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

____________________________________
Signature

____________________________________
Name Printed

There are 14 parts on this exam. I will score every part out of 10 points and count your best 10 scores.
1. Consider \( \mathbf{X} = (X_1, X_2, X_3) \) taking values in \( \{0,1\}^3 \). Below is a table giving values of a joint probability mass function for \( \mathbf{X} \).

\[
\begin{array}{c|c|c}
X_2 = 1 & \frac{1}{12} & \frac{1}{12} \\
X_2 = 0 & \frac{1}{12} & \frac{1}{12} \\
X_1 = 0 & & \\
X_1 = 1 & & \\
X_3 = 0 & & \\
X_3 = 1 & & \\
\end{array}
\]

Suppose that conditioned on \( X_3 = 1 \), \( X_1 = 1 \) with probability \( \frac{2}{3} \) and \( X_2 = 1 \) with probability \( \frac{2}{3} \).

a) Finish filling in the above table of (joint) probabilities for \( \mathbf{X} = \mathbf{x} \) in such a way that \( X_1 \perp X_2 \mid X_3 \).

b) After doing part a), show that \( X_1 \) and \( X_2 \) are not independent.
2. Consider an undirected graph where the vertices are associated with corners of a cube as below.

Maximal cliques here are pairs of vertices that share an edge.

We'll suppose that random variables $X_{000}, X_{100}, X_{010}, X_{110}, X_{001}, X_{101}, X_{011}, X_{111}$ have a joint pdf on $\mathbb{R}^8$

$$f(x) \propto \prod_{(x,x')} \exp\left(-\left((x-x')^2 - x^2 - (x')^2\right)\right)$$

Further suppose that conditioned on the variables $X_{000}, X_{100}, X_{010}, X_{110}, X_{001}, X_{101}, X_{011}, X_{111}$, variables $Y_{000}, Y_{100}, Y_{010}, Y_{110}, Y_{001}, Y_{101}, Y_{011}, Y_{111}$ are independent and normal with standard deviation 1 and means given by corresponding $X$'s ($EY_{000} = X_{000}, \ldots, EY_{111} = X_{111}$).

**a)** Describe how (should you need to) you might generate a realization from $f(x)$. (Convince me that your method would really work.)

**b)** If one observes $Y_{000} = 2, Y_{100} = 3, Y_{010} = 7, Y_{110} = 2, Y_{001} = 3, Y_{101} = 4, Y_{011} = 6, Y_{111} = 2$, and uses a Gibbs algorithm to sample the posterior of $X$, exactly how does one update $X_{000}$ to $X_{000}'$ based on the values of the $j$th iterates of all other $X$'s? (Exactly what distribution is sampled to create the update?)
3. Consider a two-level approximation to a full Polya tree process prior for distributions on \((0, \infty)\) that uses a parameter measure \(H\) that is Exponential with mean 1, two initial partitions that are

\[ B_0 = (0, -\ln(0.5)), \text{ and } B_1 = (-\ln(0.5), \infty), \text{ and} \]

\[ B_{00} = (0, -\ln(0.75)), B_{01} = [-\ln(0.75), -\ln(0.5)), B_{10} = [-\ln(0.5), -\ln(0.25)), \text{ and } B_{11} = [-\ln(0.25), \infty) \]

and parameters

\[ \alpha_0 = \alpha_1 = 1 \quad \text{and} \quad \alpha_{00} = \alpha_{01} = \alpha_{10} = \alpha_{11} = 4 \]

Suppose that an iid sample of size \(n = 4\) from a distribution \(P\) that has the approximate Polya tree process prior described above produces 1 observation in \(B_{00}\), 2 observations in \(B_{01}\), and 1 observation in \(B_{11}\).

**a)** What is the posterior mean of \(P((0.2, 0.3))\), the \(P\) probability assigned to the interval \((0.2, 0.3)\)?

(Hint: \(P((0.2, 0.3)) = P((0.2, -\ln(0.75)]) + P([-\ln(0.75), 0.3])) \))

**b)** Provide either a pencil and paper calculation or some **BUGS** code that you could use to evaluate the posterior variance of \(P((0.2, 0.3))\).
4. A place that the Dirichlet process and its finite stick-breaking relatives have been used in data analysis is in "clustering" (which amounts to attempting to break data cases up into more or less homogeneous groups/clusters). This question concerns stick-breaking-based clustering for the $n = 20$ values from $\mathbb{R}$ presented below (in ordered fashion).

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_i$</td>
<td>3.4</td>
<td>4.4</td>
<td>4.5</td>
<td>5.1</td>
<td>5.4</td>
<td>5.6</td>
<td>5.8</td>
<td>5.9</td>
<td>5.9</td>
<td>6.0</td>
<td>6.2</td>
<td>6.2</td>
<td>6.2</td>
<td>6.4</td>
<td>6.9</td>
<td>7.2</td>
<td>7.5</td>
<td>8.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Some BUGS code is below (in two-column format). It was used to implement a model for a random distribution $P_N$ that is a truncated version of a DP with parameter measure $\alpha$ that is

$$\text{priormass} \cdot \mathcal{N}(6, 10)$$

(10 is the variance of the normal distribution). Conditioned on $P_N$ some set of latent means $\mu_1, \mu_2, \ldots, \mu_n$ are iid $P_N$, and conditioned on $P_N$ and the latent means, observed (original/unordered) $Y_i$ are independent $\mathcal{N}(\mu_i, \sigma_{\text{radius}}^2)$ variables. $Y$'s sharing a common (latent) mean might be thought of as ideally clustered together. (The parameter $\sigma_{\text{radius}}$ is a kind of "radius" that functions to control the extent of the latent clusters.) One might use a posterior from this kind of modeling to investigate what is the "right number of clusters" (as represented by the posterior distribution over the number of different values among the $\mu_i$) and the likelihood that various pairs of observations "belong together" (as represented by the posterior probability that they share the same latent mean).

```R
model {
  for (i in 1:N) {
    mu[i] ~ dnorm(6,.1)
    theta[i] ~ dbeta(1, priormass)
  }

  for (i in 2 : (N - 1)) {
    sum[i] <- sum[i-1] + theta[i]*(1-sum[i-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[])
    Y[i] ~ dnorm(mu[ind[i]], tau_radius)
  }

  for (i in 1:n) {
    for (j in 1:n) {
      I[i,j] <- equals(mu[ind[i]], mu[ind[j]])
    }
  }

  for (i in 1:N) {
    tot[i,1] <- equals(mu[i], mu[ind[1]])
    for (j in 2:n) {
      tot[i,j] <- tot[i,j-1]+equals(mu[i], mu[ind[j]])
    }
  }

  act[1] <- 1-equals(tot[1,n],0)
  for (i in 2:N) {
    act[i] <- act[i-1]+(1-equals(tot[i,n],0))
  }

  active <- act[N]
}

list(N = 10, priormass=10, tau_radius=1, Y=c(3.4,4.4,4.5,5.1,5.4,5.5,5.6,5.8,5.8,5.9,5.9,6.0,6.2,6.2,6.4,6.9,7.2,7.5,8.2), n=20)
```
a) Below are some summaries of running the code with several different values of \textit{tauradius}. Describe/interpret what happens with increasing \textit{tauradius}. Explain why what you see makes sense.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figures}
\caption{Graphs showing posterior distribution of variable $I[9, j]$ for different values of tauradius.}
\end{figure}

b) What pattern would you expect/hope to see in the posterior means of variable $I[9, j]$ (see the code) as $j$ runs from 1 to 20, if the modeling process here is to be considered successful? Be explicit and explain yourself.
c) Below are values of $\tau_{radius}$ and corresponding estimated posterior means of $I_{[13,14]}$. What do these suggest to you regarding an appropriate value for $\tau_{radius}$ in this modeling?

<table>
<thead>
<tr>
<th>$\tau_{radius} = 1$</th>
<th>$\tau_{radius} = 4$</th>
<th>$\tau_{radius} = 16$</th>
<th>$\tau_{radius} = 49$</th>
<th>$\tau_{radius} = 100$</th>
<th>$\tau_{radius} = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.463</td>
<td>.619</td>
<td>.769</td>
<td>.833</td>
<td>.914</td>
<td>.998</td>
</tr>
</tbody>
</table>
5. This question concerns the use of 1-D linear regression splines and Bayes analysis. For "knots" $c_1 < c_2$ both belonging to an interval of interest $[a, b]$, and a predictor variable $x$ taking values in $[a, b]$, the function (of $x$)

$$f(x | \beta_0, \beta_1, \gamma_1, \gamma_2) = \beta_0 + \beta_1 x + \gamma_1 (x - c_1) I[x \geq c_1] + \gamma_2 (x - c_2) I[x \geq c_2]$$

(*)

is a continuous piecewise-linear function (with derivative $\beta_1$ for $x < c_1$, $\beta_1 + \gamma_1$ for $c_1 < x < c_2$, and $\beta_1 + \gamma_1 + \gamma_2$ for $x > c_2$) called a linear regression spline. For fixed knots, the variables

$$z(x) = (x - c_1) I[x \geq c_1] \quad \text{and} \quad w(x) = (x - c_2) I[x \geq c_2]$$

are known predictors, and the model

$$y_i = \beta_0 + \beta_i x_i + \gamma_i z(x_i) + \gamma_2 w(x_i) + \epsilon_i$$

for $\epsilon_i$'s iid $N(0, \sigma^2)$ is an ordinary linear model (a multiple regression model) like all those met in Stat 500 and early Stat 511. A MLR program provides standard frequentist analysis of $n$ pairs $(x_i, y_i)$. But fitting not only the parameters $\beta_i, \beta_1, \gamma_1, \gamma_2$, and $\sigma$, but also the knots as well takes one into the realm of non-linear regression analysis (and this problems is a fairly unpleasant case).

Here we consider the analysis of $n = 21$ data pairs where each $x_i \in [0,1]$. These are plotted to the right along with fitted means from a **BUGS** analysis of the nonlinear regression model. Some code used to do the fitting is below (in two-column format) and some **BUGS** output is provided on the next page.

```plaintext
model {

for (i in 1:4) {
  beta[i] ~ dflat()
}

for (i in 1:2) {
  phi[i] ~ dbeta(1,1)
}

logsigma~dflat()
sigma<-exp(logsigma)
tau<-exp(-2*logsigma)

for (i in 1:21) {
  +beta[3]*max(0,x[i]-knot[1])
  +beta[4]*max(0,x[i]-knot[2])
y[i]~dnorm(mu[i],tau)
}

for (i in 1:21) {
  ynew[i]~dnorm(mu[i],tau)
}

}

list(x=c(0,.05,.10,.15,.20,.25,.30,.35,.40,.45,.50,.55,.60,.65,.70,.75,.80,.85,.90,.95,1.0), y=c(0.0243, 0.1097, -0.0482, 0.0390, 0.1191, 0.1272, 0.1009, 0.2987, 0.2728, 0.2547, 0.5261, 0.5446, 0.6659, 0.8744, 0.9388, 1.3079, 1.3176, 1.4383,1.7490, 1.8725, 2.1309))

list(beta=c(0,0,0,0),phi=c(.5,.5), logsigma=0)
```

![Image](image.png)
| BUGS OUTPUT |
|---|---|---|---|---|---|---|---|
|  | **mean** | **sd** | **MC_error** | **val2.5pc** | **median** | **val97.5pc** | **sample** |
| beta[1] | 0.0215 | 0.0677 | 0.0019 | -0.0977 | 0.0172 | 0.1659 | 270640 |
| beta[3] | 1.632 | 4.992 | 0.1981 | -12.3 | 1.993 | 17.75 | 270640 |
| beta[4] | 2.901 | 5.67 | 0.2288 | -6.866 | 1.539 | 1.715 | 270640 |
| mu[1] | 0.0215 | 0.0677 | 0.0019 | -0.0977 | 0.0172 | 0.1659 | 270640 |
| mu[2] | 0.0167 | 0.0497 | 0.0013 | -0.0874 | 0.0182 | 0.1109 | 270640 |
| mu[3] | 0.0321 | 0.0455 | 0.0009 | -0.0681 | 0.0351 | 0.1136 | 270640 |
| mu[4] | 0.0586 | 0.0423 | 0.0007 | -0.0358 | 0.0622 | 0.1326 | 270640 |
| mu[5] | 0.0917 | 0.0392 | 0.0006 | 0.0052 | 0.095 | 0.1608 | 270640 |
| mu[6] | 0.1294 | 0.0386 | 0.0005 | 0.0437 | 0.1327 | 0.1974 | 270640 |
| mu[7] | 0.1721 | 0.04 | 0.0006 | 0.085 | 0.175 | 0.243 | 270640 |
| mu[8] | 0.221 | 0.0421 | 0.0006 | 0.1338 | 0.2232 | 0.2969 | 270640 |
| mu[9] | 0.2729 | 0.0513 | 0.0009 | 0.1619 | 0.2784 | 0.3589 | 270640 |
| mu[10] | 0.3317 | 0.0604 | 0.0012 | 0.2073 | 0.3389 | 0.4322 | 270640 |
| mu[11] | 0.4228 | 0.0493 | 0.0007 | 0.3276 | 0.4224 | 0.5177 | 270640 |
| mu[12] | 0.5414 | 0.0467 | 0.0009 | 0.4476 | 0.5427 | 0.6294 | 270640 |
| mu[13] | 0.6894 | 0.0487 | 0.0011 | 0.5884 | 0.6922 | 0.7768 | 270640 |
| mu[14] | 0.8555 | 0.045 | 0.001 | 0.7589 | 0.8589 | 0.9335 | 270640 |
| mu[15] | 1.026 | 0.0385 | 0.0008 | 0.9423 | 1.03 | 1.094 | 270640 |
| mu[16] | 1.199 | 0.032 | 0.0005 | 1.131 | 1.2 | 1.257 | 270640 |
| mu[17] | 1.371 | 0.0286 | 0.0003 | 1.313 | 1.371 | 1.426 | 270640 |
| mu[18] | 1.543 | 0.0296 | 0.0003 | 1.484 | 1.543 | 1.601 | 270640 |
| mu[19] | 1.715 | 0.0348 | 0.0006 | 1.646 | 1.715 | 1.784 | 270640 |
| mu[20] | 1.887 | 0.0427 | 0.0009 | 1.804 | 1.887 | 1.973 | 270640 |
| mu[21] | 2.059 | 0.0521 | 0.0012 | 1.959 | 2.059 | 2.165 | 270640 |
| sigma | 0.0804 | 0.0155 | 0.0002 | 0.0567 | 0.0782 | 0.1169 | 270640 |
| ynew[1] | 0.0217 | 0.1061 | 0.0019 | -0.1838 | 0.0197 | 0.2375 | 270640 |
| ynew[2] | 0.0164 | 0.0959 | 0.0013 | -0.176 | 0.0174 | 0.2044 | 270640 |
| ynew[3] | 0.0322 | 0.0937 | 0.0009 | -0.1563 | 0.0331 | 0.2154 | 270640 |
| ynew[4] | 0.0588 | 0.0921 | 0.0008 | -0.1256 | 0.0594 | 0.2397 | 270640 |
| ynew[5] | 0.0918 | 0.0909 | 0.0006 | -0.0896 | 0.0923 | 0.2704 | 270640 |
| ynew[6] | 0.1295 | 0.0905 | 0.0006 | -0.0503 | 0.13 | 0.3082 | 270640 |
| ynew[7] | 0.172 | 0.0912 | 0.0006 | -0.0094 | 0.1726 | 0.3512 | 270640 |
| ynew[8] | 0.221 | 0.0918 | 0.0006 | 0.0375 | 0.2215 | 0.4012 | 270640 |
| ynew[9] | 0.2729 | 0.0965 | 0.0009 | 0.0771 | 0.2744 | 0.46 | 270640 |
| ynew[10] | 0.3319 | 0.1018 | 0.0012 | 0.1264 | 0.334 | 0.527 | 270640 |
| ynew[11] | 0.4228 | 0.0955 | 0.0008 | 0.234 | 0.4225 | 0.6121 | 270640 |
| ynew[12] | 0.541 | 0.0939 | 0.0009 | 0.3557 | 0.5408 | 0.7265 | 270640 |
| ynew[13] | 0.6895 | 0.0953 | 0.0011 | 0.5001 | 0.6899 | 0.8777 | 270640 |
| ynew[14] | 0.8555 | 0.0936 | 0.0011 | 0.669 | 0.8561 | 1.038 | 270640 |
| ynew[15] | 1.026 | 0.0905 | 0.0008 | 0.8454 | 1.027 | 1.204 | 270640 |
| ynew[16] | 1.199 | 0.0878 | 0.0005 | 1.024 | 1.199 | 1.372 | 270640 |
| ynew[17] | 1.371 | 0.0866 | 0.0003 | 1.199 | 1.371 | 1.542 | 270640 |
| ynew[18] | 1.543 | 0.087 | 0.0003 | 1.37 | 1.543 | 1.715 | 270640 |
| ynew[19] | 1.715 | 0.0891 | 0.0006 | 1.539 | 1.715 | 1.892 | 270640 |
| ynew[20] | 1.887 | 0.0923 | 0.0009 | 1.705 | 1.887 | 2.07 | 270640 |
| ynew[21] | 2.06 | 0.0969 | 0.0012 | 1.868 | 2.059 | 2.253 | 270640 |
a) As indicated earlier, the function of \( x \) in (*) on page 8 is piecewise-linear. Look again at the plot of posterior means on page 8 (the crosses are the fitted means). Does the posterior mean function appear to be piecewise-linear with (with two "corners")? Is this an indication of something wrong in the posterior simulation? Explain.

b) Is there evidence of problems with the Bayes analysis in the posterior predictive simulations summarized on the previous page? Explain.

c) It appears that the fitted posterior mean function on the previous page may be convex (or close to being so), i.e. have an increasing first derivative. What modification of the \texttt{BUGS} code would guarantee that the posterior mean function is convex for any data set? (It is sufficient that conditional on the knot values the posterior mean function is convex for any data set.)
A classical hierarchical data set originally appearing in Snedecor and Cochran concerns measured glycogen content of rat livers. Rats were each given one of 3 different treatments, 2 rats per treatment. Rats were sacrificed and livers cut into 3 pieces each. Then 2 analyses were made for glycogen content on each piece (for $3 \times 2 \times 3 \times 2 = 36$ measured contents). With $y_{ijkl} = \text{measurement } l \text{ from piece } k \text{ from rat } j \text{ given treatment } i$ we will consider a Bayes analysis based on a model

$$y_{ijkl} = \mu_i + \rho_j + \theta_{ijk} + \epsilon_{ijkl}$$

where $\mu_1, \mu_2,$ and $\mu_3$ are fixed parameters, the $\rho_j$ are iid $N\left(0, \sigma^2_{\rho}\right)$ random rat effects, independent of iid $N\left(0, \sigma^2_{\theta}\right)$ random piece effects $\theta_{ijk},$ independent of random analysis errors $\epsilon_{ijkl}$ that are iid $N\left(0, \sigma^2\right)$. Next is some **BUGS** code and output for this problem.

```r
model {
  for (i in 1:3) {
    mu[i] ~ dflat()
  }
  for (j in 1:6) {
    rho[j] ~ dnorm(0,taurho)
  }
  for (k in 1:18) {
    theta[k] ~ dnorm(0,tautheta)
  }
  for (i in 1:36) {
    mean[i]<-
    mu[treat[i]]+rho[rat[i]]+theta[piece[i]]
  }
  for (i in 1:36) {
    y[i]~dnorm(mean[i],tau)
  }
  sigmarho~dunif(0,100)
  taurho<-1/(sigmarho*sigmarho)
  sigmatheta~dunif(0,100)
  tautheta<-1/(sigmatheta*sigmatheta)
  logsigma~dflat()
  sigma<-exp(2*logsigma)
  tau<-1/(sigma*sigma)
}
```

List of data:

- `y`: 36 glycogen content measurements
- `treat`: treatment assignment for each rat
- `rat`: rat number for each measurement
- `piece`: piece number for each measurement

List of parameters:


Output summary:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>MC_error</th>
<th>2.5pc</th>
<th>Median</th>
<th>97.5pc</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>d12</td>
<td>-10.51</td>
<td>15.8</td>
<td>0.1011</td>
<td>-41.72</td>
<td>-10.52</td>
<td>20.86</td>
<td>999001</td>
</tr>
<tr>
<td>d13</td>
<td>5.581</td>
<td>16.66</td>
<td>0.1223</td>
<td>-25.48</td>
<td>5.397</td>
<td>38.03</td>
<td>999001</td>
</tr>
<tr>
<td>d23</td>
<td>16.09</td>
<td>16.62</td>
<td>0.1289</td>
<td>-15.25</td>
<td>15.89</td>
<td>48.76</td>
<td>999001</td>
</tr>
<tr>
<td>mu[1]</td>
<td>140.5</td>
<td>11.25</td>
<td>0.07198</td>
<td>118.4</td>
<td>140.5</td>
<td>162.8</td>
<td>999001</td>
</tr>
<tr>
<td>mu[2]</td>
<td>151.0</td>
<td>11.36</td>
<td>0.07772</td>
<td>128.8</td>
<td>151.0</td>
<td>173.2</td>
<td>999001</td>
</tr>
<tr>
<td>mu[3]</td>
<td>134.9</td>
<td>11.87</td>
<td>0.09302</td>
<td>111.8</td>
<td>135.1</td>
<td>156.7</td>
<td>999001</td>
</tr>
<tr>
<td>sigmarho</td>
<td>11.87</td>
<td>10.91</td>
<td>0.08049</td>
<td>1.903</td>
<td>8.669</td>
<td>43.09</td>
<td>999001</td>
</tr>
<tr>
<td>sigmatheta</td>
<td>3.976</td>
<td>1.865</td>
<td>0.00398</td>
<td>0.488</td>
<td>3.902</td>
<td>7.984</td>
<td>999001</td>
</tr>
<tr>
<td>sigma</td>
<td>4.896</td>
<td>0.852</td>
<td>0.00136</td>
<td>3.521</td>
<td>4.798</td>
<td>6.824</td>
<td>999001</td>
</tr>
</tbody>
</table>
a) Are there clear differences indicated among the three treatment effects? Is it clear whether the largest part of the response variability for a given treatment comes from rat-to-rat differences, piece-to-piece differences, or from analysis-to-analysis differences? Explain both your answers.

b) What prior distributions were used for standard deviations $\sigma_\rho, \sigma_\alpha$, and $\sigma$? These are in accord with Gelman's recommendations for this kind of linear mixed model. Why does he recommend two different kinds of priors be used?