DD2434/FDD3434 Machine Learning, Advanced Course
Module 1 Exercise Solutions

November 2022

1 Conjugate priors

1.1: Let $X = (X_1, .., X_N)$ be i.i.d. where $X_n|m, \theta \sim Binomial(m, \theta)$ and $\theta \sim Beta(\alpha, \beta)$. Show that the posterior $p(\theta|X, m)$ follows a Beta-distribution, i.e. that the Beta is conjugate prior to the Binomial with known m. What are the parameters of the posterior? Compare with the Wikipedia Conjugate prior table.

Solution: We only need the posterior up to a multiplicative constant, in order to rewrite the full pdf of the posterior to the same form as the prior. First, we employ the Bayesian theorem and discard $p(X)$ since it is a multiplicative constant w.r.t. our variable of interest, θ:

$$\begin{aligned} p(\theta|X) = \frac{p(X,\theta)}{p(X)} \propto p(X,\theta) = p(X|\theta)p(\theta) = \prod_{n=1}^{N} p(X_n|\theta) p(\theta) \\ = \{\text{Wikipedia for pdf/pmfs}\} = \prod_{n=1}^{N} \left[\binom{m}{X_n} \theta^{X_n} (1-\theta)^{m-X_n} \right] \cdot \frac{\Gamma(\alpha+\beta}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \end{aligned}$$

Here, we can further discard factors that don't depend on θ since we know we only have to find the posterior up to a normalizing constant. So we simplify further:

$$\propto \prod_{n=1}^{N} \left[\theta^{X_n} (1-\theta)^{m-X_n} \right] \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

Now, we aim to make the full expression in the form of the red part, since those factors correspond to the prior. We proceed by rewriting the likelihood (blue) factors in a form similar to the prior (red) factors by applying the product to each blue factor:

$$= \theta^{\sum_n X_n} (1-\theta)^{\sum_n (m-X_n)} \cdot \theta^{\alpha-1} (1-\theta))^{\beta-1}$$

Grouping similar terms together,

$$= \theta^{\sum_n X_n + \alpha - 1} (1-\theta)^{\sum_n (m-X_n) + \beta - 1}$$

By comparing to the prior, we can easily conclude that $\theta|X \sim Beta\big(\alpha + \sum_n X_n, \beta + \sum_n (m - X_n)\big)$.

1.2: Let $D = (d_1, ..., d_N)$ be i.i.d. with $d_n|\lambda \sim Poisson(\lambda)$ and $\lambda \sim Gamma(\alpha, \beta)$. Show that the posterior $p(\lambda|D)$ follows a Gamma-distribution, i.e. that the Gamma is conjugate prior to the Poisson distribution. What are the parameters of the posterior? Compare with the Wikipedia Conjugate prior table.

Solution:

$$p(\lambda|D) = \frac{p(D,\lambda)}{p(D)} \propto p(D,\lambda) = p(D|\lambda)p(\lambda) = \prod_{n=1}^{N} p(d_n|\lambda)p(\lambda)$$
$$= \{\text{Wikipedia for pdf/pmfs}\} = \prod_{n=1}^{N} \frac{\lambda^{d_n} e^{-\lambda}}{d_n!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

We discard the factors that don't depend on λ:

$$\prod_{n=1}^{N} \lambda^{d_n} e^{-\lambda} \cdot \lambda^{\alpha-1} e^{-\beta\lambda} \tag{1}$$

Now, we make the full expression in the form of the red part (the prior):

$$(e^{-\lambda})^N \lambda^{\sum_{n=1}^N d_n} \cdot \lambda^{\alpha-1} e^{-\beta\lambda} = \{\text{regrouping the factors }\} = \lambda^{\sum_{n=1}^N d_n} \lambda^{\alpha-1} e^{-N\lambda} e^{-\beta\lambda}$$
$$= \lambda^{\sum_{n=1}^N d_n + \alpha - 1} e^{-(\beta+N)\lambda}$$

By comparing to equation 1 we can easily conclude that $\lambda|D \sim Gamma(\sum_{n=1}^N d_n + \alpha, \beta + N)$.

1.3: Let $X = (X_1, ..., X_N)$ be i.i.d. with $X_n|\mu, \tau \sim Normal(\mu, \frac{1}{\tau})$ and $(\mu, \tau) \sim NormalGamma(\mu_0, \lambda, \alpha, \beta)$. Show that the posterior $p(\mu, \tau|X)$ follows a NormalGamma-distribution, i.e. that the NormalGamma is conjugate prior to the Normal distribution with unknown mean and precision. What are the parameters of the posterior? Compare with the Wikipedia Conjugate prior table.

Solution: We jump into the log of the likelihood and prior, and remove additive constants with respect to μ and τ:

$$\log p(\mu, \tau|X) \stackrel{\pm}{=} \sum_{n=1}^{N} \log p(X_n|\mu, \tau) + \log p(\mu, \tau) = \{\text{Wikipedia for pdfs}\}$$
$$= \sum_{n=1}^{N} \log[\frac{\sqrt{\tau}}{\sqrt{2\pi}} \cdot e^{-\frac{\tau}{2}(x_n - \mu^2)}] + \log[\frac{\beta^\alpha \sqrt{\lambda}}{\Gamma(\alpha)\sqrt{2\pi}} \tau^{(\alpha - \frac{1}{2})} \cdot e^{-\beta\tau} \cdot e^{-\frac{\lambda\tau}{2}(\mu - \mu_0)^2}]$$
$$\stackrel{\pm}{=} \sum_{n=1}^{N} \frac{1}{2} \log \tau - \frac{\tau}{2}(x_n - \mu)^2 + (\alpha - \frac{1}{2}) \log \tau - \beta\tau - \frac{\lambda\tau}{2}(\mu - \mu_0)^2 (\star)$$

Again, our goal is to rewrite the expression in the form of the prior (red). We continue our simplification by pushing the sum to act on the terms of the likelihood and start grouping terms by their dependence on μ and τ (see hand-written notes next page):

$$(\ast) = \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}\left(x_n^2 - 2x_n\mu + \mu^2\right) + \left(\alpha - \tfrac{1}{2}\right)\log\tau - \beta\tau - \frac{\lambda\tau}{2}\left(\mu^2 - 2\cdot\mu\,\mu_0 + \mu_0^2\right)$$

$$= \left(\frac{N}{2} + \alpha - \frac{1}{2}\right)\log\tau - \beta\tau - \frac{\tau}{2}\left(\mu^2(N+\lambda) - 2\cdot\mu\left(\sum_{n=1}^{N}x_n + \mu_0\lambda\right) + \left(\sum_{n=1}^{N}x_n^2 + \lambda\mu_0^2\right)\right)$$

$$= \left(\frac{N}{2} + \alpha - \frac{1}{2}\right)\log\tau - \beta\tau - \frac{\tau}{2}(N+\lambda)\underbrace{\left(\mu^2 - \frac{2\cdot\mu}{(N+\lambda)}\left(\sum^{N} x_n + \mu_0\lambda\right) + \frac{\sum x_n^2 + \lambda\mu_0^2}{N+\lambda}\right)}_{=:\,\bullet}$$

We want $\bullet$ in form of prior in $(\ast)$, i.e. $\frac{\tau^*}{2}(\mu - \mu^*)^2$

In order to get $(\mu - \mu^*)^2$ we must complete the square of $\bullet$

We do this in 2 steps:

1. Take out $\frac{\sum x_n^2 + \lambda\mu_0^2}{N+\lambda}$ and match it with "$-\beta\tau$" term

2. Trick: add and subtract term needed to complete the square

1 $\Rightarrow$ $(\ast) = \left(\frac{N}{2} + \alpha - 1\right)\log\tau - \tau\left(\beta + \frac{1}{2}\left(\sum x_n^2 + \lambda\mu_0^2\right)\right) - \frac{\tau}{2}(N+\lambda)\left(\mu^2 - 2\mu\cdot\underbrace{\frac{(\sum x_n + \mu_0\lambda)}{N+\lambda}}_{c}\right)$

2 The term c^2 is the missing component to complete the square. Notice that it is also multiplied with $-\frac{\tau}{2}(N+\lambda)$. So we add and subtract: $\frac{\tau}{2}(N+\lambda)\cdot c^2$ to $(\ast)$ (i.e. adding zero) and pass the positive term to the $\tau(\beta + \frac{1}{2}\ldots)$-term and the negative term to complete the square:

$$(\ast) = \left(\frac{N}{2} + \alpha - 1\right)\log\tau - \tau\left(\beta + \frac{1}{2}\left(\sum x_n^2 + \lambda\mu_0^2\right)\right) + \frac{\tau}{2}(N+\lambda)\cdot c^2$$

$$- \frac{\tau}{2}(N+\lambda)\left(\mu^2 - 2\mu\cdot c + c^2\right) =$$

$$\underbrace{\left(\frac{N}{2} + \alpha - 1\right)}_{\alpha^*}\log\tau - \tau\underbrace{\left(\beta + \frac{1}{2}\left(\sum x_n^2 + \lambda\mu_0^2 - \frac{(\sum x_n + \mu_0\lambda)^2}{N+\lambda}\right)\right)}_{\beta^*} - \frac{\tau}{2}\underbrace{(N+\lambda)}_{\lambda^*}\left(\mu - \underbrace{\frac{(\sum x_n + \mu_0\lambda)}{N+\lambda}}_{\mu_0^*}\right)^2$$