PROBABILITY DISTRIBUTIONS AND THE DEFINITE INTEGRAL

John E. Sasser

Abstract

Beginning with the most basic and elementary notions in probability, a mathematical model is developed for predicting the outcome of experiments based on assumptions about certain associated events occurring. This model is stated in the form of a continuous function and its accompanying graph. Then using this function and its graph, it is shown how the area under the graph can be used to calculate probabilities for various events.

Discussion

Probability theory attempts to develop models for predicting the outcomes of experiments based on assumptions about the likelihood of certain associated events occurring. Those assumptions are stated in the form of a function called a probability distribution. First, we shall develop the basic principles of probability theory that let us determine a function to use for this purpose. Secondly, we shall show how the definite integral is used to calculate probabilities for various events.

We shall develop these principles of probability theory that will lead us to a probability distribution function through a series of carefully chosen questions. We begin with the question:

QUESTION 1

HOW MANY DIFFERENT ORDERINGS, P, ARE POSSIBLE WITH N OBJECTS?

SOLUTION

Objects	Number of objects	Permutations	Number of P
A	1	{A}	1
A,K	2	{(A,K),(K,A)}	2
A,K,Q	3	{(A,K,Q),(K,A,Q),(Q,A,K), (A,Q,K),(K,Q,A),(Q,K,A)}	6
A,K,Q,J	4	4! = 4*3*2*1	24
	:	:	:
	:	:	:
A,K,Q,J, …	N	N! = N(N-1)(N-2)…(3)(2)(1)	N!

Table 1. Solution 1.

Thus, there are N! different orderings possible with N objects.

> **QUESTION 2**
>
> **IF ONE HAS FOUR FACE CARDS PLACED FACE-DOWN, WHAT ISTHE PROBABILITY OF TURNING THE CARDS FACE-UP IN THE FOLLOWING ORDER: A, K, Q, J?**

SOLUTION

Since there are 24 possible outcomes, the probability of selecting this single event (randomly of course) is:

P(A,K,Q,J) = 1/24

In Question 2, we have associated a probability value with a *single event.*

Now consider the set {A,B,C,D}. Rank the objects by two's:

Rank	1	2	3	4	5	6	7	8	9	10	11	12
1	A	A	A	B	B	B	C	C	C	D	D	D
2	B	C	D	A	C	D	A	B	D	A	B	C

Table 2. Ranking four objects by two's.

> **QUESTION 3**
>
> **IF A COMMITTEE MUST RANK TWO APPLICANTS OUT OF A TOTAL OF FOUR THEY CONSIDER TO BE THE BEST FOR THE JOB, WHAT IS THE PROBABILITY THAT APPLICANTS A AND C WILL BE SELECTED WITH A RANKED FIRST?**

SOLUTION

If we are concerned with order, i.e., "what is the probability that applicants A and C being selected with applicant A ranked first, and applicant C ranked second, then we have one possible event out of 12, or P(event)=1/12.

We denote 4 objects taken 2 at a time: $_4P_2$ = 4!/(4-2)! = 4!/2! = (4*3*2*1)/(2*1) = 12.

> **FORMULA 1**
> **in general**
>
> $$nPx = \frac{n!}{(n-x)!}$$

In QUESTION 3, we have associated a probability value with a *subset of events*.

In both QUESTION 2 and QUESTION 3, we have been concerned with the number of ways objects of a set can be ordered, i.e., the number of permutations.

Now, we may not be concerned with the order the committee ranks the two applicants. That is, we may be concerned only with which two have been selected. In this case two sets of objects are considered to be identical if they contain exactly the same elements, no matter how these objects are arranged.

A set of objects where order is unimportant is called a combination.

Notice in Table 2 that #2 and #7 represent the same two applicants, just in different order. This is also true for #'s 1 and 4; #'s 3 and 10; #'s 5 and 8; #'s 6 and 11 and #'s 9 and 12.

Thus, we have these six combinations:

Different Combinations					
1	2	3	4	5	6
A	A	A	B	B	C
B	C	D	C	D	D

Table 3. Combinations

There are always fewer combinations than permutations for a given n and x, since different orderings do not count as combinations, but do count as permutations.

We saw in Table 2 that 4 objects taken 2 at a time gives us 12 permutations:

nPx = n!/(n-x)!=4!/(4-2)!=(4*3*2*1)/(2*1) = 12 permutations.

And we saw in Table 3, above, that 4 objects taken 2 at a time gives us 6 combinations:

nCx $= {}_4C_2$ = 6 combinations.

Now, 6*2! = 12

or, NCx * x! = nPx

which implies NCx $= \frac{{}_nP_x}{x!}$

but, nPx $= \frac{n!}{(n-x)!} * \frac{1}{x!}$

so nCx $= \frac{n!}{[x!(n-x)!]}$

FORMULA 2
Thus

$$nCx = \frac{n!}{[x!(n-x)!]}$$

QUESTION 4

WHAT IS THE PROBABILITY (ALL THINGS BEING EQUAL) THAT APPLICANT A WILL GET THE JOB?

SOLUTION

$_nC_x$ = n!/[x!(n-x)!]

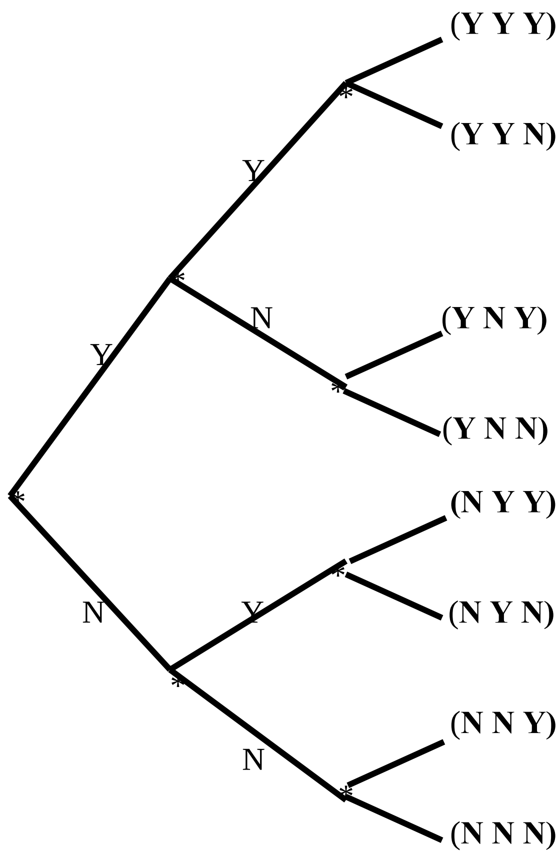
$_4C_2$ = 4!/[2!(4-2)!] = {(4*3*2*1)/[2*1(2*1)]} = 12/2 = 6.

All things being equal, the probability that applicant A gets the job is 1/6.

QUESTION 5

THREE PEOPLE ARE ASKED IF THEY VOTED FOR GOVERNOR MOORE. DETERMINE THE PROBABILITY OF THE OUTCOME "TWO VOTED FOR, ONE AGAINST.

From the diagram, which follows on the next page, we know that there are eight different sample point:



P(one sample pt.) = 1/8

DIAGRAM 1

The following sample points have two YES and one NO:

Relevant pts = (YYN), (YNY), (NYY) = 3.

In these circumstances the probability of an event can be determined by:

P(Event) = (No. of relevant points) X P(1 sample pt.)

Thus, P(2Y,1N) = (3)(1/8) = 3/8 = 0.375

OR, P(Event) = (No. of relevant points) X P(1 sample pt.)
= $_nC_x$ * P(1 sample pt.)
= {n!/[x!(n-x)!]} * (1/8)
= {3!/[2!(3-2)!]} * (1/8)
= (6/2) * (1/8)
= 3/8 = 0.375

.

Again, in QUESTION 5, we associated a probability value with a subset of events.

Finally, let us consider all possible events in an experiment.

QUESTION 6 SUPPOSE WE ROLL TWO DICE. WHAT IS THE PROBABILITY THAT WE WOULD ROLL A NINE?

SOLUTION

```
                    (3,6)
        6 ---|----|----*----|----|----|-
                         (4,5)
        5 ---|----|----|----*----|----|-
                              (5,4)
        4 ---|----|----|----|----*----|-
Die A                                 (6,3)
        3 ---|----|----|----|----|----*-

        2 ---|----|----|----|----|----|-

        1 ---|----|----|----|----|----|-
             1    2    3    4    5    6
                    Die B
```

DIAGRAM 2

The probability of any value x is given by the number of sample points for which the number equals x, divided by the total number of sample points:

For {x=9} => {(3,6),(4,5),(5,4),(6,3)} = 4 points.

Thus, P(9) = 4/36 = 0.111

Let us determine the probabilities of all possible events and graph this distribution:

P(2) = 1/36 = 0.0278
P(3) = 2/36 = 0.0556
P(4) = 3/36 = 0.0833
P(5) = 4/36 = 0.1111
P(6) = 5/36 = 0.1389
P(7) = 6/36 = 0.1666
P(8) = 5/36 = 0.1389
P(9) = 4/36 = 0.1111
P(10)= 3/36 = 0.0833
P(11)= 2/36 = 0.0556
P(12)= 1/36 = 0.0278

Total= 1.0000

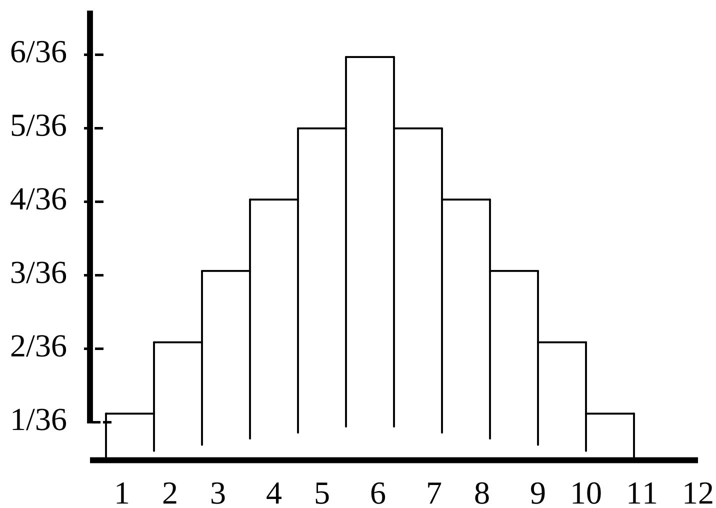


DIAGRAM 3

QUESTION 7

SUPPOSE FOUR COINS ARE RANDOMLY SELECTED FROM A BOX CONTAINING 70% DIMES AND 30% NICKELS. WHAT IS THE PROBABILITY THAT THREE OF THE FOUR COINS SELECTED WILL BE DIMES?

The probability that the first coin selected is a dime is .7. The probability that the first two coins are dimes is .7*.7 and the probability that the first three coins are dimes is .7*.7*.7. And the probability of selecting a nickel is .3. Thus, the probability of selecting three dimes and one nickel is:

7 * .7 * .7 * .3 = 1029/10000=0.1029.

There are four relevant points (combinations):

{d,d,d,n}, {d,d,n,d}, {d,n,d,d}, {n,d,d,d}

P(Event) = (No. of relevant pts) * P(one such point)

$= \quad nCx \quad * \; [(7/10)^3][(3/10)^1]$

$= \quad {}_4C_3 \quad * \; [(7/10)^3][(3/10)^1]$

We will let π represent the probability of success, i.e., in this case, .70. Thus, 1 - π is the probability of failure, i.e., in this case, .30.

For the P(one such sample pt.) we must find the probability of any one ordering of outcomes where there are x successes. If there are x successes in n trials, there must be (n-x) failures.

$$\underbrace{(\pi)\,(\pi)\,(\pi)\ \ldots\ (\pi)}_{\text{x successes}} \cdot \underbrace{(1-\pi)\,(1-\pi)\,(1-\pi)\ \ldots\ (1-\pi)}_{\text{(n-x) failures}}$$

$$(\pi)^x \quad \cdot \quad (1-\pi)^{n-x}$$

Thus, P(Event) = $_nC_x * [(\pi)^x(1-\pi)^{n-x}]$

FORMULA 3

Therefore

$$P(Event) =\{n!/[x!(n-x)!]\}*\{(\pi)^x(1-\pi)^{n-x}\}$$

$$= \{4!/[3!(4-3)!]\}*\{(.70)^3(1-.70)^{4-3}\} = 0.4116$$

Diagram 2 is the sample space and diagram 3 is the <u>probability distribution</u> for random variable X in Question 6.

For the experiment in Question 6 we can define many events and determine their probabilities. For example:

$P(X = 2,3,\text{or}4) = 1/36 + 2/36 + 3/36 = 1/6$
$P(X \le 5) = 1/36 + 2/36 + 3/36 + 4/36 = 10/36$
$P(3 < X \le 5) = P(X = 4) + P(X = 5) = 3/36 + 4/36 = 7/36$

DEFINITION 1

If X is a random variable on the sample space $\{x_1,x_2,..,x_n\}$ and p_i is the probability $p_i=p\{X=x_i\}$ for $i =1,2,...,n$, then the *expected value* of X, E[X], is:

$E[X] = x_1p_1 + x_2p_2 + ...+ x_np_n$

For the experiment in Question 6 the expected value of X is:

$E[X] = 2(1/36) + 3(2/36) + 4(3/36) + 5(4/36) + 6(5/36) + 7(6/36) + 8(5/36) + 9(4/36) +10(3/36) + 11(2/36) + 12(1/36) = 7.$

DEFINITION 2
Variance

The variance of X, V[X], is the expected squared deviation of the values of X around their expected value E[X]:

$$V[X] = E[(X-E[X])^2] = (X_1-E[X])^2P_1 + (X_2-E[X])_2P_2 + ... + (X_n-E[X]))^2P_n.$$

For the experiment in Question 6 the variance of X is:

$V[X] = (2-7)2(1/36) + (3-7)2(2/36) + (4-7)2(3/36) + (5-7)2(5/36) + (6-7)2(5/36) + (10-7)2(3/36) + (11-7)2(2/36) +(12-7)2(1/36) = 35/6.$

DEFINITION 3

Standard Deviation

**The standard deviation,
Sigma, of a random variable X is:**

sigma = s = SQR(V[X])

For the experiment in Question 6 the standard deviation is:

s = SQR(V[X] = SQR(35/6) = 2.415.

In many situations the outcome of an experiment can be one of infinitely many real numbers.

A random variable X that can equal any number in an interval (its sample space) is a continuous random variable.

It is apparent we cannot assign individual probabilities to each of the numbers in such a sample space as we did in Definitions 1, 2 and 3 above.

Thus, for continuous probabilities, P(a < X > b), we define a *<u>probability density function</u>* as:

DEFINITION 4

Probability Density Function

P(a < X < b) = $\int_a^b f(x)\,dx$ **f(x) dx**

In diagram 3, there exist a graph, y=f(x), [see diagram 3(a) below] so that the total area (1.000) of the sum of the rectangles is beneath the curve:

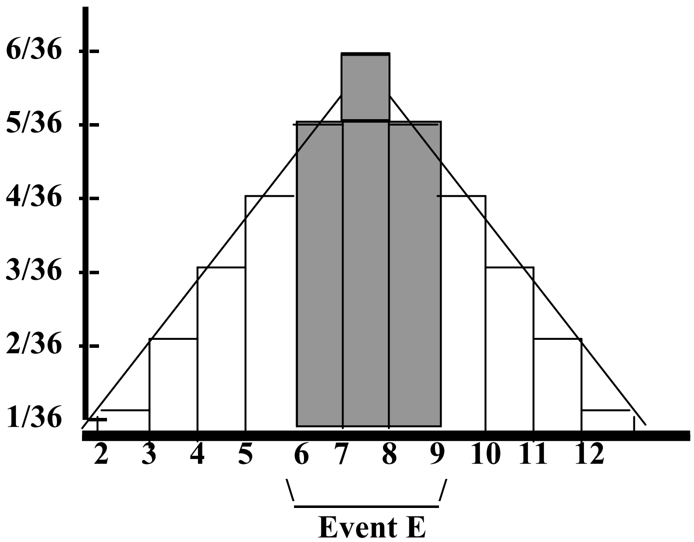


Diagram 3(a)

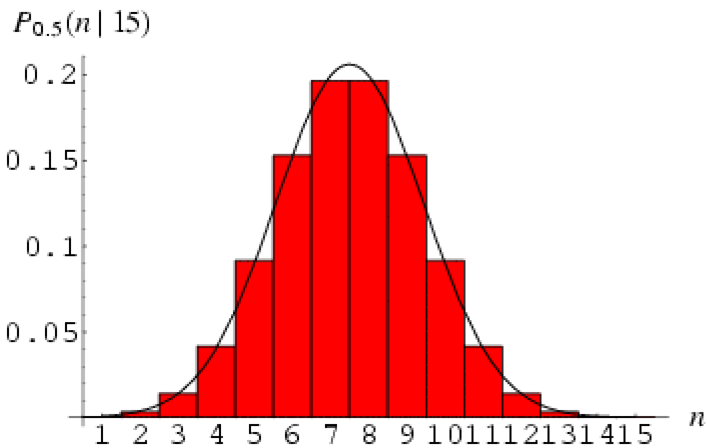


Diagram 3(b)

Notice that the area under the curve that forms a triangle with the x-axis is NOT equal to 1. Rather:

$$A=(1/2)bh = (11/2)(6/35) = 11/12.$$

However, the area under a bell shaped curve, y=f(x) does equal 1.

Now consider the Event E.

In the discrete case [sum of the areas of the three rectangles in diagram 3(a)] the probability of Event:

E = {6,7,8} is: $P(6 < X < 8) = P(6) + P(7) + P(8)$.

In the continuous case (bell-shaped area) the probability of Event E = [a,b] is:

$$P\{a < X < b\} = \int_a^b f(x)dx$$

So we see that a probability density function determines the probability of an event associated with a random variable.

When we specify the cumulative probability associated with all values of the random variable less than or equal to a given number we define a similar, but different, function called a distribution function, F(X):

$F(X) = P(X < x)$, for all x in the sample space for X.

Thus, we have the following relationship between the continuous probability density function, f, and the associated distribution function, F:

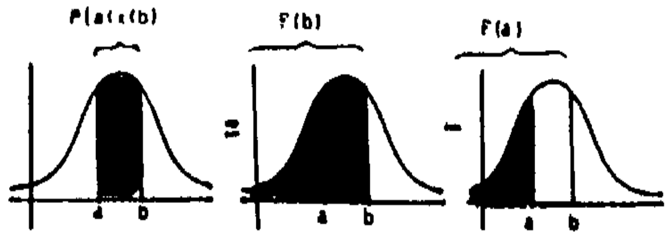
$$\begin{aligned}\int_a^b f(x)dx &= P(a \le x \le b)\\ &= P(x \le b) - P(x \le a)\\ &= F(b) - F(a).\end{aligned}$$

That is:

FORMULA 4
The Definite Integral
$\int_a^b f(x)dx = F(b) - F(a).$

The distribution function F gives the area $\int_a^b f(x)dx$ of the region bounded by the probability density function on an interval [a, b].

Formula 4 is illustrated in Diagram 4 on the next page:



$$P[a \le x \le b] = \int_a^b f(x)\,dx = F(b) - F(a)$$

Diagram 4

QUESTION 8

FIND THE NUMBER a SO THAT THE FUNCTION
F(X)=aX(X-4)
IS A PROBABILITY DENSITY
FUNCTION ON THE SAMPLE SPACE [0,4].

SOLUTION

Since the probability associated with an entire sample space is always one, the integral of f over the entire sample space must equal one. So we must find "A" such that:

$$\int_0^4 ax(x-4)\,dx = 1.$$

$$\int_0^4 ax(x-4)\,dx = a * o\,(x^2 - 4x)\,dx$$

$$= a \bullet \left(\frac{x^3}{3} - 2x^2 \right)_0^4$$

$$= a \bullet \left(\frac{64}{3} - 32 \right)$$

$$= a \bullet \left(-10\frac{2}{3} \right)$$

$$= -10\frac{2}{3} \bullet a = 1$$

thus, a = $1/(\frac{-32}{3}) = \frac{-3}{32}$

Therefore: $f(x) = (\frac{-3x}{32})(x-4) = (\frac{-3}{32})(x^2 - 4x)$

The probability density function:

$$(\frac{-3}{32})(x^2 - 4x)$$

is illustrated in diagram 5.

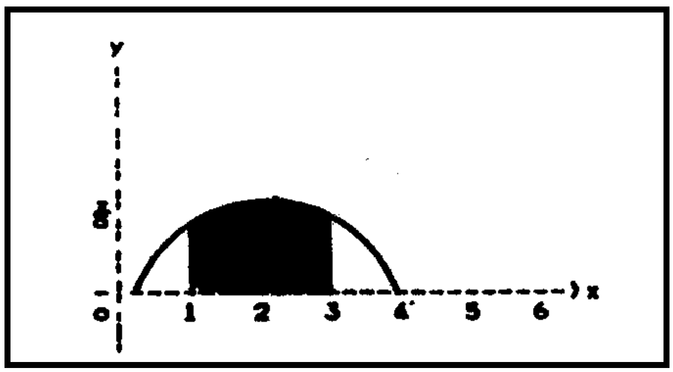


Diagram 5

QUESTION 9

FOR THE PROBABILITY DENSITY FUNCTION IN QUESTION 8 FIND P(1 < X < 3)

SOLUTION

According to Formula 4 we have:

$$P(1 < x < 3) = \int_1^3 \left(x^4 - 4x\right)dx = \frac{11}{16}$$

QUESTION 10

FIND THE EXPECTED VALUE OF THE RANDOM VARIABLE X WITH PROBABILITY DENSITY FUNCTION:
F(X) = (-3X/32)(X – 4)

SOLUTION

Let us extend Definition 1 to a continuous random variable X defined on a finite interval [a,b], with continuous probability density function f:

$$E(X) = \int_a^b x f(x)\,dx$$

$$E(X) = \int_0^4 x \frac{-3x}{32}(x-4)\,dx = 2.$$

Since this function is a parabola with axis of symmetry x=2, which is the midpoint of the sample space [0,4], we are not surprised that the expected value is 2.

Extending Definition 2 we find the variance associated with the continuous random variable X on sample space [a,b] to be:

$$V[X] = E([X-E(X)]^2)$$

$$= \int_a^b [X-E(X)]2\ f(x)\ dx$$

Now, $E(X) = 2$ **and** $f(x) = (-3x/32)(x-4)$

so, $\int_0^4 (x-2)^2 * (-3x/32)(x-4)\ dx = .8$

Thus, .8 is the expected value of the square of the deviation of X from its expected value.

The standard deviation for the above function would be:

$$\sigma = SQR(.8) = 0.894$$

The following example illustrates how the area under the graph of a continuous function can be used to calculate probabilities for various events.

EXAMPLE

A biologist determines that the random variable X giving the time at which a sleeping animal awakens during a 4 minute experiment has probability density function:

$$f(x)=\left(\frac{5}{4}\right)\left[\frac{1}{(1+x)}\right]^2, 0\leq x\leq 4.$$

Find:

a. the probability that the animal awakens during the first two minutes of the experiment.

b. the probability that the animal awakens after one minute but before the experiment ends.

SOLUTION:

a.) $\int_0^2 \frac{5}{4}\left[\frac{1}{(x+1)}\right]_0^2 dx$

$$= -\frac{5}{4}\,[1/(x+1)]\Big|_0^2 = 5/6 = 0.8333\ldots$$

- -

b.) $= -\frac{5}{4}\,[1/(x+1)]\Big|_1^4 = 3/4 = 0.75$

REFERENCES

E.S. Pearson and M.G. Kendall, Studies in the History of Statistics and Probability. Hafner Pub. Co., Conn., 1970.

H. Eves, An Introduction to the History of Mathematics. 4th edition. Holt, Rinehart, and Winston, New York, 1976.

I. Hacking, The Emergence of Probability. Cambridge Univ. Press London, 1975.

I. Tod hunter, A History of the Mathematical Theory of Probability. Chelsea Pub. Co., New York, 1949.

L. E. Maistrov, Probability Theory - A Historical Sketch. Academic Press, New York, 1974.

M. Kline, Mathematical Thought From Ancient to Modern Times. Oxford Univ. Press, New York, 1972.