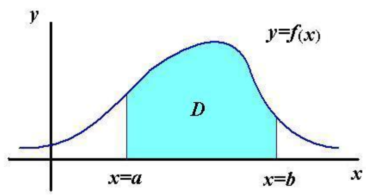
The probability density function (PDF) of a continuous random variable

- A random variable X is continuous if $P(X = x) = 0$ for every x.
- For a function f and for $b > a$, the notation $\int_a^b f(x)dx$ is called the integral of f from a to b (f 從 a 到 b 的積分).
  - If $f \geq 0$, $\int_a^b f(x)dx$ is the area of region D in the following graph



- Some facts about integration.
  - $\int_a^b f(x)dx$ can be defined for f that is not nonnegative on $[a, b]$.
  - When $f(x) \leq 0$ for $x \in [a, b]$, $\int_a^b f(x)dx = -\int_a^b [-f(x)]dx$.
  - $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$.
- PDF (機率密度函數) of a continuous random variable. For a continuous random variable X, if there exists a non-negative function f_X such that

$$P(a < X < b) = \int_a^b f_X(x)dx \text{ for all } a < b, \tag{1}$$

  then f is called the probability density function (PDF) of X.
  - Note: (1) implies that $\int_{-\infty}^{\infty} f_X(x)dx = 1$.
- Compute an integral $\int_a^b h(x)dx$ using R. Suppose that the function h has been defined in R. Then running the R command

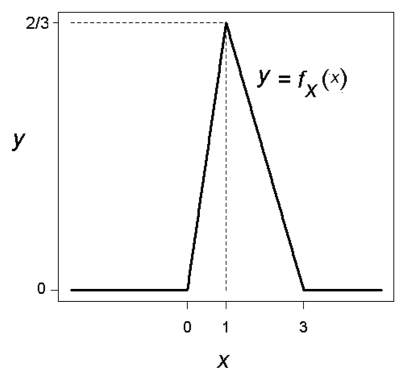
```
integrate(h, a, b)$value
```

  gives $\int_a^b h(x)dx$. Note that in `integrate(h, a, b)`, `a` can be `-Inf` (for $a = -\infty$) and `b` can be `Inf` (for $b = \infty$). For example,
  - `integrate(h, -Inf, b)$value` gives $\int_{-\infty}^b h(x)dx$.
  - `integrate(h, a, Inf)$value` gives $\int_a^{\infty} h(x)dx$.

- Example 1. Suppose that X has PDF f_X, where

$$f_X(x) = \begin{cases} 0 & \text{if } x < 0; \\ 2x/3 & \text{if } 0 \le x \le 1; \\ (3-x)/3 & \text{if } 1 \le x \le 3; \\ 0 & \text{if } x > 3. \end{cases}$$

The graph of $y = f_X(x)$ is given below.



Find $P(1 < X < 3)$ and $P(X < 0)$.

Sol.

$$P(1 < X < 3) = \int_1^3 f_X(x)dx = \int_1^3 \frac{3-x}{3}dx = \frac{1}{2}\cdot(3-1)\cdot\left(\frac{2}{3}\right) = \frac{2}{3}.$$

$$P(X < 0) = \int_{-\infty}^0 f_X(x)dx = \int_{-\infty}^0 0dx = 0.$$

We can also compute $P(1 < X < 3)$ using R:

```
f <- function(x){ (3-x)/3 }
integrate(f,1,3)$value
```

- If X is continuous with PDF f_X, then

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

In particular, the mean of X can be calculated using

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

and $Var(X) = E(X^2) - [E(X)]^2$, where

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

- Example 2. The output after running the R commands

```
h <- function(x){ 2*x/3 }
integrate(h, 0, 1)$value - 1/3
h <- function(x){ (3-x)/3 }
integrate(h, 1, 3)$value - 2/3

h <- function(x){ x*2*x/3 }
integrate(h, 0, 1)$value -2/9
h <- function(x){ x*(3-x)/3 }
integrate(h, 1, 3)$value -10/9

h <- function(x){ (x^2)*2*x/3 }
integrate(h, 0, 1)$value - 1/6
h <- function(x){ (x^2)*(3-x)/3 }
integrate(h, 1, 3)$value
```

is

```
[1] 5.551115e-17
[1] 1.110223e-16
[1] 2.775558e-17
[1] -2.220446e-16
[1] 2.775558e-17
[1] 2
```

Find $E(X)$, $E(X^2)$ and $Var(X)$ for the X in Example 1 using the above R output. Note that in the R output,

```
2.775558e-17
```

means 2.775558×10^{-17}.

Sol. Let f_X be a PDF of X, then

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_0^1 x \cdot \frac{2x}{3} dx + \int_1^3 x \cdot \frac{3-x}{3} dx \\
&\approx 2/9 + 10/9 = 4/3,
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
&= \int_0^1 x^2 \cdot \frac{2x}{3} dx + \int_1^3 x^2 \cdot \frac{3-x}{3} dx \\
&\approx 1/6 + 2 = 13/6,
\end{aligned}
$$

and

$$Var(X) = E(X^2) - [E(X)]^2 \approx 13/6 - (4/3)^2 = 7/18.$$