International Journal of Pure and Applied Mathematics
Volume 56 No. 2 2009, 217-221

INNER PRODUCT OF RANDOM VECTORS

Eungchun Cho
Division of Mathematics and Sciences
Kentucky State University
Frankfort, KY, 40601, USA
e-mail: eung.cho@kysu.edu

Abstract: Distribution function of the inner product of random vectors uniformly distributed on the unit sphere in E^{n+1} is derived. The distribution function of the inner product is useful in analyzing the autocorrelation function of time series.

AMS Subject Classification: 62H10
Key Words: autocorrelation function, gamma function, hypersphere, inner product, moment, random vector, variance

1. Introduction

Most pairs of random vectors in high dimensional Euclidean space are almost orthogonal. We make this statement precise by deriving the pdf (probability density function) of the inner product of random unit vectors in the $n+1$ dimensional Euclidean space E^{n+1}. We also derive a general formula for the moments of the inner product.

Geometrically, the correlation coefficient of random variables X and Y is the cosine of the angle between $X-E(X)$ and $Y-E(Y)$, that is, the inner product of the unit vectors representing them in the variable space. Our results are applicable for studying time series [4] and the hypothesis testing of correlatedness [3] between random variables. A common practice of using $[-2/\sqrt{n}, 2/\sqrt{n}]$ as the 95 % confidence interval for the acf (autocorrelation function) of a time

Received: August 28, 2009

series of length n being white Gaussian noise is only approximately correct. We give precise numerical values for the confidence interval.

2. Inner Product of Random Unit Vectors in E^{n+1}

Let U and V be random vectors uniformly distributed on the standard unit n-sphere

$$S^n = \{x \in E^{n+1} : |x| = 1\}.$$

Let $Z = \langle U, V \rangle$, the inner product of U and V. It is a random variable on the outcome space $S^n \times S^n$ with values in $[-1, 1]$. The pdf and and the moments of Z are given in the following theorems.

Theorem 1. *The pdf of Z, denoted by f_n, is given as*

$$f_n(z) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi}} \sqrt{1-z^2}^{\,n-2}, & \text{for } -1 < z < 1, \\ 0, & \text{elsewhere}, \end{cases} \tag{1}$$

where $n = 1, 2, 3, \ldots$

Proof. Let u and v be unit vectors in E^{n+1}. We may assume (after a rotation) $u = (0, \ldots, 0, 1)$ so that $\langle u, v \rangle = v_{n+1}$. $F_n(z) = Pr(Z \le z)$, the cumulative distribution function of Z, is given as

$$F_n(z) = \frac{\text{vol}\,(R_z)}{\text{vol}\,(S^n)}, \tag{2}$$

where $R_z = \{v \in S^n : v_{n+1} \le z\}$ for $z \in \mathbf{R}$. Recall the volume of the n-sphere with radius r is given as

$$\text{vol}\,(S^n_r) = \frac{2\sqrt{\pi}^{\,n+1}}{\Gamma(\frac{n+1}{2})}\, r^n.$$

Substitution of the volume of R_z

$$\begin{aligned} \text{vol}\,(R_z) &= \int_{-1}^{z} \text{vol}\,(S^{n-1}_{\sqrt{1-x^2}}) \frac{dx}{\sqrt{1-x^2}} & (3) \\ &= \int_{-1}^{z} \frac{2\sqrt{\pi}^{\,n}}{\Gamma(\frac{n}{2})} \sqrt{1-x^2}^{\,n-2}\, dx & (4) \end{aligned}$$

into (2) yields

$$F_n(z) = \int_{-1}^{z} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi}} \sqrt{1-x^2}^{\,n-2}\, dx,$$

hence

$$f_n(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}} \sqrt{1-z^2}^{\,n-2}. \qquad \square$$

Remarks. Note $f_2(z)$ is constant, that is, Z is uniformly distributed for $n = 2$. We can also write

$$f_n(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \sqrt{1-z^2}^{\,n-2}.$$

Substitution of $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(x+1) = x\,\Gamma(x)$ into the equation (1) gives

$$\begin{aligned}
f_1 &= \frac{1}{\pi\sqrt{1-z^2}}, \\
f_2 &= \frac{1}{2}, \\
f_n &= \frac{2\cdot 4\cdots(n-1)}{1\cdot 3\cdots(n-2)\,\pi} \sqrt{1-z^2}^{\,n-2} \quad \text{for } n = 3, 5, 7, \ldots, \\
&= \frac{1}{2}\,\frac{3\cdot 5\cdots(n-1)}{2\cdot 4\cdots(n-2)} \sqrt{1-z^2}^{\,n-2} \quad \text{for } n = 4, 6, \ldots .
\end{aligned}$$

for $-1 < z < 1$. Obviously, all the odd order moments of Z vanish as f_n is even. The even order moments are given in the following

Theorem 2. *Under the same hypothesis, the even order moments*

$$E(Z^{2m}) = \frac{\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{1}{2}+m\right)}{\Gamma\left(\frac{n+1}{2}+m\right)\Gamma(\frac{1}{2})},$$

where $m = 1, 2, 3 \ldots.$

Proof. By definition,

$$\begin{aligned}
E\left(Z^{2m}\right) &= \int_{-1}^{1} f_n(z)\, z^{2m}\, dz \\
&= \int_{-1}^{1} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \sqrt{1-z^2}^{\,n-2} z^{2m} dz \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos t)^{n-1} (\sin t)^{2m}\, dt
\end{aligned}$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{\Gamma\left(\frac{2m+1}{2}\right)}{\Gamma\left(\frac{2m+n+1}{2}\right)}.$$

We used the following identity in the last step to derive the formula.

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos t)^r (\sin t)^s \, dt = \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{s+1}{2}\right)}{\Gamma\left(\frac{r+s+2}{2}\right)}. \qquad \square$$

Substitution of

$$\Gamma\left(\frac{2m+1}{2}\right) = \frac{2m-1}{2} \cdot \frac{2m-3}{2} \cdots \frac{3}{2} \cdot \frac{1}{2} \Gamma\left(\frac{1}{2}\right),$$

$$\Gamma\left(\frac{2m+n+1}{2}\right) = \frac{2m+n-1}{2} \cdot \frac{2m+n-3}{2} \cdots \frac{n+3}{2} \cdot \frac{n+1}{2} \Gamma\left(\frac{n+1}{2}\right)$$

yields

$$E\left(Z^{2m}\right) = \frac{1 \cdot 3 \cdots (2m-1)}{(n+1)(n+3)\cdots(n+2m-1)}.$$

In particular, $E(Z) = 0$, $E(Z^2) = 1/(n+1)$, and $E(Z^4) = 3/(n+1)(n+3)$.

f_n obviously is not Gaussian. The kurtosis $B_2 = 3(n+1)/(n+3)$ of f_n is less than 3 for all n, that is, f_n has a flatter top (platykurtic). However, the common practice of using $\pm 2\sigma = \pm 2/\sqrt{n}$ for the 95% confidence interval of $Z = 0$ is problematic, specially for small values of n. The confidence interval should be narrower. Even for n large, the interval $[-2/\sqrt{n}, 2/\sqrt{n}]$ covers more than 95% as the following numerical estimation shows. The table below gives the values of

$$p = \int_{\frac{-2}{\sqrt{n}}}^{\frac{2}{\sqrt{n}}} f_{n-1}(x) dx.$$

n	4	5	6	7	8	9	10
p	1.0000	0.98391	0.97496	0.96998	0.96688	0.96469	0.96321

n	1000	1200	1400	1600	1800	2000
p	0.95455	0.95457	0.95457	0.95453	0.95456	0.95455

Acknowledgments

Eungchun Cho's work at Seoul National University was supported by The Korea Research Foundation and The Korean Federation of Science and Technology Societies Grant funded by Korea Government (MOEHRD, Basic Research Promotion Fund).

References

[1] M. Abramowitz, I.A. Stegun, Ed-s., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9-th Edition, New York, Dover (1972), 255-258.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second Edition, New York, McGraw-Hill (1984), 145-149.

[3] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Linear Correlation. 4.5 in Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Second Edition, Cambridge, England, Cambridge University Press (1992), 630-633.

[4] R.H. Shumway, D.S. Stoffer, *Time Series Analysis and its Applications, with R Examples*, Second Edition, New York, Springer-Verlag (2000).

[5] Eric W. Weisstein, *MathWorld – A Wolfram Web Resource*, http://mathworld.wolfram.com/TimeSeriesAnalysis.html