Robot Learning

Probability

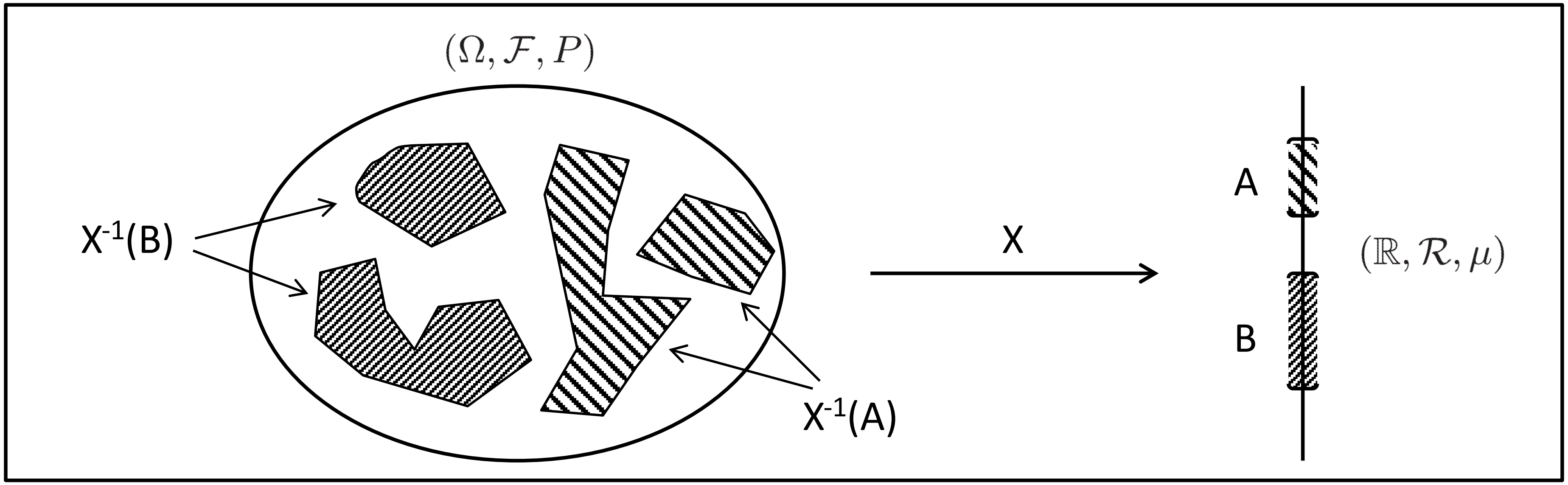
Prof. Songhwai Oh
ECE, SNU

REVIEW OF PROBABILITY THEORY

Random Variables

- **Probability space**: $(\Omega, \mathcal{F}, P)$
  - Ω: Set of outcomes
  - $\mathcal{F}$: Set of events (σ-field)
  - $P : \mathcal{F} \to [0, 1]$: Function which assigns probabilities to events
- **Random variable**: A function $X : \Omega \to \mathbb{R}$ is a **random variable** if for every Borel set $B \subset \mathbb{R}$, we have

$$X^{-1}(B) = \{\omega \in \Omega \; : \; X(\omega) \in B\} \in \mathcal{F}. \tag{1}$$



Gaussian Random Variable

- X is a **Gaussian random variable** if X is a random variable having the following probability density function (**Gaussian or normal distribution**):

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

  - Mean: $\mathbb{E}(X) = \int x f(x) dx = \mu$
  - Variance: $\mathbf{var}(X) = \mathbb{E}\left(X - \mathbb{E}(X)\right)^2 = \sigma^2$
  - Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$

- **Central limit theorem**: Let $X_1, X_2, \ldots$ be independent and identically distributed with $\mathbb{E}(X_i) = \mu$ and $\mathbf{var}(X_i) = \sigma^2 < \infty$. If $S_n = X_1 + \cdots + X_n$, then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} X,$$

where X has the standard normal distribution, i.e., $X \sim \mathcal{N}(0, 1)$.

Note: It can be generalized such that a sum of a large number of small errors has approximately a normal distribution (under some technical conditions).

Multivariate Gaussian

- A random vector $\mathbf{x} = [X_1 \ \ldots \ X_n]^T$ is said to be **multivariate Gaussian** if every linear combination of the components of X is a Gaussian random variable.
  - That is, for any a_i, $\sum_{i=1}^{n} a_i X_i$ is a Gaussian random variable.
  - We also say $X_1, \ldots, X_n$ are **jointly Gaussian**.
- **Multivariate Gaussian density function**:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right) \tag{1}$$

$\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$, where μ is the mean vector and Σ is the covariance matrix.

$$\mu = \mathbb{E}(\mathbf{x}) = \begin{bmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{bmatrix} \qquad \Sigma = \mathbf{cov}(\mathbf{x}) = \mathbb{E}\left(\left((\mathbf{x}-\mu)(\mathbf{x}-\mu)^T\right)\right)$$

Conditional Density of Multivariate Gaussian

Theorem: If $\mathbf{x} \in \mathbb{R}^r$ and $\mathbf{y} \in \mathbb{R}^m$ are jointly Gaussian with $n = r + m$, mean vector $[\mathbb{E}(\mathbf{x})^T \ \mathbb{E}(\mathbf{y})^T]^T$, and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix},$$

then the conditional probability density function $p(\mathbf{x}|\mathbf{y})$ is also a Gaussian random vector with mean $\mathbb{E}(\mathbf{x}|\mathbf{y})$ and covariance matrix $\Sigma_{x|y}$, where

$$\begin{aligned} \mathbb{E}(\mathbf{x}|\mathbf{y}) &= \mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \mathbb{E}(\mathbf{y})) \\ \Sigma_{x|y} &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}. \end{aligned}$$

Proof:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \quad (1)$$

$$= \frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\begin{bmatrix} \mathbf{x} - \mathbb{E}(\mathbf{x}) \\ \mathbf{y} - \mathbb{E}(\mathbf{y}) \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x} - \mathbb{E}(\mathbf{x}) \\ \mathbf{y} - \mathbb{E}(\mathbf{y}) \end{bmatrix}\right)}{\frac{1}{(2\pi)^{m/2}|\Sigma_{yy}|^{1/2}} \exp\left(-\frac{1}{2}[\mathbf{y} - \mathbb{E}(\mathbf{y})]^T \Sigma_{yy}^{-1} [\mathbf{y} - \mathbb{E}(\mathbf{y})]\right)} \quad (2)$$

Proof Continued

$$p(\mathbf{x}|\mathbf{y}) = \frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\begin{bmatrix} \mathbf{x}-\mathbb{E}(\mathbf{x}) \\ \mathbf{y}-\mathbb{E}(\mathbf{y}) \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x}-\mathbb{E}(\mathbf{x}) \\ \mathbf{y}-\mathbb{E}(\mathbf{y}) \end{bmatrix}\right)}{\frac{1}{(2\pi)^{m/2}|\Sigma_{yy}|^{1/2}} \exp\left(-\frac{1}{2}\left[\mathbf{y}-\mathbb{E}(\mathbf{y})\right]^T \Sigma_{yy}^{-1} \left[\mathbf{y}-\mathbb{E}(\mathbf{y})\right]\right)} \tag{1}$$

$$= \frac{1}{(2\pi)^{r/2}\left(|\Sigma|/|\Sigma_{yy}|\right)^{1/2}} \exp\left(-\frac{1}{2}A\right), \tag{2}$$

where

$$A = \begin{bmatrix} \mathbf{x}-\mathbb{E}(\mathbf{x}) \\ \mathbf{y}-\mathbb{E}(\mathbf{y}) \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x}-\mathbb{E}(\mathbf{x}) \\ \mathbf{y}-\mathbb{E}(\mathbf{y}) \end{bmatrix} - \left[\mathbf{y}-\mathbb{E}(\mathbf{y})\right]^T \Sigma_{yy}^{-1} \left[\mathbf{y}-\mathbb{E}(\mathbf{y})\right]. \tag{3}$$

We now need to compute two terms:

- $|\Sigma|/|\Sigma_{yy}| = \det(\Sigma)/\det(\Sigma_{yy})$
- A

Proof Continued (Determinant)

We can easily verify that

$$\begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \underbrace{\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}}_{\Sigma} \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix}$$

Hence,

$$\det(\Sigma) = \det(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})\det(\Sigma_{yy}) \tag{1}$$

$$\frac{\det(\Sigma)}{\det(\Sigma_{yy})} = \det(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}). \tag{2}$$

So we have

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{(2\pi)^{r/2}\left(|\Sigma|/|\Sigma_{yy}|\right)^{1/2}} \exp\left(-\frac{1}{2}A\right) \tag{3}$$

$$= \frac{1}{(2\pi)^{r/2}\left(\det(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})\right)^{1/2}} \exp\left(-\frac{1}{2}A\right). \tag{4}$$

Proof Continued (A)

If $XYZ = W$ and matrices are invertible,

- By inverting both sides, we get $Z^{-1}Y^{-1}X^{-1} = W^{-1}$.
- Hence, $Y^{-1} = ZW^{-1}X$.

Since we have (where $S_{yy} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$)

$$\begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \underbrace{\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}}_{\Sigma} \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix} = \begin{bmatrix} S_{yy} & 0 \\ 0 & \Sigma_{yy} \end{bmatrix}, \tag{1}$$

$$\Sigma^{-1} = \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix} \begin{bmatrix} S_{yy}^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix}. \tag{2}$$

Now let $\mathbf{x}' = \mathbf{x} - \mathbb{E}(\mathbf{x})$ and $\mathbf{y}' = \mathbf{y} - \mathbb{E}(\mathbf{y})$. Then

$$A = \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} - \mathbf{y}'^T \Sigma_{yy}^{-1} \mathbf{y}'^T \tag{3}$$

$$= \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}^T \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix} \begin{bmatrix} S_{yy}^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}$$
$$- \mathbf{y}'^T \Sigma_{yy}^{-1} \mathbf{y}'^T \tag{4}$$

Proof Continued (A)

$$
\begin{aligned}
A &= \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix}^T \begin{bmatrix} I & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & I \end{bmatrix} \begin{bmatrix} S_{yy}^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} \\
&- \mathbf{y}'^T \Sigma_{yy}^{-1} \mathbf{y}'^T \\
&= \begin{bmatrix} \mathbf{x}' - \Sigma_{xy}\Sigma_{yy}^{-1}\mathbf{y}' \\ \mathbf{y}' \end{bmatrix}^T \begin{bmatrix} S_{yy}^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}' - \Sigma_{xy}\Sigma_{yy}^{-1}\mathbf{y}' \\ \mathbf{y}' \end{bmatrix} - \mathbf{y}'^T \Sigma_{yy}^{-1} \mathbf{y}'^T \\
&= \left(\mathbf{x}' - \Sigma_{xy}\Sigma_{yy}^{-1}\mathbf{y}'\right)^T S_{yy}^{-1} \left(\mathbf{x}' - \Sigma_{xy}\Sigma_{yy}^{-1}\mathbf{y}'\right) \\
&= \left(\mathbf{x} - \left(\mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}\left(\mathbf{y} - \mathbb{E}(\mathbf{y})\right)\right)\right)^T S_{yy}^{-1} \left(\mathbf{x} - \left(\mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}\left(\mathbf{y} - \mathbb{E}(\mathbf{y})\right)\right)\right)
\end{aligned}
$$

where $S_{yy} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$.
Hence,

$$
p(\mathbf{x}|\mathbf{y}) = \frac{1}{(2\pi)^{r/2}|\Sigma_{x|y}|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y})\right)^T \Sigma_{x|y}^{-1}\left(\mathbf{x} - \mathbb{E}(\mathbf{x}|\mathbf{y})\right)\right), \quad (1)
$$

i.e., $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbb{E}(\mathbf{x}|\mathbf{y}), \Sigma_{x|y})$, where

$$
\begin{aligned}
\mathbb{E}(\mathbf{x}|\mathbf{y}) &= \mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}\left(\mathbf{y} - \mathbb{E}(\mathbf{y})\right) \quad (2) \\
\Sigma_{x|y} &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}. \quad (3)
\end{aligned}
$$

Matrix Inversion Lemma

Suppose A_{11} and A_{22} are invertible and let $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. Then

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

If $XYZ = W$ and matrices are invertible, $Y^{-1} = ZW^{-1}X$. Hence,

$$\begin{aligned} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix} \end{aligned}$$

We also have

$$\begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{bmatrix} \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix} \begin{bmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \end{bmatrix} \end{aligned}$$

Matrix Inversion Lemma

Hence, we must have

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

or

$$\begin{aligned} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} &= A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} \\ (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} &= A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1} &= A_{11}^{-1} - A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} \\ (A_{11} + A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} &= A_{11}^{-1}A_{12}(A_{22} + A_{21}A_{11}^{-1}A_{12})^{-1} \end{aligned}$$

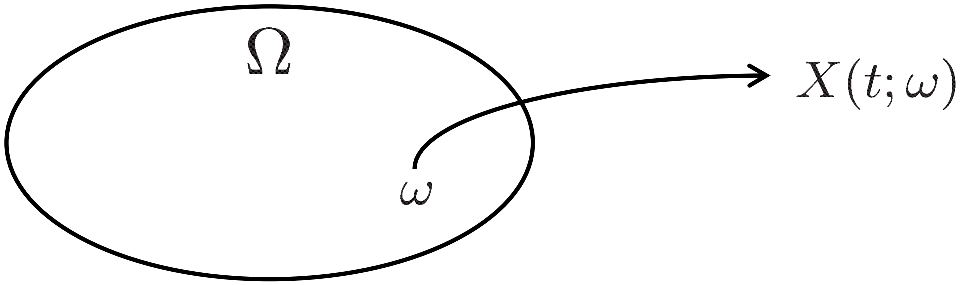
Matrix Inversion Lemma

$$\begin{aligned} (A - BC^{-1}D)^{-1} &= A^{-1} + A^{-1}B(C - DA^{-1}B)^{-1}DA^{-1} \\ (A - BC^{-1}D)^{-1}BC^{-1} &= A^{-1}B(C - DA^{-1}B)^{-1} \end{aligned}$$

$$\begin{aligned} (A + BC^{-1}D)^{-1} &= A^{-1} - A^{-1}B(C + DA^{-1}B)^{-1}DA^{-1} \\ (A + BC^{-1}D)^{-1}BC^{-1} &= A^{-1}B(C + DA^{-1}B)^{-1} \end{aligned}$$

Random Processes

- A **random process** $X(t)$ is a collection of random variables, one for each t, defined on sample space Ω.



Two interpretations:

- For fixed t, $X(t;\omega)$ is a function of ω, i.e., $X(t,\cdot) : \Omega \to \mathbb{R}$. Hence, $X(t,\cdot)$ is a random variable.
- For fixed ω, $X(\cdot;\omega) : \mathbb{R} \to \mathbb{R}$ is a sample path function.

The distribution of a random process is specified by a collection of cumulative distribution functions (CDFs). More precisely, for all $k \in \mathbb{N}$ and for all $t_1, \ldots, t_k$, we need to specify the joint CDF of $X(t_1), \ldots, X(t_k)$.

Gaussian Processes

- **Gaussian process**: A random process $X(t)$ is a **Gaussian process** if for all $k \in \mathbb{N}$ and for all $t_1, \ldots, t_k$, a random vector formed by $X(t_1), \ldots, X(t_k)$ is jointly Gaussian.
- The joint density is completely specified by
  - Mean: $m(t) = \mathbb{E}(X(t))$, where m is known as a mean function.
  - Covariance: $k(t,s) = \mathbf{cov}(X(t), X(s)) = \mathbb{E}\left((X(t) - m(t))(X(s) - m(s))\right)$, where k is known as a covariance function.
- Notation: $X(t) \sim \mathcal{GP}(m(t), k(t,s))$
- Example: $X(t) = tA$, where $A \sim \mathcal{N}(0,1)$ and $t \in \mathbb{R}$.

Summary (up to now)

- **Gaussian random variable**: $X \sim \mathcal{N}(\mu, \sigma^2)$.
- **Central limit theorem**: a sum of a large number of small errors has approximately a normal distribution.
- **Multivariate Gaussian random vector**: $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right) \tag{1}$$

- **Conditional distribution of jointly Gaussian random vectors**: $\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\mathbb{E}(\mathbf{x}|\mathbf{y}), \Sigma_{x|y})$, where

$$\mathbb{E}(\mathbf{x}|\mathbf{y}) = \mathbb{E}(\mathbf{x}) + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{y} - \mathbb{E}(\mathbf{y})) \tag{2}$$

$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}. \tag{3}$$

- **Gaussian process**: $X(t) \sim \mathcal{GP}(\mu(t), k(t,s))$
- **Matrix inversion lemma**