Contents

Arrangements:

- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ takes k items from n disregarding order
- $n!$ finds the arrangements of length n using each item in n once
- $n^{(k)} = \frac{n!}{(n-k)!}$ finds the arrangements of length k using each item from n at most once
- n^k finds the arrangements of length k using the items from n as often as wanted

INTRODUCTION TO PROBABILITY

1.1 Definitions of Probability

Def Randomness: Caused by (1) variability in population and (2) variability in processes

Def Sample Space (S): All distinct possible outcomes to a random experiment
Note: Continues in 2.1

Def Probability: Can be defined in 3 ways:

1. **Def** Classical: Provided all points in S are equal,

$$\frac{\text{number of ways the event can occur}}{\text{number of outcomes in } S}$$

   Note: Continues in 2.1

2. **Def** Relative Frequency: The portion of times an event has happened after repetitions of an experiment.

3. **Def** Subjective Probability: How sure an individual is that an event will happen.

Def Probability Model:

- The sample space is defined
- A set of events (subset of S) is defined
- A mechanism for assigning probabilities is defined

MATHEMATICAL PROBABILITY MODELS

2.1 Sample Spaces and Probability

Def Sample Space (S) Continued: All distinct possible outcomes to a random experiment

- In a single trial, one and only one outcome can occur
- The sample space does not need to be uniquely defined ($S = \{1, 2, 3, 4, 5, 6\}$ or $S = \{\text{Even}, \text{Odd}\}$)
- May be discrete ($S = \{1, 2, 3 \ldots\}$ or $S = \left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8} \ldots\right\}$) or non-discrete ($S = \{x : x > 0\}$)

Def Simple Event: The subset of the event $A \in S$ contains only a single point where S is discrete ($A = \{a_1\}$)

Def Compound Event: The subset of the event $A \in S$ contains two or more simple events ($A = \{a_1, a_2 \ldots\}$)

Def Probability Distribution : The probability distribution on S is the set of probabilities $\{P(a_i), i = 1, 2, \ldots\}$ where the following conditions hold:

- $0 \leq P(a_i) \leq 1$
- $\sum_{\text{all} i} P(a_i) = 1$

Def Probability: The probability of an event A occurring is

$$P(A) = \sum_{a \in A} P(a)$$

Def Odds: The odds of an event A occurring is

$$\frac{P(A)}{1 - P(A)}$$

Note: The odds against the event is the reciprocal

PROBABILITY AND COUNTING TECHNIQUES

3.1 Addition and Multiplication Rules

Def Uniform Probability Model: A set where each simple event has probability $\frac{1}{n}$

Def Addition Rule: Suppose we can do job 1 in p ways and job 2 in q ways. Then we can do either job 1 **OR** job 2 (but not both), in $p+q$ ways

Def Multiplication Rule: Suppose we can do job 1 in p ways and, for each of these ways, we can do job 2 in q ways. Then we can do both job 1 **AND** job 2 in $p \times q$ ways

3.2 Counting Arrangements or Permutations

Def Permutations: A sample space which is a set of arrangements or sequences

Def n to k factors: A product is said to have n to k factors if

$$n^{(k)} = \frac{n!}{(n-k)!}$$

Def Stirling's Approximation: An approximation to $n!$ for large n values

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

Note: For $n \geq 8$, the error is less than 0.01

Def Complement: The complement of A, denoted $\overline{A}$, is the set of all outcomes in S that are not in A

3.3 Counting Subsets or Combinations

Def Combinatorial: "n choose k" is sued to denote the number of subsets (with no order) of size k that can be selected from the set of n objects

$$\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!}$$

Note: Properties of $\binom{n}{k}$ are as follows

- $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
- $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!}$
- $\binom{n}{k} = \binom{n}{n-k}$ for all $k = 0, 1, \ldots n$
- If we define $0! = 1$, then the formulas hold with $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = \binom{n-1}{n-k} + \binom{n-1}{k}$
- **Binomial Theorem:** $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$

PROBABILITY RULES AND CONDITIONAL PROBABILITY

4.1 General Methods

Proved Set-Theoretic Rules:

1. $P(S) = 1$
2. For an event A, $0 \leq P(A) \leq 1$
3. For $A \subseteq B$, $P(A) \leq P(B)$

Def Union: If A or B occur (inclusive), the event occurred

$$A \cup B$$

Def Intersection: If A and B occur, the event occurred

$$A \cap B$$

Note: Often shortened to AB

Def Complement: If A did not occur, the event occurred

$$\overline{A}$$

Note: $\overline{S} = \emptyset$

Proved De Morgan's Laws:

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$

4.2 Rules for Unions of Events

Proved Addition Law of Probability or the Sum Rule:

4. **A** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proved Probability of the Union of Three Event:

4. **B** $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Proved Probability of the Union of n Events:

4. **C** $P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_i P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) - \dots$

Def Mutually Exclusive: Events A and B are mutually exclusive if

$$A \cap B = \emptyset$$

Proved Probability of the Union of Two Mutually Exclusive Events: Let A, B be mutually exclusive, then

5. **A** $P(A \cup B) = P(A) + P(B)$

Proved Probability of the Union of n Mutually Exclusive Events: Let $A_a, A_2, \dots, A_n$ be mutually exclusive, then

5. **B** $P(A \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$

Proved Probability of the Complement of an Event:

6. $P(A) = 1 - P(\overline{A})$

4.3 Intersections of Events and Independence

Def Independent Events: Events are independent if and only if

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

Def Mutually Independent: Events $A_1, A_2, \dots A_n$ are mutually independent if and only if

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2)\dots P(A_n)$$

for all sets $\{i_1, i_2, \dots i_k\}$ of distinct subscripts chosen from $(1, 2, \dots, n)$

Note: Often referred to as "independent"

4.4 Conditional Probability

Def Conditional Events: If an event B occurred, the probability that A occurs is the conditional probability of A given B

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

Note: If A and B are independent, then $P(A \mid B) = P(A)$

Theorem 10: The events A and B are independent if and only if

$$P(A \mid B) = P(A) \text{ or } P(B \mid A) = P(B)$$

4.5 Product Rules, Law of Total Probability and Bayes' Theorem

Proved Product Rules: Let $P(A) > 0, P(A \cap B) > 0, P(A \cap B \cap C) > 0$

7.
- $P(AB) = P(A)P(B \mid A)$
- $P(ABC) = P(A)P(B \mid A)P(C \mid AB)$
- $P(ABCD) = P(A)P(B \mid A)P(C \mid AB)P(D \mid ABC)$

Proved Law of Total Probability: Let $A_1, A_2, \dots A_k$ be a partition of the sample space into mutually exclusive (disjoint) events. Let B be an event in S. Then

$$P(B) = P(BA_1) + P(BA_2) + \dots + P(BA_k) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$

Proved Bayes' Theorem: Let $P(B) > 0$. Then

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid \overline{A})P(\overline{A}) + P(B \mid A)P(A)}$$

4.6 Useful Series and Sums

Geometric Series

$$\text{if } t \neq 1, \text{ then } \sum_{i=0}^{n-1} t^i = 1 + t + t^2 + \dots t^{n-1} = \frac{1-t^n}{1-t}$$

$$\text{if } |t| < 1, \text{ then } \sum_{x=0}^{\infty} t^x = 1 + t + t^2 + \cdots = \frac{1}{1-t}$$

and thus with higher derivatives,

$$\text{if } |t| < 1, \text{ then } \sum_{x=0}^{\infty} x t^{x-1} = \frac{1}{(1-t)^2}$$

Binomial Theorem

$$\text{if } n \in \mathbb{N} \text{ and } t \in \mathbb{R}, \text{ then } (1+t)^n = 1 + \binom{n}{1} t + \binom{n}{2} t^2 + \cdots + \binom{n}{n} t^n = \sum_{x=0}^{n} \binom{n}{x} t^x$$

$$\text{if } n \notin \mathbb{N} \text{ and } |t| < 1, \text{ then } (1+t)^n = \sum_{x=0}^{\infty} \binom{n}{x} t^x$$

Multinomial Theorem

$$\text{if } n \in \mathbb{N}, \text{ then } (t_1 + t_2 + \cdots + t_k)^n = \sum \frac{n!}{x_1! x_2! \dots x_k!} t_1^{x_1} t_2^{x_2} \dots t_k^{x_k}$$

where the summation is over all non-negative integers such that $x_1 + x_2 + \cdots + x_k = n$

Note: See page 62

Hypergeometric Identity

$$\sum_{x=0}^{\infty} \binom{a}{x} \binom{b}{n-x} = \binom{a+b}{n}$$

Exponential Series Let $f(x) = e^x$ and $f^{(k)}(0) = 1$ for $k = 1, 2, \dots$

$$\text{if } t \in \mathbb{R}, \text{ then } e^t = \frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}$$

Special Integer Series

- $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$
- $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$
- $1^3 + 2^3 + 3^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$

DISCRETE RANDOM VARIABLES

5.1 Random Variables and Probability Functions

Def Random Variable: A function that assigns areal number to each point in a sample space S

$$X = x_1, x_2, x_3, \ldots$$

Note: The full sample space must be a union of the events of each element in X

Def Discrete Random Variable: Takes values in a countable set

Def Continuous Random Variable: Takes values in an interval, not countable

Def Probability Function: Let X be a discrete random variable with range$(X) = A$

$$f(x) = P(X = x)$$

1. $f(x) \geq 0, \forall x \in A$
2. $\sum_{x \in A} f(x) = 1$

Note: Make sure to state the domain of the function

Def Probability Distribution: The set of pairs

$$\{(x, f(x)) : x \in A\}$$

Def Cumulative Distribution Function: The sum of all previous probability functions

$$F(x) = \sum_{u \leq x} f(u) = P(X \leq x) \text{ for all } x \in \mathbb{R}$$

1. $F(x)$ is non-decreasing
2. $0 \leq F(x) \leq 1$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

5.2 Discrete Uniform Distribution

Physical Setup For a range of X of $\{a, \ldots, b\}$, where each integer is equally probable

Note: With replacement

Note: Parameters a, and b

Probability Function For $b - a + 1$ values in the set, each has $\frac{1}{b-a+1}$, thus

$$f(x) = P(X = x) = \begin{cases} \frac{1}{b-a+1} & \text{for } x \in \{a, \ldots, b\} \\ 0 \end{cases}$$

5.3 Hypergeometric Distribution

Physical Setup A collection of N objects with r of S and $N - r$ of F. X is the number of successes obtained

Note: Without replacement

Note: Parameters r, N, and n

Probability Function For $x \geq \max(0, n - N + r)$ and $x \leq \min(r, n)$

$$f(x) = P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

5.4 Binomial Distribution

Physical Setup A experiment with outcome $P(S) = p$ and $P(F) = 1 - p$ repeated for n independent times (Bernoulli Trials).

$$\sim Binomial(n, p)$$

Note: With replacement

Note: Parameters n, and p

Note: If $p = 0$ or $p = 1$, then X is said to be a degenerate random variable

Probability Function For $0 \leq x \leq n$ and $0 < p < 1$

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

5.5 Negative Binomial Distribution

Physical Setup A experiment with outcome $P(S) = p$ and $P(F) = 1 - p$ repeated for until S is obtained for the k^{th} time. X is the number of failures before the k^{th} success

$$\sim NegativeBinomial(k, p)$$

Note: With replacement
Note: Parameters k, and p

Probability Function For $0 \leq x$ and $0 < p < 1$

$$f(x) = P(X = x) = \binom{x + k - 1}{x} p^k (1 - p)^x$$

5.6 Geometric Distribution

Physical Setup The negative Binomial Distribution with $k = 1$

$$\sim Geometric(p)$$

5.7 Poisson Distribution from Binomial

Physical Setup Restrict the product $np = \mu$, then take the Binomial Distribution as $n \to \infty$ (thus $p \to 0$)

$$\sim Poisson(\mu)$$

Note: Used when n is large and p is small
Note: If n is large and p is large, switch "failure" vs "success" Note: If $\mu = 0$ then is said to be a degenerate distribution

Probability Function When $np = \mu$, For $x \geq 0$

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}$$

5.8 Poisson Distribution from Poisson Process

def Order Notation: $g(\Delta t) = o(\Delta t)$ as $\Delta t \to 0$ means g approaches 0 faster than Δt approaches 0

$$\frac{g(\Delta t)}{\Delta t} \to 0 \text{ as } \Delta t \to 0$$

Physical Setup A situation where certain events occur at random points of time and follow the Poisson Process

1. **Independence**: Occurrences in non-overlapping intervals are independent
2. **Individuality**: Events do not occur in clusters, that is

$$P(2 \text{ or more events in } (t, t+\Delta t)) = o(\Delta t) \text{ as } \Delta t \to 0$$

3. **Homogeneity/Uniformity**: The probability of one occurrence in an interval $(t, t+\Delta t)$ is $\lambda \Delta t$ for small Δt

Note: λ is the intensity or rate of occurrence parameter, thus λt is the average number of occurrences per t units of time

Note: If n is large and p is large, switch "failure" vs "success"

Probability Function Let $f_t(x)$ be the probability of x occurrences over the interval t. For $x \geq 0$

$$f_t(x) = f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

no thanks

EXPECTED VALUE AND VARIANCE

7.1 Summarizing Data on Random Variables

def Frequency Distribution: The number of times each value of X occurred

def Sample Mean: The average for a particular sample, the mean of n outcomes $x_1, \dots x_n$ for random variable X is

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

def Median: The value such that half of results are below and half the results are above when arranged in numerical order

def Mode: The value which occurs the most often.
Note: There is no guarantee of a single mode

7.2 Expectation of a Random Variable

def Expected Value: Let X be a discrete random variable with $range(X) = A$ and probability function $f(x)$, then

$$\mu = E(X) = \sum_{x \in A} x f(x)$$

Proved Theorem 17: Let X be a discrete random variable with $range(X) = A$ and probability function $F(x)$. Then the expected value of some $g(X)$ of X is

$$E[g(X)] = \sum_{x \in A} g(x) f(x)$$

Note: $E[g(X)]$ is the average value (expected value) of $g(X)$ in an infinite series of repetitions of the process where X is defined

Proved Linearity Properties of Expectation: For constants a, b

$$E[ag(X) + b] = aE[g(X)] + b$$

7.3 Means and Variances of Distributions

Proved Expected value of a Binomial random variable: Let $X \approx Binomial(n, p)$

$$E(X) = np$$

Proved Expected value of the Poisson random variable: Let X have a Poisson distribution

$$E(X) = \lambda t$$

Proved Expected value of the Hypergeometric random variable: Let X have a Hypergeometric distribution

$$E(X) = \frac{nr}{N}$$

Proved Expected value of the Negative Binomial random variable: Let X have a Negative Binomial distribution

$$E(X) = \frac{k(1-p)}{p}$$

def Variance: The average square distance from the mean, that is

$$\sigma^2 = Var(X) = E\left[(X-\mu)^2\right]$$

(1): $Var(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$
(2): $Var(X) = E[X(X-1)] + E(X) - [E(X)]^2 = E[X(X-1)] + \mu - \mu^2$

def Standard Deviation: The square root of the variance, that is

$$\sigma = sd(X) = \sqrt{Var(X)} = \sqrt{E\left[(X-\mu)^2\right]}$$

Proved Variance of a Binomial random variable: Let $X \approx Binomial(n, p)$

$$Var(X) = np(1-p)$$

Proved Variance of a Poisson random variable: Let X have a Poisson distribution

$$Var(X) = \mu$$

(2): The variance is equal to the mean

Proved If a, b are constants, $Y = aX + b$, and $\mu_X = E(X), \sigma_X^2 = Var(X), E(Y) = \mu_Y, Var(Y) = \sigma_Y^2$, then

$$\mu_Y = E(Y) = aE(X) + b = a\mu_X + b$$

and

$$\sigma_Y^2 = Var(Y) = a^2 Var(X) = a^2\sigma_X^2$$

CONTINUOUS RANDOM VARIABLES

8.1 Terminology and Notation

def Continuous Random Variables: Have a range of all possible values over an interval (or collection of intervals)

def Cumulative Distribution Function:

1. $F(x)$ is defined for all real x
2. $F(x)$ is non-decreasing over all real x
3. $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$
4. $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = F(b) - F(a)$

def Probability Density Function: The likely hood of small intervals around specific x values

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

1. $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)\mathrm{d}x$
2. $f(x) > 0$
3. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{\text{all } x} f(x)\mathrm{d}x = 1$
4. $F(x) = \int_{\infty}^{x} f(u)\mathrm{d}u$

def Quantiles and Percentiles: For a cumulative distribution function $F(x)$, the p^{th} quantile is the value $q(p)$ such that $P[X \leq q(p)] = p$
Note: $q(p)$ is the 100^{th} percentile of distribution
Note: $m = q(0.5)$ is the median of distribution

def Expected Value: For a continuous random variable,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(X)\mathrm{d}x$$

8.2 Continuous Uniform Distribution

Physical Setup Over an interval $[a, b]$, each subinterval of a fixed length is equally likely

$$\sim Uniform(a, b)$$

Note: Parameters $b > a$

Probability Density Function

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Mean

$$E(X) = \frac{b+a}{2}$$

Variance

$$Var(X) = \frac{(b-a)^2}{12}$$

Def Gamma Function: For $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \mathrm{d}y$$

1. For $\alpha > 0$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
2. For $\alpha \in \mathbb{N}$, $\Gamma(\alpha) = (\alpha - 1)!$
3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

8.3 Exponential Distribution

Physical Setup The time it takes it takes between occurrences of an event in the Poisson process

$$\sim Exponential(\theta)$$

Note: Parameters $\lambda > 0$ is the average rate of occurrence
Note: Parameters $\theta > 0$ is the waiting time until an occurrence

Probability Density Function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases} \text{ or } f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases} \text{ or } F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\frac{x}{\theta}} & x > 0 \end{cases}$$

Mean

$$E(X) = \frac{1}{\lambda} \text{ or } \theta$$

Variance

$$Var(X) = \theta^2$$

Def Memoryless Property: The probability you have to wait c unit of time does not depend on how long you have been waiting, that is

$$P(X > c + b \mid X > b) = P(X > c)$$

8.4 Computer Generation of Random Variables

Theorem 24 If F is an arbitrary cumulative distribution function and $U \sim Uniform(0, 1)$ then $X = F^{-1}(U)$ has cumulative distribution function $F(x)$

8.5 Normal Distribution

Physical Setup A "bell curve", where X is a physical dimension of some kind

$$X \sim N(\mu, \sigma^2)$$

Note: Parameters $x, \mu \in \mathbb{R}$
Note: Parameters $\sigma \in \mathbb{R}^+$

Gaussian Distribution Similar but with σ instead of σ^2

$$X \sim G(\mu, \sigma)$$

Probability Density Function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Cumulative Distribution Function

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}y$$

Mean

$$E(X) = \mu$$

Variance

$$Var(X) = \sigma^2$$

Def Standard Normal Distribution: A normal distribution with $\mu = 0$ and $\sigma = 1$

$$N(0, 1)$$

Theorem 25 Let $X \sim N(\mu, \sigma^2), Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0, 1)$ and

$$P(X \leq x) = P(Z \leq \frac{x-\mu}{\sigma})$$

MULTIVARIATE DISTRIBUTIONS

9.1 Basic Terminology and Techniques

Def Joint Probability Function: For discrete random variables X, Y, the probability both occur

$$f(x,y) \geq 0 \text{ and } \sum_{\text{all } (x,y)} f(x,y) = 1$$

Def Marginal Probability: For a joint probability function $f(x,y)$, the probability when interested in only one random variable

$$f_1(x) = \sum_{\text{all } y} f(x,y)$$

Def Independent Random Variables: For a joint probability function $f(x,y)$, being independent means that

$$f(x,y) = f_1(x)f_2(y)$$

or generalized to

$$f(x_1, x_2, \ldots x_n) = f_1(x_1)f_2(x_2)\ldots f_n(x_n)$$

Def Conditional Probability: For a joint probability function $f(x,y)$, if $f_2(y) > 0$ then the conditional probability function of X given Y is

$$f_1(x \mid y) = \frac{f(x,y)}{f_2(y)}$$

Theorem 29 If $X \sim Poisson(\mu_1)$ and $Y \sim Poisson(\mu_2)$ independently, then

$$T = X + Y \sim Poisson(\mu_1 + \mu_2)$$

Theorem 30 If $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ independently, then

$$T = X + Y \sim Binomial(n + m, p)$$

9.2 Multinomial Distribution

Physical Setup An experiment with k distinct outcomes with probability $p_!, p_2, \dots p_k$, repeated n times. Let X_i be the number of times i outcome occurs

$$(X_1, X_2, \dots X_k) \sim Multinomial(n, p_1, p_{2,k})$$

Note: $p_1 + p_2 + \dots + p_k = 1$
Note: $X_1 + X_2 + \dots + X_k = n$

Joint Probability Function

$$f(x_1, x_2, \dots x_p) = \frac{n!}{x_1!x_2! \dots x_p!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

9.3 Markov Chains

Def Markov Chain: A sequence of discrete random variables $X_1, X_2, \dots$ which take integer states $1, 2, \dots N$. There exists a certain transition probability matrix P, such that for all $i = 1, 2, \dots N, j = 1, 2, \dots N$

$$P(X_{n+1} = j \mid X_n = i) = P_{ij}$$

Note: Markov chains only depend on present state, not past states

Did not complete reading, was optional for stat 230

9.4 Covariance and Correlation

Mean

$$E(g(X_1, X_2, \dots X_n)) = \sum_{\text{all } (x_1, x_2, \dots, x_n)} g(x_1, x_2, \dots x_n) f(x_1, x_2, \dots x_n)$$

Proved Property of Multivariate Expectation:

$$E[ag_1(X_1, X_2) + bg_2(X_1, X_2)] = aE[g_1(X_1, X_2)] + bE[g_2(X_1, X_2)]$$

Def Covariance: A way to measure the relation between X and Y, denoted as

$$\sigma_{XY} = Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

Note: > 0 means positively correlated, < 0 negative means negatively correlated

Theorem 35 If X and Y are independent then $Cov(X, Y) = 0$

Theorem 36 If X and Y are independent then,

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$$

Def Correlation Coefficient: A way to measure the strength of the relation between X and Y, covariance scaled to $[-1, 1]$ denoted as

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Note: Since p has same sign as $Cov(X,Y)$, Theorems 35 and 36 hold
Note: As $p \to \pm 1$, the relation becomes linear

9.5 Mean and Variance of a Linear Combination of Random Variables

Proved Results for Means:

1. $E(aX + bY) = aE(X) + bE(Y)$
2. $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{n=1}^{n} E(X_i)$
3. if $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ then $E(X) = \mu$

Proved Results for Covariance:

1. $Cov(X, X) = Var(X)$
2. $Cov(aX+bY, cU+dV) = acCov(X,U)+adCov(X,V)+bcCov(Y,U)+bdCov(Y,V)$

Proved Variance of a linear combination:

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

Proved Variance of a sum of independent random variables: Assume X and Y are independent,

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y)$$

Proved Variance of a general linear combination of random variables:

$$Var\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_ia_jCov(X_i, X_j)$$

Proved Variance of a linear combination of independent random variables: Assume $X_1, X_2, \dots X_n$ are independent,

1. $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$
2. If $X_1, X_2, \dots X_n$ have the same variance, then $Var(\overline{X}) = \frac{\sigma^2}{n}$

9.6 Linear Combinations of Independent Normal Random Variables

Theorem 38 Linear Combinations of Independent Normal Random Variables:

1. Let $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$
2. Let $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ be independent, then $aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
3. Let $X_1, X_2, \dots X_n$ be independent $\sim N(\mu, \sigma^2)$ variables, then $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ and $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

9.7 Indicator Random Variables

Def Indicator Random Variables: Define a variable X_i where $X_i = 0$ indicates the trial was a failure, while $X_i = 1$ indicates the trial was a success

Proved Variance of a Hypergeometric random variable: Let X have a Hypergeometric distribution

$$Var(X) = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

CENTRAL LIMIT THEOREM/MOMENT GENERATING FUNCTIONS

10.1 Central Limit Theorem

Theorem 39 Central Limit Theorem: Let $X_1, X_2, \dots X_n$ be independent random variables with the same distribution, mean (μ), and variance (σ^2), then as $n \to \infty$

$$S_n = \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Note: Better approximation for larger n

Note: Better approximation when the distribution X_i is symmetric

Def Continuity Correction: Changing the bounds of $P(10 \leq S_{100} \leq 20)$ to be offset by .5, thus $P(9.5 \leq S_{100} \leq 20.5)$. Sign of offset is decided by whether left/right hand Riemann sum corrects value (if past expected value make positive)

Note: Should not be applied to a continuous distribution

Theorem 40 Normal Approximation to Poisson: Let $X \sim Poisson(\mu)$, then as $\mu \to \infty$, the cdf

$$Z = \frac{X - \mu}{\sqrt{\mu}} \sim N(0, 1)$$

Note: $X \sim N(\mu, \mu)$

Theorem 41 Normal Approximation to Binomial: Let $X \sim Binomial(n, p)$, then as $n \to \infty$, the random variable

$$W = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$$

Note: $X \sim N(np, np(1-p))$

10.2 Moment Generating Functions

Def Moment Generating Function: For a discrete random variable X and $a > 0$, the moment generating function is defined as

$$M(t) = E(e^{tX}) = \sum_{x \in \text{ all}} e^{tx} f(x) < \infty$$

For a continuous random variable X and $a > 0$, the moment generating function is

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) \mathrm{d}x < \infty$$

Note: The k^{th} moment is $E(X^k)$

Theorem 43: Let X have the moment generating function $M(t)$ for $t \in [-a, a]$, then

$$E(X^k) = M^{(k)}(0)$$

Proved MGF of Binomial: Let $X \sim Binomial(n, p)$, then

$$M(t) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (pe^t + 1 - p)^n$$

Proved MGF of Poisson: Let $X \sim Poisson(\mu)$, then

$$M(t) = e^{-\mu + \mu e^t}$$

Theorem 44 Uniqueness Theorem for Moment Generating Functions: Let X, Y have moment generating functions $M_X(t), M_Y(t)$, if $M_X(t) = M_Y(t)$ for all $t \in \mathbb{R}$ then X and Y have the same distribution

Proved MGF of Normal: Let $X \sim N(\mu, \sigma^2)$, then

$$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Proved MGF of Exponential: Let $X \sim Exponential(\theta)$, then

$$M(t) = \frac{1}{1 - \theta t} \text{ for } t < \frac{1}{\theta}$$

10.3 Multivariate Moment Generating Functions

Def Joint Moment Generating Function: For random variables X, Y, the joint moment generating function is defined as

$$M(s,t) = E(e^{sX+tY})$$

Note: If X, Y are independent, then $M(s,t) = M_X(s)M_Y(t)$

Theorem 47: The moment generating function of the sum of independent random variables is the product of individual moment generating functions

Theorem 48: Let $X_i = N(\mu_i, \sigma_i^2)$ be independent where $a_1, a_2, \dots a_n \in \mathbb{R}$ then

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Was optional for STAT 230