M378K Introduction to Mathematical Statistics
Problem Set #11
De Moivre-Laplace.

Problem 11.1. *You are given a TRUE/FALSE exam with 30 questions. Suppose that you need to answer 21 questions correctly in order to pass. You have no idea what the class is about and decide to toss a fair coin to answer all the questions; you circle TRUE if the outcome is tails and you circle FALSE if the outcome is heads. What is your approximation of the probability p that you manage to pass the exam using this strategy?*

For $Y \sim Binomial(n, p)$ we know that its probability mass function is:

$$p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n$$

Moreover, its expectation and its variance are

$$\mathbb{E}[Y] = np \quad \text{and} \quad Var[Y] = np(1-p).$$

Now, consider a sequence of binomial random variables $Y_n \sim Binomial(n, p)$. Note that, while the number of trials n varies, the probability of success in every trial p remains the same for all n. The *normal approximation to the binomial* is a theorem which states that

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \overset{\mathcal{D}}{\Rightarrow} N(0,1) \quad \text{as } n \to \infty.$$

Practically, this means that Y_n is "approximately" normal with mean np and variance $np(1-p)$ for "large" n. The usual rule of thumb is that both $np > 10$ and $n(1-p) > 10$.

Another practical adjustment needs to be made due to the fact that discrete distributions of Y_n are approximated by a continuous (normal) distribution. This adjustment is usually referred to as the **continuity correction**. More specifically, provided that the conditions above are satisfied, for every integer $a < b$, we have that

$$\begin{aligned}
\mathbb{P}[a \leq Y_n \leq b] &= \mathbb{P}\left[a - \tfrac{1}{2} < Y_n < b + \tfrac{1}{2}\right] \\
&= \mathbb{P}\left[\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}} < \frac{Y_n - np}{\sqrt{np(1-p)}} < \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right] \\
&\approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

where Φ, as usual, stands for the cummulative distribution function of the standard normal distribution.

For more about the history of the theorem and ideas for its proof, go to:
Wikipedia: de Moivre-Laplace.

Solution: Let us denote the number of correct answers you get using the coin-toss strategy by X. Then, $X \sim b(30, 1/2)$. The mean of X is $30 \cdot \frac{1}{2} = 15$ and its variance is $30 \cdot \frac{1}{2} \cdot \frac{1}{2} = 7.5$. So, the standard deviation of X is $\sqrt{7.5} \approx 2.74$. We can express the probability p as

$$p = \mathbb{P}[X \geq 21] = \mathbb{P}[X \geq 20.5] = \mathbb{P}[\frac{X - 15}{2.74} \geq \frac{20.5 - 15}{2.74}] \approx \mathbb{P}[\frac{X - 15}{2.74} \geq 2].$$

This probability is approximately $\Phi(+\infty) - \Phi(2) = 1 - \Phi(2) = 1 - 0.97725 = 0.02275$.

Problem 11.2. *A new addition of Kafka's "Metamorphosis" has 72 pages. The printing press often malfunctions and introduces typos. The number of typos on each page has a Poisson distribution with mean* $\ln(3)$ *and is independent of the number of typos on other pages (or other books). A book is thrown away if it contains typos on more than 32 pages. Use the normal approximation to estimate the proportion of books that get thrown away.*

Solution: By the formula for the Poisson distribution, the probability that a page contains no typos is $p = e^{-\ln(3)} = \frac{1}{3}$. The number of pages with typos per book is, therefore, binomially distributed with the mean $\mu = 72 \times \frac{2}{3} = 48$ and standard deviation $\sigma = \sqrt{72 \times \frac{2}{3} \times \frac{1}{3}} = 4$.

In the normal approximation with continuity correction, we are looking for the probability

$$\begin{aligned}\mathbb{P}\left[Z \geq \frac{32 - \frac{1}{2} - 48}{4}\right] &= \mathbb{P}[Z \geq] \\ &= 1 - \Phi(-4.125) \cong 1.\end{aligned}$$

We would, therefore, expect approximately all of the books to be thrown away. It is a really lousy printing press!

Note: If you use ‘1-pbinom(31, 72, 2/3)‘ in **R**, you get 0.9999669.