Bayes Homework 1

Vayser Kirill

October 2021

1 Problem 1 [25 pts]

Let $\delta(x)$ be the Dirac delta function. Consider the function

$$f(x) = \int_0^\infty \delta\left(\sum_{i=1}^n \theta_i - x\right) \prod_{i=1}^n \theta_i^{u_i - 1} d\theta_1 \dots d\theta_n,\ u_i > 0\ \forall i.$$

By using forward and inverse Laplace transform show that

$$f(x) = \frac{\prod_{i=1}^n \Gamma(u_i)}{\Gamma\left(\sum_{i=1}^n u_i\right)} x^{\sum_{i=1}^n u_i - 1}.$$

Solution

Let us calculate Laplace transform for both representations of $f(x)$:

$$F_1(s) = \int\limits_0^\infty e^{-sx} f(x) dx = \int_0^\infty e^{-sx} \delta\left(\sum_{i=1}^n \theta_i - x\right) \prod_{i=1}^n \theta_i^{u_i-1} d\theta_1 \dots d\theta_n dx =$$

$$= \int_0^\infty e^{-s\sum_{i=1}^n \theta_i} \prod_{i=1}^n \theta_i^{u_i-1} d\theta_1 \dots d\theta_n = \int_0^\infty \prod_{i=1}^n e^{-s\theta_i} \theta_i^{u_i-1} d\theta_1 \dots d\theta_n =$$

$$= \prod_{i=1}^n \int_0^\infty e^{-s\theta_i} \theta_i^{u_i-1} d\theta_i = \prod_{i=1}^n \int_0^\infty e^{-s\theta_i} (s\theta_i)^{u_i-1} \frac{ds\theta_i}{s^{\sum_{i=1}^n u_i}} =$$

$$= \prod_{i=1}^n \frac{\Gamma(u_i)}{s^{u_i}}$$

Let us denote the second representation

$$f(x) = \frac{\prod_{i=1}^n \Gamma(u_i)}{\Gamma\left(\sum_{i=1}^n u_i\right)} x^{\sum_{i=1}^n u_i - 1} = Ax^{\sum_{i=1}^n u_i - 1}$$

Then

$$F_2(s) = \int\limits_0^\infty e^{-sx} f(x) dx = A \int_0^\infty e^{-sx} x^{\sum_{i=1}^n u_i - 1} dx = A \int_0^\infty e^{-sx} (sx)^{\sum_{i=1}^n u_i - 1} \frac{dsx}{s^{\sum_{i=1}^n u_i}} =$$

$$= \frac{\prod_{i=1}^n \Gamma(u_i)}{\Gamma\left(\sum_{i=1}^n u_i\right)} \frac{\Gamma(\sum_{i=1}^n u_i)}{s^{\sum_{i=1}^n u_i}} = \prod_{i=1}^n \frac{\Gamma(u_i)}{s^{u_i}}$$

Since $F_1(s) = F_2(s)$ and Laplace transformation is bijection, then representations of $f(x)$ are equal.

2 Problem 2 [25 pts]

Consider the following divergence between distributions p and q:

$$D_\alpha(p\|q) = \frac{\int \alpha p(x) + (1-\alpha)q(x) - [p(x)]^\alpha [q(x)]^{1-\alpha} dx}{\alpha(1-\alpha)}, \ \alpha \in (0;1).$$

1. Prove that $D_\alpha(p\|q) \geqslant 0$. And $D_\alpha(p\|q) = 0$ if and only if $p(x) = q(x)$ for any x.

2. Prove that

$$\lim_{\alpha \to 0} D_\alpha(p\|q) = \mathrm{KL}(q\|p)$$
$$\lim_{\alpha \to 1} D_\alpha(p\|q) = \mathrm{KL}(p\|q)$$

Solution

1. Let us simplify the expression :

$$D_\alpha(p\|q) = \frac{1}{\alpha(1-\alpha)} \left(1 - \int [p(x)]^\alpha [q(x)]^{1-\alpha} dx\right)$$

Let $p(x) = z(x)q(x)$. Then

$$\int [p(x)]^\alpha [q(x)]^{1-\alpha} dx = \int z(x)^\alpha q(x) dx$$

Because of Jensen inequality and concavity of $f(x) = x^\alpha, \ \alpha < 1$

$$\int z(x)^\alpha q(x) dx \leq \left(\int z(x) q(x) dx\right)^\alpha = \left(\int p(x) dx\right)^\alpha = 1$$

And therefore $D_\alpha(p\|q) \geq 0$. Also, since power of α is a nonlinear function, equality in Jensen inequality, according to its properties, is achieved only if $z(x) = 1$ almost everywhere. It means that $D_\alpha(p\|q) = 0$ if and only if $p(x) = q(x)$ almost everywhere.

2. With $\alpha \to 0$ both numerator and denominator will seek to 0 also. It makes it possible to use L'Hopital rule :

$$\begin{aligned}
\lim_{\alpha \to 0} D_\alpha(p\|q) &= \lim_{\alpha \to 0} \frac{\int \alpha p(x) + (1-\alpha)q(x) - [p(x)]^\alpha [q(x)]^{1-\alpha} dx}{\alpha(1-\alpha)} = \\
&= \lim_{\alpha \to 0} \frac{\frac{d}{d\alpha} \int \alpha p(x) + (1-\alpha)q(x) - [p(x)]^\alpha [q(x)]^{1-\alpha} dx}{\frac{d}{d\alpha}\alpha(1-\alpha)} = \\
&= \lim_{\alpha \to 0} \frac{\frac{d}{d\alpha} \int -[p(x)]^\alpha [q(x)]^{1-\alpha} dx}{1-2\alpha} = \lim_{\alpha \to 0} \frac{\frac{d}{d\alpha} \int -[z(x)]^\alpha [q(x)] dx}{1-2\alpha} = \\
&= \lim_{\alpha \to 0} \frac{\int -\log z(x) [z(x)]^\alpha [q(x)] dx}{1-2\alpha} = \int -\log \frac{p(x)}{q(x)} [q(x)] dx = \\
&= \int \log \frac{q(x)}{p(x)} [q(x)] dx = \mathrm{KL}(q\|p)
\end{aligned}$$

For the second point we could use that if we swap p and q with and α with $1-\alpha$, then we get the same problem, so the second point could be done analogically.

Aspiration for this point of the 2 problem was from talks from seminars and here

3 Problem 3 [25 pts]

Consider the following exponential family of distributions:

$$p(x;\beta) = \frac{1}{Z(\beta)}\pi(x)\exp(-\beta h(x)),\ \beta \in [\beta_n;\beta_0]\,.$$

$Z(\beta)$ – the normalizing constant. $x \in \mathbb{R}$.

Prove the following identity:

$$\log Z(\beta_n) - \log Z(\beta_0) = \int_{\beta_n}^{\beta_0} \langle h(x)\rangle_{p(x;\beta)} d\beta$$

Solution:

It is well known that for exponential family the following property holds:

$$\nabla \log Z(\beta) = \mathbb{E}h(x)$$

It follows from the moment-generating function of exponential family distributions. However, since it is crucial for this problem, we will show it explicitly:

$$\mathbb{E}h(x) = \int h(x)\frac{1}{Z(\beta)}\pi(x)\exp(-\beta h(x))dx = \int h(x)\frac{\nabla \log Z(\beta)}{\nabla Z(\beta)}\pi(x)\exp(-\beta h(x))dx$$

Since $Z(\beta) = \int \pi(x)\exp(-\beta h(x))dx$, then $\nabla Z(\beta) = \int -h(x)\pi(x)\exp(-\beta h(x))dx$. Placing it in the statement for $\mathbb{E}h(x)$ we get:

$$\mathbb{E}h(x) = \int h(x)\frac{\nabla \log Z(\beta)}{\nabla Z(\beta)}\pi(x)\exp(-\beta h(x))dx =$$
$$= \int h(x)\frac{\nabla \log Z(\beta)}{\int -h(x)\pi(x)\exp(-\beta h(x))dx}\pi(x)\exp(-\beta h(x))dx = -\nabla \log Z(\beta)$$

Then, integrating from β_n to β_0 we obtain:

$$\int_{\beta_n}^{\beta_0} \mathbb{E}h(x)d\beta = \int_{\beta_n}^{\beta_0} -\nabla \log Z(\beta)d\beta = \log Z(\beta_n) - \log Z(\beta_0)$$

4 Problem 4 [25 pts]

Consider the model with the following likelihood function:

$$p\left(X_1 = x_1, \dots X_K = x_K \mid \pi\right) = \frac{N!}{x_1! x_2! \dots x_K!} \pi_1^{x_1} \pi_2^{x_2} \dots \pi_K^{x_K}, \quad \sum_{k=1}^{K} x_k = N, \quad \sum_{k=1}^{K} \pi_k = 1$$

Let us choose the following prior for the model parameters π:

$$p\left(\pi_1, \dots \pi_k \mid \alpha^{(1)}, \alpha^{(2)}, \gamma\right) = \gamma p_1\left(\pi_1, \dots \pi_K \mid \alpha^{(1)}\right) + (1-\gamma) p_2\left(\pi_1, \dots \pi_K \mid \alpha^{(2)}\right),$$

where $p_i\left(\pi_1, \dots \pi_K \mid \alpha^{(i)}\right) \sim \mathrm{Dir}\left(\alpha^{(i)}\right),\ \alpha \in \mathbb{R}^k,\ \gamma \in (0;1)$.

Given i.i.d. observations $X = (X_1, \dots, X_k)$, calculate:

1. MLE estimate of the model parameters
2. Posterior distribution and its expectation
3. Predictive distribution

Solution

1. Likelihood maximization problem:

$$\max \frac{N!}{X_1! X_2! \dots X_K!} \pi_1^{X_1} \pi_2^{X_2} \dots \pi_K^{X_K}$$
$$s.t. \sum_{k=1}^{K} \pi_k = 1$$

Removing constant and applying logarithm, we obtain equivalent problem:

$$\max \sum_{k=1}^{K} X_k \log \pi_k$$
$$s.t. \sum_{k=1}^{K} \pi_k = 1$$

We solve it through lagrangian:

$$\sum_{k=1}^{K} X_k \log \pi_k + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) = 0$$

$$i\text{-th derivative} \frac{X_i}{\pi_i} = -\lambda$$

$$\pi_i = -\frac{X_i}{\lambda}$$

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} -\frac{X_k}{\lambda} = -\frac{N}{\lambda} = 1 \Rightarrow \lambda = -N$$

$$\pi_i = -\frac{X_i}{\lambda} = \frac{X_i}{N}$$

2. Prior distribution for Dirichlet is

$$Dir(\alpha) = \frac{1}{\mathbb{B}(\alpha)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$$

Posterior distribution

$$\begin{aligned}
P(\pi \mid X) = \frac{P(X \mid \pi)P(\pi)}{P(X)} = & \\
&= \frac{\pi_1^{X_1}\pi_2^{X_2}\dots\pi_K^{X_K}\left(\gamma\left(\frac{1}{\mathbb{B}(\alpha^1)}\prod\limits_{k=1}^{K}\pi_k^{\alpha_k^1-1}\right) + (1-\gamma)\left(\frac{1}{\mathbb{B}(\alpha^2)}\prod\limits_{k=1}^{K}\pi_k^{\alpha_k^2-1}\right)\right)}{\int \pi_1^{X_1}\pi_2^{X_2}\dots\pi_K^{X_K}\left(\gamma\left(\frac{1}{\mathbb{B}(\alpha^1)}\prod\limits_{k=1}^{K}\pi_k^{\alpha_k^1-1}\right) + (1-\gamma)\left(\frac{1}{\mathbb{B}(\alpha^2)}\prod\limits_{k=1}^{K}\pi_k^{\alpha_k^2-1}\right)\right)d\pi_1\dots d\pi_K} = \\
&= \frac{\gamma\left(\frac{1}{\mathbb{B}(\alpha^1)}\prod\limits_{k=1}^{K}\pi_k^{X_k+\alpha_k^1-1}\right) + (1-\gamma)\left(\frac{1}{\mathbb{B}(\alpha^2)}\prod\limits_{k=1}^{K}\pi_k^{X_k+\alpha_k^2-1}\right)}{\gamma\frac{\mathbb{B}(X+\alpha^1)}{\mathbb{B}(\alpha^1)} + (1-\gamma)\frac{\mathbb{B}(X+\alpha^2)}{\mathbb{B}(\alpha^2)}} = \\
&= \frac{Z_1 Dir(X+\alpha^1) + Z_2 Dir(X+\alpha^2)}{Z_1 + Z_2},
\end{aligned}$$

where $Z_i = \gamma^{2-i}(1-\gamma)^{i-1}\frac{\mathbb{B}(X+\alpha^i)}{\mathbb{B}(\alpha^i)}$

Expectation:

$$\begin{aligned}
\mathbb{E}\pi_k &= \frac{Z_1}{Z_1+Z_2}\mathbb{E}_{Dir(X+\alpha^1)}\pi_k + \frac{Z_2}{Z_1+Z_2}\mathbb{E}_{Dir(X+\alpha^2)}\pi_k = \\
&= \frac{Z_1}{Z_1+Z_2}\frac{X_k+\alpha_k^1}{N+\sum\limits_{k=1}^{K}\alpha_k^1} + \frac{Z_2}{Z_1+Z_2}\frac{X_k+\alpha_k^2}{N+\sum\limits_{k=1}^{K}\alpha_k^2}
\end{aligned}$$

3. Predictive distribution:

$$\begin{aligned}
P(X^* \mid X) = \int P(X^*|\pi)P(\pi \mid X)d\pi = & \\
&= \int \frac{N!}{X_1^*!X_2^*!\dots X_K^*!}\pi_1^{X_1^*}\dots\pi_K^{X_K^*}\frac{Z_1 Dir(X+\alpha^1) + Z_2 Dir(X^*+\alpha^2)}{Z_1+Z_2} = \\
&= A\left[\int \gamma\left(\frac{1}{\mathbb{B}(\alpha^1)}\prod_{k=1}^{K}\pi_k^{X_k^*+X_k+\alpha_k^1-1}\right)d\pi + \int(1-\gamma)\left(\frac{1}{\mathbb{B}(\alpha^2)}\prod_{k=1}^{K}\pi_k^{X_k^*+X_k+\alpha_k^2-1}\right)d\pi\right] = \\
&= A\left[\gamma\frac{\mathbb{B}(X^*+X+\alpha^1)}{\mathbb{B}(\alpha^1)} + (1-\gamma)\frac{\mathbb{B}(X^*+X+\alpha^2)}{\mathbb{B}(\alpha^2)}\right]
\end{aligned}$$

where

$$A = \frac{N!}{X_1^*!X_2^*!\dots X_K^*!}\frac{1}{\gamma\frac{\mathbb{B}(X+\alpha^1)}{\mathbb{B}(\alpha^1)} + (1-\gamma)\frac{\mathbb{B}(X+\alpha^2)}{\mathbb{B}(\alpha^2)}}$$