流れ星の観測待ち時間の推定

１．流れ星の観測待ち時間の推定

流れ星の 5 分ごとの観測データを考える．

表 1（流れ星のデータ）

回数 (n)	1	2	3	4	5	6	7	8	9	10
個数 (x_n)	1	0	0	2	0	1	3	0	1	0

そしてポアソン分布$P_O(0.8)$の事前分布はガンマ分布とする．以下において，表 1（流れ星のデータ） から，次にn個の流れ星を約 90%の確率で観測できるためには，何分待つ必要があるかという問題を考える．

ガンマ分布$Ga(n,\lambda)$の確率密度関数は，

$$f(x|n,\lambda) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} \tag{1}$$

であった．特に，$n=1$のとき，

$$f(x|1,\lambda) = \lambda e^{-\lambda x} \tag{2}$$

これは指数分布$Ga(1,\lambda)$と呼ばれた．

指数分布$Ga(n,\lambda)$の意味を考えてみよう．そのために，$f(x)=\lambda e^{-\lambda x}$と置くと，

$$f'(x) = -\lambda^2 e^{-\lambda x} = -\lambda f(x)$$

となる．これより，

$$f(x) = -\lambda \int_0^x f(t)dt + C$$

を得る．$x=0$のとき$f(0)=\lambda$より$C=\lambda$となる．よって，

$$f(x) = \lambda\left(1 - \int_0^x f(x)dt\right) \tag{3}$$

を得た．ここで，Δxを微小時間xとし，$\lambda\Delta x$をΔx内である事象が起こった確率と考えると，

$$f(x)\Delta x = \left(1 - \int_0^x f(x)dt\right) \times (\lambda \Delta x)$$

は，時刻xから$x+\Delta x$の間に事象が起こる確率を意味する．よって，

$$\int_0^x f(x)\,dx = \int_0^x \lambda e^{-\lambda x}\,dx$$

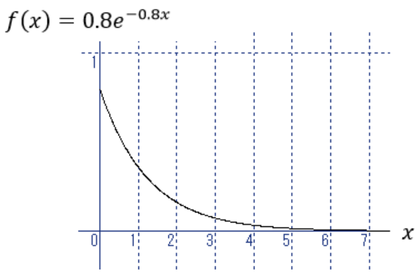
は，x時間までの間に事象が起こらない確率を意味する．以上のことを以下にまとめる．

(1) 指数分布$Ga(n,\lambda)$の確率密度関数$f(x)=\lambda e^{-\lambda x}$は，単位時間あたりに平均$\lambda$回起こる事象において，その事象の発生間隔が$x$単位時間である確率を表す.

(2) $Ga\left(1,\lambda\right)$の平均$V[X]=1/\lambda$は，その事象の平均発生間隔時間を表す.

(3) $\int_0^x \lambda e^{-\lambda x}\,dx$ は，x時間までの間に事象が起こらない確率を意味する.

表 1（流れ星のデータ）から，次に流れ星を観測するまでの待ち時間は，指数分布$Ga\left(1,0.8\right)$に従い，そのグラフは以下である.



平均待ち時間は，平均$E[X]$に単位時間である 5 分を掛けてやればよく

$$E[X]\times 5=\frac{5}{0.8}=6.25\text{分後}$$

となる. このことから，2.9分までの間に流れ星が発生しない確率を予測してみると，

$$\int_0^{2.9} 0.8e^{-0.8t}dt\sim 0.90$$

が得られる. したがって，流れ星を約 90%の確率で観測するには，$2.9\times 5=14.5$ 分待てばよい.

問題 1. 以下の流れ星の 5 分ごとの観測データ（表2）について，以下の問いに答えよ.

表2（流れ星のデータ）

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

(1) 流れ星の個数の分布はポアソン分布$P_O(\lambda)$と考えたとき，5 分ごとの平均発生個数λを求めよ.

(2) 流れ星を観測するまでの平均待ち時間を求めよ.

(3) 3 分までの間に流れ星が発生しない確率を求めよ.

２．n個の流れ星の観測待ち時間の推定

では，表 8.1（流れ星のデータ）から，次にn個の流れ星を見つめるためには，平均何分待つ必要があるか考えてみよう．

X_1, X_2は互いに独立で，確率密度関数はそれぞれ，$f_{X_1}\left(x_1\right), f_{X_2}\left(x_2\right)$とする．このとき，$Y = X_1 + X_2$の確率密度関数$f_Y(y)$を求める．そのために

$$P(Y \leq y) = \int_{-\infty}^{y} f_Y(u)du$$

を$F_Y(y)$と置く．

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(X_1 + X_2 \leq y\right) = P\left(X_2 \leq y - X_1\right) \\ &= \int_{-\infty}^{\infty} f_{x_1}\left(x_1\right)\left\{\int_{-\infty}^{y-x_1} f_{X_2}\left(x_2\right)dx_2\right\}dx_1 \\ &= \int_{-\infty}^{\infty} f_{x_1}\left(x_1\right)\left\{\int_{-\infty}^{y} f_{X_2}\left(u - x_1\right)du\right\}dx_1 \qquad \left(x_2 \coloneqq u - x_1\right) \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{y} f_{X_2}\left(u - x_1\right)f_{x_1}\left(x_1\right)dx_1\right\}du \end{aligned}$$

したがって，$f_Y(y)$は$f_{X_1}\left(x_1\right)$と$f_{X_2}\left(x_2\right)$の畳み込み，すなわち

$$f_Y(y) = f_{X_2}\left(u - x_1\right)f_{x_1}\left(x_1\right)dx_1 = f_{X_1} * f_{X_2}$$

であることがわかった．

さて，2 個の流れ星の待ち時間の分布を求めよう．1 個の流れ星の待ち時間は，指数分布$G_a(n, \lambda)$に従い，その確率密度関数は$f(x|1, \lambda) = \lambda e^{-\lambda x}$であったので，$\mathrm{Y} = \mathrm{X} + \mathrm{X} = 2\mathrm{X}$の確率密度関数$f(y|2, \lambda)$を求めればよい．(4)より，

$$f\left(y|2, \lambda\right) = \lambda e^{-\lambda x} * \lambda e^{-\lambda x} = \lambda^2 \int_0^y e^{-\lambda(y-x)} e^{-\lambda x} dx = \lambda^2 e^{-\lambda y} \int_0^y dx = \lambda^2 y e^{-\lambda y}$$

次に，3 個の流れ星の待ち時間の分布を求めよう．そのためには，$\mathrm{Y} = 2\mathrm{X} + \mathrm{X}$の確率密度関数$f(y|3, \lambda)$を求めればよい．(4)より，

$$f\left(y|3, \lambda\right) = \lambda^2 x e^{-\lambda x} * \lambda e^{-\lambda x} = \lambda^3 \int_0^y x e^{-\lambda(y-x)} e^{-\lambda x} dx = \lambda^3 e^{-\lambda y} \int_0^y x dx = \frac{\lambda^3}{2} y^2 e^{-\lambda y}$$

同様に，4 個の流れ星の待ち時間の分布を求めるには，$\mathrm{Y} = 3\mathrm{X} + \mathrm{X}$の確率密度関数$f(y|4, \lambda)$を求めればよく，(8.4)より，

$$f\left(y|4, \lambda\right) = \frac{\lambda^3}{2} x^2 e^{-\lambda x} * \lambda e^{-\lambda x} = \frac{\lambda^4}{2} \int_0^y x^2 e^{-\lambda(y-x)} e^{-\lambda x} dx = \frac{\lambda^4}{2} e^{-\lambda y} \int_0^y x^2 dx = \frac{\lambda^4}{3!} y^3 e^{-\lambda y}$$

となる．これを繰り返して，n個の流れ星の待ち時間の分布の確率密度関数$f(x|n, \lambda)$

$$f(x|n, \lambda) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$$

であることがわかる．これはガンマ分布$G_a(n, \lambda)$の確率密度関数である．

以上のことから以下のことがいえた．

(1) ガンマ分布$G_a(n,\lambda)$の確率密度関数$f(x)=\frac{\lambda^n}{\Gamma(n)}x^{n-1}e^{-\lambda x}$は，単位時間あたりに平均$\lambda$回起こる事象において，その事象が$n$回発生する間隔が$x$単位時間である確率を表す． (2) $G_a(n,\lambda)$の平均$V[X]=n/\lambda$は，その事象がn回発生するまでの平均発生間隔の時間を表す．

例えば，3個の流れ星が発生するまでの平均待ち時間は，平均$E[X]$に単位時間である 5 分を掛けてやればよく

$$E[X]\times 5=\frac{3}{0.8}\times 5=18.75\text{分後}$$

となる．このことから，$x=6.7$分までの間に流れ星が 3 個発生しない確率を予測してみると，

$$\int_0^{6.7}\frac{0.8^3}{\Gamma(3)}t^2e^{-0.8t}dt\sim 0.90$$

を得る．したがって，3個の流れ星を約 90%の確率で観測するには，$6.7\times 5=33.5$ 分待てばよい．

問題２．表２の流れ星データをもとに，以下の問いに答えよ．

(1) ２個の流れ星が発生するまでの平均待ち時間を求めよ．

(2) ５分までの間に流れ星が２個発生しない確率を求めよ．

解答

問題１．(1) $\lambda=0.7$　(2) 7.14分　(3) 0.88

問題２．(1) 14.3分　(2) 0.86