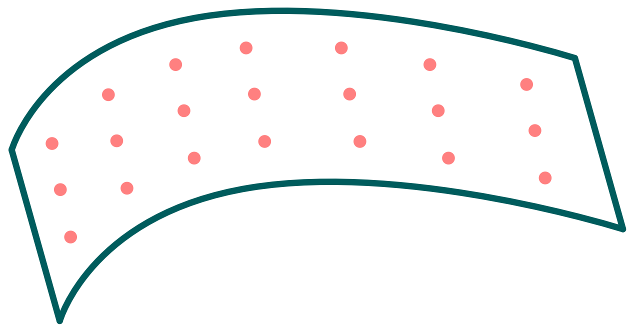
Multivariate Gaussians and Gaussian processes

A: The multivariate Gaussian

Example: Modeling a network of sensors

Increasingly, large structures (bridges, skyscrapers) contain arrays of embedded sensors to warn of excessive strain or small cracks.

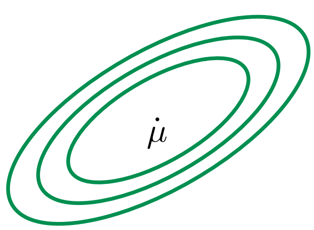


Say there are d sensors, each with a real-valued reading.

1. Model distribution of sensor readings under normal conditions.
2. Accommodate a moderate number of sensor failures.

How to exploit correlations between sensors?

The multivariate Gaussian



$N(\mu, \Sigma)$: Gaussian in $\mathbb{R}^d$

- mean: $\mu \in \mathbb{R}^d$
- covariance: $d \times d$ matrix Σ

Generates points $X = (X_1, X_2, \ldots, X_d)$.

- μ is the vector of coordinate-wise means:

$$\mu_1 = \mathbb{E}X_1,\ \mu_2 = \mathbb{E}X_2, \ldots,\ \mu_d = \mathbb{E}X_d.$$

- Σ is a matrix containing all pairwise covariances:

$$\Sigma_{ij} = \Sigma_{ji} = \text{cov}(X_i, X_j) \quad \text{if } i \neq j$$
$$\Sigma_{ii} = \text{var}(X_i)$$

Density $\quad p(x) = \dfrac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\dfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

Special case: spherical Gaussian

The X_i are independent and all have the same variance σ^2.

1. What is the covariance matrix Σ, and what is its inverse Σ^{-1}?

2. Simplify the density

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Spherical Gaussian: summary

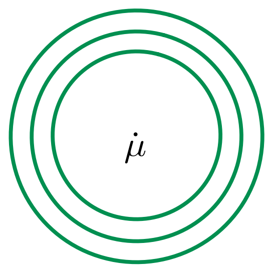
The X_i are independent and all have the same variance σ^2.

$$\Sigma = \sigma^2 I_d = \text{diag}(\sigma^2, \sigma^2, \ldots, \sigma^2) \quad \text{(diagonal elements } \sigma^2\text{, rest zero)}$$

Each X_i is an independent univariate Gaussian $N(\mu_i, \sigma^2)$:

$$\Pr(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{\|x-\mu\|^2}{2\sigma^2}\right) = \prod_{i=1}^{d} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu_i)^2/2\sigma^2}\right)$$

Density at a point depends only on its distance from μ:



Special case: diagonal Gaussian

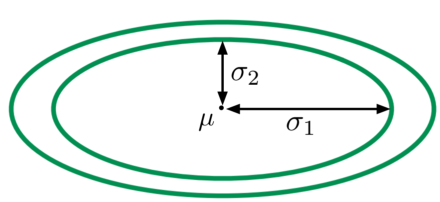
The X_i are independent, with variances σ_i^2. Thus

$$\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_d^2) \quad \text{(off-diagonal elements zero)}$$

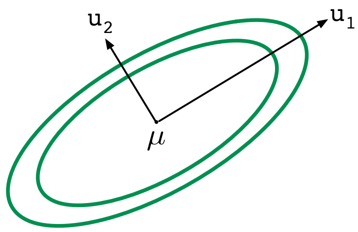
Each X_i is an independent univariate Gaussian $N(\mu_i, \sigma_i^2)$:

$$p(x) = \frac{1}{(2\pi)^{d/2}\sigma_1 \cdots \sigma_d} \exp\left(-\sum_{i=1}^{d} \frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)$$

Contours of equal density are **axis-aligned ellipsoids** centered at μ:



The general Gaussian $N(\mu, \Sigma)$ in $\mathbb{R}^d$



$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

- **Eigenvectors** of Σ are $u_1, \ldots, u_d$
- Corresponding **eigenvalues** $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$

Rotated version of a diagonal Gaussian $N(\mu, \Lambda)$:

B: Useful properties of the Gaussian

Closure under linear transformation

Lemma. If $X \sim N(\mu, \Sigma)$ in $\mathbb{R}^d$ and A is a $k \times d$ matrix, then AX is a Gaussian in $\mathbb{R}^k$.

What are the mean and covariance of AX?

Characteristic function

The **characteristic function** of random vector $Z \in \mathbb{R}^d$ is the function $\phi : \mathbb{R}^d \to \mathbb{C}$,

$$\phi(t) = \mathbb{E}e^{it \cdot Z} = \mathbb{E}\cos(t \cdot Z) + i\mathbb{E}\sin(t \cdot Z).$$

This always exists.

- **Fourier inversion**: two random vectors have the same distribution iff they have the same characteristic function.
- Thus two distributions on $\mathbb{R}^d$ are equal iff all their one-dimensional projections are equal.

Closure under linear transformation (cont'd)

A quick calculation shows that $N(\mu, \Sigma)$ has characteristic function

$$\phi(t) = \exp\left(it \cdot \mu - \frac{1}{2} t^T \Sigma t\right).$$

Prove: any linear transformation of a Gaussian is Gaussian.

Back to sensor example

Sensor readings $X_1, \ldots, X_d$: model by a Gaussian $N(\mu, \Sigma)$.

(1) How to set μ and Σ?

(2) What is the distribution of sensor X_i?

(3) What is the joint distribution of a set of sensors $S \subset [d]$?

Distribution under conditioning

(4) Suppose sensors S fail. What is the conditional distribution of X_S given the readings x_T of the remaining sensors $T = [d] \setminus S$?

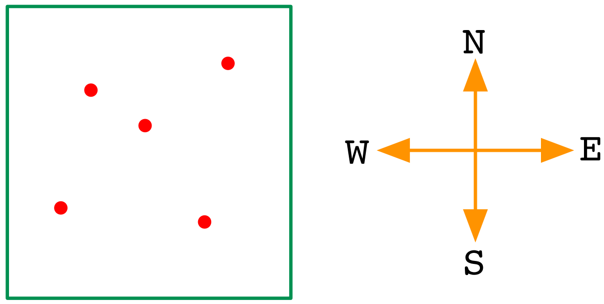
Lemma. The conditional distribution $X_S | X_T = x_T$ is Gaussian, with

$$\text{mean } \mu_S + \Sigma_{ST}\Sigma_{TT}^{-1}(x_T - \mu_T) \text{ and covariance } \Sigma_{SS} - \Sigma_{ST}\Sigma_{TT}^{-1}\Sigma_{TS}.$$

C: Gaussian processes

Example: County pollution levels

Want a pollution map for San Diego county:



At any given time, have readings $y \in \mathbb{R}$ at a few locations $x \in [0,1]^2$:

- Fixed county facilities
- Individual citizen reports

Use these to create a full pollution map for the county.

Given readings $(x_1, y_1), \ldots, (x_n, y_n)$, infer $f : [0,1]^2 \to \mathbb{R}$.

A prior on smooth functions

Let $\mathcal{X}$ be an input space, e.g. $[0,1]^2$.

A Gaussian in infinite dimension: a distribution over all functions $f : \mathcal{X} \to \mathbb{R}$.

- A **Gaussian process** (GP) on $\mathcal{X}$ is a collection of random variables indexed by $\mathcal{X}$ such that any finite subset of them has a Gaussian distribution.
  Pick any $x_1, \ldots, x_s \in \mathcal{X}$.
  If f is sampled from a GP then $(f(x_1), \ldots, f(x_s))$ has a Gaussian distribution.
- A Gaussian process is specified by
  - Mean function $m : \mathcal{X} \to \mathbb{R}$
  - Covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- If $f \sim GP(m, k)$ then for any finite subset $S \subset \mathcal{X}$, we have $f_S \sim N(m_S, k_{SS})$.
  Here f_S is a shorthand for $(f(x) : x \in S)$.

Example

A distribution over functions on the line:

- $\mathcal{X} = \mathbb{R}$
- $m(x) \equiv 0$
- $k(x, x') = \exp(-(x - x')^2)$

Sampling a function from a GP

For $t = 1, 2, 3, \ldots$:

- Pick some $x_t \in \mathcal{X}$
- The joint distribution $(f(x_1), \ldots, f(x_t))$ is Gaussian
  Thus so is the conditional distribution $f(x_t)|f(x_1), \ldots, f(x_{t-1})$
- Sample $f(x_t)$ from this conditional distribution

Conditioning formula: for $S \subset \mathcal{X}$,

$$f(x)|f_S \;\sim\; N\left(m(x) + k(x,S)k(S,S)^{-1}f_S, k(x,x) - k(x,S)k(S,S)^{-1}k(S,x)\right).$$

Example (cont'd)

$$\mathcal{X} = \mathbb{R},\ m(x) \equiv 0,\ k(x, x') = \exp(-\|x - x'\|^2)$$

Several draws