Pearson New International Edition

Probability and Statistics

Morris DeGroot Mark Schervish

Fourth Edition

ALWAYS LEARNING

PEARSON

Pearson New International Edition

Probability and Statistics

Morris DeGroot Mark Schervish

Fourth Edition

PEARSON

if $0 < y < 1$. Using properties of power series from calculus, we know that the derivative of $g(y)$ can be found by differentiating the individual terms of the power series. That is,

$$g'(y) = \sum_{x=0}^{\infty} x y^{x-1} = \sum_{x=1}^{\infty} x y^{x-1},$$

for $0 < y < 1$. But we also know that $g'(y) = 1/(1-y)^2$. The last sum in Eq. (4.5.3) is $g'(0.999) = 1/(0.001)^2$. It follows that

$$E(X) = 0.001 \frac{1}{(0.001)^2} = 1000.$$ ◂

Minimizing the Mean Absolute Error

Another possible basis for predicting the value of a random variable X is to choose some number d for which $E(|X - d|)$ will be a minimum.

Definition 4.5.3 Mean Absolute Error/M.A.E. The number $E(|X - d|)$ is called the *mean absolute error* (M.A.E.) of the prediction d.

We shall now show that the M.A.E. is minimized when the chosen value of d is a median of the distribution of X.

Theorem 4.5.3 Let X be a random variable with finite mean, and let m be a median of the distribution of X. For every number d,

$$E(|X - m|) \le E(|X - d|). \tag{4.5.4}$$

Furthermore, there will be equality in the relation (4.5.4) if and only if d is also a median of the distribution of X.

Proof For convenience, we shall assume that X has a continuous distribution for which the p.d.f. is f. The proof for any other type of distribution is similar. Suppose first that $d > m$. Then

$$\begin{aligned}
E(|X - d|) - E(|X - m|) &= \int_{-\infty}^{\infty} (|x - d| - |x - m|) f(x)\, dx \\
&= \int_{-\infty}^{m} (d - m) f(x)\, dx + \int_{m}^{d} (d + m - 2x) f(x)\, dx + \int_{d}^{\infty} (m - d) f(x)\, dx \\
&\ge \int_{-\infty}^{m} (d - m) f(x)\, dx + \int_{m}^{d} (m - d) f(x)\, dx + \int_{d}^{\infty} (m - d) f(x)\, dx \\
&= (d - m)[\Pr(X \le m) - \Pr(X > m)]. \qquad (4.5.5)
\end{aligned}$$

Since m is a median of the distribution of X, it follows that

$$\Pr(X \le m) \ge 1/2 \ge \Pr(X > m). \tag{4.5.6}$$

The final difference in the relation (4.5.5) is therefore nonnegative. Hence,

$$E(|X - d|) \ge E(|X - m|). \tag{4.5.7}$$

Furthermore, there can be equality in the relation (4.5.7) only if the inequalities in relations (4.5.5) and (4.5.6) are actually equalities. A careful analysis shows that these inequalities will be equalities only if d is also a median of the distribution of X.

The proof for every value of d such that $d < m$ is similar. ■

Example 4.5.8 Last Lottery Number. In Example 4.5.5, in order to compute the median of X, we must find the smallest number x such that the c.d.f. $F(x) \geq 0.5$. For integer x, we have

$$F(x) = \sum_{n=1}^{x} 0.001(0.999)^{n-1}.$$

We can use the popular formula

$$\sum_{n=0}^{x} y^n = \frac{1 - y^{x+1}}{1 - y}$$

to see that, for integer $x \geq 1$,

$$F(x) = 0.001 \frac{1 - (0.999)^x}{1 - 0.999} = 1 - (0.999)^x.$$

Setting this equal to 0.5 and solving for x gives $x = 692.8$; hence, the median of X is 693. The median is unique because $F(x)$ never takes the exact value 0.5 for any integer x. The median of X is much smaller than the mean of 1000 found in Example 4.5.7. ◀

The reason that the mean is so much larger than the median in Examples 4.5.7 and 4.5.8 is that the distribution has probability at arbitrarily large values but is bounded below. The probability at these large values pulls the mean up because there is no probability at equally small values to balance. The median is not affected by how the upper half of the probability is distributed. The following example involves a symmetric distribution. Here, the mean and median(s) are more similar.

Example 4.5.9 Predicting a Discrete Uniform Random Variable. Suppose that the probability is 1/6 that a random variable X will take each of the following six values: 1, 2, 3, 4, 5, 6. We shall determine the prediction for which the M.S.E. is minimum and the prediction for which the M.A.E. is minimum.

In this example,

$$E(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5.$$

Therefore, the M.S.E. will be minimized by the unique value $d = 3.5$.

Also, every number m in the closed interval $3 \leq m \leq 4$ is a median of the given distribution. Therefore, the M.A.E. will be minimized by every value of d such that $3 \leq d \leq 4$ and only by such a value of d. Because the distribution of X is symmetric, the mean of X is also a median of X. ◀

Note: When the M.A.E. and M.S.E. Are Finite. We noted that the median exists for every distribution, but the M.A.E. is finite if and only if the distribution has a finite mean. Similarly, the M.S.E. is finite if and only if the distribution has a finite variance.

Summary

A median of X is any number m such that $\Pr(X \leq m) \geq 1/2$ and $\Pr(X \geq m) \geq 1/2$. To minimize $E(|X - d|)$ by choice of d, one must choose d to be a median of X. To minimize $E[(X - d)^2]$ by choice of d, one must choose $d = E(X)$.

Exercises

1. Prove that the 1/2 quantile as defined in Definition 3.3.2 is a median as defined in Definition 4.5.1.

2. Suppose that a random variable X has a discrete distribution for which the p.f. is as follows:

$$f(x) = \begin{cases} cx & \text{for } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{otherwise.} \end{cases}$$

Determine all the medians of this distribution.

3. Suppose that a random variable X has a continuous distribution for which the p.d.f. is as follows:

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Determine all the medians of this distribution.

4. In a small community consisting of 153 families, the number of families that have k children ($k = 0, 1, 2, \ldots$) is given in the following table:

Number of children	Number of families
0	21
1	40
2	42
3	27
4 or more	23

Determine the mean and the median of the number of children per family. (For the mean, assume that all families with four or more children have only four children. Why doesn't this point matter for the median?)

5. Suppose that an observed value of X is equally likely to come from a continuous distribution for which the p.d.f. is f or from one for which the p.d.f. is g. Suppose that $f(x) > 0$ for $0 < x < 1$ and $f(x) = 0$ otherwise, and suppose also that $g(x) > 0$ for $2 < x < 4$ and $g(x) = 0$ otherwise. Determine: **(a)** the mean and
(b) the median of the distribution of X.

6. Suppose that a random variable X has a continuous distribution for which the p.d.f. f is as follows:

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the value of d that minimizes
(a) $E[(X - d)^2]$ and **(b)** $E(|X - d|)$.

7. Suppose that a person's score X on a certain examination will be a number in the interval $0 \le X \le 1$ and that X has a continuous distribution for which the p.d.f. is as follows:

$$f(x) = \begin{cases} x + \frac{1}{2} & \text{for } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine the prediction of X that minimizes **(a)** the M.S.E. and **(b)** the M.A.E.

8. Suppose that the distribution of a random variable X is symmetric with respect to the point $x = 0$ and that $E(X^4) < \infty$. Show that $E[(X - d)^4]$ is minimized by the value $d = 0$.

9. Suppose that a fire can occur at any one of five points along a road. These points are located at -3, -1, 0, 1, and 2 in Fig. 4.9. Suppose also that the probability that each of these points will be the location of the next fire that occurs along the road is as specified in Fig. 4.9.

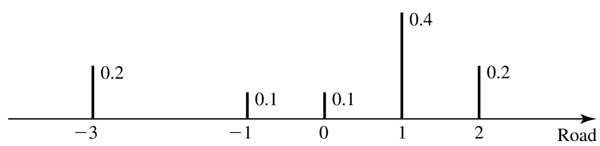

Figure 4.9 Probabilities for Exercise 9.

a. At what point along the road should a fire engine wait in order to minimize the expected value of the square of the distance that it must travel to the next fire?

b. Where should the fire engine wait to minimize the expected value of the distance that it must travel to the next fire?

10. If n houses are located at various points along a straight road, at what point along the road should a store be located in order to minimize the sum of the distances from the n houses to the store?

11. Let X be a random variable having the binomial distribution with parameters $n = 7$ and $p = 1/4$, and let Y be a random variable having the binomial distribution with parameters $n = 5$ and $p = 1/2$. Which of these two random variables can be predicted with the smaller M.S.E.?

12. Consider a coin for which the probability of obtaining a head on each given toss is 0.3. Suppose that the coin is to be tossed 15 times, and let X denote the number of heads that will be obtained.

a. What prediction of X has the smallest M.S.E.?

b. What prediction of X has the smallest M.A.E.?

13. Suppose that the distribution of X is symmetric around a point m. Prove that m is a median of X.

14. Find the median of the Cauchy distribution defined in Example 4.1.8.

15. Let X be a random variable with c.d.f. F. Suppose that $a < b$ are numbers such that both a and b are medians of X.

a. Prove that $F(a) = 1/2$.

b. Prove that there exist a smallest $c \le a$ and a largest $d \ge b$ such that every number in the closed interval $[c, d]$ is a median of X.

c. If X has a discrete distribution, prove that $F(d) > 1/2$.

16. Let X be a random variable. Suppose that there exists a number m such that $\Pr(X < m) = \Pr(X > m)$. Prove that m is a median of the distribution of X.

17. Let X be a random variable. Suppose that there exists a number m such that $\Pr(X < m) < 1/2$ and $\Pr(X > m) < 1/2$. Prove that m is the unique median of the distribution of X.

18. Prove the following extension of Theorem 4.5.1. Let m be the p quantile of the random variable X. (See Definition 3.3.2.) If r is a strictly increasing function, then $r(m)$ is the p quantile of $r(X)$.

4.6 Covariance and Correlation

When we are interested in the joint distribution of two random variables, it is useful to have a summary of how much the two random variables depend on each other. The covariance and correlation are attempts to measure that dependence, but they only capture a particular type of dependence, namely linear dependence.

Covariance

Example 4.6.1 Test Scores. When applying for college, high school students often take a number of standardized tests. Consider a particular student who will take both a verbal and a quantitative test. Let X be this student's score on the verbal test, and let Y be the same student's score on the quantitative test. Although there are students who do much better on one test than the other, it might still be reasonable to expect that a student who does very well on one test to do at least a little better than average on the other. We would like to find a numerical summary of the joint distribution of X and Y that reflects the degree to which we believe a high or low score on one test will be accompanied by a high or low score on the other test. ◀

When we consider the joint distribution of two random variables, the means, the medians, and the variances of the variables provide useful information about their marginal distributions. However, these values do not provide any information about the relationship between the two variables or about their tendency to vary together rather than independently. In this section and the next one, we shall introduce summaries of a joint distribution that enable us to measure the association between two random variables, determine the variance of the sum of an arbitrary number of dependent random variables, and predict the value of one random variable by using the observed value of some other related variable.

Definition 4.6.1 Covariance. Let X and Y be random variables having finite means. Let $E(X) = \mu_X$ and $E(Y) = \mu_Y$ The *covariance of X and Y*, which is denoted by $\text{Cov}(X, Y)$, is defined as

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)], \qquad (4.6.1)$$

if the expectation in Eq. (4.6.1) exists.

It can be shown (see Exercise 2 at the end of this section) that if both X and Y have finite variance, then the expectation in Eq. (4.6.1) will exist and $\text{Cov}(X, Y)$ will be finite. However, the value of $\text{Cov}(X, Y)$ can be positive, negative, or zero.

Example 4.6.2

Test Scores. Let X and Y be the test scores in Example 4.6.1, and suppose that they have the joint p.d.f.

$$f(x, y) = \begin{cases} 2xy + 0.5 & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

We shall compute the covariance $\text{Cov}(X, Y)$. First, we shall compute the means μ_X and μ_Y of X and Y, respectively. The symmetry in the joint p.d.f. means that X and Y have the same marginal distribution; hence, $\mu_X = \mu_Y$. We see that

$$\begin{aligned} \mu_X &= \int_0^1 \int_0^1 [2x^2y + 0.5x]dydx \\ &= \int_0^1 [x^2 + 0.5x]dx = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}, \end{aligned}$$

so that $\mu_Y = 7/12$ as well. The covariance can be computed using Theorem 4.1.2. Specifically, we must evaluate the integral

$$\int_0^1 \int_0^1 \left(x - \frac{7}{12}\right)\left(y - \frac{7}{12}\right)(2xy + 0.5)\, dy\, dx.$$

This integral is straightforward, albeit tedious, to compute, and the result is $\text{Cov}(X, Y) = 1/144$. ◂

The following result often simplifies the calculation of a covariance.

Theorem 4.6.1

For all random variables X and Y such that $\sigma_X^2 < \infty$ and $\sigma_Y^2 < \infty$,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \tag{4.6.2}$$

Proof It follows from Eq. (4.6.1) that

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y. \end{aligned}$$

Since $E(X) = \mu_X$ and $E(Y) = \mu_Y$, Eq. (4.6.2) is obtained. ■

The covariance between X and Y is intended to measure the degree to which X and Y tend to be large at the same time or the degree to which one tends to be large while the other is small. Some intution about this interpretation can be gathered from a careful look at Eq. (4.6.1). For example, suppose that $\text{Cov}(X, Y)$ is positive. Then $X > \mu_X$ and $Y > \mu_Y$ must occur together and/or $X < \mu_X$ and $Y < \mu_Y$ must occur together to a larger extent than $X < \mu_X$ occurs with $Y > \mu_Y$ and $X > \mu_X$ occurs with $Y < \mu_Y$. Otherwise, the mean would be negative. Similarly, if $\text{Cov}(X, Y)$ is negative, then $X > \mu_X$ and $Y < \mu_Y$ must occur together and/or $X < \mu_X$ and $Y > \mu_Y$ must occur together to larger extent than the other two inequalities. If $\text{Cov}(X, Y) = 0$, then the extent to which X and Y are on the same sides of their respective means exactly balances the extent to which they are on opposite sides of their means.

Correlation

Although $\text{Cov}(X, Y)$ gives a numerical measure of the degree to which X and Y vary together, the magnitude of $\text{Cov}(X, Y)$ is also influenced by the overall magnitudes of X and Y. For example, in Exercise 5 in this section, you can prove that $\text{Cov}(2X, Y) = 2\,\text{Cov}(X, Y)$. In order to obtain a measure of association between X and Y that is *not driven by arbitrary changes in the scales* of one or the other random variable, we define a slightly different quantity next.

Definition 4.6.2 Correlation. Let X and Y be random variables with finite variances σ_X^2 and σ_Y^2, respectively. Then the *correlation of X and Y*, which is denoted by $\rho(X, Y)$, is defined as follows:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \tag{4.6.3}$$

In order to determine the range of possible values of the correlation $\rho(X, Y)$, we shall need the following result.

Theorem 4.6.2 Schwarz Inequality. For all random variables U and V such that $E(UV)$ exists,

$$[E(UV)]^2 \le E(U^2)E(V^2). \tag{4.6.4}$$

If, in addition, the right-hand side of Eq. (4.6.4) is finite, then the two sides of Eq. (4.6.4) equal the same value if and only if there are nonzero constants a and b such that $aU + bV = 0$ with probability 1.

Proof If $E(U^2) = 0$, then $\Pr(U = 0) = 1$. Therefore, it must also be true that $\Pr(UV = 0) = 1$. Hence, $E(UV) = 0$, and the relation (4.6.4) is satisfied. Similarly, if $E(V^2) = 0$, then the relation (4.6.4) will be satisfied. Moreover, if either $E(U^2)$ or $E(V^2)$ is infinite, then the right side of the relation (4.6.4) will be infinite. In this case, the relation (4.6.4) will surely be satisfied.

For the rest of the proof, assume that $0 < E(U^2) < \infty$ and $0 < E(V^2) < \infty$. For all numbers a and b,

$$0 \le E[(aU + bV)^2] = a^2E(U^2) + b^2E(V^2) + 2abE(UV) \tag{4.6.5}$$

and

$$0 \le E[(aU - bV)^2] = a^2E(U^2) + b^2E(V^2) - 2abE(UV). \tag{4.6.6}$$

If we let $a = [E(V^2)]^{1/2}$ and $b = [E(U^2)]^{1/2}$, then it follows from the relation (4.6.5) that

$$E(UV) \ge -[E(U^2)E(V^2)]^{1/2}. \tag{4.6.7}$$

It also follows from the relation (4.6.6) that

$$E(UV) \le [E(U^2)E(V^2)]^{1/2}. \tag{4.6.8}$$

These two relations together imply that the relation (4.6.4) is satisfied.

Finally, suppose that the right-hand side of Eq. (4.6.4) is finite. Both sides of (4.6.4) equal the same value if and only if the same is true for either (4.6.7) or (4.6.8). Both sides of (4.6.7) equal the same value if and only if the rightmost expression in (4.6.5) is 0. This, in turn, is true if and only if $E[(aU + bV)^2] = 0$, which occurs if and only if $aU + bV = 0$ with probability 1. The reader can easily check that both sides of (4.6.8) equal the same value if and only if $aU - bV = 0$ with probability 1. ∎