THE ART OF
COMPUTER PROGRAMMING
VOLUME 4  PRE-FASCICLE 5A

MATHEMATICAL
PRELIMINARIES
REDUX

DONALD E. KNUTH  Stanford University

ADDISON–WESLEY
April 8, 2014

See also http://www-cs-faculty.stanford.edu/~knuth/sgb.html for information about The Stanford GraphBase, including downloadable software for dealing with the graphs used in many of the examples in Chapter 7.

See also http://www-cs-faculty.stanford.edu/~knuth/mmixware.html for downloadable software to simulate the MMIX computer.

Copyright © 2013 by Addison–Wesley

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior consent of the publisher, except that the official electronic file may be used to print single copies for personal (not commercial) use.

Zeroth printing (revision 11), 06 April 2014

April 8, 2014
PREFACE

We—or the Black Chamber—have a little agreement with [Knuth]:
he doesn't publish the real Volume 4 of The Art of Computer Programming,
and they don't render him metabolically challenged.
— CHARLES STROSS, The Atrocity Archive (2001)

This booklet contains draft material that I'm circulating to experts in the
field, in hopes that they can help remove its most egregious errors before too
many other people see it. I am also, however, posting it on the Internet for
courageous and/or random readers who don't mind the risk of reading a few
pages that have not yet reached a very mature state. Beware: This material
has not yet been proofread as thoroughly as the manuscripts of Volumes 1, 2,
3, and 4A were at the time of their first printings. And those carefully-checked
volumes, alas, were subsequently found to contain thousands of mistakes.

Given this caveat, I hope that my errors this time will not be so numerous
and/or obtrusive that you will be discouraged from reading the material carefully.
I did try to make the text both interesting and authoritative, as far as it goes.
But the field is vast; I cannot hope to have surrounded it enough to corral it
completely. So I beg you to let me know about any deficiencies that you discover.

To put the material in context, this pre-fascicle contains an exposition of
mathematical material (mostly about probability theory) that I plan to include
at the beginning of Volume 4B. Its raison d'être is explained below, in an excerpt
from the preface to that volume.

* * *

Probability theory has made huge strides since I “completed” my college
education in 1963; hence I'm basically self-taught with respect to these new-
fangled ideas, and I fear that in many respects my knowledge lags behind that
of today's students. I've tried my best to get the story right, yet I fear that in
many respects I'm woefully ignorant.

For example, I urgently need your help with respect to some exercises that I
made up as I was preparing this material. I certainly don't like to receive credit
for things that have already been published by others, and most of these results
are quite natural “fruits” that were just waiting to be “plucked.” Therefore
please tell me if you know who deserves to be credited, with respect to the ideas
found in exercises 6, 9, 10, 19, 32, 33, 38, 73, 88, and 96.
* * *

Special thanks are due to Persi Diaconis, Svante Janson, Sheldon Ross, Ernst Schulte-Geers, and ... for their detailed comments on my early attempts at exposition, as well as to numerous other correspondents who have contributed crucial corrections.

* * *

I happily offer a “finder’s fee” of $2.56 for each error in this draft when it is first reported to me, whether that error be typographical, technical, or historical. The same reward holds for items that I forgot to put in the index. And valuable suggestions for improvements to the text are worth 32¢ each. (Furthermore, if you find a better solution to an exercise, I’ll actually do my best to give you immortal glory, by publishing your name in the eventual book:-)

Cross references to yet-unwritten material sometimes appear as ‘00’; this impossible value is a placeholder for the actual numbers to be supplied later.

Happy reading!

Stanford, California

21 October 2012

D. E. K.
Part of the Preface to Volume 4B

During the years that I’ve been preparing Volume 4, I’ve often run across basic techniques of probability theory that I would have put into Section 1.2 of Volume 1 if I’d been clairvoyant enough to anticipate them in the 1960s. Finally I realized that I ought to collect most of them together in one place, near the beginning of Volume 4B, because the story of these developments is too interesting to be broken up into little pieces scattered here and there.

Therefore this volume begins with a special section entitled “Mathematical Preliminaries Redux,” and future sections use the abbreviation ‘MPR’ to refer to its equations and its exercises.
In books of this nature I can only suggest you keep it as simple as the subject will allow.

— Kode Vicious (2012)
MATHEMATICAL PRELIMINARIES REDUX

Many parts of this book deal with discrete probabilities, namely with a finite or countably infinite set $\Omega$ of atomic events $\omega$, each of which has a given probability $\Pr(\omega)$, where

$$0 \leq \Pr(\omega) \leq 1 \quad \text{and} \quad \sum_{\omega \in \Omega} \Pr(\omega) = 1. \quad (1)$$

This set $\Omega$, together with the function $\Pr$, is called a “probability space.” For example, $\Omega$ might be the set of all ways to shuffle a pack of 52 playing cards, with $\Pr(\omega) = 1/52!$ for every such arrangement.

An event is, intuitively, a proposition that can be either true or false with certain probability. It might, for instance, be the statement “the top card is an ace,” with probability $1/13$. Formally, an event $A$ is a subset of $\Omega$, namely the set of all atomic events for which the corresponding proposition $A$ is true; and

$$\Pr(A) = \sum_{\omega \in A} \Pr(\omega) = \sum_{\omega \in \Omega} \Pr(\omega)[\omega \in A]. \quad (2)$$

A random variable is a function that assigns a value to every atomic event. We typically use uppercase letters for random variables, and lowercase letters for the values that they might assume; thus, we might say that the probability of the event $X = x$ is $\Pr(X = x) = \sum_{\omega \in \Omega} \Pr(\omega)[X(\omega) = x]$. In our playing card example, the top card $T$ is a random variable, and we have $\Pr(T = Q \spadesuit) = 1/52$. (Sometimes, as here, the lowercase-letter convention is ignored.)

The random variables $X_1, \ldots, X_k$ are said to be independent if

$$\Pr(X_1 = x_1 \text{ and } \cdots \text{ and } X_k = x_k) = \Pr(X_1 = x_1) \cdots \Pr(X_k = x_k) \quad (3)$$

for all $(x_1, \ldots, x_k)$. For example, if $F$ and $S$ denote the face value and suit of the top card, $T$, clearly $F$ and $S$ are independent. Hence in particular we have $\Pr(T = Q \spadesuit) = \Pr(F = Q) \Pr(S = \spadesuit)$. But $T$ is not independent of the bottom card, $B$; indeed, we have $\Pr(T = t \text{ and } B = b) \neq 1/52^2$ for any cards $t$ and $b$.

A system of $n$ random variables is called $k$-wise independent if no $k$ of its variables are dependent. With pairwise (2-wise) independence, for example, we could have variable $X$ independent of $Y$, variable $Y$ independent of $Z$, and variable $Z$ independent of $X$; yet all three variables needn’t be independent (see exercise 6). Similarly, $k$-wise independence does not imply $(k + 1)$-wise independence. But $(k + 1)$-wise independence does imply $k$-wise independence.

The conditional probability of an event $A$, given an event $B$, is

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A \text{ and } B)}{\Pr(B)}, \quad (4)$$

April 8, 2014
when $\Pr(B) > 0$, otherwise it’s $\Pr(A)$. Imagine breaking the whole probability space $\Omega$ into two parts, $\Omega = B$ and $\Omega' = \Omega \setminus B$, with $\Pr(\Omega') = \Pr(B)$ and $\Pr(\Omega') = 1 - \Pr(B)$. If we assign new probabilities to atomic events by the rules

$$
\Pr'(\omega) = \frac{\Pr(\omega) \mid \omega \in \Omega'}{\Pr(\Omega')}, \quad \Pr''(\omega) = \frac{\Pr(\omega) \mid \omega \in \Omega''}{\Pr(\Omega'')},
$$

we obtain new probability spaces $\Omega'$ and $\Omega''$, allowing us to contemplate a world where $B$ is always true and another world where $B$ is always false. It’s like taking two branches in a tree, each of which has its own logic. Conditional probability is important for the analysis of algorithms because algorithms often get into different states where different probabilities are relevant. Notice that we always have

$$
\Pr(A) = \Pr(A \mid B) \cdot \Pr(B) + \Pr(A \mid \Omega') \cdot \Pr(\Omega').
$$

(5)

The events $A_1, \ldots, A_k$ are said to be independent if the random variables $[A_1], \ldots, [A_k]$ are independent. (Bracket notation applies in the usual way to events-as-statements, not just to events-as-subsets: $[A] = 1$ if $A$ is true, otherwise $[A] = 0$.) Exercise 20 proves that this happens if and only if

$$
\Pr\left( \bigcap_{j \in J} A_j \right) = \prod_{j \in J} \Pr(A_j), \quad \text{for all } J \subseteq \{1, \ldots, k\}. \quad (6)
$$

In particular, events $A$ and $B$ are independent if and only if $\Pr(A \mid B) = \Pr(A)$.

When the values of a random variable $X$ are real numbers or complex numbers, we’ve defined its expected value $E X$ in Section 1.2.10: We said that

$$
E X = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega) = \sum_x x \Pr(X = x),
$$

provided that this definition makes sense when the sums are taken over infinitely many nonzero values. A simple but extremely important case arises when $A$ is any event, and when $X = [A]$ is a binary random variable representing the truth of that event; then

$$
E[A] = \sum_{\omega \in \Omega} [A](\omega) \Pr(\omega) = \sum_{\omega \in \Omega} [\omega \in A] \Pr(\omega) = \sum_{\omega \in A} \Pr(\omega) = \Pr(A).
$$

(8)

We’ve also noted that the expectation of a sum, $E(X_1 + \cdots + X_k)$, always equals the sum of the expectations, $(E X_1) + \cdots + (E X_k)$, whether or not the random variables $X_j$ are independent. Furthermore the expectation of a product, $E X_1 \cdots X_k$, is the product of the expectations, $(E X_1) \cdots (E X_k)$, if those variables do happen to be independent. In Section 3.3.2 we defined the covariance,

$$
covar(X, Y) = E((X - E X)(Y - E Y)) = (E X Y) - (E X)(E Y),
$$

(9)

which tends to measure the way $X$ and $Y$ depend on each other. The variance, $\text{var}(X)$, is $\text{covar}(X, X)$; the middle formula in (9) shows why it is nonnegative whenever the random variable $X$ takes on only real values.

All of these notions of expected value carry over to conditional expectation,

$$
E(X \mid A) = \sum_{\omega \in A} X(\omega) \frac{\Pr(\omega)}{\Pr(A)} = \sum_x x \frac{\Pr(X = x \text{ and } A)}{\Pr(A)},
$$

(10)

April 8, 2014
conditioned on any event \( A \), when we want to work in the probability space for which \( A \) is true. One of the most important formulas, analogous to (5), is
\[
E[X] = \sum_y E(X \mid Y = y) \Pr(Y = y)
\]
\[
= \sum_y \sum_x x \Pr(X = x \mid Y = y) \Pr(Y = y).
\]

(11)

Furthermore there's also another important kind of conditional expectation: When \( X \) and \( Y \) are random variables, it's often helpful to write \( E[X \mid Y] \) for "the expectation of \( X \) given \( Y \)." Using that notation, Eq. (11) becomes simply
\[
E[X] = E(E(X \mid Y)).
\]

(12)

This is a truly marvelous identity, great for hand-waving and for impressing outsiders—except that it can be confusing until you understand what it means.

In the first place, if \( Y \) is a Boolean variable, \( E(X \mid Y) \) might look as if it means \( E(X \mid Y=1) \), thus asserting that \( Y \) is true, just as \( E(X \mid A) \) asserts the truth of \( A \) in (10). Wrong.

In the second place, you might think of \( E(X \mid Y) \) as a function of \( Y \). Well, yes; but the best way to understand \( E(X \mid Y) \) is to regard it as a random variable. That's why we're allowed to compute its expected value in (12).

All random variables are functions of the atomic events \( \omega \). The value of \( E(X \mid Y) \) at \( \omega \) is the average of \( X(\omega') \) over all events \( \omega' \) such that \( Y(\omega') = Y(\omega) \):
\[
E(X \mid Y)(\omega) = \sum_{\omega' \in \Omega} X(\omega') \Pr(\omega') \mid Y(\omega') = Y(\omega) / \Pr(Y = Y(\omega)).
\]

(13)

Similarly, \( E(X_1, \ldots, X_r) \) averages over events with \( Y_j(\omega') = Y_j(\omega) \) for \( 1 \leq j \leq r \).

For example, suppose \( X_1 \) through \( X_n \) are binary random variables constrained by the condition that \( \nu(X_1 \ldots X_n) = X_1 + \cdots + X_n = m \), where \( m \) and \( n \) are constants with \( 0 \leq m \leq n \); all \( \binom{n}{m} \) such bit vectors \( X_1 \ldots X_n \) are assumed to be equally likely. Clearly \( E(X_1) = m/n \). But what is \( E(X_2 \mid X_1) \)? If \( X_1 = 0 \), the expectation of \( X_2 \) is \( m/(n-1) \); otherwise that expectation is \( (m-1)/(n-1) \); consequently \( E(X_2 \mid X_1) = (m-X_1)/(n-1) \). And what is \( E(X_3 \mid X_1, X_2) \)? The answer is easy, once you get used to the notation: if \( \nu(X_1 \ldots X_{k-1}) = r \), then \( X_k \ldots X_n \) is a random bit vector with \( \nu(X_k \ldots X_n) = m-r \); hence the average value of \( X_k \) will be \( (m-r)/(n+1-k) \) in that case. We conclude that
\[
E(X_k \mid X_1, \ldots, X_{k-1}) = \frac{m - \nu(X_1 \ldots X_{k-1})}{n+1-k}, \quad \text{for } 1 \leq k \leq n.
\]

(14)

The random variables on both sides of these equations are the same.

**Inequalities.** In practice we often want to prove that certain events are rare, in the sense that they occur with very small probability. Conversely, our goal is sometimes to show that an event is not rare. And we're in luck, because mathematicians have devised several fairly easy ways to derive upper bounds or lower bounds on probabilities, even when the exact values are unknown.
We’ve already discussed the most important technique of this kind in Section 1.2.10. Stated in highly general terms, the basic idea can be formulated as follows: Let $f$ be any nonnegative function such that $f(x) \geq s > 0$ when $x \in S$. Then
\[
\Pr(X \in S) \leq \frac{E(f(X))}{s},
\]
provided that $\Pr(X \in S)$ and $E(f(X))$ both exist. For example, $f(x) = |x|$ yields
\[
\Pr(|X| \geq m) \leq \frac{E(|X|)}{m}
\]
whenever $m > 0$. The proof is amazingly simple, because we obviously have
\[
E(f(X)) \geq \Pr(X \in S) \cdot s + \Pr(X \notin S) \cdot 0.
\]
Formula (15) is often called Markov’s inequality, because A. A. Markov discussed the special case $f(x) = |x|^2$ in Izvĕstĭa Imp. Akad. Nauk (6) 1 (1907), 707–716. If we set $f(x) = (x - E(X))^2$, we get the famous 19th-century inequality of Bienaymé and Chebyshev:
\[
\Pr(|X - E(X)| \geq r) \leq \frac{\text{var}(X)}{r^2}.
\]
The case $f(x) = e^{ax}$ is also extremely useful.

Another fundamental estimate, known as Jensen’s inequality [Acta Mathematica 30 (1906), 175–193], applies to convex functions $f$; we’ve seen it so far only as a “hint” to exercise 6.2.2–36(1). The real-valued function $f$ is said to be convex in an interval $I$ of the real line, and $-f$ is said to be concave in $I$, if
\[
f(px + qy) \leq pf(x) + qf(y)
\]
whenever $p \geq 0, q \geq 0$, and $p+q = 1$. This condition turns out to be equivalent to saying that $f''(x) \geq 0$ for all $x \in I$, if $f$ has a second derivative $f''$. For example, the functions $e^{ax}$ and $x^{2n}$ are convex for all constants $a$ and all nonnegative integers $n$; and if we restrict consideration to positive values of $x$, then $f(x) = x^n$ is convex for all integers $n$ (notably $f(x) = 1/x$ when $n = -1$). The functions $\ln(1/x)$ and $x \ln x$ are also convex for $x > 0$. Jensen’s inequality states that
\[
f(E(X)) \leq E(f(X))
\]
when $f$ is convex in the interval $I$ and the random variable $X$ takes values only in $I$. (See exercise 42 for a proof!) For example, we have $1/E(X) \leq E(1/X)$ and $\ln E(X) \geq E \ln X$ and $(E(X) \ln X \leq E(X \ln X)$, when $X$ is positive. Notice that (20) actually reduces to the very definition of convexity, (19), in the special case when $X = x$ with probability $p$ and $X = y$ with probability $q$.

Third and fourth on our list of remarkably useful inequalities are two classical results that apply to any random variable $X$ whose values are nonnegative integers:
\[
\Pr(X > 0) \leq E(X); \quad \text{("the first moment principle")}
\]
\[
\Pr(X > 0) \geq (E(X))^2/(E(X)^2). \quad \text{("the second moment principle")}
\]
Formula (21) is obvious, because the left side is $p_1 + p_2 + p_3 + \cdots$ when $p_k$ is the probability that $X = k$, while the right side is $p_1 + 2p_2 + 3p_3 + \cdots$.

April 8, 2014
Formula (22) isn’t quite so obvious; it is \( p_1 + p_2 + p_3 + \cdots \) on the left and 
\[ (p_1 + 2p_2 + 3p_3 + \cdots)^2 / (p_1 + 4p_2 + 9p_3 + \cdots) \] on the right. However, as we saw with Markov’s inequality, there is a remarkably simple proof, once we happen to discover it:

\[
E X^2 = E (X^2 | X > 0) \Pr(X > 0) + E (X^2 | X = 0) \Pr(X = 0) \\
= E (X^2 | X > 0) \Pr(X > 0) \\
\geq (E (X | X > 0))^2 \Pr(X > 0) = (E X)^2 / \Pr(X > 0). \tag{23}
\]

In fact this proof shows that the second moment principle is valid even when \( X \) is not restricted to integer values (see exercise 46). Furthermore the argument can be strengthened to show that (22) holds even when \( X \) can take arbitrary negative values, provided only that \( E X > 0 \) (see exercise 47).

Exercise 54 applies (21) and (22) to the study of random graphs.

Another important inequality, which applies in the special case where \( X = X_1 + \cdots + X_m \) is the sum of binary random variables \( X_j \), was introduced more recently by S. M. Ross [Probability, Statistics, and Optimization (New York: Wiley, 1994), 185–190], who calls it the “conditional expectation inequality”:

\[
\Pr(X > 0) \geq \sum_{j=1}^{m} \frac{E X_j}{E (X | X_j = 1)}. \tag{24}
\]

Ross showed that the right-hand side of this inequality is always at least as big as the bound \((E X)^2 / (E X^2)\) that we get from the second moment principle (see exercise 50). Furthermore, (24) is often easier to compute, even though it may look more complicated at first glance.

For example, his method applies nicely to the problem of estimating a reliability polynomial \( f(p_1, \ldots, p_n) \), when \( f \) is a monotone Boolean function; here \( p_j \) represents the probability that component \( j \) of a system is “up.” We observed in Section 7.1.4 that reliability polynomials can be evaluated exactly, using BDD methods, when \( n \) is reasonably small; but approximations are necessary when \( f \) gets complicated. The simple example \( f(x_1, \ldots, x_5) = x_1 x_2 x_3 \lor x_2 x_3 x_4 \lor x_4 x_5 \) illustrates Ross’s general method: Let \( Y_1, \ldots, Y_5 \) be independent binary random variables, with \( E Y_j = p_j \); and let \( X = X_1 + X_2 + X_3 \), where \( X_1 = Y_1 Y_2 Y_3 \), \( X_2 = Y_2 Y_3 Y_4 \), and \( X_3 = Y_4 Y_5 \) correspond to the prime implicates of \( f \). Then

\[
\Pr(X > 0) = \Pr(f(Y_1, \ldots, Y_5) = 1) = E f(Y_1, \ldots, Y_5) = f(p_1, \ldots, p_5),
\]

because the \( Y_j \)’s are independent. And we can evaluate the bound in (24) easily:

\[
\Pr(X > 0) \geq \frac{p_1 p_2 p_3}{1 + p_4 + p_4 p_6} + \frac{p_2 p_3 p_4}{p_1 + 1 + p_6} + \frac{p_4 p_6}{p_1 p_2 p_3 + p_2 p_6 + 1}. \tag{25}
\]

If, for example, each \( p_j \) is 0.9, this formula gives \( \approx 0.848 \), while \((E X)^2 / (E X^2) \approx 0.847\); the true value, \( p_1 p_2 p_3 + p_2 p_3 p_4 + p_4 p_6 - p_1 p_2 p_4 - p_1 p_3 p_4 - p_2 p_3 p_6 \), is 0.9558.

Many other important inequalities relating to expected values have been discovered, of which the most significant for our purposes in this book is the FKG inequality discussed in exercise 61. It yields easy proofs that certain events are correlated, as illustrated in exercise 62.
Martingales. A sequence of dependent random variables can be difficult to analyze, but if those variables obey invariant constraints we can often exploit their structure. In particular, the “martingale” property, named after a classic betting strategy (see exercise 67), proves to be amazingly useful when it applies. Joseph L. Doob featured martingales in his pioneering book *Stochastic Processes* (New York: Wiley, 1953), and developed their extensive theory.

The sequence \( \langle Z_n \rangle = Z_0, Z_1, Z_2, \ldots \) of real-valued random variables is called a **martingale** if it satisfies the condition

\[
E(Z_{n+1} \mid Z_0, \ldots, Z_n) = Z_n \quad \text{for all } n \geq 0. \tag{26}
\]

(We also implicitly assume, as usual, that the expectations \( E Z_n \) are well defined.) For example, when \( n = 0 \), the random variable \( E(Z_1 \mid Z_0) \) must be the same as the random variable \( Z_0 \) (see exercise 63).

Figure 1 illustrates George Pólya’s famous “urn model” [F. Eggenberger and G. Pólya, *Zeitschrift für angewandte Math. und Mech.* 3 (1923), 279–289], which is associated with a particularly interesting martingale. Imagine an urn that initially contains two balls, one red and one black. Repeatedly remove a randomly chosen ball from the urn, then replace it and contribute a new ball of the same color. The numbers \( (r, b) \) of red and black balls will follow a path in the diagram, with the respective local probabilities indicated on each branch.

One can show without difficulty that all \( n+1 \) nodes on level \( n \) of Fig. 1 will be reached with the same probability, \( 1/(n+1) \). Furthermore, the probability that a red ball is chosen when going from any level to the next is always 1/2. Thus the urn scheme might seem at first glance to be rather tame and uniform. But in fact the process turns out to be full of surprises, because any inequity between red and black tends to perpetuate itself. For example, if the first ball chosen is black, so that we go from \( (1, 1) \) to \( (1, 2) \), the probability is only \( 2\ln 2 - 1 \approx .036 \) that the red balls will ever overtake the black ones in the future (see exercise 88).

One good way to analyze Pólya’s process is to use the fact that the ratios \( r/(r+b) \) form a martingale. Each visit to the urn changes this ratio either to \( (r+1)/(r+b+1) \) (with probability \( r/(r+b) \)) or to \( r/(r+b+1) \) (with probability \( b/(r+b) \)); so the expected new ratio is \( (r+b^2+r)/(b(r+b)(r+b+1)) = r/(r+b) \), no different from what it was before. More formally, let \( X_0 = 1 \), and for \( n > 0 \) let \( X_n \) be the random variable “[the nth ball chosen is red]”. Then there are \( X_0 + \cdots + X_n \) red balls and \( X_0 + \cdots + X_n + 1 \) black balls at level \( n \) of Fig. 1; and the sequence \( \langle Z_n \rangle \) is a martingale if we define

\[
Z_n = (X_0 + \cdots + X_n)/(n+2). \tag{27}
\]

In practice it’s usually most convenient to define martingales \( Z_0, Z_1, \ldots \) in terms of auxiliary random variables \( X_0, X_1, \ldots \), as we’ve just done. The sequence \( \langle Z_n \rangle \) is said to be a **martingale with respect to the sequence** \( \langle X_n \rangle \) if \( Z_n \) is a function of \( (X_0, \ldots, X_n) \) that satisfies

\[
E(Z_{n+1} \mid X_0, \ldots, X_n) = Z_n \quad \text{for all } n \geq 0. \tag{28}
\]
Furthermore we say that a sequence $\langle Y_n \rangle$ is **fair with respect to the sequence $\langle X_n \rangle$** if $Y_n$ is a function of $(X_0, \ldots, X_n)$ that satisfies the simpler condition

$$E(Y_{n+1} | X_0, \ldots, X_n) = 0 \quad \text{for all } n \geq 0; \quad (29)$$

and we call $\langle Y_n \rangle$ fair whenever

$$E(Y_{n+1} | Y_0, \ldots, Y_n) = 0 \quad \text{for all } n \geq 0. \quad (30)$$

Exercise 77 proves that (28) implies (26) and that (29) implies (30); thus an auxiliary sequence $\langle X_n \rangle$ is sufficient but not necessary for defining martingales and fair sequences.

Whenever $\langle Z_n \rangle$ is a martingale, we obtain a fair sequence $\langle Y_n \rangle$ by letting $Y_0 = Z_0$ and $Y_n = Z_n - Z_{n-1}$ for $n > 0$, because the identity $E(Y_{n+1} | Z_0, \ldots, Z_n) = E(Z_{n+1} - Z_n | Z_0, \ldots, Z_n) = Z_n - Z_{n-1}$ shows that $\langle Y_n \rangle$ is fair with respect to $\langle Z_n \rangle$. Conversely, whenever $\langle Y_n \rangle$ is fair, we obtain a martingale $\langle Z_n \rangle$ by letting $Z_n = Y_0 + \cdots + Y_n$, because the identity $E(Z_{n+1} | Y_0, \ldots, Y_n) = E(Z_n + Y_{n+1} | Y_0, \ldots, Y_n) = Z_n$ shows that $\langle Z_n \rangle$ is a martingale with respect to $\langle Y_n \rangle$. In other words, fairness and martingales are essentially equivalent. The $Y$’s represent unbiased “tweaks” that change one $Z$ to its successor.

It’s easy to construct fair sequences. For example, every sequence of independent random variables with mean 0 is fair. And if $\langle Y_n \rangle$ is fair with respect to $\langle X_n \rangle$, so is the sequence $\langle Y'_n \rangle$ defined by $Y'_n = f_n(X_0, \ldots, X_{n-1}) Y_n$ when $f_n(X_0, \ldots, X_{n-1})$ is almost any function whatsoever! (We need only keep $f_n$ small enough that $E(Y'_n)$ is well defined.) In particular, we can let

$$f_n(X_0, \ldots, X_{n-1}) = 0 \quad \text{for all large } n,$

thereby making $\langle Z_n \rangle$ eventually fixed.

A sequence of functions $N_n(x_0, \ldots, x_{n-1})$ is called a stopping rule if each value is either 0 or 1 and if $N_n(x_0, \ldots, x_{n-1}) = 0$ implies $N_{n+1}(x_0, \ldots, x_n) = 0$. We can assume that $N_0 = 1$. The number of steps before stopping, with respect to a sequence of random variables $\langle X_n \rangle$, is then the random variable

$$N = N_1(X_0) + N_2(X_0, X_1) + N_3(X_0, X_1, X_2) + \cdots \quad (31)$$

(Intuitively, $N_n(x_0, \ldots, x_{n-1})$ means [the values $X_0 = x_0$, \ldots, $X_{n-1} = x_{n-1}$ do not stop the process]; hence it’s really more about “going” than “stopping.”)

Any martingale $Z_n = Y_0 + \cdots + Y_n$ with respect to $\langle X_n \rangle$ can be adapted to
stop with this strategy if we change it to \( Z'_n = Y'_0 + \cdots + Y'_n \), where \( Y'_n = N_n(x_0, \ldots , x_{n-1})X_n \). Gamblers who wish to “quit when ahead” are using the stopping rule \( N_{n+1}(x_0, \ldots , x_n) = [Z'_n \leq 0] \), when \( Z'_n \) is their current balance.

Notice that if the stopping rule always stops after at most \( m \) steps—in other words, if the function \( N_m(x_0, \ldots , x_{m-1}) \) is identically zero—then we have \( Z'_m = Z'_N \), because \( Z'_n \) doesn’t change after the process has stopped. Therefore \( E Z'_N = E Z'_m = E Z'_0 = E Z_0 \). No stopping rule can change the expected outcome of a martingale, when the number of steps is bounded.

An amusing game of chance called Ace Now illustrates this optional stopping principle. Take a deck of cards, shuffle it and place the cards face down, then turn them face up one at a time as follows: Just before seeing the \( n \)th card, you are supposed to say either “Stop” or “Deal,” based on the cards you’ve already observed. (If \( n = 52 \) you must say “Stop.”) After you’ve decided to stop, you win \$12 if the next card is an ace; otherwise you lose \$1. What is the best strategy for playing this game? Should you hold back until you have a pretty good chance at the \$12? What is the worst strategy? Exercise 82 has the answer.

**Tail inequalities from martingales.** The essence of martingales is equality of expectations. Yet martingales turn out to be important in the analysis of algorithms because we can use them to derive inequalities, namely to show that certain events occur with very small probability.

To begin our study, let’s introduce inequality into Eq. (26): A sequence \( \langle Z_n \rangle \) is called a submartingale if it satisfies

\[
E(Z_{n+1} \mid Z_0, \ldots, Z_n) \geq Z_n \quad \text{for all } n \geq 0.
\]  

(32)

Similarly, it’s called a supermartingale if ‘\( \geq \)’ is changed to ‘\( \leq \)’ in the left-hand part of this definition. (Thus a martingale is both sub- and super-.) In a submartingale we have \( E Z_0 \leq E Z_1 \leq E Z_2 \leq \cdots \), by taking expectations in (32). A supermartingale, similarly, has ever smaller expectations as \( n \) grows. One way to remember the difference between submartingales and supermartingales is to observe that their names are the reverse of what you might expect.

Submartingales are significant largely because of the fact that they’re quite common. Indeed, if \( \langle Z_n \rangle \) is any martingale and if \( f \) is any convex function, then \( \langle f(Z_n) \rangle \) is a submartingale (see exercise 84). For example, the sequences \( \langle |Z_n| \rangle \) and \( \langle \max(Z_n, c) \rangle \) and \( \langle Z_n^2 \rangle \) and \( \langle e^{Z_n} \rangle \) all are submartingales whenever \( \langle Z_n \rangle \) is known to be a martingale. If, furthermore, \( Z_n \) is always positive, then \( \langle Z_n^3 \rangle \) and \( \langle 1/Z_n \rangle \) and \( \langle \ln(1/Z_n) \rangle \) and \( \langle Z_n \ln Z_n \rangle \), etc., are submartingales.

If we modify a submartingale by applying a stopping rule, it’s easy to see that we get another submartingale. Furthermore, if that stopping rule is guaranteed to quit within \( m \) steps, we’ll have \( E Z_m \geq E Z_N = E Z'_m = E Z'_n \). Therefore no stopping rule can increase the expected outcome of a submartingale, when the number of steps is bounded.

That comparatively simple observation has many important consequences. For example, exercise 86 uses it to give a simple proof of the so-called “maximal

April 8, 2014
inequality*: If \( \langle Z_n \rangle \) is a nonnegative submartingale then

\[
\Pr(\max(Z_0, Z_1, \ldots, Z_n) \geq x) \leq \frac{E(Z_n)}{x}, \quad \text{for all } x > 0. \tag{33}
\]

Special cases of this inequality are legion. For instance, martingales \( \langle Z_n \rangle \) satisfy

\[
\Pr(\max(\{Z_0, |Z_1|, \ldots, |Z_n|\}) \geq x) \leq \frac{E(|Z_n|)}{x}, \quad \text{for all } x > 0; \tag{34}
\]
\[
\Pr(\max(Z^2_0, Z^2_1, \ldots, Z^2_n) \geq x) \leq \frac{E(Z^2_n)}{x}, \quad \text{for all } x > 0. \tag{35}
\]

Relation (35) is known as Kolmogorov’s inequality, because A. N. Kolmogorov proved it when \( Z_n = X_1 + \cdots + X_n \) is the sum of independent random variables with \( E(X_k) = 0 \) and \( \text{var}(X_k) = \sigma^2_k \) for \( 1 \leq k \leq n \) [Math. Annalen 99 (1928), 309–311]. In that case \( \text{var}(Z_n) = \sigma^2_1 + \cdots + \sigma^2_n = \sigma^2 \), and the inequality can be written

\[
\Pr(|X_1| < t\sigma, |X_1 + X_2| < t\sigma, \ldots, |X_1 + \cdots + X_n| < t\sigma) \geq 1 - 1/t^2. \tag{36}
\]

Chebyshev’s inequality gives only \( \Pr(|X_1 + \cdots + X_n| < t\sigma) \geq 1 - 1/t^2 \), which is a considerably weaker result.

Another important inequality applies in the common case where we have good bounds on the terms \( Y_1, \ldots, Y_n \) that enter into the standard representation \( Z_n = Y_0 + Y_1 + \cdots + Y_n \) of a martingale. This one is called the Hoeffding–Azuma inequality, after papers by W. Hoeffding [J. Amer. Statistical Association 58 (1963), 13–30] and K. Azuma [Tohoku Math. Journal (2) 19 (1967), 357–367]. It reads as follows: If \( \langle Y_n \rangle \) is any fair sequence with \( a_n \leq Y_n \leq b_n \), then

\[
\Pr(Y_1 + \cdots + Y_n \geq x) \leq e^{-2x^2/(b_1-a_1)^2 + \cdots + (b_n-a_n)^2}). \tag{37}
\]

The same bound applies to \( \Pr(Y_1 + \cdots + Y_n \leq -x) \), since \( -b_n \leq -Y_n \leq -a_n \); so

\[
\Pr(|Y_1 + \cdots + Y_n| \geq x) \leq 2e^{-2x^2/(b_1-a_1)^2 + \cdots + (b_n-a_n)^2}). \tag{38}
\]

Exercise 90 breaks the proof of this result into small steps. In fact, the proof even shows that \( a_n \) and \( b_n \) may be functions of \( \{Y_0, \ldots, Y_{n-1}\} \).

Applications. The Hoeffding–Azuma inequality is useful in the analysis of many algorithms because it applies to “Doob martingales,” a very general class of martingales that J. L. Doob featured as Example 1 in his Stochastic Processes (1953), page 92. (In fact, he had already considered them many years earlier, in Trans. Amer. Math. Soc. 47 (1940), 486.) Doob martingales arise from any sequence of random variables \( \langle X_n \rangle \), independent or not, and from any other random variable \( Q \): We simply define

\[
Z_n = E(Q|X_0, \ldots, X_n). \tag{39}
\]

Then, as Doob pointed out, the resulting sequence is a martingale (see exercise 91). In our applications, \( Q \) is an aspect of some algorithm that we wish to study, and the variables \( X_0, X_1, \ldots \) reflect the inputs to the algorithm. For example, in an algorithm that uses random bits, the \( X \)'s are those bits.

Consider a hashing algorithm in which \( t \) objects are placed into \( m \) random lists, where the \( n \)th object goes into list \( X_n \); thus \( 1 \leq X_n \leq m \) for \( 1 \leq n \leq t \), and we assume that each of the \( m^t \) possibilities is equally likely. Let \( Q(x_1, \ldots, x_t) \) be

April 8, 2014
the number of lists that remain empty after the objects have been placed into lists \( x_1, \ldots, x_t \), and let \( Z_n = E(Q \mid X_1, \ldots, X_n) \) be the associated Doob martingale. Then \( Z_0 = E(Q) \) is the average number of empty lists; and \( Z_t = Q(X_1, \ldots, X_t) \) is the actual number, in any particular run of the algorithm.

What fair sequence corresponds to this martingale? If \( 1 \leq n \leq t \), the random variable \( Y_n = Z_n - Z_{n-1} \) is \( f_n(x_1, \ldots, x_n) \), where \( f_n(x_1, \ldots, x_n) \) is the average of

\[
\Delta(x_1, \ldots, x_t) = \sum_{x=1}^{m} \Pr(X_n = x) \left( Q(x_1, \ldots, x_{n-1}, x_n, x_{n+1}, \ldots, x_t) - Q(x_1, \ldots, x_{n-1}, x, x_{n+1}, \ldots, x_t) \right)
\]

taken over all \( m^{t-n} \) values of \((x_{n+1}, \ldots, x_t)\).

In our application the function \( Q(x_1, \ldots, x_t) \) has the property that

\[
|Q(x_1, \ldots, x_{n-1}, x', x_{n+1}, \ldots, x_t) - Q(x_1, \ldots, x_{n-1}, x, x_{n+1}, \ldots, x_t)| \leq 1
\]

for all \( x \) and \( x' \), because a change to any one hash address always changes the number of empty lists by either 1, 0, or -1. Consequently, for any fixed setting of the variables \((x_1, \ldots, x_{n-1}, x_{n+1}, \ldots, x_t)\), we have

\[
\max_{x_n} \Delta(x_1, \ldots, x_t) \leq \min_{x_n} \Delta(x_1, \ldots, x_t) + 1.
\]

The Hoeffding–Azuma inequality (37) therefore allows us to conclude that

\[
\Pr(Z_t - Z_0 \geq x) = \Pr(Y_1 + \cdots + Y_t \geq x) \leq e^{-2x^2/\beta}.
\]

Furthermore, \( Z_0 \) in this example is \( m(m - 1)s/m^t \), because exactly \( (m - 1)s \) of the \( m^t \) possible hash sequences leave any particular list empty. And the random variable \( Z_t \) is the actual number of empty lists when the algorithm is run. Hence we can, for example, set \( x = \sqrt{t \ln f(t)} \) in (43), thereby proving that

\[
\Pr(Z_t \geq (m - 1)s/m^t - \sqrt{t \ln f(t)}) \leq 1/f(t)^2, \quad \text{whenever } f(t) > 1.
\]

The same upper bound applies to \( \Pr(Z_t \leq (m - 1)s/m^t - \sqrt{t \ln f(t)}) \).

Notice that the inequality (41) was crucial in this analysis. Therefore the strategy we’ve used to prove (43) is often called the “method of bounded differences.” In general, a function \( Q(x_1, \ldots, x_t) \) is said to satisfy a Lipschitz condition in coordinate \( n \) if we have

\[
|Q(x_1, \ldots, x_{n-1}, x, x_{n+1}, \ldots, x_t) - Q(x_1, \ldots, x_{n-1}, x', x_{n+1}, \ldots, x_t)| \leq c_n
\]

for all \( x \) and \( x' \). (This terminology mimics a well-known but only slightly similar constraint that was introduced long ago into functional analysis by Rudolf Lipschitz [Crelle 63 (1864), 296-308].) Whenever condition (45) holds, for a function \( Q \) associated with a Doob martingale for independent random variables \( X_1, \ldots, X_t \), we can prove that \( \Pr(Y_1 + \cdots + Y_t \geq x) \leq \exp(-2x^2/c_n^2 + \cdots + c_t^2)) \).

Let’s work out one more example, due to Colin McDiarmid [London Math. Soc. Lecture Notes 141 (1989), 148-188, §8(a)]; a gain we consider independent integer-valued random variables \( X_1, \ldots, X_t \) with \( 1 \leq X_n \leq m \) for \( 1 \leq n \leq t; \)

April 8, 2014
but this time we allow each $X_i$ to have a different probability distribution. Furthermore we define $Q(x_1, \ldots, x_t)$ to be the minimum number of bins into which objects of sizes $x_1, \ldots, x_t$ can be packed, where each bin has capacity $m$.

This bin-packing problem sounds a lot harder than the hashing problem that we just solved. Indeed, the task of evaluating $Q(x_1, \ldots, x_t)$ is well known to be NP-complete [see M. R. Garey and D. S. Johnson, *SICOMP* 4 (1975), 397–411].

Yet $Q$ obviously satisfies the condition (45) with $c_n = 1$ for $1 \leq n \leq t$. Therefore the method of bounded differences tells us that inequality (43) is true, in spite of the apparent difficulty of this problem!

The only difference between this bin-packing problem and the hashing problem is that we’re clueless about the value of $Z_n$. Nobody knows how to compute $E Q(X_1, \ldots, X_t)$, except for very special distributions of the random variables. However—and this is the magic of martingales—we do know that, whatever the value is, the actual numbers $Z_n$ will be tightly concentrated around that average.

If all the $X_i$’s have the same distribution, the values $\beta_t = E Q(X_1, \ldots, X_t)$ satisfy $\beta_{t+t'} \leq \beta_t + \beta_{t'}$, because we could always pack the $t$ and $t'$ items separately. Therefore, by the subadditive law (see the answer to exercise 2.5–39), $\beta_t/t$ approaches a limit $\beta$ as $t \to \infty$. Still, however, random trials won’t give us decent bounds on that limit, because we have no good way to compute the $Q$ function.

---

**Statements that are almost sure, or quite sure.** Probabilities that depend on an integer $n$ often have the property that they approach 0 or 1 as $n \to \infty$, and special terminology simplifies the discussion of such phenomena. If, say, $A_n$ is an event for which $\lim_{n \to \infty} \Pr(A_n) = 1$, it’s convenient to express this fact in words by saying, “$A_n$ occurs almost surely, when $n$ is large.” (Indeed, we usually don’t bother to state that $n$ is large; if we already understand that $n$ is approaching infinity in the context of the current discussion.)

For example, if we toss a fair coin $n$ times, we’ll find that the coin almost surely comes up heads more than $.49n$ times, but fewer than $.51n$ times.

Furthermore, we’ll occasionally want to express this concept tersely in formulas, by writing just ‘a.s.’ instead of spelling out the words “almost surely.” For instance, the statement just made about $n$ coin tosses can be formulated as

$$0.49n < X_1 + \cdots + X_n < 0.51n \text{ a.s.},$$

(46)

if $X_1, \ldots, X_n$ are independent binary random variables, each with $E X_j = 1/2$. In general a statement such as “$A_n$ a.s.” means that $\lim_{n \to \infty} \Pr(A_n) = 1$; or, equivalently, that $\lim_{n \to \infty} \Pr(\overline{A_n}) = 0$.

If $A_n$ and $B_n$ are both a.s., then the combined event $C_n = A_n \cap B_n$ is also a.s., regardless of whether those events are independent. The reason is that $\Pr(C_n) = \Pr(A_n \cup B_n) \leq \Pr(A_n) + \Pr(B_n)$, which approaches 0 as $n \to \infty$.

Thus, to prove (46) we need only show that $X_1 + \cdots + X_n > 0.49n$ a.s. and that $X_1 + \cdots + X_n < 0.51n$ a.s., or in other words that $\Pr(X_1 + \cdots + X_n \leq 0.49n)$

April 8, 2014
and \( \Pr(X_1 + \cdots + X_n \geq .51n) \) both approach 0. Those probabilities are actually equal, by symmetry between heads and tails; so we need only show that \( p_n = \Pr(X_1 + \cdots + X_n \leq .49n) \) approaches 0. And that’s no sweat, because we know from exercise 1.2.10–21 that \( p_n \leq e^{-0.001n} \).

In fact, we’ve proved more: We’ve shown that \( p_n \) is superpolynomially small, namely that

\[
p_n = O(n^{-K}) \quad \text{for all fixed numbers } K. \tag{47}
\]

When the probability of an event \( A_n \) is superpolynomially small, we say that \( A_n \) holds “quite surely,” and abbreviate that by ‘q.s.’. In other words, we’ve proved

\[
.49n < X_1 + \cdots + X_n < .51n \quad \text{q.s.} \tag{48}
\]

We’ve seen that the combination of any two a.s. events is a.s.; hence the combination of any finite number of a.s. events is also a.s. That’s nice, but q.s. events are even nicer: The combination of any polynomial number of q.s. events is also q.s. For example, if \( n^4 \) different people each toss \( n \) coins, it is quite sure that every one of them, without exception, will obtain between .49n and .51n heads!

(When making such asymptotic statements we ignore the inconvenient truth that our bound on the failure of the assertion, \( 2n^4 e^{-0.001n} \) in this case, becomes negligible only when \( n \) is greater than 700,000 or so.)

**EXERCISES**

1. **[M21]** (Nontransitive dice.) Suppose three biased dice with the respective faces

\[
A = \begin{array}{ccc}
\begin{array}{ccc}
\times & \times & \times \\
\mathbf{1} & \mathbf{2} & \mathbf{3}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{4} & \mathbf{5} & \mathbf{6}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{7} & \mathbf{8} & \mathbf{9}
\end{array}
\end{array}, \\
B = \begin{array}{ccc}
\begin{array}{ccc}
\times & \times & \times \\
\mathbf{1} & \mathbf{2} & \mathbf{3}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{4} & \mathbf{5} & \mathbf{6}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{7} & \mathbf{8} & \mathbf{9}
\end{array}
\end{array}, \\
C = \begin{array}{ccc}
\begin{array}{ccc}
\times & \times & \times \\
\mathbf{1} & \mathbf{2} & \mathbf{3}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{4} & \mathbf{5} & \mathbf{6}
\end{array} & \begin{array}{ccc}
\times & \times & \times \\
\mathbf{7} & \mathbf{8} & \mathbf{9}
\end{array}
\end{array}
\]

are rolled independently at random.

a) Show that \( \Pr(A > B) = \Pr(B > C) = \Pr(C > A) = 5/9 \).

b) Find dice with \( \Pr(A > B), \Pr(B > C), \Pr(C > A) \) all greater than \( 5/9 \).

c) If Fibonacci dice have \( F_m \) faces instead of just six, show that we could have

\[
\Pr(A > B) = \Pr(B > C) = F_{m-1}/F_m \quad \text{and} \quad \Pr(C > A) = F_{m-1}/F_m \pm 1/F_m^2.
\]

2. **[M32]** Prove that the previous exercise is asymptotically optimum, in the sense that \( \min(\Pr(A > B), \Pr(B > C), \Pr(C > A)) < 1/\phi \), regardless of the number of faces.

3. **[22]** (Lake Wobegon dice.) Continuing the previous exercises, find three dice such that \( \Pr(A > \frac{1}{3}(A + B + C)) \geq \Pr(B > \frac{1}{3}(A + B + C)) \geq \Pr(C > \frac{1}{3}(A + B + C)) \geq 16/27 \). Each face of each die should be \( \blacksquare \) or \( \emptyset \) or \( \mathcal{C} \) or \( \clubsuit \) or \( \heartsuit \) or \( \spadesuit \).

4. **[22]** (Nontransitive Bingo.) Each player in the game of NanoBingo has a card containing four numbers from the set \( S = \{1, 2, 3, 4, 5, 6\} \), arranged in two rows. An announcer calls out the elements of \( S \), in random order; the first player whose card has a horizontal row with both numbers called shouts “Bingo!” and wins. (Or victory is
shared when there are multiple Bin goes.) For example, consider the four cards
\[
A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 4 \\ 2 & 6 \end{pmatrix}.
\]
If the announcer calls “6, 2, 5, 1” when A plays against B, then A wins; but the sequence “1, 3, 2” would yield a tie. One can show that \(\Pr(A \text{ beats } B) = \frac{12}{25}, \Pr(B \text{ beats } A) = \frac{13}{25}, \text{ and } \Pr(A \text{ and } B \text{ tie}) = \frac{12}{25}.\) Determine the probabilities of all possible outcomes when there are (a) two (b) three (c) four different players using those cards.

5. [HM22] (T. M. Cover, 1989.) Common wisdom asserts that longer games favor the stronger player, because they provide more evidence of the relative skills.

However, consider an \(n\)-round game in which Alice scores \(A_1 + \cdots + A_n\) points while Bob scores \(B_1 + \cdots + B_n\) points, where each of \(A_1, \ldots, A_n\) are independent random variables representing Alice’s strength, and each of \(B_1, \ldots, B_n\) independently represent Bob’s (and are independent of the \(A_i\)’s). Suppose Alice wins with probability \(P_n\).

a) Show that it’s possible to have \(P_1 = .99\) but \(P_{1000} < .0001.\)

b) Let \(m_k = 2^{k^2}, n_k = 2^{k^2+k}, \text{ and } q_k = 2^{-k^2}/D, \) where \(D = 2^{-\infty} + 2^{-1} + 2^{-4} + 2^{-9} + \cdots \approx 1.56447.\) Suppose \(A_i\) and \(B_i\) are zero except that \(A_i = m_k\) with probability \(q_k\) when \(k \geq 0\) is even, \(B_i = n_k\) with probability \(q_k\) when \(k \geq 1\) is odd. What are \(\Pr(A > B), \Pr(A < B), \text{ and } \Pr(A = B)\)?

c) With the distributions in (b), prove that \(P_{n_k} \to [k \text{ even}]\) as \(k \to \infty.\)

6. [M20] Consider \(n \geq 2\) random Boolean (or binary) variables \(X_1, \ldots, X_n\) with the following joint distribution: The vector \(x_1 \cdots x_n\) occurs with probability \(1/(n-1)^2\) if \(x_1 + \cdots + x_n = 2,\) with probability \((n-2)/(2n-2)\) if \(x_1 + \cdots + x_n = 0,\) and with probability 0 otherwise. Show that these variables are pairwise independent (that is, \(X_i\) is independent of \(X_j\) when \(i \neq j\)); but they are not \(k\)-wise independent for \(k > 2.\)

7. [M22] Continuing the previous exercise, find a symmetrical joint distribution that is \(k\)-wise independent for \(k = 2\) and \(k = 3\) but not \(k = 4.\)

8. [M30] (Ernst Schulte-Geers, 2012.) Generalizing exercises 6 and 7, construct a symmetrical distribution that has \(k\)-wise but not \((k+1)\)-wise independence, given \(k \geq 1.\)

9. [M20] Suppose the Boolean vector \(x_1 \cdots x_n\) occurs with probability \((2+(-1)^{\nu x})/2^{n+1},\) where \(\nu x = x_1 + \cdots + x_n.\) For what \(k\) is this distribution \(k\)-wise independent?

10. [M20] Find a distribution of Boolean vectors \(x_1 \cdots x_n\) such that any two variables are dependent; yet if we know the value of any \(x_j,\) the remaining variables are \((n-1)\)-wise independent. Hint: The answer is so simple, you might feel hornswoggled.

11. [M20] If \(X_1, \ldots, X_{2n}\) are independent random variables with the same discrete distribution, and if \(\alpha\) is any real number whatsoever, prove that
\[
\Pr\left(\left|\frac{X_1 + \cdots + X_{2n}}{2n} - \alpha\right| \leq \left|\frac{X_1 + \cdots + X_n}{n} - \alpha\right|\right) > \frac{1}{2}.
\]

12. [18] Which of the following four statements are equivalent to the statement that \(\Pr(A \mid B) > \Pr(A)?\) (i) \(\Pr(B \mid A) > \Pr(B);\) (ii) \(\Pr(A \mid B) > \Pr(A \mid B);\) (iii) \(\Pr(B \mid A) > \Pr(B \mid A);\) (iv) \(\Pr(A \mid B) > \Pr(A \mid B) \). 

13. [15] True or false? \(\Pr(A \mid C) > \Pr(A)\) if \(\Pr(A \mid B) > \Pr(A)\) and \(\Pr(B \mid C) > \Pr(B)\).

14. [16] (Thomas Bayes, 1763.) Prove the “chain rule” for conditional probability:
\[
\Pr(A_1 \cap \cdots \cap A_n) = \Pr(A_1) \Pr(A_2 \mid A_1) \cdots \Pr(A_n \mid A_1 \cap \cdots \cap A_{n-1}).
\]

15. [12] True or false? \(\Pr(A \mid B \cap C) \Pr(B \mid C) = \Pr(A \cap B \mid C).\)

April 8, 2014
16. [M15] Under what circumstances is \( \Pr(A|B) = \Pr(A \cup C | B) \)?

17. [15] Evaluate the conditional probability \( \Pr(T \text{ is an ace} | B = Q \bullet) \) in the playing card example of the text, where \( T \) and \( B \) denote the top and bottom cards.

18. [20] Let \( M \) and \( m \) be the maximum and minimum values of the random variable \( X \). Prove that \( \text{var}(X) \leq (M - E(X)) (E(X) - m) \).

19. [HM25] Let \( X \) be a random nonnegative integer, with \( \Pr(X = x) = 1/2^{x+1} \), and suppose that \( X = (\ldots X_2 X_1 X_0) \) and \( X + 1 = (\ldots Y_2 Y_1 Y_0) \); in binary notation.
   a) What is \( E(X) \)? Hint: Express this number in the binary number system.
   b) Prove that the random variables \( \{X_0, X_1, \ldots, X_{n-1}\} \) are independent.
   c) Find the mean and variance of \( S = X_0 + X_1 + X_2 + \cdots \).
   d) Find the mean and variance of \( R = X_0 \oplus X_1 \oplus X_2 \oplus \cdots \).
   e) Let \( \pi = (11, p, p, 11, 11, 11) \). What is the probability that \( X_n = p_n \) for all \( n \geq 0 \)?
   f) What is \( E(Y_0) \)? Show that \( Y_0 \) and \( Y_1 \) are not independent.
   g) Find the mean and variance of \( T = Y_0 + Y_1 + Y_2 + \cdots \).

20. [M18] Let \( X_1, \ldots, X_k \) be binary random variables for which we know that \( \text{cov}(X_j X_j) = E(X_j) \), for all \( j \in \{1, \ldots, k\} \). Prove that the \( X_j \)'s are independent.

21. [M20] Find a small-as-possible example of random variables \( X \) and \( Y \) that satisfy \( \text{cov}(X,Y) = 0 \), that is, \( E(XY) = (E(X))(E(Y)) \), although they aren't independent.

22. [M20] Use Eq. (8) to prove the “union inequality”
   \[
   \Pr(A_1 \cup \cdots \cup A_n) \leq \Pr(A_1) + \cdots + \Pr(A_n).
   \]

23. [M21] If each \( X_k \) is an independent binary random variable with \( E X_k = p \), the cumulative binomial distribution \( B_{m,n}(p) \) is the probability that \( X_1 + \cdots + X_n \leq m \). Thus it’s easy to see that \( B_{m,n}(p) = \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} \).
   a) Show that \( B_{m,n}(p) \) is also equal to \( \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k} \), for \( 0 \leq m \leq n \).
   b) Consider the random variables \( j_1, j_2, \ldots, j_T \) defined by the rule that \( j_i = 0 \) if and only if \( j \) has one of the \( T \) values \( \{j_1, j_2, \ldots, j_T\} \), where \( 1 \leq j_1 < j_2 < \cdots < j_T \leq n \). What is \( \Pr(T \geq r \text{ and } j_r = s) \)?

24. [HM27] The cumulative binomial distribution also has many other properties.
   a) Prove that \( B_{m,n}(p) = (n-m) \binom{n}{m} \int_p^1 x^m (1-x)^{n-1-m} \, dx \), for \( 0 \leq m < n \).
   b) Use the formula to prove that \( B_{m,n}(m/n) > \frac{1}{e} \), for \( 0 \leq m < n/2 \). Hint: Show that \( \int_0^{m/n} x^m (1-x)^{n-1-m} \, dx \leq \int_1^{1/m} x^m (1-x)^{n-1-m} \, dx \).
   c) Show furthermore that \( B_{m,n}(m/n) > \frac{1}{e} \) when \( n/2 \leq m \leq n \). [Thus \( m \) is the median value of \( X_1 + \cdots + X_n \), when \( p = m/n \) and \( m \) is an integer.]

25. [M25] Suppose \( X_1, X_2, \ldots \) are independent random binary variables, with means \( E X_k = p_k \). Let \( \binom{n}{k} \) be the probability that \( X_1 + \cdots + X_n = k \); thus \( \binom{n}{k} = p_k (\binom{n-1}{k-1}) + q_k (\binom{n-1}{k}) \).
   a) Prove that \( \binom{n}{k} \geq (\binom{n}{k+1}) \), if \( p_k \leq (k+1)/(n+1) \) for \( 1 \leq j \leq n \).
   b) Furthermore \( \binom{n}{k} \leq (\binom{n}{k}) \binom{n}{k+1} \), if \( p_k \leq k/n \) for \( 1 \leq k \leq n \).

26. [M27] Continuing exercise 25, prove that \( \binom{n}{k} \binom{n}{k+1} \geq (\binom{n-1}{k-1}) (1 + \frac{1}{k+1}) (1 + \frac{1}{n-k}) \) for \( 0 < k < n \). Hint: Consider \( r_{n,k} = \binom{n}{k} / \binom{n}{k+1} \).

27. [M22] Find an expression for the generalized cumulative binomial distribution \( \sum_{k=0}^n \binom{n}{k} \) that is analogous to the alternative formula in exercise 23.

April 8, 2014
28. [HM28] (W. Hoeffding, 1956.) Let $X = X_1 + \cdots + X_n$ and $p_n + \cdots + p_n = np$ in exercise 25, and suppose that $E(g(X)) = \sum_{k=0}^{n} g(k)(\binom{n}{k})$ for some function $g$.
   a) Prove that $E(g(X)) \leq \sum_{k=0}^{n} g(k)(\binom{n}{k})p^k(1-p)^{n-k}$ if $g$ is convex in $[0 \ldots n]$.
   b) If $g$ isn’t convex, show that the maximum of $E(g(X))$, over all choices of $\{p_1, \ldots, p_n\}$
      with $p_1 + \cdots + p_n = np$, can always be attained by a set of probabilities for which
      at most three distinct values $\{0, a, 1\}$ occur among the $p_i$.
   c) Furthermore $\sum_{k=0}^{m}(\binom{n}{k}) \leq B_{m,n}(p)$, whenever $p_1 + \cdots + p_n = np \geq m + 1$.

29. [HM29] (S. M. Sambells, 1965.) Continuing exercise 28, prove that we have
      $B_{m,n}(p) \geq \frac{(1-p)(m+1)/((1-p)m+1)}{(1-p)(m+1)/((1-p)m+1)} ^m$ whenever $np \leq m + 1$.

30. [HM34] Let $X_1, \ldots, X_n$ be independent random variables whose values are non-
      negative integers, where $EX_k = 1$ for all $k$, and let $p = Pr(X_1 + \cdots + X_n \leq n)$.
   a) What is $p$, if each $X_k$ takes only the values 0 and 1? 
   b) Show that, in any set of distributions that minimize $p$, each $X_k$ assumes only two
      values, 0 and $m_k$, where $1 \leq m_k \leq n + 1$.
   c) Furthermore we have $p > 1/e$, whenever each $X_k$ has the same distribution.

31. [M20] Assume that $A_1, \ldots, A_n$ are random events such that, for every subset
      $I \subseteq \{1, \ldots, n\}$, the probability $Pr(\bigcap_{i \in I} A_i)$ that $A_i$ for $i \in I$ occur simultaneously
      is $\pi_I$; here $\pi_I$ is a number with $0 \leq \pi_I \leq 1$, and $\pi_\emptyset = 1$. Show that the probability
      of any combination of the events, $Pr(f([A_1], \ldots, [A_n]))$ for any Boolean function $f$, can be
      found by expanding $f$'s multilinear reliability polynomial $f([A_1], \ldots, [A_n])$ and replacing
      each term $\prod_{i \in I} [A_i]$ by $\pi_I$. For example, the reliability polynomial of $x_1 \oplus x_2 \oplus x_3$
      is $x_1 + x_2 + x_3 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3 + 4x_1x_2x_3$; hence $Pr([A_1] \oplus [A_2] \oplus [A_3])$
      $= \pi_1 + \pi_2 + \pi_3 - 2\pi_1\pi_2 - 2\pi_1\pi_3 - 2\pi_2\pi_3 + 4\pi_1\pi_2\pi_3$. (Here $\pi_{\emptyset}$ is short for $\pi_{\{1,2,3\}}$, etc.)

32. [M21] Not all sets of numbers $\pi_I$ in the preceding exercise can arise in an actual
      probability distribution. For example, if $I \subseteq J$ we must have $\pi_I \geq \pi_J$. What is a
      necessary and sufficient condition for the $2^n$ values of $\pi_I$ to be legitimate?

33. [M20] Suppose $X$ and $Y$ are binary random variables whose joint distribution
      is defined by the probability generating function $G(w, z) = E(w^X z^Y) = pw + qz + rzw$, where
      $p, q, r > 0$ and $p + q + r = 1$. Use the definitions in the text to compute the
      probability generating function $E(z^{E(X|Y)})$ for the conditional expectation $E(X|Y)$.

34. [M17] Write out an algebraic proof of (12), using the definitions (7) and (13).

35. [M20] True or false? (a) $E(E(X|Y)|Y) = E(X|Y)$; (b) $E(E(X|Y)|Z) = E(X|Z)$.

36. [M21] Simplify the formulas (a) $E[f(X)|X]$; (b) $E[f(Y)|E(g(X)|Y)]$.

37. [M20] Suppose $X_1, \ldots, X_n$ is a random permutation of $\{1, \ldots, n\}$, with every
      permutation occurring with probability $1/n!$. What is $E(X_k | X_1, \ldots, X_{k-1})$?

38. [M26] Let $X_1, \ldots, X_n$ be a random restricted growth string of length $n$, each with
      probability $1/2^n$ (see Section 7.2.1). What is $E(X_k | X_1, \ldots, X_{k-1})$?

39. [HM21] A hen lays $N$ eggs, where $Pr(N = n) = e^{-n}n^m/n!$ obeys the Poisson
      distribution. Each egg hatches with probability $p$, independent of all other eggs. Let
      $K$ be the resulting number of chicks. Express (a) $E(K|N)$, (b) $EK$, and (c) $E(N|K)$
      in terms of $N, K, \mu$, and $p$.

40. [M16] Suppose $X$ is a random variable with $X \leq M$, and let $m$ be any value with
      $m < M$. Show that $Pr(X > m) \geq (E X - m)/(M - m)$.

41. [HM21] Which of the following functions are convex in the set of all real
      numbers $x$? (a) $x^a$, where $a$ is a constant; (b) $\sum_{k=0}^{n} x^k$, where $n \geq 0$ is an integer;
      (c) $e^{x}$; (d) $f(x) = x \in I + \infty [x \notin I]$, where $f$ is convex in the interval $I$.

April 8, 2014
42. [HM21] Prove Jensen’s inequality (20).

43. [M18] Use (12) and (20) to strengthen (20): If \( f \) is convex in \( I \) and if the random variable \( X \) takes values in \( I \), then \( f(EX) \leq E(f(EX | Y)) \leq E(f(X)) \).

44. [M25] If \( f \) is convex on the real line and if \( EX = 0 \), prove that \( E(f(aX)) \leq E(f(bX)) \) whenever \( 0 \leq a \leq b \).

45. [M18] Derive the first moment principle (21) from Markov’s inequality (15).

46. [M15] Explain why \( E(X^2 | X > 0) \geq (E(X | X > 0))^2 \) in (23).

47. [M15] If \( X \) is random and \( Y = \max(0, X) \), show that \( EY \geq EX \) and \( EY^2 \leq EX^2 \).

48. [M20] Suppose \( X_1, \ldots, X_n \) are independent random variables with \( E X_k = 0 \) and \( E X_k^2 = \sigma_k^2 \) for \( 1 \leq k \leq n \). Chebyshev’s inequality tells us that \( \Pr(|X_1 + \cdots + X_n| \geq a) \leq \frac{\sigma_1^2 + \cdots + \sigma_n^2}{a^2} \); show that the second moment principle gives a somewhat better one-sided estimate, \( \Pr(X_1 + \cdots + X_n \geq a) \leq \frac{\sigma_1^2 + \cdots + \sigma_n^2}{a^2 + \sigma_1^2 + \cdots + \sigma_n^2} \), if \( a \geq 0 \).

49. [M20] If \( X \) is random and \( \geq 0 \), prove that \( \Pr(X = 0) \leq (EX^2)/EX^2 - 1 \).

50. [M20] Let \( X = X_1 + \cdots + X_n \) be the sum of binary random variables, with \( E X_j = p_j \). Let \( J \) be independent of the \( X_j \)'s, and uniformly distributed in \{1, \ldots, m\}.
   a) Prove that \( \Pr(X > 0) = \sum_{j=1}^m \Pr(X_j/X | X_j > 0) \cdot \Pr(X_j > 0) \).
   b) Therefore (24) holds. Hint: Use Jensen’s inequality with \( f(x) = 1/x \).
   c) What are \( \Pr(X = 1) \) and \( \Pr(J = j | X = 1) \)?
   d) Let \( t_j = E(X | J = j \text{ and } X = 1) \). Prove that \( EX^2 = \sum_{j=1}^m p_j t_j \).
   e) Jensen’s inequality now implies that the right side of (24) is \( \geq (EX^2)/EX^2 \).

51. [M21] Show how to use the conditional expectation inequality (24) to obtain also an upper bound on the value of a reliability polynomial, and apply your method to the case illustrated in (25).

52. [M21] What lower bound does inequality (24) give for the reliability polynomial of the symmetric function \( S_{\geq k}[x_1, \ldots, x_n] \), when \( p_1 = \cdots = p_n = p \)?

53. [M20] Use (24) to obtain a lower bound for the reliability polynomial of the non-monotonic Boolean function \( f(x_1, \ldots, x_n) = x_1 x_2 \bar{x}_3 \lor x_2 x_3 \bar{x}_4 \lor \cdots \lor x_n x_1 \bar{x}_2 \).

54. [M22] Suppose each edge of a random graph on the vertices \{1, \ldots, n\} is present with probability \( p \), independent of every other edge. If \( u \), \( v \), \( w \) are distinct vertices, let \( X_{uvw} \) be the probability that \{\( u, v, w \)\} is a 3-clique, namely the probability that \( u \leftrightarrow v, u \leftrightarrow w, \text{ and } v \leftrightarrow w \). Also let \( X = \sum_{1 \leq u < v < w \leq n} X_{uvw} \) be the total number of 3-cliques. Use the (a) first and (b) second moment principle to derive bounds on the probability that the graph contains at least one 3-clique.

55. [23] Evaluate the upper and lower bounds in the previous exercise numerically in the case \( n = 10 \), and compare them to the true probability, when (a) \( p = 1/2 \); (b) \( p = 1/10 \).

56. [HM20] Evaluate the upper and lower bounds of exercise 54 asymptotically when \( p = \lambda/n \) and \( n \to \infty \).

57. [M21] Obtain a lower bound for the probability in exercise 54(b) by using the conditional expectation inequality (24) instead of the second moment principle (22).

58. [M22] Generalizing exercise 54, find bounds on the probability that a random graph on \( n \) vertices has a \( k \)-clique, when each edge has probability \( p \).

April 8, 2014
59. [HM25] (The four functions theorem.) The purpose of this exercise is to prove an inequality that applies to four sequences \( \{a_n\}, \{b_n\}, \{c_n\}, \{d_n\}\) of nonnegative numbers:

\[
a_j b_k \leq c_{j|k} d_{j|k} \quad \text{for } 0 \leq j, k < \infty \quad \text{implies} \quad \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_j b_k \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} c_{j|k} d_{j|k}. \tag{*}
\]

(The sums will be \(\infty\) if they don’t converge.) Although the inequality might appear at first to be merely a curiosity, of interest only to a few lovers of esoteric formulas, we shall see that it’s a fundamental result with many applications of great importance.

a) Prove the special case where \(a_j = b_j = c_j = d_j = 0\) for \(j \geq 2\), namely that

\[
a_0 b_0 \leq c_0 d_0, \quad a_0 b_1 \leq c_1 d_0, \quad a_1 b_0 \leq c_1 d_0, \quad \text{and} \quad a_1 b_1 \leq c_1 d_1\]

implies \((a_0 + a_1)(b_0 + b_1) \leq (c_0 + c_1)(d_0 + d_1)\).

Can equality hold in the first four relations but not in the last one? Can equality hold in the last relation but not in the first four?

b) Use that result to prove (*) when \(a_j = b_j = c_j = d_j = 0\) for all \(j \geq 2^n\), given \(n > 0\).

c) Conclude that (*) is true in general.

60. [M21] If \(\mathcal{F}\) is a family of sets, and if \(\alpha\) is a function that maps sets into real numbers, let \(\alpha(\mathcal{F}) = \sum_{S \in \mathcal{F}} \alpha(S)\). Suppose \(\mathcal{F}\) and \(\mathcal{G}\) are finite families of sets for which nonnegative set functions \(\alpha, \beta, \gamma, \) and \(\delta\) have been defined with the property that

\[
\alpha(S) \beta(T) \leq \gamma(S \cup T) \delta(S \cap T) \quad \text{for all } S \in \mathcal{F} \text{ and } T \in \mathcal{G}.
\]

a) Use exercise 59 to prove that \(\alpha(\mathcal{F}) \beta(\mathcal{G}) \leq \gamma(\mathcal{F} \cup \mathcal{G}) \delta(\mathcal{F} \cap \mathcal{G})\).

b) In particular, \(|\mathcal{F}| |\mathcal{G}| \leq |\mathcal{F} \cup \mathcal{G}| \cdot |\mathcal{F} \cap \mathcal{G}|\) for all families \(\mathcal{F}\) and \(\mathcal{G}\).

61. [M28] Consider random sets in which \(S\) occurs with probability \(\mu(S)\), where

\[
\mu(S) \geq 0 \quad \text{and} \quad \mu(S \cup T) \leq \mu(S) \mu(T) \mu(S \cap T) \quad \text{for all sets } S \text{ and } T. \tag{**}
\]

Assume also that \(U = \bigcup_{\mu(S) > 0} S\) is a finite set.

a) Prove the FKG inequality (which is named for C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre): If \(f\) and \(g\) are real-valued set functions, then

\[
f(S) \leq f(T) \quad \text{and} \quad g(S) \leq g(T) \quad \text{for all } S \subseteq T \quad \text{implies} \quad E(fg) \geq E(f)E(g).
\]

Here, as usual, \(E(f)\) stands for \(\sum_S \mu(S) f(S)\). The conclusion can also be written ‘covar\((f, g) \geq 0\)’, using the notation of \(\langle g \rangle\); we say that \(f\) and \(g\) are “positively correlated” when this is true. (The awkward term “nonnegatively correlated” would be more accurate, because \(f\) and \(g\) might actually be independent.) Hint: Prove the result first in the special case that both \(f\) and \(g\) are nonnegative.

b) Furthermore,

\[
f(S) \geq f(T) \quad \text{and} \quad g(S) \geq g(T) \quad \text{for all } S \subseteq T \quad \text{implies} \quad E(fg) \geq E(f)E(g); \quad f(S) \leq f(T) \quad \text{and} \quad g(S) \geq g(T) \quad \text{for all } S \subseteq T \quad \text{implies} \quad E(fg) \leq E(f)E(g).\]

c) It isn’t necessary to verify condition (** for all sets, if (** is known to hold for sufficiently many pairs of “neighboring” sets. Given \(\mu\), let’s say that set \(S\) is supported if \(\mu(S) \neq 0\). Prove that (** holds for all \(S\) and \(T\) whenever the following three conditions are satisfied: (i) If \(S\) and \(T\) are supported, so are \(S \cup T\) and \(S \cap T\).

(ii) If \(S\) and \(T\) are supported and \(S \subseteq T\), the elements of \(T \setminus S\) can be labeled \(t_1, \ldots, t_k\) such that each of the intermediate sets \(S \cup \{t_1, \ldots, t_j\}\) is supported, for \(1 \leq j \leq k\). (iii) Condition (** holds whenever \(S = R \cup s\) and \(T = R \cup t\) and \(s, t \notin R\).

April 8, 2014
d) The multivariate Bernoulli distribution \( B(p_1, \ldots, p_m) \) on subsets of \( \{1, \ldots, m\} \) is 
\[
\mu(S) = \left( \prod_{j=1}^{m} p_{j}^{\lfloor j \in S \rfloor} \right) \left( \prod_{j=1}^{m} (1 - p_{j})^{\lfloor j \not\in S \rfloor} \right),
\]
given \( 0 \leq p_1, \ldots, p_m \leq 1 \). (Thus each element \( j \) is included independently with probability \( p_j \), as in exercise 25.) Show that this distribution satisfies \((***)\).

c) Describe other simple distributions for which \((***)\) holds.

62. [M20] Suppose the \( m = \binom{n}{2} \) edges \( E \) of a random graph \( G \) on \( n \) vertices are chosen with the Bernoulli distribution \( B(p_1, \ldots, p_m) \). Let \( f(E) = |G \text{ is connected}| \) and \( g(E) = |G \text{ is 4-colorable}| \). Prove that \( f \) is negatively correlated with \( g \).

63. [M17] Suppose \( Z_0 \) and \( Z_1 \) are random ternary variables with \( \Pr(Z_0 = a \text{ and } Z_1 = b) = p_{a,b} \) for \( 0 \leq a, b \leq 2 \), where \( p_0 + p_1 + \cdots + p_2 = 1 \). What can you say about those probabilities \( p_{a,b} \) when \( \Pr(Z_1|Z_0) = Z_0 \)?

64. [M22] (a) If \( \Pr(Z_{n+1} | Z_n) = Z_n \) for all \( n \geq 0 \), is \( \langle Z_n \rangle \) a martingale? (b) If \( \langle Z_n \rangle \) is a martingale, is \( \Pr(Z_{n+1} | Z_n) = Z_n \) for all \( n \geq 0 \)?

65. [M21] If \( \langle Z_n \rangle \) is a martingale, show that any subsequence \( \langle Z_{m(n)} \rangle \) is also a martingale, where the nonnegative integers \( \langle m(n) \rangle \) satisfy \( m(0) < m(1) < m(2) < \cdots \).

66. [M22] Find all martingales \( Z_0, Z_1, \ldots \) such that each random variable \( Z_n \) assumes only the values \( \pm n \).

67. [M20] The Equitable Bank of El Dorado features a money machine such that, if you insert \( k \) dollars, you receive \( 2k \) dollars back with probability exactly \( 1/2 \); otherwise you get nothing. Thus you either gain \$k \) or lose \$k \), and your expected profit is \$0. (Of course these transactions are all done electronically.)

a) Consider, however, the following scheme: Insert \$1; if that loses, insert \$2; if that also loses, insert \$4; then \$8, etc. If you first succeed after inserting \( 2^n \) dollars, stop and take the \( 2^{n+1} \) dollars. What is your expected net profit at the end?

b) Continuing (a), what’s the expected total amount that you put into the machine?

c) If \( Z_n \) is your net profit after \( n \) trials, show that \( \langle Z_n \rangle \) is a martingale.

68. [HM23] When J. H. Quick (a student) visited El Dorado, he decided to proceed by making repeated bets of \$1 each, and to stop when he first came out ahead. (He was in no hurry, and was well aware of the perils of the high-stakes strategy in exercise 67.)

a) What martingale \( \langle Z_n \rangle \) corresponds to this more conservative strategy?

b) Let \( N \) be the number of bets that Quick made before stopping. What is the probability that \( N = n \)?

c) What is the probability that \( N \geq n \)?

d) What is \( \operatorname{EN} N \)?

e) What is the probability that \( \min(Z_0, Z_1, \ldots) = -m \)? (Possible "gambler’s min."")

f) What is the expected number of indices \( n \) such that \( Z_n = -m \), given \( m \geq 0 \)?

69. [M20] Section 1.2.5 discusses two basic ways by which we can go from permutations of \( \{1, \ldots, n-1\} \) to permutations of \( \{1, \ldots, n\} \): “Method 1” inserts \( n \) among the previous elements in all possible ways; “Method 2” puts a number \( k \) from 1 to \( n \) in the final position, and adds 1 to each previous number that was \( \geq k \).

Show that, using either method, every permutation can be associated with a node of Fig. 1, using a rule that obeys the probability assumptions of Pólya’s urn model.

70. [M25] If Pólya’s urn model is generalized so that we start with \( c \) balls of different colors, is there a martingale that generalizes Fig. 1?
71. [M21] (G. Pólya.) What is the probability of going from node \((r,b)\) to node \((r',b')\) in Fig. 1, given \(r, r', b,\) and \(b'\) with \(r' \geq r\) and \(b' \geq b\)?

72. [M21] Let \(X_n\) be the red-ball indicator for Pólya's urn, as discussed in the text. What is \(E(X_{n_1} X_{n_2} \cdots X_{n_m})\) when \(0 < n_1 < n_2 < \cdots < n_m\)?

73. [M24] The ratio \(Z_n = r/(n+2)\) at node \((r, n+2-r)\) of Fig. 1 is not the only martingale definable on Pólya's urn. For example, \(r^n/n+1\) is another; so is \(r^{n+1}/2^n\).

Find the most general martingale \(\langle Z_n \rangle\) for this model. Given any sequence \(a_0, a_1, \ldots\), show that there's exactly one suitable function \(Z_n = f(r,n)\) such that \(f(1,k) = a_k\).

74. [M20] (Bernard Friedman's urn.) Instead of contributing a ball of the same color, as in Fig. 1, suppose we use the opposite color. Then the process changes to

![Diagram]

and the probabilities of reaching each node become quite different. What are they?

75. [M25] Find an interesting martingale for Bernard Friedman's urn.

76. [M20] If \(\langle Z_n \rangle\) and \(\langle Z'_n \rangle\) are martingales, is \(\langle Z_n + Z'_n \rangle\) a martingale?

77. [M21] Prove or disprove: If \(\langle Z_n \rangle\) is a martingale with respect to \(\langle X_n \rangle\), then \(\langle Z_n \rangle\) is a martingale with respect to itself (that is, a martingale).

78. [M20] A sequence of random variables \(\langle V_n \rangle\) for which \(E(V_{n+1} | V_0, \ldots, V_n) = 1\) is called “multiplicatively fair.” Show that \(Z_n = V_0 V_1 \ldots V_n\) is a martingale in such a case. Conversely, does every martingale lead to a multiplicatively fair sequence?

79. [M20] (De Moivre's martingale.) Let \(X_1, X_2, \ldots\) be a sequence of independent coin tosses, with \(Pr(\text{"heads" occurred on the } n\text{th toss}) = Pr(X_n = 1) = p\) for each \(n\). Show that \(Z_n = (q/p)^{n-1}\) defines a martingale, where \(q = 1 - p\).

80. [M20] Are the following statements true or false for every fair sequence \(\langle Y_n \rangle\)?
   (a) \(E(Y_2 Y_3) = 0\). (b) \(E(Y_2 Y_3) = 0\). (c) \(E(Y_1 Y_2 \cdots Y_m) = 0\) if \(n_1 < n_2 < \cdots < n_m\).

81. [M21] Suppose \(E(X_{n+1} | X_0, \ldots, X_n) = X_n + X_{n+1}\) for \(n \geq 0\), where \(X_{-1} = 0\). Find sequences \(a_n\) and \(b_n\) of coefficients so that \(Z_n = a_n X_n + b_n X_{n-1}\) is a martingale, where \(Z_0 = X_0\) and \(Z_1 = 2X_0 - X_1\). (We might call this a “Fibonacci martingale.”)

82. [M20] In the game of Ace Now, let \(X_n = [\text{the } n\text{th card is an ace}]\), with \(X_0 = 0\).
   a) Show that \(Z_n = (4 - X_1 - \cdots - X_n) / (52 - n)\) satisfies (a8) for \(0 \leq n < 52\).
   b) Consequently \(E Z_N = 1/13\), regardless of the stopping rule employed.
   c) Hence all strategies are equally good (or bad); you win $0 on average.

83. [HM23] A sequence \(\langle X_n \rangle\) of independent and nonnegative random variables, let \(S_n = X_1 + \cdots + X_n\). If \(S_n(x_0, \ldots, x_{n-1})\) is any stopping rule and if \(N\) is defined by (31), prove that \(E S_N = E \sum_{k=1}^{N} E X_k\) (In particular, if \(E X_n = EX_1\) for all \(n > 0\) we have “Wald's equation,” which states that \(E S_N = (EN)(EX_1)\).

April 8, 2014
84. [HM20] Let \( f(x) \) be a convex function for \( a \leq x \leq b \), and assume that \( \langle Z_n \rangle \) is a martingale such that \( a \leq Z_n \leq b \) for all \( n \geq 0 \). (Possibly \( a = -\infty \) and/or \( b = +\infty \).)
   a) Prove that \( \langle f(Z_n) \rangle \) is a submartingale.
   b) What can you say if the sequence \( \langle Z_n \rangle \) is assumed only to be a submartingale?
85. [M20] Suppose there are \( R_n \) red balls and \( B_n \) black balls at level \( n \) of Pólya’s urn (Fig. 1). Prove that the sequence \( \langle R_n /B_n \rangle \) is a submartingale.
   ▶ 86. [M22] Prove (33) by inventing a suitable stopping rule \( N_{n+1} (Z_0, \ldots , Z_n) \).
87. [M17] What does the maximal inequality (33) reveal about the chances that Pólya’s urn will hold thrice as many red balls as black balls at some point?
   ▶ 88. [HM20] Let \( S = \sup Z_n \) be the least upper bound of \( Z_n \) as \( n \to \infty \) in Fig. 1.
   a) Prove that \( S > 1/2 \) with probability \( \ln 2 \approx .693 \).
   b) Similarly, show that \( \Pr (S > 2/3) = \ln 3 - \pi /\sqrt{27} \approx .494 \).
   c) Generalize to \( \Pr (S > (t-1)/t) \) for all \( t \geq 2 \). Hint: Use exercise 7.2.1.6-36.
89. [M16] Let \( (X_1, \ldots , X_n) \) be random variables that have the Bernoulli distribution \( B(p_1, \ldots , p_n) \). Use (37) to show that \( \Pr (X_1 + \cdots + X_n \geq p_1 + \cdots + p_n + x) \leq e^{-2x^2/n} \).
90. [HM25] The Hoeffding-Azuma inequality (32) can be derived as follows:
   a) Show first that \( \Pr (Y_1 + \cdots + Y_n \leq t) \leq e^{(t/n^2)} \) for all \( t > 0 \).
   b) If \( 0 \leq p \leq 1 \) and \( q = 1 - p \), show that \( e^{qt} \leq e^{t(1 + e^n)} \) when \( 0 \leq t \leq q \) and \( t > 0 \), where \( f(t) = -pt + \ln(q + pe^t) \) and \( g(t) = -pt + \ln(e^t - 1) \).
   c) Prove that \( f(t) \leq t^2 /8 \). Hint: Use Taylor’s formula, Eq. 1.2.11.3-(5).
   d) Consequently \( a \leq Y \leq b \) implies \( e^{qt} \leq e^{(b-a)^2t/8} + Yh(t) \) for some function \( h(t) \).
   e) Let \( c = (c_1^2 + \cdots + c_n^2) /2 \), where \( c_k = b_k - a_k \).
   f) Prove that \( E(e^{Y_t + \cdots + Y_n}) \leq e^{t^2/4} \).
   g) We obtain (37) by choosing the best value of \( t \).
91. [M18] Prove that Doob’s general formula (39) always defines a martingale.
92. [M20] Let \( \langle Q_n \rangle \) be the Doob martingale that corresponds to Pólya’s urn (27) when \( Q = X_m \) for some fixed \( m \geq 0 \). Calculate \( Q_0, Q_1, Q_2, \) etc.
93. [M20] Solve the text’s hashing problem under the more general model considered in the bin-packing problem: Each variable \( X_n \) has probability \( p_{n,k} \) of being equal to \( k \), for \( 1 \leq n \leq t \) and \( 1 \leq k \leq m \). What formula do you get instead of (44)?
94. [M22] Where is the fact that the variables \( \{X_1, \ldots , X_n\} \) are independent used in the previous exercise?
95. [M20] True or false? “Pólya’s urn q.s. accumulates more than 100 red balls.”
96. [HM22] Let \( X \) be the number of heads seen in \( n \) flips of an unbiased coin. Decide whether each of the following statements about \( X \) is a.s., q.s., or neither, as \( n \to \infty \):
   (i) \( |X - n/2| < \sqrt{n \ln n} \).
   (ii) \( |X - n/2| < \sqrt{n \ln n} \).
   (iii) \( |X - n/2| < \sqrt{n \ln \ln n} \).
   (iv) \( |X - n/2| < \sqrt{n} \).
97. [HM21] Suppose \( \lfloor n^{1+\delta} \rfloor \) items are hashed into \( n \) bins, where \( \delta \) is a positive constant. Prove that every bin q.s. gets between \( \frac{1}{4} n^\delta \) and \( 2n^\delta \) of them.
98. [M21] Many algorithms are governed by a loop of the form
   \[ X \leftarrow n; \quad \text{while } X > 0, \text{ set } X \leftarrow X - F(X) \]
   where \( F(X) \) is a random integer in the range \( [1 \ldots X] \). We assume that each integer \( F(X) \) is completely independent of any previously generated values, subject only to the requirement that \( E F(j) \geq g \), where \( 0 < g \leq g \leq \cdots \leq g_n \).

April 8, 2014
Prove that the loop sets $X \leftarrow X - F(X)$ at most $1/g_1 + 1/g_2 + \cdots + 1/g_n$ times, on the average. (“If one step reduces by $g_i$, then perhaps (1/$g_i$)th of a step reduces by 1.”)

99. [HM36] Show that the result in the previous exercise holds even when the range of $F(X)$ is $(-\infty, X]$, given $0 < g_1 \leq \cdots \leq g_n \leq g_{n+1} \leq \cdots$. (Thus $X$ might increase.)

*Every man must judge for himself between conflicting vague probabilities.*

— CHARLES DARWIN, letter to N. A. von Mengden (5 June 1879)
ANSWERS TO EXERCISES

It isn’t that they can’t see the solution.
It is that they can’t see the problem.

MATHEMATICAL PRELIMINARIES REDUX

1. (a) \(A\) beats \(B\) in 5+0+5+5+0+5 cases out of 36; \(B\) beats \(C\) in 4+2+4+4+2+4; \(C\) beats \(A\) in 2+2+2+6+2+6.

(b) The unique solution, without going to more than six spots per face, is

\[
A = \begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix}, \quad B = \begin{bmatrix}
5 & 6 \\
7 & 8
\end{bmatrix}, \quad C = \begin{bmatrix}
9 & 10 \\
11 & 12
\end{bmatrix}.
\]

(c) \(A = \{F_{m-2} \times 1, F_{m-1} \times 4\}, B = \{F_{m} \times 3\}, C = \{F_{m-1} \times 2, F_{m-2} \times 5\}\) makes \(\Pr(C > A) = F_{m-2}F_{m-1}F_{m} / F_{m}^{3}\) and we have \(F_{m-2}F_{m} = F_{m-1}F_{m} = (-1)^{m} \cdot [\text{Similarly, with } n \text{ faces and } A = \{\lfloor n/\phi \rfloor \times 1, \lfloor n/\phi \rfloor \times 4\}, \text{ etc. the probabilities are } 1/\phi - O(1/n).\text{ See R. P. Savage, Jr., AMM 101 (1994), 429–436.}\]

2. Let \(\Pr(A > B) = A, \Pr(B > C) = B, \Pr(C > A) = C\). We can assume that no \(x\) appears on more than one die; if it did, we could replace it by \(x + \epsilon\) in \(A\) and \(x - \epsilon\) in \(C\) (for small enough \(\epsilon\)) without decreasing \(A, B,\) or \(C\). So we can list the face elements in nondecreasing order and replace each one by the name of its die; for example, the previous answer (b) yields CBBC.AAAACCCCCBBBA. Clearly \(AB, BC,\) and \(CA\) are never consecutive in an optimal arrangement of this kind: \(BA\) is always better than \(AB\).

Suppose the sequence is \(C = B^{k_{1}}A^{n_{1}} \cdots B^{k_{2}}A^{n_{2}}\) where \(c_{i} > 0\) for \(1 \leq i \leq k\) and \(b_{i}, a_{i} > 0\) for \(1 \leq i < k\). Let \(a_{i} = a_{i}/(a_{1} + \cdots + a_{k}), \quad b_{i} = b_{i}/(b_{1} + \cdots + b_{k}), \quad \gamma_{i} = c_{i} / (c_{1} + \cdots + c_{k});\text{ then } A = a_{1}b_{1} + a_{2}b_{2} + \cdots, B = b_{1}a_{1} + b_{2}a_{2} + \cdots, C = a_{1} + a_{2} + a_{3} + \cdots.\text{ We will show that } \min(A, B, C) \leq 1/\phi \text{ when the } a_i\text{'s, } b_i\text{'s, and } \gamma_i\text{'s are nonnegative real numbers; then it is } < 1/\phi \text{ when they are rational.}\)

The key idea is that we can assume \(k \leq 2\) and \(a_{2} = 0.\) Otherwise the following transformation leads to a shorter array without decreasing \(A, B,\) or \(C:\)

\[
\gamma_{2}^{'} = \lambda_{2}, \quad \gamma_{1}^{'} = \gamma_{1} + \gamma_{2} - \gamma_{2}^{'} - \gamma_{2}, \quad \beta_{2}^{'} = \lambda_{2}, \quad \beta_{1}^{'} = \beta_{1} + \beta_{2} - \beta_{2}^{'} - \alpha_{1}, \quad \alpha_{1} = \alpha_{1}/\lambda, \quad \alpha_{2} = \alpha_{1} + \alpha_{2} - \alpha_{1}^{'}.
\]

Indeed, \(A' = A, C' = C,\) and \(B' - B = (1 - \lambda)(\beta_{1} - \lambda_{2})\gamma_{2}^{'}\), and we can choose \(\gamma_{i}^{'} = 0\) thus:

Case 1: \(\beta_{1} \geq \beta_{2}.\text{ Choose } \lambda = \alpha_{1}/(\alpha_{1} + \alpha_{2}), \text{ making } \alpha_{2}^{'} = 0.\)

Case 2: \(\beta_{1} < \beta_{2} \text{ and } \gamma_{1}/\gamma_{2} \leq \beta_{1}/\beta_{2}.\text{ Choose } \lambda = 1 + \gamma_{1}/\gamma_{2}, \text{ making } \gamma_{1}^{'} = 0.\)

Case 3: \(\beta_{1} < \beta_{2} \text{ and } \gamma_{1}/\gamma_{2} > \beta_{1}/\beta_{2}.\text{ Choose } \lambda = 1 + \beta_{1}/\beta_{2}, \text{ making } \beta_{1}^{'} = 0.\)

Finally, then, \(A = \beta_{1}, B = 1 - \beta_{1}\gamma_{2}, C = \gamma_{2};\text{ they can't all be greater than } 1/\phi.\)

[Similarly, with \(n\) dice, the asymptotic optimum probability \(p_{n}\) satisfies \(p_{n} = \alpha_{2}^{(n)} = 1 - \alpha_{1}^{(n-1)} - \alpha_{2}^{(n)} = \cdots = 1 - \alpha_{1}^{(2)} - \alpha_{2}^{(3)} = \alpha_{1}^{(2)}\). One can show that \(f_n(1 - p_n) = 0,\)

April 8, 2014
where $f_{k+1}(x) = f_k(x) + x f_{k-1}(x)$. $f_0(x) = 1, f_1(x) = 1 - x$. Then $f_k(x^2)$ is expressible as the Chebyshev polynomial $x^{n+1} U_{n+1}(\frac{x}{\sqrt{2}})$; and we have $p_n = 1 - 1/4 \cos^2 \pi/(n+2)$. See Z. Usiskin, *Annals of Mathematical Statistics* 35 (1964), 857–862; S. Trybula, *Zastosowania Matematyki* 8 (1965), 143–156.

3. Brute force (namely a program) finds eight solutions, of which the simplest is

$$A = \begin{bmatrix} \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix}, \quad B = \begin{bmatrix} \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix},$$

all with respective probabilities $\frac{17}{27}, \frac{10}{27}, \frac{10}{27}$. [If $\ast$ is also allowed, the unique solution is]

$$A = \begin{bmatrix} \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix}, \quad B = \begin{bmatrix} \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix}, \quad C = \begin{bmatrix} \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix}$$

has the property that every roll has exactly one die below the average and two above, with each of $A$, $B$, $C$ equally likely to be below; hence all three probabilities are $2/3$. See J. Moraleda and D. G. Stork, *College Mathematics Journal* 43 (2012), 152–159.]  

4. (a) The permutation (1 2 3 4 5 6) takes $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$. So $B$ versus $C$ is like $A$ versus $B$, etc. Also $Pr(A$ beats $C) = Pr(C$ beats $A) = Pr(B$ beats $D) = Pr(D$ beats $B) = \frac{3}{10}$; $Pr(A$ and $C$ tie) = $Pr(B$ and $D$ tie) = $\frac{3}{10}$.

(b) Assume by symmetry the players are $A$, $B$, $C$. Then the biggers are $(A, B, C, AB, AC, BC, ABC)$ with respective probabilities $(168, 216, 168, 48, 72, 36, 12)/720$.

(c) It’s $(A, AB, AC, ABC, ABCD)$ with probabilities $(120, 24, 48, 12, 0)/720$.

5. (a) If $A_k = 1001$ with probability $\frac{99}{1000}$, otherwise $A_k = 0$, but $B_k = 1000$ always, then $P_{1000} = \frac{99}{1000} \approx 0.0099$. (This example gives the smallest possible $P_{1000}$, because $Pr(A_1 - B_1) + \cdots + (A_n - B_n) > 0 \geq Pr((A_1 > B_1) \cdots (A_n > B_n)) = P_n$.)

(b) Let $E = \sum_{k=1}^{n} q_k \approx 0.67915$ be the probability that $B = 0$. Then $Pr(A > B) = \sum_{k=1}^{n} q_k (E + \sum_{j=0}^{k-1} q_{j+1}) \approx 0.574458$; $Pr(A < B) = \sum_{k=0}^{n} q_{k+1} (1 - E + \sum_{j=0}^{k-1} q_{j+1}) \approx 0.30807$; and $Pr(A = B) = Pr(A = B = 0) = E(1 - E) \approx 0.21790$ is also the probability that $AB = 0$.

(c) During the first $n_k$ rounds, the probability that either Alice or Bob has scored more than $m_k$ is at most $n_k (q_{k+1} + q_{k+2} + \cdots) = O(2^{-k})$; and the probability that neither has ever scored $m_k$ is $(1 - q_{k+1})^k < \exp(-q_k n_k) = \exp(-2^k/D)$. Also $m_k > n_k n_{k+1}$ when $k > 1$. Thus Alice “quite surely” wins when $k$ is even, but loses when $k$ is odd, as $k \to \infty$. [The American Statistician 43 (1989), 277–278.]

6. The probability that $X_i = 1$ is clearly $p_1 = 1/(n-1)$; hence $X_i = 0$ with probability $p_0 = (n-2)/(n-1)$. And the probability that $X_i = X_j = 1$ when $i < j$ is $p_1^2$. Thus (see exercise 20), $(X_i, X_j)$ will equal $(0, 1)$, $(1, 0)$, or $(0, 0)$ with the correct probabilities $p_{00}, p_{10}, p_{01}$. But $X_i = X_j = X_k = 1$ with probability $0$ when $i < j < k$.

7. $Pr(X_1 \cdots X_n = x_1 \cdots x_n) = a_{x_1 + \cdots + x_n}/(n - 2)^3$, where $a_0 = 2 \binom{n - 2}{3}$, $a_1 = \binom{n - 2}{2}$, $a_2 = 1$, otherwise $a_j = 0$.

8. Let $f_m(n) = \sum_{j=0}^{m} \binom{n}{j} (-1)^j (n + 1 - m)^{n-j}$, and define probabilities via $a_j = f_{n-j}(n-j)$ as in answer 7. (In particular, we have $f_0(n) = 1, f_1(n) = 0, f_2(n) = \binom{n - 2}{2}$, $f_3(n) = 2 \binom{n - 5}{5}, f_4(n) = 3 \binom{n - 3}{3} + \binom{n - 3}{2}^2$.) This definition is valid if we can prove that $f_m(n) \geq 0$ for $n \geq m$, because of the identity $\sum_{j=0}^{m} \binom{n}{j} f_{m-j}(n-j) = (n + 1 - m)^m$.

To prove that inequality, Schulte-Geers notes (see CMath(5-19)) that $f_m(n) = \sum_{k=0}^{m} \binom{m-k}{n-k}(n-m)^{n-k} = \sum_{k=0}^{m} (-1)^k (n-m)^{n-k} k^{n-k}$; these terms pair up nicely to yield $\sum_{k=0}^{m-1} k^{n-k} (n-m)^{n-k}$ for even $m$; and $\sum_{k=0}^{m} (-1)^k (n-m)^{n-k} k^{n-k}$ for odd $m$.

April 8, 2014

Chebyshev polynomial
Usiskin
Trybula
Moraleda
Stork
quite surely
Schulte-Geers
He also points out that there’s a much simpler asymmetrical solution: We can let
\( X_1, \ldots, X_{n-1} \) be independent random uniform bits, and set \( X_n = X_1 \oplus \cdots \oplus X_k \)

9. If \( 0 < k < n \), the probability that \( k \) of the variables have any particular setting is \( 1/2^k \), because the remaining variables have even parity as often as odd parity. So there’s \((n-1)\)-wise independence, but not \(n\)-wise.

10. Give probability \( 1/2 \) to 0 . . . 0 and 1 . . . 1; all other vectors have probability 0.

11. We can assume that \( n = 1 \), because \((X_1 + \cdots + X_n)/n\) and \((X_{n+1} + \cdots + X_{2n})/n\) are independent random variables with the same discrete distribution. Then \( \Pr(|X_1 + X_2 - 2\alpha| < 2|X_1 - \alpha| \geq \Pr(|X_1 - \alpha| + |X_2 - \alpha| \leq 2|X_1 - \alpha|) = \Pr(|X_1 - \alpha| \leq |X_1 - \alpha|) = (1 + \Pr(X_1 = X_2))/2 > 1/2 \). [This exercise was suggested by T. M. Cover.]

12. Let \( w = \Pr(A \text{ and } B) \), \( x = \Pr(A \text{ and } \bar{B}) \), \( y = \Pr(\bar{A} \text{ and } B) \), \( z = \Pr(\bar{A} \text{ and } \bar{B}) \). All five statements are equivalent to \( wz > xy \), or to \( \left| \frac{w}{z} \right| > 0 \), or to \( \mathbb{A} \text{ and } B \) are strictly correlated [see exercise 61]. [This exercise was suggested by E. Georgiadis.]

13. False in many cases. For example, take \( \Pr(\hat{A} \text{ and } \bar{B} \text{ and } \bar{C}) = \Pr(\hat{A} \text{ and } B \text{ and } C) = 2/7 \), and all other probabilities 1/7.

14. Induction on \( n \). [Philosophical Transactions 53 (1763), 370–418, proof of Prop. 6.]

15. If \( \Pr(C) > 0 \), this is the chain rule, conditional on \( C \). But if \( \Pr(C) = 0 \), it’s false by our conventions, unless \( A \) and \( B \) are independent.

16. If and only if \( \Pr(\overline{A \cap B} \cap C) = 0 \neq \Pr(B) \) or \( \Pr(\overline{A \cap C}) = 0 \).

17. 4/51, because four of the cards other than \( 5 \) are aces.

18. Since \( (M - X)(X - m) \geq 0 \), we have \((M \cdot E)X - (E \cdot X) + m \cdot E - m \cdot M \geq 0 \).

[See C. Davis and R. Blatia, AMM 107 (2000), 353–356, for generalizations.]

19. (a) The binary values of \( \Pr(X_n = 1) = E(X_n) \) for \( n = 0, 1, 2, \ldots \) are respectively \((0, 0, 0, 0) \), \((0, 0, 1, 1) \), \((0, 0, 1, 0) \), \((0, 0, 0, 1) \), \((0, 0, 0, 0) \), \((0, 0, 0, 0) \), \( \ldots \); thus they’re complemented reflections of the “magic masks” 7.13–(17). The answer is therefore \( (2^n - 1)/2^{n+1} - 1 = 1/(2^{n+1}) \).

(b) \( \Pr(X_0X_1 \ldots X_{n-1} = 00 \ldots 01) = 2^{n-1}/(2^n - 1) \) can be “read off” from the magic masks by ANDing and complementing. [See E. Lukacs, Characteristic functions (1960), 119, for related theory.]

(c) The infinite sum \( S \) is well defined because \( \Pr(S = \infty) = 0 \). Its expectation \( E(S) = \sum_{n=0}^{\infty} 1/(2^n + 1) \approx 0.5966 \) corresponds to the case \( z = 1/2 \) in answer 7.1.3–41(c).

By independence, \( S = \sum_{n=0}^{\infty} \Pr(X_n) = \sum_{n=0}^{\infty} 2^n/(2^n + 1)^2 \approx 0.44148 \).

(d) The parity number \( E R = \Pr \left((1, 1, \ldots, 1) \right) \) has the decimal value

\[ 0.41245403364010759778336136823845528308958 \]

and can be shown to equal \( \frac{1}{4} - \frac{1}{4} P \) where \( P = \prod_{n=0}^{\infty} \left(1 - 1/2^n\right) \) R. W. Gosper and R. Schoenfield, MTT AI Laboratory Memo 239 (29 February 1972), Hack 122, which is transcendental [K. Mahler, Mathematische Annalen 101 (1929), 342–366; 103 (1930), 532]. (Furthermore it turns out that \( 1/P - 1/2 = \prod_{n=0}^{\infty} 1/2^{n+1} \). Since \( R \) is binary, \( \text{var}(R) = (E(R)^2 - E(R)) \approx 0.24336 \).

(e) Zero (because \( \pi \) is irrational, hence \( p_0 + p_1 + \cdots = \infty \)). However, if we ask the analogous question for Euler’s constant \( \gamma \) instead of \( \pi \), nobody knows the answer.

(f) \( E X_n = 2 E X_n; \) in fact, \( \Pr(Y_0Y_1Y_2 \ldots = x_0x_1x_2 \ldots) \), for any infinite string \( x_0x_1x_2 \ldots \), is equal to \( 2 \Pr(X_0X_1X_2 \ldots = x_0x_1x_2 \ldots) \) mod 1, because we shift the binary representation one place to the left (and drop any carry). Thus in particular,
\[ EY_m Y_n = 2EX_m X_n = \frac{1}{4} EY_m EY_n \]  
when \( m \neq n \); \( Y_m \) and \( Y_n \) are negatively correlated because \( \text{cov}(Y_m, Y_n) = -\frac{1}{4} EY_m EY_n \).

(g) Clearly \( ET = 2ES \). Also \( ET^2 = 2ES^2 \), because \( EY_m Y_n = 2EX_m X_n \) for all \( m \) and \( n \). So \( \text{var}(T) = 2(\text{var}(S) + (ES)^2) - (2ES)^2 = 2\text{var}(S) - 2(ES)^2 \approx 0.17237 \).

20. Let \( p_j = EY_j \). We must prove, for example, that \( E(X_1(1 - X_2)(1 - X_3)X_4) = p_1(1 - p_2)(1 - p_3)p_4 \) when \( k \geq 4 \). But this is \( E(X_1X_2 - X_1X_4 - X_1X_3X_4) = p_1p_2 - p_1p_3p_4 + p_1p_4p_2 - p_1p_3p_4p_2 \).

21. From the previous exercise we know that they can't both be binary. Let \( X \) be binary and \( Y \) ternary, taking the values \( (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2) \) with probabilities respectively proportional to \( \{a, b, 3a - b - 3d, d, 1, 1\} \). Then \( EXY = 3/D \), \( EX = 2/D \), and \( EY = 3/2 \), where \( D = 4a + 2b + 4d + 2 \).

22. By (8) we have \( \text{Pr}(A_1 \cup \cdots \cup A_n) = E[A_1 \cup \cdots \cup A_n] = \text{E}[\{A_1, \ldots, A_n\}] \leq E[A_1 + \cdots + A_n] = \text{Pr}(A_1) + \cdots + \text{Pr}(A_n) \).

23. The hinted probability is \( \text{Pr}(X_s = 0 \text{ and } X_t + \cdots + X_{s-1} = s - r) \), so it equals \( (r-1)^{n-r}(1-p)^r \). To get \( B_{m,n}(p) \), sum it for \( r = n-m \) and \( n-m \leq n \). For an algebraic rather than probabilistic/combinatorial proof, see CMath, Exercise 8.17.

24. (a) The derivative of \( B_{m,n}(x) = \sum_{k=0}^{m} \binom{n}{k} x^k (1-x)^{n-k} \) is

\[
B_{m,n}(x) = \sum_{k=0}^{m} \binom{n}{k} x^k (1-x)^{n-k} - (n-k)x^k (1-x)^{n-1-k}
\]

\[ = n \sum_{k=0}^{m-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} - \sum_{k=0}^{m} \binom{n-1}{k} x^k (1-x)^{n-1-k} \]

\[ = -n \binom{n-1}{m} x^m (1-x)^{n-1-m}. \]

[See Karl Pearson, *Biometrika* 16 (1924), 202–203.]

(b) The hint, which says that \( \int_0^1 x^{a+b} dx < \int_0^1 x^a dx \) when \( 0 \leq a \leq b \), will prove that \( 1 - B_{m,n}(m/n) < B_{m,n}(m/n) \). It suffices to show that \( \int_0^1 x^a (1-x)^b dx < \int_0^1 x^a (1-x)^b dx \), because we have \( \int_0^1 x^a (1-x)^b dx < \int_0^1 x^a (1-x)^b dx \), and observe that \( (a - \epsilon)^n (a + \epsilon)^n \) is less than or equal to \( (a + \epsilon)^n (a - \epsilon)^n \) for \( 0 \leq \epsilon \leq a \), because the quantity

\[
\left( \frac{a - \epsilon}{a + \epsilon} \right)^n = e^{(a-1)(n-1)/n(1+n/\alpha)} \exp\left( -2e \left( 1 + e^2 \left( \frac{2a}{3a^2} + \frac{e^4}{5a^4} + \cdots \right) \right) \right)
\]

increases when \( a \) increases.

(c) Let \( t_k = \binom{n}{k} m^n (n-m)^{n-k} \). When \( m \geq n/2 \) we can show that \( 1 - B_{m,n}(m/n) = \sum_{k=0}^{m} t_k/n^m < B_{m,n}(m/n) = \sum_{k=0}^{m} t_k/n^m \), because \( t_{m+d} < t_{m+1-d} \) for \( 1 \leq d \leq n-m \). For \( r_d = t_{m+d} / t_{m+1-d} \), we have \( r_1 = m/(m+1) < 1 \); also

\[
\frac{r_d+1}{r_d} = \frac{(n-m+d)(n-m-d)m^2}{(m+1+d)(m+1-d)(n-m)^2} < 1,
\]

because \( (m+1)^2 - d^2)(n-m)^2 = ((n-m)^2 - d^2)m^2 = (2m+1)(n-m)^2 + (2m-n)mn^2 \).

[Peter Neumann proved in *Wissenschaftliche Zeitschrift der Technischen Universität Dresden* 15 (1966), 223–226, that \( m \) is the median. The argument in part (c) is due to Nick Lord, in *The Mathematical Gazette* 94 (2010), 331–332.]
25. (a) \((\binom{n}{k}) = \sum_{j=0}^{\min(k, n-k)} (\binom{n-k}{j}) - \binom{n}{n-k-1},\) summing over all partitions of \(\{1, \ldots, n\}\) into disjoint sets \(I \cup J \cup \{1\},\) where \(|I| = k, |J| = n-k-1, p_j = \prod_{i \in I} p_i, q_j = \prod_{i \in J} q_i.\) And \(q_j/(n-k) = p_j/(k+1) \geq 0 \iff p_j \leq (k+1)/(n+1).\)

(b) Given \(p_1, \ldots, p_{n-1},\) the quantity \((\binom{n}{k})\) is maximized when \(p_n = p,\) by (a). The same argument applies symmetrically to all indices \(j.\)

26. The inequality is equivalent to \(r_{n,k}^2 \geq r_{n,k-1}r_{n,k+1},\) which was stated without proof on pages 242–245 of Newton’s *Arithmetica Universalis* (1707), then finally proved by Sylvester many years later [Proc. London Math. Soc. 1 (1865), 1–16]. We have
\[
m_{n,k} = kp_n r_{n-1,k-1} + (n-k)q_n r_{n-1,k}; \text{ hence } n^2 (r_{n,k}^2 - r_{n,k-1}r_{n,k+1}) = (p_n r_{n-1,k-1} - q_n r_{n-1,k})^2 + (k^2 - 1)p_n^2 A + (k - 1)(n - 1 - k)p_n q_n B + (n - k)^2 q_n^2 C,\]
where \(A = r_{n-1,k-1} - r_{n-1,k-2}r_{n-1,k}, B = r_{n-1,k-1}r_{n-1,k} - r_{n-1,k-2}r_{n-1,k+1},\) and \(C = r_{n-1,k} - r_{n-1,k-1}r_{n-1,k+1}\) are nonnegative, by induction on \(n.\)

27. \(\sum_{k=0}^{n} \binom{n-k}{k} = \sum_{k=0}^{m} \binom{n-k}{k} \left(1 - p_{n-k+k}\right),\) by the same argument as before.

28. (a) \((\binom{n}{k}) = (\binom{n}{k} - \binom{n-1}{k})A + (\binom{n-1}{k} - \binom{n-2}{k})C\) and \(E g(X) = \sum_{k=0}^{n} \binom{n}{k} h_k,\) where \(A = (1 - p_{n-1})(1 - p_n), \ B = 1 - A - C, \text{ and } h_k = A g(k) + B g(k+1) + C g(k+2).\) If the \(p_j\)'s aren't all equal, we may assume that \(p_{n-1} < p < p_n.\) Setting \(p_{n-1} = p + \epsilon \text{ and } p_n = p - \epsilon,\) \(\epsilon > 0,\) changes \(A, B, C\) to \(A = A + \delta, B = B - 2\delta, C = C + \delta,\) \(\delta = (p_n - p)(p - p_{n-1}).\)

(b) Suppose \(E g(X)\) is maximum, and that \(r\) of the \(p_j\)'s are 0 and \(s\) of them are 1. Let \(a\) satisfy \((n - r - s)a + s = np\) and assume that \(0 < p_{n-1} < a < p_n < 1.\) As in part (a) we can write \(E g(X) = a A + B + C g(k+2).\) And if \(\alpha - 2\beta + \gamma > 0,\) we could increase it with a similar transformation, using \(\delta = -\min(p_{n-1}, 1 - p_n).\) Therefore \(\alpha - 2\beta + \gamma = 0,\) and we can repeat the transformation of (a) until every \(p_j\) is 0, 1, or \(a.\)

(c) For this function \(g(k) = 0 \leq k \leq m\) we have \(\alpha - 2\beta + \gamma = \binom{n}{k} - \binom{n}{k-1};\) this difference cannot be positive if the choice of \(p_1, \ldots, p_n\) is optimum. In particular we cannot have \(s \geq m.\) If \(r > 0\) we can make \(p_{n-1} = 0 \text{ and } p_n = a,\) so that \(\binom{n}{k} = \binom{n-1}{k} + (1-a)^{n-k} \text{ and } \binom{n}{k-1} = \binom{n-1}{k-1} + (1-a)^{n-k-1} - \binom{n}{k} a^{n-k} - \binom{n}{k-1} a^{n-k-1} + a^{n-k} - a^{n-k-1} - \binom{n}{k} a^{n-k}.\) But then the ratio \(\binom{n}{k} / \binom{n}{k-1} = (n-r+m)a / ((n-s)(1-a)) > 1,\) \(\text{ if } r > 0.\) Similarly if \(s > 0\) we can set \(p_{n-1} = p_n = (a, 1),\) getting the ratio \(\binom{n}{k} / \binom{n}{k-1} = (n-r+m)a / ((n-s)(1-a)) > 1.\) In this case \(\binom{n}{k} > \binom{n}{k-1}\) if and only if \(np = m+1;\) we can transform without changing \(E g(X),\) until \(s = 0\) and each \(p_j = p.\)

Reference: *Annals of Mathematical Statistics* 27 (1956), 713–721. The coefficients \((\binom{n}{k})\) also have many other important properties; see exercise 7.2.1.5-63, and the survey by J. Pitman in *J. Combinatorial Theory* A77 (1997), 29–303.

29. The result is obvious when \(m = 0\) or \(n;\) and there's a direct proof when \(m = n - 1: B_{m,n}(p) = 1 - p^m \geq 1 - p^{m+1} \geq (1 - p)n(1 - p) > 0.\) So the result is also clear when \(p = 0 \text{ or } 1.\)

If \(p = (m+1)/n,\) where \(B_{m,n}(p) = (1 - p^m)(m+1)/(1 - p) = (n-m)(m+1)/n.\) So if \(m > 0 \text{ and } p = m/(n-1),\) we can apply exercise 28(c) with \(p_1 = \cdots = p_{n-1} = p = m / (n-1) = 1.

\[
B_{m,n}(p) \geq \sum_{k=0}^{m} \binom{m}{k} p^{k-1}(1 - p)^{m-k} = B_{m,n-1}(p).
\]

April 8, 2014
When $1 \leq m < n - 1$, let $Q_{m,n}(p) = B_{m,n}(p) - R_{m,n}(p)$. The derivative

$$Q_{m,n}(p) = (n-m)\binom{n}{m}(1-p)^{m-1}(1-p)^{n-m+1},$$

where $A = (m+1)^{m-1}/\binom{n}{m} > 1$ and $F(p) = p^{n+1}/(1-p)^n + 1$, begins positive at $p = 0$, eventually becomes negative but then is positive again at $p = 1$. (Notice that $F(0) = 0$, and $F(p)$ increases dramatically until $p = (m+1)/(n+1)$; then it decreases to $F(1) = 1$.) The facts that $Q_{m,n}(\frac{m+1}{n}) > 0$ and $Q_{m,n}(0) = Q_{m,n}(1)$ now complete the proof, because $Q_{m,n}(p)$ changes sign only once in $[0, \frac{m+1}{n}]$. [Annals of Mathematical Statistics 36 (1965), 1272-1278.]

30. (a) $\text{Pr}(X_k = 0) = n/(n+1)$; hence $p = n^2/(n+1)^2 > 1/e \approx 0.368$.

(b) Let $p_{km} = \text{Pr}(X_k = m)$. Assume that these probabilities are fixed for $1 \leq k < n$, and let $x_m = p_{km}$. Then $x_0 = x_2 = 2x_3 + 3x_4 + \cdots$; and we want to minimize

$$p = \sum_{m=1}^n (A_m + (m-1)A_0)x_m$$

in nonnegative variables $x_1, x_2, \ldots$, where $A_m = \text{Pr}(X_1 + \cdots + X_{n-1} \leq n - m)$, subject to the condition $\sum_{m=1}^n m x_m = 1$. Since all coefficients of $p$ are nonnegative, the minimum is achieved when all $x_m$ for $m \geq 1$ are zero except for some value $m = m_n$, which minimizes $(A_m + (m-1)A_0)/m$. And $m_n \leq n+1$, because $A_m = 0$ whenever $m > n$. Similarly $m_1, \ldots, m_{n-1}$ also exist.

(c) We want to minimize $p(t) = \sum_{m=1}^n (\binom{n}{m} t^{n-m})$, where $m = \cdots = m_n = t$.

When $|t| = q - 1$ we have $t > n/q$; hence $p(t) > s(q) = \sum_{k=0}^{\min(n,t)} \binom{n}{k} q^{n-k}/n^k$.

Now $s(q) = B_{q,n}(q/n) - \binom{n}{q} q^q/(n-q)^q$ [see exercise 23]. Since $B_{q,n}(q/n) > 1/2$, it follows that $s(q) > p(n+1) = n^q/(n+1)^q$ whenever $10 < q < n - 10$ and $n \geq 100$. Furthermore a careful analysis shows that $s(1) < s(2) < \cdots < s(10)$ whenever $n \geq 100$. After checking all cases for $n < 100$ by computer, we conclude that $p(n+1)$ is optimum.

The following alternative solution by Ernst Schulte-Geers needs no computer assistance. The inequality of Samuels in exercise 29 implies that

$$B_{m,n}(p) \geq \left(1 - \frac{1}{f(m,n,p) + 1}\right)^n$$

for $p \leq \frac{m+1}{n}$, where

$$f(m,n,p) = \frac{(m+1)(1-p)}{(n-m)p},$$

because we can set $x = ((1-p)(n+1))/((1-p)(m+1))$ in the arithmetic–geometric mean inequality $x^m/n^m \leq (m-x)/(n-x)$, now if $m_1 = \cdots = m_n = t \leq n+1$, let $p = 1/t$ and $m = [(n+1)/p] + 1$. Then $p \leq (m+1)/(n+1)$ and $f(m,n,p) \geq n$; hence $B_{m,n}(p) \geq n^q/(n+1)^q$.

The first solution above is by John H. Elton (to appear), who has verified computationally that the joint distributions in (a) are optimum when $n \leq 20$. Are those distributions in fact optimum for all $n$? Uriel Feige has conjectured more generally that $\text{Pr}(X_1 + \cdots + X_n \geq \frac{n+1}{2}) > 1/e$ whenever $X_1, \ldots, X_n$ are independent nonnegative random variables with $\text{E}X_k \leq 1$; see SICOMP 35 (2006), 964-984.]

31. Suppose, for example, that $n = 4$. The reliability polynomial is the sum of the reliability polynomials for the minterms of $f$; so it suffices to show that the result is true for functions like $x_1 \land x_2 \land \bar{x}_3 \land \bar{x}_4 = x_1(1-x_2)(1-x_3)\cdot x_4$. And it’s clear that $\text{Pr}(A_1 \cap A_2 \cap A_3 \cap A_4) = \text{Pr}(A_1 \cap A_2 \cap A_3 \cap A_4)$.

(See exercise 7.1.1-12; also recall the inclusion-exclusion principle.)

32. The $2^n$ minterm probabilities in the previous answer must all be nonnegative, and the only sum to 1. We’ve already stipulated that $\pi_i = 1$, so the sum-to-1 condition is automatically satisfied. (The condition stated in the exercise when $I \subseteq J$ is necessary but not sufficient; for example, $\pi_{12}$ must be $\geq \pi_1 + \pi_2 - 1$.)

April 8, 2014
33. The three events \((X, Y) = (1, 0), (0, 1), (1, 1)\) occur with probabilities \(p, q, r\), respectively. The value of \(E[X | Y] = 1\), \(r/(q+r)\), \(r/(q+r)\) in those cases. Hence the answer is \(px + (q + r)z^{r/(q+r)}\). (This example demonstrates why univariate generating functions are not used in the study of conditional random variables such as \(E[X | Y]\). But we do have the simple formula \(E[X | Y = k] = \left([z^k] \frac{\partial}{\partial z} G(1, z)\right) / ([z^k] G(1, z))\).

34. The right-hand side is 
\[
\sum_{\omega} E(X | Y = Y) Pr(\omega) = \sum_{\omega} Pr(\omega) \sum_{\omega} X(\omega') Pr(\omega') | Y = Y(\omega') | Pr(Y = Y(\omega'))
\]
\[
= \sum_{\omega} Pr(\omega) \sum_{\omega} X(\omega') Pr(\omega') | Y = Y(\omega') | Pr(Y = Y(\omega'))
\]
\[
= \sum_{\omega} X(\omega') Pr(\omega') \sum_{\omega} Pr(\omega) | Y = Y(\omega') | Pr(Y = Y(\omega')).
\]

35. Part (a) is the special case \(Y = Z\) of part (b). And both parts are true because (b) is just (12) in the probability spaces conditioned by \(Z\). (It’s likewise true that \(E(E(X | Y) | Z) = E(E(X | Y, Z) | Z)\); this identity underlies exercise 91.)

36. (a) \(E(f(X))\); (b) \(E(f(Y) | g(X))\). Generalizing (12). Proof: \(E(f(Y) | g(X) | Y) = \sum_y f(g(y)) E(g(y) | X = y) Pr(Y = y) = \sum_y f(g(y)) E(g(y) | X = y) = E(f(Y) | g(X)).\)

37. If we’re given the values of \(X_1, \ldots, X_{k-1}\), the value of \(X_k\) is equally likely to be any of the \(n + 1 - k\) values in \(\{1, \ldots, n\} \setminus \{X_1, \ldots, X_{k-1}\}\). Hence its average value is 
\[(1 + \cdots + n - X_k - \cdots - X_{k-1})/(n + 1 - k).\] We conclude that \(E(X_k | X_1, \ldots, X_{k-1}) = (n(n + 1)/2 - X_k - \cdots - X_{k-1})/(n + 1 - k).\) Incidentally, the sequence \(Z_0, Z_1, \ldots,\) defined by \(Z_j = (n + j)X_1 + (n + j - 1)X_2 + \cdots + (n - j)X_{n + 1 - j} + (j + 1)n(n + 1)/2\) for \(0 \leq j \leq n\) and \(Z_n = -n - 1\) is itself a martingale.

38. Let \(t_{m,n}\) be the number of restricted growth strings of length \(m + n\) that begin with \(01 \cdots (m-1)\) (the number of set partitions of \(\{1, \ldots, m + n\}\) in which each of \(\{1, \ldots, m\}\) appears in a different block.) The generating function \(\sum_{n \geq 0} t_{m,n} z^n/n!\) turns out to be \(exp(e^z - 1 + mz)\); hence \(t_{m,n} = \sum_k \binom{m}{n}^k (m^n - k)\).

Suppose \(M = \max(X_1, \ldots, X_{k-1}) + 1\). Then \(Pr(X_k = j) = t_{M,n-k}/t_{M,n}\) for \(0 \leq j \leq M\), and \(t_{M,n-k}/t_{M,n}\) for \(j = M\). Hence \(E(X_k | X_1, \ldots, X_{k-1}) = \binom{M}{k} t_{M,n-k}/t_{M,n}\).

39. (a) Since \(E(K | N = n) = pn\) we have \(E(K | N) = pn\).

(b) Hence \(E(k) = \mathbb{E}E(K | N) = E(pn) = pn\).

(c) Let \(p_k = Pr(N = n, K = k) = (e^{-\mu} \mu^k/n!) \binom{n}{k} (1 - p)^{n-k}(1 - p)^k/p^k\times f(n - k)\), where \(f(n) = (1 - p)^\mu n!\). Then \(E(N | K = k) = \sum_n n p_n/k \sum_n p_n\). Since \(n f(n - k) = k f(n - k) + (n - k) f(n - k)\) and \(n f(n) = (1 - p) f(n)\), the answer is \(k(1 + p)\mu^n\); hence \(E(N | K) = K + (1 - p)\mu^n\). [G. Grimmett and D. Stirzaker, Probability and Random Processes (Oxford: 1982), 33.7.]

40. If \(p = Pr(X > m)\) clearly \(EX < (1 - p)m + pm\). [We also get this result from (15), by taking \(S = x \mid x \leq m\), \(f(x) = M - x, s = M - m\).]

41. (a) Convex when \(a \geq 1\) or \(a = 0\); otherwise neither convex nor concave. (However, \(x^a\) is concave when \(0 < a < 1\) and convex when \(a > 0\), if we consider only positive values of \(x\).) (b) Convex when \(n\) is even or \(n = 1\); otherwise neither convex nor concave. (This function is \(\int_0^x e^{n-1} e^{-x} (n-1)\), according to 1.2.11.3-15; so \(f''(x)/x > 0\) when \(n \geq 3\) is odd.) (c) Convex. [In fact \(f(|x|)\) is convex whenever \(f(z)\) has a power series with nonnegative coefficients, convergent for all \(z\).] (d) Convex, provided of course that we allow \(f\) to be infinite in the definition (19). April 8, 2014
42. We can show by induction on \( n \) that \( f(p_1 x_1 + \cdots + p_n x_n) \leq p_1 f(x_1) + \cdots + p_n f(x_n) \), when \( p_1, \ldots, p_n \geq 0 \) and \( p_1 + \cdots + p_n = 1 \), as in exercise 6.2.2–36. The general result follows by taking limits as \( n \to \infty \). [The quantity \( p_1 x_1 + \cdots + p_n x_n \) is called a “convex combination” of \( \{x_1, \ldots, x_n\} \); similarly, \( EX \) is a convex combination of \( X \) values. Jensen actually began his study by assuming only the case \( p = q = \frac{1}{2} \) of (19).]

43. \( f(EX) = f(E(E(X | Y))) \leq E(f(E(X | Y))) \leq E(E(f(X) | Y)) = E(f(X)). \) [S. M. Ross, *Probability Models for Computer Science* (2002), Lemma 3.2.1.]

44. The function \( f(xy) \) is convex in \( y \) for any fixed \( x \). Therefore \( g(y) = E f(xY) \) is convex in \( y \). It’s a convex combination of convex functions. Also \( g(y) \geq f(EX) = f(0) = g(0) \) by \( (\alpha) \). Hence \( 0 \leq a \leq b \) implies \( g(0) \leq g(a) \leq g(b) \) by convexity of \( g \). [S. Boyd and L. Vandenberghe, *Convex Optimization* (2004), exercise 3.10.]

45. \( \Pr(X > 0) = \Pr(|X| \geq 1) \); set \( m = 1 \) in \( (\phi) \).

46. \( EX^2 \geq (EX)^2 \) in any probability distribution, by Jensen’s inequality, because squaring is convex. We can also prove it directly, since \( EX^2 - (EX)^2 = E(X - EX)^2 \).

47. We always have \( Y \geq X \) and \( Y^2 \leq X^2 \). (Consequently \( (22) \) yields \( \Pr(X > 0) = \Pr(Y > 0) \geq (EY)^2/(EY^2) \geq (EX)^2/(EX^2) \) when \( EX \geq 0 \).

48. \( \Pr(a - X_1 - \cdots - X_n > 0) \geq a^2/(\sigma_1^2 + \cdots + \sigma_n^2) \), by exercise 47. [This inequality was also known to Chebyshev; see J. Math. Pures et Appl. (2) 19 (1874), 157–160. In the special case \( n = 1 \) it is equivalent to “Cantelli’s inequality.”

\[
\Pr(X \geq EX + a) \leq \var(X)/(\var(X) + a^2), \quad \text{for } a \geq 0;
\]

see *Atti del Conferenza Internazionale dei Matematici* 6 (Bologna: 1928), 47–59, [6–57].

49. \( \Pr(X = 0) = 1 - \Pr(X > 0) \leq (\frac{EX - (EX)^2}{EX}) = (\frac{EX^2 - (EX)^2}{EX^2}) = (\frac{EX^2}{EX^2})^2 - 1 \). Some authors call this inequality the “second moment principle,” but it is strictly weaker than \( (22) \).

50. (a) Let \( Y_j = X_j \) if \( X_j > 0 \), otherwise \( Y_j = 0 \). Then \( Y_1 + \cdots + Y_m = [X > 0] \).

Hence \( \Pr(X > 0) = \sum_{j=1}^m \Pr(Y_j) \); and \( \Pr(X > 0) \). [This identity requires only that \( X_j \geq 0 \), is elementary yet nonlinear, so it apparently lay undiscovered for many years. See D. Akhous, *Discrete Math.* 76 (1989), 168.]

(b) Since \( X_j \in [0, 1] \), we have \( \Pr(X_j > 0) = EX_j = p_j \); and \( EX_j/X_j = 0 = \sum_{j=1}^m \frac{p_j}{m} = E(X/X) \).

(c) \( \Pr(J = 1) = \sum_{j=1}^m \Pr(J = 1) = \sum_{j=1}^m \frac{p_j}{m} = E(X/X) = \Pr(X_j = 1) \).

(d) Since \( J \) is independent we have \( t_j = E(X/X) = 1 \).

(e) The right side is \( (EX/X) = \sum_{j=1}^m (p_j/EX/X) = (EX/X) / \sum_{j=1}^m (p_j/EX/X) \).

51. If \( g(q_1, \ldots, q_m) = 1 - f(p_1, \ldots, p_m) \) is the dual of \( f \), where \( q_j = 1 - p_j \), a lower bound on \( g \) gives an upper bound on \( f \). For example, when \( f \) is \( p_1 p_2 p_3 \lor p_2 p_3 p_4 \lor p_3 p_4 \), \( g \) is \( q_1 q_2 + q_2 q_3 + q_3 q_4 + q_4 q_5 + q_5 q_6 \); so the inequality \( (24) \) gives \( g(q_1, \ldots, q_6) \geq q_1 q_2 / (1 + q_2 + q_3 q_5 + q_5 q_6) \). In particular, \( g(1, 1, 1, 1, 1, 1) > 0.039 \) and \( f(9, 9, 9, 9, 9) < 0.961 \).

52. \( \binom{n}{k} p^k (1-p)^{n-k} \).

53. \( f(p_1, \ldots, p_n) \geq p_1 p_2 (1-p_3) + p_2 p_3 (1-p_4) + \cdots + p_n p_1 (1-p_2) (1-p_3) + p_1 p_2 p_3 (1-p_4) + \cdots + p_n p_1 p_2 (1-p_3) \).

Monotonicity is not required when applying this method; nor need the implications be prime. The result is exact when the implications are disjoint.

April 8, 2014

convex combination
Jensen
Ross
Boyd
Vandenberghe
Jensen’s inequality
convex
Chebyshev
Cantelli’s inequality
second moment principle
Akhous
dual of a Boolean function
prime implications
54. (a) $Pr(X > 0) \leq E[X] = \binom{w}{2}p^2$, because $E[X_{uvw}] = p^2$ for all $u < v < w$.

(b) $Pr(X > 0) \geq (E[X])^2/(E[X^2])$, where the numerator is the square of (a) and the denominator can be shown to be $\binom{w}{2}p^2 + 12\binom{w}{3}p^3 + 30\binom{w}{4}p^4 + 20\binom{w}{5}p^5$. For example, the expansion of $X^2$ contains 12 terms of the form $X_{uvw}X_{u'v'w'}$ with $u < v < w < u'$, and each of those terms has expected value $p^5$.

55. A BDD for the corresponding Boolean function of $\binom{w}{5}$ = 45 variables has about 1.4 million nodes, and allows us to evaluate the true probability $(1 - p)^{15}G(p/(1 - p))$ exactly, where $G(z)$ is the corresponding generating function (see exercise 7.1.4–25). The results are: (a) $30/37 \approx .811 < 35165158461687345 \approx .999 < 15$; (b) $10/109 \approx .092 < 418024678447086252691034958990199032987399/(4 \times 10^{10}) \approx .105 < .12$.

56. The upper bound is $\mu = \frac{\lambda}{\theta}$; the lower bound divides this by $1 + \mu$. The exact asymptotic value can be obtained using the principle of inclusion and exclusion and its "bracketing" property, as in Eq. 7.2.1.4–(7); the result is $1 - e^{-\theta}$. See P. Erdős and A. Rényi, Magyar Tudományos Akadémia Mat. Kut. Int. Közl. 5 (1960), 17–61, 33

57. To compute $E[X|X_{uvw} = 1]$ we sum $Pr(X_{u'v'w'}|X_{uvw} = 1)$ over all $\binom{w}{3}$ choices of $u' < v' < w'$. If $\{u', v', w'\} \cup \{u, v, w\}$ has $t$ elements, this probability is $p^{3-t(t-1)/2}$, and there are $\binom{3}{2}$ such cases. Consequently we get

$$Pr(X > 0) \geq \binom{w}{5}p^5/\left( \binom{w}{3}p^3 + 3\binom{w}{2}p^3 + 3\binom{w}{1}p^2 + \binom{w}{0}p^0 \right).$$

In this problem the lower bound turns out to be the same using either inequality; but the derivation here was easier.

58. $Pr(X > 0) \leq \binom{w}{5}p^{(k-1)/2}$. The lower bound, using the conditional expectation inequality as in the previous answer, divides this by $\sum_{k=0}^{5} \binom{w}{k}p^{(k-1)/2} - t(t-1)/2$.

59. (a) The hypotheses imply that $a_0a_1b_0b_1 \leq c_0c_1d_0d_1$. The key observation is that

$$c_1d_0(c_0c_1d_0d_1) - (a_0a_1b_0b_1) =$$

$$c_1d_0(c_0c_1d_0d_1 - a_0a_1b_0b_1) = (c_1d_0 - a_0a_1b_0b_1)(c_0c_1d_0d_1 - a_0a_1b_0b_1).$$

Thus the result holds when $c_1d_0 \neq 0$. If $c_1 = 0$ we have $a_0a_1b_0b_1 + a_1b_0 + a_1b_1 = a_0a_1b_0b_1 + a_1b_0$, and a similar argument applies to the case $d_0 = 0$.

All four hypotheses hold with equality when $a_0 = b_0, a_1 = b_1$, and the other variables are 1, yet the conclusion is that $1 \leq 2$. Conversely, when $b_1 = c_1 = 2$ and the other variables are 1, we have $a_0b_1 < c_1d_0$ but conclude only that $6 \leq 6$.

(b) Let $A_k = \sum_{i=0}^{k} a_{ji}b_{ji} \leq 2^{(k-1)}$ for $l = 0$ and $l = 1$, and define $B_k, C_k, D_k$ similarly from $b_{ji}a_{ji}, c_{ji}a_{ji}$, $d_{ji}a_{ji}$. The hypotheses for $m = 2$ and $k = 0$ prove $A_0B_0 \leq C_0D_0$, by induction on $k$. Hence, by part (a), we have the desired inequality $(A_0 + A_1)(B_0 + B_1) \leq (C_0 + C_1)(D_0 + D_1)$. [This result is due to R. Ahsiswe and D. E. Daykin, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 43 (1978), 183–185, who stated it in the language of the next exercise.]

(c) Now $A_n = a_0 + \cdots + a_{n-1}$, and define $B_n, C_n, D_n$ similarly. If $A = \infty$, we'll have $A_n > C_nD_n$ for some $n$. But $C_nD_n \geq C_nD_n$, contra (b).

60. (a) We can consider each set to be a subset of the nonnegative integers. Let $\overline{S} = \alpha(S)$, $\overline{\beta(S)} = \beta(S) \in \mathbb{F}$, $\overline{g(S)} = \gamma(S)$, $\phi(S) = \delta(S)$, $\phi(S) = \gamma(S)$, $\overline{\delta(S)} = \delta(S)$, $\overline{\delta(S)} = \delta(S)$. Then $\overline{\delta(S)} = \delta(S)$, $\overline{\delta(S)} = \delta(S) = \delta(S)$ for all sets $S$.
61. (a) In the hinted case we can let $\alpha(S) = f(S)\mu(S)$, $\beta(S) = g(S)\mu(S)$, $\gamma(S) = f(S)g(S)\mu(S)$, $\delta(S) = \mu(S)$; the four functions theorem yields the result. The general case follows because we have $E(fg) - E(f)E(g) = E(\tilde{f}) - E(\tilde{f})E(\tilde{g})$, where $\tilde{f} = f(S) - f(\emptyset)$ and $\tilde{g} = g(S) - g(\emptyset)$. [See Commun. Math. Physics 22 (1971), 89–103.]

(b) Changing $f(S)$ to $\theta f(S)$ and $g(S)$ to $g(\theta g)$ changes $E(fg) - E(f)E(g)$ to $\theta(\theta f g) - \theta f \theta g(\theta g)$, for all real numbers $\theta$ and $\phi$.

(c) If $S$ and $T$ are supported, then $R = S \cap T$ and $U = S \cup T$ are supported. Furthermore we can write $S = S \cap \{s_1, \ldots, s_r\}$ and $T = T \cap \{t_1, \ldots, t_s\}$ where the sets $S_i = S \cap \{s_1, \ldots, s_i\}$ and $T_j = T \cap \{t_1, \ldots, t_j\}$ are supported, as are their unions $U_{i,j} = S \cup T_j$ for $0 \leq i \leq k$ and $0 \leq j \leq l$. By (iii) we know that $\mu(U_{i+1,j})/\mu(U_{i,j}) \leq \mu(U_{i+1,j+1})/\mu(U_{i,j+1})$ when $0 \leq i < k$ and $0 \leq j < l$. Multiplying these inequalities for $0 \leq i < k$, we obtain $\mu(U_{k,j})/\mu(U_{0,j}) \leq \mu(U_{k,0})/\mu(U_{0,0}) \leq \mu(U_{0,0})/\mu(U_{0,0}) = \mu(U)/\mu(T)$. Therefore

$$\frac{\mu(S)}{\mu(T)} \leq \frac{\mu(R)}{\mu(T)} \leq \frac{\mu(S)}{\mu(T)}.$$}

[Note: Random variables with this distribution are often confusingly called “Poisson trials,” a term that conflicts with the (quite different) Poisson distribution of exercise 39.]

(e) Choose $c$ in the following examples so that $\sum_S \mu(S) = 1$. In each case the supported subsets are subsets of $U = \{1, \ldots, m\}$.

(i) Let $\mu(S) = c(1/r_1 \cdot 1/r_2 \cdot \ldots \cdot 1/r_m)$, where

$$0 < r_1, \ldots, r_m \leq \leq r_m.$$ \, 

(ii) Let $\mu(S) = c_1$ when $S = \{1, \ldots, j\}$ and $1 \leq j \leq m$, otherwise $\mu(S) = 0$. (If $p_1 = \cdots = p_m$, $\mu(S) = 1$, the FKG inequality reduces to Chebychev’s monotonic inequality of exercise 1.23–31.)

(iii) Let

$$\mu(S) = c_1 \mu_1(S \cap U_1) \mu_2(S \cap U_2) \ldots \mu_k(S \cap U_k),$$

where each $\mu_j$ is a distribution on $U_j \subseteq U$ that satisfies (ss). The subsets $U_1, \ldots, U_k$ needn’t be disjoint.

(iv) Let $\mu(S) = c \mu(f(S))$, where $f$ is a submodular set function on the supported subsets of $U$: $f(S \cup T) + f(S \cap T) \leq f(S) + f(T)$. (See Section 7.6.)

62. A Boolean function is essentially a set function whose values are 0 or 1. In general, under the Bernoulli distribution or any other distribution that satisfies the condition of exercise 61, the FKG inequality implies that any monotone increasing Boolean function is positively correlated with any other monotone increasing Boolean function, but negatively correlated with any monotone decreasing Boolean function. In this case, $f$ is monotone increasing but $g$ is monotone decreasing. Adding an edge doesn’t disconnect a graph; deleting an edge doesn’t invalidate a 4-coloring.

[Notice that when $f$ is a Boolean function, $E(f)$ is the probability that $f$ is true under the given distribution. The fact that $\text{covar}(f, g) \leq 0$ in such a case is equivalent to saying that the conditional probability $\Pr(f | g) \leq \Pr(f)$]

63. If $\omega$ is the event $(Z_0 = a, Z_1 = b)$, we have $\Pr(\omega) = a$ and $E(1 | Z_0 = a) = (p_0 + 2p_1)/2$. Hence $p_0 = p_1 = p_2 = 0$, and $p_0 = p_1$ if these conditions are necessary and sufficient for $E(1 | Z_0 = Z_0) = Z_0$.

64. (a) No. Consider the probability space consisting of just three events $(Z_0, Z_1, Z_2) = (0, 0, 0), (1, 1, 2), (0, 2, 0)$, each with probability $1/3$. Call those events $a, b, c$. Then $E(Z_1 | Z_0 = 0) = 0 = Z_0(a, b, c) = E(Z_1 | Z_0 = 1) = 2 = Z_1(a, b, c) = E(Z_1 | Z_0 = 2) = 2 = Z_1(c)$.

(b) Yes. We have $\sum_{n+1}(z_{n+1} - z_n) \Pr(Z_0 = a, \ldots, Z_{n+1} = z_{n+1}) = 0$ for all fixed $(z_0, \ldots, z_n)$. Sum these to get $\sum_{n+1}(z_{n+1} - z_n) \Pr(Z_n = z_n, Z_{n+1} = z_{n+1}) = 0$.

65. Observe first that $E(Z_n | Z_0, \ldots, Z_k) = E(\sum_{n+1} | Z_0, \ldots, Z_k) = Z_n, \ldots, Z_k)$ whenever $k < n$. Thus $E(Z_{n+1} | Z_0, \ldots, Z_{n+1}) = Z_{n+1}$ for all $n \geq 0$. Hence $E(Z_{m(n+1)} | Z_{m(0)}, \ldots, Z_{m(n)}) = Z_{m(n+1)} = Z_{m(n)}$ as in the previous exercise.
66. We need to specify the joint distribution of \( \{Z_1, \ldots, Z_n\} \), and it’s not difficult to see that there is only one solution. Let \( p(\sigma_1, \ldots, \sigma_n) = \text{Pr}(Z_1 = \sigma_1, \ldots, Z_n = \sigma_n \mid n) \) when \( \sigma_1, \ldots, \sigma_n \) are each \pm 1. The martingale law \( p(\sigma_1, \ldots, \sigma_1, 0) = p(\sigma_1, \ldots, \sigma_1, 1) = p(\sigma_1, \ldots, \sigma_n) = p(\sigma_1, \ldots, \sigma_n, 0) = n \). Theorem gives \( p(\sigma_1, \ldots, \sigma_n) = (1 + 2n \sum_{k=1}^{n-1} (1 + 2k[\sigma_k \sigma_{k+1} > 0])/2n) \). When \( n = 3 \), for example, the eight possible cases \( z_1 z_2 z_3 = 123, 123, \ldots, 123 \) occur with probabilities \( (15, 3, 1, 5, 1, 3, 1, 5)/48 \).

67. (a) You “always” (with probability 1) make \( 2n^2 - (1 + 2 + \cdots + 2^n) = 1 \) dollar.
(b) Your total payments are \( X = X_0 + X_1 + \cdots \) dollars, where \( X_n = 2^n \) with probability \( 2^n \), otherwise \( X_n = 0 \). So \( E X_n = 1, \text{ and } E X = E X_0 + E X_1 + \cdots = \infty \).
(c) Let \( (T_n) \) be a sequence of uniformly random bits; and define the fair sequence \( Y_n = (-1)^{n+1} 2^n \) if there is no nth bet. Then \( Z_n = Y_0 + \cdots + Y_n \).

The famous adventurer Casanova lost a fortune in 1754 using this strategy, which he called “the martingale” in his autobiography Histoire de ma vie. A similar betting scheme had been proposed by Nicolas Bernoulli (see P. R. de Montmort, Essai d’Analyse sur les Jeux de Hazard, second edition (1713), page 402); and the perplexities of (a) and (b) were studied by his cousin Daniel Bernoulli, whose important paper in Commentarii Academii Scientiarum Imperialis Petropolitanae 5 (1731), 175-192, has caused this scenario to become known as the St. Petersburg paradox.

68. (a) Now \( Z_n = Y_1 + \cdots + Y_n \), where \( Y_n = (-1)^n N \geq n \). Again \( \text{Pr}(Z_n > 1) = 1 \).
(b) The generating function \( g(z) \) equals \( z(1 + g(z)^2)/2 \), so he must win \$2 if the first bet loses. Hence \( g(z) = 1 - \sqrt{1 - z^2} \); and the desired probability is \( [z^n] C_{n-1/2}/[z^{n+1}]/(2^n) \), where \( C_k \) is the Catalan number \( (k^2) \).
(c) \( \text{Pr}(N > n) = [z^n] (1 - g(z)) \), or \( [z^n] (1 + z)/\sqrt{1 - z^2} = \left( \binom{n+1}{n+1/2} \right)^2 \).
(d) \( E N = g'(1) = \infty \). (It’s also \( \sum_{n=1}^{\infty} \text{Pr}(N > n) \), where \( \text{Pr}(N > n) \sim 1/\sqrt{\pi n} \).
(e) Let \( p_n = \text{Pr}(Z_n > m) \) for all \( n \geq 0 \). Clearly \( p_0 = 1/2 \) and \( p_n = (1 + p_{n-1})/2 \) for \( m > 0 \); this recurrence has the solution \( p_n = (m + 1)/(m + 2) \). The answer is \( 1/((m + 1)(m + 2)) \); it’s another probability distribution with infinite mean.
(f) The generating function \( g_n(z) \) for the number of times \( m \) is hit satisfies \( g_{n+1}(z) = (1 + g_{n+1}(z))/2 \) for \( m > 0 \). So \( g_m(z) = h_m(z)/h_{m+1}(z) \) for \( m > 0 \), where \( h_m(z) = 2m - 2(m + 1)z \), and \( g_1(z) = 2. \) [A distribution with finite mean! See W. Feller, An Introduction to Probability Theory 2, second edition (1971), XII.12.]

69. Each permutation of \( n \) elements corresponds to a configuration of \( n + 1 \) balls in the urn. For Method 1, the number of corresponding “red balls” is the position of \( m \) in Method 2, it is the value in position 1. For example, we’d put 3 1 2 4 into node (2, 3) with respect to Method 1 but into (3, 2) with respect to Method 2. (In fact, Methods 1 and 2 construct permutations that are inverses of each other.)

70. Start with the permutation \( 1 \ldots c-1 \) at the root, and use Method 1 of the previous exercise to generate all \( n!/(c-1)! \) permutations in which these elements retain that order. A permutation \( j \) in position \( P_j \) for \( 1 \leq j < c \) stands for \( P_j = P_{j+1} \) balls of color \( j \), where \( P_0 = 0 \) and \( P_c = n + 1 \); for example, if \( c = 3 \), the permutation 3 1 4 2 would correspond to node (2, 2, 1). The resulting tuples \( (A_1, \ldots, A_n)/(n + 1) \), then form a martingale for \( n = c, c + 1, \ldots \), uniformly distributed (for each \( n \)) among all \( \binom{n}{c-1} \) compositions of \( n + 1 \) into \( c \) positive parts.

We can also use this setup to deal with Fólya’s two-color model when there are \( r \) red balls and \( b \) black balls at the beginning. Imagine \( r + b \) colors, then identify the
first $r$ of them with red. This model was first studied by D. Blackwell and D. Kendall, *J. Applied Probability* 1 (1964), 284–296.

71. If $m = r' - r$ and $n = b' - b$ we must move $m$ times to the right and $n$ times to the left; there are $\binom{m+n}{n}$ such paths. Every path occurs with the same probability, because the numerators of the fractions are $r \cdot (r+1) \cdot \ldots \cdot (r' - 1) \cdot b \cdot (b+1) \cdot \ldots \cdot (b' - 1) = r^{r'} b^{b'}$ in some order, and the denominators are $(r+b) \cdot (r+b+1) \cdot \ldots \cdot (r'+b' - 1) = (r+b)^{m+n}$.

The answer, $\binom{m+n}{n} r^{r'} b^{b'} (r+b)^{m+n}$, reduces to $1/(r'+b' - 1)$ when $r = b = 1$.

72. Since all paths have the same probability, this expected value is the same as $E(X_1 X_2 \ldots X_m)$, which is obviously $1/(m+1)$. (Thus the $X$s are very highly correlated: This expected value would be $1/2^m$ if they were independent. Notice that the probability of an event such as $(X_2 = 1, X_3 = 0, X_0 = 1)$ is $E(X_2 (1-X_3)X_0) = 1/3 - 1/4$.)

The far-reaching ramifications of such exchangeable random variables are surveyed in O. Kallenberg’s book *Probabilistic Symmetries and Invariance Principles* (2005).]

73. $f(r, n) = r \binom{m+n}{n} \sum_k \binom{n}{k} (-1)^k q_{k, n-r+k}$, where $q_k = a_k / (k+1)$.

This answer can be obtained the hard way, as the author first did, by calculating $f(r, n)$ for numerous small values of $n$ and $r$, then psyching out a pattern, and eventually proving the general formula by induction. But there’s a higher-level way to proceed: Consider the Doob martingale where $Q_k$ is the random variable $X_1 = \cdots = X_m = 0$ and $X_{m+1} = 1$. Then $E(Q_{k+1} | X_1, \ldots, X_n) = 0$ except that $f(1, k) = (k+1) / (m+1) (m+2))$ for $0 \leq k < m$. Hence we know that $q_k = [k < m]$ implies $f(r, k) = [r = 1] | n < m | (n+1)$. The equations $\frac{1}{r} f(1, n) + \frac{1}{r+1} f(2, n) = f(1, n-1)$ now yield $f(2, n) = n(n+1) (q_{n-1} - q_n)$, etc.

74. Node $(r, n+2-r)$ on level $n$ is reached with probability $\binom{n}{r} / n!$, proportional to an Eulerian number (see Section 5.1.3). (Indeed, we can associate the permutations of $\{1, \ldots, n+1\}$ that have exactly $r$ inversions with this node, using Method 1 as in exercise 69.)

Reference: *Communications on Pure and Applied Mathematics* 2 (1949), 59–70.

75. As before, let $R_n = X_0 + \cdots + X_n$ be the number of red balls at level $n$. Now we have $E(X_{n+1} | X_0, \ldots, X_n) = 1 - R_n / (n+2)$. Hence $E(R_{n+1} | R_n) = (n+1) R_n / (n+2) + 1$, and the definition $Z_n = (n+1) R_n / (n+2) + 1 / n+1 / 2$ is a natural choice.

Different paths have quite different probabilities in this model, hence there’s no reason why we should have a single value of $Z_n$ for each node on level $n$ of the diagram. For example, the martingale $Z_n = (1 + H_{n+1}) S_n - T_n$, where $S_n = \sum_{k=0}^{n} (X_k - 1/2)$ and $T_n = \sum_{k=1}^{n} (X_k - 1/2) H_{k+1}$, depends on the entire path and converges better as $n \to \infty$.

76. No. For example, let $Z_0 = X$, $Z_1 = Y$, and $Z_1 = Z_1 = X + Y$, where $X$ and $Y$ are independent with $E(X) = E(Y) = 0$. Then $E(Z_1 | Z_0) = Z_0$ and $E(Z_1' | Z_0') = Z_0'$, but $E(Z_1 + Z_1' | Z_0 + Z_0') = 2(Z_0 + Z_0')$. (On the other hand, if $(Z_n)$ and $(\langle Z_n \rangle)$ are both martingales with respect to some common sequence $(X_n)$, then $(Z_n + \langle Z_n \rangle)$ is also.)

77. $E(Z_{n+1} | Z_0, \ldots, Z_n) = E(E(Z_{n+1} | Z_0, \ldots, Z_n, X_0, \ldots, X_n) | Z_0, \ldots, Z_n)$, which equals $E(E(Z_{n+1} | X_0, \ldots, X_n) | Z_0, \ldots, Z_n)$ because $Z_n$ is a function of $X_0, \ldots, X_n$; and that equals $E(Z_n | Z_0, \ldots, Z_n) = Z_n$. (Furthermore $\langle Z_n \rangle$ is a martingale with respect to, say, a constant sequence. But not with respect to every sequence.)

A similar proof shows that any sequence $(Y_n)$ that is fair with respect to $\langle X_n \rangle$ is also fair with respect to itself.

78. $E(Z_{n+1} | V_0, \ldots, V_n) = E(E(Z_{n+1} | V_0, \ldots, V_n) | V_0, \ldots, V_n) = Z_n$.

The converse holds with $V_0 = Z_0$ and $V_n = Z_n / Z_{n-1}$ for $n > 0$, provided that $Z_{n-1} = 0$ implies $Z_n = 0$, and that we define $V_0 = 1$ at that happens.

April 8, 2014

Blackwell
Kendall
correlated
exchangeable random variables
Kallenberg
Doob martingale
Eulerian number
runs
fair with respect to
79. \( Z_n = V_0 V_1 \ldots V_n \), where \( V_0 = 1 \) and each \( V_n \) for \( n > 0 \) is independently equal to \( q/p \) (with probability \( p \)) or to \( p/q \) (with probability \( q \)). Since \( E(V_n) = q + p = 1 \), \( \{ V_n \} \) is multiplicatively fair. [See A. de Moivre, *The Doctrine of Chances* (1718), 102–154.]

80. (a) True; in fact \( E(f(Y_0 \ldots Y_{n-1})Y_n) = 0 \) for any function \( f \).
   (b) False: For example, let \( Y_5 = \pm 1 \) if \( Y_5 > 0 \), otherwise \( Y_5 = 0 \). (Hence permutations of a fair sequence needn't be fair. The statement is, however, true if the \( Y \)'s are independent with mean zero.)
   (c) False if \( n_1 = 0 \) and \( m = 1 \) (or if \( m = 0 \); otherwise true. (Sequences that satisfy \( E(Y_{n_1} - EY_{n_1}) \ldots (Y_{n_m} - EY_{n_m})) = E(Y_{n_1} - EY_{n_1}) \ldots E(Y_{n_m} - EY_{n_m}) \) are called *totally uncorrelated*. Such sequences, with \( EY = 0 \) for all \( n \), are not always fair; but fair sequences are always totally uncorrelated.)

81. Assuming that \( X_0, \ldots, X_n \) can be deduced from \( Z_0, \ldots, Z_n \), we have \( a_nX_n + b_nX_{n-1} = Z_n = E(Z_{n+1} \mid Z_0, \ldots, Z_n) = E(a_{n+1}X_{n+1} + b_{n+1}X_n \mid X_0, \ldots, X_n) = a_{n+1}(X_n + X_{n-1}) + b_{n+1}X_n \) for \( n \geq 1 \). Hence \( a_{n+1} = b_n \), \( b_{n+1} = a_n - a_{n+1} = b_{n-1} - b_n \); and we have \( a_n = F_{n-1}, b_n = F_{n-2} \) by induction, verifying the assumption.


82. (a) \( Z_n = A_n/C_n \), where \( A_n = 4 - X_1 - \cdots - X_n \) is the number of aces and \( C_n \) is the number of cards remaining after you've seen \( n \) cards. Hence \( E Z_{n+1} = (A_{n+1}/C_{n+1})(A_n/C_n) = (1 - 1/A_n)(A_n/C_n) = A_n/C_n\) \((A_n/C_n) \), \( C_n = A_n/C_n \). (In every generalization of Polya's urn for which the \( n \)th step adds \( k \) balls of the chosen color, the ratio \( \text{red/ (red + black) } \) is always a martingale, even when \( k \) is negative, as long as \( k \) balls of the chosen color remain. This exercise represents the case \( k = -1 \).)

(b) This is the optional stopping principle in a bounded-time martingale.

(c) \( Z_N = A_N/C_N \) is the probability that an ace will be next. [*Ace Now* is a variant of R. Connelly's game "Say Red"; see *Fallbears Review* 9 (1974), 702.]

83. \( Z_n = \sum_{i=1}^{n-1}(X_i - E X_i) \) is a martingale, for which we can study the bounded stopping rules \( \min(m, N) \) for any \( m \). But Svante Janson suggests a direct computation, beginning with the formula \( S_n = \sum_{i=1}^{n-1}X_i[N \geq n] \) where \( N \) might be \( \infty \). We have \( E(X_n[N \geq n]) = E(E X_n[N \geq n]) \), because \( N \geq n \) is a function of \( \{X_0, \ldots, X_{n-1}\} \), hence independent of \( X_n \). And since \( X_n \geq 0 \), we have \( E S_n = \sum_{i=1}^{n-1}E(X_i[N \geq n]) = \sum_{i=1}^{n-1}E(E X_i)[N \geq n] = E \sum_{i=1}^{n-1}(E X_i)[N \geq n] \), which is \( E(\sum_{i=1}^{n-1}E X_i) \). The equation might be \( \infty = \infty \).]

[Wald's original papers, in *Annals of Mathematical Statistics* 15 (1944), 283–296, 16 (1945), 287–293, solved a somewhat different problem and proved more.]

84. (a) We have \( f(Z_n) = f(E(Z_{n+1} \mid Z_0, \ldots, Z_n)) \leq E(f(Z_{n+1}) \mid Z_0, \ldots, Z_n) \) by Jensen's inequality. And the latter is \( E(f(Z_{n+1}) \mid f(Z_0), \ldots, f(Z_n)) \) as in answer 77. [Incidentally, D. Gilat has shown that every nonnegative submartingale is \( \{|Z_n|\} \) for some martingale \( \{Z_n\} \); see *Annals of Probability* 5 (1977), 475–481.]

(b) Again we get a submartingale, provided that we also have \( f(x) \leq f(y) \) for \( a \leq x \leq y \leq b \). [J. L. Doob, *Stochastic Processes* (1953), 295–296.]

85. Since \( \langle B_x \rangle \mid \langle R_x \rangle = (1 - R_x/B_x) \) is a martingale by (27), and since \( f(x) = 1/x \) is convex for positive \( x \), \( \langle R_x + B_x \rangle/B_x = (R_x/B_x) + 1 \) is a submartingale by exercise 84. (A direct proof could also be given.)

86. The rule \( N_{n+1}(Z_0, \ldots, Z_n) = \max(Z_0, \ldots, Z_n) \times x + n + 1 \) is bounded. If \( \max(Z_0, \ldots, Z_{n-1}) < x \) then we have \( Z_N < x \), where \( N \) is defined by (32); similarly, if \( \max(Z_0, \ldots, Z_{n-1}) \geq x \) then \( Z_N \geq x \). Hence \( \Pr(\max(Z_0, \ldots, Z_n) \geq x) = (E Z_n)/x \) by Markov's inequality, and \( E Z_N \leq E Z_n \) in a submartingale.

April 8, 2014 de Moivre

independent totally uncorrelated uncorrelated

Fibonacci numbers

MacQueen

Polya’s urn

optimal stopping principle

Connelly

Say Red

Janson

Wald

Jensen’s inequality

Gilat

one negative submartingale

Doob

Markov’s inequality
87. This is the probability that $Z_n$ becomes $3/4$, which also is $\Pr(\max(Z_0, \ldots, Z_n) \geq 3/4)$. But $E Z_n = 1/2$ for all $n$, hence (33) tells us that it is at most $(1/2)(3/4) = 2/3$.

The exact value can be calculated as in the following exercise. It turns out to be

$$\sum_{k=0}^{\infty} \frac{\left(\frac{3}{2}\right)^k}{\left(\frac{4k+2}{4k+3}\right)^2} = \frac{1}{2} H_{1/3} - \frac{1}{2} H_{1/2} + \frac{1}{3} = \frac{\pi}{2} - \frac{1}{2} \ln 2 \approx 0.439.$$  

88. (a) We have $S > 1/2$ if and only if there comes a time when there are more red balls than black balls. Since that happens if and only if the process passes through one of the nodes $(2, 1), (3, 2), (4, 3), \ldots$, the desired probability is $p_1 + p_2 + \cdots$, where $p_k$ is the probability that node $(k+1, k)$ is hit before any of $(j+1, j)$ for $j < k$.

All paths from the root to $(k+1, k)$ are equally likely, and the number of paths meeting our restrictions is equivalent to the paths in 7.2.16–(8). Thus we can use Eq. 7.2.16–(23) to show that $p_k = 1/(2k-1) = 1/2k$; and $1-1/2+1/3-1/4+\cdots = \ln 2$.

(b, c) If $p_k$ is the probability of hitting node $(t-1)k+1, k)$ before any previous $((t-1)j+1, j)$, a similar calculation using the $t$-ary ballot numbers $C_k^n$ yields $p_k = (t-1)(1/(tk-1) - 1/(tk))$. Then $\sum_{k=1}^{\infty} p_k = 1 + (1-1/t)H_{1/t}$ (see Appendix A).

Notes: We have $\Pr(S = 1/2) = 1 - \ln 2$, since $S$ is always $\geq 1/2$. But we cannot claim that $\Pr(S \geq 2/3)$ is the sum of cases that pass through $(2, 1), (4, 2), (6, 3), \ldots$, because the supremum might be $2/3$ even though the value $2/3$ is never reached. Those cases occur with probability $\pi/\sqrt{27}$; hence $\Pr(S = 2/3) \geq 2\pi/\sqrt{27} - \ln 3 \approx .111$. A determination of the exact value of $\Pr(S = 2/3)$ is beyond the scope of this book, because we’ve avoided the complications of measure theory by defining probability only in discrete spaces; we can’t consider a limiting quantity such as $S$ to be a random variable, by our definitions! But we can assign a probability to the event that max$(Z_0, Z_1, \ldots, Z_n) > x$, for any given $x$ and $n$, and we can reason about the limits of such probabilities.

With the help of deeper methods, Ernst Schulte-Geers and Wolfgang Stadje have proved that the supremum is a.s. reached. Hence $\Pr(S = 2/3) = 2\pi/\sqrt{27} - \ln 3$; indeed, $\Pr(S$ is rational) = 1, since only rationals are reached; and $\Pr(S = (t-1)/t) = (2-3/t)H_{1/t} - (1+2/t)H_{1-2/t} - (t-2)/(t-1)$. (See arXiv:1401.1805 [math.PR].)

89. Set $Y_n = X_n - p_n, a_n = -p_n, b_n = 1 - p_n$. (Incidentally, exercise 1.2.10–22 gives an upper bound for this quantity that has quite a different form.)

90. (a) Apply Markov’s inequality to $\Pr(e^{|Y|} + \cdots + Y_n \geq e^{t|x|})$.
(b) $e^{x} \leq e^{(y-x)}(y-\epsilon) + e^{\epsilon} (y+\epsilon) = e^{(y-\epsilon)} + ye^{\epsilon} (t) \because$ because the function $e^{x}$ is convex.

(c) We have $f'(y) = -p + pe^t/(q + pe^t)$ and $f''(y) = ppe^t/(q + pe^t)^2$; hence $f(0) = f'(0) \leq 1/A, \therefore$ the geometric mean of $q$ and $pe^t$, $(q + pe^t)^{1/2}$, is less than or equal to the arithmetic mean, $(q + pe^t)/2$.

(d) Set $c = b-a, p = a/c, q = b/c, Y = Y/c, t = ct, h(t) = e^{\epsilon(\frac{t|x|}{c})}.$

(e) In $E(c^{(2x)^2/4} + Y_h(t) \cdots c^{(2x)^2/4} + Y_h(t))$ the terms involving $h(t)$ all drop out, because $(Y_n)$ is fair. So we’re left with the constant term, $e^{|x|c/2}$.

(f) Let $t = 2c|x|, to make $e^{|x|c/2} = t - x^2/c$.

91. $E(Z\geq n | X_0, \ldots, X_n) = E(E(Q | X_0, \ldots, X_n, X_{n+1}) | X_0, \ldots, X_n)$, which equals $E(E(Q | X_{n+1}) | X_0, \ldots, X_n)$, which is $E(Q | X_0, \ldots, X_n)$ by (23). Apply exercise 77.

92. $Q_0 = E X_m = 1/2$. If $n < m$ we have $Q_n = E X_m | X_0, \ldots, X_n$, which is the same as $E(X_{n+1} | X_0, \ldots, X_n)$ (see exercise 72); and this is $(1 + X_1 + \cdots + X_n)/(n+2)$, which is the same as $\frac{Z_n}{n+2}$.

93. Everything goes through exactly as before, except that we must replace the quantity ($m-1)^{n-1}$ by the generalized expected value, which is $\sum_{k=1}^{m} (1 - p_{n+k})$.

94. If the $X$’s are dependent, the Doob martingale still is well defined; but when we write its fair sequence as an average of $\Delta(x_1, \ldots, x_t)$ there is no longer a nice

April 8, 2014
formula such as (40). In any formula for $\Delta$ that has the form $\sum_{x} p_x (Q(\ldots x_n \ldots) - Q(\ldots x_{n+1} \ldots))$, $\Pr(X_n = x_n, X_{n+1} = x_{n+1}, \ldots) / (\Pr(X_n = x_n) \Pr(X_{n+1} = x_{n+1}, \ldots))$ must equal $\sum_{x} p_x$, so it must be independent of $x_n$. Thus (41) can’t be used.

95. False; the probability of only one red ball at level $n$ is $1/(n+1) = \Omega(n^{-1})$. But there are a.s. more than 100 red balls, because that happens with probability $(n-99)/(n+1)$.

96. Exercise 1.2.10–21, with $en$ equal to the bound on $|X - n/2|$, tells us that (i) is q.s. and that (ii), (iii) are a.s. To prove that (iv) isn’t a.s., we can use Stirling’s approximation to show that $(n^{n/2})/2^n$ is $\Theta(n^{-1/2})$ when $k = \sqrt{n}$; consequently $\Pr(|X| < \sqrt{n}) = \Theta(1)$. A similar calculation shows that (ii) isn’t q.s.

97. We need to show only that a single bin q.s. receives that many. The probability generating function for the number of items $H$ that appear in any particular bin is $G(z) = (n - 1 + z)/n$, where $N = [n^{1+\delta}]$. If $r = \frac{1}{2} n^{\delta}$, we have

$$\Pr(H \leq r) \leq \left(\frac{1}{2}\right)^{-r} G\left(\frac{1}{2}\right) = 2^{-r} \left(1 - \frac{1}{2n}\right)^{2nr} \leq 2^{-r} \left(1 - \frac{1}{2n}\right)^{2n^{\delta}-1} \leq 2^{-r+1} e^{-r},$$

by 1.2.10–(24). And if $r = 2n^{\delta}$ we have

$$\Pr(H \geq r) \leq 2^{-r} G(2) = 2^{-r} \left(1 + \frac{1}{n}\right)^{nr/2} \leq 2^{-r} \left(1 + \frac{1}{n}\right)^{nr/2} \leq 2^{-r} e^{r/2},$$

by 1.2.10–(25). Both are exponentially small. [See Knuth, Motwani, and Pittel, Random Structures & Algorithms 1 (1990), 1–14, Lemma 1.]

98. Let $E_n$ be the expected number of reduction steps, and suppose $F(n) = k$ with probability $p_k$, where $\sum_{k=1}^{n} p_k = 1$ and $\sum_{k=1}^{n} k p_k = g \geq g_k$. (The values of $p_1, \ldots, p_n$, and $g$ might be different, in general, every time we compute $F(n)$.)

Let $\Sigma^b_n = \sum_{k=1}^{n} 1/g_k$. Clearly $E_0 = 0$. And if $n > 0$, we have by induction

$$E_n = 1 + \sum_{k=1}^{n} p_k E_{n-k} \leq 1 + \sum_{k=1}^{n} p_k \Sigma^b_{n-k} = 1 + \sum_{k=1}^{n} p_k (\Sigma^b_n - \Sigma^b_{n-k+1})$$

$$= \Sigma^b_n + 1 - \sum_{k=1}^{n} p_k \Sigma^b_{n-k+1} \leq \Sigma^b_n + 1 + \sum_{k=1}^{n} p_k \frac{k}{g_k} \leq \Sigma^b_n.$$


99. The same proof would work, provided that induction could be justified, if we were to do the sums from $k = -\infty$ to $n$ and define $\Sigma^b_n = -\sum_{k=-\infty}^{n-1} 1/g_k$ when $a > b$. (For example, that definition gives $-\Sigma^b_{n+1} = 1/g_{n+1} + 1/g_{n+2} \leq 2/g_n$.)

And indeed, it becomes a proof, by induction on $m$, that we have $\min(m, E_n) \leq \Sigma^b_n$ for all $n$, if we begin with $\min(m, E_n) = 1 + \sum_{k=-\infty}^{n} p_k \min(m-1, E_{n-k})$. [This problem is exercise 1.6 in Randomized Algorithms by Motwani and Raghavan (1995).]
INDEX AND GLOSSARY

He writes indexes to perfection.
— OLIVER GOLDSMITH, Citizen of the World (1762)

When an index entry refers to a page containing a relevant exercise, see also the answer to that exercise for further information. An answer page is not indexed here unless it refers to a topic not included in the statement of the exercise.

<table>
<thead>
<tr>
<th>Entry</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$ (Euler's constant)</td>
<td>24</td>
</tr>
<tr>
<td>$\nu(x)$ (sew executive sum)</td>
<td>13</td>
</tr>
<tr>
<td>$\pi$ (circle ratio)</td>
<td>14</td>
</tr>
<tr>
<td>$\phi$ (golden ratio)</td>
<td>12, 22</td>
</tr>
<tr>
<td>a.s.: Almost surely</td>
<td>11–12, 20, 35, 36</td>
</tr>
<tr>
<td>Ace Now</td>
<td>8, 19</td>
</tr>
<tr>
<td>Abhmede, Rudolph</td>
<td>30</td>
</tr>
<tr>
<td>Aldous, David John</td>
<td>29</td>
</tr>
<tr>
<td>Almost sure</td>
<td>11</td>
</tr>
<tr>
<td>Analysis of algorithms</td>
<td>9, 20</td>
</tr>
<tr>
<td>Arithmetic–geometric mean inequality</td>
<td>27, 35</td>
</tr>
<tr>
<td>Asymptotic methods</td>
<td>11, 12, 16</td>
</tr>
<tr>
<td>Atomic events</td>
<td>1</td>
</tr>
<tr>
<td>Azuma, Kazuo (吾妻一興)</td>
<td>9, 20</td>
</tr>
<tr>
<td>Ballot numbers</td>
<td>35</td>
</tr>
<tr>
<td>Bayes, Thomas</td>
<td>13</td>
</tr>
<tr>
<td>BDD (binary decision diagram)</td>
<td>5, 30</td>
</tr>
<tr>
<td>Bell, Eric Temple, numbers $\omega$</td>
<td>15</td>
</tr>
<tr>
<td>Bernoulli, Daniel</td>
<td>32</td>
</tr>
<tr>
<td>Bernoulli, Jacques (Jakob = James)</td>
<td>32</td>
</tr>
<tr>
<td>distribution, multivariate</td>
<td>14, 18, 20</td>
</tr>
<tr>
<td>Bernoulli, Nicolas (Nikolaus)</td>
<td>32</td>
</tr>
<tr>
<td>Beta distribution</td>
<td>14</td>
</tr>
<tr>
<td>Bhattacharya, Rajendra (बाहुभाल राजेंद्र), 24</td>
<td></td>
</tr>
<tr>
<td>Bienaymé, Irène Jules, inequality</td>
<td>4</td>
</tr>
<tr>
<td>Bin-packing problem</td>
<td>11, 20</td>
</tr>
<tr>
<td>Binary notation</td>
<td>14</td>
</tr>
<tr>
<td>Binary random variables, 2, 3, 5, 13–15, 20</td>
<td></td>
</tr>
<tr>
<td>Bingo</td>
<td>12–13</td>
</tr>
<tr>
<td>Binomial distribution</td>
<td>14, 28</td>
</tr>
<tr>
<td>cumulative</td>
<td>14–15, 27</td>
</tr>
<tr>
<td>Bit vectors</td>
<td>3, 9, 13–14</td>
</tr>
<tr>
<td>Blackwell, David Harold</td>
<td>33</td>
</tr>
<tr>
<td>Boolean functions</td>
<td>5, 15, 29, 31</td>
</tr>
<tr>
<td>dual of</td>
<td>29</td>
</tr>
<tr>
<td>monotone</td>
<td>5, 31</td>
</tr>
<tr>
<td>symmetric</td>
<td>16</td>
</tr>
<tr>
<td>Boolean random variables</td>
<td>see Binary random variables</td>
</tr>
<tr>
<td>Boolean vectors</td>
<td>see Bit vectors</td>
</tr>
<tr>
<td>Boyce, Stephen Poythress</td>
<td>29</td>
</tr>
<tr>
<td>Bracket notation</td>
<td>2</td>
</tr>
<tr>
<td>Bracketing property</td>
<td>30</td>
</tr>
<tr>
<td>Brown, John O'Connor</td>
<td>22</td>
</tr>
<tr>
<td>Caneelli, Francesco Paolo, inequality</td>
<td>29</td>
</tr>
<tr>
<td>Casanova de Seingalt, Giacomo Girolamo</td>
<td>32</td>
</tr>
<tr>
<td>Catalan, Eugène Charles, numbers</td>
<td>32</td>
</tr>
<tr>
<td>Chain rule</td>
<td>13, 24</td>
</tr>
<tr>
<td>Chebyshev (Теодоричев, Пафунтий Львович = Чебышёв, Пафунтий Львович)</td>
<td>29</td>
</tr>
<tr>
<td>inequality</td>
<td>4, 9, 16</td>
</tr>
<tr>
<td>monotonic inequality</td>
<td>31</td>
</tr>
<tr>
<td>polynomials</td>
<td>23</td>
</tr>
<tr>
<td>Chesterton, Gilbert Keith</td>
<td>22</td>
</tr>
<tr>
<td>Chicks</td>
<td>15</td>
</tr>
<tr>
<td>Circle ratio ($\pi$)</td>
<td>14</td>
</tr>
<tr>
<td>Cliques</td>
<td>16</td>
</tr>
<tr>
<td>CMath: Concrete Mathematics, a book</td>
<td></td>
</tr>
<tr>
<td>by R. L. Graham, D. E. Knuth, and O. Patashnik</td>
<td>23, 25</td>
</tr>
<tr>
<td>Coin tosses</td>
<td>11–12, 19, 20</td>
</tr>
<tr>
<td>Compositions</td>
<td>32</td>
</tr>
<tr>
<td>Concave functions</td>
<td>4, 28</td>
</tr>
<tr>
<td>Conditional expectation</td>
<td>2–3, 16–19</td>
</tr>
<tr>
<td>inequality</td>
<td>5, 16, 30</td>
</tr>
<tr>
<td>Conditional probability</td>
<td>1, 13–14, 31</td>
</tr>
<tr>
<td>Connelly, Robert, Jr.</td>
<td>34</td>
</tr>
<tr>
<td>Convex combinations</td>
<td>29</td>
</tr>
<tr>
<td>Convex functions</td>
<td>4, 8, 15–16, 20, 29, 34, 35</td>
</tr>
<tr>
<td>Correlated random variables</td>
<td>17–18, 24, 33</td>
</tr>
<tr>
<td>Correlation inequalities</td>
<td>17</td>
</tr>
<tr>
<td>Covariance</td>
<td>2, 14, 17, 21, 31</td>
</tr>
<tr>
<td>Cover, Thomas Merrill</td>
<td>13, 24</td>
</tr>
<tr>
<td>Cumulative binomial distribution</td>
<td>14–15, 27</td>
</tr>
<tr>
<td>Darwin, Charles Robert</td>
<td>21</td>
</tr>
<tr>
<td>Davis, Chandler</td>
<td>24</td>
</tr>
<tr>
<td>Daykin, David Edward</td>
<td>30</td>
</tr>
<tr>
<td>de Méré, Abraham</td>
<td>34</td>
</tr>
<tr>
<td>martingale</td>
<td>19</td>
</tr>
<tr>
<td>de Montmort, Pierre Rémond</td>
<td>32</td>
</tr>
<tr>
<td>Diaconis, Persi Warren</td>
<td>iv</td>
</tr>
<tr>
<td>Discrete probabilities</td>
<td>1</td>
</tr>
<tr>
<td>Doob, Joseph Leo</td>
<td>6, 9, 34</td>
</tr>
<tr>
<td>martingales</td>
<td>9–10, 20, 33, 35</td>
</tr>
<tr>
<td>Dual of a Boolean function</td>
<td>29</td>
</tr>
<tr>
<td>Eggenberger, Florian</td>
<td>6</td>
</tr>
<tr>
<td>Elton, John Hancock</td>
<td>27</td>
</tr>
<tr>
<td>Enveloping series</td>
<td>30</td>
</tr>
</tbody>
</table>

April 8, 2014

37
Erdős, Pál (= Paul), 30.
Euler, Leonhard (Εὐλερος, Λεόναρς = Εўлер, Левонарь), constant γ, 24.
Eulerian numbers, 33.
Events, 1-3.
Exchangeable random variables, 33.
Expected value, 2-5, 14-16, see also Conditional expectation.
Fair sequences, 7, 10, 19, 34.
with respect to a sequence, 7, 33.
Families of sets, 17.
Feige, Uriel (_Flagel), iv, 27.
Feller, William (=[ Vilim = Willy = William]), 32.
Fibonacci, Leonardo of Pisa (= Leonardo filio Bonacci Pisanus), dice, 12.
martingale, 19.
numbers, 12, 34.
First moment principle, 4, 16.
FKG inequality, 5, 17, 31.
Forni, Cornelis Marius, 17.
Four functions theorem, 17, 30.
Friedman, Bernard, urn, 19.
Games, 4, 8, 13.
Garey, Michael Randolph, 11.
Generating functions, 15, 28, 29, 32.
Geometric mean and arithmetic mean, 27, 35.
Georgiadis, Evangelos (Γεώργιαδης, Γεώργιαδος), 21.
Gilat, David, 34.
Gini, Jean, 17.
Golden ratio (ϕ), 12, 22.
Goldsmith, Oliver, 37.
Gosper, Ralph William, Jr., 24.
Graham, Ronald Lewis (Γραμμα), 37.
Grinstead, Geoffrey Richard, 28.
HAKMEM, 24.
Harmonic numbers, fractional, 35.
Hashing, 9-10, 20.
Hoeffding, Wassily, 9, 15.
inequality, 9-10, 20.
Inclusion and exclusion, 27, 30.
Incomplete beta function, 14.
Independent events, 2.
Independent random variables, 1, 7, 9, 10, 13-15, 20, 34.
k-wise, 1, 13.
Infinite mean, 32, 34.
Integer multilinear representation, see Reliability polynomials.
International, ii, iii.
Inverses, 32.

James White, Phyllis Dorothy, 11.
Janson, Carl Svante, iv, 34.
Jensen, Johan Ludvig William Valdemar, 29.
inequality, 4, 16, 29, 31.
Johnson, David Stifler, 11.
Joint distribution, 13, 31.
k-cliques, 16.
k-wise independence, 1, 13.
Kallenberg, Olav Herbert, 33.
Karap, Richard Manning, 36.
Kasteleyn, Pieter Willem, 17.
Kendall, David George, 33.
Knuth, Donald Ervin (Κουνθ), i, iii, iv, 33, 36, 37.
Kolmogorov, Andrei Nikolayevich (Ко́лмогоро́в, Андре́й Никола́евич), 9.
inequality, 9.

Lake Wobegon dice, 12.
Large deviations, see Tail inequalities.
Lipschitz, Rudolph Otto Sigismund, condition, 10.
Loop, running time of, 20.
Lord, Nicholas John, 25.
Lukács, Eugene (= Jenő), 24.

MacQueen, James Buford, 34.
Magic masks, 24.
Mahler, Kurt, 24.
Markov (= Markoff), Andrei Andreyevich (Марков, Андре́й Андре́евич), the elder, 4.
inequality, 4, 5, 16, 34, 35.
Martingale differences, see Fair sequences.
Martingale differences, 6-11, 18-20, 28.
with respect to a sequence, 7, 19, 33.
Maximal inequality, 8-9, 20.
McDiarmid, Colin John Hunter, 10.
Median value of a random variable, 14.
Mengden, Nicolai Alexandrovitch von (Менден, Николай Александрович), 11.
Method of bounded differences, 10.
Minterns, 27.
Moivre, Abraham de, 19, 34.
Monotone Boolean functions, 5, 31.
Montmor, Pierre Rémond de, 32.
Moore, Harold Calvin Marston, constant, 24.
Moottani, Rajeev (राजीव मोट्टानी), 36.
MPR: Mathematical Preliminaries Redux, v.
Multiplicatively fair sequences, 19.
Multivariate Bernoulli distribution, 14, 18, 20.
Multivariate total positivity, see FKG inequality.

April 8, 2014
NanoHing, 12–13.
Negative binomial distribution, cumulative, 14.
Negatively correlated random variables, 18, 25.
Neumann, Peter, 25.
Neville-Neil, George Vernon, III, 30.
Newton, Isaac, 26.
Nonnegative submartingales, 9, 34.
Nonnegatively correlated random variables, 17.
Nontransitive dice, 12.
NP-complete problems, 11.
One-sided estimates, 16.
Optional stopping principle, 8, 34.
\& (power set, the family of all subsets), 30.
Pairwise independent random variables, 1, 13.
Paradoxes, 12, 13, 32.
Parity number, 24.
Parity of a binary integer, 13.
Patashnik, Oren, 37.
Pearson, Karl (= Karl), 25.
\( H (x) \), 14.
Pittel, Boris Gershon [Vacarevau, Bogdan Tepeşonau], 36.
Playing cards, 1, 8, 14, 19.
Poisson, Siméon Denis, distribution, 15, 31.
trials, 31.
Pólya, György (= George), 6, 19.
urn model, iv, 6–7, 18–20, 34.
Positively correlated random variables, 17, 24.
Power series, 28.
Prime implicants of a Boolean function, 5, 29.
Probability estimates, 3–5, 8–9, 16.
Probability generating functions, 15, 36.
Probability spaces, 1–2.
q.s.: Quite surely, 12, 20.
Quick, Jonathan Horatio, 18.
Quite sure events, 12, 23.
Raghavan, Prabhakar (רגרבנאו פרבאקה), 36.
Random bits, 2, 3, 5, 9, 13–15, 32.
Random graphs, 16, 18.
Random permutations, 15.
Random variables, 1–21.
Reliability polynomials, 5, 15, 16.
Rémond de Montmorin, Pierre, 32.
Rényi, Alfréd, 30.
Restricted growth strings, 15.
Roes, Sheldon Mark, iv, 5, 29.
Runs of a permutation, 33.
\( S_{n} \) (a symmetric threshold function), 16.
Samuels, Stephen Mitchell, 15, 27.
Savage, Richard Preston, Jr., 22.
Say Red, 34.
Schroeppe, Richard Crabtree, 24.
Schulte-Geers, Ernst Franz Fred, iv, 13, 20–24, 27.
Second moment principle, 4, 16, 29.
Set partitions, 26, 28.
Sets, represented as integers, 30.
Shuffles, 1.
Sideways sum \( \{x\} \), 13.
St. Petersburg paradox, 32.
Stadje, Gerrit Wolfgang, 35.
Sticker, David Robert, 28.
Stork, David Godfrey, 23.
Stross, Charles David George, iii.
Subadditive law, 11.
Submartingales, 8–9, 20.
Submodular set functions, 31.
Subsequence of a martingale, 18.
Summation by parts, 34.
Supermartingales, 8.
Supported sets, 17.
Symmetric Boolean functions, 16.
\( t \)-ary ballot numbers, 35.
Tail inequalities, 4, 8–11, 15, 20.
Taylor, Brook, formula, 20.
Thae, Axel, constant, 24.
Totally uncorrelated sequences, 34.
Transcendental numbers, 24.
Triangles (3-cliques), 16.
Trybula, Stanislav, 23.
Uncorrelated sequences, 34.
Union inequality, 14.
Upfal, Eli (יעפה אלי), 36.
Urn models, 6–7, 18–20, 34.
Usiskin, Zalman Philip, 23.
Vandenberghhe, Lieven Lodewijk André, 29.
Variance, 2, 4, 9, 14, 31.
Vicious, Kode (pen name of George Vernon Neville-Neil III), vi.
von Mengden, Nikolai Alexandrovich (פונצ מינגדן, ניקולאי אלכסנדרוביץ), 21.
Wald, Abraham (= Ahraham), 34.
equation, 19.
Wigderson, Avi (ויגדרסון אבי), 36.