Solution to Some Problems on "Central Limit Theorem"

5.10. If X is the number of sixes in n rolls of a (fair) six-sided die, then $X \sim$ Binomial$(n,\ p = 1/6)$. The chances of seeing 15 to 20 sixes in 100 rolls are given by

$$P(X = 15) + P(X = 16) + \cdots + P(X = 20) = 0.5607,$$

as you can easily check using R (or by hand). But the question asks us to use a normal approximation to this binomial distribution – a special case of the CLT. First we check that it is reasonable to use the approximation: the usual conditions $np > 5$ and $n(1-p) > 5$ are both satisfied here. We want to approximate

$$P(15 \leq X \leq 20) = P(X \leq 20) - P(X \leq 14),$$

where $E(X) = np = 100/6 = 16.6667$ and $\text{Var}(X) = np(1-p) = 500/36 = 13.8889$.

Application of the normal approximation to the distribution of X yields

$$P(X \leq 20) = P\left(\frac{X - 16.6667}{\sqrt{13.8889}} \leq \frac{20 - 16.6667}{\sqrt{13.8889}}\right) \cong P(Z \leq 0.8944) \approx 0.8144.$$

Similarly, $P(X \leq 14) \cong P(Z \leq -0.7155) \approx 0.2372$, so $P(15 \leq X \leq 20) \approx 0.8144 - 0.2372 = 0.5772$, a reasonable approximation to 0.5607 (about 2.9% relative error).

Because we are using a continuous distribution (the normal) to approximate a discrete distribution (the binomial distribution of X, a random variable with the integers between 0 and 100 as possible values), the approximation will be improved by use of a "continuity correction". We write

$$P(15 \leq X \leq 20) = P(14.5 \leq X \leq 20.5) = P(X \leq 20.5) - P(X < 14.5).$$

Proceeding as above, we obtain $P(X \leq 20.5) \cong P(Z \leq 1.0286) \approx 0.8482$ and $P(X < 14.5) \cong P(Z < -0.5814) \approx 0.2805$, so $P(15 \leq X \leq 20) \approx 0.5677$, a considerably improved approximation (about 1.2% relative error).

For the second part of the question, we are interested in calculating the chances that the sum of the face values showing on the 100 rolls is less than 300. If we let

$$S = \sum_{i=1}^{100} X_i,$$

where $X_i = 1, 2, \ldots, 6,$ each with probability 1/6, is the face value that appears on the i^{th} roll, then this is simply $P(S < 300)$. This probability is difficult to calculate exactly (!!), but it is easy to use the CLT to approximate it.

The "discrete uniform" distribution of X_i has $E(X_i) = 3.50$ and $\text{Var}(X_i) = 2.9167$. The X_i's are independent and identically distributed random variables, so the CLT yields

$$P(S < 300) = P\left(\frac{S - 3.50 \times 100}{\sqrt{2.9167 \times 100}} < \frac{300 - 3.50 \times 100}{\sqrt{2.9167 \times 100}}\right) \cong P(Z < -2.9277) \approx 0.0017.$$

Note that if we instead approximate $P(S \le 299)$, we get a slightly different answer: $P(S \le 299) \cong P(Z \le -2.9863) \approx 0.0014$.

Again, we are using a continuous distribution to approximate a discrete distribution (the distribution of S, a sum of 100 discrete uniform random variables), so we might want to use a "continuity correction" to improve the approximation. We write $P(S < 300) = P(S \le 299.5)$ and proceed as above to obtain

$$P(S < 300) \cong P(Z \le -2.9570) \approx 0.0016.$$

It shouldn't be too surprising that the "continuity correction" has little effect in this case because S takes on a very large number of possible values (any positive integer between 100 and 600). Hence the distribution of S can be very well-approximated as a continuous distribution.

Note that use of the "continuity correction" leads to exactly the same approximation for $P(S < 300)$ and for $P(S \le 299)$; we approximate both of these probabilities by $P(S \le 299.5)$.

5.12. Let the round-off for the i^{th} number be given by $R_i \sim \text{Unif}[-\frac{1}{2}, \frac{1}{2}]$. It follows that $E(R_i) = 0$ and $\text{Var}(R_i) = \sigma^2 = 1/12$. If

$$S_n = \sum_{i=1}^{n} R_i$$

is the sum of n such (independent and identically distributed) round-offs then, by the CLT,

$$P(S_n > x) = P(\frac{S_n}{\sigma\sqrt{n}} > \frac{x}{\sigma\sqrt{n}}) \cong P(Z > \frac{x}{\sigma\sqrt{n}}).$$

In particular,

$$P(S_{100} > x) \cong P(Z > \frac{x}{\sqrt{100/12}}).$$

Substituting we get

$$\begin{aligned} P(S_{100} > 1) &\cong P(Z > 0.3464) \approx 0.3645, \\ P(S_{100} > 2) &\cong P(Z > 0.6928) \approx 0.2442, \\ P(S_{100} > 5) &\cong P(Z > 1.7321) \approx 0.0416. \end{aligned}$$

Note that the question could be interpreted as asking for the chances that the total round-off is larger **in magnitude** than the specified numbers; that is, $P(|S_n| > x)$. In that case, because $E(S_n) = 0$, simply multiply the above answers by two.

5.16. Let $S = X_1 + \ldots + X_{20}$, where the X_i's are independent and identically distributed with density function $f(x) = 2x$ for $0 \le x \le 1$. To use the CLT, we need $E(X_i)$ and $\text{Var}(X_i)$: $E(X_i) = \int_0^1 2x^2\,dx = 2/3$ and $E(X_i^2) = \int_0^1 2x^3\,dx = 1/2$, so $\text{Var}(X_i) = 1/18$. Then, by the CLT,

$$S \approx N\Big(20 \times (2/3),\, 20 \times (1/18)\Big),$$

so

$$P(S \le 10) = P\left(\frac{S - 40/3}{\sqrt{20/18}} \le \frac{10 - 40/3}{\sqrt{20/18}}\right) \cong P\left(Z \le -3.1623\right) \approx 0.0008.$$

5.18. Let X_i be the weight of the i^{th} package. The X_i's are assumed to be independent, with $\mu_{X_i} = 15$ and $\sigma_{X_i} = 10$. For $S = \sum_{i=1}^{100} X_i$, we want to approximate $P(S > 1700)$. By the CLT,

$$P(S > 1700) = P(\frac{S - 100(15)}{10\sqrt{100}} > \frac{1700 - 1500}{10\sqrt{100}}) \cong P(Z > 2) \approx 0.0228.$$