# Statistics 3657 : Pdf, Cdf and Quantiles

In this section we consider a continuous r.v., say $X$. It has a cdf defined in the same way as for discrete rv's. Let $F$ be the cdf. Its properties are determined by the Axioms and rules of probability. However in this case of a continuous rv it has additional properties

1. $F$ is differentiable, except at some isolated points. Let $f(x)$ be the derivative at the point $x$.

2. $F$ can be recovered from $f$, that is

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

   for all $x \in (-\infty, \infty)$.

$f$ is called the probability density function of $F$. $F$ and $f$ are equivalent in the sense that either can be used to calculate

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(y) dy \ .$$

Either determines the distribution of the rv $X$.

For those values of $x$ for which $F'(x)$ (derivative of $F$ at $x$) does not exist, we can assign a value to $f(x)$ that is arbitrary or convenient for us. Usually this is done to give a simpler expression, for example to be continuous or right or left continuous.

Property 2 is really needed. Consider the following cdf of a discrete rv $X$ which takes the value 2 with probability 1. It has cdf

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

Then for all $x$ except $x = 2$,

$$f(x) = \lim_{\delta \to 0} \frac{F(x+\delta) - F(x)}{\delta} = 0 \ .$$

Then, since integrals are the same for functions that might differ at a given fixed argument, we have

$$G(x) = \int_{-\infty}^{x} f(y) dy = 0$$

for all $x$. Clearly $G \neq F$.

For most continuous distributions it is more convenient to work with the pdf rather than the cdf. However the cdf is needed either explicitly or through integrating the pdf to determine probabilities concerning the rv $X$. Later when we study transformations our goal will be to determine the distribution of the new resulting r.v., which may be in either the CDF or an equivalent form. However our goal will generally be to write the distribution as a pdf in the continuous case or pmf in the discrete case.

Properties of a PDF :

1. $f(x) \geq 0$.

2. $$\int_{-\infty}^{\infty} f(y)dy = 1$$

In fact any function $f$ satisfying these two properties can be used as a pdf. However there are some particular pdfs that are common and have names. **The student should read the text for these particular pdfs which we will use throughout the course**.

**Exponential pdf**

The exponential, parameter $\lambda$ pdf is given by

$$f(x) = ce^{-\lambda x} I_{[0,\infty)}(x) \tag{1}$$

where $I$ is an indicator function.

$$I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

In other contexts we $I_A(x) = I(x \in A)$. These often allow us to write a function in a single line.

If $\lambda \leq 0$ then from (1) we have

$$\int_0^\infty e^{-\lambda x} dx = \infty$$

since the integrand is greater or equal to 1 for all $x \geq 0$. Thus (1) cannot be a pdf if $\lambda \leq 0$. If $\lambda > 0$ then (1) integrates to 1 if $c = \lambda$. Thus the parameter space for the family of exponential distributions is $\Lambda = \{\lambda : \lambda > 0\} = (0, \infty)$.

It then has CDF $F$ given by

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Notice this cdf is continuous. This function $F$ has derivative for all values of $x$ except $x = 0$, which is on the boundary of the support of $f$. By direct calculation one verifies that if $x \neq 0$ then $F'(x) = f(x)$ above. If $x = 0$ then

$$\lim_{\delta \to 0-} \frac{F(0+\delta) - F(0)}{\delta} = 0$$

and

$$\lim_{\delta \to 0+} \frac{F(0+\delta) - F(0)}{\delta} = \lambda$$

Thus the left hand and right hand limits are not equal and the derivative of $F$ does not exist at $x = 0$. Thus if we take any function $g$ such that $g(x) = f(x)$ for $x \neq 0$ then it can also be a pdf of $F$; that is $g$ obeys the two requirements of being a pdf. By convention and convenience we will take $g$ to be either left continuous or right continuous, so we use either (1) or

$$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x) \ .$$

**A No Name Example**

Consider a function $g$ with support [0,2] given by

$$g(x) = x + x^2$$

and a pdf $f$ proportional to $g$. This means there is a constant $c$ (constant with respect to the argument of the function $f$) such that $f(x) = cg(x)$ for all $x$. Clearly $g$ is a non-negative function, and hence we need to find a positive value of $c$. The student should do this at home.

Then

$$f(x) = \begin{cases} \frac{3}{14}\left(x + x^2\right) & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

In a more compact form we write this as

$$f(x) = \frac{3}{14}\left(x + x^2\right) \mathrm{I}(0 \le x \le 2)$$

or

$$f(x) = \frac{3}{14}\left(x + x^2\right) \mathrm{I}_{[0,2]}(x)$$

where I is an indicator function. The CDF is then given by

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \frac{3}{14}\left(x^2/2 + x^3/3\right) & \text{if } 0 < x \le 2 \\ 1 & \text{if } x > 2 \end{cases}$$

The $p$-th quantile is found by solving for $x$ in the equation $p = F(x)$. Since $0 < p < 1$ then the solution $x$ must be $0 < x < 2$ and must solve

$$\frac{3}{14}\left(x^2/2 + x^3/3\right) = p\ .$$

The table below gives some solutions for selected $p$. These solutions have been obtained by the solution (in $x$) of a cubic equation. They can also be obtained by numerical methods, such as an interval halving method (see the course text for a description of this method).

| | |
|---|---|
| 0.10 | 0.783076 |
| 0.25 | 1.149380 |
| 0.50 | 1.521960 |
| 0.75 | 1.787230 |
| 0.90 | 1.919550 |

Note that if we change the support we may not get a pdf. For example

$$g(x) = \left(x + x^2\right) \mathrm{I}_{[-1,1]}(x)$$

cannot be rescaled to produce a pdf. Why?

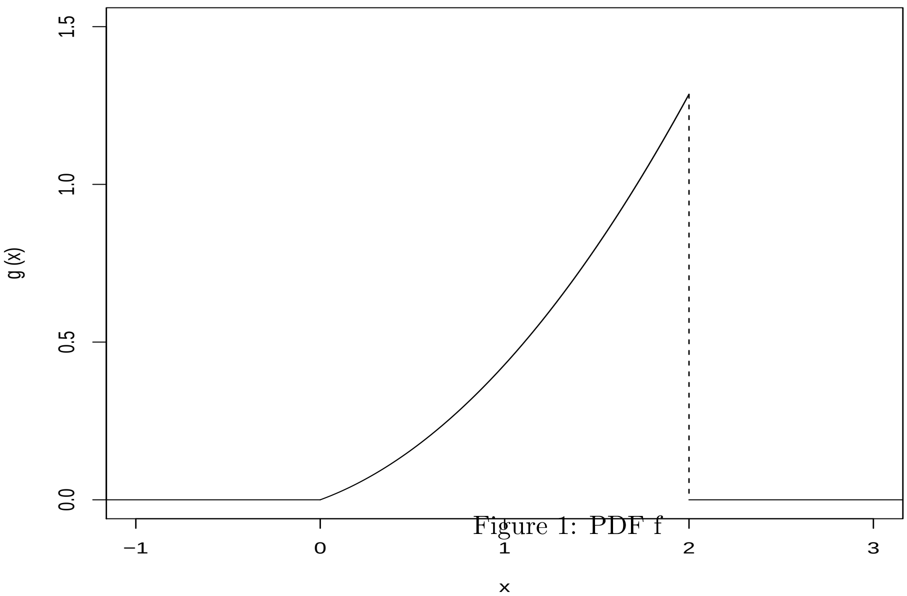


Figure 1: PDF f

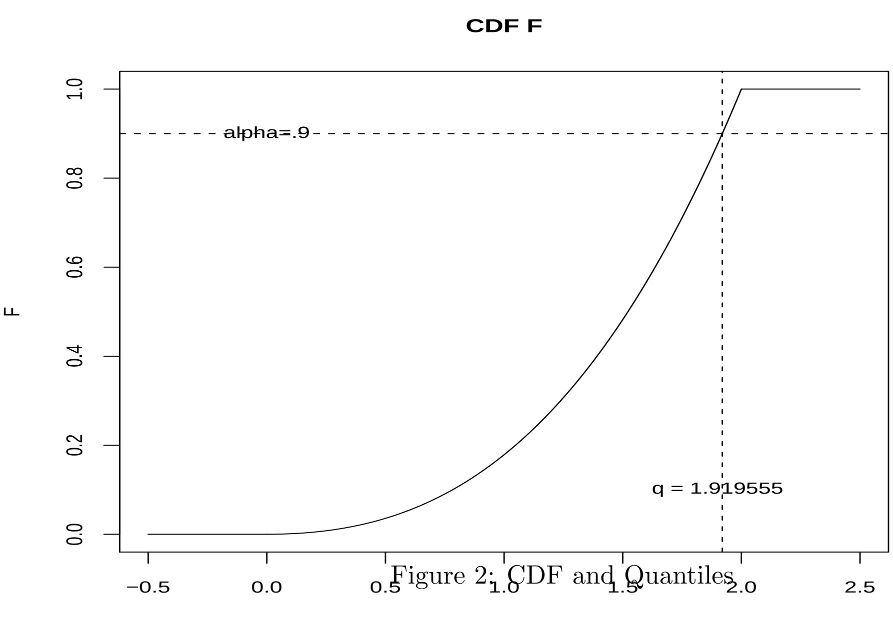


Figure 2: CDF and Quantiles

**Standard Normal Distribution**

Consider the probability density function $\phi$

$$\phi(x) = ce^{-\frac{x^2}{2}} \ , \ -\infty < x < \infty \ .$$

The constant $c$ can be determined as

$$c = \frac{1}{\sqrt{2\pi}}$$

and so the standard normal pdf is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \ .$$

The student should see Rice, Third Edition, Problem 2.5.51 in order to find $c$. Actually the student should modify this problem to find $c$ by the same technique. Specifically let

$$a = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \ .$$

The value of $a$ is finite; use one of the problems on the prerequisites review assignment to verify this. Then

$$\begin{aligned} a^2 &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \end{aligned}$$

Now change variables to polar co-ordinates and find the expression for $a^2$. After the polar co-ordinate change of variables the integration becomes quite nice and simple. The student should complete this at home.

The cdf is

$$\Phi(x) = \int_{-\infty}^{x} \phi(y) dy$$

Since this distribution comes up often it has its own special function name.

The cdf does not have a closed form expression. The integral must be calculated numerically or tabulated. Such tables are given in nearly every introductory statistics text, and is programmed in R and other common statistical languages and packages.