CENTRAL LIMIT THEOREM & NORMAL APPROXIMATIONS [DEVORE 5.4]

- **EXPECTED VALUE & VARIANCE OF iid RV's:** Given random sample $X_1, \ldots, X_n$ and $c_1, \ldots, c_n \neq 0$. Then:
  - ⋆ $\mathbb{E}[c_1X_1 + \cdots + c_nX_n] = c_1\mathbb{E}[X_1] + \cdots + c_n\mathbb{E}[X_n]$
  - ⋆ $\mathbb{V}[c_1X_1 + \cdots + c_nX_n] = c_1^2\mathbb{V}[X_1] + \cdots + c_n^2\mathbb{V}[X_n]$
- **SAMPLE MEAN PROPERTIES:** Given a random sample from a distribution with mean μ & variance σ^2. Then:

$$\mu_{\overline{X}} = \mathbb{E}[\overline{X}] = \mu \qquad \sigma^2_{\overline{X}} = \mathbb{V}[\overline{X}] = \sigma^2/n \qquad \sigma_{\overline{X}} = \sigma/\sqrt{n}$$

- **SAMPLE TOTAL PROPERTIES:** Given a random sample from a distribution with mean μ & variance σ^2. Then:

$$\mu_{X_1+\cdots+X_n} = \mathbb{E}[X_1 + \cdots + X_n] = n\mu \qquad \sigma^2_{X_1+\cdots+X_n} = \mathbb{V}[X_1 + \cdots + X_n] = n\sigma^2 \qquad \sigma_{X_1+\cdots+X_n} = \sigma\sqrt{n}$$

- **SAMPLE MEAN & TOTAL MEAN OF RANDOM SAMPLE FROM NORMAL POPULATION:**
  Let random sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$. Then:
  - For any sample size $n > 1$, $\overline{X} \sim \text{Normal}(\mu, \sigma^2/n)$
  - For any sample size $n > 1$, $X_1 + \cdots + X_n \sim \text{Normal}(n\mu, n\sigma^2)$

- **CENTRAL LIMIT THEOREM:**

  Let $X_1, \ldots, X_n$ be a random sample from a non-normal distribution with mean μ and variance σ^2. Then:
  - The sample mean $\overline{X}$ is approximately normal as follows: $\overline{X} \overset{approx}{\sim} \text{Normal}(\mu, \sigma^2/n)$
  - The sample total $X_1 + \cdots + X_n$ is approximately normal as follows: $X_1 + \cdots + X_n \overset{approx}{\sim} \text{Normal}(n\mu, n\sigma^2)$

  Requirement for this normal approximation to be valid: $n > 30$

  The larger the sample size n, the better the approximation.

- **NORMAL APPROXIMATION TO THE BINOMIAL:** Let $X \sim \text{Binomial}(n, p)$. Then:

$$X \overset{approx}{\sim} \text{Normal}(\mu = np, \sigma^2 = npq) \qquad (\text{where } q = 1 - p)$$

$$\mathbb{P}(X \leq x) = \text{Bi}(x; n, p) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right) \qquad (\text{where } q = 1 - p)$$

  Requirement for this normal approximation to be valid: $\min\{np, nq\} \geq 10$ (i.e. $np \geq 10$ and $nq \geq 10$)

  Remember that normal distributions are symmetric. If $\min\{np, nq\} < 10$, then the binomial distribution is too skewed.

  The **continuity correction term** "+ 0.5" improves the approximation.

- **NORMAL APPROXIMATION TO THE POISSON:** Let $X \sim \text{Poisson}(\lambda)$. Then:

$$X \overset{approx}{\sim} \text{Normal}(\mu = \lambda, \sigma^2 = \lambda)$$

$$\mathbb{P}(X \leq x) = \text{Pois}(x; \lambda) \approx \Phi\left(\frac{x + 0.5 - \lambda}{\sqrt{\lambda}}\right)$$

  Requirement for this normal approximation to be valid: $\lambda > 20$

  Remember that normal distributions are symmetric. If $\lambda \leq 20$, then the binomial distribution is too skewed.

  The **continuity correction term** "+ 0.5" improves the approximation.

EX 5.4.1: Let random sample $X_1, X_2, \ldots, X_{20} \overset{iid}{\sim} \text{Normal}(\mu = 3, \sigma^2 = 10)$.

(a) What is the probability that the sample mean is at most four?

(b) What is the probability that the sample mean is between 2.5 and 3.8?

(c) What is the probability that the sample mean is within two standard deviations of the population mean?

EX 5.4.2: Let $X_1, X_2, \ldots, X_{60}$ be a random sample of some non-normal population with mean $\mu = 10$ & variance $\sigma^2 = 80$.

(a) What is the probability that the sample mean is at most eight?

(b) What is the probability that the sample mean is at least 12.5?

(c) What is the probability that the sample mean is outside 1.5 standard deviations of the population mean?

EX 5.4.3: Suppose there are fifty plane flights from the Dallas airport to the Lubbock airport in a single day. Moreover, suppose that for every ten flights on a given day, three of them arrive late to the gate.

(a) What is the approximate probability that exactly ten flights are late in a day?

(b) What is the approximate probability that at most ten flights are late in a day?

(c) What is the approximate probability that at least ten flights are late in a day?

EX 5.4.4: A study of a bank's teller lines indicated that all day four customers are expected to be waiting per hour. Assume that the number of customers waiting in line is modeled by a Poisson distribution.

(a) What is the approximate probability that upon visiting the bank from 9am to 7pm that thirty customers are waiting?

(b) What is the approx. probability that upon visiting the bank from 9am to 7pm that at least 30 customers are waiting?