Properties of Dot Product of Random Vectors

Yeqi Bai
School of Computer Science and Engineering
Nanyang Technological University
yeqi001@ntu.edu.sg

1 Introduction

This article provides a proof for the relationship mentioned in [1]:

Lemma 1 *Assume that the components of q and k are independent random variables with mean 0 and variance 1. Then their dot product, $q \cdot k = \sum_{i=1}^{d_k} q_i k_i$, has mean 0 and variance d_k.*

The structure of this article can be summarized as followed: (1) Section 2 introduces the basic properties of mean and variance of random variables; (2) Section 3 and 4 revisit properties of sum and product of random variables; and (3) Section 5 provides a final proof to Lemma 1.

2 Foundations

2.1 Variance

The variance of a random variable X is defined as:

$$\begin{aligned}
\mathrm{var}[X] &= \mathrm{E}[(X - \mathrm{E}(X))^2] \\
&= \mathrm{E}[X^2 + \mathrm{E}[X]^2 - 2X\,\mathrm{E}[X]] \\
&= \mathrm{E}[X^2] + \mathrm{E}[X]^2 - 2\,\mathrm{E}[X]^2 \\
&= \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned} \tag{1}$$

2.2 Covariance

The covariance of two random variables X and Y is defined as:

$$\begin{aligned}
\mathrm{cov}[X, Y] &= \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))] \\
&= \mathrm{E}[XY - X\,\mathrm{E}[Y] - \mathrm{E}[X]Y + \mathrm{E}[X]\,\mathrm{E}[Y]] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] - \mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[X]\,\mathrm{E}[Y] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]
\end{aligned} \tag{2}$$

Note also that:

$$\mathrm{cov}[X, X] = \mathrm{var}[X] \tag{3}$$

The covariance of two independent random variables is 0.

3 Product of Random Variables

3.1 Mean of Product of Random Variables

By the law of total expectation, the mean of the product of two random variables X and Y can be derived as:

$$\begin{aligned} \mathrm{E}[XY] &= \mathrm{E}\left[\mathrm{E}[XY|Y]\right] \\ &= \mathrm{E}\left[Y \cdot \mathrm{E}[X|Y]\right] \end{aligned} \tag{4}$$

When X and Y are independent, $\mathrm{E}[X|Y] = \mathrm{E}[X]$, the above equation can be simplified as:

$$\begin{aligned} \mathrm{E}[XY] &= \mathrm{E}\left[Y \cdot \mathrm{E}[X]\right] \\ &= \mathrm{E}[X] \cdot \mathrm{E}[Y] \end{aligned} \tag{5}$$

3.2 Variance of Product of Random Variables

The variance of the product of two random variables X and Y can be formulated as:

$$\mathrm{var}[XY] = \mathrm{E}[X^2Y^2] - \mathrm{E}[XY]^2 \tag{6}$$

According to Equation 2 and 1:

$$\begin{aligned} \mathrm{E}[X^2Y^2] &= \mathrm{cov}[X^2, Y^2] + \mathrm{E}[X^2]\,\mathrm{E}[Y^2] \\ &= \mathrm{cov}[X^2, Y^2] + \big(\mathrm{E}[X]^2 + \mathrm{var}[X]\big) \cdot \big(\mathrm{E}[Y]^2 + \mathrm{var}[Y]\big) \end{aligned} \tag{7}$$

and:

$$\mathrm{E}[XY]^2 = \left(\mathrm{cov}[X, Y] + \mathrm{E}[X]\,\mathrm{E}[Y]\right)^2 \tag{8}$$

Afterwards, we substitute Equation 7 and 8 into Equation 6, and obtain:

$$\mathrm{var}[XY] = \mathrm{cov}[X^2, Y^2] + \big(\mathrm{E}[X]^2 + \mathrm{var}[X]\big) \cdot \big(\mathrm{E}[Y]^2 + \mathrm{var}[Y]\big) - \left(\mathrm{cov}[X, Y] + \mathrm{E}[X]\,\mathrm{E}[Y]\right)^2 \tag{9}$$

When X and Y are independent, $\mathrm{cov}[X^2, Y^2] = \mathrm{cov}[X, Y] = 0$, Equation 9 reduces to:

$$\begin{aligned} \mathrm{var}[XY] &= \big(\mathrm{E}[X]^2 + \mathrm{var}[X]\big) \cdot \big(\mathrm{E}[Y]^2 + \mathrm{var}[Y]\big) - \left(\mathrm{E}[X]\,\mathrm{E}[Y]\right)^2 \\ &= \mathrm{E}[X]^2\,\mathrm{var}[Y] + \mathrm{E}[Y]^2\,\mathrm{var}[X] + \mathrm{var}[X]\,\mathrm{var}[Y] \end{aligned} \tag{10}$$

In the case that both X and Y has 0 mean, the above can be further reduced to:

$$\mathrm{var}[XY] = \mathrm{var}[X]\,\mathrm{var}[Y] \tag{11}$$

4 Sum of Random Variables

In this section, we revise the properties of the sum of several random variables. In particular, we study a random variable Z given by

$$Z = \sum_{i=1}^{n} X_i \tag{12}$$

4.1 Mean of Sum of Random Variables

According to the linearity of expectation:

$$\mathrm{E}[Z] = \sum_{i=1}^{n} \mathrm{E}[X_i] \tag{13}$$

4.2 Variance of Sum of Random Variables

The variance of multiple random variables can be derived as:

$$\begin{aligned}\mathrm{var}(Z) &= \mathrm{cov}\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}[X_i, X_j]\end{aligned} \tag{14}$$

Provided independence between each X_i, the above equation can be simplified as:

$$\begin{aligned}\mathrm{var}(Z) &= \sum_{i=1}^{n} \mathrm{cov}[X_i, X_i] \\ &= \sum_{i=1}^{n} \mathrm{var}[X_i]\end{aligned} \tag{15}$$

5 Dot Product of Random Vectors

Lemma 1 states that both q and k are vector with dimension d_k, whose components are independent random variables with the following properties:

$$\begin{aligned}&\mathrm{E}[q_i] = \mathrm{E}[k_i] = 0 \\ &\mathrm{var}[q_i] = \mathrm{var}[k_i] = 1\end{aligned} \tag{16}$$

where $i \in [0, d_k]$.

With the help of properties revised in Section 3 and 4, the mean of the dot product $q \cdot k$ is

$$\begin{aligned}\mathrm{E}[q \cdot k] &= \mathrm{E}\left[\sum_{i=1}^{d_k} q_i k_i\right] \\ &= \sum_{i=1}^{d_k} \mathrm{E}\left[q_i k_i\right] \\ &= \sum_{i=1}^{d_k} \mathrm{E}[q_i]\,\mathrm{E}[k_i] \\ &= 0\end{aligned} \tag{17}$$

Similarly, we formulate the variance of $q \cdot k$, based on the properties in Section 3 and 4:

$$\begin{aligned}\mathrm{var}[q \cdot k] &= \mathrm{var}\left[\sum_{i=1}^{d_k} q_i k_i\right] \\ &= \sum_{i=1}^{d_k} \mathrm{var}\left[q_i k_i\right] \\ &= \sum_{i=1}^{d_k} \mathrm{var}[q_i]\,\mathrm{var}[k_i] \\ &= \sum_{i=1}^{d_k} 1 \\ &= 1\end{aligned} \tag{18}$$

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.