

Multivariate normal distribution

1 Bivariate normal distribution

Bivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

Contour of bivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

2 Multivariate normal distribution

Multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

Multivariate normal distribution - Joint MGF

X and Y form the same multivariate normal \mathbf{x} , where $Cov(X, Y) = 0$

3 How to generate random samples from $\mathcal{N}(\mu, \Sigma)$

How to generate random samples from $\mathcal{N}(\mu, \Sigma)$

Example - How to generate random samples from $\mathcal{N}(\mu, \Sigma)$

4 Joint, marginal, and conditional distribution of $\mathcal{N}(\mu, \Sigma)$

Properties of Σ and Λ

Joint, marginal, and conditional distribution of $\mathcal{N}(\mu, \Sigma)$ - Part 1

Joint, marginal, and conditional distribution of $\mathcal{N}(\mu, \Sigma)$ - Part 2

Bivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

PDF

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1} (\mathbf{x}-\mu)}$$

where

mean μ

covariance matrix Σ

determinant of the covariance matrix Σ $|\Sigma|$

Covariance matrix and its inverse

$$\begin{aligned} \Sigma &= \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \Rightarrow |\Sigma| = (1 - \rho^2)\sigma_x^2\sigma_y^2 \\ &\Rightarrow \Sigma^{-1} = \frac{1}{(1 - \rho^2)\sigma_x^2\sigma_y^2} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix} \end{aligned}$$

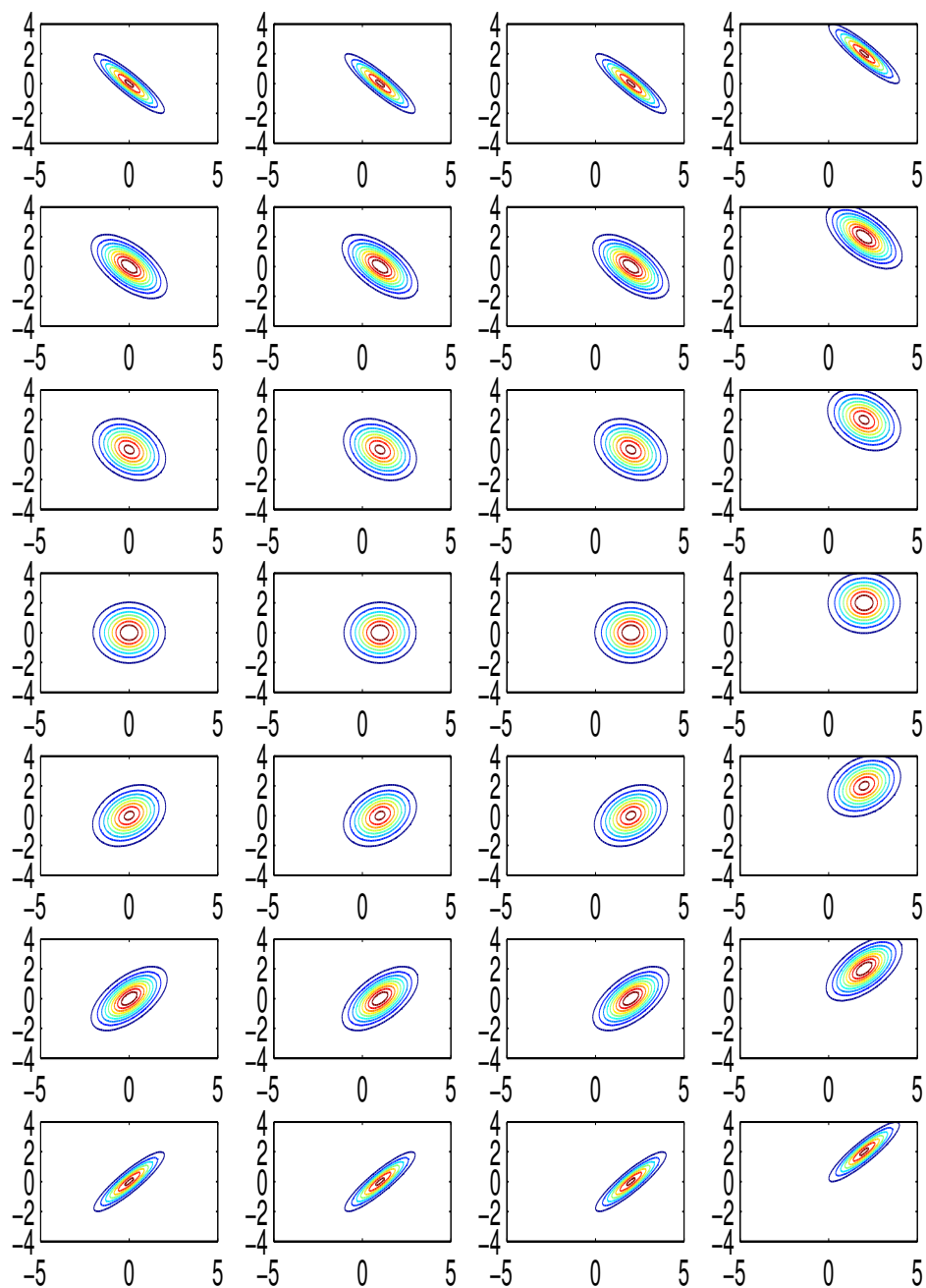
Another form of PDF

With $\tilde{x} = \frac{x - \mu_x}{\sigma_x}$, $\tilde{y} = \frac{y - \mu_y}{\sigma_y}$

$$(\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) = \frac{\tilde{x}^2 + \tilde{y}^2 - 2\rho\tilde{x}\tilde{y}}{1 - \rho^2}$$

$$f(\mathbf{x}) = \frac{1}{2\pi} e^{-\frac{\tilde{x}^2 + \tilde{y}^2}{2}} \quad \text{if } X \text{ and } Y \text{ are independent}$$

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-\frac{\tilde{x}^2 + \tilde{y}^2 - 2\rho\tilde{x}\tilde{y}}{2(1 - \rho^2)}} \quad \text{in general}$$



```
clear all; close all; clc;

for i=1:7
    for j=1:4

        % Choose mean
        Choose_mean = j;
        switch Choose_mean
            case 1; mu = [0 0];
            case 2; mu = [1 0];
            case 3; mu = [2 0];
            case 4; mu = [2 2];
        end

        % Choose covariance matrix
        Choose_covariance_matrix = i;
        switch Choose_covariance_matrix
            case 1; Sigma = [1 -0.9; -0.9 1];
            case 2; Sigma = [1 -0.6; -0.6 1];
            case 3; Sigma = [1 -0.3; -0.3 1];
            case 4; Sigma = [1 0; 0 1];
            case 5; Sigma = [1 0.3; 0.3 1];
            case 6; Sigma = [1 0.6; 0.6 1];
            case 7; Sigma = [1 0.9; 0.9 1];
        end

        x = -5:0.1:5;
        y = -4:0.1:4;
        [X,Y] = meshgrid_Lee(x,y);
        F = mvnpdf([X(:) Y(:)],mu,Sigma);
        F = reshape(F,length(x),length(y));

        subplot(7,4,4*i+j-4)
        % mesh(X,Y,F);
        contour(X,Y,F);

    end
end
```

Multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

PDF

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)' \Sigma^{-1} (\mathbf{x}-\mu)}$$

where

mean μ

covariance matrix Σ

determinant of the covariance matrix Σ $|\Sigma|$

Definition

Multivariate normal $\mathbf{x} = [x_1, x_2, \dots, x_d]'$ is given by

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dn} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}}_{\mathbf{z}} + \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}}_{\boldsymbol{\mu}},$$

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu},$$

where z_k are iid standard normal and where \mathbf{A} and $\boldsymbol{\mu}$ are constants.

Computation of mean and covariance matrix

$$\mathbb{E}\mathbf{x} = \mathbb{E}(\mathbf{A}\mathbf{z} + \boldsymbol{\mu}) = \mathbf{A}\mathbb{E}\mathbf{z} + \boldsymbol{\mu} = \mathbf{A}\mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu}$$

$$\Sigma = \mathbb{E}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T = \mathbb{E}(\mathbf{A}\mathbf{z})(\mathbf{A}\mathbf{z})^T = \mathbf{A}(\mathbb{E}\mathbf{z}\mathbf{z}^T)\mathbf{A}^T = \mathbf{A}\mathbf{I}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T$$

Joint MGF of multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

$$\phi(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad \text{for } x \sim \mathcal{N}(\mu, \sigma^2)$$

$$\phi(\mathbf{t}) = e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} \quad \text{for } \mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$$

With $\mathbf{z} = (z_k)^T$ iid $N(0, 1^2)$

$$\mathbf{x} \sim \mathcal{N}(\mu, \Sigma) \Rightarrow \mathbf{x} = \mathbf{A}\mathbf{z} + \mu \quad \text{where } \mathbf{A}\mathbf{A}^T = \Sigma$$

$$\mathbf{t}^T \mathbf{x} = \mathbf{t}^T (\mathbf{A}\mathbf{z} + \mu) = \sum_k a_k z_k + b \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

where

$$\mu_1 = \mathbb{E} \mathbf{t}^T (\mathbf{A}\mathbf{z} + \mu) = \mathbf{t}^T (\mathbf{A} \mathbb{E} \mathbf{z} + \mu) = \mathbf{t}^T (\mathbf{A} \mathbf{0} + \mu) = \mathbf{t}^T \mu$$

$$\sigma_1^2 = \mathbb{E} (\mathbf{t}^T \mathbf{A} \mathbf{z}) (\mathbf{t}^T \mathbf{A} \mathbf{z})^T = \mathbf{t}^T \mathbf{A} \mathbb{E} (\mathbf{z} \mathbf{z}^T) \mathbf{A}^T \mathbf{t} = \mathbf{t}^T \mathbf{A} \mathbf{I} \mathbf{A}^T \mathbf{t} = \mathbf{t}^T \mathbf{A} \mathbf{A}^T \mathbf{t} = \mathbf{t}^T \Sigma \mathbf{t}$$

$$\phi_{\mathbf{x}}(\mathbf{t}) = \mathbb{E} e^{\mathbf{t}^T \mathbf{x}} = \mathbb{E} e^{\mathbf{t}^T (\mathbf{A}\mathbf{z} + \mu)} = \phi_{\mathcal{N}(\mu_1, \sigma_1^2)}(1) = e^{\mu_1 + \frac{1}{2} \sigma_1^2} = e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}}$$

Properties of multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$

- (1) μ and Σ completely determine the multivariate normal distribution
- (2) If off diagonals of Σ are all 0, then all the components of \mathbf{x} are independent
- (3) If for fixed i , $\Sigma_{ij} = 0$ for all $j \neq i$, then \mathbf{x}_i is independent to \mathbf{x}_j , $j \neq i$

Suppose two multivariate normal random variables \mathbf{x} and \mathbf{y} have common mean μ and covariance matrix Σ . Then, their joint MGFs of \mathbf{x} and \mathbf{y} are identical:

$$\phi_{\mathbf{x}}(\mathbf{t}) = e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} = \phi_{\mathbf{y}}(\mathbf{t})$$

Hence, they have same distribution.

Suppose a multivariate normal random variable \mathbf{x} have a covariance matrix Σ , whose off diagonals are all 0. Then, joint MGFs of \mathbf{x} and \mathbf{y} , where \mathbf{y}_i are independent $\mathcal{N}(\mu_i, \sigma_i^2)$, are identical:

$$\phi_{\mathbf{x}}(\mathbf{t}) = \phi_{\mathcal{N}(\mu, \Sigma)}(\mathbf{t}) = e^{\mathbf{t}^T \mu + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t}} = \underbrace{e^{\mu_1 t_1 + \frac{1}{2} \sigma_1^2 t_1^2}}_{\phi_{N(\mu_1, \sigma_1^2)}(t_1)} \underbrace{e^{\mu_2 t_2 + \frac{1}{2} \sigma_2^2 t_2^2}}_{\phi_{N(\mu_2, \sigma_2^2)}(t_2)} \dots \underbrace{e^{\mu_d t_d + \frac{1}{2} \sigma_d^2 t_d^2}}_{\phi_{N(\mu_d, \sigma_d^2)}(t_d)} = \phi_{\mathbf{y}}(\mathbf{t})$$

Hence, they have same distribution. In particular, **all the components x_i of the multivariate normal random variable \mathbf{x} are independent.**

Suppose a multivariate normal random variable \mathbf{x} have a covariance matrix Σ such that for fixed i , $\Sigma_{ij} = 0$ for all $j \neq i$. Then, joint MGFs of \mathbf{x} and \mathbf{y} , where the mean and covariance matrix of \mathbf{y} are identical to those of \mathbf{x} and where in addition \mathbf{y}_i is independent to \mathbf{y}_j , $j \neq i$, are identical. Hence, they have same distribution. In particular, **\mathbf{x}_i is independent to \mathbf{x}_j , $j \neq i$.**

X and Y from the same multivariate normal \mathbf{x} , where $Cov(X, Y) = 0$

$$X, Y \text{ independent} \Rightarrow Cov(X, Y) = 0$$

Converse is not true in general. However, if $[X, Y]^T$ bivariate normal, its a completely different story.

$$\begin{array}{ll} Cov(X, Y) = 0 \Rightarrow X, Y \text{ independent (wrong)} \\ X, Y \text{ normal, } Cov(X, Y) = 0 \Rightarrow X, Y \text{ independent (wrong)} \\ [X, Y]^T \text{ bivariate normal, } Cov(X, Y) = 0 \Rightarrow X, Y \text{ independent (correct)} \end{array}$$

Let X be standard normal. Independent to X we flip a fair coin and record its output S as 1 if head and -1 if tail. Now, define Y as

$$Y = S \cdot X$$

Y is standard normal.

$$\begin{aligned} P(Y \leq y) &= P(S = 1)P(X \leq y) + P(S = -1)P(X \geq -y) \\ &= \frac{1}{2} \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds + \frac{1}{2} \int_{-y}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \\ f_Y(y) &= \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \end{aligned}$$

$$\begin{aligned} EXY &= P(S = 1)E(XY|S = 1) + P(S = -1)E(XY|S = -1) \\ &= P(S = 1)E(X^2) + P(S = -1)E(-X^2) = \frac{1}{2}E(X^2) - \frac{1}{2}E(X^2) = 0 \end{aligned}$$

$$Cov(X, Y) = 0$$

But, by the construction of Y , X and Y are not independent. For example, if $X = 2$, then Y is either 2 or -2 .

How to generate random samples from $\mathcal{N}(\mu, \Sigma)$

Cholesky decomposition of covariance or correlation matrix Σ

$$\Sigma = \mathbf{L}\mathbf{U}$$

where \mathbf{L} is an lower triangular matrix, \mathbf{U} is an upper triangular matrix, and

$$\mathbf{L} = \mathbf{U}^T$$

[Matlab] chol

$$\mathbf{U} = \text{chol}(\text{Sigma})$$

How to generate random samples from $\mathcal{N}(\mu, \Sigma)$

[Step 1] Do Cholesky decomposition, i.e., $\mathbf{U} = \text{chol}(\text{Sigma})$.

[Step 2] With iid standard normal samples \mathbf{z}

$$\mathbf{x} = \mu + \mathbf{L}\mathbf{z} = \mu + \mathbf{U}^T\mathbf{z}$$

$$\mathbb{E}\mathbf{x} = \mathbb{E}(\mu + \mathbf{L}\mathbf{z}) = \mu + \mathbf{L}\mathbb{E}\mathbf{z} = \mu + \mathbf{L}\mathbf{0} = \mu$$

$$\mathbb{E}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T = \mathbb{E}(\mathbf{L}\mathbf{z})(\mathbf{L}\mathbf{z})^T = \mathbf{L}(\mathbb{E}\mathbf{z}\mathbf{z}^T)\mathbf{L}^T = \mathbf{L}\mathbf{I}\mathbf{U} = \mathbf{L}\mathbf{U} = \Sigma$$

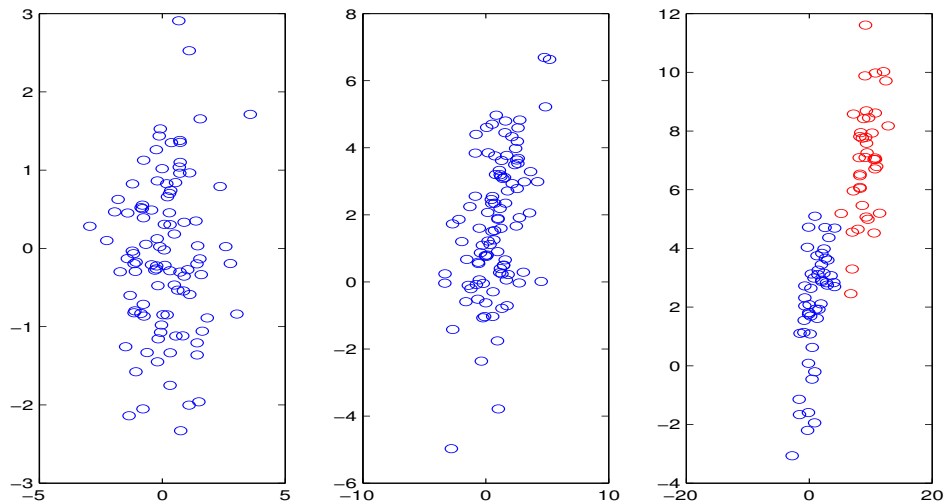


Figure 1: 100 standard normal samples in the plane (left), 100 samples from $N(\mu, \Sigma)$ (center), and 50 (blue) samples from $N(\mu_1, \Sigma_1)$ and 40 (red) samples from $N(\mu_2, \Sigma_2)$ (right).

```
clear all; close all; clc; rng('default');

subplot(131)
n=100; x=randn(n,2); plot(x(:,1),x(:,2),'o')

subplot(132)
n=100;
Mu=[1 2]'; Si=[3 2; 2 5];
U=chol(Si);
x= repmat(Mu,1,n)+U'*randn(2,n);
plot(x(1,:),x(2:,:), 'o')

subplot(133)
n1=50;
Mu1=[1 2]'; Si1=[3 2; 2 5];
U1=chol(Si1); x1=repmat(Mu1,1,n1)+U1'*randn(2,n1);
plot(x1(1,:),x1(2:,:), 'o'); hold on
n2=40;
Mu2=[9 7]'; Si2=[3 1; 2 3];
U2=chol(Si2); x2=repmat(Mu2,1,n2)+U2'*randn(2,n2);
plot(x2(1,:),x2(2:,:), 'or')
```

Properties of Σ and Λ

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$$

$$\Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\Sigma\Lambda = \begin{bmatrix} \Sigma_{11}\Lambda_{11} + \Sigma_{12}\Lambda_{21} & \Sigma_{11}\Lambda_{12} + \Sigma_{12}\Lambda_{22} \\ \Sigma_{21}\Lambda_{11} + \Sigma_{22}\Lambda_{21} & \Sigma_{21}\Lambda_{12} + \Sigma_{22}\Lambda_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0}_{12} \\ \mathbf{0}_{21} & \mathbf{I}_{22} \end{bmatrix}$$

$$\begin{array}{c} \text{Remove } \Lambda_{21} \\ \Rightarrow \end{array} \Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\Sigma_{11}\Lambda_{12} = -\Sigma_{12}\Lambda_{22}$$

$$\Sigma\Lambda = \begin{bmatrix} \Sigma_{11}\Lambda_{11} + \Sigma_{12}\Lambda_{21} & \Sigma_{11}\Lambda_{12} + \Sigma_{12}\Lambda_{22} \\ \Sigma_{21}\Lambda_{11} + \Sigma_{22}\Lambda_{21} & \Sigma_{21}\Lambda_{12} + \Sigma_{22}\Lambda_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0}_{12} \\ \mathbf{0}_{21} & \mathbf{I}_{22} \end{bmatrix}$$

$$\Sigma_{21}\Lambda_{12} = -\Sigma_{22}\Lambda_{22} + \mathbf{I}_{22}$$

$$\Sigma\Lambda = \begin{bmatrix} \Sigma_{11}\Lambda_{11} + \Sigma_{12}\Lambda_{21} & \Sigma_{11}\Lambda_{12} + \Sigma_{12}\Lambda_{22} \\ \Sigma_{21}\Lambda_{11} + \Sigma_{22}\Lambda_{21} & \Sigma_{21}\Lambda_{12} + \Sigma_{22}\Lambda_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0}_{12} \\ \mathbf{0}_{21} & \mathbf{I}_{22} \end{bmatrix}$$

$$\Lambda_{11}^{-1}\Lambda_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$$

$$\begin{aligned} \Lambda_{11}^{-1}\Lambda_{12} &= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})\Lambda_{12} \\ &= \Sigma_{11}\Lambda_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Lambda_{12} \\ &= -\Sigma_{12}\Lambda_{22} - \Sigma_{12}\Sigma_{22}^{-1}(-\Sigma_{22}\Lambda_{22} + \mathbf{I}_{22}) \\ &= -\Sigma_{12}\Sigma_{22}^{-1} \end{aligned}$$

Joint, marginal, and conditional distribution of $\mathcal{N}(\mu, \Sigma)$ - Part 1

Joint

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

Marginal

$$\mathbf{x}_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$$

Conditional

$$\mathbf{x}_1 | \mathbf{x}_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$$

where

$$\begin{aligned} \mu_{1|2} &= \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (\mathbf{x}_2 - \mu_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2) \\ \Sigma_{1|2} &= \Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \end{aligned}$$

With $\mathbf{y}_i = \mathbf{x}_i - \mu_i$

$$\begin{aligned} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) &= \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T \end{bmatrix} \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \\ &= \mathbf{y}_1^T \Lambda_{11} \mathbf{y}_1 + \mathbf{y}_1^T \Lambda_{12} \mathbf{y}_2 + \mathbf{y}_2^T \Lambda_{21} \mathbf{y}_1 + \mathbf{y}_2^T \Lambda_{22} \mathbf{y}_2 \\ &\propto \mathbf{y}_1^T \Lambda_{11} \mathbf{y}_1 + \mathbf{y}_1^T \Lambda_{12} \mathbf{y}_2 + \mathbf{y}_2^T \Lambda_{21} \mathbf{y}_1 \\ &\propto (\mathbf{y}_1 - \alpha)^T \Lambda_{11} (\mathbf{y}_1 - \alpha) \end{aligned}$$

$$-\mathbf{y}_1^T \Lambda_{11} \alpha = \mathbf{y}_1^T \Lambda_{12} \mathbf{y}_2 \Rightarrow -\Lambda_{11} \alpha = \Lambda_{12} \mathbf{y}_2 \Rightarrow \alpha = -\Lambda_{11}^{-1} \Lambda_{12} \mathbf{y}_2$$

$$\mathbf{y}_1 - \alpha = \mathbf{x}_1 - \mu_1 + \Lambda_{11}^{-1} \Lambda_{12} \mathbf{y}_2 = \mathbf{x}_1 - (\mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (\mathbf{x}_2 - \mu_2)) := \mathbf{x}_1 - \mu_{1|2}$$

$$\Sigma_{1|2} = \Lambda_{11}^{-1} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Joint, marginal, and conditional distribution of $\mathcal{N}(\mu, \Sigma)$ - Part 2

Joint

$$\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_{\varepsilon}) \quad \varepsilon \text{ independent to } \mathbf{x}$$

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mathbf{A}\mu_{\mathbf{x}} + \mathbf{b} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{x}} & \Sigma_{\mathbf{x}}\mathbf{A}^T \\ \mathbf{A}\Sigma_{\mathbf{x}} & \mathbf{A}\Sigma_{\mathbf{x}}\mathbf{A}^T + \Sigma_{\varepsilon} \end{bmatrix} \right)$$

Precision

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\varepsilon}^{-1} \mathbf{A} & -\mathbf{A}^T \Sigma_{\varepsilon}^{-1} \\ -\Sigma_{\varepsilon}^{-1} \mathbf{A} & \Sigma_{\varepsilon}^{-1} \end{bmatrix}$$

Marginal

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}\mu_{\mathbf{x}} + \mathbf{b}, \mathbf{A}\Sigma_{\mathbf{x}}\mathbf{A}^T + \Sigma_{\varepsilon})$$

Conditional

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$$

where

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) \\ &= \mu_{\mathbf{x}} + \Sigma_{\mathbf{x}}\mathbf{A}^T(\mathbf{A}\Sigma_{\mathbf{x}}\mathbf{A}^T + \Sigma_{\varepsilon})^{-1}(\mathbf{y} - (\mathbf{A}\mu_{\mathbf{x}} + \mathbf{b})) \end{aligned}$$

$$\begin{aligned} \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{\mathbf{x}} - \Sigma_{\mathbf{x}}\mathbf{A}^T(\mathbf{A}\Sigma_{\mathbf{x}}\mathbf{A}^T + \Sigma_{\varepsilon})^{-1}\mathbf{A}\Sigma_{\mathbf{x}} \end{aligned}$$

or using Woodbury identity $(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}$

$$\mu_{1|2} = \Sigma_{1|2}(\mathbf{A}^T \Sigma_{\varepsilon}^{-1}(\mathbf{y} - \mathbf{b}) + \Sigma_{\mathbf{x}}^{-1}\mu_{\mathbf{x}})$$

$$\Sigma_{1|2}^{-1} = \Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\varepsilon}^{-1} \mathbf{A}$$

With $\tilde{\mathbf{x}} = \mathbf{x} - \mu_{\mathbf{x}}$, $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{A}\mu_{\mathbf{x}} - \mathbf{b}$

$$\begin{aligned}
\log p(\mathbf{x}, \mathbf{y}) &= \log p(\mathbf{x}) + \log p(\mathbf{y}|\mathbf{x}) \\
&\propto -\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{x}})^T \Sigma_{\mathbf{x}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}}) - \frac{1}{2}(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b})^T \Sigma_{\epsilon}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{b}) \\
&= -\frac{1}{2}\tilde{\mathbf{x}}^T \Sigma_{\mathbf{x}}^{-1}\tilde{\mathbf{x}} - \frac{1}{2}(\tilde{\mathbf{y}} - \mathbf{A}\tilde{\mathbf{x}})^T \Sigma_{\epsilon}^{-1}(\tilde{\mathbf{y}} - \mathbf{A}\tilde{\mathbf{x}}) \\
&= -\frac{1}{2} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} \Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A} & -\mathbf{A}^T \Sigma_{\epsilon}^{-1} \\ -\Sigma_{\epsilon}^{-1} \mathbf{A} & \Sigma_{\epsilon}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \\
&:= -\frac{1}{2} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \\
&:= -\frac{1}{2} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}
\end{aligned}$$

$$\Rightarrow \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A} & -\mathbf{A}^T \Sigma_{\epsilon}^{-1} \\ -\Sigma_{\epsilon}^{-1} \mathbf{A} & \Sigma_{\epsilon}^{-1} \end{bmatrix}$$

$$\Rightarrow \quad \Sigma_{1|2}^{-1} = \Lambda_{11} = \Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A}$$

$$\begin{aligned}
\Rightarrow \quad \mu_{1|2} &= \mu_1 - \Lambda_{11}^{-1} \Lambda_{12}(\mathbf{x}_2 - \mu_2) \\
&= \Sigma_{1|2} \Sigma_{1|2}^{-1} \mu_{\mathbf{x}} + \Sigma_{1|2} \mathbf{A}^T \Sigma_{\epsilon}^{-1}(\mathbf{y} - \mathbf{A}\mu_{\mathbf{x}} - \mathbf{b}) \\
&= \Sigma_{1|2}(\Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A})\mu_{\mathbf{x}} + \Sigma_{1|2} \mathbf{A}^T \Sigma_{\epsilon}^{-1}(\mathbf{y} - \mathbf{A}\mu_{\mathbf{x}} - \mathbf{b}) \\
&= \Sigma_{1|2} [(\Sigma_{\mathbf{x}}^{-1} + \mathbf{A}^T \Sigma_{\epsilon}^{-1} \mathbf{A})\mu_{\mathbf{x}} + \mathbf{A}^T \Sigma_{\epsilon}^{-1}(\mathbf{y} - \mathbf{A}\mu_{\mathbf{x}} - \mathbf{b})] \\
&= \Sigma_{1|2} [\Sigma_{\mathbf{x}}^{-1} \mu_{\mathbf{x}} + \mathbf{A}^T \Sigma_{\epsilon}^{-1}(\mathbf{y} - \mathbf{b})]
\end{aligned}$$