Rob 501 - Mathematics for Robotics
Recitation #7

Nils Smit-Anseeuw (Courtesy: Abhishek Venkataraman, Wubing Qin)

Nov 6, 2018

1 Probability

1. **Definition:** An experiment is the execution of any process resulting in a measurable outcome, called a sample, denoted as ω.
2. **Definition:** The sample space for an experiment is the set of all possible outcomes, denoted as Ω.
3. **Definition:** An event A is a collection of possible outcomes $\omega \in \Omega$, possessing some characteristics.
4. **Definition:** A σ-field or σ-algebra is a collection of events, denoted as $\mathcal{F} = \{A_1, A_2, \dots\}$ which has some special properties:
   - $\Omega \in \mathcal{F}$.
   - If $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$.
   - If $\{A_1, A_2\} \subset \mathcal{F}$, then $A_1 \cup A_2 \subset \mathcal{F}$.
5. **Definition:** Let F be a σ-field, defined on a sample space ω. Then a probability measure is a set function $P : \mathcal{F} \to [0, 1]$ with two special properties:
   - $P(\Omega) = 1$.
   - For any set of disjoint (i.e., mutually exclusive) events $\{A_1, A_2, \dots, A_N\} \subset F$,
   $$P(A_1 \cup A_2 \cup \ldots \cup A_N) = \sum_{i=1}^{N} P(A_i).$$

   The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space.
6. **Definition:** The conditional probability of event A_1, given event A_2, is
   $$P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)}.$$
7. **Total probability theorem:** Suppose $\{A_1, A_2, \dots, A_n\}$ are disjoint events in probability space $(\Omega, \mathcal{F}, P)$, and partition Ω, i.e.,
   - $\forall\, i, j,\ A_i \cap A_j = \emptyset$,
   - $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

   Then for some other event B,
   $$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

8. **Bayes' Theorem:** Suppose A_1 and A_2 are two events in probability space $(\Omega, \mathcal{F}, P)$, and that $P(A_2) > 0$. Then

$$P(A_1|A_2) = \frac{P(A_2|A_1)P(A_1)}{P(A_2)}.$$

Ex: Given a digital signal transmission line, consists of a transmitter and a receiver. The signal transmitted only takes logic '0' and '1'. Let X be the signal transmitted by the transmitter, and Y be the signal received by the receiver. Suppose we know the property of the signal that needs to be transmitted, i.e.,$P(X = 0) = p$, and the reliability of the transmission line, i.e., $P(Y = 0|X = 0) = q$ and $P(Y = 1|X = 1) = r$.

(a) Try to understand the concepts of experiment, samples, sample space, event, σ-field, and probability measure.

Experiment: The transmitter transmits a signal X and the receiver receives Y, the outcome is denoted as $\omega = (X, Y)$.
Sample: $\omega_1 = (0,0)$, $\omega_2 = (0,1)$, $\omega_3 = (1,0)$, $\omega_4 = (1,1)$.
Sample space: $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$.
Event: $A_1 = \{\text{Error occurs in transmission}\} = \{(0,1), (1,0)\}$,
$A_2 = \{\text{Data transceived correctly}\} = \{(0,0), (1,1)\}$,
$X_0 = \{\text{Data transmitted is } 0\} = \{(0,0), (0,1)\}$,
$X_1 = \{\text{Data transmitted is } 1\} = \{(1,0), (1,1)\}$,
$Y_0 = \{\text{Data received is } 0\} = \{(0,0), (1,0)\}$,
$Y_1 = \{\text{Data received is } 1\} = \{(0,1), (1,1)\}$.
σ-field: The following are several possible σ-fields.
(1) $\mathcal{F} = \{\emptyset, \Omega\}$,
(2) $\mathcal{F} = \{\emptyset, \{(0,0)\}, \{(0,1), (1,0), (1,1)\}, \Omega\}$,
(3) $\mathcal{F} = \{\emptyset, \{(0,0)\}, \{(0,1), (1,0), (1,1)\}, \{(0,0), (0,1)\}, \{(1,0), (1,1)\}, \Omega\}$,
(4) $\mathcal{F} = 2^\Omega$ is the largest possible σ-field.
Probability measure: $P(Y = 0|X = 0) = q$ and $P(Y = 1|X = 1) = r$.

(b) What is the probability of receiving a '0', i.e., $P(Y = 0)$?
Experiment: The transmitter transmits a signal, the outcome is X.
Sample space: $\Omega = \{\{X = 0\}, \{X = 1\}\}$. So $\{A_1 = \{X = 0\}, A_2 = \{X = 1\}\}$ is a partition of the sample space Ω. According to the total probability theorem,

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1) = qp + (1 - r)(1 - p).$$

(c) What is the probability of receiving a '1', i.e., $P(Y = 1)$?
According to the total probability theorem,

$$P(Y = 1) = P(Y = 1|X = 0)P(X = 0) + P(Y = 1|X = 1)P(X = 1) = (1 - q)p + r(1 - p).$$

(d) If the receiver receives a '0', what is the probability of that the signal transmitted is also '0', i.e., $P(X = 0|Y = 0)$?
According to the Bayes' theorem,

$$P(X = 0|Y = 0) = \frac{P(Y = 0|X = 0)P(X = 0)}{P(Y = 0)} = \frac{qp}{qp + (1 - r)(1 - p)}.$$

(e) If the receiver receives a '1', what is the probability of the signal transmitted also being '1', i.e., $P(X = 1|Y = 1)$?

According to the Bayes' theorem,

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1)} = \frac{r(1-p)}{(1-q)p + r(1-p)}.$$

(f) If we know the signal the receiver receives is y, what is the probability of that the signal transmitted is the same as the signal received, i.e., $P(X = Y)$?

$$P(X = Y) = P(Y = 0|X = 0)P(X = 0) + P(Y = 1|X = 1)P(X = 1) = qp + r(1-p).$$

2 Random variables

1. **Definition:** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A scalar random variable is a mapping $X : \Omega \to R$, i.e., $x = X(\omega)$.

2. **Definition:** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A vector random variable is a mapping $X : \Omega \to R^n$, i.e., $x = X(\omega)$.

3. **Definition:** Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then the Cumulative Distribution Function (CDF) of a scalar random variable $X \in \mathbb{R}$ is

$$F_X(x) = P(X(\omega) \leq x),$$

and the Probability Density Function (PDF) is

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

If $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \in \mathbb{R}^n$ is an n-dimensional vector random variable, then the CDF is

$$F_X(x) = F_{X_1 X_2 \cdots X_n}(x_1, x_2, \dots, x_n) = P(X_i(\omega) \leq x_i, \forall\, i = 1, \dots, n),$$

and the PDF is

$$f_X(x) = f_{X_1 X_2 \cdots X_n}(x_1, x_2, \dots, x_n) = \frac{\partial F_X(x)}{\partial x_1 \partial x_2 \dots \partial x_n}.$$

4. **Properties of CDF:** If $F_X(x)$ is the CDF of a scalar random variable $X \in \mathbb{R}$, then

   - $F_X(x)$ is non-decreasing with respect to x,
   - $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$,
   - $F_X(x)$ is right-continuous, i.e., $F_X(x) = \lim_{\delta \to 0^+} F_X(x + \delta)$.

5. **Definition:** Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $Z = \begin{bmatrix} X \\ Y \end{bmatrix} \in \mathbb{R}^n$ is a vector random variable, where $X \in \mathbb{R}^{n_1}$, $Y \in \mathbb{R}^{n_2}$ and $n = n_1 + n_2$. Suppose the CDF of Z is $F_Z(z)$ and the PDF of Z is $f_Z(z)$. The marginal PDF of X is

$$f_X(x) = \int_{\mathbb{R}^{n_2}} f_Z(z)\mathrm{d}y = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{XY}(x, y)\mathrm{d}y_1 \cdots \mathrm{d}y_{n_2}.$$

The conditional PDF of X given $Y = a$ (a is a given constant)

$$f_X(x|Y = a) = \left.\frac{f_Z(z)}{f_Y(y)}\right|_{y=a} = \left.\frac{f_{XY}(x, y)}{f_Y(y)}\right|_{y=a}.$$

6. **Definition:** Let $X : \Omega \to \mathbb{R}^n$ be a random variable with PDF $f_X(x)$. Then for some mapping $\psi : \mathbb{R}^n \to \mathbb{R}^m$, the quantity

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}^n} \psi(X) f_X(x)\mathrm{d}(x)$$

is called the expectation or expected value of $\psi(X)$. More specifically,

- If $\psi(X) = X$, then $\mu = \mathbb{E}[X]$ is called the <u>mean</u> value of X.
- If $\psi(X) = XX^\top$, then $\mathbb{E}[XX^\top]$ is called the <u>second moment</u> of X.
- If $\psi(X) = (X-\mathbb{E}[X])(X-\mathbb{E}[X])^\top$, then $\Sigma = \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^\top]$ is called the <u>covariance matrix</u> of X, sometimes also called second central moment.

7. **Properties:** Let $X : \Omega \to \mathbb{R}^n$ be a random variable with mean μ and covariance matrix Σ. Then

   - The second moment is $\mathbb{E}[XX^\top] = \Sigma + \mu\mu^\top$.

     Special case: When $n = 1$, i.e., X is a scalar random variable, $\Sigma = \sigma^2$ is the variance. The second moment $\mathbb{E}[X^2] = \sigma^2 + \mu^2$.
   - Given another random variable $Y = AX + B$ where $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^m$, then
     (a) The mean of Y is $\mathbb{E}[Y] = \mathbb{E}[AX + B] = A\,\mathbb{E}[X] + B$.
     (b) The covariance of Y is $\Sigma_Y = A\Sigma A^\top$.

8. normal distribution

   - Scalar case: $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{(x-\mu)^2}{2\sigma^2} \Big\}$$

   - Vector case: $X \sim N(\mu, \Sigma)$

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\Big\{ -\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu) \Big\}$$

Theorem: Given a random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, where $X_1 \in \mathbb{R}^{n_1}$ and $X_2 \in \mathbb{R}^{n_2}$, satisfying normal distribution with $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$, where $\mu_1 \in \mathbb{R}^{n_1}$, $\mu_2 \in \mathbb{R}^{n_2}$, $\Sigma_{11} \in \mathbb{R}^{n_1\times n_1}$, $\Sigma_{12} \in \mathbb{R}^{n_1\times n_2}$, $\Sigma_{21} \in \mathbb{R}^{n_2\times n_1}$, $\Sigma_{22} \in \mathbb{R}^{n_2\times n_2}$. Then

(a) The marginal PDF of X_1 is also normal distribution and $X_1 \sim N(\mu_1,\ \Sigma_{11})$.
(b) The marginal PDF of X_2 is also normal distribution and $X_2 \sim N(\mu_2,\ \Sigma_{22})$.
(c) The Conditional PDF of $X_1|X_2 = x_2$ is also normal distribution and $(X_1|X_2 = x_2) \sim N(\mu_{1|2},\ \Sigma_{1|2})$, where

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

Ex: Given a 2-dimensional random variable $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and its PDF

$$f_X(x) = \begin{cases} c, & ||x||_1 \le 1, \\ 0, & ||x||_1 > 1. \end{cases}$$

- Find the value of c.

  Remember that $||x||_1 \le 1$ is equivalent to $|x_1| + |x_2| \le 1$.
  If $f_X(x)$ is the PDF of X, then

  $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_X(x) \mathrm{d}x_1 \mathrm{d}x_2 = 1.$$

  Thus, we can get $c = 1/2$.

- Find the marginal PDF $f_{X_2}(x_2)$.

  Based on the definition, we have

  $$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2) \mathrm{d}x_1$$

  There are two cases:

  Case 1: If $|x_2| > 1$, then $||x||_1 = |x_1| + |x_2| > 1$, $f_{X_1X_2}(x_1, x_2) = f_X(x) = 0$, thus $f_{X_2}(x_2) = 0$.
  Case 2:If $|x_2| \le 1$, then

  $$f_{X_1X_2}(x_1, x_2) = \begin{cases} \frac{1}{2} & \text{if } |x_1| + |x_2| \le 1, \\ 0, & \text{if } |x_1| + |x_2| > 1 \end{cases}$$

  So,

  $$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1X_2}(x_1, x_2) \mathrm{d}x_1 = \int_{-(1-|x_2|)}^{1-|x_2|} \frac{1}{2} \mathrm{d}x_1 = 1 - |x_2|.$$

  Thus,

  $$f_{X_2}(x_2) = \begin{cases} 1 - |x_2|, & \text{if } |x_2| \le 1, \\ 0, & \text{if } |x_2| > 1. \end{cases}$$

- Find the conditional PDF $f_{X_1}(x_1|X_2 = a)$ where a is a constant and $|a| < 1$.

  Based on definition of conditional probability,

  $$f_{X_1}(x_1|X_2 = a) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}\Big|_{x_2=a}$$

  We already know that

  $$f_{X_2}(a) = 1 - |a|$$

  $$f_{X_1X_2}(x_1, a) = \begin{cases} \frac{1}{2} & \text{if } |x_1| + |a| \le 1, \\ 0, & \text{if } |x_1| + |a| > 1 \end{cases}$$

  Thus,

  $$f_{X_1}(x_1|X_2 = a) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}\Big|_{x_2=a} = \begin{cases} \frac{1}{2(1-|a|)} & \text{if } |x_1| \le 1 - |a|, \\ 0, & \text{if } |x_1| > 1 - |a| \end{cases}$$