HCMUS
Viet Nam National University
Ho Chi Minh City
University of Science

Khoa Toán - Tin học
Fac. of Math. & Computer Science

CHƯƠNG 2: PHÂN PHỐI XÁC SUẤT

Lưu Giang Nam

Bộ môn Ứng dụng Tin học
Khoa Toán - Tin học
Trường Đại học KHTN, ĐHQG TPHCM

09/2025

Mục lục

1 Biến nhị phân
- Biến nhị phân và Phân phối nhị thức
- Phân phối Beta

2 Biến đa thức

3 Phân phối Gauss
- Giới thiệu
- Phân phối Gauss có điều kiện
- Phân phối Gaussian biên
- Lý thuyết Bayes cho biến Gauss
- Maximum Likelihood cho Phân phối Gaussian
- Phân phối T-student

4 Họ phân phối mũ

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phần 1

Biến nhị phân

Phần 1

Biến nhị phân

Mục 1: Biến nhị phân và Phân phối nhị thức

Biến nhị phân (Binary Variables)

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Xét một biến ngẫu nhiên nhị phân $x \in \{0, 1\}$ - ví dụ, kết quả khi tung đồng xu:

$$p(x = 1|\mu) = \mu, \quad p(x = 0|\mu) = 1 - \mu, \quad 0 \leq \mu \leq 1 \tag{1}$$

- Trong đó, μ thể hiện xác suất thành công (ra mặt "1"), còn $1 - \mu$ là xác suất thất bại (ra "0").
- Phân phối xác suất của biến này được gọi là **phân phối Bernoulli**:

$$\text{Bern}(x|\mu) = \mu^x (1 - \mu)^{1-x} \tag{2}$$

- Trung bình và phương sai của biến Bernoulli là:

$$\mathbb{E}[x] = \mu, \quad \text{var}[x] = \mu(1 - \mu) \tag{3}$$

- Điều này có nghĩa là nếu ta lặp lại phép thử nhiều lần, tần suất trung bình của "1" sẽ xấp xỉ μ.

Likelihood and Log-Likelihood

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Giả sử ta có tập dữ liệu $D = \{x_1, x_2, \ldots, x_N\}$ gồm N phép thử độc lập. Khi đó, xác suất quan sát toàn bộ dữ liệu (hay còn gọi là **hàm hợp lý - likelihood**) được viết:

$$p(D|\mu) = \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n} \tag{4}$$

$$\ln p(D|\mu) = \sum_{n=1}^{N} [x_n \ln \mu + (1-x_n)\ln(1-\mu)] \tag{5}$$

- Để tìm giá trị μ hợp lý nhất cho dữ liệu, ta thực hiện **ước lượng cực đại hợp lý (MLE)**:

$$\mu_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{m}{N} \tag{6}$$

- Nghĩa là: xác suất thành công tốt nhất ước lượng được chính là *tần suất quan sát được trong dữ liệu*.

Phân phối nhị thức (Binomial Distribution)

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Nếu ta thực hiện N phép thử Bernoulli độc lập, số lần thành công m (tức là số lần $x = 1$) tuân theo **phân phối nhị thức**:

$$\text{Bin}(m|N, \mu) = \binom{N}{m} \mu^m (1 - \mu)^{N-m} \tag{7}$$

- Các đại lượng kỳ vọng và phương sai của m lần thành công là:

$$\mathbb{E}[m] = N\mu, \quad \text{var}[m] = N\mu(1 - \mu) \tag{8}$$

- Nghĩa là nếu ta lặp lại thí nghiệm nhiều lần, trung bình số lần thành công sẽ bằng $N\mu$, còn phương sai thể hiện độ dao động quanh giá trị trung bình này.

Phần 1

Biến nhị phân

Mục 2: Phân phối Beta

Phân phối Beta (The Beta Distribution)

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Trong phần trước, ta đã thấy rằng ước lượng cực đại hợp lý (MLE) cho tham số μ trong phân phối Bernoulli có thể dẫn đến kết quả phi lý khi số lượng mẫu nhỏ.
- Ví dụ, nếu ta tung đồng xu 3 lần và đều ra mặt sấp, thì $\mu_{\text{ML}} = 1$, nghĩa là xác suất mặt sấp $= 100\%$.
- Trực giác cho thấy đây là kết luận quá tự tin - ta cần một cách kết hợp **niềm tin ban đầu (prior belief)** với dữ liệu quan sát.
- Để mô hình hóa niềm tin ban đầu về $\mu \in [0, 1]$, ta sử dụng **phân phối Beta**, được định nghĩa như sau:

$$\text{Beta}(\mu|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1} \tag{9}$$

- Trong đó $a > 0$ và $b > 0$ là hai **tham số hình dạng (shape parameters)**, điều khiển hình dạng của phân phối.

ính chất của phân phối Beta

- Phân phối Beta cho phép mô hình hóa nhiều mức độ tin tưởng khác nhau về giá trị μ:
  - Khi $a = b = 1$: Beta trở thành phân phối **đồng nhất**, biểu thị rằng ta không thiên vị giá trị nào.
  - Khi $a > b$: phân phối nghiêng về phía μ lớn (nghiêng về "mặt sấp").
  - Khi $a < b$: phân phối nghiêng về phía μ nhỏ (nghiêng về "mặt ngửa").
- Trung bình và phương sai của phân phối Beta là:

$$\mathbb{E}[\mu] = \frac{a}{a+b}, \quad \mathrm{var}[\mu] = \frac{ab}{(a+b)^2(a+b+1)} \tag{10}$$

- Khi a, b càng lớn, phương sai càng nhỏ - nghĩa là ta càng "tự tin" hơn về prior của mình.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Phân phối hậu nghiệm (Posterior)

- Trong Bayes, mục tiêu là tính toán phân phối của tham số sau khi quan sát dữ liệu:

$$p(\mu|D) \propto p(D|\mu)p(\mu) \tag{11}$$

- Với $p(\mu)$ là prior Beta và $p(D|\mu)$ là likelihood Bernoulli:

$$p(\mu|D) \propto \mu^{m+a-1}(1-\mu)^{N-m+b-1} \tag{12}$$

- Dạng trên cũng là một phân phối Beta, nên ta có:

$$p(\mu|D) = \text{Beta}(\mu|a+m,\, b+N-m) \tag{13}$$

- Ta nói rằng phân phối Beta là **đồng dạng (conjugate prior)** với Bernoulli vì posterior có cùng dạng với prior.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Kỳ vọng hậu nghiệm và diễn giải trực giác

- Kỳ vọng của phân phối hậu nghiệm là:

$$\mathbb{E}[\mu|D] = \frac{a+m}{a+b+N} \tag{14}$$

- Công thức này có thể xem như một **trung bình có trọng số** giữa prior và dữ liệu:

$$\mathbb{E}[\mu|D] = \lambda \frac{a}{a+b} + (1-\lambda)\frac{m}{N}, \quad \lambda = \frac{a+b}{a+b+N} \tag{15}$$

- Khi N lớn (nhiều dữ liệu), trọng số λ giảm, nghĩa là dữ liệu chi phối nhiều hơn.
- Khi N nhỏ, prior vẫn ảnh hưởng mạnh, giúp ta tránh **overfitting**.

Ví dụ: Tung đồng xu 3 lần được 3 mặt sấp

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Chọn prior Beta$(2, 2)$ - tương đương với niềm tin ban đầu rằng đồng xu khá cân bằng.
- Quan sát $N = 3$ phép thử và được $m = 3$ mặt sấp:

$$p(\mu|D) = \text{Beta}(\mu|5, 2) \tag{16}$$

- Trung bình hậu nghiệm:

$$\mathbb{E}[\mu|D] = \frac{5}{7} \approx 0.714 \tag{17}$$

- So sánh:
  - MLE: $\mu_{\text{ML}} = 1.0$ (quá tự tin, overfitting)
  - Bayes: $\mathbb{E}[\mu|D] = 0.714$ (hợp lý hơn, phản ánh cả prior và dữ liệu)

Tóm tắt ý nghĩa của phân phối Beta

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Phân phối Beta là lựa chọn tự nhiên để mô hình hóa niềm tin về tham số xác suất $\mu \in [0, 1]$.
- Cung cấp cách tiếp cận thống nhất giữa:
  - prior (niềm tin ban đầu),
  - dữ liệu quan sát,
  - và posterior (niềm tin sau quan sát).
- Giúp tránh overfitting, đặc biệt khi dữ liệu ít.
- Đây là nền tảng của **suy luận Bayes** và mở rộng sang các mô hình phức tạp hơn (ví dụ: Beta-Binomial, Bayesian Regression).

Phần 2

Biến đa thức

Biểu diễn và định nghĩa

- Khi một biến rời rạc có K trạng thái loại trừ lẫn nhau, ta thường dùng **mô hình 1-K** tức là có K số nhưng toàn 0 trừ một số 1:

$$x = (x_1, \ldots, x_K)^T, \qquad x_k \in \{0, 1\}, \quad \sum_{k=1}^{K} x_k = 1 \tag{18}$$

- Ví dụ: với $K = 6$ và trạng thái thứ 3 được quan sát,

$$x = (0, 0, 1, 0, 0, 0)^T. \tag{19}$$

- Gọi $\mu_k = p(x_k = 1)$ với ràng buộc

$$\mu_k \geq 0, \quad \sum_{k=1}^{K} \mu_k = 1. \tag{20}$$

Biểu diễn và định nghĩa

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Với vector 1-K, phân phối của một quan sát x có dạng:

$$p(x|\mu) = \prod_{k=1}^{K} \mu_k^{x_k}. \tag{21}$$

- Phân phối chuẩn hóa vì khi cộng trên tất cả các K vectơ 1-K ta được 1:

$$\sum_{x \text{ 1-K}} p(x|\mu) = \sum_{k=1}^{K} \mu_k = 1. \tag{22}$$

- Kỳ vọng của vector x là:

$$\mathbb{E}[x|\mu] = \mu = (\mu_1, \ldots, \mu_K)^T. \tag{23}$$

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Hàm hợp lý

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Cho dữ liệu $D = \{x_1, \ldots, x_N\}$ độc lập, hàm hợp lý:

$$p(D|\mu) = \prod_{n=1}^{N}\prod_{k=1}^{K}\mu_k^{x_{nk}} = \prod_{k=1}^{K}\mu_k^{\sum_n x_{nk}}. \tag{24}$$

- Định nghĩa số lần xuất hiện trạng thái k:

$$m_k = \sum_{n=1}^{N} x_{nk}, \qquad \sum_{k=1}^{K} m_k = N. \tag{25}$$

- Khi đó ta có:

$$p(D|\mu) = \prod_{k=1}^{K}\mu_k^{m_k}. \tag{26}$$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Ước lượng cực đại (MLE)

- Log-hợp lý sử dụng Lagrange multiplier λ cho ràng buộc $\sum_k \mu_k = 1$):

$$\mathcal{L}(\mu, \lambda) = \sum_{k=1}^{K} m_k \ln \mu_k + \lambda \left(\sum_{k=1}^{K} \mu_k - 1 \right). \tag{27}$$

- Lấy đạo hàm theo μ_k và cho bằng 0:

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \frac{m_k}{\mu_k} + \lambda = 0 \quad \Rightarrow \quad \mu_k = -\frac{m_k}{\lambda}. \tag{28}$$

- Dùng ràng buộc $\sum_k \mu_k = 1$ suy ra $\lambda = -N$, do đó

$$\boxed{\mu_k^{\mathrm{ML}} = \frac{m_k}{N}.} \tag{29}$$

- Giải thích: MLE là tần suất quan sát của trạng thái k.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Phân phối đa biến

- Khi chỉ quan tâm đến vector đếm $(m_1, \ldots, m_K)$, phân phối xác suất là:

$$\text{Mult}(m_1, \ldots, m_K | \mu, N) = \frac{N!}{m_1!\, m_2! \cdots m_K!} \prod_{k=1}^{K} \mu_k^{m_k}. \quad (30)$$

- Ràng buộc: $m_k \geq 0$ và $\sum_k m_k = N$.
- Kỳ vọng và phương sai/covariance:

$$\mathbb{E}[m_k] = N\mu_k, \qquad \text{Var}[m_k] = N\mu_k(1 - \mu_k), \quad (31)$$

$$\text{Cov}[m_i, m_j] = -N\mu_i\mu_j \quad (i \neq j). \quad (32)$$

- Ý nghĩa: tổng các biến đếm gây phụ thuộc và covariance âm giữa các thành phần khác nhau.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Phân phối Dirichlet - tiên nghiệm liên hợp

- Ta chọn tiên nghiệm cho μ là Dirichlet:

$$\mathrm{Dir}(\mu|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \mu_k^{\alpha_k - 1}, \tag{33}$$

$$\text{với } \alpha_k > 0, \quad \alpha_0 = \sum_{k=1}^{K} \alpha_k. \tag{34}$$

- Trực giác về α:
  - α_k có thể hiểu như *pseudo-counts* (số quan sát ảo) cho trạng thái k.
  - Nếu mọi $\alpha_k = 1$: prior là đồng nhất.
  - Nếu $\alpha_k < 1$: prior đẩy mật độ về các đỉnh.
  - Nếu $\alpha_k > 1$: prior tập trung hơn vào miền trong.

Phần 3

Phân phối Gauss

Phần 3

Phân phối Gauss

Mục 1: Giới thiệu

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Phân phối Gaussian

- Phân phối Gaussian (còn gọi là Normal distribution) là một mô hình quan trọng cho các biến ngẫu nhiên liên tục.
- Với một biến đơn x, phân phối có dạng:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \tag{35}$$

- Trong đó:
  - μ: giá trị trung bình (mean)
  - σ^2: phương sai (variance)

Phân phối Gaussian đa biến

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Với vector $x \in \mathbb{R}^D$, ta có:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{36}$$

- μ: vector trung bình kích thước D
- Σ: ma trận hiệp phương sai $D \times D$, xác định hình dạng của phân phối
- $|\Sigma|$: định thức của Σ

Tính chất và ý nghĩa

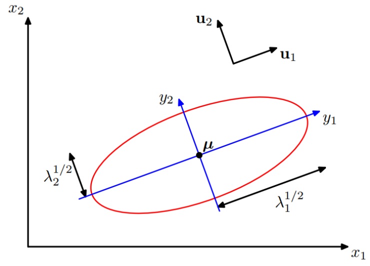
Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Gaussian xuất hiện trong nhiều ngữ cảnhddặc biệt là xuất hiện trong `Định lý giới hạn trung tâm (Central Limit Theorem)`: Nếu $x_1, \ldots, x_N$ là các biến ngẫu nhiên độc lập phân bố đều trên $[0, 1]$, thì trung bình $\frac{1}{N}\sum x_i$ sẽ có xu hướng tuân theo phân phối Gaussian khi $N \to \infty$.
- Phần mũ trong hàm mật độ có dạng:

$$\Delta^2 = (x-\mu)^T \Sigma^{-1} (x-\mu) \tag{37}$$

  với Δ là **khoảng cách Mahalanobis** giữa x và trung tâm μ.
- Nếu $\Sigma = I$, khoảng cách Mahalanobis trở thành khoảng cách Euclid.

Tính chất và ý nghĩa



Hình: Các đường đồng mật độ là các **ellipsoid** trong không gian x.

Phân tích giá trị riêng của ma trận hiệp phương sai

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

- Giải phương trình riêng:

$$\Sigma u_i = \lambda_i u_i \tag{38}$$

- Các u_i tạo thành hệ trực chuẩn, λ_i là các trị riêng (đều dương nếu Σ xác định dương).
- Ta có:

$$\Sigma = \sum_{i=1}^{D} \lambda_i u_i u_i^T, \quad \Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T \tag{39}$$

Biến đổi tọa độ và ý nghĩa hình học

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Đặt $y_i = u_i^T(x - \mu)$, hay viết gọn $y = U(x - \mu)$.
- Trong hệ tọa độ y, phân phối trở thành tích các Gaussian với biến độc lập:

$$p(y) = \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_j)^{1/2}} \exp\left(-\frac{y_j^2}{2\lambda_j}\right) \tag{40}$$

- Mỗi trục u_i tương ứng một phương độc lập, được co giãn theo $\lambda_i^{1/2}$.
- Kỳ vọng (mean): $\mathbb{E}[x] = \mu$. Ma trận hiệp phương sai:

$$\text{cov}[x] = \mathbb{E}[(x - \mu)(x - \mu)^T] = \Sigma \tag{41}$$

- Ma trận Σ quy định **hướng lan rộng** và **mức độ tương quan** giữa các thành phần của x.

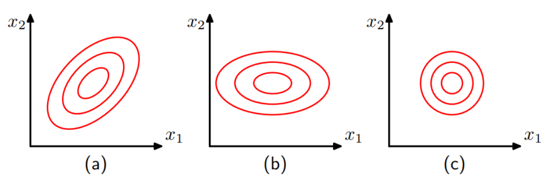
Các dạng ma trận hiệp phương sai

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Tổng số tham số của Gaussian tổng quát:

$$D(D+3)/2$$

- Một số dạng rút gọn:
  a) **Tổng quát**: Σ đầy đủ - ellipsoid nghiêng.
  b) **Đường chéo**: $\Sigma = \text{diag}(\sigma_i^2)$ - ellipsoid trục song song.
  c) **Đẳng hướng (isotropic)**: $\Sigma = \sigma^2 I$ - hình cầu.



Giới hạn của Gaussian

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Gaussian là **phân phối đơn đỉnh** (unimodal).
- Không thể mô tả tốt các phân phối đa đỉnh (multimodal).
- Với số chiều lớn, việc ước lượng Σ trở nên tốn kém vì số tham số tăng bậc $O(D^2)$.
- Giải pháp: dùng **mô hình có ẩn (latent variable)** như Mixture of Gaussians.

Tổng kết và ứng dụng

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Gaussian là nền tảng cho nhiều mô hình trong học máy:
  - Hồi quy tuyến tính (Linear Regression)
  - PCA và các mô hình giảm chiều
  - Kalman Filter, HMM Gaussian
  - Gaussian Mixture Model (GMM)
- Hiểu và thao tác thành thạo Gaussian là bước quan trọng để hiểu:
  - Các mô hình xác suất phức tạp hơn
  - Các kỹ thuật suy luận và tối ưu trong không gian nhiều chiều

Phần 3

Phân phối Gauss

Mục 2: Phân phối Gauss có điều kiện

Ý tưởng tổng quát

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Một tính chất quan trọng của phân phối Gaussian đa biến:
  - Nếu hai nhóm biến là **jointly Gaussian** (phân phối chung Gaussian), thì phân phối **có điều kiện** của một nhóm khi biết nhóm kia cũng là Gaussian.
- Tương tự, phân phối biên (marginal) của mỗi nhóm cũng là Gaussian.
- Đây là nền tảng cho các mô hình tuyến tính Gaussian (linear-Gaussian models).

Phân hoạch biến và tham số

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Giả sử $\mathbf{x}$ là vector D chiều có phân phối:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{42}$$

- Ta chia $\mathbf{x}$ thành hai phần:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix} \tag{43}$$

- $\mathbf{x}_a$ gồm M phần tử đầu, $\mathbf{x}_b$ gồm $D - M$ phần tử còn lại.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Ma trận hiệp phương sai và ma trận chính xác

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Định nghĩa ma trận **precision matrix** (ma trận chính xác):

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} \tag{44}$$

- Phân hoạch (partitioned form) tương ứng:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix} \tag{45}$$

- Chú ý rằng $\boldsymbol{\Lambda}_{aa}$ **không phải** là nghịch đảo của $\boldsymbol{\Sigma}_{aa}$.

Dạng toàn phương của phân phối Gaussian

- Xét hàm mũ trong phân phối Gaussian:

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \tag{46}$$

- Sau khi thay dạng phân hoạch, ta được:

$$\begin{aligned}-\frac{1}{2}(\mathbf{x}_a-\boldsymbol{\mu}_a)^T\boldsymbol{\Lambda}_{aa}(\mathbf{x}_a-\boldsymbol{\mu}_a)-(\mathbf{x}_a-\boldsymbol{\mu}_a)^T\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b-\boldsymbol{\mu}_b)\\ -\frac{1}{2}(\mathbf{x}_b-\boldsymbol{\mu}_b)^T\boldsymbol{\Lambda}_{bb}(\mathbf{x}_b-\boldsymbol{\mu}_b)\end{aligned} \tag{47}$$

- Đây là một **dạng toàn phương** theo $\mathbf{x}_a$ $\Rightarrow$ phân phối có điều kiện $p(\mathbf{x}_a|\mathbf{x}_b)$ cũng là Gaussian.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Phần bù bình phương (Completing the square)

- Phép phần bù bình phương (Completing the square) cho phép ta xác định trung bình và hiệp phương sai từ dạng toàn phương. Dạng tổng quát:

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \text{const} \tag{48}$$

- Từ đó, ta có thể so sánh hệ số để suy ra $\boldsymbol{\Sigma}$ và $\boldsymbol{\mu}$ cho phân phối có điều kiện. Từ phép biến đổi trên, ta thu được:

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Lambda}_{aa}^{-1} \tag{49}$$

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) \tag{50}$$

- Biểu thức trở nên đơn giản và trực quan hơn nếu so với Chuyển sang không gian covariance.

Chuyển sang không gian covariance

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Ta có công thức nghịch đảo cho ma trận phân hoạch:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix} \tag{51}$$

- Trong đó:

$$M = (A - BD^{-1}C)^{-1} \tag{52}$$

- M^{-1} được gọi là **Schur complement** của D trong ma trận ban đầu. Dựa vào các quan hệ trên, ta có:

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \tag{53}$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} \tag{54}$$

- Nhận xét:
  - Trung bình có điều kiện là **hàm tuyến tính** theo $\mathbf{x}_b$.
  - Hiệp phương sai có điều kiện **không phụ thuộc** vào $\mathbf{x}_a$.

Phần 3

Phân phối Gauss

Mục 3: Phân phối Gaussian biên

Tổng quan

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Ở phần trước, ta đã biết rằng:

  Nếu $p(\mathbf{x}_a, \mathbf{x}_b)$ là Gaussian $\Rightarrow p(\mathbf{x}_a|\mathbf{x}_b)$ cũng là Gaussian.

- Giờ ta xem xét **phân phối biên**:

$$p(\mathbf{x}_a) = \int p(\mathbf{x}_a, \mathbf{x}_b)\, d\mathbf{x}_b \tag{55}$$

- Mục tiêu: chứng minh rằng phân phối này cũng là Gaussian, và tìm trung bình, hiệp phương sai của nó.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Ý tưởng tính toán

- Giống như phần điều kiện, ta bắt đầu từ **dạng toàn phương** của hàm mũ trong phân phối chung:

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}(\mathbf{x}-\boldsymbol{\mu}) \tag{56}$$

- Sử dụng phân hoạch ma trận:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}$$

- Ta sẽ tách riêng các hạng tử phụ thuộc vào $\mathbf{x}_b$, sau đó phân bù bình phương để dễ tích phân theo $\mathbf{x}_b$.

Tách các hạng tử phụ thuộc vào $\mathbf{x}_b$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Chỉ giữ lại các phần có chứa $\mathbf{x}_b$:

$$-\frac{1}{2}\mathbf{x}_b^T \boldsymbol{\Lambda}_{bb}\mathbf{x}_b + \mathbf{x}_b^T \mathbf{m} \tag{57}$$

- Với:

$$\mathbf{m} = \boldsymbol{\Lambda}_{bb}\boldsymbol{\mu}_b - \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a) \tag{58}$$

- Sau khi phần bù bình phương, ta được:

$$-\frac{1}{2}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m})^T \boldsymbol{\Lambda}_{bb}(\mathbf{x}_b - \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m}) + \frac{1}{2}\mathbf{m}^T \boldsymbol{\Lambda}_{bb}^{-1}\mathbf{m} \tag{59}$$

Tích phân trên $\mathbf{x}_b$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Phần phụ thuộc $\mathbf{x}_b$ có dạng Gaussian chưa chuẩn hóa:

$$\int \exp\left[-\frac{1}{2}(\mathbf{x}_b - \mathbf{\Lambda}_{bb}^{-1}\mathbf{m})^T \mathbf{\Lambda}_{bb}(\mathbf{x}_b - \mathbf{\Lambda}_{bb}^{-1}\mathbf{m})\right] d\mathbf{x}_b \quad (60)$$

- Do đây là tích phân của Gaussian, nên kết quả chỉ còn phần hằng số: $(2\pi)^{D_b/2}|\mathbf{\Lambda}_{bb}^{-1}|^{1/2}$. Phần còn lại (phụ thuộc $\mathbf{x}_a$) nằm ở hạng tử $\frac{1}{2}\mathbf{m}^T\mathbf{\Lambda}_{bb}^{-1}\mathbf{m}$.
- Kết hợp các hạng tử chứa $\mathbf{x}_a$, ta thu được:

$$-\frac{1}{2}\mathbf{x}_a^T(\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab}\mathbf{\Lambda}_{bb}^{-1}\mathbf{\Lambda}_{ba})\mathbf{x}_a + \mathbf{x}_a^T(\mathbf{\Lambda}_{aa}\boldsymbol{\mu}_a + \mathbf{\Lambda}_{ab}\boldsymbol{\mu}_b) \quad (61)$$

- Dạng này là hàm mũ của một phân phối Gaussian khác $\Rightarrow$ $p(\mathbf{x}_a)$ cũng là Gaussian.

Trong không gian precision

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- So sánh với dạng tổng quát:

$$-\frac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_a)^T \boldsymbol{\Sigma}_a^{-1}(\mathbf{x}_a - \boldsymbol{\mu}_a)$$

- Suy ra:

$$\boldsymbol{\Sigma}_a = (\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba})^{-1} \quad (62)$$

$$\mathbb{E}[\mathbf{x}_a] = \boldsymbol{\mu}_a \quad (63)$$

- Kết quả này thể hiện với phân phối biên, trung bình không thay đổi, chỉ hiệp phương sai được hiệu chỉnh.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Trong không gian covariance

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Từ quan hệ giữa $\boldsymbol{\Sigma}$ và $\boldsymbol{\Lambda}$:

$$\begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix} \tag{64}$$

- Dùng công thức Schur complement, ta được:

$$(\boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba})^{-1} = \boldsymbol{\Sigma}_{aa} \tag{65}$$

- Do đó:

$$\boxed{\mathbb{E}[\mathbf{x}_a] = \boldsymbol{\mu}_a, \quad \mathrm{cov}[\mathbf{x}_a] = \boldsymbol{\Sigma}_{aa}}$$

Tổng kết: Gaussian phân hoạch

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Cho phân phối chung:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} \tag{66}$$

- Phân hoạch:

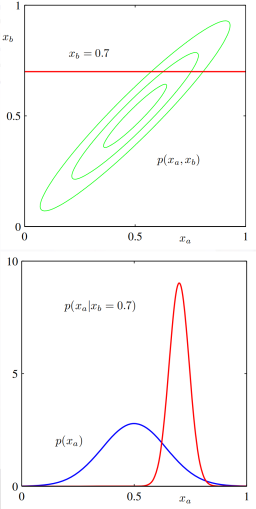
$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

- Kết quả:

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Lambda}_{aa}^{-1}),$$
$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b),$$

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})$$

Phân phối biên vs có điều kiện Gaussian hai chiều



Minh họa: Hình bên trái thể hiện các đường đẳng mật độ (contours) của phân phối Gaussian hai chiều $p(x_a, x_b)$. Đường đỏ $x_b = 0.7$ cho biết lát cắt cố định x_b, từ đó suy ra phân phối có điều kiện $p(x_a|x_b = 0.7)$.

- $p(x_a)$ (đường xanh) là kết quả *tích phân toàn bộ* theo x_b, thể hiện mức độ bất định lớn hơn.
- $p(x_a|x_b = 0.7)$ (đường đỏ) là lát cắt tại $x_b = 0.7$, có phương sai nhỏ hơn do đã biết thông tin về x_b.
- Khi x_a và x_b tương quan dương, $\mu_{a|b}$ dịch theo hướng cùng chiều với $x_b - \mu_b$.

Phần 3

Phân phối Gauss

Mục 4: Lý thuyết Bayes cho biến Gauss

Mục tiêu và mô hình

- Mục tiêu: cho $p(x)$ và $p(y|x)$ đều là Gaussian, tìm $p(y)$ và $p(x|y)$. Mô hình tuyến tính Gaussian:

$$p(x) = \mathcal{N}(x \mid \mu, \Lambda^{-1}), \qquad p(y \mid x) = \mathcal{N}(y \mid Ax + b,\ L^{-1}) \quad (67)$$

  với μ, A, and b là những tham số trung bình, Λ và L là các ma trận chính xác (precision matrix). Nếu $x \in \mathbb{R}^M$, $y \in \mathbb{R}^D$, thì $A \in \mathbb{R}^{D\times M}$.

- Giải pháp: xây dựng phân phối chung $p(x, y)$, rồi suy ra biên (marginal) và điều kiện.

Lý thuyết Bayes cho biến Gauss

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Thay biểu thức Gaussian:

$$\begin{aligned} \ln p(x,y) &= \ln p(x) + \ln p(y \mid x) \\ &= -\tfrac{1}{2}(x-\mu)^T \Lambda (x-\mu) - \tfrac{1}{2}(y - Ax - b)^T L (y - Ax - b) \end{aligned} \tag{68}$$

- Đây là một dạng **đa thức bậc hai** theo hai biến x và y $\Rightarrow$ $p(x,y)$ là phân phối Gaussian.
- Thu gom các hạng bậc hai:

$$\ln p(x,y) = -\frac{1}{2} \begin{pmatrix} y \\ x \end{pmatrix}^T \begin{pmatrix} L & -LA \\ -A^T L & \Lambda + A^T L A \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix}. \tag{69}$$

- Do đó, ma trận chính xác (precision) của $z = (y,x)^T$ là

$$R = \begin{pmatrix} L & -LA \\ -A^T L & \Lambda + A^T L A \end{pmatrix}. \tag{70}$$

- Lưu ý: R là đối xứng, dương xác định nếu L và Λ dương xác định.

Ma trận hiệp phương sai và Trung bình

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Áp dụng công thức nghịch đảo ma trận phân khối (Schur complement) ta thu:

$$\operatorname{cov}[z] = R^{-1} = \begin{pmatrix} L^{-1} + A\Lambda^{-1}A^T & A\Lambda^{-1} \\ \Lambda^{-1}A^T & \Lambda^{-1} \end{pmatrix}. \tag{71}$$

với $\operatorname{cov}[y] = L^{-1} + A\Lambda^{-1}A^T$, $\operatorname{cov}[x] = \Lambda^{-1}$ và $\operatorname{cov}[y, x] = A\Lambda^{-1}$.

- Các phần tử tuyến tính trong $\ln p(x, y)$ là:

$$x^T\Lambda\mu - x^TA^TLb + y^TLb = \begin{pmatrix} y \\ x \end{pmatrix}^T \begin{pmatrix} Lb \\ \Lambda\mu - A^TLb \end{pmatrix}. \tag{72}$$

- Do đó ta có $\mathbb{E}[z]$ và rút gọn:

$$\mathbb{E}[z] = R^{-1}\begin{pmatrix} Lb \\ \Lambda\mu - A^TLb \end{pmatrix} = \mathbb{E}[z] = \begin{pmatrix} A\mu + b \\ \mu \end{pmatrix}. \tag{73}$$

Phân phối biên $p(y)$ và phân phối hậu nghiệm $p(x \mid y)$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Từ khối tương ứng của cov và mean, ta có:

$$\mathbb{E}[y] = A\mu + b, \quad \mathrm{cov}[y] = L^{-1} + A\Lambda^{-1}A^T. \tag{74}$$

- Vậy ta có Phân phối biên $p(y)$:

$$p(y) = \mathcal{N}(y \mid A\mu + b,\ L^{-1} + A\Lambda^{-1}A^T). \tag{75}$$

- Giải thích: phương sai của y gồm hai nguồn - noise trong $p(y \mid x)$ (độ không chắc chắn còn lại) và biến thiên do x.
- Kết quả lý thuyết Bayes cho phân phối Gaussian tuyến tính:

$$p(x \mid y) = \mathcal{N}\left(x \mid \Sigma\left[A^T L(y-b) + \Lambda\mu\right],\ \Sigma\right) \tag{76}$$

$$\text{với} \qquad \Sigma = (\Lambda + A^T L A)^{-1}. \tag{77}$$

Diễn giải trực quan & mối liên hệ với Bayes

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- $p(x)$ là **prior** (niềm tin ban đầu về x).
- $p(y \mid x)$ là **likelihood** (quan sát y cho mỗi x).
- $p(x \mid y)$ là **posterior** (cập nhật niềm tin về x sau khi biết y).
- Posterior precision $\Sigma^{-1} = \Lambda + A^T L A$ = prior precision + thông tin. Nếu L lớn (noise nhỏ), dữ liệu y ảnh hưởng mạnh; nếu L nhỏ (noise lớn), prior chi phối.
- Trường hợp đặc biệt: $A = I$ và $b = 0$ ta có: $y = x +$ noise.
- Khi đó:

$$p(y) = \mathcal{N}(y \mid \mu,\ L^{-1} + \Lambda^{-1}) \tag{78}$$

- Đây là quy tắc cộng phương sai khi cộng hai biến Gaussian độc lập.
- Posterior:

$$\Sigma = (\Lambda + L)^{-1}, \quad \mathbb{E}[x \mid y] = \Sigma(Ly + \Lambda\mu). \tag{79}$$

Ví dụ minh họa

- Giả sử $M = D = 1$, $\mu = 0$, $A = 2$, $b = 0$, $\Lambda = 1/\sigma_x^2$, $L = 1/\sigma_y^2$.
- Cov của y:

$$\mathrm{Var}(y) = \sigma_y^2 + A^2\sigma_x^2. \tag{80}$$

- Posterior variance:

$$\Sigma = (\Lambda + A^T L A)^{-1} = \left(\frac{1}{\sigma_x^2} + \frac{A^2}{\sigma_y^2}\right)^{-1}. \tag{81}$$

- Posterior mean:

$$\mathbb{E}[x \mid y] = \Sigma\left(\frac{A}{\sigma_y^2} y + \frac{1}{\sigma_x^2}\mu\right). \tag{82}$$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Tổng kết

- Với mô hình tuyến tính Gaussian, ta có công thức đóng cho:
  - $p(y) = \mathcal{N}(A\mu + b,\ L^{-1} + A\Lambda^{-1}A^T)$.
  - $p(x \mid y) = \mathcal{N}(\Sigma[A^T L(y - b) + \Lambda\mu],\ \Sigma)$ với $\Sigma = (\Lambda + A^T LA)^{-1}$.
- Những kết quả này là nền tảng cho Bayesian linear regression, Kalman filtering và nhiều mô hình Gaussian khác.

Phần 3

Phân phối Gauss

Mục 5: Maximum Likelihood cho Phân phối Gaussian

Maximum Likelihood cho Phân phối Gaussian

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Giả sử ta có bộ dữ liệu $X = (x_1, x_2, \ldots, x_N)^T$, trong đó mỗi $x_n \in \mathbb{R}^D$. Mỗi quan sát x_n được giả định lấy mẫu độc lập từ phân phối Gaussian đa biến:

$$p(x_n|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu)\right) \tag{83}$$

- Mục tiêu: tìm μ_{ML} và Σ_{ML} sao cho $p(X|\mu, \Sigma)$ là lớn nhất. Do các quan sát độc lập, ta có:

$$p(X|\mu, \Sigma) = \prod_{n=1}^{N} p(x_n|\mu, \Sigma) \tag{84}$$

- Lấy log để đơn giản hoá phép nhân:

$$\ln p(X|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^T \Sigma^{-1}(x_n - \mu) \tag{85}$$

Ước lượng cực đại cho μ

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

- Hàm $\ln p(X|\mu, \Sigma)$ phụ thuộc vào dữ liệu chỉ thông qua:

$$\sum_{n=1}^{N} x_n, \quad \sum_{n=1}^{N} x_n x_n^T \tag{86}$$

- Hai đại lượng này được gọi là **sufficient statistics** — thống kê đầy đủ của Gaussian. Chúng chứa toàn bộ thông tin cần thiết để ước lượng μ và Σ. Lấy đạo hàm theo μ:

$$\frac{\partial}{\partial \mu} \ln p(X|\mu, \Sigma) = \sum_{n=1}^{N} \Sigma^{-1}(x_n - \mu) \tag{87}$$

- Đặt đạo hàm bằng 0:

$$\sum_{n=1}^{N} (x_n - \mu) = 0 \quad \Rightarrow \quad \mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{88}$$

- Kết luận: trung bình cực đại khả năng chính là trung bình mẫu.

Ước lượng cực đại cho Σ

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Thay μ_{ML} vào hàm log-likelihood và lấy đạo hàm theo Σ^{-1}:

$$\frac{\partial}{\partial \Sigma^{-1}} \ln p(X|\mu_{ML}, \Sigma) = \frac{N}{2}\Sigma - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T \tag{89}$$

- Đặt đạo hàm bằng 0 ta được:

$$\Sigma_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T \tag{90}$$

- Đây là ma trận hiệp phương sai mẫu (sample covariance) nhưng chia cho N.

Hiệu chỉnh độ chệch cho Σ

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Với phân phối (μ, Σ):

$$\mathbb{E}[\mu_{ML}] = \mu, \qquad \mathbb{E}[\Sigma_{ML}] = \frac{N-1}{N}\Sigma \tag{91}$$

  - μ_{ML} là **ước lượng không chệch** (unbiased estimator).
  - Σ_{ML} là **ước lượng bị chệch** (biased estimator).

- Để có ước lượng không chệch, ta hiệu chỉnh:

$$\tilde{\Sigma} = \frac{1}{N-1}\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T \tag{92}$$

- Khi đó:

$$\mathbb{E}[\tilde{\Sigma}] = \Sigma \tag{93}$$

- Đây là công thức thường dùng trong thống kê thực nghiệm.

Tổng kết

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- **Phương pháp cực đại khả năng (MLE):**

$$\mu_{ML} = \frac{1}{N}\sum x_n, \quad \Sigma_{ML} = \frac{1}{N}\sum (x_n - \mu_{ML})(x_n - \mu_{ML})^T \tag{94}$$

- **Ước lượng không chệch (unbiased):**

$$\tilde{\Sigma} = \frac{1}{N-1}\sum (x_n - \mu_{ML})(x_n - \mu_{ML})^T \tag{95}$$

- Cả hai đều được sử dụng trong thực hành, tuỳ theo mục đích ước lượng.

Phần 3

Phân phối Gauss

Mục 6: Phân phối T-student

Ý tưởng xuất phát

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Ta có **phân phối tiên nghiệm liên hợp (conjugate prior)** cho độ chính xác (precision) τ của Gaussian là phân phối Gamma.

$$\mathrm{Gam}(\lambda|a,b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} e^{-b\lambda} \tag{96}$$

với $\Gamma(a)$ là hàm Gamma[1]. Khi đó xét mô hình:

$$x|\mu,\tau \sim \mathcal{N}(x|\mu,\tau^{-1}), \quad \tau \sim \mathrm{Gam}(\tau|a,b) \tag{97}$$

- Nếu ta tích phân τ, ta thu được phân phối biên theo x:

$$p(x|\mu,a,b) = \int_0^\infty \mathcal{N}(x|\mu,\tau^{-1})\mathrm{Gam}(\tau|a,b)\,d\tau \tag{98}$$

- Kết quả của tích phân này là **Student's t-distribution**.

[1] Gamma function

Khai triển tích phân

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

$$p(x|\mu, a, b) = \int_0^\infty \frac{b^a e^{-b\tau} \tau^{a-1}}{\Gamma(a)} \left(\frac{\tau}{2\pi}\right)^{1/2} e^{-\frac{\tau}{2}(x-\mu)^2} d\tau \quad (99)$$

- Đây là tích phân của hai hàm: Gaussian theo τ^{-1} và Gamma theo τ. Đổi biến $z = \tau[b + (x-\mu)^2/2]$ để đưa về dạng quen thuộc. Sau các bước biến đổi, ta có:

$$p(x|\mu, a, b) = \frac{b^a \Gamma(a + \frac{1}{2})}{\Gamma(a)\sqrt{\pi}\,[b + (x-\mu)^2/2]^{a+1/2}} \quad (100)$$

- Đặt lại các tham số mới: $\nu = 2a, \lambda = \frac{a}{b}$, khi đó, ta thu được:

$$\mathrm{St}(x|\mu, \lambda, \nu) = \frac{\Gamma(\frac{\nu}{2} + \frac{1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\pi\nu/\lambda}} \left(1 + \frac{\lambda(x-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad (101)$$

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Ý nghĩa của các tham số

- μ: trung tâm (mean) của phân phối.
- λ: "độ chính xác" (precision), điều khiển độ rộng của phân phối.
- ν: **bậc tự do (degrees of freedom)** — điều khiển độ nặng của phần đuôi (tail heaviness).

Giới hạn đặc biệt

$$\begin{aligned} \nu = 1 \quad &\Rightarrow \quad \text{Phân phối Cauchy} \\ \nu \to \infty \quad &\Rightarrow \quad \text{Phân phối Gaussian } \mathcal{N}(x|\mu, \lambda^{-1}) \end{aligned} \tag{102}$$

Diễn giải Bayesian

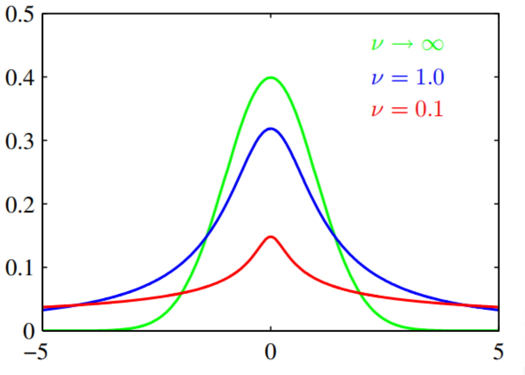
Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

- Từ biểu thức tích phân:

$$\mathrm{St}(x|\mu,\lambda,\nu) = \int_0^\infty \mathcal{N}(x|\mu,(\eta\lambda)^{-1})\mathrm{Gam}(\eta|\tfrac{\nu}{2},\tfrac{\nu}{2})\,d\eta \quad (103)$$

- Ta thấy t-distribution là **trung bình vô hạn** của các Gaussian có cùng μ nhưng khác độ chính xác.
- Phần đuôi (tails) của Student's t phân rã chậm hơn Gaussian. Điều này làm cho nó **ít nhạy cảm hơn với các ngoại lệ (outliers)**. Phân phối t giữ được hình dạng "mạnh mẽ" ngay cả khi dữ liệu có vài điểm xa trung bình.
  - Gaussian: nhạy cảm với outliers
  - Student's t: bền vững hơn.

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ



- Khi ν nhỏ, t-distribution có đuôi dài hơn $\rightarrow$ độ "robust" cao hơn.
- Khi ν tăng, phân phối dần trở nên giống Gaussian.

Phân phối Student's t đa biến

- Tổng quát hoá cho vector $x \in \mathbb{R}^D$:

$$\mathrm{St}(x|\mu, \Lambda, \nu) = \int_0^\infty \mathcal{N}(x|\mu, (\eta\Lambda)^{-1})\mathrm{Gam}(\eta|\tfrac{\nu}{2}, \tfrac{\nu}{2})\, d\eta \quad (104)$$

- Kết quả của tích phân:

$$\mathrm{St}(x|\mu, \Lambda, \nu) = \frac{\Gamma\left(\frac{D+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{|\Lambda|^{1/2}}{(\pi\nu)^{D/2}} \left(1 + \frac{\Delta^2}{\nu}\right)^{-\frac{D+\nu}{2}} \quad (105)$$

với $\Delta^2 = (x-\mu)^T\Lambda(x-\mu)$

- Các tính chất quan trọng:

$$\mathbb{E}[x] = \mu \,(\text{nếu } \nu > 1), \;\; \mathrm{cov}[x] = \frac{\nu}{\nu-2}\Lambda^{-1} \,(\text{nếu } \nu > 2), \;\; \mathrm{mode}[\quad (106)$$

- Khi ν nhỏ, phương sai tăng lên đáng kể $\rightarrow$ đuôi dài hơn.
- Khi $\nu \rightarrow \infty$, $\mathrm{cov}[x] \rightarrow \Lambda^{-1}$ và ta thu lại phân phối Gaussian.

Phần 4

Họ phân phối mũ

Định nghĩa

Một họ phân phối đối với biến x, tham số bởi η (tham số tự nhiên – *natural parameters*), được định nghĩa như sau:

$$p(x|\eta) = h(x)g(\eta)\exp\left(\eta^T u(x)\right) \tag{107}$$

trong đó:

- η: vector **tham số tự nhiên** (natural parameter vector),
- $u(x)$: **thống kê đầy đủ** (sufficient statistic),
- $g(\eta)$: hệ số chuẩn hoá (normalizing coefficient), thoả mãn:

$$g(\eta)\int h(x)\exp\left(\eta^T u(x)\right)dx = 1 \tag{108}$$

Ví dụ 1: Phân phối Bernoulli

Với: $p(x|\mu) = \mu^x(1-\mu)^{1-x}$. Ta có thể viết lại:

$$p(x|\mu) = (1-\mu)\exp\left[x \ln \frac{\mu}{1-\mu}\right] \tag{109}$$

Xác định được:

$$\eta = \ln \frac{\mu}{1-\mu}, \quad \sigma(\eta) = \frac{1}{1+\exp(-\eta)} \quad \text{(hàm sigmoid logistic)} \tag{110}$$

Do đó:

$$p(x|\eta) = \sigma(-\eta)\exp(\eta x) \tag{111}$$

với $u(x) = x, \quad h(x) = 1, \quad g(\eta) = \sigma(-\eta).$

Ví dụ 2: Phân phối đa thức

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

$$p(x|\mu) = \prod_{k=1}^{M} \mu_k^{x_k} = \exp\left(\sum_{k=1}^{M} x_k \ln \mu_k\right) \tag{112}$$

Nếu đặt $\eta_k = \ln \mu_k$ thì:

$$p(x|\eta) = \exp(\eta^T x) \tag{113}$$

với:

$$u(x) = x, \quad h(x) = 1, \quad g(\eta) = 1$$

và điều kiện ràng buộc $\sum_k \mu_k = 1$.

Biểu diễn Softmax

Biến nhị phân
Biến nhị phân và Phân phối nhị thức
Phân phối Beta
Biến đa thức
Phân phối Gauss
Giới thiệu
Phân phối Gauss có điều kiện
Phân phối Gaussian biên
Lý thuyết Bayes cho biến Gauss
Maximum Likelihood cho Phân phối Gaussian
Phân phối T-student
Họ phân phối mũ

Khi loại bỏ μ_M bằng điều kiện $\sum_{k=1}^{M} \mu_k = 1$, ta có:

$$\ln \frac{\mu_k}{1 - \sum_{j=1}^{M-1} \mu_j} = \eta_k \tag{114}$$

suy ra:

$$\mu_k = \frac{\exp(\eta_k)}{1 + \sum_{j=1}^{M-1} \exp(\eta_j)} \tag{115}$$

Do đó:

$$p(x|\eta) = \left(1 + \sum_{k=1}^{M-1} \exp(\eta_k)\right)^{-1} \exp(\eta^T x) \tag{116}$$

Đây là dạng **chuẩn hoá mũ (softmax form)**.

Ví dụ 3: Phân phối Gaussian

Với phân phối Gauss một chiều:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] \tag{117}$$

Ta viết lại dưới dạng họ mũ:

$$p(x|\eta) = h(x)g(\eta)\exp(\eta^T u(x)) \tag{118}$$

trong đó:

$$\eta = \begin{bmatrix} \mu/\sigma^2 \\ -1/(2\sigma^2) \end{bmatrix}, \quad u(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

,

$$h(x) = (2\pi)^{-1/2}, \quad g(\eta) = (-2\eta_2)^{1/2} \exp\left(\frac{\eta_1^2}{4\eta_2}\right)$$

Ước lượng cực đại (Maximum Likelihood)

Từ điều kiện chuẩn hoá:

$$g(\eta) \int h(x) \exp(\eta^T u(x)) dx = 1 \tag{119}$$

Lấy đạo hàm theo η: $-\nabla \ln g(\eta) = \mathbb{E}[u(x)]$ Với N dữ liệu độc lập $\{x_n\}$:

$$p(X|\eta) = \left(\prod_{n=1}^{N} h(x_n)\right) g(\eta)^N \exp\left(\eta^T \sum_{n=1}^{N} u(x_n)\right) \tag{120}$$

Giải điều kiện $\nabla_\eta \ln p(X|\eta) = 0$ ta có:

$$-\nabla \ln g(\eta_{ML}) = \frac{1}{N} \sum_{n=1}^{N} u(x_n) \tag{121}$$

Do đó, ước lượng cực đại η_{ML} chỉ phụ thuộc vào dữ liệu qua $\sum_{n=1}^{N} u(x_n)$, đây chính là **thống kê đầy đủ (sufficient statistic)**.

Good luck!