60850 Midterm 1 Solutions[1]

1. Question 1

- Let $X\colon\Omega\to\mathbf{R}$ be a random variable such that
$$\mathbf{P}(X>0)>0.$$
Show that there exists a positive integer n such that $\mathbf{P}(X>1/n)>0$.
- Let $X_1,X_2,\ldots\Omega\to\mathbf{R}$ be random variables such that $\mathbf{E}X_i=0$ and $\mathbf{E}X_i^2=1$ for all $i\geq 1$. Show that
$$\mathbf{P}(X_n>n\text{ for infinitely many }n\geq 1)=0.$$

Solution. Note that $\{X>1\}\supseteq\{X>1/2\}\supseteq\{X>1/3\}\supseteq\cdots$. So, from Continuity of the Probability Law,

$$\lim_{n\to\infty}\mathbf{P}(X>1/n)=\mathbf{P}(\cap_{n=1}^{\infty}\{X>1/n\})=\mathbf{P}(X>0)>0.$$

In the last line, we used our assumption. So, by definition of the limit, there exists a positive integer n such that $\mathbf{P}(X>1/n)>0$.

We now do the second part. Note that $\mathrm{var}(X_n)=\mathbf{E}X_n^2-(\mathbf{E}X_n)^2=1$ for all $n\geq 1$. From Chebyshev's inequality,

$$\mathbf{P}(X_n>n)\leq\frac{\mathrm{var}(X_n)}{n^2}=\frac{1}{n^2},\qquad\forall\, n\geq 1.$$

Therefore,

$$\sum_{n\geq 1}\mathbf{P}(X_n>n)\leq\sum_{n\geq 1}\frac{1}{n^2}<\infty.$$

So, from the Borel-Cantelli Lemma, the second claim follows.

2. Question 2

Let $X_1,X_2,\ldots$ be independent identically distributed random variables with $\mathbf{P}(X_1=1)=\mathbf{P}(X_1=-1)=1/2$. Let $a_1,a_2,\ldots\in\mathbf{R}$. Show that, for any $n\geq 1$,

$$\mathbf{P}\Big(\sum_{i=1}^{n}a_iX_i\geq t\Big)\leq e^{-\frac{t^2}{2\sum_{i=1}^{n}a_i^2}},\qquad\forall\, t\geq 0.$$

(You can use the following inequality without proof: $\cosh(x)\leq e^{x^2/2}$, $\forall\; x\in\mathbf{R}$.)

Solution. By dividing $a_1,\ldots,a_n$ by a constant, we may assume $\sum_{i=1}^{n}a_i^2=1$. Let $\alpha>0$. Using the (exponential) moment method as in Markov's inequality, and $\alpha t\geq 0$,

$$\mathbf{P}(\sum_{i=1}^{n}a_iX_i\geq t)=\mathbf{P}(e^{\alpha\sum_{i=1}^{n}a_iX_i}\geq e^{\alpha t})\leq e^{-\alpha t}\mathbf{E}e^{\alpha\sum_{i=1}^{n}a_iX_i}=e^{-\alpha t}\prod_{i=1}^{n}\mathbf{E}e^{\alpha a_iX_i}.$$

The last equality used independence of $X_1,X_2,\ldots$. Using an explicit computation and the above cosh inequality,

$$\mathbf{E}e^{\alpha a_iX_i}=(1/2)(e^{\alpha a_i}+e^{-\alpha a_i})=\cosh(\alpha a_i)\leq e^{\alpha^2a_i^2/2},\qquad\forall\, i\geq 1.$$

[1] March 10, 2018, © 2018 Steven Heilman, All Rights Reserved.

In summary, for any $t \geq 0$

$$\mathbf{P}(\sum_{i=1}^{n} a_i X_i \geq t) \leq e^{-\alpha t} e^{\alpha^2 \sum_{i=1}^n a_i^2/2} = e^{-\alpha t + \alpha^2/2}.$$

Since $\alpha > 0$ is arbitrary, we choose α to minimize the right side. This minimum occurs when $\alpha = t$, so that $-\alpha t + \alpha^2/2 = -t^2/2$, giving the desired bound.

3. Question 3

Let n be a positive integer. Suppose $X_1, X_2, \ldots$ are independent random variables that are uniformly distributed in the set $\{1, 2, \ldots, n\}$. We can think of $\{1, 2, \ldots, n\}$ as a set of baseball cards, and for any $i \geq 1$, X_i is a uniformly random baseball card that you have found. Your goal is to collect all of the n baseball cards.

For any $0 \leq j \leq n$, let T_j be the first time that you have collected exactly j baseball cards. That is, T_j is the smallest integer k such that $\{X_1, \ldots, X_k\}$ consists of exactly k distinct elements. For example, $T_0 = 0$, $T_1 = 1$, T_2 is 2 when $X_2 \neq X_1$, T_2 is 3 when $X_2 = X_1$ and $X_3 \neq X_1$, and so on. (You may assume that $\mathbf{P}(T_j < \infty) = 1$ for all $0 \leq j \leq n$.)

For any $1 \leq j \leq n$, let $Y_j := T_j - T_{j-1}$ be the time it takes to go from a collection of $j-1$ distinct baseball cards to a collection of j distinct baseball cards.

Show that Y_2 and Y_3 are independent, geometric random variables with parameters $\frac{n-1}{n}$ and $\frac{n-2}{n}$, respectively.

(Recall that a geometric random variable Y with parameter $0 < p < 1$ is a positive-integer valued random variable such that $\mathbf{P}(Y = k) = (1-p)^{k-1} p$ for any $k \geq 1$.)

Solution. Let a, b be positive integers. To prove independence, we need to show that

$$\mathbf{P}(Y_2 = a, Y_3 = b) = \mathbf{P}(Y_2 = a)\mathbf{P}(Y_3 = b).$$

In fact, to complete the problem, it suffices to show that

$$\mathbf{P}(Y_2 = a, Y_3 = b) = \frac{(n-1)(n-2)}{n^2} \frac{1}{n^{a-1}} \Big(\frac{2}{n}\Big)^{b-1},$$

since summing over $b \geq 1$ shows that Y_2 is geometric, and similarly for summing over $a \geq 1$. Note that

$$\{Y_2 = a, Y_3 = b\} = \{X_j = X_1 \,\forall\, 1 \leq j \leq a-1,\ X_a \neq X_1,$$
$$X_k \in \{X_1, X_a\} \,\forall a < k < b + a,\ X_{a+b} \notin \{X_1, X_a\}\}.$$

Since the random variables $X_1, X_2, \ldots$ are independent and uniformly distributed in $\{1, \ldots, n\}$, we then have

$$\mathbf{P}(Y_2 = a, Y_3 = b) = \Big(\prod_{j=1}^{a-1} \mathbf{P}(X_j = 1)\Big)\mathbf{P}(X_a \neq 1)\Big(\prod_{j=a+1}^{a+b-1} \mathbf{P}(X_j \in \{1, 2\})\Big)\mathbf{P}(X_b \notin \{1, 2\})$$

$$= \Big(\prod_{j=1}^{a-1} \frac{1}{n}\Big)\frac{n-1}{n}\Big(\prod_{j=a+1}^{a+b-1} \frac{2}{n}\Big)\frac{n-2}{n} = \frac{1}{n^{a-1}}\Big(\frac{2}{n}\Big)^{b-1}\frac{(n-1)(n-2)}{n^2}.$$

4. Question 4

We continue the definitions and notation from the previous problem. In this problem, you may assume that $Y_1, \ldots, Y_n$ are independent random variables and for any $1 \leq j \leq n$, Y_j is a geometric random variable with parameter $\frac{n-j+1}{n}$.

You may use the following fact: a geometric random variable Y with parameter $0 < p < 1$ has mean $\frac{1}{p}$ and variance $\frac{1-p}{p}$.

Note that $T_n = Y_1 + \cdots + Y_n$.

- Show that $\mathbf{E}T_n = n \log n + O(n)$ and $\mathrm{var}(T_n) = O(n^2)$.
- Conclude that
$$\frac{T_n}{n \log n}$$
converges in probability to 1 as $n \to \infty$.

(Hint: Can you bound $\mathbf{P}(|T_n - n \log n + O(n)| > tn)$?)

So, if you want to complete a set of 100 distinct baseball cards, you would need to randomly sample about $100 \log 100 \approx 460$ baseball cards.

(As usual, $O(a)$ denotes any quantity whose absolute value is bounded by a constant multiple ca of a.)

Solution. By integral Comparison (as in Calculus 2), we have

$$\begin{aligned}\mathbf{E}T_n &= \sum_{i=1}^{n} \mathbf{E}Y_i = \sum_{i=1}^{n} \frac{n}{n-i+1} = \sum_{i=1}^{n} \frac{n}{i} = n \sum_{i=1}^{n} \frac{1}{i} \\ &= n\left(\int_1^n \frac{1}{x} dx + O(1)\right) = n \log(n) + O(n).\end{aligned}$$

Also, by independence, we have

$$\mathrm{var}T_n = \sum_{i=1}^{n} \mathrm{var}Y_i = \sum_{i=1}^{n} \frac{i-1}{n} \frac{n}{n-i+1} = \sum_{i=1}^{n} \frac{i-1}{n-i+1} \leq \sum_{i=1}^{n} n = O(n^2).$$

So, from Chebyshev's Inequality, for any $t > 0$

$$\mathbf{P}\left(|T_n - n \log n + O(n)| > tn\right) \leq \frac{\mathrm{var}(T_n)}{t^2 n^2} \leq O(t^{-2}).$$

That is,

$$\mathbf{P}\left(\left|\frac{T_n}{n \log n} - 1 + O(1/\log n)\right| > \frac{t}{\log n}\right) \leq O(t^{-2}).$$

Or, replacing t with $t \log n$,

$$\mathbf{P}\left(\left|\frac{T_n}{n \log n} - 1 + O(1/\log n)\right| > t\right) \leq O((t \log n)^{-2}).$$

In particular, for any $t > 0$, $\lim_{n\to\infty} \mathbf{P}(|\frac{T_n}{n \log n} - 1| > t) = 0$.

5. Question 5

Find a sequence of random variables $X_1, X_2, \ldots : \Omega \to \mathbf{R}$ such that

- $X_1, X_2, \ldots$ converges in probability to 0.
- $X_1, X_2, \ldots$ does **not** converge almost surely to 0.
- $X_1, X_2, \ldots$ does **not** converge in L_2 to 0.

Prove that your example of $X_1, X_2, \ldots$ satisfies the above three properties.

(As usual, it might be easiest to use $\Omega = [0,1]$ with $\mathbf{P}$ Lebesgue measure on Ω.)

Solution. We use a traveling bump with growing height. Let $n \geq 1$ and let $j = j(n) \geq 0$ be the unique integer such that $2^j \leq n < 2^{j+1}$. Define

$$X_n := 2^j 1_{[\frac{n-2^j}{2^j}, \frac{n-2^j+1}{2^j}]}.$$

Note that $\{X_n \neq 0\} = [\frac{n-2^j}{2^j}, \frac{n-2^j+1}{2^j}]$, and this interval has width 2^{-j}. Since $j(n) \to \infty$ as $n \to \infty$, we therefore have

$$\lim_{n\to\infty} \mathbf{P}(X_n \neq 0) = \lim_{j\to\infty} 2^{-j} = 0.$$

Therefore, $X_1, X_2, \ldots$ converges in probability to 0.

Now, $X_1, X_2, \ldots$ does not converge in L_2 to 0 since

$$\mathbf{E}X_n^2 = 2^{2j}2^{-j} = 2^j \to \infty \text{ as } n \to \infty.$$

Finally, $X_1, X_2, \ldots$ does not almost surely to 0 since $\cup_{2^k \leq n < 2^{k+1}}[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}] = [0,1]$. So, if $t \in [0,1]$ is fixed, the sequence of numbers $X_1(t), X_2(t), \ldots$ has an infinite number of $0's$ and an infinite number of integers larger than 1. That is, $\lim_{n\to\infty} X_n(t)$ does not exist, for all $t \in [0,1]$, so that

$$\mathbf{P}(\lim_{n\to\infty} X_n = 0) = 0 \neq 1.$$