Bayesian reasoning

Ekaterina Lobacheva

Research Fellow at Samsung-HSE Laboratory

DeepBayes

Bayesian reasoning: problem set

The problem set is available here:

tiny.cc/BR_problems

Problem 1: basic Bayesian reasoning

Setting

The Dark Mark

- stays with 20% probability if the maker 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 1/101

Question

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

Problem 1: basic Bayesian reasoning

- $x \in \{0, 1\}$ — the Dark Lord (1 means that he is alive)
- $y \in \{0, 1\}$ — the Dark Mark (1 means that it is still visible)

Setting: $p(y = 1 \mid x = 1) = 1, \quad p(y = 1 \mid x = 0) = 0.2, \quad p(x = 1) = 1/101$

Question: $p(x = 1 \mid y = 1) = \, ?$

Problem 1: basic Bayesian reasoning

- $x \in \{0, 1\}$ — the Dark Lord (1 means that he is alive)
- $y \in \{0, 1\}$ — the Dark Mark (1 means that it is still visible)

Setting: $p(y = 1 \mid x = 1) = 1, \quad p(y = 1 \mid x = 0) = 0.2, \quad p(x = 1) = 1/101$

Question: $p(x = 1 \mid y = 1) = \,?$

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

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

Problem 2: frequentist framework

Setting

- $X = \{x_1, \dots, x_N\}$ — independent dice rolls
- $N_k = \sum_{n=1}^{N} \mathbb{I}(x_n = k)$ — counts
- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$

Question

Maximum likelihood estimate for $\theta_{ML} = \arg\max_{\theta \in S_K} \log p(X \mid \theta)$

Problem 2: frequentist framework

θ is restricted to a simplex. To omit the inequality restrictions change parameterization to $\mu_k = \log \theta_k, \quad \mu_k \in \mathbb{R}$

The Lagrangian has the form:

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

Problem 2: frequentist framework

θ is restricted to simplex. To omit the inequality restrictions change parameterization to $\mu_k = \log \theta_k, \quad \mu_k \in \mathbb{R}$

The Lagrangian has the form:

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

Differentiation:

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

$$\Rightarrow \quad \theta_k = \frac{N_k}{\sum_{l=1}^{K} N_l}$$

Problem 3: Bayesian framework

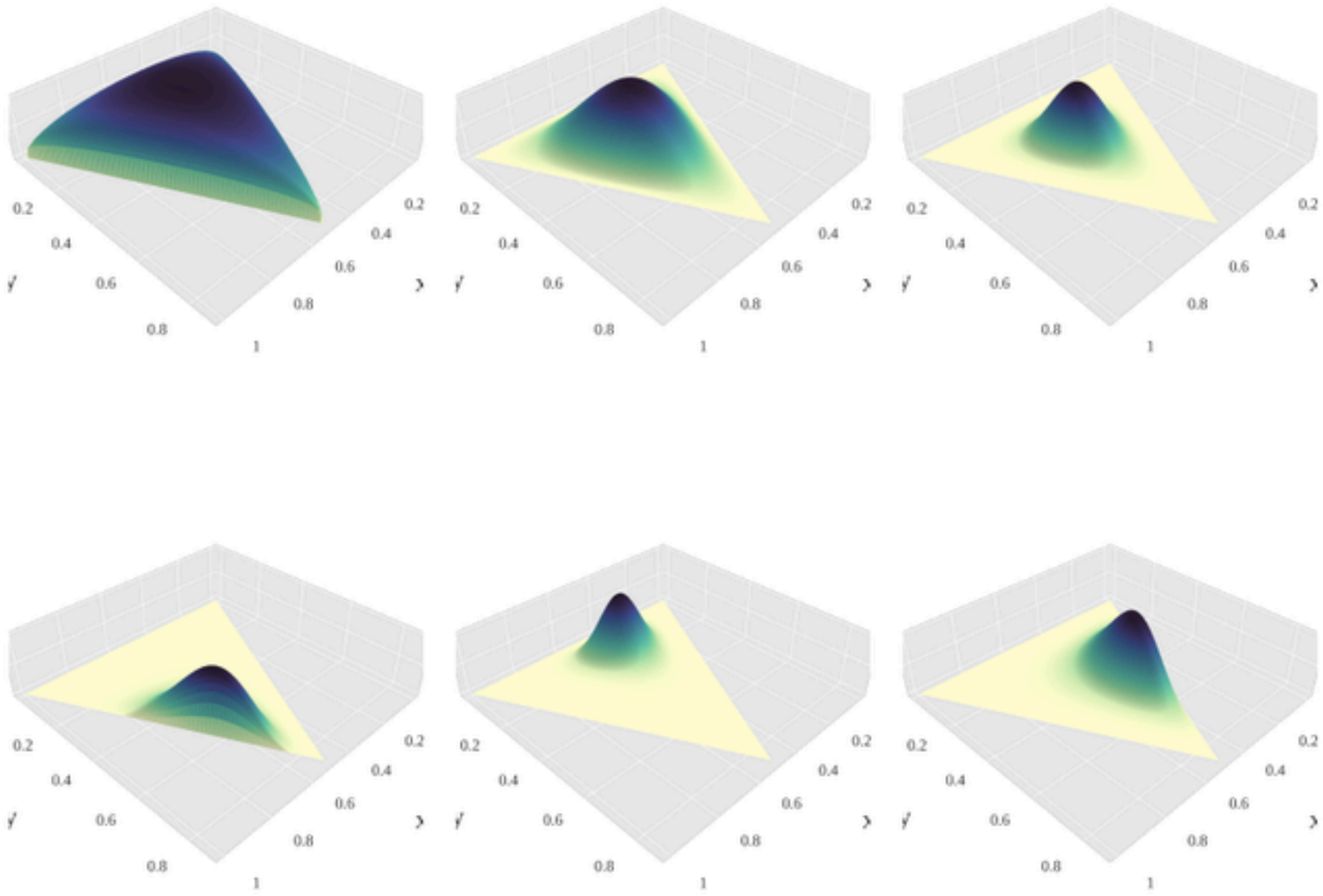
Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior: $$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- Compute the posterior $p(\theta | X, \alpha)$
- Compare $\mathbb{E}_{p(\theta|X,\alpha)}\theta$ and θ_{ML}
- Compute the predictive posterior $p(x_{N+1} = k | X, \alpha)$

Dirichlet distribution



Beta distribution is a special case of Dirichlet distribution:

$$\mathrm{Dir}(\theta \mid \alpha) \propto \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

$$Beta(\theta \mid a, b) \propto \theta^{a-1}(1-\theta)^{b-1}$$

https://en.wikipedia.org/wiki/Dirichlet distribution

Problem 3: Bayesian framework

Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior:
$$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- Compute the posterior $p(\theta|X, \alpha)$
- Compare $\mathbb{E}_{p(\theta|X,\alpha)}\theta$ and θ_{ML}
- Compute the predictive posterior $p(x_{N+1} = k|X, \alpha)$

Problem 3: Bayesian framework

Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior: $$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- Compute the posterior $p(\theta | X, \alpha)$
- Compare $\mathbb{E}_{p(\theta|X,\alpha)}\theta$ and θ_{ML}
- Compute the predictive posterior $p(x_{N+1} = k | X, \alpha)$

Problem 3: Bayesian framework

Probabilistic model: $$p(X, \theta) = p(X \mid \theta) p(\theta) = Dir(\theta \mid \alpha) \prod_{k=1}^{K} p\left(x_k \mid \theta\right)$$

Here different constants are denoted with
the same letter C for demonstration reasons.

Problem 3: Bayesian framework

Probabilistic model: $$p(X,\theta) = p(X \mid \theta)p(\theta) = Dir(\theta \mid \alpha)\prod_{k=1}^{K} p\left(x_k \mid \theta\right)$$

Prior: $$p(\theta) = \frac{1}{B(\alpha_1,\ldots,\alpha_K)}\prod_{k=1}^{K}\theta_k^{\alpha_k-1} = C\prod_{k=1}^{K}\theta_k^{C}$$

Here different constants are denoted with
the same letter C for demonstration reasons.

Problem 3: Bayesian framework

Probabilistic model: $$p(X, \theta) = p(X \mid \theta) p(\theta) = Dir(\theta \mid \alpha) \prod_{k=1}^{K} p\left(x_k \mid \theta\right)$$

Prior: $$p(\theta) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} = C \prod_{k=1}^{K} \theta_k^{C}$$

Posterior: $$p(\theta \mid X) \propto p(X \mid \theta) p(\theta) = \prod_{k=1}^{K} \theta_k^{N_k} \cdot \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} =$$

$$= C \prod_{k=1}^{K} \theta_k^{C}$$

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 3: Bayesian framework

Probabilistic model: $$p(X, \theta) = p(X \mid \theta)p(\theta) = Dir(\theta \mid \alpha) \prod_{k=1}^{K} p\left(x_k \mid \theta\right)$$

Prior: $$p(\theta) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} = \boxed{C \prod_{k=1}^{K} \theta_k^{C}}$$

Posterior: $$p(\theta \mid X) \propto p(X \mid \theta)p(\theta) = \prod_{k=1}^{K} \theta_k^{N_k} \cdot \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} =$$

$$= \boxed{C \prod_{k=1}^{K} \theta_k^{C}}$$

conjugate

Here different constants are denoted with the same letter C for demonstration reasons.

Problem 3: Bayesian framework

Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior: $$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- $\boxed{\text{Compute the posterior } p(\theta|X, \alpha)}$
- Compare $\mathbb{E}_{p(\theta|X,\alpha)}\theta$ and θ_{ML}
- Compute the predictive posterior $p(x_{N+1} = k|X, \alpha)$

Problem 3: Bayesian framework

Likelihood and prior are conjugate $\longrightarrow$ posterior is Dirichlet

$$p(\theta \mid X) \propto p(X \mid \theta)p(\theta) = \prod_{k=1}^{K} \theta_k^{N_k} \cdot \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \propto$$

$$\propto \prod_{k=1}^{K} \theta_k^{N_k + \alpha_k - 1}$$

$$p(\theta \mid X) = Dir(\theta \mid \alpha'), \quad \alpha' = (\alpha_1 + N_1, \ldots, \alpha_K + N_K)$$

Problem 3: Bayesian framework

Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior: $$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- Compute the posterior $p(\theta|X, \alpha)$
- $\boxed{\text{Compare } \mathbb{E}_{p(\theta|X,\alpha)}\theta \text{ and } \theta_{ML}}$
- Compute the predictive posterior $p(x_{N+1} = k|X, \alpha)$

Problem 3: Bayesian framework

Maximum likelihood estimate: $$\theta_k = \frac{N_k}{\sum_{l=1}^{K} N_l}$$

Expectation of the posterior: $$\mathbb{E}_{p(\theta|X)}\theta_k = \frac{\alpha_k + N_k}{\sum_{l=1}^{K} \alpha_l + N_l}$$

Small K $\longrightarrow$ Bayesian estimate is mostly based on prior

Large K $\longrightarrow$ Bayesian estimate is very similar to ML estimate

Problem 3: Bayesian framework

Setting

- $p(X \mid \theta) = \prod_{k=1}^{K} \theta_k^{N_k}$ — multinomial likelihood, $\theta \in S_K$
- Dirichlet prior:

$$\mathrm{Dir}(\theta \mid \alpha) = \frac{1}{B(\alpha_1, \ldots, \alpha_K)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

Questions

- Check that likelihood and prior are conjugate
- Compute the posterior $p(\theta | X, \alpha)$
- Compare $\mathbb{E}_{p(\theta|X,\alpha)}\theta$ and θ_{ML}
- Compute the predictive posterior $p(x_{N+1} = k | X, \alpha)$

Problem 3: Bayesian framework

$$p(x_{N+1} = k \mid X, \alpha) = \int_{S_K} p(x_{N+1} = k \mid \theta) p(\theta \mid X, \alpha) \mathrm{d}\theta =$$

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

$$= \frac{\Gamma(\alpha_1 + N_1) \ldots \Gamma(\alpha_k + N_k + 1) \ldots \Gamma(\alpha_K + N_K)}{\Gamma(\alpha_1 + N_1) \ldots \Gamma(\alpha_k + N_k) \ldots \Gamma(\alpha_K + N_K)} \cdot \frac{\Gamma(\sum_l (\alpha_l + N_l))}{\Gamma(\sum_l (\alpha_l + N_l) + 1)} =$$

$$= \frac{\alpha_k + N_k}{\sum_k \alpha_k + N}$$