4 Continuous Random Variables and Probability Distributions

4.2 Cumulative Distribution Functions and Expected Values

The Cumulative Distribution Function

The Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ for a discrete rv X gives, for any specified number x, the probability $P(X \leq x)$.

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

The cdf of a continuous rv gives the same probabilities $P(X \leq x)$ and is obtained by integrating the pdf $f(y)$ between the limits $-\infty$ and x.

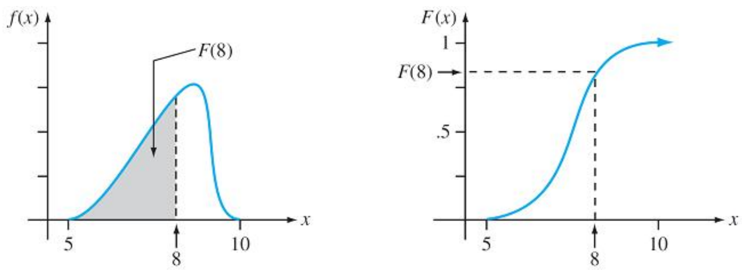
The Cumulative Distribution Function

Definition

The **cumulative distribution function** $F(x)$ for a continuous rv X is defined for every number x by

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

For each x, $F(x)$ is the 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.



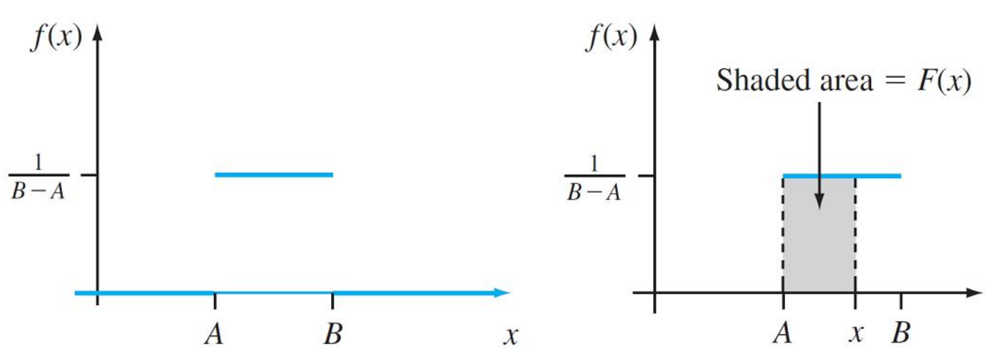
A pdf and associated cdf

Figure 4.5

Example 6

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

The density function is shown in Figure 4.6.



The pdf for a uniform distribution

Figure 4.6

Example 6

cont'd

For $x < A$, $F(x) = 0$, since there is no area under the 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. Finally for $A \leq x \leq B$,

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

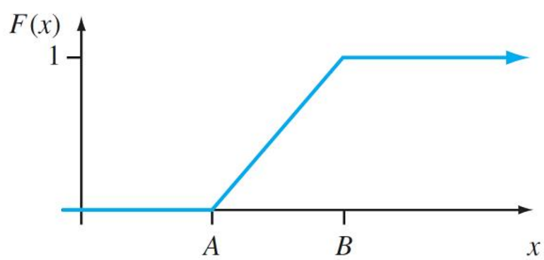
Example 6

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.



The cdf for a uniform distribution

Figure 4.7

Using $F(x)$ to Compute Probabilities

Using $F(x)$ to Compute Probabilities

The importance of the cdf here, just as for discrete rv'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 rv with pdf $f(x)$ and cdf $F(x)$. Then for any number a,

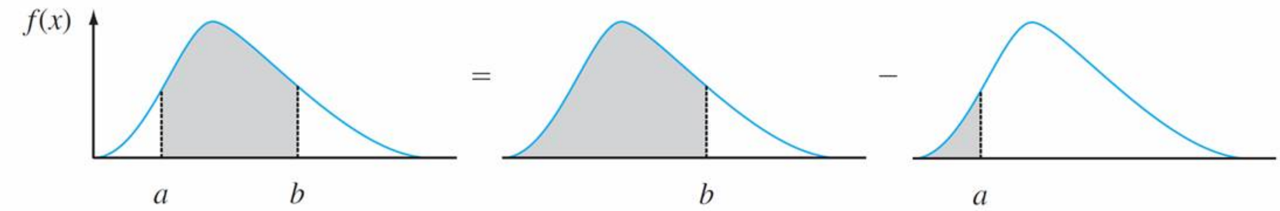
$$P(X > a) = 1 - F(a)$$

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

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

Using $F(x)$ to Compute Probabilities

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



Computing P(a ≤ X ≤ b) from cumulative probabilities

Figure 4.8

This is different from what is appropriate for a 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.

Example 7

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

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

For any number *x* between 0 and 2,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} \left(\frac{1}{8} + \frac{3}{8}y \right) dy = \frac{x}{8} + \frac{3}{16}x^2$$

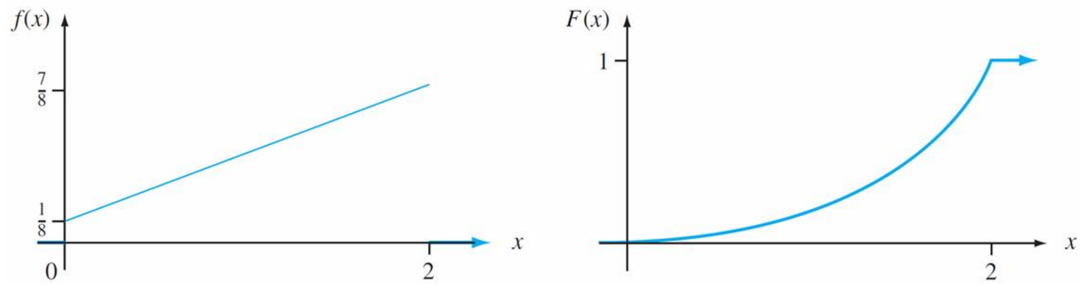
Example 7

cont'd

Thus

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

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



The pdf and cdf for Example 4.7

Figure 4.9

Example 7

cont'd

The probability that the load is between 1 and 1.5 is

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

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

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

$$= .297$$

The probability that the load exceeds 1 is

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

$$= 1 - F(1)$$

Example 7

cont'd

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

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

$$= .688$$

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

Obtaining *f*(*x*) from *F*(*x*)

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

For X discrete, the pmf is obtained from the cdf by taking the difference between two $F(x)$ values. The continuous analog of a difference is a derivative.

The following result is a consequence of the Fundamental Theorem of Calculus.

Proposition

If X is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every x at which the derivative $F'(x)$ exists, $F'(x) = f(x)$.

Example 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 $A < x < B$,

$$F'(x) = \frac{d}{dx}\left(\frac{x - A}{B - A}\right) = \frac{1}{B - A} = f(x)$$

Percentiles of a Continuous Distribution

Percentiles of a Continuous Distribution

When we say that an individual's test score was at the 85th percentile of the 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.

Percentiles of a Continuous Distribution

Proposition

0.0 – 1.0

Let p be a number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous rv X, denoted by $\eta(p)$, is defined by

พท.ใต้กราฟ ของ pdf

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

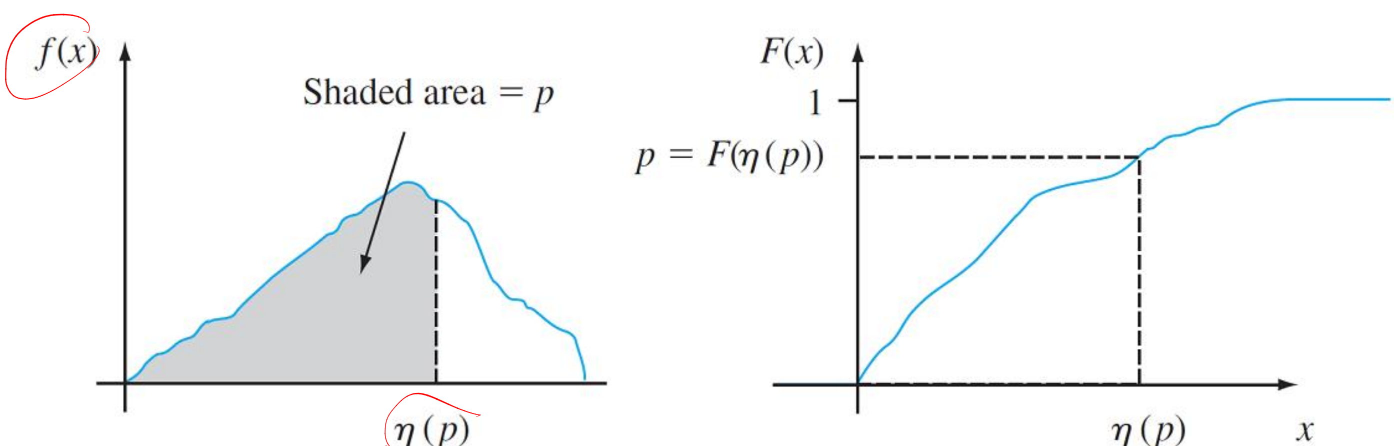
CDF

According to Expression (4.2), $\eta(p)$ is that value on the measurement axis such that 100p% of the area under the graph of $f(x)$ lies to the left of $\eta(p)$ and 100(1 – p)% lies to the right.

Percentiles of a Continuous Distribution

Thus $\eta(.75)$, the 75th percentile, is such that the area under the graph of $f(x)$ to the left of $\eta(.75)$ is .75.

Figure 4.10 illustrates the definition.



ค่า The (100p)th percentile of a continuous distribution

Figure 4.10

Example 9

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

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

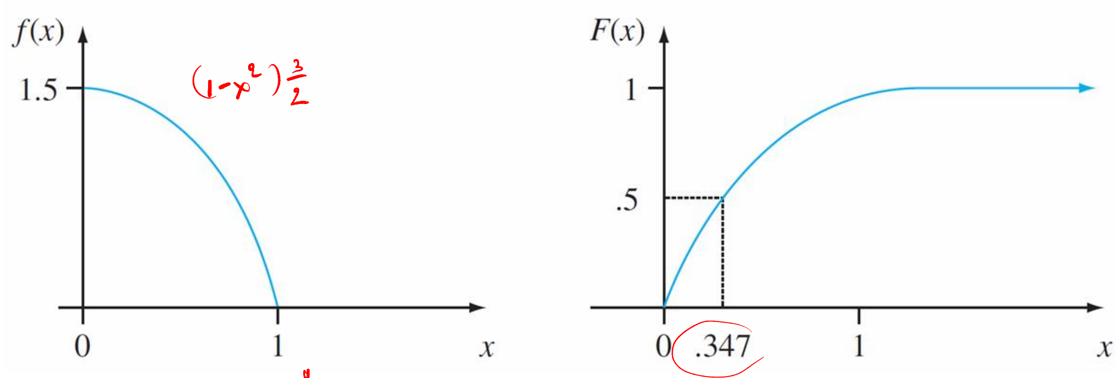
The cdf of sales for any x between 0 and 1 is

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

Example 9

cont'd

The graphs of both $f(x)$ and $F(x)$ appear in Figure 4.11.



$(1-x^2)\frac{3}{2}$

ขายได้เยอะ

x เป็น 0

ขายไม่ได้

The pdf and cdf for Example 4.9

Figure 4.11

Example 9

cont'd

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

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

that is,

$x^3 - 3x + 2p = 0$

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

For the 50th percentile, $p = .5$, and the equation to be solved is $\eta^3 - 3\eta + 1 = 0$; the solution is $\eta = \eta(.5) = .347$. If the distribution remains the same from week to week, then in the long run 50% of all weeks will result in sales of less than .347 ton and 50% in more than .347 ton.

Percentiles of a Continuous Distribution

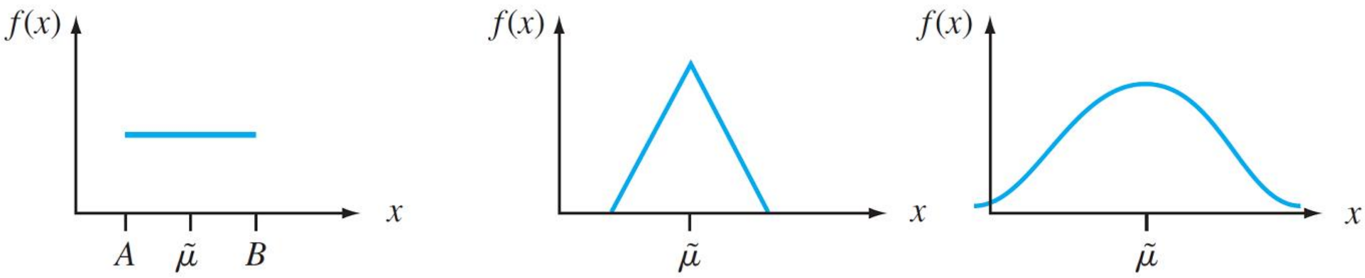
Definition

The **median** of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $.5 = F(\tilde{\mu})$ That is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.

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

Percentiles of a Continuous Distribution

Figure 4.12 gives several examples. The error in a measurement of a physical quantity is often assumed to have a symmetric distribution.



Medians of symmetric distributions

Figure 4.12

Expected Values

Expected Values

For a 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 the pmf by the pdf to get a continuous weighted average.

Definition

The **expected** or **mean value** of a continuous rv$X$ with pdf $f(x)$ is

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

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

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

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

Expected Values

When the pdf $f(x)$ specifies a model for the distribution of values in a numerical population, then μ is the population mean, which is the most frequently used measure of population location or center.

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

If we think of $h(X)$ as a new rv Y, techniques from mathematical statistics can be used to derive the pdf of Y, and $E(Y)$ can then be computed from the definition.

Expected Values

Fortunately, as in the discrete case, there is an easier way to compute $E[h(X)]$.

Proposition

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

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

Example 11

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

Let X = the proportion of the 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.)

Example 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

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

Example 11

cont'd

$$= \int_0^1 \max(x, 1 - x) \cdot 1\, dx$$

$$= \int_0^{1/2} \max(x, 1 - x) \cdot 1\, dx + \int_{1/2}^1 x \cdot 1\, dx$$

$$= \frac{3}{4}$$

Expected Values

For $h(X)$, a linear function, $E[h(X)] = E(aX + b) = aE(X) + b$.

In the discrete case, the 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

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

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

Expected Values

The variance and standard deviation give quantitative measures of how much spread there is in the distribution or population of x values.

Again σ is roughly the size of a typical deviation from . Computation of σ^2 is facilitated by using the same shortcut formula employed in the discrete case.

Proposition

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

Example 12

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_{0}^{1} x^2 \quad \frac{3}{2} (1 - x^2)\, dx$$

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

Example 12

cont'd

$$V(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2$$

$$= \frac{19}{320}$$

$$= .059$$

and $\sigma_X = .244$

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

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