.08851	0.1313	10001	120000
rate[5]	0.659	0.3383	0.001675	0.0845	0.6356	1.629	10001	120000
rate[6]	0.6488	0.1377	4.869E-4	0.4068	0.6394	0.9445	10001	120000
rate[7]	1.025	0.8629	0.004406	0.07156	0.7215	2.871	10001	120000
rate[8]	1.018	0.8514	0.004517	0.07161	0.7202	2.846	10001	120000
rate[9]	1.797	0.7943	0.003817	0.5449	1.852	3.473	10001	120000
rate[10]	2.082	0.4331	0.001352	1.316	2.054	3.01	10001	120000
rate[11]	2.461	5.592	0.01751	0.06336	0.6487	19.95	10001	120000



b) What is compact terminology for the prior that has been used for the rates in this second analysis? (Be explicit about parameters, etc.) The estimated posterior densities here are substantially "bumpier" than those on page 8. Why (at least in retrospect and given the nature of the prior) does this not surprise you? Consider, for example, the density for $rate_{11}$. What does this represent and why do you expect it to be "bumpy"? Looking at the times for $i = 5, 7, 8,$ and 9, why are you not surprised that densities for these are the "bumpiest" of those above?

6. Consider a 3-level approximation to a full Polya tree process prior for distributions on $(0,1)$ that uses a mean distribution H that is uniform on $(0,1)$, the natural sequence of partitions

$$B_0 = \left(0, \frac{1}{2}\right] \text{ and } B_1 = \left(\frac{1}{2}, 1\right)$$

$$B_{00} = \left(0, \frac{1}{4}\right], B_{01} = \left(\frac{1}{4}, \frac{1}{2}\right], B_{10} = \left(\frac{1}{2}, \frac{3}{4}\right], \text{ and } B_{11} = \left(\frac{3}{4}, 1\right)$$

$$B_{000} = \left(0, \frac{1}{8}\right], B_{001} = \left(\frac{1}{8}, \frac{1}{4}\right], B_{010} = \left(\frac{1}{4}, \frac{3}{8}\right], B_{011} = \left(\frac{3}{8}, \frac{1}{2}\right], B_{100} = \left(\frac{1}{2}, \frac{5}{8}\right], B_{101} = \left(\frac{5}{8}, \frac{3}{4}\right], B_{110} = \left(\frac{3}{4}, \frac{7}{8}\right], \text{ and } B_{111} = \left(\frac{7}{8}, 1\right)$$

and parameters

$$\alpha_0 = \alpha_1 = 1, \ \alpha_{00} = \alpha_{01} = \alpha_{10} = \alpha_{11} = 4, \text{ and}$$
$$\alpha_{000} = \alpha_{001} = \alpha_{010} = \alpha_{011} = \alpha_{100} = \alpha_{101} = \alpha_{110} = \alpha_{111} = 9$$

An iid sample $Y_1, Y_2, \ldots, Y_{10}$ from a distribution P (itself with the approximate Polya tree prior described above) produces values in the table below (along with the information as to which sets of the 3rd level partition contain the observations).

Observation	.33	.63	.27	.58	.69	.38	.70	.71	.53	.52
Partition Element	B_{010}	B_{101}	B_{010}	B_{100}	B_{101}	B_{011}	B_{101}	B_{101}	B_{100}	B_{100}

What are 1) the posterior distribution of $P(B_{010})$ and 2) the posterior mean of $P((0,.3))$?