

Probability Distributions

Random Variables

EXAMPLE: Consider rolling a fair die twice.

$$S = \{(i, j) : i, j \in \{1, \dots, 6\}\}$$

Suppose we are interested in computing the sum, i.e. we have placed a bet at a craps table. Let X be the sum. Then $X \in \{2, 3, \dots, 12\}$ is random as it depends upon the outcome of the experiment. It is a *random variable*. We can compute probabilities associated with X .

$$P(X = 2) = P\{(1, 1)\} = 1/36$$

$$P(X = 3) = P\{(1, 2), (2, 1)\} = 2/36$$

$$P(X = 4) = P\{(1, 3), (2, 2), (3, 1)\} = 3/36$$

Can write succinctly

Sum, i	2	3	4	5	6	...	12
$P(X = i)$	1/36	2/36	3/36	4/36	5/36	...	1/36

DEFINITION: Let S be a sample space. Then a function $X : S \rightarrow \mathbb{R}$ is a random variable.

EXAMPLE: Consider a bin with 5 white and 4 red chips. Let S be the sample space associated with selecting three chips, *with replacement*, from the bin. Let X be the number of white chips chosen. The sample space is

$$S = \{(a, b, c) : a, b, c \in \{W, R\}\}.$$

We have

$$X(R, R, R) = 0$$

$$X(W, R, R) = X(R, W, R) = X(R, R, W) = 1$$

$$X(W, W, R) = X(W, R, W) = X(R, W, W) = 2$$

$$X(W, W, W) = 3$$

We now find probabilities:

$$P(X = 0) = P(R, R, R) = \left(\frac{4}{9}\right)^3$$

$$P(X = 1) = P((W, R, R) \cup (R, W, R) \cup (R, R, W))$$

$$= P(W, R, R) + P(R, W, R) + P(R, R, W) = 3 \cdot \frac{5}{9} \cdot \left(\frac{4}{9}\right)^2$$

$$P(X = 2) = 3 \cdot \left(\frac{5}{9}\right)^2 \cdot \frac{4}{9}$$

$$P(X = 3) = \left(\frac{5}{9}\right)^3$$

X	0	1	2	3
$P(X = i)$	$\left(\frac{4}{9}\right)^3$	$3 \cdot \frac{5}{9} \cdot \left(\frac{4}{9}\right)^2$	$3 \cdot \left(\frac{5}{9}\right)^2 \cdot \frac{4}{9}$	$\left(\frac{5}{9}\right)^3$

Note that

$$\left(\frac{4}{9}\right)^3 + 3 \cdot \frac{5}{9} \cdot \left(\frac{4}{9}\right)^2 + 3 \cdot \left(\frac{5}{9}\right)^2 \cdot \frac{4}{9} + \left(\frac{5}{9}\right)^3 = 1$$

The two tables above illustrate what we mean by a **probability distribution**.

The Binomial Distribution

Consider n independent repeated trials of a random variable. Let the probability of a success be the same for each trial. Let X be the number of successes in the n trials. Then

$$f(k) = P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \text{ if } k \in \{0, 1, 2, \dots, n\}$$

EXAMPLE: If the probability is 0.70 that any one registered voter (randomly selected from official rolls) will vote in a given election, what is the probability that two of five registered voters will vote in the election?

Solution: We have

$$f(2) = \binom{5}{2} (0.70)^2 (1 - 0.70)^{5-2} = 10(0.70)^2 (0.30)^3 = 0.132$$

EXAMPLE: There is a product that is defective with a probability of 0.01. They are sold in packages of 10. The company offers money back if two or more are defective. What is the probability that a given package will be returned for cash?

Solution: Let X denote the number of defective products in a given package. We want to know $P\{X \geq 2\}$. Clearly, X is a binomial random variable. It is obviously easier to compute $P\{X \geq 2\} = 1 - P\{X \leq 1\}$. Thus, we need

$$P\{X = 0\} = \binom{10}{0} \cdot 0.01^0 \cdot 0.99^{10} = 0.90438, \quad P\{X = 1\} = \binom{10}{1} \cdot 0.01^1 \cdot 0.99^9 = 0.09135$$

Therefore, the desired probability is

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} \approx 0.0043$$

Thus, 0.4 percent of the packages are eligible for recall.

EXAMPLE: The probability is 0.30 that a person shopping at a certain supermarket will take advantage of its special promotion of ice cream. Find the probabilities that among six persons shopping at this market there will be 0, 1, 2, 3, 4, 5, or 6 who will take advantage of this promotion.

EXAMPLE: The probability is 0.30 that a person shopping at a certain supermarket will take advantage of its special promotion of ice cream. Find the probabilities that among six persons shopping at this market there will be 0, 1, 2, 3, 4, 5, or 6 who will take advantage of this promotion.

Solution: We have

$$f(0) = \binom{6}{0}(0.30)^0(0.70)^6 = 0.118$$

$$f(1) = \binom{6}{1}(0.30)^1(0.70)^5 = 0.303$$

$$f(2) = \binom{6}{2}(0.30)^2(0.70)^4 = 0.324$$

$$f(3) = \binom{6}{3}(0.30)^3(0.70)^3 = 0.185$$

$$f(4) = \binom{6}{4}(0.30)^4(0.70)^2 = 0.060$$

$$f(5) = \binom{6}{5}(0.30)^5(0.70)^1 = 0.010$$

$$f(6) = \binom{6}{6}(0.30)^6(0.70)^0 = 0.001$$

all rounded to three decimals.

Hypergeometric Distribution

Suppose that n objects are to be chosen from a set of a objects of one kind (successes) and b objects of another kind (failures), the selection is without replacement, and we are interested in the probability of getting x successes and $n-x$ failures. Then for sampling without replacement the probability of “ x successes in n trials” is

$$f(x) = \frac{\binom{a}{x} \cdot \binom{b}{n-x}}{\binom{a+b}{n}} \quad \text{for } x = 0, 1, 2, \dots, n$$

where x cannot exceed a and $n-x$ cannot exceed b . This is the formula for the **hypergeometric distribution**.

EXAMPLE: A mailroom clerk is supposed to send 6 of 15 packages to Europe by airmail, but he gets them all mixed up and randomly put airmail postage on 6 of the packages. What is the probability that only 3 of the packages that are supposed to get sent by air get airmail postage?

Solution: We have

$$f(3) = \frac{\binom{6}{3} \cdot \binom{9}{6-3}}{\binom{6+9}{6}} = \frac{20 \cdot 84}{5,005} \approx 0.336$$

EXAMPLE: Among an ambulance service's 16 ambulances, five emit excessive amounts of pollutants. If eight of the ambulances are randomly picked for inspection, what is the probability that this sample will include at least three of the ambulances that emit excessive amounts of pollutants?

Solution: We have

$$f(3) = \frac{\binom{5}{3} \cdot \binom{11}{8-3}}{\binom{5+11}{8}} = \frac{10 \cdot 462}{12,870} = 0.359$$

$$f(4) = \frac{\binom{5}{4} \cdot \binom{11}{8-4}}{\binom{5+11}{8}} = \frac{5 \cdot 330}{12,870} = 0.128$$

$$f(5) = \frac{\binom{5}{5} \cdot \binom{11}{8-5}}{\binom{5+11}{8}} = \frac{1 \cdot 165}{12,870} \approx 0.013$$

then

$$f(3) + f(4) + f(5) = 0.359 + 0.128 + 0.013 = 0.500$$

This result suggests that the inspection should, perhaps, have included more than eight of the ambulances.

The Poisson Distribution

When n is large and p is small, binomial probabilities are often approximated by means of the formula

$$f(x) = \frac{(np)^x \cdot e^{-np}}{x!} \quad \text{for } x = 0, 1, 2, 3, \dots$$

which is a special form of the **Poisson distribution**.

REMARK: It is safe to use the Poisson approximation to the binomial distribution only when

$$n \geq 100 \quad \text{and} \quad np \leq 10$$

EXAMPLE: A community of 100,000 people is struck with a fatal disease in which the chance of death is 0.0001. What is the probability that at most 8 people will die?

EXAMPLE: A community of 100,000 people is struck with a fatal disease in which the chance of death is 0.0001. What is the probability that at most 8 people will die?

Solution: This is binomial with parameters $n = 10^5$ and $p = 10^{-4}$. So, if X is the random variable giving number of deaths, then

$$\begin{aligned} P\{X \leq 8\} &= \sum_{k=0}^8 \binom{10^5}{k} (10^{-4})^k (1 - 10^{-4})^{10^5 - k} = \sum_{k=0}^8 \frac{10^5!}{k!(10^5 - k)!} (10^{-4})^k (1 - 10^{-4})^{10^5 - k} \\ &= \frac{10^5!}{k!(10^5 - 0)!} (10^{-4})^0 (1 - 10^{-4})^{10^5 - 0} + \frac{10^5!}{1!(10^5 - 1)!} (10^{-4})^1 (1 - 10^{-4})^{10^5 - 1} + \dots \\ &\quad + \frac{10^5!}{7!(10^5 - 7)!} (10^{-4})^7 (1 - 10^{-4})^{10^5 - 7} + \frac{10^5!}{8!(10^5 - 8)!} (10^{-4})^8 (1 - 10^{-4})^{10^5 - 8} \end{aligned}$$

Computationally this formula is too complicated. Another option is to use the Poisson approximation with $\lambda = np = 10$:

$$P\{X \leq 8\} \approx \sum_{k=0}^8 e^{-10} \frac{10^k}{k!} = e^{-10} \frac{10^0}{0!} + e^{-10} \frac{10^1}{1!} + e^{-10} \frac{10^2}{2!} + \dots + e^{-10} \frac{10^8}{8!} = 0.3328197$$

EXAMPLE: Disney World averages 5 injuries per day. What is the probability that there will be more than 6 injuries tomorrow?

Solution: It seems like there is not enough information to be able to answer this question. However, it is known that

$$\text{mean of a binomial distribution} = np$$

therefore $np = 5$. We also know that n is big, p is small. Therefore,

$$P\{X \geq 7\} = 1 - P\{X < 7\} \approx 1 - \sum_{k=0}^6 e^{-5} \frac{5^k}{k!} = 1 - \left(e^{-5} \frac{5^0}{0!} + e^{-5} \frac{5^1}{1!} + \dots + e^{-5} \frac{5^6}{6!} \right) = 0.2378$$

Note that it would be impossible to solve this using the binomial view because we don't know p or n .

EXAMPLE: Suppose that there are, on average, 1.2 typos on a page of a given book. What is the probability that there are more than two errors on a given page?

Solution: We model the number of errors on a given page, X , as a Poisson random variable with $np = 1.2$. Therefore

$$\begin{aligned} P\{X \geq 3\} &= 1 - P\{X \leq 2\} = 1 - \sum_{k=0}^2 e^{-1.2} \frac{1.2^k}{k!} \\ &= 1 - \left(e^{-1.2} \frac{1.2^0}{0!} + e^{-1.2} \frac{1.2^1}{1!} + e^{-1.2} \frac{1.2^2}{2!} \right) = 0.120512 \end{aligned}$$