ガンマ分布とベータ分布

ベイズ推定で必要なガンマ分布とベータ分布を理解しよう．

1 ガンマ分布

正の定数λに関するポアソン分布$P_O(\lambda)$の確率密度関数は，

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

であり，その平均$E[X]$と分散$V[X]$は，

$$E[X] = \lambda, \quad V[X] = \lambda$$

であった．

例えば，自動車の通行台数が単位時間につき平均 5 台である道路において，単位時間に通過する自動車の台数xは，ポアソン分布$P_O(5)$に従うとみなすことはよくある．この場合の平均$E[X]$と分散$V[X]$は，$E[X] = 5,\ V[X] = 5$となる．

一方，ガンマ関数は，

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt \qquad (\alpha > 0)$$

であった．これを用いて，今度はαとλを正の定数として

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

という関数を考える．

命題 1

$$\int_0^\infty f(x|\alpha,\lambda)\,dx = \int_0^\infty \left(\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\right)dx = 1$$

（証明）$t = \lambda x$と置くと，

$$\int_0^\infty f(x|\alpha,\lambda)\,dx = \int_0^\infty \left(\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\right)dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-\lambda x}\,dx$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}e^{-t}\,dt = \frac{1}{\Gamma(\alpha)}\Gamma(\alpha) = 1$$

Q.E.D.

命題 1 により，$f(x|\alpha,\lambda)$は確率密度関数となる．この分布をガンマ分布$\boldsymbol{Ga(\alpha,\lambda)}$という．特に，確率密度関数$f(x|1,\lambda)$をもつ分布を，**指数分布**という．

命題２ ガンマ分布$Ga(\alpha,\lambda)$の平均$E[X]$と分散$V[X]$は，それぞれ以下となる．

$$E[X]=\frac{\alpha}{\lambda},\qquad V[X]=\frac{\alpha}{\lambda^2}$$

（証明）$t=\lambda x$と置く．

$$E[X]=\int_0^\infty xf(x|\alpha,\lambda)\,dx=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^\alpha e^{-\lambda x}\,dx=\frac{1}{\lambda}\cdot\frac{1}{\Gamma(\alpha)}\int_0^\infty t^\alpha e^{-t}\,dt$$

$$=\frac{1}{\lambda}\cdot\frac{1}{\Gamma(\alpha)}\left([-t^\alpha e^{-t}]_0^\infty+\alpha\int_0^\infty t^{\alpha-1}e^{-t}\,dt\right)=\frac{\alpha}{\lambda}\cdot\frac{1}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}e^{-t}\,dt=\frac{\alpha}{\lambda}$$

$V[X]=E[X^2]-(E[X])^2$を用いる．

$$E[X^2]=\int_0^\infty x^2f(x|\alpha,\lambda)\,dx=\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+1}e^{-\lambda x}\,dx=\frac{1}{\lambda^2}\cdot\frac{1}{\Gamma(\alpha)}\int_0^\infty t^{\alpha+1}e^{-t}\,dt$$

$$=\frac{\alpha+1}{\lambda^2}\cdot\frac{1}{\Gamma(\alpha)}\int_0^\infty t^\alpha e^{-t}\,dt=\frac{\alpha(\alpha+1)}{\lambda^2}\cdot\frac{1}{\Gamma(\alpha)}\int_0^\infty t^{\alpha-1}e^{-t}\,dt=\frac{\alpha(\alpha+1)}{\lambda^2}$$

よって，

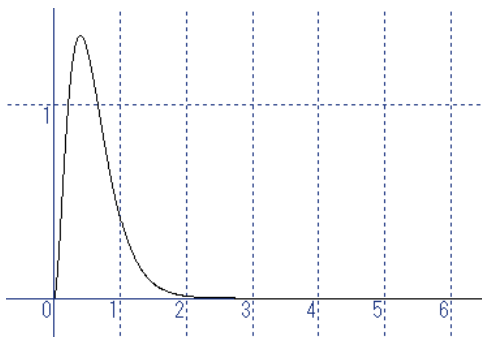
$$V[X]=E[X^2]-(E[X])^2=\frac{\alpha(\alpha+1)}{\lambda^2}-\frac{\alpha^2}{\lambda^2}=\frac{\alpha}{\lambda^2}$$

Q.E.D.

例１．以下のグラフは，ガンマ分布$Ga(3,5)$のグラフで，

$$E[X]=\frac{3}{5},\qquad V[X]=\frac{3}{25}$$

である．

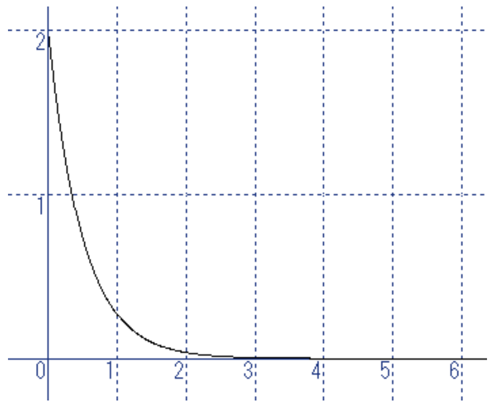


ガンマ分布$Ga(3,5)$のグラフ

例２．右のグラフは，指数分布$Ga(1,2)$のグラフで，

$$E[X]=\frac{1}{2},\qquad V[X]=\frac{1}{4}$$

である．



指数分布$Ga(1,2)$のグラフ

問題１．ガンマ分布$Ga(4,7)$の平均$E[X]$と分散$V[X]$を求めよ．

ガンマ分布$Ga(\alpha,\lambda)$の確率密度関数は，

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

である．ガンマ分布$Ga(\alpha,\lambda)$は，観測時間$1/\lambda$程度の間に１回発生するランダムな事象において，それがα回起こるまでの時間xを表すものとして扱われる．

例３．ある流星群は，観測時間2分程度の間に１個発生するとされる．このとき 8 分間の間に流れ星を5個見れる確率$P(0\le x\le 8)$を，ガンマ分布を使って求めよ．数値積分は小数第３位までの近似値でよい．

（解答）題意より，$\lambda=0.5,\alpha=5$なので，8 分間の間に5個見れるまでの時間の確率分布は$Ga(5,0.5)$に従う．したがって，

$$P(0\le x\le 8)=\int_0^8\frac{0.5^5}{\Gamma(5)}x^4e^{-0.5x}dx=\frac{0.5^5}{24}\int_0^8 x^4e^{-0.5x}dx\sim 0.371$$

である．

例４．ある Web ページには，1 分間に 5 人の来訪者がある．この Web ページに 18 分の間に 100 人が訪れるまでの確率$P(0\le x\le 18)$を，ガンマ分布を使って求めよ．数値積分は小数第３位までの近似値でよい．

（解答）題意より，$\lambda=5,\alpha=100$なので，1 時間の間に 100 人が訪れるまでの時間の確率分布は$Ga(100,5)$に従う．したがって，

$$P(0\le x\le 18)=\int_0^{18}\frac{5^{100}}{\Gamma(100)}x^{99}e^{-5x}dx=\frac{5^{100}}{99!}\int_0^{18}x^{99}e^{-5x}dx\sim 0.158$$

である．

問題２．ある物質の発生現象を観測したら，0.5 分程度の間に 1 回発生することがわかった．3 分の間に 5 回発生するまでの確率$P(0\le x\le 3)$を，ガンマ分布を使って求めよ．数値積分は小数第３位までの近似値でよい．

2 ベータ分布

ベータ関数は,

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx \qquad (a,b>0)$$

であった．これを用いて,

$$f(x|a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

という関数を考える．

命題 3

$$\int_0^1 f(x|a,b)\,dx = 1$$

（証明）

$$\int_0^1 f(x|a,b)\,dx = \frac{1}{B(a,b)}\int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{1}{B(a,b)}B(a,b) = 1$$

Q.E.D.

命題 3 により，$f(x|p,q)$は確率密度関数となる．この分布をベータ分布$\boldsymbol{Beta(a,b)}$という．

命題 4 ベータ分布$Beta(a,b)$の平均$E[X]$と分散$V[X]$は，それぞれ

$$E[X] = \frac{a}{a+b}, \qquad V[X] = \frac{ab}{(a+b)^2(a+b+1)}$$

（証明）

$$E[X] = \int_0^1 xf(x|a,b)\,dx = \frac{1}{B(a,b)}\int_0^1 x^a(1-x)^{b-1}\,dx = \frac{B(a+1,b)}{B(a,b)}$$

$$= \frac{\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} = \frac{a!\,(a+b-1)!}{(a-1)!\,(a+b)!} = \frac{a}{a+b}$$

$V[X] = E[X^2] - (E[X])^2$を用いる．

$$E[X^2] = \int_0^1 x^2 f(x|a,b)\,dx = \frac{1}{B(a,b)}\int_0^1 x^{a+1}(1-x)^{b-1}\,dx = \frac{B(a+2,b)}{B(a,b)}$$

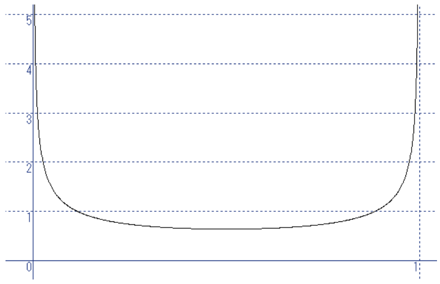
$$= \frac{\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+2)} = \frac{(a+1)!\,(a+b-1)!}{(a-1)!\,(a+b+1)!} = \frac{(a+1)a}{(a+b+1)(a+b)}$$

よって,

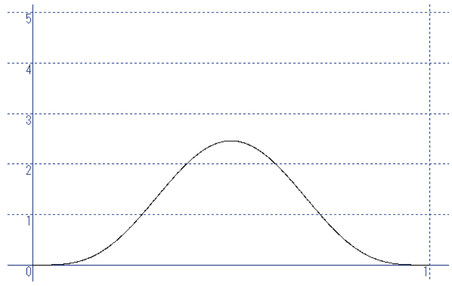
$$V[X] = E[X^2] - (E[X])^2 = \frac{(a+1)a}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2(a+b+1)}$$

Q.E.D.

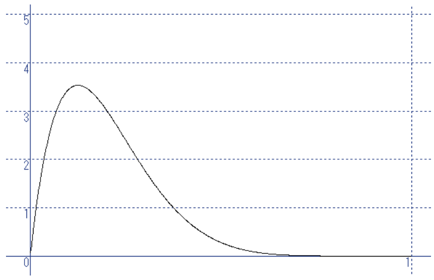
ベータ分布のグラフは，様々な形をもつ．以下にその例を紹介する．



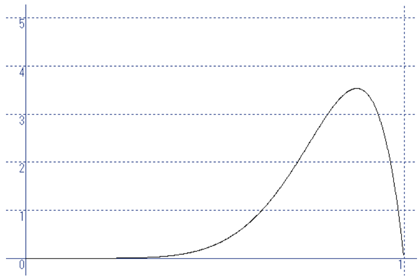
$Beta(0.5,0.5)$



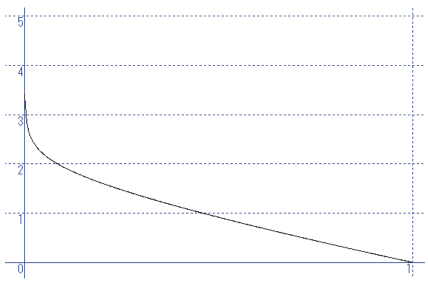
$Beta(5,5)$



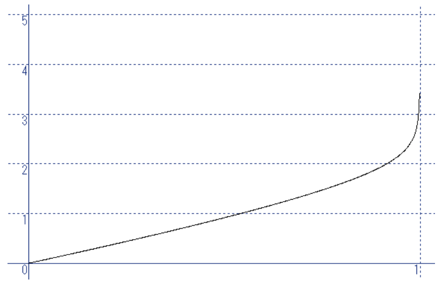
$Beta(2,8)$



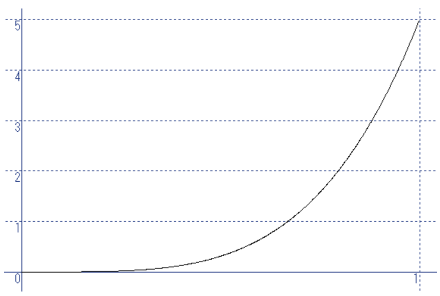
$Beta(8,2)$



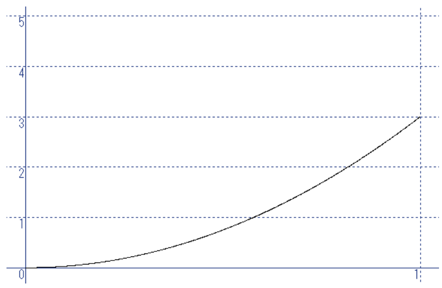
$Beta(0.9,2)$



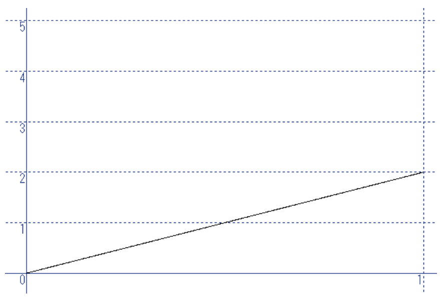
$Beta(2,0.9)$



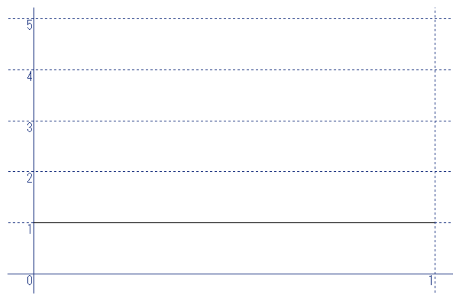
$Beta(5,1)$



$Beta(3,1)$



$Beta(2,1)$



$Beta(1,1)$

上記のように，ベータ分布のグラフは様々であることから，確率モデルを立てたいときに類似的なグラフと考えて恣意的に利用することがある．

問題３．次のベータ分布の平均$E[X]$と分散$V[X]$を求めよ．

（1）$Beta(2,8)$　（2）$Beta(8,2)$　（3）$Beta(2,1)$　（4）$Beta(1,1)$

問題４．あるベータ分布$Beta(a,b)$の平均は$E[X]=\frac{2}{5}$で分散は$V[X]=\frac{1}{25}$である．a,bをそれぞれ求めよ．

解答

問題１．$E[X]=\frac{4}{7},V[X]=\frac{4}{49}$　問題２．$Ga(5,2)$で$P(0\leq x\leq 3)=\int_0^3\frac{2^5}{\Gamma(5)}x^4e^{-2x}dx=\frac{2^5}{4!}\int_0^3 x^4e^{-2x}dx\sim 0.715$

問題 3.（1）$E[X]=\frac{1}{5},V[X]=\frac{4}{275}$　（2）$E[X]=\frac{4}{5},V[X]=\frac{4}{275}$　（3）$E[X]=\frac{2}{3},V[X]=\frac{1}{18}$　（4）$E[X]=\frac{1}{2},V[X]=\frac{1}{12}$　問題 4.　$a=2,b=3$