Probability

Basic Concepts

- Probability: quantifies the likelihood that an event occurs ranging from 0 to 1 (impossible to certain)
- Sample space (S): set of all possible outcomes
- Event (A): subset of the sample space containing outcomes of focus
- Types of events:
  - Simple Events: Events with a single outcome (e.g., rolling a 3 on a die).
    - $E = \{s_i\} \quad \text{where} \quad s_i \in S$
  - Compound Events: Events that combine multiple outcomes (e.g., rolling an even number).
    - $P(A^c) = 1 - P(A)$
  - Independent Events: The occurrence of one event does not affect the other (e.g., flipping a coin and rolling a die).
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Dependent Events: The occurrence of one event affects the probability of the other (e.g., drawing cards from a deck without replacement).
    - $P(A \cap B) = P(A) \cdot P(B|A) \quad \text{or} \quad P(A \cap B) = P(B) \cdot P(A|B)$
  - Mutually Exclusive Events: Events that cannot occur simultaneously (e.g., rolling a 2 or a 3 on a die).
    - $P(A \cup B) = P(A) + P(B)$

Axiom of Probability

Kolmogorov's axioms of probability

1. Non-negativity: For any event A, the probability P(A) is non-negative:
   a. P(A) ≥ 0
   b. Ensures meaningful interpretations in context
2. Normalization: The probability of the entire sample space S is 1:
   a. P(S) = 1
   b. One out of all the possible outcomes must occur

3. Additivity: For any two mutually exclusive events, A and B, the probability of their union is the sum of their probabilities
    a. $P(A \cup B)=P(A)+P(B)$ if $A \cap B=\emptyset$
    b. Foundational concept for probability spaces with random variables

Properties of Probability

- Complement rule: the probability of the complement of an event A (A does not occur) is equal to 1 minus the probability of A
    - $P(A') = 1 - P(A)$
    - From axiom 1 & 2
- Additive rule (union of 2 events): incorporates intersection of events A and B to avoid double counting into the formula for probability of the union of 2 events
    - $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
    - From axiom 3
- Law of Total Probability (LTP): calculates the probability A given all the possible ways B that A can occur
    - $$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$
- Multiplicative rule (independence of events): if the occurrence of events A and B have no affect on each other, their joint probability is their product
    - $P(A \cap B)=P(A) \cdot P(B)$
- Conditional probability: defines the probability of an event A occurring given an event B occurred
    - $$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
    - $P(B) > 0$
    - Bayes theorem: $$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$
- Monotonicity: the probability of an event in a subset must be less than to equal to the probability of the superset
    - $P(A) \leq P(B)$ given $A \subseteq B$

Random Variables

- Random variable (X): assigns a numerical value to each outcome in a sample space
- Discrete random variable: random variable with a countable number of values
- Continuous random variable: random variable with an uncountable number of values
- Probability mass function (PMF): for discrete variables, describes the likelihood that the variable is equal to a particular value
  - $p(x)=P(X=x)$
- Probability density function (PDF): for continuous variables, describes the likelihood that the variable falls within a particular range
- Cumulative distribution function (CDF): the probability that a random variable takes on a value less than or equal to x
  - $F(x)=P(X\leq x)$

Expected Value

- Expected value: the measure of the center of the distribution for a random variable X
  - Discrete: $E(X) = \sum x \cdot p(x)$
  - Continuous: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$

Variance

- Variance: measures the spread of the random variable around the mean
  - $Var(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2$
- Standard deviation: how far spread out the set of data is in relation to the mean
  - $\sigma = \sqrt{Var(X)}$

Common Distributions

Discrete distributions: Binomial, Poisson, Geometric, Negative Binomial

- Binomial Distribution: Models the number of successes in n independent Bernoulli trials (2 possible outcomes with same probability every tiem)
    - $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Poisson Distribution: Models the number of events in a fixed interval of time/space
    - $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
    - Lambda = average rate
- Geometric Distribution: Models the number of trials until the first success
    - $P(X = k) = (1-p)^{k-1} p$

Continuous distributions: Normal, Exponential, Uniform, Beta, Gamma

- Normal Distribution: Characterized by its bell-shaped curve, described by its mean μ and standard deviation σ
    - PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Exponential Distribution: Models the time between events in a Poisson process
    - PDF for x ≥ 0: $f(x) = \lambda e^{-\lambda x}$
- Uniform Distribution: All outcomes are equally likely within a specified range

Joint, Marginal, and Conditional Distributions

- Joint Distribution: Describes the probability of two or more random variables occurring simultaneously
    - Can be represented in a joint probability table for discrete variables
- Marginal Distribution: The probability distribution of a subset of random variables obtained by summing/integrating out the other variables
    - Discrete: $P_X(x) = \sum_y P(X = x, Y = y)$

- Continuous: $$f_X(x) = \int f_{X,Y}(x, y)\, dy$$
- Conditional Distributions: The distribution of one variable given another
  - P(X|Y)

Moment Generating Functions

Definition and properties

Use in finding moments and distributions

- Moment generating function (MGF): calculates the expected value of moments to understand behavior
  - Discrete: $$M_X(t) = \mathbb{E}[e^{tX}] = \sum_x e^{tx} P(X = x)$$
  - Continuous: $$M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$
- Used to prove central limit theorem by analyzing convergence in distribution
- Used to identify the distribution of a random variable
- Used to understand the behavior of random variables

Law of Large Numbers

Weak and strong law of large numbers

Implications for sample averages

- Weak law: the sample mean converges in probability to the expected value as the sample size increases for any ϵ>0
  - $P(|\bar{X}_n - \mu| \geq \epsilon) \to 0 \text{ as } n \to \infty$
- Strong law: the sample mean converges almost surely to the expected value
  - $$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

- Intuitively: the larger the sample size, the more accurate the sample mean is to the true population mean
- Conditions:
    - Random variables must be independent of each other
    - Random variables must be identically distributed
    - For strong: random variables should have finite mean and variance

Central Limit Theorem

- Definition: the distribution of the sum (or average) of a large number of independent, identically distributed random variables approaches a normal distribution, regardless of the original distribution's shape
    - Can apply normal distribution techniques if n is sufficiently large
    - $$\lim_{n\to\infty} P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$
- Limitations:
    - Sample size must be sufficiently large; generally > 30
    - Population must have finite variance
    - Samples must be independent of each other

Statistical Inference

- Point Estimation: providing a single value estimate of a population parameter
    - Ex: sample mean as population mean
- Interval Estimation: providing a range of values within which the parameter is expected to lie
    - confidence intervals
- Hypothesis Testing: procedure for testing claims about a population parameter based on sample data
    - Alternative hypothesis attempts to reject null hypothesis with predefines confidence
    - Test statistics and p-values determine significance

Bayes Statistics

- Bayes Theorem: determines the conditional probability of an event
  - $$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$
  - Prior probability P(H): believed probability of the hypothesis before evidence
  - Marginal likelihood P(E): total probability of the evidence (sum of hypotheses)
    - $$P(E) = \sum_H P(E|H)P(H)$$
  - Likelihood P(E|H): probability of observing the evidence given that the hypothesis is true
  - Posterior probability P(H|E): probability of the hypothesis given evidence
- Parameters are treated as random variables with a probability distribution that is updated by Bayes theorem with additional data collection
- Advantages:
  - Include historical data
  - Produce full probability distributions (posterior distributions)
  - Flexibility
  - Predictive power
  - Decision theory – help make decisions under uncertainty
- Challenges:
  - Choosing which to use as prior – impacts bias
  - Computationally complex
  - Susceptible to overfitting

Markov Process

Definition and examples

Markov chains and their properties

Applications in various fields

- Markov chain: sequence of random variables where the probability of the next state depends only on the current state
- Monte Carlo Sampling: the use of random sampling to estimate numerical results

R packages

- Dplr: data summarization for empirical probabilities, conditional probabilities, expected value, simulation & aggregation, grouping
    - mutate() adds new variables that are functions of existing variables
    - select() picks variables based on their names
    - filter() picks cases based on their values
    - summarise() reduces multiple values down to a single summary
    - arrange() changes the ordering of the rows
- stats
    - probability distributions
    - statistical inference functions
    - random sampling & simulations
    - descriptive statistics & summary