Carleton University
Department of Systems and
Computer Engineering

Stochastic Processes
94.553 Fall 2000
A. H. Banihashemi

Chapter 1 Summary

(An Introduction to the Notion of Probability)

1 Random experiments

An *experiment* is a procedure which results in an *outcome.* The outcome of an experiment depends on the conditions under which the experiment is performed. In a *deterministic* experiment, the observed result is not subject to chance. If we repeat a deterministic experiment under exactly the same conditions, we expect the same result. In a *random* experiment the outcome is always subject to chance. If the experiment is repeated the outcome may be different.

Many random experiments can be defined in terms of an "urn model". The urn contains a number of balls, which are identified in some way, for example, by a color scheme or by a numbering scheme. Consider an urn containing three balls, labeled 0, 1, and 2.

An urn experiment is defined as follows.

1. Shake the urn.
2. Draw a ball from the urn (without looking into the urn, of course).
3. Record the ball label.
4. Return the ball to the urn.

The last step corresponds to the situation in which we perform "sampling with replacement;" this is in contrast to the situation in which we do "sampling without replacement," and don't return the ball to the urn.

Assume that the "urn experiment" (with replacement) is performed seven times, and the following sequence of outcomes are obtained:

$$\{2, 0, 0, 0, 0, 1, 2\}. \tag{1}$$

2 Outcomes and Sample Spaces

In a given random experiment, the result of the experiment is called the *outcome* of the experiment. The set of all possible outcomes is called the *sample space* corresponding to the given experiment[1], and is often denoted as $\mathcal{S}$. In the urn experiment, the sample space $\mathcal{S}$ is given by

$$\mathcal{S} = \{0, 1, 2\}.$$

[1] Sometimes it might be convenient to include impossible outcomes in a given sample space. This poses no problem: we simply assign zero probability to these outcomes. (See later in the course.)

In a die-rolling experiment, the sample space $\mathcal{S}$ can be denoted by

$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}.$$

3 Counting the Outcomes in a Sequence of Trials, and Relative Frequency

Suppose, now, that we perform a random experiment with sample space $\mathcal{S}$. If we perform this experiment n times, i.e., we perform n trials, we get a sequence of outcomes

$$\{o_1, o_2, o_3, \ldots, o_n\}. \tag{2}$$

We denote by $N_k(n)$ the number of times outcome k occurs in the list (2) for all $k \in \mathcal{S}$.

For example, in the urn experiment, $\mathcal{S} = \{0, 1, 2\}$, and the sequence of outcomes was given in (1). For this particular sequence of seven trials,

$$N_0(7) = 4, \quad N_1(7) = 1, \quad N_2(7) = 2.$$

A couple of observations about the numbers $N_k(n)$ can be made:

$$0 \leq N_k(n) \leq n, \tag{3}$$

$$\sum_{k \in \mathcal{S}} N_k(n) = n. \tag{4}$$

Often, we are interested in the *proportion* of trials that had a given outcome k. This number is called the *relative frequency* of outcome k, and is defined as

$$f_k(n) = \frac{1}{n} N_k(n).$$

In the urn experiment,

$$f_0(7) = 4/7, \quad f_1(7) = 1/7, \quad f_2(7) = 2/7.$$

The notion of relative frequency is an important one in probability; in fact, the axioms of probability are modeled based on the properties of relative frequency, some of which are noted below.

4 Events

An outcome is an element of the sample space; an *event* is a subset of the sample space. Thus, an event is a collection of outcomes. Events allow us to group together several outcomes into a single set of interest. For example, in a die-rolling experiment, the event "the outcome is even" is the set $\{2, 4, 6\}$. In the urn experiment, the event "the outcome is even" is the set $E = \{0, 2\}$.

The following table lists all of the $2^{|\mathcal{S}|}$ events associated with the urn experiment:

$\{0, 1, 2\}$	the sample space itself is an event;
$\{0, 1\}, \{0, 2\}, \{1, 2\}$	events with two outcomes, including event E;
$\{0\}, \{1\}, \{2\}$	"singletons" are events;
$\phi = \{\}$	the empty or "null event" is useful, but never occurs.

Just as we can count the number of times that a particular outcome occurs, we can count the number of times that a particular event occurs. If A is an event, we denote by $N_A(n)$ the number of times that A occurs in n trials. Similarly,

$$f_A(n) = \frac{1}{n} N_A(n)$$

denotes the relative frequency of occurrence of the event A. In the urn experiment $f_E(7) = 6/7$. Note that we always have $f_{\mathcal{S}}(n) = 1$ and $f_\phi(n) = 0$.

5 Statistical Regularity

Even though we cannot predict the outcome of a given trial, we would notice after many trials that the relative frequencies of the various outcomes in the urn experiment are quite close to 1/3. Intuitively, if the experiment were to be performed n times, we would expect that approximately $n/3$ of the outcomes would be 0, approximately $n/3$ of the outcomes would be 1, and approximately $n/3$ of the outcomes would be 2. (Usually, not *exactly* $n/3$, though.) Furthermore, one would expect that, as n becomes large, the relative frequencies would "settle down" to 1/3. This phenomenon, in which the relative frequencies "converge" to a fixed value is called *statistical regularity.*

The phenomenon of statistical regularity can be observed experimentally. Figs. 1.3 and 1.4 of the textbook illustrate the relative frequencies of the various outcomes after n trials, for various values of n. One can see a "convergence" of the relative frequencies to the value of 1/3, for large values of n. This "limiting value" of the relative frequency of a particular outcome is called the *probability* of that outcome. That is, (for now) we define

$$p_k = \lim_{n \to \infty} f_k(n)$$

to be the probability of the outcome k, without worrying too much about how this limit can be carried out[2]. Since we think of probabilities as the limiting case of relative frequencies, probabilities will share many of the properties of relative frequencies.

In the urn experiment, $p_0 = p_1 = p_2 = 1/3$. On the other hand, if the urn experiment were changed so that the urn initially contained two balls labeled 0, one ball labeled 1, and one ball labeled 2, we would have $p_0 = 1/2$, and $p_1 = p_2 = 1/4$. Thus, changing the experimental conditions changes the probabilities.

6 Properties of Relative Frequency

Let k be a typical outcome in a sample space $\mathcal{S}$, and let $f_k(n) = N_k(n)/n$ be the relative frequency of outcome k after n trials of the random experiment. We can list a few of the most important properties of $f_k(n)$.

1. $0 \leq f_k(n) \leq 1$.
2. $\sum_{k \in \mathcal{S}} f_k(n) = 1$.

[2]In fact, it can't be carried out, since it is impossible to carry out an infinite number of trials. To get around this, we define probabilities *axiomatically* later in the course.

3. Let A be the event $\{i\} \cup \{j\}$, where i and j are two *different* outcomes from the sample space, i.e., $i \neq j$. Then

$$f_A(n) = f_i(n) + f_j(n).$$

3′. Property 3 can be generalized as follows. Let A and B be two *disjoint* events, i.e., two events having no outcomes in common ($A \cap B = \phi$), and let C be the event $A \cup B$. Then

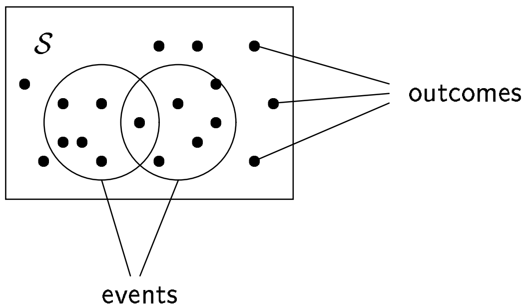
$$f_C(n) = f_A(n) + f_B(n).$$

Chapter 2 Summary

(Basic Concepts of Probability Theory)

7 Sample Spaces

- sample space $\mathcal{S}$
- outcomes or elementary events: $\in \mathcal{S}$
- events: $\subset \mathcal{S}$
- $\mathcal{S}$ finite, countable infinite = discrete
- $\mathcal{S}$ uncountably infinite = continuous



8 Set Operations

- union ($A \cup B$), intersection ($A \cap B$), complement ($A^c = \mathcal{S} \setminus A$), containment or "$A$ implies B" ($A \subset B$)
- commutativity and associativity of union and intersection
- distributive laws
- DeMorgan's rules
- mutually exclusive (disjoint) events: $A \cap B = \phi$

9 The Axioms of Probability

A probability law is a real-valued function P defined on the set of events which satisfies

I. $0 \leq P[A]$, for any event A;

II. $P[\mathcal{S}] = 1$;

III. given events $A_1, A_2, \ldots$ with $A_i \cap A_j = \phi$, $i \neq j$, then $P[\bigcup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$. In particular, if $A \cap B = \phi$, then $P[A \cup B] = P[A] + P[B]$.

Corollaries

1. $P[A^c] = 1 - P[A]$
2. $P[A] \leq 1$
3. $P[\phi] = 0$
4. If $A_1, \ldots, A_n$ is a collection of pairwise disjoint events, then

$$P[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} P[A_i].$$

5. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ (see page 34 of the textbook for generalization).
6. (Union bound) $P[A \cup B] \leq P[A] + P[B]$.
7. If $A \subset B$, then $P[A] \leq P[B]$.

10 Probability Assignments

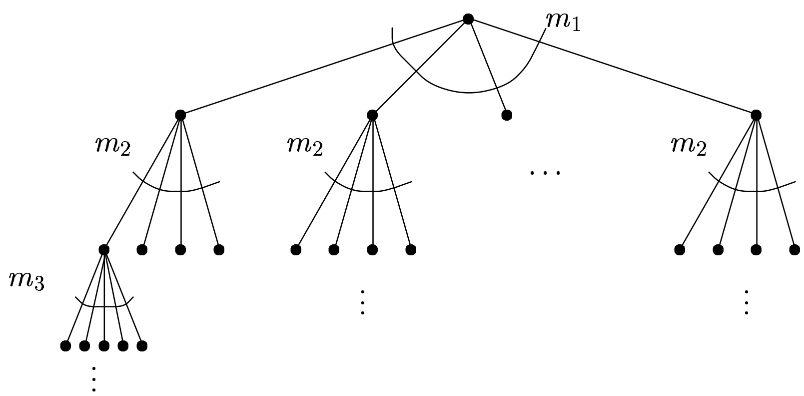
- Discrete Sample Spaces: probabilities of elementary events (events containing a single outcome) determine all the other probabilities. In many cases, the outcomes are *equally likely.* In such cases,

$$P(A) = \frac{|A|}{|\mathcal{S}|}, \tag{5}$$

  where $|A|$ denotes the cardinality of A.
- Continuous Sample Spaces: probabilities are assigned to intervals of real line, or regions in the plane, or more generally, n-dimensional regions in R^n.

11 Counting

Fundamental Rule:



A number of multiple choices are to be made. There are m_1 possibilities for the first choice, m_2 for the second, m_3 for the third, etc. If these choices can be combined freely, then the total number of possibilities for the whole set of choices is equal to

$$m_1 \times m_2 \times m_3 \times \cdots.$$

Sampling problems—Sample k from a population of n distinct objects.

1. Sampling with replacement and with ordering:

$$n^k \text{ distinct ordered } k\text{-tuples.}$$

2. Sampling without replacement and with ordering:

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!} \text{ distinct ordered } k\text{-tuples.}$$

   This is also denoted by P_k^n or $(n)_k$, referred to as the permutation of k objects out of n (distinct) objects. For $k = n$, we get the no. of permutations of n objects which is equal to $n!$.

3. Sampling without replacement and without ordering (partitioning n distinct objects into two subsets of sizes k and $n-k$): "n choose k"

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ distinct possibilities.}$$

   $\binom{n}{k}$ is called the *binomial coefficient.* Note that $\binom{n}{k}$ is equal to the number of subsets of size k from a set of size n. It is also denoted by C_k^n, referred to as the combination of k objects out of n (distinct) objects.

   Binomial theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

$$a = b = 1 \quad \rightarrow \quad \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

$$a = -1, b = 1 \quad \rightarrow \quad \sum_{\substack{k=0 \\ k \text{ even}}}^{n} \binom{n}{k} = \sum_{\substack{k=1 \\ k \text{ odd}}}^{n} \binom{n}{k} = 2^{n-1}$$

   Partitioning n distinct objects into ξ subsets $B_1, \ldots, B_\xi$, where $|B_i| = k_i, i = 1, \ldots, \xi$, and $k_1 + \ldots + k_\xi = n$:

$$\frac{n!}{k_1! k_2! \cdots k_\xi!}\,. \tag{6}$$

   Equation (6) is called the *multinomial coefficient.*

4. Sampling with replacement and without ordering:

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

   This is the number of subsets of size k from a set of size n, where in subsets, the elements are allowed to occur with multiplicity greater than one. This is also equivalent to the number of different arrangements of k 'x's and $n-1$ '/'s (slash symbols).

12 Conditional Probability

A and B are events, $P[B] > 0$. *Given* that B occurred, what is the probability that A occurred? Answer:

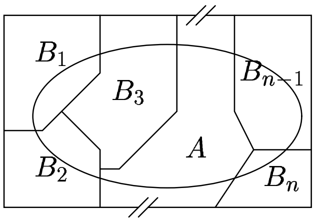
$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

Note that $P[B|A] = \frac{P[A \cap B]}{P[A]}$, hence

$$P[A \cap B] = P[A|B]P[B] = P[B|A]P[A].$$

Suppose $B_1, B_2, \ldots, B_n$ are disjoint events that *partition* a sample space $\mathcal{S}$, i.e., $B_i \cap B_j = \phi, \;\; i \neq j$, and $\cup_{i=1}^{n} B_i = \mathcal{S}$. Then, for any event A, the "total probability theorem" states

$$\begin{aligned} P[A] &= P[A|B_1]P[B_1] + P[A|B_2]P[B_2] + \cdots + P[A|B_n]P[B_n] \\ &= \sum_{i=1}^{n} P[A|B_i]P[B_i]. \end{aligned}$$



Illustrating the total probability theorem.

Bayes' Rule:

$$\begin{aligned} P[B_j|A] &= \frac{P[B_j \cap A]}{P[A]} \\ &= \frac{P[A|B_j]P[B_j]}{\sum_{i=1}^{n} P[A|B_i]P[B_i]} \end{aligned}$$

13 Independence of Events

A and B are *independent* if and only if $P[A \cap B] = P[A]P[B]$. (For the generalization to more than two events, see page 59 of the textbook). If $P[B] \neq 0$ and A and B are independent, then

$$P[A|B] = P[A],$$

i.e., occurrence of B does not affect $P[A]$ (and vice-versa).

Clearly, independent experiments result in independent events.

14 Sequential Experiments

Sequences of Independent Experiments

Bernoulli trial: perform an experiment once and note if a particular event A occurs. Define "SUCCESS" if A occurs and "FAILURE" if A does not occur. Let $P[A] = P[\text{SUCCESS}] = p$; then $P[\text{FAILURE}] = 1 - p$.

Binomial Probability Law

Perform n independent Bernoulli trials and note the number of successes. The probability of having exactly k successes is

$$p_n(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Multinomial Probability Law

Let $B_1, \ldots, B_M$ be a partition of the sample space $\mathcal{S}$ of some random experiment and let $P[B_i] = p_i$. Thus $p_1 + \cdots + p_M = 1$. Perform n independent repetition of the experiment. Let k_i be the number of times event B_i occurs. The probability of having the vector $(k_1, \ldots, k_M)$ is

$$P[(k_1, \ldots, k_M)] = \frac{n!}{k_1! k_2! \cdots k_M!} p_1^{k_1} p_2^{k_2} \cdots p_M^{k_M} \;,$$

where $k_i \geq 0, \forall i$, and $k_1 + k_2 + \cdots + k_M = n$.

Geometric Probability Law

Repeat independent Bernoulli trials until the first success and record the number of trials needed. Then the probability that exactly m trials are needed is

$$\begin{aligned} p(m) &= p(1-p)^{m-1}, \quad m = 1, 2, 3, \ldots, \quad \text{and,} \\ P[m > K] &= (1-p)^K \end{aligned}$$

Sequences of Dependent Experiments

Let $A_1, \ldots, A_n$ denotes a sequence of dependent events. It is usually helpful to use the following *chain rule* to compute the probability of the sequence.

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) P(A_2|A_1) \cdots P(A_n|A_1 \cap \ldots \cap A_{n-1}) \,.$$

Often, sequences of dependent experiments can be described via the "state" of the experiment, in which the probabilities of the experimental outcomes depend only on the state. Such sequences are called *Markov chains*. For a Markov chain, a "trellis diagram" can be used to compute the probability of any sequence of outcomes.

Acknowledgement: I would like to thank Professor Frank Kschischang from University of Toronto for providing me with an initial version of summaries for Chapters 1 to 4 of the textbook.