確率と統計(O)
「連続型確率分布の例(第6章)」

- 担当教員: 杉山 将 (計算工学専攻)
- 居室: W8E-406
- 電子メール: sugi@cs.titech.ac.jp

- 授業のウェブサイト:
  http://sugiyama-www.cs.titech.ac.jp/~sugi/

講義計画(シラバス)

- 確率と統計の基礎
- 確率変数，確率分布
- 積率，積率母関数
- 離散型の確率分布の例
- 連続型の確率分布の例
- 確率不等式， 擬似乱数
- 多次元の確率分布
- 大数の法則，中心極限定理
- 統計的推定，仮説検定

主な連続型の確率分布

- 正規分布
- 一様分布
- ガンマ分布
- 指数分布
- ベータ分布
- コーシー分布

ガンマ分布

- ガンマ分布(gamma distribution): $\alpha, \lambda > 0$

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

- ガンマ関数: $\Gamma(\alpha) = \displaystyle\int_0^{\infty} x^{\alpha-1} e^{-x} dx > 0$

- 単位時間に平均 λ 回起こる事象が, α 回起こるまでの時間 x の分布

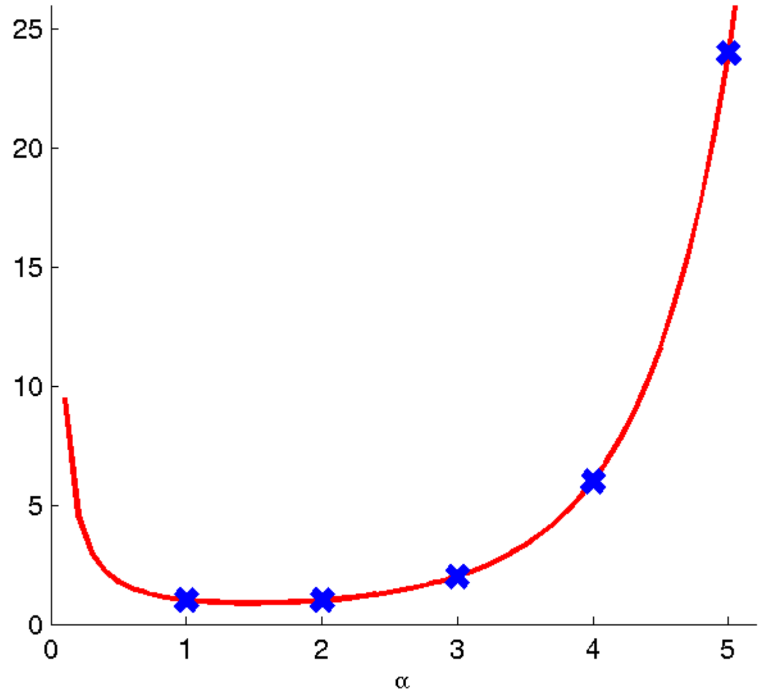
- ポアソン分布:単位時間中に平均 λ 回起こる事象が, 単位時間中にx 回起こる確率

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

ガンマ関数

- ガンマ関数は階乗の一般化：
  整数の α に対して

$$\Gamma(\alpha) = (\alpha - 1)!$$



ガンマ分布(続き)

- $f(x)$ が確率密度関数であることの証明
  - $f(x) \geq 0$ は明らか
  - $y = \lambda x$ とおけば,

$$\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y} \frac{1}{\lambda} dy$$

$$= \frac{1}{\lambda^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy$$

$$= \frac{\Gamma(\alpha)}{\lambda^\alpha}$$

従って,

$$\int_0^\infty f(x) dx = 1$$

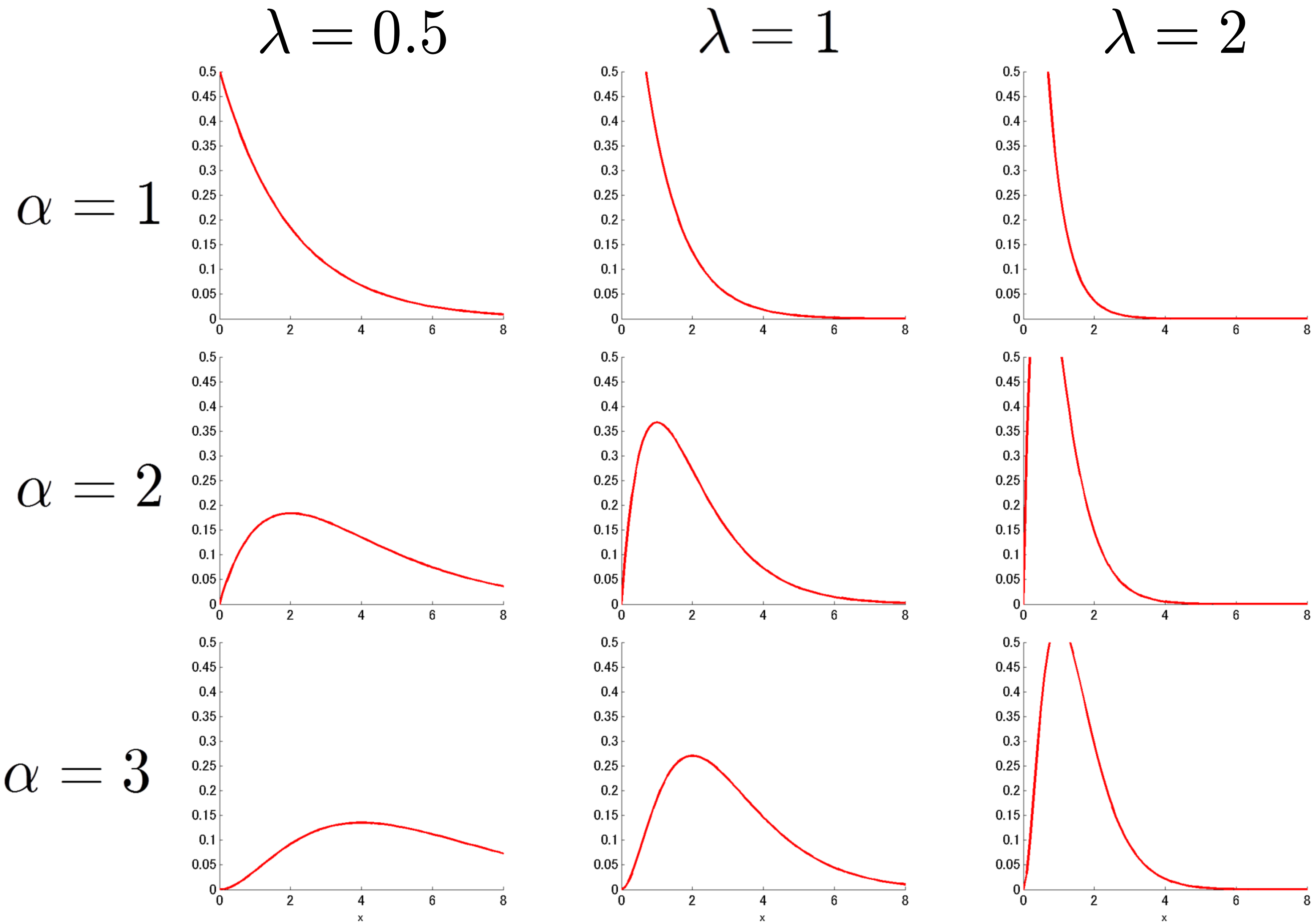
ガンマ分布の性質

- 期待値: $E(X) = \alpha/\lambda$
- 分散: $V(X) = \alpha/\lambda^2$
- 積率母関数: $M_X(t) = \left(\dfrac{\lambda}{\lambda - t}\right)^{\alpha}$

証明は演習！

- α, λ で指定されるガンマ分布を $Ga(\alpha, \lambda)$ で表す

ガンマ分布の例



指数分布

- 指数分布(exponential distribution)
  - 単位時間に平均 λ 回起こる事象が，初めて起こるまでの時間 x の分布

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

- $\alpha = 1$ のガンマ分布に対応

カイニ乗分布

- 確率変数 X_i が独立に標準正規分布 $N(0,1)$ に従うとき，

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

はガンマ分布 $Ga(n/2, 1/2)$ に従う

- このガンマ分布 $Ga(n/2, 1/2)$ を特に，自由度 n の χ^2 (カイ二乗) 分布(chi-squared distribution)と呼ぶ

$$f(x) = \begin{cases} \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

ベータ分布

- ベータ分布(Beta distribution): $\alpha, \beta > 0$

$$f(x) = \begin{cases} x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta) & (0 < x < 1) \\ 0 & (\text{otherwise}) \end{cases}$$

- ベータ関数:

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- 独立に一様分布 $U(0,1)$に従う$\alpha+\beta-1$個の確率変数を大きさの順に並べ替えたとき，小さい方から α 番目の確率変数はベータ分布に従う
- $f(x)$が確率密度関数であることは明らか

ベータ分布の性質

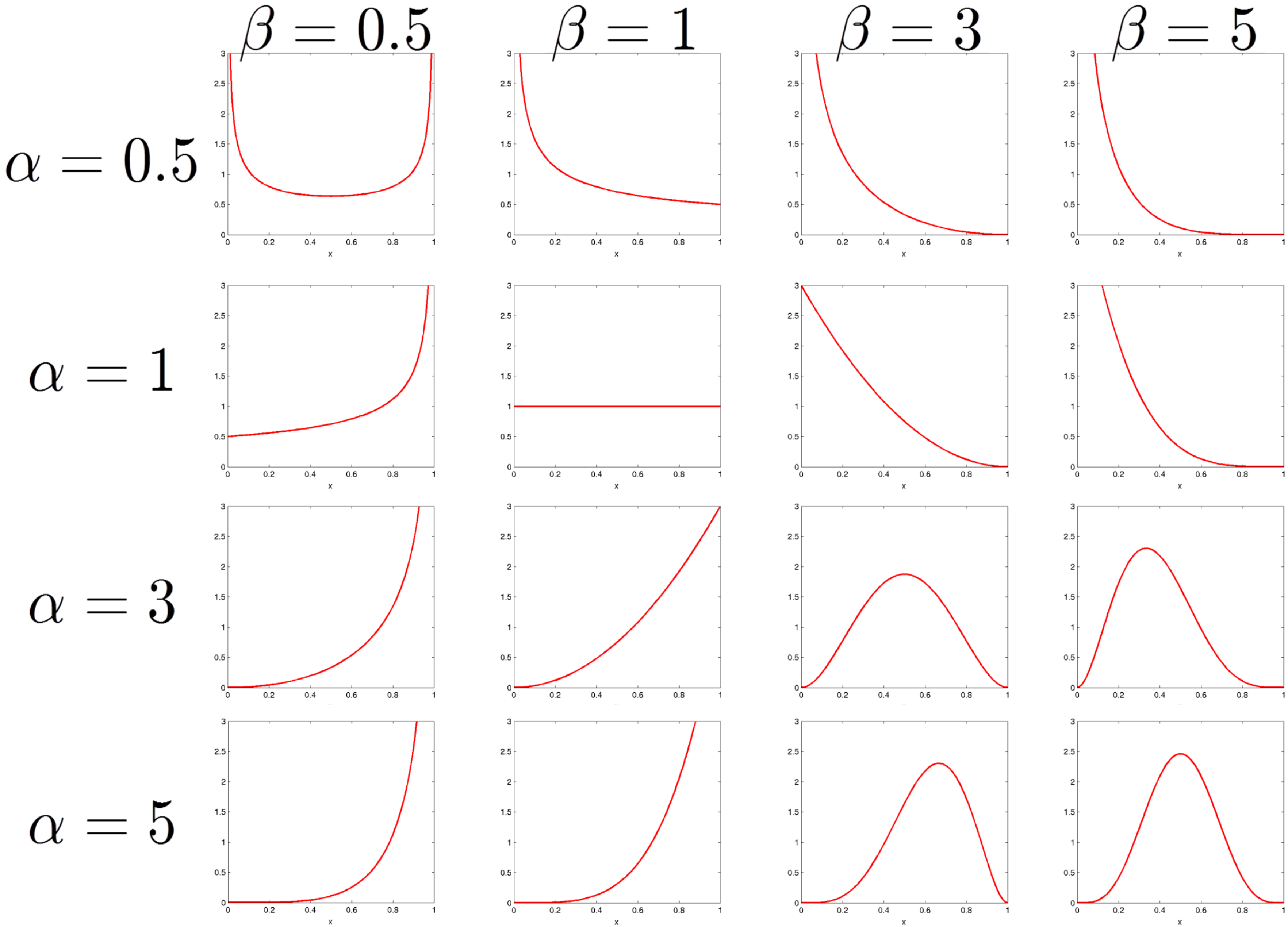
- 期待値: $$E(X) = \frac{\alpha}{\alpha + \beta}$$
- 分散: $$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$
- α, β で指定されるベータ分布を $Be(\alpha, \beta)$ で表す
- $\alpha = \beta = 1$ のとき，特に連続一様分布(uniform distribution of continuous type)と呼び，$U(0, 1)$ で表す

$$f(x) = \begin{cases} 1 & (0 < x < 1) \\ 0 & (\text{otherwise}) \end{cases}$$

証明

- $$E(X) = \frac{1}{B(\alpha,\beta)} \int_0^1 x x^{\alpha-1}(1-x)^{\beta-1} dx$$
  $$= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx$$
  $$(1-x)^{\beta-1} = \frac{d}{dx}\left\{-\frac{(1-x)^{\beta}}{\beta}\right\}$$
  $$= \frac{1}{B(\alpha,\beta)} \left\{ \left[x^{\alpha}\left(-\frac{(1-x)^{\beta}}{\beta}\right)\right]_0^1 - \int_0^1 \alpha x^{\alpha-1}\left(-\frac{(1-x)^{\beta}}{\beta}\right) dx \right\}$$
  $$= \frac{\alpha}{\beta}\frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}(1-x) dx$$
  （部分積分）
  $$= \frac{\alpha}{\beta}\frac{1}{B(\alpha,\beta)} \left\{ \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx - \int_0^1 x x^{\alpha-1}(1-x)^{\beta-1} dx \right\}$$
  $$= \frac{\alpha}{\beta}\{1 - E(X)\}$$
  これより，$E(X) = \dfrac{\alpha}{\alpha+\beta}$
- 分散の証明は宿題

ベータ分布の例



コーシー分布

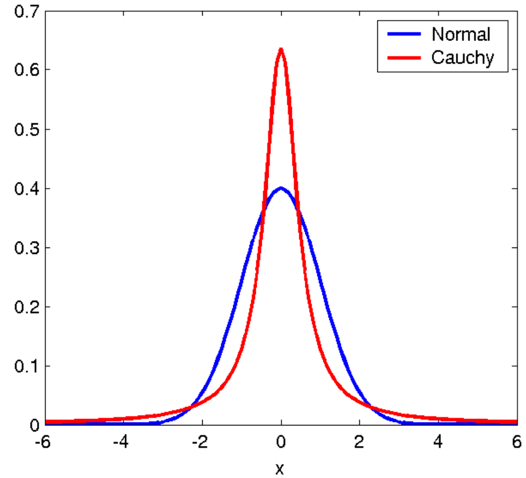
- コーシー分布(Cauchy distribution)　$\alpha > 0$

$$f(x) = \frac{\alpha}{\pi(\alpha^2 + (x-\lambda)^2)}$$

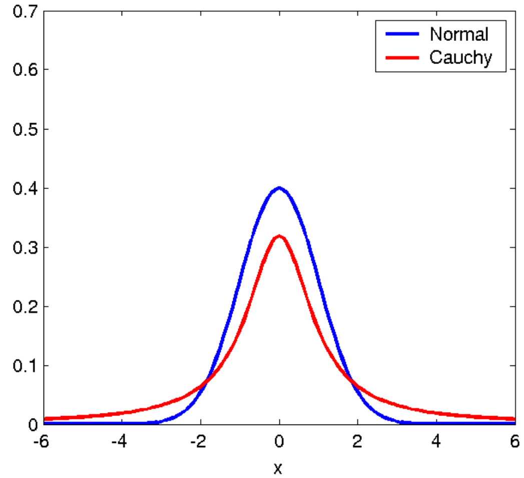
- 標準正規分布に独立に従う確率変数 X, Y の比 X/Y は $\alpha = 1, \lambda = 0$ のコーシー分布に従う

コーシー分布の性質

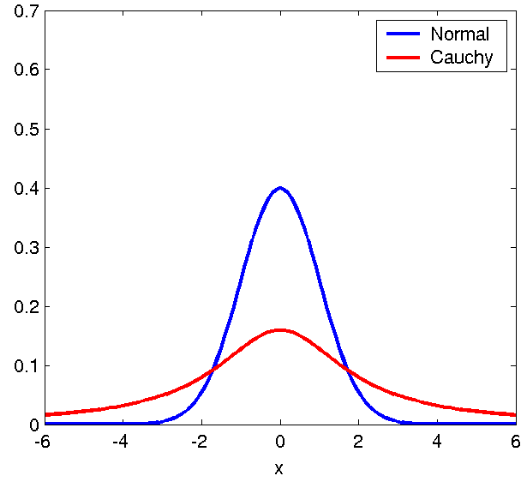
- コーシー分布は見た目が正規分布と似ているが，
  正規分布と異なり期待値と分散が存在しない



$\alpha = 0.5$



$\alpha = 1$



$\alpha = 2$

コーシー分布の期待値について

- 定義より，コーシー分布の期待値は $\alpha = 1$

$$\int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} dx$$

- 被積分関数は奇関数なので一見期待値はゼロに見えるが，実際には期待値は存在しない．なぜなら以下の積分が共に存在しないからである

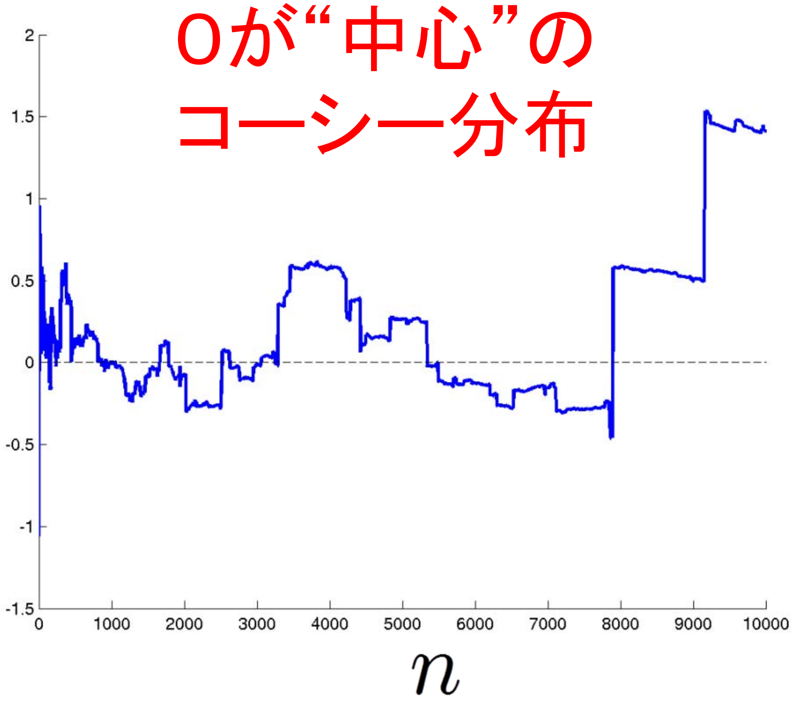
$$\int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} dx \qquad \int_{0}^{\infty} \frac{x}{\pi(1+x^2)} dx$$

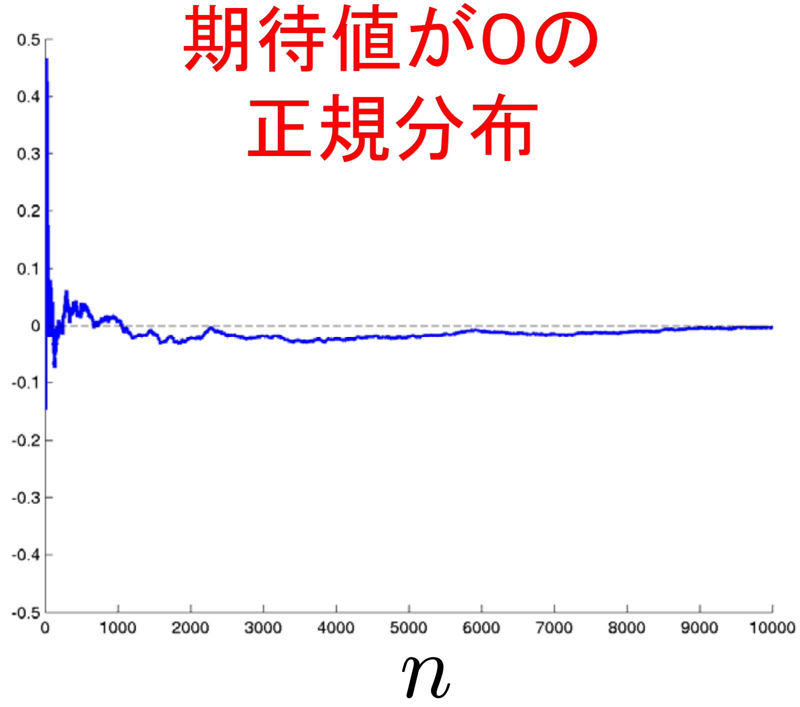
- この事は，後に示す大数の法則(標本平均は，標本数を増やせば本当の期待値に収束する)からもわかる

$$\frac{1}{n}\sum_{i=1}^{n} x_i \to \mu$$

コーシー分布の期待値について(続き)

■ 標本平均の変化 $\displaystyle \frac{1}{n}\sum_{i=1}^{n} x_i$





標本平均は収束しない

→ 平均は存在しない

標本平均は0に収束

まとめ

- 連続分布
  - ガンマ分布
  - カイ二乗分布
  - ベータ分布
  - コーシー分布

宿題

ベータ分布 $Be(\alpha, \beta)$ の分散が次式で与えられることを示せ

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

ヒント: $E(X^2)$ を求め,$V(X) = E(X^2) - \{E(X)\}^2$を使う