Mathematics and numerics for data assimilation and state estimation – Lecture 2

Mathematics for Uncertainty Quantification | RWTH AACHEN UNIVERSITY

Summer semester 2020

Overview

1. Summary of lecture 1
2. Discrete random variables
   - Independence of random variables and events
   - Expected value and moments
3. Conditional probability and expectation

On ubungs, presentation and lectures

- 10:30-12:00 on most Fridays.
- Structure: 5-10 questions, which I will put up in pdf form on course webpage and on Moodle. Roughly 30 minutes work in groups or alone, where I will be present for discussions, thereafter solutions in plenary by me and/or you.
- No hand-ins, unless you want to (i.e., only for feedbac, kdoes not affect grade).
- The only "graded" part of the course, in the form of bonus points, is the presentation early July, and, of course, the final exam.
- Presentations can be done alone or in groups of maximum 2 people.
- Lectures after July 17th moved to first week of June.

Overview

1. Summary of lecture 1
2. Discrete random variables
   - Independence of random variables and events
   - Expected value and moments
3. Conditional probability and expectation

Measurable spaces and probability measures

- introduced a probabilty space $(\Omega, \mathcal{F}, \mathbb{P})$
- discrete random variable $X : \Omega \to A = \{a_1, a_2, \dots, \}$ satisfies the event constraints

$$X^{-1}(a) = \{\omega \in \Omega \mid X(\omega) = a\} \in \mathcal{F} \quad \text{for all} \quad a \in A.$$

- X can be represented by a simple function

$$X(\omega) = \sum_{a \in A} a \mathbb{1}_{X=a}(\omega). \quad \text{where } \mathbb{1}_{X=a}(\omega) := \begin{cases} 1 & \text{if } X(\omega) = a \\ 0 & \text{otherwise} \end{cases}$$

Overview

1 Summary of lecture 1

2 Discrete random variables
- Independence of random variables and events
- Expected value and moments

3 Conditional probability and expectation

Discrete random variables 2

Example 1 (Coin toss, $X \sim \text{Bernoulli}(p)$)

- image-space outcomes $A = \{0, 1\}$,
- $$\Omega = \{Heads, Tails\}, \qquad \mathcal{F} = \{\emptyset, \{Heads\}, \{Tails\}, \Omega\}$$
- $X(Heads) = 1$ and $X(Tails) = 0$ and

$$\mathbb{P}(X = 1) = \mathbb{P}(X^{-1}(1)) = \mathbb{P}(Heads) = p, \quad \mathbb{P}(X = 0) = \mathbb{P}(Tails) = 1 - p.$$

Comment from last lecture: image-outcomes $\{a_1, a_2, \ldots,\}$ may not be associated uniquely to (probability-space) outcomes in Ω.

Larger set of outcomes in Ω than in A

Alternative, and admittedly confusing, probability space for the same rv as in the preceding example:

Example 2 (Coin toss, $X \sim \text{Bernoulli}(p)$)

- image-space outcomes $A = \{0, 1\} \subset \mathbb{R}$,
- $\Omega = \{Heads, Tails, Nose\}$ and

$$\mathcal{F} = \{\emptyset, \{Nose\}, \{Heads\}, \{Tails\}, \{Nose, Heads\}, \{Nose, Tails\}, \{Heads, Tails\}, \Omega\}$$

- $X^{-1}(1) = \{Heads, Nose\}$ and $X^{-1}(0) = \{Tails\}$ and

$$\mathbb{P}(X = 1) = \mathbb{P}(X^{-1}(1)) = \mathbb{P}(\{Heads, Nose\}) = p,$$
$$\mathbb{P}(X = 0) = \mathbb{P}(Tails) = 1 - p.$$

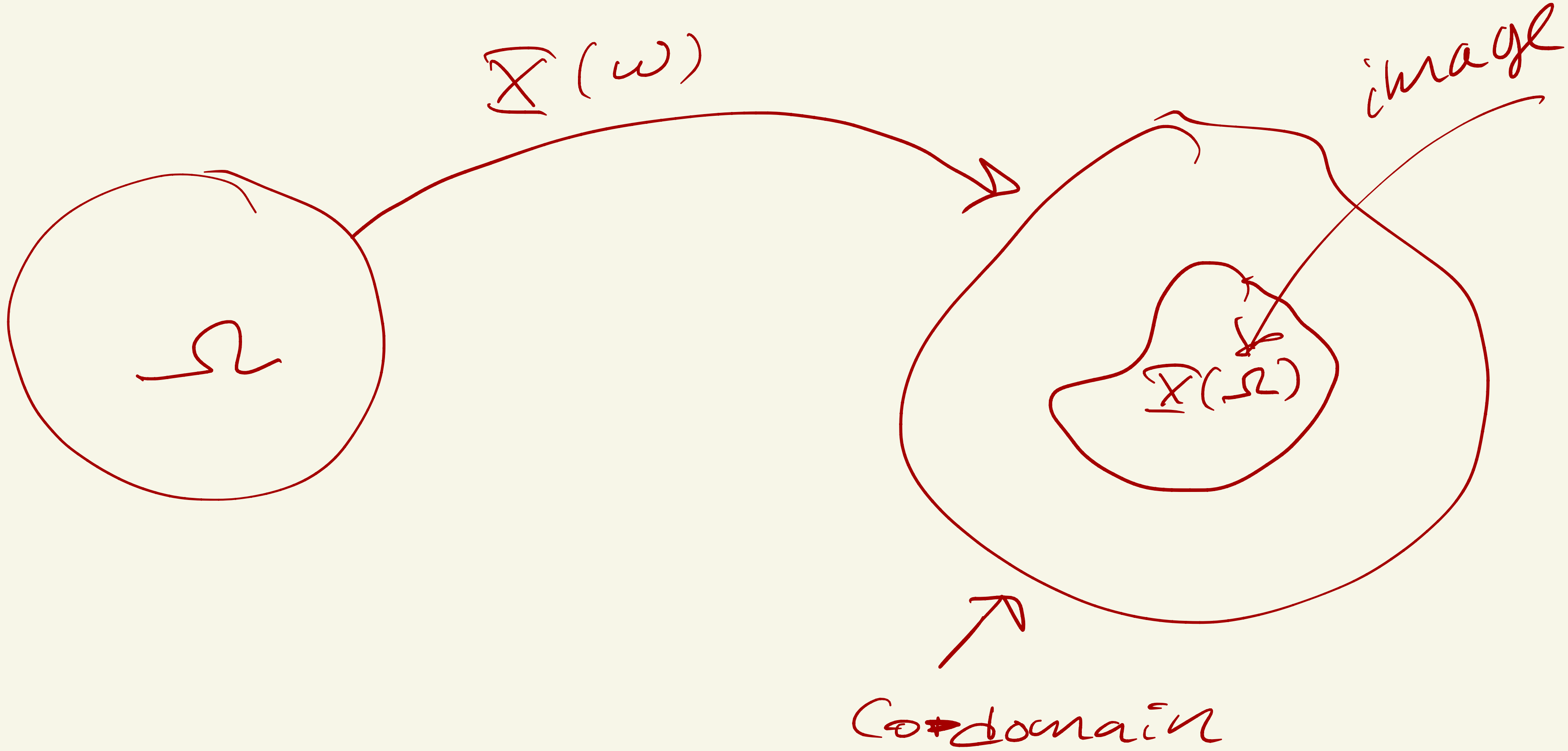
Motivation: if, for instance, you want to represent both a coin toss and a three-sided-die toss in the same probability space.

Joint rv

$\subset \{a_1, a_2, \dots\}$

If $X : \Omega \to A$ and $Y : \Omega \to B = \{b_1, b_2, \dots\}$ are two discrete rv on the same probability space, then

- $(X, Y) : \Omega \to A \times B$ is also a discrete rv with countable set of outcomes
$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$
- with joint distribution:
$$\mathbb{P}_{(X,Y)}((a, b)) = \mathbb{P}(X = a, Y = b).$$
- Question: why is $\mathbb{P}(X = a, Y = b)$ defined? Answer: when we say X and Y are defined on the same probability space, this entails that
$$\{X = a\}, \{Y = b\} \in \mathcal{F} \quad \underbrace{\Longrightarrow}_{\text{since } \mathcal{F} \text{ is } \sigma-\text{algebra}} \quad \{X = a\} \cap \{Y = b\} \in \mathcal{F},$$
and
$$\mathbb{P}(X = a, Y = b) = \mathbb{P}(\{X = a\} \cap \{Y = b\}).$$


$X(\omega)$
Ω
image
$X(\Omega)$
Codomain

Definition 3 (Independence of two rv)

If $X : \Omega \to A$ and $Y : \Omega \to B = \{b_1, b_2, \ldots\}$ are two discrete rv on the same probability space[a] are said to be independent random variables if

$$\mathbb{P}(X = a, Y = b) = \mathbb{P}(X = a)\mathbb{P}(Y = b), \qquad \forall a \in A \quad b \in B.$$

Notation: $X \perp Y$.

[a] From now on, it will be implicitly assumed that all rv are defined on the same probability space, unless otherwise stated.

Example 4

Given independent coin tosses $X_k \sim \mathit{Bernoulli}(1/2)$ for $k = 1, 2$, describe the smallest possible σ-algebra on which the rv (X_1, X_2) is defined.

Solution:

$(X_1, X_2)(\Omega) = \{(0,0), (1,0), (0,1), (1,1)\}$

$X^{-1}((0,0)), X^{-1}((1,0)), X^{-1}((0,1)), X^{-1}((1,1)) \in \mathcal{F}$

and these events are disjoint

suggestion

$$X^{-1}((0,0)) = \{\text{Tails}, \text{Tails}\} = \{TT\}$$

$$X^{-1}((1,0)) = \{\text{Heads}, \text{Tails}\} = \{HT\}$$

$$X^{-1}((0,1)) = \{\text{Tails}, \text{Heads}\} = \{TH\}$$

$$X^{-1}((1,1)) = \{\text{Heads}, \text{Heads}\} = \{HH\}$$

Answer: smallest σ-algebra containing the above sets and $\emptyset$, and Ω

$$\mathcal{F} = \{\emptyset, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}, \{HH, HT, TH\}, \{HH, HT, TT\}, \{HH, TH, TT\}, \{HT, TH, TT\}, \{HH, HT, TH, TT\}\}.$$

Example 5 (one coin toss and one three-sided-die toss)

- Consider $X : \Omega \to \{0, 1\}$ and and $Y : \Omega \to \{1, 2, 3\}$ both defined on the probability space from Example 2.
- Recall that $X^{-1}(1) = \{Heads, Nose\}$ and $X^{-1}(0) = \{Tails\}$ and let us assume that

$$\mathbb{P}(X = 1) = 1/2, \quad \mathbb{P}(X = 0) = 1/2$$

  and that $Y^{-1}(1) = \{Heads\}$, $Y^{-1}(2) = \{Nose\}$ and $Y^{-1}(3) = \{Tails\}$.

- Quation: For $p = 1/2$, what is

$$\mathbb{P}(X = 0, Y \in \{1, 2\}) = \mathbb{P}(\{Tails\} \cap (\{Heads\} \cup \{Nose\}))$$
$$= \mathbb{P}(\emptyset) = 0$$

- Question: Are X and Y independent?

$X \perp Y \iff \mathbb{P}(X=a, Y=b) = \mathbb{P}(X=a)\,\mathbb{P}(Y=b)$

On the one hand

$$\mathbb{P}(X=0, Y\in\{1,2\}) = 0$$

on the other

$$\mathbb{P}(X=0)\,\mathbb{P}(Y\in\{1,2\}) = \mathbb{P}(\text{Tails})\,\mathbb{P}(\{\text{Heads}, \text{Noses}\})$$

$$= \frac{1}{4}$$

so $\mathbb{P}(X=0, Y\in\{1,2\}) \neq \mathbb{P}(X=0)\,\mathbb{P}(Y\in\{1,2\})$

$\Rightarrow Y \not\perp X$.

Independence of multiple rv

$\{a_1, a_2, \ldots\}$

Definition 6

Let $X_k : \Omega \to A_k$ for $k = 1, 2, \ldots, N$, be a finite sequence of discrete rv. Then $X_1, X_2, \ldots, X_N$ are independent provided

$$\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N) = \prod_{k=1}^{N} \mathbb{P}(X_k = a_k) \tag{1}$$

for all $a_1 \in A_1$, $a_2 \in A_2$, $\ldots$, $a_n \in A_N$.

Extension: A **countable** sequence of discrete rv $X_1, X_2, \ldots$ are independent provided every finite subsequence $\{X_{k_j}\}_j$ satisfies (1).

$$X_k = \Omega \to A_k = \{0, 1\}$$

Example 7

Let $X_i \sim Bernoulli(p)$ for $i = 1, \ldots, N$ with joint distribution

$$\mathbb{P}\left(X_1 = a_1, X_2 = a_2, \ldots, X_N = a_N\right) = p^{\sum_{k=1}^{N} a_k}(1-p)^{N-\sum_{k=1}^{N} a_k}$$

for any $a_1, \ldots, a_N \in \{0, 1\}$. Then $X_1, X_2, \ldots$ are independent and identically distributed (iid).

$(*)\ \mathbb{P}(X_k = a_k) = p^{a_k}(1-p)^{1-a_k}$ for $k = 1, \ldots, N$

$$\Rightarrow \mathbb{P}(X_1 = a_1, \ldots, X_N = a_N) = \prod_{k=1}^{N} \underline{\mathbb{P}(X_k = a_k)}$$

Example 8 (Functions of joint discrete rv are also discrete rv)

Let $X_i \sim \mathit{Bernoulli}(p)$ be independent for $i = 1, 2, \ldots, N$ and

$$S_N = f(X_1, \ldots, X_N) := \sum_{i=1}^{N} X_i.$$

Then

$$\mathbb{P}(S_N = k) = \binom{N}{k}(1-p)^{N-k}p^k$$

S_N is called the **Binomial distribution** with degrees of freedom N and p, and we write $S_N \sim B(N, p)$.

Comment: the number of different ways the event $\{S_N = k\}$ when flipping N independent coins once equals factor in the $k+1$-th summand of

$$((1-p)+p)^N = (1-p)^N + \binom{N}{1}p(1-p)^{N-1} + \ldots + \binom{N}{k}p^k(1-p)^{N-k} + \ldots$$

Independence of events

Equation (1) is on the form:

$$\mathbb{P}\left(\bigcap_{k=1}^{N}\{X_k = a_k\}\right) = \mathbb{P}(\text{intersection of events}) = \text{Product of}\Big[\mathbb{P}(\text{each event})\Big]$$

Definition 9

A finite sequence of events $H_1, H_2, \ldots, H_N$ that belongs to $\mathcal{F}$ are independent provided

$$\mathbb{P}\left(\bigcap_{k=1}^{N} H_k\right) = \prod_{k=1}^{N} \mathbb{P}(H_k) \tag{2}$$

A **countable** sequence of events A_1, A_2, belonging to $\mathcal{F}$ are independent provided finite subsequence $\{A_{k_j}\}_j$ satisfies (2).

Connection between independence of rv and independence of events

$$\mathbb{1}_{H_1} \perp \mathbb{1}_{H_2} \Leftrightarrow \mathbb{P}(\mathbb{1}_{H_1}=a, \mathbb{1}_{H_2}=b) = \mathbb{P}(\mathbb{1}_{H_1}=a)\mathbb{P}(\mathbb{1}_{H_2}=b)$$
$$\forall a,b \in \{0,1\}$$

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we can assign an rv to each event $H \in \mathcal{F}$ as follows

$$\mathbb{1}_H(\omega) := \begin{cases} 1 & \omega \in H \\ 0 & \text{otherwise} \end{cases}.$$

Easy consequence of preceding definition: $\mathbb{1}_{H_1}$ and $\mathbb{1}_{H_2}$ are independent if and only if

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(H_1)\mathbb{P}(H_2).$$

$$\mathbb{1}_H : \Omega \to \{0,1\}$$

$$\left.\begin{aligned} &\mathbb{1}_H^{-1}(0) = H^c \in \mathcal{F} \\ &\mathbb{1}_H^{-1}(1) = H \in \mathcal{F} \end{aligned}\right\} \text{verification } \mathbb{1}_H \text{ an rv.}$$

$$\mathbb{1}_H \sim \text{Bernoulli}(\mathbb{P}(H))$$

Expectation of rv

Definition 10

For a discrete rv $X : \Omega \to A \subset \mathbb{R}^d$, the expectation X is defined as

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \sum_{a\in A} a\mathbb{P}(X = a)$$

Motivation of the above integral:

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\Omega \sum_{a\in A} a \, \mathbb{1}_{X=a}(\omega)\mathbb{P}(d\omega)$$

$$= \sum_{a\in A} a \int_\Omega \mathbb{1}_{X=a} \mathbb{P}(d\omega) = \dots$$

next page

- The condition

$$\mathbb{E}[|X|] = \sum_{a\in A} |a|\mathbb{P}(X = a) < \infty$$

  is a sufficent condition for $\mathbb{E}[X]$ being defined and bounded.

- Example for $X \sim \mathit{Beronoulli}(p)$

$$\mathbb{E}[X] =? \; 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) = 0\cdot(1-p) + 1\cdot p = p$$

$$\dots = \sum_{a \in A} a \int_\Omega \mathbb{1}_{X=a}(\omega)\, \mathbb{P}(d\omega)$$

$$= \sum_{a \in A} a \int_{\{X=a\}} \mathbb{P}(d\omega)$$

$$= \sum_{a \in A} a\, \mathbb{P}(X=a)$$

Expectation of rv

Definition 11

For a discrete rv $X : \Omega \to A \subset \mathbb{R}^d$, the expectation X is defined as

$$\mathbb{E}[X] := \int_\Omega X(\omega)\mathbb{P}(d\omega) = \sum_{a\in A} a\mathbb{P}\left(X = a\right)$$

- The condition
$$\mathbb{E}[|X|] = \sum_{a\in A} |a|\mathbb{P}\left(X = a\right) < \infty$$
is a sufficent condition for $\mathbb{E}[X]$ being defined and bounded.
- For mappings $f : \mathbb{R}^d \to \mathbb{R}^k$ and rv $f(X)$ the above definition readily extends:
$$\mathbb{E}[f(X)] = \sum_{a\in A} f(a)\mathbb{P}\left(X = a\right).$$
- Example for $X \sim Beronoulli(p)$
$$\mathbb{E}[X] =$$

$f(X)(\omega) = \sum_{a \in A} f(a) \mathbb{1}_{\{X = a\}}(\omega)$

Properties of the expectation

- For mappings $f : \mathbb{R}^d \to \mathbb{R}^k$ and rv $f(X)$, the expectation becomes

$$\mathbb{E}[f(X)] = \sum_{a \in A} f(a) \mathbb{P}(X = a) .$$

- For a pair of rv $X : \Omega \to A \subset \mathbb{R}^d$ and $Y : \Omega \to B \subset \mathbb{R}^d$, it holds for any $c \in \mathbb{R}$, that

$$\mathbb{E}[X + cY] = \mathbb{E}[X] + c\,\mathbb{E}[Y]$$

provided $\mathbb{E}[|X|] + \mathbb{E}[|Y|] < \infty$ (sufficient condition).

Motivation:

$$\int_\Omega X(\omega) + cY(\omega)\, \mathbb{P}(d\omega) = \int_\Omega X(\omega)\, \mathbb{P}(d\omega) + c \int_\Omega Y(\omega)\, \mathbb{P}(d\omega)$$

Properties of the expectation 2

- Probability of events can be expressed through expectations:

$$\mathbb{P}(H) = \int_H \mathbb{P}(d\omega) = \int_\Omega \mathbb{1}_H(\omega)\,\mathbb{P}(d\omega) = \mathbb{E}[\mathbb{1}_H]$$

  for any $H \in \mathcal{F}$.

- Expectation of discrete rv of the form $f(X, Y)$ where $X : \Omega \to A$ and $Y : \Omega \to B$:

$$\mathbb{E}[f(X, Y)] = \sum_{a\in A,\, b\in B} f(a,b)\,\mathbb{P}(X=a, Y=b)$$

Variance of an rv

$k \in \mathbb{R}$

- For $X : \Omega \to A \subset \mathbb{R}$
$$F(k) = \mathbb{E}[(X-k)^2]$$
is the squared deviation of X from k in expectation.
- For $\mu := \mathbb{E}[X]$, and provided $\mathbb{E}[X^2] < \infty$, it can be shown that
$$\underline{\pm\mu} \quad F(\mu) \leq F(k) \quad \text{for all} \quad k \in \mathbb{R},$$

$F(k) = \mathbb{E}[(X-k)^2] = \mathbb{E}[((X-\mu)+(\mu-k))^2]$

$= \mathbb{E}[(X-\mu)^2] + 2(\mu-k)\mathbb{E}[X-\mu] + (\mu-k)^2 = \ldots$

- Which motivates the variance of X:
$$\mathrm{Var}(X) := \mathbb{E}[(X-\mu)^2]$$

cont next page

- For $X \sim \mathit{Bernolli}(p)$, $\mu = p$ and
$$\mathrm{Var}(X) =$$

Observe that

$$\mathbb{E}[X-\mu] = \mu - \mu = 0$$

$$\Rightarrow \underset{F(K)}{\ldots} = \mathbb{E}[(X-\mu)^2] + 0 + (\mu - K)^2$$

$$\mathbb{E}[(X-K)^2] = \mathbb{E}[(X-\mu)^2] + \underbrace{(\mu-K)^2}_{\geq 0}$$

$$\geq \mathbb{E}[(X-\mu)^2] = F(\mu)$$

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_{a\in A} (a-\mu)^2\, \mathbb{P}(X=a)$$

$X \sim \mathrm{Bernoulli}(p)$, $\mathbb{E}[X] = p$

$$\mathrm{Var}(X) = (0-p)^2\, \mathbb{P}(X=0) + (1-p)^2\, \mathbb{P}(X=1) = p^2(1-p) + (1-p)^2 p$$

$$= p(1-p)$$

Notation with same meaning

For events $H_1, H_2, \ldots \in \mathcal{F}$, the following notation is used interchangeably in the literature

$$\mathbb{P}\left(H_1 H_2 \ldots H_n\right) = \mathbb{P}\left(H_1, H_2, \ldots, H_n\right) = \mathbb{P}\left(\bigcap_{j=1}^{n} H_j\right).$$

And since

$$\mathbb{1}_{\bigcap_{j=1}^{n} H_j} = \prod_{i=1}^{n} \mathbb{1}_{H_j}.$$

$$\bigcap_{j=1}^{n} H_j \longleftrightarrow \prod_{i=1}^{n} \mathbb{1}_{H_j}$$

$$\bigcup_{j=1}^{n} H_j \underset{\leq}{\longleftrightarrow} \sum_{j=1}^{n} \mathbb{1}_{H_j}$$

we have that

$$\mathbb{P}\left(\bigcap_{j=1}^{n} H_j\right) = \mathbb{E}[\mathbb{1}_{\bigcap_{j=1}^{n} H_j}] = \mathbb{E}[\prod_{i=1}^{n} \mathbb{1}_{H_j}].$$

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \leq \underbrace{\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{1}_{A_j}\right]}$$