Exercises IV

1. To say 'variables are exchangeable' is equivalent to saying that 'they are *conditionally independent*, given a parameter p'. They are not unconditionally independent, i.e. covariance $Cov(X_i, X_j) > 0$. Show this with binary variables X_i which are conditionally independently distributed as Bernoulli(p), and where p has Beta(α, β)-distribution. Hint: Use the formula for covariance below, similar to the formula of total variance, to show $\mathrm{Cov}(X_i, X_j) > 0$:

$$Cov(X_i, X_j) = Cov(E(X_i \mid p), E(X_j \mid p)) + E(Cov(X_i, X_j \mid p))$$

$$= Cov(p, p) + E(0) = Var(p) = \alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)] > 0$$

(This exemplifies the fact that although X_i are independent of each other conditionally, when p is given: $P(X_i \mid p, X_j, j \neq i) = P(X_i \mid p)$, they are not independent when p is unknown. Then, for predicting next X_i we can obviously learn from observing previous X_j. Only **if** p would be perfectly known, observed data would have no role in our predictions).

2. Bayes factor measures the *change* between prior odds and posterior odds, due to data. Posterior probability simply tells the current degree of belief, given all data and priors we had. Assume the model for X as Normal distribution $\mathrm{N}(\mu, 1)$. A null hypothesis is $H_0 : \mu = 0$, versus alternative hypothesis $H_1 : \mu = 2$. Given that we observed $X = 1$, what is Bayes factor? Assume priors $q_0 = P(\mu = 0) = 0.5 = P(\mu = 2) = q_1$, what is posterior probability of each hypothesis? How large should X increase before alternative hypothesis has posterior probability over 90%? What the Bayes factor would then be?

$$BF = \frac{P(X \mid H_0)}{P(X \mid H_1)} = \frac{\exp(-0.5X^2)}{\exp(-0.5(X-2)^2)} = \exp(2 - 2X) = 1 \text{ when } X = 1$$

The Bayes factor tells that, compared to the prior odds ($q_0/q_1 = 1$) an observation of $X = 1$ does not add any information, and the posterior odds remain the same as prior odds. Since posterior odds is $O = P/(1-P)$, we get (if $X = 1$) posterior probability $P = O/(1+O) = 0.5$, the same as prior probability.

Posterior probability of the alternative hypothesis is

$$P(H_1 \mid X) = \frac{q_1 P(X \mid H_1)}{q_1 P(X \mid H_1) + q_0 P(X \mid H_0)} = \frac{\exp(-0.5(X-2)^2)}{\exp(-0.5(X-2)^2) + \exp(-0.5X^2)} > 0.9$$

$$\Leftrightarrow 0.1 \exp(-0.5(X-2)^2) > 0.9 \exp(-0.5X^2)$$

$$\Leftrightarrow \exp(-0.5(X^2 - 4X + 4))/\exp(-0.5X^2) > 9$$

$$\Leftrightarrow \exp(2X - 2) > 9$$

$$\Leftrightarrow X > (\log(9) + 2)/2 \approx 2.098$$

Bayes factor at $X = 2.098$ would be $\exp(2 - 2 * 2.098) \approx 0.11$.

3. Assume variables X_i have Poisson(λ) model, and that the prior of λ is Gamma(α, β). In lecture notes, the posterior distribution of λ was calculated in the case of single observation X. Compute the

posterior when we have $X_1, \dots, X_n$ observations.

$$\pi(\lambda \mid X_1, \dots, X_n) \propto \prod_{i=1}^{n} \lambda^{X_i} \exp(-\lambda) \times \lambda^{\alpha-1} \exp(-\lambda\beta)$$
$$= \lambda^{\sum_i X_i + \alpha - 1} \exp(-\lambda(n+\beta))$$
$$\propto \text{Gamma}(\sum_i X_i + \alpha, n + \beta)$$

4. Solve the posterior predictive distribution for exponential model $\pi(X) = \theta \exp(-\theta X)$, assuming the posterior distribution $\pi(\theta \mid \text{data})$ is represented by Gamma(α, β)-density: $(\beta^\alpha/\Gamma(\alpha))\theta^{\alpha-1}\exp(-\beta\theta)$. In other words, calculate $\int_0^\infty \pi(X \mid \theta)\pi(\theta \mid \alpha, \beta)\mathbf{d}\theta$ with these functions. (In the manipulations, you may need this result of *gamma-functions*: $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$).

$$\pi(X) = \int_0^\infty \theta \exp(-\theta X)(\beta^\alpha/\Gamma(\alpha))\theta^{\alpha-1}\exp(-\beta\theta)\mathbf{d}\theta$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(\beta+X)^{\alpha+1}} \int_0^\infty \underbrace{\frac{(\beta+X)^{\alpha+1}}{\Gamma(\alpha+1)} \theta^{\alpha+1-1} \exp(-(\beta+X)\theta)\,\mathbf{d}\theta}_{\text{Gamma}(\alpha+1,\beta+X)\text{-density}}$$
$$= \frac{\alpha\beta^\alpha}{(\beta+X)^{\alpha+1}} \quad \text{(because integral of probability density = 1)}$$

And if the Gamma(α, β)-density was actually a posterior distribution, resulting from earlier data X_{old}, (from $\exp(\theta)$-distribution), then it must have been of the form Gamma$(\alpha_0 + n, \beta_0 + n\bar{X}_{\text{old}})$, where $\bar{X}_{\text{old}} = \frac{1}{n}\sum_i X_{\text{old},i}$, the mean of earlier data. The posterior predictive distribution would then be

$$\pi(X \mid X_{\text{old}}) = \frac{(\alpha_0+n)(\beta_0+n\bar{X}_{\text{old}})^{(\alpha_0+n)}}{(\beta_0+n\bar{X}_{\text{old}}+X)^{\alpha_0+n+1}} \quad \text{for } X > 0.$$

...And if the prior parameters were $\alpha_0 = \beta_0 = 0$, (improper prior), then it becomes

$$\pi(X \mid X_{\text{old}}) = \frac{n^{(n+1)}(\bar{X}_{\text{old}})^n}{(n\bar{X}_{\text{old}}+X)^{n+1}}$$

5. Intro to Monte Carlo: posterior distributions can be approximated by drawing large number of random draws from the distribution, then drawing histograms. Assume the posterior of p is Beta$(x+1, n-x+1)$, and choose some values for n, x. A sample of 1000 draws is generated in R by `p <- rbeta(1000,x+1,n-x+1)` . With this sample, we could easily visualize the distribution of any transformation $\sqrt{p}$ or p^2, etc, by plotting the histogram of this from the sample, e.g. `hist(sqrt(p))` or `hist(p^2)`. Try this in the computer class with R. **Optional:** if you still have time, open the OpenBUGS-software, locate the menu bar, and click `File`, and `New` to open a window. Write the following lines there

```
model{
p ~ dbeta(alpha,beta); alpha <- x+1; beta <- n-x+1
x <- 7; n <- 20
p2 <- p*p
}
```

Then click `Model` and within that `Specification` to open 'specification tool window'. Click `check model`, then `compile`, then `gen inits`. Click again `Model` and within that `Update` to open 'Update Tool window' and click `update`. See iteration counter running. Click `Inference` from the menu bar and within that `Samples` to open 'Sample Monitor Tool window'. Write `p` in the 'node' and click `set`. Do the same for `p2`. Click again `update` from 'Update Tool'. Go back to 'Sample Monitor Tool' and select `p` from the 'node'. Click `density` to plot the approximated density from the generated sample. Select and plot `p2` too.