Math 302, PSET 5

(1) An evil mathematician has trapped you in a dungeon behind 5 doors. Every door is locked with a keypad, in which you must enter a number between 1 and 6000. You enter one random number (with replacement) into the keypad every second. The lock will open if you enter one of 15 special numbers the evil mathematician has selected for each lock, and once you open one lock, you move to the next door and keypad.

(a) Use an exponential random variable to approximate the probability that it'll take you longer than 25 minutes to open the first door.

(b) Use a Poisson random variable to approximate the probability that after one hour, you have escaped the dungeon.

Solution:

(a) On average, we will have $\lambda = 60\frac{15}{6000} = 0.15$ successes per minute. We therefore model the time until the first success (in minutes) by an $Exp(0.15)$ random variable X. (Geometric is approximately exponential.) The probability in question is then

$$\mathbb{P}(X > 25) = 1 - \mathbb{P}(X \leq 25) = 1 - (1 - e^{-0.15*25}) = e^{-3.75} \approx 0.02$$

(b) We model number of doors opened after one hour by a $Poisson(60\lambda)$ random variable Y. (Binomial is approximately Poisson.) The probability in question is then

$$\mathbb{P}(Y \geq 5) = 1 - \mathbb{P}(Y \leq 4) = 1 - e^{-9}(1 + 9 + \frac{9^2}{2!} + \frac{9^3}{3!} + \frac{9^4}{4!}) \approx 0.95$$

(2) Let X be a Poisson random variable with unknown parameter λ.

(a) Which $n = n(\lambda) \geq 0$ is the most likely value of X, i.e. maximizes $\mathbb{P}(X = n)$?

(b) Suppose the experiment described by X has returned the value $n \geq 0$. Which parameter $\lambda = \lambda(n)$ maximizes $\mathbb{P}(X = n)$?

Solution:

(a) Let $f(n) = P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!}$. Then we have

$$\frac{f(n+1)}{f(n)} = \frac{\lambda}{n+1}$$

If $\lambda < 1$ then $f(n+1)/f(n) < 1$, hence $f(n)$ is maximal when $n = 0$. If $\lambda \in \mathbb{N}_0$ then $f(n+1)/f(n) = 1$ when $n = \lambda - 1$, hence $f(n)$ is maximal when $n = \lambda - 1$. If $\lambda > 1$ then $f(n+1)/f(n)$ is closest to 1 when $n = floor(\lambda)$, hence $f(n)$ is maximal when $n = \lfloor \lambda \rfloor$.

(b) $f(\lambda) = P(X = n) = e^{-\lambda}\frac{\lambda^n}{n!}$.

$$f'(\lambda) = 0 \implies \frac{1}{n!}e^{-\lambda}\lambda^{n-1}(n - \lambda) = 0$$

So $f(\lambda)$ is maximal when $\lambda = n$. (Clearly $f'(\lambda) > 0$ if $\lambda < n$ and $f'(\Lambda) < 0$ if $\lambda > n$, so f is increasing on the interval $[0, n]$ and decreasing on $[n, \infty)$. Thus there is really a global maximum at $\lambda = n$)

(3) For each of the sequences of random variables given, decide if it converges in distribution/probability/almost surely, and if so, identify the limiting distribution or random variable.

(a) $W_n \sim$ Uniform$(0, n)$ for $n \geq 1$.

(b) $X_n \sim$ independent Normal$(0, n^{-2})$ for $n \geq 1$.

(c) Let Y_1 be a Uniform$(0, 1)$ random variable, and for $n \geq 2$, define random variables Y_n on the same probability space by $Y_n = 1 - Y_{n-1}$.

(d) $Z_n \sim$ independent Bernoulli(n^{-1}) for $n \geq 1$.

Solution: Recall that a.s. convergence implies convergence in probability implies convergence almost surely, so conversely, if a sequence doesn't converge in distribution, then it can't converge in probability or almost surely; and if a sequence doesn't converge in probability, then it doesn't converge almost surely.

(a) W_n doesn't converge in distribution, since the PDF of W_n is 1 on $(0, n)$ and 0 elsewhere, which converges as $n \to \infty$ to the function that is 1 everywhere, which is not the PDF of any random variable (since it doesn't integrate to 1).

(b) X_n converges in distribution to the 0 random variable, since the CDF of X_n is $F_{X_n}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-n^2x^2/2}$, which converges to the function that is 0 for $x < 0$ and 1 for $x > 0$, i.e. the CDF of the random variable that has all its mass at 0. The X_n also converge to 0 in probability, since by Chebychev's inequality,

$$\mathbb{P}(|X_n| > \epsilon) \leq \frac{\mathrm{Var}(X_n)}{\epsilon} = \frac{1}{n^2\epsilon} \to 0$$

as $n \to \infty$. Finally, the X_n do converge almost surely. To see why, let $A_n = \{|X_n| > \frac{1}{\log n}\}$, and apply Chebychev:

$$\mathbb{P}(A_n) \leq \frac{(\log n)^2}{n^2}.$$

Let $N = \sum_{n=1}^{\infty} 1_{A_n}$, i.e. N counts how many of the A_n occur. By the above, and by linearity of expectation,

$$\mathbb{E}N \leq \sum_{n \geq 1} \frac{(\log n)^2}{n^2} < \infty.$$

Thus $\mathbb{P}(N = \infty) = 0$, i.e. N is finite almost surely. (If N took value ∞ with positive probability, its expectation would be infinite.) It follows that A_n occurs for only finitely many n, i.e. $|X_n| < (\log n)^{-1}$ for n sufficiently large with probability 1. Thus $X_n \to 0$ almost surely. (Note: this is a well-known method, called the 'first Borel-Cantelli lemma.')

(c) Note that Y_n has Uniform$(0, 1)$ distribution for every n, since if $U \sim$ Uniform$(0, 1)$, then also $1 - U \sim$ Uniform$(0, 1)$. So Y_n converges in distribution to a Uniform$(0, 1)$. However, Y_n does not converge in probability (or almost surely). Note that the sequence has the form $(Y_1, 1-Y_1, Y_1, 1-Y_1, \ldots)$, and suppose the limit in probability is some random variable Y. Then

$$\mathbb{P}(|Y_n - Y| > \epsilon) = \begin{cases} \mathbb{P}(|Y_1 - Y| > \epsilon), & n \text{ odd} \\ \mathbb{P}(|1 - Y_1 - Y| > \epsilon), & n \text{ even} \end{cases}$$

Since the probability doesn't depend on n in either case, we must have both $Y = Y_1$ and $Y = 1 - Y_1$ with probability 1, a contradiction. Thus no such Y exists.

(d) This is an example where Z_n converges in probability but not almost surely. Applying Markov's inequality immediately gives convergence in probability to 0:

$$\mathbb{P}(Z_n > \epsilon) \leq \frac{n^{-1}}{\epsilon} = \frac{1}{n\epsilon} \to 0$$

as $n \to \infty$. To see that this sequence does not converge almost surely, we can try to use an idea similar to in (b): let $N = \sum_{n \geq 1} Z_n$ be the number of Z_n's that are 1, and observe that

$$\mathbb{E}N = \sum_{n \geq 1} \frac{1}{n} = \infty.$$

As we have seen, this is not enough to conclude that $N = \infty$ with positive probability! (Finite random variables can have infinite expectation.) What works instead is to rely on the independence assumption. Note that

$$\{N < \infty\} \subset \bigcup_{n \geq 1} \{Z_n = Z_{n+1} = \ldots = 0\},$$

since if $N < \infty$ then all the Z_n must be zero past some point. By countable additivity and independence,

$$\begin{aligned}
\mathbb{P}(N < \infty) &\leq \sum_{n \geq 1} \mathbb{P}(Z_n = Z_{n+1} = \ldots = 0) \\
&= \sum_{n \geq 1} \prod_{m=n}^{\infty} \mathbb{P}(Z_m = 0) \\
&= \sum_{n \geq 1} \prod_{m=n}^{\infty} \left(1 - \frac{1}{m}\right)
\end{aligned}$$

The inner product is zero for each n, because

$$\prod_{m=n}^{k} = \frac{n-1}{n} \cdot \frac{m}{m+1} \cdots \frac{k-1}{k} = \frac{n-1}{k},$$

so taking the limit as $k \to \infty$ yields

$$\prod_{m=n}^{\infty} = \lim_{k \to \infty} \frac{n-1}{k} = 0.$$

This holds for every n, so plugging into the previous calculation, $\mathbb{P}(N < \infty) = 0$, i.e. $N = \infty$ with probability 1. It follows that the sequence of Z_n's has infinitely many 1s almost surely, so $\mathbb{P}(\lim_{n \to \infty} Z_n \text{ is undefined}) = 1$, i.e. Z_n does not converge to 0 almost surely. (Note: this is a well-known method, called the 'second Borel-Cantelli lemma.')

(4) Let $X_1, X_2, \ldots$ be iid random variables with $X_1 \geq a$ almost surely for some $a > 0$ and $\mathbb{E}X_1 < \infty$.
   (a) Use a method similar to the one used in class to show that

$$P_n = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$$

   converges almost surely to a constant random variable.
   (b) Assuming $X_n \sim \text{Geo}(1/2)$, find an expression for the limiting constant of the form

$$\prod_{n=1}^{\infty} a_n,$$

   and compute its approximate value.

Solution: (a) Take logs: $\log P_n = \frac{1}{n}\sum_{i=1}^n \log X_i$. Since $X_1 \geq a > 0$, and the X_i are iid, $\log X_i$ is well defined for each i, and

$$\mathbb{E}\log X_1 = \int_a^\infty f(x)\log x\,dx \leq \int_a^\infty x f(x)\,dx < \infty,$$

since $\log(x) < x$ for $x > 0$. In order to apply the LLN, we should also check that $\mathbb{E}\log X_1$ is not $-\infty$. Indeed, $\log X_1 \geq 0$ almost surely if $a > 1$, and if $a \leq 1$, we have

$$\mathbb{E}\log X_1 \geq \log(a)\mathbb{P}(\log X_1 < 0) > -\infty,$$

using the assumption $a > 0$. Now apply the LLN to the sequence $\log X_i$ to obtain

$$\log P_n = \frac{1}{n}\sum_{i=1}^n \log X_i \to_{a.s.} \mathbb{E}\log X_1.$$

Taking exponentials, we get that $P_n \to_{a.s.} e^{\mathbb{E}\log X_1}$.

(b) For $X_1 \sim \mathrm{Geo}(1/2)$, $\mathbb{E}\log X_1 = \sum_{n\geq 1} 2^{-n}\log n \approx .5078$. Thus P_n converges almost surely to $\approx e^{\mathbb{E}\log X_1} \approx 1.662$. To get an explicit form for the limit, we just have to play with exponentials a bit:

$$e^{\sum_{n\geq 1} 2^{-n}\log n} = e^{\sum_{n\geq 1}\log n^{2^{-n}}} = \prod_{n\geq 1} e^{\log n^{2^{-n}}} = \prod_{n\geq 1} n^{2^{-n}}.$$

So $a_n = n^{2^{-n}}$.

(5) Suppose X_n are non-negative random variables with $\mathbb{E}[X_n] > 0$ for all n. Show that $S_n = X_1 + X_2 + X_3 + \cdots + X_n$ converges in probability to ∞. (In other words, show that $\lim_{n\to\infty}\mathbb{P}(S_n > s) = 1$ for every $s > 0$.)

By the WLLN, $\frac{1}{n}S_n \to_p \mathbb{E}[X_1]$, i.e. $\lim_{n\to\infty}\mathbb{P}(|S_n - n\mathbb{E}X_1| > n\epsilon) \to 0$ as $n \to \infty$ for each $\epsilon > 0$. Taking $\epsilon = \frac{1}{2}\mathbb{E}X_1$, and re-arranging the inequality inside the probability, we obtain that

$$\mathbb{P}(S_n < \frac{n}{2}\mathbb{E}X_1) \to 0 \text{ as } n \to \infty.$$

Since $\frac{n}{2}\mathbb{E}X_1 \to \infty$ as $n \to \infty$, the statement above implies the result.

Challenge: Use the SLLN to show that $S_n \to \infty$ almost surely.

(6) Let X_n be iid Poisson(1) random variables, and let $Z_n = \frac{X_1+X_2+\cdots+X_n-n}{\sqrt{n}}$.
- (a) Show that Z_n converges in distribution to a Normal$(0,1)$ random variable by using the central limit theorem.
- (b) Use your result from part (a) to show that

$$\lim_{n\to\infty} e^{-n}\sum_{k=0}^n \frac{n^k}{k!} = 1/2.$$

[Hint: identify the LHS as a Poisson probability, then use (a).]

Solution: (a) Since $\mathbb{E}X_1 = \mathrm{Var}(X_1) = 1$, applying the CLT directly gives the distributional convergence.

(b) The expression in the limit is $\mathbb{P}(\text{Poisson}(n) \le n) = \mathbb{P}(X_1 + X_2 + \cdots + X_n \le n)$, since the sum of n Poisson(1) iid random variables has Poisson(n) distribution. Thus by part (a), the limit expression is

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le 0) = \mathbb{P}(\text{N}(0,1) \le 0) = 1/2.$$

Extra: use a similar method to compute, for any $\lambda > 0$

$$\lim_{n\to\infty} \frac{1}{n^n} \sum_{k=0}^{\lambda} \binom{n}{k} (n-1)^k.$$

(7) A surveyor is measuring the height of a cliff known to be about 1000 feet. She assumes her instrument is properly calibrated and that her measurement errors are independent, with mean $\mu = 0$ and variance $\sigma^2 = 10$. She plans to take n measurements and form the average. Estimate how large n should be to be 95% sure that the average falls within 1 foot of the true value

(a) using Chebychev's inequality.

(b) using the normal approximation.

Now estimate, in the same two ways, what value σ^2 should have if she wants to only make $n = 10$ measurements.

Solution:

(a) Let $X = (1/n)(X_1 + X_2 + ... + X_n)$ where X_i is the ith measurement. $Var(X) = Var(X_i)/n$. By Chebyshev; $\mathbb{P}(|X - E[X]| \ge a) \le .\frac{Var(X)}{na^2}$. Hence

$$\frac{10}{n1^2} \ge 10 \implies \mathbb{P}(|X - E[X]| \ge 1)$$

Solving for n, we see that at least 200 measurements should be taken for use to get the desired accuracy.

Similarly, if we set $n = 10$, then

$$0.05 \le \frac{10}{n} \implies \mathbb{P}(|X - E[X]| \ge 1)$$

Solving for $Var(X_i)$, we see that the variance of the measurements should be less that 0.5 feet if we want to be 95% sure of the true height of the cliff, with only 10 measurements.

(b) By the central limit theorem we have that

$$\sqrt{n}\frac{X - E[X]}{\sqrt{Var(X_i)}} \implies Z \sim N(0,1) \quad (\textit{in distribution})$$

So

$$\begin{aligned}\mathbb{P}(|X - E[X]| \ge 1) &= 1 - \mathbb{P}(-1 < X - E[X] < 1) \\ &= 1 - \mathbb{P}\Big(\frac{-\sqrt{n}}{\sqrt{Var(X_i)}} < \frac{\sqrt{n}(X - E[X])}{\sqrt{Var(X_i)}} < \frac{\sqrt{n}}{\sqrt{Var(X_i)}}\Big) \\ &\approx 2\Big(1 - \Phi\Big(\frac{\sqrt{n}}{\sqrt{10}}\Big)\Big)\end{aligned}$$

Therefore $\mathbb{P}(|X - E[X]| \ge 1) \le 0.05$ occurs (approximately) when $\phi(\frac{\sqrt{n}}{\sqrt{10}}) \ge 0.975 \implies (\sqrt{n}/\sqrt{10}) \ge 1.96 \implies n \ge 39$

Similarly, setting $n = 10$ we use exactly the same method and find $(\sqrt{10}/\sqrt{var(X_1)}) \geq 1.96 \implies Var(X_i) \geq 5.102$

(8) A fair die is rolled 24 times. Use the CLT to estimate the probability that
 (a) The sum is greater than 89.
 (b) The sum is equal to 84.

Solution:

(a) Let X_i denote the outcome of the i'th roll. $X = X_1 + X_2 + ... + X_{24}$. $E(X_i) = (1/6)(1+2+3+4+5+6) = 3.5$. $E({X_i}^2) = (1/6)(1+4+9+16+25+36) = 91/6$. By independence of the X_i and linearity; $E(X) = 84$, $Var(X) = 70$. By the CLT;

$$\begin{aligned}\mathbb{P}(X > 86) &= \mathbb{P}(X > 86.5)\\ &= \mathbb{P}(\frac{X - 24E(X_i)}{\sqrt{24var(X_i)}} \geq \frac{86.5 - 24E(X_i)}{\sqrt{24var(X_i)}})\\ &\approx \mathbb{P}(Z \geq 0.2988)\\ &1 - \Phi(0.2988) \approx 0.383\end{aligned}$$

(b)

$$\begin{aligned}\mathbb{P}(X = 84) &= \mathbb{P}(85 > X > 83)\\ &= \mathbb{P}(84.5 \geq X \geq 83.5)\\ &= \mathbb{P}(\frac{84.5 - 24E(X_i)}{\sqrt{24var(X_i)}} \geq \frac{X - 24E(X_i)}{\sqrt{24var(X_i)}} \geq \frac{83.5 - 24E(X_i)}{\sqrt{24var(X_i)}})\\ &\approx \mathbb{P}(0.0597 \geq Z \geq -0.0597)\\ &= 2\Phi(0.0597) - 1\\ &\approx 0.048\end{aligned}$$

Additional exercises: Anderson 7.41, 9.20, 9.21, 9.25, 9.27, 9.29, 9.31