

CSC 665: Calibration Homework

Chicheng Zhang

September 8, 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 $\nabla f(x) \triangleq (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d})$ is the gradient of f at x , and

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

is the Hessian of f at x .

Problem 1

Denote by $B(n, p)$ the binomial distribution with n being the number of trials, and p being the success probability of each trial, and denote by $N(\mu, \sigma^2)$ 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 $Y = -1$, $X \sim B(3, \frac{2}{3})$; given $Y = +1$, $X \sim B(2, \frac{1}{3})$. 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 $Y = -1$, $X \sim N(\mu_-, \sigma^2)$; given $Y = +1$, $X \sim N(\mu_+, \sigma^2)$. 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

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

(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 $P(Y = +1|x) = \lim_{\epsilon \rightarrow 0} P(Y = +1|X \in [x - \epsilon, x + \epsilon])$.)

Solution

1. (a) The joint probability table is

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = -1$	$\frac{1}{54}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{4}{27}$
$Y = +1$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{18}$	0

(b)

$$P(Y = 1|X = 3) = \frac{P(Y = 1, X = 3)}{P(X = 3)} = \frac{P(Y = 1, X = 3)}{P(Y = 1, X = 3) + P(Y = -1, X = 3)} = \frac{0}{0 + \frac{4}{27}} = 0.$$

(c)

$$P(Y = -1|X = 1) = \frac{P(Y = -1, X = 1)}{P(X = 1)} = \frac{P(Y = -1, X = 1)}{P(Y = 1, X = 1) + P(Y = -1, X = 1)} = \frac{\frac{1}{9}}{\frac{1}{9} + \frac{2}{9}} = \frac{1}{3}.$$

2. Recall that by the definition of normal distribution, $p_{+1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_+)^2}{2\sigma^2}\right\}$, and $p_{-1}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu_-)^2}{2\sigma^2}\right\}$.

$$\text{Therefore, } p_{-1}(x)/p_{+1}(x) = \exp\left\{-\frac{(x-\mu_-)^2 - (x-\mu_+)^2}{2\sigma^2}\right\} = \exp\left\{-\frac{(\mu_+ - \mu_-)(x - \frac{\mu_+ + \mu_-}{2})}{\sigma^2}\right\}.$$

In addition, $P(Y = +1) = P(Y = -1) = \frac{1}{2}$. Therefore,

$$\begin{aligned} P(Y = +1|x) &= \frac{P(Y = +1)p_{+1}(x)}{P(Y = +1)p_{+1}(x) + P(Y = -1)p_{-1}(x)} \\ &= \frac{p_{+1}(x)}{p_{+1}(x) + p_{-1}(x)} \\ &= \frac{1}{1 + \frac{p_{-1}(x)}{p_{+1}(x)}} \\ &= \frac{1}{1 + \exp\left\{-\frac{(\mu_+ - \mu_-)(x - \frac{\mu_+ + \mu_-}{2})}{\sigma^2}\right\}}. \end{aligned}$$

Problem 2

1. Suppose $D = U([0, 1])$, 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,

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

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 (e.g. 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 1 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 standard normal distribution $N(0, 1)$, what function does F_n converge to?

Solution

1. See Figure 1 below. (The exact results depend on the random points drawn - they can vary case by case.) They are different. As S_1 and S_2 are two different samples, their induced CDFs would also be different.

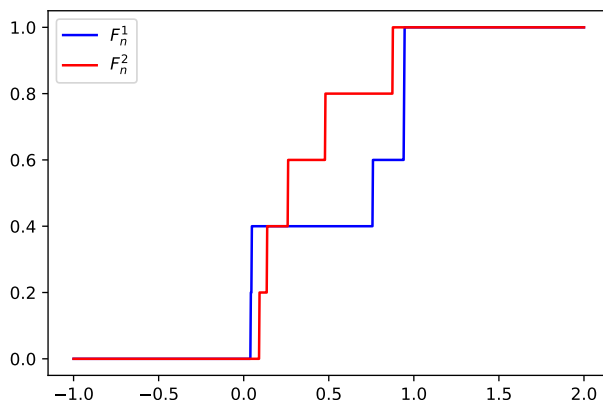


Figure 1: A random draw of F_n^1 and F_n^2 with sample size $n = 5$.

2. See Figures 2 and 3 below. Yes, F_n^1 and F_n^2 do become closer as n increases. (Of course, in some extremely unlikely draws of random samples, they *can* become farther away as n increases.)
3. By the (weak) law of large numbers, for every t in \mathbb{R} , $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i \leq t) \rightarrow F(t) = \mathbb{P}(X \leq t)$ (in probability).

$$F(t) = \mathbb{P}(X \leq t) = \begin{cases} 0 & t \leq 0 \\ t & 0 < t < 1 \\ 1 & t \geq 1 \end{cases}.$$

4. When X is drawn from the standard normal distribution,

$$F(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

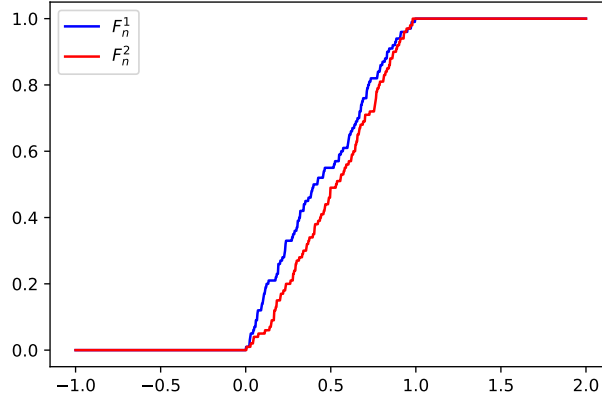


Figure 2: A random draw of F_n^1 and F_n^2 with sample size $n = 100$.

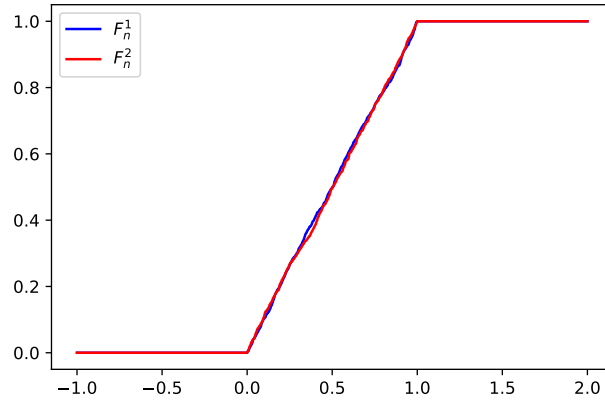


Figure 3: A random draw of F_n^1 and F_n^2 with sample size $n = 1000$.

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 $(0, 1)$,

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

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

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

$$\text{kl}(p, q) \geq 2(p - q)^2.$$

Furthermore, show that

$$\text{kl}(p, q) \geq \frac{(p - q)^2}{2q}.$$

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

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

and the relative entropy between p and q as:

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

Verify that

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

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

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

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

Solution

1. Observe that $h'(x) = (1 + \ln x) - (1 + \ln(1 - x)) = \ln x - \ln(1 - x)$.

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

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

2. Observe that $h(x)$ is twice differentiable on $(0, 1)$. Then, by Taylor's Theorem, for any p, q ,

$$\text{kl}(p, q) = h(p) - h(q) - h'(q)(p - q) = \frac{1}{2}h''(\xi)(p - q)^2,$$

for some $\xi = (1 - t)p + tq$, where $t \in [0, 1]$. As both p and q are in $(0, 1)$, $\xi \in (p, q) \subset (0, 1)$.

Now, observe that for all x in $(0, 1)$, $h''(x) = \frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)} \geq \frac{1}{((1-x+x)/2)^2} = 4$. This implies that

$$\text{kl}(p, q) = \frac{1}{2}h''(\xi)(p - q)^2 \geq \frac{1}{2} \cdot 4(p - q)^2 = 2(p - q)^2.$$

The second inequality follows from the fact that $h''(\xi) = \frac{1}{\xi(1-\xi)} \geq \frac{1}{\xi} \geq \frac{1}{q}$. This implies that

$$\text{kl}(p, q) = \frac{1}{2}h''(\xi)(p - q)^2 \geq \frac{(p - q)^2}{2q}.$$

3. Observe that $\nabla H(q) = (\frac{\partial H}{\partial x_1}, \dots, \frac{\partial H}{\partial x_m}) = (1 + \ln q_1, \dots, 1 + \ln q_m)$.

$$\begin{aligned}
H(p) - H(q) - \langle \nabla H(q), p - q \rangle &= \sum_{i=1}^m p_i \ln p_i - \sum_{i=1}^m q_i \ln q_i - \sum_{i=1}^m (p_i - q_i)(1 + \ln q_i) \\
&= \sum_{i=1}^m (p_i \ln p_i - q_i \ln q_i - (p_i - q_i) \ln q_i) - \sum_{i=1}^m (p_i - q_i) \\
&= \sum_{i=1}^m p_i \ln \frac{p_i}{q_i} - \sum_{i=1}^m (p_i - q_i),
\end{aligned}$$

As both p_i and q_i are in Δ^{m-1} , $\sum_{i=1}^m p_i = \sum_{i=1}^m q_i = 1$. Therefore,

$$H(p) - H(q) - \langle \nabla H(q), p - q \rangle = \sum_{i=1}^m p_i \ln \frac{p_i}{q_i} = \text{KL}(p, q).$$

4. Observe that $H(p)$ is twice differentiable on $(0, 1)^m$. Then, by Talyor's Theorem, for any p, q in $(0, 1)^m$,

$$H(p) - H(q) - \langle \nabla H(q), p - q \rangle = \frac{1}{2} (p - q)^\top \nabla^2 H(\xi) (p - q).$$

for some $\xi = (1 - t)p + tq$ and $t \in [0, 1]$.

Here, the entries of Hessian of H can be written as the following:

$$(\nabla^2 H(x))_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} \frac{1}{x_i}, & i = j \\ 0, & i \neq j \end{cases}.$$

Now, for p and q in Δ^{m-1} , we have that

$$\begin{aligned}
\text{KL}(p, q) &= H(p) - H(q) - \langle \nabla H(q), p - q \rangle \\
&= \frac{1}{2} (p - q)^\top \nabla^2 H(\xi) (p - q) \\
&= \frac{1}{2} \sum_{i=1}^m \frac{(p_i - q_i)^2}{\xi_i}, \tag{1}
\end{aligned}$$

for some $\xi = (1 - t)p + tq$ and $t \in [0, 1]$. Specifically, all ξ_i 's are nonnegative, and $\sum_{i=1}^m \xi_i = (1 - t) \sum_{i=1}^m p_i + t \sum_{i=1}^m q_i = 1$. As all the terms in the sum are nonnegative, $\text{KL}(p, q) \geq 0$.

Now, applying the (variant of) Cauchy-Schwarz Inequality, we get

$$\begin{aligned}
\sum_{i=1}^m \frac{(p_i - q_i)^2}{\xi_i} &= \left(\sum_{i=1}^m \xi_i \right) \cdot \left(\sum_{i=1}^m \frac{(p_i - q_i)^2}{\xi_i} \right) \\
&\geq \left(\sum_{i=1}^m |p_i - q_i| \right)^2.
\end{aligned}$$

Plugging this inequality into Equation (1), we conclude that $\text{KL}(p, q) \geq \frac{1}{2} (\sum_{i=1}^m |p_i - q_i|)^2$.