1 Central Limit Theorem

Previously, we discussed the distribution of the sample mean ($\bar{X}$) when the underlying distribution of $X_1, X_2, ..., X_n$ was Normal. In practice however the underlying distribution isn't always Normal however, so what do we do?

Theorem 1 (Central Limit Theorem). *Let $X_1, X_2, ...X_n$ be i.i.d. with $E[X_i], V[X_i] < \infty$. Then*

$$U_n = \frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}}$$

as a sequence of random variables converges to the standard normal distribution. That is:

$$\lim_{n\to\infty} F_{U_n}(u) = \lim_{n\to\infty} P(U_n < u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t \text{ for all } u$$

- In practice, we never have an infinite sample size
- But this limit does mean that if our sample is large enough, we can *approximate* U by the standard normal distribution
- How large is large enough can vary from situation to situation. For this class we will assume that for any situation, 50 or more data points will be sufficient.

1.1 Examples

1. Let $X_1, X_2, ..., X_{100}$ be i.i.d. $U(0, 1)$.
   We know that
   $$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - \frac{1}{2}}{\sqrt{1/(12 * 100)}}$$
   Is *approximately* normally distributed with a mean of 0, and a standard deviation of 1.

2. Let $X_1, X_2, ..., X_{100}$ be i.i.d. $Exp(1)$.
   We know that
   $$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - 1}{\sqrt{1/100}}$$
   Is *approximately* normally distributed with a mean of 0, and a standard deviation of 1.

3. Let $X_1, X_2, ..., X_{100}$ be i.i.d. $Bernoulli(.5)$.
   We know that
   $$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - \frac{1}{2}}{\sqrt{1/(4 * 100)}}$$
   Is *approximately* normally distributed with a mean of 0, and a standard deviation of 1.

Note for the last example we actually know what the distribution of $\sum_{i=1}^{100} X_i$ is (it is Binomial(100,1/2)). This implies that we can actually approximate the binomial distribution hen n is large with a normal distribution.

2 Exercises

1. Let$X_1, X_2, ..., X_{121}$ be i.i.d. $\Gamma(2, 1)$. Construct a statistic that is approximately distributed $N(0, 1)$
2. Let$X_1, X_2, ..., X_{144}$ be i.i.d. $Beta(2, 2)$. Construct a statistic that is approximately distributed $N(0, 1)$
3. Let$X_1, X_2, ..., X_{100}$ be i.i.d. $Exp(\delta)$, where δ is assumed to be known. Construct a statistic that is approximately distributed $N(0, 1)$
4. Let$X_1, X_2, ..., X_{100}$ be i.i.d. $Poisson(\lambda)$, where λ is assumed to be known. Construct a statistic that is approximately distributed $N(0, 1)$

2.1 Solutions

1. We know that since $X_1 \sim \Gamma(2, 1)$ $E[X_1] = 2$ and $V[X_1] = 2$. Therefore:
$$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - 2}{\sqrt{2/121}}$$
is approximately distributed $N(0, 1)$
2. We know that since $X_1 \sim Beta(2, 2)$ $E[X_1] = \frac{1}{2}$ and $V[X_1] = \frac{1}{20}$. Therefore:
$$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - \frac{1}{2}}{\sqrt{1/(20 * 144)}}$$
is approximately distributed $N(0, 1)$
3. We know that since $X_1 \sim Exp(\delta)$ $E[X_1] = \delta$ and $V[X_1] = \delta^2$. Therefore:
$$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - \delta}{\sqrt{\delta^2/100}}$$
is approximately distributed $N(0, 1)$
4. We know that since $X_1 \sim Poisson(\lambda)$ $E[X_1] = \lambda$ and $V[X_1] = \lambda$. Therefore:
$$\frac{\bar{X} - E[X_1]}{\sqrt{V[X_1]/n}} = \frac{\bar{X} - \lambda}{\sqrt{\lambda/100}}$$
is approximately distributed $N(0, 1)$