Journal of
Mathematical
Inequalities
Volume 3, Number 1 (2009), 15–20

GRÜSS'S INEQUALITY, ITS PROBABILISTIC INTERPRETATION, AND A SHARPER BOUND

RIČARDAS ZITIKIS

(*communicated by C. E. M. Pearce*)

Abstract. Motivated by statistical and actuarial applications of Grüss's inequality, we argue that the inequality can be sharpened if there is additional information about the mean values of the two functions in Grüss's inequality. In this sense, our research deviates from a large body of literature where Grüss's inequality has been sharpened by imposing more smoothness on the functions.

1. Introduction

Grüss (1935) proved that if two functions $f(x)$ and $g(x)$ are such that $a \leqslant f(x) \leqslant A$ and $b \leqslant g(x) \leqslant B$ for all x in an interval $[x_1, x_2]$, then

$$\left| \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} f(x)g(x)dx - \frac{1}{(x_2-x_1)^2} \int_{x_1}^{x_2} f(x)dx \int_{x_1}^{x_2} g(x)dx \right| \leqslant \frac{1}{4}(A-a)(B-b). \quad (1.1)$$

Inequality (1.1) can be reformulated in probabilistic terms as follows. Let ξ be a random variable uniformly distributed on the interval $[x_1, x_2]$. That is, the probability density function of ξ is equal to $(x_2 - x_1)^{-1}$ when $x \in [x_1, x_2]$ and 0 for all other $x \in \mathbf{R}$. The difference inside the absolute values on the left-hand side of inequality (1.1) can now be written as $\mathbf{E}[f(\xi)g(\xi)] - \mathbf{E}[f(\xi)]\mathbf{E}[g(\xi)]$, which is the covariance $\mathbf{Cov}[f(\xi), g(\xi)]$. Consequently, inequality (1.1) is equivalent to

$$\left| \mathbf{Cov}[f(\xi), g(\xi)] \right| \leqslant \frac{1}{4}(A-a)(B-b). \quad (1.2)$$

In general, the constant $1/4$ cannot be replaced by any smaller one. It is worth also noting that when the functions $f(x)$ and $g(x)$ are non-decreasing, or both are non-increasing, then the random variables $X = f(\xi)$ and $Y = g(\xi)$ are called comonotonic and play an important role in actuarial science (see, e.g., Denuit, Dhaene, Goovaerts, and Kaas, 2005).

We can formulate Grüss's inequality for general random variables X and Y without assuming any specific structure of the random variables. Namely, let $(\Omega, \mathscr{A}, \mathbf{P})$ be

Mathematics subject classification (2000): 26D15, 60E15.
Keywords and phrases: Grüss inequality, covariance, mean, variance, inner space, Hilbert space, Hoeffding inequality.

Paper JMI-03-02

a probability triplet, and let X and Y be random variables such that $a \leqslant X \leqslant A$ and $b \leqslant Y \leqslant B$ for some finite $a \leqslant A$ and $b \leqslant B$. Then (see Theorem 1 by Dragomir, 1999)

$$\big|\mathbf{Cov}[X,Y]\big| \leqslant \frac{1}{4}(A-a)(B-b), \tag{1.3}$$

which is the 'probabilistic interpretation' of Grüss's inequality that we refer to in the title of this paper.

NOTE 1.1. The aforementioned Theorem 1 by Dragomir (1999) concerns with general inner spaces. Hence, in order to derive bound (1.3) from the theorem, we choose the space $L_2(\Omega,\mathscr{A},\mathbf{P})$ of all random variables with finite second moments $\mathbf{E}[X^2] = \int_\Omega X^2(\omega)\mathbf{P}(d\omega)$ equipped with the inner product $\langle X,Y\rangle = \mathbf{E}[XY]$. For further generalizations of Grüss's inequality in inner spaces, see Dragomir (2005) and Ma (2007).

In the next section we shall present a problem that has motivated our interest in Grüss's inequality. There we also obtain a sharper inequality assuming additional information about the location of the means $\mathbf{E}[X]$ and $\mathbf{E}[Y]$. Section 3 concludes the paper with afterthoughts.

2. A sharper Grüss-type bound

Numerous statistical and actuarial applications rely on constructing confidence intervals for the unknown mean $\mathbf{E}[X]$, also known as the net premium in actuarial science, using the empirical mean $n^{-1}\sum_{i=1}^n X_i$ of independent (or dependent) copies $X_1,\dots,X_n$ of X. Given a confidence level, say $(1-\alpha)100\%$, the margin of error of the asymptotic confidence interval for $\mathbf{E}[X]$ is $z_{1-\alpha/2}\sqrt{\sigma^2[X]/n}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th standard normal quantile, whose numerical value is known given $\alpha \in (0,1)$. Hence, to determine the minimal sample size n needed to achieve a specified margin of error, we need a good estimate of the variance $\sigma^2[X]$, which is unknown. However, since in many applications we may reasonably assume that $a \leqslant X \leqslant A$ for some finite constants a and A, from Grüss's inequality (1.3) with $X \equiv Y$, and also using the equation $\sigma^2[X] = \mathbf{Cov}[X,X]$, we obtain the bound

$$\sigma^2[X] \leqslant \frac{1}{4}(A-a)^2. \tag{2.1}$$

Two notes follow before we resume our main discussion.

NOTE 2.1. Bound (2.1) is the best possible in the sense that there is a random variable, say X_0, for which inequality (2.1) is an equality. Indeed, rephrasing Dragomir's (1999, pp. 76–77) example, let ε be a Bernoulli random variable taking on two values ± 1 with the same probability $1/2$, and let X_0 be defined by $X_0 = 2^{-1}(A+a) + 2^{-1}(A-a)\varepsilon$. Note that $X_0 = a$ and $X_0 = A$ with the same probability $1/2$, which provides a convenient redefinition of X_0. The mean of this random variable is $\mathbf{E}[X_0] = (A+a)/2$, and the second moment $\mathbf{E}[X_0^2] = (A^2 + a^2)/2$. Consequently, the variance $\sigma^2[X_0]$, which is $\mathbf{E}[X_0^2] - (\mathbf{E}[X_0])^2$, can be written as follows:

$$\sigma^2[X_0] = \frac{1}{4}(A-a)^2.$$

This equation establishes the optimality of bound (2.1). Note in passing that we have just proved that $\sup_{a \leqslant X \leqslant A} \sigma^2[X] = \sigma^2[X_0]$.

NOTE 2.2. Bound (2.1) holds for every random variable X such that $a \leqslant X \leqslant A$. If, however, we also happen to know, or assume, that the mean $\mathbf{E}[X]$ is zero, then we have that $\sigma^2[X] = -aA - \mathbf{E}[(A-X)(X-a)]$ and so

$$\sigma^2[X] \leqslant -aA. \tag{2.2}$$

(Note that when $\mathbf{E}[X] = 0$, then $a \leqslant 0$ and $A \geqslant 0$, and so $-aA \geqslant 0$.) Bound (2.2) is sharper than bound (2.1) except in the case $a = -A$ when the two bounds coincide.

The above notes lead us naturally to the following problem: Suppose that in view of our understanding of the physical phenomenon at hand, or some other considerations, we know that the mean $\mathbf{E}[X]$ is in an interval $[\mu_1, \mu_2] \subseteq [a, A]$. Under this additional information, what upper bound can we have for $\sigma^2[X]$? An answer to this question is provided in Theorem 2.1 below. Before formulating it we first note that the inclusions $\mathbf{E}[X] \in [a, A]$ and $\mathbf{E}[Y] \in [b, B]$ always hold. Furthermore, since $\mathbf{Cov}[X, Y] = 0$ when $a = A$ and/or $b = B$, we can and thus do assume without loss of generality that $a < A$ and $b < B$.

THEOREM 2.1. *Assume that* $\mathbf{E}[X] \in [\mu_a, \mu_A]$ *and* $\mathbf{E}[Y] \in [\mu_b, \mu_B]$ *for some intervals* $[\mu_a, \mu_A] \subseteq [a, A]$ *and* $[\mu_b, \mu_B] \subseteq [b, B]$. *Then we have that*

$$\big|\mathbf{Cov}[X, Y]\big| \leqslant (1 - \mathfrak{A})\,(1 - \mathfrak{B})\,\frac{1}{4}(A-a)(B-b), \tag{2.3}$$

where $\mathfrak{A}$ *and* $\mathfrak{B}$ *are what we call 'information coefficients', defined by the equations*

$$\mathfrak{A} = 1 - \frac{2}{A-a} \sup_{x \in [\mu_a, \mu_A]} \sqrt{(A-x)(x-a)}\,,$$
$$\mathfrak{B} = 1 - \frac{2}{B-b} \sup_{x \in [\mu_b, \mu_B]} \sqrt{(B-x)(x-b)}\,.$$

Proof. The proof is elementary and starts with $|\mathbf{Cov}[X, Y]| \leqslant \sigma[X]\sigma[Y]$, which is of course a consequence of the Cauchy–Bunyakovsky-Schwarz inequality. Next we note that $\sigma^2[X] \leqslant (A - \mathbf{E}[X])(\mathbf{E}[X] - a)$ and, likewise, $\sigma^2[Y] \leqslant (B - \mathbf{E}[Y])(\mathbf{E}[Y] - b)$. Since $\mathbf{E}[X] \in [\mu_a, \mu_A]$ and $\mathbf{E}[Y] \in [\mu_b, \mu_B]$, we have that

$$\begin{aligned} |\mathbf{Cov}[X, Y]| &\leqslant \sqrt{(A - \mathbf{E}[X])(\mathbf{E}[X] - a)}\sqrt{(B - \mathbf{E}[Y])(\mathbf{E}[Y] - b)} \\ &\leqslant \sup_{x \in [\mu_a, \mu_A]} \sqrt{(A-x)(x-a)} \sup_{x \in [\mu_b, \mu_B]} \sqrt{(B-x)(x-b)}. \end{aligned}$$

Bound (2.3) follows.

NOTE 2.3. The 'information coefficients' $\mathfrak{A}$ and $\mathfrak{B}$ are always in the interval $[0,1]$. The fact that the coefficients do not exceed 1 immediately follows from their definitions. To see that the coefficients are non-negative, we just need to notice that the function $x \mapsto (A-x)(x-a)$ achieves its maximum at $x=(a+A)/2$, which is in the interval $[a,A]$.

NOTE 2.4. When $[\mu_a,\mu_A]=[a,A]$, which means that there is no additional information about the mean $\mathbf{E}[X]$ since it is always in $[a,A]$, then $\mathfrak{A}=0$. Likewise, when $[\mu_b,\mu_B]=[b,B]$, then $\mathfrak{B}=0$. In summary, when there is no 'useful' additional information about X and Y, that is, when $\mathfrak{A}=0$ and $\mathfrak{B}=0$, then we have Grüss's bound (1.3).

NOTE 2.5. The case $(a+A)/2 \in [\mu_a,\mu_A]$ gives $\mathfrak{A}=0$. Hence, in this case knowing that $\mathbf{E}[X] \in [\mu_a,\mu_A]$ is not useful in the current context. However, when the interval $[\mu_a,\mu_A]$ does not cover the point $(a+A)/2$, then $\mathfrak{A}>0$ and thus bound (2.3) is sharper than Grüss's bound (1.3). In this case, therefore, knowing that $\mathbf{E}[X] \in [\mu_a,\mu_A]$ is a useful bit of information.

NOTE 2.6. If we know that $\mathbf{E}[X]=0$, then we can choose $\mu_a=0=\mu_A$, which gives the equation

$$\mathfrak{A}=1-\frac{2}{A-a}\sqrt{-aA}.$$

If, in addition, $\mathbf{E}[Y]=0$, then we have an analogous equation for $\mathfrak{B}$. Hence, when the means of X and Y are both equal to zero, then bound (2.3) reduces to

$$\big|\mathbf{Cov}[X,Y]\big| \leqslant \sqrt{-aA}\sqrt{-bB}. \tag{2.4}$$

When $X=Y$, bound (2.4) reduces to bound (2.2).

NOTE 2.7. In view of Theorem 2.1, we can now reflect upon Grüss's original inequality (1.1) and see that it can be sharpened depending on the values of $(x_2-x_1)^{-1}\int_{x_1}^{x_2} f(x)dx$ and $(x_2-x_1)^{-1}\int_{x_1}^{x_2} g(x)dx$. Likewise, in the notation of Dragomir (1999) and Ma (2007), the upper bound of Grüss's inequality in inner spaces can be sharpened depending on the values of the inner products $\langle x,e\rangle$ and $\langle y,e\rangle$.

NOTE 2.8. There are of course many other extensions and generalizations of Grüss's inequality. For recent contributions to the area, we refer to, e.g., Dragomir (2005), Cerone (2006), Dragomir (2007), Elezović, Marangunić and Pečarić (2007), as well as to the references therein.

3. Afterthoughts

In our considerations so far, we have not assumed any particular dependence structure between the random variables X and Y. One can of course take a different route and specify a dependence structure of some particular theoretical or practical interest. For example, when X and Y are 'perfectly' positively dependent (i.e.,

$X = Y$) or 'perfectly' negatively dependent (i.e., $X = -Y$), then bound (1.3) says that $|\mathbf{Cov}[X,Y]| \leqslant (A-a)^2/4$. On the other hand, when X and Y are uncorrelated, then $\mathbf{Cov}[X,Y] = 0$ by definition. This opens up yet another venue for sharpening Grüss's inequality.

NOTE 3.1. Dependence structures between two random variables are frequently defined in terms of their joint distribution function $(s,t) \mapsto \mathbf{P}[X \leqslant s, Y \leqslant t]$. This function and the covariance $\mathbf{Cov}[X,Y]$ are related to each other via Hoeffding's (1940) equality

$$\mathbf{Cov}[X,Y] = \int_{\mathbf{R}}\int_{\mathbf{R}} \Big(\mathbf{P}[X \leqslant s, Y \leqslant t] - \mathbf{P}[X \leqslant s]\,\mathbf{P}[Y \leqslant t]\Big) ds\,dt; \tag{3.1}$$

see Block and Fang (1988) for a multivariate version. Among various ways for specifying the joint distribution function, we can utilize the notion of copula, which has been actively explored and utilized in various theoretical and applied contexts (see, e.g., Denuit, Dhaene, Goovaerts, and Kaas, 2005; Nelsen, 2006).

Next, let $X = f(\xi)$ and $Y = g(\xi)$, where the random variable ξ is same as in Section 1. Then the integrand in (3.1) can be written as

$$\begin{aligned} &\frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \mathbf{1}\{f(x) \leqslant s, g(x) \leqslant t\} dx \\ &\qquad - \left(\frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \mathbf{1}\{f(x) \leqslant s\} dx\right)\left(\frac{1}{x_2 - x_1}\int_{x_1}^{x_2} \mathbf{1}\{g(x) \leqslant t\} dx\right), \end{aligned} \tag{3.2}$$

where the indicator $\mathbf{1}\{S\}$ is equal to 1 if statement S is correct and 0 otherwise. Replacing the integrand in (3.1) by quantity (3.2), we arrive at an alternative representation of the covariance $\mathbf{Cov}[f(\xi), g(\xi)]$ in (1.1). Note, for example, that

$$(x_2 - x_1)^{-1}\int_{x_1}^{x_2} \mathbf{1}\{f(x) \leqslant s\} dx$$

can be interpreted as the amount of 'time' that the function $f(x)$ spends below the level s relative to the length of the 'time' interval $[x_1, x_2]$. Analogous interpretations hold for the two other integrals in (3.2).

Acknowledgements. The author is grateful to an anonymous referee for constructive criticism and insightful comments that have helped in revising the paper. The research has been supported by a Discovery Research Grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

[1] BLOCK, H. W. AND FANG, Z., *A multivariate extension of Hoeffding's lemma*, Annals of Probability, **16** (1988), 1803–1820.

[2] CERONE, P., *On a Čebyšev-type functional and Grüss-like bounds*, Mathematical Inequalities and Applications, **9** (2006), 87–102.

[3] DENUIT, M., DHAENE, J., GOOVAERTS, M. AND KAAS, R., *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, Wiley, Chichester, 2005.

[4] DRAGOMIR, S. S., *A generalization of Grüss's inequality in inner product spaces and applications*, Journal of Mathematical Analysis and Applications, **237** (1999), 74–82.

[5] DRAGOMIR, S. S., *Advances in Inequalities of the Schwarz, Grüss and Bessel Type in Inner Product Spaces*, Nova Science Publishers, Hauppauge, New York, 2005.

[6] DRAGOMIR, S. S., *Sharp Grüss-type inequalities for functions whose derivatives are of bounded variation*, Journal of Inequalities in Pure and Applied Mathematics, **8**, Article 117 (2007), 13 pp.

[7] ELEZOVIĆ, N., MARANGUNIĆ, LJ. AND PEČARIĆ, J., *Some improvements of Grüss type inequality*, Journal of Mathematical Inequalities, **1** (2007), 425–436.

[8] GRÜSS, G., *Über das maximum des absoluten betrages von* $\frac{1}{b-a}\int_a^b f(x)g(x)\,dx - \frac{1}{(b-a)^2}\int_a^b f(x)\,dx\int_a^b g(x)\,dx$, Mathematische Zeitschrift, **39** (1935), 215–226.

[9] HOEFFDING, W., *Masstabinvariante korrelationstheorie*, Schriften des Matematischen Instituts für Angewandte Matematik der Universität Berlin, **5** (1940), 179–233.

[10] MA, J., *An identity in real inner product spaces*, Journal of Inequalities in Pure and Applied Mathematics, **8**, Article 48 (2007), 4 pp.

[11] NELSEN, R. B., *An Introduction to Copulas*. (2nd edition.) Springer, New York, 2006.

(Received January 6, 2008)
(Revised July 24, 2008)

Ričardas Zitikis
Department of Statistical and Actuarial Sciences
University of Western Ontario, London
Ontario N6A 5B7, Canada
e-mail: zitikis@stats.uwo.ca