# 「確率分布と統計的な推測」〜補充 ver1.13



拙著「確率分布と統計的な推測」では扱っていない確率密度関数についてまとめたプリントです。センター試験数学II・B の選択問題で確率分布を選択するためには是非読んでおいて下さい。

# §1.　確率密度関数の基本

## 1.1　確率密度関数の定義

この節の概要
相対度数を表すヒストグラムを一般化して，**確率密度関数**を定義する。

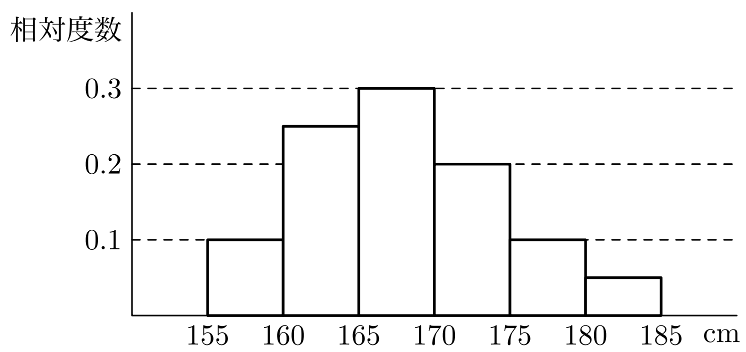
### 1.1.1　相対度数のヒストグラムから確率密度関数へ

ある高校の 3 年生 100 人の身長は次のようになった。

| 身長の階級 (cm) | 度数 | 相対度数 |
|---|---|---|
| 155 以上 160 未満 | 10 | 0.1 |
| 160 〜 165 | 25 | 0.25 |
| 165 〜 170 | 30 | 0.3 |
| 170 〜 175 | 20 | 0.2 |
| 175 〜 180 | 10 | 0.1 |
| 180 〜 185 | 5 | 0.05 |
| 合計 | 100 | 1.00 |

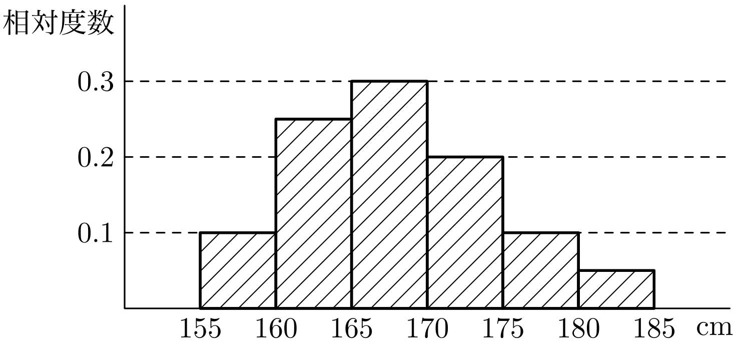
この場合の「度数」は人数，**「相対度数」は全体の人数 100 人に対する割合**だ。

この相対度数を**ヒストグラム**で表すと次のようになる。(ヒストグラムは数学 I「データの分析」で扱った柱状グラフだ。)

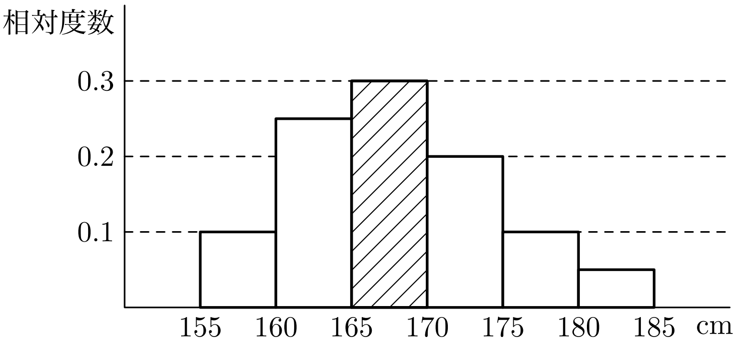


さて，このヒストグラムを次のように解釈してみよう。

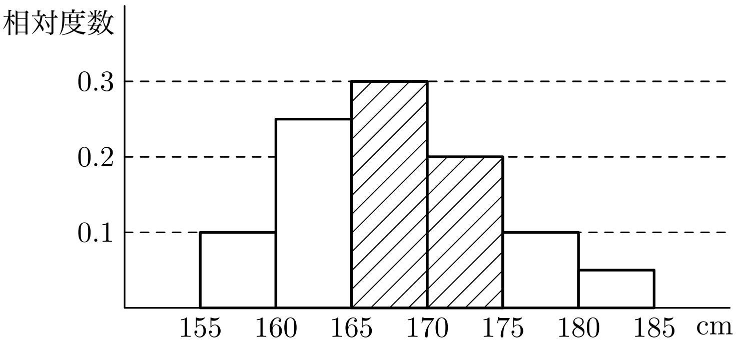
step1. ヒストグラム全体の面積 (次の図の斜線部) を 1 とする。**「1」は相対度数の合計**のことだ。



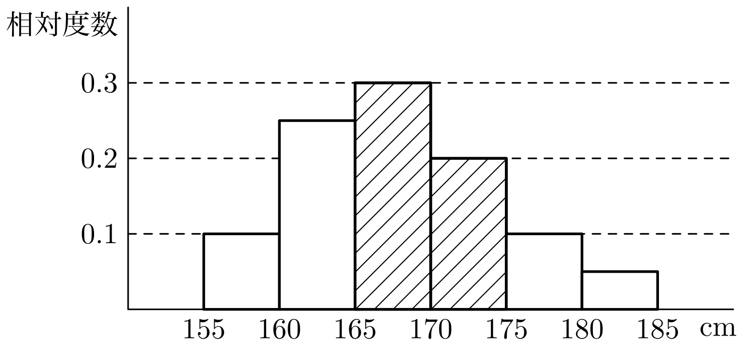

step2. すると，例えば「165cm 以上 170cm 未満」の相対度数 (全体に対する割合) である 0.3 は，次の図の**斜線部の面積**を表すことになる。




step3. さらに例えば「165cm 以上 175cm 未満」の相対度数 (全体に対する割合) である $0.3 + 0.2 = 0.5$ は，次の図の**斜線部の面積**を表すことになる。



step4. ではこの高校の 3 年生 100 人から無作為に一人を選ぶとき，「165cm 以上 175cm 未満」である**確率**はどうなるだろう。それは「165cm 以上 175cm 未満」の相対度数 (全体に対する割合) であるから，次の図の**斜線部の面積**で表されることになる。(先ほどの図と同じ。(^_^;))

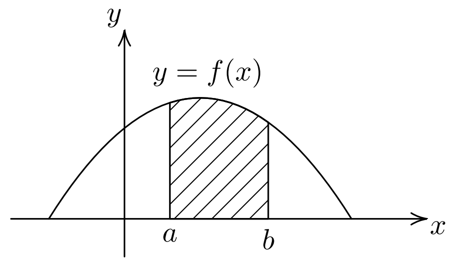



以上を参考に，**確率密度関数** (相対度数に対するヒストグラムを一般化したもの) を次のように定義する。

**確率密度関数の定義**

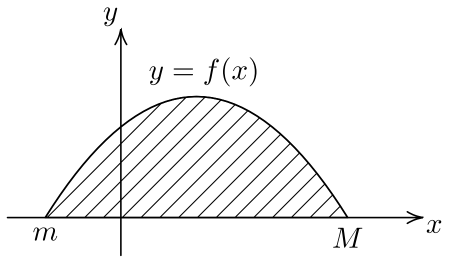
次の性質を満たす $f(x)$ を，確率変数 $X$ に対する**確率密度関数**という。

- すべての実数 $x$ に対して $f(x) \geqq 0$ (ヒストグラムの『柱』は横軸より上に伸ばすことに相当)
- $a \leqq X \leqq b$ となる確率 $P(a \leqq X \leqq b)$ は，$y = f(x)$ のグラフと $x$ 軸，直線 $x = a$，$x = b$ で囲まれた斜線部 (もちろん $a \leqq x \leqq b$ の部分) の面積で表される。次の図はその例。(上記の step4 に相当)

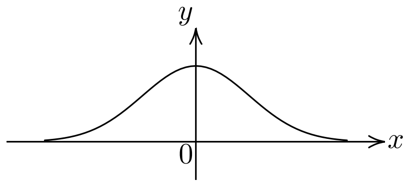



この確率変数 $X$ はとり得る値が連続して変化するから，**連続型確率変数**と言う。

$X$ の最小値を $m$，最大値を $M$ とすると $P(m \leqq X \leqq M) = 1$ (全事象の確率は 1 ということ) であるから，上の例では次の図の斜線部全体の面積が 1 になる。このことは**確率密度関数を決定する問題で手がかり**になる。



標準正規分布の分布曲線 (次の図) は $y = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ のグラフであったが (「確率分布と統計的な推測」第 1 章第 6 節参照)，この関数も確率密度関数である。というか，確率密度関数の最も重要な例である。




### 1.1.2　確率密度関数を考える目的

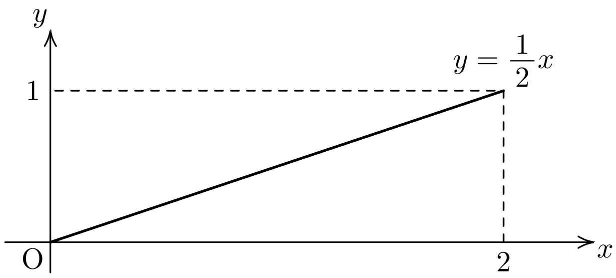
図形の面積は積分で求められることを数学 II の「微分・積分」で学ぶ。この微分・積分というのはものすごく強力な計算方法であるから，これを利用すると統計についてよく分かるはずだ。

それが確率密度関数を導入する目的なのだ。

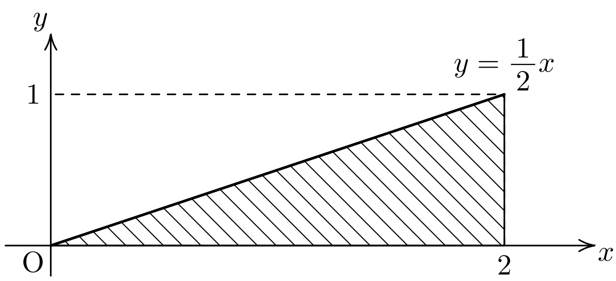
標準正規分布の様々な性質もそれによって解明されたのだ・・・ということの詳しい説明はこのプリントの範囲を超えてしまう。是非，大学で学んで欲しい。

### 1.1.3　確率密度関数の具体例 1

$X$ のとり得る範囲は $0 \leqq X \leqq 2$ であり，その確率密度関数を $y = \dfrac{1}{2}x$ とする。



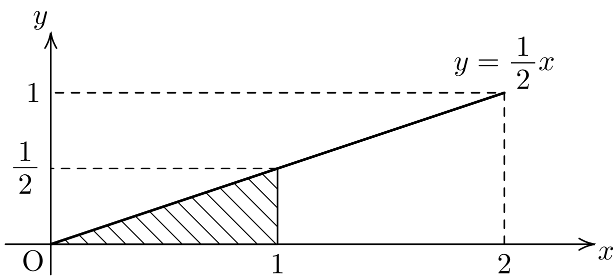

(1) $0 \leqq X \leqq 2$ となる確率は 1 であり，斜線部の面積である。



(2) $0 \leqq X \leqq 1$ となる確率は，次の図の斜線部の面積であるから，

$$\frac{1}{2}\cdot 1\cdot \frac{1}{2}=\frac{1}{4}$$

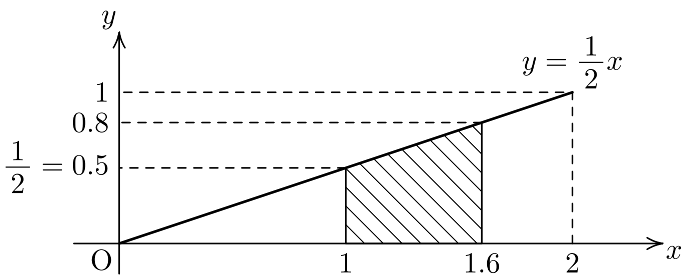
となる。




(3) $1 \leqq X \leqq 1.6$ となる確率は，次の図の斜線部の面積であるから，台形の面積公式より

$$\frac{1}{2}(0.5+0.8)\cdot 0.6=0.39$$

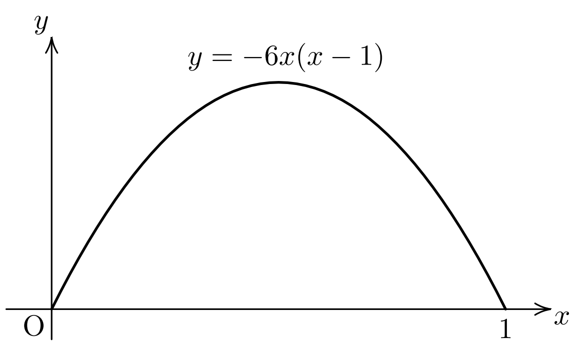
となる。




### 1.1.4 確率密度関数の具体例 2

今度は，面積の計算に積分を使う例を見よう。

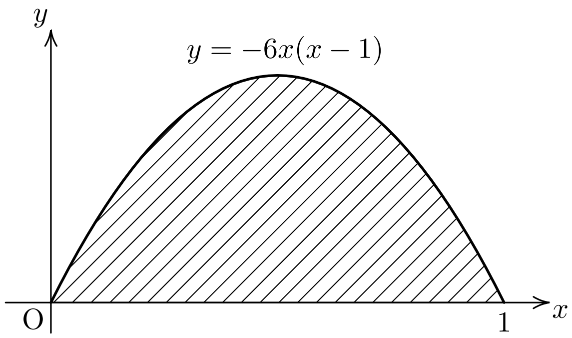
$X$ のとり得る範囲は $0 \leqq X \leqq 1$ であり，その確率密度関数を $y=-6x(x-1)$ とする。



(1) $0 \leqq X \leqq 1$ となる確率は 1 であり，斜線部の面積である。

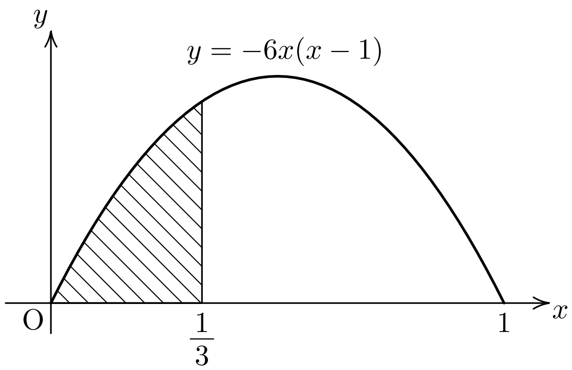
実際，公式 $\displaystyle\int_a^b (x-a)(x-b)\,dx = -\frac{1}{6}(b-a)^3$ を用いて

$$\int_0^1 -6x(x-1)\,dx = \frac{6}{6}(1-0)^3 = 1$$



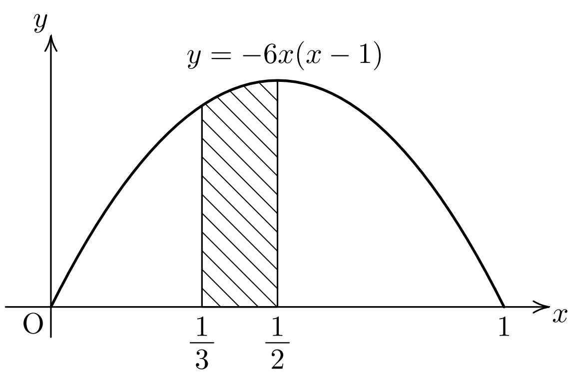

(2) $0 \leqq X \leqq \dfrac{1}{3}$ となる確率は，次の図の斜線部の面積であるから，

$$\int_0^{\frac{1}{3}} -6x(x-1)\,dx = \Big[-2x^3+3x^2\Big]_0^{\frac{1}{3}} = -\frac{2}{27}+\frac{1}{3} = \frac{7}{27}$$




(3) $\dfrac{1}{3} \leqq X \leqq \dfrac{1}{2}$ となる確率は，次の図の斜線部の面積であるから

$$\int_{\frac{1}{3}}^{\frac{1}{2}} -6x(x-1)\,dx = \Big[-2x^3+3x^2\Big]_{\frac{1}{3}}^{\frac{1}{2}} = -\frac{1}{4}+\frac{3}{4}-\left(-\frac{2}{27}+\frac{1}{3}\right) = \frac{13}{54}$$



### 1.1.5　練習問題

**練習問題 1**

確率変数 $X$ のとり得る範囲は $0 \leqq X \leqq 2$ であり，その確率密度関数は

$$f(x) = ax(2-x) \quad (a \text{ は正の定数})$$

と表されている。このとき，次の問いに答えよ。

(1) $0 \leqq X \leqq 2$ となる確率は

$$\int_0^2 f(x)\,dx = \boxed{\text{ア}}$$

となるから，$a = \dfrac{\boxed{\text{イ}}}{\boxed{\text{ウ}}}$ である。

(2) $0 \leqq X \leqq \dfrac{1}{2}$ となる確率は $\dfrac{\boxed{\text{エ}}}{\boxed{\text{オカ}}}$ である。

また，$\dfrac{1}{2} \leqq X \leqq 1$ となる確率は $\dfrac{\boxed{\text{キク}}}{\boxed{\text{ケコ}}}$ である。

**解 答**

(1) $0 \leqq X \leqq 2$ となる確率は，全事象の確率であるから

$$\int_0^2 f(x)\,dx = \overset{ア}{\boxed{\mathbf{1}}}$$

よって

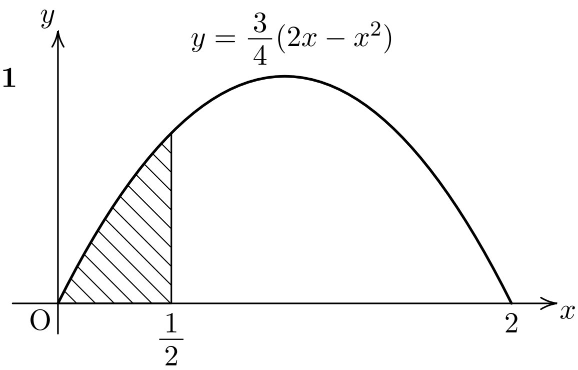
$$\int_0^2 \underbrace{-ax(x-2)}_{f(x)}\,dx = \frac{a}{6}(2-0)^3 = \frac{4}{3}a = 1$$

したがって，$a = \dfrac{\overset{イ}{\boxed{\mathbf{3}}}}{\underset{ウ}{\boxed{\mathbf{4}}}}$

(2) (1) より $f(x) = \dfrac{3}{4}x(2-x) = \dfrac{3}{4}(2x-x^2)$

$0 \leqq X \leqq \dfrac{1}{2}$ となる確率は次の図 1 の斜線部の面積である。

**図 1**

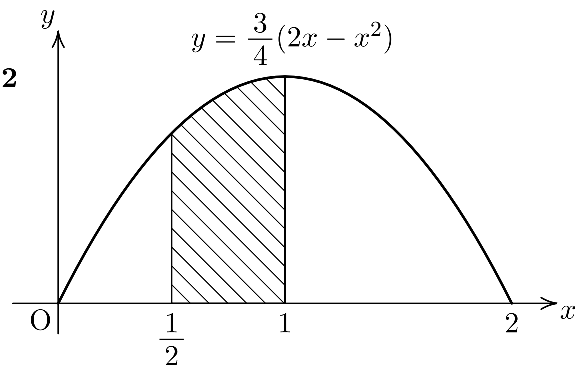



つまり

$$\int_0^{\frac{1}{2}} f(x)\,dx = \int_0^{\frac{1}{2}} \frac{3}{4}(2x-x^2)\,dx = \frac{3}{4}\left[x^2 - \frac{1}{3}x^3\right]_0^{\frac{1}{2}} = \frac{\overset{エ}{\boxed{\mathbf{5}}}}{\underset{オカ}{\boxed{\mathbf{32}}}}$$

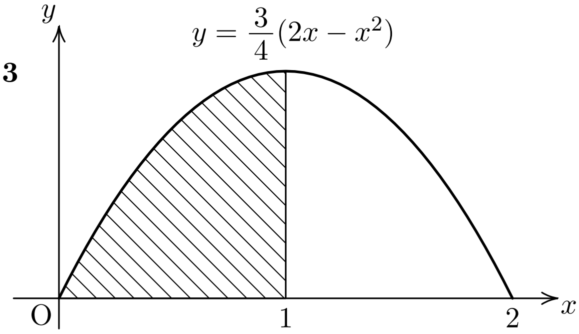
$\dfrac{1}{2} \leqq X \leqq 1$ となる確率は，次の図 2 の斜線部の面積である。

**図 2**



$y = f(x) = \dfrac{3}{4}(2x - x^2)$ のグラフは放物線であり，軸 $x = 1$ について対称であるから次の図 3 の斜線部の面積は $\dfrac{1}{2}$ になる。( $\displaystyle\int_0^2 f(x)\,dx = 1$ の半分！)

**図 3**




以上より，$\dfrac{1}{2} \leqq X \leqq 1$ となる確率は

(図 2 の斜線部の面積) = (図 3 の斜線部の面積) − (図 1 の斜線分の面積) $= \dfrac{1}{2} - \dfrac{5}{32} = \dfrac{\overset{\text{キク}}{\boxed{\mathbf{11}}}}{\underset{\text{ケコ}}{\boxed{\mathbf{32}}}}$

## 1.2 確率密度関数と平均

この節の概要

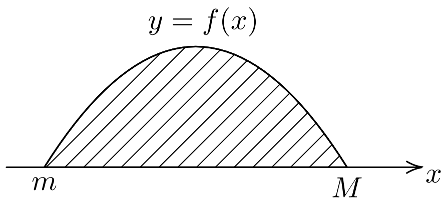
確率密度関数を用いて連続型確率変数の平均を定義する。

### 1.2.1 $\int f(x)\,dx$ という記号の意味

連続型確率変数 $X$ に対する確率密度関数を $f(x)$ とする。
$X$ のとり得る範囲を $m \leqq X \leqq M$ とすると，前節で見たように全事象の確率は

$$\int_m^M f(x)\,dx = 1$$

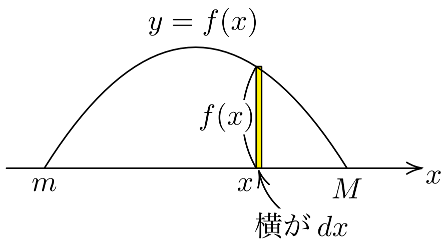
であり，これは $y = f(x)$ のグラフを次の図のようなものだとすると斜線部の面積が 1 となることを意味した。




$\int_m^M f(x)\,dx$ という記号はライプニッツが作った。その意図は

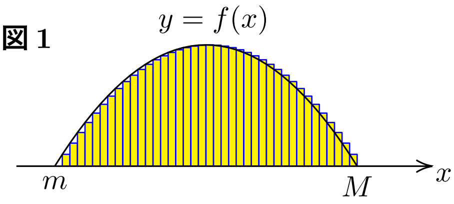
縦が $f(x)$ で横が $dx$　($dx$ とはものすごく 0 に近い数という意味) という長方形 (下の図の黄色い部分のようなもの) の面積 $f(x)\,dx$ をすべて足せ ($\int$)

と言うことだ. $\int$ は「和」を表すラテン語 summa (英語だと sum) の頭文字 s だ。s は当時 $\int$ と書いたのだそうだ。




「このような長方形をすべて足す」というのは，次の図のような長方形全体の面積を足すということだ。(柱の横幅である $dx$ はものすごく 0 に近いものを考えのだが，目に見える程度にしてある。(^^;)

**図 1**




(注．これで $y = f(x)$ のグラフと $x$ 軸とで囲まれた図形の面積が求められる，というのは数学 III で学ぶ**区分求積法**の考え方だ。)

上の図の「柱が並んでいる様子」は**ヒストグラムに見える**だろう。

今考えているヒストグラムは相対度数のヒストグラムなので「1 つの柱に対する $X$ の値」は「柱の左下の $x$(柱の幅 $dx$ が非常に狭いので右下でもよい)」だとし，$X$ がその値になる確率は柱の面積 $f(x)dx$ で表される。(次の図を見よ。)

**図 2**

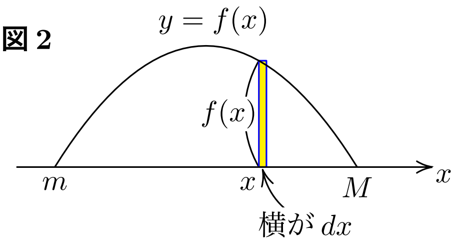



**図 1** のヒストグラム全体の面積は，**図 2** の柱の面積「確率 $f(x)dx$」の合計であるから，全事象の確率 1 になる。つまり

$$\int_m^M f(x)\,dx = 1$$

と言う確率密度関数の性質は，「相対度数のヒストグラム全体の面積は全事象の確率 1」ということをライプニッツの考え方で定積分によって表したものである。

### 1.2.2 確率密度関数から連続型確率変数の平均を定める

確率変数の平均の定義を確認しよう。(『確率分布と統計的な推測』第 1 章第 1 節参照)

**取り得る値が有限個である確率変数 $X$ の平均の定義**

確率変数 $X$ の取り得る値が $x_1, x_2, x_3, \cdots, x_n$ であるとし，

$$P(X = x_k) = p_k \quad (X = x_k \text{となる確率だよ})$$

とする．このとき $X$ の**平均 (平均値，期待値)** は，

$$E(X) = \sum_{k=1}^{n} x_k p_k = x_1 p_1 + x_2 p_2 + \cdots + x_n p_n \qquad \cdots ①$$

つまり，$X$ の平均は

「($X$ の取り得る値) × (その値を取る確率)」をすべてたしたもの ⋯②

となる．

この定義①は連続型確率変数 $X$ には使えない。$X$ の取り得る値が無限にあるからだ。
**そこで②に注目する。**
連続型確率変数 $X$ に対する確率密度関数を $f(x)$ とし，$X$ の取りうる範囲が $m \leqq X \leqq M$ としよう。

**図 2(再掲)**

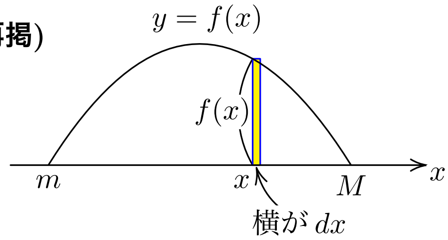



**図 2** のように $X$ の値が $x$ になる確率は $f(x)dx$ となるから，「($X$ の取り得る値) × (その値を取る確率)」とは，$\boldsymbol{xf(x)dx}$ となる。それをすべて足すとは，ライプニッツの考え方により

$$\int_m^M \boldsymbol{xf(x)\,dx}$$

となる。これを連続型確率変数 $X$ の平均の定義とする。

**連続型確率変数の平均の定義**

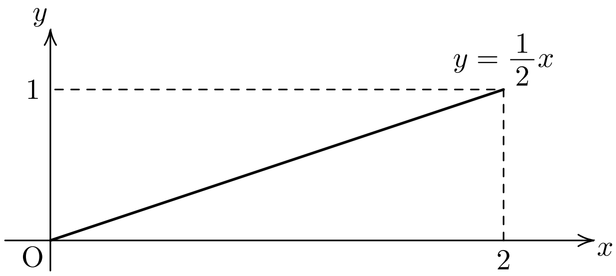
連続型確率変数 $X$ に対する確率密度関数を $f(x)$ とし，$X$ の取りうる範囲が $m \leqq X \leqq M$ とするとき，$X$ の**平均** $E(X)$ を

$$E(X) = \int_m^M xf(x)\,dx$$

と定める。

### 1.2.3 連続型確率変数の平均の例 1

第 1.1.3 節での連続型確率変数 $X$ を考えよう。つまり，$X$ のとり得る範囲は $0 \leqq X \leqq 2$ であり，その確率密度関数を $y = \dfrac{1}{2}x$ とする。

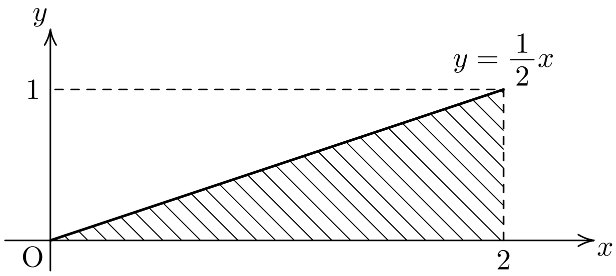



この $X$ の平均 $E(X)$ は

$$E(X) = \int_0^2 x \cdot \frac{1}{2}x\,dx = \left[\frac{1}{6}x^3\right]_0^2 dx = \frac{4}{3} \quad (\text{答})$$

### 1.2.4 平均の計算テクニック 1〜三角形の重心

上記の場合，確率密度関数のグラフとと $x$ 軸で**はさまれる部分**が三角形になっている。(次図の斜線部)




そして，上で求めた $E(X) = \dfrac{4}{3}$ は**この三角形の重心の $x$ 座標**である。実際，三角形の頂点が $(0, 0)$，$(2, 0)$，$(2, 1)$ であるから重心の $x$ 座標は

$$\frac{0+2+2}{3} = \frac{4}{3}$$

となり，確かに $E(X)$ に等しい。

このことは次の事実から説明できる。

平均 $E(X)$ の図形的意味

$X$ の確率密度関数が $f(x)$ であり，$X$ の取り得る範囲が $m \leqq X \leqq M$ のとき，**領域 $D$「$0 \leqq y \leqq f(x)$ かつ $m \leqq X \leqq M$」の重心の $x$ 座標が $E(X)$ である。**

今の例では領域 $D$ は上図の斜線部の三角形なので，$E(X)$ がその重心の $x$ 座標に等しくなったのだ。

「三角形の重心は分かるけど，一般的な図形の重心とは何なの？」と思うだろうが，実は領域 $D$ の重心の $x$ 座標の定義が平均 $E(X)$ なのだ。それを説明してもセンター試験では役に立たないのでここでは説明しない。(^^;

センター試験で役に立つ可能性があるのは重心に関する次の性質だ。

**重心の性質**

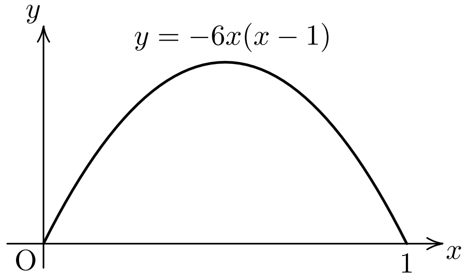
- 三角形の重心の $x$ 座標は，3 頂点を $(x_k, y_k)$　$(k=1,2,3)$ として，$\dfrac{x_1+x_2+x_3}{3}$ である。
- 直線 $x=a$ について対称な図形の重心は，対称軸 $x=a$ 上にある。

領域 $D$ が**三角形**となる場合は，重心の座標は重心の公式により簡単に求められる。**この性質は 2017 年センター試験数学 II・B の確率分布の問題に使うと計算が簡単になった。**

ただし，三角形**以外**の多角形の重心は，三角形のようには求められないことに注意しよう。領域 $D$ が三角形以外の多角形の場合は $E(X)$ の定義にしたがって積分を計算しよう。

### 1.2.5　連続型確率変数の平均の例 2

第 1.1.4 節での連続型確率変数 $X$ を考えよう。つまり，$X$ のとり得る範囲は $0 \leqq X \leqq 1$ であり，その確率密度関数を $y=-6x(x-1)$ とする。

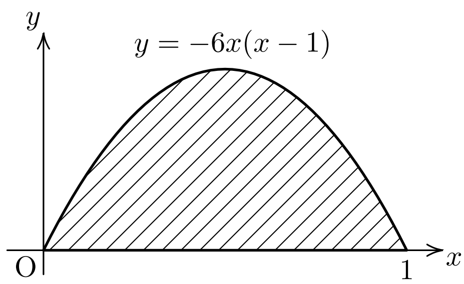



このとき $X$ の平均 $E(X)$ は

$$E(X) = \int_0^1 x \times \{-6x(x-1)\}\,dx = -6\int_0^1 (x^3-x^2)\,dx = -6\left[\frac{x^4}{4}-\frac{x^3}{3}\right]_0^1 = \frac{1}{2} \quad (\text{答})$$

### 1.2.6　平均の計算テクニック 2〜対称性の利用

上記の例の場合，p.14 で説明した領域 $D$ は次の図の斜線部であり，放物線 $y=-6x(x-1)$ の軸 $x=\dfrac{1}{2}$ について対称である。



したがって領域 $D$ の重心は直線 $x = \dfrac{1}{2}$ の上にあるから

$$E(X) = (D \text{の重心の } x \text{ 座標}) = \frac{1}{2}$$

とすぐに分かる。

### 1.2.7 練習問題

**練習問題 2**

確率変数 $X$ のとり得る範囲は $0 \leqq X \leqq 1$ であり，その確率密度関数は

$$f(x) = ax^2(1-x) \quad (a \text{ は正の定数})$$

と表されている。このとき，次の問いに答えよ。

(1) $0 \leqq X \leqq 1$ となる確率は

$$\int_0^1 f(x)\,dx = \boxed{\text{ア}}$$

となるから，$a = \boxed{\text{イウ}}$ である。

(2) $X$ の平均 $E(X)$ は $\dfrac{\boxed{\text{エ}}}{\boxed{\text{オ}}}$ となる。

**解 答**

(1) $0 \leqq X \leqq 1$ となる確率は，全事象の確率であるから

$$\int_0^1 f(x)\,dx = \overset{\text{ア}}{\boxed{\mathbf{1}}}$$

よって

$$\int_0^1 \underbrace{a(x^2 - x^3)}_{f(x)}\,dx = a\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1 = \frac{a}{12} = 1$$

したがって，$a = \overset{\text{イウ}}{\boxed{\mathbf{12}}}$

(2) (1) より $f(x) = 12x^2(1-x) = 12(x^2 - x^3)$

$X$ の平均 $E(X)$ は

$$E(X) = \int_0^1 xf(x)\,dx = \int_0^1 12(x^3 - x^4)\,dx = 12\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1 = \frac{\overset{\text{エ}}{\boxed{\mathbf{3}}}}{\underset{\text{オ}}{\boxed{\mathbf{5}}}}$$

**練習問題 3**

$a$ を正の定数とし，確率変数 $X$ のとり得る範囲は $-1 \leqq X \leqq a$ であり，その確率密度関数は

$$f(x) = \begin{cases} \dfrac{1}{2}(x+1) & (-1 \leqq x \leqq 0) \\ \dfrac{1}{2a}(a-x) & (0 \leqq x \leqq a) \end{cases}$$

と表されている。このとき，次の問いに答えよ。

(1) $-1 \leqq X \leqq a$ となる確率は

$$\int_{-1}^{a} f(x)\,dx = \boxed{\text{ア}}$$

となるから，$a = \boxed{\text{イ}}$ である。

(2) $X$ の平均 $E(X)$ は $\dfrac{\boxed{\text{ウ}}}{\boxed{\text{エ}}}$ となる。

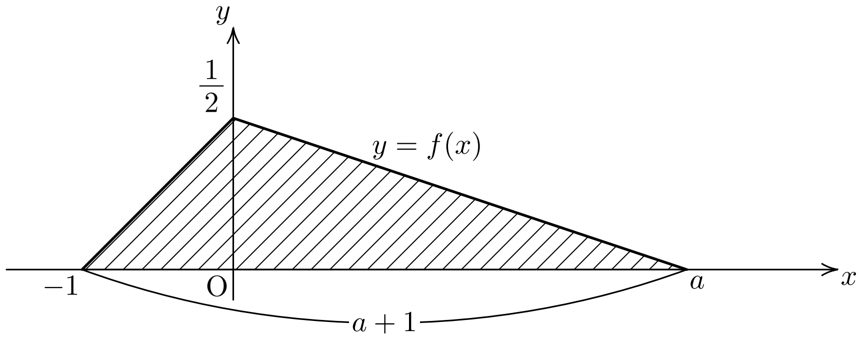
(3) 確率変数 $Y$ を $Y = 3X + 1$ とすると，$Y$ の平均は $\boxed{\text{オ}}$ である。

**解 答**

(1) $-1 \leqq X \leqq a$ となる確率は，全事象の確率であるから

$$\int_0^1 f(x)\,dx = \overset{ア}{\boxed{\mathbf{1}}}$$

これは次の図の斜線部の三角形の面積である。




よって

$$\frac{1}{2}(a+1)\frac{1}{2} = 1$$

したがって，$a = \overset{イ}{\boxed{\mathbf{3}}}$

(2) 斜線部が三角形であるから，$X$ の平均 $E(X)$ はその重心の $x$ 座標である。よって

$$E(X) = \frac{-1+0+3}{3} = \frac{\overset{ウ}{\boxed{\mathbf{2}}}}{\underset{エ}{\boxed{\mathbf{3}}}}$$

(3) $Y = 3X + 1$ より

$$E(Y) = E(3X+1) = 3E(X) + 1 = 3 \cdot \frac{2}{3} + 1 = \overset{オ}{\boxed{\mathbf{3}}}$$