

Sums and Products of Jointly Distributed Random Variables: A Simplified Approach

Sheldon H. Stein
Cleveland State University

Journal of Statistics Education Volume 13, Number 3 (2005), jse.amstat.org/v13n3/stein.html

Copyright © 2005 by Sheldon H. Stein, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Covariance; Joint probability distribution; Means; Variances.

Abstract

Three basic theorems concerning expected values and variances of sums and products of random variables play an important role in mathematical statistics and its applications in education, business, the social sciences, and the natural sciences. A solid understanding of these theorems requires that students be familiar with the proofs of these theorems. But while students who major in mathematics and other technical fields should have no difficulties coping with these proofs, students who major in education, business, and the social sciences often find it difficult to follow these proofs. In many textbooks and courses in statistics which are geared to the latter group, mathematical proofs are sometimes omitted because students find the mathematics too confusing. In this paper, we present a simpler approach to these proofs. This paper will be useful for those who teach students whose level of mathematical maturity does not include a solid grasp of differential calculus.

1. Introduction

The following three theorems play an important role in all disciplines which make use of mathematical statistics:

Let X and Y be two jointly distributed random variables. Then

THEOREM 1. $E(X + Y) = E(X) + E(Y)$

THEOREM 2. $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$

THEOREM 3. If X and Y are statistically independent, then $E(XY) = E(X) E(Y)$

where $E(\cdot)$ represents expected value, $\text{var}(\cdot)$ is variance, and $\text{cov}(\cdot)$ is the covariance.

In mathematical statistics, one relies on these theorems to derive estimators and to examine their properties. But textbooks and courses differ with regard to the extent to which they cover these theorems and their proofs. The best coverage is found in texts like [Hogg and Tannis \(2001\)](#) and [Mendenhall, Scheaffer, and Wackerly \(1997\)](#). But texts written for “Statistics 101” courses or for students who are not majors in mathematics or statistics, like [Becker \(1995\)](#), [Newbold \(1991\)](#), and [Mansfield \(1994\)](#) tend to have sparser coverage. Also, in some fields, like

finance and econometrics, the subject matter relies rather heavily on these theorems and their proofs. But other fields seem to take a more casual approach to the mathematical foundations of statistical theory. Thus, in a statement put out by the European Union-India Cross Cultural Innovation Network, snowwhite.it.brighton.ac.uk/research/euindia/knowledgebase/indiapg/curriculum.htm we find the following statement:

“The traditional phobia for statistics among biology students arises out of the association of mathematics and statistics. That, however, is required only for the proofs of the formulae and not of their use. The availability of large amounts of data and good computational tools is what is required to make sense of statistics.”

But something is lost when the student does not understand the proofs of these theorems. For example, some of my students have argued with me in class, when the textbook was not immediately available, that Theorems 1 and 2 do require that X and Y be statistically independent and that Theorem 3 is always true. So without understanding the proofs, can one really understand their content and the statistical concepts that they support?

But even students who must master the content of these theorems are often uncomfortable with their formal proofs. The proofs of these three theorems make use of double summation notation in the case of discrete random variables. I suspect that a large part of the problem that students have with the proofs is a result of the fact many of them suffer from what might be termed "double summation notation anxiety." I have found that by simplifying the standard proofs, the theorems and their proofs become more accessible. While the simplifications involve some loss of generality, I believe on the basis of my personal experience that the tradeoff is well worth it.

2. Joint Probability Distributions

To understand [Theorems 1, 2, and 3](#), one must first understand what is meant by a joint probability distribution. While some textbooks, written for mathematics and statistics majors (e.g. [Hogg and Tannis \(2001\)](#) and [Mendenhall, Scheaffer, and Wackerly\(1997\)](#)) as well as other majors (like [Mirer \(1995\)](#)) illustrate the concept of a joint probability distribution by using an example for the discrete case, many (e.g. [Becker \(1995\)](#), [Kmenta \(1997\)](#) and [Newbold \(1991\)](#)) do not, which makes it more difficult for the student to grasp the material.

Consider two random variables X and Y distributed as in [Table 1](#):

Table 1. Joint Probability Distribution of X and Y

$X \backslash Y$	y_1	y_2	...	y_n	$P(X)$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$...	$p(x_1, y_n)$	$P(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$...	$p(x_2, y_n)$	$P(x_2)$
...
x_m	$p(x_m, y_1)$	$p(x_m, y_2)$...	$p(x_m, y_n)$	$P(x_m)$
$P(Y)$	$P(y_1)$	$P(y_2)$...	$P(y_n)$	1

where $p(x_i, y_j)$ is the joint probability that $X = x_i$ and $Y = y_j$. Also, let $P(x_i)$ be the simple probability that $X = x_i$ and let $P(y_j)$ be the simple probability that $Y = y_j$. $P(x_i)$ and $P(y_j)$ are also called marginal probabilities because they appear on the margins of the table. Note that the simple probability that X assumes the value x_i is

$$P(x_i) = p(x_i, y_1) + p(x_i, y_2) + \dots + p(x_i, y_n),$$

since when i assumes the value 1, j assumes a value between 1 and n . Similarly

$$P(y_j) = p(x_1, y_j) + p(x_2, y_j) + \dots + p(x_m, y_j).$$

The sum of all of the joint probabilities $p(x_i, y_j)$ is of course equal to unity, as is the sum of all of the marginal probabilities of X and the sum of all of the marginal probabilities of Y .

Let us consider the standard proof of [Theorem 1](#):

$$\begin{aligned} E(X + Y) &= \sum_i \sum_j (x_i + y_j) p(x_i, y_j) \\ &= \sum_i \sum_j x_i p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j) \\ &= \sum_i x_i \sum_j p(x_i, y_j) + \sum_j y_j \sum_i p(x_i, y_j) \\ &= \sum_i x_i P(x_i) + \sum_j y_j P(y_j) \\ &= E(X) + E(Y) \end{aligned}$$

because $\sum_j p(x_i, y_j) = P(x_i)$ and $\sum_i p(x_i, y_j) = P(y_j)$. The proofs of [Theorems 2 and 3](#) follow a similar format (See [Kmenta \(1997\)](#).)

My experience has been that most students do not feel comfortable with these proofs. And in many textbooks, the proof is presented without a joint probability table which compounds the difficulty for students. In either case, many students will memorize Theorems 1, 2, and 3 and learn their proofs if necessary.

3. Theorem 1: An Alternative Proof

I have found that most of my students can understand the proof of [Theorem 1](#) (and the other theorems) if it is presented in the following manner. Let us assume that random variables X and Y take on just two values each, as in [Table 2](#):

Table 2. Simplified Joint Probability Distribution of X and Y

$X \backslash Y$	y_1	y_2	$P(X)$
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$P(x_1)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$P(x_2)$
$P(Y)$	$P(y_1)$	$P(y_2)$	1

The definition of an expected value of a random variable like X is, as we know

$$E(X) = x_1P(x_1) + x_2P(x_2).$$

In terms of the [table](#) above, the simple probabilities are of course equal to the sums of the relevant joint probabilities:

$$P(X_1) = [p(x_1, y_1) + p(x_1, y_2)],$$

$$P(X_2) = [p(x_2, y_1) + p(x_2, y_2)].$$

Hence, it follows that

$$E(X) = x_1p(x_1, y_1) + x_1p(x_1, y_2) + x_2p(x_2, y_1) + x_2p(x_2, y_2).$$

In other words, we multiply each joint probability by its X value and then sum. Many students have an easier time with this mode of presentation than the equivalent one expressed using double summation notation.

To determine $E(X + Y)$, students are told that for each of the four interior cells of the joint probability distribution, we

1. add the values of X and Y corresponding to that cell;
2. multiply that sum by the corresponding joint probability; and
3. sum these products across all elements of the table.

Hence,

$$E(X + Y) = (x_1 + y_1)p(x_1, y_1) + (x_1 + y_2)p(x_1, y_2) + (x_2 + y_1)p(x_2, y_1) + (x_2 + y_2)p(x_2, y_2).$$

The expected value of $X + Y$ is just a weighted average of the four possible values of $x_i + y_j$ with the joint probabilities serving as the weights.

By expanding the above expression and collecting terms, we obtain

$$E(X+Y) = x_1[p(x_1, y_1) + p(x_1, y_2)] + x_2[p(x_2, y_1) + p(x_2, y_2)] + y_1[p(x_1, y_1) + p(x_2, y_1)] + y_2[p(x_1, y_2) + p(x_2, y_2)].$$

Because

$$P(x_1) = [p(x_1, y_1) + p(x_1, y_2)],$$

$$P(x_2) = [p(x_2, y_1) + p(x_2, y_2)],$$

$$P(y_1) = [p(x_1, y_1) + p(x_2, y_1)],$$

$$P(y_2) = [p(x_1, y_2) + p(x_2, y_2)],$$

we can write

$$E(X + Y) = [x_1P(x_1) + x_2P(x_2)] + [y_1P(y_1) + y_2P(y_2)] = E(X) + E(Y).$$

This proof is straightforward, intuitive, and does not require the use of double or even single summation signs. If one wants to make the proof a little more general, one can allow for a third value of one of the random variables. Once the student understands the simple proof above, it is, hopefully, a simple step forward to the more general proof.

4. Theorem 2

The proofs of [Theorems 2 and 3](#) can be constructed using the same approach. Because the variance of any random variable Z is $E[Z - E(Z)]^2$ the variance of $X + Y$ is again derived using each cell of the joint probability table:

$$\text{VAR}(X + Y) = [(x_1 + y_1) - E(X + Y)]^2 p(x_1, y_1) + [(x_1 + y_2) - E(X + Y)]^2 p(x_1, y_2) + [(x_2 + y_1) - E(X + Y)]^2 p(x_2, y_1) + [(x_2 + y_2) - E(X + Y)]^2 p(x_2, y_2).$$

By applying [Theorem 1](#), we can rewrite this as

$$\text{VAR}(X + Y) = [(x_1 - E(X)) + (y_1 - E(Y))]^2 p(x_1, y_1) + [(x_1 - E(X)) + (y_2 - E(Y))]^2 p(x_1, y_2) + [(x_2 - E(X)) + (y_1 - E(Y))]^2 p(x_2, y_1) + [(x_2 - E(X)) + (y_2 - E(Y))]^2 p(x_2, y_2).$$

Now applying $(a+b)^2 = a^2 + 2ab + b^2$, and collecting terms:

$$\begin{aligned} \text{VAR}(X+Y) = & \{[x_1 - E(X)]^2[p(x_1, y_1) + p(x_1, y_2)]\} + \{[x_2 - E(X)]^2[p(x_2, y_1) + p(x_2, y_2)]\} + \{[y_1 - E(Y)]^2[p(x_1, y_1) + p(x_2, y_1)]\} \\ & + \{[y_2 - E(Y)]^2[p(x_1, y_2) + p(x_2, y_2)]\} \\ & + 2\{[(x_1 - E(X))(y_1 - E(Y))]p(x_1, y_1) + [(x_1 - E(X))(y_2 - E(Y))]p(x_1, y_2) + [(x_2 - E(X))(y_1 - E(Y))]p(x_2, y_1) + [(x_2 - E(X))(y_2 - E(Y))]p(x_2, y_2)\}. \end{aligned}$$

Again, since

$$P(x_1) = [p(x_1, y_1) + p(x_1, y_2)],$$

$$P(x_2) = [p(x_2, y_1) + p(x_2, y_2)],$$

$$P(y_1) = [p(x_1, y_1) + p(x_2, y_1)],$$

$$P(y_2) = [p(x_1, y_2) + p(x_2, y_2)],$$

the expression above becomes $\text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y)$ where the expression in the third set of braces is, by definition, the covariance of X and Y .

5. Theorem 3

To obtain $E(XY)$, in each cell of the joint probability distribution table, we multiply each joint probability by its corresponding X and Y values:

$$E(XY) = x_1 y_1 p(x_1, y_1) + x_1 y_2 p(x_1, y_2) + x_2 y_1 p(x_2, y_1) + x_2 y_2 p(x_2, y_2).$$

At this point, the assumption of statistical independence of X and Y is utilized. If X and Y are statistically independent, then

$$p(x_i, y_j) = P(x_i)P(y_j).$$

Hence

$$E(XY) = x_1 y_1 P(x_1)P(y_1) + x_1 y_2 P(x_1)P(y_2) + x_2 y_1 P(x_2)P(y_1) + x_2 y_2 P(x_2)P(y_2).$$

Factoring out the $x_1 P(x_1)$ expression that is common to the first two terms and the $x_2 P(x_2)$ expression common to the third and fourth terms, we have

$$E(XY) = x_1 P(x_1)[y_1 P(y_1) + y_2 P(y_2)] + x_2 P(x_2)[y_1 P(y_1) + y_2 P(y_2)].$$

Factoring out the common term, $[y_1 P(y_1) + y_2 P(y_2)]$.

$$E(XY) = [x_1 P(x_1) + x_2 P(x_2)][y_1 P(y_1) + y_2 P(y_2)] = E(X)E(Y).$$

This proof is much simpler than the general proof that is found in statistics textbooks. As the student sees the role of the assumption of statistical independence in the proof of [Theorem 3](#) and the lack of such a role in the proofs of [Theorems 1 and 2](#), the students' understanding of these three theorems is complete.

6. Numerical Examples

The following numerical examples of the theorems is very useful in clarifying the material discussed above. Three cases will be presented—one in which the covariance between two random variables is positive, another in which the covariance is negative, and a third in which two random variables are statistically independent, which gives rise to a zero covariance. Consider an experiment where two fair coins are tossed in the air simultaneously. The four possible outcomes of this chance experiment are HH, HT, TH, and TT where H represents heads and T represents tails. Consider [Table 3](#) below where the random variable X represents the number of heads in an outcome and Y represents the number of tails:

Table 3. Generation of Random Variables X and Y and Their Sum

Outcome	Probability	X	Y	X + Y
HH	0.25	2	0	2
HT	0.25	1	1	2
TH	0.25	1	1	2
TT	0.25	0	2	2

Clearly, X and Y have the same distribution, as seen in [Table 4](#).

Table 4. Probability Distribution of X (Also Y)

X (also Y)	Probability
0	0.25
1	0.50
2	0.25

In [Table 5](#), the joint probability distribution of X and Y is presented:

Table 5. Joint Probability Distribution of X and Y

--	--	--	--	--	--

X\Y	0	1	2	P(X)
0	0	0	0.25	0.25
1	0	0.50	0	0.50
2	0.25	0	0	0.25
P(Y)	0.25	0.50	0.25	1

We can now illustrate [Theorems 1, 2, and 3](#) by using either the simple or joint probabilities of X and Y. The expected value of X (and also of Y) is $(0)(0.25) + (1)(0.50) + (2)(0.25)$, or 1.0. The variance of X (and also of Y) is $(0-1)^2(0.25) + (1-1)^2(0.50) + (2-1)^2(0.25)$ or 0.5. Notice that the simple probability distributions from [Table 4](#) are the same as the marginal probability distributions of [Table 5](#). In [Table 3](#), we see that the expected value of $X + Y$ is obviously 2 and the variance of $X + Y$ is obviously 0 because the number of heads plus the number of tails in each possible outcome is always 2. This is consistent with [Theorem 1](#) since $E(X)$ and $E(Y)$ are both 1. The covariance of X and Y [i.e. $[(2-1)(0-1)(0.25) + (1-1)(1-1)(0.50) + (0-1)(2-1)(0.25)]$] is -0.5. Hence, [Theorem 2](#), which says that the variance of $X + Y$ is equal to the variance of X plus the variance of Y plus two times the covariance of X and Y, which add up to zero, is consistent with $X + Y$ being 2 for all possible outcomes.

Note also that [Theorems 1 and 2](#) find support in this example even though random variables X and Y are not statistically independent (since the joint probabilities are not equal to the product of their corresponding marginal probabilities). But Theorem 3 does not hold as $E(XY)$ is equal to 0.5, which is not equal to the product of $E(X)$ and $E(Y)$, which is 1.

Let us now consider [Table 6](#) in which the random variables R and S are defined in a different way. Random variable R is defined purely on the basis of the toss of the first coin and random variable S is defined purely on the basis of the second coin toss. If the outcome of the first coin toss is heads, we assign random variable R the value 1 while if the first coin toss is a tails, we assign R the value 0 without regard to the outcome of the second toss. Similarly, random variable S is defined solely on the basis of the toss of the second coin, without regard to the outcome of the first coin toss. If the outcome of the second coin toss is heads, we assign random variable S the value 1 while if the second coin toss is a tails, we assign S the value 0. Hence, random variables R and S are statistically independent.

Table 6. Generation of Random Variables R and S and Their Sum

Outcome	Probability	R	S	R + S
HH	0.25	1	1	2
HT	0.25	1	0	1
TH	0.25	0	1	1
TT	0.25	0	0	0

Note in [Table 7](#) that R and S have the same distribution.

Table 7. Probability Distribution of R (Also S)

R (also S)	Probability
-------------------	--------------------

0	0.50
1	0.50

The expected value of R is 0.5 and the expected value of S is also 0.5; the variance of R is 0.25 and the variance of S is also 0.25.

From [Table 6](#) we also derive [Table 8](#) which shows the distribution of the random variable $R + S$:

Table 8. Probability Distribution of $R + S$

$R + S$	Probability
0	0.25
1	0.50
2	0.25

The expected value of $R + S$ is clearly 1 and its variance is 0.5. From [Table 6](#), we also derive [Table 9](#) which presents the joint probability distribution table of random variables R and S . The covariance of R and S is zero, which is a consequence of the fact that R and S are statistically independent. We know this because each joint probability is equal to the product of the corresponding marginal probabilities. Since the expected values of R and S are each 0.5 and since the variances of R and S are each 0.25 with the covariance of R and S being zero, these results are consistent with [Theorems 1 and 2](#).

Table 9. Joint Probability Distribution of R and S

$R \backslash S$	0	1	$P(R)$
0	0.25	0.25	0.50
1	0.25	0.25	0.50
$P(S)$	0.50	0.50	1

[Theorem 3](#) requires that $E(RS)$ be equal to $E(R)E(S)$ since R and S are statistically independent. Given our values of $E(R)$ and $E(S)$, $E(RS)$ should be equal to 0.25. In [Table 9](#), it is obvious that this is the case.

Finally, we construct an example where random variables are defined in a coin flipping experiment where the covariance between the two random variables are positive. Consider [Table 10](#) below. The generation of random variable X was presented in [Table 3](#). Random variable W is defined by assigning a value of 2 if both coins turn up heads, -2 if both coins turn up tails, and zero if one coin turns up heads and the other turns up tails.

Table 10. Generation of Random Variables X and W and Their Sum

Outcome	Probability	X	W	$X + W$
HH	0.25	2	2	4

HT	0.25	1	0	1
TH	0.25	1	0	1
TT	0.25	0	-2	-2

The probability distribution for X was shown in [Table 4](#) with a mean of 1 and a variance of 0.5 while the probability distribution for W is shown in [Table 11](#). Its mean is equal to 0

Table 11. Probability Distribution of W

W	Probability
-2	0.25
0	0.50
2	0.25

and its variance is 2. The distribution of $X + W$ is shown in [Table 12](#):

Table 12. Probability Distribution of $X + W$

$X + W$	Probability
-2	0.25
1	0.50
4	0.25

$X + W$ has an expected value of 1 and a variance of 4.5. The joint probability distribution of X and W is found in [Table 13](#):

Table 13. Joint Probability Distribution of X and W

$X \backslash W$	-2	0	2	$P(X)$
0	0.25	0	0	0.25
1	0	0.50	0	0.50
2	0	0	0.25	0.25
$P(W)$	0.25	0.50	0.25	1

The covariance of X and W in this example turns out to be 1.0. The expected value of $X + W$, which is 1, is equal to the sum of the expected value of X , which is 1 and the expected value of W , which is zero. The variance of $X + W$, which is 4.5 is equal to the variance of X , which is 0.5 plus the variance of W , which is 2, plus two times the covariance of X and W , or 2. But because X and W are not statistically independent, which we can see by comparing the joint probabilities with the products of the concomitant marginal probabilities, [Theorem 3](#)

need not apply, and it does not. $E(X)E(W)$ is equal to zero but cursory examination of [Table 13](#) reveals that $E(XW)$ is equal to 1.

7. Conclusion

In teaching a course in money and financial markets, I went over the proofs of [Theorems 1, 2, and 3](#) in the manner presented above. While I did it in order to facilitate the understanding of the application of those theorems to the study of risk, diversification, and the capital asset pricing model, my students reported that it also helped them understand the discussion of those theorems in an econometrics course that they were taking simultaneously. Having taught that econometrics course in prior years, I remember the tension in the classroom when I went over the proofs of the three theorems above in the conventional manner with the double summation notation. Unless one understands these theorems and their proofs, a course in econometrics becomes an exercise in rote memorization rather than in understanding. The same can be said for other courses using these theorems. Thus, the techniques outlined in this paper are useful whenever the objective of a course is not only to teach students how to do statistics, but also to understand statistics.

Acknowledgement

I would like to give thanks to two anonymous referees for useful comments and suggestions and to my graduate students at Cleveland State University for providing the inspiration to write this paper.

References

- Becker, W. E. (1995), *Statistics for Business and Economics*, Cincinnati, OH: Southwestern.
- Hogg, R. and Tanis, E (2001), *Probability and Statistics*, Upper Saddle River, NJ: Prentice Hall.
- Kmenta, J. (1997), *Elements of Econometrics*, 2nd Ed., New York: Macmillan.
- Mansfield, E. (1994), *Statistics for Business and Economics: Methodology and Applications*, New York: Norton.
- Mendenhall, W., Scheaffer, R., and D. Wackerly (1997), *Mathematical Statistics with Applications*, Boston: Duxbury Press.
- Mirer, T. (1995). *Economic Statistics and Econometrics*, Englewood Cliffs, NJ: Prentice Hall.
- Newbold, P. (1991), *Statistics for Business and Economics*, Englewood Cliffs, NJ: Prentice Hall.

Sheldon H. Stein
Department of Economics
Cleveland State University
Cleveland, OH
U.S.A.
S.Stein@csuohio.edu