数理・データサイエンス教材案（パレート分布と伝送遅延 RTT）

2019 年 7 月 30 日

提案者：松田 修

確率密度関数が$y=\frac{ab^a}{x^{a+1}}$である分布を**パレート分布**$\mathrm{Par}(a,b)$という。

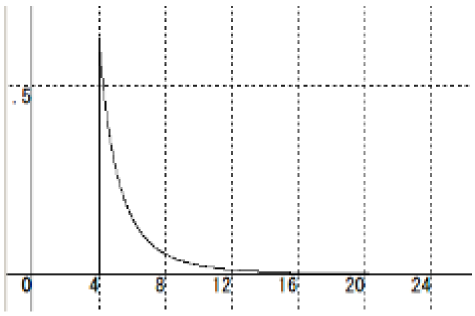
ただし，$a>0, b>0, x\geq b$とする。$\mathrm{Par}(a,b)$の平均は$\frac{ab}{a-1}$である。

パレート分布は，xがbに近いときの確率は高いが，以後，xを大きくしてもyの値はなかなか減少していかないといった特徴を持つ。

下の図は，$a=2.5, b=4.0$としたパレート分布$\mathrm{Par}(2.5,4.0)$を表すグラフで，

$$y=\frac{2.5\times 4^{2.5}}{x^{3.5}}$$

である。



である。 平均μは，以下のようにして得られる。

$$\mu=\int_b^\infty xy\,dx=\int_b^\infty \frac{ab^a}{x^a}dx=-\frac{ab^a}{a-1}\left[\frac{1}{x^{a-1}}\right]_b^\infty=\frac{ab}{a-1}\quad（ただし，a>1）$$

$\mathrm{Par}(2.5,4.0)$の平均は，$\mu=\frac{20}{3}$である。さらに，$\mu_1=\mu+1$として，

$$P(4\leq X\leq \mu_1)=\int_4^{\mu_1}\frac{2.5\times 4^{2.5}}{x^{3.5}}dx\sim 0.803,$$

$$P\big(4\leq X\leq 5(\mu_1-4)\big)=\int_4^{5(\mu_1-4)}\frac{2.5\times 4^{2.5}}{x^{3.5}}dx\sim 0.978$$

$$\frac{P(5\leq X\leq \mu_1)}{P\big(5\leq X\leq 5(\mu_1-5)\big)}=0.821$$

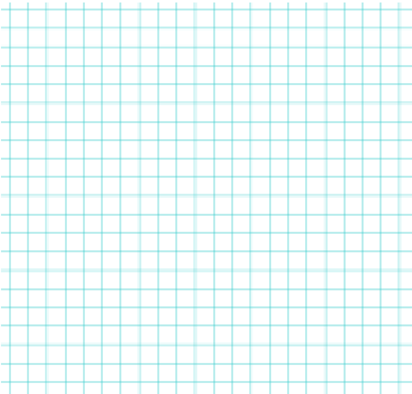
である。これは，$4\leq X\leq 5(\mu_1-4)$の前半の２割区間 $4\leq X\leq \mu_1$の確率$P(4\leq X\leq \mu_1)$が$P(4\leq X\leq 5(\mu_1-4))$の約８割であることを示している。パレート分布は以下のパレートの法則と比較するために用いられることが多い。

パレートの法則：マーケティングにおいて「２：８の法則」とも呼ばれ，顧客全体の２割である優良顧客が全体の売り上げの８割をあげているという法則。

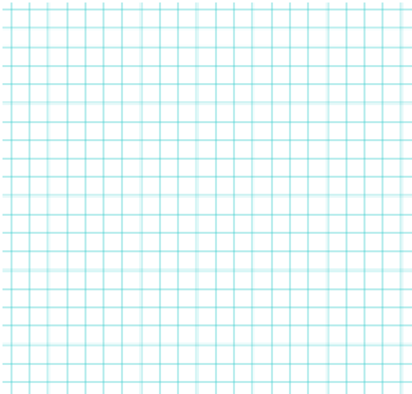
問題１．以下は，あるネット回線の伝送遅延(Round Trip Time)のデータで，0.5msの間隔で測定したものの一部である。表中の相対遅延度数fとは，10.0 ~ 10.01の間の遅延回数を全測定回数で割ったものである。

RTT [ms] の階級値 x	10.0	14.0	18.0	22.0	26.0
相対遅延度数 f	0.093	0.034	0.016	0.009	0.005
$p = f/0.5$					
べき関数 y					
誤差 $\lvert p - y \rvert$					

(1) pの欄を埋めて，データ(x, p)のヒストグラムを図Ａに描け。



図A



図B

(2) (x, p) はべき関数 $y = \frac{C}{x^k}$に従うと考え，誤差$\lvert f - y \rvert < 0.02$となるように，$C$と$k$を決定し， yの欄と，誤差 $\lvert f - y \rvert$ の欄を埋めよ。

(3) 関数 $y = \frac{C}{x^k}$ に従う分布(x, p)は，パレート分布$\mathrm{Par}(a, b)$であることを確かめ，a, bを小数点第２位まで決定せよ。

(4) (3)から得られたパレート分布$\mathrm{Par}(a, b)$のグラフの概形を図Ｂに描け。

(5) 上で得られたパレート分布$\mathrm{Par}(a, b)$について，次を小数第３位まで計算せよ。

$$\mathrm{P}(9.75 \leq X \leq 10.25), \qquad \mathrm{P}(10.25 \leq X \leq 11.75)$$

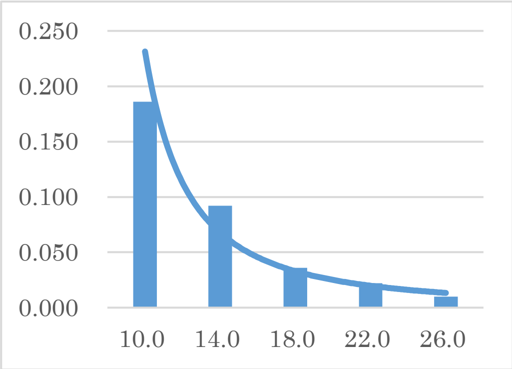
(6) パレート分布$\mathrm{Par}(a, b)$の平均μ，すなわち RTT の平均 μ を求めよ。

(7) パレート分布$\mathrm{Par}(a, b)$の平均μを用いて，次を小数第３位まで計算せよ。

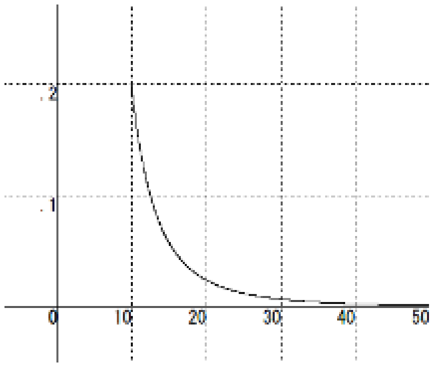
$$\frac{\mathrm{P}(b \leq X \leq \mu + 1)}{\mathrm{P}\big(b \leq X \leq 5(\mu + 1 - b)\big)}$$

解答例

(1)



図A



図B

(2) $(x, y) = (10.0, 0.186), (12.0, 0.108)$を用いて，連立方程式

$$0.186 \times 10.0^k = C,\ \ 0.01 \times 26.0^k = C$$

これより，$k = \dfrac{\log\dfrac{0.186}{0.010}}{\log\dfrac{26.0}{10.0}} \sim 3.059,\ \ C = 0.186 \times 10.0^{3.06} \sim 213.0$

しかし，$x = 14$のとき，$|p - y| > 0.02$となる。$k = 3.02$とすると，$|f - y| < 0.02$を満たす。

よって，$y = \dfrac{213.0}{x^{3.02}}$と推測する。

RTT [ms] の階級値 x	10.0	14.0	18.0	22.0	26.0
相対遅延度数 f	0.093	0.046	0.018	0.011	0.005
$p = f/0.5$	0.186	0.092	0.036	0.022	0.010
べき関数 y	0.204	0.074	0.035	0.019	0.011
誤差 $\lvert p - y\rvert$	0.018	0.018	0.001	0.003	0.001

(3) $a + 1 = k = 3.02,\ ab^a = C = 213$より，$a = 2.02,\ b = \left(\frac{213}{2.02}\right)^{\frac{1}{2.02}} = 10.03$ である。よって，$y = \frac{213}{x^{3.02}}$ はパレート分布Par(2.02,10.03)である。

(5) $\mathrm{P}(9.75 \leq X \leq 10.25) = 213\displaystyle\int_{9.75}^{10.25} x^{-3.02}dx = \frac{213}{-2.02}[x^{-2.02}]_{9.75}^{10.25} = 0.102$

$$\mathrm{P}(21.75 \leq X \leq 22.25) = \frac{213}{-2.02}[x^{-2.02}]_{21.75}^{22.25} = 0.009$$

(6) $\mu = \dfrac{ab}{a-1} = \dfrac{2.02 \times 10.03}{1.02} = 19.86\ \mathrm{ms}$

(7) $\mu+1=20.86,\ 5(\mu+1-b)=54.15$

$$\mathrm{P}(10.03 \le X \le 20.86) = \int_{10.03}^{20.86} \frac{213}{x^{3.02}} \sim 0.772,$$

$$\mathrm{P}(9.85 \le X \le 54.15) = \int_{10.03}^{54.15} \frac{213}{x^{3.02}} \sim 0.967$$

$$\frac{\mathrm{P}(b \le X \le \mu+1)}{\mathrm{P}\big(b \le X \le 5(\mu+1-b)\big)} \sim 0.798$$