General Multivariate Normal Distribution $N(\mu, \Sigma)$

- Recall standard normal $Z \sim N(0, I_n)$: pdf $\boxed{f_Z(z) = \frac{1}{(2\pi)^{n/2}} e^{-\|z\|_2^2/2}}$, $z \in \mathbb{R}^n$ (*)

Def (General normal)
$X = (X_1, \dots, X_n) \in \mathbb{R}^n$ is a normal r.vector if X is an affine transformation of a standard normal r.vector $Z \sim N(0, I_n)$, i.e.
$$X = AZ + \mu \quad \text{for some invertible } A \in \mathbb{R}^{n \times n},\ \mu \in \mathbb{R}^n.$$

- $\mathbb{E}X = \mu$,
- $\mathrm{Cov}(X) = \mathbb{E}(AZ)(AZ)^T = \mathbb{E}[AZZ^TA^T] = A \cdot \underbrace{\mathbb{E}[ZZ^T]}_{= I_d} \cdot A^T = \boxed{AA^T =: \Sigma}$
- pdf? Assume $\mu = 0$ for simplicity

$$\forall B \in \mathcal{B}(\mathbb{R}^n): \quad P\{\underset{AZ}{X} \in B\} = P\{Z \in A^{-1}(B)\} = \int_{A^{-1}(B)} f_Z(z)\,dz$$

Change of var:
$x = Az$
$dx = |A|\,dz$
determinant

$$= \int_B \underbrace{f_Z(A^{-1}x)\,|A|^{-1}}\,dx$$

$$\Rightarrow f_X(x) = f_Z(A^{-1}x)\,|A|^{-1} \overset{(*)}{=} \frac{1}{(2\pi)^{n/2}|A|} \exp\left(-\|A^{-1}x\|_2^2/2\right)$$

- Let's express this in terms of $\Sigma = AA^T$.

$$|A| = \sqrt{|AA^T|} = \sqrt{|\Sigma|}.$$

det(AB) = det(A) det(B)

$$\|A^{-1}x\|_2^2 = (A^{-1}x)^T(A^{-1}x) = x^T(A^{-1})^TA^{-1}x = x^T(AA^T)^{-1}x = x^T\Sigma^{-1}x$$

$$\Rightarrow f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x^T\right)$$

- General case, for $\mu \neq 0$: apply the shift $\Rightarrow$

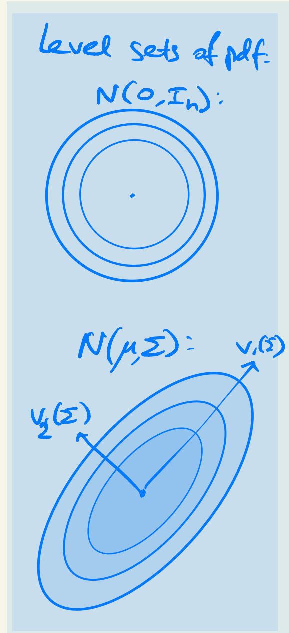
$$\boxed{f_X(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}, \quad x \in \mathbb{R}^n \qquad (**)$$

- Compare to pdf of $X \sim N(\mu, \sigma^2)$ in 1D:
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-\mu)^2/2\sigma^2\right).$$

Vice versa: $\forall$ r.v. X with pdf (**) can be written as $X = AZ + \mu$ where $Z \sim N(0, I_n)$ and A is invertible (take $A = \sqrt{\Sigma}$)

SUMMARY

- A $\underline{\text{general normal r.vector}}$ X is an affine transformation of a standard normal r.v. $Z \sim N(0, I_n)$.
- The pdf of X is (**) p.59
  $\Downarrow$
- The distribution of X is determined by mean $\mu = \mathbb{E}X$ and covariance $\Sigma = \text{Cov}(X)$. It is denoted $N(\mu, \Sigma)$.



Def. • R.variables $X_1, \dots, X_n$ are $\underline{\text{jointly}}$ $\underline{\text{normal}}$ if the random vector $X = (X_1, \dots, X_n)$ has normal distr.

• Their joint distr. is determined by the $\underline{\text{means}}$ and $\underline{\text{covariances}}$

• **Ex** $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$, $\text{Var}(X_1) = 1$, $\text{Var}(X_2) = 4$, $\rho_{X_1, X_2} = 1/2$

Find the pdf of (X_1, X_2).

$$\frac{1}{2} = \rho_{X_1, X_2} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \cdot \text{Var}(X_2)}} = \frac{\text{Cov}(X_1, X_2)}{1 \cdot 2} \Rightarrow \text{Cov}(X_1, X_2) = 1.$$

$$\Rightarrow \Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix} \Rightarrow \Sigma^{-1} = \frac{1}{3} \begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}, \quad |\Sigma| = 3.$$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{3}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) = \boxed{\frac{1}{2\pi\sqrt{3}} \exp\left(-\frac{4x_1^2 + x_2^2 - 2x_1 x_2}{6}\right)}.$$

Prop Jointly normal r.v's are independent $\Leftrightarrow$ they are uncorrelated.

Proof $(\Rightarrow)$ is always true.

$(\Leftarrow)$ Uncorrelated $\Rightarrow \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ where $\sigma_i^2 = \text{Var}(X_i)$

$$\Rightarrow f_X(x) \text{ factors}: \quad f_X(x) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-(x_i - \mu_i)^2 / 2\sigma_i^2\right)}_{= f_{X_i}(x_i)} \Rightarrow \text{indep.}$$

$\square$

Prop $\forall$ r.v's $Y_1, \dots, Y_n$ $\exists$ jointly normal r.v's $X_1, \dots, X_n$ with $\mathbb{E}X_i = \mathbb{E}Y_i$ and $\text{Cov}(X_i, X_j) = \text{Cov}(Y_i, Y_j)$.

"Gaussian model"

• **Remark:** $\exists$ normal r.v's X_1, X_2 (but not jointly normal), that are uncorrelated but not independent.

HW

HW: $Z \in \mathbb{R}^n$ is normal, $B \in \mathbb{R}^{m \times n}$ full rank $\Rightarrow BZ \in \mathbb{R}^m$ is normal.

CONSTRUCTING INDEPENDENT RANDOM VARIABLES

- Why do indep. r.v's $X_1, X_2, \dots$ exist? Finite # first.

Prop (Coupling)
1. $\forall$ prob. measure ν on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ $\exists$ r.v's $X_1, \dots X_n$ with joint distribution ν.

$\Downarrow$

2. $\forall$ prob. measures $\mu_1, \dots, \mu_n$ on $(\mathbb{R}, \mathcal{B})$ $\exists$ indep r.v's $X_1, \dots, X_n$ with marginal distributions $\mu_1, \dots, \mu_n$

Proof 1. Define $X(\omega) := \omega$, $\omega \in \mathbb{R}^n$ $\Rightarrow$ X has distribution ν:

$$\underbrace{\mathbb{P}}_{=\nu}\{\underbrace{X(\omega)}_{=\omega} \in B\} = \nu(B) \qquad \forall\ B \in \mathcal{B}(\mathbb{R}^n).$$

2. Apply part 1 for $\nu := \mu_1 \times \dots \times \mu_n$ $\Rightarrow$ QED by Prop p. 40. $\square$

- How to construct ∞ sequence $X_1, X_2, \dots$ with given joint distributions of $(X_1, \dots, X_n)$ $\forall n$? In particular, independent?

- Extend $\nu_n = \mu_1 \times \dots \times \mu_n$ to $\nu = \mu_1 \times \mu_2 \times \dots$?

Kolmogorov's Extension Thm

- $\boxed{\mathbb{R}^{\mathbb{N}}}$:= sequences of real numbers, i.e. $\omega \in \mathbb{R}^{\mathbb{N}}$ if $\omega = (\omega_1, \omega_2, \dots)$ with $\omega_i \in \mathbb{R}$.

- **Cylinder σ-algebra** $\mathcal{B}(\mathbb{R}^{\mathbb{N}}) := \sigma(\text{cylinders})$

  where a cylinder is a set of the form

$$A = I_1 \times I_2 \times \dots \times I_n \times \mathbb{R} \times \mathbb{R} \times \dots \subset \mathbb{R}^{\mathbb{N}}$$
$$= \{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_1 \in I_1, \dots, \omega_n \in I_n\} \qquad \text{for some } n \in \mathbb{N}$$

  where $I_i = (a_i, b_i)$ are intervals (endpoints could be $= \pm\infty$)

- A sequence of measures $\nu_1, \nu_2, \dots$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is **consistent** if
$$\nu_n(I_1 \times \dots \times I_n) = \nu_{n+1}(I_1 \times \dots \times I_n \times \mathbb{R})$$
  $\forall n \in \mathbb{N}$ and $\forall$ intervals I_i.

THM (Kolmogorov's Extension Thm)
$\forall$ consistent sequence of probability measures $\nu_1, \nu_2, \ldots$
can be uniquely extended to a probability measure ν on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$.

Proof is based on

nonnegative & countably additive on $\mathcal{A}$

a family that is closed under complement & finite unions.

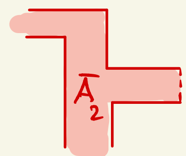
THM (Caratheodory Extension Thm)
$\forall$ premeasure on an algebra $\mathcal{A}$
can be extended to a measure on $\sigma(\mathcal{A})$

If the premeasure is σ-finite, the extension is unique (π-λ thm)

Proof of K.E.T. We will apply C.E.T for the algebra $\mathcal{A}$ generated by the cylinders, i.e.

$$A \in \mathcal{A} \iff A = \bar A_n \times \mathbb{R} \times \mathbb{R} \times \ldots$$

for some $n \in \mathbb{N}$, where $\bar A_n =$ finite union of $\overset{\text{products of intervals, e.g.}}{\overbrace{\text{boxes}}}$



- Define ν on $\mathcal{A}$ by $\nu(\bar A_n \times \mathbb{R} \times \mathbb{R} \times \cdots) := \nu_n(\bar A_n)$.

  Well defined by the consistency assumption (check!)

- Obviously $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ generated by cylinders.
- Hence, by C.E.T. it is enough to check that ν is a premeasure on $\mathcal{A}$, i.e. is countably additive on $\mathcal{A}$. So, enough to show:

$$\nu\Big(\bigsqcup_{i=1}^{\infty} B_i\Big) \overset{?}{=} \sum_{i=1}^{\infty} \nu(B_i) \quad \text{whenever } B_i \in \mathcal{A} \text{ and } \bigsqcup_{i=1}^{\infty} B_i \in \mathcal{A}. \qquad (1)$$

- We already know that ν is finitely additive on $\mathcal{A}$:

$$\nu\Big(\bigsqcup_{i=1}^{n} B_i\Big) = \sum_{i=1}^{n} \nu(B_i) \quad \forall n \in \mathbb{N} \qquad (2)$$

because on these sets ν coinsides with $\nu_{N(n)}$ by consistency assumption.

- LHS(2) $= \nu_{N(n)}(\text{" — "}) \le 1 \Rightarrow$ RHS(2) converges to $\sum_1^{\infty} \nu(B_i)$.

  Subtract (2) from (1) $\Rightarrow$ enough to show that

$$\underbrace{\nu\Big(\bigsqcup_{i=n+1}^{\infty} B_i\Big)}_{= A_n} \xrightarrow{?} 0$$

Note: $A_n \in \mathcal{A}$ by (1), and $A_n \downarrow \emptyset$ i.e. $\bigcap_{n=1}^{\infty} A_n = \emptyset$

$\Downarrow$

- It is enough to show: $A_n \in \mathcal{A},\ A_n \downarrow \emptyset \overset{?}{\Rightarrow} \nu(A_n) \to 0$

↯ $\exists\, A_n \in \mathcal{A}:\ A_n \downarrow \emptyset$ and $\nu(A_n) > \varepsilon > 0 \quad \forall n$

- WLOG A_n are of the following form (otherwise repeat A_i's as needed)

$$
\begin{array}{lll}
a_1 \in A_1 = \underset{a_1(1)}{\boxed{\bar{A}_1}} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \cdots \\
\quad\ \ \cup \\
a_2 \in A_2 = \underset{a_2(1)\qquad a_2(2)}{\boxed{\qquad \bar{A}_2 \qquad}} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \cdots \\
\quad\ \ \cup \\
a_3 \in A_3 = \underset{a_3(1)\qquad a_3(2)\qquad a_3(3)}{\boxed{\qquad\qquad \bar{A}_3 \qquad\qquad}} \times \mathbb{R} \times \mathbb{R} \times \cdots \\
\quad\ \ \cup \\
\quad\ \ \cdots
\end{array}
$$

where all $\bar{A}_n \subset \mathbb{R}^n$ are finite unions of rectangles.

- First assume that all $\bar{A}_n$ are compact

- Diagonal argument: choose $a_n \in A_n \quad \forall n$.

  – First coordinate: $a_n \in A_n \subset A_1 \quad \forall n \geq 1 \Rightarrow a_n(1) \in \bar{A}_1$

    Compactness $\Rightarrow \exists$ convergent subsequence $a_{n_k}(1) \to x(1) \in \bar{A}_1$

  – Second coordinate: $a_{n_k} \in A_{n_k} \subset A_2 \ \forall n_k \geq 2 \Rightarrow \left(a_{n_k}(1), a_{n_k}(2)\right) \in \bar{A}_2$

    Compactness $\Rightarrow \exists$ convergent subsequence $\left(a_{n_{k_j}}(1), a_{n_{k_j}}(2)\right) \to (x(1), x(2)) \in \bar{A}_2$

  ...

  $\Rightarrow \exists\ x = (x(1), x(2), \dots) \in \mathbb{R}^{\mathbb{N}}$ s.t. $(x(1), \dots, x(n)) \in \bar{A}_n \quad \forall n$

  $\Rightarrow x \in A_n \ \ \forall n \Rightarrow x \in \bigcap_{n=1}^{\infty} A_n \Rightarrow A_n \not\downarrow \emptyset$ ↯.

- For general $\bar{A}_n$, use regularity to approximate A_n by compact sets:

  – Borel measures ν_n are inner-regular, i.e. $\exists$ compact $\bar{K}_n \subset \bar{A}_n$ s.t.

$$\nu_n(\bar{A}_n \setminus \bar{K}_n) \leq \varepsilon / 2^{n+1} \qquad \forall n.$$

  $K_n := \bar{K}_1 \times \mathbb{R} \times \mathbb{R} \times \dots$ like A_n's. Then

  – To make these sets decrease like A_n's, let

$$B_n := K_1 \cap \dots \cap K_n$$

  Then B_n decrease, and $\nu(B_n) \geq \overbrace{\nu(A_n)}^{\geq\, \varepsilon} - \overbrace{\nu(A_n \setminus B_n)}^{\leq\, \sum_{n=1}^{\infty} \varepsilon/2^{n+1} \text{ (check!)}} \geq \varepsilon/2$.

  – Run the diagonal argument for $B_n = \underset{\text{compact}}{\boxed{\bar{B}_n}} \times \mathbb{R} \times \mathbb{R} \dots \Rightarrow$ QED.