第 6 章 連続分布

浅川 伸一

2006 年 05 月 18 日

課題に対するコメント。著しく間違っている回答は見受けられなかった。

インターネットの普及により、必要なものは必要とするときに検索できる。したがって、以前行われていたような計算練習、公式の暗記は不必要である。

1. 我々はワードプロセッサの普及により、文書を清書するという作業から開放された。
2. 我々は、Web の検索エンジンを駆使することにより、枝葉末節に至る記銘力から開放された。
3. 我々は統計を学ぶ上で、筆算による計算ミスから開放されている。

後に残された、より本質的な概念の獲得、より深い知識を身に付けるべきである。

要求される能力としては、

1. どのように検索したら自分の探している情報を検索できるか
2. 溢れる情報の中から真に価値あるものを見つけ出すことができるか、

が必要とされると考える。

記憶すべき公式が多いと諦めてはいけない。まずは全体を俯瞰して本質を捉えることを考えて欲しい。もちろん公式を覚えていれば、理解を促す手助けにはなる。しかし、細かいことは気にせずに、まずは全体像をつかむことに専心すべきである。

6 連続型の分布

これまで扱ってきた分布は、確率変数が離散的な場合であった。コインの裏表など。ところが世の中には連続量で記述した方が良い場合も存在する。例えば身長の分布。連続型の分布の場合、確率密度関数 probability desnsity function : pdf $f(x)$ はすべての x について

$$f(x) \ > 0, \text{かつ} \int_{-\infty}^{\infty} f(x)\, dx \leq 0 \tag{1}$$

が成り立っていなければならない。これは二項分布 $B(n,p)$ において、すべての確率が 0 より大きく、$\sum_x B(x,p)=1$ であることと対応する。

連続型の確率変数の期待値は、

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

となる。ちなみに離散型の場合は $\sum x\ f(x)$ であった。さらに分散は

$$V(x)=\int (x-\mu)^2 f(x)\ dx \tag{3}$$

である。

離散型の分布関数 cdf は、

$$F(x_i)=P(x\le x_i)=f(x_1)+f(x_2)+\cdots+f(x_i) \tag{4}$$

と表記された。このことと同じく連続型の確率変数の場合は積分記号を用いて

$$F(y)=P(x\le y)=\int_{-\infty}^{y} f(x)\ dx \tag{5}$$

のように表される。離散的な分布においては、確率密度関数 pdf の値が直接、その変数 x の確率を表現していたが、連続型の分布の場合は、ある一定区間の積分した値として表現されることに注意が必要である。

連続型の分布においては、確率密度関数 pdf $f(x)$ を積分すると、累積分布関数 cdf $F(x)$ となり、反対に累積分布関数を微分すると確率密度関数となる。

$$f(x)=\frac{d}{dx}F(x) \tag{6}$$

6.1　三角分布

例題

問: 関数 $y=f(x)=cx$ が区間 $[0,1]$ において確率密度関数 pdf となるような c を定めよ。このときの確率密度関数 pdf と累積分布関数 cdf を描け。さらにこの分布の期待値と分散を求めよ。

答: ある関数が定義された区間で確率密度関数となるためには、その区間内で確率密度関数を積分すると 1 になることを利用する。

$$\begin{aligned}\int_0^1 c\ x\ dx &= \left[\frac{1}{2}\ c\ x^2\right]_{x=0}^{1} && (7)\\ &= \frac{1}{2}\ c\ 1^2-\frac{1}{3}\ c\ 0^2 && (8)\\ &= \frac{1}{2}\ c=1 && (9)\end{aligned}$$

より $c=2$ となるので密度関数 pdf は

$$f(x) = 2\ x \tag{10}$$

である。cdf は pdf を積分して

$$F(x) = \int 2\ x\ dx = x^2 \tag{11}$$

となる。期待値は

$$\begin{aligned} E(x) = \int_0^1 x\ f(x)\ dx &= \int_0^1 x\ 2\ x\ dx & (12)\\ &= \int_0^1 2\ x^2\ dx & (13)\\ &= \left[\frac{2}{3}x^3\right]_{x=0}^1 & (14)\\ &= \frac{2}{3} & (15) \end{aligned}$$

分散は

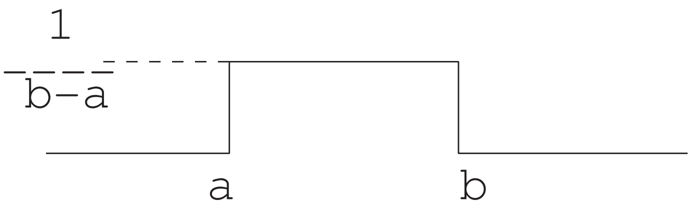
$$\begin{aligned} V(x) &= \int_0^1\ (x-\mu)^2\ \ f(x)\ dx & (16)\\ &= \int_0^1 x^2\ f(x)\ dx - \{E(x)\}^2 & (17)\\ &= \int_0^1 x^2\ 2x\ dx - \{E(x)\}^2 & (18)\\ &= \left[\frac{2}{3}\ x^3\right]_{x=0}^1 - \left(\frac{2}{3}\right)^2 & (19)\\ &= \frac{2}{3} - \left(\frac{2}{3}\right)^2 & (20)\\ &= \frac{2}{9} & (21) \end{aligned}$$

6.2 一様分布 uniform distribution

区間 $[a, b]$ において密度関数が一定の値をとる場合一様関数という。(エクセルの rand() 関数がこれに該当する) このとき pdf の値は一定であり、定義からその値は

$$f(x) = c = \int_a^b c\ dx = c\,[x]_{x=a}^b = c \cdot (b-a) = 1 \tag{22}$$

$f(x) = \dfrac{1}{b-a}$ となる。cdf は pdf を積分して



$$F(x) = \int_{x=a}^{x} \frac{1}{b-a}\,dx = \frac{1}{b-a}\,[x]_a^x \tag{23}$$

$$= \frac{x-a}{b-a} \tag{24}$$

期待値は

$$E(x) = \int_{x=a}^{b} x\,\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b \tag{25}$$

$$= \frac{1}{b-a}\left[\frac{1}{2}b^2 - \frac{1}{2}a^2\right] \tag{26}$$

$$= \frac{1}{2}\frac{1}{b-a}\left[(b-a)(b+a)\right] \tag{27}$$

$$= \frac{b+a}{2} \tag{28}$$

分散は

$$V(x) = \int_a^b x^2 - (fracb+a2)^2 \tag{29}$$

$$= \frac{1}{b-a}\left[\frac{1}{3}x^3\right]_a^b - (fracb+a2)^2 \tag{30}$$

$$= \frac{1}{b-a}\frac{1}{3}\left(b^3-a^3\right) - (fracb+a2)^2 \tag{31}$$

$$= \frac{1}{3}\frac{1}{b-a}(b-a)\left(b^2+ba+a^2\right) - (fracb+a2)^2 \tag{32}$$

$$= \frac{1}{3}\left(b^2+ba+a^2\right) - \frac{1}{4}\left(a^2+2ab+b^2\right) \tag{33}$$

$$= \frac{1}{12}\left(4b^2+4ba+4a^2-3a^2-3\times 2ab-3b^2\right) \tag{34}$$

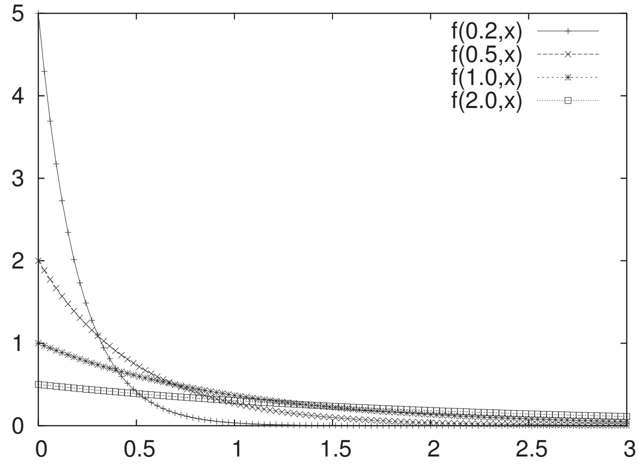
$$= \frac{1}{12}\left(b^2+a^2-2ab\right) \tag{35}$$

$$= \frac{(b-a)^2}{12} \tag{36}$$

6.3 指数分布 exponential distribution

$$f(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}} \tag{37}$$

このとき、β は指数分布のパラメータとよばれ、ゼロより大きい値をとる。ここに、期待値 (平均) は β、分散は β^2 である。累積分布関数は $1-e^{(-\frac{x}{\beta})}$



となる。なお、指数分布の定義式を $f(x)=\theta e^{-\theta x}$ と表している文献もある。この場合 $\theta=1/\beta$ という関係がある。

pdf を 0 から ∞ までの区間で積分して 1 になることを確認する。

$$\int_0^\infty \theta e^{-\theta x} dx = \theta \int_0^\infty e^{-\theta x}\, dx \tag{38}$$

$$= \theta \left[-\frac{1}{\theta} e^{-\theta x}\right]_{x=0}^{\infty} \tag{39}$$

$$= \theta \left[-\frac{1}{\theta} e^{-\theta\times\infty} - \left(-\frac{1}{\theta}\right) e^{\theta\times 0}\right] \tag{40}$$

ここで指数関数のグラフからも $e^{-\theta\times\infty}=0$、$e^{\theta\times 0}=1$ となることが予想されるので、

$$\int_0^\infty \theta e^{-\theta x} dx = \theta\left[\frac{1}{\theta}\right] = 1 \tag{41}$$

となる。次に累積分布関数を求める。

$$F(x) = \int_0^x \theta e^{-\theta x}\, dx \tag{42}$$

$$= \left[-e^{\theta x}\right]_{x=0}^{x} \tag{43}$$

$$= -e^{-\theta x} - \left(-e^{-\theta\times 0}\right) \tag{44}$$

$$= 1 - e^{-\theta x} \tag{45}$$

期待値 $E(x)$ は

$$
\begin{aligned}
E(x) &= \int_0^\infty x\ e^{-\theta x}\ dx && (46)\\
&= \left[-\frac{1}{\theta}e^{-\theta x}\right]_{x=0}^\infty && (47)\\
&= -\frac{1}{\theta}e^{-\theta\times\infty} - \left(-\frac{1}{\theta}e^{-\theta\times 0}\right) && (48)
\end{aligned}
$$

ここで再び $e^{-\theta\times\infty}=0$、$e^{\theta\times 0}=1$ であることを考慮して

$$
\begin{aligned}
E(x) &= \int_0^\infty x\ e^{-\theta x}\ dx && (49)\\
&= -\frac{1}{\theta}\times 0 - \left(-\frac{1}{\theta}\times 1\right) && (50)\\
&= \frac{1}{\theta} && (51)
\end{aligned}
$$

分散は

$$
V(x) = \int_0^\infty x^2\ E(x) - (E(x))^2 = \int_0^\infty x^2\ \theta e^{-\theta x}\ dx - \left(\frac{1}{\theta}\right)^2 \tag{52}
$$

ここで部分積分を用いて

$$
\begin{aligned}
f(x) = x^2 &\quad\rightarrow\quad f'(x) = 2x && (53)\\
g'(x) = \theta\ e^{-\theta x} &\quad\rightarrow\quad g(x) = -e^{-\theta x} && (54)
\end{aligned}
$$

であるから、直上の式の右辺第二項は

$$
\begin{aligned}
&\int_0^\infty x^2\ \theta e^{-\theta x}\ dx && (55)\\
= &\left[x^2\left(-e^{-\theta x}\right)\right]_{x=0}^\infty - \int_0^\infty 2x\times\left(-e^{-\theta}\right)\ dx - \left(\frac{1}{\theta}\right)^2 && (56)
\end{aligned}
$$

この式 (56) の第一項は積分の概念を拡張しなければならないが、ここは数値計算で逃げる。式 (56) の第ニ項

$$
\int_0^\infty 2x\times\left(-e^{-\theta}\right)\ dx \tag{57}
$$

だけ取り出して、再び部分積分をする。定数 2 を除けば

$$
\begin{aligned}
f(x) = x &\quad\rightarrow\quad f'(x) = 1 && (58)\\
g'(x) = \theta\ e^{-\theta x} &\quad\rightarrow\quad g(x) = -\frac{1}{\theta}e^{-\theta x} && (59)
\end{aligned}
$$

なので、式 (57) は

$$
\left[x\left(-\frac{1}{\theta}e^{-\theta x}\right)\right]_{x=0}^\infty - \int_0^\infty -\frac{1}{\theta}e^{-\theta x}\ dx \tag{60}
$$

ここで、前と同じ様に積分概念を拡張して、式 (60) の第一項は 0 第二項は

$$-\int_0^\infty \frac{1}{\theta}e^{-\theta x}dx = +\frac{1}{\theta}\int_0^\infty e^{-\theta x}dx \tag{61}$$

ここで上式の積分は密度関数 pdf の $1/\theta$ であるからこの式全体では $\left(\frac{1}{\theta}\right)^2$ である。この結果を式 (56) に戻すと

$$\begin{aligned}
& \left[x^2\left(-e^{-\theta x}\right)\right]_{x=0}^\infty - \int_0^\infty 2x \times \left(-e^{-\theta}\right)\, dx - \left(\frac{1}{\theta}\right)^2 && (62)\\
= \quad & 0 + 2 \times \left(\frac{1}{\theta}\right)^2 - \left(\frac{1}{\theta}\right)^2 && (63)\\
= \quad & \left(\frac{1}{\theta}\right)^2 && (64)
\end{aligned}$$