Chapter 8 – Normal Random Variables

The *Normal* or Gaussian pdf with mean μ and variance σ^2 is given at the top of p. 203, and repeated here:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty$$

We write $X \sim N(\mu,\sigma^2)$. The cdf, $F(x) = \int_{-\infty}^{x} f(t)dt$, has no closed form solution; cumulative probabilities are tabled (front cover and pp.508-9 of textbook) for the *Standard Normal distribution* whose pdf is:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty$$

Here $Z = (X-\mu)/\sigma$, and since $X \sim N(\mu,\sigma^2)$, it follows that $Z \sim N(0,1)$. On p. 221, it is pointed out that the cdf of the Standard Normal distribution is usually denoted $\Phi(z)$.

Important:

- **On pp. 206-7, it is shown that the mean of a $N(\mu,\sigma^2)$ RV is μ;**
- **On pp. 207-8, it is shown that the variance of a $N(0,1)$ RV is 1, and so it follows (p. 209) that the variance of a $N(\mu,\sigma^2)$ RV is σ^2;**
- **$\Phi(-1.96) \approx 0.025$ and $\Phi(1.96) \approx 0.975$, so we use the value '1.96' instead of '2' for 95% CI's (confidence intervals);**
- **Many things/phenomena in nature follow the Normal curve**

Example – X = Brain weight $\sim N(1400.48, 106.328^2)$. Randomly pick one such brain, and find (a) $\Pr\{1300 < X < 1500\}$, (b) the 31st percentile of brain weights, and (c) the proportion of brains with weights exceeding 1600 grams.

(a) $\Pr\{1300 < X < 1500\} = \Pr\{-0.95 < Z < 0.94\} = \Phi(0.94) - \Phi(-0.95) = 0.82639 - 0.17106 = 0.65533$.

(b) Z = -0.50 is close enough since $\Phi(-0.50) = 0.30854$ is the closest entry to 0.31 = 31%. Then, X = 1400.48 – 0.50*106.328 = 1347.3g
(c) Pr{X > 1600} = Pr{Z > 1.88} = 1 - $\Phi(1.88)$ = 1 – 0.96995 = 0.03005.
Note: to get more precise answers above, use Minitab or a calculator.

Provided we use the *Continuity Correction* (CC), the Normal curve can be used to *approximate the Binomial* for large n and small π. Since for the Binomial the mean is $\mu = n\pi$ and the variance is $\sigma^2 = n\pi(1-\pi)$, these are the values we will use for the Normal approximation.

Code errors example (p.229) – K = number of signals misread, here $\pi = 0.03$ and n = 1000 (bits), so $\mu = 30$ and $\sigma = 5.3944$. We want Pr{K ≤ 35}, which from Minitab is actually 0.846076, but let's use the normal approximation to illustrate.

- **Without CC: Since the standard score is $(35-30)/5.3944 = 0.9269 \approx 0.93$, Pr{K ≤ 35} ≈ Pr{Z ≤ 0.93} ≈ $\Phi(0.93)$ = 0.8238 (off by 2.2%)**
- **With CC: Here we use $(35.5-30)/5.3944 = 1.0196 \approx 1.02$, Pr{K ≤ 35} ≈ Pr{Z ≤ 1.02} ≈ $\Phi(1.02)$ = 0.8461 (really close!)**

You could even use the Normal Approximation to estimate the value of a Binomial exact probability such as Pr{K = 13} as on p.230, but use the CC. But remember to heed the conditions on p.231: Normal Approximation okay provided both $n\pi \geq 5$ and $n(1-\pi) \geq 5$.

The CLT (Central Limit Theorem) on p.234 tells us for independent RVs – none of which dominate the others – the sum of these RVs approaches the Normal curve as the number in the summand grows. It's not surprising, then, that the Binomial distribution is often well approximated by the Normal one (since a Binomial RV is the sum of independent Bernoulli RVs).

In-class exercises: #8.35 on p.237 and #8.23, #8.25 of p.232 if time.