流れ星の出現とベイズ推定

１．流れ星の出現とポアソン分布

以下の表は流れ星の5分ごとの観測データである．

表 1（流れ星のデータ）

n回目	1	2	3	4	5	6	7	8	9	10
個 数（x_n）	1	0	0	2	0	1	3	0	1	0

統計学的にあるデータがポアソン分布に従うための３つの条件があり，それは以下である．
（条件１）希少性がある．すなわち同じタイミングで 2 個以上起こらない．
（条件２）独立性がある．今事象が起こったことは，それ以前の事象が起こったことに依存しない．
（条件３）定常性がある．単位時間内にある事象が起こる確率は常に一定である．

ここでは，流れ星が発生する確率がポアソン分布$P_O(\lambda)$に従うと仮定して，以下に，平均発生個数λを**再尤度推定**で推定し，5分間に流れ星が発生する確率Pを求めよう．

まず，再尤度推定によるλの推定式は以下である．

$$\lambda = \frac{1}{10}\sum_{n=1}^{10} x_n = \frac{1+2+1+3+1}{10} = 0.8$$

したがって，流れ星が発生する確率はポアソン分布$P_O(0.8)$に従い，その確率度密度関数は

$$f(x|0.8) = \frac{e^{-0.8}\cdot 0.8^x}{x!}$$

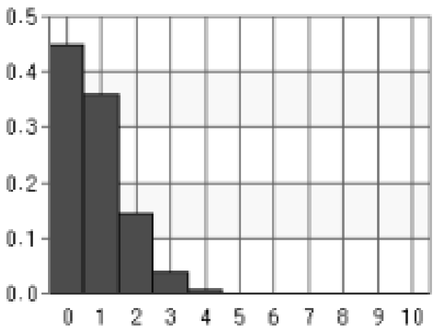
である．よって，

$$f(0|0.8) = \frac{e^{-0.8}\cdot 0.8^0}{0!} = e^{-0.8}$$

である．したがって，5分間に流れ星が発生する確率Pは

$$P = 1 - P(0|0.8) = 0.55$$

となる．



$P_O(0.8)$のグラフ

問題１．以下の表は流れ星の5分ごとの観測データである．流れ星が発生する確率がポアソン分布$P_O(\lambda)$に従うと仮定して，平均発生個数λを再尤度推定で推定し，5分間に2個以上の流れ星が発生する確率Pを小数点以下第３位まで求めよ．

表 2（流れ星のデータ）

n回目	1	2	3	4	5	6	7	8	9	10
個 数（x_n）	2	0	2	1	0	3	4	2	1	1

２．ポアソン分布のベイズ推定

扱う統計の母集団がポアソン分布と考えられる場合にベイズ推定を行うときは，仮にλをその母集団の平均出現回数とし，尤度を

$$f(x|\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}$$

とする．ここで， xは１回の観測で，ある事象が起こる回数である．したがって，確率分布に関するベイズの定理より

$$f(\lambda|x)=\frac{f(x|\lambda)f(\lambda)}{\int_0^\infty f(x|\lambda)f(\lambda)dx} \tag{1}$$

が得られるが，これは１回の観測で得られた結果xをもとに，λに関する確率密度関数$f(\lambda|x)$が得られたことを意味する．

以下において，２回目以降の観測結果を推定につなげるするために式(1)を

$$f\left(\lambda|x^{(1)}\right)=\frac{f\left(x^{(1)}|\lambda\right)f\left(\lambda|x^{(0)}\right)}{\int_0^\infty f\left(x^{(1)}|\lambda\right)f\left(\lambda|x^{(0)}\right)dx} \tag{2}$$

と書き直しておく．そして，

$$f\left(\lambda|x^{(2)}\right)=\frac{f\left(x^{(2)}|\lambda\right)f\left(\lambda|x^{(1)}\right)}{\int_0^\infty f\left(x^{(2)}|\lambda\right)f\left(\lambda|x^{(1)}\right)dx} \tag{3}$$

を定義する．これを，λに関する確率密度関数$f\left(\lambda|x^{(2)}\right)$のベイズ推定という．

以下同様にして，帰納的にλに関する確率密度関数$f\left(\lambda|x^{(n)}\right)$のベイズ推定を定義していく．すなわち，

$$f\left(\lambda|x^{(n)}\right)=\frac{f\left(x^{(n)}|\lambda\right)f\left(\lambda|x^{(n-1)}\right)}{f\left(x^{(n)}\right)},\qquad f\left(x^{(n)}\right)=\int_0^\infty f\left(x^{(n)}|\lambda\right)f\left(\lambda|x^{(n-1)}\right)dx \tag{4}$$

と定義する．

３．ポアソン分布の共役事前分布であるガンマ分布について

母集団がポアソン分布$P_O(\lambda)$に従っているとする．そして事前分布をガンマ分布$Ga(a,b)$とする．このとき，事後分布もガンマ分布となる．このことを説明する．

まず，この場合の確率分布のベイズ推定式は(4)より

$$f\left(\lambda|x^{(1)}\right)=\frac{f\left(x^{(1)}|\lambda\right)}{f\left(x^{(1)}\right)},\quad f\left(x^{(1)}\right)=\int_0^\infty f\left(x^{(1)}|\lambda\right)f(\lambda)d\lambda \tag{5}$$

で，尤度は

$$f\left(x_1|\lambda\right)=\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}$$

である．また，仮定より

$$f(\lambda)=f(\lambda|a,b)=\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-\lambda b} \tag{6}$$

である．よって，

$$f\left(x^{(1)}|\lambda\right)f(\lambda)=\left(\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\right)\left(\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-\lambda b}\right)\propto\lambda^{a+x_1-1}e^{-\lambda(b+1)}$$

を得る．さらに

$$f\left(x^{(1)}\right)=\int_0^\infty f\left(x^{(1)}|\lambda,a,b\right)f(\lambda)d\lambda\propto\int_0^\infty\lambda^{a+x_1-1}e^{-\lambda(b+1)}d\lambda$$

ここで，$t=\lambda\left(b+1\right)$と置くと，

$$f\left(x^{(1)}\right)\propto\int_0^\infty\lambda^{a+x_1-1}e^{-t}\frac{dt}{b+1}=\frac{1}{b+1}\Gamma\left(a+x_1\right)$$

となる．(5)と(6)より

$$f\left(\lambda|x^{(1)}\right)\propto\frac{\lambda^{a+x_1-1}e^{-\lambda(b+1)}}{\frac{1}{b+1}\Gamma\left(a+x_1\right)}=\frac{b+1}{\Gamma\left(a+x_1\right)}\lambda^{a+x_1-1}e^{-\lambda(b+1)}$$

を得る．$f\left(\lambda|x^{(1)}\right)$は，ガンマ分布$Ga\left(a+x_1,b+1\right)$の確率密度関数を意味する．

同様にして

$$f\left(x^{(2)}|\lambda\right)f\left(\lambda|x^{(1)}\right)=\left(\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\right)\left(\frac{b+1}{\Gamma\left(a+x_1\right)}\lambda^{a+x_1-1}e^{-\lambda(b+1)}\right)\propto\lambda^{a+x_1+x_2-1}e^{-\lambda(b+2)}$$

$$f\left(x^{(2)}\right)\propto\int_0^\infty\lambda^{a+x_1+x_2-1}e^{-\lambda(b+2)}d\lambda=\frac{1}{b+2}\Gamma\left(a+x_1+x_2\right)$$

より

$$f\left(\lambda|x^{(2)}\right)\propto\frac{b+2}{\Gamma\left(a+x_1+x_2\right)}\lambda^{a+x_1+x_2-1}e^{-\lambda(b+2)}$$

を得て，$f\left(\lambda|x^{(2)}\right)$はガンマ分布$Ga\left(a+x_1+x_2,b+2\right)$の確率密度関数であることがわかる．

これを繰り返して，

$$f\left(\lambda|x^{(n)}\right) \propto \frac{b+n}{\Gamma(a+\sum x_i)} \lambda^{a+\sum x_i -1} e^{-\lambda(b+n)}$$

が得られ，$f\left(\lambda|x^{(n)}\right)$というガンマ分布$Ga(a+\sum x_i, b+n)$の確率密度関数が得られる．

ポアソン分布$P_O(\lambda)$の事前分布をガンマ分布$Ga(a,b)$とする．このとき，事後パラメータがn個のデータであるときの事後分布は，ガンマ分布 $Ga(a+\sum x_i, b+n)$とできる．すなわち，ポアソン分布の事前分布であるガンマ分布は，共役事前分布である．したがって 事後分布の平均と分散を $$E[\lambda] = \frac{a+\sum x_i}{b+n}, \qquad V[\lambda] = \frac{a+\sum x_i}{(b+n)^2}$$ と考えることができる．

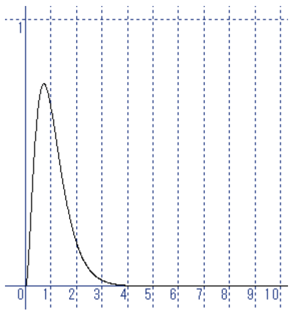
４．λのEPA推定法

ポアソン分布$P_O(\lambda)$の λの推定として扱われるEPA 推定値λ_{epa}（事後分布から得られる平均値λ）とMAP推定値λ_{map}（事後分布の極大値となるλ）を，例を用いて説明する．

（例）ある確率分布の尤度はポアソン分布$P_O(\lambda)$とする．$a=3, b=2.8$のとき，事前分布であるガンマ分布$Ga(3,2.8)$の確率密度関数は，(6)より

$$f(\lambda) = f(\lambda|3,2.8) = \frac{2.8^3}{\Gamma(3)} \lambda^2 e^{-2.8\lambda}$$

で，平均と分散はそれぞれ$E[\lambda] = \frac{a}{b}, V[\lambda] = \frac{a}{b^2} \sim 0.38$であり，グラフは以下となる．



そして，事後パラメータは以下のデータとする．

i	1	2	3	4	5	6	7	8	9	10
x_i	1	0	0	2	0	1	3	0	1	0

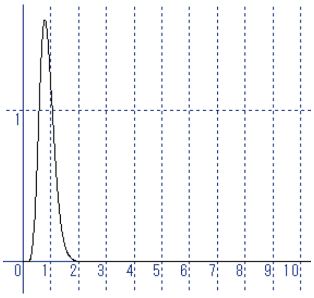
このとき，

$$n = 10, \quad \sum_{i=1}^{10} x_i = 8, \quad a + \sum_{i=1}^{10} x_i = 11, \quad b+n = 12.8$$

そこで事後分布を$Ga(11,12.8)$として考える．このとき，確率密度関数$f\left(\lambda|x^{(10)}\right)$は

$$f\left(\lambda|x^{(10)}\right)=\frac{12.8^{11}}{\Gamma(11)}\lambda^{10}e^{-12.8\lambda}$$

で，このグラフは以下である．



そして，平均と分散はそれぞれ

$$E[\Lambda]=\frac{11}{12.8}\sim 0.86,\quad V[\lambda]=\frac{11}{12.8^2}\sim 0.07$$

である，したがって，EPA 推定値λ_{epa}（事後分布から得られる平均値λ）は

$$\lambda_{epa}=0.86$$

となる．

また，

$$\frac{d}{d\lambda}f\left(\lambda|x^{(10)}\right)=-\frac{1}{12}\lambda^9(63964329\lambda-49972132)e^{-12.8\lambda}$$

これより，MAP推定値λ_{map}（事後分布の極大値となるλ）は

$$\lambda_{map}=\frac{49972132}{63964329}=0.781$$

となる．

確率$p(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05)$を求めると，

$$p(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05)=\frac{12.8^{11}}{\Gamma(11)}\int_{0.731}^{0.831}\lambda^{10}e^{-12.8\lambda}d\lambda=0.159$$

となる．

問題２．ある確率分布の尤度はポアソン分布$P_O(\lambda)$とし，事前分布であるガンマ分布$Ga(4,1.2)$とする．以下の問いに答えよ．

（１）$Ga(4,1.2)$の確率密度関数$f(\lambda)$を求めよ．

（２）事後パラメータから，

$$n=12,\quad \sum_{i=1}^{12}x_i=9$$

が得られた．事後分布$Ga\left(a',b'\right)$のa',b'を求めよ．

（３）EPA推定値λ_{epa}を小数点以下第３位まで求めよ．

（４）MAP推定値λ_{map}を小数点以下第３位まで求めよ．

（５）確率$p\left(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05\right)$を小数点以下第３位まで求めよ．

総合問題．以下の表は流れ星の5分ごとの観測データである．ただし，尤度はポアソン分布$P_O(\lambda)$に従うと仮定する．

表2（流れ星のデータ）

n回目	1	2	3	4	5	6	7	8	9	10
個 数（x_n）	2	0	2	1	0	3	4	2	1	1

（１）再尤度推定によるλを求めよ．

（２）再尤度推定によるλから流れ星が発生する確率を小数点以下第３位まで求めよ．

（３）事前分布をガンマ分布$Ga(4,1)$としたときの事後分布$Ga\left(a',b'\right)$のa',b'を求めよ．

（４）事前分布をガンマ分布$Ga(4,1)$としたときのλのEPA推定値λ_{epa}を小数点以下第３位まで求めよ．

（５）事前分布をガンマ分布$Ga(4,1)$としたときのλのMAP推定値λ_{map}を小数点以下第３位まで求めよ．

（６）確率$p\left(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05\right)$を小数点以下第３位まで求めよ．

解答

問題１．$\lambda=1.6\quad P=0.798$

問題２．⑴ $f(\lambda)=f(\lambda|4,1.2)=\frac{1.2^4}{\Gamma(4)}\lambda^3e^{-1.2\lambda}$　⑵$a'=13,\ b'=13.2$ ⑶$\lambda_{epa}=0.985$ ⑷ $\lambda_{map}=0.909$

⑸ $\mathrm{p}\left(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05\right)=0.150$

総合問題．⑴ 1.2　⑵ $P=0.699$　⑶$a'=16,\ b'=11$ ⑶$\lambda_{epa}=1.455$ ⑷ $\lambda_{map}=1.364$

⑸ $p\left(\lambda_{map}-0.05\leq\lambda\leq\lambda_{map}+0.05\right)=0.112$