Control of Stochastic Systems
Lecture 1
Introduction Course and Probability

Jan H. van Schuppen

13 February 2025
Delft University of Technology,
Delft, The Netherlands

Outline

Outline

Introduction to course

Course

- Control of discrete-time stochastic systems (TUD WI4221).
- Lecturer Jan H. van Schuppen.
- http://diamhomes.ewi.tudelft.nl/ jhvanschuppen/courseguide2025.html
- Evaluation of course:
  Weekly homework sets (50%) and oral exam (50%).
- Major topics of the course:
  - Probability and stochastic processes.
  - Stochastic systems.
  - Stochastic realization.
  - Control with complete observations.
  - Filtering.
  - Control with partial observations.

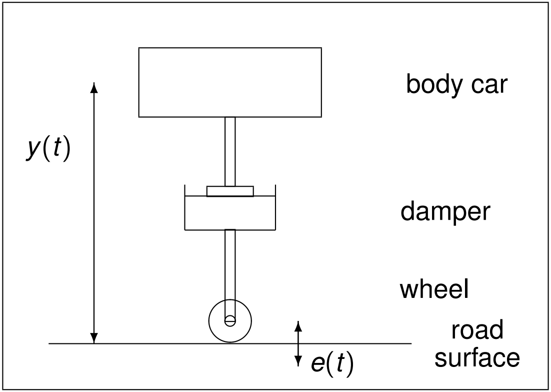
Example. Control of a shock absorber (1)

Problem. Control of a shock absorber
Control the variable damping
of a shock absorber of a motor vehicle
such that the ride is comfortable to the passengers.

Project was carried out by Fabien Campillo (INRIA Sophia Antipolis)
with his research advisor Etienne Pardoux.

Example. Control of a shock absorber (2)

Engineering model of body of car, a damper, and a wheel



Example. Control of a shock absorber (3)

Control system of damped mass

The system is excited by fluctuations of the road surface.
The continuous-time system for this example,

$$m\frac{d^2y(t)}{dt^2} + u(t)\,\frac{dy(t)}{dt} + ky(t) + F\mathrm{sign}\left(\frac{dy(t)}{dt}\right)$$

$$= -m\frac{d^2v(t)}{dt^2},$$

$$\mathrm{sign}(x) = \begin{cases} +1 & \text{if } 0 < x, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Process v related to Brownian motion noise.

Example. Control of a shock absorber (4)

Transformation to state space system representation

$$
\begin{aligned}
x_1(t) &= y(t),\ x_2(t) = dy(t)/dt, \\
dx(t) &= \begin{bmatrix} x_2(t) \\ \frac{-1}{m}[u(t)x_2(t) + kx_1(t) + F\text{sign}(x_2(t))] \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dw(t) \\
&= f(x(t), u(t))dt + Mdw(t),\ x(0) = x_0, \\
f_1(x,u) &= x_2, \\
f_2(x,u) &= \frac{-1}{m}[u\ x_2 + k\ x_1 + F\ \text{sign}(x_2)], \\
f(x,u) &= \begin{bmatrix} f_1(x,u) \\ f_2(x,u) \end{bmatrix}.
\end{aligned}
$$

Example. Control of a shock absorber (5)

Control Synthesis

Control objective. Minimize the acceleration of the body mass of the car.

$$f_2(x,u) = \frac{d^2 y(t)}{dt^2} + \frac{d^2 v(t)}{dt^2}.$$

Define a control law $g \in G$ and the closed-loop system by the formulas,

$$\begin{aligned} g : X = \mathbb{R}^{n_x} &\to U = \mathbb{R}^1, \\ dx^g(t) &= f(x^g(t), g(x^g(t)))dt + Mdw(t),\ x^g(0) = x_0, \\ J_{ac}(g) &= \limsup_{t\to\infty} \frac{1}{t} E\left[\int_0^t f_2(x^g(s), g(x^g(s)))ds\right]. \end{aligned}$$

Problem of optimal control. Solve,

$$\inf_{g\in G} J_{ac}(g); \quad \text{determine } J^*_{ac} \in \mathbb{R},\ g^* \in G \text{ such that,}$$

$$J^*_{ac} = \inf_{g\in G} J_{ac}(g) = J_{ac}(g^*).$$

Example. Control of a shock absorber (6)

Control Design

Numerical approximation of optimal control law.

Cost	Control law
2.93	constant control
2.68	g_1
2.37	g_2
2.22	g_a^* optimal control law, numerically approximated

$$g_1(x) = [-k\ x_1 - F\ \mathrm{sign}(x_2)]/x_2 \quad \Rightarrow \quad f_2(x, g_1(x)) = 0;$$
$$g_2(x) = [a + b\ x_1\ \mathrm{sign}(x_2)]^+, \quad a,\ b \in \mathbb{R}.$$

One optimizes the average cost over the parameters, $(a,\ b)$, which produces the cost of g_2 listed in the table above.
This approach was satisfactory to control designers.

Introduction to course

Presentation of course contents

- Audience of course consists of engineers and of mathematicians.
- For mathematics students:
  Presentation of probability in a mathematical formulation.
- For engineering students:
  Emphasis on concepts and theorems, and engineering understanding.
- The course aims at both groups of students.
- Chapter 2 of the book on probability, is like an encyclopedia,
  read what you need or are interested in.

Lecture 1 learning goals

- Gaussian random variables.
- Conditional expectation
  of Gaussian and of finite-valued random variables.

Outline

Probability distribution functions

Notation of sets

$$\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}, \quad \text{integers},$$
$$\mathbb{Z}_+ = \{1, 2, \ldots\}, \quad \text{strictly-positive integers},$$
$$\mathbb{N} = \{0, 1, 2, \ldots\}, \quad \text{natural numbers},$$
$$\mathbb{Z}_m = \{1,\ 2, \ldots,\ m\}, \quad \text{first } m \text{ integers, } m \in \mathbb{Z}_+,$$
$$\mathbb{N}_m = \{0,\ 1,\ 2, \ldots,\ m\}, \quad \text{first } m \text{ natural numbers},$$
$$\mathbb{R} \quad \text{real numbers},$$
$$\mathbb{R}_+ = [0, \infty) \subset \mathbb{R}, \quad \text{positive real numbers},$$
$$\mathbb{R}_{s+} = (0, \infty) \subset \mathbb{R}, \quad \text{strictly-positive real numbers}.$$

Probability distribution functions

Def. Probability distribution function on $\mathbb{R}$

Define a probability distribution function (pdf) as a function such that,

$$
\begin{aligned}
& f : \mathbb{R} \to \mathbb{R}_+, \\
(1)\ \ & \text{increasing } u \le v \ \Rightarrow\ f(u) \le f(v); \\
(2)\ \ & \text{limits } \lim_{u \to -\infty} f(u) = 0,\ \lim_{u \to +\infty} f(u) = 1; \\
(3)\ \ & \text{right continuous, } \lim_{v \downarrow u} f(v) = f(u).
\end{aligned}
$$

$$f(u^-) = \lim_{v \uparrow u,\ v<u} f(v) \le f(u) = f(u^+) := \lim_{v \downarrow u,\ v>u} f(v).$$

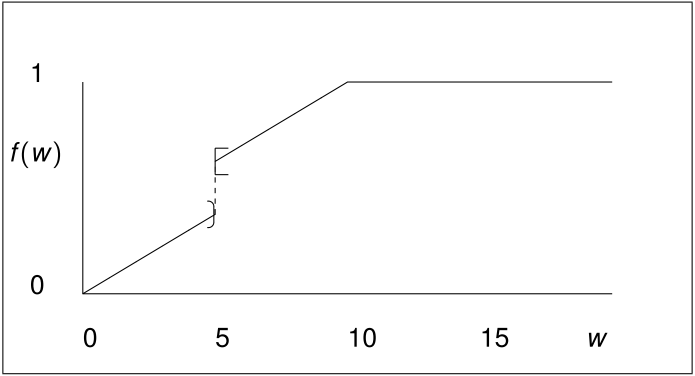
See figure next slide.

Pdfs on $\mathbb{R}^n$ defined differently. See book (Ash,1972).
Class of probability distribution is a convex set:

$$\forall\ f_1,\ f_2 \text{ pdf and } \forall\ c \in [0,\ 1] \ \Rightarrow\ c f_1 + (1-c) f_2 \text{ is a pdf.}$$

Probability distribution functions

Fig. Probability distribution function on $\mathbb{R}$



Probability distribution functions

Definition. Subclasses of pdfs

(a) Discrete pdf

$$f(u) = \sum_{k \in \mathbb{Z}} p(k) I_{[u_k,\infty)}(u), \quad I_{[u_k,\infty)}(u) = \begin{cases} 0, & -\infty < u < u_k, \\ 1, & u_k \le u < \infty. \end{cases}$$

frequency function $p : \mathbb{Z} \to \mathbb{R}_+, \ \sum_{k \in \mathbb{Z}} p(k) = 1,$

$\{u_k \in \mathbb{R}, \ k \in I \subseteq \mathbb{Z}\}$

strictly increasing: $\forall \ k \in I \backslash \{1\}, \ u_{k-1} < u_k.$

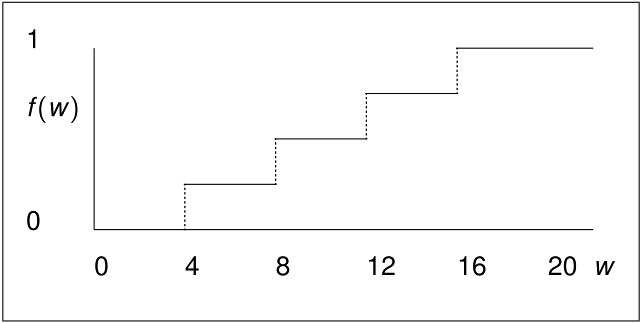
(b) Absolute continuous pdf

$$f(u) = \int_{-\infty}^{u} p(v) dv, \quad p : \mathbb{R} \to \mathbb{R}_+, \ \int_{-\infty}^{+\infty} p(v) dv = 1,$$

p probability density function.

Probability distribution functions

Fig. Discrete probability distribution function on $\mathbb{R}$



Probability distribution functions

- Poisson pdf on $\mathbb{N} = \{0, 1, \ldots, \}$ with rate parameter $\lambda \in \mathbb{R}_+$,

$$p(k) = \lambda^k \exp(-\lambda)/k!, \;\; \forall\, k \in \mathbb{N}.$$

- Gamma pdf on $\mathbb{R}_+$ with parameters $\lambda, r \in \mathbb{R}_+$,

$$p(v) = \lambda^{-r} v^{r-1} \exp(-\lambda^{-1} v)/\Gamma(r), \;\; \Gamma(r) = \int_0^\infty v^{r-1} \exp(-v) dv.$$

- Gaussian pdf on $\mathbb{R}$ with parameters $(m,\; q) \in \mathbb{R} \times (0, \infty)$,

$$p(v) = \exp(-(v - m)^2/2q)\; (2\pi q)^{-1/2}.$$

- Gaussian pdf on $\mathbb{R}^n$ with $(m,\; Q) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}_{spds}$, $0 \prec Q$,

$$p(v_1, \ldots, v_n) = \exp(-\frac{1}{2}(v - m)^T Q^{-1}(v - m))\; [(2\pi)^n \det(Q)]^{-1/2}.$$

Outline

Motivation of measure theoretic probability

Unsuccessful definition

Initial attempt to define a probability measure:

$$f : \text{Pwrset}(\Omega) \to [0,1], \;\; \text{Pwrset}(\Omega) = \{A \subseteq \Omega\}.$$

satisfying properties not listed here.
It was proven that such an object cannot exist!

Successfull definition

Different approach, restrict in definition of P attention
from all subsets of Ω
to a strict subset $F \subsetneq \text{Pwrset}(\Omega)$.
For the subset F, one uses the concept of a σ-algebra.
This is mathematically correct and exists!
Note that σ-algebras are not needed if Ω is a fintie set.

σ-Algebras

Definition. σ-Algebra of subsets of Ω

$F \subseteq \text{Pwrset}(\Omega)$ such that

(1) $\Omega \in F$;

(2) $A \in F \;\Rightarrow\; A^c \in F$;

(3) $\{A_k \in F,\; k \in \mathbb{Z}_+\} \;\Rightarrow\; \cup_{k\in\mathbb{Z}_+} A_k \in F$.

(Ω, F) called a measurable space.

$G \subseteq F$ called a sub-σ-algebra of F

if (1) G is a σ-algebra and (2) $G \subseteq F$.

Definition. Family of subsets

Consider a set Ω. A family of subsets $\{A_k \subseteq \Omega, k \in \mathbb{Z}_+\}$ is called:

disjoint if $\forall k,\; m \in \mathbb{Z}_+,\; k \neq m \;\Rightarrow\; A_k \cap A_m = \emptyset$;

partition if (1) disjoint and (2) $\cup_{k\in\mathbb{Z}_+} A_k = \Omega$.

σ-Algebras

Examples of σ-Algebras

(1) $F_0 = \{\emptyset, \Omega\}$ called the trivial σ-algebra.

(2) $\{\emptyset, A, A^c, \Omega\}, \ \forall A \subseteq \Omega$.

Proposition

Consider Ω and a family $\{A_i \subseteq \Omega, \ i \in I\}$.

there exists a smallest σ-algebra $F(\{A_i, \ i \in I\})$
such that $\forall i \in I, \ A_i \in F(\{A_i, \ i \in I\})$.

Call $F(\{A_i, \ i \in I\})$
the σ-algebra generated by the collection $\{A_i \subseteq \Omega, \ i \in I\}$.

Note that the index set I need not be countable!

σ-Algebras

Examples of σ-algebras

$$
\begin{aligned}
F(A) &= \{\emptyset, A, A^c, \Omega\}, \quad \forall A \subseteq \Omega. \\
B(\mathbb{R}) &= F(G), \text{ where,} \\
G &= \{(a, b) \subset \mathbb{R} | a < b\}, \text{ set of open intervals of } \mathbb{R}, \\
B(\mathbb{R}) &\quad \text{called the Borel } \sigma\text{-algebra of } \mathbb{R}, \\
(\mathbb{R},\ B(\mathbb{R})) &\quad \text{called a Borel space.} \\
B(\mathbb{R}^n) &\quad \text{defined similarly, } n \in \mathbb{Z}_+, \\
(\mathbb{R}^n,\ B(\mathbb{R}^n)) &\quad \text{called a Borel space.} \\
F_1 \vee F_2 &= \sigma(\{F_1, F_2\}), \text{ notation.}
\end{aligned}
$$

Emile Borel (1871–1956; mathematician born in France).

Outline

Probability measures

Def. Probability measure

Consider measurable space (Ω, F).

$$
\begin{aligned}
& P : F \to \mathbb{R}_+ \text{ called a measure if it is } \sigma\text{-additive:} \\
& \forall \, \{A_k \in F, k \in \mathbb{Z}_+\} \text{ disjoint} \\
\Rightarrow \quad & P(\cup_{k \in \mathbb{Z}_+} A_k) = \sum_{k \in \mathbb{Z}_+} P(A_k); \\
& P : F \to \mathbb{R}_+ \text{ called a probability measure} \\
& \text{if (1) it is a measure and (2) } P(\Omega) = 1.
\end{aligned}
$$

Call (Ω, F, P) a probability space if:
(1) (Ω, F) is a measurable space and
(2) P is a probability measure on (Ω, F).

Probability measures

Theorem. Probability measure on the real numbers

There exists a probability measure $P : B(\mathbb{R}) \to [0, 1]$ on $(\mathbb{R},\ B(\mathbb{R}))$
if and only if
there exists a pdf $f : \mathbb{R} \to \mathbb{R}_+$ such that

$$P((a, b]) = f(b) - f(a),\ \forall a, b \in \mathbb{R},\ a < b.$$

Theorem. Properties of probability measure

Properties of a probability measure (Ω, F, P).

(a) $P(\emptyset) = 0$.

(b) Monotonicity. $A_1 \subseteq A_2$ implies that $P(A_1) \leq P(A_2)$.

(c) Subadditivity. $\{A_k \in F, k \in \mathbb{Z}_+\}$, not necessarily disjoint, implies that $P(\cup_{k\in\mathbb{Z}_+} A_k) \leq \sum_{k\in\mathbb{Z}_+} P(A_k)$.

(d) $0 \leq P(A) \leq 1$ for all $A \in F$.

Probability measures

Definition. Independent σ-algebras

$$
\begin{array}{rl}
(\Omega, F, P) & m \in \mathbb{Z}_+,\ \mathbb{Z}_m = \{1,\ 2,\ \ldots,\ m\}, \\
\{F_k \subseteq F, k \in \mathbb{Z}_m\} & \text{finite collection of sub-}\sigma\text{-algebras,} \\
 & \text{is called independent with respect to } P \text{ if} \\
 & \forall \; \{A_k \in F_k, k \in \mathbb{Z}_m\},
\end{array}
$$

$$
P(\cap_{k \in \mathbb{Z}_m} A_k) = \prod_{k \in \mathbb{Z}_m} P(A_k).
$$

Any infinite family $\{F_i,\ i \in I\}$ is defined to be independent
if every finite subfamily is independent.

Remark. Definition of independence

In the literature,
independence is defined for a probability distribution function,
and then generalized.

Outline

Random variables

Definition. Random variable

Consider a probability space (Ω, F) and a measurable space (X, G).
Define a random variable (rv) as a function,

$$x : \Omega \to X \text{ such that,}$$
$$\text{if } \forall\, A \in G,\ x^{-1}(A) = \{\omega \in \Omega | x(\omega) \in A\} \in F.$$

Motivation, condition necessary to define $P(\{\omega \in \Omega | x(\omega) \in A\})$.

Proposition. Real-valued random variables

Assume that $(X, G) = (\mathbb{R}, B(\mathbb{R}))$.
The function $x : \Omega \to \mathbb{R}$ is a random variable if and only if,

$$\begin{aligned} x^{-1}((-\infty, w]) &= \{\omega \in \Omega | x(\omega) \in (-\infty, w]\} \\ &= \{\omega \in \Omega | x(\omega) \leq w\} \in F,\ \ \forall\, w \in \mathbb{R}, \end{aligned}$$

Random variables

Def. Indicator random variable

Indicator function $I_A : \Omega \to \mathbb{R}$ of a subset $A \subset \Omega$

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \Leftrightarrow \omega \in A^c. \end{cases}$$

Indicator I_A is a random variable if and only if $A \in F$.

Def. Simple random variable

A simple random variable $x : \Omega \to \mathbb{R}$ is defined
as a finite sum of products of a real number with an indicator function,

$$x(\omega) = \sum_{k=1}^{n} c_k I_{A_k}(\omega), \ \ \{c_k \in \mathbb{R}, k \in \mathbb{Z}_n\}, \ \{A_k \in F, k \in \mathbb{Z}_n\}, \ n \in \mathbb{Z}_+.$$

There exists a representation of any simple random variable such that $\{A_k \in F, k \in \mathbb{Z}_n\}$ is a partition of Ω and $\{c_k \in \mathbb{R}, k \in \mathbb{Z}_n\}$ are distinct.

Random variables – Modeling

Def. Binary-valued random variable

Used in information theory and communication theory.

$$\begin{aligned} x : \Omega &\to \{0,\ 1\}, \\ P(\{\omega \in \Omega |\ x(\omega) = 1\}) &= P(\{x = 1\}) = q \in [0,\ 1], \\ P(\{\omega \in \Omega |\ x(\omega) = 0\}) &= 1 - q \in [0,\ 1], \\ P(\{\omega \in \Omega |\ x(\omega) = 0 \text{ or } 1\}) &= (1 - q) + q = 1; \\ &\text{special cases} \\ (1 - q_a,\ q_a) &= (0.5,\ 0.5), \\ (1 - q_b,\ q_b) &= (0.4,\ 0.6). \end{aligned}$$

Random variables – Modeling

Def. Random variable for outcome of throw of a die

$$
\begin{aligned}
x : \Omega &\to \{1,\ 2,\ 3,\ 4,\ 5,\ 6\} = \mathbb{Z}_6, \\
1/6 &= P(\{\omega \in \Omega |\ x(\omega) = i\}) \in [0,\ 1],\ \forall\ i \in \mathbb{Z}_6, \\
1 &= \sum_{i=1}^{6} P(\{\omega \in \Omega |\ x(\omega) = i\}), \\
1/3 &= P(\{x = 2 \text{ or } 4\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.
\end{aligned}
$$

This is an ideal model of a die.
In practice probability of any particular outcome $i \in \mathbb{Z}_6$ not exactly $1/6$.

Random variables – Modeling

Def. Random variable for life time of a lamp

Assume life time has
an exponential probability distribution with parameter $h \in (0, \infty)$.

$$\begin{aligned} x &: \Omega \to \mathbb{R}_+, \\ & P(\{\omega \in \Omega | \; x(\omega) \leq 864 \text{ hours}\}) \\ &= \int_0^{864} f_x(dw) = \int_0^{864} p_x(w) \, dw \\ p_x(w) &= \exp(-w/h)/h, \quad p_x : \mathbb{R}_+ \to \mathbb{R}_+, \\ 1 &= \int_0^{\infty} \exp(-w/h) \, dw/h. \end{aligned}$$

p_x called the exponential probability density function.
This pdf is a special case of the Gamma probability density function.

Random variables – Modeling

Def. Scalar Gaussian Random variable

Could model the current of an electric circuit.
Gaussian random variable with parameters $m \in \mathbb{R}$ and $q \in (0, \infty)$.

$$\begin{aligned}
& x : \Omega \to \mathbb{R}, \\
& P(\{\omega \in \Omega |\ x(\omega) \leq -2\}) \\
= & \int_{-\infty}^{-2} f_x(dw) = \int_{-\infty}^{-2} p_x(w)\, dw \\
p_x(w) = & (2\pi q)^{-1/2} \exp(-(w-m)^2/(2q)),\ p_x : \mathbb{R} \to \mathbb{R}_+; \\
1 = & \int_{-\infty}^{+\infty} p_x(w)\, dw.
\end{aligned}$$

Random variables

Proposition. Random variables from others

If $x, y : \Omega \to \mathbb{R}$ are random variables
then so are $x + y$, $x - y$, $x \times y$, and

$$\begin{aligned} x \wedge y(\omega) &= \min\{x(\omega),\ y(\omega)\}, \\ x \vee y(\omega) &= \max\{x(\omega),\ y(\omega)\}, \\ x^+ &= \max\{x, 0\}, \quad x^- = -\min\{x, 0\} \geq 0, \quad \text{consequently,} \\ x &= x^+ - x^-; \\ xy^{-1} & \quad \text{if real number specified on the subset } \{\omega \in \Omega | y(\omega) = 0\}. \end{aligned}$$

Definition. Equality almost surely

The random variables $x, y : \Omega \to \mathbb{R}$ are said to be
equal almost surely, notation $x = y$ a.s., if

$$P(\{\omega \in \Omega | x(\omega) = y(\omega)\}) = 1.$$

Random variables and σ-algebras

Definition. σ-algebra generated by a random variable

Consider (Ω, F), (X, G), and $x : \Omega \to X$.
Define the sets,

$$x^{-1}(A) = \{\omega \in \Omega |\ x(\omega) \in A\},$$
$$x^{-1}(G) = \{x^{-1}(A) \in F |\ \forall\ A \in G\},$$

note abuse of notation!

Then $x^{-1}(G)$ is a σ-algebra.
Notation $F^x = F(x) = x^{-1}(G)$.

Random variables

Definition. Probability measure induced by a random variable

The random variable $x : \Omega \to \mathbb{R}$
induces a probability measure P_x on the range space according to

$$
\begin{aligned}
&P_x : B(\mathbb{R}) \to [0,1], \\
P_x(A) = {} & P(x^{-1}(A)) = P(\{\omega \in \Omega | x(\omega) \in A\}); \\
&(\Omega, F, P),\ (\mathbb{R}, B(\mathbb{R}), P_x),\ \ \text{note the two probability spaces;} \\
&(\Omega, F, P) \mapsto^{x}\ (\mathbb{R}, B(\mathbb{R}), P_x),\ \text{note the transformation;} \\
f_x(w) = {} & P_x((-\infty, w]),
\end{aligned}
$$

for any probability distribution function f_x.

Random variables

Definition. Measurability of a random variable with respect to a σ-algebra

Consider (Ω, F), (X, G), $x : \Omega \to X$.
Call x measurable with respect to the sub-σ-algebra $H \subseteq F$ if,

$$F^x = x^{-1}(G) \subseteq H \quad (\Leftrightarrow \ \forall\, A \in G,\ x^{-1}(A) \in H).$$

Definition. Borel measurable function

Call a function $h : \mathbb{R}^m \to \mathbb{R}^n$ a Borel measurable function if

$$h^{-1}(A) = \{x \in \mathbb{R}^m |\ h(x) \in A\} \in B(\mathbb{R}^m),\ \ \forall\, A \in B(\mathbb{R}^n).$$

Proposition. Measurability related to a function.

Consider $x, y : \Omega \to \mathbb{R}$.
If y is measurable with respect to F^x
then there exists a Borel measurable function $h : \mathbb{R} \to \mathbb{R}$
such that $y = h(x)$. Interpretation!

Random variables

Exposition on expectation may be found in other books.

Definition. Characteristic function
of a real valued random variable $x : \Omega \to \mathbb{R}^n$
having a probability density function $p_x : \mathbb{R}^n \to \mathbb{R}_+$
is defined as,

$$E[\exp(iw^T x)] = \int_{\mathbb{R}^n} \exp(iw^T v) p_x(v) dv, \ \forall\, w \in \mathbb{R}^n.$$

A characteristic function is a Fourier transform
of the associated probability density function. Note that

$$1 = |\exp(iw^T v)| \ \Rightarrow$$

$$E|\exp(iw^T x)| = \int_{\mathbb{R}^n} |\exp(iw^T v)| p_x(v) dv = \int p_x(v) dv = 1.$$

Outline

Gaussian random variables

Definition. Gaussian random variable
A Gaussian random variable with parameters m_x, Q_x is defined to be

$$x : \Omega \to \mathbb{R}^{n_x},\ (m_x,\ Q_x) \in (\mathbb{R}^{n_x} \times \mathbb{R}^{n_x \times n_x}_{pds}),\ \text{if}$$

$$E[\exp(iw^T x)] = \exp(iw^T m_x - \frac{1}{2} w^T Q_x w),\ \ \forall\ w \in \mathbb{R}^{n_x};$$

$$x \in G(m_x, Q_x),\ \text{notation;}$$

$$(x_1, \ldots, x_n) \text{ called jointly Gaussian if}$$

$$x = (x_1, x_2, \ldots, x_n)^T \in G(m_x,\ Q_x).$$

If $0 \prec Q_x$
then a Gaussian random variable
has a Gaussian probability density function on $\mathbb{R}^{n_x}$ and conversely.
If $Q_x = 0$
then $x = m_x \in \mathbb{R}^{n_x}$, hence any constant is a Gaussian random variable.

Gaussian random variables

Proposition

$$x : \Omega \to \mathbb{R}^{n_x},\ x \in G(m_x, Q_x),\ A \in \mathbb{R}^{n_y \times n_x},\ b \in \mathbb{R}^{n_y},\ \Rightarrow$$
$$y = Ax + b \in G(Am_x + b, AQ_xA^T),\ y : \Omega \to \mathbb{R}^{n_y}.$$

In words, any affine function y of a Gaussian random variable x

$$y = Ax + b,$$

is a Gaussian random variable.

Gaussian random variables

Proposition. Decomposition
Consider a Gaussian random variable

$$x : \Omega \to \mathbb{R}^{n_x},\ n_x \in \mathbb{Z}_+,\ x \in G(0, Q_x).$$

Then,

$$\exists\ n_v \in \mathbb{N}_n,\ \exists\ v : \Omega \to \mathbb{R}^{n_v},\ v \in G(0, I_{n_v}),$$
$$\exists\ M \in \mathbb{R}^{n_x \times n_v},\ \text{such that } Q_x = MM^T,\ \text{and}$$
$$x = M\, v \text{ a.s.}$$

Thus, there exists v and M
such that the above representation holds.

Gaussian random variables

Def. Tuple of Gaussian random variables

$$\begin{bmatrix} x \\ y \end{bmatrix} = (x,\ y) \in G(m_{(x,\ y)},\ Q_{(x,y)}),$$

$$x : \Omega \to \mathbb{R}^{n_x},\ y : \Omega \to \mathbb{R}^{n_y},$$

$$m_{(x,\ y)} = \begin{bmatrix} m_x \\ m_y \end{bmatrix} \in \mathbb{R}^{n_x+n_y},$$

$$Q_{(x,\ y)} = \begin{bmatrix} Q_x & Q_{x,y} \\ Q_{x,y}^T & Q_y \end{bmatrix} \in \mathbb{R}_{pds}^{(n_x+n_y)\times(n_x+n_y)}.$$

Call $m_{(x,\ y)}$ mean value of $(x,\ y)$,
Q_x the variance matrix of x, Q_y the variance matrix of y,
$Q_{x,y}$ the covariance matrix of x and y,
$Q_{(x,y)}$ the variance matrix of $(x,\ y)$.
Condition on matrix $Q_{(x,y)}$ being positive-definite symmetric,
is necessary for a proper definition.

Gaussian random variables

Proposition. Equivalent condition for independence

$$
\begin{aligned}
& x : \Omega \to \mathbb{R}^{n_x},\; y : \Omega \to \mathbb{R}^{n_y}, \text{ random variables}, \\
& (x, y) \in G(m_{(x,y)}, Q_{(x,y)}), \\
Q_{(x,\,y)} = & \begin{bmatrix} Q_x & Q_{x,y} \\ Q_{x,y}^T & Q_y \end{bmatrix}.
\end{aligned}
$$

The rvs x and y are independent Gaussian random variables
if and only if $F^x,\; F^y$ are independent σ-algebras
if and only if $Q_{x,y} = 0$.

Proposition

Consider the independent Gaussian random variables,

$$x : \Omega \to \mathbb{R}^{n_x},\; y : \Omega \to \mathbb{R}^{n_y}.$$

Then $z = (x, y)^T \in G$ is a Gaussian random variable.

Gaussian random variables

Example. Gaussian signal and noise representation

Consider the random variables

$$
\begin{aligned}
& y = Cx + w, \\
& \qquad n_y,\ n_x \in \mathbb{Z}_+,\ C \in \mathbb{R}^{n_y \times n_x}, \\
& \qquad w : \Omega \to \mathbb{R}^{n_y},\ x : \Omega \to \mathbb{R}^{n_x},\ y : \Omega \to \mathbb{R}^{n_y}, \\
& \qquad w \in G(0, Q_w),\ x \in G(0, Q_x),\ F^x,\ F^w \text{ independent.}
\end{aligned}
$$

Call y the observation, x the signal, and w the noise.

Then $(y,\ x,\ w)$ are jointly Gaussian random variables and y is a Gaussian random variable.
Note that $(x,\ w)$ are jointly Gaussian random variables by the previous proposition.

Gaussian Random Variables

Def. Triple of Gaussian random variables

$$\begin{aligned} &(x,\ y,\ z) \in G(0,\ Q_{(x,\ y,\ z)}), \\ &x : \Omega \rightarrow \mathbb{R}^{n_x},\ y : \Omega \rightarrow \mathbb{R}^{n_y},\ z : \Omega \rightarrow \mathbb{R}^{n_z}, \end{aligned}$$

$$Q_{(x,\ y,\ z)} = \begin{bmatrix} Q_x & Q_{x,y} & Q_{x,z} \\ Q_{x,y}^T & Q_y & Q_{y,z} \\ Q_{x,z}^T & Q_{y,z}^T & Q_z \end{bmatrix} \in \mathbb{R}_{pds}^{(n_x+n_y+n_z)\times(n_x+n_y+n_z)}.$$

$Q_{x,y}$ called covariance matrix of x and y,
$Q_{x,z}$ called covariance matrix of x and z, etc.
Condition on matrix $Q_{(x,\ y,\ z)}$ being positive-definite symmetric,
is necessary for the pdf to be well defined.

Finite-Valued Random Variables

Def. Representation of a finite-valued random variable

Consider a finite-valued random variable,

$$\begin{aligned}
& x : \Omega \to X = \{a_1,\ a_2,\ a_3,\ \ldots,\ a_{n_{i_x}}\} \subset \mathbb{R}^{n_x},\ n_x \in \mathbb{Z}_+;\ \text{define,} \\
i_{x,j} &= I_{\{\omega \in \Omega |\ x(\omega) = a_j\}} \in \{0,\ 1\},\ \ i_x : \Omega \to \mathbb{R}^{n_{i_x}},\ n_{i_x} \in \mathbb{Z}_+, \\
i_x &= \left[i_{x,1},\ i_{x,2},\ \ldots,\ i_{x,n_{i_x}}\right]^T \in \{0,\ 1\}^{n_{i_x}} \subset \mathbb{R}^{n_{i_x}}, \\
C_x &= \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_{n_{i_x}} \end{bmatrix} \in \mathbb{R}^{n_x \times n_{i_x}};\ \text{notation,} \\
m_{i_x} &= p_x = E[i_x] \in [0,\ 1]^{n_{i_x}} \subset \mathbb{R}^{n_{i_x}}, \\
Q_{i_x} &= E[(i_x - m_{i_x})(i_x - m_{i_x})^T] \in \mathbb{R}^{n_{i_x} \times n_{i_x}}; \\
& \text{assume that, } \forall\ i,\ j \in \mathbb{Z}_{n_{i_x}},\ a_i \neq a_j.
\end{aligned}$$

Call i_x the vector indicator representation of the random variable x and n_{i_x} the atom multiplicity of x.

Remark

Notation of i_x slightly different from that of lecture notes.

Finite-Valued Random Variables

Proposition. Elementary properties of a finite-valued random variable

(a) $1_{n_{i_x}}^T \; i_x = 1$.

(b) Note the representation of x as a simple random variable,

$$x = C_x \; i_x = \sum_{j=1}^{n_{i_x}} a_j \; i_{x,j}.$$

(c) $Q_{i_x} = \mathrm{Diag}(p_x) - m_{i_x} m_{i_x}^T \in \mathbb{R}_{pds}^{n_{i_x} \times n_{i_x}}$.

(d) $F^x = F^{i_x}$. Assumption used that $\{a_j \in \mathbb{R}^{n_x}, \; \forall \, j \in \mathbb{Z}_{n_{i_x}}\}$ values are pairwise different.

Outline

Conditional expectation

Definition. Conditional expectation of a random variable given a σ-algebra

(a) Consider a positive random variable,

$$\begin{aligned} &(\Omega, F),\ G \subseteq F \text{ sub-}\sigma\text{-algebra of } F, \\ &x : \Omega \to \mathbb{R}_+,\ E[x] < \infty. \end{aligned}$$

There exists a random variable,

$$\begin{aligned} & E[x|G] : \Omega \to \mathbb{R}_+ \text{ such that,} \\ (1)\quad & E[x|G] \text{ is } G\text{-measurable,} \\ (2)\quad & E[x\ I_A] = E[E[x|G]\ I_A],\ \forall A \in G;\ \text{ hence } E[E[x|G]] = E[x] < \infty. \end{aligned}$$

$E[x|G]$ is unique up to an almost sure modification:
If $y : \Omega \to \mathbb{R}_+$ satisfies (1) and (2) then $y = E[x|G]$ a.s.
Call $E[x|G]$ the conditional expectation of the positive random variable x given or conditioned on the σ-algebra G.

Conditional expectation

Definition (Continued)

(b) For an integrable random variable there exists a random variable called the conditional expectation of the random variable with respect to the σ-algebra G if,

$$x : \Omega \to \mathbb{R},\ E|x| < \infty,$$

$$x = x^+ - x^-,\ E[x^+] + E[x^-] = E|x| < \infty,$$

$$\Rightarrow\ E[x^+] < \infty,\ E[x^-] < \infty,\ \text{hence it follows from (a) that,}$$

$$\Rightarrow\ E[x^+|G] < \infty,\ E[x^-|G] < \infty,$$

$$E[x|G] = E[x^+|G] - E[x^-|G];\ \text{ then,}$$

(1) $E[x|G]$ is G-measurable,

(2) $E[I_A E[x]] = E[I_A E[x|G]],\ \forall\, A \in G.$

Examples of conditional expectation follow after a theorem with properties of conditional expectation.

Conditional expectation

Theorem. Properties of conditional expectation (1)

Consider

$$(\Omega, F, P),\ G,\ G_1,\ G_2 \subseteq F,$$
$$x,\ y : \Omega \to \mathbb{R},\ E|x| < \infty,\ E|y| < \infty.$$

Properties of the conditional expectation operator:

(a) Linearity

$$E[x + y|G] = E[x|G] + E[y|G],$$
$$E[c\,x|G] = c\,E[x|G],\ \forall\, c \in \mathbb{R}.$$

(b) Order preservation

$$x \leq y \Rightarrow\ E[x|G] \leq E[y|G].$$

Conditional expectation

Theorem. Properties of conditional expectation (2)

(c) Measurability

$$\begin{aligned} y \text{ is } G \text{ measurable, and } E|x\ y| < \infty \ \Rightarrow \\ E[x\ y|\ G] = y\ E[x|\ G]; \\ E[y|\ G] = y, \text{ in particular.} \end{aligned}$$

(d) Reconditioning

$$\begin{aligned} G_1 \subseteq G_2 \ \Rightarrow \\ E[x|\ G_1] = E[\ E[x|\ G_2]|\ G_1]; \\ E[\ E[x|\ G]] = E[x], \text{in particular.} \end{aligned}$$

Conditional expectation

Theorem. Properties of conditional expectation (3)

(e) Independence,

$$F^x, G \text{ independent implies that}$$
$$E[x|G] = E[x].$$

(f) Equivalence of independence

$$x : \Omega \to \mathbb{R}^{n_x},\ n_x \in \mathbb{Z}_+;$$
$$F^x,\ G \ \text{ are independent}$$
$$\text{if and only if}$$
$$E[\exp(iw^T x)|G] = E[\exp(iw^T x)],\ \forall w \in \mathbb{R}^{n_x}.$$

Conditional expectation

Theorem. Conditional expectation of Gaussian random variables
Consider

$$
\begin{aligned}
& x : \Omega \to \mathbb{R}^{n_x},\ y : \Omega \to \mathbb{R}^{n_y}, \\
& (x, y) \in G\left(\begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} Q_x & Q_{xy} \\ Q_{xy}^T & Q_y \end{bmatrix} \right), \\
& 0 \prec Q_y \text{ assumed;}
\end{aligned}
$$

$$
\begin{aligned}
E[x|F^y] &= m_x + Q_{xy} Q_y^{-1} (y - m_y), \\
& E[(x - E[x|F^y])(x - E[x|F^y])^T | F^y] \\
& = E[(x - E[x|F^y])(x - E[x|F^y])^T] \\
& = Q_x - Q_{xy} Q_y^{-1} Q_{xy}^T = \widetilde{Q} \in \mathbb{R}_{pds}^{n_x \times n_x}, \\
E[\exp(iw^T x)|F^y] &= \exp(iw^T E[x|F^y] - \frac{1}{2} w^T \widetilde{Q} w),\ \forall\ w \in \mathbb{R}^{n_x}.
\end{aligned}
$$

Please read proof in book. Note that $\widetilde{Q}$ does not depend on y!

Conditional expectation

Theorem. Conditional expectation of a simple random variable

Consider

$$y = C_y\ i_y = \sum_{k=1}^{n_{i_y}} C_{y,k}\ i_{y,k},\ \ \text{a simple random variable,}$$

$$x, y : \Omega \to \mathbb{R},\ \ E|x| < +\infty,\ C_y \in \mathbb{R}^{1 \times n_{i_y}},$$

assume that,

$$E[i_{y,k}] = P(\{\omega \in \Omega |\ y(\omega) = C_{y,k}\}) > 0,\ \ \forall\ k \in \mathbb{Z}_{n_{i_y}};\ \text{then,}$$

$$E[x|F^y] = C_{x|y}\ i_y = \sum_{k=1}^{n_{i_y}} C_{x|y,k}\ i_{y,k},$$

$$C_{x|y,k} = E[x\ i_{y,k}]/E[i_{y,k}],\ \forall\ k \in \mathbb{Z}_{n_{i_y}},\ \ C_{x|y} \in \mathbb{R}^{1 \times n_{i_y}}.$$

Conditional expectation

Example. Conditional expectation of a simple random variable

Consider two finite-valued random variables with

$$
\begin{aligned}
(\Omega, F, P), \Omega = \mathbb{Z}_9 &= \{1,\ 2,\ 3,\ \ldots, 9\}, \\
F^x &= \{\{1,\ 2,\ 3\},\ \{4,\ 5,\ 6\},\ \{7,\ 8,\ 9\}\} = \{A_1,\ A_2,\ A_3\}, \\
F^y &= \{\{1,\ 4,\ 7\},\ \{2,\ 5\},\ \{3\},\ \{6,\ 8,\ 9\}\} = \{B_1,\ B_2,\ B_3,\ B_4\}, \\
1/9 &= P(\{i\}),\ \forall\ i \in \mathbb{Z}_9,\ \ \text{uniform measure on a finite set}, \\
x &= I_{\{1,\ 2,\ 3\}} = I_{A_1}, \\
y &= 4\ I_{B_1} + 3 I_{B_2} + 2 I_{B_3} + 0 I_{B_4}, \\
E[x|\ F^y] &= (1/3) I_{B_1} + (1/2) I_{B_2} + I_{B_3} + 0 I_{B_4}.
\end{aligned}
$$

The proof of the conditional expectation is a simple computation.

Conditional expectation

Lecture 1. Topics of importance

- Conditional expectation
  of a Gaussian random variable and
  of a finite-valued random variable.
- Gaussian random variables and their properties.
- The theoretical framework
  of σ-algebras, probability measures, and random variables,
  for students interested in theory.