Common Claim Frequency Distributions

- Binomial Distribution, $X \sim \text{binomial}(n,p)$:

  $$p(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x},$$

  where $x \in 0,1,2,\ldots,n$, $n \in 1,2,\ldots$, and $p \in [0,1]$,

  $E[X] = np$, and $Var[X] = np(1-p)$.

  If $X \sim \text{binomial}(1,p)$, then $X \sim \text{Bernoulli}(p)$.

  If $X_1 \sim \text{binomial}(n_1,p)$ and $X_2 \sim \text{binomial}(n_2,p)$ are independent random variables, then $X_1 + X_2 \sim \text{binomial}(n_1 + n_2, p)$.

- Poisson Distribution, $X \sim \text{Poisson}(\lambda)$:

  $$p(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!},$$

  where $x \in 0,1,2,\ldots$, and $\lambda \in (0,+\infty)$,

  $E[X] = \lambda$, and $Var[X] = \lambda$.

  If $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ are independent random variables, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

- Negative Binomial Distribution, $X \sim \text{negative binomial}(r,p)$:

  $$p(x|r,p) = \binom{r+x-1}{r-1} p^r (1-p)^x,$$

  where $x \in 0,1,2,\ldots$, $r \in 1,2,\ldots$, and $p \in [0,1]$.

  $E[X] = \dfrac{r(1-p)}{p}$, and $Var[X] = \dfrac{r(1-p)}{p^2}$.

  If $r = 1$, then $X \sim \text{geometric}(p)$.

  If $X_1 \sim \text{negative binomial}(r_1,p)$ and $X_2 \sim \text{negative binomial}(r_2,p)$ are independent random variables, then $X_1 + X_2 \sim \text{negative binomial}(r_1 + r_2, p)$.

The Normal Approximation

- Central Limit Theorem:

  If $X_1, X_2, \ldots, X_n \sim$ i.i.d. $F(x)$ such that $\mu = E[X]$ and $\sigma^2 = Var[X] < +\infty$ both exist, then

  $$\Pr\left\{ \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z \right\} \to \Pr\{Z \le z\} = \Phi(z)$$

  as $n \to +\infty$.

- The central limit theorem often allows one to assume that the sample mean, $\overline{X}$, is approximately $N\left(\mu, \frac{\sigma^2}{n}\right)$, or that the sum $\sum_{i=1}^{n} X_i$ is approximately $N\left(n\mu, n\sigma^2\right)$.

  The normal approximation becomes more accurate as the sample size, n, increases. For many applications, n should be > 25.

  For a fixed sample size, the accuracy of the normal approximation is generally better for random variables X_i that are less skewed.

- Continuity Correction for Discrete Distributions:

  Suppose that one wishes to use the normal approximation to find

  $$\Pr\{a \le Y \le b\}, \quad \Pr\{a \le Y < b\}, \quad \Pr\{a < Y \le b\}, \quad \text{or} \quad \Pr\{a < Y < b\},$$

  where Y is a discrete random variable defined on the integers, and a and b are both integers.

  In modeling Y as a continuous normal random variable, it is more precise to recognize the discrete nature of a and b by either adding or subtracting $\frac{1}{2}$ from these values when normalizing Y.

- Let $E[Y] = \mu$ and $Var[Y] = \sigma^2$. Then:

  $$\Pr\{a \le Y \le b\} \approx \Pr\left\{ \frac{a - \frac{1}{2} - \mu}{\sigma} \le Z \le \frac{b + \frac{1}{2} - \mu}{\sigma} \right\} = \Phi\left(\frac{b + \frac{1}{2} - \mu}{\sigma} \right) - \Phi\left(\frac{a - \frac{1}{2} - \mu}{\sigma} \right), \text{ or}$$

  $$\Pr\{a < Y < b\} \approx \Pr\left\{ \frac{a + \frac{1}{2} - \mu}{\sigma} \le Z \le \frac{b - \frac{1}{2} - \mu}{\sigma} \right\} = \Phi\left(\frac{b - \frac{1}{2} - \mu}{\sigma} \right) - \Phi\left(\frac{a + \frac{1}{2} - \mu}{\sigma} \right).$$