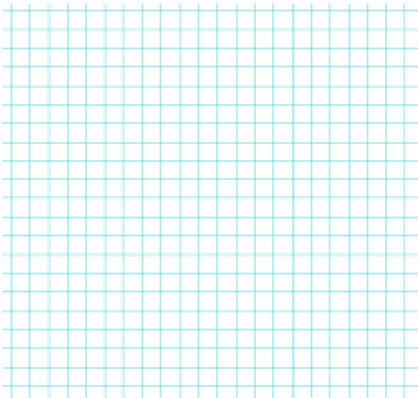
数理・データサイエンス教材案（人気投票とべき関数）

2019 年 7 月 31 日
提案者：松田 修

問題１．ある年，女性タレント 320 人の人気投票が行なわれた。得票総数は 3,260,000 票で，上位の得票率のサンプルデータは以下である。

順位 x	1	20	40	60	80
タンレント	RS	AT	MF	KK	AM
得票率 P	0.0745	0.0102	0.0071	0.0051	0.0040
べき関数値 y					
誤差 $\lvert P-y\rvert$					

(1) データ(x,P)を下の方眼になるべく正確にプロットし，近似曲線を描け。



(2) (x,P)は，べき関数$y=\frac{C}{x^k}$に従うと考え，誤差$\lvert p-y\rvert<0.06$となるように，Cとkを決定し， yの欄と，誤差 $\lvert p-y\rvert$ の欄を埋めよ。

(3) $y=\frac{C}{x^k}$ に従う分布は(x,P)，パレート分布$\mathrm{Par}(a,b)$でないことを示せ。

(4) (2)から得られた関数$y=\frac{C}{x^k}$が確率密度関数となるためのxの範囲を決定せよ。

(5) (x,P)を確率密度関数$y=\frac{C}{x^k}$から得られる分布と考え，次を小数第３位まで計算せよ。

$$\mathrm{P}(0.5\le X\le 1.5),\qquad \mathrm{P}(1.5\le X\le 2.5)$$

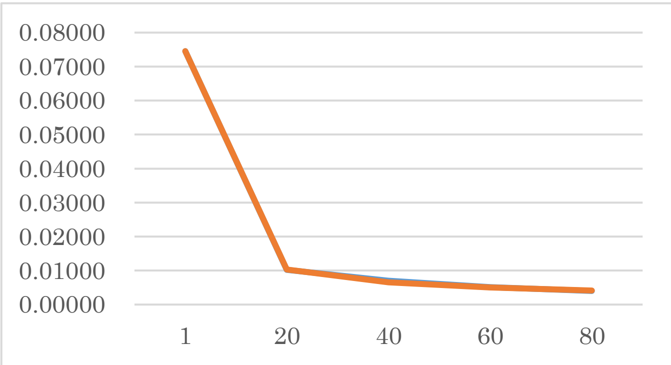
(6) $y=\frac{C}{x^k}$の平均μを求めよ。

(7) $(r-0.5):(u-0.5)=2:10$であるrに対して，確率 $\mathrm{P}(0.5\le X\le r)$ を小数第３位まで計算し，パレートの法則と比較せよ。

解答例

順位 x	1	20	40	60	80
タンレント	RS	AT	MF	KK	AM
得票率 P	0.0745	0.0102	0.00725	0.0051	0.0040
推測得票率 y	0.0745	0.0103	0.0065	0.0050	0.0041
誤差 $\lvert P-y \rvert$	0	0.0001	0.0005	0.0001	0.0001

(1)



(2) $(x,y)=(1,0.0745),(80,0.0040)$を用いて，連立方程式

$$0.0745\times 1^k=C,\quad 0.0040\times 80^k=C$$

これより，$C=0.0745, k=\log_{80}\frac{0.0745}{0.004}\sim 0.667$，すなわち，$y=\frac{0.07454}{x^{0.667}}$を得る。しかし，$|P-y|\geq 0.0006$である。$k=0.66$とすると，$y=\frac{0.07454}{x^{0.66}}$，$|P-y|<0.006$となる。

(3) $a+1=k, C=ab^a$より，$a=-0.34<0$であるため，$y=\frac{0.075}{x^{0.66}}$はパレート分布ではない。

(4) $y=\dfrac{0.07454}{x^{0.66}}$が確率密度関数となるために，$\displaystyle\int_{0.5}^{u}\frac{0.07454}{x^{0.66}}dx=1$となる$u$を求める。

$$\frac{0.07454}{0.34}[x^{0.34}]_{0.5}^{u}=\frac{0.07454}{0.34}(u^{0.34}-0.5)=1 \text{ より，} u=\left(\frac{0.37727}{0.07454}\right)^{\frac{1}{0.34}}\sim 117.86 \text{ となる。}$$

したがって，$1\leq x\leq\left(\frac{0.37727}{0.07454}\right)^{\frac{1}{0.34}}$とすればよい。

(5) $$\mathrm{P}(0.5\leq X\leq 1.5)=\int_{0.5}^{1.5}\frac{0.07454}{x^{0.66}}dx=\frac{0.07454}{0.34}[x^{0.34}]_{0.5}^{1.5}\sim 0.078$$

$$\mathrm{P}(59.5\leq X\leq 60.5)=\frac{0.07454}{0.34}[x^{0.34}]_{59.5}^{60.5}\sim 0.0050$$

(6)

$$\mu = \int_{0.5}^{u} xy\,dx = 0.07454\int_{0.5}^{u} x^{0.34}\,dx = \frac{0.07454}{1.34}[x^{1.34}]_{0.5}^{u} = 33.16$$

(7) $(r-0.5):(u-0.5)=2:10$ より，$r=\dfrac{u-0.5}{5}+0.5\sim 23.97$ である。よって，

$$\mathrm{P}(0.5 \leq X \leq 23.97) = \int_{0.5}^{23.97}\frac{0.07454}{x^{0.66}}dx = \frac{0.07454}{0.34}[x^{0.34}]_{0.5}^{23.97} \sim 0.472$$

である。パレートの法則が，２：８の法則であるのに対し，この分布は全区間$0.5 \leq X \leq u$の前半の２割区間$0.5 \leq X \leq r$が全体の約 4 割 7 分程度であることを示している。