STAT 2857A – Lecture 25 Examples and Exercises

Solutions

Example 25.1

a) We know that:

$$\mu_X = 174.2$$
$$\sigma_X^2 = 42.36$$
$$\sigma_X = 6.508.$$

Since $X_1, \ldots, X_{25}$ are independent we know that T_{25} has mean

$$\begin{aligned}
\mu_T &= E(\sum_{i=1}^{25} X_i) \\
&= \sum_{i=1}^{25} E(X_i) \\
&= 25(174.2) \\
&= 4355
\end{aligned}$$

and variance

$$\begin{aligned}
\sigma_T^2 &= V(\sum_{i=1}^{25} X_i) \\
&= \sum_{i=1}^{25} V(X_i) \\
&= 25(42.36) \\
&= 1059.
\end{aligned}$$

The final piece is that the total is normally distributed. Hence:

$$T_{25} \sim \text{Normal}(4355, 1059).$$

b) Consider that the average (i.e., the sample mean) is a linear function of T_{25}

$$\bar{X}_{25} = \frac{T_{25}}{25}.$$

So,

$$\begin{aligned} E(\bar{X}_{25}) &= \frac{E(T_{25})}{25} \\ &= \frac{4355}{25} \\ &= 174.2 \end{aligned}$$

and

$$\begin{aligned} V(\bar{X}_{25}) &= V\left(\frac{T_{25}}{25}\right) \\ &= \frac{V(T_{25})}{25^2} \\ &= \frac{1059}{625} \\ &= 1.694. \end{aligned}$$

And the final piece – $\bar{X}_{25}$ is normally distributed! So

$$\bar{X}_{25} \sim \text{Normal}(174.2, 1.694).$$

c) Finally, consider that

$$Z = \frac{(\bar{X}_{25} - 174.20)}{\sqrt{1.69}} = \frac{\bar{X}_{25} - E(\bar{X}_{25})}{\sqrt{V(\bar{X}_{25})}}.$$

Hence, $Z \sim \text{Normal}(0, 1)$.

Example 25.2

a) Consider that $\bar{X}_n = \sum_{i=1}^n X_i/n$. Then the possible values of $\bar{X}_n$ are $0, 1/n, 2/n, \ldots, (n-1)/n, 1$ and $T_n = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$. So

$$\begin{aligned} P(\bar{X}_n = \bar{x}) &= P(\sum_{i=1}^n X_n = n\bar{x}) \\ &= P(T_n = n\bar{x}) \\ &= \binom{n}{n\bar{x}} p^{n\bar{x}}(1-p)^{n-n\bar{x}} \end{aligned}$$

for $\bar{x} \in \{0, 1/n, 2/n, \ldots, (n-1)/n, 1\}$. It may seem confusing to see $n\bar{x}$ in the binomial, but note that this value is always going to be an integer.

b) Consider that

$$\mu = E(X_i) = p$$

and

$$\sigma = V(X_i) = p(1-p).$$

So

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Applying the central limit theorem directly implies that

$$Z \mathrel{\dot\sim} \text{Normal}(0,1)$$

meaning that

$$P\left(\frac{(\bar{X}_n - E(\bar{X}_n))}{\sqrt{V(\bar{X}_n)}} \leq z\right) \approx P(Z \leq z)$$

where Z is standard normal, provided that n is big enough.

In fact, we have already seen this result. Let T_n be the total number of successes, $T_n = \sum_{i=1}^n X_i$. Then

$$T_n \sim \text{Binomial}(n,p).$$

Manipulating Z, we get that

$$\begin{aligned}
P(Z \leq z) &= P\left(\frac{(\bar{X}_n - p)}{\sqrt{p(1-p)/n}} \leq z\right) \\
&= P\left(\frac{(n\bar{X}_n - np)}{\sqrt{np(1-p)}} \leq z\right) \\
&= P\left(\frac{(T_n - np)}{\sqrt{np(1-p)}} \leq z\right) \\
&\approx P\left(\frac{(T_n + .5 - np)}{\sqrt{np(1-p)}} \leq z\right)
\end{aligned}$$

since $.5 - np \approx -np$ if n is big. But this is simply the normal approximation to the binomial which we have seen before. If $T_n \sim \text{Binomial}(n,p)$ and n is large enough, then

$$\frac{(T_n + .5 - np)}{\sqrt{np(1-p)}} \mathrel{\dot\sim} \text{Normal}(0,1).$$

The normal approximation is simply the central limit theorem in disguise.

Exercise 25.3

See slides.

Exercise 25.4

Let $X_1, \dots, X_{50}$ denote the length of string on each of the balls and $\bar{X}_{50}$ the average length of string. Then the total is $T_n = 50\bar{X}_{50}$. We want to find the value t such that

$$\begin{aligned}
P(T_n \le t) &= .95 \\
P\left(\frac{T_n}{50} \le \frac{t}{50}\right) &= .95 \\
P\left(\bar{X}_{50} \le \frac{t}{50}\right) &= .95 \\
P\left(\frac{\bar{X}_{50} - 101}{.2/\sqrt{50}} \le \frac{t/50 - 101}{.2/\sqrt{50}}\right) &= .95 \\
P\left(Z \le \frac{t/50 - 101}{.2/\sqrt{50}}\right) &\approx .95
\end{aligned}$$

where $Z \sim N(0, 1)$ by the CLT. Then

$$P(Z \le 1.645) = .95$$

so we set

$$\frac{t/50 - 101}{.2/\sqrt{50}} = 1.645$$

and solve for t which gives

$$t = 50[1.645(.2/\sqrt{50}) + 101] = 5052.326.$$

The 95-th percentile of the total amount of string in a box of 50 spools is approximately 5052.326~m.

Exercise 25.1

a) The distribution is unimodal (there is only one peak) and right-skewed (the tail on the right of the peak extends farther then the tail on the left of the peak).

b) The central limit theorem tells us that if the sample size is large enough (the rule of thumb is $n > 30$), then $\bar{W}_n$ will be approximately normally distributed with mean

$$E(\bar{W}_n) = 91.45$$

and variance

$$V(\bar{W}_n) = \frac{1970.83}{n}.$$

Alternatively, we can say that

$$Z_n = \frac{\bar{W}_n - 91.45}{\sqrt{1970.83/n}}$$

is approximately standard normal. The two are equivalent.

c) The approximation means that we can approximate probabilities for $\bar{W}_n$ with probabilities computed from a normal distribution. Specifically, we can approximate the cdf of $\bar{W}_n$ by

$$P(\bar{W}_n \le w) \approx P\left(Z \le \frac{w - 91.45}{\sqrt{1970.83/n}}\right)$$

where Z is a standard normal random variable ($Z \sim \text{Normal}(0, 1)$). Technically,

$$\lim_{n\to\infty} P(\bar{W}_n \le w) = P\left(Z \le \frac{w - 91.45}{\sqrt{1970.83/n}}\right).$$

d) The approximation implies that

$$\begin{aligned}
P(\mu_W - \epsilon < \bar{W}_n < \mu_W + \epsilon) &= P(91.45 - \epsilon < \bar{W}_n < 91.45 + \epsilon) \\
&= P(\bar{W}_n \le 91.45 + \epsilon) - P(91.45 - \epsilon \le \bar{W}_n) \\
&= P\left(Z \le \frac{91.45 + \epsilon - 91.45}{\sqrt{1970.83/n}}\right) - P\left(Z \le \frac{91.45 - \epsilon - 91.45}{\sqrt{1970.83/n}}\right) \\
&= P\left(\frac{-\epsilon}{\sqrt{1970.83/n}} \le Z \le \frac{\epsilon}{\sqrt{1970.83/n}}\right)
\end{aligned}$$

where Z is standard normal. Then note that

$$\lim_{n\to\infty} \frac{-\epsilon}{\sqrt{1970.83/n}} = \lim_{n\to\infty} -\frac{\sqrt{n}\epsilon}{\sqrt{1970.83}} = -\infty$$

and similarly

$$\lim_{n\to\infty} \frac{\epsilon}{\sqrt{1970.83/n}} = \infty.$$

Hence,

$$\lim_{n\to\infty} P(\mu_W - \epsilon < \bar{W}_n < \mu_W + \epsilon) = P(-\infty < Z < \infty) = 1.$$

This means that for any small interval about the mean, $(\mu - \epsilon, \mu + \epsilon)$, the probability that $\bar{W}_n$ is inside this interval will be very close to 1 if n is large enough, and the probability will increase as n increases. We say that $\bar{W}_n$ converges to μ_W in probability.