AN INEQUALITY IN PROBABILITY THEORY

G. A. HUNT

1. Let \(x\) and \(y\) be random variables with finite expectations. We shall say that \(x\) dominates \(y\) if \(E\{\phi(x)\} \geq E\{\phi(y)\}\) whenever \(\phi\) is a continuous convex function on the real line \(\mathbb{R}^1\). (The expectations \(E\{\phi(x)\}\) and \(E\{\phi(y)\}\) are always well defined if \(+\infty\) is admitted as a value.)

Assume now that \(X_1\) and \(X_2\) are independent and dominate respectively the independent random variables \(Y_1\) and \(Y_2\). Let \(\phi\) be a continuous convex function on \(\mathbb{R}^2\) and denote by \(F_i\) and \(G_i\) the distribution functions of \(X_i\) and \(Y_i\). We have

\[
\int dF_1(u_1) \int \phi(u_1, u_2) dF_2(u_2) \geq \int dF_1(u_1) \int \phi(u_1, u_2) dG_2(u_2) \\
= \int dG_2(u_2) \int \phi(u_1, u_2) dF_1(u_1) \\
\geq \int dG_2(u_2) \int \phi(u_1, u_2) dG_1(u_1),
\]

so that \(E\{\phi(x_1, x_2)\} \geq E\{\phi(y_1, y_2)\}\). In the same manner one can verify more generally that

\[
E\{\phi(x_1, \ldots, x_n)\} \geq E\{\phi(y_1, \ldots, y_n)\}
\]

provided that

(i) \(\phi\) is a continuous convex function on \(\mathbb{R}^n\),
(ii) \(x_1, \ldots, x_n\) are independent and \(y_1, \ldots, y_n\) are independent,
(iii) all expectations \(E\{x_i\}\) and \(E\{y_i\}\) exist, and
(iv) \(X_i\) dominates \(Y_i\) for \(1 \leq i \leq n\).

2. We shall give two illustrations how (1) may be used. First, however, we prove a lemma.

**Lemma.** Let \(E\{y\} = 0\) and let \(|y| \leq 1\) with probability one. Then \(y\) is dominated by every symmetric random variable \(x\) such that \(|x| \leq a\). Let \(a = E\{|x|\}\) and let \(z\) be a random variable with the distribution \(\Pr\{z = -a\} = \Pr\{z = +a\} = 1/2\). It will obviously suffice to prove that \(x\) dominates \(z\) and that \(z\) dominates \(y\).

Received by the editors September 10, 1951 and, in revised form, August 30, 1954.
Observe that the relations to be proved concern only distribution functions. Thus we may take the probability space to be the unit square $0 < \omega, \eta < 1$ with ordinary Lebesgue measure, and $x$ a function only of $\omega$ and increasing in $\omega$. Let us define $\tilde{x}(\omega, \eta)$ by

$$
\tilde{x}(\omega, \eta) = \begin{cases} x(\eta/2), & 0 < \omega < 1/2, \\ x(1/2 + \eta/2), & 1/2 \leq \omega < 1. 
\end{cases}
$$

It is clear that $\tilde{x}$ has the same distribution function as $x$, and a simple computation shows that $\int_0^1 \tilde{x}(\omega, \eta) d\eta$ is $-a$ or $+a$ according as $\omega < 1/2$ or $\omega \geq 1/2$. Consequently if $\phi$ is convex we have, making use of Jensen's inequality,

$$
\mathcal{E}\{\phi(x)\} = \int_0^1 d\omega \int_0^1 \phi(\tilde{x}(\omega, \eta)) d\eta 
\leq \int_0^1 \phi \left( \int_0^1 \tilde{x}(\omega, \eta) d\eta \right) d\omega 
= (1/2)\phi(-a) + (1/2)\phi(+a) 
= \mathcal{E}\{\phi(z)\}.
$$

The proof that $z$ dominates $y$ proceeds similarly. First define

$$
\tilde{z}(\omega, \eta) = \begin{cases} -a, & 0 < \eta < (1/2)(1 - y(\omega)/a), \\ +a, & (1/2)(1 - y(\omega)/a) \leq \eta < 1. 
\end{cases}
$$

Then verify that $\tilde{z}$ has the same distribution function as $z$ and that $\int_0^1 \tilde{z}(\omega, \eta) d\eta = y(\omega)$ for all $\omega$. A computation like that in (2) completes the proof.

It is usually most convenient to take for $x$ either a Gaussian variable with mean 0 and variance $\pi/2$ or a variable whose distribution is $\Pr \{x = -1\} = \Pr \{x = +1\} = 1/2$.

3. As a first application we obtain a generalization of Khinchin's inequalities.

**Theorem.** Let $(y_n)$ be a sequence of independent random variables satisfying $\mathcal{E}\{y_n\} = 0$ and $|y_n| \leq 1$, and let $(a_n)$ be a sequence of constants. Then

$$
\mathcal{E}\left\{ \sup_n \left| \sum_{1 \leq k \leq n} a_k y_k \right|^{2q} \right\} \leq 2( \sum_{1}^{n} a_k^2 ) 1 \cdot 3 \cdot 5 \cdots (2q - 1)
$$

for every positive integer $q$.  

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
It is enough to prove (3) for \( n \) ranging over the finite set \( \{1, \cdots, N\} \) as the full assertion is then obtained by letting \( N \) become large. The inequality has already been proved in \([2]\) and \([3]\) when the \( y_n \) are the Rademacher functions. On noting that \( \sup_{n \leq N} \left| \sum_{1 \leq k \leq n} a_k u_k \right|^{2q} \) is a convex function of the point \((u_1, \cdots, u_N)\) of \( R^N \), we see that (3) follows from (1) and the lemma of §2.

Inequality (3) would still hold if we were to assume that each of the \( y_n \), instead of being bounded by 1, is dominated by a Gaussian variable with mean zero and variance 1.

4. Our second application is to a subject treated by Goldstine and von Neumann in \([1]\).

Numerical calculation of a matrix yields a result which differs from the true matrix by an error matrix whose elements are composed of rounding-off errors. Since the magnitude of the error is often measured by the bound of the error matrix, it becomes of some moment to estimate the probability of large values of the bound. The estimates are easy to make if the elements of the error matrix are all Gaussian; I shall perform the computation in such a way that the results are valid for certain other types of random matrices.

Let \( X \) be an \( n \times n \) matrix whose elements \( x_{ij} \) are independent Gaussian variables with means 0 and variances 1. Our first step is to find an upper bound for the probability that \( \|X\| \) be greater than \( (2\pi n)^{1/2} \), where \( \pi \) is some number greater than 1.

Let \( \lambda_1 > \lambda_2 > \cdots > \lambda_n \) be the squares of the absolute values of the characteristic values of \( X \). Fisher and Hsu (see Wilks' Mathematical statistics, pages 261–265) have shown that the joint density function \( k_n(\lambda_1, \cdots, \lambda_n) \) of the \( \lambda_i \) vanishes unless \( \lambda_1 > \lambda_2 > \cdots > \lambda_n \) and then has the expression

\[
\pi^{n/2} 2^{-n/2} \prod_i \left[ \Gamma(i/2) \right]^{-2} \exp \left( -\frac{1}{2} \sum_i \lambda_i \right) \prod_{i < j} (\lambda_i - \lambda_j) \prod_i \lambda_i^{-1/2}.
\]

We need only the relation

\[
k_n(\lambda_1, \cdots, \lambda_n) = (2\pi)^{1/2} 2^{-n} \left[ \Gamma(n/2) \right]^{-1/2} \lambda_1^{-1/2} e^{-\lambda_1/2} \prod_{j>1} (\lambda_1 - \lambda_j) k_{n-1}(\lambda_2, \cdots, \lambda_n).
\]

The function

\[
\psi(u) = u^{-r} e^{u/2}
\]

(defined for positive \( u \)) is convex and attains its minimum for \( u = 2\pi n \). We have
\begin{align*}
&\int_{2^n}^{\infty} d\lambda_1 \int_{0}^{\lambda_1} d\lambda_2 \cdots \int_{0}^{\lambda_{n-1}} d\lambda_n \psi(\lambda_1) k_n(\lambda_1, \ldots, \lambda_n) \\
&= \frac{(2\pi)^{1/2}}{2^n [\Gamma(n/2)]^2} \int_{2^n}^{\infty} d\lambda_1 \int_{0}^{\lambda_1} d\lambda_2 \cdots \int_{0}^{\lambda_{n-1}} d\lambda_n \lambda_1^{n-rn-1/2} \\
&\cdot \prod_{j>1} (\lambda_1 - \lambda_j) k_{n-1}(\lambda_2, \ldots, \lambda_n) \\
&\leq \frac{(2\pi)^{1/2}}{2^n [\Gamma(n/2)]^2} \int_{2^n}^{\infty} \lambda_1^{n-rn-3/2} d\lambda_1 \int_{0}^{\infty} d\lambda_2 \cdots \int_{0}^{\infty} d\lambda_n k_{n-1}(\lambda_2, \ldots, \lambda_n) \\
&= \frac{(2\pi)^{1/2}(2rn)^{n-rn-1/2}}{2^n [\Gamma(n/2)]^2(rn - n + 1/2)}.
\end{align*}

The inequality comes from replacing each \( \lambda_j \) by zero in the product \( \prod (\lambda_1 - \lambda_j) \) and then extending the range of integration on \( \lambda_2 \) to infinity. The succeeding equality follows from the fact that \( k_{n-1} \) is a probability density function.

Observing that \( \lambda_1 \) is nothing but \( \|X\|^2 \), we conclude from (4) that

\begin{align*}
\Pr \left\{ \|X\|^2 > 2rn \right\} \\
&\leq \frac{1}{\psi(2rn)} \int_{2rn}^{\infty} d\lambda_1 \int_{0}^{\lambda_1} d\lambda_2 \cdots \int_{0}^{\lambda_{n-1}} \psi(\lambda_1) k_n(\lambda_1, \ldots, \lambda_n) d\lambda_n \\
&\leq \frac{(2\pi)^{1/2}(2rn)^{n-1/2}e^{-rn}}{2^n [\Gamma(n/2)]^2(rn - n + 1/2)} \\
&\leq \frac{1}{4(r - 1)(rn)^{1/2}} \left( \frac{2r}{e^r - 1} \right)^n.
\end{align*}

Stirling's formula has been used in the last step.

The inequality (1) does not imply that (4) and its consequence (5) remain true when \( X \) is replaced by a matrix whose elements are independent and dominated by Gaussian variables, for \( \psi(\|X\|^2) \) is not a convex function of the matrix \( X \). Relation (1) may be used, however, if one argues this way.

Let \( a = 2rn + 1 + (4rn + 1)^{1/2} \). The tangent to the graph of \( \psi \) at the point \( (a, \psi(a)) \) meets the \( u \) axis at \( u = 2rn \). So the function \( \phi(u) \) which is zero for \( u \leq 2rn \), coincides with \( \psi(u) \) for \( u \geq a \), and is linear and continuous in the interval \( 2rn \leq u \leq a \), is an increasing convex function. Since \( u^2 \) is an increasing convex function on \( 0 \leq u < \infty \) and \( \|X\|^2 \) is a convex function on \( R^{n^2} \), the composite function \( \phi(\|X\|^2) \) is convex on \( R^{n^2} \) and always less than \( \psi(\|X\|^2) \). Thus if \( Y = (y_{ij}) \) with the \( y_{ij} \) independent and dominated by the Gaussian variables \( x_{ij} \), we have
\[ \Pr \{ \|Y\|^2 \geq a \} \leq \frac{1}{\phi(a)} \mathcal{E} \{ \phi(\|Y\|^2) \} \]

\[ \leq \frac{1}{\psi(2\pi r)} \mathcal{E} \{ \phi(\|Y\|^2) \} \]

\[ \leq \frac{1}{\psi(2\pi r)} \int_0^\infty d\lambda_1 \int_0^{\lambda_1} d\lambda_2 \cdots \int_0^{\lambda_{n-1}} d\lambda_n \psi(\lambda_1) k_n(\lambda_1, \ldots, \lambda_n) \]

\[ \leq \frac{1}{4(r - 1)(r\pi)^{1/2}} \left( \frac{2r}{e^{r-1}} \right)^n. \]

**References**


**Cornell University**