4. Distribution Functions

In real-world problems, we are frequently faced with one or more quantities that do not have fixed values. For example, *the number of students who take Calculus but failed for the first time; the light bulbs produced by a certain factory that could survive for more than 8000 hours.*

- Definition: Let S be the sample space of an experiment. A real-valued function X: S→ R is called a random variable (r.v.) of the experiment if, for each interval $I \subseteq R$, {s: X(s)∈ I} is an event. In probability theory, the set {s: X(s)∈ I} is often abbreviated as $\{X \in I\}$, or simply as X∈ I.
- Definition: If X is a random variable, then the function F defined on $(-\infty, +\infty)$ by $F(x) = P(X \leq x)$ is called the *distribution function* of X or *cumulative distribution function (cdf).*

Example

- In rolling two fair dice, the sample (outcome) space S can be denoted as
- S= $\{(x,y) | 1 \leq x, y \leq 6\}$
- Let X be a random variable defined as the sum of face values, X: $S \rightarrow I$ with
  X$\left((x,y)\right) = x + y$, X can only assume the values 2, 3, 4, $\cdots$, 12 with probabilities
- P(X=2)=P({(1,1)})=$\frac{1}{36}$, P(X=3)=P({1,2),(2,1)})=$\frac{2}{36}$, P(X=4)=P({(1,3),(2,2),(3,1)})=$\frac{3}{36}$,
- P(X=5)=P({(1,4),(2,3),(3,2),(4,1)})=$\frac{4}{36}$, P(X=6)=P({(1,5),(2,4),(3,3),(4,2),(5,1)})=$\frac{5}{36}$,
- P(X=7)=P({(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)})=$\frac{6}{36}$,
- P(X=8)=P({(2,6),(3,5),(4,4),(5,3),(6,2)})=$\frac{5}{36}$, P(X=9)=P({(3,6),(4,5),(5,4),(6,3)})=$\frac{4}{36}$,
- P(X=10)=P({(4,6),(5,5),(6,4)})=$\frac{3}{36}$, P(X=11)=P({(5,6),(6,5)})=$\frac{2}{36}$, P(X=12)=P({(6,6)})=$\frac{1}{36}$.
- Then, P(X$\leq$ 6) $= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{15}{36}$

Example

- Suppose that three cards are drawn from an ordinary deck of 52 cards, one by one, *at random and with replacement*. Let X be the number of spades drawn; then X is a random variable. If an outcome of spades is denoted by s, and other outcomes are represented by t, then X is a real-valued function defined on the sample space S,

S={(s,s,s), (t,s,s), (s,t,s), (s,s,t), (s,t,t), (t,s,t), (t,t,s), (t,t,t)}

By X(s,s,s)=3, X(s,t,s)=2, X(s,s,t)=2, X(s,t,t)=1, X(t,t,t)=0, and so on.

P(X=0)=P({(t,t,t)})=$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{27}{64}$,

P(X=1)=P({(s,t,t), (t,s,t), (t,t,s)})=3$\times$ ($\frac{1}{4} \times \frac{3}{4} \times \frac{3}{4}$)=$\frac{27}{64}$,

P(X=2)=P({(s,s,t), (s,t,s), (t,s,s)})=$\frac{9}{64}$,   P(X=3)=P({(s,s,s)})=$\frac{1}{64}$

Example

- Suppose that three cards are drawn from an ordinary deck of 52 cards, one by one, *at random and without replacement*. Let X be the number of ***spades*** drawn; then X is a random variable.

- P(X=0)= $\frac{\binom{39}{3}}{\binom{52}{3}}$, P(X=1)=$\frac{\binom{13}{1}\binom{39}{2}}{\binom{52}{3}}$, P(X=2)=$\frac{\binom{13}{2}\binom{39}{1}}{\binom{52}{3}}$, P(X=3)=$\frac{\binom{13}{3}\binom{39}{0}}{\binom{52}{3}}$.

- P(X=i)=$\frac{\binom{13}{i}\binom{39}{3-i}}{\binom{52}{3}}$, $i = 0, 1, 2, 3.$

Example

- Example In U.S.A., the number of twin births is approximately 1 in 90. Let X be the number of births in a certain hospital until the first twins are born. X is a random variable. Denote twin births by T and single births by N. Then X is a real-values function defined on the sample space.
- S={*T, NT, NNT, NNNT,* $\cdots\cdots\}$ $by\ X(N^{i-1}\mathrm{T})$=$i,\ i=1,2,3,\cdots$
- The set of all possible values of X is {1, 2, 3, $\cdots\}$ and
- P($X=i$) $=\left(\frac{89}{90}\right)^{i-1}\left(\frac{1}{90}\right)$. (called a *Geometric Distribution*)

Example

- Let X:S→ R, Y:S→ R be random variables over the same sample space S. Then we can define new random variables based on X and Y as functions of variables, such as sinX, cosY, X^2+Y^2, and etc.
- Example The diameter of a flat metal disk manufactured by a factory is a random number between 4 inches and 4.5 inches. What is the probability that the area of such a flat disk chosen at random is at least 4.41π inches?
- Ans: Let D be the diameter of the metal disk (in inches) selected at random. Then $D \in (4.0, 4.5)$, and the area of the metal disk is $\pi\left(\frac{D}{2}\right)^2$ inches^2, P($\frac{\pi D^2}{4}$>4.41π)=P(D^2>17.64)=P(D>4.2)=$\frac{4.5-4.2}{4.5-4.0}$=$\frac{3}{5}$

Example

- Example A random number is selected from the interval $(0, \frac{\pi}{2})$, what is the probability that its sine value is greater cosine value?
- Ans: P(sinX > cosX)=P(tanX > 1) = P(X > $\frac{\pi}{4}$) = $\frac{\frac{\pi}{2}-\frac{\pi}{4}}{\frac{\pi}{2}-0} = \frac{1}{2}$
- Example Let X be a random variable with E(X)=3 and E[(X-3)(4-X)]=-15. Find Var(-3X+8).
- Ans: Since E[(X-3)(4-X)]=E[-X^2+7X-12]=-15, thus, E(X^2)=24, then

  Var(-3X+8)=$(-3)^2$Var(X)=9(E(X^2)-E^2(X))=9(24-9)=135.

Example

- E01. Let S be the sample space of results for taking a driver's licence test, that is, S={success, failure}. X can be defined as X: $\mathrm{S} \to \mathrm{R}$, with X(success)=1, X(failure)=0.
- E02. Let S be the sample space of results by rolling two fair dice, then S=$\{(x, y) | 1 \leq x, y, \leq 6\}$, and define X(s=$(x, y)$)=$x + y$. Then,

- P(X=t)=$\frac{6-|7-t|}{36}$, t=2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.
- F(X$\leq 5) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36}$ = $\frac{10}{36}$

4.2 Distribution Functions

- Random variables are often used for the calculation of the probabilities of events. For example, in the experiment of rolling two dice, if we are interested in a sum of at least 8, we define X to be the sum and calculate P(X≥8); we can also calculate the probability of the event that the sum is not greater than 7 by $P(X \leq 7)$.
- Definition: If X is a r.v., then the function F defined on $(-\infty, +\infty)$ by $F(t)=P(X \leq t)$ is called the (*cumulative*) distribution function of *X*.
- A distribution function has the following properties.
- (1) F is nondecreasing: If $t<u$, $F(t) \leq F(u)$.
- (2) $\lim_{t \to -\infty} F(t)=0$, and $\lim_{t \to \infty} F(t)=1$.
- (3) F is right continuous, that is, F(t+)=F(t).

Exercises

- 8 (p.155) Let X be a random variable with distribution function F. For p (0<p<1), Q_p is said to be a *quantile* of order p or *p-quantile* if
- $F(Q_p-) \le p \le F(Q_p)$. In a certain country, the rate at which the price of oil per gallon changes from one year to another has the distribution function:
- $F(x) = \frac{1}{1+e^{-x}}, \ -\infty < x < \infty.$

The first quartile $Q_{0.25}$=-ln3 by solving $F(Q_{0.25}) = 0.25$, the median $Q_{0.50} = 0$ by solving $F(Q_{0.50}) = 0.50$, and the 3rd quartile $Q_{0.75} = \ln 3 by\ solving\ F(Q_{0.75}) = 0.75$ of F.

Exercises on P.157

- B.17 Let X be a randomly selected point from (0,3). What is the probability that $X^2 - 5\mathrm{X} + 6 = (\mathrm{X}-2)(\mathrm{X}-3) > 0$?

  Ans: P(X>3 or X<2)=2/3.

- B.18 Let X be a random point selected from the interval (0,1). Calculate F, the distribution function Y=$\frac{X}{1+X}$, and sketch its graph.

  Ans: F(y)=P(Y$\leq y$) =P($\frac{X}{1+X} \leq y) = P(X \leq \frac{y}{1-y}$)=$\frac{y}{1-y}$ for $\frac{1}{2} < y < 1$.

- Quiz 2. Suppose the distribution function of a random variable X is given by
- F(t)=0 if t<1 and $1 - \frac{1}{t^2}$ if t≥1, then
- P(X>3)=1-F(3)=1-($1 - \frac{1}{3^2}) = \frac{1}{9}$ and P(X>5|X>3)=$\frac{1-F(5)}{1-F(3)}$=$\frac{1/25}{1/9}$=$\frac{9}{25}$