ST559 Final

Nick Sun

June 11, 2020

Abstract

Four questions in total: One question about Bayesian ANOVA, another with a Poisson-Gamma Bayesian model, and a last question with a multinomial-dirichlet model.

1 Q1

The following data provides the % silver of Byzantine coins from four different periods of time in the reign of King Manuel I, Comnenus. We are interested in seeing if there is a significant differneces in the silver content of coins minted early or later in Manuel's reign. Use Bayesian methods while specifying your assumptions, prior distribution, the likelihood, and comparison of silver content in the coins.

Period	Measurements
1	5.9,6.8,6.4,7.0,6.6,7.7,7.2,6.9,6.2
2	6.9,9.0,6.6,8.1,9.3,9.2,8.6
3	4.9,5.5,4.6,4.5
4	5.3,5.6,5.5,5.1,6.2,5.8,5.8

Assumptions We start by assuming that each of the observations in the four coin periods follows this model:

$$y_{ij} \sim N(\mu_i, \sigma^2), i = 1, 2, 3, 4$$

and each observation within a coin period is independent from the others.

Prior A reasonable prior that we have been using for Bayesian ANOVA is

$$\pi(\mu_1, \ldots, \mu_4, \sigma^2) \propto \frac{1}{\sigma^2}$$

Likelihood The likelihood function is going to be equal to

$$f(data|\mu_1, \ldots, \mu_4, \sigma^2) = \left(\sqrt{2\pi}\sigma\right)^{-N} exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{4}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{4}\sum_{j=1}^{n_i}(\bar{y}_i - \mu_i)^2\right]\right)$$

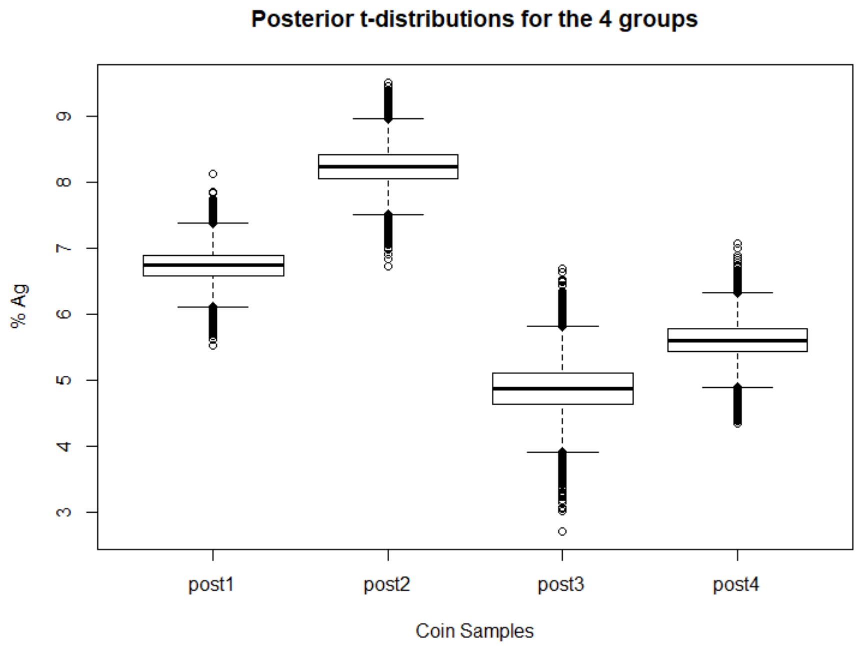
where $N = 27$, the total number of coins, and $MSW = \frac{\sum_{i=1}^{4}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2}{N-4} = .4789$.

Posterior Previous work has shown that the posterior μ parameters follow a multivariate t-distribution and that the individual μ parameters marginal posterior are location-scale t-distribution.

$$\mu_i \sim \bar{y}_i + \sqrt{\frac{MSW}{n_i}}\frac{z}{V/(N-4)}$$

where $z \sim N(0, 1)$ and $V \sim \chi^2_{N-4}$.

Conclusion With $nsim = 100000$, we get the following posterior distributions:



We can see that the coins in period 1 are similar to period 2 and the coins in period 3 are similar to period 4. However, there is little overlap between the coins in periods and 1 and 2 and the coins in periods 3 and 4. Taking a numerical summary of period 2 and period 3, we see that there is in fact no overlap between the ranges of those location-scale t-distributions.

This gives us evidence that there is a difference in $\mu_i, i = 1, 2, 3, 4$ and furthermore that the percent silver differs in coins between the beginning and start of Manuel's reign.

```
coindata <- list(
    s1 = c(5.9, 6.8, 6.4, 7.0, 6.6, 7.7, 7.2, 6.9, 6.2),
    s2 = c(6.9, 9.0, 6.6, 8.1, 9.3, 9.2, 8.6),
    s3 = c(4.9, 5.5, 4.6, 4.5),
    s4 = c(5.3, 5.6, 5.5, 5.1, 6.2, 5.8, 5.8)
)

ybars <- unlist(lapply(coindata, mean))
ni <- unlist(lapply(coindata, length))
N <- 9+7+4+7
ssw <- Reduce("+", lapply(coindata, function(x) {sum((x - mean(x))^2)}))
msw <- ssw/(N-4)

nsim <- 100000
post.t <- data.frame(post1 = vector(mode = "numeric", length = nsim),
                     post2 = vector(mode = "numeric", length = nsim),
                     post3 = vector(mode = "numeric", length = nsim),
                     post4 = vector(mode = "numeric", length = nsim))
for (i in 1:4) {
    post.t[,i] <- ybars[i] + sqrt(msw/ni[i])*(rnorm(nsim)/sqrt(rchisq(nsim,
       N-4)/(N-4)))
}

boxplot(post.t,
        main = "Posterior t-distributions for the 4 groups",
        xlab = "Coin Samples",
        ylab = "% Ag")
summary(post.t)
```

2 Q2

We have data on 8 plots of 50 cabbage plants, with 4 plots randomly assigned to one of two treatments. Each data point is the number of loopers in the plot, a kind of pest.

If we denote the first treatment data as $X_1, \ldots, X_n$ and the second treatment data as $Y_1, \ldots, Y_n$ with the two samples being treated as independent from each other, suppose that all samples follow a Poisson distribution with parameters λ_1 and λ_2 respectively.

Treatment	Loopers
1	11, 4, 4, 5
2	6, 4, 3, 6

2.1 part a.

Derive the likelihood function.

answer Since $X_1, \ldots, X_n$ is independent from $Y_1, \ldots, Y_n$, we have

$$f(data|\lambda_1, \lambda_2) = \frac{1}{\prod x_i \prod y_i} exp\left(-n\lambda_1 - n\lambda_2\right) \lambda_1^{\sum_{i=1}^{n} x_i} \lambda_2^{\sum_{j=1}^{n} y_j}$$

2.2 part b.

Consider the following prior distribution.

$$\pi(\lambda_1, \lambda_2) \propto \lambda_1^{a-1} \lambda_2^{b-1} exp\left(-c\lambda_1 - d\lambda_2\right)$$
$$a, b, c, d > 0$$

Find the posterior distribution. Is this a conjugate prior?

answer

$$f(data|\lambda_1, \lambda_2)\pi(\lambda_1, \lambda_2) \propto \lambda_1^{\sum x_i + a - 1} \lambda_2^{\sum y_j + b - 1} exp\left(-\lambda_1(n + c) - \lambda_2(n + d)\right)$$

This is equivalent to a bivariate Gamma distribution, so yes this is a conjugate prior.

2.3 part c.

How would you construct a noninformative prior for λ_1 and λ_2? Are there any values of a, b, c, d that would correspond to a noninformative prior?

answer I would use Jeffrey's prior to construct a noninformative prior. For a single sample with one λ parameter, we have

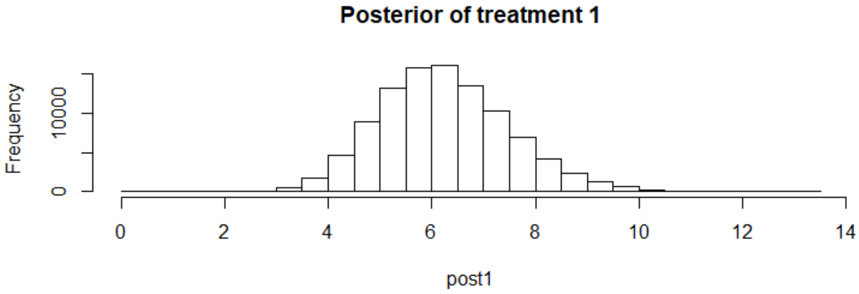
$$lnf(data|\lambda) = -n\lambda + ln(\lambda) \sum x_i - ln(\prod x_i!)$$
$$\frac{\partial^2}{\partial \lambda^2} f(data|\lambda) = -\frac{\sum x_i}{\lambda^2}$$
$$-E[\frac{\partial^2}{\partial \lambda^2} f(data|\lambda)] = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}$$
$$J(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

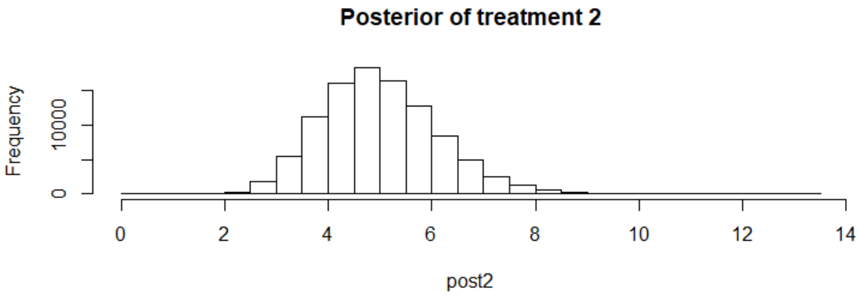
Since the two samples are independent, we can then establish the joint Jeffrey's prior of λ_1 and λ_2 as $\frac{1}{\sqrt{\lambda_1}} \frac{1}{\sqrt{\lambda_2}}$. This is equivalent to $a = b = .5$, $c = d = 0$, although since c and d must be positive, we can instead make c and d very small values.

2.4 part d.

Take $a = b = 1$ and $c = d = .001$. Use the data given above to compute the posterior mean and variance λ_1 and λ_2.

answer By generating 100000 gamma random variables for each treatment group, I get the following histograms.





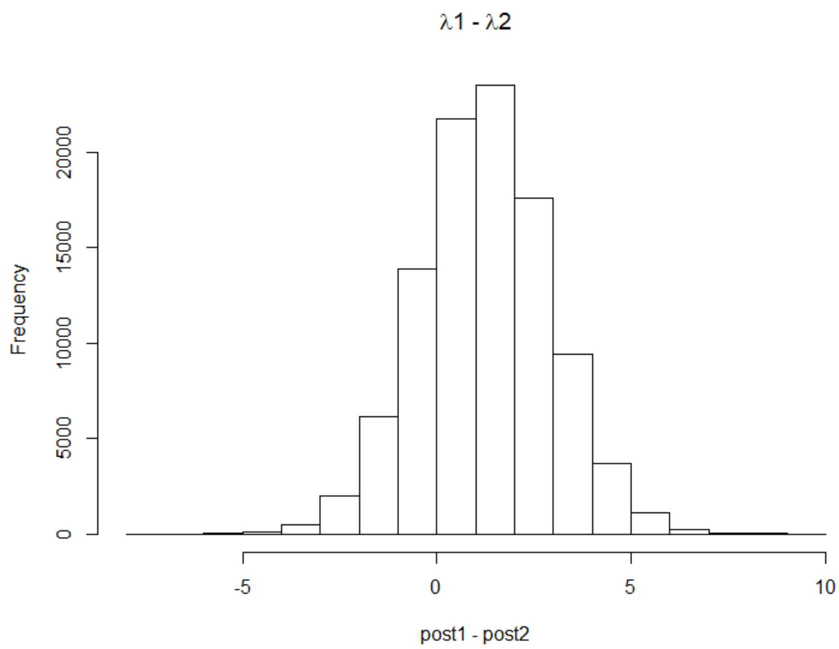
The posterior means and variances are

Treatment	Mean	Variance
1	6.243	1.55
2	4.998	1.245

2.5 part e.

Compute $P(\lambda_1 > \lambda_2 | data)$

answer For this problem, I calculated the equivalent $P(\lambda_1 - \lambda_2 > 0 | data)$.

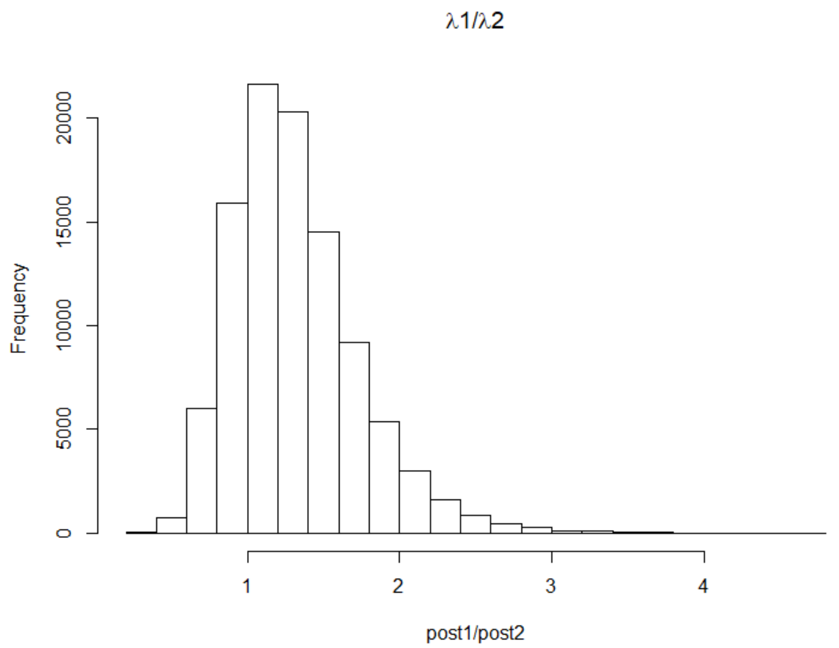


Calculating which of these differences is greater than 0 gives .7736

2.6 part f.

Provide a 95% Bayesian interval for $\theta = \frac{\lambda_1}{\lambda_2}$.

answer Here is the posterior distribution:



The 2.5% and 97.5% quantiles of this posterior distribution are $(.695, 2.299)$
All the code used for this problem is provided below:

```
treatment1 <- c(11, 4, 4, 5)
treatment2 <- c(6, 4, 3, 6)
```

```
post1 <- rgamma(nsim, shape = sum(treatment1) + 1, scale = (4 + .001)^(-1))
post2 <- rgamma(nsim, shape = sum(treatment2) + 1, scale = (4 + .001)^(-1))

mean(post1); mean(post2)
var(post1); var(post2)

par(mfrow = c(2, 1))
hist(post1, main = "Posterior of treatment 1", breaks = seq(0, 13.5, .5))
hist(post2, main = "Posterior of treatment 2", breaks = seq(0, 13.5, .5))
par(mfrow = c(1, 1))

hist(post1 - post2, main = expression(lambda*"1 - "*lambda*"2"))
sum(post1 - post2 > 0)/nsim

hist(post1/post2, main = expression(lambda*"1/"*lambda*"2"))
quantile(post1/post2, c(.025, .975))
```

3 Q3

Derive a noninformative prior for the ratio parameter θ.

answer I did this by reparameterizing the likelihood by $\theta = \frac{\lambda_1}{\lambda_2} \to \theta\lambda_2 = \lambda_1$. This gives us the following likelihood function and the partial derivatives to put into the Fisher information matrix.

$$
\begin{aligned}
f(data|\lambda_1, \lambda_2) &= \frac{1}{\prod x_i! \prod y_j!} exp\left(-n\lambda_2(\theta+1)\right) \theta^{\sum x_i} \lambda_2^{\sum x_i + \sum y_j} \\
\frac{\partial^2}{\partial\theta^2}\ell(\theta, \lambda_2) &= -\frac{\sum x_i}{\theta^2} \\
\frac{\partial^2}{\partial\lambda_2^2}\ell(\theta, \lambda_2) &= -\frac{\sum x_i + \sum y_j}{\lambda_2^2} \\
\frac{\partial^2}{\partial\theta\lambda_2}\ell(\theta, \lambda_2) &= -n \\
\frac{\partial^2}{\partial\lambda_2^2}\ell(\theta, \lambda_2) &= -n \\
I(\theta, \lambda_2) &= \begin{pmatrix} \frac{n\lambda_2}{\theta} & n \\ n & \frac{n(\theta+1)}{\lambda_2} \end{pmatrix}
\end{aligned}
$$

The determinant of the Fisher information matrix is then $\frac{n^2}{\theta}$. We then take the square root of this value to get $J(\theta) \propto \frac{n}{\sqrt{\theta}}$.

If we plug in our known value of $n = 4$, we get $J(\theta) \propto \frac{4}{\sqrt{\theta}}$. This is technically the joint noninformative prior for θ, λ_2, but this noninformative prior only depends on θ.

4 Q4

Do animals bite more during a full moon? The lunar cycle was divided into 10 periods, and the number of bites in four periods is shown in the following tables.

Lunar Day	28,29,1	2,3,4	5,6,7	8,9,10
Bite cases	269	155	142	146

If we let $n_1, \ldots, n_4$ denote the nuumber of admissions to the medical facility in the four periods with $n_1 + \ldots + n_4 = 712$, then $n_1, \ldots, n_4$ follows a multinomial distribution

$$
p(n_1, \ldots, n_4) = \binom{712}{n_1, \ldots, n_4} \theta_1^{n_1} \ldots \theta_4^{n_4}
$$

Furthermore, suppose that the prior distribution for $(\theta_1, \ldots, \theta_4)$ is a Dirichlet distribution with density $\pi(\theta_1, \ldots, \theta_4) \propto \theta_1^2\theta_2\theta_3\theta_4$

4.1 part a.

Find the prior mean and prior variances of $\theta_1, \ldots, \theta_4$

answer We are given a Dirichlet distribution with $\alpha = (3, 2, 2, 2)$. Simulating 100000 random draws from this distribution, we get the following means and variances:

	1	2	3	4
Mean	.333	.221	.223	222
Variance	.022	.017	.017	.017

4.2 part b.

Find the posterior distribution

answer We know that the Dirichlet distribution is a conjugate prior with the multinomial distribution. Therefore, the posterior distribution is also going to be Dirichlet.

$$\begin{aligned} P(p_1, \ldots, p_4|data) &\propto P(data|p_1, \ldots, p_4)\pi(p_1, \ldots, p_4) \\ &\propto p_1^{x_1+a_1-1} \ldots p_4^{x_4+a_1-1} \end{aligned}$$

With our data and prior, the posterior is going to be Dirichlet(271, 156, 143, 147).

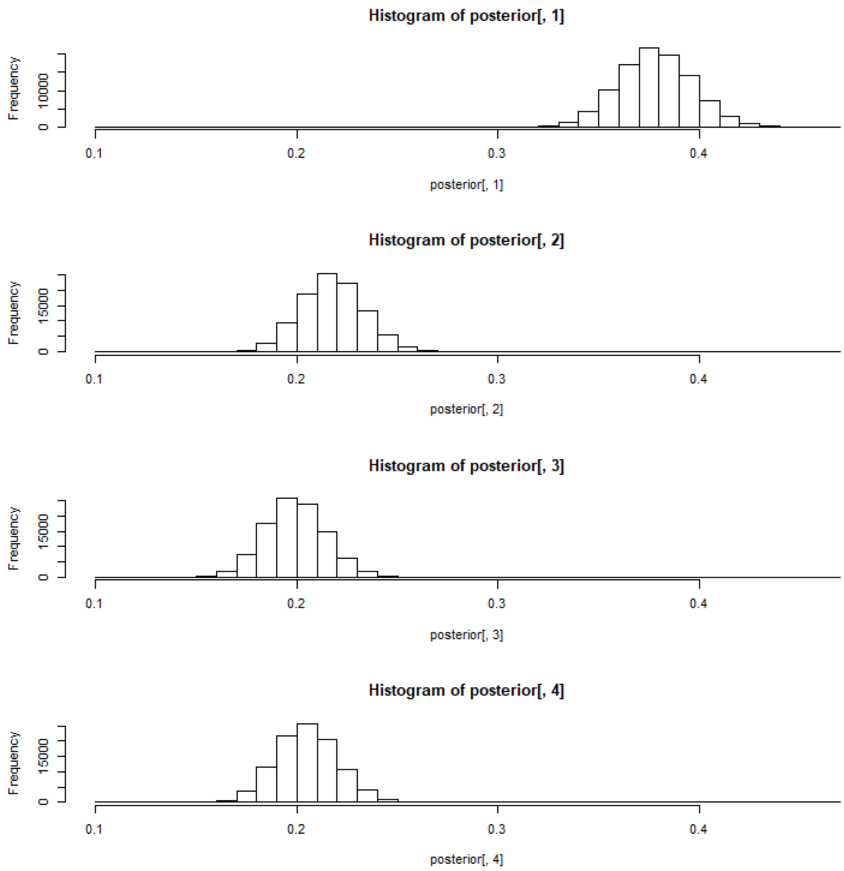
4.3 part c.

Find the posterior means and variances of $\theta_1, \ldots, \theta_4$. What conclusions can you draw?

answer Generating 100000 random draws from our posterior, we get the following means and variances:

	1	2	3	4
Mean	.377	.218	.199	.205
Variance	.0003	.0002	.0002	.0002

We notice that the mean of p_1 posterior distribution is much larger than the other p parameters. Additionally, the variances of these marginal posterior distributions are very small so there is almost no overlap between the posterior of p_1 and the other marginal posterior distributions. This gives us evidence that p_1 is higher from the other parameters and there are more bites on lunar days 28, 29, 1.



4.4 part d.

Find a 95% Bayesian interval for θ_1

answer Using our marginal posterior, we get $(.342, .413)$.

4.5 part e.

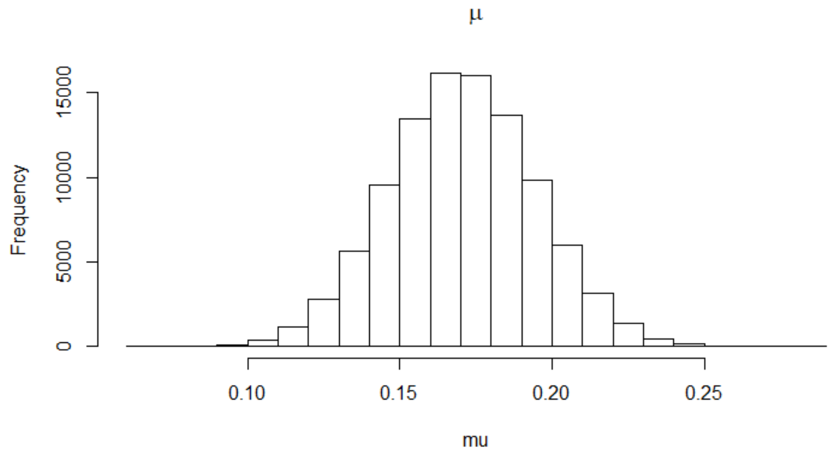
Consider the quantity of interest $\mu = \theta_1 - (\theta_2 + \theta_3 + \theta_4)/3$. Find the posterior mean and variance of μ.

answer Calculating this using the posteriors we simulated, I arrived at a mean of .170 and the variance of .00058

4.6 part f.

Compute $P(\mu > 0|data)$

answer I calculated a probability of 1. An accompanying histogram shows that no random samples from the posterior was lower than .069



All the code used for this question is provided here:

```
library(gtools)
bites <- c(269, 155, 142, 146)

prior <- rdirichlet(nsim, alpha = c(3, 2, 2, 2))
apply(prior, 2, mean)
apply(prior, 2, var)

posterior <- rdirichlet(nsim, alpha = bites + c(3, 2, 2, 2) - rep(1, 4))
apply(posterior, 2, mean)
apply(posterior, 2, var)

par(mfrow=c(4,1))
hist(posterior[,1], breaks = seq(.10, max(posterior[,1])+.01, .01))
hist(posterior[,2], breaks = seq(.10, max(posterior[,1])+.01, .01))
hist(posterior[,3], breaks = seq(.10, max(posterior[,1])+.01, .01))
hist(posterior[,4], breaks = seq(.10, max(posterior[,1])+.01, .01))
par(mfrow=c(1,1))

quantile(posterior[,1], c(.025, .975))

mu <- posterior[,1] - apply(posterior[,c(2, 3, 4)], 1, sum)/3
mean(mu); var(mu)

hist(mu, main = expression(mu))
sum(mu > 0)/nsim
```