Homework 1

Andrey Savinov

September 20, 2019

1 Problem

1.
$$H(p_1) = -\int_X p_1 \log p_1 dx = -\int_X \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1)) \log(\exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx =$$
$$= F(\lambda_1)\int_X \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx - \int_X \langle t(x), \lambda_1\rangle \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx =$$
$$= F(\lambda_1) - \langle \lambda_1, \nabla F(\lambda_1)\rangle$$

   Because of, $F(\lambda_1) = \log \int_X \exp(\langle t(x), \lambda_1\rangle)dx$
   and $\nabla_{\lambda_1} F(\lambda_1) = \exp(-F(\lambda_1)) \int_X t(x) \exp(\langle t(x), \lambda_1\rangle)dx = \int_X t(x) \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx$

2.
$$KL(p_1||p_2) = \int_X p_1 \log \frac{p_1}{p_2} dx = \int_X p_1 \log p_1 dx - \int_X p_1 \log p_2 dx = -H(p_1) - \int_X p_1 \log p_2 dx$$
$$\int_X p_1 \log p_2 dx = \int_X \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1)) \log(\exp(\langle t(x), \lambda_2\rangle - F(\lambda_2))dx =$$
$$= -F(\lambda_2)\int_X \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx + \int_X \langle t(x), \lambda_2\rangle \exp(\langle t(x), \lambda_1\rangle - F(\lambda_1))dx = -F(\lambda_2) + \langle \lambda_2, \nabla F(\lambda_1)\rangle$$

   Using previous result from 1
$$KL(p_1||p_2) = -F(\lambda_1) + \langle \lambda_1, \nabla F(\lambda_1)\rangle + F(\lambda_2) - \langle \lambda_2, \nabla F(\lambda_1)\rangle = F(\lambda_2) - F(\lambda_1) + \langle \lambda_1 - \lambda_2, \nabla F(\lambda_1)\rangle$$

3. Let denote normalization constant as $Z = \int_X p_1^\alpha(x, \lambda_1) p_2^{1-\alpha}(x, \lambda_2)dx$. Then,
$$KL(p_1||q) = \int_X p_1 \log \frac{p_1 Z}{p_1^\alpha p_2^{1-\alpha}} dx = \int_X p_1 \log Z dx + \int_X p_1 \log p_1 dx - \alpha \int_X p_1 \log p_1 dx - (1-\alpha)\int_X p_1 \log p_2 dx =$$
$$= \log Z - H(p_1) + \alpha H(p_1) + (1-\alpha)F(\lambda_2) - (1-\alpha)\langle \lambda_2, \nabla F(\lambda_1)\rangle = \log Z + (1-\alpha)(F(\lambda_2) - F(\lambda_1)) + (1-\alpha)\langle \lambda_1 - \lambda_2, \nabla F(\lambda_1)\rangle$$
$$Z = \int_X \exp(\langle t(x), \alpha\lambda_1 + (1-\alpha)\lambda_2\rangle - \alpha F(\lambda_1) - (1-\alpha)F(\lambda_2))dx$$
$$\log Z = -\alpha F(\lambda_1) - (1-\alpha)F(\lambda_2) + \log \int_X \exp(\langle t(x), \alpha\lambda_1 + (1-\alpha)\lambda_2\rangle)dx = -\alpha F(\lambda_1) - (1-\alpha)F(\lambda_2) + F(\alpha\lambda_1 + (1-\alpha)\lambda_2)$$
$$KL(p_1||q) = -\alpha F(\lambda_1) - (1-\alpha)F(\lambda_2) + F(\alpha\lambda_1 + (1-\alpha)\lambda_2) + (1-\alpha)(F(\lambda_2) - F(\lambda_1)) + (1-\alpha)\langle \lambda_1 - \lambda_2, \nabla F(\lambda_1)\rangle =$$
$$= F(\alpha\lambda_1 + (1-\alpha)\lambda_2) - F(\lambda_1) + (1-\alpha)\langle \lambda_1 - \lambda_2, \nabla F(\lambda_1)\rangle$$

2 Problem

The probability density function of α is

$$f(x) = \begin{cases} 0, \; x \notin [a,b] \\ \frac{1}{b-a}, \; x \in [a,b]. \end{cases}$$

$G(y)$ is cumulative distribution function of α^2. $G(y) = 0, \; y < 0$. For $y \geq 0$,

$$G(y) = P(\alpha^2 \leq y) = \int_{x^2<y} f(x)dx = \int_{-\sqrt{y}}^{\sqrt{y}} f(x)dx.$$

$g(y) = G^{'}(y)$ is probability density function of α^2.

$$g(y) = \begin{cases} 0, \; y < 0 \\ \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y})), \; y \geq 0. \end{cases}$$

3 Problem

1. MLE estimate of the model parameters

   m - particular sample, M - number of samples: $x^m = [x_1^m, x_2^m, \ldots, x_K^m], \; m = 1, \ldots, M$

$$P(X|\pi) = \prod_{m=1}^{M} N! \prod_{i=1}^{K} \frac{\pi_i^{x_i^m}}{x_i^m!}$$

$$\pi_{MLE} = \arg\max_{\pi} P(X|\pi) = \arg\max_{\pi} \log P(X|\pi)$$

$$\log P(X|\pi) = \sum_{m=1}^{M} [\log N! + \sum_{i=1}^{K} (x_i^m \log \pi_i - \log x_i^m!)].$$

   Method of Lagrange multipliers:

$$L(\pi, \lambda) = \sum_{m=1}^{M} [\log N! + \sum_{i=1}^{K} (x_i^m \log \pi_i - \log x_i^m!)] + \lambda(1 - \sum_{i=1}^{K} \pi_i) \to \max$$

$$\frac{\partial L}{\partial \pi_j} = \frac{1}{\pi_j} \sum_{m=1}^{M} x_j^m - \lambda = 0, \; j = 1, \ldots, K$$

$$1 - \sum_{i=1}^{K} \pi_i = 0$$

$$\lambda = \sum_{i=1}^{K} \sum_{m=1}^{M} x_i^m$$

$$\pi_j^{MLE} = \frac{\sum_{m=1}^{M} x_j^m}{\sum_{i=1}^{K} \sum_{m=1}^{M} x_i^m}, \; j = 1, \ldots, K$$

2. Posterior distribution and its expectation

$$P(\pi|X,\alpha) \sim P(X|\pi)P(\pi|\alpha) = \left[\prod_{m=1}^{M} N! \prod_{i=1}^{K} \frac{\pi_i^{x_i^m}}{x_i^m!}\right] \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1} \sim const \prod_{i=1}^{K} \pi_i^{\alpha_i - 1 + \sum_{m=1}^{M} x_i^m}$$

   Again Dirichlet distribution.

$$P(\pi|X,\alpha) = \frac{\Gamma(\sum_{i=1}^{K} [\alpha_i + \sum_{m=1}^{M} x_i^m])}{\prod_{i=1}^{K} \Gamma(\alpha_i + \sum_{m=1}^{M} x_i^m)} \prod_{i=1}^{K} \pi_i^{\alpha_i - 1 + \sum_{m=1}^{M} x_i^m}$$

Let's denote $\mu_i = \alpha_i + \sum_{m=1}^{M} x_i^m$

$$\langle \pi_j \rangle_{P(\pi|X,\alpha)} = \frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \int_\Pi \prod_{i \neq j} \pi_i^{\mu_i - 1} \pi_j^{\mu_j - 1 + 1} d\pi = \frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \frac{\prod_{i \neq j} \Gamma(\mu_i) \Gamma(\mu_j + 1)}{\Gamma(\sum_{i=1}^{K} \mu_i + 1)}$$

Due to property of Gamma function: $\Gamma(z+1) = z\Gamma(z)$:

$$\frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \frac{\prod_{i=1}^{K} \Gamma(\mu_i) \mu_j}{\Gamma(\sum_{i=1}^{K} \mu_i) \sum_{i=1}^{K} \mu_i} = \frac{\mu_i}{\sum_{i=1}^{K} \mu_i} = \frac{\alpha_i + \sum_{m=1}^{M} x_i^m}{\sum_{i=1}^{K} [\alpha_i + \sum_{m=1}^{M} x_i^m]} = \langle \pi_j \rangle_{P(\pi|X,\alpha)}$$

3. MAP estimate

$$\pi_{MAP} = \arg\max_\pi P(\pi|X,\alpha) = \arg\max_\pi \log P(\pi|X,\alpha).$$

$$\log P(\pi|X,\alpha) = \log \Gamma(\sum_{i=1}^{K} \mu_i) - \log \prod_{i=1}^{K} \Gamma(\mu_i) + \sum_{i=1}^{K} (\mu_i - 1) \log \pi_i$$

Method of Lagrange multipliers:

$$L(\pi, \lambda) = \log \Gamma(\sum_{i=1}^{K} \mu_i) - \log \prod_{i=1}^{K} \Gamma(\mu_i) + \sum_{i=1}^{K} (\mu_i - 1) \log \pi_i + \lambda(1 - \sum_{i=1}^{K} \pi_i) \to \max$$

$$\frac{\partial L}{\partial \pi_j} = \frac{1}{\pi_j}(\mu_j - 1) - \lambda = 0, \ j = 1, \ldots, K$$

$$1 - \sum_{i=1}^{K} \pi_i = 0$$

$$\lambda = \sum_{i=1}^{K} (\mu_i - 1)$$

$$\pi_j^{MAP} = \frac{\mu_j - 1}{\sum_{i=1}^{K} (\mu_i - 1)} = \frac{\alpha_j - 1 + \sum_{m=1}^{M} x_j^m}{\sum_{i=1}^{K} [\alpha_i - 1 + \sum_{m=1}^{M} x_i^m]}, \ j = 1, \ldots, K$$

4. Predictive distribution. Notation $\mu_i = \alpha_i + \sum_{m=1}^{M} x_i^m$.

$$P(x^*|X,\alpha) = \int_\Pi P(x^*|\pi) P(\pi|X,\alpha) d\pi = \int_\Pi N! \prod_{i=1}^{K} \frac{\pi_i^{x_i^*}}{x_i^*!} \frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \prod_{i=1}^{K} \pi_i^{\mu_i - 1} d\pi =$$

$$= N! \frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \left[\prod_{i=1}^{K} \frac{1}{x_i^*!} \right] \int_\Pi \prod_{i=1}^{K} \pi_i^{\mu_i + x_i^* - 1} d\pi = N! \frac{\Gamma(\sum_{i=1}^{K} \mu_i)}{\prod_{i=1}^{K} \Gamma(\mu_i)} \left[\prod_{i=1}^{K} \frac{1}{x_i^*!} \right] \frac{\prod_{i=1}^{K} \Gamma(\mu_i + x_i^*)}{\Gamma(\sum_{i=1}^{K} [\mu_i + x_i *])}$$

$$P(x^*|X,\alpha) = N! \frac{\Gamma(\sum_{i=1}^{K} [\alpha_i + \sum_{m=1}^{M} x_i^m])}{\prod_{i=1}^{K} \Gamma(\alpha_i + \sum_{m=1}^{M} x_i^m)} \frac{\prod_{i=1}^{K} \Gamma(\alpha_i + \sum_{m=1}^{M} x_i^m + x_i^*)}{\Gamma(\sum_{i=1}^{K} [\alpha_i + \sum_{m=1}^{M} x_i^m + x_i *])} \left[\prod_{i=1}^{K} \frac{1}{x_i^*!} \right]$$

All sums like $\sum_{i=1}^{K} \sum_{m=1}^{M} x_i^m$ can be simplified up to:

$$\sum_{i=1}^{K} \sum_{m=1}^{M} x_i^m = \sum_{m=1}^{M} N = MN$$

5. Re-do points (2-4) with slightly another prior: $P(\pi|\alpha^{(1)}, \alpha^{(2)}) = \gamma P_1(\pi, \alpha^{(1)}) + (1-\gamma) P_2(\pi, \alpha^{(2)}), \ P_i \sim (\alpha^{(i)},)$

- Posterior distribution and its expectation

$$P(\pi|X,\alpha^{(1)},\alpha^{(2)}) \sim P(X|\pi)[\gamma P(\pi|\alpha^{(1)}) + (1-\gamma)P(\pi,\alpha^{(2)})]$$

Due to result of point 2 normalization is done automatically, $P(\pi|X,\alpha)$ is given in point 2:

$$P(\pi|X,\alpha^{(1)},\alpha^{(2)}) = \gamma P(\pi|X,\alpha^{(1)}) + (1-\gamma)P(\pi|X,\alpha^{(2)})$$

$$\langle \pi_j \rangle_{P(\pi|X,\alpha^{(1)},\alpha^{(2)})} = \gamma \langle \pi_j \rangle_{P(\pi|X,\alpha^{(1)})} + (1-\gamma)\langle \pi_j \rangle_{P(\pi|X,\alpha^{(2)})}$$

$\langle \pi_j \rangle_{P(\pi|X,\alpha)}$ is given in point 2.

- MAP estimate

$$\pi_{MAP} = \arg\max_{\pi} P(\pi|X,\alpha_1,\alpha_2) = \arg\max_{\pi} \log P(\pi|X,\alpha_1,\alpha_2) = \arg\max_{\pi} \log\left[\gamma P(\pi|X,\alpha^{(1)}) + (1-\gamma)P(\pi|X,\alpha^{(2)})\right].$$

Unfortunately, I don't know how to optimize logarithm of sum without numerical method.

- Predictive distribution

$$P(x^*|X,\alpha^{(1)},\alpha^{(2)}) = \int_{\Pi} P(x^*|\pi)P(\pi|X,\alpha^{(1)},\alpha^{(2)})d\pi = \int_{\Pi} P(x^*|\pi)[\gamma P(\pi|X,\alpha^{(1)}) + (1-\gamma)P(\pi|X,\alpha^{(2)})]d\pi$$

$$P(x^*|X,\alpha^{(1)},\alpha^{(2)}) = \gamma P(x^*|X,\alpha^{(1)}) + (1-\gamma)P(x^*|X,\alpha^{(2)})$$

$P(x^*|X,\alpha)$ is given in point 4.

4 Problem

Since $f^*(x) = x\log x$, then $D_{f^*}(p||q) = \langle f^*(\frac{p}{q})\rangle_q = \int q\frac{p}{q}\log\frac{p}{q}dx = \int p\log\frac{p}{q}dx = KL(p||q)$. I will use log-sum inequality for non-negative numbers:

$$\sum_{i=1}^{n} a_i \log\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \le \sum_{i=1}^{n} a_i \log\frac{a_i}{b_i}.$$

Let p_1, p_2, q_1, q_2 - probability density functions, $\lambda \in (0,1)$.

$$D_{f^*}(\lambda p_1+(1-\lambda)p_2||\lambda q_1+(1-\lambda)q_2) = \langle f^*(\frac{p}{q})\rangle_q = KL(\lambda p_1+(1-\lambda)p_2||\lambda q_1+(1-\lambda)q_2) = \int [\lambda p_1+(1-\lambda)p_2]\log\frac{\lambda p_1+(1-\lambda)p_2}{\lambda q_1+(1-\lambda)q_2}dx \le$$

$$\le \int \lambda p_1 \log\frac{\lambda p_1}{\lambda q_1}dx + \int (1-\lambda)p_2\log\frac{(1-\lambda)p_2}{(1-\lambda)q_2}dx = \lambda D_{f^*}(p_1||q_1) + (1-\lambda)D_{f^*}(p_2||q_2).$$

Then for $f^*(x) = x\log x$, $D_{f^*}(p||q)$ - is convex over both p,q.

5 Problem

Since $p_i(x)$ is probability distribution, $\int_X p_i(x)dx = 1$.

$$\int_X p_i(x)dx = \int_X \frac{1}{Z(\beta_i)}\pi(x)\exp(-\beta_i h(x))dx = 1$$

Consequently, $Z(\beta_i) = \int_X \pi(x)\exp(-\beta_i h(x))dx$ and $\nabla_\beta Z(\beta_i) = -\int_X h(x)\pi(x)\exp(-\beta_i h(x))dx = -Z(\beta_i)\int_X h(x)p_i(x)dx$

$$-\frac{\nabla_\beta Z(\beta_i)}{Z(\beta_i)} = \int_X h(x)p_i(x)dx = \langle h(x)\rangle_{p_{\beta_i}}$$

$$\int_{\beta_n}^{\beta_0} -\frac{\nabla_\beta Z(\beta)}{Z(\beta)}d\beta = \int_{\beta_n}^{\beta_0}\langle h(x)\rangle_{p_\beta}d\beta$$

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