Thoery and Applications of Probability and Statistics

Xi Tan (tan19@purdue.edu)

April 8, 2013

Contents

Part I: Probability **6**

1 Combinatorial Analysis **7**

1.1 Axioms . . . 7

1.2 Binomial Coefficient and Its Applications . . . 7

1.2.1 Binomial Coefficient . . . 7

1.2.2 Bernoulli Distribution . . . 9

1.2.3 The i.i.d. Case: Binomial Distribution . . . 9

1.2.4 The Batch Mode Case: Hypergeometric Distribution . . . 9

1.3 Multinomial Coefficient and Its Applications . . . 9

1.3.1 Multinomial Coefficient . . . 9

1.3.2 Categorical Distribution . . . 10

1.3.3 Multinomial Distribution . . . 10

1.4 Multiset Coefficient and Its Applications . . . 10

1.4.1 Multiset Coefficient . . . 10

1.5 Selected Topics . . . 11

1.5.1 Double Factorial . . . 11

1.5.2 Stirling Numbers . . . 11

1.5.3 The Bertrand's Ballot Problem . . . 11

1.5.4 Catalan Number . . . 14

2 Probability **14**

2.1 Axioms . . . 14

2.1.1 Law of Total Probability . . . 14

2.1.2 Law of Total Variance . . . 14

2.1.3 Law of Total Covariance . . . 14

2.1.4 Law of Total Expectation . . . 14

2.1.5 Law of Total Cumulance . . . 14

2.1.6 Probability Inequalities . . . 14

2.2 Definitions of Sample Space, Events, and Probability . . . 14

2.3 Types of Probabilities: Frequentism and Bayesian . . . 15

2.4 Probability Redefined: Formal Definition with Measure Theory . 15

3 Conditional Probability and Independence 15
3.1 Conditional Probability 16
3.2 Conditional Expectation 16
3.3 Conditional Independence 16

4 Random Variables 16
4.1 Continuous Random Variables 18
4.2 Discrete Random Variables 18
4.3 Joint Distributed Random Variables 18
4.4 Random Vectors/Matrices 18
4.5 Stochastic Processes 18
4.6 Random Field 18
4.7 Convolutions: Sum of Normally Distributed Random Variables . 18
4.8 Product Distribution 18
4.9 Ratio Distribution 18

5 Quantitative Measure and Characteristic Functions 18
5.1 Describing Shape of a Distribution: Skewness, Kurtosis 18
5.2 Describing a Sample: Mean, Variance 18
5.3 Degrees of Freedom, Mean, Variance, and Moment, Central Moment, Cumulant, Law of the unconscious statistician 18
5.4 Percentile and Median 18
5.5 Coefficient of Variation 18
5.6 Covariance and Correlation 18
5.7 Moment Generating Function 18
5.8 Characteristic Function 18

6 Limit Theorems 18

7 Selected Topics in Probability 18
7.1 Ordered Statistics 18
7.2 Copula 18

8 Summary of Common Distributions 18

Part II: Statistics 18

9 Collecting Data: Experiments and Surveys 18
9.1 Design of Experiments 18
9.2 Statistical Survey 18
9.3 Opinion Poll 18
9.4 Sampling 18
9.4.1 Sampling Distribution 18
9.4.2 Sampling: Stratified Sampling, Quota Sampling 18
9.4.3 Biased Sample: Spectrum Bias, Survivorship Bias 18

10 Describing Data **18**
10.1 Average: Mean, Median, and Mode 18
10.2 Measures of Scale: Variance, Standard Deviation, Geometric Standard Deviation, and Median Absolute Deviation 18
10.3 Correlation and Dependence 18
10.4 Outlier 18
10.5 Statistical Graphics: Histogram, Frequency Distribution, Quantile, Survival Function, and Failure Rate 18

11 Filtering Data **18**
11.1 Recursive Bayesian Estimation 18
11.1.1 Kalman Filter 18
11.1.2 Particle Filter 18
11.2 Moving Average 18

12 Modeling Sample Data **18**
12.1 Regression Analysis 18
12.1.1 Analysis of Variance (ANOVA) 18
12.1.2 General Linear Model 18
12.1.3 Generalized Linear Model 18
12.2 Density Estimation 18
12.2.1 Kernel Density Estimation 18
12.2.2 Multivariate Kernel Density Estimation 18
12.3 Time Series 18
12.4 Multivariate Analysis 18
12.4.1 Principal Component Analysis (PCA) 18
12.4.2 Factor Analysis 18
12.4.3 Cluster Analysis 18
12.4.4 Correspondence Analysis 18
12.4.5 Multidimensional Scaling 18
12.5 Robust Statistics 18

13 Modeling Population Data: Statistical Inference **18**
13.1 Bayesian Inference 18
13.1.1 Bayes' Theorem, Bayes Estimator, Prior Distribution, Posterior Distribution, Conjugate Prior, and All That 18
13.2 Frequentist Inference 18
13.2.1 Statistical Hypothesis Testing: Null, Alternative, P-value, Significance level, power, likelihood-ratio test, goodness-of-fit, confidence interval, M-estimator, Trimmed Estimator 18
13.3 Non-parametric Statistics 18
13.3.1 Nonparametric Regression, Kernel Methods 18

14 Making Decisions: Decision Theory **18**
14.1 Optimal Decision, Type I and Type II errors 18
14.2 Correlation and Causation 18

15 Statistics Theory **19**
15.1 Mathematical Statistics 19
15.1.1 Degrees of Freedom 19
15.1.2 sufficient, complete, and etc. 19
15.1.3 Likelihood Function 19
15.1.4 Exponential Family 19
15.2 Statistical Learning Theory 19
15.3 Interpretations of Probability 19
15.3.1 Cox's Theorem 19
15.3.2 Principle of Maximum Entropy 19

Preface

This booklet is divided into 7 Chapters. The first chapter introduces the definitions of basic concepts, such as event, sample space, and probability space. Followed in the next chapter, we will discuss the relationship between two or more events when they interplay with each other. The third chapter formally brings in random variables and vectors, as a basis to develop their quantitative measure and characteristic functions later in chapter four. Chapter five includes some well-known limit theorems, which is useful for asymptotic analysis. The last two chapters will discuss several selected topics in probability theory, and provide a summary of common distributions.

Part I: Probability

1 Combinatorial Analysis

1.1 Axioms

There are two important rules in combinatorics: the rule of sum, and the rule of product.

The rule of sum says, if we have a ways to finish a task using one method and alternatively, b ways to finish the same task using another method, then there are ab ways of finish this task. More generally,

$$|S_1 \cup S_2 \cup \ldots \cup S_n| = |S_1| + |S_2| + \ldots + |S_n| \tag{1}$$

One extension of the rule of sum is the inclusion-exclusion principle, which does not require sets A_i to be disjoint. This does include the rule of sum, in that if sets A_i are disjoint, the terms from the second to the last are all zero.

$$\begin{aligned} |\bigcup_{i=1}^{n} A_i| = & \sum_{i=1}^{n} |A_i| - \sum_{1 \le i < j \le n} |A_i \cap A_j| + \sum_{1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| - \cdots \\ & + (-1)^{n-1} |A_1 \cap \cdots \cap A_n| \end{aligned} \tag{2}$$

The rule of product says, if finishing one task requires two steps, and there are a ways to choose in the first step and b ways to choose in the second step, then there are ab ways to finish this task. More generally,

$$|S_1 \times S_2 \times \cdots \times S_n| = |S_1| \cdot |S_2| \cdots |S_n| \tag{3}$$

1.2 Binomial Coefficient and Its Applications

1.2.1 Binomial Coefficient

We list here some of the useful binomial identities, all numbers are nature number (not including 0).

$$\binom{n}{k} = \binom{n}{n-k} \tag{4}$$

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{5}$$

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} \tag{6}$$

From the famous Pascal's rule,

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \tag{7}$$

There is another form which is equivalent to equation 7,

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \tag{8}$$

Here is an example that uses the **logarithmic differentiation**, $f' = f\dot{[}\ln(f)]'$.

$$\frac{d}{dt}\binom{t}{k} = \binom{t}{k}\sum_{i=0}^{k-1}\frac{1}{t-i} \tag{9}$$

A list of series that involves binomial coefficients,

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n \tag{10}$$

$$\sum_{k=0}^{n}k\binom{n}{k} = n2^{n-1} \tag{11}$$

$$\sum_{k=0}^{n}k^2\binom{n}{k} = (n+n^2)2^{n-2} \tag{12}$$

These can all be obtained by examining the function value or derivatives of the function $(1+x)^\alpha$, where α could be any real number, and $|x| < 1$.

There are some identities that could be proved using combinatorial analysis, such as **double counting**. Here is an example,

$$\sum_{k=1}^{n}\binom{n}{k}\binom{k}{q} = 2^{n-q}\binom{n}{q} \tag{13}$$

The left side of equation 13 counts the number of ways of selecting k elements first, and then choosing q elements from the resulting subset. These q elements could be identical for different k. The right hand side of the equation says this is equivalent to first choosing q elements directly from the set, and merging them into one of the 2^{n-q} subset of the set containing all but those selected q elements.

Another example is,

$$\sum_{m_1=0}^{n_1}\binom{n_1}{m_1}\binom{n_2}{m-m_1} = \binom{n}{m} \tag{14}$$

This simply means choosing $m = m_1 + m_2$ objects from a set of $n = n_1 + n_2$ objects is equivalent to choosing m_1 objects from n_1 objects, and m_2 objects from n_2 objects.

Sometimes, knowing the bounds and asymptotic formulae could be helpful.

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \frac{n^k}{k!} \le \left(\frac{n\cdot e}{k}\right)^k \tag{15}$$

$$\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}, \text{ as } n \to \infty \tag{16}$$

1.2.2 Bernoulli Distribution

1.2.3 The i.i.d. Case: Binomial Distribution

1.2.4 The Batch Mode Case: Hypergeometric Distribution

1.3 Multinomial Coefficient and Its Applications

1.3.1 Multinomial Coefficient

The notion of **multinomial coefficient** is a generalization of binomial coefficient, which is defined in the multinomial theorem:

$$(x_1 + x_2 + \ldots + x_r)^n = \sum_{(n_1,\ldots,n_r):n_1+\ldots,n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$$

We call $\binom{n}{n_1,n_2,\ldots,n_r}$ the multinomial coefficient.

Problem 1.1 A set of n distinct items is to be divided into r distinct groups of respective sizes $n_1, n_2, \ldots, n_r$, where $\sum_{i=1}^r n_i = n$. How many different divisions are possible?

Note that there are $\binom{n}{n_1}$ possible choices for the first group; for each choice of the first group there are $\binom{n}{n-n_1}$ possible choices for the second group; and so on. Hence it follows that there are

$$\begin{aligned}
&\binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-n_2-\ldots-n_{r-1}}{n_r} \\
&= \frac{n!}{(n-n_1)!n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!n_2!}\ldots\frac{(n-n_1-n_2-\ldots-n_{r-1}!}{(0)!n_r!} \\
&= \frac{n!}{n_1!n_2!\ldots n_r!}
\end{aligned}$$

possible divisions.

Alternatively, we can first permute these n items, where there are $n!$ such orderings. The first n_1 elements are assigned to group 1, the next n_2 elements are assigned to group 2, and so on. However, for example, keeping all but n_i group fixed, this method would generate $n_i!$ equivalent divisions (note the order within a group does not matter). Therefore, we need to cancel out the equivalent-group effect by dividing $n_1!n_2!\ldots n_r!$. Finally, the multinomial coefficient is,

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\ldots n_r!} \tag{17}$$

1.3.2 Categorical Distribution

1.3.3 Multinomial Distribution

1.4 Multiset Coefficient and Its Applications

1.4.1 Multiset Coefficient

The notion of multiset (or bag) is a generalization of the notion of set in which members are allowed to appear more than once.

The number of times an element belongs to the multiset is the **multiplicity** of that member. The total number of elements in a multiset, including repeated memberships, is the **cardinality** of the multiset. For example, in the multiset {a, a, b, b, b, c} the multiplicities of the members a, b, and c are respectively 2, 3, and 1, and the cardinality of the multiset is 6.

The number of multisets of cardinality k, with elements taken from a finite set of cardinality n, is called the **multiset** coefficient or **multiset number**, and is denoted as $\left(\!\binom{n}{k}\!\right)$. It is equivalent to asking, with replacement, the number of all possible combinations of making k draws from a urn with n distinguishable balls labeled $1 \dots n$.

Problem 1.2 With replacement, how many possible combinations to make k draws from a urn with n distinguishable balls labeled $1 \dots n$?

If we translate this problem directly to the combinatorial language, we may end up with counting the total number of the multinomial coefficients (actually, it is also correct). To solve this problem, let's first see a similar example.

It is NOT the sum of all the multinomial coefficients, which can be seen by computing $(1 + \cdots + 1)^k$, that is, $\sum \binom{k}{N_1, N_2, \ldots, N_n} = n^k$.

Problem 1.3 Suppose k balls that are indistinguishable from each other are to be distributed into n distinguishable (non-empty) urns, how many different outcomes are possible?

We note that this problem is equivalent to selecting $n-1$ of the $k-1$ spaces between (fixed) adjacent objects as our dividing points, e.g., $OOO|OOO|OO$. We count the ”bars” as urns (in total n) and big ”O” as balls (in total k). Therefore, there are $\binom{k-1}{n-1}$ such outcomes.

Objects being adjacent ensures the non-emptiness.

Now, we are ready to go back to our original problem. Problem 1.2 could be asked this way: if however, we allow empty urns, how many outcomes are possible?

One possible way of borrowing the non-empty case solution to solve the empty one is by starting with $k+n$ balls (instead of k), and place them into n urns, however at last remove one ball from each urn (so in total n). Because some of the urns may only contain one ball, this would give us the number of orderings $\binom{n+r-1}{r-1}$. Another way to look at it is to allow all $k+n-1$ positions (not gaps) available to both symbols, and we count all balls between two bars the same type. Hence, the number of all possible combinations is $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$. Note, this scheme is allowing emptiness, because two “bars” can be adjacent to each other.

It is of probability 1 to remove one ball from each urn.

Another beautiful explanation is to construct an equivalent mapping. Note, the rule of drawing a series of k numbers $a_1, a_2, \ldots, a_k$ from the set $\{1, 2, \ldots, n\}$ with repetition is

$$1 \leq a_1 \leq a_2 \leq \ldots \leq a_k \leq n$$

Now, a new series of k numbers $b_1, b_2, \ldots, b_k$ can be constructed as follows

$$\begin{array}{ccccccc} 1 \leq a_1 < & a_2{+}1 < & \ldots < & a_k{+}k-1 \leq n+k-1 \\ \downarrow & \downarrow & & \downarrow \\ b_1 & b_2 & \ldots & b_k \end{array}$$

Note that $b_1 < b_2 < \ldots < b_k$. This is a model without replacement, and is a one-to-one mapping of the original problem. Under the model of drawing k times from $n + k - 1$ balls without replacement, the number of all possible combinations is $\binom{n+k-1}{k}$, which is the same as what we obtained earlier.

A last thing worth noting is that, as we mentioned earlier, the number of all multinomial coefficients is also $\binom{k+n-1}{n-1}$.

1.5 Selected Topics

1.5.1 Double Factorial

1.5.2 Stirling Numbers

1.5.3 The Bertrand's Ballot Problem

The Bertrand's ballot problem was first introduced by Joseph Bertrand in 1887, in the form of: "In an election where candidate A receives p votes and candidate B receives q votes with $p > q$, what is the probability that A will be strictly ahead of B throughout the count?"

J. Bertrand himself gave the solution $\frac{p-q}{p+q}$ by using mathematical induction in the original paper. First of all, let's consider the initial case. At the first count, the vote can be either for candidate A or B. If it *could* be for candidate A, the simplest scenario is $p = 1, q = 0$, which is of probability 1 for candidate A to win the vote. This agrees with the formula. If it *could* be for candidate B, the simplest scenario is $p = 2, q = 1$, now there are $\binom{2+1}{1} = 3$ counting orders, i.e., AAB, ABA, or BAA. Note "AAA" is the only favorable order out of three possible orders. This again agrees with the formula. Assume the theorem is true when $p = a - 1$ and $q = b$ (last vote would be for candidate A), and when $p = a$ and $q = b - 1$ (last vote would be for candidate B), which is the two possible scenarios at the second to last count. Now, considering the case with $p = a$ and $q = b$, the last vote is either for candidate A with probability $a/(a + b)$, or for candidate B with probability $b/(a + b)$. So the probability of candidate A to always lead the count is:

$$\frac{a}{a+b}\frac{(a-1)-b}{(a-1)+b} + \frac{b}{a+b}\frac{a-(b-1)}{a+(b-1)} = \frac{a-b}{a+b}$$

The word "could" means it is possible for an event to happen, but does not indicate its necessity.

$\frac{(a-1)-b}{(a-1)+b}$ and $\frac{a-(b-1)}{a+(b-1)}$ are conditional probabilities, which are conditioned on the last count.

This proves that the theorem is true for all $p > q \geq 0$.

Désiré André in the same year gave an elegant proof of this problem, and the method is now called ”the principle of reflection”, or ”the André's reflection method”, or ”the Bertrand's ballot theorem”. There are three facts need to noted. One is that, the sequences start with A or B with probability $p/(p+q)$ and $q/(p+q)$, respectively. The second fact is that, sequences start with B will for sure to tie at some point because A will finally win, and they are all unfavorable, because A already ”loses” at the first count. The third is that, sequences that start with A can be classified into two cases, one is the case that A leads the counting process from beginning to the end which is the favorable case, the other is the case when the sequence will tie at some point which is the unfavorable case, importantly, the number of sequences in this latter case is the same as that of the sequences starting with B, because there is a bijection mapping. The probability is $1 - 2 \times \frac{q}{p+q} = \frac{p-q}{p+q}$.

One possible mapping is to denote the sequence as LBR, where ”L” and ”R” is the left and right part of the first tie position (must be a ”B”), and then converting each character in L to its alternative, e.g., AABBABAA would become BBABABAA.

The number of the unfavorable cases is $2 \times \binom{p+q-1}{q-1}$, of which half starts with ”A” and another half starts with ”B”, as explained above. The number of favorable cases is $\binom{p+q-1}{p-1} - \binom{p+q-1}{q-1}$. Actually, $\frac{\binom{p+q-1}{p-1}-\binom{p+q-1}{q-1}}{\binom{p+q}{p}} = \frac{p-q}{p+q}$.

Consider now the problem to find the probability that the second candidate is never ahead (i.e. ties are allowed); the solution is $\frac{p+1-q}{p+1}$. This is simply seen by awaring the following equivalent description:

- same as the basic version, ties are NOT allowed; but,
- there are $p+1$ votes for candidate A and q votes for candidate B;
- the first vote is for candidate A;

The probability can then be computed as:

$$\begin{aligned} P(\text{A winning } \textit{with} \text{ ties}) &= P(\text{A winning } \textit{without} \text{ ties} \mid \text{the first vote is A}) \\ &= \frac{P(\text{A winning without ties } \textit{and} \text{ the first vote is A})}{P(\text{the first vote is A})} \\ &= \frac{(p+1-q)/(p+1+q)}{(p+1)/(p+1+q)} = \frac{p+1-q}{p+1} \end{aligned}$$

Another way to look at this problem is to model it as the following: represent a voting sequence as a lattice path on the Cartesian plane and,

- Start the path at (0, 0);
- Each time a vote for the first candidate is received move right 1 unit;
- Each time a vote for the second candidate is received move up 1 unit.

Each such path corresponds to a unique sequence of votes and will end at (p, q). A sequence is “good” exactly when the corresponding path never goes above the diagonal line $y = x$; equivalently, a sequence is “bad” exactly when the corresponding path touches the line $y = x + 1$. For each “bad” path P, define a new path P' by reflecting the part of P up to the first point it touches the line across it. P' is a path from (-1, 1) to (p, q). The same operation applied again restores the original P. This produces a one-to-one correspondence between the "bad" paths and the paths from (-1, 1) to (p, q). The number of these paths is $\binom{p+q}{q-1}$. So the probability asked is $\frac{\binom{p+q}{q}-\binom{p+q}{q-1}}{\binom{p+q}{p}} = \frac{p+1-q}{p+1}$.

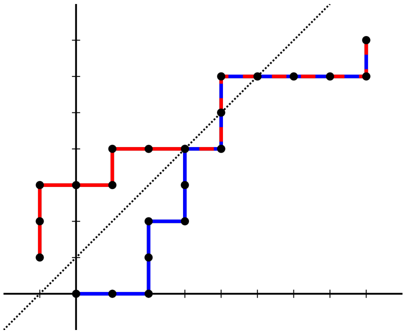

Figure 1: Bertrand's Ballot Problem allowing Ties

An interesting application of this is the famous Catalan number formula, which can be introduced under the random walk story. A random walk on the integers is to take n steps of unit length, beginning at the origin and ending at the point m, that never become negative. Assuming n and m have the same parity and $n \geq m \geq 0$, this number is, according to the Bertrand's Ballot problem allowing ties,

$$\binom{n}{\frac{n+m}{2}} - \binom{n}{\frac{n+m}{2}+1} = \frac{m+1}{\frac{n+m}{2}+1}\binom{n}{\frac{n+m}{2}}$$

Here, $p+q=n$ and $p-q=m$, compared to our used settings. When $m=0$ and n is even, this gives the Catalan number $\frac{1}{\frac{n}{2}+1}\binom{n}{\frac{n}{2}}$.

Let's tweak this problem a little bit more. Let's say candidate A starts at a "bonus" votes, not 0. That is to say, the system goes from $(0,a)$ to $(p+q, p+a-q)$. If we do not allow ties, all paths hit the x-axis will be unfavorable, and the number of these paths equals the number of paths from $(0,a)$ to its "mirror" point $(p+q, -p-a+q)$. So, there are in total $\binom{p+q}{p+a}$ unfavorable paths, and the probability is $1-\frac{\binom{p+q}{p+a}}{\binom{p+q}{p}}$. Note when $a=0$, there is already a tie, and the question really should be asked as: "What if the first count is A, and then the process never has a tie". So the probability should compute as:

$$1-\frac{p}{p+q}\frac{\binom{p-1+q}{p-1}}{\binom{p-1+q}{p-1}} = \frac{p-q}{p+q}$$

It should have no problem with $a>0$.

If we allow ties, the "mirror" point should be reflected against $y=-1$, so it becomes $(p+q, -2-p-a+q)$. Now, suppose we go up u steps and go down d

steps. Solving the following equations:

$$\begin{aligned} u+d &= p+q \\ u+a-d &= -2-p-a+q \end{aligned}$$

gives us $u = q-a-1$ and $d = p+a+1$. So there are $\binom{p+q}{p+a+1}$ paths unfavorable, and the probability is then $1 - \frac{\binom{p+q}{p+a+1}}{\binom{p+q}{p}}$. When $a = 0$, this becomes $\frac{p+1-q}{p+1}$, which agrees with what we obtained earlier.

In summary, if one starts at $(0, a)$ and ends at $(p+q, p+a-q)$, the number of unfavorable paths is $\binom{p+q}{p+a}$ if not allowing ties, $\binom{p+q}{p+a+1}$ if allowing ties.

1.5.4 Catalan Number

In section 1.5.3, we first met Catalan number from the generalized Bertrand's Ballot Problem. We write here again the definition of Catalan number, with an intuitive interpretation.

The Catalan number is defined as,

$$C_n = \frac{1}{n+1}\binom{2n}{n} \tag{18}$$

The underlying story reads: Given two urns, one with n red balls and the other with n black balls, we want to draw one ball at a time (either red or black), such that at no time the number of pre-specified color is less than its alternative.

Since the Catalan number is associated with two equal-sized sets, it is oftentimes co-occurrent with the words "pair", "full binary", and etc.

2 Probability

2.1 Axioms

2.1.1 Law of Total Probability

2.1.2 Law of Total Variance

2.1.3 Law of Total Covariance

2.1.4 Law of Total Expectation

2.1.5 Law of Total Cumulance

2.1.6 Probability Inequalities

2.2 Definitions of Sample Space, Events, and Probability

Definition 2.1 A sample space S is a set of all possible outcomes of an experiment. An outcome is also called a sample point.

Note, when an experiment consists of several repetitions, each one of them is called a *trial*. As an example, if one decides to toss a coin 42 times, we can call each toss a trial of the experiment composed of 42 ones.

Definition 2.2 An event E is a subset of the sample space S. If the outcome of the experiment is contained in E, then we say that E occurred.

Definition 2.3 In short, a probability space is a measure space such that the measure of the whole space is equal to one. The expanded definition is following: a probability space is a triple $(\Omega,\ \mathcal{F},\ P)$ consisting of:

- the sample space Ω — an arbitrary non-empty set,
- the $\sigma - algebra$ $\mathcal{F} \in 2^{\Omega}$ (also called $\sigma - filed$) — a set of subsets of Ω, called events, such that:
  - $\mathcal{F}$ contains the empty set: $\emptyset \in \mathcal{F}$,
  - $\mathcal{F}$ is closed under complements: if $A \in \mathcal{F}$, then also $(\Omega \setminus A) \in \mathcal{F}$,
  - $\mathcal{F}$ is closed under countable unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then also $(\bigcup A_i) \in \mathcal{F}$,
- the probability measure $P : \mathcal{F} \to [0, 1]$ — a function on $\mathcal{F}$ such that:
- P is countably additive: if $\{A_i\} \in \mathcal{F}$ is a countable collection of pairwise disjoint sets, then $P(\bigcup A_i) = \sum P(A_i)$, where $\bigcup$ denotes the disjoint union,
- the measure of entire sample space is equal to one: $P(\Omega) = 1$.

2.3 Types of Probabilities: Frequentism and Bayesian

2.4 Probability Redefined: Formal Definition with Measure Theory

3 Conditional Probability and Independence

Definition 3.1 If $P(F) > 0$, then

$$P(E|F) = \frac{P(E, F)}{P(F)} \tag{19}$$

$P(E|F)$ is called the conditional probability of E given F. Conditional probability agrees with Definition 2.3, and should be treated in the same way.

Definition 3.2 The multiplication rule

$$P(E_1, E_2, \ldots, E_n) = P(E_1)P(E_2|E_1) \ldots P(E_n|E_1, \ldots, E_{n-1}) \tag{20}$$

Definition 3.3 Bayes' Formula

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \tag{21}$$

where $P(A_i)$ is sometimes called the prior distribution, $P(B|A_i)$ the likelihood function, and $P(A_i|B)$ the posterior distribution. The partition function $P(B)$ can be computed using the law of total probability

$$P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j) \tag{22}$$

Definition 3.4 Two events E and F are said to be *independent* if

$$P(E, F) = P(E)P(F) \tag{23}$$

A set of events are independent if every finite subset of these events is independent.

3.1 Conditional Probability

3.2 Conditional Expectation

3.3 Conditional Independence

4 Random Variables

The Laplace distribution can be written as an infinite mixture of Gaussians with variance w distribution accorindg to an exponential distribution. An exponential distribution can be written as a χ^2 distribution with two degrees of freedom.

4.1 Continuous Random Variables

4.2 Discrete Random Variables

4.3 Joint Distributed Random Variables

4.4 Random Vectors/Matrices

4.5 Stochastic Processes

4.6 Random Field

4.7 Convolutions: Sum of Normally Distributed Random Variables

4.8 Product Distribution

4.9 Ratio Distribution

5 Quantitative Measure and Characteristic Functions

5.1 Describing Shape of a Distribution: Skewness, Kurtosis

5.2 Describing a Sample: Mean, Variance

5.3 Degrees of Freedom, Mean, Variance, and Moment, Central Moment, Cumulant, Law of the unconscious statistician

5.4 Percentile and Median

5.5 Coefficient of Variation

5.6 Covariance and Correlation

5.7 Moment Generating Function

5.8 Characteristic Function

6 Limit Theorems

7 Selected Topics in Probability

7.1 Ordered Statistics

7.2 Copula

8 Summary of Common Distributions

Part II: Statistics

9 Collecting Data: Experiments and Surveys

9.1 Design of Experiments

1. A causes B.
2. B causes A.
3. A and B both partly cause each other.
4. A and B are both caused by a third factor, C.
5. The observed correlation was due purely to chance.

15 Statistics Theory

15.1 Mathematical Statistics

15.1.1 Degrees of Freedom

15.1.2 sufficient, complete, and etc.

15.1.3 Likelihood Function

15.1.4 Exponential Family

15.2 Statistical Learning Theory

15.3 Interpretations of Probability

15.3.1 Cox's Theorem

15.3.2 Principle of Maximum Entropy