# (Conditional) Independence

Events $A$ and $B$ are said to be Conditionally Independent given information $C$ if

$$P(A, B|C) = P(A|C)P(B|C)$$

Conditional independence of $A$ and $B$ given $C$ is often expressed as

$$A \perp B|C$$

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero

## Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
    - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$
- A **continuous** random variable's support is a continuum (e.g., [0,1])

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$
- A **continuous** random variable's support is a continuum (e.g., [0,1])
  - Each possible value has a probability **density** $p(X = x)$ (or just $p(x)$ for short)

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$
- A **continuous** random variable's support is a continuum (e.g., [0,1])
  - Each possible value has a probability **density** $p(X = x)$ (or just $p(x)$ for short)
  - The probability *mass* of any value on the continuum is zero, regardless of the density!

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$
- A **continuous** random variable's support is a continuum (e.g., [0,1])
  - Each possible value has a probability **density** $p(X = x)$ (or just $p(x)$ for short)
  - The probability *mass* of any value on the continuum is zero, regardless of the density!
  - Properness is characterized in terms of an integral: $\int_x P(X = x)\,\mathrm{d}x = 1$ (note the derivative!)

## Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
  - Each possible value has a probability **mass** $P(X = x)$ (or just $P(x)$ for short)
  - Properness is characterized in terms of a sum: $\sum_x P(X = x) = 1$
- A **continuous** random variable's support is a continuum (e.g., [0,1])
  - Each possible value has a probability **density** $p(X = x)$ (or just $p(x)$ for short)
  - The probability *mass* of any value on the continuum is zero, regardless of the density!
  - Properness is characterized in terms of an integral: $\int_x P(X = x)\,\mathrm{d}x = 1$ (note the derivative!)
  - Remember that probability densities have units (the inverse of the unit of the continuum), and the densities can exceed 1 per unit!

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
- A **continuous** random variable's support is a continuum (e.g., [0,1])

- (We are presently eliding over cases where a random variable can have both mass and density on different sets of values)

# Discrete vs continuous random variables

- The **support** of a random variable is the set of values for which its probability is non-zero
- A **discrete** random variable's support is a finite or countably infinite number of values
- A **continuous** random variable's support is a continuum (e.g., [0,1])

- (We are presently eliding over cases where a random variable can have both mass and density on different sets of values)
- Unless I mention otherwise, things I say will hold for both discrete and continuous random variables, and I will freely use sums or integrals with the implicit understanding that what I say applies to both cases

# Mean, variance, and standard deviation

- (Population) **mean**, or **expected value**:

$$E[X] = \sum_x x\, P(X = x) \qquad \text{(discrete)}$$

$$E[X] = \int_x x\, p(X = x)\, dx \qquad \text{(continuous)}$$

# Mean, variance, and standard deviation

- (Population) **mean**, or **expected value**:

$$E[X] = \sum_x x\, P(X = x) \qquad \text{(discrete)}$$

$$E[X] = \int_x x\, p(X = x)\, \mathrm{d}x \qquad \text{(continuous)}$$

- (Population) **variance**:

$$\mathrm{Var}[X] = \sum_x (x - E[X])^2\, P(X = x) \qquad \text{(discrete)}$$

$$\mathrm{Var}[X] = \int_x (x - E[X])^2\, p(X = x)\, \mathrm{d}x \qquad \text{(continuous)}$$

# Mean, variance, and standard deviation

- (Population) **mean**, or **expected value**:

$$E[X] = \sum_x x\, P(X = x) \qquad \text{(discrete)}$$

$$E[X] = \int_x x\, p(X = x)\, \mathrm{d}x \qquad \text{(continuous)}$$

- (Population) **variance**:

$$\mathrm{Var}[X] = \sum_x (x - E[X])^2\, P(X = x) \qquad \text{(discrete)}$$

$$\mathrm{Var}[X] = \int_x (x - E[X])^2\, p(X = x)\, \mathrm{d}x \qquad \text{(continuous)}$$

- (Population) **standard deviation**, often notated $\sigma$, is just the (positive) square root of the variance:

$$(\sigma_X =)\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}$$

# Linearity of the expectation

- A crucial property we will often use is linearity of the expectation: variables is the sum of the expectation of random variables:

$$E[X_1 + X_2 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i] \quad \text{(for random variables } \{X_i\}\text{)}$$

$$E[cX] = cE[X] \quad \text{(for any constant } c\text{)}$$

# Linearity of the expectation

- A crucial property we will often use is linearity of the expectation: variables is the sum of the expectation of random variables:

$$E[X_1 + X_2 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i] \quad \text{(for random variables } \{X_i\}\text{)}$$
$$E[cX] = cE[X] \quad \text{(for any constant } c\text{)}$$

- Note that this requires no further assumptions (e.g., the random variables don't need to be independent)

# Covariance and correlation

- The **covariance** between two random variables is how much they vary together:

$$\mathrm{Cov}(X, Y) = \int_{x,y} (x - E[X])(y - E[Y]) \mathrm{d}x \mathrm{d}y$$

# Covariance and correlation

- The **covariance** between two random variables is how much they vary together:

$$\mathrm{Cov}(X, Y) = \int_{x,y} (x - E[X])(y - E[Y])\mathrm{d}x\mathrm{d}y$$

- The **correlation** between two random variables, sometimes notated $\rho_{XY}$ for random variables $X$ and $Y$, is their covariance "standardized" by their standard deviations:

$$(\rho_{XY} =)\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov(X,Y)}}{\sigma_X \sigma_Y}$$