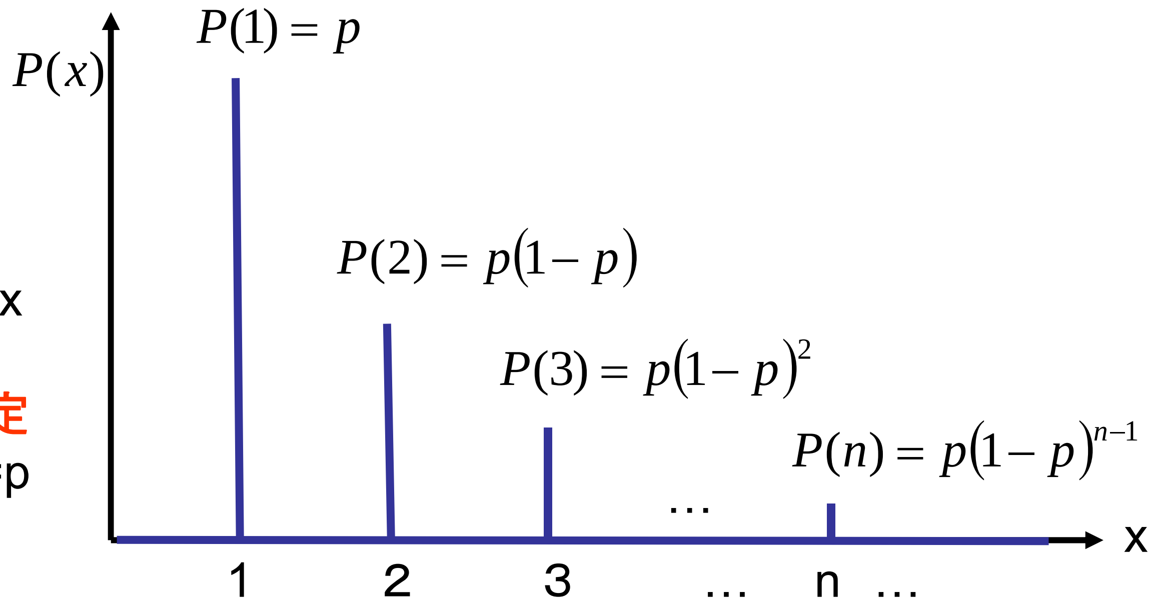
代表的な確率分布関数・確率密度関数について補足

幾何分布や指数分布のように「時間」に関する分布(密度)関数の場合は、
考え方に注意すること → **時刻0の時点で**未来を考えた分布である!

幾何分布
(geometric distribution)

成功する確率が p の事象が
初めて成功するまでの試行回数 x

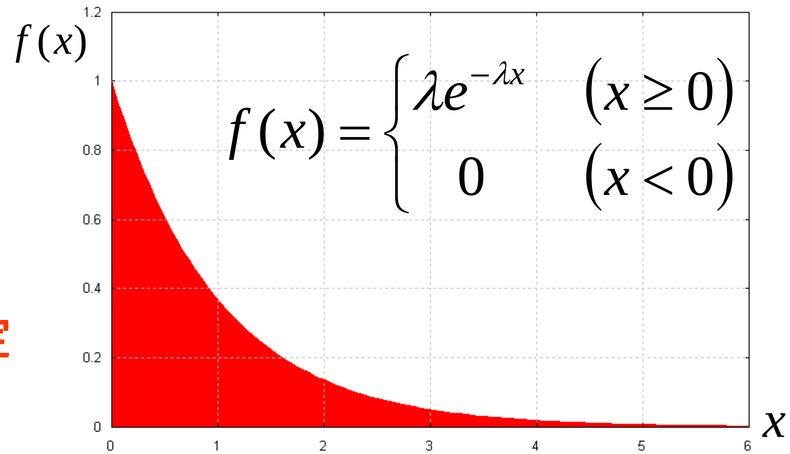
1回あたりの発生確率は p で一定
＝1試行あたりの平均発生件数=p
(平均発生時間間隔1／p)



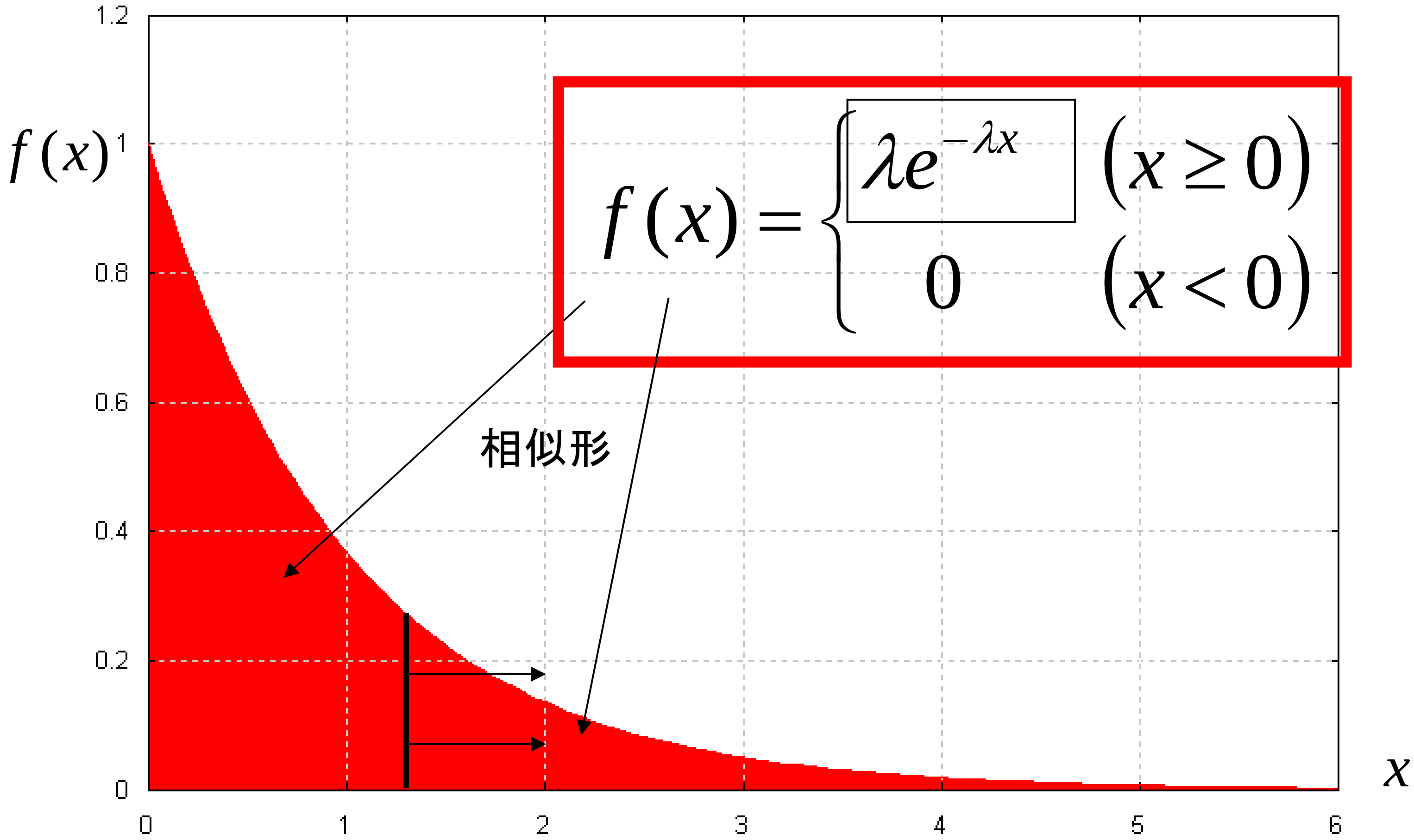
指数分布
(exponential distribution)

確率変数 x は連続値で、時間を表す

単位時間あたりの平均発生件数はλ で一定
(平均発生時間間隔 1／λ)



単位時間あたりの平均発生件数はλ で一定
であるため、時刻 0 からスタートし、現時刻が t である場合、 この時点で未来を考えると（条件付確率密度を計算すると）もとの確率密度関数と同じになる。

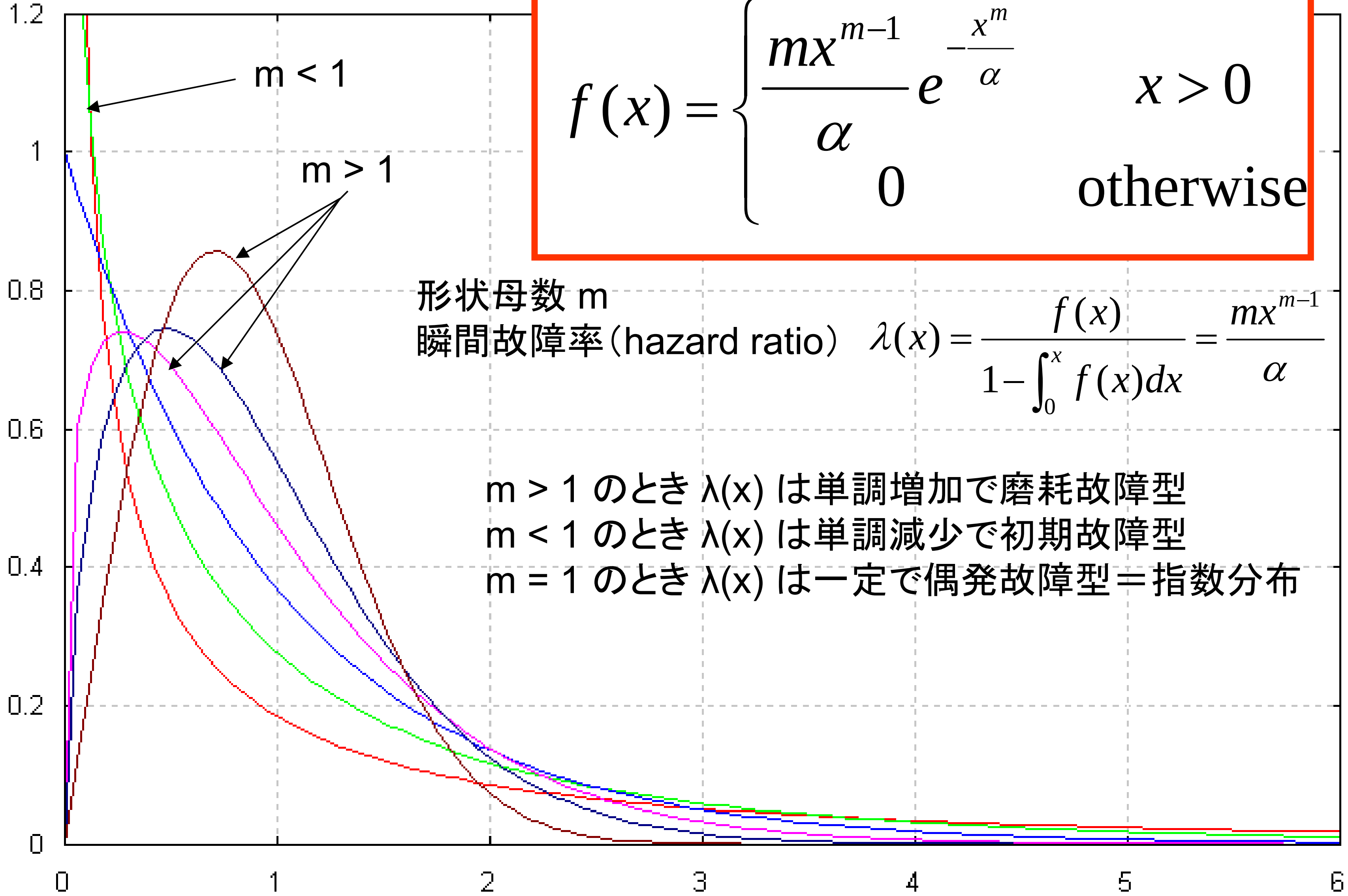


ある時刻 t まで事象が発生しなかった場合、それ以降についても密度関数は同じ
（時刻tまで事象が発生しなかった場合の条件付き確率密度関数＝もとの関数と同じ）

【参考】ワイブル分布（Weibull distribution）

機械の故障発生時間のモデルによく用いられる
指数分布の発展形

$$f(x)=\begin{cases}\dfrac{mx^{m-1}}{\alpha}e^{-\frac{x^m}{\alpha}} & x>0\\ 0 & \text{otherwise}\end{cases}$$



形状母数 m

瞬間故障率（hazard ratio） $\lambda(x)=\dfrac{f(x)}{1-\int_0^x f(x)dx}=\dfrac{mx^{m-1}}{\alpha}$

m > 1 のとき λ(x) は単調増加で磨耗故障型
m < 1 のとき λ(x) は単調減少で初期故障型
m = 1 のとき λ(x) は一定で偶発故障型＝指数分布

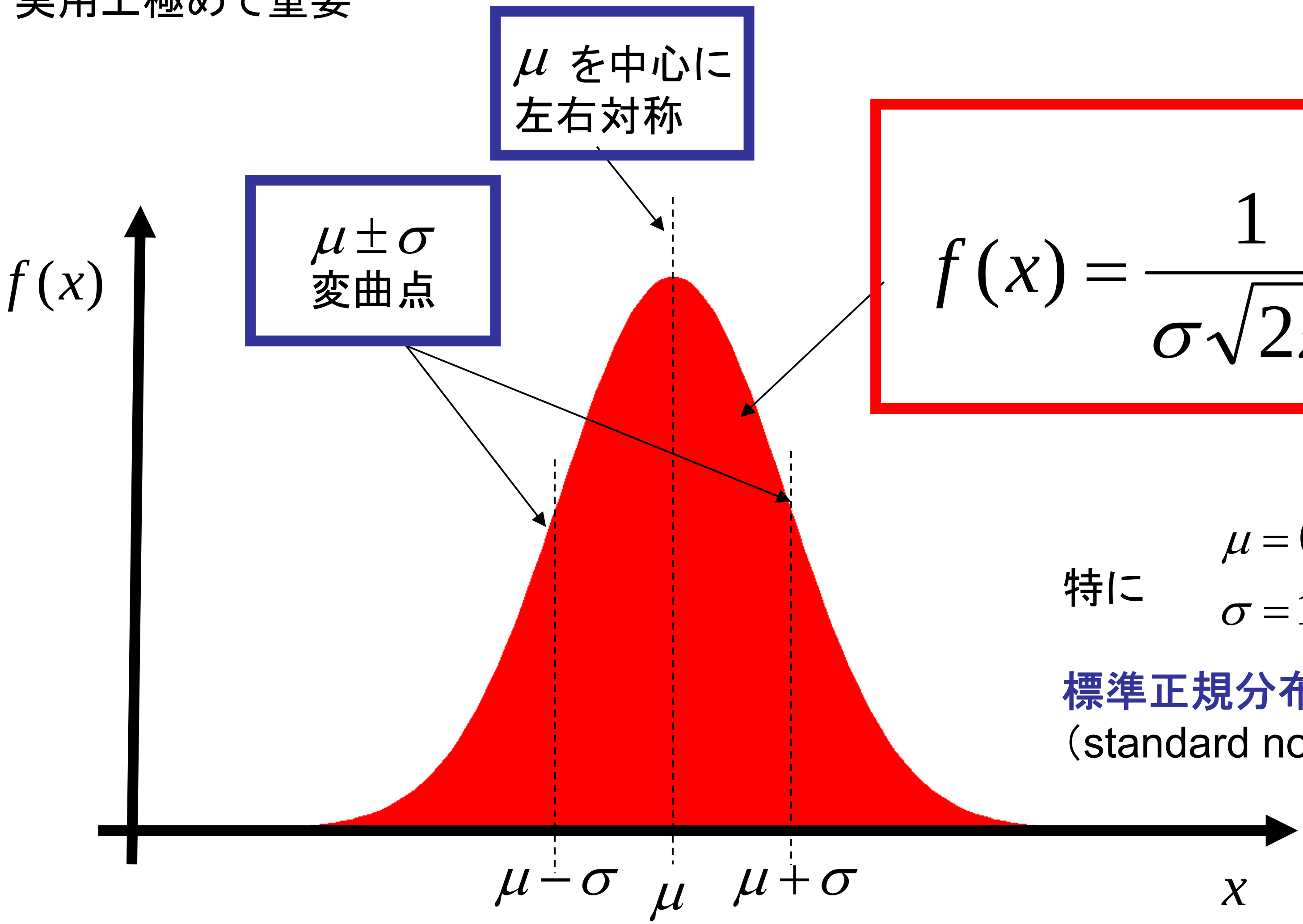
代表的な連続形分布関数：正規分布

Gaussian distribution, normal distribution

$$N(\mu, \sigma^2)$$

期待値 μ
分散 σ^2

多数の分布の和として出てくるような現象の分布：
実用上極めて重要



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

特に $\mu = 0$, $\sigma = 1$ の場合を

標準正規分布と呼ぶ
（standard normal distribution）

九州大学 工学部地球環境工学科
船舶海洋システム工学コース

海事統計学　第7回　（担当：木村）

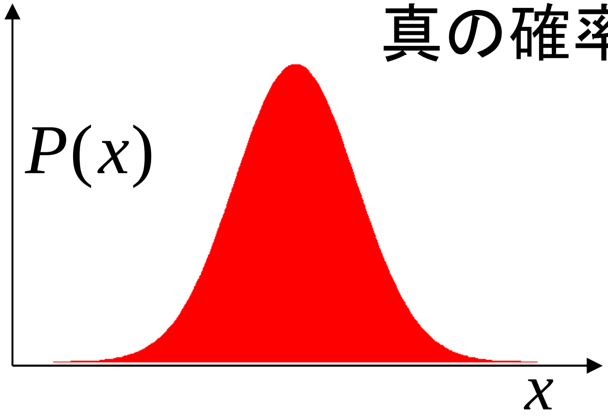
データからの確率モデル構築：最尤推定

授業の資料等は

http://sysplan.nams.kyushu-u.ac.jp/gen/index.html

確率モデルの推定（パラメトリック推定）

環境



真の確率モデル（母関数）　例：正規分布

$$P(x) = N(\mu, \sigma^2)$$

母平均（期待値）および母分散などの
パラメータは観測者にとって未知

観測者は真のモデルが
正規分布であることは
知っていると仮定

観測データ
標本(sample)　$x_1, x_2, \cdots, x_n$

観測者

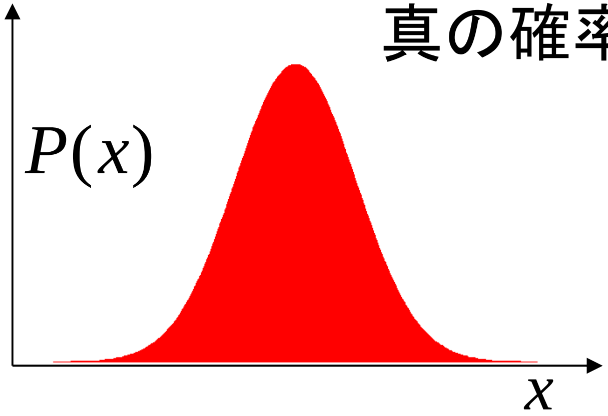
標本を用いて、未知パラメータを推定する：　推定値　$\hat{\mu}, \hat{\sigma}$

単一のパラメータ推定値を求める　→

未知パラメータが存在する区間を求める　→　区間推定

確率モデルの推定（パラメトリック推定）

環境



真の確率モデル（母関数）　例：正規分布

$$P(x) = N(\mu, \sigma^2)$$

母平均（期待値）および母分散などの
パラメータは観測者にとって未知

観測者は真のモデルが
正規分布であることは
知っていると仮定

観測データ
標本(sample)　$x_1, x_2, \cdots, x_n$

観測者

標本を用いて、未知パラメータを推定する：　推定値　$\hat{\mu}, \hat{\sigma}$

単一のパラメータ推定値を求める　→　**点推定**（最尤推定法など）

未知パラメータが存在する区間を求める　→　**区間推定**

最尤推定法(maximum likelihood method)とは？

値 x が生じる確率
（または確率密度）

指数分布ならλ
正規分布ならμとσ

（独立な試行）

推定モデル $\hat{P}(x)$ から観測データ $x_1, x_2, \cdots, x_n$ が
生成される確率 $\boxed{}$ が最大となるように推定パラメータ θ を決定する

$$L(\theta) = \hat{P}(x_1)\hat{P}(x_2)\cdots\hat{P}(x_n)$$

この関数を最大化するパラメータを
見つければ良いが、乗算なので
解析しにくい

対数をとる

$$\ln L(\theta) = \ln\hat{P}(x_1) + \ln\hat{P}(x_2) + \cdots + \ln\hat{P}(x_n)$$

$\boxed{}$

線形になっているので、解析しやすい

尤度を最大化する θ と対数尤度を
最大化する θ の値は同じになる

対数尤度を最大化
するパラメータθを
探索するのが一般的

最尤推定法(maximum likelihood method)とは？

値 x が生じる確率
（または確率密度）

指数分布ならλ
正規分布ならμとσ

（独立な試行）

推定モデル $\hat{P}(x)$ から観測データ $x_1, x_2, \cdots, x_n$ が
生成される確率(**尤度**)が最大となるように推定パラメータ θ を決定する

$$L(\theta) = \hat{P}(x_1)\hat{P}(x_2)\cdots\hat{P}(x_n)$$

この関数を最大化するパラメータを
見つければ良いが、乗算なので
解析しにくい

対数をとる

$$\ln L(\theta) = \ln \hat{P}(x_1) + \ln \hat{P}(x_2) + \cdots + \ln \hat{P}(x_n)$$

対数尤度

線形になっているので、解析しやすい

尤度を最大化する θ と対数尤度を
最大化する θ の値は同じになる

対数尤度を最大化
するパラメータθを
探索するのが一般的

例）Bernoulli分布の最尤推定

真の確率分布モデル（母関数）： 確率 p で x=1, 確率 1-p で x=0 となる分布 F(x)

母関数からの標本： x1=0, x2=1, x3=1, x4=0,…,xn=1
サンプルnコのうち、x=0 となる場合がk回

このとき、未知パラメータpを最尤推定する

対数尤度を計算すると、

$$\ln L(\theta) = \ln F(x_1) + \ln F(x_2) + \cdots \ln F(x_n)$$
$$= k\left(\ln(1-p)\right) + (n-k)\left(\ln(p)\right)$$

これを最大化するpを求めるためpで微分

$$\ln F(x) = \begin{cases} \ln(1-p) & \text{where} \quad x=0 \\ \ln(\mathrm{p}) & \text{where} \quad x=1 \end{cases}$$

$$\frac{\partial}{\partial p} \ln F(x) = \begin{cases} \dfrac{-1}{1-p} & \text{where} \quad x=0 \\ \dfrac{1}{p} & \text{where} \quad x=1 \end{cases}$$

$$\frac{\partial}{\partial p} \ln L(\theta) =$$

最大点では傾きゼロ
これを解くと、
p=（x=1となる場合の数）／（全標本数）
という常識的な式を得る

例）Bernoulli分布の最尤推定

真の確率分布モデル（母関数）: 確率 p で x=1, 確率 1-p で x=0 となる分布 F(x)

母関数からの標本: x1=0, x2=1, x3=1, x4=0,…,xn=1
サンプルnコのうち、x=0 となる場合がk回

このとき、未知パラメータpを最尤推定する

対数尤度を計算すると、

$$\ln L(\theta) = \ln F(x_1) + \ln F(x_2) + \cdots \ln F(x_n)$$
$$= k\left(\ln(1-p)\right) + (n-k)\left(\ln(p)\right)$$

これを最大化するpを求めるためpで微分

$$\ln F(x) = \begin{cases} \ln(1-p) & \text{where} \quad x = 0 \\ \ln(\mathrm{p}) & \text{where} \quad x = 1 \end{cases}$$

$$\frac{\partial}{\partial p} \ln L(\theta) = \boxed{k\left(\frac{-1}{1-p}\right) + (n-k)\frac{1}{p}}$$

$$\frac{\partial}{\partial p} \ln F(x) = \begin{cases} \dfrac{-1}{1-p} & \text{where} \quad x = 0 \\ \dfrac{1}{p} & \text{where} \quad x = 1 \end{cases}$$

$$\boxed{\begin{array}{l} k\left(\dfrac{-1}{1-p}\right) + (n-k)\dfrac{1}{p} = 0 \\ p = \dfrac{n-k}{n} \end{array}}$$

最大点では傾きゼロ
これを解くと、
p=（x=1となる場合の数）／（全標本数）
という常識的な式を得る

例）k値離散分布の最尤推定

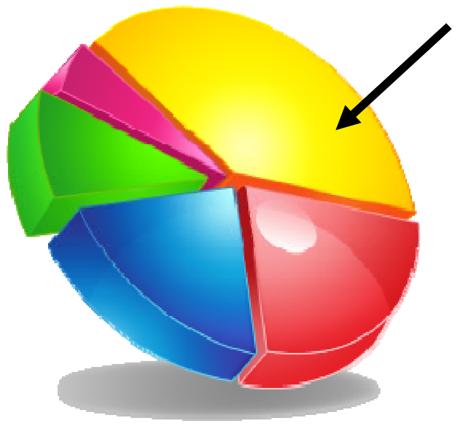
1 から k までの整数の値をとる確率変数 X があり、ある値 i をとる確率は P_i と表される。

この分布からT個のデータ $x_1 x_2 \cdots x_T$ が観測されたとする。
ただしある値 i が観測された回数を n_i と表す。このとき尤度は

$$L = P_{x_1} P_{x_2} \cdots P_{x_T} = \prod_{i=1}^{k} P_i^{n_i}$$

で与えられる。これを最大にする P_i を求める

ただし、P_i は条件として
$$\begin{cases} P_i \geq 0, \quad i = 1,2,\cdots,k \\ \sum_{i=1}^{k} P_i = 1, \quad i = 1,2,\cdots,k \end{cases}$$
を満たしていなければならない



制約条件付き最大値問題

↓

ラグランジュ未定乗数法を使って解く

（参考）ラグランジュ未定乗数法とは？

N 個の変数 $\mathbf{X}=[\theta_1,\theta_2,\cdots\theta_N]$ について、M 個の制約条件 $g_i(\mathbf{X})=0 \quad (i=1,2,\cdots M)$ が与えられている。
この制約のもとで、ある関数 $f(\mathbf{X})$ が極値をとるような変数 $\mathbf{X}$ を求める。

M 個の別の未知変数 $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ を使って、以下の関数 $F(\mathbf{X},\boldsymbol{\lambda})$ を考える:

$$F(\mathbf{X},\boldsymbol{\lambda}) = f(\mathbf{X}) + \sum_{i=1}^{M} \lambda_i\, g_i(\mathbf{X}) \qquad \text{（式１）}$$

この関数 $F(\mathbf{X},\boldsymbol{\lambda})$ の極値条件は、

$$\frac{\partial}{\partial\theta_k} F(\mathbf{X},\boldsymbol{\lambda}) = \frac{\partial}{\partial\theta_k} f(\mathbf{X}) + \frac{\partial}{\partial\theta_k}\sum_{i=1}^{M} \lambda_i\, g_i(\mathbf{X}) = 0 \qquad (k=1,2,\cdots N)$$

$$\frac{\partial}{\partial\lambda_i} F = g_i(\mathbf{X}) = 0 \qquad (i=1,2,\cdots M)$$

これらを満たす解の中に、求める解が存在する。

ここで式（１）をラグランジュ関数、$\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ をラグランジュの未定乗数という。

（参考）ラグランジュ未定乗数法とは？

N 個の変数 $\mathbf{X}=[\theta_1,\theta_2,\cdots\theta_N]$ について、M 個の制約条件 $g_i(\mathbf{X})=0 \quad (i=1,2,\cdots M)$ が与えられている。
この制約のもとで、ある関数 $f(\mathbf{X})$ が極値をとるような変数 $\mathbf{X}$ を求める。

M 個の別の未知変数 $\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ を使って、以下の関数 $F(\mathbf{X},\boldsymbol{\lambda})$ を考える:

$$F(\mathbf{X},\boldsymbol{\lambda}) = \underbrace{f(\mathbf{X})}_{\text{最大化・最小化したい値}} + \sum_{i=1}^{M}\lambda_i \underbrace{g_i(\mathbf{X})}_{\text{制約条件}} \qquad \text{（式１）}$$

この関数 $F(\mathbf{X},\boldsymbol{\lambda})$ の極値条件は、

$$\frac{\partial}{\partial\theta_k}F(\mathbf{X},\boldsymbol{\lambda}) = \frac{\partial}{\partial\theta_k}f(\mathbf{X}) + \frac{\partial}{\partial\theta_k}\sum_{i=1}^{M}\lambda_i g_i(\mathbf{X}) = 0 \qquad (k=1,2,\cdots N)$$

$$\frac{\partial}{\partial\lambda_i}F = g_i(\mathbf{X}) = 0 \qquad (i=1,2,\cdots M)$$

これらを満たす解の中に、求める解が存在する。

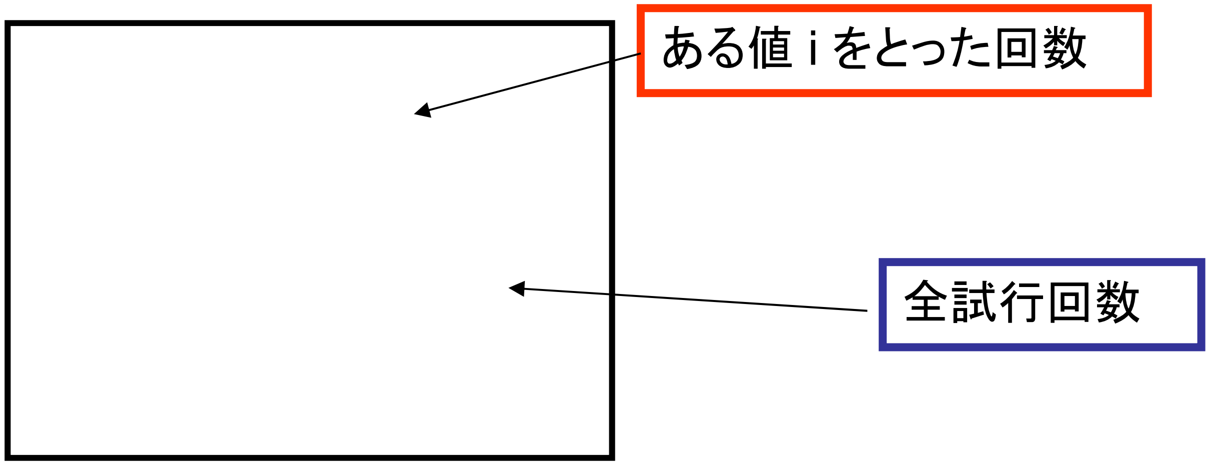
ここで式（１）をラグランジュ関数、$\boldsymbol{\lambda}=[\lambda_1,\lambda_2,\cdots\lambda_M]$ をラグランジュの未定乗数という。

k値離散分布の最尤推定

ラグランジュ関数

$$F = \log L + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{i=1}^{k} P_i \right) - 1 \right)$$

と置いて、極値条件から連立方程式を解くと、



を得る。

k値離散分布の最尤推定

ラグランジュ関数

$$F = \log L + \sum_{i=1}^{S} \lambda_i \left(\left(\sum_{i=1}^{k} P_i \right) - 1 \right)$$

と置いて、極値条件から連立方程式を解くと、

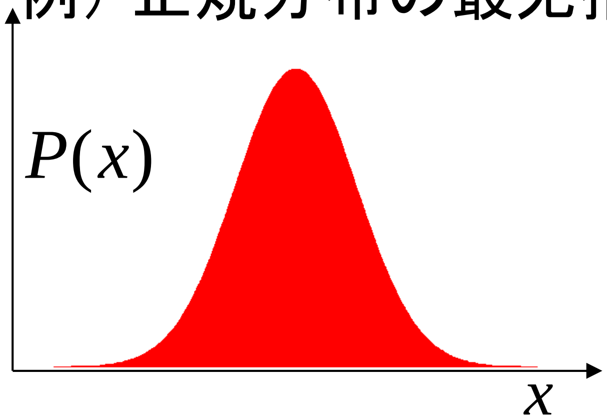
$$\boxed{P_i = \frac{n_i}{\sum_{j=1}^{k} n_j}}$$

ある値 i をとった回数

全試行回数

を得る。

例）正規分布の最尤推定



真の確率モデル:正規分布

$$P(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

標本を用いて、未知パラメータ $\hat{\mu}, \hat{\sigma}$ を最尤推定した場合、

右式のように標本平均と標本分散より推定値が得られることを示せ。

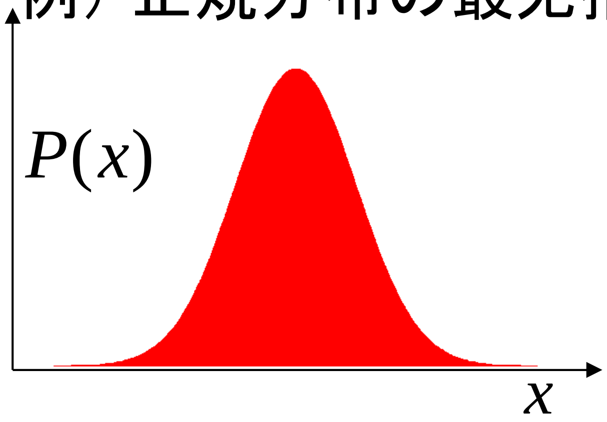
$$\begin{cases} \hat{\mu} = \dfrac{x_1 + x_2 + \cdots x_n}{n}, & (1) \\ \hat{\sigma}^2 = \dfrac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_n - \hat{\mu})^2}{n} & (2) \end{cases}$$

（ヒント） 微分公式

$$\frac{\partial}{\partial\theta} \ln f(\theta) = \frac{1}{f(\theta)} \frac{\partial}{\partial\theta} f(\theta)$$

確率密度関数の場合は、密度関数値で尤度を計算する:

例）正規分布の最尤推定



真の確率モデル:正規分布

$$f(x) = N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

標本を用いて、未知パラメータ $\hat{\mu}, \hat{\sigma}$ を最尤推定した場合、

右式のように標本平均と標本分散より推定値が得られることを示せ。

$$\begin{cases} \hat{\mu} = \dfrac{x_1 + x_2 + \cdots x_n}{n}, \\ \hat{\sigma}^2 = \dfrac{(x_1 - \hat{\mu})^2 + (x_2 - \hat{\mu})^2 + \cdots + (x_n - \hat{\mu})^2}{n} \end{cases}$$

（ヒント） 微分公式

$$\frac{\partial}{\partial \theta} \ln f(\theta) = \frac{1}{f(\theta)} \frac{\partial}{\partial \theta} f(\theta)$$

確率密度関数の場合は、密度関数値で尤度を計算する:

$$\ln L(\mu, \sigma) = \ln f(x_1) + \ln f(x_2) + \cdots + \ln f(x_n)$$

$$= \sum_{i=1}^{n} \ln f(x_i) = \sum_{i=1}^{n} \ln \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right)$$

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma) = \sum_{i=1}^{n} -\frac{(x_i - \mu)}{\sigma^2} = 0$$

を解くと、

$$\sum_{i=1}^{n} (x_i - \mu) = 0$$

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

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

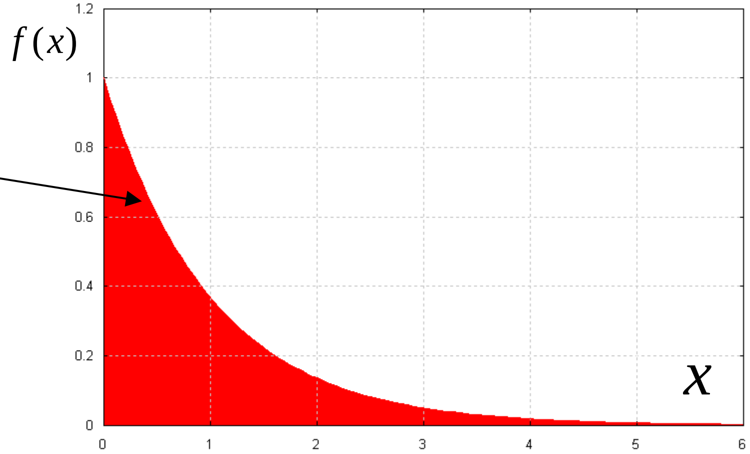
分散についても同様

よってμの最尤推定値は
標本平均で与えられる

【練習問題】

指数分布の
確率密度関数

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



これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

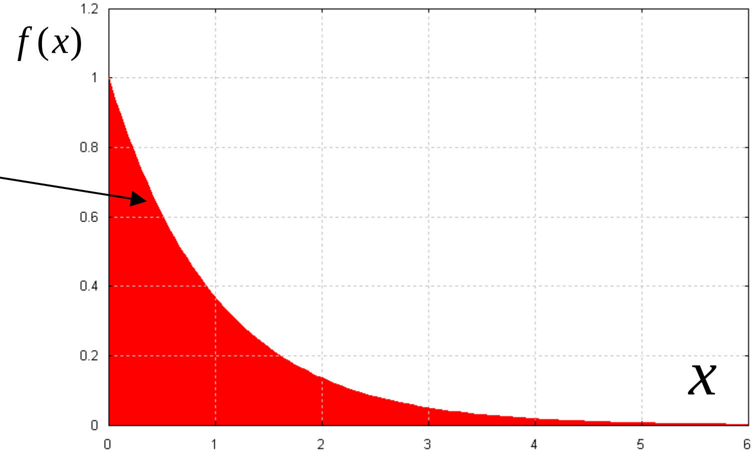
パラメータλを最尤推定せよ。

確率密度関数の場合は、密度関数値で尤度を計算する:

【練習問題】

指数分布の確率密度関数

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



これから得られる観測データ

標本(sample) $x_1, x_2, \cdots, x_n$ のとき、

パラメータλを最尤推定せよ。

確率密度関数の場合は、密度関数値で尤度を計算する：

$$\ln L(\lambda) = \ln f(x_1) + \ln f(x_2) + \cdots + \ln f(x_n)$$

$$= \sum_{i=1}^{n} \ln f(x_i) = \sum_{i=1}^{n} \ln \lambda \exp(-\lambda x_i)$$

$$\frac{\partial}{\partial \lambda} \ln L(\lambda) = \sum_{i=1}^{n} \frac{1}{\lambda \exp(-\lambda x_i)} \left(\exp(-\lambda x_i) + \lambda \exp(-\lambda x_i)(-x_i) \right)$$

$$= \sum_{i=1}^{n} \frac{1}{\lambda} - x_i = 0$$

よって

$$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

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

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

よってλの最尤推定値は
標本平均の逆数で与えられる

【復習】 確率変数のとりうる値が離散かつ無限個の場合

ポアソン分布（Poisson distribution）

単位時間あたり平均で λ 回発生する事象が、

単位時間に x 回（ゼロを含む）発生する確率

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

ポアソン分布の現れる例：

備考：$x=0$ のとき $x\,!=1$

●１分間に放射性物質から放射される粒子が平均２個観測されるとき、
１分間に１個も観測されない確率は？

●ある地域において、年間の交通事故件数が平均730件のとき、
１日に発生する事故の件数が0件である確率は？

$$\lambda = 2 \qquad P(0) = \frac{e^{-2}2^{0}}{0\,!} = e^{-2} = 0.135$$

ポアソン分布について、

$$\frac{\partial}{\partial\lambda}\ln\frac{e^{-\lambda}\lambda^{x}}{x\,!} = \frac{1}{\dfrac{e^{-\lambda}\lambda^{x}}{x\,!}}\frac{-e^{-\lambda}\lambda^{x}+e^{-\lambda}x\lambda^{x-1}}{x\,!} = -1+\frac{x}{\lambda}$$

データからパラメータλを
最尤推定するとどうなる？
処理手順を考えてみよう

まとめ

最尤推定法(maximum likelihood method)とは？

観測データを最も尤もらしく説明する確率モデルのパラメータを決める

モデルの尤もらしさ　＝　モデルからデータが生成される確率（**尤度**）

モデルが確率密度関数の場合は、
確率密度関数値で尤度を計算する

尤度≒対数尤度

（対数）尤度を最大化するパラメータを求める

【演習問題】　2018.04.24　学籍番号　氏名

1 から 3 までの整数値をとる確率変数Xがあり、それぞれの値をとる確率は P_1, P_2, P_3
と表される。 この分布から、確率変数Xが１の値をとった回数が n_1 、２の値をとった回数が n_2
3の値をとった回数が n_3 というデータが観測されたとする。
また確率は $P_1 = P_2$ であることが分かっているものとする。
このとき確率 P_1 をデータから最尤推定せよ。

【演習問題】　　　　　　　　　　　　　　学籍番号　　　　　　　氏名

1 から 3 までの整数値をとる確率変数Xがあり、それぞれの値をとる確率は P_1, P_2, P_3 と表される。 この分布から、確率変数Xが１の値をとった回数がn_1 、２の値をとった回数がn_2 3の値をとった回数が n_3 というデータが観測されたとする。
また確率は $P_1 = P_2$ であることが分かっているものとする。
このとき確率 P_1 をデータから最尤推定せよ。

$$\left.\begin{aligned} P_1 &= P_2 \\ P_1 + P_2 + P_3 &= 1 \end{aligned}\right\} \longrightarrow$$

$$\begin{aligned} \ln L(p_1) &= n_1 \ln p_1 + n_2 \ln p_2 + n_3 \ln p_3 \\ &= (n_1 + n_2)\left(\ln p_1\right) + n_3\left(\ln(1-2p_1)\right) \end{aligned}$$

この対数尤度を最大化するP1を求める

$$\frac{\partial}{\partial p_1} \ln L(p_1) = (n_1 + n_2)\frac{1}{p_1} + n_3\left(\frac{-2}{1-2p_1}\right) = 0$$

この方程式をP1について解くと

$$p_1 = \frac{n_1 + n_2}{2(n_1 + n_2 + n_3)}$$