GENERALIZATIONS OF CAUCHY-SCHWARZ IN PROBABILITY THEORY

R. HASEGAWA AND B. KARAMAKAR

ABSTRACT. We explore two generalizations of the Cauchy-Schwarz - Bessel’s inequality and the Selberg inequality - and their application to probability theory. We then give a tautological proof of the De Caen-Selberg Inequality and a proof of the second Borel-Cantelli Lemma with negative dependence. We finish with a suggestion of how linear operator theory can help us understand the tightness of many Selberg-type bounds.

1. IMPROVING ON THE CAUCHY-SCHWARZ INEQUALITY

While studying any well structured space knowing how to control the size of ‘correlation’ between two elements is an essential tool. If we are confined to a vector space $X$ with an inner product, where we measure this correlation by the inner product $(\cdot, \cdot)$, the first such control is given by the triangle inequality. A better and hugely more successful tool is the Cauchy-(Buniakowsky-)Schwarz inequality. The trick to using the Cauchy-Schwarz inequality effectively is to choose the right pair of vectors on which to apply it. A clue as to which pair one should choose is to keep in mind that the equality holds when one vector is a constant multiple of the other. To emphasize this point a bit more, consider the case in which $x \in \mathbb{R}^n$ and $\{y_1, y_2, \ldots, y_k\} \subset \mathbb{R}^n$ is a collection of orthonormal vectors. Then, if we are interested in controlling the size of $\sum_{j=1}^{k} |(x, y_j)|^2$ a direct use of the Cauchy-Schwarz inequality would give us a bound of $k||x||^2$. But, if we write $x = \sum_{j=1}^{k} (x, y_j) y_j + z$, where $z$ is orthogonal to each of the $y_j$’s then we get the inequality,

\[
\sum_{j=1}^{k} |(x, y_j)|^2 \leq ||x||^2 .
\]

We immediately see a gain of a factor of $k$. This gain comes from the fact that a fixed vector $x$ cannot be simultaneously close in direction to multiple $y_j$’s. But, $x$ can be well approximated by a linear combination of the $y_j$’s. Considering this, we use the Cauchy-Schwarz inequality on the $\langle x, \sum_{j=1}^{k} (x, y_j) y_j \rangle$ to get the better bound in (1) rather than using Cauchy-Schwarz on each of the terms individually. The inequality in Equation 1 is called Bessel’s inequality and is the primary tool in Hilbert space theory to establish the existence of a basis.

R. Hasegawa, B. Karmakar: The Wharton School, Department of Statistics, Huntsman Hall 4th Floor, University of Pennsylvania, Philadelphia, PA 19104. Email address: raiden@wharton.upenn.edu, bikramk@wharton.upenn.edu.
If we relax the orthogonality of the $y_j$’s, a generalization due to Bombieri (Montgomery, 1968) is the following,

$$\sum_{j=1}^{k} |\langle x, y_j \rangle|^2 \leq \|x\|^2 \max_{1 \leq i \leq k} \left\{ \sum_{j=1}^{k} |\langle x, y_j \rangle| \right\}.$$ 

In this note we shall focus on a different generalization of Bessel’s inequality, known as the Selberg inequality and its applications in probability. A graphical comparison of these three inequalities - Cauchy-Schwarz, Bessel’s, and Selberg - is shown in Figure 1.

2. Selberg Inequality

**Theorem 1** (Selberg Inequality in Hilbert Space). Suppose $y_1, \ldots, y_n$ are an arbitrary collection of non-zero elements in Hilbert space $\mathcal{H}$. Then

$$\sum_{j=1}^{n} \frac{|\langle x, y_j \rangle|^2}{\sum_{k=1}^{n} |\langle y_j, y_k \rangle|} \leq \|x\|^2 \text{ for all } x \in \mathcal{H},$$

with equality if and only if $x = \sum_{j=1}^{n} \alpha_j y_j$ and for each pair $(j, k)$ either $\langle y_j, y_k' \rangle = 0$ or $|\alpha_j| = |\alpha_k|$ with $\langle \alpha_j y_j, \alpha_k y_k \rangle \geq 0$.

In probability theory and the study of strong laws, we are often interested in determining the lower bound on the probability of the union of some arbitrary, dependent collection of events. Noticing that indicator functions of measurable sets belong to $L^2(dP)$ where $(\Omega, \mathcal{F}, P)$ is a probability triple, we leverage the Selberg inequality to give us a very direct path to the lower bound of De Caen (1997):

**Lemma 2** (De Caen-Selberg Inequality). Let $A_1, \ldots, A_n$ be a collection of arbitrary positive events in $(\Omega, \mathcal{F}, P)$. Then,

$$\sum_{j=1}^{n} \frac{P(A_j)^2}{\sum_{k=1}^{n} P(A_j \cap A_k)} \leq P \left( \bigcup_{j=1}^{n} A_j \right).$$

**Proof.** The proof is straightforward if not obvious. Observe that $\mathbb{1}_{A_j} \in L^2(dP)$ for all $j \in \{1, \ldots, n\}$ with $\langle X, Y \rangle \overset{def}{=} E(XY)$ when $\mathcal{H} = L^2(dP)$. Finally, note that $P(\bigcup_{i=1}^{n} A_i \cap A_k) = P(A_k)$ for $k \in \{1, \ldots, n\}$. The result now follows immediately from Theorem 1 with $x = \mathbb{1}_{\bigcup_{i=1}^{n} A_i}$ and $y_j = \mathbb{1}_{A_j}$. \qed
This is a nice, if simple, result that looks similar to the familiar Erdős-Chung inequality (Chung and Erdos, 1952),

\[ \frac{\sum_{i,j} P(A_i) P(A_j)}{\sum_{i,j} P(A_i \cap A_j)} \leq P \left( \bigcup_{j=1}^{n} A_j \right) . \]

What we find to be the most compelling conceptual distinction is that, in some approximate sense, Lemma 2 disentangles the contribution each set \( A_j \) makes to the union of all the events. Interestingly, we actually get a bit more juice out of Lemma 2 than we can from Erdős-Chung. An easy variation of Cauchy-Schwarz is all that we need to see the improvement.

**Lemma 3.** Consider \( \{a_1, \ldots, a_n\} \subset \mathbb{R}^+ \) and \( \{b_1, \ldots, b_n\} \subset \mathbb{R}^+ \). Then we have that

\[ \left( \sum_{i=1}^{n} b_i \right)^2 \leq \sum_{i=1}^{n} \frac{b_i^2}{a_i} . \]

**Proof.** Let \( a'_i = \sqrt{a_i} \) and \( b'_i = b_i/\sqrt{a_i} \) for \( i = 1, 2, \ldots, n \). Then applying Cauchy-Schwarz to \( \sum a'_i b'_i \) we get

\[ \left( \sum_{i=1}^{n} b_i \right)^2 = \left( \sum_{i=1}^{n} a'_i b'_i \right)^2 \leq \left( \sum_{i=1}^{n} b_i^2 \right) \left( \sum_{i=1}^{n} a_i^2 \right) = \left( \sum_{i=1}^{n} \frac{b_i^2}{a_i} \right) \left( \sum_{i=1}^{n} a_i \right) . \]

\[ \square \]

If we let \( b_i = P(A_i) \) and \( a_i = \sum_{k=1}^{n} P(A_i \cap A_k) \) then we see that

\[ \frac{\sum_{i,j} P(A_i) P(A_j)}{\sum_{i,j} P(A_i \cap A_j)} \leq \sum_{j=1}^{n} \sum_{k=1}^{n} P(A_j \cap A_k) . \]

Thus, any result that is proved using Erdős-Chung inequality can be improved using this better bound. In Section 4 we give one such case.

### 3. De Caen-Selberg Inequality from Tautology

Up until this point, we have taken the Selberg inequality in general Hilbert space as given and from that derived a refinement of the Erdős-Chung lower bound (lemma 2). We might ask ourselves if there is a simpler and more self-contained path to the De Caen-Selberg inequality. As in Steele (2014), we begin with a tautology

\[ \sum_{i=1}^{n} I_{A_i} \alpha_i = I_B \sum_{i=1}^{n} I_{A_i} \alpha_i . \]

Here we let \( B = \bigcup_{i=1}^{n} A_i \) and \( \alpha_1, \ldots, \alpha_n \) be any sequence of non-negative reals. Later, we will pick these \( \alpha_i \) and see that a very naive choice will do. Applying Cauchy-Schwarz we find that

\[ E \left[ \sum_{i=1}^{n} I_{A_i} \alpha_i \right] = E \left[ I_B \right]^{1/2} E \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} I_{A_i} I_{A_j} \alpha_i \alpha_j \right]^{1/2} . \]

We can pull the \( \alpha_i \)'s out of the inner sum on the RHS by observing that \( \alpha_i \alpha_j \leq \frac{\alpha_i^2}{2} + \frac{\alpha_j^2}{2} \) and that there is some nice symmetry in the summand (i.e. \( P(A_i \cap A_j) = \)
$P(A_j \cap A_i)$ for all $i, j$). Plugging in this estimate and writing our expectations of indicator functions as probabilities we find that

$$\sum_{i=1}^{n} P(A_i) \alpha_i \leq P(B)^{1/2} \left[ \sum_{i,j=1}^{n} \left\{ P(A_i \cap A_j) \left( \frac{\alpha_i^2}{2} + \frac{\alpha_j^2}{2} \right) \right\} \right]^{1/2}$$

(8)

Using the tried and true principle of giving ourselves what we want, we let $\alpha_i = P(A_i)/\sum_{j=1}^{n} P(A_i \cap A_j)$. Substituting our choice of $\alpha_i$ into (8) yields the De Caen-Selberg inequality:

$$\sum_{i=1}^{n} \frac{P(A_i)^2}{\sum_{j=1}^{n} P(A_i \cap A_j)} \leq P(B)^{1/2} \left[ \sum_{i=1}^{n} \frac{P(A_i)^2}{\sum_{j=1}^{n} P(A_i \cap A_j)} \right]^{1/2}$$

(9)

$$\Rightarrow \sum_{i=1}^{n} \frac{P(A_i)^2}{\sum_{j=1}^{n} P(A_i \cap A_j)} \leq P(B).$$

(10)

This path to the De Caen-Selberg lower bound is nice for a couple of reasons. First, it relies only on a tautology, two applications of Cauchy-Schwarz, and a naive choice for the $\alpha_i$’s. Any talk about Hilbert spaces is avoided. Second, since the $\alpha_i$’s are arbitrary, it suggests many other possible lower bounds. For example, letting $\alpha_i = 1$ for all $i$ gives us Erdős-Chung as seen in Steele (2014). In fact, we could even optimize over $\alpha_1, \ldots, \alpha_n$ which would be equivalent to finding the orthogonal projection of $1_B$ onto $1_{A_1}, \ldots, 1_{A_n}$ (Feng et al., 2010). This would be the tightest bound we could achieve but the choice of $\alpha_i$’s in De Caen-Selberg leads to some convenient cancellations and a very interpretable result. The lower bound is simply a weighted sum of the $P(A_i)$’s.

4. A Proof of BC(II) Under Negative Dependence

In this section we prove a version of BC(II) under negative dependence. Originally proposed by Petrov (2002) and proved using (4), we provide a more concise and straightforward proof using the De Caen-Selberg inequality.

**Theorem 4** (Second Borel-Cantelli with weak repulsion). Let $A_1, A_2, \ldots$ be a sequence events satisfying (a) $\sum_{i=1}^{\infty} P(A_i) = \infty$ and (b) $P(A_k \cap A_l) \leq C P(A_k) P(A_l)$ for all $k, l > L$ such that $k \neq l$ for some constants $C \geq 1$ and $L$. Then

$$P(\limsup A_n) \geq 1/C.$$  

(11)

**Proof.** Consider $n > m > L$. The weak repulsion hypothesis gives us

$$\sum_{k=m}^{n} P(A_j \cap A_k) \leq P(A_j) + C \cdot \sum_{k=m, k \neq j}^{n} P(A_j) P(A_k),$$

where $C$ is a constant such that $P(A_k \cap A_l) \leq C P(A_k) P(A_l)$ for all $k, l > L$ such that $k \neq l$. This allows us to bound the probability of the intersection of events $A_k$ for $k > m$ using the given conditions.
for all $j$ such that $m \leq j \leq n$. Combining this with Lemma 2 we find that

\begin{equation}
P \left( \bigcup_{j=m}^{n} A_j \right) \geq \sum_{j=m}^{n} P(A_j)^2 + C \cdot \sum_{k=m}^{n} P(A_j)P(A_k)\end{equation}

\begin{equation}
= \frac{\sum_{j=m}^{n} P(A_j)}{1 + C \cdot \sum_{k=m}^{n} P(A_k)}.
\end{equation}

Taking limits as $n \to \infty$, applying the Monotone Convergence Theorem to the LHS, and appealing to hypothesis (a) we have

\begin{equation}
P \left( \bigcup_{j=m}^{\infty} A_j \right) \geq \frac{1}{C}.
\end{equation}

Finally, since $\limsup A_n := \cap_{m=1}^{\infty} \cup_{k=m}^{\infty} A_k$ we can take limits as $m \to \infty$ inside the probability yielding

\begin{equation}
P(\limsup A_n) \geq \frac{1}{C}.
\end{equation}

There are many other generalizations of BC(II) that relax the independence condition and impose some dependence restriction on the collection of sets. Many of the existing proofs of these results employ the weaker Erdős-Chung.

5. A Few Final Words

We return to general Hilbert space for one final view of the Selberg inequality. If for a finite set of vectors $y = \{y_1, y_2, \ldots, y_n\}$, we write a linear functional $S_y$ on the Hilbert space $H$ that maps any vector $x$ to the quantity on the LHS of the Selberg inequality, then Selberg states that $S_y$ is a bounded (continuous) operator which has operator norm 1 (or $\|S_y x\| \leq \|x\|$). As we’ve seen with the Selberg inequality, identifying such square root type inequalities with bounded linear operators can be very useful. In general, if one can find a linear operator $T$ such $\|Tx\|$ is the quantity we are interested in controlling, then an immediate inequality is that $\|Tx\| \leq \|T\|_{op}\|x\| = \|TT^*\|_{op}^{1/2}\|x\|$. This way of exploring an inequality is called the $TT^*$ method. Here, $\|T\|_{op} := \sup_{\|y\|=1,\|z\|=1} \langle Tx, y \rangle$ and $T^*$ is the adjoint of $T$. We direct you to Fujii (1991) for more on this line of thinking.

References


