Exercises

E1. Let $\theta \in (0,1)$ be a probability relating to a Bernoulli random variable. The *logit-transformation* is defined as $\phi = g(\theta) = \log(\frac{\theta}{1-\theta})$, i.e. this is a re-parametrization in terms of *log-odds* for an event. Check how the new parameter depends on the probability θ, e.g. by plotting a graph of their relation. Write down the log-likelihood function for n conditionally independent $X_i \sim$Bernoulli(θ), $i = 1, ..., n$, *using the logit parametrization* and find out the maximum likelihood estimate of ϕ. Show that ϕ arises as the *natural parameter* when the probability mass function of a Bernoulli(θ) random variable X is presented in the *exponential family form.* The logit-transformation is widely used to fit models explaining dependence of a success probability on some explanatory variables (logistic regression).

E2. You observe one real number x and wish to compare standard normal distribution to the standard Cauchy distribution. The density function of the latter equals

$$p(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty$$

How does the Bayes factor behave as a function of x?

E3. The first six pages of a draft version of the course material contained the following number of typos

$$3, 4, 2, 1, 2, 3.$$

What would you conclude about the predictive distribution of the number of typos on page seven, e.g. by using a Poisson model with a Gamma prior?

E4. Let $X_1, X_2, ..., X_n$ be conditionally independent random variables, such that $X_i \sim N(\mu, \sigma^2), i = 1, ..., n$, where σ^2 is a known quantity, that is we assume that the random variables are normally distributed with an unknown mean and known variance. Show that $X_1, (X_3 + X_{n-1})/2$ and $(1/n)\sum_{i=1}^n X_i$ are all *unbiased* estimators of μ. Compare their mean squared errors (MSE). Remember that MSE of an estimator $g(T)$ of θ is defined as the expectation $E(g(T) - \theta)^2$, where $g(T)$ is some function of the data $X_1, X_2, ..., X_n$.

E5. In some statistical problems involving a Binomial likelihood, the number of trials is an unknown parameter. Let $Y_1, ..., Y_m$ be conditionally independent random variables, such that $Y_i \sim Binomial(n, \theta), i = 1, ..., n$, where θ is known and n is unknown. Write down the log-likelihood function and investigate its behavior numerically for the data $y_1 = 16, y_2 = 18, y_3 = 22, y_4 = 25, y_1 = 27$, when θ is assumed to equal 1/3. Investigate also the log-likelihood function behavior when both n and θ are unknown for these data.