Stat 501

Independence of Events and Random Variables, Applications in Modeling

Dr. Maher Boudabra

Course Outline

Lecture 1: Independence of Events

What is Independence?

Intuitive Definition

Two events are independent if the occurrence of one does not affect the probability of the other occurring.

Mathematical Definition

Events A and B are independent if and only if:

$$P(A \cap B) = P(A) \cdot P(B)$$

Example

Tossing a fair coin twice:

$A =$ first toss is heads, $B =$ second toss is heads

$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A)P(B)$

Conditional Probability and Independence

Recall Conditional Probability (Week 2)

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{for } P(B) > 0$$

Theorem

Events A and B are independent if and only if:

$$P(A|B) = P(A) \quad \textit{or} \quad P(B|A) = P(B)$$

Example

Drawing cards from a deck:

$A =$ first card is ace, $B =$ second card is ace

With replacement: $P(B|A) = P(B)$ (independent)

Without replacement: $P(B|A) \neq P(B)$ (dependent)

Independence vs. Mutual Exclusivity

Mutually Exclusive Events

- $A \cap B = \emptyset$
- $P(A \cap B) = 0$
- Cannot occur together
- Visually disjoint

Independent Events

- $P(A \cap B) = P(A)P(B)$
- Can occur together
- No relationship between occurrences
- Visually may overlap

Important Distinction

Mutually exclusive events are **never** independent (unless one has probability 0)!

$$\text{If } P(A) > 0 \text{ and } P(B) > 0, \text{ then } P(A \cap B) = 0 \neq P(A)P(B) > 0$$

Pairwise Independence

Definition

Events $A_1, A_2, \ldots, A_n$ are pairwise independent if:

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{for all } i \neq j$$

Example

Consider a fair coin tossed twice:

$A =$ first toss is H, $B =$ second toss is H, $C =$ both tosses are same

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

$$P(A \cap C) = \frac{1}{4} = P(A)P(C)$$

$$P(B \cap C) = \frac{1}{4} = P(B)P(C)$$

So A, B, and C are pairwise independent.

Mutual Independence

Definition

Events $A_1, A_2, \ldots, A_n$ are mutually independent if for every subset $I \subseteq \{1, 2, \ldots, n\}$:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Example (Continued)

For A, B, and C from previous slide:

$$P(A \cap B \cap C) = P(\text{HH}) = \frac{1}{4}$$

But:

$$P(A)P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

So A, B, and C are pairwise but not mutually independent.

Testing for Mutual Independence

Procedure

To verify mutual independence of n events, we must check:

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n - n - 1$$

conditions (all possible intersections).

Example

For 3 events A, B, C, we need to check:

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(B \cap C) = P(B)P(C)$$
$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Common Mistake

Pairwise independence does NOT imply mutual independence!

Conditional Independence

Definition

Events A and B are conditionally independent given event C if:

$$P(A \cap B|C) = P(A|C) \cdot P(B|C)$$

or equivalently (when $P(B \cap C) > 0$):

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

Example

Weather and car accidents:
$A =$ car accident occurs, $B =$ rainy weather
Given $C =$ icy roads, A and B may be independent
(The roads being icy explains both the accident and the need for rain)

Properties of Conditional Independence

Key Properties

1. Conditional independence does not imply unconditional independence
2. Unconditional independence does not imply conditional independence
3. Independence can be created or destroyed by conditioning

Example (Simpson's Paradox)

A medical treatment may appear:

- Harmful when looking at overall data
- Beneficial when conditioning on patient severity
- The conditioning variable (severity) affects both treatment assignment and outcome

Lecture 2: Random Variables and Their Independence

What is a Random Variable?

Definition

A random variable is a function that assigns a numerical value to each outcome in a sample space.

$$X : \Omega \to \mathbb{R}$$

Types of Random Variables

- **Discrete**: Takes on countable values (e.g., number of heads in coin tosses)
- **Continuous**: Takes on values in a continuum (e.g., height, weight)
- **Mixed**: Has both discrete and continuous components

Example

- Let X be the number of heads in 3 coin tosses (discrete)
- Let Y be the time until a radioactive atom decays (continuous)
- Let Z be the waiting time in a queue with probability p of no wait (mixed)

Probability Distributions

Probability Mass Function (PMF)

For discrete random variables:

$$p_X(x) = P(X = x)$$

Properties: $p_X(x) \geq 0$ and $\sum_x p_X(x) = 1$

Cumulative Distribution Function (CDF)

For any random variable:

$$F_X(x) = P(X \leq x)$$

Properties: Non-decreasing, right-continuous, $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$

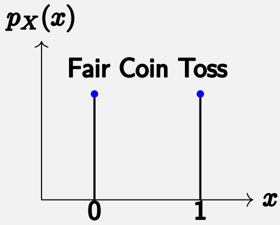
Example

Fair coin toss (X = 1 for heads, 0 for tails):

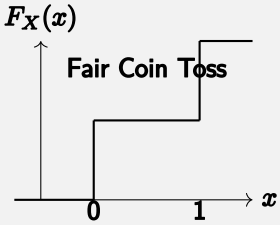
$$p_X(0) = 0.5, \quad p_X(1) = 0.5$$

Visualizing Probability Distributions

Discrete Distribution (PMF)



Cumulative Distribution (CDF)



Probability Density Function (PDF)

For continuous random variables:

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Properties: $f_X(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x) dx = 1$

Independent Random Variables

Definition

Random variables X and Y are independent if for all x, y:

$$P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y)$$

or equivalently:

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$$

Discrete Case

For discrete random variables:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

Continuous Case

For continuous random variables:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

Testing for Independence of Random Variables

Methods to Check Independence

1. Check if joint CDF factors: $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
2. Check if joint PMF/PDF factors: $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ or $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
3. Check if conditional distribution equals marginal distribution
4. Check if events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all x, y

Example

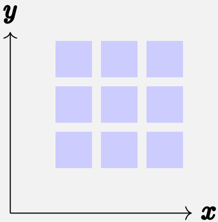
Consider two random variables with joint distribution:

X	Y = 0	Y = 1	Y = 2
0	0.10	0.20	0.10
1	0.15	0.30	0.15

Check if $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all cells.

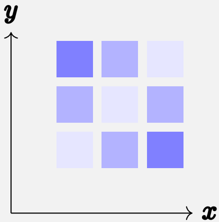
Visualizing Independence of Random Variables

Independent RVs



Uniform density

Dependent RVs



Non-uniform density

Key Insight

For independent random variables, the joint distribution is the product of marginals.

For dependent random variables, the joint distribution has a more complex structure.

Functions of Independent Random Variables

Theorem

If X and Y are independent random variables, and g and h are functions, then:

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$$

provided the expectations exist.

Special Cases

- $E[XY] = E[X]E[Y]$ (if X and Y are independent)
- $M_{X+Y}(t) = M_X(t)M_Y(t)$ (moment generating functions)
- The joint PDF/PMF factors: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

Important Note

$E[XY] = E[X]E[Y]$ does NOT imply independence!
This is called uncorrelatedness, which is weaker than independence.

Conditional Independence for Random Variables

Definition

Random variables X and Y are conditionally independent given Z if:

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z) \cdot P(Y \leq y | Z = z)$$

for all x, y, z (or almost all z in the continuous case).

Example

Test scores and study time:

$X =$ test score, $Y =$ hours studied, $Z =$ student ability

Given student ability, test score and study time may be independent

(Ability affects both how much one studies and how well one performs)

Markov Property

A Markov process has the property that the future is conditionally independent of the past given the present.

Bernoulli Trials

Definition

A sequence of experiments is called Bernoulli trials if:

1. Each trial has two possible outcomes (success/failure)
2. The probability of success remains constant (p)
3. The trials are independent

Example

Tossing a fair coin 10 times:

- Each toss is independent
- Probability of heads remains 0.5
- The number of heads follows a binomial distribution

Binomial Distribution

If X is the number of successes in n Bernoulli trials:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

IID Random Variables

Definition

Random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) if:

1. They are mutually independent
2. They all have the same probability distribution

Importance in Statistics

- Many statistical methods assume i.i.d. data
- Simplifies analysis and derivation of properties
- Foundation for laws of large numbers and central limit theorem

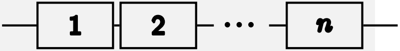
Example

Measuring the height of randomly selected individuals:
$X_1, X_2, \ldots, X_n$ are i.i.d. random variables
(Assuming the population distribution doesn't change and individuals are selected independently)

Lecture 3: Applications in Modeling

Reliability Systems

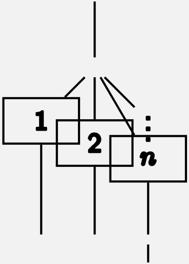
Series System



$$R_{\text{series}} = \prod_{i=1}^{n} R_i$$

System fails if any component fails

Parallel System



$$R_{\text{parallel}} = 1 - \prod_{i=1}^{n} (1 - R_i)$$

System fails only if all components fail

k-out-of-n Systems

Definition

A system with n components works if at least k components work.

Example

An airplane with 4 engines can fly if at least 2 engines work (2-out-of-4 system).

Reliability Calculation

If components are independent with reliability p:

$$R = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Special Cases

- $k = 1$: Parallel system
- $k = n$: Series system
- $k = \lceil n/2 \rceil$: Majority voting system

Acceptance Sampling

Concept

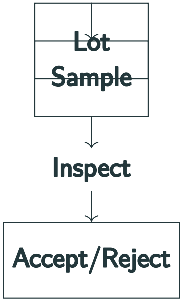
- Inspect a sample from a lot
- Decide whether to accept or reject the entire lot
- Based on number of defective items found

Probability of Acceptance

If defect rate is p and we sample n items independently:

$$P(\text{accept}) = \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}$$

where c is the acceptance number (maximum allowed defects)



Operating Characteristic (OC) Curve

Definition

The OC curve shows the probability of accepting a lot as a function of the defect rate p.

Properties

- Decreasing function of p
- Steepness depends on sample size n
- Acceptance number c affects the shape

Designing Sampling Plans

Choose n and c to achieve:

- High probability of accepting good lots
- Low probability of accepting bad lots

Medical Diagnostic Tests

Measures of Test Performance

- **Sensitivity**: $P(\text{test}+|\text{disease}) = \frac{\text{TP}}{\text{TP+FN}}$
- **Specificity**: $P(\text{test}-|\text{no disease}) = \frac{\text{TN}}{\text{TN+FP}}$
- **Positive Predictive Value**: $P(\text{disease}|\text{test}+)$
- **Negative Predictive Value**: $P(\text{no disease}|\text{test}-)$

Bayes' Theorem Application

$$P(\text{disease}|\text{test}+) = \frac{P(\text{test}+|\text{disease})P(\text{disease})}{P(\text{test}+)}$$

where

$$P(\text{test}+) = P(\text{test}+|\text{disease})P(\text{disease}) + P(\text{test}+|\text{no disease})P(\text{no disease})$$

Multiple Tests and Conditional Independence

Independent Tests

If tests are independent given disease status:

$$P(\text{test1+}, \text{test2} + |\text{disease}) = P(\text{test1} + |\text{disease}) \cdot P(\text{test2} + |\text{disease})$$

Example

Two independent tests with sensitivity 0.9 each:

$P(\text{both} + |\text{disease}) = 0.9 \times 0.9 = 0.81$

$P(\text{at least one} + |\text{disease}) = 1 - 0.1 \times 0.1 = 0.99$

Serial Testing

- Both tests must be positive for diagnosis
- Increases specificity, decreases overall sensitivity

Parallel Testing

- Either test positive leads to diagnosis
- Increases sensitivity, decreases overall specificity

Bayesian Networks

Definition

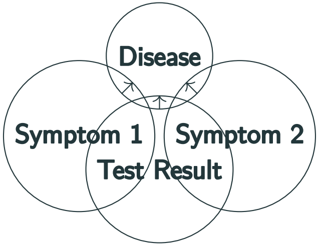
A Bayesian network is a directed acyclic graph representing conditional dependencies among random variables.

Key Property

Each variable is conditionally independent of its non-descendants given its parents.

Joint Distribution Factorization

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{parents}(X_i))$$



Applications of Bayesian Networks

Medical Diagnosis

- Symptoms are conditionally independent given disease
- Efficient computation of probabilities
- Update beliefs as new evidence arrives

Machine Learning

- Naive Bayes classifiers
- Hidden Markov models
- Probabilistic graphical models

Risk Assessment

- Financial risk modeling
- Engineering failure analysis
- Environmental risk assessment

Entropy and Information

- Entropy measures uncertainty in a random variable
- For independent events, information adds up

Joint Entropy of Independent Variables

If X and Y are independent:

$$H(X,Y) = H(X) + H(Y)$$

where H represents entropy.

Mutual Information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

For independent X and Y: $I(X;Y) = 0$

Example

If weather and stock market are independent:

Uncertainty about both = uncertainty about weather + uncertainty

Summary and Conclusion

Summary of Key Concepts

Independence of Events

- $P(A \cap B) = P(A)P(B)$
- $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- Pairwise vs. mutual independence
- Conditional independence

Independence of Random Variables

- $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
- $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ (continuous case)
- $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ (discrete case)
- IID assumption in statistics

Summary of Applications

Modeling Applications

- Reliability systems (series, parallel, k-out-of-n)
- Quality control (acceptance sampling)
- Medical testing (multiple tests, conditional independence)
- Bayesian networks (graphical models)
- Information theory (entropy, mutual information)

Importance in Statistics

- Foundation for statistical inference
- Simplifies complex probability calculations
- Enables powerful theorems (LLN, CLT)
- Basis for many statistical models and methods

Next Steps in STAT501

Week 4: Random Variables

- Discrete and continuous random variables
- Probability mass functions (PMF)
- Probability density functions (PDF)
- Cumulative distribution functions (CDF)

Important Connection

Understanding independence is crucial for:

- Joint distributions of multiple random variables
- Sums of independent random variables
- The binomial distribution (sum of independent Bernoulli trials)