STAT 230
Probability

PROFESSOR N. MOHAMMAD • WINTER 2014 • UNIVERSITY OF WATERLOO

Last Revision: April 6, 2015

Table of Contents

Future Modifications

1 Introduction

Definition 1.1. Probability is the study of randomnesses and uncertainty.

- Variability in population, processes or phenomena

Definition 1.2. Classic Interpretatation makes assumptions about the physical world to deduce probability.

$$\frac{\text{\# ways an event can occur}}{\text{Total \# of outcomes}}$$

Definition 1.3. Relative-frequency interpretation is when the probability of specific outcome is defined as the proportion of times it occurs over the long run. May only be used if the experiment may be repeated.

Definition 1.4. Frequency is just the amount of times an event occurs while **relative frequency** is a fraction of the amount of times an event occurs over the total amount of possible outcomes.

Definition 1.5. Personal-probability interpretation is the degree to which a given individual believes the event wil happen.

> **Quote. Coherent** means that personal probability of one event does not contradict personal probability of another.

Example 1.1. If the probability of finding a parking space is 0.2, the probability of not finding one should be 0.8.

1.1 Probability Models

Definition 1.6. Experiment is any action, phenomenon, or process whose outcome is subject to uncertainty.

Definition 1.7. Trial is a single repetition of an experiment.

Definition 1.8. Sample space, denoted by S, is the set of possible distinct outcomes. In a single trial, only one outcome may occur. The sample space may be either **discrete** or **continuous**.

Definition 1.9. An **event** is any subset of outcomes contained in the sample space S.

- **Simple** - one outcome

- **Compound** - more than one outcome

Probability notation

$P(\text{event})$ is used to denote the probability of an event occurring. The **compliment** is the probability of an event not happening and is denoted as $P(A^c)$ or $P(\overline{A})$

Definition 1.10. The odds in **favour** of an event is the odds of an event occurring compared to its compliment.

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

The odds against is the reciprocal of odds in favour.

Theorem 1.1.1.

$$\sum_{i=0}^{k} P(A_i) = 1$$

The set $P(A_i), i = 1, 2, \ldots$ is the probability distribution on S.

Theorem 1.1.2. If A and B are two events wirh $A \subseteq B$, then $P(A) \leq P(B)$

Definition 1.11. Two events are **mutually exclusive** or **disjoint** if they cannot happen simultaneously.

$$A \cap B = \emptyset$$

1.2 Counting Techniques

Counting Principle • Permutations • Combinations

Definition 1.12. Addition rule: When there are m ways to perform A, and n ways to perform B, there are $m + n$ ways to perform A **OR** B.

Definition 1.13. Product rule: Where there are p ways to perform A, and q ways to perform B, there are $m \times n$ ways to perform A **AND** B.

Definition 1.14. Uniform Probability Model:

$$P(A) = \frac{\text{Outcomes in } A}{\text{Outcomes in } S}$$

Definition 1.15. Permutation is an arrangement of elements in an ordered list.

Note. The amount of ways to arrange n items (all of them have to be used) is $n!$

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

Definition 1.16. Any unordered sequence of k objects taken from a set of n distinct objects is called a **combination**.

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

Example 1.2. Binomial Theorem:

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$$

2 Conditional Probability

Definition 2.1. The **union** of two events A and B, $A \cup B$ is the set containing all outcomes in either A or B.

Definition 2.2. The **intersection** of two events A and B, $A \cup B$ is the set containing all outcomes that are in both A and B.

Definition 2.3. The **empty set**, $\emptyset$ is the set containing no outcomes.

Definition 2.4. The probability of **either** event happening is the sum of their individual probabilities:

$$P(A \cup B) = P(A) + P(B)$$

only if A and B are mutually exclusive.

Note. For non-mutually exclusive events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cup C)] + P(A \cap B \cap C)$$

Definition 2.5. Two events are **independent** if and only if

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

Definition 2.6. The conditional probability of A given B is defined by

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

Note. If two events A and B are independent, then

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

Quote. Show all steps to combinations and permutation. (formulas)

Definition 2.7. A set of events is **exhaustive** when at least one of the events must occur.

2.1 Disease Problem

Definition 2.8. Terminology:

- **False positive** is when test indicates positive and is wrong (true status is negative). $(T \mid D^c)$

- **False negative** is when a test indicates a negative status and is wrong (true status is positive). $(T^c \mid D)$

- **Sensitivity/ True positive** is when a test indicates a positive status and is correct. $(T \mid D)$

- textbfSpecificity/ True negative is when a test indicates a negative status and is wrong (true status is negative) $(T^c \mid D^c)$

Example 2.1. A cheap blood test for HIV has the following characteristics:

- False positive rate is 0.5%

- False negative rate is 2%

- Around 0.04% of Canadian males are infected with HIV

Determine the probability that if a male tests positive for HIV, he actually has HIV.

Solution. Let D represent has disease and T represent test positive. We are looking for $P(D \mid T)$. This is equivalent to

$$P(D \mid T) = \frac{P(DT)}{P(T)}$$

Given information:

$$P(D) = 0.0004, P(D^c) = 0.9996$$

$$P(T^c \mid D) = 0.02, P(T \mid D) = 0.98$$

$$P(T \mid D^c) = 0.005, P(T^c \mid D^c) = 0.995$$

Now for the calculations

$$P(DT) = P(T \mid D) \cdot P(D) = 0.98 \cdot 0.0004 = 0.000392$$

$$P(T) = P(DT) + P(D^cT) = P(DT) + P(T \mid D^c) \cdot P(D^c) = 0.000392 + 0.005 \cdot 0.9996 = 0.00539$$

$$P(D \mid T) = \frac{0.000392}{0.00539} = 0.0727$$

Theorem 2.1.1. Bayes's Theorem states: Let $A_1, A_2, \ldots, A_k$ be mutually exclusive with prior probabilities $P(A_i) \, i = 1, 2, \ldots, k$. Then for any other event B where $P(B) > 0$, the positerior probability of A_j given that B has occurred is

$$P(A_j|B) = \frac{P(A_J \cap B)}{P(B)}$$

3 Distributions

Definition 3.1. A **variable** is any characteristic whose value may change from one object to another in the population. They are denoted by lowercase letters.

Types of Variables

Definition 3.2. Numeric variable is a quantitative variable.

- **Continuous variable**: Can take values consisting of an entire interval ith infinite number of real values. (eg: time)

- **Discrete variable**: Can only take finite number of real values (even if limits may approach infinity). (eg number of people)

Definition 3.3. Categorical variables have values that describe a quality or characeristic of a data unit.

- **Ordinal variables** take on values that can be ordered/ranked. (grades, clothing size)

- **Nominal variables** take on value that are not able to be sequentialized. (gender)

Definition 3.4. Univariate Data only involve a single variable. Main purpose of these is for central tendency analysis.

Definition 3.5. Bivariate data involves two variables, and deals with causes or relationships. Analysis of variables simultaneously for correlation.

Definition 3.6. Multivariate data involve more than two variables for more in depth analysis.

Definition 3.7. A **random variable** is a function whose domain is the sample space and whose range is the set of possible values of the variable. It maps each outcome in the sample space with an outcome based on what the r.v. represents.

Example 3.1. Let X denote number of heads in two coin flips. $P(X = 0) = \frac{1}{4}$.

$$\text{Sample space: } \{TT, HT, TH, HH\}$$

$$X = \begin{cases} 0: & TT \\ 1: & HT, TH \\ 2: & HH \end{cases}$$

Definition 3.8. The **probability function** of a random variable, X is a function

$$f(x) = P(X = x) \qquad \forall x \in \mathbb{S}$$

Properties include:

$$f(x) \geq 0 \qquad \forall x \in \mathbb{S}$$
$$\sum_{x \in \mathbb{S}} f(x) = 1$$

Definition 3.9. A **probability distribution** of X is a description of the probabilities associated with all the possible values of X. It shows how the total probably of 1 is distributed among the various values of X.

Definition 3.10. The **cumulative distribution function** is defined by

$$F(x) = P(X \leq x) = \sum_i^x f(i)$$

For a number x, $F(x)$ is the probability that the observed value will be at most x.

Method 3.1. To calculate the cdf, first $P(x)$ must be determined for all values of x. Then add each $f(x)$ to all the previous ones to obtain $P(x)$. For example if the range was $\{0, 1, 2\}$

$$F(1) = P(X \leq 1) = P(1) + P(0)$$

$$F(x) = \begin{cases} \# : & x < 0 \\ \# : & 0 \leq x < 1 \\ \# : & 1 \leq x < 2 \end{cases}$$

Note. For any number a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

$a-$ represents the largest X value that is strictly less than a.

Note. In addition: $f(x) = F(x) - F(x - 1)$

3.1 Uniform Distribution

Definition 3.11. Uniform distribution is where each outcome has equal probability. If the range was from a to b,

$$P(X = x) = \begin{cases} \frac{1}{b-a+1} = \frac{1}{n} & : x \in [a, b] \\ 0 & : x \notin [a, b] \end{cases}$$

Examples of this distribution include rolling a fair die, flipping a fair coin.

3.2 Binomial Distribution

Definition 3.12. Binomial distributions are based on the probability experiments for which the results of each trial can be either a success or faillure. Requirements include:

- Experiment is repeated for fixed number of trials

- Only two possible outcomes for each trial

- Trials are independent

- Probability of success is fixed

- If sampling is without replacement, only use binomial distribution if the sample size is at most 5% of the population size.

Formula

- p denotes the probability of success in a trial.

- $q = 1 - p$ denotes the probability of failure in a trial.

- X is the range of successes in n tries (where $x = 0, \cdots, n$)

$$P(X = x) = b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}, \ 0 \le x \le n$$

We write $x \sim \text{Binomial}(n, p)$.

Definition 3.13. The **Hypergeometric Distribution** is where a sample is taken from a finite population without replacement. Similar to binomial distribution, each trial has the possibility of success or failure. N is the population, n is the sample selected, r is the total amount of successes available in the population, and x represents the amount of successes within the selected sample.

$$P(X = x) = \frac{\binom{r}{x}\binom{n-1}{n-x}}{\binom{N}{n}}, \ x = 0, 1, \ldots, n$$

Quote. Choose x successes from the success samples, and out of the remaining failures, $N - r$, choose $n - x$ (failures in sample), all over the total amount of ways to select n things in the population N.

Definition 3.14. The **negative binomial distribution** is based on an experiment satisfying thes conditions (similar to binomial)

- Independent trials

- Only success or failure

- P(success) is constant

- **Experiment continues until k successes have occurred**

The rv in this distribution is the number of failures obtained before the kth success occurs.

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

Definition 3.15. The **geometric distribution** is a special case of negative binomial where k is fixed at 1.

$$f(x) = P(X = x) = p(1 - p)^x$$

Definition 3.16. The **Poisson Distribution** is used to model the number of events occurring within a given time. It can be used as a limiting case of the binomial distribution where n is very large and p is very small.

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, 2, \cdots, \infty$$

Note. $\lambda = np$ from a binomial distribution.

Quote. This approximation can be used as a substitute if $n > 50$ and $np < 5$.

Definition 3.17. Poisson Process: must satisfy three conditions.

- Independence

- Individuality (chance of 2 or more events occuring in the same interval close to 0)

- Homogencity: Events occur at uniform or homogeneous rate λ over time

$\mu = \lambda t$

4 Value and Variance

Definition 4.1. A **histogram** displays information on the frequency and relative frequency of each x in the data set.

Characteristics

- **Symmetry**: if the graph is symmetrical.

- **Unimodal**: if there is a single prominent peak. (Local maximum)

- **Bimodal**: Two prominent peaks

- **Multimodal**: More than two prominent peaks.

Definition 4.2. Skewedness refers to whether the data is pulled to wards a side. Could be positively skewed (towards left) or negatively skewed (towards right).

Definition 4.3. The **arithmetic mean** denotes the arithmetic average value of observations. (Average)
The **sample mean** measure the location, balance, and center of a sample.

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

The **population mean** is a weighted average in a probability distribution.

$$\mu = \sum_{i=1}^{N} x_i \cdot f(x)$$

Definition 4.4. The **median** is the middle value in the **ordered sample**. If there are two, take the average of the two.

Definition 4.5. The **mode** is the value that occurs most often.

Definition 4.6. Outliers are values that are very far from the rest of the data. Outlists affect the mean, but generally do not affect median and mode.

Definition 4.7. Variance is a measure of the spread of the recorded values on a variable. It is a measure of dispersion. The **sample variance** is

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

For N observations in a population, the **population variance** is

$$\sigma^2 = \sum_{i=1}^{N}(x_i - \mu)^2 \cdot f(x) = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Definition 4.8. $E(X)$ is the expected value of a distribution and is usually denoted as μ_x.

Variance of X, denoted as σ^2 is

$$\sigma^2 = V(X) = \sum(x - \mu)^2 f(x) = E(X - \mu)^2 = \sum x^2 f(x) - \mu^2 = E(X^2) - [E(X)]^2$$

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$
$$\mathrm{Var}(x) = E(X(X - 1)) + \mu - \mu^2$$

If X is a discrete random variable with the probability function $f(x)$, then the expected value of a function $g(x)$ applied to all the values of X is

$$E(g(x)) = \sum g(x)f(x)$$

Properties of Expected Value

$$E(a\,g(X) + b) = aE(g(X)) + b$$

Rules of Variance

$$V(aX + b) = a^2\sigma_x^2$$
$$\sigma_{ax+b} = |a|\sigma_x$$

Note. The addition of a constant does not affect variance because it only changes location of mean value, but entire graph is shifted so spread is the same.

If $X \sim \text{Bin}(n, p)$ then

$$\mu_x = E(X) = np$$
$$\sigma^2 = Var(X) = np(1 - p)$$

If X follows hypergeometric:

$$p = \frac{r}{N}$$
$$E(X) = np$$
$$V(X) = \frac{N - n}{N - 1} np(1 - p)$$

If X follows Poisson distribution

$$\mu = \lambda$$
$$\sigma^2 = \lambda$$

If X follows uniform distribution

$$\mu = \frac{a + b}{2}$$
$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12}$$

If X follows negative binomial

$$\mu = \frac{kq}{p}$$
$$\sigma = \frac{kq}{p^2}$$

cpt 7

5   Continuous Probability Distributions

Definition 5.1. A random variable that can accept any value within an interval of numbers is **continuous**. Since there is an infinite amount of numbers between any two intervals, $P(X = x) = 0$ for each x within the range.

Definition 5.2. A **probability distribution** for a **continuous variable** X is a function $f(x)$, and it measures an interval such that for any two numbers $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

Note. Since $\int_a^a f(x) = 0$, $P(a \leq X \leq b) = P(a < X < b)$.

Example 5.1. Suppose that X is a continuous rv with a probability density function

$$f(x) = \begin{cases} C(4x - 2x^2) : 0 < x < 2 \\ 0 : \text{otherwise} \end{cases}$$

- Find the value of C

- Find $P(X > 1)$

Solution.

$$C \int_0^2 (4x - 2x^2) \, dx = 1$$

$$C\left(2x^2 - \frac{2x^3}{3}\right)\Big|_0^2 = 1$$

$$C = \frac{3}{8}$$

$$f(x) = \frac{3}{8}(4x - 2x^2), \, 0 < x < 2$$

Therefore

$$P(X > 1) = \int_1^2 f(x) \, dx = \frac{3}{8} \int_1^2 (4x - 2x^2) \, dx = \frac{1}{2}$$

Example 5.2. Let the continuous rv X denote the current measured in a thin wire in milliamperes. Assume that the range of X is $[0.200]$ mA, and the probability density function is

$$f(x) = \begin{cases} 0.05 : 0 \leq x \leq 20 \\ 0 : \text{otherwise} \end{cases}$$

What is the probability that a current measurement is less than 10 milliamperes?

Solution.

$$P(X < 10) = \int_0^{10} 0.05 \, dx = 0.05x \Big|_0^{10} = 0.5$$

Definition 5.3. The **cumulative distribution function**, $F(x)$ for a **continuous random variable**,X, is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y) \, dy$$

Properties of cdf

- $\lim_{x \to \pm \infty} F(x) = 0$

- $F(x)$ is a non-decreasing function of x.

- $P(a < X \leq b) = F(b) - F(a)$

- $P(X = a) = 0$

Example 5.3. Utilizing the previous battery example, the cdf is

$$F(x) = \int_0^x f(u)\, du = 0.05x$$

$$F(x) = \begin{cases} 0 : x < 0 \\ 0.05x : 0 \le x < 20 \\ 1 : x \ge 20 \end{cases}$$

Note. If X is a continuous rv with pdf $f(x)$, and cdf $F(x)$, then at every number for which $F'(x)$ exists,

$$F'(x) = f(x)$$

Definition 5.4. The **percentile** refers to the value for which a percentage of data falls below. For example being in the 99th percentile means that 99% of people in the sample space are below you.

Note. Sometimes it is useful to change the variable to solve a specific problem. The process is

- Write the cdf of Y as a function of X

- Use $F_X(x)\ to\ find\ F_Y(y)$ If pdf is needed, differentiate $F_Y(y)$

- Find the range of values of y

Example 5.4.

$$f(x) = \begin{cases} \frac{1}{4} : 0 \le x < 4 \\ 0 : \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 : x < 0 \\ \frac{x}{4} : 0 \le x < 4 \\ 1 : x \ge 4 \end{cases}$$

Let $Y = \frac{1}{X}$. Find $f(y)$

Solution.

$$F_Y(y) = P(Y \le y) = P\left(\frac{1}{X} \le y\right) = P\left(X \ge \frac{1}{y}\right)$$
$$= 1 - P\left(X < \frac{1}{y}\right)$$
$$= 1 - F_X\left(\frac{1}{y}\right)$$

Substitute $\frac{1}{y}$ for x in $F_X(x)$

$$F_Y(y) = 1 - \frac{\frac{1}{y}}{4} = 1 - \frac{1}{4y}$$

To obtain $f_Y(y)$, simply differentiate the cdf. New bounds are determined by subbing in y for x.

$$0 < x \leq 4 \implies 0 < \frac{1}{y} \leq 4 \implies \infty > y \geq \frac{1}{4} \implies y \geq \frac{1}{4}$$

$$f_Y(y) = \begin{cases} \frac{1}{4y^2} : y \geq \frac{1}{4} \\ 0 : \text{otherwise} \end{cases}$$

Definition 5.5. The **expected value** of a continuous random variable X is,

$$E(X) = \int_{-\infty}^{\infty} x f(x) \, dx$$

Similarly, if $f(x)$ and $g(x)$ are functions of X, then

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) \, dx$$

Example 5.5. Find $E(X)$ if the probability density function is $f(x) = 2x$, $0 \leq x \leq 1$.

Solution.

$$E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_0^1 2x \cdot x \, dx = \frac{2x^3}{3} \Big|_0^1 = \frac{2}{3}$$

Example 5.6. Suppose

$$f(x) = \begin{cases} 1 : 0 \leq x \leq 1 \\ 0 : \text{otherwise} \end{cases}$$

$$h(x) = \begin{cases} 1 - x : 0 \leq x < \frac{1}{2} \\ x : \frac{1}{2} \leq x \leq 1 \end{cases}$$

Find E(h(x))

Solution.

$$E(h(x)) = \int_0^{\frac{1}{2}} (1 - x) \cdot 1 \, dx + \int_{-\frac{1}{2}}^1 x \cdot 1 \, dx$$

$$E(h(x)) = \left(x - \frac{x^2}{2} \right) \Big|_0^{\frac{1}{2}} + \left(\frac{x^2}{2} \right) \Big|_{\frac{1}{2}}^1$$

Definition 5.6. If the random variable X is continuous with the pdf $f(x)$, then the variance is

$$\text{Var}(X) = E((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \mu^2 = E(X^2) - E(\mu^2)$$

Definition 5.7. A continuous random variable has **uniform distribution** from a to b if all subinterals of a fixed length are equally likely

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

$\mu = E(X) = \frac{(a+b)}{2}$ and $\sigma^2 = V(X) = \frac{(b-a)^2}{2}$

Example 5.7. If X is uniformly distributed over $(0, 10)$, find $P(3 < X < 8)$.

Solution.

$$P)3 < X < 8) = \int_3^8 \frac{1}{10}\, dx = \frac{1}{2}$$

Definition 5.8. The normal distribution with parameter values $\mu = 0, \sigma = 1$ is the **standard normal distribution**. A random variable that follows this distribution is a standard normal random variable, and is denoted as Z.

Note. To convert a normal distribution to a standard score (z-score),

$$z = \frac{\text{observed mean} - \text{mean}}{\text{standard distribution}}$$

$$z = \frac{x - \mu}{\sigma}$$

Note. Since our z table does not have negatives, when calculating $\phi(-x)$, use $1 - \phi(x)$ since the distribution is symmetrical.

5.1 Exponential Distribution

Definition 5.9. X is said to follow **exponential distribution** if the pdf of X is

$$f(x; \lambda) = \lambda e^{\lambda x}$$

The cdf for this function is $F(x; \lambda) = 1 - e^{-\lambda x}$.

It is used to describe the time or distance until some event happens.

$$E(X) = \mu = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Note. If the **average rate of occurrence** in a Poisson process is given, λ is used. If the **average waiting time for an occurrence** then θ is used. We use $\theta = \frac{1}{\lambda}$

Example 5.8. The time between arrivals of cars at a gas pump station follows an exponential distribution with an average time between arrivals of 3 minutes. What is the probability that the time between two successive arrivals will be 2 mins or less.

Solution. Must calculate $P(X \leq 2)$. From the question, $\theta = 3$, so $F(x) = 1 - e^{-\frac{x}{3}}$. $F(2) = 1 - e^{-\frac{2}{3}} = 0.4866$.

Note. Exponential distributions are memoryless, so for example the probability of waiting 15 mins given that you have already waited 5 mins is the same as the probability without the initial waiting.

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

6   Multivariate Distributions

Definition 6.1. Let X and Y represent two discrete random variables in the sample space $\mathbb{S}$. The **joint probability function** $f(x, y)$ is

$$f(x, y) = P(X = x \text{ and } Y = y) = P(X = x, Y = y)$$

Note. If in a joint probability distribution, you are only asked to find one variable (eg: $P(X > 1)$), then you have to include all the possible values for the other variable (eg: $\sum^{\text{all } y} P(X > 1.Y = i)$).

Definition 6.2. The **marginal probability functions** of X and Y, denoted $f_X(x)$ and $f_Y(y)$ are given by

$$f_X(a) = \sum_{}^{\text{all y}} f(x, y)$$

$$f_Y(b) = \sum_{}^{\text{all x}} f(x, y)$$

Example 6.1. Let the joint distribution of X and Y be given by

$$f(x, y) = \frac{x + y}{30} \qquad x = 0, 1, 2, 3 \qquad y = 0, 1, 2$$

Find the marginal distribution function of X and Y.

Solution.

$$f_X(x) = \sum_{y=0}^{2} \frac{x + y}{30} = \frac{1}{30}((x + 0) + (x + 1) + (x + 2)) = \frac{x + 1}{10}$$

$$f_X(x) = \begin{cases} \frac{x+1}{10} : & x = 0, 1, 2, 3 \\ 0 : & \text{otherwise} \end{cases}$$

Similarly, by computation,

$$f_Y(y) = \begin{cases} \frac{2y+3}{15} : & y = 0, 1, 2 \\ 0 : & \text{otherwise} \end{cases}$$

Note. An alternative solution would be to make a table and sum up the rows and columns, but that is a lot of work.

Definition 6.3. The random variables $X, Y, \ldots$ are **independent random variables** if for every subset of the variables, the joint probability function of the subset is equal to the product of the marginal probability functions.

$$f(x, y) = f(x) \cdot f(y) \qquad \forall x, y \in \mathbb{S}$$

Definition 6.4. The conditional probability function of X given $Y = y$ is given by

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

and similarly

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

Example 6.2. Suppose the joint probability function of X and Y is given by

$$f(0,0) = 0.4, f(0,1) = 0.2, f(1,0) = 0.1, f(1,1) = 0.3$$

Calculate the conditional probability function of X given that $Y = 1$.

Solution.

$$f_Y(1) = \sum_{}^{\text{all x}} f(x,1) = f(0,1) + f(1,1) = 0.5$$

$$f_{X|Y}(0|1) = \frac{f(0,1)}{f_Y(1)} = \frac{0.2}{0.5} = \frac{2}{5}$$

$$f_{X|Y}(1|1) = \frac{f(1,1)}{f_Y(1)} = \frac{0.3}{0.5} = \frac{3}{5}$$

p33, bi,poisson distribution with multivariable, and change of variable

6.1   Multinomial Distribution

Suppose an experiment is repeated independently n times with k distinct types of outcome each time. Each experiment can result in any one of k possible outcomes, with probabilities $p_1, \ldots, p_k$ each time. If X denotes the number of times the ith type occurs, then $(X_1, \ldots, X_k)$ has a multinomial distribution.

$$f(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x1} \cdot p_2^{x2} \cdots p_k^{xk}$$

Example 6.3. Suppose that a fair die is rolled twelve times. The probability that 1 appears 4 times, 2 and 3 appear 3 times each, 4 and 5 appear once each, and 6 not at all is

$$\frac{12!}{4!3!3!1!1!0!} \left(\frac{1}{6}\right)^4 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0$$

$$= \frac{12!}{4!3!3!} \left(\frac{1}{6}\right)^{12}$$

Some stuff in the middle

Definition 6.5. Let P be a $n \times n$ matrix. P is a **transition probability matrix** when the conditional probabilities are given by corresponding elements of the matrix

Actually a lot of stuff in the middle

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

Example 6.4. Suppose that the probability that tomorrow is rainy given that today is not raining is α. The probability that tomorrow is dry given that today is rainy is β. The random variable $X_n, n = 0, 1, 2, \ldots$ form a **Markov chain** with 2 possible states and having probability transition matrix

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

The rows represent the current day, columns represent tomorrow, with first row and column representing dry and 2nd row and column representing rainy.

Quote. Every row in the Markov chain must add up to 1.

Definition 6.6. A **stohastic matrix** has the property that all its rows add up to 1.

Definition 6.7. A **double stochastic matrix** has the property that both all its rows and columns add up to 1.

Definition 6.8. If X is a discrete random variable with probability function $f(x) = P(X = x)$, then the expected value is

$$E(X) = \mu_X = \sum_{}^{\text{all x}} x f(x)$$

$$E(g(X)) = \sum_{}^{\text{all x}} g(x) f(x)$$

Similarly,

$$E(XY) = \sum_{}^{\text{all(x,y)}} xy f(x, y)$$

$$E(g, X, Y) = \sum_{}^{\text{all(x,y)}} g(x, y) f(x, y)$$

$$E(X + Y) = E(X) + E(Y)$$

If X and Y are independent, then for any functions h and g,

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

$$E(ag(X, Y) + bh(X, Y)) = aE(g(X, Y)) + bE(h(X, Y))$$

Definition 6.9. Deterministic relationship has no variance nor uncertainty. If the value of one variable is known, the other variable can be determined exactly. (Eg: Unit conversions). Represented as $y = \beta_0 + \beta_1 x$

Definition 6.10. Statistical relationship has both variance and uncertainty. For a fixed value of x there is unertainty in the value of the second variable $y/$

Definition 6.11. Regression analysis is the part of statistics that investigates the relationship between two or more variables related in nondeterministic fashion. For this course, **covariance** and **correlation coefficient** are analyzed.

Definition 6.12. The **covariance** between two random variables is

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

$$= \sum (x - \mu_X)(y - \mu_Y) f(x, y)$$

$$\text{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

This measures how much two random variables change together.

Note. If X and Y are independent, then $\text{Cov}(X, Y) = 0$, but the converse is not true.

Definition 6.13. The **correlation coefficient**, denoted as $\text{Corr}(X, Y)$, $\rho_{X,Y}$ or just ρ is defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

If a and c have the same sign (both positive/negative),

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

The relationship is strong if $|\rho| > 0.8$, moderate if $0.5 < |\rho| < 0.8$, and weak if $|\rho| < 0.5$. ρ measures the degree of **linear relationship** between X and Y.

Definition 6.14. Given a collection of n random variables $X_1, \ldots, X_n$, the random variable

$$Y = a_1 X_1 + \cdots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a linear combination of the X_is. The mean is given by

$$E(a_1 X_1 + \cdots + a_n X_n) = a_1 E(X_1) + \cdots + a_n E(X_n) = a_1 \mu_1 + \cdots + a_n \mu_n$$

Theorem 6.1.1. Variance of a linear combination of random variables:
If $X_1, \ldots, X_n$ are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

$$V(a_1 X_1 + \cdots + a_n X_n) = a_1^2 V(X_1) + \cdots + a_n^2 V(X_n) = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2$$

If $X_1, \ldots, X_n$ are dependent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

If X_1 and X_2 are independent,

$$V(X_1 - X_2) = V(X_1) + V(X_2)$$

The women problem on pg 12 9c