Blackboard exam 5, Mathematics Mathematics 3501, Section ETY6
Starts: 6:55 pm, Thurs, May 6; ends: 7:45 pm (late submission loses points).
Instructor: Attila Máté

1. A fair die is tossed 450 times. Let X denote the number of times that a number ≤ 4 is obtained.

a) Write a formula for the probability $\mathrm{P}(X = k)$ $(0 \le k \le 450)$.

Solution. X has a binomial distribution $X \sim \mathrm{Bin}(450, 2/3)$. That is,

$$P(X = k) = \binom{450}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{450-k}.$$

b) Write the sum describing the probability $\mathrm{P}(X \le 305)$

Solution. We have

$$\mathrm{P}(X \le 305) = \sum_{k=0}^{305} \mathrm{P}(X = k) = \sum_{k=0}^{305} \binom{450}{k} \left(\frac{2}{3}\right)^k \left(\frac{1}{3}\right)^{450-k}.$$

c) Find a numerical approximation for the probability $\mathrm{P}(X \le 305)$; make sure to write the formula you are using to obtain the approximation. (Use an approximation that dates back to before computers, and do not evaluate the formula you wrote in Part *b* on computer. You can use the table given at the website for the course.)

Solution. We use normal approximation with continuity correction. A $\mathrm{Bin}(n, p)$ distribution has expectation np and standard deviation $\sqrt{np(1-p)}$. Accordingly, writing $\mathrm{D}(X)$ for the standard deviation of X, we have

$$\mathrm{E}(X) = 450 \cdot \frac{2}{3} = 300 \quad \text{and} \quad \mathrm{D}(X) = \sqrt{450 \cdot \frac{2}{3} \cdot \frac{1}{3}} = 10.$$

Thus, the distribution of X can be approximated by a that of a normal variable $Y \sim \mathcal{N}(300, 10^2)$, i.e., by a normal variable Y having expectation 300 and standard deviation 10. Hence, using continuity correction,

$$\mathrm{P}(X \le 305) \approx \mathrm{P}(Y \le 305.5) = \mathrm{P}\left(\frac{Y-300}{10} \le \frac{305.5-300}{10}\right) = \mathrm{P}\left(\frac{Y-300}{10} \le .55\right) = \Phi(.55) \approx .7088;$$

here, the Φ denotes the distribution function of the standard normal distribution $\mathcal{N}(0,1)$, and the last exact equality holds since $(Y-300)/10$ has a standard normal distribution.

Calculating the sum part *b*) directly, with a short program written in the programming language Maxima, we get a more precise answer

$$\mathrm{P}(X \le 305) \approx 0.7074816661881398.$$

2. Let n be a positive integer. Show that

$$\sum_{k=2n}^{4n} \binom{4n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{4n-k} < \frac{3}{4n}.$$

(Hint: an argument similar to the one used in the proof of the Weierstrass approximation theorem in the notes can be used.)

Solution. Let X be a random variable with distribution $\mathrm{Bin}(4n, 1/4)$. The sum in the problem can be described as the sum

$$\sum_{k=2n}^{4n} \mathrm{P}(X = k) = \mathrm{P}(X \ge 2n) < P(X = 0) + \mathrm{P}(X \ge 2n) = \mathrm{P}(|X - n| \ge n);$$

for the last equation, note that $|X-n| \geq n$ holds exactly if $-n \geq X-n$ or $X-n \geq n$, i.e., exactly if $X=0$ or $X \geq 2n$, given that X can only assume nonnegative values. We are going to use Chebyshev's inequality. According to this, for a random variable X with expectation m and variance σ^2, for every $\epsilon > 0$ we have

$$\mathrm{P}(|X-m| \geq \epsilon) \leq \sigma^2/\epsilon^2.$$

For our choice of X we have $m = n$ and $\sigma^2 = 3n/4$. With the choice $\epsilon = n$, Chebyshev's inequality then gives

$$\mathrm{P}(|X-n| \geq n) \leq \frac{3n/4}{n^2} = \frac{3}{4n}.$$

Thus, the required inequality follows; we have strict inequality, since we dropped $\mathrm{P}(X=0)$ in the calculation above.

First Note. The inequality is not particularly sharp. Calculating the ratio

$$F(n) = \sum_{k=2n}^{4n} \binom{4n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{4n-k} \Bigg/ \frac{3}{4n}$$

of the two sides of the inequality for various values of n, we have $F(1) \approx 0.348958$, $F(2) \approx 0.303507$, $F(3) \approx 0.217609$, $F(4) \approx 0.144693$, $F(5) \approx 0.092429$, $F(15) \approx 5.338854 \cdot 10^{-4}$, $F(20) \approx 3.498013 \cdot 10^{-5}$, ...

Second Note. In Subsection 38.2 on p. 103 in the notes for the course, we proved the inequality

$$\mathrm{P}(|X-nx| \geq n\delta) \leq \frac{n}{n^2\delta^2} = \frac{1}{n\delta^2}$$

for a random variable X having distribution $\mathrm{Bin}(n,x)$ (note that this X is different from the random variable X used above) and for any $\delta > 0$. In doing so, we used the inequality $V(X) < n$. If we had used the more precise value $\mathrm{V}(X) = nx(1-x)$, we would have obtained the slightly stronger inequality

$$\mathrm{P}(|X-nx| \geq n\delta) \leq \frac{x(1-x)}{n\delta^2}.$$

With $x = 1/4$ and $\delta = 1/4$ this says

$$\mathrm{P}(|X-n/4| \geq n/4) \leq \frac{3}{n}.$$

This is equivalent to

$$\sum_{\substack{k:0\leq k\leq n\\ |k-n/4|\geq n/4}} \mathrm{P}(X=k) \leq \frac{3}{n}.$$

Noting that

$$P(X=k) = \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k},$$

this implies

$$\sum_{\substack{k:0\leq k\leq n\\ |k-n/4|\geq n/4}} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} \leq \frac{3}{n}.$$

Replacing n with $4n$, this becomes

$$\sum_{\substack{k:0\leq k\leq 4n\\ |k-n|\geq n}} \binom{4n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{4n-k} \leq \frac{3}{4n}.$$

Note that the inequality $|k-n| \geq n$ for the integer k with $0 \leq k \leq 4n$ means that either $k = 0$ or $2n \leq k \leq 4n$, and omitting the term of the sum for $k = 0$ (which is positive), we obtain the inequality

$$\sum_{k=2n}^{4n} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{4n-k} < \frac{3}{4n};$$

strictly less, because we omitted a positive term on the left-hand side. This inequality is identical to the inequality to be proved above. This shows that the inequality in the problem is in essentially[1] a special case of an inequality proved in the notes.

[1]That is, aside from the slight sharpening obtained by using the exact value for $\mathrm{V}(X)$ here.