Lecture 1: Probability Measure and Random Variables

Tianpei Xie

Jul. 19th., 2015

Contents

1 Probability Measure

1.1 Definitions

- **Definition** [Resnick, 2013, Billingsley, 2008]
  A ***probability space*** is a triple $(\Omega, \mathscr{F}, \mathcal{P})$ where

  1. Ω is ***the sample space*** corresponding to ***outcomes*** of some (perhaps hypothetical) experiment.
  2. $\mathscr{F}$ is the σ-algebra of subsets of Ω. These subsets are called ***events***.
  3. $\mathcal{P}$ is a ***probability measure***; that is, $\mathcal{P}$ is a function with domain $\mathscr{F}$ and range $[0, 1]$ such that
     (a) ***Non-Negative***: $\mathcal{P}(A) \geq 0$ for all $A \in \mathscr{F}$.
     (b) ***Countably Additive***: If $\{A_n\} \subset \mathscr{F}$ are *disjoint*, then
     $$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathcal{P}\left(A_n\right).$$
     (c) ***Finiteness***: $\mathcal{P}(\Omega) = 1$.

- **Proposition 1.1** *The following properties are important*

  1. ***Complements***: $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$;
  2. $\mathcal{P}(\emptyset) = 0$;
  3. ***Finite subadditivity***: *for any collection of* $\{A_k : 1 \leq k \leq n\} \subset \mathscr{F}$,
     $$\mathcal{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} \mathcal{P}(A_k);$$
  4. ***Monotonicity***: *If* $A \subset B$, *then* $\mathcal{P}(A) \leq \mathcal{P}(B)$;
  5. ***Countably Subadditivity***: *for any collection of* $\{A_k : k \geq 1\} \subset \mathscr{F}$,
     $$\mathcal{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathcal{P}(A_k);$$

- **Remark** [Billingsley, 2008]

  1. Ω is called the *sample space*, and a point $\omega \in \Omega$ is referred as a *sample point.*
  2. Each sample point is associated with an outcome of some experiment. It can be interpreted as a *trigger* from which an experiments start; or a *probe* from which an observation is made.
  3. The σ-algebra $\mathscr{F}$ encodes all possible information conveyed in outcomes of all possible experiments. A measureable set $E \in \mathscr{F}$ is called an *event.* In terms of this, $\mathscr{F}$ is the collection of all possible events associated with all experiments.

4. Each event is associated with a measure of "*possibility volume*", which is a probability measure of that event.

- **Proposition 1.2** ***(Monotone Continuity)*** *[Resnick, 2013, Billingsley, 2008]*
  - *If $A_n \uparrow A$, for $A_n \in \mathscr{F}$, then $\mathcal{P}(A_n) \uparrow \mathcal{P}(A)$;*
  - *If $A_n \downarrow A$, for $A_n \in \mathscr{F}$, then $\mathcal{P}(A_n) \downarrow \mathcal{P}(A)$;*

  Proof: – Suppose $A_1 \subset A_2 \cdots$ and $A = \bigcup_{n=1}^{\infty} A_n$, so $\lim_{n\to\infty} \uparrow A_n = A$. Define $\{B_k\} \subset \mathscr{F}$ such that

$$\begin{aligned} &B_1 = A_1 \\ &B_k = A_k - A_{k-1}; \quad k > 1. \end{aligned}$$

  So

$$B_i \cap B_j = \emptyset; \text{ and } \bigcup_{k=1}^{n} B_k = \bigcup_{k=1}^{n} A_k = A_n.$$

  Therefore,

$$\begin{aligned} \mathcal{P}(A) &= \mathcal{P}\left(\bigcup_{k\geq 1} A_k\right) \\ &= \mathcal{P}\left(\bigcup_{k\geq 1} B_k\right) \\ &= \sum_{k=1}^{\infty} \mathcal{P}(B_k) = \lim_{n\to\infty} \uparrow \sum_{k=1}^{n} \mathcal{P}(B_k) \\ &= \lim_{n\to\infty} \uparrow \mathcal{P}\left(\bigcup_{k=1}^{n} B_k\right) \\ &= \lim_{n\to\infty} \uparrow \mathcal{P}(A_n). \end{aligned}$$

  For the second part, it is similar. ∎

- **Proposition 1.3** ***(Fatou Lemma)*** *[Resnick, 2013, Billingsley, 2008]*

$$\begin{aligned} \mathcal{P}\left(\liminf_{n\to\infty} A_n\right) &\leq \liminf_{n\to\infty} \mathcal{P}(A_n) \\ &\leq \limsup_{n\to\infty} \mathcal{P}(A_n) \\ &\leq \mathcal{P}\left(\limsup_{n\to\infty} A_n\right). \end{aligned}$$

Proof: See that

$$
\begin{aligned}
\mathcal{P}\left(\liminf_{n\to\infty} A_n\right) &= \mathcal{P}\left(\lim_{k\to\infty}\uparrow\left\{\bigcap_{n\ge k} A_n\right\}\right)\\
&= \lim_{k\to\infty}\uparrow\mathcal{P}\left(\bigcap_{n\ge k} A_n\right) \qquad \text{(by monotone continuity)}\\
&\le \liminf_{k\to\infty}\mathcal{P}\left(A_k\right) \qquad \text{(by monotonicity } \mathcal{P}(\bigcap_{n\ge k} A_n) \le \mathcal{P}(A_k))\\
&\le \limsup_{k\to\infty}\mathcal{P}(A_k) \qquad \text{(by definition)}\\
&\le \lim_{k\to\infty}\downarrow\mathcal{P}\left(\bigcup_{n\ge k} A_n\right) \qquad \text{(by monotonicity } \mathcal{P}(\bigcup_{n\ge k} A_n) \ge \mathcal{P}(A_k))\\
&= \mathcal{P}\left(\lim_{k\to\infty}\downarrow\left\{\bigcup_{n\ge k} A_n\right\}\right) \qquad \text{(by monotone continuity)}\\
&= \mathcal{P}\left(\limsup_{n\to\infty} A_n\right) \quad \blacksquare
\end{aligned}
$$

- **Definition** Let $\Omega = \mathbb{R}$, and suppose $\mathcal{P}$ is a *probability measure on* $\mathbb{R}$. Define $F : \mathbb{R} \to [0, 1]$ by

$$F(x) = \mathcal{P}((-\infty, x]), \quad x \in \mathbb{R}.$$

  F satisfies *the following conditions*

  1. F is ***right continuous***,
  2. F is ***monotone non-decreasing***,
  3. F has ***limits*** at $\pm\infty$

$$
\begin{aligned}
F(+\infty) &:= \lim_{x\to+\infty} F(x) = 1\\
F(-\infty) &:= \lim_{x\to-\infty} F(x) = 0
\end{aligned}
$$

  The function F defined above is called a ***(probability) distribution function***. We abbreviate distribution function by *df*.

- **Definition** (***Outer Regularity***) [Folland, 2013]
  Let μ be a ***Borel*** measure on X and E a *Borel subset* of X. The measure μ is called ***outer regular*** on E if

$$\mu(E) = \inf\{\mu(U) : U \supseteq E, U \text{ is open}\}$$

- **Definition** (***Inner Regularity***) [Folland, 2013]
  Let μ be a ***Borel*** measure on X and E a *Borel subset* of X. The measure μ is called ***inner regular*** on E if

$$\mu(E) = \sup\{\mu(C) : C \subseteq E, C \text{ is compact}\}$$

- **Definition** If μ is *outer* and *inner regular* on *all Borel sets*, μ is called ***regular***.

- **Remark** ***Baire measure*** is equivalent to a ***regular Borel measure (Randon measure)*** in the context of ***compact space*** X.

- **Definition** (***Radon Measure***) [Folland, 2013]
  A ***Radon measure*** μ on X is a *Borel measure* that is

  1. ***finite*** on *all* ***compact*** *sets*; i.e. for any ***compact subset*** $K \subseteq X$,
  $$\mu(K) < \infty.$$
  2. ***outer regular*** on *all Borel sets*; i.e. for any *Borel set* E
  $$\mu(E) = \inf\left\{\mu(U) : E \subseteq U, U \text{ is open}\right\}.$$
  3. ***inner regular*** on all *open sets*; i.e. for any *open set* E
  $$\mu(E) = \sup\left\{\mu(C) : C \subseteq E, C \text{ is compact and Borel}\right\}.$$

1.2 Dynkin's π-λ System

- **Remark** (***Beyond*** σ***-Algebra***)
  A σ-algebra is a collection of subsets of Ω satisfying certain closure properties, namely ***closure*** *under* ***complementation*** *and* ***countable union***. We will have need of collections of sets satisfying ***different closure axioms***. We define a *structure* $\mathscr{G}$ to be a collection of subsets of Ω satisfying certain specified closure axioms.

- **Definition** [Resnick, 2013, Billingsley, 2008]

  1. π***-system*** ($\mathscr{G}$ is a π-system, if it is ***closed*** *under* ***finite intersections***: $A, B \in \mathscr{G}$ implies $A \cap B \in \mathscr{G}$).
  2. λ***-system*** (synonyms: σ***-additive class***, ***Dynkin class***): $\mathscr{G}$ contains Ω and is ***closed*** under the formation of ***complements*** and of ***finite*** *and* ***countable disjoint unions***:
     (a) $\Omega \in \mathscr{G}$.
     (b) $A \in \mathscr{G}$ then $A^c = \Omega \setminus A \in \mathscr{G}$
     (c) $A_1, A_2, \ldots \in \mathscr{G}$ and $A_n \cap A_m = \emptyset$ for $m \neq n$ imply
     $$\bigcup_{n=1}^{\infty} A_n \in \mathscr{G}.$$

     Because of the ***disjointness condition*** in (3), *the definition of* λ*-system* is ***weaker*** (more inclusive) than that of σ*-algebra*. Although a σ*-algebra* is a λ*-system*, the ***reverse*** is not true.

- **Lemma 1.4** *[Resnick, 2013, Billingsley, 2008]*
  A class that is ***both*** *a* π***-system*** *and a* λ***-system*** *is a* σ*-algebra.*

- Many ***uniqueness arguments*** depend on the following theorem:

  Theorem 1.5 *(**Dynkin's** π-λ **Theorem**) [Resnick, 2013, Billingsley, 2008]*

  1. *If $\mathscr{P}$ is a π**-system** and $\mathscr{G}$ is a λ**-system**, then $\mathscr{P} \subseteq \mathscr{G}$ implies $\sigma(\mathscr{P}) \subseteq \mathscr{G}$.*
  2. *If $\mathscr{P}$ is a π**-system***
  $$\sigma(\mathscr{P}) = \mathscr{G}(\mathscr{P}),$$
  *that is, **the minimal** σ**-field over** $\mathscr{P}$ equals **the minimal** λ**-system over** $\mathscr{P}$.*

- **Remark** *Dynkin's theorem* is a remarkably flexible device for performing set inductions which is ideally suited to probability theory.

- **Corollary 1.6** *(**Uniquness Condition of Probability Measure**) [Resnick, 2013, Billingsley, 2008]*
  *Suppose that $\mathcal{P}_1$ and $\mathcal{P}_2$ are probability measures on $\sigma(\mathscr{P})$, where $\mathscr{P}$ is a π**-system**. If $\mathcal{P}_1$ and $\mathcal{P}_2$ **agree** on $\mathscr{P}$, then they **agree** on $\sigma(\mathscr{P})$.*

- The following shows that *probability measure* on $\mathbb{R}$ is ***uniquely determined*** by *its distribution function.*

  Corollary 1.7 *[Resnick, 2013, Billingsley, 2008]*
  *Let $\Omega = \mathbb{R}$. Let $\mathcal{P}_1, \mathcal{P}_2$ be two probability measures on $(\mathbb{R}, \mathscr{F}(\mathbb{R}))$ such that their **distribution functions are equal**:*
  $$F_1(x) = \mathcal{P}_1((-\infty, x]) = F_2(x) = \mathcal{P}_2((-\infty, x]), \quad \forall x \in \mathbb{R}.$$
  Then
  $$\mathcal{P}_1 \equiv \mathcal{P}_2$$
  on $\mathscr{F}(\mathbb{R})$.

- **Definition** (***Monotone Classes***)
  A class $\mathscr{M}$ of subsets of Ω is ***monotone*** if it is ***closed*** under the formation of ***monotone unions*** and ***intersections***:

  1. $A_1, A_2, \ldots \in \mathscr{M}$ and $A_n \uparrow A$ imply $A \in \mathscr{M}$;
  2. $A_1, A_2, \ldots \in \mathscr{M}$ and $A_n \downarrow A$ imply $A \in \mathscr{M}$.

- **Theorem 1.8** *(**Halmos's Monotone Class Theorem**) [Resnick, 2013]*
  *If $\mathscr{F}_0$ is a **field** and $\mathscr{M}$ is a **monotone class**, then $\mathscr{F}_0 \subseteq \mathscr{M}$ implies $\sigma(\mathscr{F}_0) \subseteq \mathscr{M}$.*

2 Random Variables

2.1 Pre-image

- **Remark** Suppose Ω and Ω' are two sets. Frequently $\Omega' = \mathbb{R}$. Suppose
  $$X : \Omega \to \Omega'$$

meaning X is a function with domain Ω and range Ω'. Then X determines a ***preimage***

$$X^{-1} : 2^{\Omega'} \to 2^{\Omega}$$

defined by

$$X^{-1}(A') = \left\{\omega \in \Omega : X(\omega) \in A'\right\}$$

for $A' \subseteq \Omega'$. X^{-1} *preserves* complementation, union and intersections.

- **Proposition 2.1** *(σ-**Algebra Preserved by Preimage**) [Resnick, 2013]*
  *If $\mathscr{B}$ is a σ-**algebra** of subsets of Ω', then $X^{-1}(\mathscr{B})$ is a σ-**algebra** of subsets of Ω.*

- **Proposition 2.2** *[Resnick, 2013]*
  If $\mathscr{C}$ is a collection of subsets in Ω', then

$$X^{-1}\left(\sigma(\mathscr{C})\right) = \sigma\left(X^{-1}(\mathscr{C})\right),$$

  *that is, **the pre-image of the σ-algebra** generated by $\mathscr{C}$ in Ω' is the same as the σ-algebra generated by pre-image of $\mathscr{C}$.*

2.2 Measurable Functions as Random Variable

- **Definition** (***Random Element and Random Variables***)
  Given $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{B})$ are two *measureable space*, a map $X : \Omega \to \Omega'$ is a *measurable map* (or $(\mathscr{F}/\mathscr{B})$ *measurable*) if

$$X^{-1}(\mathscr{B}) \subset \mathscr{F}.$$

  X is called a ***random element*** in Ω', and denoted as

$$X \in \mathscr{F}/\mathscr{B},$$
$$\text{or } X : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{B})$$

  If $(\Omega', \mathscr{B}) = (\mathbb{R}, \mathcal{B})$, $\mathcal{B} = \mathscr{B}(\mathbb{R})$ is *Borel σ-algebra* on $\mathbb{R}$, X is called a ***random variable***.

- **Definition** (***Distribution of Random Variable***)
  Given the probability space $(\Omega, \mathscr{F}, \mathcal{P})$ and suppose $X : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{B})$ is measurable, then the set function

$$\mathcal{P}_X \equiv \mathcal{P} \circ X^{-1}$$
$$\Rightarrow \mathcal{P}_X(B) = \mathcal{P}(X^{-1}(B)) \quad \text{for all } B \in \mathscr{B}$$

  is called the ***induced probability*** or ***the distribution for random variable X***.

  Given random variable X, we obtain *an* ***induced probability space*** $(\Omega', \mathscr{B}, \mathcal{P}_X)$ on the image set.

- **Remark** (***Pushforward Measure***)

  Definition For a *continous* map $T : \mathcal{X} \to \mathcal{Y}$, the ***push-forward operator*** is defined as $T_{\#} : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{Y})$ that satisfies

$$\left(T_{\#}\alpha\right)(B) := \alpha\left(\{\boldsymbol{x} : T(\boldsymbol{x}) \in B \subset \mathcal{Y}\}\right) = \alpha(T^{-1}(B))$$

  where the ***push-forward measure*** $\beta := T_{\#}\alpha \in \mathcal{M}(\mathcal{Y})$ of some $\alpha \in \mathcal{M}(\mathcal{X})$, $T^{-1}(\cdot)$ is the *pre-image* of T, and $\mathcal{M}(\mathcal{X})$ is the set of ***Radon measures*** on the space $\mathcal{X}$.

Thus *the* ***distribution of random variable*** X is ***the pushforward measure*** of $\mathcal{P}$ by *random map* X:

$$\mathcal{P}_X = X_{\#}\mathcal{P}.$$

- **Remark** Usually we write

$$\mathcal{P} \circ X^{-1}(B) = \mathcal{P}\left(\{\omega : X(\omega) \in B\}\right) = \mathcal{P}(X \in B)$$

If X is a random variable, $\mathcal{P}_X$ is an *induced probability measure* on $\mathbb{R}$:

$$\mathcal{P} \circ X^{-1}((-\infty, x]) = \mathcal{P}(X \leq x)$$

- **Remark** [Billingsley, 2008]
  - We can interpret each random variable $X : \Omega \to \mathbb{R}$ as the result of a ***random experiment*** whose *outcome measurement* is a real number. When the experiment design is complete, the random variable as a $\mathscr{F}$-measurealbe function is fixed, and the outcome for each run is associated with a specific *sample point* $\omega \in \Omega$.
  - The σ-algebra generated by a random variable X, $\sigma(X)$, encodes ***all possible information*** conveyed by the ***outcome*** *of experiment* X. In communication, where X is the message, all information of the message can be encoded in $\sigma(X)$. The set $[X \in A] \equiv \{\omega : X(\omega) \in A\} \in \sigma(X)$ incorporates all possible realizations whose outcomes lie in A.
  - Moreover, $\sigma(X) \subset \mathscr{F}$ provides a specific structure in $\mathscr{F}$ that is induced by the given random variable X. Here, $\sigma(X) \subset \sigma(X, Z)$ indicates that the there is, in general, finer information structure contained in experiments yielding multiple outcome $(X(\omega), Z(\omega)), \omega \in \Omega$ than those yielding a simple outcome $X(\omega)$. Finer means more detailed information is available to be explored.
  - In terms of this, the overall σ-algebra $\mathscr{F}$ just encode all possible information conveyed by any feasible experiments.
  - The distribution of random variable as an induced probability measure $\mathcal{P} \equiv \mathbb{P} \circ X^{-1}$ is then a measure of all possible outcomes in real $\mathbb{R}$, which is generated by experiments of X. Here $\mathbb{P}$ is a probability measure of event in sample space.

    Note that the induced probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$ contained all information regarding the random experiment. It allow as to "forget" the original space $(\Omega, \mathscr{F}, \mathbb{P})$.
  - Sometime, we can specify a fixed sample point $\omega \in \Omega$ from a given outcome of the experiment X, then for any *event* $E \in \sigma(X)$, we can reveal whether or not $\omega \in E$, but still have no information about the event itself.

- **Proposition 2.3** ***(Test for Measurability)****[Resnick, 2013]*
Suppose

$$X : \Omega \to \Omega'$$

where $(\Omega, \mathscr{F})$, and $(\Omega', \mathscr{B})$ are two measurable spaces. Suppose $\mathscr{C}$ ***generates*** *$\mathscr{B}$; that is*

$$\mathscr{B} = \sigma(\mathscr{C}).$$

Then X is ***measurable if and only if***

$$X^{-1}(\mathscr{C}) \subset \mathscr{B}.$$

- **Corollary 2.4** *(**Special Case of Random Variables**) [Resnick, 2013]*
  The real valued function

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

  is a random variable if and only if

  $$X^{-1}((-\infty, \lambda]) = [X \le \lambda] \in \mathscr{B}, \quad \forall\, \lambda \in \mathbb{R}.$$

2.3 Measurability and Limits

- **Proposition 2.5** *[Resnick, 2013]*
  Let $X_1, X_2, \ldots$ be random variables defined on $(\Omega, \mathscr{F})$. Then
  - $\inf_{n\ge 1} X_n$ *and* $\sup_{n \ge 1} X_n$ *are random variables;*
  - $\liminf\limits_{n\to\infty} X_n$ *and* $\limsup\limits_{n\to\infty} X_n$ *are random variables;*
  - *If* $\lim\limits_{n\to} X_n(\omega)$ *exists for all ω, then* $\lim\limits_{n\to} X_n$ *is a random variable;*
  - *The set on which $\{X_n : n \ge 1\}$ has a limit is measureable; that is,*

    $$\left\{\omega \;:\; \lim_{n\to} X_n(\omega) \;\textit{exists}\right\} \in \mathscr{F}.$$

  Proof:
  - Given that $X_k \in \mathscr{F}/\mathcal{B}$, $k \ge 1$, the event

    $$\left\{\omega : \inf_{n\to\infty} X_n(\omega) \in (-\infty, \lambda]\right\} = \bigcup_{n \ge 1} \{\omega : X_n(\omega) \in (-\infty, \lambda]\} \in \mathscr{F}, \text{for any } \lambda \in \mathbb{R}$$

    since $\{\omega : X_n(\omega) \in (-\infty, \lambda]\} \in \mathscr{F}$.

    Also

    $$\left\{\omega : \sup_{n\to\infty} X_n(\omega) \in (-\infty, \lambda]\right\} = \bigcap_{n \ge 1} \{\omega : X_n(\omega) \in (-\infty, \lambda]\} \in \mathscr{F}, \text{for any } \lambda \in \mathbb{R}.$$

  - The event

    $$\begin{aligned}\left\{\omega : \liminf_{n\to\infty} X_n(\omega) \in (-\infty, \lambda]\right\} &= \left\{\omega : \sup_{k\ge 1} \inf_{n \ge k} X_n(\omega) \in (-\infty, \lambda]\right\} \\ &= \bigcap_{k\ge 1} \bigcup_{n \ge k} \{\omega : X_n(\omega) \in (-\infty, \lambda]\} \in \mathscr{F}, \text{for any } \lambda \in \mathbb{R}.\end{aligned}$$

Similarly,

$$\left\{\omega : \limsup_{n\to\infty} X_n(\omega) \in (-\infty, \lambda]\right\} = \left\{\omega : \inf_{k\geq 1}\sup_{n\geq k} X_n(\omega) \in (-\infty, \lambda]\right\}$$
$$= \bigcup_{k\geq 1}\bigcap_{n\geq k}\{\omega : X_n(\omega) \in (-\infty, \lambda]\} \in \mathscr{F}, \text{for any } \lambda \in \mathbb{R}$$

– If $\lim_{n\to} X_n(\omega)$ exists for all ω, then

$$\lim_{n\to} X_n = \limsup_{n\to\infty} X_n = \liminf_{n\to\infty} X_n,$$

which is a random variable.

– Consider the complement

$$\begin{aligned}\left\{\omega \;:\; \lim_{n\to} X_n(\omega) \text{ exists}\right\}^c &= \left\{\omega \;:\; \limsup_{n\to\infty} X_n(\omega) > \liminf_{n\to\infty} X_n(\omega)\right\}\\
&= \bigcup_{r\in\mathcal{Q}}\left\{\omega \;:\; \limsup_{n\to\infty} X_n(\omega) > r \geq \liminf_{n\to\infty} X_n(\omega)\right\}\\
&= \bigcup_{r\in\mathcal{Q}}\left(\left[\left\{\omega \;:\; \limsup_{n\to\infty} X_n(\omega) \leq r\right\}^c\right]\bigcap\left[\left\{\omega \;:\; \liminf_{n\to\infty} X_n(\omega) \leq r\right\}\right]\right)\\
&= \bigcup_{r\in\mathcal{Q}}\bigcap_{k\geq 1}\left(\bigcup_{n\geq k}\{X_n(\omega) > r\} \cap \bigcup_{n\geq k}\{X_n(\omega) \leq r\}\right)\\
&\in \mathscr{F},\end{aligned}$$

since $\limsup_{n\to\infty} X_n$, $\liminf_{n\to\infty} X_n$ are both measureable. ■

2.4 σ-Algebra Generated by Random Variables

- **Definition** (***σ-Algebra Generated by Random Variable***)
  Let $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a ***random variable***. The ***σ-algebra generated by random variable***, *denoted* $\sigma(X)$, is defined as
  $$\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R})).$$
  Another *equivalent* description of $\sigma(X)$ is
  $$\sigma(X) := \{[X \in A]\,,\; A \in \mathcal{B}(\mathbb{R})\}\,,$$
  where
  $$[X \in A] \equiv X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}\,.$$

- **Remark** ($\sigma(X) =$ ***Information about X in Probability Space***)
  This is the σ-algebra generated by ***information*** *about* X, which is a way of ***isolating*** *that* ***information*** *in the probability space that* **pertains** *to* X.

- **Definition** (***σ-Algebra Generated by Random Element***)
  Suppose
  $$X : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{B})$$
  is a random element. Then we define
  $$\sigma(X) = X^{-1}(\mathscr{B}).$$
  as ***σ-algebra generated by random element***.

- **Remark** (***Measurable with respect to Sub σ-Algebra***)
  $\mathscr{F}' \subset \mathscr{F}$, we say X is ***measureable*** with respect to $\mathscr{F}'$, written $X \in \mathscr{F}'$ if and only if $\sigma(X) \subset \mathscr{F}'$.

- **Definition** (***Smallest σ-Algebra Containing*** $\sigma(X_t)$)
  Let $X_t : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{B})$ for each t in some index set T, then denote
  $$\sigma(X_t, t \in T) = \bigvee_{t \in T} \sigma(X_t)$$
  the ***smallest σ-algebra*** *containing all* $\sigma(X_t)$.

- **Remark** (***Increasing Family of σ-Algebras in Stochastic Process***)
  In *stochastic process theory*, we frequently keep track of ***potential information*** that can be *revealed* to us by *observing the evolution of a stochastic process* by ***an increasing family of σ-algebras***.

  If $\{X_n, n \geq 1\}$ is a ***(discrete time) stochastic process***, we may define
  $$\mathscr{F}_n := \sigma\left(X_1, \ldots, X_n\right), \quad n \geq 1.$$
  Thus, $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ and we think of $\mathscr{F}_n$ as the ***information potentially available at time*** n. This is a way of cataloguing what information is contained in *the probability model*. ***Properties*** of the stochastic process are sometimes expressed in terms of $\{\mathscr{F}_n, n \geq 1\}$.

  For instance, one formulation of ***the Markov property*** is that *the conditional distribution of* X_{n+1} *given* $\mathscr{F}_n$ is the ***same*** as *the conditional distribution of* X_{n+1} *given* X_n.
  $$\mathcal{P}(X_{n+1} | \mathscr{F}_n) = \mathcal{P}(X_{n+1} | X_n)$$

- **Proposition 2.6** *[Resnick, 2013]*
  Suppose X is a random variable and $\mathscr{C}$ is a class of subsets of $\mathbb{R}$. such that
  $$\sigma(\mathscr{C}) = \mathcal{B}(\mathbb{R}).$$
  Then
  $$\sigma(X) = \sigma\left([X \in B] : B \in \mathscr{C}\right).$$
  A special case of this result is
  $$\sigma(X) = \sigma\left([X \leq \lambda], \lambda \in \mathbb{R}\right).$$

- **Example** The followings are $\sigma(X)$ from some special random variables:
  1. For ***constant function*** $X(\omega) = a \in \mathbb{R}$ for all ω, then generated σ-algebra
  $$\sigma(X) = \{\emptyset, \Omega\}.$$
  2. For ***indicator function*** $X(\omega) = \mathbb{1}\{\omega \in A\}$, the generated σ-algebra
  $$\sigma(X) = \{\emptyset, A, A^c, \Omega\}.$$
  Since $X^{-1}(1) = A$, and $X^{-1}(0) = A^c$, so $X^{-1}(B) = \emptyset$, $\{0,1\} \cap B = \emptyset$ and $X^{-1}(B) = \Omega$, $\{0,1\} \subset B$; similarly, $X^{-1}(B) = A$, $\{0,1\} \cap B = \{1\}$ and $X^{-1}(B) = A^c$, $\{0,1\} \cap B = \{0\}$.
  3. If $(X_1, X_2, \ldots)$ is a ***stochastic process***, then
  $$\mathscr{F}_n \equiv \sigma(X_1, \ldots, X_n)$$
  is the σ-algebra generated by collection of subsets (n-dimensional *cylinder sets*)
  $$\left\{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in A'\right\} \in \mathscr{F}, \text{ for } A' \in \mathcal{B}(\mathbb{R}^n).$$
  This collects all information from 0 to n. See that
  $$\sigma(X_1, \ldots, X_n) \subset \sigma(X_1, \ldots, X_n, X_{n+1}).$$

3 Probability Measures on Product Spaces

3.1 Product Spaces

- **Definition** (***Product Space***)
  Let Ω_1, Ω_2 be two sets. Define ***the product space***
  $$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_i \in \Omega_i, i = 1, 2\}$$
  and define *the* ***coordinate*** *or* ***projection maps*** by $(i = 1, 2)$
  $$\begin{aligned} \pi_i : \Omega_1 \times \Omega_2 &\to \Omega_i \\ (\omega_1, \omega_2) &\mapsto \omega_i \end{aligned}$$
  so that If $A \subset \Omega_1 \times \Omega_2$ define
  $$\begin{aligned} A_{\omega_1} &= \{\omega_2 : (\omega_1, \omega_2) \in A\} = \pi_2(A) \subset \Omega_2 \\ A_{\omega_2} &= \{\omega_1 : (\omega_1, \omega_2) \in A\} = \pi_1(A) \subset \Omega_1. \end{aligned}$$
  A_{ω_i} is called ***the section of A at ω_i***.
- **Definition** (***Function on Product Space***)
  Now suppose we have a function X with ***domain*** $\Omega_1 \times \Omega_2$ *and range* equal to some set S. It does no harm to think of S as a *metric space*. Define ***the section of the function X*** as
  $$X_{\omega_1}(\omega_2) = X(\omega_1, \omega_2)$$

so $X_{\omega_1} \circ \pi_2 = X$ for

$$X_{\omega_1} : \Omega_2 \to S.$$

We think of ω_1 as ***fixed*** and ***the section*** is a function *of varying* ω_2. Call X_{ω_1} ***the section of*** X ***at*** ω_1.

- **Lemma 3.1** ***(Sectioning Sets)*** *[Resnick, 2013]*
  Sections of measurable *sets are* ***measurable****. If* $A \in \mathscr{F}_1 \times \mathscr{F}_2$*, then for all* $\omega_1 \in \Omega_1$

$$A_{\omega_1} \in \mathscr{F}_2.$$

- **Corollary 3.2** *[Resnick, 2013]*
  Sections of measurable *functions are* ***measurable****. That is, if*

$$X : (\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2) \to (S, \mathscr{S})$$

  then

$$X_{\omega_1} \text{ is } \mathscr{F}_2\text{-measurable.}$$

3.2 Probability Measure on Product Spaces

- **Definition** (***Transition Function / Transition Kernel***)
  Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces. A map

$$K : \Omega_1 \times \mathscr{F}_2 \to [0, 1]$$

  is called ***a transition function*** ***(or transition kernel)*** if it satisfies the following conditions:

  1. for each ω_1, $K(\omega_1, \cdot)$ is a ***probability measure*** on $\mathscr{F}_2$, and
  2. for each $A_2 \in \mathscr{F}_2$, $K(\cdot, A_2)$ is a $\mathscr{F}_1/\mathcal{B}([0, 1])$-***measurable function***.

- **Proposition 3.3** ***(Joint Probability from Transition Kernel)*** *[Resnick, 2013]*
  Let $\mathcal{P}_1$ *be a* ***probability measure*** *on* $\mathscr{F}_1$*, and suppose*

$$K : \Omega_1 \times \mathscr{F}_2 \to [0, 1]$$

  is a ***transition function****. Then* K *and* $\mathcal{P}_1$ ***uniquely*** *determine a* ***probability*** *on* $\mathscr{F}_1 \times \mathscr{F}_2$ *via the formula*

$$\mathcal{P}(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2)\mathcal{P}_1(d\omega_1)$$

  for all $A_1 \times A_2 \in \mathscr{F}_1 \times \mathscr{F}_2$*. This probability measure on product space* $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2)$ *is called* ***the joint probability****.*

- **Proposition 3.4** ***(Marginal Random Variable)*** *[Resnick, 2013]*
  Let $\mathcal{P}_1$ *be a* ***probability measure*** *on* $(\Omega_1, \mathscr{F}_1)$ *and suppose* $K : \Omega_1 \times \mathscr{F}_2 \to [0, 1]$ *is a* ***transition kernel****. Define* $\mathcal{P}$ *on* $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2)$ *by*

$$\mathcal{P}(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2)\mathcal{P}_1(d\omega_1).$$

Assume

$$X : (\Omega_1 \times \Omega_2, \mathscr{F}_1 \times \mathscr{F}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

and furthermore suppose $X \geq 0$ *or* $X \in L^1(\mathcal{P})$ *is* ***integrable***. *Then*

$$Y(\omega_1) = \int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2).$$

has the properties

1. Y *is well defined.*
2. Y *is* $\mathscr{F}_1$***-measurable***.
3. $Y \geq 0$ *or* $Y \in L^1(\mathcal{P}_1)$ *is* ***integrable***,

and furthermore

$$\int_{\Omega_1 \times \Omega_2} X d\mathcal{P} = \int_{\Omega_1} Y(\omega_1) \mathcal{P}_1(d\omega_1) = \int_{\Omega_1} \left[\int_{\Omega_2} K(\omega_1, d\omega_2) X_{\omega_1}(\omega_2) \right] \mathcal{P}_1(d\omega_1).$$

- **Theorem 3.5** ***(Fubini Theorem)*** *[Resnick, 2013]*
  Let $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2$ *be* ***product measure***. *If* X *is* $(\mathscr{F}_1 \times \mathscr{F}_2)$*-measurable and is either non-negative or* ***integrable*** *with respect to* $\mathcal{P}$, *then*

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} X d\mathcal{P} &= \int_{\Omega_1} \left[\int_{\Omega_2} X_{\omega_1}(\omega_2) \mathcal{P}_2(d\omega_2) \right] \mathcal{P}_1(d\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} X_{\omega_2}(\omega_1) \mathcal{P}_1(d\omega_1) \right] \mathcal{P}_2(d\omega_2). \end{aligned}$$

References

Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

Sidney I Resnick. *A probability path*. Springer Science & Business Media, 2013.