Exercise Sheet 5: Data Science Methods

Technische Universität München

Winter semester 2024-2025

Instructions

You can hand in your solutions during the next tutorial on 30/01/25 or submit them earlier at the group office 1269 in the Physics Department.

1 Exercise: Conjugate Distributions

1.1 Let the likelihood for one observation x be $p\left(x|\mu,\sigma^2\right) \sim \mathcal{N}\left(\mu,\sigma^2\right)$, for given σ^2. Furthermore, choose a normally distributed prior $p\left(\mu|\sigma^2\right) \sim \mathcal{N}\left(\mu_0,\sigma_0^2\right)$, where μ_0 and σ_0^2 are known. Compute the posterior distribution $p\left(\mu|x,\sigma^2\right)$.

Solution: From Bayes' theorem we know

$$p\left(\mu|x,\sigma^2\right) = \frac{p\left(x|\mu,\sigma^2\right)p(\mu|\sigma^2)}{p(x|\sigma^2)}.$$

Note that the numerator and denominator have the form

$$p\left(x|\mu,\sigma^2\right)p(\mu|\sigma^2) = N\exp\left(-a\mu^2 - b\mu - c\right),$$

$$p(x|\sigma^2) = \int \mathrm{d}\mu p\left(x|\mu,\sigma^2\right)p(\mu|\sigma^2) = N\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a} - c\right),$$

where we just solved a Gaussian integral in the last step and

$$a = \frac{\sigma_0^2 + \sigma^2}{2\sigma_0^2\sigma^2},$$

$$b = -\frac{\sigma_0^2 x + \sigma^2\mu_0}{\sigma_0^2\sigma^2}.$$

There is no need to determine N and c, since they cancel when building the ratio for the posterior:

$$p\left(\mu|x,\sigma^2\right) = \sqrt{\frac{a}{\pi}}\exp\left[-a\left(\mu + \frac{b}{2a}\right)\right],$$

which is a Gaussian distribution with

$$\mu_p = -\frac{b}{2a} = \frac{\sigma_0^2 x + \sigma^2\mu_0}{\sigma_0^2 + \sigma^2} = \frac{x/\sigma^2 + \mu_0/\sigma_0^2}{1/\sigma^2 + 1/\sigma_0^2},$$

$$\sigma_p^2 = \frac{1}{2a} = \frac{\sigma_0^2\sigma^2}{\sigma_0^2 + \sigma^2} = \frac{1}{1/\sigma^2 + 1/\sigma_0^2}.$$

1.2 Now generalize your result for multiple observations $\vec{x} = (x_1,\ldots,x_n)$ (i.i.d.) instead of only one observation x, i.e. determine the posterior $p\left(\mu|\vec{x},\sigma^2\right)$.

Solution: We proceed as before, writing $p\left(x|\mu,\sigma^2\right)p(\mu|\sigma^2) = N\exp\left(-a\mu^2 - b\mu - c\right)$. Only this time, we have $x \to \overline{x} \equiv \frac{\sum_{i=1}^n x_i}{n}$ and $\sigma^2 \to \sigma^2/n$, which can be easily checked by writing out the expression above. Hence,

$$\mu_p = \frac{n\overline{x}/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2},$$

$$\sigma_p^2 = \frac{1}{n/\sigma^2 + 1/\sigma_0^2}.$$

2 Exercise: Jeffreys Prior

The **Fisher information matrix** for a vector of parameters $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ is defined as:

$$\mathcal{I}_{ij}(\vec{\theta}) = \mathbb{E}\left[\frac{\partial \log f\left(x \mid \vec{\theta}\right)}{\partial \theta_i}\frac{\partial \log f\left(x \mid \vec{\theta}\right)}{\partial \theta_j}\right],$$

where $f\left(x \mid \vec{\theta}\right)$ is the likelihood function.

The **Jeffreys prior** for $\vec{\theta}$ is then defined as:

$$\pi\left(\vec{\theta}\right) \propto \sqrt{\det \mathcal{I}\left(\vec{\theta}\right)}.$$

Now, determine the Jeffreys prior, assuming that the likelihood is given by a normal distribution with unknown mean μ and variance σ^2.

Solution: The log-likelihood reads $\ell \equiv \log f(x|\mu, \sigma^2) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$. The derivatives of ℓ w.r.t the parameters are

$$\frac{\partial \ell}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4}.$$

Therefore,

$$\mathcal{I}_{\mu\mu}(\mu, \sigma^2) = \mathbb{E}\left[\frac{(x-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2},$$

$$\mathcal{I}_{\mu\sigma^2}(\mu, \sigma^2) = \mathbb{E}\left[\frac{(x-\mu)^3}{2\sigma^4} - \frac{(x-\mu)}{2\sigma^4}\right] = 0,$$

$$\mathcal{I}_{\sigma^2\sigma^2}(\mu, \sigma^2) = \mathbb{E}\left[\frac{1}{4\sigma^4} - \frac{(x-\mu)^2}{2\sigma^6} + \frac{(x-\mu)^4}{4\sigma^8}\right] = \frac{1}{4\sigma^4} - \frac{1}{2\sigma^4} + \frac{3}{4\sigma^4} = \frac{1}{2\sigma^4},$$

where we have used that $\mathbb{E}\left[(x-\mu)^2\right] = \sigma^2$ and $\mathbb{E}\left[(x-\mu)^n\right] = 0$ for all odd n, since the latter is an integral of an odd function over a symmetric interval. We then have

$$\pi\left(\mu, \sigma^2\right) \propto \sqrt{\det \mathcal{I}\left(\mu, \sigma^2\right)} = \text{const.}\left(\frac{1}{\sigma^2}\right)^{3/2}.$$

3 Exercise: Maximum a Posteriori (MAP) Estimator

The Maximum a Posteriori (MAP) estimator finds the parameter value that maximizes the posterior probability.

1.1 Argue that for a uniformly distributed prior, the MAP estimator is equivalent to the MLE in classical statistics.

Solution: For a uniform prior, the posterior is proportional to the likelihood. Hence, maximizing the posterior is equivalent to maximizing the likelihood.

1.2 Consider again the situation of Exercise 1 with multiple observations $\vec{x} = (x_1, \ldots, x_n)$. Determine the MAP estimator and compare it to the well-known MLE from the frequentist approach.

Solution: The posterior is proportional to

$$p\left(\mu|\vec{x}, \sigma^2\right) \propto p\left(\vec{x}|\mu, \sigma^2\right) p(\mu|\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2 - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right).$$

Computing the log of the posterior ℓ gives

$$\ell \propto -\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2 - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}.$$

Taking the derivative w.r.t. μ, setting it equal to zero and solving for μ, yields

$$\mu_{\mathrm{MAP}} = \frac{\frac{\overline{x}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2}}{n/\sigma^2 + 1/\sigma_0^2}.$$

This is the sum of the contributions from the likelihood (the MLE) and from the prior, weighted by the inverse of the uncertainties. This should coincide with the result of Exercise 1, since we know the mode of a normal distribution lies at the expected value, here μ_p. Note also that, as expected, this reduces to $\hat{\mu}_{\mathrm{MLE}} = \overline{x}$ by removing the factor coming from the prior.