CSC 665: Calibration Homework

Chicheng Zhang

August 26, 2019

Please complete the following set of exercises **on your own**. The homework is due **on Sep 3, in class**. You may find the following version of Taylor's Theorem in multivariate calculus helpful:

Theorem 1. *Suppose f is twice differentiable in $\mathbb{R}^d$. Then given two points a, b in $\mathbb{R}^d$, there exists some t in $[0,1]$, such that*

$$f(b) = f(a) + \langle \nabla f(a), b-a \rangle + \frac{1}{2}(b-a)^\top \nabla^2 f(\xi)(b-a),$$

where $\xi = ta + (1-t)b$. Here $\nabla f(x) \triangleq (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d})$ is the gradient of f at x, and

$$\nabla^2 f(x) \triangleq \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ & \cdots & \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

is the Hessian of f at x.

Problem 1

Denote by $\mathrm{B}(n,p)$ the binomial distribution with n being the number of trials, and p being the success probability of each trial, and denote by $\mathrm{N}(\mu,\sigma^2)$ the normal distribution with mean μ and variance σ^2.

1. Suppose Y is a random variable such that $P(Y=+1) = P(Y=-1) = \frac{1}{2}$. In addition, given Y, X has the following conditional probability distribution: given $Y=-1$, $X \sim \mathrm{B}(3, \frac{2}{3})$; given $Y=+1$, $X \sim \mathrm{B}(2, \frac{1}{3})$. Calculate:
   (a) the joint probability table of (X,Y);
   (b) $P(Y=1|X=3)$;
   (c) $P(Y=-1|X=1)$.

2. Suppose Y is a random variable such that $P(Y=+1) = P(Y=-1) = \frac{1}{2}$. In addition, suppose given Y, X has the following conditional probability distribution: given $Y=-1$, $X \sim \mathrm{N}(\mu_-, \sigma^2)$; given $Y=+1$, $X \sim \mathrm{N}(\mu_+, \sigma^2)$. Define

   $$P(Y=+1|x) \triangleq \frac{P(Y=+1)p_{+1}(x)}{P(Y=+1)p_{+1}(x) + P(Y=-1)p_{-1}(x)}.$$

   where p_{+1} and p_{-1} are the conditional probability density functions of X given $Y=+1$ and $Y=-1$ respectively. Show that

   $$P(Y=+1|x) = \frac{1}{1+\exp\Big(-\frac{\mu_+-\mu_-}{\sigma^2}\cdot(x-\frac{\mu_++\mu_-}{2})\Big)}.$$

   (Remark: $P(Y=+1|x)$ has the intuitive interpretation that it is the conditional probability of $Y=+1$ given $X=x$. It can be shown rigorously that $P(Y=+1|x) = \lim_{\epsilon\to 0} P(Y=+1|X\in[x-\epsilon, x+\epsilon])$.)

Problem 2

1. Suppose $D = \mathrm{U}([0,1])$, i.e. the uniform distribution over the $[0,1]$ interval. Consider a set of samples $S = (X_1, \dots, X_n)$ drawn identically and independently from distribution D.

   Write a program that plots the empirical *cumulative distribution function* (CDF) of the sample S, that is,

   $$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i \le t), t \in \mathbb{R},$$

   where

   $$\mathbf{1}(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases}$$

   is the indicator function. You may use any programming languages you like (I recommend using Python and Jupyter notebook).

   Draw two sets of samples S_1 and S_2 of size $n = 5$. Plot F_n^1, the CDF of S_1, and plot F_n^2, the CDF of S_2. Are they different? Why?

2. Repeat the same experiment in for $n = 100$ and $n = 1000$. Do the F_n^1 and F_n^2 functions become closer as n increases?

3. In the above experiment, as n goes to infinity, what function does F_n converge to? Can you derive a formula for that function (denoted as F)?

4. Suppose D is the the standard normal distribution $\mathrm{N}(0,1)$, what function does F_n converge to?

Problem 3

1. Define function

   $$h(x) \triangleq x \ln x + (1-x)\ln(1-x), x \in (0,1).$$

   Show that for any p and q in $(0,1)$,

   $$h(q) - h(p) - h'(p)(q-p) = p \ln \frac{p}{q} + (1-p)\ln \frac{1-p}{1-q}.$$

   (Remark: the expression on the right hand side is often called the *binary relative entropy*, denoted as $\mathrm{kl}(p,q)$; the h function is often called the *negative binary entropy*.)

2. Suppose $0 < p < q < 1$. Use Taylor's Theorem to show that

   $$\mathrm{kl}(p,q) \ge 2(p-q)^2.$$

   Furthermore, show that

   $$\mathrm{kl}(p,q) \ge \frac{(p-q)^2}{2q}.$$

3. Define the m-dimensional probability simplex Δ^{m-1} as $\left\{p \in \mathbb{R}^m : \text{ for all } i, p_i \ge 0, \sum_{i=1}^m p_i = 1\right\}$. For two vectors p, q in Δ^{m-1}, define the negative entropy of p as:

   $$\mathrm{H}(p) \triangleq \sum_{i=1}^{m} p_i \ln p_i,$$

and the relative entropy between p and q as:

$$\mathrm{KL}(p, q) \triangleq \sum_{i=1}^{m} p_i \ln \frac{p_i}{q_i}.$$

Verify that

$$\mathrm{H}(q) - \mathrm{H}(p) - \left\langle \nabla \mathrm{H}(p), q - p \right\rangle = \mathrm{KL}(p, q).$$

4. Using Taylor's Theorem, show that for any p, q in Δ^{m-1}, $\mathrm{KL}(p, q) \geq 0$. Furthermore, show that $\mathrm{KL}(p, q) \geq \frac{1}{2}(\sum_{i=1}^{m} |p_i - q_i|)^2$.

   Hint: at some point, you may want to use the following variant of Cauchy-Schwarz inequality:

$$(\sum_{i=1}^{m} y_i) \cdot (\sum_{i=1}^{m} \frac{x_i^2}{y_i}) \geq (\sum_{i=1}^{m} |x_i|)^2.$$