T-61.5040 Oppivat mallit ja menetelmät
T-61.5040 Learning Models and Methods
Pajunen, Viitaniemi

Solutions to exercise 7, 2.3.2007

Problem 1. i) According to the problem statement, the coins are independent. The problem statement must be interpreted in such way that this is the case. The independence can be achieved by regarding a and b as known constants (for unknown a and b this would not be the case). The following probabilities (in subproblem i)) include implicitly conditioning on these values.

The likelihood is a Binomial distribution, so

$$p(y|\theta_i) = Bin(y|n, \theta_i)$$

The posterior is

$$p(\theta_i|y_i) \propto Bin(y_i|n, \theta_i)Beta(\theta_i|a, b) \propto \theta_i^{y_i}(1-\theta_i)^{n-y_i}\theta_i^{a-1}(1-\theta_i)^{b-1}$$

The posterior is therefore $Beta(\theta_i|a + y_i, b + n - y_i)$

ii) Now the likelihood is $Bin(y = y_1 + y_2|2n, \theta)$ and the prior is $Beta(\theta|a, b)$. The same calculation as above gives

$$p(\theta|y_1, y_2) = Beta(\theta|a + y_1 + y_2, b + 2n - y_1 - y_2)$$

iii) $p(\theta_1, \theta_2, a, b|y_1, y_2) = p(\theta_1, \theta_2|a, b, y_1, y_2)p(a, b|y_1, y_2)$

The first term is simply the product of two posteriors from part i), so

$$p(\theta_1, \theta_2|a, b, y_1, y_2) = p(\theta_1|y_1, a, b)p(\theta_2|y_2, a, b) = Beta(\theta_i|a+y_1, b+n-y_1)Beta(\theta_2|a+y_2, b+n-y_2).$$

The term $p(a, b|y_1, y_2)$ is more difficult. To compute it, use the product rule to obtain

$$p(\theta_1, \theta_2, a, b|y_1, y_2) = p(\theta_1, \theta_2|a, b, y_1, y_2)p(a, b|y_1, y_2)$$
$$\implies p(a, b|y_1, y_2) = p(\theta_1, \theta_2, a, b|y_1, y_2)/p(\theta_1, \theta_2|a, b, y_1, y_2)$$

Here the term $p(\theta_1, \theta_2|a, b, y_1, y_2)$ was just computed above (Beta times Beta). The term $p(\theta_1, \theta_2, a, b|y_1, y_2)$ can be computed as

$$p(\theta_1, \theta_2, a, b|y_1, y_2) \propto p(y_1|\theta_1, a, b)p(y_2|\theta_2, a, b)p(\theta_1, \theta_2|a, b)p(a, b)$$
$$= Bin(y_1|n, \theta_1)Bin(y_2|n, \theta_2)Beta(\theta_1|a, b)Beta(\theta_2|a, b)Exp(a|1)Exp(b|1)$$

and thus $p(a, b|y_1, y_2)$ can be computed: The denominator has the product of Beta distributions and the numerator has the product of Binomial and Beta distributions. The θ_i terms cancel out. Also the Binomial constants $\binom{n}{y_i}$ can be dropped. The resulting distribution is

$$p(a, b|y_1, y_2) \propto Exp(a|1)Exp(b|1)\frac{[\Gamma(a+b)]^2\Gamma(a+y_1)\Gamma(b+n-y_1)\Gamma(a+y_2)\Gamma(b+n-y_2)}{[\Gamma(a)\Gamma(b)\Gamma(a+b+n)]^2}$$

This problem demonstrates how to do Bayesian Inference on hierarchical data. In part i) the problem splits into two subproblems, because the prior parameters a, b are known and observing $y_1 = 12$ gives no information about θ_2. But in part iii) observing $y_1 = 12$ gives information about the values a and b, which then affect θ_2.

Problem 2.

i) The model is $p(y|\mu, \sigma^2) = N(y|\mu, \sigma^2)$. In last week's exercises we showed that the Jeffrey's prior for the mean of a Normal distribution is constant, and for the variance it is $p(\sigma^2) \propto \sigma^{-2}$. Thus the product of Jeffrey's priors is now $p(\mu, \sigma^2) \propto \sigma^{-2}$. The Bayes' theorem gives

$$p(\mu, \sigma^2|y) \propto p(y|\mu, \sigma^2)p(\mu, \sigma^2) \propto \sigma^{-1}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)\sigma^{-2} = \sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

ii) The conditional posterior $p(\mu|\sigma^2, y)$ answers the question "What is the mean μ, when data y is observed and the variance σ^2 is known?". This was answered last week for the case of normal data model with known variance and a normal prior for the mean. Now we can regard the constant prior of μ as an infinitely flat normal distribution. The posterior is then a normal distribution with mean given by a weighted average of prior mean and data. The weights are the prior precision and the data precision σ^{-2}. The uniform prior has zero precision and thus the posterior mean is y. The posterior precision is the sum of prior and data precisions. Again, prior precision is zero so the posterior variance is σ^2. So

$$p(\mu|\sigma^2, y) = N(\mu|y, \sigma^2).$$

iii) Write the integral explicitly as

$$p(\sigma^2|y) = \int p(\mu, \sigma^2|y)d\mu$$
$$\propto \int \sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)d\mu$$
$$= \sigma^{-3}\sqrt{2\pi\sigma^2}\int \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)d\mu$$
$$= \sigma^{-2}\sqrt{2\pi} \propto \sigma^{-2}$$

and thus the posterior of σ^2 is of the same form as the prior.

iv)

$$p(\mu|y) \propto \int_0^\infty \sigma^{-3}\exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)d\sigma^2$$

Substitute $z = \frac{(y-\mu)^2}{2\sigma^2} = A\sigma^{-2}$. Then the integration limits are switched, and

$$dz = -A\sigma^{-4}d\sigma^2 \Rightarrow -A^{-1}\sigma^4 dz = d\sigma^2.$$

Also $z^{-1/2} = A^{-1/2}\sigma$ Then the integral is

$$\begin{aligned} p(\mu|y) &\propto \int_0^\infty \sigma^{-3} \exp(-z) A^{-1}\sigma^4 dz \\ &= \int_0^\infty A^{-1}\sigma \exp(-z) dz \\ &= A^{-1/2} \int z^{-1/2} \exp(-z) dz \\ &= A^{-1/2}\Gamma(1/2) \end{aligned}$$

The Gamma integral is constant with respect to μ, so the posterior is

$$p(\mu|y) \propto A^{-1/2} = \left[\frac{(y-\mu)^2}{2}\right]^{-1/2} \propto \frac{1}{|y-\mu|}$$

Problem 3.

i) We are estimating the unknown mean θ_i of a Normal distribution with a known variance σ^2. The prior for θ_i is $N(\mu, \tau^2)$ which is known. The result was obtained before (Exercises 6, Problem 1) and is

$$p(\theta_1|\mu, \sigma, \tau, D) = N\left(\theta_1 \left| \frac{\mu/\tau^2 + (\sum x_i)/\sigma^2}{1/\tau^2 + n/\sigma^2}, (1/\tau^2 + n/\sigma^2)^{-1}\right.\right)$$

Similarly for θ_2 (in which case the number of observations is m).

ii) Now we are estimating the unknown mean of $N(\mu, \tau^2)$ when τ is known. The "data"are the known values θ_1, θ_2. Since μ has zero prior precision (infinite variance), the result is

$$p(\mu|\theta_1, \theta_2, \sigma, \tau, D) = N\left(\mu|(\theta_1 + \theta_2)/2, \tau^2/2\right)$$

iii) This time the variance σ^2 is unknown, but the mean is known for each observation. The prior is $p(\sigma^2) \propto \sigma^{-2}$. This can be written as $p(\sigma^2) = IG(\sigma^2|0, 0)$. Then use the hint given in the problem to compute

$$p(\sigma^2|\theta_1, \theta_2, \mu, \tau, D) = IG\left(\sigma^2|(n+m)/2, (n+m)v/2\right)$$

where

$$v = \frac{1}{n+m}\left(\sum_i (x_i - \theta_1)^2 + \sum_j (y_j - \theta_2)^2\right)$$

iv) Again, τ^2 is the unknown variance and μ is the known mean of a Normal distribution. The "data"is θ_1, θ_2, both known. The prior for τ^2 is $p(\tau^2) \propto (\tau^2)^{-1/2}$. Non-rigorously this is $p(\tau^2) = IG(\tau^2| -1/2, 0)$. Then the posterior is

$$p(\tau^2|\theta_1, \theta_2, \mu, \sigma, D) = IG\left(\tau^2 \left| 1/2, \frac{1}{2}[(\theta_1 - \mu)^2 + (\theta_2 - \mu)^2]\right.\right).$$

Problem 4.

The posterior

$$p(\theta|y) = \int p(\theta, \sigma_1^2, \dots, \sigma_n^2|y) d\sigma_1^2 \dots d\sigma_n^2$$

requires the joint posterior $p(\theta, \sigma_1^2, \dots, \sigma_n^2|y)$. It is

$$p(\theta, \sigma_1^2, \dots, \sigma_n^2|y) \propto \prod_i p(y_i|\theta, \sigma_i^2) p(\sigma_i^2) = \prod_i N(y_i|\theta, \sigma_i^2) p(\sigma_i^2) = \prod_i G_i.$$

The term G_i is

$$G_i \propto \sigma_i^{-8} \exp\left(-1/2\sigma_i^{-2}(y_i - \theta)^2\right) \exp(-2\sigma_i^{-2}).$$

Each G_i contains just the parameters σ_i^2 and θ, so to integrate out the variances, we can do it term by term:

$$J_i = \int G_i d\sigma_i^2 \propto \int_0^\infty \sigma_i^{-8} \exp\left(-\sigma_i^{-2}(\frac{1}{2}(y_i - \theta)^2 + 2)\right) d\sigma_i^2$$

Let us change variables by setting $z = \sigma_i^{-2}\left[\frac{1}{2}(y_i - \theta)^2 + 2\right]$. For the differentials then

$$dz/d\sigma_i^2 = -\sigma_i^{-4}\left[\frac{1}{2}(y_i - \theta)^2 + 2\right]$$

and the integration limits will change, too. Substituting these into the integral we get

$$J_i = \int_\infty^0 -\exp(-z)\sigma_i^{-4}[]^{-1} dz = \int_0^\infty \exp(-z)\sigma_i^{-2}[]^{-2} z dz = \int_0^\infty \exp(-z)[]^{-3} z^2 dz,$$

where we have used the shorthand $[] = \left[\frac{1}{2}(y_i - \theta)^2 + 2\right]$. The term $[]$ does not depend on σ_i^2, so it can be taken out of the integral. The rest is a Gamma integral $\int_0^\infty z^2 \exp(-z) dz$, which equals $\Gamma(3) = 2!$, independent of θ. The posterior of θ is then

$$p(\theta|y) \propto \prod_i \left[\frac{1}{2}(y_i - \theta)^2 + 2\right]^{-3}.$$

Given the data and $\theta = 0$, the posterior value is $2^{-15}10^{-3} \approx 3 \cdot 10^{-8}$, and for $\theta = 1$ it is $[5/2]^{-15}[13/2]^{-3} \approx 4 \cdot 10^{-9}$. Therefore $\theta = 0$ has a higher posterior value.

For comparison, we also determine whether $\theta = 0$ or $\theta = 1$ results in larger value of likelihood $p(y|\theta, \sigma^2)$ if the variance σ^2 is constant for all observations ($\sigma_i^2 = \sigma^2$). The likelihood is a normal distribution and by the symmetry of the distribution around its mean we can find the maximum likelihood estimate and see whether it is closer to $\theta = 0$ or $\theta = 1$. The likelihood is

$$\prod_i p(y_i|\theta, \sigma^2) \propto \prod_i \exp\left(-\frac{1}{2}(y_i - \theta)\sigma^{-2}\right) = \exp\left(-\frac{1}{2}\sum_i (y_i - \theta)^2 \sigma^{-2}\right)$$

whose maximum is found at

$$\frac{\partial}{\partial\theta}\left[-\frac{1}{2}\sigma^{-2}\sum_i (y_i-\theta)^2\right] = 0 \Rightarrow \theta = \frac{1}{n}\sum_i y_i$$

which equals the mean 4/6 of the observations. This is closer to $\theta = 1$, so the likelihood (or posterior probability with constant prior) is higher for $\theta = 1$.

Comments: this is an example of a multivariate model which can be solved and marginalized in closed form. It also illustrates the flexibility of Bayesian inference: we could easily allow the variance to depend on the sample y_i. With a suitable prior for σ_i^2, the result is a posterior which has some robustness against outliers. This means that the single value $y = 4$ did not make the more probable θ equal one, as opposed to the standard model with fixed variance.