ECE276A: Sensing & Estimation in Robotics
Lecture 2: Probability Theory

Lecturer:

Nikolay Atanasov: natanasov@ucsd.edu

Teaching Assistants:

Siwei Guo: s9guo@eng.ucsd.edu

Anwesan Pal: a2pal@eng.ucsd.edu

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

Project 1 is out

- Click on P1 on the schedule and enter the password
- **Start early**! We have not covered everything but:
  - Make sure you can open and download everything
  - Try loading, viewing, and labeling the data
  - Work on the first 2 problems
  - Practice your python skills
  - Start your report
- TA sessions: Friday, 2:00 - 3:30 pm, Jacobs Hall 2315
- First TA session – python tutorial

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**: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}(B) \neq 0$
- **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**: $\mathbb{P}(\bigcap_i A_i) = \prod_i \mathbb{P}(A_i)$
  - 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, disjoint events are always dependent

Measure Space

- σ-**algebra**: a collection of subsets of Ω closed under complementation and countable unions.
- **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.
- A measure μ is σ-**finite** on $(\Omega, \mathcal{F})$, if Ω can be obtained as the countable union $\cup_n A_n$ of sets $A_n \in \mathcal{F}$ of finite measure, $\mu(A_n) < \infty$.

Probability Space

- **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.
- **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 $[0, 1)$.

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
$X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P} \circ X^{-1})$:

  - $f(x) \geq 0$
  - $\int f(y)dy = 1$
  - $F(x) = \int_{-\infty}^{x} f(y)dy = \mathbb{P}(X \leq x)$

Discrete RV
$X : (\Omega, 2^{\Omega}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}, \mathbb{P} \circ X^{-1})$:

  - $f(x) = \mathbb{P}(X = x) \geq 0$
  - $\sum_i f(i) = 1$
  - $F(x) = \sum_{i\in\mathbb{Z}, i\leq x} f(i) = \mathbb{P}(X \leq x)$

  - $\mathbb{P}(X = x) = F(x) - F(x^-) = \lim_{\epsilon\to 0} \int_{x-\epsilon}^{x} f(y)dy = 0$
  - $\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$

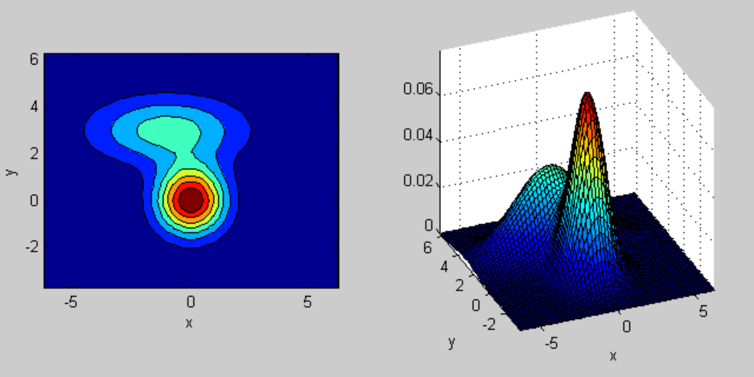
Expectation

- **Lebesgue Integration**: The integral $\int g d\mu$ of a measurable function g on measurable space $(\Omega, \mathcal{F})$ with a σ-finite measure μ can be defined. In the case that μ has a pdf p, the Lebesgue integral is equivalent to a Riemann integral: $\int g d\mu = \int g(x)p(x)dx$.
- **Expectation**: Given a random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^n, \mathcal{B}^n, \mathbb{P} \circ X^{-1})$ and a measurable function $g : (\mathbb{R}^n, \mathcal{B}^n, \mathbb{P} \circ X^{-1}) \to (\mathbb{R}^m, \mathcal{B}^m, \mathcal{L})$, the expectation of $g(X)$ is defined as follows:
$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} g(x)d\mathbb{P}(X^{-1}(x)) = \int_{\mathbb{R}^m} y d\mathcal{L}(y)$$
When X has a pdf p and g has a pdf l, the above simplifies to:
$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} g(x)p(x)dx = \int_{\mathbb{R}^m} y l(y)dy$$
- **Expectation of an Indicator**: $\mathbb{E}[\mathbb{1}_A] = \int \mathbb{1}_A(\omega)d\mathbb{P}(\omega) = \mathbb{P}(A)$
- **Variance of a random variable X**: $Var[X] := \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right] = \mathbb{E}\left[XX^T\right] - \mathbb{E}[X]\mathbb{E}[X]^T$

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** $X \sim \mathcal{NM}(\{\alpha_k\}, \{\mu_k\}, \{\Sigma_k\})$
  - 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: $\mathbb{E}[X] = \int x p(x) dx = \sum_k \alpha_k \mu_k =: \bar{\mu}$
  - variance: $\mathbb{E}\left[XX^T\right] - \mathbb{E}[X]\mathbb{E}[X]^T = \sum_k \alpha_k \left(\Sigma_k + \mu_k \mu_k^T\right) - \bar{\mu}\bar{\mu}^T$

Mixture of two 2-D Gaussians



Other Distributions

- **Uniform continuous random variable** $X \sim \mathcal{U}(a, b)$
  - paramteres: $-\infty < a < b < \infty$
  - pdf: $p(x) = \frac{1}{b-a}$ for $x \in [a, b]$
  - expectation: $\mathbb{E}[X] = \frac{1}{2}(a + b)$
  - variance: $Var[X] = \frac{1}{12}(b - a)^2$
- **Chi-Square random variable** $X \sim \chi^2(k)$
  - paramteres: degrees of freedom $k \in \mathbb{N}$
  - pdf: $p(x) = \frac{1}{2^{k/2}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$ for $x \geq 0$
  - expectation: $\mathbb{E}[X] = k$
  - variance: $Var[X] = 2k$
  - if $Y \sim \mathcal{N}(\mu, \Sigma)$, then $X := \|Y - \mu\|_\Sigma^2 \sim \chi^2(n)$
- The **Gamma function** $\Gamma(\alpha) := \int_0^\infty y^{\alpha-1} e^{-y} dy$, $\alpha > 0$, satisfies $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(k + 1) = k\Gamma(k)$

Composition and Norm

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

$$\mathbb{P}(Y \in A) = \mathbb{P}(X \in f^{-1}(A)) = \int_{f^{-1}(A)} p_x(x)dx$$

$$\overset{\text{change of}}{\underset{\text{variables}}{=}} \int_A \boxed{\frac{1}{\det\left(\frac{df}{dx}(f^{-1}(y))\right)} p_x(f^{-1}(y))} dy$$

- **L^p-Space**: Let (S, Σ, μ) be a measure space and $1 \leq p < \infty$. Then, $f \in L^p(\mu)$ if f is measurable wrt S and

$$\boxed{\|f\|_p := \left(\int_S |f|^p d\mu\right)^{\frac{1}{p}} < \infty}$$

- **Differential Entropy** of a continuous random variable X with pdf p is: $h(p) := -\int p(x) \log p(x) dx$

- **Kullback-Leibler (KL) divergence** from pdf p to pdf q is: $d_{\mathcal{KL}}(p||q) := \int p(x) \log \frac{p(x)}{q(x)} dx$

Example: 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:

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

Inequalities

- **Markov's Inequality** $\boxed{0 \le X, f} \quad \Rightarrow \quad \boxed{\mathbb{P}(X \in A) \le \frac{\mathbb{E}f(X)}{\inf_{x \in A} f(x)}}$

  . Examples: $\boxed{\mathbb{P}(X \ge a) \le \frac{\mathbb{E}X}{a}}$ and $\boxed{\mathbb{P}\left(|X - \mathbb{E}X| \ge a\right) \le \frac{Var(X)}{a}}$

- $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if one of the following holds:
  - $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y), \quad \forall x, y \in \mathbb{R}^n, \lambda \in [0,1]$
  - $f(y) \ge f(x) + \nabla f(x)^T (y - x), \quad \forall x, y \in \mathbb{R}^n$
  - $\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbb{R}^n$

- **Jensen's Inequality** Let X be random variable and $\psi : \mathbb{R} \to \mathbb{R}$ be a convex function. Then, $\boxed{\mathbb{E}\psi(X) \ge \psi(\mathbb{E}(X))}$ provided both exist.
  - $|\mathbb{E}|X_t| - \mathbb{E}|X|| \le \mathbb{E}||X_t| - |X|| \le \mathbb{E}|X_t - X|$, thus convergence in L^1 implies convergence in expectation but *not the converse*
  - $(\mathbb{E}|X|)^2 \le \mathbb{E}X^2$
  - $f\left(\frac{\sum_i a_i z_i}{\sum_i a_i}\right) \le \frac{\sum_i a_i f(z_i)}{\sum_i a_i}$ for convex f, points $\{z_i\}$ in f's domain, and $a_i \ge 0$

Inequalities

- **Hölder's Inequality** (generalizes **Cauchy-Schwarz**) Let (S, Σ, μ) be a measure space and let $p, q \in [1, \infty]$ with $\boxed{1/p + 1/q = 1}$. Then, for measurable f, g on S: $\boxed{\|fg\|_1 \leq \|f\|_p \|g\|_q}$.
  - If $f, g \in L^2(\mu)$, then $|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$
  - $\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}$
  - $\mathbb{E}(|X|^r) \leq (\mathbb{E}(|X|^s))^{\frac{r}{s}}$
  - $\displaystyle\sum_{k=1}^{\infty} |x_k y_k| \leq \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{\infty} |y_k|^q\right)^{\frac{1}{q}}$
  - $|a^T b| \leq \|a\|_2 \|b\|_2$

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
- Two random variables X and Y are orthogonal if $\mathbb{E}[X^T Y] = 0$

Set of Random Variables

- **Total Probability Theorem** Given two random variables X_1, X_2 with a joint pdf p, one can obtain the marginal pdf p_1 of X_1 as follows: $p_1(x_1) := \int p(x_1, x_2) dx_2$.
- **Conditional expectation** Let X be an RV on $(\Omega, \mathcal{F}_0, \mathbb{P})$ with $\mathbb{E}|X| < \infty$ and let $\mathcal{F} \subseteq \mathcal{F}_0$. Then, $Y := \mathbb{E}[X \mid \mathcal{F}]$ is an RV that satisfies:
  - (Measurability Axiom) $Y \in \mathcal{F}$,
  - (Integral Axiom) $\int_G Y d\mathbb{P} = \int_G X d\mathbb{P}$ for all $G \in \mathcal{F}$.

  Y exists and is unique up to values on a set of measure zero. The following notation is common:
  - $\mathbb{E}[X \mid Z] := \mathbb{E}[X \mid \sigma(Z)]$
  - $\mathbb{P}(A \mid B) := \mathbb{E}[\mathbb{1}_A \mid \sigma(B)]$

Set of Random Variables

- **Conditional distribution** If (X, Y) has a pdf f on $\mathbb{R}^2$ and $\mathbb{E}|g(X)| < \infty$, then $\mathbb{E}[g(X) \mid \sigma(Y)] = h(Y)$ for $h(y) := \int g(x) \frac{f(x,y)}{\int f(x,y) dx} dx$. Note that this defines the pdf of X conditioned on $Y = y$ as $\boxed{p(x|y) := \frac{f(x,y)}{\int f(x,y) dx}}$

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

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y)$$

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

Set of Random Variables

- **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:

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

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

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