10 Normal Distribution Approximations

10.1 Normal Approximation to Binomial

- Normal arises when adding many similar independent quantities.
- Binomial is the sum of independent binomial random variables

Let $X \sim Binomial(n,p)$. We know $\mu = np$, and $\sigma^2 = np(1-p)$. Normal approximation says that $P(X \leq x)$ can be approximated by a normal

$$X \sim N(np, np(1-p)) \tag{1}$$

And can be evaluated using normal tables by standardization if n is sufficiently large

$$P(X \leq x) = P\left(Z \leq \frac{x - np}{\sqrt{np(1-p)}}\right) \tag{2}$$

We need to make sure that the $\mu = np$ and $\sigma^2 = np(1-p)$ is greater than 5 ($\mu, \sigma^2 > 5$)

The reason for the mean and the variance to be greater than 5, is because the approximation only works when the distribution is near the center.

Example 4.17 Digital communication channel receives X number of bits as error and can be modelled as a $Binomial(16000000, 0.00001)$.

Ans. The answer to this question is very difficult to compute

$$P(X \leq 150) = \sum_{x=0}^{150} \binom{16000000}{x} (0.00001)^x (0.99999)^1 6000000$$

There is one additional note about the normal approximation to the binomial distribution: we have to get a bit more accurate. We add and subtract 0.5 from x in order to account for a wider distribution.

$$P(X \leq x) = P(X \leq x + 0.5) = P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) \tag{3}$$

$$P(x \leq X) = P(x - 0.5 \leq X) = P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \leq Z\right) \tag{4}$$

Example 4.18 Recall Example 4.17 above. We can compute that much more easily with a normal distribution approximation

We know $\mu = np = 160$, $\sigma = \sqrt{np(1-p)} = 12.6490$, thus

$$P(X \leq 150) = P\left(Z \leq \frac{150 - 160}{12.6490}\right) = P(Z \leq -0.7906) = 0.2146$$

With correction

$$P(X \leq 150) = P\left(\frac{150 + 0.5 - 160}{12.6490}\right) = P(Z \leq -0.7510) = 0.2263$$

The exact answer computed by Matlab is 0.288, so we see that the correction is a better result.

10.2 Normal Approximation to Hypergeometric

Recall that a binomial distribution approximates a Hypergeometric when the sample size n is much smaller than the population size N. Text suggests $\frac{n}{N} < 0.1$. More conservative range requires $\frac{n}{N} < 0.05$. Here $p = \frac{K}{N}$. So here $p = \frac{K}{N}$. So if $\frac{n}{N} < 0.05$ and n is large, we can use the normal approximation to the Hypergeometric. Look at examples on the next page.