Zhengyan Lin
Zhidong Bai

Probability Inequalities

Science Press Beijing
Springer

Zhengyan Lin
Zhidong Bai

Probability Inequalities

Zhengyan Lin
Zhidong Bai

Probability Inequalities

Science Press
Beijing

Springer

Authors
Zhengyan Lin
Department of Mathematics
Zhejiang University, China
zlin@zju.edu.cn

Zhidong Bai
School of Mathematics and Statistics
Northeast Normal University, China
baizd@nenu.edu.cn
Department of Statistics and Applied Probabilty
National University of Singapore, Singapore
stabaizd@leonis.nus.edu.sg

ISBN 978-7-03-025562-4
Science Press Beijing

ISBN 978-3-642-05260-6 e-ISBN 978-3-642-05261-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009938102

Cover design: Frido Steinen-Broo, EStudio Calamar, Spain

Printed on acid-free paper

Springer is a part of Springer Science+Business Media (www.springer.com)

Preface

In almost every branch of quantitative sciences, inequalities play an important role in its development and are regarded to be even more important than equalities. This is indeed the case in probability and statistics. For example, the Chebyshev, Schwarz and Jensen inequalities are frequently used in probability theory, the Cramer-Rao inequality plays a fundamental role in mathematical statistics. Choosing or establishing an appropriate inequality is usually a key breakthrough in the solution of a problem, e.g. the Berry-Esseen inequality opens a way to evaluate the convergence rate of the normal approximation.

Research beginners usually face two difficulties when they start researching—they choose an appropriate inequality and/or cite an exact reference. In literature, almost no authors give references for frequently used inequalities, such as the Jensen inequality, Schwarz inequality, Fatou Lemma, etc. Another annoyance for beginners is that an inequality may have many different names and reference sources. For example, the Schwarz inequality is also called the Cauchy, Cauchy-Schwarz or Minkovski-Bnyakovski inequality. Bennet, Hoeffding and Bernstein inequalities have a very close relationship and format, and in literature some authors cross-cite in their use of the inequalities. This may be due to one author using an inequality and subsequent authors just simply copying the inequality's format and its reference without checking the original reference. All this may distress beginners very much.

The aim of this book is to help beginners with these problems. We provide a place to find the most frequently used inequalities, their proofs (if not too lengthy) and some references. Of course, for some of the more popularly known inequalities, such as Jensen and Schwarz, there is no necessity to give a reference and we will not do so.

The wording "frequently used" is based on our own understanding. It can be expected that many important probability inequalities are not collected in this work. Any comments and suggestions will be appreciated.

The writing of the book is supported partly by the National Science Foundation of China.

The authors would like to express their thanks to Ron Lim Beng Seng for improving our English in this book.

Zhengyan Lin

May, 2009

Contents

Chapter 1

Elementary Inequalities of Probabilities of Events

In this Chapter, we shall introduce some basic inequalities which can be found in many basic textbooks on probability theory, such as Feller (1968), Loéve (1977), etc.

We shall use the following popularly used notations. Let Ω be a space of elementary events, $\mathscr{F}$ be a σ-algebra of subsets of Ω, P be a probability measure defined on the events in $\mathscr{F}$. $(\Omega, \mathscr{F}, P)$ is so called a probability space. The events in $\mathscr{F}$ will be denoted by $A_1, A_2, \cdots$ or $A, B, \cdots$ etc. $A \bigcup B, AB$ (or $A \bigcap B$), $A - B$ and $A \Delta B$ denote the union, intersection, difference and symmetric difference of A and B respectively. A^c denotes the complement of A and $\varnothing$ denotes the empty set.

1.1 Inclusion-exclusion Formula

Let $A_1, A_2, \cdots, A_n$ be n events, then we have

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) = & \sum_{i=1}^n P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) + \cdots \\ & + (-1)^{k-1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} P(A_{i_1} \cdots A_{i_k}) \\ & + \cdots + (-1)^{n-1} P(A_1 \cdots A_n). \end{aligned}$$

Proof. When $n = 2$, it is trivially known that

$$\begin{aligned} P\left(A_1 \bigcup A_2\right) &= P(A_1) + P(A_2 - A_1 A_2) \\ &= P(A_1) + P(A_2) - P(A_1 A_2). \end{aligned} \tag{1}$$

We show the formula by induction. Assume that the formula holds for n. We will show that it holds also for $n+1$. In fact, by (1) and the induction hypothesis,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} A_i A_{n+1}\right) \\ &= \sum_{i=1}^{n+1} P(A_i) - \sum_{1 \leqslant i<j \leqslant n} P(A_i A_j) + \cdots + (-1)^{n-1} P(A_1 \cdots A_n) \\ &\quad - \left\{ \sum_{i=1}^{n} P(A_i A_{n+1}) - \sum_{1 \leqslant i<j \leqslant n} P(A_i A_j A_{n+1}) \right. \\ &\quad \left. + \cdots + (-1)^{n-1} P(A_1 \cdots A_n A_{n+1}) \right\} \\ &= \sum_{i=1}^{n+1} P(A_i) - \sum_{1 \leqslant i<j \leqslant n+1} P(A_i A_j) + \cdots + (-1)^{n} P(A_1 \cdots A_{n+1}). \end{aligned}$$

1.2 Corollaries of the Inclusion-exclusion Formula

From the inclusion-exclusion formula, it is easy to deduce the following two conclusions.

1.2.a. When $A_1, \cdots, A_n$ are exchangeable, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i-1} \binom{n}{i} P(A_1, \cdots, A_i).$$

Remark. A set of events $\{A_1, \cdots, A_n\}$ is said to be exchangeable if the probability of the intersection of any subset depends only on the size of the subset, that is, for any integers $1 \leqslant i_1 < \cdots < i_j \leqslant n$ and $1 \leqslant j \leqslant n, P(A_{i_1} A_{i_2} \cdots A_{i_j}) = p_j$.

1.2.b. When $A_1, \cdots, A_n$ are independent and $p = P(A_i)$, we have

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i-1} \binom{n}{i} p^i.$$

1.3 Further Consequences of the Inclusion-exclusion Formula

The following inequalities are also consequences of the inclusion-exclusion formula.

1.3.a. $\sum_{i=1}^{n} P(A_i) - \sum_{1\leqslant i<j\leqslant n} P(A_iA_j) \leqslant P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i)$.

Remark. The right hand side (RHS) of the above inequality can be improved to

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(A_1A_i).$$

Proof. When $n=2$, the inequality with the improved RHS reduces to the inclusion-exclusion formula. Now, by induction we have

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\left(\bigcup_{i=1}^{n} A_i\right)\bigcap A_{n+1}\right) \\ &\leqslant \sum_{i=1}^{n} P(A_i) - \sum_{i=2}^{n} P(A_1A_i) + P(A_{n+1}) - P(A_1A_{n+1}). \end{aligned}$$

This proves the improved right hand side. Similarly, by induction, we have

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\left(\bigcup_{i=1}^{n} A_i\right)\bigcap A_{n+1}\right) \\ &\geqslant \sum_{i=1}^{n} P(A_i) - \sum_{i<j\leqslant n} P(A_iA_j) + P(A_{n+1}) - \sum_{i=1}^{n} P(A_iA_{n+1}). \end{aligned}$$

This proves the left hand side (LHS) of the inequality.

1.3.b. $|P(AB) - P(A)P(B)| \leqslant \frac{1}{4}$.

Proof.

$$\begin{aligned} |P(AB) - P(A)P(B)| &= |P(A)P(AB) + P(A^c)P(AB) \\ &\quad - P(A)P(AB) - P(A)P(A^cB)| \\ &= |P(A^c)P(AB) - P(A)P(A^cB)|. \end{aligned}$$

Since A^c and AB are disjoint, $P(A^c)P(AB) \leqslant 1/4$ (by noticing that $\max_{0<p<1} p(1-p) = 1/4$). Similarly, $P(A)P(A^cB) \leqslant 1/4$ as desired.

Remark. The difference $P(AB) - P(A)P(B)$ can be regarded as the covariance of the indicators I_A and I_B. The inequality 1.3.b can be easily proved by using the Cauchy-Schwarz inequality. Here, we proved the inequality by deliberately avoiding the use of moments.

1.3.c. $|P(A) - P(B)| \leqslant P(A\Delta B)$, $(A\Delta B = (A-B)\bigcup(B-A))$

Proof. By 1.3.a,

$$P(A\Delta B) \geqslant P(A-B) = P(A) - P(AB) \geqslant P(A) - P(B).$$

By the symmetry of A and B, 1.3.c is proved.

1.3.d. (Boole inequality). $P(AB) \geqslant 1 - P(A^c) - P(B^c)$.

Proof. $P(AB) + P(B^c) \geqslant P(A) = 1 - P(A^c)$.

1.3.e. Let $\limsup\limits_{n\to\infty} A_n = \bigcap\limits_{N=1}^{\infty}\bigcup\limits_{n=N}^{\infty} A_n, \liminf\limits_{n\to\infty} A_n = \bigcup\limits_{N=1}^{\infty}\bigcap\limits_{n=N}^{\infty} A_n$. Then

$$\begin{aligned} P(\liminf_{n\to\infty} A_n) \leqslant \liminf_{n\to\infty} P(A_n) &\leqslant \limsup_{n\to\infty} P(A_n) \\ &\leqslant P(\limsup_{n\to\infty} A_n) \leqslant \lim_{N\to\infty}\sum_{n=N}^{\infty} P(A_n). \end{aligned}$$

Proof. For any positive integer N, we have

$$\bigcap_{n=N}^{\infty} A_n \subset A_N \subset \bigcup_{n=N}^{\infty} A_n,$$

which simply implies

$$P\left(\bigcap_{n=N}^{\infty} A_n\right) \leqslant P(A_N) \leqslant P\left(\bigcup_{n=N}^{\infty} A_n\right) \leqslant \sum_{n=N}^{\infty} P(A_n).$$

Letting $N \to \infty$, we obtain the desired inequalities.

1.3.f. If $P(A) \geqslant 1-\varepsilon, P(B) \geqslant 1-\varepsilon$ for some $0 < \varepsilon < \frac{1}{2}$, then $P(AB) \geqslant 1 - 2\varepsilon$.

Proof. $P(AB) = P(A) + P(B) - P(A\bigcup B) \geqslant 1 - 2\varepsilon$.

1.3.g (Bonferroni inequality). Let $P_{[m]}(P_m)$ be the probability that exactly (at least, correspondingly) m events among $A_1, \cdots, A_n$ occur simultaneously. Putting

$$S_m = \sum_{1\leqslant i_1<\cdots<i_m\leqslant n} P(A_{i_1}\cdots A_{i_m}).$$

Then

$$S_m - (m+1)S_{m+1} \leqslant P_{[m]} \leqslant S_m, \qquad S_m - mS_{m+1} \leqslant P_m \leqslant S_m.$$

Proof. Let $A_{[m]}$ ($A_{(m)}$) denote the event that exactly (at least, correspondingly) m events among $A_1, \cdots, A_n$ happen simultaneously. Then we have

$$A_{[m]} \subset A_{(m)} = \bigcup_{1\leqslant i_1<\cdots<i_m\leqslant n} A_{i_1}\cdots A_{i_m}.$$

The RHS of the above inequalities follows by using the semi-additivity of probability measure.

On the other hand, we have

$$A_{[m]} \supset \bigcup_{1\leqslant i_1<\cdots<i_m\leqslant n} A_{i_1}\cdots A_{i_m} - \bigcup_{1\leqslant i_1<\cdots<i_{m+1}\leqslant n} A_{i_1}\cdots A_{i_{m+1}}.$$

This implies $p_{[m]} \geqslant p_m - S_{m+1}$. Also, for each $i_1 < \cdots < i_{m+1}$, the probability $P(A_{i_1}\cdots A_{i_{m+1}})$ is included at most in each of $P(A_{i'_1}\cdots A_{i'_m})$ in S_m, where $(i'_1,\cdots,i'_m)$ is a subset of $(i_1,\cdots,i_{m+1})$. Among the $m+1$, one needs to contribute to p_m. Therefore,

$$S_m - p_m \geqslant mS_{m+1}.$$

The LHS of 1.3.g then follow.

Remark. In fact, the inequalities 1.3.g can be proved from the following identities:

$$P_{[m]} = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} + \cdots + (-1)^{n-m}\binom{n}{m} S_n,$$

$$P_m = S_m - \binom{m}{m-1} S_{m+1} + \binom{m+1}{m-1} S_{m+2} + \cdots + (-1)^{n-m}\binom{n-1}{m-1} S_n,$$

$$S_m = \sum_{i=m}^{n}\binom{i}{m} P_{[i]} \quad \text{and} \quad S_m = \sum_{i=m}^{n}\binom{i-1}{m-1} P_i.$$

By definition, we have

$$P_{[i]} = \sum_{F\in\mathscr{F}_i} P\Big(\bigcap_{j\in F} A_j \bigcap_{\ell\in F^c} A_\ell^c\Big),$$

where $\mathscr{F}_i$ is the collection of all subsets of size i of the set $\{1,2,\cdots,n\}$. Note that for each $\tilde{F}\in\mathscr{F}_m$, the set $\bigcap_{t\in\tilde{F}} A_t$ can be written as the union of disjoint subsets $\bigcap_{j\in F} A_j \bigcap_{\ell\in F^c} A_\ell^c$, for all $i \geqslant m$ where $F\in\mathscr{F}_i$ and $F\subset\tilde{F}$. This proves that

$$S_m = \sum_{i=m}^{n}\binom{i}{m} P_{[i]},$$

which implies $S_m = \sum\limits_{i=m}^{n}\binom{i-1}{m-1} P_i$ by noticing that $P_{[i]} = P_i - P_{i+1}$.

Substituting the expression of S_m in terms of $P_{[i]}$ into the RHS of the first identity, we obtain

$$S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} + \cdots + (-1)^{n-m} \binom{n}{m} S_n$$

$$= \sum_{j=m}^{n} (-1)^{j-m} \binom{j}{m} \sum_{i=j}^{n} \binom{i}{j} P_{[i]}$$

$$= \sum_{i=m}^{n} \binom{i}{m} P_{[i]} \sum_{j=m}^{i} \binom{i-m}{j-m} (-1)^{j-m} = P_{[m]}.$$

By the same approach, one can prove the second identity.

1.4 Inequalities Related to Symmetric Difference

1.4.a. $P\Big\{\Big(\bigcup_n A_n\Big)\Delta\Big(\bigcup_n B_n\Big)\Big\} \leqslant P\Big\{\bigcup_n \Big(A_n \Delta B_n\Big)\Big\} \leqslant \sum_n P(A_n \Delta B_n).$

Proof. The left inequality follows from $\Big(\bigcup_n A_n\Big)\Delta\Big(\bigcup_n B_n\Big) \subset \bigcup_n (A_n \Delta B_n)$ by the definition of the symmetric difference. The right one follows from 1.3.a.

1.4.b. $P\{(A_1 - A_2)\Delta(B_1 - B_2)\} \leqslant P(A_1 \Delta B_1) + P(A_2 \Delta B_2).$

Proof. The inequality follows from the observation

$$(A_1 - A_2)\Delta(B_1 - B_2) \subset (A_1 \Delta B_1) \bigcup (A_2 \Delta B_2).$$

1.5 Inequalities Related to Independent Events

1.5.a. Let $\{A_n\}$ be a sequence of mutually independent events. Then

$$1 - P\left(\bigcup_{k=1}^{n} A_k\right) \leqslant \exp\left\{-\sum_{k=1}^{n} P(A_k)\right\},$$

$$1 - P\left(\bigcup_{k=1}^{\infty} A_k\right) \leqslant \lim_{n\to\infty} \exp\left\{-\sum_{k=1}^{n} P(A_k)\right\}.$$

Remark. The inequalities are useful in the proof of the Borel-Cantelli lemma.

Proof. The conclusions follow from an application of the inequality $1 - x \leqslant \mathrm{e}^{-x}$ for real x to the RHS of the identity

$$1 - P\left(\bigcup_{k=1}^{n} A_k\right) = P\left(\bigcap_{k=1}^{n} A_k^c\right) = \prod_{k=1}^{n} (1 - P(A_k)).$$

1.5.b. Suppose that A and B are independent, $AB \subset D$ and $A^cB^c \subset D^c$. Then $P(AD) \geqslant P(A)P(D)$.

Proof.

$$\begin{aligned} P(AD) &= P(ADB) + P(ADB^c) = P(AB) + P(AB^c) - P(AD^cB^c) \\ &= P(A)P(B) + P(AB^c) - P(D^cB^c) + P(A^cD^cB^c) \\ &= P(A)P(B) + P(AB^c) - P(D^cB^c) + P(A^cB^c) \\ &= P(A)P(B) + P(B^c) - P(D^cB^c) \\ &\geqslant P(A)P(BD) + P(A)P(B^cD) = P(A)P(D). \end{aligned}$$

1.5.c (Feller-Chung). Let $A_0 = \varnothing, \{A_n\}$ and $\{B_n\}$ be two sequences of events. Suppose that either

(i) B_n is independent of $A_n A_{n-1}^c \cdots A_0^c$ for all $n \geqslant 1$, or

(ii) B_n is independent of $\{A_n, A_n A_{n+1}^c, A_n A_{n+1}^c A_{n+2}^c, \cdots\}$ for all $n \geqslant 1$. Then

$$P\left(\bigcup_{n=1}^{\infty} A_n B_n\right) \geqslant \inf_{n\geqslant 1} P(B_n) P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Proof. In case (i),

$$\begin{aligned} P\left\{\bigcup_{n=1}^{\infty} A_n B_n\right\} &= P\left\{\bigcup_{n=1}^{\infty} B_n A_n \bigcap_{j=0}^{n-1} (B_j A_j)^c\right\} = \sum_{n=1}^{\infty} P\left\{B_n A_n \bigcap_{j=0}^{n-1} (B_j A_j)^c\right\} \\ &\geqslant \sum_{n=1}^{\infty} P\left\{B_n A_n \bigcap_{j=0}^{n-1} A_j^c\right\} = \sum_{n=1}^{\infty} P(B_n) P\left\{A_n \bigcap_{j=0}^{n-1} A_j^c\right\} \\ &\geqslant \inf_{n\geqslant 1} P(B_n) P\left(\bigcup_{n=1}^{\infty} A_n\right), \end{aligned}$$

and in case (ii),

$$\begin{aligned} P\left\{\bigcup_{j=1}^{n} A_j B_j\right\} &= \sum_{j=1}^{n} P\left\{A_j B_j \bigcap_{i=j+1}^{n} (A_i B_i)^c\right\} \geqslant \sum_{j=1}^{n} P\left\{A_j B_j \bigcap_{i=j+1}^{n} A_i^c\right\} \\ &= \sum_{j=1}^{n} P(B_j) P\left\{A_j \bigcap_{i=j+1}^{n} A_i^c\right\} \geqslant \inf_{1\leqslant j\leqslant n} P(B_j) P\left\{\bigcup_{j=1}^{n} A_j\right\}. \end{aligned}$$

1.6 Lower Bound for Union (Chung-Erdös)

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \left(\sum_{i=1}^{n} P(A_i)\right)^2 \Big/ \left(\sum_{i=1}^{n} P(A_i) + 2 \sum_{1\leqslant i<j\leqslant n} P(A_i A_j)\right).$$

Proof. Define random variables $X_k(\omega), \omega \in \Omega$, by

$$X_i(\omega) = \begin{cases} 0, & \text{if } \omega \notin A_i, \\ 1, & \text{if } \omega \in A_i. \end{cases}$$

Then

$$2 \sum_{1\leqslant i<j\leqslant n} P(A_i A_j) = E(X_1 + \cdots + X_n)^2 - E(X_1^2 + \cdots + X_n^2).$$

By the Cauchy-Schwarz inequality (see 8.4.b), we have

$$(E(X_1 + \cdots + X_n))^2 \leqslant P(X_1 + \cdots + X_n > 0) E(X_1 + \cdots + X_n)^2.$$

Note that $EX_i = EX_i^2 = P(A_i), P(X_1 + \cdots + X_n > 0) = P\left(\bigcup_{i=1}^{n} A_i\right)$ by definition. Combining the above two relations yields the desired inequality.

References

Chow Y S, Teicher H. 1997. Probability Theory: Independence, Interchangeability, Martingales. 2nd ed.. New York: Springer-Verlag.

Chung K L. 1974. A Course in Probability Theory. 2nd ed.. New York: Academic Press.

Feller W. 1968. An Introduction to Probability Theory and Its Applications. 3rd ed.. New York: Wiley.

Hardy G H, Littlewood J E, Polya G. 1952. Inequalities. Cambridge: Cambridge University Press.

Loéve M. 1977. Probability Theory. 4th ed.. New York: Springer-Verlag.

Petrov V V. 1995. Limit Theorems of Probability Theory: Sequences of Independent Random Variables. Oxford & New York: Clarendon Press & Oxford University Press.

Shorack G R, Wellner J A. 1986. Empirical Processes with Applications to Statistics. New York: Wiley.

Chapter 2

Inequalities Related to Commonly Used Distributions

Commonly used distributions play an important role in applied statistics, statistical computing and applied probability. So, inequalities related to these distributions are of great interest in these areas.

Let ξ be a random variable (r.v.). Then its distribution function (d.f.) is defined by $F(x) = P(\xi < x)$ and its probability density function (pdf.) $p(x)$ (if it exists) is defined to be a measurable function such that $F(x) = \int_{-\infty}^{x} p(y)\mathrm{d}y$. Write

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\mathrm{d}t \qquad \text{and} \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$$

for the standard normal d.f. and pdf. respectively,

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \cdots, n, \quad 0 < p < 1, \quad q = 1 - p$$

for the binomial distribution with parameters n and p,

$$p(k; \lambda) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}, \qquad k = 0, 1, \cdots, \quad \lambda > 0$$

for the Poisson distribution with parameter λ.

2.1 Inequalities Related to the Normal d.f.

2.1.a. $\frac{1}{\sqrt{2\pi}}(b-a)\exp\{-(a^2 \vee b^2)/2\} \leqslant \Phi(b) - \Phi(a) \leqslant \frac{1}{\sqrt{2\pi}}(b-a), -\infty < a < b < \infty$.

Proof. It follows from the fact that $\mathrm{e}^{-x^2/2}$ on $[a, b]$ is between $\exp\{-(a^2 \vee b^2)/2\}$ and 1.