Probability and Statistics for Economics Cheat Sheet

Author: Sai Zhang (email me or check my Github page)

Random experiments

The outcome in a random experiment is **unpredictable**:

- outcome is too complicates or poorly understood
- outcome is designed to be unpredictable
- coincidences, or independent chains of events

Two examples

Random card shuffle experiment: take top card from a deck and insert randomly, to complete the shuffle of n cards, we need

$$T = n + \frac{n}{2} + \cdots + \frac{n}{n-1} + 1 = n\log n$$

shuffles.

Random number generator:

$$x_{n+1} = \frac{ax_n + b}{c} - \left[\frac{ax_n + b}{c}\right]$$

the remainder after dividing by c, hence $x_{n+1} \in [0, c-1]$, let $u_{n+1} = \frac{x_{n+1}}{c}$, x_0, a, b, c all be integers. For very large a and good choice of b, c, the sequence $u_1, u_2, \cdots$ is like a sequence of numbers randomly picked from $[0,1]$

Probabilities

Probability is a number in $[0,1]$ that measures the likelihood of an outcome or a set of outcomes.

Ways of assigning probabilities:

- **symmetry**: assume all outcomes are equally likely
- **experimental method**: relative frequency in repeated random experiment
- **subjective method**: assign probabilities using knowledge of random experiment
- **market method**

Elements of probability space

- **outcome space** Ω and outcomes $\omega \in \Omega$
- **event** E, $E \subset \Omega$
- **probability function/measure** $P : \mathcal{A} \to [0,1]$: a function from **a collection** $\mathcal{A}$ of subsets of Ω to the interval $[\mathbf{0}, \mathbf{1}]$.

Classes of events

Events $E_1, E_2, \cdots$ are just sets. They also follow the algebras of sets.

Some set algebras

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cup A^C = U$, $A \setminus B = A \cap B^C$

$(\bigcup_{i=1}^{\infty} A_i)^C = \bigcap_{i=1}^{\infty} A_i^C$

$(\bigcap_{i=1}^{\infty} A_i)^C = \bigcup_{i=1}^{\infty} A_i^C$

$A \cup B = U, A \cap B = \varnothing \Leftrightarrow B = A^C$

$(A^C)^C = A$

$A \subseteq B \Leftrightarrow A \cap B = A \Leftrightarrow A \cup B = B \Leftrightarrow A \setminus B = \varnothing \Leftrightarrow B^C \subseteq A^C$

Two special relations:

- **disjoint**: $E_1 \cap E_2 = \varnothing$
- **partition**: $\bigcup_{i=1}^{\infty} E_i = \Omega$, $\{E_i\}$ are pairwise disjoint

σ−field and Borel σ−field

Definition of σ−field

$\mathcal{A}$ (a collection of subsets of Ω) is a σ−field if:

1. $\varnothing \in \mathcal{A}$
2. $E \in \mathcal{A} \Rightarrow E^C \in \mathcal{A}$
3. $E_1, E_2, \cdots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$

It is easy to see that: $\Omega \in \mathcal{A}$ (by 1), $\bigcap E_i^C \in \mathcal{A}$, $\bigcup E_i^C \in \mathcal{A}$, $\bigcap E_i \in \mathcal{A}$ as well

Two important σ−field:

- Trivial σ−field: $\mathcal{A} = \{\varnothing, \Omega\}$
- Largest σ−field: **powerset** of Ω, $\mathcal{P}(\Omega)$

Generating classes

When the powerset has too many events to be assigned probabilities to each event, we start from a set of events $\mathcal{E}$ that we want to assign probabilities to. This $\mathcal{E}$ is a **generating class**.

σ−field generated by $\mathcal{E}$

The **smallest** σ−field that contains $\mathcal{E}$ is

$$\sigma(\mathcal{E}) = \{E \subseteq \Omega \mid E \in \mathcal{A}, \forall \mathcal{A} \supseteq \mathcal{E}\}$$

where $\mathcal{A}$ can be any σ−field (including the powerset).

Another way to write this definition is: for all the σ−fields that contain $\mathcal{E}$, $\{\mathcal{A}_i \mid \mathcal{E} \subseteq \mathcal{A}_i\}$, we have:

$$\sigma(\mathcal{E}) = \bigcap \mathcal{A}_i$$

This definition is based on the fact: **any intersection of σ−fields is a σ−field**, here is a proof:

Proof: $\cap$ of two σ−fields is a σ−field

We can check the three definitions of σ−field: for $\mathcal{F}_1, \mathcal{F}_2$

1. $\varnothing \in \mathcal{F}_1, \varnothing \in \mathcal{F}_2 \Rightarrow \varnothing \in \mathcal{F}_1 \cap \mathcal{F}_2$
2. $E \in \mathcal{F}_1 \cap \mathcal{F}_2 \Rightarrow E^C \in \mathcal{F}_1, E^C \in \mathcal{F}_2 \Rightarrow E^C \in \mathcal{F}_1 \cap \mathcal{F}_2$
3. $E_1, \cdots \in \mathcal{F}_1 \cap \mathcal{F}_2 \Rightarrow E_1, \cdots \in \mathcal{F}_1, E_1, \cdots \in \mathcal{F}_2 \Rightarrow\Rightarrow \bigcup E_i \in \mathcal{F}_1, \bigcup E_i \in \mathcal{F}_2 \Rightarrow \bigcup E_i \in \mathcal{F}_1 \cap \mathcal{F}_2$

Borel σ−field

Borel σ−field $\mathcal{B}$ is a σ−field on $\mathbb{R}$, its generating class is the set of all **open subsets** of $\mathbb{R}$. But $\mathcal{B}$'s generating class is not unique, a particularly important one is $\mathcal{E} = \{(-\infty, x] \mid x \in \mathbb{R}\}$. Here is a proof of this generating class can actually generate $\mathcal{B}$.

Proof: $\sigma(\mathcal{E}) = \mathcal{B}$

The proof is done in 2 steps:

- prove $\sigma(\mathcal{E}) \subseteq \mathcal{B}$: $\mathcal{E} = (-\infty, x] = \bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right) \in \mathcal{B}$. And by the definition of σ−field generated by $\mathcal{E}$, we know that $\sigma(\mathcal{E}) = \bigcap \mathcal{A}_i \subseteq \mathcal{E}$, hence $\sigma(\mathcal{E}) \subseteq \mathcal{B}$
- prove $\mathcal{B} \subseteq \sigma(\mathcal{E})$: each open set B in $\mathbb{R}$ can be written as $B = \cup_{i=1}^{\infty} (a_i, b_i)$. For each (a,b), we can rewrite it as $(a,b) = (-\infty, b) \cap (-\infty, a]^C$, where $(-\infty, b) = \bigcup_{n=1}^{\infty} \left(\infty, b - \frac{1}{n}\right]$. By the definition of σ−field, $\bigcup_{n=1}^{\infty} \left(\infty, b - \frac{1}{n}\right] \cap (-\infty, a]^C \in \sigma(\mathcal{E})$, hence $B \in \sigma(\mathcal{E})$.

This is actually a special case of **generating class arguments**, which is used to show that all sets in a σ−field $\mathcal{A}$ have a certain property. It follows:

- for subsets of Ω that have the property, define the collection of them as $\mathcal{E}$
- show that $\mathcal{A} \subseteq \sigma(\mathcal{E})$
- show that for $\mathcal{A}_0 = \{A_{\text{property}} \in \mathcal{A}\}$ with $\mathcal{E} \subseteq \mathcal{A}_0$, $\mathcal{A}_0$ is a sigma field (Since $\sigma(\mathcal{E})$ is the smallest σ−field that contains $\mathcal{E}$, $\sigma(\mathcal{E}) \subseteq \mathcal{A}_0$)
- $\mathcal{A} \subseteq \sigma(\mathcal{E}) \subseteq \mathcal{A}_0 \subseteq \mathcal{A} \Rightarrow \mathcal{A} = \mathcal{A}_0$

Often it is difficult to show that $\mathcal{A}_0$ is a σ−field, hence we introduce λ−systems.

λ−systems

$\mathcal{A}_0$ is a λ−system if

1. $\Omega \in \mathcal{A}_0$
2. If $D_1, D_2 \in \mathcal{A}_0$ and $D_2 \subseteq D_1$, then $D_1 \setminus D_2 = D_1 \cap D_2^C \in \mathcal{A}_0$
3. If D_n is an increasing sequence of sets in $\mathcal{A}_0$, then $\bigcup_{i=1}^{\infty} D_i \in \mathcal{A}_0$

Two theorems link λ−systems and σ−fields:

Theorems of λ−systems/σ−fields

- **Theorem 1**: If $\mathcal{E}$ is closed under finite intersections, and if $\mathcal{A}_0$ is a λ−system with $\mathcal{E} \subseteq \mathcal{A}_0$, then $\sigma(\mathcal{E}) \subseteq \mathcal{A}_0$.
- **Theorem 2**: a λ−system $\mathcal{A}_0$ is a σ−field $\Leftrightarrow$ $\mathcal{A}_0$ is closed in finite intersections.

Choice of σ−field

The choice of σ−field is usually determined by the nature of the outcome space:

- discrete (countable) outcome space: **powerset of** Ω shall be chosen.
- continuous outcome space (real line): **Borel σ−field** shall be chosen.

Probability measure

Definition of probability measure

A probability measure is a function $P : \mathcal{A} \Rightarrow \mathbb{R}$ with a σ−field $\mathcal{A}$:

1. $\forall E \in \mathcal{A}, P(E) \geq 0$
2. $P(\Omega) = 1$
3. If $E_1, E_2, \cdots$ are pairwise disjoint (and countable), $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

The usual way to specify probabilities using a probability model is to first assign probabilities to some simple collection of events $\mathcal{E}$, these assignments can be extended to $\sigma(\mathcal{E})$, this extension can usually be shown to be unique, giving a probability measure on $\sigma(\mathcal{E})$.

Discrete Ω: Dice roll

For the problem of rolling a single dice, the probability measure can be defined as

- σ−field $\mathcal{A}$: powerset of Ω
- probability measure: $P(E) = \frac{\#E}{6}$

This construction can be extended to all discrete outcome spaces: starting by assigning probabilities p_i to single outcomes and define $P(E) = \sum_{i \in E} p_i$.

Here is a special case of discrete Ω: random experiments with equally likely outcomes, i.e., $\Omega = \{\omega_i\}_{i=1}^{I}, P(\omega_i) = \frac{1}{I}, P(E) = \frac{\#E}{I}$.

4 cases of equally likely Ω

Consider the problem of selecting k elements from n distinct elements, the number of selections N_s has 4 cases:

- Ordered without replacement: $N_s = \frac{n!}{(n-k)!}$
- Ordered with replacement: $N_s = n^k$
- Unordered without replacement: $N_s = \frac{n!}{k!(n-k)!} = \binom{n}{k}$
- Unordered with replacement: $S_n = \binom{n+k-1}{k}$

Continuous Ω: Borel σ−field $\mathcal{B}$

For the probability measure on $\mathcal{B}$ can be defined as

- probabilities on $\mathcal{E} = \{(-\infty, x] \mid x \in \mathbb{R}\}$

$\mathcal{E} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ is closed under finite intersection, hence, we need to show $\forall B \in \mathcal{B}, P(B) = P'(B)$, that is, $\mathcal{E}$ has a unique probability measure.

Consider $\mathcal{B}_0 = \{B \in \mathcal{B} \mid P(B) = P'(B)\}$, which is a λ−system. For an increasing sequence B_n of events in $\mathcal{B}_0$, $P(B) = P(B_1 \cup (B_2 \setminus B_1) \cup \cdots \cup (B_n \setminus B_{n-1}) = P(B_1) + (P(B_2) - P(B_1)) + \cdots + (P(B_n) - P(B_{n-1})) = \lim P(B_n)$ then $P(B) = \lim P(B_n) = \lim P'(B_n) = P'(B)$. Now $\mathcal{E} \subseteq \mathcal{B}_0$ is closed under finite intersections, then by Thm.1 of λ−system, $\sigma(\mathcal{E}) \subseteq \mathcal{B}_0 \subseteq \mathcal{B}$, then $\mathcal{B} = \sigma(\mathcal{E})$, hence $\forall B \in \mathcal{B}, P(B) = P'(B)$.

Properties of probability measures

- $P(E^C) = 1 - P(E)$, $P(\varnothing) = 0$, $P(E) \leq 1$
- $P(E_2 \setminus E_1) = P(E_2 \cap E_1^C) = P(E_2) - P(E_2 \cap E_1)$
- $P(E_1 \cup E_2) = P(E_1 \cup (E_2 \setminus E_1)) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$
- $E_1 \subseteq E_2 \Rightarrow E_2 = E_1 \cup (E_2 \setminus E_1) \Rightarrow P(E_1) \leq P(E_2)$
- $P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$
- Bonferroni inequality: $P(E_2 \cap E_1) \geq P(E_1) + P(E_2) - 1$
- Law of total probability: for a partition of Ω, $\{E_i\}$, $P(A) = \sum_{i=1}^{\infty} P(A \cap E_i)$
- Boole's inequality: $P(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} P(E_i)$

The proof of all these results relies on the step to transform the object set to a **union of pairwise disjointed** sets.

For all subsets of $\mathbb{R}$, it is impossible to assign probabilities to them. The proof takes advantage of a series of uncountable, disjointed subsets centered at irrational numbers.

Random variable

Definition of random varibles

For a probability space $(\Omega, \mathcal{A}, P)$, a random varialbe X is a function $X : \Omega \to \mathbb{R}$ s.t. $\forall B \in$ Borel $\sigma-$field $\mathcal{B}$, $E = \{\omega \mid X(\omega) \in B\} \in \mathcal{A}$.

Another way of stating it is: X takes a value in $B \Leftrightarrow w \in E$ i.e., event E happens. Thus, $\Pr(X \in B) = \Pr(E)$

2 things to keep in mind

- $E = X^{-1}(B)$ (X^{-1} does **NOT** necessarily exist)
- X is **Borel measurable**

Properties of X^{-1}

- $X^{-1}(B^C) = (X^{-1}(B))^C$
- $X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} X^{-1}(B_i)$

Borel Measurability

Borel measurability of a random variable X (function) is established separately by the countability of outcome space Ω.

Countable Ω

For a countable Ω, $X : \Omega \to \mathbb{R}$ is Borel measurable if $\sigma-$field on Ω, $\mathcal{A}$, is the **powerset** of Ω

Uncountable Ω

Let $\mathcal{E} \subseteq \mathcal{B}$ be a generating class of Borel $\sigma-$field $\mathcal{B}$, for an uncountable Ω and its $\sigma-$field $\mathcal{A}$, X is Borel measurable if $\forall E \in \mathcal{E}, X^{-1}(E) \in \mathcal{A}$.

We use a **generating class argument** to prove this:

Step 1: Define $\mathcal{C} = \left\{B \in \mathcal{B} \mid X^{-1}(B) \in \mathcal{A}\right\}$, since $\forall E \in \mathcal{E}, X^{-1}(E) \in \mathcal{A}$, we know $\mathcal{E} \subseteq \mathcal{C}$.

Step 2: Check $\mathcal{C}$ is a $\sigma-$field:

i $\varnothing \in \mathcal{C}$

ii $B \in \mathcal{C} \Rightarrow X^{-1}(B) \in \mathcal{A} \Rightarrow X^{-1}(B^C) = X^{-1}(B)^C \in \mathcal{A} \Rightarrow B^C \in \mathcal{C}$

iii $B_1, B_2, \cdots \in \mathcal{C} \Rightarrow X^{-1}(B_1), \cdots \in \mathcal{A} \Rightarrow X^{-1}(\bigcup_{i=1}^{\infty} B_i) = \bigcup_{i=1}^{\infty} X^{-1}(B_i) \in \mathcal{A}$

Step 3: Since $\mathcal{B}$ is the smallest $\sigma-$field containing $\mathcal{E}$ (definition of generating class), $\mathcal{C}$ is a $\sigma-$field containing $\mathcal{E}$, $\mathcal{B} \subseteq \mathcal{C}$; by the definition of $\mathcal{C}$, $\mathcal{C} \subseteq \mathcal{B}$. Hence $\mathcal{C} = \mathcal{B}$, meaning that $\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{A}$, and X is Borel measurable.

Borel measurability has some important applications:

Continuous $X : \mathbb{R} \to \mathbb{R}$

For the probability space $(\mathbb{R}, \mathcal{B}, P)$ and a **continuous** $X : \mathbb{R} \to \mathbb{R}$ is Borel measurable.

X_{sup} and X_{inf} of X_n

For a sequence of random variables, $\{X_n\}_{n=1}^{\infty}$, $X_{\text{sup}} = \sup_n X_n$ **and** $X_{\text{inf}} = \inf_n X_n$ **are also Borel measurable**.

A brief proof: Take (x, ∞) as a generating class for $\mathcal{B}$, X_n being random variables $\Rightarrow \{w \mid X_n(w) > x\} \in \mathcal{A} \Rightarrow \bigcup_n \{w \mid X_n(w) > x\} = \left\{w \mid X_{\text{sup}}(w) > x\right\} \in \mathcal{A}$. Taking $(-\infty, x]$ as the generating class can proof for X_{inf}.

$\lim_{n\to\infty} X_n = X$

Again, for a sequence of random variables, $\{X_n\}_{n=1}^{\infty}$, **if** $\lim_{n\to\infty} X_n = X$ **exists,** X **is a random variable**.

A brief proof: we know

$$\liminf_{n\to\infty} X_n(\omega) = \sup_n \inf_{m \ge n} X_m(\omega)$$

$$\limsup_{n\to\infty} X_n(\omega) = \inf_n \sup_{m \ge n} X_m(\omega)$$

then it is easy to see $\liminf_{n\to\infty} X_n \le X = \lim_{n\to\infty} X_n \le \limsup_{n\to\infty} X_n$. If $\lim_{n\to\infty} X_n = X$ exists, the three limits are all equal and all Borel measurable.

Algebras of random variables

If X, Y are random variables, $Z = X + Y$ **is Borel measurable**.

A brief proof: take the generating class $\mathcal{E} = \{(x, \infty)\}$, then for $A = \{\omega \mid Z(\omega) > z\}$, we have $X(\omega) + Y(\omega) > z \Leftrightarrow X(\omega) > z - Y(\omega)$.

Here is the trick: We can always find a rational number r s.t. $X(\omega) > r > z - Y(\omega)$, and $A = \bigcup_r (\{\omega \mid X(\omega) > r\} \cap \{\omega \mid Y(\omega) > z - r\})$.

Since rational numbers are **countable**, X and Y are both Borel measurable, A is in $\mathcal{A}$, hence Z is Borel measurable.

For the collection of all Borel measurable functions $\mathcal{M} = \{X : \Omega \Rightarrow \mathbb{R}\}$ ($\mathcal{M}^+$ for non-negative functions), We focus on a special class:

Simple functions

$$X(\omega) = \sum_i^n \alpha_i I_{A_i}(\omega)$$

where

- I_A is the indicator function of event A
- $A_i \in \mathcal{A}, i = 1, \cdots, n$ are a partition of Ω
- α_i are constants

Why are they important: each non-negative borel measurable function can be approximated by an **increasing sequence** of simple functions!

Theorem of simple functions

For $\forall X \in \mathcal{M}^+$,

$$X_n(\omega) = 2^{-n} \sum_{i=1}^{4^n} I_{X \ge \frac{i}{2^n}}(\omega)$$

satisfies $0 \le X_1(\omega) \le \cdots \le X_n(\omega)$ and $X_n(\omega) \to X(w), \forall \omega \in \Omega$.

Proof of $I_A(\omega)$ approximation

Step 1: increasing

Let $\sum_{i=1}^{4^n} I_{X \ge \frac{i}{2^n}}(\omega) = C$ then it is easy to show that $\sum_{i=1}^{4^{n+1}} I_{X \ge \frac{i}{2^{n+1}}}(\omega) \ge 2C$, hence $X_n = \frac{C}{2^n}$, $X_{n+1} \ge \frac{2C}{2^{n+1}} = X_n$

Step 2: approximation There are two scenarios:

(a) $X(\omega) = \infty$. It will give $X_n(\omega) = 2^{-n} \sum_{i=1}^{4^n} I_{X \ge \frac{i}{2^n}}(\omega) = 2^n$, hence $\lim_{n\to\infty} X_n(\omega) = \infty = X$

(b) $X(\omega) < \infty$. For sufficiently large n, we can find a $k \in \{0, 1, \cdots, 4^n - 1\}$ such that $k2^{-n} \le X < (k+1)2^{-n}$, and $X_n = k2^{-n}$. Hence $|X - X_n| \le 2^{-n} \to 0 \Rightarrow X_n \to X$

Expectation and integration

Start from the simple case: $X(\omega) = \sum_{i=1}^n \alpha_i I_{A_i}(\omega) \Rightarrow E(X) = \sum_{i=1}^n \alpha_i P(A_i) \Rightarrow E(X) = \int_\Omega X(\omega) dP(\omega) = \int X dP$.

For general random variables, we take advantage of the **increasing sequences of simple function** X_n **that asymptotically approaches** X.

Def. of $E(X)$: non-negative X

For simple functions X_s,

$$E(X) = \int X dP = \sup_{X_s} \{E(X_s) \mid X \ge X_s\}$$

Properties of $E(X)$

- $\forall A \in \mathcal{A}, E(I_A) = P(A)$
- $E(\mathbf{0} = 0)$, where $\mathbf{0}$ is the null function that assigns 0 to all $\omega \in \Omega$
- **linearity**: $\forall \alpha, \beta \ge 0$ and nonnegative Borel measurable functions X, Y

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$$

- If $\forall \omega \in \Omega, X(\omega) \le Y(\omega)$, $E(Y) = E(X) + E(Y - X) \ge E(X)$

For **arbitrary** random variable X, we can write $X(\omega) = X_+(\omega) - X_-(\omega)$ with $X_+(\omega) = \max\{X(\omega), 0\}$ and $X_-(\omega) = -\min\{X(\omega), 0\}$. Now X_+ and X_- are both **non-negative**, and:

Def. of $E(X)$: arbitrary X

For random variable $X = X_+ - X_-$,

$$E(X) = E(X_+) - E(X_-) = \int_\Omega X_+ dP - \int_\Omega X_- dP$$

For $E(X)$ to be well-defined, we need $E(|X|) < \infty$, i.e., X is **integrable**.

By introducing Jensen's inequality: $\forall \lambda \in (0, 1)$, and a convex f:

$$f(\lambda x_1 + (1 - \lambda) x_2) \le \lambda f(x_1) + (1 - \lambda) f(x_2)$$

we have the following properties for a convex function f:

$E(f(x))$ for a convex f

For a convex $f : \mathbb{R} \to \mathbb{R}$, we have:

- f is **Borel measurable**: let $E = \{x \mid f(x) \in (-\infty, t]\}$, since f is convex, $\forall x_1, x_2 \in E, f(\lambda x_1 + (1 - \lambda) x_2) \le \lambda f(x_1) + (1 - \lambda) f(x_2) \le t$. Hence, E is an interval on $\mathbb{R}$, f is Borel measurable.
- $\forall x, x_0, f(x) \ge f(x_0) + \alpha(x - x_0)$ where α is a constant (may depend on x_0).

And, derive **two features of** $E(f(x))$:

- $E(f(X)_-) < \infty$ if $E(|X|) < \infty$: we know $f(x) \ge f(x_0) + \alpha(x - x_0) \ge -|f(x_0)| - |\alpha|(|x| + |x_0|)$, hence $E(|X|) < \infty \Rightarrow E(f(X)_-) < E(|X|) < \infty$
- Take $x_0 = E(X)$ get $E(f(X)) > f(E(X)) + \alpha(E(X) - E(X)) \Rightarrow E(f(X)) \ge f(E(X))$

Lebesgue integrals

In $E(X) = \int X dP$, P assigns probability 1 to Ω. **Lebesgue measure** assigns probability 0 to $\varnothing$.

Lebesgue measure ($\Omega = \mathbb{R}$) is defined as $m([a, b]) = b - a$ and Lebesgue measure of a point is 0. Hence $m((a, b)) = b - a$. Open intervals (a, b) are a generating class, m can be uniquely extended to all sets in the Borel $\sigma-$field $\mathcal{B}$.

Lebesgue integral is then $\int_{-\infty}^{\infty} f(x) dx$ (NOT $\int_{\mathbb{R}} f dm$).

Lebesgue and Riemann integral

The relation between Lebesgue integral and Riemann integral:

- Same notation: $\int_{\infty}^{\infty} f(x) dx$
- If f is integrable ($\int_{-\infty}^{\infty} |f(x)| dx < \infty$), and Riemann integral exists, the two are the same.
- If $\int_{-\infty}^{\infty} f(x)_+ dx = \int_{-\infty}^{\infty} f(x)_- dx = \infty$, Lebesgue integral is NOT defined, but Riemann integral $\lim_{t\to\infty} \int_{-t}^{t} f(x) dx$ may exist.

The last group of properties of expectations regards convergence:

Convergence properties of $E(X)$

Three properties are especially important:

(1) **Monotone convergence**: if X_n is an **increasing sequence of non-negative** random variables, i.e. $0 \le X_1 \le X_2 \le \cdots \le X_n$, then

$$\lim_{n\to\infty} X_n = X \Rightarrow \lim_{n\to\infty} E(X_n) = E(X)$$

(2) **Fatou's Lemma**: for $X_1, \cdots, X_n \ge 0$,

$$E(\liminf_{n\to\infty} X_n) \le \liminf_{n\to\infty} E(X_n)$$

(3) **Dominated convergence**: if $X_1, \cdots, X_n$ are integrable ($E(|X_i|) < \infty$), and there is a **non-negative, integrable** random variable Y s.t. $|X_i(\omega)| \le Y(\omega), \forall \omega \in \Omega, \forall i = 1, \cdots, n$, then

$$\lim_{n\to\infty} X_n = X \Rightarrow \lim_{n\to\infty} E(X_n) = E(X)$$

(1) is used to prove (2), (2) is used to prove (3):

Proof of $\rightarrow$ properties of $E(X)$

Monotone convergence:

Fatou's Lemma:

Dominated convergence:

Sets of measure 0

For integrals/expectations, the zero probability events $\{E \in \mathcal{A} \mid P(E) = 0\}$ can be neglected. This gives:

Expectation equality $E(X) = E(Y)$

If random variables X and Y satisfy that for event $E = \{\omega \mid X(\omega) \neq Y(\omega)\}$, $P(E) = 0$, then $E(X) = E(Y)$.

This is very easy to prove:

Proof of $E(X) = E(Y)$

We have

$$E(X) = E(X \cdot I_{X=Y}) + E(X \cdot I_{X \neq Y})$$
$$E(Y) = E(Y \cdot I_{Y=X}) + E(Y \cdot I_{Y \neq X})$$

$E(X \cdot I_{X \neq Y} = E(Y \cdot I_{Y \neq X} = 0$, because it is an integral over E and $P(E) = 0$.

Distribution of random variables

For a random variable $X : \Omega \to \mathbb{R}$, we can replace $(\Omega, \mathcal{A}, P)$ by $(\mathbb{R}, \mathcal{B}, P_X)$, get:

Distribution of X

The **distribution** of X is the measure $P_X : \mathcal{B} \to \mathbb{R}$, defined by

$$P_X(B) = P(X^{-1}(B)) = \Pr(X \in B)$$

The **joint distribution** of a vector of random variables $X = (X_1, \cdots, X_N)^T$ is defined as

$$P_X(C) = P(X^{-1}(C)) = \Pr(X \in C)$$

where $C \in \mathcal{B}^N$, $\mathcal{B}^N$ is the Borel σ−field in $\mathbb{R}^N$, generated by the sets $B_1 \times \cdots \times B_N = \{(x_1, \cdots, x_N) \mid x_i \in B_i, B_i \in \mathcal{B}\}$

Distribution function

For the probability space of random variable X: $(\mathbb{R}, \mathcal{B}, P_X)$, P_X is determined by assigning probabilities to $\mathcal{B}$'s generating class $\mathcal{E} - \{(-\infty, x] \mid x \in \mathbb{R}\}$, hence

Distribution function $F_X(x)$

The **distribution function** F_X of X is defined as

$$F_X(x) = P_X((-\infty, x])$$

This is just the CDF. $F_X : \mathbb{R} \to [0,1]$ and uniquely determines P_X. We can also write it as $F_X(x) = E(I_{X \leq x})$.

Properties of F_X

F_X has the following properties:

(a) **Non-decreasing**:

$$x \geq y \Rightarrow F_X(x) \geq F_X(y)$$

Proof: $x \geq y \Rightarrow (-\infty, y] \subseteq (-\infty, x] \Rightarrow F_X(y) = P_X((-\infty, y]) \leq F_X(x) = P_X((\infty, x])$

(b) **Continuous from the right**:

$$\lim_{y \to x+} F_X(y) = F_X(x)$$

Proof: $F_X(x + 1/n) = E(I_{X \leq x+1/n})$. $I_{X \leq x+1/n} \leq 1$, $\lim_{n\to\infty} I_{X \leq x+1/n} = I_{x \leq x}$, by dominance convergence, $\lim_{n\to\infty} F_X(x+1/n) = F_X(x)$. Same logic gives $\Pr(X < x) = \lim_{y\to x-} F_X(y) = F_X(x-)$.

(c) If $\Pr(X = \infty) = \Pr(X = -\infty) = 0$ then, $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.
Proof: use $\lim_{n\to\infty} I_{X \leq -n} = 0$.

The properties give some facts about the continuity of F_X and probability of points:

- F_X is always right continuous, but **not always continuous**.
- Since $I_{X=x} = \lim_{n\to\infty} I_{x-1/n < X \leq x}$, we have $\Pr(X = x) = \lim_{n\to\infty}(F_X(x) - F_X(x - 1/n)) = F_X(x) - F_X(x-)$.
- If $\Pr(X = x) > 0$, F_X is **discontinuous** in x, the jump at x is $\Pr(X = x)$.

Function of random variable

With a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$, $Y = f \circ X : \Omega \to \mathbb{R}$ is a random variable, i.e., Borel measurable.

distribution of $Y = f \circ X$

$Y : \mathbb{R} \to \mathbb{R}$ is defined on $(\mathbb{R}, \mathcal{B}, P_X$, hence:

$$\Pr(Y \in B) = P_X(\{x | f(x) \in B\}) = P(\{\omega | f(X(\omega)) \in B\})$$

and Y is a simple function on $\mathbb{R}$ and Ω:

$$Y = \sum_{i=1}^n \alpha_i I_{B_i}(x) = \sum_{i=1}^n \alpha_i I_{X^{-1}(B_i)}(\omega)$$

where $\{B_i\}$ is partition of $\mathbb{R}$, $\{X^{-1}(B_i)\}$ is hence a partition of Ω.

expectation of $Y = f \circ X$

The expectation of $Y = f \circ X$ is

$$E(Y) = \int_{\mathbb{R}} f(x) \mathrm{d}P_X(x) = \sum_{i=1}^n \alpha_i P_X(B_i) = \sum_{i=1}^n \alpha_i P(X^{-1}(B_i)) = \int_{\Omega} f(X(\omega)) \mathrm{d}P(\omega) = E(f(X))$$

It is in general true that: integral of f on $(\mathbb{R}, \mathcal{B}, P_X)$ is equal to integral of $f(X)$ on $(\Omega, \mathcal{A}, P)$.
Since $F_X(x)$ and $P_X(x)$ both determine the distribution of X, $\int_{\mathbb{R}} f(x) dP_X = \int_{\mathbb{R}} f(x) dF_X$.

Absolute continuity of P

The goal: to calculate probability by **summation** $P = \sum_i I_A(x_i) f(x_i)$ or **integration** $P = \int I_A(x) f(x) dx$.

We start by finding a measure μ for P, μ and P are both defined on $(\Omega, \mathcal{A})$, and μ is easier to compute.

Requirement of μ

1. $\forall E \in \mathcal{A}, \mu(E) \geq 0$
2. $\mu(\varnothing) = 0$ (instead of $P(\Omega) = 1$)
3. If $E_1, E_2, \cdots$ are pairwise disjoint (and countable), $\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$

* μ is σ−**finite**: $\exists \{A_i\}^{\infty}$, a countable partition of Ω, with $\mu(A_i) < \infty$. **Probability measure P is always** σ−**finite** since $P(\Omega) = 1$

The most important condition of this *transfer* of measure is **absolutely continuous**:

P is absolutely continuous

For P, μ defined on $\mathcal{A}$, P is **absolutely continuous** w.r.t. μ if

$$\mu(A) = 0 \Rightarrow P(A) = 0$$

μ is the dominating measure, $P \ll \mu$

Now, we want to show $P \ll \mu \Rightarrow \int I_A \mathrm{d}P = \int I_A f(\omega) d\mu$, where f is the density of P w.r.t. μ. We need the following theorem:

Theorem: measure change P to μ

For probability measure P and measure μ defined both on $(\Omega, \mathcal{A})$, $\exists N$ with $\mu(N) = 0$ and a **non-negative Borel measurable** $f : \Omega \to \mathbb{R}$, s.t. for **non-negative Borel measurable** $g : \Omega \to \mathbb{R}$, we have:

$$\int_{\Omega} g(\omega) \mathrm{d}P(\omega) = \underbrace{\int_{\Omega} g(\omega) I_N(\omega) \mathrm{d}P(\omega)}_{=0 \text{ if } P \ll \mu} + \int_{\omega} g(\omega) f(\omega) \mathrm{d}\mu(\omega)$$

And for any other $\tilde{N}$ and $\tilde{f}$ also satisfy this, they must be the *same* w.r.t. P and μ:

$$P(N \setminus \tilde{N}) = P(\tilde{N} \setminus N) = 0$$
$$\mu(N \setminus \tilde{N}) = \mu(\tilde{N} \setminus N) = 0$$
$$P(\{x \mid f(x) \neq \tilde{f}(x)\}) = \mu(\{x \mid f(x) \neq \tilde{f}(x)\}) = 0$$

The idea is to split up the integral into two regions: N and N^C. **Absolute continuity** gives the integral on N is 0, **the theorem** gives that the integral on N^C w.r.t. P can be changed into an integral w.r.t. μ with the density f.
Two important target measures are **counting measure** and **Lebesgue measure**. They are both defined on Borel σ−field $\mathcal{B}$, hence the outcome space must be $\mathbb{R}$.

Two dominating measures: m and ν

For a probability P_X on the space $(\mathbb{R}, \mathcal{B}, P_X)$, two most common dominating measures are:

- **Lebesgue measure** m for **continuous** X: for P_X that assigns probability 0 to countable sets of **points** in $\mathbb{R}$.
- **counting measure** ν for **discrete** X: for P_X that assigns probability 0 to sets in $\mathcal{B}$ that **do NOT contain any of the outcomes of the countable image of** X.
  If the image of X is the integers $\mathcal{I}$, the for P_X assigning probability 0 to sets in $\mathcal{R}$ that do NOT containing integers, $\nu(B) = \#B \cap \mathcal{I}$ can be used.
- **Mixed measure** $m + \nu$ for mixed **discrete-continuous** X: image of X is the union of an interval and a countable set, $P_X(B) = 0$ when B is neither an interval nor the target countable set (integers, etc.).

With dominating measures defined, we have the integrals as

For $P_X = \int I_B(x) \mathrm{d}P(x)$

For the three cases, we have:

(a) $P_X \ll m$:

$$P_X(B) = \int_{\mathbb{R}} I_B(x) f_X(x) \mathrm{d}x$$

(b) $P_X \ll \nu$:

$$P_X(B) = \int_{\mathbb{R}} I_B(x) f_X(x) \mathrm{d}\nu(x) = \sum_{B \cap \mathcal{I}} f_X(i) = \sum_{B \cap \mathcal{I}} P_X(\{i\}) = \sum_{B \cap \mathcal{I}} \Pr(X = i)$$

(c) $P_X \ll m + \nu$:

$$P_X(B) = \int_{\mathbb{R}} I_B(x) f_X(x) \mathrm{d}(m + \nu) = \int_{\mathbb{R}} I_B(x) f_X(x) \mathrm{d}x + \int_{\mathbb{R}} I_B(x) f_X(x) \mathrm{d}\nu$$

Where $f_X(x) = \Pr(X = x)$

For the density function $f_X(x) : \mathbb{R} \to \mathbb{R}$ defined on $(\mathbb{R}, \mathcal{B})$, it is actually a **Borel measurable random variable**.

Properties of $f_X(x)$

density function $f_X(x)$ has the following properties:

- non-negative: $\forall x \in \mathbb{R}, f_X(x) \geq 0$
- integrate to 1 If X is continuous:

$$1 = P_X((-\infty, \infty)) = \int_{\mathbb{R}} f_X(x) \mathrm{d}x$$

If X is discrete:

$$1 = P_X((-\infty, \infty)) = \int_{\mathbb{R}} f_X(x) \mathrm{d}\nu = \sum_{i=-\infty}^{\infty} f_X(i) = 1$$

How can we find $f_X(x) = \Pr(X) = x$?