1. Discrete Distributions


Basic Theory

As usual, we start with a random experiment with probability measure $\mathbb{P}$ on an underlying sample space Ω. A random variable X for the experiment that takes values in a countable set S is said to have a **discrete distribution**. Typically, $S \subseteq \mathbb{R}^n$ for some n, so in particular, if $n > 1$, X is vector-valued. In the picture below, the blue dots are intended to represent points of positive probability.
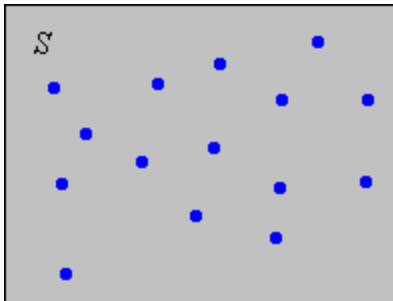

Discrete Density Functions

The (discrete) **probability density function** of X is the function f from S to $\mathbb{R}$ defined by

$$f(x) = \mathbb{P}(X = x), \quad x \in S$$

⊞ 1. Show that f satisfies the following properties. *Hint*: recall the axioms of a probability measure.

   a. $f(x) \geq 0, \ x \in S$
   b. $\sum_{x \in S} f(x) = 1$
   c. $\sum_{x \in A} f(x) = \mathbb{P}(X \in A), \ A \subseteq S$

Property (c) is particularly important since it shows that the probability distribution of a discrete random variable is completely determined by its probability density function. Conversely, any function that satisfies properties (a) and (b) is a (discrete) probability density function, and then property (c) can be used to construct a discrete probability distribution on S. Technically, f is the density of X relative to counting measure on S.
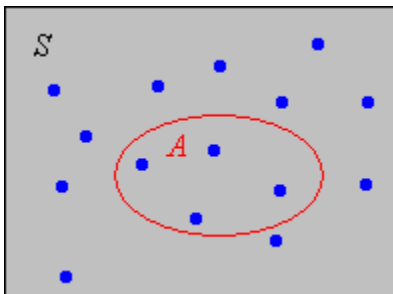
As noted before, S is typically a countable subset of some larger set, such as $\mathbb{R}^n$ for some n. We can always extend f, if we want, to the larger set by defining $f(x) = 0$ for $x \notin S$. Sometimes this extension simplifies formulas and notation.

An element $x \in S$ that maximizes the probability density function f is called a **mode** of the distribution. When there is only one mode, it is sometimes used as a measure of the *center* of the distribution.

Interpretation

A discrete probability distribution is equivalent to a **discrete mass distribution**, with total mass 1. In this analogy, S is the (countable) set of point masses, and $f(x)$ is the mass of the point at $x \in S$. Property (c) in Exercise 1 simply means that the mass of a set A can be found by adding the masses of the points in A.

For a probabilistic interpretation, suppose that we create a new, compound experiment by repeating the original experiment indefinitely. In the compound experiment, we have a sequence of independent random variables $(X_1, X_2, ...)$ each with the same distribution as X; in statistical terms, we are sampling from the distribution of X. Define

$$f_n(x) = \frac{1}{n} \#(\{i \in \{1, 2, ..., n\} : X_i = x\}), \quad x \in S$$

This is the relative frequency of x in the first n runs. Note that for each x, $f_n(x)$ is a random variable for the compound experiment. By the law of large numbers, $f_n(x)$ should converge to $f(x)$, in some sense, as $n \to \infty$. The function $f_n(x)$ is called the empirical density function; such functions are displayed in most of the simulation applets that deal with discrete variables.

Constructing Probability Density Functions

⊞ 2. Suppose that g is a nonnegative function defined on a countable set S. Let
$$c = \sum_{x \in S} g(x)$$

Show that if $0 < c < \infty$, then $f(x) = \frac{1}{c} g(x), \ x \in S$ defines a discrete probability density function on S.

Note that $c = 0$ if and only if $g(x) = 0$ for every $x \in S$. At the other extreme $c = \infty$ could only occur if S is infinite. When $0 < c < \infty$, so that we can construct the density function f, c is sometimes called the **normalizing constant**. This result is useful for constructing density functions with desired functional properties (domain, shape, symmetry, and so on).

Conditional Densities

The probability density function of a random variable X is based, of course, on the underlying probability measure $\mathbb{P}$ on the sample space Ω. This measure could be a conditional probability measure, conditioned on a given event $E \subseteq \Omega$ (with $\mathbb{P}(E) > 0$). The usual notation is

$$f(x|E) = \mathbb{P}(X = x|E), \quad x \in S$$

The following exercise shows that, except for notation, no new concepts are involved. Therefore, all results that hold for

discrete probability density functions in general have analogies for conditional discrete probability density functions.

3. Show that for fixed E, the function $x \mapsto f(x|E)$ is a discrete probability density function. That is, show that it satisfies properties (a) and (b) of Exercise 1, and show that property (c) becomes

$$\mathbb{P}(X \in A|E) = \sum_{x \in A} f(x|E), \quad A \subseteq S$$

4. Suppose that $B \subseteq S$ and $\mathbb{P}(X \in B) > 0$ Show that the conditional density of X given $X \in B$ is

$$f(x|X \in B) = \begin{cases} \dfrac{f(x)}{\mathbb{P}(X \in B)}, & x \in B \\ 0, & x \in B^c \end{cases}$$

Conditioning and Bayes' Theorem

Suppose that X is a random variable with a discrete distribution on a countable set S, and that $E \subseteq \Omega$ is an event in the experiment. Let f denote the probability density function of X.

5. Note that $\{\{X = x\} : x \in S\}$ is a partition of the sample space Ω.

Partition of the sample space

Because of this result, the versions of the conditioning rule and Bayes' theorem given in the following exercises follow immediately from the corresponding results in the section on Conditional Probability. Only the notation is different.

6. Show that:

$$\mathbb{P}(E) = \sum_{x \in S} f(x)\mathbb{P}(E|X = x)$$

This result is useful, naturally, when the distribution of X and the conditional probability of E given the values of X are known. We say that we are **conditioning** on X.

7. Prove **Bayes' Theorem**, named after Thomas Bayes:

$$f(x|E) = \frac{f(x)\mathbb{P}(E|X = x)}{\sum_{y \in S} f(y)\mathbb{P}(E|X = y)}, \quad x \in S$$

Bayes' theorem is a formula for the conditional probability density function of X given E. Again, it is useful, when the quantities on the right are known. In the context of Bayes' theorem, the (unconditional) distribution of X is referred to as the **prior** distribution and the conditional distribution as the **posterior** distribution. Note that the denominator in Bayes' formula is $\mathbb{P}(E)$ and is simply the normalizing constant.

Examples and Special Cases

Our first three models below--the discrete uniform distribution, hypergeometric distributions, and Bernoulli trials are very important. When working the computational problems that follow, try to see if the problem fits one of these models.

Discrete Uniform Distributions

8. An element X is chosen at random from a finite set S. The phrase *at random* means that all outcomes are equally likely.

   a. Show that X has probability density function $f(x) = \frac{1}{\#(S)}, \ x \in S$

   b. Show that $\mathbb{P}(X \in A) = \frac{\#(A)}{\#(S)}, \ A \subseteq S$

The distribution in the last exercise is called the **discrete uniform distribution** on S. Many random variables that arise in sampling or combinatorial experiments are transformations of uniformly distributed variables.

9. Suppose that n elements are chosen at random, with replacement from a set D with m elements. Let X denote the ordered sequence of elements chosen. Argue that X is uniformly distributed on the set $S = D^n$, and hence has probability density function

$$\mathbb{P}(X = x) = \frac{1}{m^n}, \quad x \in S$$

10. Suppose that n elements are chosen at random, without replacement from a set D with m elements. Let X denote the ordered sequence of elements chosen. Argue that X is uniformly distributed on the set S of permutations of size n chosen from D, and hence has probability density function

$$\mathbb{P}(X = x) = \frac{1}{m^{(n)}}, \quad x \in S$$

11. Suppose that n elements are chosen at random, without replacement, from a set D with m elements. Let W denote the unordered set of elements chosen. Show that W is uniformly distributed on the set T of combinations of size n chosen from D, and hence has probability density function

$$\mathbb{P}(W = w) = \frac{1}{\binom{m}{n}}, \quad w \in T$$

12. Suppose that X is uniformly distributed on a finite set S and that B is a nonempty subset of S. Show that the conditional distribution of X given $X \in B$ is uniform on B.

Hypergeometric Distributions

13. Suppose that a population consists of m objects; r of the objects are type 1 and $m - r$ are type 0. A sample of n objects is chosen at random (without replacement). Let Y denote the number of type 1 objects in the sample. Show that Y has probability density function.

$$f(k) = \frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}}, \quad k \in \{0, 1, \ldots, n\}$$

14. Suppose that a population consists of m objects; r of the objects are type 1, s are type 2, and $m - r - s$ are type 0. A sample of n objects is chosen at random (without replacement). Let Y denote the number of type 1 objects in the sample and Z the number of type 2 objects in the sample. Show that (Y, Z) has probability density function.

$$g(j, k) = \frac{\binom{r}{j}\binom{s}{k}\binom{m-r-s}{n-j-k}}{\binom{m}{n}}, \quad (j, k) \in \{0, 1, ..., n\}^2$$

The distribution defined by the density function in Exercise 13 is the **hypergeometric distributions** with parameters m, r, and n. The distribution defined by the density function in Exercise 14 is the **bivariate hypergeometric distribution** with parameters m, r, s, and n. Clearly, the same general pattern applies to populations with even more types. The hypergeometric distribution and the multivariate hypergeometric distribution are studied in detail in the chapter on Finite Sampling Models. This chapter contains a rich variety of distributions that are based on discrete uniform distributions.

Bernoulli Trials

A **Bernoulli trials sequence** is a sequence $(X_1, X_2, ...)$ of independent, identically distributed indicator variables. Random variable X_i is the outcome of trial i, and in the usual terminology of reliability, 1 denotes **success** while 0 denotes **failure**, The process is named for Jacob Bernoulli. Let $p = \mathbb{P}(X_i = 1)$ denote the success parameter of the process.

15. Show that $(X_1, X_2, ..., X_n)$ has probability density function

$$f(x_1, x_2, ..., x_n) = p^k (1-p)^{n-k} \text{ for } (x_1, x_2, ..., x_n) \in \{0, 1\}^n \text{ where } k = x_1 + x_2 + \cdots + x_n$$

16. Let Y denote the number of successes in the first n trials. Show that Y has probability density function

$$g(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, ..., n\}$$

The distribution defined by the probability density function in the last exercise is called the **binomial distribution** with parameters n and p. The binomial distribution is studied in detail in the chapter on Bernoulli Trials.

17. Consider again a sequence of Bernoulli trials with success parameter p. Let U denote the trial number of the first success and V the number of failures before the first success. Show that

  a. $\mathbb{P}(U = n) = (1-p)^{n-1} p$ for $n \in \mathbb{N}_+$.
  b. $\mathbb{P}(V = n) = (1-p)^n p$ for $n \in \mathbb{N}$.
  c. $U = V + 1$

The distribution defined by the density function in part (a) is the **geometric distribution** on $\mathbb{N}_+$. The distribution defined by the density function in part (b) is the **geometric distribution** on $\mathbb{N}$. In both cases, p is the parameter of the distribution. The geometric distribution is studied in detail in the chapter on Bernoulli Trials.

Balls and Urns

18. An urn contains 30 red and 20 green balls. A sample of 5 balls is selected at random, without replacement. Let Y denote the number of red balls in the sample.

    a. Compute the probability density function of Y.

    b. Graph the density function and identify the mode(s).

    c. Find $\mathbb{P}(Y > 3)$.

19. In the ball and urn experiment, select sampling without replacement and set $m = 50$, $r = 30$, and $n = 5$. Run the experiment 1000 times, updating every 10 runs, and note the apparent convergence of the empirical density function of Y to the theoretical density function.

20. An urn contains 30 red and 20 green balls. A sample of 5 balls is selected at random, *with* replacement. Let Y denote the number of red balls in the sample.

    a. Compute the density function of Y explicitly.

    b. Graph the density function and identify the mode(s).

    c. Find $\mathbb{P}(Y > 3)$.

21. In the ball and urn experiment, select sampling with replacement and set $m = 50$, $r = 30$, and $n = 5$. Run the experiment 1000 times, updating every 10 runs, and note the apparent convergence of the empirical density function of Y to the theoretical density function.

Coins and Dice

22. Suppose that a coin with probability of heads $p = 0.4$ is tossed 5 times. Let Y denote the number of heads.

    a. Compute the density function of Y explicitly.

    b. Graph the density function and identify the mode.

    c. Find $\mathbb{P}(Y > 3)$

23. In the coin experiment, set $n = 5$ and $p = 0.4$. Run the experiment 1000 times, updating every 10 runs, and note the apparent convergence of the empirical density function of X to the probability density function.

24. Suppose that two fair, standard dice are tossed and the sequence of scores (X_1, X_2) recorded. Let $Y = X_1 + X_2$, denote the sum of the scores, $U = \min\{X_1, X_2\}$, the minimum score, and $V = \max\{X_1, X_2\}$ the maximum score.

    a. Find the probability density function of (X_1, X_2)

    b. Find the probability density function of Y.

    c. Find the probability density function of U.

    d. Find the probability density function of V.

    e. Find the probability density function of (U, V).

    f. Find the probability density function U given $Y = 8$.

25. In the dice experiment, select $n = 2$ fair dice. Select the following random variables and note the shape and location of the density function. Run the experiment 1000 times, updating every 10 runs. For each of the following variables, note the apparent convergence of the empirical density function to the density function.

    a. Y, the sum of the scores.
    b. U, the minimum score.
    c. V, the maximum score.

26. In the die-coin experiment, a fair, standard die is rolled and then a fair coin is tossed the number of times showing on the die. Let N denote the die score, X the sequence of coin scores, and Y the number of heads. Note that the values of X are sequences of varying lengths.

    a. Find the probability density function of N.
    b. Find the probability density function of X.
    c. Find the probability density function of Y.
    d. Find the conditional density function of N given $Y = 2$.

27. Run the die-coin experiment 1000 times, updating after each run. For the number of heads, note the apparent convergence of the empirical density function to the probability density function.

28. Suppose that a bag contains 12 coins: 5 are fair, 4 are biased with probability of heads $\frac{1}{3}$; and 3 are two-headed. A coin is chosen at random from the bag and tossed 2 times. Let V denote the probability of heads of the selected coin and let Y denote the number of heads.

    a. Find the probability density function of V.
    b. Find the probability density function of Y.
    c. Find the conditional probability density function of V given $Y = 2$.

Compare Exercise 26 and Exercise 28. In the first exercise, we toss a coin with a *fixed* probability of heads a *random* number of times. In second exercise, we effectively toss a coin with a *random* probability of heads a *fixed* number of times.

29. In the coin-die experiment, a fair coin is tossed. If the coin lands tails, a fair die is rolled. If the coin lands heads, an ace-six flat die is tossed (faces 1 and 6 have probability $\frac{1}{4}$ each, while faces 2, 3, 4, 5 have probability $\frac{1}{8}$ each). Find the density function of the die score.

30. Run the coin-die experiment 1000 times, updating every 10 runs. Compare the empirical density of the die score with the theoretical density in the last exercise.

31. A coin with probability of heads $p \in (0, 1)$ is tossed until heads occurs. Let N denote the number of tosses.

a. Find the probability density function of N.

b. Compute $\mathbb{P}(N > n)$

c. Compute $\mathbb{P}(N$ is even$)$

d. Find the conditional probability density function of N given than N is even.

32. In the negative binomial experiment, set $k = 1$ and $p = 0.2$. Run the experiment 1000 times, updating every 10 runs. Note the apparent convergence of the empirical density function of the number of trials to the probability density function.

Cards

33. Recall that a **poker hand** consists of 5 cards chosen at random and without replacement from a standard deck of 52 cards. Let X denote the number of spades in the hand and Y the number of hearts in the hand.

a. Find the probability density function of X

b. Find the probability density function of Y

c. Find the probability density function of (X, Y)

34. Recall that a **bridge hand** consists of 13 cards chosen at random and without replacement from a standard deck of 52 cards. An **honor card** is a card of denomination ace, king, queen, or jack. In the most common **point counting system**, an ace is worth 4 points, a king is worth 3 points, a queen is worth 2 points, and a jack is worth 1 point. Let N denote the number of honor cards in the hand, and V the point value of the hand.

a. Find the probability density function of N

b. Find the probability density function of Y

Reliability

35. Suppose that in a batch of 500 components, 20 are defective and the rest are good. A sample of 10 components is selected at random and tested. Let X denote the number of defectives in the sample.

a. Find the probability density function of X.

b. Find the probability that the sample contains at least one defective component.

36. A plant has 3 assembly lines that produce a certain type of component. Line 1 produces 50% of the components and has a defective rate of 4%; line 2 has produces 30% of the components and has a defective rate of 5%; line 3 produces 20% of the components and has a defective rate of 1%. A component is chosen at random from the plant and tested.

a. Find the probability that the component is defective.

b. Given that the component is defective, find the conditional probability density function of the line that produced the component.

Recall that in the standard model of **structural reliability**, a systems consists of n components, each of which, independently of the others, is either working for failed. Let X_i denote the state of component i, where 1 means working and 0 means failed. Thus, the state vector is $X = (X_1, X_2, ..., X_n)$. The system as a whole is also either working or failed, depending only on the states of the components. Thus, the state of the system is an indicator random variable $Y = Y(X)$ that depends on the states of the components according to a **structure function**.

The **reliability** of a device is the probability that it is working. Let $p_i = \mathbb{P}(X_i = 1)$ denote the reliability of component i, so that $p = (p_1, p_2, ..., p_n)$ is the vector of component reliabilities. Because of the independence assumption, the system reliability depends only on the component reliabilities, according to a **reliability function** $r(p) = \mathbb{P}(Y = 1)$. Note that when all component reliabilities have the same value p, the states of the components form a sequence of n Bernoulli trials. In this case, the system reliability is, of course, a function of the common component reliability p.

37. Suppose that the component reliabilities all have the same value p. Find the probability density function of the state vector X.

38. Suppose again that the component reliabilities all have the same value p . Let Y denote the number of working components.

   a. Find the probability density function of Y.
   b. Find the reliability of the k **out of** n **system**. This system is working if and only if at least k of the n components are working.

The Poisson Distribution

39. Let $f(n) = e^{-a} \dfrac{a^n}{n!}$, $n \in \mathbb{N}$ where $a > 0$ is a parameter.

   a. Show that f is a probability density function for each $a > 0$
   b. Show that $f(n) > f(n-1)$ if and only if $n < a$.
   c. Show that the distribution has a single mode at $\lfloor a \rfloor$ if a is not an integer, and has two modes at $a - 1$ and a if a is an integer.

The distribution defined by the density in the previous exercise is the **Poisson distribution** with parameter a, named after Simeon Poisson. The Poisson distribution is studied in detail in the Chapter on Poisson Processes, and is used to model the number of "random points" in a region of time or space. The parameter a is proportional to the size of the region of time or space.

40. Suppose that the number of misprints N on a web page has the Poisson distribution with parameter 2.5.

    a. Find the mode.

    b. Find $\mathbb{P}(N > 4)$.

41. In the Poisson process, set $r = 1$ and $t = 2.5$. Run the simulation 1000 times updating every 10 runs. Note the apparent convergence of the empirical density function to the probability density function.

A Zeta Distribution

42. Let $g(n) = \dfrac{1}{n^2}$ for $n \in \mathbb{N}_+$.

    a. Find the probability density function f that is proportional to g. *Hint*: $\sum_{n=1}^{\infty} \dfrac{1}{n^2} = \dfrac{\pi^2}{6}$.

    b. Find the mode of the distribution.

    c. Find $\mathbb{P}(N \leq 5)$ where N has probability density function f.

The distribution defined in the previous exercise is a member of the **zeta** family of distributions. Zeta distributions are used to model sizes or ranks of certain types of objects, and are studied in more detail in the chapter on Special Distributions.

Benford's Law

43. Let $f(d) = \log(d + 1) - \log(d) = \log\left(1 + \dfrac{1}{d}\right)$ for $d \in \{1, 2, ..., 9\}$. (The logarithm function is the base 10 **common logarithm**, not the base e **natural logarithm**.)

    a. Show that f is a probability density function.

    b. Compute the values of f explicitly, and sketch the graph.

    c. Find $\mathbb{P}(X \leq 3)$ where X has probability density function f.

The distribution defined in the previous exercise is known as **Benford's law**, and is named for the American physicist and engineer Frank Benford. This distribution governs the leading digit in many real sets of data. Benford's law is studied in more detail in the chapter on Special Distributions.

Miscellaneous Problems

44. Let $f(n) = \dfrac{1}{n} - \dfrac{1}{n+1}$, $n \in \{1, 2, ...\}$.

    a. Show that f is a probability density function.

    b. Find $\mathbb{P}(3 \leq N \leq 7)$, where N is a random variable with density function f.

45. Let $g(n) = n(10 - n)$ for $n \in \{1, 2, ..., 9\}$.

    a. Find the probability density function f that is proportional to g.

    b. Find the mode of the distribution.

    c. Find $\mathbb{P}(3 \leq N \leq 6)$ where N has probability density function f.

46. Let $g(n) = n^2(10 - n)$ for $n \in \{1, 2, ..., 9\}$.

    a. Find the probability density function f that is proportional to g.

    b. Find the mode of the distribution.

    c. Find $\mathbb{P}(3 \leq N \leq 6)$ where N has probability density function f.

47. Let $g(x, y) = x + y$, $(x, y) \in \{1, 2, 3\}^2$.

    a. Sketch the domain of g.

    b. Find the probability density function f that is proportional to g.

    c. Find the mode of the distribution.

    d. Find $\mathbb{P}(X > Y)$ where (X, Y) is a random vector with the probability density function f.

48. Let $g(x, y) = xy$ for $(x, y) \in \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$.

    a. Sketch the domain of g.

    b. Find the probability density function f that is proportional to g.

    c. Find the mode of the distribution.

    d. Find $\mathbb{P}((X, Y) \in \{(1, 2), (1, 3), (2, 2), (2, 3)\})$ where (X, Y) is a random vector with the probability density function f.

Data Analysis Exercises

49. In the M&M data, let R denote the number of red candies and N the total number of candies. Compute and graph the empirical density of each of the following:

    a. R

    b. N

    c. R given $N > 57$

50. In the Cicada data, let G denotes gender, S denotes species type, and W denotes body weight (in grams).

Compute the empirical density of each of the following:

a. G

b. S

c. (G, S)

d. G given $W > 0.20$ grams.