The Cumulative Distribution Function for a Random Variable X

Each continuous random variable X has an associated probability density function (pdf) $f(x)$. It "records" the probabilities associated with X as areas under its graph. More precisely,

"the probability that a value of X is between a and b" $= P(a \leq X \leq b) = \int_a^b f(x)\,dx$.
For example,

$$
\begin{aligned}
&P(1 \leq X \leq 3) \qquad\quad = \int_1^3 f(x)\,dx \\
&P(3 \leq X) = P(3 \leq X < \infty) = \int_3^\infty f(x)\,dx \\
&P(X \leq -1) = P(-\infty < X \leq -1) = \int_{-\infty}^{-1} f(x)\,dx
\end{aligned}
$$

i) Since probabilities are always between 0 and 1, it must be that $f(x) \geq 0$
(so that $\int_a^b f(x)\,dx$ can never give a "negative probability"), and
ii) Since a "certain" event has probability 1,
$P(-\infty < X < \infty) = 1 = \int_{-\infty}^{\infty} f(x)\,dx$ = total area under the graph of $f(x)$

The properties i) and ii) are necessary for a function $f(x)$ to be the pdf for some random variable X.

We can also use property ii) in computations: since

$$
\begin{aligned}
&\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{3} f(x) + \int_3^\infty f(x)\,dx = 1 \\
&P(X \leq 3) = \int_{-\infty}^{3} f(x)\,dx = 1 - \int_3^\infty f(x)\,dx = 1 - P(X \geq 3)
\end{aligned}
$$

The pdf is discussed in the textbook.

There is another function, the cumulative distribution function (cdf) which records the same probabilities associated with X, but in a different way. The cdf $F(x)$ is defined by

$$F(x) = P(X \leq x).$$

$F(x)$ gives the "accumulated" probability "up to x." We can see immediately how the pdf and cdf are related:

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$ (since "x" is used as a variable in the upper limit of integration, we use some other variable, say "t", in the integrand)

Notice that $F(x) \geq 0$ (since it's a probability), and that

a) $\lim\limits_{x\to\infty} F(x) = \lim\limits_{x\to\infty} \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{\infty} f(t)\,dt = 1$ and
b) $\lim\limits_{x\to-\infty} F(x) = \lim\limits_{x\to-\infty} \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-\infty} f(t)\,dt = 0$, and that

c) $F'(x) = f(x)$ (by the Fundamental Theorem of Calculus)

Item c) states the connection between the cdf and pdf in another way:

<u>the cdf</u> $F(x)$ <u>is an antiderivative of the pdf</u> $f(x)$ (the particular antiderivative where the constant of integration is chosen to make the limit in a) true)

and therefore

$P(a \le X \le b) = \int_a^b f(x)\,dx = F(x)|_a^b = F(b) - F(a) = P(X \le b) - P(X \le a)$

<u>Example</u>: Suppose X has an exponential density function. As discussed in class,

$$f(x) = \begin{cases} 0 & x < 0 \\ ce^{-cx} & x \ge 0 \end{cases} \text{ (where } c = \tfrac{1}{\mu}\text{)}$$

If $x \ge 0$, $\int_{-\infty}^x f(t)\,dt = \int_0^x f(t)\,dt = \int_0^x ce^{-ct}\,dt = -e^{-ct}|_0^x = 1 - e^{-cx}$, so

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-cx} & x \ge 0 \end{cases}$$

If X has mean $\mu = 3$, say, then $c = \frac{1}{\mu} = \frac{1}{3}$.

If we want to know $P(X \le 4)$, we can either compute $\int_{-\infty}^4 f(x)\,dx = \int_{-\infty}^4 \frac{1}{3}e^{-(1/3)x}dx \approx 0.736403$, or (now that we have the formula for $F(x)$ we can simply compute $F(3) = 1 - e^{-(1/3/)\cdot 4} = 1 - e^{-4/3} \approx 0.736403$.

(*The graphs of $f(x)$ and $F(x)$ are shown on the last page before exercises. In the figure, notice the values of* $\lim_{x\to\infty} F(x)$ *and* $\lim_{x\to-\infty} F(x)$).

<u>Example</u>: If X is a normal random variable with mean $\mu = 0$ and standard deviation $\sigma = 1$, then its pdf is $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, and its cdf $F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-t^2/2}\,dt$. Because there is no "elementary" antiderivative for $e^{-t^2/2}$, its not possible to find an "elementary" formula for $F(x)$. However, for any x, the value of $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-t^2/2}\,dt$ can be estimated, so that a graph of $F(x)$ can be drawn. (*See figure on the last page before exercises.*)

Example: More generally, probability calculations involving a normal random variable X are computationally difficult because again there's no elementary formula for the cumulative distribution function $F(x)$ – that is, an antiderivative for the probability density function :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

Therefore it's not possible to find an exact value for

$$P(a \le X \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}dx = F(b) - F(a)$$

Suppose X is a normal random variable with mean $\mu = 1.9$ and standard deviation $\sigma = 1.7$. If we want to find $P(-3 \le X \le 2)$, we need to estimate

$$\frac{1}{(1.7)\sqrt{2\pi}}\int_{-3}^{2} e^{-(x-1.9)^2/2(1.7)^2}\,dx = F(2) - F(-3).$$

This can be done with Simpson's Rule. However, such calculations are so important that the TI83-Plus Calculator has a built in way to make the estimate:

Punch keys *2nd* $DISTR$
Choose item 2 on the menu: *normalcdf*
On the screen you see normalcdf (
Fill in normalcdf $(-3, 2, 1.9, 1.7)$
and the TI-83 gives the approximate value of the integral above: 0.521480

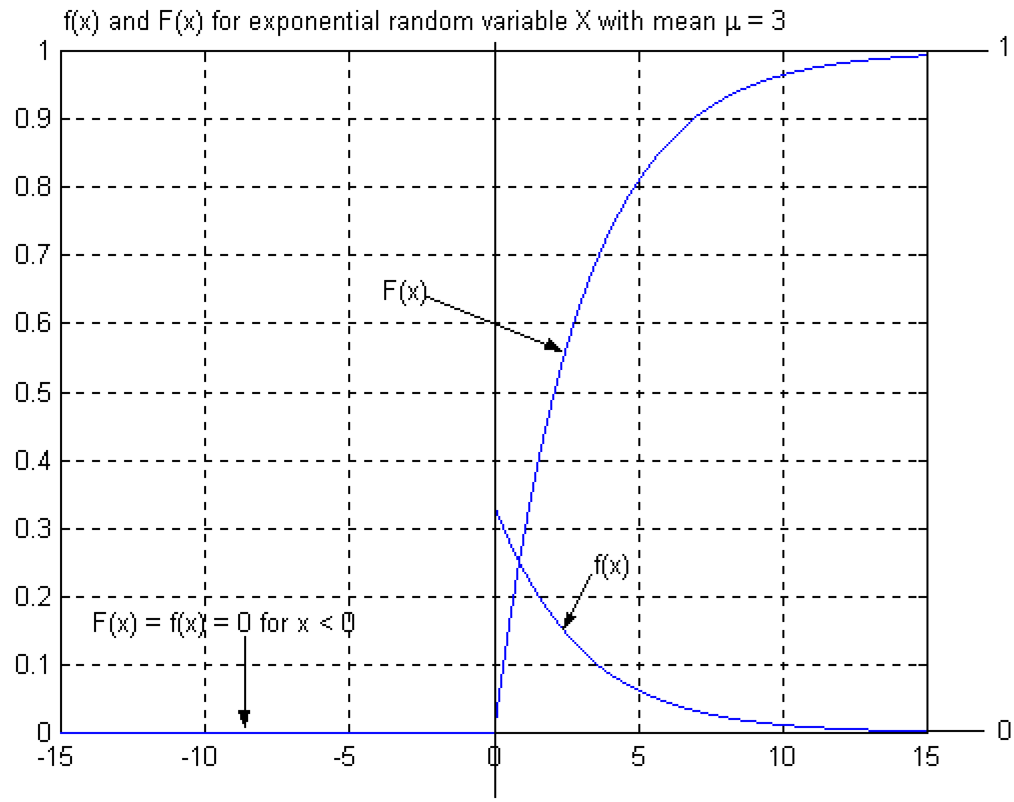
The general syntax for the command is normalcdf (*lowerlimit,upperlimit,*μ, σ)

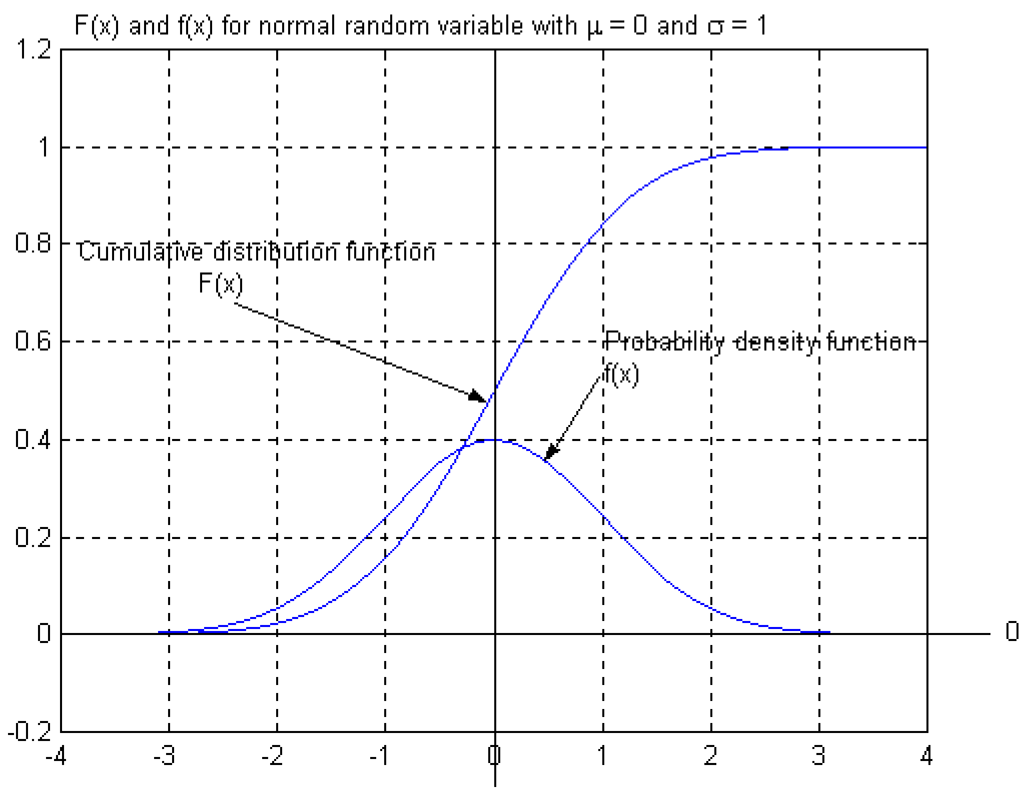
If you enter only normalcdf (*lowerlimit,upperlimit*) then the TI-83 assumes $\mu = 0, \sigma = 1$ as the default values

Note that using the values for μ, σ example given above:

$P(\mu - \sigma \le X \le \mu + \sigma) \approx$ normalcdf$(.2, 3.6, 1.9, 1.7) \approx 0.6827$
$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx$ normalcdf$(-1.5, 5.3, 1.9, 1.7) \approx 0.9545$
$P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx$ normalcdf$(-3.2, 7, 1.9, 1.7) \approx 0.9973$

In fact (as may have been mentioned in class) these probabilities come out the same for any normal random variable, no matter what the values of μ and σ: for example, the probability that any normal random variable takes on a value between $\pm$ one standard deviation of its mean is ≈ 0.6827.


f(x) and F(x) for exponential random variable X with mean μ = 3
F(x)
f(x)
F(x) = f(x) = 0 for x < 0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-15
-10
-5
0
5
10
15
1
0


F(x) and f(x) for normal random variable with μ = 0 and σ = 1
Cumulative distribution function
F(x)
Probability density function
f(x)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-4
-3
-2
-1
0
1
2
3
4
0

Exercises:

1. A certain “uniform” random variable X has pdf $f(x) = \begin{cases} 1/5 & 2 \leq x \leq 7 \\ 0 & \text{otherwise.} \end{cases}$

   a) What is $P(0 \leq X \leq 3)$?

   b) Write the formula for its cdf $F(x)$

   c) What is $F(3) - F(0)$?

2. A certain kind of random variable as density function $f(x) = \frac{1}{\pi(1+x^2)}$.

   a) What is $P(X \geq -1)$?

   b) Write the formula for its cdf $F(x)$

   c) Write a formula using $F(x)$ that gives the answer to part a). Check that it agrees with your numerical answer in a).