

Probability and random variables

✓ Basic probability rules:

- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A | B) = \frac{P(A \cap B)}{P(B)}$

✓ Independence:

- $P(A | B) = P(A)$
- $P(A \cap B) = P(A)P(B)$

✓ Discrete random variable:

- $\mu = \sum x_i p_i$
- $\sigma^2 = \sum (x_i - \mu)^2 p_i = \sum x_i^2 p_i - \mu^2$

✓ Binomial distribution:

- $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- $\mu = np, \sigma = \sqrt{np(1-p)}$
- Conditions (BINS): binary outcomes, independent trials, a fixed number of trials, the same probability of success

✓ Geometric distribution:

- $P(X = k) = (1 - p)^{k-1} p$
- $\mu = \frac{1}{p}, \sigma = \frac{\sqrt{1-p}}{p}$

✓ Bayes' theorem: $P(A | B) = \frac{P(B | A)P(A)}{P(B)}$

✓ Linear transformations:

- $E(aX + b) = aE(X) + b$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$

Sampling distributions

✓ **Central limit theorem:** for large n (typically $n \geq 30$), the sampling distribution of $\bar{x}$ is approximately normal whatever the population's shape, if the observations are independent

✓ **Unbiased estimator:** the mean of its sampling distribution equals the parameter

✓ **Sampling variability:** a statistic varies from sample to sample; $\hat{p}$ and $\bar{x}$ are unbiased estimators of p and μ

✓ **10% condition:** when sampling without replacement, keep n at no more than 10% of the population to treat observations as independent

✓ Sampling distribution of a proportion: $\hat{p} \sim$

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

✓ **Conditions:** $np \geq 10$ and $n(1-p) \geq 10$, with a random sample of no more than 10% of the population

✓ Sampling distribution of a mean: $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$