ECE276A: Sensing & Estimation in Robotics
Lecture 1: Linear Algebra and Probability Theory (Review)

Instructor:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Qiaojun Feng: qif007@eng.ucsd.edu
Tianyu Wang: tiw161@eng.ucsd.edu
Ibrahim Akbar: iakbar@eng.ucsd.edu
You-Yi Jau: yjau@eng.ucsd.edu
Harshini Rajachander: hrajacha@eng.ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Linear Algebra Review

Vectors

- A **vector** $\mathbf{x} \in \mathbb{R}^d$ with d dimensions is a collection of scalars $\mathbf{x}_i \in \mathbb{R}$ for $i = 1, \ldots, d$ organized is a column:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_d \end{bmatrix} \qquad \mathbf{x}^T = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_d \end{bmatrix}$$

- A **norm** on a vector space V over a subfield F is a function $\|\cdot\| : V \to \mathbb{R}$ such that for all $a \in F$ and all $\mathbf{x}, \mathbf{y} \in V$:
    - $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ (absolute homogeneity)
    - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)
    - $\|\mathbf{x}\| \geq 0$ (non-negativity)
    - $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$ (definiteness)
- The **Euclidean norm** of a vector $\mathbf{x} \in \mathbb{R}^d$ is $\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}^T\mathbf{x}}$ and satisfies:
    - $\max_{1 \leq i \leq d} |\mathbf{x}_i| \leq \|\mathbf{x}\|_2 \leq \sqrt{d} \max_{1 \leq i \leq d} |\mathbf{x}_i|$
    - $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$ (Cauchy-Schwarz Inequality)

Matrices

- A **matrix** $A \in \mathbb{R}^{m \times n}$ is a rectangular array of scalars $A_{ij} \in \mathbb{R}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$
- The entries of the **transpose** $A^T \in \mathbb{R}^{n \times m}$ of a matrix $A \in \mathbb{R}^{m \times n}$ are $A^T_{ij} = A_{ji}$. The transpose satisfies: $(AB)^T = B^T A^T$
- The **trace** of a matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal entries:

$$\mathrm{tr}(A) := \sum_{i=1}^{n} A_{ii} \qquad\qquad \mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$$

- The **determinant** of a matrix $A \in \mathbb{R}^{n \times n}$ is:

$$\det(A) := \sum_{j=1}^{n} A_{ij}\mathbf{cof}_{ij}(A) \qquad\qquad \det(AB) = \det(A)\det(B) = \det(BA)$$

  where $\mathbf{cof}_{ij}(A)$ is the **cofactor** of the entry A_{ij} and is equal to $(-1)^{i+j}$ times the determinant of the $(n-1) \times (n-1)$ submatrix that results when the i^{th}-row and j^{th}-col of A are removed. This recursive definition uses the fact that the determinant of a scalar is the scalar itself.

Matrices

- The **adjugate** is the transpose of the cofactor matrix:

$$\mathbf{adj}(A) := \mathbf{cof}(A)^T$$

- The **inverse** A^{-1} of A exists iff $\det(A) \neq 0$ and satisfies:

$$A^{-1} = \frac{\mathbf{adj}(A)}{\det(A)} \qquad\qquad (AB)^{-1} = B^{-1}A^{-1}$$

- If $A \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$ is a nonzero vector such that:

$$Aq = \lambda q \qquad \text{for } \lambda > 0$$

  then q is an **eigenvector** corresponding to the **eigenvalue** λ.

- A real matrix can have complex eigenvalues and eigenvectors, which appear in conjugate pairs. The n eigenvalues of $A \in \mathbb{R}^{n\times n}$ are precisely the n roots of the **characteristic polynomial** of A:

$$p(s) := \det(sI - A)$$

Matrices

- The roots of a polynomial are continuous functions of its coefficients and hence the eigenvalues of a matrix are continuous functions of its entries.

$$\operatorname{tr}(A) := \sum_{i=1}^{n} \lambda_i \qquad \det(A) := \prod_{i=1}^{n} \lambda_i$$

- The product $x^T Qx$ for $Q \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$ is called a **quadratic form** and Q can be assumed **symmetric**, $Q = Q^T$, because:

$$\frac{1}{2}x^T(Q + Q^T)x = x^T Qx, \qquad \forall x \in \mathbb{R}^n$$

- A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is **positive semidefinite** if $x^T Qx \geq 0$ for all $x \in \mathbb{R}^n$.
- A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is **positive definite** if it is positive semidefinite and if $x^T Qx = 0$ implies $x = 0$
- All eigenvalues of a symmetric matrix are **real**. Hence, all eigenvalues of a positive semidefinite matrix are non-negative and all eigenvalues of a positive definite matrix are positive.

Matrices

- The **Schur complement** of block D of $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is $S_D = A - BD^{-1}C$
- A symmetric matrix $M = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}$ is positive semidefinite **if and only if** both A and S_A are positive semidefinite (or both D and S_D are positive semidefinite).
- **Square completion**:

$$\frac{1}{2}x^T A x + b^T x + c = \frac{1}{2}\left(x + A^{-1}b\right)^T A \left(x + A^{-1}b\right) + c - \frac{1}{2} b^T A^{-1} b$$

Matrix Inversion Lemma

- **Woodbury matrix identity**:

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B\left(CA^{-1}B + D^{-1}\right)^{-1}CA^{-1}$$

- **Block matrix inversion**:

$$\begin{aligned}
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}^{-1} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}^{-1} \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}^{-1} \\
&= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1} & -\left(A - BD^{-1}C\right)^{-1}BD^{-1} \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1} & D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1}BD^{-1} \end{bmatrix}
\end{aligned}$$

Derivatives (numerator layout)

- **Derivatives by scalar**

$$\frac{d\mathbf{y}}{dx} = \begin{bmatrix} \frac{d\mathbf{y}_1}{dx} \\ \vdots \\ \frac{d\mathbf{y}_m}{dx} \end{bmatrix} \qquad \frac{d\mathbf{Y}}{dx} = \begin{bmatrix} \frac{d\mathbf{Y}_{11}}{dx} & \cdots & \frac{d\mathbf{Y}_{1n}}{dx} \\ \vdots & \ddots & \vdots \\ \frac{d\mathbf{Y}_{m1}}{dx} & \cdots & \frac{d\mathbf{Y}_{mn}}{dx} \end{bmatrix}$$

- **Derivatives by vector**

$$\frac{dy}{d\mathbf{x}} = \underbrace{\begin{bmatrix} \frac{dy}{dx_1} & \cdots & \frac{dy}{dx_p} \end{bmatrix}}_{[\nabla_{\mathbf{x}} y]^T \text{ (gradient transpose)}} \qquad \frac{d\mathbf{y}}{d\mathbf{x}} = \underbrace{\begin{bmatrix} \frac{d\mathbf{y}_1}{d\mathbf{x}_1} & \cdots & \frac{d\mathbf{y}_1}{d\mathbf{x}_p} \\ \vdots & \ddots & \vdots \\ \frac{d\mathbf{y}_m}{d\mathbf{x}_1} & \cdots & \frac{d\mathbf{y}_m}{d\mathbf{x}_p} \end{bmatrix}}_{\text{Jacobian}} \qquad \frac{d\mathbf{Y}}{d\mathbf{x}} \in \mathbb{R}^{m \times n \times p}$$

- **Derivatives by matrix**

$$\frac{dy}{d\mathbf{X}} = \begin{bmatrix} \frac{dy}{d\mathbf{X}_{11}} & \cdots & \frac{dy}{d\mathbf{X}_{p1}} \\ \vdots & \ddots & \vdots \\ \frac{dy}{d\mathbf{X}_{1q}} & \cdots & \frac{dy}{d\mathbf{X}_{pq}} \end{bmatrix} \qquad \frac{d\mathbf{y}}{d\mathbf{X}} \in \mathbb{R}^{m \times p \times q} \qquad \frac{d\mathbf{Y}}{d\mathbf{X}} \in \mathbb{R}^{m \times n \times p \times q}$$

Probability Theory Review

Events

- **Experiment**: any procedure that can be repeated infinitely and has a well-defined set of possible outcomes.
- **Sample space** Ω: the set of possible outcomes of an experiment.
  - $\Omega = \{HH, HT, TH, TT\}$
  - $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$
- **Event** A: a subset of the possible outcomes Ω
  - $A = \{HH\},\ B = \{HT, TH\}$
- **Probability of an event**: $\mathbb{P}(A) = \frac{N_A}{N} = \frac{\#\text{possible occurances of } A}{\#\text{all possible outcomes}}$

Probability Axioms

- **Probability Axioms**:
  - $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\Omega) = 1$
  - If $\{A_i\}$ are disjoint ($A_i \cap A_j = \emptyset$), then $\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i)$
- **Corollary**:
  - $\mathbb{P}(\emptyset) = 0$
  - $\max\{\mathbb{P}(A), \mathbb{P}(B)\} \leq \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B)$
  - $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

Set of Events

- **Conditional Probability**: $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$
- **Total Probability Theorem**: If $\{A_1, \ldots, A_n\}$ is a partition of Ω, i.e., $\Omega = \bigcup_i A_i$ and $A_i \cap A_j = \emptyset, i \neq j$, then:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$$

- **Bayes Theorem** If $\{A_1, \ldots, A_n\}$ is a partition of Ω, then:

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}$$

- **Independent events**: $\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i)$
  - observing one does not give any information about another
  - in contrast, disjoint events never occur together: one occuring tells you that others will not occur and hence, <u>disjoint events are always dependent</u>

Measure and Probability Space

- σ-**algebra**: a collection of subsets of Ω closed under complementation and countable unions.
- **Borel σ-algebra** $\mathcal{B}$: the smallest σ-algebra containing all open sets from a topological space. Necessary because there is no valid translation invariant way to assign a finite measure to all subsets of $[0, 1)$.
- **Measurable space**: a tuple $(\Omega, \mathcal{F})$, where Ω is a sample space and $\mathcal{F}$ is a σ-algebra.
- **Measure**: a function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$ and countable additivity $\mu(\cup_i A_i) = \sum_i \mu(A_i)$ for disjoint A_i.
- **Probability measure**: a measure that satisfies $\mu(\Omega) = 1$.
- **Probability space**: a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a sample space, $\mathcal{F}$ is a σ-algebra, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure.

Random Variable

- **Random variable** X: an $\mathcal{F}$-measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$, i.e., a function $X : \Omega \to \mathbb{R}$ s.t. the preimage of every set in $\mathcal{B}$ is in $\mathcal{F}$.

- **Distribution function** $F(x)$ **of a random variable** X: a function $F(x) := \mathbb{P}(X \leq x)$ that is non-decreasing, right-continuous, and $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

- **Density/mass function** $f(x)$ **of a random variable** X

Continuous RV
$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P} \circ X^{-1})$:
  - $f(x) \geq 0$
  - $\int f(y)dy = 1$
  - $F(x) = \int_{-\infty}^{x} f(y)dy = \mathbb{P}(X \leq x)$
  - $\mathbb{P}(X = x) = F(x) - F(x^-) = \lim_{\epsilon\to 0} \int_{x-\epsilon}^{x} f(y)dy = 0$
  - $\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$

Discrete RV
$X : (\Omega, 2^\Omega, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P} \circ X^{-1})$:
  - $f(x) = \mathbb{P}(X = x) \geq 0$
  - $\sum_i f(i) = 1$
  - $F(x) = \sum_{i\in\mathbb{Z}, i\leq x} f(i) = \mathbb{P}(X \leq x)$

Expectation and Variance

- Given a random variable X with pdf p and a measurable function g, the **expectation** of $g(X)$ is:

$$\mathbb{E}\left[g(X)\right] = \int g(x)p(x)dx$$

- The **variance** of $g(X)$ is:

$$\begin{aligned} Var[g(X)] &= \mathbb{E}\left[\left(g(X) - \mathbb{E}[g(X)]\right)\left(g(X) - \mathbb{E}[g(X)]\right)^T\right] \\ &= \mathbb{E}\left[g(X)g(X)^T\right] - \mathbb{E}[g(X)]\mathbb{E}[g(X)]^T \end{aligned}$$

- The **variance** of a sum of random variables is:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j\neq i} Cov(X_i, X_j)$$

$$Cov(X_i, X_j) = \mathbb{E}\left((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)^T\right) = \mathbb{E}(X_i X_j^T) - \mathbb{E}X_i\mathbb{E}X_j^T$$

Set of Random Variables

- The **joint distribution** of random variables $\{X_i\}_{i=1}^n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ defines their simultaneous behavior and is associated with a cumulative distribution function $F(x_1, \ldots, x_n) := \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$. The CDF $F_i(x_i)$ of X_i defines its **marginal distribution**.
- Random variables $\{X_i\}_{i=1}^n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are **jointly independent** iff for all $\{A_i\}_{i=1}^n \subset \mathcal{F}$, $\mathbb{P}(X_i \in A_i, \forall i) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$
- Let X and Y be random variables and suppose $\mathbb{E}X$, $\mathbb{E}Y$, and $\mathbb{E}XY$ exist. Then, X and Y are **uncorrelated** iff $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$ or equivalently $Cov(X, Y) = 0$.
- Independence implies uncorrelatedness

Change of Density

- **Convolution**: Let X and Y be independent random variables with pdfs f and g, respectively. Then, the pdf of $Z = X + Y$ is given by the convolution of f and g:

$$[f * g](z) := \int f(z - y)g(y)dy$$

- **Change of Density**: Let $Y = f(X)$. Then, with $dy = \left|\det\left(\frac{df}{dx}(x)\right)\right| dx$:

$$\begin{aligned}\mathbb{P}(Y \in A) = \mathbb{P}(X \in f^{-1}(A)) &= \int_{f^{-1}(A)} p_x(x)dx \\ &= \int_A \underbrace{\frac{1}{\left|\det\left(\frac{df}{dx}(f^{-1}(y))\right)\right|} p_x(f^{-1}(y))}_{p_y(y)} dy\end{aligned}$$

Conditional and Total Probability

- **Total Probability Theorem**: If two random variables X, Y have a joint pdf p, the marginal pdf of X is:

$$p(x) = \int p(x, y) dy$$

- **Conditional Distribution**: If two random variables X, Y have a joint pdf p, the pdf of X conditioned on $Y = y$ is

$$p(x|y) := \frac{p(x, y)}{\int p(x, y) dx}$$

- **Bayes Theorem**: The conditional, marginal, and joint pdfs of X and Y are related:

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y) \quad \Rightarrow \quad \boxed{p(x|y) = \frac{p(y|x)p(x)}{\int p(y \mid x')p(x')dx'}}$$

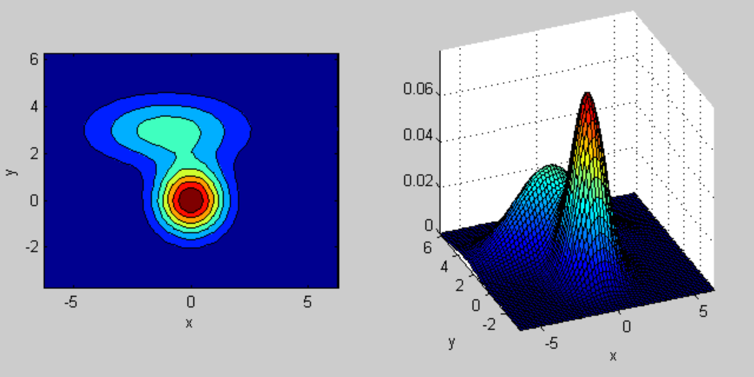
Gaussian Distribution

- The **Mahalaonobis distance** for vector $x \in \mathbb{R}^n$ and symmetric positive-definie matrix $S \in \mathbb{S}^n_{>0}$ is: $\|x\|_S^2 := x^T S^{-1} x$

- **Gaussian random variable** $X \sim \mathcal{N}(\mu, \Sigma)$
  - paramteres: **mean** $\mu \in \mathbb{R}^n$, **covariance** $\Sigma \in \mathbb{S}^n_{\succeq 0}$
  - pdf: $\phi(x; \mu, \Sigma) := \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$
  - expectation: $\mathbb{E}[X] = \int x\phi(x; \mu, \Sigma)dx = \mu$
  - variance: $Var[X] = \Sigma$

- **Gaussian mixture** $X \sim \mathcal{N}\mathcal{M}(\{\alpha_k\}, \{\mu_k\}, \{\Sigma_k\})$
  - parameters: **weights** $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$,
    means $\mu_k \in \mathbb{R}^n$, **covariances** $\Sigma_k \in \mathbb{S}^n_{\succeq 0}$
  - pdf: $p(x) := \sum_k \alpha_k \phi(x; \mu_k, \Sigma_k)$
  - expectation: $\mathbb{E}[X] = \int xp(x)dx = \sum_k \alpha_k \mu_k =: \bar{\mu}$
  - variance: $\mathbb{E}\left[XX^T\right] - \mathbb{E}[X]\mathbb{E}[X]^T = \sum_k \alpha_k \left(\Sigma_k + \mu_k \mu_k^T\right) - \bar{\mu}\bar{\mu}^T$

PDF of a Mixture of Two 2-D Gaussians



Examples

Matrix Calculus

- $\frac{d}{dX_{ij}} X = e_i ej^T$
- $\frac{d}{dx} Ax = A$
- $\frac{d}{dx} x^T Ax = x^T (A + A^T)$
- $\frac{d}{dx} M^{-1}(x) = -M^{-1}(x) \frac{dM(x)}{dx} M^{-1}(x)$
- $\frac{d}{dX} \operatorname{tr}(AX^{-1}B) = -(X^{-1}BAX^{-1})^T$
- $\frac{d}{dX} \log \det X = X^{-T}$

Matrix Calculus

- $\frac{d}{dx}Ax = \begin{bmatrix} \frac{d}{dx_1}\sum_{j=1}^n A_{1j}x_j & \cdots & \frac{d}{dx_n}\sum_{j=1}^n A_{1j}x_j \\ \vdots & \ddots & \vdots \\ \frac{d}{dx_1}\sum_{j=1}^n A_{mj}x_j & \cdots & \frac{d}{dx_n}\sum_{j=1}^n A_{mj}x_j \end{bmatrix} = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$
- $\frac{d}{dx}x^T Ax = x^T A^T \frac{dx}{dx} + x^T \frac{dAx}{dx} = x^T(A^T + A)$
- $M(x)M^{-1}(x) = I \quad \Rightarrow \quad 0 = \left[\frac{d}{dx}M(x)\right] M^{-1}(x) + M(x)\left[\frac{d}{dx}M^{-1}(x)\right]$
- $$\begin{aligned} \frac{d}{dX_{ij}}\operatorname{tr}(AX^{-1}B) &= \operatorname{tr}(A\frac{d}{dX_{ij}}X^{-1}B) = -\operatorname{tr}(AX^{-1}e_i e_j^T X^{-1}B) \\ &= -e_j^T X^{-1}BAX^{-1}e_i = -e_i^T\left(X^{-1}BAX^{-1}\right)^T e_j \end{aligned}$$
- $$\begin{aligned} \frac{d}{dX_{ij}}\log\det X &= \frac{1}{\det(X)}\frac{d}{dX_{ij}}\sum_{k=1}^n X_{ik}\mathbf{cof}_{ik}(X) \\ &= \frac{1}{\det(X)}\mathbf{cof}_{ij}(X) = \frac{1}{\det(X)}\mathbf{adj}_{ji}(X) = e_i^T X^{-T}e_j \end{aligned}$$

Events

- An experiment consists of randomly selecting one chip among ten chips marked $1, 2, 2, 3, 3, 3, 4, 4, 4, 4$.
  - What is a reasonable sample space for this experiment?
  - What is the probability of observing a chip marked with an even number?
  - What is the probability of observing a chip marked with a prime number?

Events

- An experiment consists of randomly selecting one chip among ten chips marked $1, 2, 2, 3, 3, 3, 4, 4, 4, 4$.
  - What is a reasonable sample space for this experiment? $\Omega = \{1, 2, 3, 4\}$
  - What is the probability of observing a chip marked with an even number?

$$\mathbb{P}(\{2, 4\}) = \mathbb{P}(\{2\} \cup \{4\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) = \frac{6}{10}$$

  - What is the probability of observing a chip marked with a prime number?

$$\mathbb{P}(\{2, 3\}) = \mathbb{P}(\{2\} \cup \{3\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) = \frac{5}{10}$$

Independent Events

- A box contains 7 green and 3 red chips.
- Experiment: select one chip, replace the drawn chip, and repeat until the color red has been observed four times
- Assuming that no draw affects or is affected by any other draw, what is the probability that the experiment terminates on the ninth draw?

Independent Events

- Let Ω denote the sample space for this experiment, which is a countably infinite set of all ordered tuples such that:
  - Each term is either g or r
  - The last component of the tuple is r
  - There are exactly four components of r in the tuple
- Let E be the set of elements in Ω which have 9 components, e.g., $(g, r, g, r, g, r, g, g, r) \in E$
- Idea:
  - Show that every singleton subset of E has the same probability p_e
  - Determine the cardinality of E so that $\mathbb{P}(E) = \sum_{e \in E} \mathbb{P}(e) = |E| p_e$
- Due to independence, for any element $e \in E$ we have:

$$\mathbb{P}(e) = \mathbb{P}\left(e_1 \cap e_2 \cap \cdots \cap e_9\right) = \prod_{i=1}^{9} \mathbb{P}(e_i) = \left(\frac{3}{10}\right)^4 \left(\frac{7}{10}\right)^5$$

- Since the last component of each 9-tuple $e \in E$ must be r, the cardinality of E is the number of ways to distribute 3 red chips among 8 slots, i.e., $|E| = \binom{8}{3}$

Expectation

- Suppose $V = (X, Y)$ is a continuous random vector with density $f_V(x, y) = 8xy$ for $0 < y < x$ and $0 < x < 1$. Let $g(x, y) := 2x + y$.
  - Determine $\mathbb{E}\left[g(V)\right]$
  - Evaluate $\mathbb{E}\left[X\right]$ and $\mathbb{E}\left[Y\right]$ by finding the marginal densities of X and Y and then evaluating the appropriate univariate integrals
  - Determine $Var\left[g(V)\right]$

Expectation

$$\mathbb{E}\left[2X+Y\right] = \int_0^1 \int_0^x (2x+y) 8xy \; dydx = \frac{32}{15}$$

$$f_X(x) = \int_0^x 8xy \; dy = 4x^3 \text{ for } 0 \leq x \leq 1$$

$$\mathbb{E}\left[X\right] = \int_0^1 x f_X(x) dx = \int_0^1 4x^4 dx = \frac{4}{5}$$

$$f_Y(y) = \int_y^1 8xy \; dx = 4y - 4y^3 \text{ for } 0 \leq y \leq 1$$

$$\mathbb{E}\left[Y\right] = \int_0^1 y f_Y(y) dy = \int_0^1 4y^2 - 4y^4 dy = \frac{8}{15}$$

$$Var\left[g(V)\right] = \mathbb{E}\left[(g(V) - \mathbb{E}\left[g(V)\right])^2\right] = \mathbb{E}\left[\left(2X+Y-\frac{32}{15}\right)^2\right]$$

$$= \int_0^1 \int_0^x \left(2x+y-\frac{32}{15}\right)^2 8xy \; dydx = \frac{17}{75}$$

Conditional Probability

- Suppose that $V = (X, Y)$ is a discrete random vector with probability mass function

$$f_V(x,y) = \begin{cases} 0.10 & \text{if } (x,y) = (0,0) \\ 0.20 & \text{if } (x,y) = (0,1) \\ 0.30 & \text{if } (x,y) = (1,0) \\ 0.15 & \text{if } (x,y) = (1,1) \\ 0.25 & \text{if } (x,y) = (2,2) \\ 0 & \text{elsewhere} \end{cases}$$

  - What is the conditional probability that V is $(0,0)$ given that V is $(0,0)$ or $(1,1)$?
  - What is the conditional probability that X is 1 or 2 given that Y is 0 or 1?
  - What is the probability that X is 1 or 2?
  - What is the probability mass function of $X \mid Y = 0$?
  - What is the expected value of $X \mid Y = 0$?

Conditional Probability

$$\mathbb{P}\left(V \in \{(0,0)\} \mid V \in \{(0,0),(1,1)\}\right) = \frac{\mathbb{P}\left(V \in \{(0,0)\} \cap \{(0,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(1,1)\}\right)}$$

$$= \frac{0.10}{0.25} = 0.4$$

$$\mathbb{P}\left(X \in \{1,2\} \mid Y \in \{0,1\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R} \mid V \in \mathbb{R} \times \{0,1\}\right)$$

$$= \frac{\mathbb{P}\left(V \in \{(1,0),(1,1)\}\right)}{\mathbb{P}\left(V \in \{(0,0),(0,1),(1,0),(1,1)\}\right)} = \frac{45}{75}$$

$$\mathbb{P}\left(X \in \{1,2\}\right) = \mathbb{P}\left(V \in \{1,2\} \times \mathbb{R}\right) = 0.7$$

$$f_{X|Y=0}(x) = \frac{f_V(x,0)}{\sum_{x'} f_V(x',0)dx'} = \frac{1}{4} f_V(x,0) = \begin{cases} 0.25 & \text{if } x = 0 \\ 0.75 & \text{if } x = 1 \end{cases}$$

$$\mathbb{E}\left[X \mid Y = 0\right] = \sum_{x \in \{0,1\}} x f_{X|Y=0}(x) = \frac{3}{4}$$

Change of Density

- Let $X \sim \mathcal{N}(0, \sigma^2)$ be a Gaussian random variable
- Let $Y = f(X)$ be a random variable defined as a nonlinear transformation of X according to the function $f(x) := \exp(x)$
- What is the pdf $p(y)$ of Y?

Change of Density

- Note that $f(x)$ is invertible $f^{-1}(y) = \log(y)$
- The infinitesimal integration volumes for y and x are related by:

$$dy = \left|\det\left(\frac{df}{dx}(x)\right)\right| dx = \exp(x) dx$$

- Using the change of density theorem:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \phi(x; 0, \sigma^2) dx = \int_{0}^{\infty} \frac{1}{\exp(\log(y))} \phi(\log(y); 0, \sigma^2) dy \\ &= \int_{0}^{\infty} \underbrace{\frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{\log^2(y)}{\sigma^2}\right)}_{p(y)} dy \end{aligned}$$

Change of Density

- Let $V := (X, Y)$ be a random vector with pdf:

$$p_V(x,y) := \begin{cases} 2y - x & x < y < 2x \text{ and } 1 < x < 2 \\ 0 & \text{else} \end{cases}$$

- Let $T := (M, N) = g(V) := \left(\frac{2X-Y}{3}, \frac{X+Y}{3}\right)$ be a function of V
- Note that $X = M + N$ and $Y = 2N - M$ and hence the pdf of V is non-zero for $0 < m < n/2$ and $1 < m + n < 2$. Also:

$$\det\left(\frac{dg}{dv}\right) = \det\begin{bmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{bmatrix} = \frac{1}{3}$$

- The pdf T is:

$$p_T(m,n) = \begin{cases} \frac{1}{\left|\det\left(\frac{dg}{dv}(m+n, 2n-m)\right)\right|} p_V(m+n, 2n-m), & \begin{array}{l} 0 < m < n/2 \text{ and} \\ 1 < m+n < 2, \end{array} \\ 0, & \text{else.} \end{cases}$$