Deep Bayes Summer School

Practical Session 1. Bayesian Reasoning

August 27th 2018

1. Albus Dumbledore believes that the Dark Lord somehow survived after his death spell bounced off on the historical night of October 31st 1981. While prophecies about the Dark Lord surviving that night are ambiguous, the Dark Mark on the hand of Severus Snape is a real piece of evidence. One would expect it to fade after the fall of the Dark Lord, but is this a sufficient evidence? Is it that unlikely for a magical mark to stay around after the maker dies?

   Suppose that, if the Dark Lord dies, the Dark Mark continues to exist with twenty percent probability. On the other hand, if the Dark Lord's sentience lives on, the Dark Mark will stay with one hundred percent chance. Additionally, let's stay the prior odds were a hundred-to-one against the Dark Lord surviving. What is the probability of Dark Lord being alive?

2. *(Multinomial likelihood)* Let $\mathcal{D} = \{x_1, \dots, x_N\}$ be N independent dice rolls. For brevity, we denote the number of times a dice comes up as face $k \in \{1, \dots, K\}$ as $N_k = \sum_{n=1}^{N} \mathbb{I}(x_i = k)$. With this notation the likelihood has the form

$$p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \theta_k^{N_k}, \tag{1}$$

   where θ_k is the probability of outcome k. Compute the maximum likelihood estimate for $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$. Do not forget $\boldsymbol{\theta} \in S_K$, i.e. $\sum_{k=1}^{K} \theta_k = 1$ and $\theta_k \geq 0$ for $k = 1, \dots, K$.

3. *(Dirichlet prior)* The conjugate prior distribution for multinomial likelihood defined in Eq. 1 is the Dirichlet distribution:

$$\mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \frac{\mathbb{I}(\boldsymbol{\theta} \in S_K)}{B(\alpha_1, \dots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1},$$

   where $\alpha_k > 0$ and $B(\alpha_1, \dots, \alpha_K) = \int_{S_K} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \mathrm{d}\boldsymbol{\theta}$ is the normalizing constant, also known as the multivariate Beta function. Check that the Dirichlet distribution is indeed the conjugate distribution by computing the posterior $p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha})$. Then, compute the posterior predictive $p(x_{N+1} = k|\boldsymbol{\alpha}) = \int_{S_K} p(x_{N+1} = k|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\theta}$.

   To simplify the answer, you may use the following expression for the multivariate Beta function

$$B(\alpha_1, \dots, \alpha_K) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K} \alpha_k)}$$

   and the multiplicative property of the Gamma function $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

4. To make a decision under uncertainty, Bayesian decision theory suggests to find a point estimate that minimizes the expected loss. For loss functions $L_0(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1 - \delta(\boldsymbol{\theta} - \boldsymbol{\theta'})$ and $L_2(\boldsymbol{\theta}, \boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{\theta'})^T (\boldsymbol{\theta} - \boldsymbol{\theta'})$ and an arbitrary posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$ find point estimate(s)

$$\mathrm{argmin}_{\boldsymbol{\theta'} \in S_K} \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})} L(\boldsymbol{\theta}, \boldsymbol{\theta'}).$$

   Next, compute these estimates for the Dirichlet-multinomial model $p(\mathcal{D}|\boldsymbol{\theta}) \, \mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha})$ from the previous problem. The answer has to be a function of prior hyperparameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ and observations $N_1, \dots, N_K$.

5.* Given the definition of the Gamma function $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha - 1} e^{-t} \mathrm{d}t$, prove that the multivariate Beta function can be expressed in terms of Gamma functions:

$$B(\alpha_1, \dots, \alpha_K) = \int_{S_K} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \mathrm{d}\boldsymbol{\theta} = \frac{\Gamma(\alpha_1) \dots \Gamma(\alpha_K)}{\Gamma(\sum_{k=1}^{K} \alpha_k)}.$$