Solutions for Problem Set 12:
Topic: Central Limit Theorem
Due: No due date

1. It is not easy to calculate the CDF of the total weight and the desired probability, but an approximate answer can be quickly obtained using the central limit theorem.

   We want to calculate $\mathbf{P}(S_{102} > 200)$, where S_{102} is the sum of the weights of 102 pretzels. The mean weight of a single pretzel is
   $$\mu_W = 2\,.$$
   The variance of the weight of a single pretzel is

   $$
   \begin{aligned}
   \sigma_W^2 &= \mathbf{E}[W^2] - \mu_W^2 \\
   &= \left(\int_0^\infty w^2 f_W(w) dw \right) - 4 \\
   &= \left(\int_1^2 w^2(w-1) dw + \int_2^3 w^2(3-w) dw \right) - 4 \\
   &= \frac{25}{6} - 4 \\
   &= \frac{1}{6}\,.
   \end{aligned}
   $$

   Thus we have

   $$
   \begin{aligned}
   \mathbf{P}(S_{102} > 200) &= 1 - \mathbf{P}(S_{102} \leq 200) \\
   &= 1 - \mathbf{P}\left(\frac{S_{102} - 102 \cdot 2}{\sqrt{102 \cdot (1/6)}} \leq \frac{200 - 102 \cdot 2}{\sqrt{102 \cdot (1/6)}} \right) \\
   &\approx 1 - \Phi\left(\frac{-4}{\sqrt{17}} \right) \\
   &= 1 - \left[1 - \Phi\left(\frac{4}{\sqrt{17}} \right) \right] \\
   &= \Phi(0.9701) \\
   &\approx 0.8340\,.
   \end{aligned}
   $$

2. The probability that you will believe the fair coin to be biased is the probability that the fair coin will come up with more than 525 heads out of the 1000 tosses. Let S be the number of times the coin comes up heads, which is a binomial random variable, with parameters $n = 1000$ and $p = 0.5$, so that $\mathbf{E}[S] = 1000 \cdot 0.5 = 500$ and $\sigma_S = \sqrt{1000 \cdot 0.5 \cdot 0.5} = 5\sqrt{10}$. Using the de Moivre - Laplace normal approximation to the binomial, we have

   $$
   \begin{aligned}
   \mathbf{P}(S > 525) &= \mathbf{P}(S \geq 525.5) \\
   &= \mathbf{P}\left(\frac{S - 500}{5\sqrt{10}} \geq \frac{525.5 - 500}{5\sqrt{10}} \right) \\
   &\approx 1 - \Phi\left(\frac{25.5}{5\sqrt{10}} \right) \\
   &= 1 - \Phi(1.6128) \\
   &\approx 0.0537\,.
   \end{aligned}
   $$

3. The probability that the flight will be overbooked is the probability that less than 20 of the 320 tickets will be canceled. Let N be the number of tickets that are canceled, which is a binomial random variable, with parameters $n = 320$ and $p = 0.1$, so that $\mathbf{E}[N] = 320 \cdot 0.1 = 32$ and $\sigma_N = \sqrt{320 \cdot 0.1 \cdot 0.9} = \sqrt{28.8}$. Using the de Moivre - Laplace normal approximation to the binomial, we have

$$
\begin{aligned}
\mathbf{P}(N < 20) &= \mathbf{P}(N \le 19.5) \\
&= \mathbf{P}\left(\frac{N - 32}{\sqrt{28.8}} \le \frac{19.5 - 32}{\sqrt{28.8}}\right) \\
&\approx \Phi\left(\frac{-12.5}{\sqrt{28.8}}\right) \\
&= 1 - \Phi\left(\frac{12.5}{\sqrt{28.8}}\right) \\
&= 1 - \Phi(2.3292) \\
&\approx 0.0099.
\end{aligned}
$$

4. The pipe sections have lengths that are independent random variables. The length of each section, X_i, has mean and variance both equal to 2.

    (a) Given 400 sections of pipe, we want the total length of pipe to be greater than w with a probability of 0.841. With so many sections, the total length is well approximated as a normal random variable. As an initial guess, the probability is $1/2$ that the total length is greater than 800, so the probability will be 0.841 for some w that is smaller than 800. Define T to be the total length $X_1 + ... + X_{400}$. A couple of salient facts: T has an expected value of 800 and a standard deviation of $20\sqrt{2}$.

    $$
    P(T > w) = P(T - 800 > w - 800) = P\left(\frac{T - 800}{20\sqrt{2}} > \frac{w - 800}{20\sqrt{2}}\right)
    $$

    The left hand side of the inequality can be approximated as a standard normal random variable, and we want it to exceed $(w - 800)/20\sqrt{2}$ with probability 0.841. That happens if:
    $$
    1 - \Phi\left(\frac{w - 800}{20\sqrt{2}}\right) = 0.841 \Rightarrow \frac{800 - w}{20\sqrt{2}} = 1
    $$

    (note the change in sign of the numerator, a change that allowed the use of a published table for Φ. Solving for w yields $\boxed{w = 772}$, which is a bit less than 800 as anticipated.

    (b) Now we want the number, n, of pipe sections such that the combined pipe length is at least 200 meters with probability 0.841. First a rough guess: with 100 sections, the expected length is 200 and the length's distribution can be approximated as a normal random variable. Therefore the length is at least 200 meters with probability $1/2$. With more than 100 sections, the length will be at least 200 meters with probability greater than $1/2$, so we anticipate an answer greater than 100.

    Note that the series $X_1 + ... + X_n$ has mean $2n$ and standard deviation $\sqrt{2n}$.

    $$
    P(X_1 + ... + X_n \ge 200) = P(X_1 + ... + X_n - 2n \ge 200 - 2n) = P\left(\frac{X_1 + ... + X_n - 2n}{\sqrt{2n}} \ge \frac{200 - 2n}{\sqrt{2n}}\right)
    $$

The left hand side of the inequality is a standard normal random variable, and it exceeds the right hand side with probability

$$1 - \Phi\left(\frac{200 - 2n}{\sqrt{2n}}\right) = \Phi\left(\frac{2n - 200}{\sqrt{2n}}\right)$$

Setting that equal to 0.841, we obtain:

$$\frac{2n - 200}{\sqrt{2n}} = 1$$

Squaring and using the quadratic formula yields $\boxed{n=108}$. This time the initial guess was more than a reassuring check; it was essential to point to the selection of the positive solution from the quadratic formula.