4 Continuous Random Variables and Probability Distributions

Part I

1

4.1 Probability Density Functions

2

Probability Density Functions

- Discrete Random Variable (rv) is one whose possible values either constitute finite set or else can be listed in infinite sequence (a list in which there is a first element, a second element, etc.).

- Random variable whose set of possible values is entire interval of numbers is not discrete.

Recall from Chapter 3 that a random variable X is continuous if

(1) possible values comprise either
 - single interval on the number line
 (for some $A < B$, any number x between A and B is possible value) or
 - a union of disjoint intervals, and

(2) $P(X = c) = 0$ for any number c that is a possible value of X.

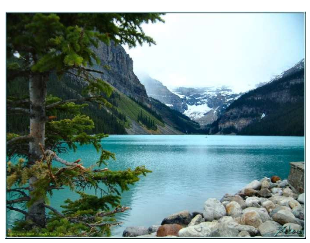
3

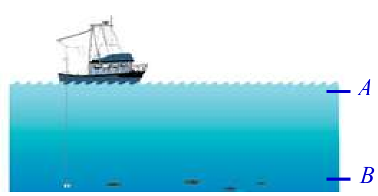
3

Example 4.1

ระบบนิเวศของทะเลสาป

- If in study of ecology of a lake, we make depth measurements at randomly chosen locations, then





X = the depth at such a location is continuous random variable

- Here A is the minimum depth in the region being sampled, and B is the maximum depth.

4

4

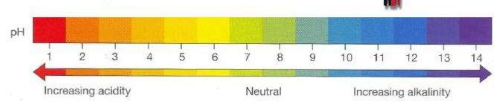
Example 4.2

- If a chemical compound is randomly selected and its pH X is determined, then X is a continuous random variable because and pH value between 0 and 14 is possible

- If more is know about the compound selected for analysis, then set of possible values might be a subinterval of [0, 14], such as $5.5 \le x \le 6.5$, but X would still be continuous







5

5

Example 4.3

- Let X represent the amount of time a randomly selected customer spends waiting for haircut before his/her haircut commences (เริ่ม/เริ่มต้น)
- Your first thought might be that X is continuous random variable, since measurement is required to determine its value

Discrete

- However, there are customers lucky enough to have no wait whatsoever before climbing into barber's chair
- So it must be the case that $P(X=0) > 0$

Continuous

- Conditional on no chairs being empty, though, waiting time will be continuous since X could then assume any value between some minimum possible time A and maximum possible time B

- This random variable is neither purely discrete nor purely continuous but instead is a mixture of the two types

6

6

Probability Density Functions

- One might argue that although in principle variables such as height, weight, and temperature are continuous, in practice the limitations of our measuring instruments restrict us to a discrete world.

- However, continuous models often approximate real-world situations very well, and continuous mathematics (calculus) is frequently easier to work with than mathematics of discrete variables and distributions.

7

7

Probability Distributions for Continuous Variables

8

8

Probability Distributions for Continuous Variables

- Suppose variable X of interest is depth of a lake at a randomly chosen point on the surface.
- Let M = Maximum depth (in meters), so that any number in interval $[0, M]$ is possible value of X.
- If we "discretize" X by measuring depth to the nearest meter, then possible values are nonnegative integers less than or equal to M (possible values $\leq M$)
- Resulting discrete distribution of depth can be pictured using probability histogram.

9

9

Probability Distributions for Continuous Variables

- If we draw histogram so that area of rectangle above any possible integer k is proportion of lake whose depth is (to the nearest meter) k, then the total area of all rectangles is 1.
- A possible histogram appears in Figure 4.1(a).

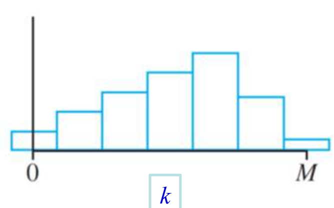


Figure 4.1 (a) Probability histogram of depth measured to the nearest meter

10

10

Probability Distributions for Continuous Variables

- If depth is measured much more accurately and the same measurement axis as in Figure 4.1(a) is used,
- each rectangle in the resulting probability histogram is much narrower,
- Total area of all rectangles is still 1.
- A possible histogram is pictured in Figure 4.1(b).

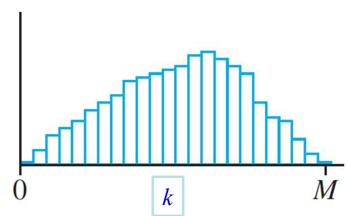


Figure 4.1(b) Probability histogram of depth measured to the nearest centimeter

11

11

Probability Distributions for Continuous Variables

- It has much smoother appearance than histogram in Figure 4.1(a).
- If we continue in this way to measure depth more and more finely,
- Resulting sequence of histograms approaches a smooth curve,
- such as is pictured in Figure 4.1(c).

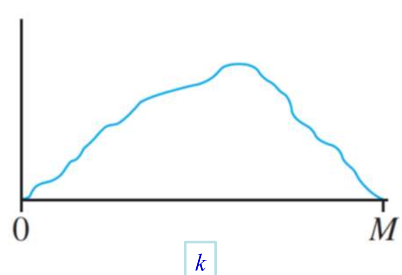
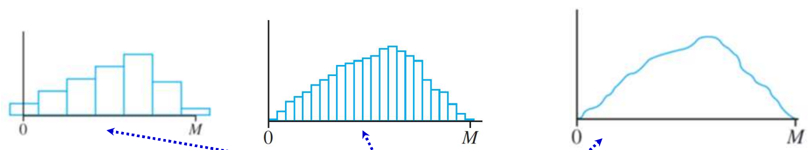


Figure 4.1(c) A limit of a sequence of discrete histograms

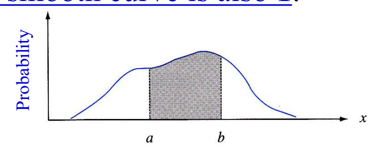
12

12

Probability Distributions for Continuous Variables



○ Because for each histogram total area of all rectangles equals **1**, total area under smooth curve is also **1**.



○ Probability that depth at randomly chosen point is between *a* and *b* is just area under the smooth curve between *a* and *b*.

○ It is exactly smooth curve of the type pictured in Figure 4.1(c) that specifies a **continuous probability distribution**.

13

13

Probability Distributions for Continuous Variables

Definition

Let X be a continuous random variable.

Then **Probability Distribution** or **Probability Density Function** (pdf) of X is a function $f(x)$ such that for any two numbers a and b with $a \le b$,

$$P(a \le X \le b) = \int_a^b f(x)dx$$

density function

density curve

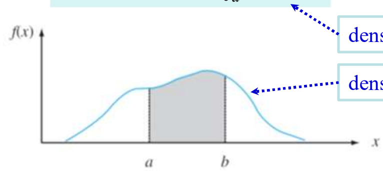


Figure 4.2 $P(a \le X \le b)$ = the area under the density curve between a and b

That is, probability that X takes on value in interval $[a, b]$ is area above this interval and under the graph of **density function**.

The graph of $f(x)$ is often referred to as ***density curve***.

14

14

Probability Distributions for Continuous Variables

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

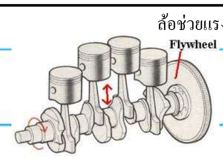
1. $f(x) \geq 0 \quad for\ all\ x$

2. $\int_{-\infty}^{\infty} f(x)dx = area\ under\ the\ entire\ graph\ of\ f(x)$
   $= 1$

15

15

Example 4.4



- Direction of an imperfection with respect to a reference line on circular object such as tire, brake rotor, or flywheel is, in general, subject to uncertainty.
- Consider reference line connecting the valve stem on a tire to center point, and
- let X be angle measured clockwise to location of imperfection.
- One possible pdf for X is

$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq x < 360 \\ 0 & otherwies \end{cases}$$

16

16

Example 4.4

cont'd

Pdf is graphed in Figure 4.3.

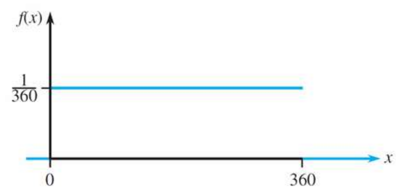


Figure 4.3 The pdf and probability from Example 4.4

Clearly $f(x) \geq 0$.

Area under the density curve is just area of a rectangle:

$$(height)(base) = \left(\frac{1}{360}\right) \cdot 360 = 1$$

17

17

Example 4.4

cont'd

- Probability that angle is between 90° and 180° is

Solution

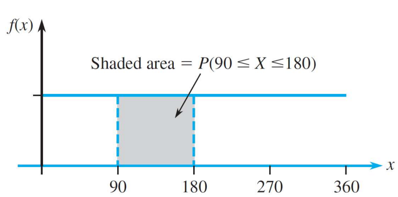


Figure 4.3 The pdf and probability from Example 4.4

19

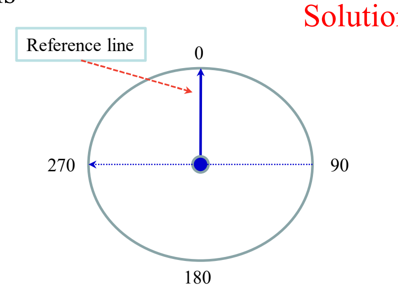
19

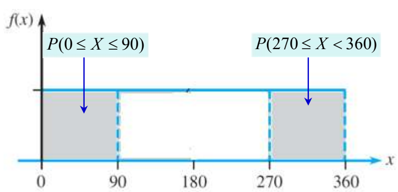
Example 4.4

cont'd

- Probability that angle of occurrence is within 90° of reference line is

Solution



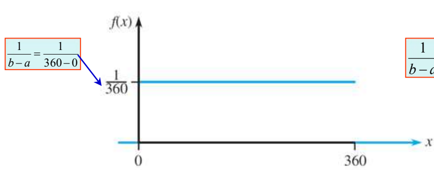


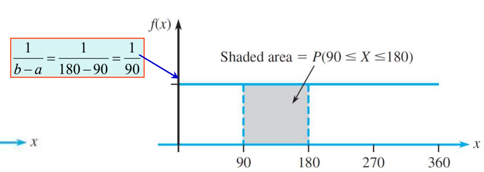
21

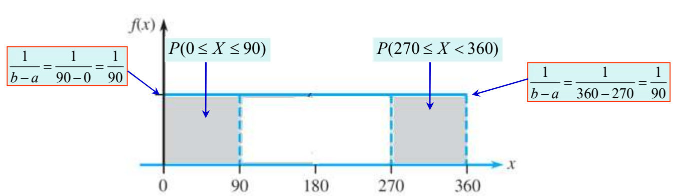
21

Probability Distributions for Continuous Variables

- Because whenever $0 \le a \le b \le 360$ in Example 4.4 and $P(a \le X \le b)$ depends only on width $b - a$ of interval, X is said to have **Uniform Distribution**.







22

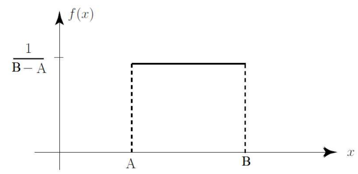
22

Probability Distributions for Continuous Variables

Definition

Continuous random variable X is said to have **Uniform Distribution** on interval $[A, B]$ if pdf of X is

$$f(x; A, B) = \begin{cases} \dfrac{1}{B-A} & A \le x \le B \\ 0 & otherwise \end{cases}$$



23

23

Probability Distributions for Continuous Variables

- When X is a discrete random variable, each possible value is assigned positive probability.

- This is not true of continuous random variable
  (that is, the second condition of the definition is satisfied)
  because area under density curve that lies above any single value is zero:

$$P(X = c) = \int_c^c f(x)dx = \lim_{\varepsilon \to 0} \int_{c-\varepsilon}^{c+\varepsilon} f(x)dx = 0$$

25

25

Probability Distributions for Continuous Variables

- The fact that $P(X = c) = 0$ when X is continuous has important practical consequence:
- Probability that X lies in some interval between a and b does not depend on whether lower limit a or upper limit b is included in probability calculation:

$$P(a \le X \le b) = P(a < X < b) = P(a < X \le b) = P(a \le X < b) \qquad (4.1)$$

$$\Downarrow$$

$$\int_a^b f(x)dx$$

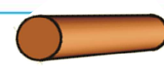
- If X is discrete and both a and b are possible values (e.g., X is binomial with $n = 20$ and $a = 5$, $b = 10$), then all four of the probabilities in (4.1) are different.

$$\left.\begin{array}{l} P(5 \le X \le 10) = \\ P(5 < X < 10) = \\ P(5 < X \le 10) = \\ P(5 \le X < 10) = \end{array}\right\} \boxed{?}$$

26

26

Probability Distributions for Continuous Variables



- Zero probability condition has physical analog.
- Consider solid circular rod with cross-sectional area = 1 in^2.

- Place rod alongside a measurement axis and suppose that density of rod at any point x is given by value $f(x)$ of density function
- Then if the rod is sliced at points a and b and this segment is removed,

$$\textit{The amount of mass removed is } \int_a^b f(x)dx$$

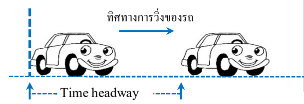
- if rod is sliced just at point c, no mass is removed.
- Mass is assigned to interval segments of rod but not to individual points.

27

27

Example 4.5

- "Time headway" in traffic flow is elapsed time between time that one car finishes passing fixed point and instant that next car begins to pass that point



ระยะเวลาห่าง (Time Headway) หมายถึง ระยะเวลาที่ห่างกันของรถยนต์ที่แล่นตามกันมา และผ่านจุดจับเวลาที่กำหนดบนถนน

- Let X = time headway for two randomly chosen consecutive cars on freeway during a period of heavy flow
- The following pdf of X is essentially the one suggested in "The Statistical Properties of Freeway Traffic" (*Transp. Res.,* vol. 11: 221–228):

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & otherwise \end{cases}$$

28

28

Example 4.5

cont'd

- The graph of $f(x)$ is given in Figure 4.4; there is no density associated with headway times less than 0.5, and
- headway density decreases rapidly (exponentially fast) as x increases from 0.5.

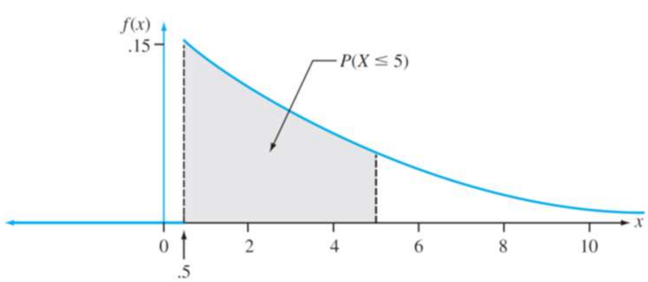


Figure 4.4 The density curve for time headway in Example 4.5

29

29

Example 4.5

cont'd

- Clearly, $f(x) \geq 0$; to show that $\int_{-\infty}^{\infty} f(x)dx = 1$,

From $f(x) = \begin{cases} \boxed{0.15e^{-0.15(x-0.5)} \quad x \geq 0.5} \\ 0 \quad otherwise \end{cases}$ $\Rightarrow$ $f(x) = 0.15\left(e^{-0.15x-(-0.15)(0.5)}\right) = 0.15\left(e^{-0.15x+0.075}\right) = 0.15e^{0.075} \cdot e^{-0.15x}$

31

31

Example 4.5

cont'd

- Probability that headway time is **<u>at most</u>** 5 sec is

$$P(X \leq 5) = \int_{-\infty}^{5} f(x)dx = \int_{0.5}^{5} 0.15e^{-0.15(x-0.5)}dx$$

$$= \int_{0.5}^{5} 0.15e^{0.075} \cdot e^{-0.15x}dx$$

$$= 0.15e^{0.075} \int_{0.5}^{5} e^{-0.15x}dx$$

$$= 0.15e^{0.075} \cdot \left(\frac{1}{-0.15} e^{-0.15x} \Big|_{x=0.5}^{x=5} \right)$$

$$= e^{0.075}\left(-e^{-(0.75)} + e^{-0.075}\right)$$

$$= 1.078(-0.472 + 0.928) = 0.491$$

$$= P(less\ than\ 5\ \text{sec}) = P(X < 5)$$

From Integrals with Exponentials

$$\int e^{ax}dx = \frac{1}{a}e^{ax}$$

e = 2.71828

32

32

4.2 Cumulative Distribution Functions and Expected Values

34

34

Cumulative Distribution Function

35

35

Cumulative Distribution Function

○ Cumulative distribution function (cdf) $F(x)$ for discrete random variable X gives, for any specified number x, $P(X \le x)$.

○ It is obtained by **summing the pmf *p*(*y*)** over all possible values y satisfying $y \le x$.

○ cdf of continuous random variable gives the same probabilities $P(X \le x)$ and is obtained by **integrating the pdf *f*(*y*)** between the limits $-\infty$ and x.

36

36

Cumulative Distribution Function

Definition

Cumulative Distribution Function $F(x)$ for continuous random variable X is defined for every number x by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy$$

For each x, $F(x)$ is area under the density curve to the left of x.

This is illustrated in Figure 4.5, where $F(x)$ increases smoothly as x increases.

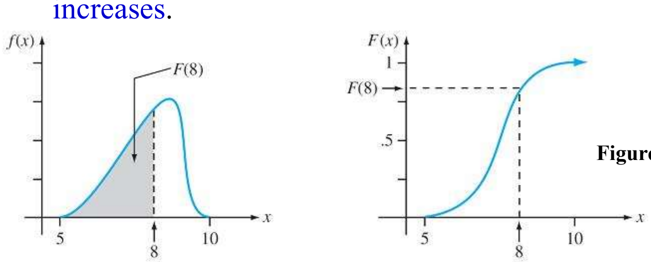


Figure 4.5 A pdf and associated cdf

37

37

Example 4.6

○ Let X, the thickness of a certain metal sheet, have uniform distribution on $[A, B]$.

○ The density function is shown in Figure 4.6.

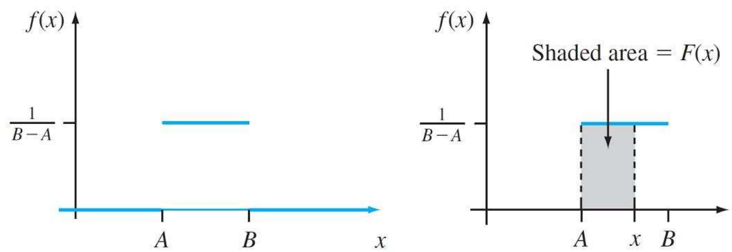


Figure 4.6 The pdf for a uniform distribution

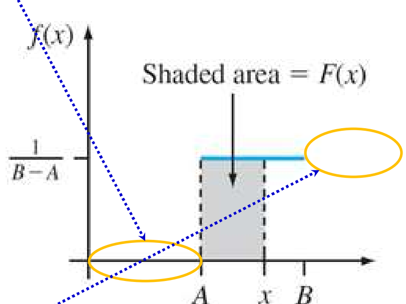
38

38

Example 4.6

cont'd

○ For $x < A$, $F(x) = 0$, since there is no area under graph of the density function to the left of such an x.



For $x \geq B$, $F(x) = 1$, since all the area is accumulated to the left of such an x.

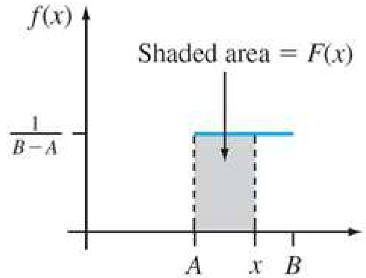
39

39

Example 4.6

cont'd

Finally for $A \leq x \leq B$,



$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A} dy = \frac{1}{B-A} \cdot y \Big|_{y=A}^{y=x} = \frac{x-A}{B-A}$$

40

Example 4.6

cont'd

- The entire cdf is

$$F(x) = \begin{cases} 0 & x < A \\ \dfrac{x-A}{B-A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

- The graph of this cdf appears in Figure 4.7.

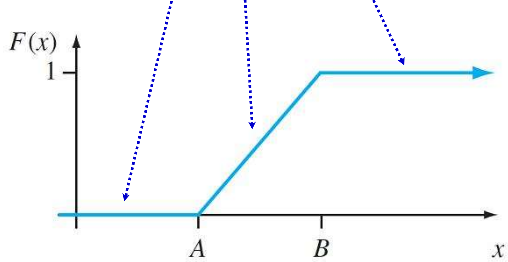


Figure 4.7 The cdf for a uniform distribution

41

Using $F(x)$ to Compute Probabilities

42

Using $F(x)$ to Compute Probabilities

- Importance of cdf here, just as for discrete random variable's, is that probabilities of various intervals can be computed from a formula for or table of $F(x)$.

Proposition

Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$.

Then for any number a,

$$P(X > a) = 1 - F(a)$$

and for any two numbers a and b with $a < b$,

$$P(a \le X \le b) = F(b) - F(a)$$

43

Using $F(x)$ to Compute Probabilities

- Figure 4.8 illustrates the second part of this proposition; desired probability is shaded area under density curve between a and b, and it equals the difference between the two shaded cumulative areas.

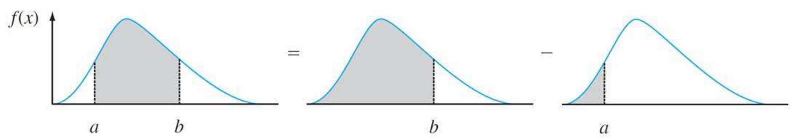


Figure 4.8 Computing $P(a \le X \le b)$ from cumulative probabilities

- This is different from what is appropriate for discrete integer valued random variable (e.g., binomial or Poisson):

$P(a \le X \le b) = F(b) - F(a - 1)$ when a and b are integers.

44

44

Example 4.7

- Suppose pdf of the magnitude X of a dynamic load on a bridge (in newtons) is

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \le x \le 2 \\ 0 & otherwise \end{cases}$$

- For any number x between 0 and 2,

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

$$= \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y \right) dy$$

$$= \frac{x}{8} + \frac{3}{16}x^2$$

45

45

Example 4.7

cont'd

Thus

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{8} + \dfrac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$

Graphs of $f(x)$ and $F(x)$ are shown in Figure 4.9.

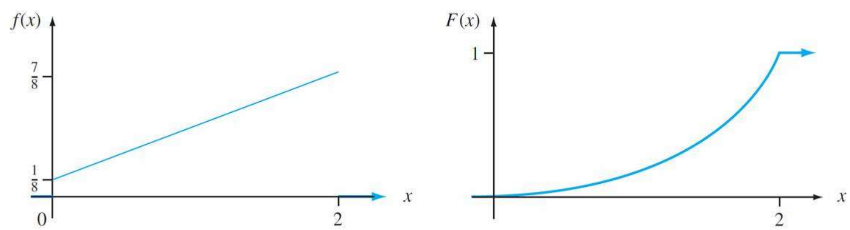


Figure 4.9 The pdf and cdf for Example 4.7

46

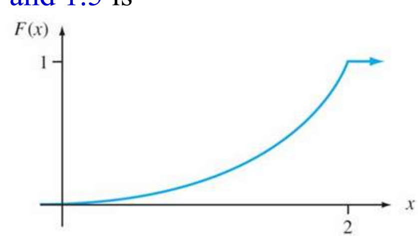
46

Example 4.7

cont'd

- Probability that load is between 1 and 1.5 is

$$F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{8} + \dfrac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$



$$P(1 \le X \le 1.5) = F(1.5) - F(1)$$

$$= \left[\frac{1}{8}(1.5) + \frac{3}{16}(1.5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right]$$

$$= \frac{19}{64}$$

$$= 0.297$$

48

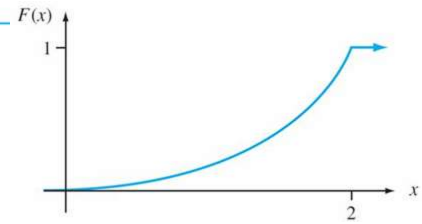
48

Example 4.7

cont'd

○ Probability that **load exceeds 1** is

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}$$



$$P(X > 1) = 1 - P(X \le 1)$$

$$= 1 - F(1)$$

$$= 1 - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right]$$

$$= \frac{11}{16} \quad = 0.688$$

Once cdf has been obtained, any probability involving X can easily be calculated without any further integration.

49

49

Obtaining $f(x)$ from $F(x)$

50

50

Obtaining $f(x)$ from $F(x)$

- For X discrete, probability mass function (pmf) is obtained from cdf by taking difference between two $F(x)$ values.
- The continuous analog of a difference is derivative.
- The following result is a consequence of Fundamental Theorem of Calculus.

Proposition

If X is continuous random variable with pdf $f(x)$ and cdf $F(x)$, then at every x at which derivative $F'(x)$ exists,

$$\boldsymbol{F'(x) = f(x).}$$

51

51

Example 4.8

When X has a uniform distribution,
$F(x)$ is differentiable except at $x = A$ and $x = B$, where the graph of $F(x)$ has sharp corners.

Since $F(x) = 0$ for $x < A$ and $F(x) = 1$ for $x > B$, $F'(x) = 0 = f(x)$ for such x.

For $A < x < B$,

$$F'(x) = \frac{d}{dx}\left(\frac{x-A}{B-A}\right)$$

$$= \frac{1}{B-A}$$

$$= f(x)$$

52

52

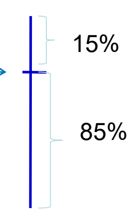
Percentiles of a Continuous Distribution

53

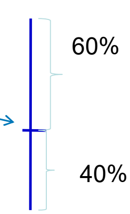
53

Percentiles of a Continuous Distribution

- we say that individual's test score was at the 85th percentile of population, we mean that 85% of all population scores were below that score and 15% were above.



- Similarly, the **40th percentile** is the score that exceeds 40% of all scores and is exceeded by 60% of all scores.



54

54

Percentiles of a Continuous Distribution

Proposition

Let p be a number between 0 and 1.

The **(100p)th percentile** of distribution of continuous random variable X, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy \qquad (4.2)$$

55

55

Percentiles of a Continuous Distribution

According to Expression (4.2),

$\eta(p)$ is that value on the measurement axis such that

100p% of area under graph of $f(x)$ lies to the left of $\eta(p)$ and $100(1-p)$% lies to the right.

- Thus $\eta(0.75)$, the 75th percentile, is such that area under graph of $f(x)$ to the left of $\eta(0.75)$ is 0.75.
- Figure 4.10 illustrates the definition.

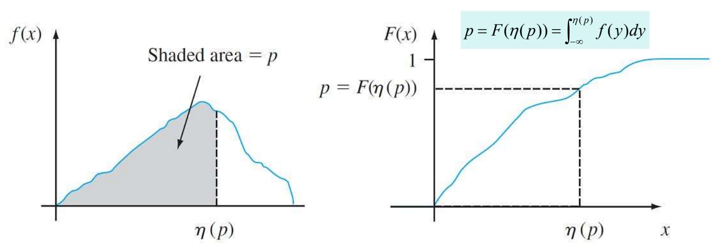


Figure 4.10 The (100p)th percentile of a continuous distribution

56

56

Example 4.9

Distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous random variable X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

57

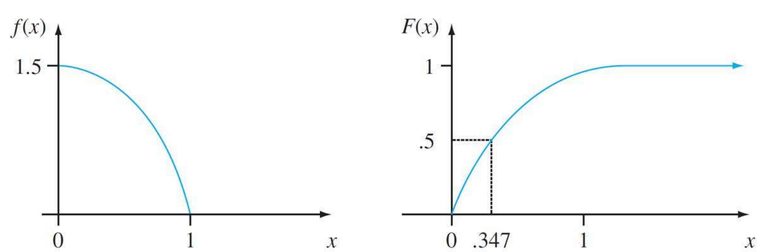
57

Example 4.9

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The cdf of sales for any x between 0 and 1 is

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)\,dy = \frac{3}{2}\left(y - \frac{y^3}{3}\right)\Bigg|_{y=0}^{y=x} = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$



58

58

Example 4.9

cont'd

The (100p)th percentile of this distribution satisfies the equation

$$F(x) = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

that is,

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

59

59

Example 4.9

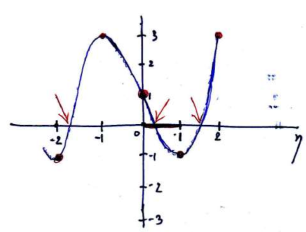
cont'd

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

For the 50th percentile, $p = 0.5$, and equation to be solved is

$$\eta^3 - 3\eta + 1 = 0$$

η	η^3-3η+1
-2	-1.0000
-1	3.0000
0	1.0000
1	-1.0000
2	3.0000



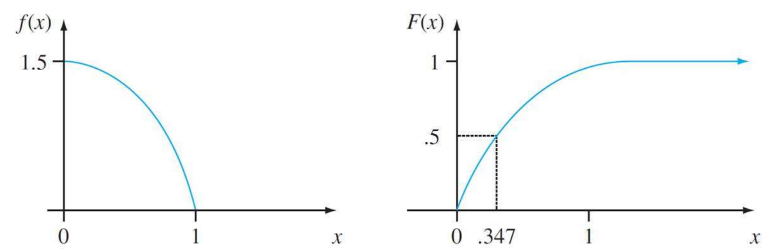
η	$(\eta^3+1)/3$
0.5	0.375000
0.375000	0.350911
0.350911	0.347737
0.347737	0.347350
0.347350	0.347303
0.347303	0.347297
0.347297	0.347296
0.347296	0.347296
0.347296	0.347296
0.347296	0.347296
0.347296	0.347296
0.347296	0.347296
0.347296	0.347296

the solution is $\eta = \eta(0.5) = 0.347$.

60

60

Example 4.9



If distribution remains the same from week to week, then in the long run

- 50% of all weeks will result in sales of less than 0.347 ton and
- 50% in more than 0.347 ton.

61

61

Percentiles of a Continuous Distribution

Definition

- **Median** of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $F(\tilde{\mu}) = 0.5$

- That is, half area under density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.

64

64

Percentiles of a Continuous Distribution

Continuous distribution whose pdf is **symmetric**—

graph of pdf to the left of some point is mirror image of graph to the right of that point—has **median $\tilde{\mu}$** equal to point of symmetry, since half area under curve lies to either side of this point.

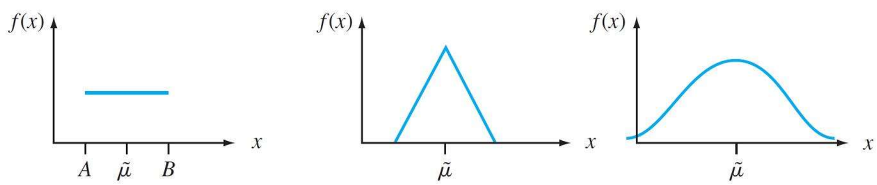


Figure 4.12 Medians of symmetric distributions

Figure 4.12 gives several examples.

Error in a measurement of physical quantity is often assumed to have symmetric distribution.

65

Expected Values

66

Expected Values

- For discrete random variable X, $E(X)$ was obtained by summing $x \cdot p(x)$ over possible X values.
- Here we replace summation by integration and pmf by pdf to get continuous weighted average.

Definition

Expected or **Mean value** of a continuous random variable X with pdf $f(x)$ is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x.f(x)dx$$

67

67

Example 10

The pdf of weekly gravel sales X was

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)\,dx$$

$$= \frac{3}{2}\int_0^1 (x - x^3)\,dx = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{3}{8}$$

68

68

Expected Values

Often we wish to compute the expected value of some function $h(X)$ of the random variable X.

If we think of $h(X)$ as new random variable Y, techniques from mathematical statistics can be used to derive pdf of Y, and $\boldsymbol{E(Y)}$ can then be computed from definition.

70

70

Expected Values

Proposition

If X is continuous random variable with pdf $f(x)$ and $h(X)$ is any function of X, then

$$\mathrm{E}[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

72

72

Example 4.11

Two species are competing in a region for control of limited amount of a certain resource.

Let X = the proportion of resource controlled by species 1 and suppose X has pdf

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

which is a uniform distribution on [0, 1]. (In her book *Ecological Diversity,* E. C. Pielou calls this the "broken- tick" model for resource allocation, since it is analogous to breaking a stick at a randomly chosen point.)

73

73

Example 4.11

cont'd

Then the species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1 - X) = \begin{cases} 1 - X & \text{if } 0 \le X < \frac{1}{2} \\ X & \text{if } \frac{1}{2} \le X \le 1 \end{cases}$$

The expected amount controlled by the species having majority control is then

$$\mathrm{E}[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} \max(x, 1 - x) \cdot f(x)\, dx$$

74

74

Example 4.11

cont'd

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad h(X) = \max(X, 1 - X) = \begin{cases} 1 - X & \text{if } 0 \le X < \frac{1}{2} \\ X & \text{if } \frac{1}{2} \le X \le 1 \end{cases}$$

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} \max(x, 1 - x) \cdot f(x)\, dx$$

75

75

Expected Values

In the discrete case, variance of X was defined as the expected squared deviation from μ and was calculated by summation.

Here again integration replaces summation.

Definition

The **variance** of a continuous random variable X with pdf $f(x)$ and mean value μ is

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E[(X - \mu)^2]$$

The **standard deviation** (SD) of X is $\sigma_X = \sqrt{V(X)}$

78

78

Variance and Standard Deviation

Variance of continuous random variable X with pdf f(x) and mean value μ is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 . f(x)dx = E[(X-\mu)^2]$$

Standard deviation (SD) of X is $\sigma = \sqrt{V(X)}$

Variance and Standard Deviation give quantitative measures of how much spread there is in distribution or population of *x* values.

The easiest way to computer σ^2 is to gain use shortcut formula.

Proposition

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

79

79

Example 12

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

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

For weekly gravel sales, we computed E(X) = $\frac{3}{8}$ Since

$$E(X^2) = \int_{-\infty}^{\infty} x^2 . f(x)dx = \int_0^1 x^2 . \frac{3}{2}\left(1-x^2\right)dx = \frac{3}{2}\int_0^1 x^2\left(1-x^2\right)dx$$

$$= \frac{3}{2}\int_0^1 \left(x^2 - x^4\right)dx = \frac{1}{5}$$

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

$$= \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{1}{5} - \frac{9}{64} = \frac{19}{320}$$

$$= 0.059$$

and $\sigma_X = \sqrt{V(X)} = \sqrt{0.059} = 0.244$

80

80

Example 12

$$E[h(X)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

cont'd

When $h(X) = aX + b$, the expected value and variance of $h(X)$ satisfy the same properties as in the discrete case:

$$E[h(X)] = a\mu + b \quad \text{and} \quad V[h(X)] = a^2 \cdot \sigma^2.$$

Example 12

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

cont'd

When $h(X) = aX + b$, the expected value and variance of $h(X)$ satisfy the same properties as in the discrete case:

$$E[h(X)] = a\mu + b \quad \text{and} \quad V[h(X)] = a^2 \cdot \sigma^2.$$