Assignments I

T1. Consider a dichotomous property of individuals in a large finite population. Let the population size be N and the size of a sample taken without replacement be n. This situation corresponds to an urn containing N balls, which are either black or white. Let θ denote the number of black balls in the urn. The probability of a sample of n balls contains x black ones is given by the hypergeometric expression

$$p(x|\theta) = \frac{\binom{\theta}{x}\binom{N-\theta}{n-x}}{\binom{N}{n}}.$$

Further, if $p(\theta = r), r = 0, ..., N$, specifies the prior probabilities for θ, we get the posterior probability of homogenous urn $\theta = N$ as

$$p(\theta|x = n) = \frac{p(x = n|\theta = N)p(\theta = N)}{\sum_{r=n}^{N} p(x = n|\theta = r)p(\theta = r)}.$$

Assume a uniform prior and calculate explicitly the above posterior probability. Analyze the situation from a general scientific perspective, does the result seem plausible for use in practice? Hint: you can utilize the following general result

$$\sum_{r=0}^{N} \binom{r}{l}\binom{N-r}{m} = \binom{N+1}{l+m+1},$$

where $l + m = n$ is the sample size for the hypergeometric model and l stands for the number of picked individuals having the characteristic of interest (*e.g.* black colour).

T2. You observe two thumbtack tosses which both land with the point down (see lecture slides for details). Investigate explicitly the predictive distribution for the outcome of the next toss, using the following four priors: Beta(1/2,1/2), Beta(1,1), Beta(0,0), Beta(10,10). How does the prior affect your predictions? NB The predictive distribution for a comparable model is derived in the lecture slides (see the spam filtering example).

T3. Generalize the previous problem (T2) to a situation where you wish to predict the number of thumbtack tosses which land with the point down in five future tosses. Calculate the expectation and variance for this quantity (Beta-Binomially distributed), either analytically or by computer simulation. Compare the results with the expectation and variance one gets from an ordinary Binomial distribution where the underlying probability of each Bernoulli event is a fixed quantity equal to the maximum likelihood estimate from the observed data.

T4. In betting situations one is often interested in odds, referring in the thumbtack tossing situation to the quantity $\theta/(1-\theta)$. Alternatively one may consider the log-odds

$$\lambda = \log \frac{\theta}{1-\theta}.$$

Show that a "uniform" distribution for λ implies the following distribution for θ

$$p(\theta) = \theta^{-1}(1-\theta)^{-1}.$$

What problems are associated with this distribution if it is used as a prior in the thumbtack tossing problem?

T5. Let $x_1, ..., x_n$ be a sample of independent random quantities from the uniform distribution on $(0, \theta)$. Show that the conditional joint distribution of the observations is

$$p(\mathbf{x}|\theta) = \begin{cases} \theta^{-n}, \theta > m \\ 0, \leq m \end{cases}, m = \max(x_1, ..., x_n).$$

What is your posterior if the prior is Pareto(α, β) distribution, *i.e.*

$$\alpha\beta^{\alpha}x^{-(\alpha+1)}, \alpha > 0, \beta > 0, x \geq \beta.$$

What happens to the posterior inference when $\alpha \to 0, \beta \to 0$? Consider the situation with $n = 10 = \max(x_1, ..., x_n)$.

T6. You observe $X = x$ to learn about parameter θ, for which you have the prior $p(\theta)$. You are uncertain about the shape of a suitable distribution, so you define the likelihood of x as a finite mixture

$$p(x|\theta) = \sum_{i=1}^{k} \pi_i p_i(x|\theta),$$

where $0 < \pi_i < 1$ and $\sum_{i=1}^{k} \pi_i = 1$. What is the shape of your posterior for θ? Consider how the information contained in the observation is reflected by the posterior.