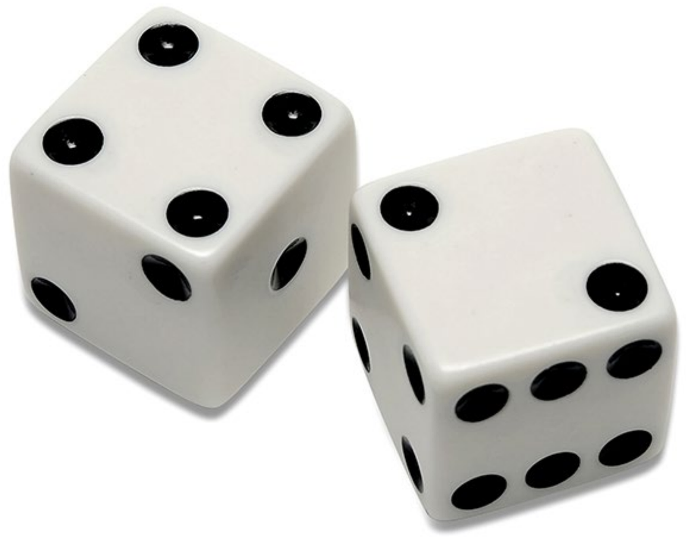
Mathematics for Robotics (ROB-GY 6013 Section A)

- **Week 11:**
  - Probability and Estimation

Rolling a Die (Discrete)



- **Sample space** Ω :
  - {"rolling a 1", "rolling a 2", "rolling a 3", "rolling a 4", "rolling a 5", "rolling a 6"}
- **Probability** of an **event**
  - P("rolling a 1") $= 1/6$
  - P("rolling an even number") $= 3/6$
- **Expected value**: e.g., average points rolled
  - Assign point value to each event
  - Sum of point value × event probability

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6}$$

Probability Distribution (Continuous)

- **Sample space** Ω :
  - Real numbers $x \in [-\infty, \infty]$
- **Probability** of an **event**
  - Integral of a **probability density function** $f(x)$

$$P(x \in [c,d]) = \int_c^d f(x)dx$$

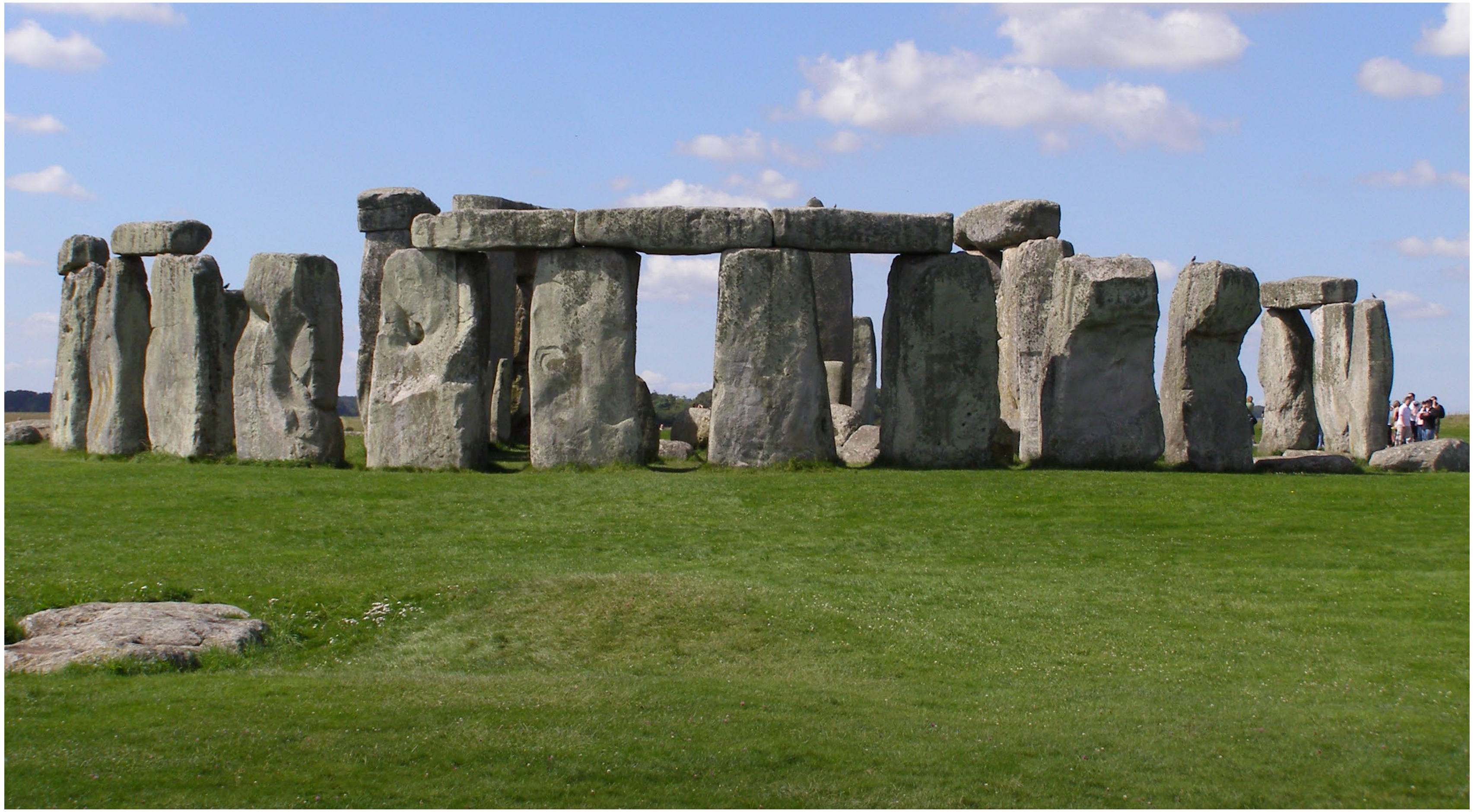
$$P(x \in [-\infty,\infty]) = 1 = \int_{-\infty}^{\infty} f(x)dx$$

- **Expectation operator**: e.g., average function value $g(x)$
  - Assign $g(x)$ to each $x \in [-\infty, \infty]$
  - Integral of $g(x) \times f(x)$

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Probability Space: The *RIGHT* way to begin

- Disclaimer: This part is just for fun.



Definition: Probability Space

- $(\Omega, \mathcal{F}, P)$ is called a **probability space**.
  - Ω is the sample space. Think of it as the set of all possible outcomes of an experiment.
  - $E \subset \Omega$ is an event
  - $\mathcal{F}$ is the collection of allowed events. It must at least contain $\emptyset$ and Ω. It is closed with respect to set complement, countable unions, and countable intersections. Such sets are called sigma algebras.
  - $P: \mathcal{F} \to [0, 1]$ is a probability measure. It has to satisfy a few basic operations:

1. $P(\emptyset) = 0$ *and* $P(\Omega) = 1$.
2. *For each* $E \in \mathscr{F}$, $0 \leq P(E) \leq 1$
3. *If the sets* $E_1, E_2, \ldots$ *are disjoint (i.e.,* $E_i \cap E_j = \emptyset$ *for* $i \neq j$*), then* $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Probability Re-Review

- **Continuous** (real-valued) **random variable** X

- **Probability density function**

Probability Re-Review

- **Continuous** (real-valued) **random variable** X
  - $X : \Omega \rightarrow \mathbb{R}$
- **Probability density function**

Probability Re-Review

- **Continuous** (real-valued) **random variable** X
  - $X : \Omega \to \mathbb{R}$
- **Probability density function**
  - $f\!: \mathbb{R} \to [0,\infty)$

Probability Re-Review

- **Continuous** (real-valued) **random variable** X
  - $X : \Omega \rightarrow \mathbb{R}$
- **Probability density function**
  - $f: \mathbb{R} \rightarrow [0,\infty)$
  - The value of f at a given point is not the probability
  - You have to **integrate** f to get the probability

Definition

- A function $X : \Omega \to \mathbb{R}$ is a **continuous random variable** with **density** $f : \mathbb{R} \to [0,\infty)$ if:

  a) it is a **random variable**, and

  b) $\forall\, x \in \mathbb{R},\ P(\{\omega \in \Omega \,|\, X(\omega) \leq x\}) = \int_{-\infty}^{x} f(\bar{x}) d\bar{x}.$

Definition

- A function $X : \Omega \rightarrow \mathbb{R}$ is a **continuous random variable** with **density** $f : \mathbb{R} \rightarrow [0,\infty)$ if:
    a) it is a **random variable**, and
    b) $\forall\, x \in \mathbb{R},\ P(\{\omega \in \Omega \,|\, X(\omega) \leq x\}) = \int_{-\infty}^{x} f(\bar{x}) d\bar{x}.$

The lower bound of $-\infty$ is for convenience. It can be replaced with $\inf\{X(\Omega) \subset R\}$.

Definition

- A function $X : \Omega \rightarrow \mathbb{R}$ is a **continuous random variable** with **density** $f : \mathbb{R} \rightarrow [0,\infty)$ if:
  a) it is a **random variable**, and
  b) $\forall\, x \in \mathbb{R},\ P(\{\omega \in \Omega \,|\, X(\omega) \leq x\}) = \int_{-\infty}^{x} f(\bar{x}) d\bar{x}.$

The lower bound of $-\infty$ is for convenience. It can be replaced with $\inf\{X(\Omega) \subset R\}$.

We can also use notation $X \sim f$ to denote the random variable and its density.
"X is distributed with density f"

Definition

- A function $X : \Omega \to \mathbb{R}$ is a **continuous random variable** with **density** $f : \mathbb{R} \to [0,\infty)$ if:

  a) it is a **random variable**, and

  b) $\forall\, x \in \mathbb{R},\ P(\{\underbrace{X \leq x}_{\text{convenient shorthand}}\}) = \int_{-\infty}^{x} f(\bar{x})\, d\bar{x}.$

The lower bound of $-\infty$ is for convenience. It can be replaced with $\inf\{X(\Omega) \subset R\}$.

We can also use notation $X \sim f$ to denote the random variable and its density.
"X is distributed with density f"

Definition

- A function $X:\Omega\rightarrow\mathbb{R}$ is a **continuous random variable** with **density** $f:\mathbb{R}\rightarrow[0,\infty)$ if:
  a) it is a **random variable**, and
  b) $\forall\ x\in\mathbb{R},\ P(\{\underbrace{X\leq x}_{\textbf{convenient shorthand}}\})=\int_{-\infty}^{x}f(\bar{x})d\bar{x}.$

The lower bound of $-\infty$ is for convenience. It can be replaced with $\inf\{X(\Omega)\subset R\}$.

We can also use notation $X\sim f$ to denote the random variable and its density.
"X is distributed with density f"

$$P(x_1\leq X\leq x_2)=\int_{x_1}^{x_2}f(x)dx$$

Examples: Uniform Random Variable

- Parameters are a, b

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & otherwise \end{cases}$$

Examples:

- Parameters are $\sigma > 0, \mu \in \mathbb{R}$

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

Examples: Gaussian or Normal Random Variable

- Parameters are $\sigma > 0$, $\mu \in \mathbb{R}$
- $N(\mu, \sigma^2)$

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

Expected Value/Expectation Operator

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f(x) dx$$

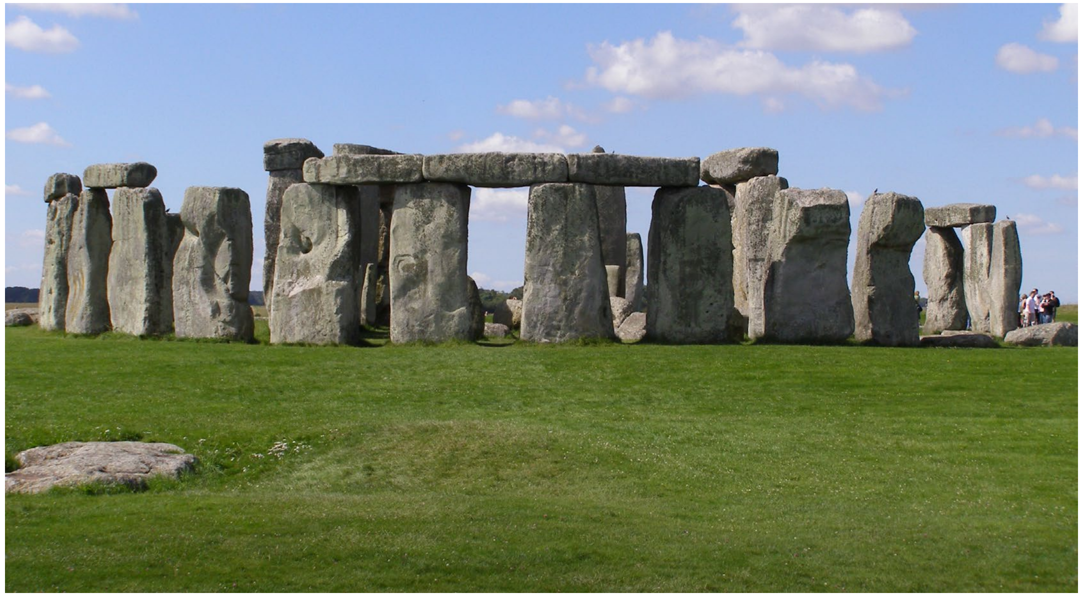
Expected Value/Expectation Operator

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- *Mean:* $\mu := \mathcal{E}\{X\} := \int_{-\infty}^{\infty} x f(x) dx$
- *Variance:* $\sigma^2 := \mathcal{E}\{(X-\mu)^2)\} := \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$ *(Var. for short)*
- *Standard Deviation:* $\sigma := \sqrt{\sigma^2}$ *(Std. Dev. for short)*

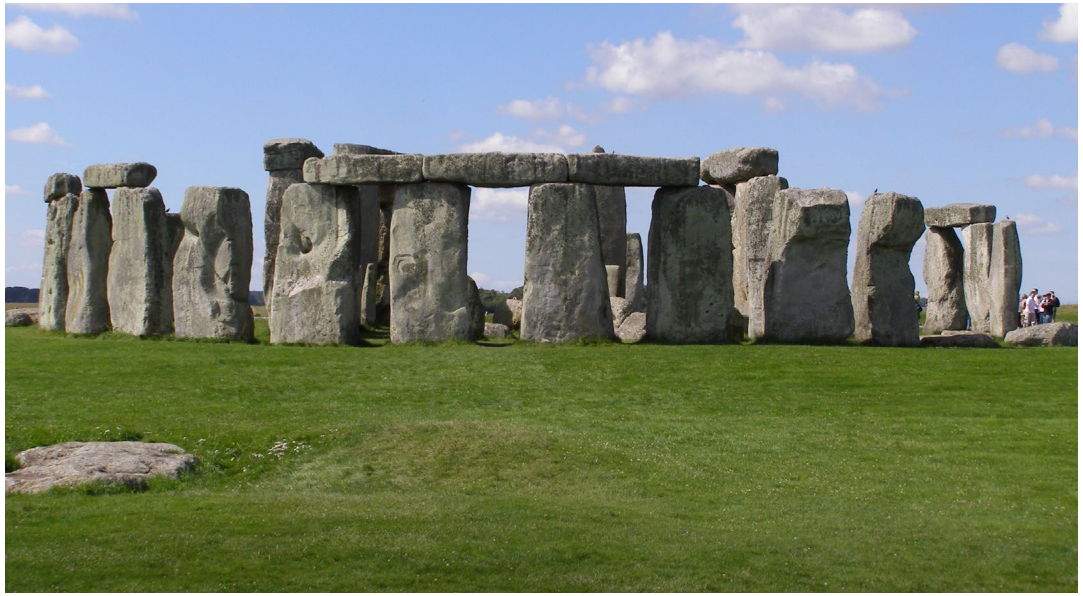
Statistical Moments

- Mean $\mu := E\{X\}$
- Variance $\sigma^2 := E\{(X-\mu)^2\}$
- Skewness $\gamma_1 := \tilde{\mu}_3 = \dfrac{E\{(X-\mu)^3\}}{\sigma^3}$
- Kurtosis $\mathrm{Kurt}[X] := \dfrac{E\{(X-\mu)^4\}}{\sigma^4}$



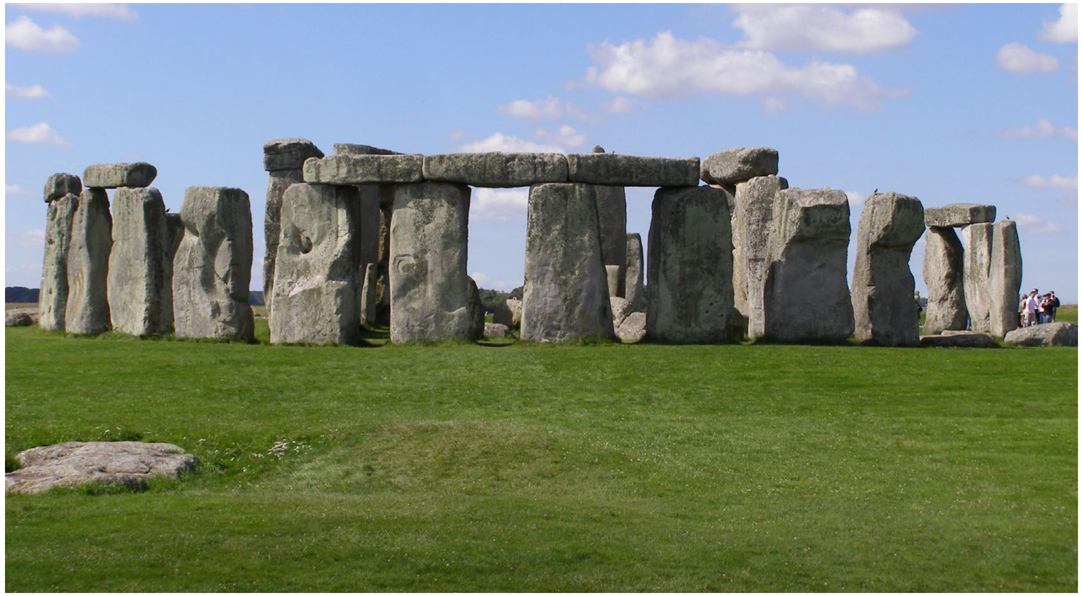
Statistical Moments

- Mean $$\mu := E\{X\}$$
- **Variance** $$\sigma^2 := E\{(X-\mu)^2\}$$
- Skewness $$\gamma_1 := \tilde{\mu}_3 = \frac{E\{(X-\mu)^3\}}{\sigma^3}$$
- Kurtosis $$\mathrm{Kurt}[X] := \frac{E\{(X-\mu)^4\}}{\sigma^4}$$



Statistical Moments

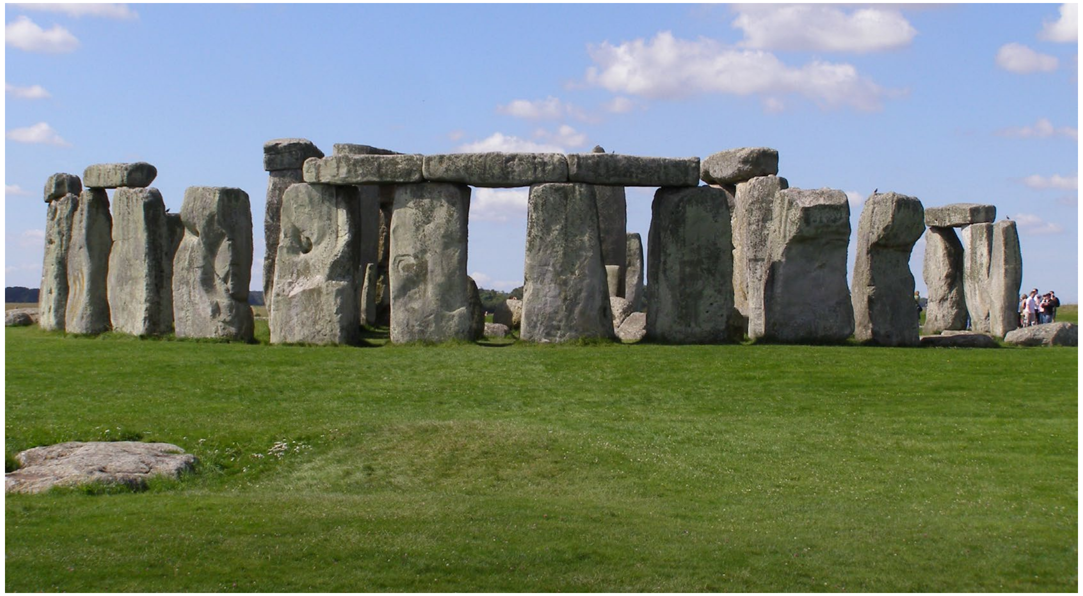
- Mean $\mu := E\{X\}$
- **Variance** $\sigma^2 := E\{(X-\mu)^2\}$
- Skewness $\gamma_1 := \tilde{\mu}_3 = \dfrac{E\{(X-\mu)^3\}}{\sigma^3}$
- Kurtosis $\mathrm{Kurt}[X] := \dfrac{E\{(X-\mu)^4\}}{\sigma^4}$



$$E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (X-\mu)^2 f(x)dx$$

Statistical Moments

- Mean $\mu := E\{X\}$
- **Variance** $\sigma^2 := E\{(X-\mu)^2\}$
- Skewness $\gamma_1 := \tilde{\mu}_3 = \dfrac{E\{(X-\mu)^3\}}{\sigma^3}$
- Kurtosis $\operatorname{Kurt}[X] := \dfrac{E\{(X-\mu)^4\}}{\sigma^4}$



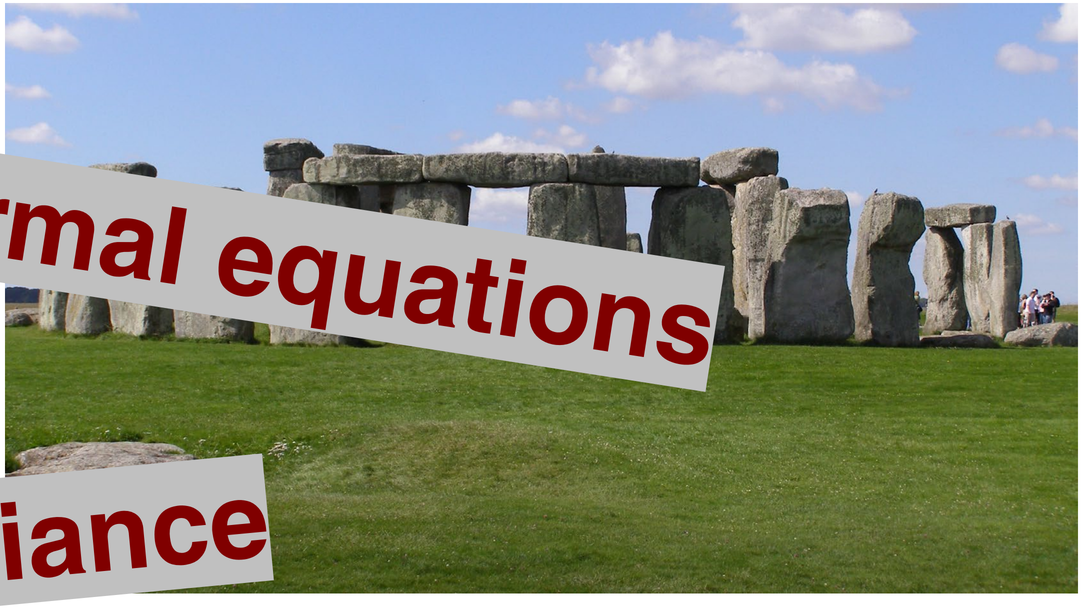
$$E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (X-\mu)^2 f(x)\,dx$$

2-norm squared for random variables

Statistical Moments

- Mean $\{X\}$
- **Variance** $:= E\{(X-\mu)^2\}$
- Skewnes $\gamma_1 := \tilde{\mu}_3 = \frac{E\{(X-\mu)^3\}}{\sigma^3}$
- Kurt $\mathrm{Kurt}[X] := \frac{E\{(X-\mu)^4\}}{}$

Minimize uncertainty



Normal equations

Minimum variance

$$E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (X$$

Weighted Least squares

2-norm squared for random variables

Random Vectors

- For example, the joint angles of a 6-DOF robot

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \qquad \mu = E\{X\} = \begin{bmatrix} E\{X_1\} \\ E\{X_2\} \\ \vdots \\ E\{X_p\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

Random Vectors are painful

Definition 5.15 *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space. A function* $X : \Omega \to \mathbb{R}^p$ *is called a* ***random vector*** *if each component of* $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ *is a random variable, that is,* $\forall\, 1 \leq i \leq p$*,* $X_i : \Omega \to \mathbb{R}$ *is a random variable.*

Consequently, $\forall x \in \mathbb{R}^p$, the set $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathscr{F}$ (i.e., it is an allowed event), where the inequality is understood **pointwise**, that is,

$$\{\omega \in \Omega \mid X(\omega) \leq x\} := \left\{ \omega \in \Omega \mid \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{bmatrix} \leq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} \right\} := \left\{ \omega \in \Omega \mid \begin{bmatrix} X_1(\omega) \leq x_1 \\ X_2(\omega) \leq x_2 \\ \vdots \\ X_p(\omega) \leq x_p \end{bmatrix} \right\} = \bigcap_{i=1}^{p} \{\omega \in \Omega \mid X_i(\omega) \leq x_i\}.$$

Random Vectors are painful

Definition 5.16 *$X : \Omega \to \mathbb{R}^p$ is a* ***continuous random vector*** *if there exists a* ***density*** *$f_X : \mathbb{R}^p \to [0, \infty)$ such that,*

$$\forall\, x \in \mathbb{R}^P, \;\; P(\{X \leq x\}) = \int_{-\infty}^{x_p} ... \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(\bar{x}_1, \bar{x}_2 ... \bar{x}_p) d\bar{x}_1 d\bar{x}_2 ... d\bar{x}_p.$$

More generally, for all $A \subset \mathbb{R}^p$ such that the indicator function I_A has bounded variation,

$$P(\{X \in A\}) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_A(\bar{x}_1, \bar{x}_2 ... \bar{x}_p) f_X(\bar{x}_1, \bar{x}_2 ... \bar{x}_p) d\bar{x}_1 d\bar{x}_2 ... d\bar{x}_p.$$

Notation 5.17 *The notation $X \sim f$ is read as X is distributed with density f or that X is a random vector with density f.*

Definition 5.18 *(Moments) Suppose $g : \mathbb{R}^p \to \mathbb{R}^k$*

$$\mathcal{E}\{g(X)\} := \int_{\mathbb{R}^p} g(x) f_X(x) dx := \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} g(x_1, ..., x_p) f_X(x_1, ..., x_p) dx_1 ... dx_p$$

Covariance matrix and Variance

- Covariance matrix is a matrix
- Variance remains a scalar

$$\Sigma := \operatorname{cov}(X) = \operatorname{cov}(X, X) = E\{(X-\mu)(X-\mu)^T\}$$

$$\operatorname{Var}(X) := \operatorname{trace}(\Sigma) = \sum_{i=1}^{p} \Sigma_{ii}\, \operatorname{cov}(X, X) = E\{(X-\mu)^T (X-\mu)\}$$

Covariance matrix and Variance

- Covariance matrix is a matrix $\Sigma_{ij} = E\{(X_i - \mu_i)(X_j - \mu_j)\}$
- Variance remains a scalar

$$\Sigma := \operatorname{cov}(X) = \operatorname{cov}(X, X) = E\{(X - \mu)(X - \mu)^T\}$$

$$\operatorname{Var}(X) := \operatorname{trace}(\Sigma) = \sum_{i=1}^{p} \Sigma_{ii} \operatorname{cov}(X, X) = E\{(X - \mu)^T (X - \mu)\}$$

Covariance matrix and Variance

- Covariance matrix is a matrix $\Sigma_{ij} = E\{(X_i - \mu_i)(X_j - \mu_j)\}$
- Variance remains a scalar

$$\Sigma := \mathrm{cov}(X) = \mathrm{cov}(X, X) = E\{(X - \mu)(X - \mu)^T\}$$

$$\mathrm{Var}(X) := \mathrm{trace}(\Sigma) = \sum_{i=1}^{p} \Sigma_{ii}\, \mathrm{cov}(X, X) = E\{(X - \mu)^T (X - \mu)\}$$

- Note is μ a vector
- Our trick is to get all relevant information from mean and covariance matrix so we can avoid working with the density, which requires a lot of integration

Covariance Matrix is positive semi-definite

Covariance Matrix is positive semi-definite

Exercise 5.19 $\Sigma := \text{cov}(X) = \text{cov}(X, X)$ *is a positive semi-definite matrix.*

Solution: For $v \in \mathbb{R}^p$, we need to show that $v^\top \Sigma v \geq 0$, where $\Sigma := \mathcal{E}\{(X - \mu) \cdot (X - \mu)^\top\}$.

$$\begin{aligned}
v^\top \Sigma v :=& v^\top \mathcal{E}\{(X - \mu) \cdot (X - \mu)^\top\} v \\
=& \mathcal{E}\{v^\top (X - \mu) \cdot (X - \mu)^\top v\} \\
=& \mathcal{E}\{((X - \mu)^\top v)^\top \cdot ((X - \mu)^\top v)\} \\
=& \mathcal{E}\{||(X - \mu)^\top v||^2\} \\
=& \int_{\mathbb{R}^p} ||(X - \mu)^\top v||^2 f_X(x) dx \\
\geq& \, 0
\end{aligned}$$

because the integral of a non-negative function over $\mathbb{R}^p$ is non-negative.

Covariance Matrix is positive semi-definite

- If the covariance matrix is **positive definite**, its inverse is the **information matrix**. The interpretation is that “high variance” means “low information” and vice versa.

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
    - *What do the eigenvalues mean?*

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
    - *What do the eigenvalues mean?*
    - *What does the corresponding density look like?*

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
  - *What do the eigenvalues mean?*
  - *What does the corresponding density look like?*
- What does diagonalizing a covariance matrix (with an orthogonal matrix) mean?

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
  - *What do the eigenvalues mean?*
  - *What does the corresponding density look like?*
- What does diagonalizing a covariance matrix (with an orthogonal matrix) mean?
  - *How does the corresponding density change in shape?*

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
  - *What do the eigenvalues mean?*
  - *What does the corresponding density look like?*
- What does diagonalizing a covariance matrix (with an orthogonal matrix) mean?
  - *How does the corresponding density change in shape?*
  - *What do the eigenvectors (that form the orthogonal matrix) mean?*

Eigenvalues and Shape of Covariance Matrix

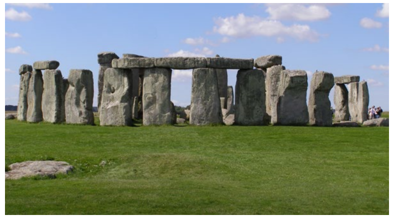
- What does a diagonal covariance matrix mean?
  - *What do the eigenvalues mean?*
  - *What does the corresponding density look like?*
- What does diagonalizing a covariance matrix (with an orthogonal matrix) mean?
  - *How does the corresponding density change in shape?*
  - *What do the eigenvectors (that form the orthogonal matrix) mean?*

- Recall: you can always diagonalize a symmetric matrix with an orthogonal matrix

Eigenvalues and Shape of Covariance Matrix

- What does a diagonal covariance matrix mean?
  - *What do the eigenvalues mean?*
  - *What does the corresponding density look like?*
- What does diagonalizing a covariance matrix (with an orthogonal matrix) mean?
  - *How does the corresponding density change in shape?*
  - *What do the eigenvectors (that form the orthogonal matrix) mean?*

- Recall: you can always diagonalize a symmetric matrix with an orthogonal matrix
- Subtle: Independence for random vectors → diagonal covariance matrix. Converse not true. Covariance matrix does not capture all information about the density.



Idea

- Least squares as a minimum distance problem

Idea

- Least squares as a minimum distance problem
- Estimation as a weighted least squares problem

Overdetermined equations

- Too many equations (find best approximation)

$$\widehat{\alpha} = \underset{\alpha \in \mathbb{R}^m}{\arg\min} \|A\alpha - b\|^2 \iff (A^\top S A)\widehat{\alpha} = A^\top S b \iff \widehat{\alpha} = (A^\top S A)^{-1} A^\top S b$$

Underdetermined equations

- Too many solutions (too few equations)
  - Find “smallest” solution

$$\widehat{x} := \arg\min_{Ax=b} ||x|| = \arg\min_{Ax=b} ||x||^2$$

Underdetermined equations

- Too many solutions (too few equations)
  - Find "smallest" solution

$$\widehat{x} := \underset{Ax=b}{\arg\min} \, ||x|| = \underset{Ax=b}{\arg\min} \, ||x||^2$$

$$\widehat{x} = S^{-1}A^\top \beta, \; AS^{-1}A^\top \beta = b \textit{ or, equivalently, } \widehat{x} = S^{-1}A^\top \left(AS^{-1}A^\top\right)^{-1} b$$

Best Linear Unbiased Estimator

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0, \operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

Best *Linear* Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0,\ \mathrm{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky$$

Best Linear ***Unbiased*** Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0, \operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0$$

holds for all $x \in \mathbb{R}^n$

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0,\ \mathrm{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky$$

$$E\{\hat{x} - x\} = 0$$

holds for all $x \in \mathbb{R}^n$

$$\mathrm{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T (\hat{x} - x)\}$$

Minimizes variance

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0, \operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \operatorname{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T (\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

$$\boxed{\hat{K} = (C^T Q^{-1} C)^{-1} C^T Q^{-1}}$$

Best Linear Unbiased Estimator (BLUE)

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon$, (C is linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *unknown, deterministic*
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0$, $\operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \operatorname{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T(\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

$$\boxed{\hat{K} = (C^T Q^{-1} C)^{-1} C^T Q^{-1} \qquad \operatorname{cov}(\hat{x} - x) = (C^T Q^{-1} C)^{-1}}$$

Minimum Variance Estimator

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0$, $\operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \operatorname{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T(\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

Minimum Variance Estimator

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C does not have to be linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0, \operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \operatorname{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T(\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

Minimum Variance Estimator

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C does not have to be linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *stochastic,* $E\{x\} = 0$, $\operatorname{cov}\{x, x\} = E\{xx^T\} = P > 0$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0$, $\operatorname{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

    state and noise are uncorrelated $E\{xx^T\} = 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \operatorname{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T(\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

Minimum Variance Estimator

- **Goal:** How to choose the **weight matrix** in an **overdetermined** problem
- **Model:** $y = Cx + \varepsilon,$ (C does not have to be linearly independent)
  - **Measurement** (model output) $y \in \mathbb{R}^m$
  - **State** (model input) $x \in \mathbb{R}^n$ *stochastic,* $E\{x\} = 0,\ \mathrm{cov}\{x, x\} = E\{xx^T\} = P > 0$
  - **Noise** (output) $\varepsilon \in \mathbb{R}^m$ *stochastic,* $E\{\varepsilon\} = 0,\ \mathrm{cov}\{\varepsilon, \varepsilon\} = E\{\varepsilon\varepsilon^T\} = Q > 0$

    state and noise are uncorrelated $E\{xx^T\} = 0$

$$\hat{x} = Ky \qquad E\{\hat{x} - x\} = 0 \qquad \mathrm{Var}(\hat{x} - x) = E\{(\hat{x} - x)^T(\hat{x} - x)\}$$

holds for all $x \in \mathbb{R}^n$

Minimizes variance

- **Find:** $\hat{K}$

$$\boxed{\hat{K} = PC^T(CPC^T + Q)^{-1} \qquad \mathrm{cov}(\hat{x} - x) = P - PC^T(CPC^T + Q)^{-1}CP}$$