Variance bounds and commutators

Koenraad M.R. Audenaert

July 10, 2009

A question

- Let A be a random variable that can assume the real values $a_1 \geq a_2 \geq \ldots \geq a_n$ and does so with probabilities p_i.
- The variance $\mathrm{Var}(A)$ of A is defined as

$$\mathrm{Var}(A) = \sum_i p_i(a_i - \sum_j p_j a_j)^2 \quad (1)$$

$$= \sum_i p_i a_i^2 - (\sum_i p_i a_i)^2. \quad (2)$$

- Question: can you find a good upper bound on $\mathrm{Var}(A)$ in terms of a_i?

Variance bounds

- The following is allegedly a 'well-known' variance bound:

$$\mathrm{Var}(A) \leq \sum_i a_i^2/2 = ||\underline{a}||^2/2.$$

- The best upper bound that uses the a_i is due to Murthy and Sethi:

$$\mathrm{Var}(A) \leq \left(\frac{a_1 - a_n}{2}\right)^2.$$

- This implies the previous bound since

$$(a_1 - a_n)^2 \leq (a_1 - a_n)^2 + (a_1 + a_n)^2 = 2(a_1^2 + a_n^2) \leq 2\sum_i a_i^2.$$

- Refs: M.N. Murthy and V.K. Sethi, Sankhya Ser B **27**, 201–210 (1965). For a very short proof see J. Muilwijk, Sankhya Ser B **28**, 183 (1966).

Why bother?

A theorem on commutators

- Here's something that I presented as an open problem in Banff in 2006:

  Theorem 1 (Böttcher and Wenzel) *For general complex $n \times n$ matrices X and Y, and for the Frobenius norm $||.||_2$,*

$$||[X, Y]||_2 \leq \sqrt{2}||X||_2||Y||_2. \qquad (3)$$

- Equality is obtained by letting X and Y be two anti-commuting Pauli matrices: $||[X, Y]||_2 = 2\sqrt{2}$ and $||X||_2 = ||Y||_2 = \sqrt{2}$.
- A less sharp bound is easy to prove, using the triangle inequality

$$||[X, Y]||_2 = ||XY - YX||_2 \leq ||XY||_2 + ||YX||_2 \leq 2||X||_2||Y||_2.$$

Proofs

- For normal matrices: rather easy.
- For non-normal real 2×2 matrices: also rather easy.
- First proof, for real case, by Seak-Weng Vong and Xiao-Qing Jin and independently also by Zhiqin Lu. Proof contains two cases:
  - First case: $n = 3$. Pages and pages of dense computations.
  - Second case: $n > 3$. Based on goodwill of reader: 'Proven similarly'
- Proof by Böttcher and Wenzel also includes complex case.
  Includes general n explicitly.
  Still rather involved, and nothing is learned apart from the validity of the theorem.

Royal Holloway University of London

What's that got to do with variance bounds?

A connection

- The commutator $[X, Y]$ and the variance of a random variable have a property in common

A connection

- The commutator $[X, Y]$ and the variance of a random variable have a property in common
- Translation invariance: with $a \in \mathbb{R}$,

$$\begin{aligned} [X - a\mathbb{1}, Y] &= [X, Y] \\ \mathrm{Var}(X - a) &= \mathrm{Var}(X) \end{aligned}$$

Royal Holloway
University of London

Why care about best constants?

Applications

- Mathematicians always care. Never know what the bound will be used for.
- Techniques used could be useful for other problems, where sharp constants do matter.
- In the inequality

$$||[X, Y]||_2 \leq \sqrt{2}||X||_2||Y||_2$$

  the matrix Y can be eliminated, giving:
- For any X with $||X||_2 = 1$,

$$||X \otimes \mathbb{1} - \mathbb{1} \otimes X^T|| \leq \sqrt{2}.$$

  with the operator norm in the LHS.

A related open problem

- A related problem concerns the existence of NPT bound entangled states (as communicated to me by Marco Piano).
- Let X and Y be general $n \times n$ matrices such that $\operatorname{Tr} X = \operatorname{Tr} Y = 0$ and

$$||X||_2^2 + ||Y||_2^2 = 1.$$

- Let $Z = X \otimes \mathbb{1} + \mathbb{1} \otimes Y$. Prove that $\sigma_1^2(Z) + \sigma_2^2(Z) \leq 2$.
- Note that equality is achieved, hence 2 is the sharpest value.
- Any higher bound would NOT work for the NPT problem!

Does anyone have objections to finding sharp constants?
If so, speak up now, or forever hold your peace!

My proof

- We have the following

$$\begin{aligned}||XY - YX||_2^2 &= \mathrm{Tr}[XYY^*X^* - XYX^*Y^* - YXY^*X^* + YXX^*Y^*] \\ &= \mathrm{Tr}[X^*XYY^* - XYX^*Y^* - YXY^*X^* + XX^*Y^*Y]\end{aligned}$$

My proof

- We have the following

$$
\begin{aligned}
||XY - YX||_2^2 &= \mathrm{Tr}[XYY^*X^* - XYX^*Y^* - YXY^*X^* + YXX^*Y^*] \\
&= \mathrm{Tr}[X^*XYY^* - XYX^*Y^* - YXY^*X^* + XX^*Y^*Y] \\
||YX^* + X^*Y||_2^2 &= \mathrm{Tr}[YX^*Y^*X + YX^*XY^* + X^*YXY^* + X^*YY^*X] \\
&= \mathrm{Tr}[X^*XY^*Y + XYX^*Y^* + YXY^*X^* + XX^*YY^*].
\end{aligned}
$$

My proof

- We have the following

$$
\begin{aligned}
||XY - YX||_2^2 &= \mathrm{Tr}[XYY^*X^* - XYX^*Y^* - YXY^*X^* + YXX^*Y^*] \\
&= \mathrm{Tr}[X^*XYY^* - XYX^*Y^* - YXY^*X^* + XX^*Y^*Y] \\
||YX^* + X^*Y||_2^2 &= \mathrm{Tr}[YX^*Y^*X + YX^*XY^* + X^*YXY^* + X^*YY^*X] \\
&= \mathrm{Tr}[X^*XY^*Y + XYX^*Y^* + YXY^*X^* + XX^*YY^*].
\end{aligned}
$$

- Taking the sum yields

$$
\begin{aligned}
&||XY - YX||_2^2 + ||YX^* + X^*Y||_2^2 \\
&= \mathrm{Tr}[X^*XYY^* + XX^*Y^*Y + X^*XY^*Y + XX^*YY^*] \\
&= \mathrm{Tr}(X^*X + XX^*)(Y^*Y + YY^*). \qquad (4)
\end{aligned}
$$

Royal Holloway University of London

- The Cauchy-Schwarz inequality yields a lower bound on the added term:

$$\begin{aligned} |\operatorname{Tr}[Y(X^*X + XX^*)]| &= |\operatorname{Tr}[(YX^* + X^*Y)X]| \\ &\leq ||YX^* + X^*Y||_2 \; ||X||_2. \end{aligned} \tag{5}$$

- Combining (4) and (5) then gives

$$\begin{aligned} ||XY - YX||_2^2 \leq\; & \operatorname{Tr}(X^*X + XX^*)(Y^*Y + YY^*) \\ & - |\operatorname{Tr}[Y(X^*X + XX^*)]|^2 / ||X||_2^2. \end{aligned}$$

- The Cauchy-Schwarz inequality yields a lower bound on the added term:

$$\begin{aligned} |\operatorname{Tr}[Y(X^*X + XX^*)]| &= |\operatorname{Tr}[(YX^* + X^*Y)X]| \\ &\leq ||YX^* + X^*Y||_2 \; ||X||_2. \end{aligned}$$

- Combining (4) and (5) then gives

$$\begin{aligned} ||XY - YX||_2^2 \leq \; & \operatorname{Tr}(X^*X + XX^*)(Y^*Y + YY^*) \\ & -|\operatorname{Tr}[Y(X^*X + XX^*)]|^2/||X||_2^2. \end{aligned}$$

- Introducing $\rho = (X^*X + XX^*)/(2||X||_2^2)$, this can be expressed as

$$||XY - YX||_2^2 \leq 4||X||_2^2 \; \left(\operatorname{Tr}[\rho(Y^*Y + YY^*)/2] - |\operatorname{Tr}[\rho Y]|^2\right). \qquad (6)$$

Note that ρ is formally a density matrix: positive semi-definite and trace 1.

- Is the second factor in the RHS bounded above by $||Y||_2^2/2$?

- Consider now the Cartesian decomposition $Y = A + iB$, where A and B are Hermitian.
- One checks that $(Y^*Y + YY^*)/2 = A^2 + B^2$.
- Therefore,

$$\mathrm{Tr}[\rho(Y^*Y + YY^*)/2] - |\,\mathrm{Tr}[\rho Y]|^2 = \mathrm{Tr}\,\rho A^2 + \mathrm{Tr}\,\rho B^2 - (\mathrm{Tr}\,\rho A)^2 - (\mathrm{Tr}\,\rho B)^2,$$

which is a sum of terms in A and in B separately.

Royal Holloway University of London

- Consider now the Cartesian decomposition $Y = A + iB$, where A and B are Hermitian.
- One checks that $(Y^*Y + YY^*)/2 = A^2 + B^2$.
- Therefore,

$$\mathrm{Tr}[\rho(Y^*Y + YY^*)/2] - |\,\mathrm{Tr}[\rho Y]\,|^2 = \mathrm{Tr}\,\rho A^2 + \mathrm{Tr}\,\rho B^2 - (\mathrm{Tr}\,\rho A)^2 - (\mathrm{Tr}\,\rho B)^2,$$

  which is a sum of terms in A and in B separately.
- Is the RHS is bounded above by $||Y||_2^2/2 = (||A||_2^2 + ||B||_2^2)/2$?
- Yes, iff $\mathrm{Tr}\,\rho A^2 - (\mathrm{Tr}\,\rho A)^2 \le ||A||_2^2/2$ for all Hermitian A.

- Let's pass to a basis in which A is diagonal:
  - Put $A = \mathrm{diag}(a_1, \ldots, a_d)$.
  - Denote the diagonal elements of ρ in that basis by p_i.
  - The p_i form a probability distribution!
- The quantity $\operatorname{Tr} \rho A^2 - |\operatorname{Tr} \rho A|^2$ then becomes a variance:

$$\sum_i p_i a_i^2 - (\sum_i p_i a_i)^2.$$

- Proof complete by applying

$$\mathrm{Var}(A) \leq \sum_i a_i^2/2 = ||A||_2^2/2.$$

Overview

- We learned that sharp commutator bounds are indeed related to variance bounds.
- Can we do better, by using better variance bounds?
- In the middle of my proof, you find:

$$||XY - YX||_2^2 \leq 4||X||_2^2 \ \left(\mathrm{Tr}[\rho(Y^*Y + YY^*)/2] - |\,\mathrm{Tr}[\rho Y]|^2\right).$$

- The RHS can be seen as a sort of quantum variance for Y.
- Can we find a generalisation of the Murthy-Sethi bound?
- Coming up next: generalisations to complex valued r.v., quantum case for normal operators, quantum case for non-normal operators.
- Various levels of completion...
- Connections with an old result of Stampfli

Complex values

- Let A be a random variable taking the *complex* values a_i with probability p_i.
- *Define* the variance of A as

$$\begin{aligned}\operatorname{Var}(A) &= \sum_i p_i |a_i - \sum_j p_j a_j|^2 \\ &= \sum_i p_i |a_i|^2 - |\sum_i p_i a_i|^2.\end{aligned}$$

- Statistical meaning: trace of covariance matrix for $(\Re A, \Im A)$.
- **Theorem.** The largest possible value for $\operatorname{Var}(A)$ is $\min_{z\in\mathbb{C}} \max_i |a_i - z|^2$.

Royal Holloway University of London

Complex values

- Interpretation: radius squared of the smallest circle in the complex plane containing the points a_i.
- Defines: *radius* and *center* of a set of points A in $\mathbb{R}^2$.
- **Lemma.** The center of a set is contained in its convex hull.
- Proof of Theorem: via Kakutani's minimax theorem
- For any distribution q, $|\sum_i q_i a_i - \sum_i p_i a_i|^2 \geq 0$. Thus

$$\sum_i p_i |a_i - \sum_j q_j a_j|^2 \geq \sum_i p_i |a_i - \sum_j p_j a_j|^2,$$

and equality is obtained for $q = p$. Hence

$$\sum_i p_i |a_i - \sum_j p_j a_j|^2 = \min_q \sum_i p_i |a_i - \sum_j q_j a_j|^2.$$

Royal Holloway University of London

- As this holds for all p, it holds for the maximisation of both sides over p:

$$\max_p \sum_i p_i |a_i - \sum_j p_j a_j|^2 = \max_p \min_q \sum_i p_i |a_i - \sum_j q_j a_j|^2.$$

- By Kakutani's minimax theorem, the min and max can be interchanged

$$\max_p \sum_i p_i |a_i - \sum_j p_j a_j|^2 = \min_q \max_p \sum_i p_i |a_i - \sum_j q_j a_j|^2.$$

- The max over p gives $\max_i |a_i - \sum_j q_j a_j|^2$.
- For any q, $\sum_j q_j a_j$ is a point in the convex hull of the a_i. Thus the min over q is a min over points in that convex hull.
- By the lemma, the min over any point is in the convex hull anyway. Thus

$$\max_p \sum_i p_i |a_i - \sum_j p_j a_j|^2 = \min_z \max_i |a_i - z|^2.$$

□

Normal matrices

- Can do the same for normal matrices. Role of a_i taken over by eigenvalues.
- Define (in defiance of the rest of the world, including statisticians and mathematical physicists) the *quantum variance* of a normal matrix X in a state determined by density matrix ρ as

$$\mathrm{Var}(X) = \mathrm{Tr}\,\rho|X - \mathrm{Tr}\,\rho X|^2 = \mathrm{Tr}\,\rho|X|^2 - |\,\mathrm{Tr}\,\rho X|^2.$$

- In terms of the eigenvalue decomposition $X = U\Lambda U^*$, with $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots)$ (complex), and putting $p_i = (U^*\rho U)_{ii}$,

$$\mathrm{Var}(X) = \sum_i p_i|\lambda_i|^2 - |\sum_i p_i\lambda_i|^2.$$

Normal matrices

- Variance bound for complex values carries over wholesale to normal matrices.
- Maximum variance is now the square of the radius of the spectrum (not to be confused with the spectral radius); denote this by $r(X)$. Hence

$$r(X) = \min_z ||X - z||.$$

- Center is in the convex hull of the spectrum, i.e. in the numerical range.
- When Y is normal, this gives us a better commutator bound:

$$\begin{aligned} ||XY - YX||_2^2 &\leq 4||X||_2^2 \left(\mathrm{Tr}[\rho Y^* Y] - |\,\mathrm{Tr}[\rho Y]|^2\right) \\ &= 4||X||_2^2 \; \mathrm{Var}_\rho(Y) \\ &\leq 4||X||_2^2 \; r^2(Y). \end{aligned}$$

Relation to norms

- We can relate the radius of a normal matrix to its unitarily invariant (UI) norms.
- Of central importance (pun intended) is the second Ky Fan norm $||X||_{(2)} = \sigma_1(X) + \sigma_2(X)$.
- **Theorem.** $r(X) = \min_z ||X - z|| = \min_z ||X - z||_{(2)}/2$.
- Thus, by taking $z = 0$, $r(X) \leq ||X||_{(2)}/2$.
- Consider the matrix $F = \mathrm{diag}(1, 1, 0, \ldots)$. For each UI norm $|||.|||$, define the F-scaled norm as $|||X|||/|||F|||$.
- The second Ky Fan norm is the smallest F-scaled norm:
- **Theorem.** For any UI norm

$$||X||_{(2)}/2 \leq |||X|||/|||F||| \leq \max(||X||, ||X||_1/2).$$

- Therefore, for any UI norm, $r(X) \leq |||X|||/|||F|||$.

Relation to old work by Stampfli

- John Holbrook pointed out to me that this definition of radius already appeared in a paper by Stampfli.

  J.G. Stampfli, The norm of a derivation, Pacific J. Math. 33, (1970)

- Also about commutators!

Relation to old work by Stampfli

- John Holbrook pointed out to me that this definition of radius already appeared in a paper by Stampfli.

  J.G. Stampfli, The norm of a derivation, Pacific J. Math. 33, (1970)

- Also about commutators!
- However, he only considered the operator norm and found:
- Let D_T be the map $X \mapsto [X, T]$, then $||D_T|| = \max_X ||[X, T]||/||X|| = 2r(T)$, where T need not be normal.

Non-normal matrices

- Work in progress
- Problems:
  - no unique definition of modulus, hence no unique definition of variance;
  - radius $r(X) = \min_z ||X - z||$ is larger than the radius of the smallest circumscribing circle for the spectrum, and even for the numerical range
  - relation radius to second Ky Fan norm?
- Positive result: optimal z is within the numerical range (already known by Stampfli)

Conclusion

- The concept of variance appeared to be of central importance in obtaining sharp norm bounds on commutators.
- Lots of things to do: variance bounds for non-normal matrices, commutator bounds for other norms, NPT problem, heck maybe $L_{p,q}$ spaces
- This is one of the topics I'd like to be working on during this thematic programme.
- Collaborators invited!