STAT 4224/5224	**Final**	**Name:**
Spring 2023	**Practice**	**Uni:**

Important:
You are allowed to bring one letter-sized page written on one side of your own notes.

1. True or False:

   a. In the Metropolis-Hastings algorithm the proposal distribution may not depend on the iteration number. F

   b. For a Metropolis-Hastings algorithm with state-independent proposals, it is important that the range of the proposal distribution be a strict subset of the range of the target distribution. F

   c. A common practice in MCMC is to discard the early iterations as a warm-up period, also called burn-in; part of the reason for this is to eliminate (or at least reduce) the dependence of the output on a user-selected starting values. T

   d. In a Metropolis algorithm with jump proposal distribution $\theta^{*} \sim N(\theta^{t-1}, \delta^2)$, a very low acceptance rate is a sign that the variance δ^2 is set too low. F

   e. In latent variable models usually one of the observable variables, say Y, is a function of the unobservable variable Z, that is $Y = f(Z)$. T

   f. Suppose we want to simulate a random sample from the following discrete distribution:

θ	1	2	3
$p(\theta)$	0.2	0.5	0.3

   Assume current $\theta^{(s)} = 2$ and we use a Metropolis-Hastings with proposal distribution which is discrete uniform distribution on $\{1, 2, 3\}$. If the Metropolis proposal is $\theta^{(*)} = 1$, then it will be accepted with probability 0.4. T

   g. A disadvantage of the mixed effects models used for hierarchical data is that they tend to overfit the data within each cluster. F

   h. In the Bayesian linear regression model $Y|\boldsymbol{\beta}, \sigma^2, X \sim N_n(X\boldsymbol{\beta}, \sigma^2 I)$, we used a multivariate normal prior on $\boldsymbol{\beta}$ and a Gamma distribution on σ^2 in order to run a Gibbs sampler to estimate the posterior distribution. F

   i. The Gibbs sampler is a special case of the Metropolis-Hastings algorithm. T

   j. In MCMC methods the strength of autocorrelation is what determines the efficiency of the algorithm. T

2. In this problem we consider the ordinal regression model fit to the following data. The outcome variable is the answer to the question "How likely are you to apply to graduate school?" and has three levels called apply, with levels "unlikely", "somewhat likely", and "very likely", coded 1, 2, and 3, respectively. We also have three variables that we will use as predictors: pared, which is a 0/1 variable indicating whether at least one parent has a graduate degree; public, which is a 0/1 variable where 1 indicates that the undergraduate institution is public and 0 private, and gpa, which is the student's grade point average.
The following R code was obtained:

```
library(rstanarm)
post0 <- stan_polr(apply ~ pared + public + gpa, data = dat,
                   prior    =    R2(0.25),    prior_counts    =
                   dirichlet(1))
summary(post0)
Estimates:
                                mean   sd   10%   50%   90%
pared                           1.0    0.3  0.7   1.0   1.3
public                         -0.1    0.3 -0.4  -0.1   0.3
gpa                             0.6    0.2  0.3   0.6   0.9
unlikely|somewhat likely        2.1    0.7  1.1   2.1   3.0
somewhat likely|very likely     4.2    0.8  3.2   4.2   5.1
```

a. Write down the estimated equation(s) for this model.

$$logit(P(Y \leq 1)) = 2.1 - 1 * PARED - (-0.1) * PUBLIC - 0.6 * GPA$$
$$logit(P(Y \leq 2)) = 4.2 - 1 * PARED - (-0.1) * PUBLIC - 0.6 * GPA$$

b. Interpret the posterior mean coefficient of the pared predictor.

For students whose parents did attend college, the odds of being more likely (i.e., very, or somewhat likely versus unlikely) to apply is 2.7 times that of students whose parents did not go to college, holding constant all other variables.

c. Which predictors are significant and with what posterior probability did you arrive at this conclusion?

Parents' education and GPA with 80% posterior probability.

3. A textile factory produces many units of fabric, using many different machines. Suppose we have a random sample of $n = 6$ units from each of $m = 4$ machines. Letting Y_{ij} denote the tensile strength of the i^{th} unit produced by the j^{th} machine, we fit the following model:

$$Y_{ij}|\theta_j, \sigma^2 \sim ind. N(\theta_j, \sigma^2)$$
$$\theta_j \sim ind. N(\mu, \tau^2)$$
$$\sigma^2, \mu, \tau^2 \sim \pi(\sigma^2, \mu, \tau^2)$$

An MCMC method was used and the output from the posterior simulation $\{(\sigma^{2(s)}, \mu^{(s)}, \tau^{2\{s\}}, s = 1, \ldots, 6)\}$ is shown below:

```
sigma2 244 156 119 134 207 234
mu      42  46  46  47  46  47
tau2   472 549 688 470 889 637
```

a) True or False: this is multilevel Bayesian model. T

b) Approximate the posterior expectation of the between-machine variance.

$$\hat{\tau}^2 = \frac{1}{6}(472 + 549 + 688 + 470 + 889 + 637) = 617.5$$

c) Approximate the posterior expectation of the within-machine variance.

$$\hat{\sigma}^2 = \frac{1}{6}(244 + 156 + 119 + 134 + 207 + 234) = 182.3$$

d) Explain how you could, from the output above and with a random number generator (like the R function rnorm or similar), generate $S = 6$ random draws from the posterior predictive distribution for the tensile strength of a unit produced by a different randomly selected machine, not one of the $m = 4$ in this data set. If the output reported above is not sufficient, specify exactly what additional posterior simulation output you would need.

y.tilde = rnorm(6, mu, sqrt(tau2 + sigma2))

4. In this problem we consider the paper quality data produced by different operators and fit the following Stan model.

```
write("
data {
  int<lower=0> N; // sample size
  int<lower=0> J; // number of groups
  int<lower=1,upper=J> predictor[N]; // group indeces
  vector[N] response; // y variable
}
parameters {
  vector[J] eta;
  real mu;
  real<lower=0> sigmaalpha;
  real<lower=0> sigmaepsilon;
}
transformed parameters {
  vector[J] a;
  vector[N] yhat;
  a = mu + sigmaalpha * eta;
  for (i in 1:N)
    yhat[i] = a[predictor[i]];
}
model {
  eta ~ normal(0, 1);
  sigmaalpha ~ exponential(1);
  response ~ normal(yhat, sigmaepsilon);
}
", "Example7.stan")
pulpdat <- list(N=nrow(pulp),
                J=length(unique(pulp$operator)),
                response=pulp$bright,
                predictor=as.numeric(pulp$operator))
mod1 <- stan_model("Example7.stan")
system.time(fit <- sampling(mod1, data=pulpdat))
print(fit, pars=c("mu","sigmaalpha","sigmaepsilon","a"))
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu	60.38	0.01	0.19	59.95	60.28	60.39	60.50	60.80	592	1.01
sigmaalpha	0.36	0.01	0.25	0.04	0.20	0.30	0.45	1.05	481	1.00
sigmaepsilon	0.36	0.00	0.07	0.25	0.31	0.35	0.40	0.54	957	1.01
a[1]	60.29	0.00	0.14	60.01	60.19	60.29	60.38	60.57	3430	1.00
a[2]	60.16	0.00	0.17	59.84	60.04	60.15	60.27	60.49	2010	1.00
a[3]	60.56	0.00	0.15	60.25	60.46	60.55	60.66	60.86	2594	1.00
a[4]	60.60	0.00	0.15	60.30	60.49	60.60	60.70	60.90	1834	1.00

a) Are there any issue with diagnostics. Explain why.

No. All Rhat values are < 1.1 and the worst ESS is 12 % (for sigmaalpha) which is still ok.

b) Write down the estimate model equation(s).

$$\hat{x}_j = a_j, j = 1, \dots, 4, a_1 = 60.29, a_2 = 60.16, a_3 = 60.56, a_4 = 60.6$$

c) Is there difference between the operators? Explain.

There is some difference as the a's are all different, but we don't have enough R output to decide if the differences are significant.

d) How does the variation between operators compare to the variation within operators?

They are estimated to be equal as both sigma's are the same 0.36

e) What other code would you need to estimate the difference between a pair of operators and test its significance?

We need simultaneous posterior CI for all 6 contrasts between the different pairs of a's.