INTRODUCTION TO THE MATHEMATICAL AND STATISTICAL FOUNDATIONS OF ECONOMETRICS

HERMAN J. BIERENS

Pennsylvania State University

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2005

Printed in the United States of America

Typeface Times New Roman PS 10/12 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Bierens, Herman J., 1943–

Introduction to the mathematical and statistical foundations of econometrics / Herman J. Bierens.

p. cm. – (Themes in modern econometrics)

Includes bibliographical references and index.

ISBN 0-521-83431-7 – ISBN 0-521-54224-3 (pb.)

1. Econometrics. I. Title. II. Series.

HB139.B527 2004
330′.01′5195 – dc22 2004040792

ISBN 0 521 83431 7 hardback
ISBN 0 521 54224 3 paperback

Contents

1 Probability and Measure

1.1. The Texas Lotto

1.1.1. Introduction

Texans used to play the lotto by selecting six different numbers between 1 and 50, which cost \$1 for each combination.[1] Twice a week, on Wednesday and Saturday at 10 P.M., six ping-pong balls were released without replacement from a rotating plastic ball containing 50 ping-pong balls numbered 1 through 50. The winner of the jackpot (which has occasionally accumulated to 60 or more million dollars!) was the one who had all six drawn numbers correct, where the order in which the numbers were drawn did not matter. If these conditions were still being observed, what would the odds of winning by playing one set of six numbers only?

To answer this question, suppose first that the order of the numbers does matter. Then the number of *ordered* sets of 6 out of 50 numbers is 50 possibilities for the first drawn number times 49 possibilities for the second drawn number, times 48 possibilities for the third drawn number, times 47 possibilities for the fourth drawn number, times 46 possibilities for the fifth drawn number, times 45 possibilities for the sixth drawn number:

$$\prod_{j=0}^{5}(50-j) = \prod_{k=45}^{50} k = \frac{\prod_{k=1}^{50} k}{\prod_{k=1}^{50-6} k} = \frac{50!}{(50-6)!}.$$

[1] In the spring of 2000, the Texas Lottery changed the rules. The number of balls was increased to fifty-four to create a larger jackpot. The official reason for this change was to make playing the lotto more attractive because a higher jackpot makes the lotto game more exciting. Of course, the actual intent was to boost the lotto revenues!

The notation $n!$, read "n factorial," stands for the product of the natural numbers 1 through n:

$$n! = 1 \times 2 \times \cdots \times (n-1) \times n \quad \text{if } n > 0, \qquad 0! = 1.$$

The reason for defining $0! = 1$ will be explained in the next section.

Because a set of six given numbers can be permutated in 6! ways, we need to correct the preceding number for the 6! replications of each unordered set of six given numbers. Therefore, the number of sets of six *unordered* numbers out of 50 is

$$\binom{50}{6} \stackrel{\text{def.}}{=} \frac{50!}{6!(50-6)!} = 15{,}890{,}700.$$

Thus, the probability of winning such a lotto by playing only one combination of six numbers is $1/15{,}890{,}700$.[2]

1.1.2. Binomial Numbers

In general, the number of ways we can draw a set of k *unordered* objects out of a set of n objects *without* replacement is

$$\binom{n}{k} \stackrel{\text{def.}}{=} \frac{n!}{k!(n-k)!}. \tag{1.1}$$

These (binomial) numbers,[3] read as "n choose k," also appear as coefficients in the binomial expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{1.2}$$

The reason for defining $0! = 1$ is now that the first and last coefficients in this binomial expansion are always equal to 1:

$$\binom{n}{0} = \binom{n}{n} = \frac{n!}{0!n!} = \frac{1}{0!} = 1.$$

For not too large an n, the binomial numbers (1.1) can be computed recursively by hand using the *Triangle of Pascal*:

2 Under the new rules (see Note 1), this probability is $1/25{,}827{,}165$.

3 These binomial numbers can be computed using the "Tools → Discrete distribution tools" menu of *EasyReg International*, the free econometrics software package developed by the author. *EasyReg International* can be downloaded from Web page http://econ.la.psu.edu/~hbierens/EASYREG.HTM

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & \dots & & \dots & & \dots & & \dots & & \dots & & 1
\end{array} \tag{1.3}
$$

Except for the 1's on the legs and top of the triangle in (1.3), the entries are the sum of the adjacent numbers on the previous line, which results from the following easy equality:

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k} \quad \text{for} \quad n \geq 2,\ k = 1, \dots, n-1. \tag{1.4}$$

Thus, the top 1 corresponds to $n = 0$, the second row corresponds to $n = 1$, the third row corresponds to $n = 2$, and so on, and for each row $n + 1$, the entries are the binomial numbers (1.1) for $k = 0, \dots, n$. For example, for $n = 4$ the coefficients of $a^k b^{n-k}$ in the binomial expansion (1.2) can be found on row 5 in (1.3): $(a + b)^4 = 1 \times a^4 + 4 \times a^3 b + 6 \times a^2 b^2 + 4 \times ab^3 + 1 \times b^4$.

1.1.3. Sample Space

The Texas lotto is an example of a statistical experiment. The set of possible outcomes of this statistical experiment is called the *sample space* and is usually denoted by Ω. In the Texas lotto case, Ω contains $N = 15{,}890{,}700$ elements: $\Omega = \{\omega_1, \dots, \omega_N\}$, where each element ω_j is a set itself consisting of six different numbers ranging from 1 to 50 such that for any pair ω_i, ω_j with $i \neq j$, $\omega_i \neq \omega_j$. Because in this case the elements ω_j of Ω are sets themselves, the condition $\omega_i \neq \omega_j$ for $i \neq j$ is equivalent to the condition that $\omega_i \cap \omega_j \notin \Omega$.

1.1.4. Algebras and Sigma-Algebras of Events

A set $\{\omega_{j_1}, \dots, \omega_{j_k}\}$ of different number combinations you can bet on is called an *event*. The collection of all these events, denoted by $\mathscr{F}$, is a "family" of subsets of the sample space Ω. In the Texas lotto case the collection $\mathscr{F}$ consists of all subsets of Ω, including Ω itself and the empty set $\emptyset$.[4] In principle, you could bet on all number combinations if you were rich enough (it would cost you \$15,890,700). Therefore, the sample space Ω itself is included in $\mathscr{F}$. You could also decide not to play at all. This event can be identified as the empty set $\emptyset$. For the sake of completeness, it is included in $\mathscr{F}$ as well.

[4] Note that the latter phrase is superfluous because $\Omega \subset \Omega$ signifies that every element of Ω is included in Ω, which is clearly true, and $\emptyset \subset \Omega$ is true because $\emptyset \subset \emptyset \cup \Omega = \Omega$.

Because, in the Texas lotto case, the collection $\mathscr{F}$ contains all subsets of Ω, it automatically satisfies the conditions

$$\text{If } A \in \mathscr{F} \quad \text{then} \quad \tilde{A} = \Omega \backslash A \in \mathscr{F}, \tag{1.5}$$

where $\tilde{A} = \Omega\backslash A$ is the *complement* of the set A (relative to the set Ω), that is, the set of all elements of Ω that are not contained in A, and

$$\text{If } A, B \in \mathscr{F} \quad \text{then} \quad A \cup B \in \mathscr{F}. \tag{1.6}$$

By induction, the latter condition extends to any finite union of sets in $\mathscr{F}$: If $A_j \in \mathscr{F}$ for $j = 1, 2, \ldots, n$, then $\cup_{j=1}^{n} A_j \in \mathscr{F}$.

Definition 1.1: *A collection $\mathscr{F}$ of subsets of a nonempty set Ω satisfying the conditions (1.5) and (1.6) is called an algebra.*[5]

In the Texas lotto example, the sample space Ω is finite, and therefore the collection $\mathscr{F}$ of subsets of Ω is finite as well. Consequently, in this case the condition (1.6) extends to

$$\text{If } A_j \in \mathscr{F} \quad \text{for} \quad j = 1, 2, 3, \ldots \quad \text{then} \quad \bigcup_{j=1}^{\infty} A_j \in \mathscr{F}. \tag{1.7}$$

However, because in this case the collection $\mathscr{F}$ of subsets of Ω is finite, there are only a finite number of distinct sets $A_j \in \mathscr{F}$. Therefore, in the Texas lotto case the countable infinite union $\cup_{j=1}^{\infty} A_j$ in (1.7) involves only a finite number of distinct sets A_j; the other sets are replications of these distinct sets. Thus, condition (1.7) does not require that all the sets $A_j \in \mathscr{F}$ are different.

Definition 1.2: *A collection $\mathscr{F}$ of subsets of a nonempty set Ω satisfying the conditions (1.5) and (1.7) is called a σ-algebra.*[6]

1.1.5. Probability Measure

Let us return to the Texas lotto example. The odds, or probability, of winning are $1/N$ for each valid combination ω_j of six numbers; hence, if you play n different valid number combinations $\{\omega_{j_1}, \ldots, \omega_{j_n}\}$, the probability of winning is n/N: $P(\{\omega_{j_1}, \ldots, \omega_{j_n}\}) = n/N$. Thus, in the Texas lotto case the probability $P(A)$, $A \in \mathscr{F}$, is given by the number n of elements in the set A divided by the total number N of elements in Ω. In particular we have $P(\Omega) = 1$, and if you do not play at all the probability of winning is zero: $P(\emptyset) = 0$.

[5] Also called a *field*.
[6] Also called a *σ-field* or a *Borel field*.

The function $P(A)$, $A \in \mathscr{F}$, is called a probability measure. It assigns a number $P(A) \in [0, 1]$ to each set $A \in \mathscr{F}$. Not every function that assigns numbers in $[0, 1]$ to the sets in $\mathscr{F}$ is a probability measure except as set forth in the following definition:

Definition 1.3: *A mapping P: $\mathscr{F} \rightarrow [0, 1]$ from a σ-algebra $\mathscr{F}$ of subsets of a set Ω into the unit interval is a probability measure on $\{\Omega, \mathscr{F}\}$ if it satisfies the following three conditions:*

$$\text{For all } A \in \mathscr{F},\ P(A) \geq 0, \tag{1.8}$$

$$P(\Omega) = 1, \tag{1.9}$$

$$\text{For disjoint sets } A_j \in \mathscr{F},\ P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j). \tag{1.10}$$

Recall that sets are *disjoint* if they have no elements in common: their intersections are the empty set.

The conditions (1.8) and (1.9) are clearly satisfied for the case of the Texas lotto. On the other hand, in the case under review the collection $\mathscr{F}$ of events contains only a finite number of sets, and thus any countably infinite sequence of sets in $\mathscr{F}$ must contain sets that are the same. At first sight this seems to conflict with the implicit assumption that countably infinite sequences of *disjoint* sets always exist for which (1.10) holds. It is true indeed that any countably infinite sequence of disjoint sets in a finite collection $\mathscr{F}$ of sets can only contain a finite number of nonempty sets. This is no problem, though, because all the other sets are then equal to the empty set $\emptyset$. The empty set is disjoint with itself, $\emptyset \cap \emptyset = \emptyset$, and with any other set, $A \cap \emptyset = \emptyset$. Therefore, if $\mathscr{F}$ is finite, then any countable infinite sequence of disjoint sets consists of a finite number of nonempty sets and an infinite number of replications of the empty set. Consequently, if $\mathscr{F}$ is finite, then it is sufficient to verify condition (1.10) for any pair of disjoint sets A_1, A_2 in $\mathscr{F}$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. Because, in the Texas lotto case $P(A_1 \cup A_2) = (n_1 + n_2)/N$, $P(A_1) = n_1/N$, and $P(A_2) = n_2/N$, where n_1 is the number of elements of A_1 and n_2 is the number of elements of A_2, the latter condition is satisfied and so is condition (1.10).

The statistical experiment is now completely described by the triple $\{\Omega, \mathscr{F}, P\}$, called the *probability space*, consisting of the sample space Ω (i.e., the set of all possible outcomes of the statistical experiment involved), a σ-algebra $\mathscr{F}$ of events (i.e., a collection of subsets of the sample space Ω such that the conditions (1.5) and (1.7) are satisfied), and a probability measure P: $\mathscr{F} \rightarrow [0, 1]$ satisfying the conditions (1.8)–(1.10).

In the Texas lotto case the collection $\mathscr{F}$ of events is an algebra, but because $\mathscr{F}$ is finite it is automatically a σ-algebra.

1.2. Quality Control

1.2.1. Sampling without Replacement

As a second example, consider the following case. Suppose you are in charge of quality control in a light bulb factory. Each day N light bulbs are produced. But before they are shipped out to the retailers, the bulbs need to meet a minimum quality standard such as not allowing more than R out of N bulbs to be defective. The only way to verify this exactly is to try all the N bulbs out, but that will be too costly. Therefore, the way quality control is conducted in practice is to randomly draw n bulbs *without* replacement and to check how many bulbs in this sample are defective.

As in the Texas lotto case, the number M of different samples s_j of size n you can draw out of a set of N elements without replacement is

$$M = \binom{N}{n}.$$

Each sample s_j is characterized by a number k_j of defective bulbs in the sample involved. Let K be the actual number of defective bulbs. Then $k_j \in \{0, 1, \ldots, \min(n, K)\}$.

Let $\Omega = \{0, 1, \ldots, n\}$ and let the σ-algebra $\mathscr{F}$ be the collection of all subsets of Ω. The number of samples s_j with $k_j = k \leq \min(n, K)$ defective bulbs is

$$\binom{K}{k}\binom{N-K}{n-k}$$

because there are "K choose k" ways to draw k unordered numbers out of K numbers without replacement and "$N - K$ choose $n - k$" ways to draw $n - k$ unordered numbers out of $N - K$ numbers without replacement. Of course, in the case that $n > K$ the number of samples s_j with $k_j = k > \min(n, K)$ defective bulbs is zero. Therefore, let

$$P(\{k\}) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{if } 0 \leq k \leq \min(n, K),$$
$$P(\{k\}) = 0 \text{ elsewhere,} \tag{1.11}$$

and for each set $A = \{k_1, \ldots, k_m\} \in \mathscr{F}$, let $P(A) = \sum_{j=1}^{m} P(\{k_j\})$. (*Exercise*: Verify that this function P satisfies all the requirements of a probability measure.) The triple $\{\Omega, \mathscr{F}, P\}$ is now the probability space corresponding to this statistical experiment.

The probabilities (1.11) are known as the *hypergeometric* (N, K, n) probabilities.

1.2.2. Quality Control in Practice[7]

The problem in applying this result in quality control is that K is unknown. Therefore, in practice the following decision rule as to whether $K \leq R$ or not is followed. Given a particular number $r \leq n$, to be determined at the end of this subsection, assume that the set of N bulbs meets the minimum quality requirement $K \leq R$ if the number k of defective bulbs in the sample is less than or equal to r. Then the set $A(r) = \{0, 1, \ldots, r\}$ corresponds to the assumption that the set of N bulbs meets the minimum quality requirement $K \leq R$, hereafter indicated by "accept," with probability

$$P(A(r)) = \sum_{k=0}^{r} P(\{k\}) = p_r(n, K), \tag{1.12}$$

say, whereas its complement $\tilde{A}(r) = \{r + 1, \ldots, n\}$ corresponds to the assumption that this set of N bulbs does not meet this quality requirement, hereafter indicated by "reject," with corresponding probability

$$P(\tilde{A}(r)) = 1 - p_r(n, K).$$

Given r, this decision rule yields two types of errors: a Type I error with probability $1 - p_r(n, K)$ if you reject, whereas in reality $K \leq R$, and a Type II error with probability $p_r(K, n)$ if you accept, whereas in reality $K > R$. The probability of a Type I error has upper bound

$$p_1(r, n) = 1 - \min_{K \leq R} p_r(n, K), \tag{1.13}$$

and the probability of a Type II error upper bound

$$p_2(r, n) = \max_{K > R} p_r(n, K). \tag{1.14}$$

To be able to choose r, one has to restrict either $p_1(r, n)$ or $p_2(r, n)$, or both. Usually it is the former option that is restricted because a Type I error may cause the whole stock of N bulbs to be trashed. Thus, allow the probability of a Type I error to be a maximal α such as $\alpha = 0.05$. Then r should be chosen such that $p_1(r, n) \leq \alpha$. Because $p_1(r, n)$ is decreasing in r, due to the fact that (1.12) is increasing in r, we could in principle choose r arbitrarily large. But because $p_2(r, n)$ is increasing in r, we should not choose r unnecessarily large. Therefore, choose $r = r(n|\alpha)$, where $r(n|\alpha)$ is the minimum value of r for which $p_1(r, n) \leq \alpha$. Moreover, if we allow the Type II error to be maximal β, we have to choose the sample size n such that $p_2(r(n|\alpha), n) \leq \beta$.

As we will see in Chapters 5 and 6, this decision rule is an example of a statistical test, where $H_0 : K \leq R$ is called the null hypothesis to be tested at

[7] This section may be skipped.

the $\alpha \times 100\%$ significance level against the alternative hypothesis $H_1 : K > R$. The number $r(n|\alpha)$ is called the critical value of the test, and the number k of defective bulbs in the sample is called the test statistic.

1.2.3. Sampling with Replacement

As a third example, consider the quality control example in the previous section except that now the light bulbs are sampled *with* replacement: After a bulb is tested, it is put back in the stock of N bulbs even if the bulb involved proves to be defective. The rationale for this behavior may be that the customers will at most accept a fraction R/N of defective bulbs and thus will not complain as long as the actual fraction K/N of defective bulbs does not exceed R/N. In other words, why not sell defective light bulbs if doing so is acceptable to the customers?

The sample space Ω and the σ-algebra $\mathscr{F}$ are the same as in the case of sampling without replacement, but the probability measure P is different. Consider again a sample s_j of size n containing k defective light bulbs. Because the light bulbs are put back in the stock after being tested, there are K^k ways of drawing an *ordered* set of k defective bulbs and $(N-K)^{n-k}$ ways of drawing an *ordered* set of $n-k$ working bulbs. Thus, the number of ways we can draw, with replacement, an ordered set of n light bulbs containing k defective bulbs is $K^k(N-K)^{n-k}$. Moreover, as in the Texas lotto case, it follows that the number of *unordered* sets of k defective bulbs and $n-k$ working bulbs is "n choose k." Thus, the total number of ways we can choose a sample with replacement containing k defective bulbs and $n-k$ working bulbs in any order is

$$\binom{n}{k} K^k(N-K)^{n-k}.$$

Moreover, the number of ways we can choose a sample of size n with replacement is N^n. Therefore,

$$\begin{aligned} P(\{k\}) &= \binom{n}{k} \frac{K^k(N-K)^{n-k}}{N^n} \\ &= \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n, \end{aligned} \tag{1.15}$$

where $p = K/N$, and again for each set $A = \{k_1, \ldots, k_m\} \in \mathscr{F}$, $P(A) = \sum_{j=1}^m P(\{k_j\})$. Of course, if we replace $P(\{k\})$ in (1.11) by (1.15), the argument in Section 1.2.2 still applies.

The probabilities (1.15) are known as the *binomial* (n, p) probabilities.

1.2.4. Limits of the Hypergeometric and Binomial Probabilities

Note that if N and K are large relative to n, the hypergeometric probability (1.11) and the binomial probability (1.15) will be almost the same. This follows from the fact that, for fixed k and n,

$$P(\{k\}) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\frac{K!(N-K)!}{K!(K-k)!(n-k)!(N-K-n+k)!}}{\frac{N!}{n!(N-n)!}}$$

$$= \frac{n!}{k!(n-k)!} \times \frac{\frac{K!(N-K)!}{(K-k)!(N-K-n+k)!}}{\frac{N!}{(N-n)!}}$$

$$= \binom{n}{k} \times \frac{\frac{K!}{(K-k)!} \times \frac{(N-K)!}{(N-K-n+k)!}}{\frac{N!}{(N-n)!}}$$

$$= \binom{n}{k} \times \frac{\left(\prod_{j=1}^{k}(K-k+j)\right) \times \left(\prod_{j=1}^{n-k}(N-K-n+k+j)\right)}{\prod_{j=1}^{n}(N-n+j)}$$

$$= \binom{n}{k} \times \frac{\left[\prod_{j=1}^{k}\left(\frac{K}{N} - \frac{k}{N} + \frac{j}{N}\right)\right] \times \left[\prod_{j=1}^{n-k}\left(1 - \frac{K}{N} - \frac{n}{N} + \frac{k}{N} + \frac{j}{N}\right)\right]}{\prod_{j=1}^{n}\left(1 - \frac{n}{N} + \frac{j}{N}\right)}$$

$$\to \binom{n}{k} p^k (1-p)^{n-k} \quad \text{if } N \to \infty \quad \text{and} \quad K/N \to p.$$

Thus, the binomial probabilities also arise as limits of the hypergeometric probabilities.

Moreover, if in the case of the binomial probability (1.15) p is very small and n is very large, the probability (1.15) can be approximated quite well by the Poisson(λ) probability:

$$P(\{k\}) = \exp(-\lambda)\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots, \tag{1.16}$$

where $\lambda = np$. This follows from (1.15) by choosing $p = \lambda/n$ for $n > \lambda$, with $\lambda > 0$ fixed, and letting $n \to \infty$ while keeping k fixed:

$$P(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} (\lambda/n)^k (1-\lambda/n)^{n-k} = \frac{\lambda^k}{k!} \times \frac{n!}{n^k(n-k)!}$$

$$\times \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k} \to \exp(-\lambda)\frac{\lambda^k}{k!} \quad \text{for} \quad n \to \infty,$$

because for $n \to \infty$,

$$\frac{n!}{n^k(n-k)!} = \frac{\prod_{j=1}^{k}(n-k+j)}{n^k} = \prod_{j=1}^{k}\left(1 - \frac{k}{n} + \frac{j}{n}\right) \to \prod_{j=1}^{k} 1 = 1$$

$$(1-\lambda/n)^k \to 1$$

and

$$(1 - \lambda/n)^n \to \exp(-\lambda). \tag{1.17}$$

Due to the fact that (1.16) is the limit of (1.15) for $p = \lambda/n \downarrow 0$ as $n \to \infty$, the Poisson probabilities (1.16) are often used to model the occurrence of *rare* events.

Note that the sample space corresponding to the Poisson probabilities is $\Omega = \{0, 1, 2, \ldots\}$ and that the σ-algebra $\mathscr{F}$ of events involved can be chosen to be the collection of *all* subsets of Ω because any nonempty subset A of Ω is either countable infinite or finite. If such a subset A is countable infinite, it takes the form $A = \{k_1, k_2, k_3, \ldots\}$, where the k_j's are distinct nonnegative integers; hence, $P(A) = \sum_{j=1}^{\infty} P(\{k_j\})$ is well-defined. The same applies of course if A is finite: if $A = \{k_1, \ldots, k_m\}$, then $P(A) = \sum_{j=1}^{m} P(\{k_j\})$. This probability measure clearly satisfies the conditions (1.8)–(1.10).

1.3. Why Do We Need Sigma-Algebras of Events?

In principle we could define a probability measure on an algebra $\mathscr{F}$ of subsets of the sample space rather than on a σ-algebra. We only need to change condition (1.10) as follows: For disjoint sets $A_j \in \mathscr{F}$ such that $\cup_{j=1}^{\infty} A_j \in \mathscr{F}$, $P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$. By letting all but a finite number of these sets be equal to the empty set, this condition then reads as follows: For disjoint sets $A_j \in \mathscr{F}, j = 1, 2, \ldots, n < \infty$, $P(\cup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P(A_j)$. However, if we confined a probability measure to an algebra, all kinds of useful results would no longer apply. One of these results is the so-called strong law of large numbers (see Chapter 6).

As an example, consider the following game. Toss a fair coin infinitely many times and assume that after each tossing you will get one dollar if the outcome is heads and nothing if the outcome is tails. The sample space Ω in this case can be expressed in terms of the winnings, that is, each element ω of Ω takes the form of a string of infinitely many zeros and ones, for example, $\omega = (1, 1, 0, 1, 0, 1 \ldots)$. Now consider the event: "After n tosses the winning is k dollars." This event corresponds to the set $A_{k,n}$ of elements ω of Ω for which the sum of the first n elements in the string involved is equal to k. For example, the set $A_{1,2}$ consists of all ω of the type $(1, 0, \ldots)$ and $(0, 1, \ldots)$. As in the example in Section 1.2.3, it can be shown that

$$P(A_{k,n}) = \binom{n}{k}(1/2)^n \quad \text{for} \quad k = 0, 1, 2, \ldots, n,$$
$$P(A_{k,n}) = 0 \quad \text{for} \quad k > n \text{ or } k < 0.$$

Next, for $q = 1, 2, \ldots$, consider the events after n tosses the average winning k/n is contained in the interval $[0.5 - 1/q, 0.5 + 1/q]$. These events correspond to the sets $B_{q,n} = \cup_{k=[n/2-n/q)]+1}^{[n/2+n/q]} A_{k,n}$, where $[x]$ denotes the smallest integer $\geq x$. Then the set $\cap_{m=n}^{\infty} B_{q,m}$ corresponds to the following event:

From the nth tossing onwards the average winning will stay in the interval $[0.5 - 1/q, 0.5 + 1/q]$; the set $\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}$ corresponds to the event there exists an n (possibly depending on ω) such that from the nth tossing onwards the average winning will stay in the interval $[0.5 - 1/q, 0.5 + 1/q]$. Finally, the set $\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}$ corresponds to the event the average winning converges to $1/2$ as n converges to infinity. Now the strong law of large numbers states that the latter event has probability 1: $P[\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}] = 1$. However, this probability is only defined if $\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m} \in \mathscr{F}$. To guarantee this, we need to require that $\mathscr{F}$ be a σ-algebra.

1.4. Properties of Algebras and Sigma-Algebras

1.4.1. General Properties

In this section I will review the most important results regarding algebras, σ-algebras, and probability measures.

Our first result is trivial:

Theorem 1.1: *If an algebra contains only a finite number of sets, then it is a σ-algebra. Consequently, an algebra of subsets of a finite set Ω is a σ-algebra.*

However, an algebra of subsets of an *infinite* set Ω is not necessarily a σ-algebra. A counterexample is the collection $\mathscr{F}_*$ of all subsets of $\Omega = (0, 1]$ of the type $(a, b]$, where $a < b$ are *rational* numbers in $[0, 1]$ together with their *finite* unions and the empty set $\emptyset$. Verify that $\mathscr{F}_*$ is an algebra. Next, let $p_n = [10^n \pi]/10^n$ and $a_n = 1/p_n$, where $[x]$ means truncation to the nearest integer $\leq x$. Note that $p_n \uparrow \pi$; hence, $a_n \downarrow \pi^{-1}$ as $n \to \infty$. Then, for $n = 1, 2, 3, \ldots, (a_n, 1] \in \mathscr{F}_*$, but $\cup_{n=1}^{\infty}(a_n, 1] = (\pi^{-1}, 1] \notin \mathscr{F}_*$ because π^{-1} is irrational. Thus, $\mathscr{F}_*$ is *not* a σ-algebra.

Theorem 1.2: *If $\mathscr{F}$ is an algebra, then $A, B \in \mathscr{F}$ implies $A \cap B \in \mathscr{F}$; hence, by induction, $A_j \in \mathscr{F}$ for $j = 1, \ldots, n < \infty$ implies $\cap_{j=1}^{n} A_j \in \mathscr{F}$. A collection $\mathscr{F}$ of subsets of a nonempty set Ω is an algebra if it satisfies condition (1.5) and the condition that, for any pair $A, B \in \mathscr{F}$, $A \cap B \in \mathscr{F}$.*

Proof: Exercise.

Similarly, we have

Theorem 1.3: *If $\mathscr{F}$ is a σ-algebra, then for any countable sequence of sets $A_j \in \mathscr{F}$, $\cap_{j=1}^{\infty} A_j \in \mathscr{F}$. A collection $\mathscr{F}$ of subsets of a nonempty set Ω is a σ-algebra if it satisfies condition (1.5) and the condition that, for any countable sequence of sets $A_j \in \mathscr{F}$, $\cap_{j=1}^{\infty} A_j \in \mathscr{F}$.*