Bayesian Reasoning

Kirill Struminsky
PhD Student at HSE

DeepBayes

The problem set is available at https://github.com/bayesgroup/deepbayes-2018

Problem 1

Insufficient Bayesian Evidence[1]

Setting

The Dark Mark

- stays with 20% probability if the makes dies
- stays with 100% probability if the maker is still alive

The Dark Lord survived his attack on Harry Potter with the small chance of 1:100

Objective

Given that Severus Snapes Dark Mark has not faded, find the probability of the Dark Lord being alive.

[1] Problem taken from http://www.hpmor.com/

Problem 1

Solution

Let $x \in \{0, 1\}$ indicate that the Dark Lord is alive, $y \in \{0, 1\}$ indicate that the Dark Mark is still visible.
Compute $p(x = 1|y = 1)$:

$$p(x = 1|y = 1) = \frac{p(y = 1|x = 1)p(x = 1)}{p(y = 1)} = \frac{p(y = 1|x = 1)p(x = 1)}{\sum_j p(y = 1|x = j)p(x = j)} =$$

$$= \frac{1 \cdot \frac{1}{101}}{1 \cdot \frac{1}{101} + \frac{1}{5} \cdot \frac{100}{101}} = \frac{1}{1 + \frac{100}{5}} = \frac{1}{21}$$

Problem 2

MLE for multinomial likelihood

Setting

- $\mathcal{D} = \{x_1, \ldots, x_N\}$ — independent dice rolls
- $N_k = \sum_{n=1}^{N} \mathbb{I}(x_i = k)$ — counts
- $p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \theta_k^{N_k}$ — likelihood, $\boldsymbol{\theta} \in S_K$

Objective

Maximum likelihood estimate for $\boldsymbol{\theta}$, $\mathrm{argmax}_{\boldsymbol{\theta} \in S_K} \log p(\mathcal{D}|\boldsymbol{\theta})$

Problem 2

Solution

$\boldsymbol{\theta}$ is restricted to simplex. Change parameterization to $\mu_k = \log \theta_k, \;\; \mu_k \in \mathbb{R}$ to omit the inequality restrictions. The Lagrangian has form

$$\mathcal{L}(\boldsymbol{\mu}, \lambda) = \log p(\mathcal{D} | \exp \boldsymbol{\mu}) + \lambda(\sum_{k=1}^{K} \exp \mu_k - 1) = \sum_{k=1}^{K} (N_k \mu_k - \lambda \exp \mu_k) - \lambda$$

After differentiation we find the singular point $\theta_k = \frac{N_k}{\sum_{l=1}^{K} N_l}, \;\; k = 1, \ldots, K$:

$$0 = \frac{\partial \mathcal{L}(\boldsymbol{\mu}, \lambda)}{\partial \mu_k} = N_k - \lambda \exp \mu_k \Rightarrow \theta_k = \exp \mu_k = \frac{N_k}{\lambda}$$

$$0 = \frac{\partial \mathcal{L}(\boldsymbol{\mu}, \lambda)}{\partial \lambda} = \sum_{k=1}^{K} \exp \mu_k - 1 \Rightarrow \lambda = \sum_{k=1}^{K} N_k$$

Problem 3

Dirichlet-multinomial model

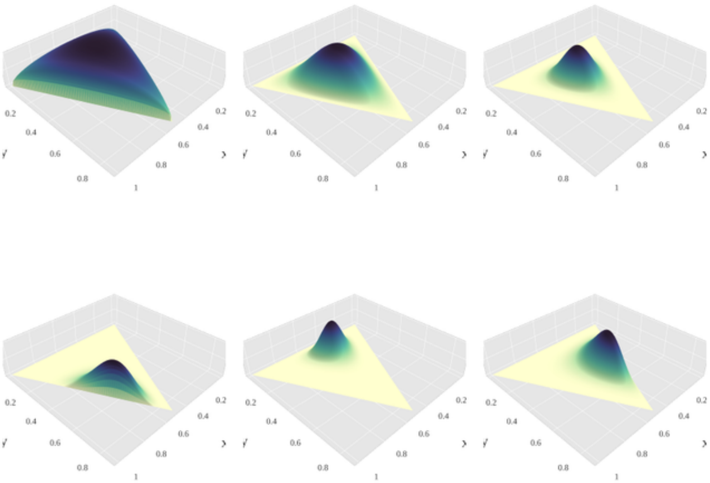
- Check that the Dirichlet prior

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

  is the conjugate prior for the multinomial likelihood $p(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \theta_k^{N_k}$ by computing the posterior $p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha})$ [2].
- Compute the posterior predictive $p(x_{N+1} = k|\boldsymbol{\alpha}) = \int_{S_K} p(x_{N+1} = k|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\theta}$.

[2] As $B(\alpha_1, \ldots, \alpha_K) = \int_{S_K} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \mathrm{d}\boldsymbol{\theta}$ we denote the normalizing constant.

Dirichlet distribution density plots

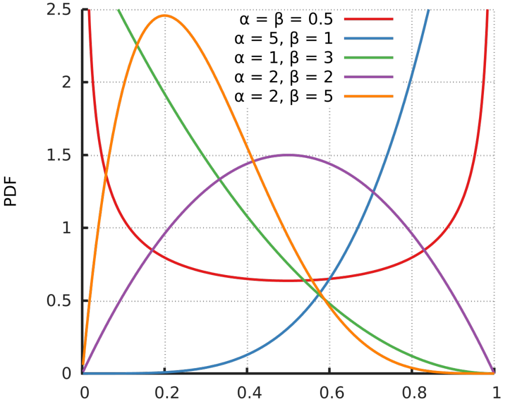


https://en.wikipedia.org/wiki/Dirichlet_distribution

Beta distribution density plots

The special case with density $p(q|\alpha, \beta) \propto q^{\alpha-1}(1-q)^{\beta-1}, \ \alpha, \beta > 0$.

Prior for an unfair coin?



https://en.wikipedia.org/wiki/Beta_distribution

Problem 3

Solution: posterior

Apply the Bayes rule $p(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\alpha}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{p(\mathcal{D}|\boldsymbol{\alpha})}$. Denominator (a.k.a. **evidence**):

$$p(\mathcal{D}|\boldsymbol{\alpha}) = \int_{S_K} p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\theta} = \frac{\int_{S_K}\prod_{k=1}^K \theta_k^{N_k}\cdot\theta_k^{\alpha_k-1}\mathrm{d}\boldsymbol{\theta}}{B(\alpha_1,\ldots,\alpha_K)} = \frac{B(\alpha_1+N_1,\ldots,\alpha_K+N_K)}{B(\alpha_1,\ldots,\alpha_K)}$$

and use the resulting expression

$$p(\boldsymbol{\theta}|\mathcal{D},\boldsymbol{\alpha}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})}{p(\mathcal{D}|\boldsymbol{\alpha})} = \frac{B(\alpha_1,\ldots,\alpha_K)\mathbb{I}(\boldsymbol{\theta}\in S_K)\prod_{k=1}^K \theta_k^{N_k}\cdot\theta_k^{\alpha_k-1}}{B(\alpha_1+N_1,\ldots,\alpha_K+N_K)B(\alpha_1,\ldots,\alpha_K)} = \mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}'),$$

where $\boldsymbol{\alpha}' = (\alpha_1+N_1,\ldots,\alpha_K+N_K)$. Was it necessary to compute $p(\mathcal{D}|\boldsymbol{\alpha})$?

Problem 3

Solution: posterior predictive

$$
\begin{aligned}
p(x_{N+1} = k|\boldsymbol{\alpha}) = \int_{S_K} p(x_{N+1} = k|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\theta} = \frac{\int_{S_K} \theta_k \prod_{k=1}^{K} \theta_k^{N_k+\alpha_k-1} \mathrm{d}\theta}{B(\alpha_1 + N_1, \ldots, \alpha_K + N_K)} \\
= \frac{B(\alpha_1 + N_1, \ldots, \alpha_k + N_k + 1, \ldots, \alpha_K + N_K)}{B(\alpha_1 + N_1, \ldots, \alpha_k + N_k, \ldots, \alpha_K + N_K)} \\
= \frac{\Gamma(\alpha_1 + N_1) \ldots \Gamma(\alpha_k + N_k + 1) \ldots \Gamma(\alpha_K + N_K)}{\Gamma(\alpha_1 + N_1) \ldots \Gamma(\alpha_k + N_k) \ldots \Gamma(\alpha_K + N_K)} \cdot \frac{\Gamma(\sum_l (\alpha_l + N_l))}{\Gamma(\sum_l (\alpha_l + N_l) + 1)} \\
= \frac{\alpha_k + N_k}{\sum_k \alpha_k + N}
\end{aligned}
$$

Problem 4

Bayesian decision theory

For loss functions

- $L_0(\boldsymbol{\theta}, \boldsymbol{\theta}') = 1 - \delta(\boldsymbol{\theta} - \boldsymbol{\theta}')$
- $L_2(\boldsymbol{\theta}, \boldsymbol{\theta}') = (\boldsymbol{\theta} - \boldsymbol{\theta}')^T(\boldsymbol{\theta} - \boldsymbol{\theta}')$

and a posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$ find point estimate(s)

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

Compute these estimates for the Dirichlet-multinomial model $p(\mathcal{D}|\boldsymbol{\theta}) \, \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha})$.

Problem 4

Solution

L_0

Compute the expectation:

$$\mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})} L_0(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int_{S_K} p(\boldsymbol{\theta}|\mathcal{D})(1 - \delta(\boldsymbol{\theta} - \boldsymbol{\theta}'))\mathrm{d}\boldsymbol{\theta} = 1 - p(\boldsymbol{\theta}'|\mathcal{D}).$$

Its minimum is achieved at any distribution mode.

L_2

The only stationary point is expectation $\boldsymbol{\theta}' = \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}\boldsymbol{\theta}$

$$\begin{aligned}\nabla_{\boldsymbol{\theta}'} \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})} L_2(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \nabla_{\boldsymbol{\theta}'} \int_{S_K} p(\boldsymbol{\theta}|\mathcal{D})(\boldsymbol{\theta} - \boldsymbol{\theta}')^T(\boldsymbol{\theta} - \boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta} \\ &= 2\int_{S_K} p(\boldsymbol{\theta}|\mathcal{D})(\boldsymbol{\theta} - \boldsymbol{\theta}')\mathrm{d}\boldsymbol{\theta} = \boldsymbol{\theta}' - \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}\boldsymbol{\theta} = 0\end{aligned}$$

Problem 4

Solution: Dirichlet-multinomial case

L_0

The optimization problem is the same as in Problem 2:

$$C \cdot \prod_{k=1}^{K} \theta_k^{\alpha_k + N_k - 1} \to \max_{\boldsymbol{\theta} \in S_k}$$

The optimal value $\boldsymbol{\theta}'$ has form $\theta'_k = \frac{\alpha_k + N_k - 1}{\sum_l \alpha_l + N - K}$.

L_2

The expectation has the same form as in the posterior predictive in Problem 3:

$$\theta'_k = \frac{\int_{S_K} \theta_k \prod_{k=1}^{K} \theta_k^{N_k + \alpha_k - 1} \mathrm{d}\theta}{B(\alpha_1 + N_1, \ldots, \alpha_K + N_K)} = \frac{\alpha_k + N_k}{\sum_l \alpha_l + N}$$