Supplementary exercise 5.40 of IPS7e

Simple random sample (SRS) of 200 households in the area of a restaurant to determine the proportion of adults committed to eating nutritious food when dining out. Denote by X the number of respondents answering yes to the question (whether they are committed to eating nutritious food when dining out). The National Restaurant Association estimate 40% of adults to have this commitment.

(a) It is probably quite reasonable to use a binomial distribution B(200,0.4) for X if the area where the survey is conducted has a large number of households. We can write $X \sim B(200, 0.4)$.

The sampling is without replacement from a finite population of unknown size, but it seems reasonable to assume the population size to be large enough to make the binomial distribution a good approximation. (Depending of course on the location of the restaurant; if it is in a small village, the sample may even exhaust the population completely). The rule of thumb is that the population should be at least 20 times as large as the sample (lecture slide 4L−10); that is, the number of households in the area should be at least 4000.

(b) The mean is $\mathrm{E}X = 200 \cdot 0.4 = 80$, by the formula for the mean of a binomial distribution.

The question asks for the probability that X lies between 75 and 85. We will get different answers if we interpret that as all values in the range 75 − 85 including 75 and 85, or as all values in the range excluding 75 and 85. This is because the binomial distribution is discrete and has non-zero probability of taking either of the values 75 and 85 exactly. The way the question is worded, I would consider both of these interpretations as valid.

For the purpose of this solution we have chosen the latter interpretation. The probability $P(75 < X < 85)$ can be computed in different ways. The table for the binomial distribution (Stevens, Table 1) only has n up to 20, so we need to use either the software or the normal approximation.

Normal approximation:
The condition for using the normal approximation is that $n \cdot p \cdot (1-p) > 104$. We compute $200 \cdot 0.4 \cdot (1–0.4) = 48 > 10$, so we can indeed use the approximation. In order to use the formula on slide 5L−7, we need to rephrase the event is question: $P(75 < X < 85) = P(76 \le X \le 84)$. To ease the notation, we also compute $\mathrm{Var}X = np(1-p) = 48$. Then we have

$$\begin{aligned} P(76 \le X \le 84) &\approx P\left(Z < \frac{84 + 0.5 - 80}{\sqrt{48}}\right) - P\left(Z < \frac{76 - 0.5 - 80}{\sqrt{48}}\right) \\ &= P(Z < 0.6495) - P(Z < -0.6495) = 0.7420 - 0.2580 = 0.4840. \end{aligned}$$

Direct (exact) calculation in $B(200, 0.4)$:
We may use the cumulative distribution function in Minitab (after entering the values 84 and 75 in a column) to obtain:

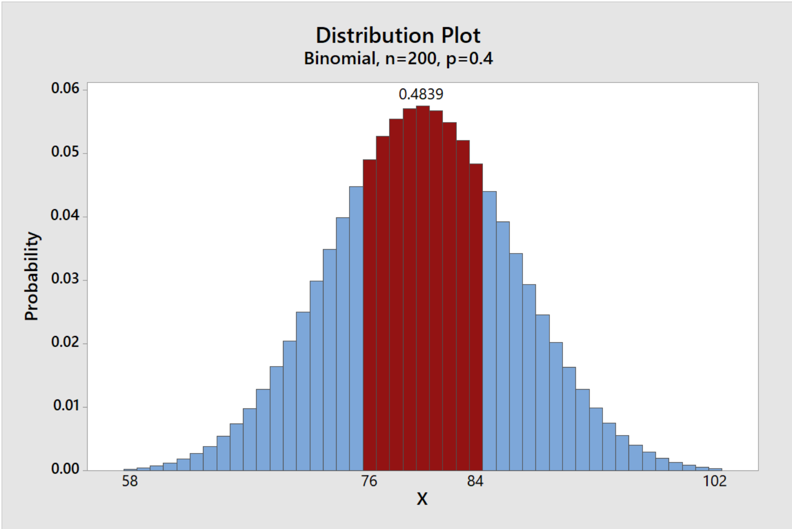
$$P(76 \le X \le 84) = P(X \le 84) - P(X \le 75) = 0.742849 - 0.258956 = 0.483893.$$

Cumulative Distribution Function

Binomial with n = 200 and p = 0.4

x	P(X ≤ x)
84	0.742849
75	0.258956

When using the Probability Distribution Plot menu directly, we get the value 0.4839.



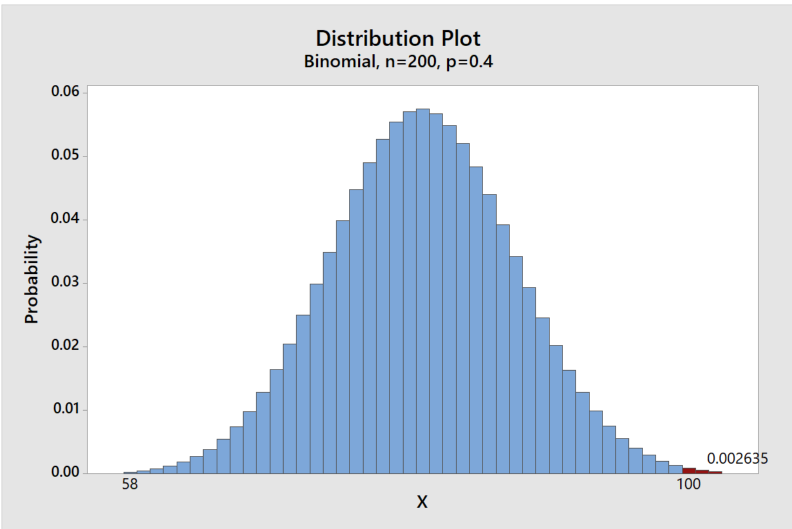
(c) As under (b), the probability $P(X \geq 100)$ can be computed in different ways. If we for the normal approximation use the formula on slide 5L−22, it is helpful to rephrase the event in question as: $P(X \geq 100) = 1 - P(X \leq 99)$. Then we have, still using $\mathrm{E}X = np = 80$ and $\mathrm{Var}X = np(1-p) = 48$,

$$\begin{aligned} P(X \geq 100) &= 1 - P(X \leq 99) \approx 1 - P\left(Z < \frac{99 + 0.5 - 80}{\sqrt{48}}\right) \\ &= 1 - P(Z < 2.8146) = P(Z < -2.8146) = 0.00244. \end{aligned}$$

Direct (exact) calculation in a B(200,0.4) using Minitab gives,

$$P(X \geq 100) = 1 - P(X \leq 99) = 1 - 0.997365 = 0.002635,$$

or again using the Probability Distribution plot menu:



This probability is low, so observing 100 out of 200 respondents being committed to nutritious food seems to indicate that the local commitment is greater than the value reported by the National Restaurant Association. Stated more precisely, the probability of observing 100 or more committed respondents by chance when the true p is 0.4, is about 0.25%. This type of reasoning is used for statistical testing; in fact, we will later see that the probability computed is the P-value (for testing the hypothesis that p equals 0.4 against a one-sided alternative hypothesis).

We see that for these calculations the normal approximations are quite close to the exact values from the binomial distribution, as we should expect when the guideline for use of the approximation is clearly met (as is the case here). In practice however, there is no need to do both calculations, and if we have access to software for exact calculations these would be our preferred choice.