CMPE 58K
Bayesian Statistics and Machine Learning
Lecture 5

Multivariate distributions: Gaussian, Bernoulli, Probability tables

Department of Computer Engineering,
Boğaziçi University, Istanbul, Turkey
Instructor: A. Taylan Cemgil

Fall 2009

Multivariate Bernoulli

- A probability table with binary states

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	p_{00}	p_{01}
$x_1 = 1$	p_{10}	p_{11}

$$
\begin{aligned}
p(x_1, x_2) &= p_{00}^{(1-x_1)(1-x_2)} p_{10}^{x_1(1-x_2)} p_{01}^{(1-x_1)x_2} p_{11}^{x_1 x_2} \\
&= \exp\left((1 - x_1 - x_2 + x_1 x_2) \log p_{00} + (x_1 - x_1 x_2) \log p_{10}\right. \\
&\quad \left. + (x_2 - x_1 x_2) \log p_{01} + x_1 x_2 \log p_{11}\right) \\
&= \exp\left(\log p_{00} + x_1 \log \frac{p_{10}}{p_{00}} + x_2 \log \frac{p_{01}}{p_{00}} + x_1 x_2 \log \frac{p_{00} p_{11}}{p_{10} p_{01}}\right)
\end{aligned}
$$

Multivariate Bernoulli

- Independent case $p(x_1, x_2) = p(x_1)p(x_2)$

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	$p_0 q_0$	$p_0 q_1$
$x_1 = 1$	$p_1 q_0$	$p_1 q_1$

$$
\begin{aligned}
p(x_1, x_2) &= \exp\left(\log p_0 q_0 + x_1 \log \frac{p_1 q_0}{p_0 q_0} + x_2 \log \frac{p_0 q_1}{p_0 q_0} + x_1 x_2 \log \frac{p_0 q_0 p_1 q_1}{p_1 q_0 p_0 q_1}\right) \\
&= \exp\left(\log p_0 q_0 + x_1 \log \frac{p_1}{p_0} + x_2 \log \frac{q_1}{q_0} + x_1 x_2 \log 1\right)
\end{aligned}
$$

- The coupling parameter is zero

Multivariate Bernoulli

$$
\begin{aligned}
p(x_1, x_2) &= \exp\left(g + h_1 x_1 + h_2 x_2 + \theta_{1,2} x_1 x_2\right) \\
h_1 &= \log \frac{p_{10}}{p_{00}} \\
h_2 &= \log \frac{p_{01}}{p_{00}} \\
\theta_{1,2} &= \log \frac{p_{00} p_{11}}{p_{10} p_{01}}
\end{aligned}
$$

- *Canonical* parameters
  - h_i : threshold
  - $\theta_{i,j}$: pairwise coupling
  - p are the *moment* parameters
- Coupling parameters are harder to interpret but provide direct information about the conditional independence structure

Example

$x_i \in \{0, 1\}$

$$
\begin{aligned}
p(x_1, x_2) &\propto \exp\left(2x_1 - 0.5x_2 - 3x_1x_2\right) \\
q(x_1, x_2) &\propto \exp\left(-2x_1 - 0.5x_2 + 0.01x_1x_2\right)
\end{aligned}
$$

- Are x_1 and x_2 independent, given p (or given q) ?
- What is $p(x_1)$ ($q(x_1)$)?

Odds and the Cross Product Ratio (cpt)

For two events A and B

$$\begin{aligned} \text{odds}(A) &= \frac{p(A)}{p(A^c)} \\ \text{odds}(A|B) &= \frac{p(A|B)}{p(A^c|B)} \\ \text{cpr}(A,B) &= \frac{p(A \cap B)p(A^c \cap B^c)}{p(A^c \cap B)p(A \cap B^c)} \\ &= \text{odds}(A|B)/\text{odds}(A|B^c) \end{aligned}$$

- When the events A and B are independent, we have $p(A|B) = p(A)$

$$\text{cpr}(A,B) = \text{odds}(A)/\text{odds}(A) = 1$$

Odds and the Cross Product Ratio (cpt)

- (Univariate) Bernoulli distribution is parametrised by the event $A : x_1 = 1$

$$p(x_1) = \exp\left(\log p_0 + x_1 \log \frac{p_1}{p_0}\right) \propto \exp\left(x_1 \log \textsf{odds}(x_1 = 1)\right)$$

- (Bivariate) Bernoulli distribution

$$\begin{aligned} p(x_1, x_2) &= \exp\left(\log p_{00} + x_1 \log \frac{p_{10}}{p_{00}} + x_2 \log \frac{p_{01}}{p_{00}} + x_1 x_2 \log \frac{p_{00} p_{11}}{p_{10} p_{01}}\right) \\ &\propto \exp\left(x_1 \log \textsf{odds}(x_1 = 1 | x_2 = 0) + x_2 \log \textsf{odds}(x_2 = 1 | x_1 = 0)\right. \\ &\quad \left. + x_1 x_2 \log \textsf{cpr}(x_1 = 1, x_2 = 1)\right) \end{aligned}$$

Multivariate Bernoulli

For trivariate case, we have pairwise and triple interactions

$$
\begin{aligned}
p(x_1, x_2, x_3) \quad = \quad & p_{000}^{(1-x_1)(1-x_2)(1-x_3)} p_{100}^{x_1(1-x_2)(1-x_3)} p_{010}^{(1-x_1)x_2(1-x_3)} p_{110}^{x_1 x_2 (1-x_3)} \\
& p_{001}^{(1-x_1)(1-x_2)x_3} p_{101}^{x_1(1-x_2)x_3} p_{011}^{(1-x_1)x_2 x_3} p_{111}^{x_1 x_2 x_3}
\end{aligned}
$$

$$
\begin{aligned}
\log p(\mathbf{x}) \quad = \quad & \log p_{000} \\
& + \log \frac{p_{100}}{p_{000}} x_1 + \log \frac{p_{010}}{p_{000}} x_2 + \log \frac{p_{001}}{p_{000}} x_3 \\
& + \log \frac{p_{000} p_{110}}{p_{010} p_{100}} x_1 x_2 + \log \frac{p_{000} p_{101}}{p_{001} p_{100}} x_1 x_3 + \log \frac{p_{000} p_{011}}{p_{001} p_{010}} x_2 x_3 \\
& - \log \frac{p_{001} p_{010} p_{100} p_{111}}{p_{000} p_{011} p_{101} p_{110}} x_1 x_2 x_3
\end{aligned}
$$

$$
\begin{aligned}
\log p(\mathbf{x}) \quad = \quad & \log p_{000} \\
& + \log \mathsf{odds}(x_1 = 1 | x_2 = 0, x_3 = 0) x_1 \\
& + \log \mathsf{odds}(x_2 = 1 | x_1 = 0, x_3 = 0) x_2 \\
& + \log \mathsf{odds}(x_3 = 1 | x_1 = 0, x_2 = 0) x_3 \\
& + \log \mathsf{cpr}(x_1 = 1, x_2 = 1 | x_3 = 0) x_1 x_2 \\
& + \log \mathsf{cpr}(x_1 = 1, x_3 = 1 | x_2 = 0) x_1 x_3 \\
& + \log \mathsf{cpr}(x_2 = 1, x_3 = 1 | x_1 = 0) x_2 x_3 \\
& + \log \frac{\mathsf{cpr}(x_1 = 1, x_2 = 1 | x_3 = 0)}{\mathsf{cpr}(x_1 = 1, x_2 = 1 | x_3 = 1)} x_1 x_2 x_3
\end{aligned}
$$

Canonical Parameters versus Moment Parameters

- Canonical parameters are harder to interpret but give information about the independence structure

- Inference in exponential family models turns out to be just conversion from a canonical representation to a moment representation. Sounds simple but can be easily intractable

The Multivariate Gaussian Distribution. $\mathcal{N}(s; \mu, P)$

μ is the mean and P is the covariance:

$$\begin{aligned}
\mathcal{N}(s; \mu, P) &= |2\pi P|^{-1/2} \exp\left(-\frac{1}{2}(s-\mu)^\top P^{-1}(s-\mu)\right) \\
&= \exp\left(-\frac{1}{2}s^\top P^{-1}s + \mu^\top P^{-1}s - \frac{1}{2}\mu^\top P^{-1}\mu - \frac{1}{2}|2\pi P|\right) \\
\log \mathcal{N}(s; \mu, P) &= -\frac{1}{2}s^\top P^{-1}s + \mu^\top P^{-1}s + \text{ const} \\
&= -\frac{1}{2}\,\mathbf{Tr}\, P^{-1}ss^\top + \mu^\top P^{-1}s + \text{ const} \\
&=^+ -\frac{1}{2}\,\mathbf{Tr}\, P^{-1}ss^\top + \mu^\top P^{-1}s
\end{aligned}$$

Notation: $\log f(x) =^+ g(x) \iff f(x) \propto \exp(g(x)) \iff \exists c \in \mathbb{R} : f(x) = c \exp(g(x))$

Gaussian potentials

Consider a Gaussian potential with mean μ and covariance Σ on x.

$$\phi(x) = \alpha\mathcal{N}(\mu, \Sigma) \quad (1)$$

$$= \alpha|2\pi\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)) \quad (2)$$

where $\int dx\phi(x) = \alpha$ and $|2\pi\Sigma|$ is a short notation for $(2\pi)^d \det \Sigma$, where Σ is $d \times d$..

- If $\alpha = 1$ the potential is normalized.
- A general Gaussian potential ϕ need not to be normalized so α is in fact an arbitrary positive constant.
- The exponent is just a quadratic form.

Canonical Form

$$\begin{aligned}\phi(x) &= \exp(\{\log \alpha - \frac{1}{2}\log|2\pi\Sigma| - \frac{1}{2}\mu^T\Sigma^{-1}\mu\} + \mu^T\Sigma^{-1}x - \frac{1}{2}x^T\Sigma^{-1}x) \\ &= \exp(g + h^T x - \frac{1}{2}x^T K x)\end{aligned}$$

- Alternative to the conventional and intuitive moment form.
- Here we represent the potential by the polynomial coefficients h and K.
- Coefficients h and K as natural parameters.

Canonical and Moment parametrisations

The moment parameters and canonical parameters are related by

$$
\begin{aligned}
K &= \Sigma^{-1} \\
h &= \Sigma^{-1}\mu \\
g &= \log\alpha - \frac{1}{2}\log|2\pi\Sigma| - \frac{1}{2}\mu^T\Sigma^{-1}\Sigma\Sigma^{-1}\mu \\
&= \log\alpha + \frac{1}{2}\log|\frac{K}{2\pi}| - \frac{1}{2}h^T K^{-1} h
\end{aligned}
$$

Jointly Gaussian Vectors

- Moment form

$$\phi(x_1, x_2) \quad = \quad \alpha \mathcal{N}(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix})$$

$$\phi = \alpha |2\pi\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix})$$

- Canonical form

$$\phi(x_1, x_2) = \exp(g + \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix})$$

- need to find a parametric representation of $K = \Sigma^{-1}$ in terms of the partitions $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$.

Partitioned Matrix Inverse

- Strategy: We will find two matrices X and Z such that W becomes block diagonal.

$$
\begin{aligned}
L\Sigma R &= W \\
\Sigma &= L^{-1} W R^{-1} \\
\Sigma^{-1} &= R W^{-1} L = K
\end{aligned}
$$

Gauss Transformations

- Add a multiple of row s to row t
- Premultiply Σ with $L(s,t)$ where

$$L_{i,j}(s,t) = \begin{cases} 1, & i = j \\ \gamma, & i = s \text{ and } j = t \\ 0, & \text{o/w} \end{cases}$$

- Example: $s = 2$, $t = 1$

$$\begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + \gamma c & b + \gamma d \\ c & d \end{pmatrix}$$

- The inverse just subtracts what is added

$$\begin{pmatrix} 1 & \gamma \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -\gamma \\ 0 & 1 \end{pmatrix}$$

Gauss Transformations

- Given Σ, add a multiple of column s to column t
- Postmultiply Σ with $R(s,t)$ where

$$R_{i,j}(s,t) = \begin{cases} 1, & i = j \\ \gamma, & j = s \text{ and } i = t \\ 0, & \text{o/w} \end{cases}$$

- Example: $s = 2$, $t = 1$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} = \begin{pmatrix} a + \gamma b & b \\ c + \gamma d & d \end{pmatrix}$$

Scalar example

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
L\Sigma &= \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a - bd^{-1}c & \cancel{b - bd^{-1}d} \\ c & d \end{pmatrix} \\
L\Sigma R &= \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix} \\
&= \begin{pmatrix} a - bd^{-1}c & 0 \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix} \\
&= \begin{pmatrix} a - bd^{-1}c & 0 \\ \cancel{c - dd^{-1}c} & d \end{pmatrix} = \begin{pmatrix} a - bd^{-1}c & 0 \\ 0 & d \end{pmatrix} = W
\end{aligned}
$$

Scalar example (cont)

$$
\begin{aligned}
\Sigma &= L^{-1} W R^{-1} \\
\Sigma^{-1} &= R W^{-1} L \\
&= \begin{pmatrix} 1 & 0 \\ -d^{-1}c & 1 \end{pmatrix} \begin{pmatrix} (a - bd^{-1}c)^{-1} & 0 \\ 0 & d^{-1} \end{pmatrix} \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} (a - bd^{-1}c)^{-1} & 0 \\ -d^{-1}c(a - bd^{-1}c)^{-1} & d^{-1} \end{pmatrix} \begin{pmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} (a - bd^{-1}c)^{-1} & -(a - bd^{-1}c)^{-1}bd^{-1} \\ -d^{-1}c(a - bd^{-1}c)^{-1} & d^{-1} + d^{-1}c(a - bd^{-1}c)^{-1}bd^{-1} \end{pmatrix}
\end{aligned}
$$

Scalar example

We could also use

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
L\Sigma &= \begin{pmatrix} 1 & 0 \\ -ca^{-1} & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \\
L\Sigma R &= \begin{pmatrix} 1 & 0 \\ -ca^{-1} & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & -a^{-1}b \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} a & 0 \\ 0 & d - ca^{-1}b \end{pmatrix} = W \\
RW^{-1}L &= \begin{pmatrix} a^{-1} + a^{-1}b\left(d - ca^{-1}b\right)^{-1} ca^{-1} & -a^{-1}b\left(d - ca^{-1}b\right)^{-1} \\ -\left(d - ca^{-1}b\right)^{-1} ca^{-1} & \left(d - ca^{-1}b\right)^{-1} \end{pmatrix}
\end{aligned}
$$

Partitioned Matrix Inverse

In matrix case, this leads to following dual factorizations of Σ as

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \\
&= \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{pmatrix}
\end{aligned}
$$

The Schur Complement

We will introduce the notation

$$\begin{aligned}\Sigma/\Sigma_{22} &= \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21} \\ \Sigma/\Sigma_{11} &= \Sigma_{22} - \Sigma_{21}(\Sigma_{11})^{-1}\Sigma_{12}\end{aligned}$$

Determinant

$$|\Sigma| = |\Sigma/\Sigma_{11}||\Sigma_{11}| = |\Sigma/\Sigma_{22}||\Sigma_{22}|$$

$$\begin{aligned}\Sigma^{-1} &= \begin{pmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \begin{pmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma/\Sigma_{11})^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}\end{aligned}$$

Partitioned Matrix Inverse

$$
\begin{aligned}
\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} (\Sigma/\Sigma_{22})^{-1} & -(\Sigma/\Sigma_{22})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma/\Sigma_{22})^{-1} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}(\Sigma/\Sigma_{22})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma/\Sigma_{11})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma/\Sigma_{11})^{-1} \\ -(\Sigma/\Sigma_{11})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & (\Sigma/\Sigma_{11})^{-1} \end{pmatrix}
\end{aligned}
$$

- Quite complicated looking formulas, but straightforward to implement
- **Caution:** $\Sigma_{11}^{-1} \neq K_{11}$ in general!

Matrix Inversion Lemma

- Read the diagonal entries

$$
\begin{aligned}
(\Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21})^{-1} &= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma/\Sigma_{11})^{-1}\Sigma_{21}\Sigma_{11}^{-1} \\
\left(A - BC^{-1}D\right)^{-1} &= A^{-1} + A^{-1}B\left(C - DA^{-1}B\right)^{-1}DA^{-1}
\end{aligned}
$$

Factorisation of Multivariate Gaussians

Consider the joint distribution over the variable

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

where the joint distribution is Gaussian $p(x) = \mathcal{N}(x; \mu, \Sigma)$ with

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_2 \end{pmatrix}$$

Factorisation of Multivariate Gaussians

Find the following

1. Conditionals
   (a) $p(x_1|x_2)$
   (b) $p(x_2|x_1)$
2. Marginals
   (a) $p(x_1)$
   (b) $p(x_2)$

Factorisation of Multivariate Gaussians

Using the partitioned inverse equations, we rearrange

$$p(x_1, x_2) \quad \propto \quad \exp\left(-\frac{1}{2} \left(\begin{array}{c} x_1 - \mu_1 \\ x_2 - \mu_2 \end{array} \right)^\top \left(\begin{array}{cc} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_2 \end{array} \right)^{-1} \left(\begin{array}{c} x_1 - \mu_1 \\ x_2 - \mu_2 \end{array} \right) \right)$$

bring the expression in form of $p(x_1)p(x_2|x_1)$ (or $p(x_2)p(x_1|x_2)$) where the marginal and conditional can be easily identified. (See also Bishop, section 2.3.)

Factorisation of Multivariate Gaussians

We have the two decompositions

$$
\begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_2 \end{pmatrix}^{-1} = \Bigg(\begin{matrix} \big(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\big)^{-1} & -\big(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\big)^{-1} \\ -\Sigma_2^{-1}\Sigma_{12}^\top\big(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\big)^{-1} & \Sigma_2^{-1} + \Sigma_2^{-1}\Sigma_{12}^\top\big(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1} \end{matrix}
$$

$$
= \Bigg(\begin{matrix} \Sigma_1^{-1} + \Sigma_1^{-1}\Sigma_{12}\big(\Sigma_2 - \Sigma_{12}^\top\Sigma_1^{-1}\Sigma_{12}\big)^{-1}\Sigma_{12}^\top\Sigma_1^{-1} & -\Sigma_1^{-1}\Sigma_{12}\big(\Sigma_2 - \\ -\big(\Sigma_2 - \Sigma_{12}^\top\Sigma_1^{-1}\Sigma_{12}\big)^{-1}\Sigma_{12}^\top\Sigma_1^{-1} & \big(\Sigma_2 - \Sigma_{12}^\top\Sigma \end{matrix}
$$

We let $s_i = x_i - \mu_i$ and use the first decomposition.

$$
p(s_1, s_2) \propto \exp\left(-\frac{1}{2} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}^\top \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^\top & \Sigma_2 \end{pmatrix}^{-1} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \right)
$$

$$
\begin{aligned}
= \quad & \exp\left(-\frac{1}{2}\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}^\top \begin{pmatrix} \left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1} & -\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1}\Sigma_{12}\Sigma \\ -\Sigma_2^{-1}\Sigma_{12}^\top\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1} & \Sigma_2^{-1} + \Sigma_2^{-1}\Sigma_{12}^\top\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-} \end{pmatrix}\right. \\
= \quad & \exp(-\frac{1}{2}s_1^\top\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1}s_1 \\
& s_2^\top\Sigma_2^{-1}\Sigma_{12}^\top\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1}s_1 \\
& -\frac{1}{2}s_2^\top\Sigma_2^{-1}\Sigma_{12}^\top\left(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top\right)^{-1}\Sigma_{12}\Sigma_2^{-1}s_2 \\
& -\frac{1}{2}s_2^\top\Sigma_2^{-1}s_2) \\
\propto \quad & \mathcal{N}(s_1; \Sigma_{12}\Sigma_2^{-1}s_2, \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top)\mathcal{N}(s_2; 0, \Sigma_2) \\
= \quad & \mathcal{N}(x_1; \mu_1 + \Sigma_{12}\Sigma_2^{-1}(x_2 - \mu_2), \Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{12}^\top)\mathcal{N}(x_2; \mu_2, \Sigma_2)
\end{aligned}
$$

This leads to a factorisation of form $p(x_2)p(x_1|x_2)$. The second decomposition will lead to the other factorisation $p(x_1)p(x_2|x_1)$.

Keywords Summary

Multivariate Bernoulli

Multivariate Gaussian

Partitioned Inverse Equations, Matrix Inversion Lemma