AI 701 Bayesian machine learning, Fall 2020

Homework assignment 1

1. (20 points) Let X be a $\mathbb{R}$-valued continuous random variable with Probability Density Function (PDF)

$$f(x) = \begin{cases} \exp(\theta - x), & x > \theta, \\ 0, & x \leq \theta. \end{cases} \tag{1}$$

   (a) (3 points) Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

   (b) (3 points) Let $X_1, \ldots, X_n$ be i.i.d. copies of X. Show that

$$\hat{\theta}_n := \frac{1}{n}\sum_{i=1}^{n}(X_i - 1), \tag{2}$$

   is an unbiased estimator of θ.

   (c) (7 points) Derive the $100(1-\alpha)\%$ confidence interval of θ for $\alpha \in (0, 1)$.

   (d) (7 points) Assume we have a set of observations $\mathcal{D} = \{10.0, 12.0, 15.0\}$. Using the fact that

$$\int_{-\infty}^{1.96} \mathcal{N}(x|0,1)\mathrm{d}x \approx 0.975, \tag{3}$$

   compute the 95% confidence interval of θ. Do you find anything counter-intuitive? If so, what is weird about the confidence interval?

2. (10 points) Let $(E, \mathcal{E}, \mu)$ be a measure space. Using the definition of measures, show that

   1. For any $A \subset B$, $\mu(A) \leq \mu(B)$.
   2. For any $A_1, \ldots, A_n$, $\mu(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n, \mu(A_i)$.

3. (5 points) Let $(X_i)_{i\geq 1}$ be an i.i.d. sequence of random variables with $\mathbb{E}[X_1] < \infty$ and $\mathrm{Var}(X_1) < \infty$. Show that

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\mathrm{p}} \mathbb{E}[X_1]. \tag{4}$$

   Hint: use Chebyshev's inequality, for a random variable X with $\mathbb{E}[X] = \mu < \infty$ and $\mathrm{Var}(X) = \sigma^2 < \infty$,

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \tag{5}$$

4. (15 points) Let $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ be a sequence of random variables defined on a common probability space. Show that if $X_n \xrightarrow{\mathrm{p}} X$ and $Y_n \xrightarrow{\mathrm{p}} Y$ for some random variables X and Y, then $X_n + Y_n \xrightarrow{\mathrm{p}} X + Y$.

5. (15 points) Let $X_i \sim \mathrm{Gamma}(\alpha_i, 1)$ for $i = 1, \ldots, n$ with PDF

$$f_X(x) = \frac{x^{\alpha_i - 1}e^{-x}}{\Gamma(\alpha_i)}, \tag{6}$$

where

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \mathrm{d}t, \tag{7}$$

denotes the gamma function. Show that

$$Y := \left(\frac{X_1}{\sum_{i=1}^n X_i}, \dots, \frac{X_n}{\sum_{i=1}^n X_i} \right), \tag{8}$$

is a Dirichlet random variable with PDF

$$f(y) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n y_i^{\alpha_i - 1}. \tag{9}$$

Hint: use the change of variable

$$(X_1, \dots, X_n) \to (Y_1, \dots, Y_{n-1}, Z), \tag{10}$$

where

$$Y_i = \frac{X_i}{Z} \text{ for } i = 1, \dots, n-1, \quad Z = \sum_{i=1}^n X_i. \tag{11}$$

6. (20 points) Consider the following Gaussian-Gaussian-inverse-Wishart model.

$$\theta := (\mu, \Sigma), \tag{12}$$

$$p(x|\theta) = \mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu) \right) \tag{13}$$

$$\begin{aligned} p(\theta) &= \mathcal{N}(\mu|m, \Sigma/r) \mathcal{W}^{-1}(\Sigma|\Psi, \nu) \\ &= \frac{r^{d/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{r}{2}(\mu - m)^\top \Sigma^{-1} (\mu - m) \right) \\ &\times \frac{|\Psi|^{\nu/2}}{2^{\nu d/2} \Gamma_d(\nu/2)} |\Sigma|^{-\frac{\nu+d+1}{2}} \exp\left(-\frac{1}{2} \mathrm{tr}(\Psi \Sigma^{-1}) \right), \end{aligned} \tag{14}$$

where

$$\Gamma_d(z) := \pi^{\frac{d(d-1)}{4}} \prod_{j=1}^d \Gamma(z + (1-j)/2). \tag{15}$$

Let $X := \{x_1, \dots, x_n\}$ be a set of observations.

(a) (7 points) Compute the posterior $p(\theta|X)$. (Hint: use the fact $x^\top A y = \mathrm{tr}(A y x^\top)$).

(b) (6 points) Compute the marginal likelihood $p(X)$.

(c) (7 points) Show that $p(x|\theta)$ and $p(\theta)$ belong to conjugate exponential families. That is, show that they can be represented as

$$p(x|\theta) = \exp(T(x)^\top \eta - \mathbb{1}^\top A(\eta) - B(x)) \tag{16}$$

$$p(\theta) = \exp(\eta^\top \chi - \zeta^\top A(\eta) - \mathbb{1}^\top C(\chi, \zeta)). \tag{17}$$

(Hint: use the fact $\mathrm{tr}(A^\top B) = \mathrm{vec}(A)^\top \mathrm{vec}(B)$.

7. (15 points) Define U_1, U_2 as follows:

$$U_1, U_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}(0,1), \quad \text{accept } (U_1, U_2) \text{ if } U_1^2 + U_2^2 \leq 1. \tag{18}$$

Show that

$$X_1 = U_1 \sqrt{\frac{-2\log(U_1^2 + U_2^2)}{U_1^2 + U_2^2}} \sim \mathcal{N}(0,1), \quad X_2 = U_2 \sqrt{\frac{-2\log(U_1^2 + U_2^2)}{U_1^2 + U_2^2}} \sim \mathcal{N}(0,1). \tag{19}$$

Hint: use the intermediate transform

$$U_1 = R\cos\Theta, \quad U_2 = R\sin\Theta, \quad 0 \leq R \leq 1. \tag{20}$$