## 1.1 Exercises

7. Using product rule, factorize the joint density $\pi(x, y, z)$ into a product of three parts.

$\pi(x, y, z) = \pi(x \mid y, z)\pi(y, z) = \pi(x \mid y, z)\pi(y \mid z)\pi(z)$

With $n$ variables, we would have

$\pi(x_1, \ldots, x_n) = \pi(x_1 \mid x_2, \ldots, x_n)\pi(x_2, \ldots, x_n) = \pi(x_1 \mid x_2, \ldots, x_n) \ldots \pi(x_{n-1} \mid x_n)$ which still would involve many terms with lots of conditions. Yet, at least some terms of the product can be easier to specify than the much more difficult joint distribution directly. To make the modeling tractable, some assumptions of conditional independence are assumed. In the most trivial case we could have the structure: $\pi(x_1, \ldots, x_n, \theta) = \pi(x_1, \ldots, x_n \mid \theta)\pi(\theta) = \prod_i \pi(x_i \mid \theta)\pi(\theta)$ which could represent the 'data model × prior'. In other context, we might factorize $\pi(x_1, \ldots, x_n) = \pi(x_1) \prod_{i=1}^{n-1} \pi(x_{n-i+1} \mid x_{n-i})$ corresponding to assumed Markov chain model of the $x_i$ variables. Each of these structures could be represented as a graph of conditional distributions which are needed for computing the posterior distribution. Using conditional independence, the difficult joint distributions can be factorized into tractable parts which can be specified based on the problem context. Divide and conquer. This is also described as 'reduction of complexity'.

2. Observed bacterial counts (cfu /10 grams, cfu = colony forming units) in 17 samples were as follows:

```
X <- c(0, 0, 0, 0, 5, 3, 0, 0, 70, 0, 0, 0, 8, 0, 0, 3, 0)
```

Assume Poisson model $X_i \sim \text{Poisson}(\lambda)$. Use Gamma$(\alpha, \beta)$ prior for $\lambda$. Compute posterior mean, mode and standard deviation of $\lambda$.

With Gamma$(\alpha, \beta)$ prior, posterior is Gamma$(89 + \alpha, 17 + \beta)$. If $(\alpha, \beta)$ small (when trying to be uninformative), then

$$\text{mean} = \frac{89 + \alpha}{17 + \beta} \approx 5.24, \text{mode} = \frac{89 + \alpha - 1}{17 + \beta} \approx 5.18, \text{SD} = \frac{\sqrt{89 + \alpha}}{17 + \beta} \approx 0.55$$

7. Janne Ahonen and Jakub Janda shared the Four Hills Tournament (Vierschanzentournee, Keski-Euroopan mäkiviikot) championship in 2006. Both scored a total of 1081.5 points from four competitions. Before the tournament, both took part in four other competitions. Their scores from all eight competitions were

```
ahonen <-c(299.7, 255.2, 281.7, 238.0,   270.9, 262.2, 255.4, 293.0)
janda <-c(238.7, 285.6, 287.1, 252.2,   262.6, 264.7, 263.2, 291.0)
```

Assuming a normal model $\text{N}(\mu_i, \sigma_i^2)$ for both jumpers $(i = 1, 2)$ and the uninformative prior $\pi(\mu_i, \sigma_i^2) \propto 1/\sigma_i^2$, a posterior density can be obtained. If both mean and variance are unknown, and the prior is uninformative, can we have realistic estimates of $(\mu_1, \mu_2)$ with these data only? Note: a 95% posterior interval of $\mu_i$ is obtained from $t_{n-1}$ distribution as $\bar{X}_i \pm 1.997 s_i/\sqrt{n}$. Compare the two jumpers using such posterior intervals.

$$\text{Ahonen: } 269.5125 \pm 1.997 * 20.92138/\sqrt{8} = [254.7410, 284.2840]$$

$$\text{Janda: } 268.1375 \pm 1.997 * 18.41831/\sqrt{8} = [255.1333, 281.1417]$$

which are nearly identical intervals for the expected score of each jumper. Based on these intervals, there seems to be such uncertainty regarding both, that we could not conclude one to be better than the other.

8. Assume a multinomial model $X \sim \text{Multi}(N,p)$ with a fairly small number, say $N = 4$. The data could then be e.g. $x = (0,1,0,0,2,0,0,1)$. Using Dirichlet prior, solve the posterior density of $p$ and discuss what would be an uninformative Dirichlet prior for $p$. How your prior will affect the result if the number of categories (dimension of $p$ and $X$) is large, and yet $N$ is small? What if we had prior information about the (unequal) proportions but if we do not wish that prior to dominate the result overwhelmingly?

With $\text{Dir}(\alpha_1, \ldots, \alpha_8)$ prior, the posterior is $\text{Dir}(\alpha_1 + x_1, \ldots, \alpha_8 + x_8)$. If $\alpha_i = 1$ (if we generalize the Bayes-Laplace prior) then the prior equals to sample size of $N_0 = 8$ which is twice the actual observed sample size of $N = 4$. The prior would dominate the result. To make it less influential we could make it correspond to $N_0 = 1$, e.g. setting $\alpha_i = 1/8$. But if we had real prior knowledge about the relative proportions of the categories, we could use those as unequal values of $\alpha_i$, so that $\sum \alpha_i = 1 = N_0$. In this case, the prior would be informative, but it would not dominate the result. The posterior would be mostly driven by data. E.g. posterior mean for $p_1$ would be $\alpha_i/(\sum_i \alpha_i + 4)$ (if $\alpha_i = 1/8$, then $(1/8)/5=1/40$) whereas the maximum likelihood estimate is 0. (But if $\alpha_i = 1$, then $1/(8+4)=1/12$, so the prior effect would be stronger).

5. Using the R-function for calculating posterior probabilities of $N$ in a binomial$(N,r)$ problem with known $X$ and $r$, calculate the 95% 'HPD interval' (although the "interval" is a set of integer values). Hint: use `A<-sort(p,index.return=TRUE)`, then `attach(A)`, and then check what is `x` and `ix`. Operate with `p` and `ix`. Assume the data: `X<- 1;r<- 0.2;M<- 100`.

The R-function for posterior probabilities:

```
p0 <- function(X,N,r){
s <- log(N)
for(i in 1:X-1){
s <- s+log(N-i)
}
s<-s+N*log(1-r)
exp(s)
}
postn <- function(X,N,M,r){
p0(X,N,r)/sum(p0(X,X:M,r))
}
```

Computation of the HPD interval:

```
X<-1; r<-0.2; M<-100;
p <- postn(X,X:M,M,0.2)
A <- sort(p,index.return=TRUE)
attach(A)
sum(p[ix[73:M]])
c(min(ix[73:M]),max(ix[73:M]))
```

3. Assume the prevalences of a disease in two different populations are $p_1$ and $p_2$. A small sample of individuals is observed from both, $(N_1, N_2)$. In these observed samples, $X_1$ and $X_2$ individuals have the disease. Choose some value for $p_1$ and $p_2$. Then choose the sample sizes $N_1, N_2$ and simulate the observed data $X_1$ and $X_2$. Then simulate the posterior (beta)density of $p_1$ and $p_2$ and the difference $q = p_1 - p_2$. Study if there was evidence for $p_1 < p_2$ by computing $P(q < 0 \mid X_1, N_1, X_2, N_2)$.

```
N1 <- 20; N2 <- 20; p1 <- 0.1; p2 <- 0.15
x1 <- rbinom(1,N1,p1)
x2 <- rbinom(1,N2,p2)
pp1 <- rbeta(1000,x1+1,N1-x1+1)
pp2 <- rbeta(1000,x2+1,N2-x2+1)
q <- pp1-pp2
sum(q<0)/1000
par(mfrow=c(2,1))
plot(density(pp1))
plot(density(pp2))
```

7. Continue Example 3.5. Compute by simulation the posterior probability that $\mu_1$ (Ahonen) is larger than $\mu_2$ (Janda). Using only the results of the first 7 competitions, simulate the posterior of $(\mu_1, \sigma_1)$ and $(\mu_2, \sigma_2)$. Then, simulate predicted values $X_1, X_2$ for the last competition. From this predictive distribution (based on 7 competitions), compute the probability that the difference of Tournament total points is less than 1, given the total sum of the previous 3 games of the Tournament. Was the actual final result surprising at that point?

To compute $P(\mu_1 - \mu_2 > 0)$ simulate $\mu_i$ from the marginal posterior distribution, which is a t-distribution. Sample $T_1$ and $T_2$ independently from standard t-distributions with $n - 1 = 7$ degrees of freedom. Then calculate $\mu_i = T_i s_i / \sqrt{n} + \bar{x}_i$, for $i = 1, 2$. Then compute the percentage of Monte Carlo samples where $\mu_1 > \mu_2$.

```
T1 <- rt(1000,7); T2 <- rt(1000,7)
mu1 <- T1*20.92138/sqrt(8)+269.5125
mu2 <- T2*18.41831/sqrt(8)+268.1375
sum(mu1>mu2)/1000
```

This seems to show about 57% probability for the expected score of Ahonen to be larger than that of Janda. If only sports was a theoretical game!

To simulate the joint posterior of $(\mu_i, \sigma_i)$ you can recall that the marginal posterior of $\sigma_i^2$ was Scaled-Inv-Gamma$(n - 1, s_i^2)$ with $s_i^2 = \frac{1}{n-1} \sum_j (x_{i,j} - \mu_i)^2$. From the table of distributions we see this is the same as Inv-Gamma$(\frac{n-1}{2}, \frac{n-1}{2} s_i^2)$. So that $\tau_i = 1/\sigma_i^2$ has distribution Gamma$(\frac{n-1}{2}, \frac{n-1}{2} s_i^2)$ which can be used for Monte Carlo. Then, the conditional density of $\mu_i$ is $N(\bar{x}_i, \sigma_i^2/n)$ which can be used for sampling $\mu_i$. From the obtained joint sample of $(\mu_i, \sigma_i^2)$ we can generate the predicted next outcome, using simply $N(\mu_i, \sigma_i^2)$.

```
s1 <- sd(ahonen[1:7]); m1 <- mean(ahonen[1:7])
s2 <- sd(janda[1:7]); m2 <- mean(janda[1:7])
```

```
tau1 <- rgamma(1000,(7-1)/2,((7-1)/2)*s1*s1)
tau2 <- rgamma(1000,(7-1)/2,((7-1)/2)*s2*s2)
sig21 <- 1/tau1; sig22 <- 1/tau2
mu1 <- rnorm(1000,m1,sqrt(sig21))
mu2 <- rnorm(1000,m2,sqrt(sig22))
ahonen8 <- rnorm(1000,mu1,sqrt(sig21))
janda8 <- rnorm(1000,mu2,sqrt(sig22))
diff <- (sum(ahonen[5:7])+ahonen8)-(sum(janda[5:7])+janda8)
sum(abs(diff)<1)/1000
```

This seems to give about 2% probability for the difference to be less than one point. Surely, it is small, but much larger than the probability of winning e.g. the biggest prize in lotto. Maybe a small bet might have been a good idea for a gambler?

11. Simulate from $\pi(p_1, p_2)$ defined as uniform over the plane $[0, 1] \times [0, 1]$ restricted in the area where $p_1 < p_2$. Hence, $\pi(p_1, p_2) = 2 \times 1_{\{p_1 < p_2\}}(p_1, p_2)$. Implement (1) a rejection sampler and (2) alternating sampler, and compare the results.

Rejection sampler:

```
p1u <- runif(1000,0,1); p2u <- runif(1000,0,1)
p1 <- p1u[p1u<p2u]; p2 <- p2u[p1u<p2u]
plot(p1,p2)
```

Alternating (Gibbs) sampler:

```
p1[1] <- 0.5; p2[1] <- 0.6
for(i in 2:1000){
p1[i] <- runif(1,0,p2[i-1])
p2[i] <- runif(1,p1[i],1)
}
plot(p1,p2)
```

12. Assume the model $Y_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$ and the prior $\pi(\mu, \sigma^2) \propto 1/\sigma^2$. Verify that the posterior is then $\pi(\mu, \sigma^2 \mid Y_{1,\ldots,n}) \propto (\sigma^2)^{-(n/2+1)} \exp(-0.5 \sum (Y_i - \mu)^2 / \sigma^2)$. If $\sigma^2$ is assumed as known, this simplifies to the case where $\pi(\mu \mid \sigma^2, Y_{1,\ldots,n}) = \mathrm{N}(\bar{Y}, \sigma^2/n)$. Verify from the table of distributions that: if $\mu$ is assumed as known, we get $\pi(\sigma^2 \mid \mu, Y_{1,\ldots,n}) = \mathrm{Inv\text{-}Gamma}(n/2, 0.5 \sum (Y_i - \mu)^2)$. In other words, for $\tau = 1/\sigma^2$: $\pi(\tau \mid \mu, Y_{1,\ldots,n}) = \mathrm{Gamma}(n/2, 0.5 \sum (Y_i - \mu)^2)$. Construct a Gibbs sampler (e.g. in R) and simulate joint posterior for $\mu, \sigma^2$, with some data $Y_{1,\ldots,n}$ you have first generated with some selected 'true values' $\mu_0, \sigma_0^2$.

This is nearly the same as the example with Ahonen & Janda, except now implemented as a Gibbs sampler instead of the 'direct' Monte Carlo sampler.

```
install.packages("gplots")
library(gplots) # these are for using hist2d
mu<-5; sigma<-10; # chosen 'true' parameters
```

```
n<-100;  # chosen number of observations
mu.initial<-0; sigma.initial<-1 # initials for MCMC
# generate 100 observations with 'true' parameters:
y<-rnorm(n, mu, sigma) #here sigma, not sigma^2
mup<-numeric(10000); sigmap<-numeric(10000); taup<-numeric(10000)
mup[1]<-mu.initial;
sigmap[1]<-sigma.initial;
taup[1]<-1/sigma.initial
for(i in 2:10000){
# sample mup[i] conditionally on sigmap[i-1]:
mup[i] <- rnorm(1, mean(y), sigmap[i-1]/n)
# sample taup[i] conditionally on mup[i]:
taup[i]<- rgamma(1, n/2, 0.5*sum((y-mup[i])^2))
sigmap[i] <- 1/taup[i] }

hist2d(mup, sigmap)
# or to get a simple scatterplot without hist2d:
plot(mup,sigmap,cex=0.5)
```

The posterior of $(\mu, \sigma^2)$ should be reasonably focused at the true values $(5, 100)$. This happens easily with enough many observations ($n = 100$ is fairly sufficient), but with smaller sample size we naturally would have more uncertainty, and the prior in this case is not informative.