Solution to Homework 1
Bayesian Methods

Anna Pavlova

Problem 1

a = apple, o = orange, l = lime, r = red, b = blue, g =green Probability of selecting an apple

$$P(a) = P(a|r)P(r) + P(a|b)P(g) + P(a|b)P(b) = 0.3 \cdot 0.2 + 0.5 \cdot 0.2 + 0.3 \cdot 0.6 = 0.34$$

If we observe that the selected fruit is in fact an orange, the probability that it came from the green box will be:

$$P(g|o) = \frac{P(o|g)P(g)}{P(o)} = \frac{0.18}{0.36} = \frac{1}{2}$$

$$P(o) = 0.4 \cdot 0.2 + 0.5 \cdot 0.2 + 0.3 \cdot 0.6 = 0.36$$

Problem 2

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu^2)}{2\sigma^2}}$$

1. If $n = m$: $\mathbb{E}[x_n^2] = \mu^2 + \sigma^2$
2. If $n \neq m$: $\mathbb{E}[x_n x_m] = \mathbb{E}[x_n]\mathbb{E}[x_m] = \mu^2$

Let us derive log likelihood

$$\log P(x) = -\frac{1}{2} N \log 2\pi\sigma^2 - \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{2\sigma^2}$$

μ:

$$\frac{\partial \log P(x)}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \frac{\partial (x_i - \mu)^2}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \frac{\partial (x_i^2 - 2x_i\mu + \mu^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (\mu - x_i) = 0$$

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

σ

$$\frac{\partial \log P(x)}{\partial (\frac{1}{\sigma^2})} = -\frac{-N\sigma^2}{2} - \frac{1}{2} \sum_{i=1}^{N} (x_i - \mu)^2 = 0$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Problem 3

For $x_1, \ldots, x_n \sim \text{Poiss}(x|\lambda)$ we get Jeffrey's prior

$$p(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{\mathbb{E}\left[\left(\frac{d}{d\lambda}\log p(x|\lambda)\right)^2\right]} = \sqrt{\mathbb{E}\left[\left(\frac{x-\lambda}{\lambda}\right)^2\right]} =$$

$$= \sqrt{\sum_{n=1}^{\infty} e^{-\lambda}\frac{\lambda^n(n-\lambda)^2}{n!\lambda^2}}$$

For the series term we have $(n-\lambda)^2 = n^2 - 2n\lambda + \lambda^2$, so

$$p(\lambda) \propto \sqrt{\frac{1}{\lambda^2}\mathbb{E}x^2 - \frac{2}{\lambda}\mathbb{E}x + 1}.$$

The mean and variance of Poisson distribution $\mathbb{E}x^2 = \lambda + \lambda^2, \mathbb{E}x = \lambda$, so that leaves $p(\lambda) \propto \sqrt{\frac{1}{\lambda}}$ This is a special case of Gamma distribution which is conjugate prior for Poisson

$$p(\lambda) \sim \mathcal{G}(\lambda|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

Posterior distribution for Poisson is Gamma distribution

$$p(\lambda|x_1,\ldots,x_n) \propto p(x|\lambda)p(\lambda) = \mathcal{G}\big(\lambda|\alpha + \sum_{k=1}^{n} x_k, \beta + n\big).$$

So since Jeffrey's prior is a special case of Gamma $\sqrt{\frac{1}{\lambda}} = \mathcal{G}(\lambda|\alpha = 1/2, \beta = 0)$, for a posterior we have

$$p(\lambda|x_1,\ldots,x_n) = \mathcal{G}\big(1/2 + \sum_{k=1}^{n} x_k, n\big)$$

Problem 4

The variables in the data set $\mathbf{X} = \{X_1, \ldots, X_n\}$ are normally distributed $\mathcal{N}(x|\mu,\tau^{-1})$. Bayes theorem reads

$$\text{Posterior}(\mu,\tau|\mathbf{X}) \propto \text{Likelihood}(\mathbf{X}|\mu,\tau)\cdot\text{Prior}(\mu,\tau)$$

We put **proportional** instead of **equals** here because this holds up to normalization parameter that does not affect the functional dependence. Since the variables in $\mathbf{X}$ are i.i.d., the likelihood decomposes to a product of same distributions

$$\text{L}(\mathbf{X}|\mu,\tau) = \prod_{i=1}^{n}\text{L}(x_i|\mu,\tau)$$

with each $\text{L}(x_i|\mu,\tau) = \mathcal{N}(x_i|\mu,\tau^{-1})$. Thus, up to some power of $\sqrt{2\pi}$ we have

$$\text{L}(\mathbf{X}|\mu,\tau) \propto \prod_{i=1}^{n}\sqrt{\tau}\exp[-\tau(x_i-\mu)^2/2] = \tau^{n/2}\exp[-\tfrac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2] \propto$$

$$\propto \tau^{n/2} \exp[-\tfrac{\tau}{2} \sum_{i=1}^{n} \left((x_i - \bar{x})^2 + (\bar{x} - \mu)^2\right)]$$

which (if we denote $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – the mean, $\sigma = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$ – the variance) gives

$$\mathrm{L}(\mathbf{X}|\mu,\tau) \propto \tau^{n/2} \exp[-\tfrac{\tau}{2} \left(n\sigma + n(\bar{x} - \mu)^2\right)]$$

With the prior being Gaussian-Gamma

$$\mathrm{P}(\mu,\tau) = \mathcal{N}\left(\mu|\mu_0, (\beta\tau)^{-1}\right)\Gamma(\tau|a,b)$$

Let us denote for simplicity

$$\mu_0 \leftrightarrow \mu, \quad \lambda_0 \leftrightarrow (\beta\tau)^{-1}$$

$$\alpha_0 \leftrightarrow a, \quad \beta_0 \leftrightarrow b$$

thus having the prior $\mathrm{P}(\mu,\tau) = \mathrm{GaussianGamma}(\mu,\tau|\mu_0,\lambda_0,\alpha_0,\beta_0)$. Posterior equals

$$\mathrm{Po}(\tau,\mu|\mathbf{X}) \propto \mathrm{L}(\mathbf{X}|\tau,\mu)\mathrm{Pr}(\tau,\mu) \propto$$

$$\propto \tau^{n/2} \exp\left[-\frac{\tau}{2}\left(n\sigma + n(\bar{x}-\mu)^2\right)\right] \tau^{\alpha_0 - \frac{1}{2}} \exp\left[-\beta_0\tau\right] \exp\left[-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}\right]$$

$$\propto \tau^{\frac{n}{2}+\alpha_0-\frac{1}{2}} \exp\left[-\tau\left(\frac{1}{2}n\sigma + \beta_0\right)\right] \exp\left[-\frac{\tau}{2}\left(\lambda_0(\mu-\mu_0)^2 + n(\bar{x}-\mu)^2\right)\right]$$

Let's complete the square in the last exponent (we wish it to look like $(\mu - \text{smth.})^2 + \text{smth.}$)

$$\lambda_0(\mu-\mu_0)^2 + n(\bar{x}-\mu)^2 = \lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + \lambda_0\mu_0^2 + n\mu^2 - 2n\bar{x}\mu + n\bar{x}^2$$

$$= (\lambda_0 + n)\mu^2 - 2(\lambda_0\mu_0 + n\bar{x})\mu + \lambda_0\mu_0^2 + n\bar{x}^2 =$$

so now we have decomposed over powers of μ

$$= (\lambda_0 + n)\left(\mu^2 - 2\frac{\lambda_0\mu_0 + n\bar{x}}{\lambda_0 + n}\mu\right) + \lambda_0\mu_0^2 + n\bar{x}^2 =$$

and here is the completion of square

$$(\lambda_0 + n)\left(\mu - \frac{\lambda_0\mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \lambda_0\mu_0^2 + n\bar{x}^2 - \frac{(\lambda_0\mu_0 + n\bar{x})^2}{\lambda_0 + n} =$$

finally equals

$$= (\lambda_0 + n)\left(\mu - \frac{\lambda_0\mu_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \frac{\lambda_0 n(\bar{x}-\mu_0)^2}{\lambda_0 + n}$$

Substituting to the last expression that we had, getting the $-\frac{\tau}{2}\left(\frac{\lambda_0 n(\bar{x}-\mu_0)^2}{\lambda_0+n}\right)$ term from the 'Gaussian' exponent to the 'Gamma' exponent, we finally have

$$\mathrm{Posterior}(\tau,\mu|\mathbf{X}) = \mathrm{GaussianGamma}\left(\frac{\mu + n\bar{x}\beta\tau}{1 + n\beta\tau}, (\beta\tau)^{-1} + n, a + \frac{n}{2}, b + \frac{1}{2}\left(n\sigma + \frac{n(\bar{x}-\mu)^2}{(1+n\beta\tau)}\right)\right)$$

the answer. We can now see how the prediction of τ,μ parameters changes with the change of $\mathbf{X}$ – explicitly with the change of n – number of observables, and with the change in mean and variance $\bar{x},\sigma$ – the first two moments of the variable set.