HALF-POINT CORRECTION IN THE CLT APPROXIMATION

When the sample $X_1, X_2, \ldots, X_n$ comes from a discrete distribution with integer values, then the sum $Y = \sum_{i=1}^{n} X_i$ is an integer as well. Assume that the mean of X_i is μ and the variance is σ^2. We can approximate the probability of Y being between two integers $k' < k''$, denoted $P\Big(k' \le Y \le k''\Big)$:

Directly with the Central Limit Theorem

$$P\Big(k' \le Y \le k''\Big) = P\Big(\frac{k' - n\mu}{\sqrt{n\sigma^2}} \le \frac{Y - n\mu}{\sqrt{n\sigma^2}} \le \frac{k'' - n\mu}{\sqrt{n\sigma^2}}\Big) \simeq \Phi(\frac{k'' - n\mu}{\sqrt{n\sigma^2}}) - \Phi(\frac{k' - n\mu}{\sqrt{n\sigma^2}})\,.$$

The **half-point correction** gives a $1/2 = 0.5$ correction above and below the integer values as follows:

$$P\Big(k' \le Y \le k''\Big) = P\Big(k' - \frac{1}{2} \le Y \le k'' + \frac{1}{2}\Big) =$$

$$P\Big(\frac{k' - \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}} \le \frac{Y - n\mu}{\sqrt{n\sigma^2}} \le \frac{k'' + \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}}\Big) \simeq \Phi(\frac{k'' + \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}}) - \Phi(\frac{k' - \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}})\,.$$

The approximation is improved.

If one of the endpoints is missing, we have simply

$$P\Big(Y \le k''\Big) = P\Big(Y \le k'' + \frac{1}{2}\Big) \simeq \Phi(\frac{k'' + \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}})$$

and of course

$$P\Big(k' \le Y\Big) = P\Big(k' - \frac{1}{2} \le Y\Big) \simeq 1 - \Phi(\frac{k' - \frac{1}{2} - n\mu}{\sqrt{n\sigma^2}})\,.$$

Finally, we must pay attention to the inequality sign. If the sign is strict, i.e. $<$ or $>$, we convert it into $\le$ or $\ge$, respectively.

$$P\Big(k' < Y \le k''\Big) = P\Big(k' + 1 \le Y \le k''\Big)$$
$$P\Big(k' \le Y < k''\Big) = P\Big(k' \le Y \le k'' - 1\Big)$$
$$P\Big(k' < Y < k''\Big) = P\Big(k' + 1 \le Y \le k'' - 1\Big)$$

and proceed like before.

Read carefully **Examples 3,4, 5** in the text in *Approximations for Discrete Distributions* (Section 5.7 in the 8th Ed.)

Exercise. Let $Y \sim Bin(25, .45)$. Compare $P(10 \le Y \le 12)$ obtained (i) exactly, from Table II; (ii) directly with the CLT; and (iii) using the half-point approximation.

Answer: (i) .4513; (ii) .3255; (iii) .4673.