Sampling with replacement

P(▲|▲) = P(▲|●)
= P(▲|■)
= P(▲)

When drawing with replacement, draws are independent

P(▲ and ▲) = P(▲|▲) × P(▲)
= P(▲) × P(▲)
= 3/10 × 3/10 = 9/100

Sampling without replacement

P(▲|▲) ≠ P(▲)

When drawing without replacement, draws are not independent

P(▲ and ▲) = P(▲|▲) × P(▲)
= 2/9 × 3/10 = 15/90 = 1/6

Relevance to statistics

- Our mathematical methods assume independent sampling, but we don't sample with replacement
- Deviation from independence is smaller when population is large
- For our purposes, we will require that the population of interest is much larger than our sample

Random Variable

Random Variable

A random process or variable with a numerical outcome

$$X = x_i$$

Random variable Possible outcome

Discrete random variable: integer outcomes
E.g., # of books, # of credit hours

Continuous random variable: decimal outcomes
E.g., cost of books, grade point average

Expected Value (Mean)

期望

"Average" outcome of a random variable:

$$E[X] = x_1 \times P(X = x_1) + \ldots + x_k \times P(X = x_k)$$
$$= \sum_{i=1}^{k} x_i P(X = x_i)$$

Compare to data set:

$$\mu = x_1 \times \frac{1}{n} + x_2 \times \frac{1}{n} + \ldots + x_n \times \frac{1}{n}$$

△ outcomes have different p

Expected Value Example

Is it worth $1 to play if you draw a card and get:

- $1 if it's a heart
- $5 if it's an ace
- $10 it's the king of spades

Event	*X*	*P(X)*	*X P(X)*
Heart (not ace)	$1	12/52	$12/52
Ace	$5	4/52	$20/52
King of spades	$10	1/52	$10/52
Other	$0	35/52	$0/52

E[X] = $42/52
≈ $0.81

Law of Large Numbers

Variability of Random Variables

$$Var(X) = (x_1 - E[X])^2 P(X = x_1) + \ldots + (x_k - E[X])^2 P(X = x_k)$$
$$= \sum_{i=1}^{k} (x_i - E[X])^2 P(X = x_i)$$
$$StdDev(X) = \sqrt{\sum_{i=1}^{k} (x_i - E[X])^2 P(X = x_i)}$$

Variability Example

Event	*X*	*P(X)*	*X P(X)*
Heart (not ace)	$1	12/52	$12/52
Ace	$5	4/52	$20/52
King of spades	$10	1/52	$10/52
Other	$0	35/52	$0/52

$$StdDev(X) = \sqrt{\left(1 - \tfrac{42}{52}\right)^2 \tfrac{12}{52} + \left(5 - \tfrac{42}{52}\right)^2 \tfrac{4}{52} + \left(10 - \tfrac{42}{52}\right)^2 \tfrac{1}{52} + \left(0 - \tfrac{42}{52}\right)^2 \tfrac{35}{52}} \approx 1.85$$

Variability of Random Variables

The sample mean of a large number of trials of a random variable X will be approximately E[X]

The sample standard deviation of a large number of trials of X will be approximately StdDev(X)

Linear Combination

Linear Combinations

Given random variables *Y* and *Z* and fixed numbers *a* and *b*, a **linear combination** is:

$$a \times Y + b \times Z$$

The expected value of a linear combination is:

$$E[a \times Y + b \times Z] = a \times E[Y] + b \times E[Z]$$

The variance of a linear combination is:

$$Var(a \times Y + b \times Z) = a^2 \times Var(Y) + b^2 \times Var(Z)$$

Linear Combination Example

On average you spend:

- 10 minutes on each statistics problem
- 15 minutes on each chemistry problem

If you're assigned 5 statistics and 4 chemistry problems, how long should you expect to spend?

E[5 × S + 4 × C]

= 5 × E[S] + 4 × E[C]

Linear Combination Example

The standard deviation of time spent is

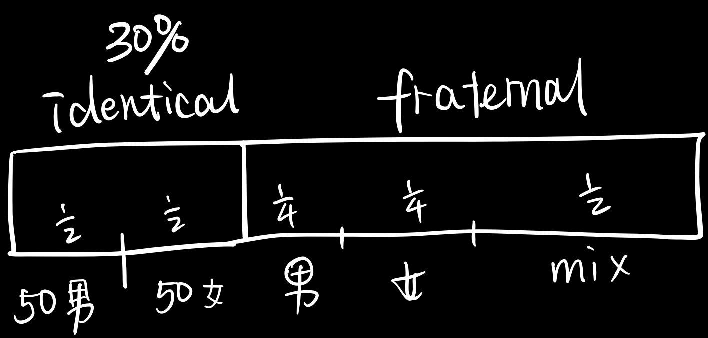
- 2 minutes on each statistics problem
- 5 minutes on each chemistry problem

If you're assigned 5 statistics and 4 chemistry problems, what is the s.d. of the assignment?

$$Var(X) = \underbrace{5 \times Var(S)}_{\text{5 copies of Var(S)}} + 4 \times Var(C) = 5 \times 4 + 4 \times 25 = 120$$

$$StdDev(X) = \sqrt{120} \approx 10.95$$

Why 5 × Var(X) instead of 25 × Var(X)?



both female

$$P(A|B) = \frac{P(A\&B)}{P(B)} = \frac{0.15}{0.15 + 0.7 * 0.25} =$$

identical

$$\frac{P(B_2 \& B_1)}{P(B_1)} = \frac{\frac{4}{12} * \frac{3}{11}}{\frac{4}{12}} = \frac{\frac{1}{11}}{\frac{1}{3}} = \frac{3}{11}$$

$$P(B_1) \times P(B_2|B_1)$$

$$= \frac{4}{12} \times \frac{3}{11}$$

$$= \frac{1}{11}$$

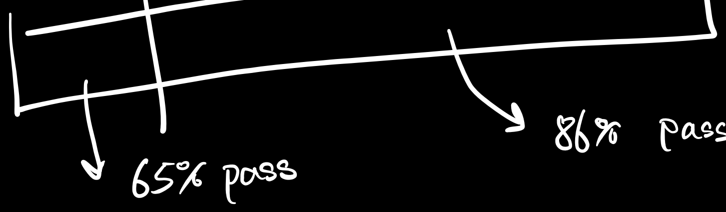
can

pass

P(A|B)

20%

80% (can)



$$\frac{16}{100} = \frac{4}{25}$$

$$\frac{2}{15}$$

$$P(A) = 0.8$$

$$P(B) = 0.65 * 0.2 + 0.86 * 08$$
$$= 0.13 + 0.688$$
$$= 0.818$$

$$P(A|B) = \frac{P(A \& B)}{P(B)}$$
$$= \frac{0.8 * 0.86}{0.818}$$
$$= 0.84$$