MATH 541a Homework 2

Qilin Ye

February 1, 2022

Problem 1

† You want to complete a set of 100 baseball cards. Cards are sold in packs of ten. Assume that each individual card in the pack has a uniformly random chance of being any element in the full set of 100 baseball cards. (In particular, there is a chance of getting identical cards in the same pack.) How many packs of cards should you buy in order to get a complete set of cards? That is, what is the expected number of cards you should buy in order to get a complete set of cards (rounded up to a multiple of ten)?

Solution. Let N_i and N be defined as suggested by the hint. Then $N_1 = 1$. Also define $N_0 = 0$. Now we compute $N_i - N_{i-1}$. If our collection now contains $i-1$ distinct cards, the probability of buying and getting a new card is $(100-(i-1))/100$. Therefore the expected n Let N_i and N be defined as suggested by the hint. Then $N_1 = 1$. Also define $N_0 = 0$. Now we compute $N_i - N_{i-1}$. If our collection now contains $i-1$ distinct cards, the probability of buying and getting a new card is $(100-(i-1))/100$. Therefore the expected number of cards to buy in order to get a new card is $100/(100-i+1)$. Thus,

$$\begin{aligned}\mathbb{E}N = \mathbb{E}N_{100} &= \sum_{i=1}^{100} \mathbb{E}(N_i - N_{i-1}) \\ &= 100 \sum_{k=0}^{99} \frac{1}{100-k} = 100 \sum_{k=1}^{100} \frac{1}{k} \approx 518.7.\end{aligned}$$

That is, we need to buy $51.87 \approx 52$ packs in order to get a complete set of cards. umber of cards to buy in order to get a new card is $100/(100-i+1)$. Thus,

$$\begin{aligned}\mathbb{E}N = \mathbb{E}N_{100} &= \sum_{i=1}^{100} \mathbb{E}(N_i - N_{i-1}) \\ &= 100 \sum_{k=0}^{99} \frac{1}{100-k} = 100 \sum_{k=1}^{100} \frac{1}{k} \approx 518.7.\end{aligned}$$

That is, we need to buy $51.87 \approx 52$ packs in order to get a complete set of cards.

Problem 2

† You are trapped in a maze. Your starting point is a room with three doors. The first door will lead you to a corridor which lets you exist the maze after 3 hours of walking. The second dour leads you through a corridor which puts you back to the starting point of the maze after seven hours of walking. The third

door leads you through a corridor which puts you back to the starting point of the maze after nine hours of walking. Each time at the starting point you choose one of the doors with equal probability. Let X be the number of hours it takes for you to exist the maze and let Y be the number of door that you initially choose.

- Compute $\mathbb{E}(X \mid Y = i)$, $i \in \{1, 2, 3\}$, in terms of $\mathbb{E}X$.
- Compute $\mathbb{E}X$.

Solution.

$$\mathbb{E}(X \mid Y = 1) = 3$$
$$\mathbb{E}(X \mid Y = 2) = 7 + \mathbb{E}X$$
$$\mathbb{E}(X \mid Y = 3) = 9 + \mathbb{E}X.$$

Then, since $\mathbb{E}X = \sum_{i=1}^{3} \mathbb{P}(Y = i)\mathbb{E}(X \mid Y = i)$, we obtain the function

$$\frac{3 + 7 + \mathbb{E}X + 9 + \mathbb{E}X}{3} = \mathbb{E}X \implies \mathbb{E}X = 19.$$

Problem 3

Let $X_1, ..., X_n$ be continuous random variables with joint PDF $f : \mathbb{R}^n \to [0, \infty)$. Assume that

$$f_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \qquad \text{for all } x_1, ..., x_n \in \mathbb{R}.$$

Show that $X_1, ..., X_n$ are independent.

Proof. Let $(x_1, ..., x_n) \in \mathbb{R}^n$. Then

$$\begin{aligned}
\mathbb{P}(X_i \leqslant x_i \text{ for all } i) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,...,X_n}(s_1, ..., s_n)\, \mathrm{d}s_n \cdots \mathrm{d}s_1 \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f_{X_i}(s_i)\, \mathrm{d}s_n \cdots \mathrm{d}s_1 && \text{(assumption)} \\
&= \left(\int_{-\infty}^{x_1} f_{X_1}(s_1)\, \mathrm{d}s_1 \right) \cdots \left(\int_{-\infty}^{x_n} f_{X_n}(s_n)\, \mathrm{d}s_n \right) && \text{(Fubini)} \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \leqslant x_i).
\end{aligned}$$

Therefore $X_1, ..., X_n$ are independent. □

Problem 4

Let $\varphi : \mathbb{R} \to \mathbb{R}$. Show that φ is convex if and only if: for any $y \in \mathbb{R}$, there exists a constant a and a function $L : \mathbb{R} \to \mathbb{R}$ defined by $L(x) = a(x - y) + \varphi(y)$ such that $L(y) = \varphi(y)$ and $L(x) \leqslant \varphi(x)$ for all $x \in \mathbb{R}$.

Proof. If φ is convex, then for any $x \in \mathbb{R}$ and $h > 0$,

$$\varphi(x) = \varphi\left(\frac{(x-h)+(x+h)}{2}\right) \leqslant \frac{\varphi(x-h)+\varphi(x+h)}{2},$$

so

$$\frac{\varphi(x)-\varphi(x-h)}{h} \leqslant \frac{\varphi(x+h)-\varphi(x)}{h}. \tag{1}$$

Furthermore, if $0 < \tilde{h} < h$, writing $x - \tilde{h}$ as $(\tilde{h}/h)(x-h) + (1-\tilde{h}/h)x$, we have

$$\varphi(x-\tilde{h}) \leqslant (\tilde{h}/h)\varphi(x-h) + (1-\tilde{h}/h)\varphi(x).$$

Multiplying both sides by h, rearranging, and then dividing by $h\tilde{h}$ gives

$$\frac{\varphi(x)-\varphi(x-h)}{h} \leqslant \frac{\varphi(x)-\varphi(x-\tilde{h})}{\tilde{h}} \qquad \text{for } 0 < \tilde{h} < h.$$

A similar argument shows

$$\frac{\varphi(x+h)-\varphi(x)}{h} \geqslant \frac{\varphi(x+\tilde{h})-\varphi(x)}{\tilde{h}} \qquad \text{for } 0 < \tilde{h} < h.$$

A bounded monotone sequence has a limit, so it is well-defined to take $\lim_{h \searrow 0}$ of (1) and obtain

$$\lim_{h\searrow 0}\frac{\varphi(x)-\varphi(x-h)}{h} \leqslant \lim_{h\searrow 0}\frac{\varphi(x+h)-\varphi(x)}{h}. \tag{2}$$

Pick any value a in-between these two limits in (2) and we have obtained our linear "tangent" function bounding φ from below.

Conversely, let $x, y \in \mathbb{R}$ and let $\lambda \in (0,1)$. Define $p := \lambda x + (1-\lambda)y$. By assumption there exists constant a such that $L(x) := a(x-p) + \varphi(p)$ is "tangent" to φ at p and bounds φ from below. Thus

$$a(x-p) + \varphi(p) \leqslant \varphi(x) \qquad \text{and } a(y-p) + \varphi(p) \leqslant \varphi(y). \tag{3}$$

Now we apply the convex combination of $\varphi(x)$ and $\varphi(y)$:

$$\begin{aligned}\lambda\varphi(x) + (1-\lambda)\varphi(y) &\geqslant \lambda(a(x-p)+\varphi(p)) + (1-\lambda)(a(y-p)+\varphi(p)) \\ &= \lambda a(1-\lambda)(x-y) + \lambda\varphi(p) - (1-\lambda)a\lambda(x-y) + (1-\lambda)\varphi(p) \\ &= \varphi(p) = \varphi(\lambda x + (1-\lambda y)).\end{aligned}$$

□

Problem 5

Prove Jensen's inequality: if φ is convex and $\mathbb{E}|X| < \infty, \mathbb{E}|\varphi(X)| < \infty$, then $\varphi(\mathbb{E}X) \leqslant \mathbb{E}\varphi(X)$. Deduce the triangle inequality $|\mathbb{E}X| \leqslant \mathbb{E}|X|$.

Proof. By the previous problem there exists a constant c and a linear function $L(x) = c(x - \mathbb{E}X) + \varphi(\mathbb{E}X)$ that bounds φ from below. Then

$$\mathbb{E}\varphi(X) \geqslant \mathbb{E}L(X) = \mathbb{E}(c(X-\mathbb{E}X) + \varphi(\mathbb{E}X)) = \varphi(\mathbb{E}X).$$

Applying Jensen's inequality to $\varphi(x) := |x|$ gives $|\mathbb{E}X| \leqslant \mathbb{E}|X|$. □

Problem 6

Prove Markov's inequality $\mathbb{P}(|X| \geqslant t) \leqslant \mathbb{E}|X|/t$ for all $t \geqslant 0$.

Proof. This follows from decomposing $|X|$ into $|X|1_{\{|X|\geqslant t\}}$ and $|X|1_{\{|X|<t\}}$:

$$\mathbb{E}|X| = \mathbb{P}(|X| \geqslant t)\mathbb{E}|X|1_{\{|X|\geqslant t\}} + \mathbb{P}(|X| < t)1_{\{|X|\leqslant t\}} \geqslant t\mathbb{P}(|X| \geqslant t) + 0. \qquad \square$$

Problem 7

Let X be a random variable and let $r > 0$. Define $M_X(t) := \mathbb{E}e^{tX}$ for $t \in \mathbb{R}$. Prove the **Chernoff bound**: for any $t > 0$,

$$\mathbb{P}(X > r) \leqslant e^{-tr} M_X(t).$$

Proof. Since the exponential function is monotone, $X > r$ if and only if $e^{tX} > e^{tr}$. Then Markov's inequality applied to e^{tX} implies

$$\mathbb{P}(X > r) = \mathbb{P}(e^{tX} > e^{tr}) \leqslant e^{-tr} M_X(t). \qquad \square$$

Problem 8

† Among 625 members of a bank chosen uniformly at random among all bank members, it was found that 25 had a savings account. Give an interval of form $[a, b]$ where $a, b \in \mathbb{Z}$ such that with about 95% certainty, if we sample 625 bank members independently and uniformly at random (from a very large bank membership), then the number of these people with savings accounts lies in the interval $[a, b]$.

Solution. Let X_i denote the status of whether the i^{th} person has a savings. Let $X_i = 1$ if yes and $= 0$ otherwise. Then given the assumptions each X_i should follow a Bernoulli distribution with parameter $25/625 = 1/25$ and variance $24/625$. Using the CLT we see that $\sum_{i=1}^{625} X_i$ roughly follows a Gaussian with mean $625 \cdot 1/25 = 25$ and standard deviating $\sqrt{625} \cdot \sqrt{24/625} = \sqrt{24}$. To have a 95% confidence interval, we want $Z \in [-2, 2]$, which corresponds to $[25 - 2\sqrt{24}, 25 + 2\sqrt{24}] \approx [15, 35]$.

Problem 9

† Suppose we run a casino and we want to test whether a particular roulette wheel is biased. Let p be the probability that red results from a spin. Let the null hypothesis be $p = 18/38$ and let $p \neq 18/38$ be the alternate hypothesis. For $i \geqslant 1$, let $X_i = 1$ if the i^{th} spin is red and $= 0$ otherwise. Let $\mu := \mathbb{E}X_1$ and $\sigma := \sqrt{\text{var}(X_1)}$.
To test the null hypothesis we spin the wheel n times. In our test, we reject the null hypothesis if $|X + \ldots + X_n - n\mu| > 2\sigma\sqrt{n}$. We set the type I error (false positive) to be 5%.
Suppose we spin the wheel $n = 3800$ times and get red 1868 times. Is the wheel biased?

Solution. Assuming the null hypothesis, each X_i is a Bernoulli trial with mean $18/38$ and variance $(18/38)(20/38)$. CLT states that $\sum_{i=1}^{3800} X_i$ roughly follows a Gaussian with mean $3800 \cdot 18/38 = 1800$ and standard

deviation $\sqrt{3800}\sqrt{(18/38)(20/38)}$. Our observed value is 1868 and it corersponds to

$$Z = \left| \frac{1868 - 1800}{\sqrt{3800}\sqrt{(18/38)(20/38)}} \right| \approx 2.21 > 2$$

so we reject the null hypothesis with $> 95\%$ certainty.

Problem 10

† A community has m facilities. Each family has at least one child. The largest family has $k > 0$ children. For each $i \in \{1, \cdots, k\}$, there are n_i families with i children so $n_1 + \cdots + n_k = m$. Choose a child randomly in the following ways.

(1) First choose one of the families uniformly at random among all the families. Then, in the chosen family, choose one of the children uniformly at random.

(2) Among all $n_1 + 2n_2 + \cdots + kn_k$ children, choose one uniformly at random.

What is the probability that the chosen child is the first-born in their family if you use method (1)? What about (2)?

Solution. For method 1, there is a probability of n_i/m to choose a family of i children. Then there is a probability of $1/i$ that the children picked is the first-born. Thus, the total probability is $m^{-1} \sum_{i=1}^{k} n_i/i$.
For the second method, we simply need to compute the number of first-born children and decide it by the total number of children. Clearly m family correspond to m first-born children, and there are $\sum_{i=1}^{k} i \cdot n_i$ children. Thus the total probability is $m / \sum_{i=1}^{k} (i \cdot n_i)$.

Problem 11

Let $0 < p \leqslant \infty$. Show that if $Y_1, Y_2, \ldots : \Omega \to \mathbb{R}$ converge to $Y : \Omega \to \mathbb{R}$ in L^p then $Y_n \to Y$ in probability. Then show that the converse is false.

Proof. Let $\epsilon > 0$ be given. For $p < \infty$, we have

$$\begin{aligned} \|Y_n - Y\|_p^p &= \int_\Omega |Y_n - Y|^p \, \mathrm{d}\mathbb{P} \\ &= \int_{\{|Y_n - Y| \leqslant \epsilon\}} |Y_n - Y|^p \, \mathrm{d}\mathbb{P} + \int_{\{|Y_n - Y| > \epsilon\}} |Y_n - Y|^p \, \mathrm{d}\mathbb{P} \\ &\geqslant \int_{\{|Y_n - Y| > \epsilon\}} |Y_n - Y|^p \, \mathrm{d}\mathbb{P}. \end{aligned}$$

This shows that

$$\mathbb{P}(\{Y_n - Y\} > \epsilon) < \frac{\|Y_n - Y\|_p^p}{\epsilon^p}.$$

Taking $n \to \infty$ finishes the proof.
(I could've used Markov's and said $\mathbb{P}(|Y_n - Y| > \epsilon) = \mathbb{P}(|Y_n - Y|^p > \epsilon^p) \leqslant \epsilon^{-p} \mathbb{E}|Y_n - Y|^p = \epsilon^{-p} \|Y_n - Y\|_p^p$, a one-liner.)
If $p = \infty$, simple note that whenever $\|Y_n - Y\|_\infty \leqslant \epsilon$ we have $\mathbb{P}(\{\omega : |Y_n(\omega) - Y(\omega)| > \epsilon\}) = 0$. Then the claim follows from the assumption that $\|Y_n - Y\|_\infty \to 0$.

For a counterexample, let $Y_n := n^{1/p} 1_{(0,1/n)}$. See below. □

Problem 12

Show that (almost sure convergence) $\not\Rightarrow$ (convergence in L^p) and also show (convergence in L^p) $\not\Rightarrow$ (almost sure convergence).

Solution. Consider $Y_n := n^{1/p} 1_{(0,1/n)}$. Clearly $Y_n \to Y$, the constant random variable taking value zero, for all $x \in \mathbb{R}$, whereas

$$\|Y_n - Y\|_p = \left(\int_{\mathbb{R}} |n^{1/p} 1_{(0,1/n)}|^p \, d\mathbb{P} \right)^{1/p} = 1^{1/p} = 1 \qquad \text{for all } n.$$

Conversely, consider the following sequence of random variables:

$$Y_1 := 1_{[0,1]}$$
$$Y_2 := 1_{[0,1/2]} \qquad Y_3 := 1_{[1/2,1]}$$
$$Y_4 := 1_{[0,1/3]} \qquad Y_5 := 1_{[1/3,2/3]} \qquad Y_6 := 1_{[2/3,1]}$$
$$\dots$$

Since any Y_j on the k^{th} line is the indicator variable on an interval of length $1/k$, we have $\|Y_j\|_p = (1/k)^{1/p} = k^{-1/p}$. As $k \to \infty$ we have $\|Y_n\|_p \to 0$, so $Y_n \to$ the zero variable in L^p. However, Y_n does not converge almost surely to Y — in fact it converges nowhere on $[0,1]$. Given any $x \in [0,1]$, on each line, at least one of the Y_j's will have $Y_j(x) = 1$, so the sequence $\{Y_n(x)\}_{n \geq 1}$ cannot possibly converge to 0.

Problem 13

† Estimate the probability that a million coin flips of fair coins will result in more than $501,000$ heads using the CLT.

Solution. Each coin flip can be viewed as a Bernoulli random variable with $p = 0.5$. Thus the mean is 0.5, the variance 0.25, and the standard deviation 0.5. Adding a million i.i.d. copies of them, we roughly have a Gaussian with mean 0.5 million and standard deviation $\sqrt{10^6} \cdot 0.5 = 500$. Thus having $> 501,000$ heads corresponds to > 2 standard deviations, i.e., $Z > 2$, which has the probability

$$\mathbb{P}(Z > 2) = \frac{1}{\sqrt{2\pi}} \int_2^\infty \exp(-s^2/2) \, ds \approx 0.0228.$$