AI 701 Bayesian machine learning, Fall 2021

Homework assignment 1

1. (15 points) Suppose we toss a coin n times. Let X be a random variable denoting the number of heads among n coin tosses. We simply assume that X follows the binomial distribution with parameter θ,

$$\mathbb{P}(X = x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}. \tag{1}$$

   (a) (5 points) We assume that θ itself is a random variable following the beta distribution with Probability Density Function (PDF) with parameters $a, b > 0$,

$$f(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{\{0\le\theta\le1\}}, \tag{2}$$

   where $\Gamma(t) = \int_0^\infty z^{t-1}e^{-z}\mathrm{d}z$ is the gamma function. Compute the marginal likelihood of x under this prior setting. That is, compute

$$\mathbb{P}(X = x; a, b) = \int_0^1 \mathbb{P}(X = x|\theta)f(\theta; a, b)\mathrm{d}\theta. \tag{3}$$

   (b) (10 points) Assume we toss the coin $n = 10$ times and see $x = 9$ heads. To see if the coin is fair, we compare two prior distributions for θ:

$$\mathcal{M}_1 : \mathbb{P}(\theta) = \delta_{1/2}(\theta), \quad \mathcal{M}_2 : f(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{\{0\le\theta\le1\}}. \tag{4}$$

   Set $a = b = 1$ and compute the Bayes factor

$$\frac{\mathbb{P}(X = x|\mathcal{M}_1)}{\mathbb{P}(X = x|\mathcal{M}_2)}. \tag{5}$$

   (Hint) Use the fact that $\Gamma(n) = (n-1)!$ for all $n \in \mathbb{N}$.

2. (15 points) Let E be a set and $\mathcal{A}$ be a collection of subsets of E. The σ-algebra generated by $\mathcal{A}$, denoted by $\sigma(\mathcal{A})$, is the intersection of all σ-algebras containing $\mathcal{A}$, or equivalently, the smallest σ-algebra containing $\mathcal{A}$.

   (a) (5 points) Let $E = \{1, 2, 3, 4\}$ and $\mathcal{A} = \{\{1\}, \{2\}\}$. Find $\sigma(\mathcal{A})$.

   (b) (10 points) Let E be a set and $\mathcal{A}, \mathcal{C}$ be collections of subsets of E. Prove the followings.

      1. $\mathcal{A} \subset \mathcal{C} \implies \sigma(\mathcal{A}) \subset \sigma(\mathcal{C})$.
      2. $\mathcal{A} \subset \sigma(\mathcal{C}),\ \mathcal{C} \subset \sigma(\mathcal{A}) \implies \sigma(\mathcal{A}) = \sigma(\mathcal{C})$.
      3. $\mathcal{A} \subset \mathcal{C} \subset \sigma(\mathcal{A}) \implies \sigma(\mathcal{A}) = \sigma(\mathcal{C})$.

3. (25 points) Let X be a $\mathbb{R}$-valued random variable. The Moment Generating Function (MGF) of X is defined as

$$M_X(t) = \mathbb{E}[e^{tX}]. \tag{6}$$

   An important property of MGF is that it uniquely determines a distribution of a random variable. That is, given two $\mathbb{R}$-valued random variables X and Y (assume that there exist MGF for both X and Y),

$$\forall t, M_X(t) = M_Y(t) \implies \forall x, F_X(x) = F_Y(x). \tag{7}$$

   One can also show that, if $(X_n)_{n\ge1}$ is a sequence of random variables whose MGF $M_{X_n}(t)$ converges to the MGF of a random variable X, then $(X_n)_{n\ge1}$ converges in distribution to X.

(a) (5 points) Compute the MGF of a standard normal random variable X with PDF,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{8}$$

(b) (5 points) For $n \in \mathbb{N}$, let $X_1, \ldots, X_n$ be i.i.d. $\mathbb{R}$-valued random variables, and let $M_{X_1}(t)$ be the MGF of X_1. Derive the MGF of $Y_n := c(X_1 + \cdots + X_n)$.

(c) (15 points) Let X be a $\mathbb{R}$-valued random variable with mean μ and variance σ^2. For $n \in \mathbb{N}$, define

$$Y_i := \frac{X_i - \mu}{\sigma}, \quad Z_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i. \tag{9}$$

Show that Z_n converges in distribution to a standard normal distribution as $n \to \infty$.
(Hint) Show that the MGF of Z_n converges to the MGF of a standard normal distribution as $n \to \infty$. Use the Taylor's theorem to analyze the limit.

4. (25 points) Let X be a random variable with finite mean μ and variance σ^2. Let $X_1, \ldots, X_n$ be i.i.d. copies of X, and consider the following estimator of σ:

$$\hat{\sigma}_n^2 := \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2, \quad \bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{10}$$

(a) (10 points) Show that $\hat{\sigma}_n^2$ is consistent; that is, show that $\hat{\sigma}_n^2 \xrightarrow{\text{p}} \sigma^2$ as $n \to \infty$.

(b) (10 points) Show that $\hat{\sigma}_n^2$ is biased; that is, show that $\mathbb{E}[\hat{\sigma}_n^2] \neq \sigma^2$.

(c) (5 points) Propose a simple fix to make $\hat{\sigma}_n^2$ unbiased.

5. (20 points) Let X be a $\mathbb{R}$-valued random variable. Show that $X_n \xrightarrow{\text{d}} c$ where c is a constant implies $X_n \xrightarrow{\text{p}} c$.