

最尤推定・最尤法

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

最尤推定・最尤法とは？

- ✓ データが与えられたときに、そのデータが従う確率分布のパラメータを求める方法
- ✓ 尤度関数 (もってもらしさをあらわす関数) を最大化する (微分して 0 とする)

コインの表が出る確率は？

- ✓あるコインを 100 回投げたら、60回表が出た
- ✓コインの表が出る確率 θ を最尤推定により求めてみる
 - ここでは、 θ がパラメータ
- ✓ θ の確率で表が出るコインを100回投げて、60回表が出る確率 L は

$$L(\theta) = {}_{100}C_{60} \theta^{60} (1-\theta)^{40}$$

- L を尤度関数と呼ぶ

尤度関数を最大にする θ を求める

- ✓ 尤度関数 L を最大にする $\theta =$ 一番もっともらしい θ
- ✓ L を最大にする θ と、 L の対数を最大にする θ は同じ
- ✓ L の対数のほうが扱いやすいため、対数に変換する

$$\begin{aligned}\log L(\theta) &= \log \left({}_{100}C_{60} \theta^{60} (1-\theta)^{40} \right) \\ &= \log \left({}_{100}C_{60} \right) + \log \left(\theta^{60} \right) + \log \left((1-\theta)^{40} \right) \\ &= \log \left({}_{100}C_{60} \right) + 60 \log(\theta) + 40 \log(1-\theta)\end{aligned}$$

θ で微分して 0

$\log L(\theta)$ が最大値を取る



$\log L(\theta)$ が極大値を取る



$\log L(\theta)$ を θ で微分したものが 0

$$\frac{d}{d\theta} \log L(\theta) = \frac{60}{\theta} - \frac{40}{1-\theta} = 0$$

$$60(1-\theta) - 40\theta = 0$$

$$\theta = 0.6$$

二項分布

- ✓結果が2通り A or B しかないときに (コインの表・裏など)、
 n 回 行ったとき A になった回数の確率分布
- ✓A になる確率を θ とすると、A が k 回おこる確率は、

$${}_n C_k \theta^k (1-\theta)^{n-k}$$

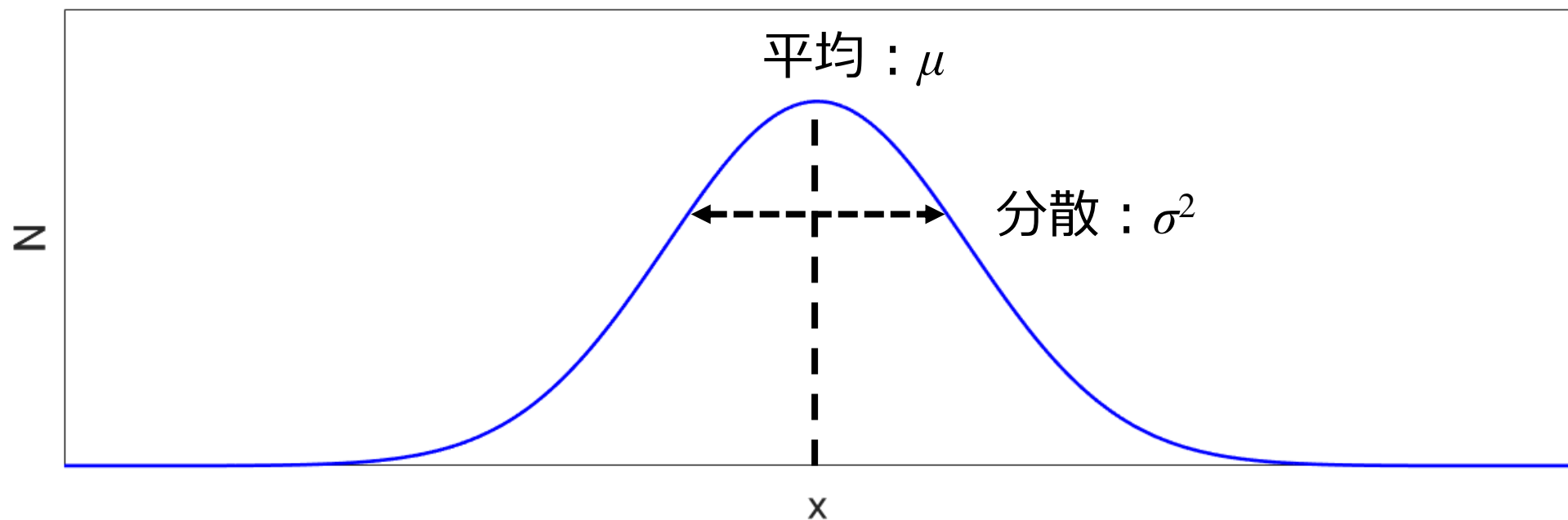
- ✓ θ がわからない (θ がパラメータの) とき、尤度関数 $L(\theta)$ は

$$L(\theta) = {}_n C_k \theta^k (1-\theta)^{n-k}$$

正規分布 (ガウス分布, Gaussian distribution)

6

✓ データが、平均値付近に一番固まっていて、ばらつきのある確率分布



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

正規分布 補足

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

✓ 適当に μ と σ を決めて、上の式にいろいろ x を代入して N を計算すると、前ページの青い線が得られる

✓ 頭の $\frac{1}{\sqrt{2\pi\sigma^2}}$ は、確率なので $\int_{-\infty}^{\infty} N(x | \mu, \sigma^2) dx = 1$

にするためのもの

最尤推定でパラメータを求める

✓ データ $x^{(1)}, x^{(2)}, \dots, x^{(m)}$ が得られたときに、これらが正規分布に従うとしてパラメータ μ, σ^2 を最尤推定により求める

- m はサンプル数
- $\theta = (\mu, \sigma^2)$

✓ 正規分布で、たとえばデータ $x^{(1)}$ になる確率は、 N に $x^{(1)}$ を代入して得られた値

✓ 正規分布で、データ $x^{(1)}, x^{(2)}, \dots, x^{(m)}$ になる確率 L (尤度関数) は、 N にそれぞれ代入して、すべてかけ合わせたもの (確率のかけ算)

$$L(\mu, \sigma^2) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x^{(i)} - \mu)^2 \right\}$$

尤度関数を最大にする μ, σ^2 を求める

✓ 尤度関数 $L(\mu, \sigma^2)$ は

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (x^{(i)} - \mu)^2\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \prod_{i=1}^m \exp\left\{-\frac{1}{2\sigma^2} (x^{(i)} - \mu)^2\right\} \end{aligned}$$

✓ 尤度関数 L を最大にする $\mu, \sigma^2 =$ 一番もつともらしい μ, σ^2

✓ L を最大にする μ, σ^2 と、 L の対数を最大にする μ, σ^2 は同じ

✓ L の対数のほうが扱いやすいため、対数に変換する

$$\begin{aligned}\log L(\mu, \sigma^2) &= \log \left(\frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \prod_{i=1}^m \exp \left\{ -\frac{1}{2\sigma^2} (x^{(i)} - \mu)^2 \right\} \right) \\ &= -\frac{m}{2} \log 2\pi\sigma^2 + \log \prod_{i=1}^m \exp \left\{ -\frac{1}{2\sigma^2} (x^{(i)} - \mu)^2 \right\} \\ &= -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m (x^{(i)} - \mu)^2\end{aligned}$$

μ で偏微分して 0

$L(\mu, \sigma^2)$ が最大値を取る



$L(\mu, \sigma^2)$ が極大値を取る



$L(\mu, \sigma^2)$ を μ で偏微分したものが 0

$$\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2) = -\frac{1}{\sigma^2} \sum_{i=1}^m (x^{(i)} - \mu) = 0$$

μ を求める

$$-\frac{1}{\sigma^2} \sum_{i=1}^m (x^{(i)} - \mu) = 0$$

$$\sum_{i=1}^m (x^{(i)} - \mu) = 0$$

$$\sum_{i=1}^m x^{(i)} - m\mu = 0$$

$$\mu = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

これは、いわゆる**サンプルの平均値**

σ^2 で偏微分して 0

$L(\mu, \sigma^2)$ が最大値を取る



$L(\mu, \sigma^2)$ が極大値を取る



$L(\mu, \sigma^2)$ を σ^2 で偏微分したものが 0

$$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2) = -\frac{m}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^m (x^{(i)} - \mu)^2 = 0$$

σ^2 を求める

$$-\frac{m}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^m (x^{(i)} - \mu)^2 = 0$$

$$-\sigma^2 m + \sum_{i=1}^m (x^{(i)} - \mu)^2 = 0$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2$$

これは、いわゆる**サンプルの分散**

(一般的には m ではなく $m-1$ で割りますが、これは、真の分散はサンプルから計算される分散の期待値に等しく、サンプルの分散を計算するときはサンプルの平均を用いているため小さく見積もられることに由来します。具体的な計算は複雑になるため省きます。)

変数の標準化を考える

✓ 標準化 (オートスケーリング)

- サンプルから平均値を引き、
- サンプルを標準偏差 (分散の平方根) で割る操作

✓ 実は標準化は、各変数のデータ分布を正規分布と仮定して、最尤推定により求められた平均値・分散を利用して行っていた

最尤推定 まとめ

- ✓与えられたデータ・情報から、それが起こる確率 (尤度関数) $L(\theta)$ を求める
 - θ は (複数の) パラメータ
- ✓尤度関数の対数 $\log L(\theta)$ を計算する
- ✓ $\log L(\theta)$ を θ で微分して 0 とし、 θ を求める
 - 微分できれば、どんな分布でも与えられたデータからパラメータの値を計算できる！