

Notes of Math 733 Probability Theory

Prof. Timo Seppäläinen

YI WEI

Sep 2024

Contents

1 Intro	1
1.1 Probability Space	1
1.2 Independence	6

1 Intro

[Date: Sep 5,2024]

1.1 Probability Space

Setup (Undergraduate level):

Ω sample space: set of all the individual outcomes

\mathcal{F} event space: appropriate collection of subsets of Ω

P : a function on a subsets of Ω , $P(A)$ = the probability of the set (event) A

Axiom 1.1.

$$P\left(\bigcup_k A_k\right) = \sum_k P(A_k) \quad \text{whenever } A_k \text{ is a pairwise disjoint sequence of events}$$

Example 1.1.

1. roll a dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ = power set of Ω = collection of all subset of Ω
2. # of customers to a service station in some fixed time interval

$$\Omega = \mathbb{Z}_{\geq 0}$$

$$\mathcal{F} = \mathcal{P}(\Omega)$$

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k \in \Omega$$

$$P(A) = \sum_{k \in A} e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } A \subseteq \Omega$$

3. Choose uniformly random real number from $[0, 1]$

$$P(x) = 0 \quad \forall x \in [0, 1]$$

if $0 \leq a < b \leq 1$:

$$P([a, b]) = b - a$$

4. Flip a fair coin for infinitely many times, 0 = heads, 1 = tails:

$$\Omega = \{0, 1\}^{\mathbb{Z}_{\geq 0}}$$

$$P\{w : x_1 = a_1, x_2 = a_2, \dots, x_n = a_n\} = 2^{-n} \quad (*)$$

From this: $P\{w\} = 0 \quad \forall w \in \Omega$

Remark. how to prove Ω is uncountable: diagonal principle

Definition 1.1. Let X be a space. A σ -algebra on X is a collection \mathcal{A} of subsets of X that satisfies these properties:

1. $\emptyset \in \mathcal{A}$
2. $A \in \mathcal{A} \implies A^C \in \mathcal{A}$
3. $\{A_k\}_{k=1}^{\infty} \implies \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

And we call (X, \mathcal{A}) is a measurable space.

Definition 1.2. Given (X, \mathcal{A}) A measure is a function $u : \mathcal{A} \rightarrow [0, \infty]$ such that:

1. $P(\emptyset) = 0$
2. $u(\bigcup_k A_k) = \sum_{k=1} u(A_k)$ for a pairwise disjoint sequence $\{A_k\}_k \subseteq \mathcal{A}$

(X, \mathcal{A}, u) is a measure space.

Definition 1.3. If X is a metric space, its Borel σ -algebra \mathcal{B}_X is by definition the smallest σ -algebra containing all the OPEN subsets of X .

Definition 1.4. Lebesgue measure m on \mathbb{R}^d is the measure that satisfies

$$m\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d (b_i - a_i)$$

Definition 1.5. A probability space (Ω, \mathcal{F}, P) is a measure space such that $P(\Omega) = 1$.

Example 1.2. Example of product σ -algebra from example 1.1. 4:

\mathcal{F} = product σ -algebra = samllest σ -algebra that contains all sets of the type

$$\{w : x_1 = a_1, \dots, x_n = a_n\} \quad , n \in \mathbb{Z}_{>0}, a_1, \dots, a_n \in \{0, 1\}.$$

P obtained from Eq. *

Definition 1.6. Let (X, \mathcal{A}) , (Y, \mathcal{B}) be measurable space, and $f : X \rightarrow Y$ be a function.

We say f is a measurable function if:

$$f^{-1}(B) = \{x \in X : f(x) \in B\} \subseteq \mathcal{A}, \quad \forall B \in \mathcal{B}$$

A random variable X is a measurable function:

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

Example 1.3. flip of a fair coin $\Omega = \{w = (x_1, x_2) : x_1, x_2 \in \{0, 1\}\}$, 0 = heads, 1 = tails:

$$X_1(w) = x_1 \quad \text{outcome of the first flip}$$

$$X_2(w) = x_2 \quad \text{outcome of the second flip}$$

We define $Y(w) = X_1(w) + X_2(w) = \#$ of tails in the two flips

The information contained in $Y(w)$ is represented by σ -algebra generated by Y defined as follows:

$$\begin{aligned} \varsigma(Y) &= \{\{Y \in B\} : B \in \mathcal{B}_{\mathbb{R}}\} \\ &= \left\{ \emptyset, \Omega, \{(0, 0)\}, \{(0, 1), (1, 0)\}, \{(1, 1)\} \text{ and the unions of these sets} \right\} \subsetneq \mathcal{F} \end{aligned}$$

[Date: Sep 10, 2024]

1. push-forward: (X, \mathcal{A}, μ) is a measure space, and (Y, \mathcal{B}) is a measurable space. And there is a $f : X \rightarrow Y$. The push-forward of μ is the measure ν on (Y, \mathcal{B}) defined by $\nu(B) = \mu(f^{-1}(B))$

Remark. Check ν is a measure.

2. Absolute continuity: Let μ, λ be measures on (X, \mathcal{A}) . Then μ is absolute continuous w.r.t λ if $\lambda(A) = 0 \implies \mu(A) = 0 \quad \forall A \in \mathcal{A}$.

Remark. $\mu \ll \lambda$. If $\mu \ll \lambda$, then there exists a measurable function $f : X \rightarrow \mathbb{R}_{\geq 0}$ s.t.

$$\mu(A) = \int_A f d\lambda \quad \forall A \in \mathcal{A}$$

This is called Radom-Nikodym derivative $f(x) = \frac{d\mu}{d\lambda}(x)$

Definition 1.7. Let $X : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a random variable. The distribution of X is the $\mu = P \circ X^{-1}$, i.e.,

$$\mu(B) = P\{w \in \Omega : X(w) \in B\} \quad \text{for } B \in \mathcal{B}$$

In short: $P\{X \in B\} = P(X \in B)$

Definition 1.8. The CDF of X is the function F on \mathbb{R} defined by

$$F(x) = P(X \leq x) = \mu(-\infty, x]$$

Definition 1.9. If $\mu \ll$ Lebesgue measure, then X has a density function f which satisfies

$$P(a < X \leq b) = \int_a^b f(x) dx = \mu(a, b] = F(b) - F(a)$$

Remark. A discrete random variable has at most countably many values, and since individual pts have positive probability

$$\mu k = P(X = k) > 0 = \text{leb}\{x\}$$

Then we know $\mu \ll \text{Leb}$ fails and X has no density function.

Definition 1.10. The expectation of a r.v. X is defined by

$$EX = \int_{\Omega} X dP$$

Remark. Abstract Lebesgue integral on (Ω, \mathcal{F}, P)

Definition 1.11. If $A \in \mathcal{F}$ is an event, its indicator random variable is

$$\mathbf{1}_A(w) = \begin{cases} 1, & \text{if } w \in A, \\ 0, & \text{if } w \notin A. \end{cases}$$

We know

$$\begin{aligned} E[\mathbf{1}_A] &= 0 \cdot P\{\mathbf{1}_A = 0\} + 1 \cdot P\{\mathbf{1}_A = 1\} \\ &= P(A) \end{aligned}$$

Example 1.4.

$$X \sim \text{Poisson}(\lambda) \implies E[g(X)] = \sum_{k=0}^{\infty} g(k) \frac{e^{-\lambda} \lambda^k}{k!}$$

$$X \sim \text{Exp}(\lambda) \implies E[g(X)] = \int_0^{\infty} g(x) \lambda e^{-\lambda x} dx$$

Theorem 1.2.

Key result:

$$E[f(X)] := \int_{\Omega} f(X) dP = \int_{\mathbb{R}} f d\mu$$

Here: X is a r.v. on (Ω, \mathcal{F}, P) , $\mu = P \circ X^{-1}$ = distribution of X , $f : \mathbb{R} \rightarrow \mathbb{R}$ is a Borel function
 $f(X(w)) = (f \circ X)(w)$

Proof. 1. $f = \mathbf{1}_B, B \in \mathcal{B}_{\mathbb{R}}$.

Remark. $\int_{\Omega} \mathbf{1}_B(X(w)) P(dw)$ (same as $dP(w)$)

$$\begin{aligned} \int_{\Omega} \mathbf{1}_B(X(w)) P(dw) &= \int_{\Omega} \mathbf{1}_{X^{-1}(B)}(w) dx \\ &= P(X^{-1}(B)) = \mu(B) = \int_{\mathbb{R}} \mathbf{1}_B d\mu \end{aligned}$$

2. $f = \sum_{i=1}^n a_i \mathbf{1}_{B_i}, a_1, \dots, a_n \in \mathbb{R}, B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$

$$\begin{aligned} \int_{\Omega} \sum_{i=1}^n a_i \mathbf{1}_{B_i}(X) dP &= \sum_{i=1}^n a_i \int_{\Omega} \mathbf{1}_{B_i}(X) dP \\ &= \sum_{i=1}^n a_i \int_{\mathbb{R}} \mathbf{1}_{B_i} d\mu \\ &= \int_{\mathbb{R}} \sum_{i=1}^n a_i \mathbf{1}_{B_i} d\mu \end{aligned}$$

3. $f \geq 0, \exists$ simple function $0 \leq f_n$

$$\begin{aligned} \int_{\Omega} f(X) dP &= \lim_{n \rightarrow \infty} \int_{\Omega} f_n(X) dP \quad (M.C.T.) \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n d\mu \\ &= \int_{\mathbb{R}} f d\mu \end{aligned}$$

Remark. $f_n(x) = \sum_{k=0}^{n(2^n-1)} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\}} + n \mathbf{1}_{\{f(x) \geq n\}}$

4. For general $f : \mathbb{R} \rightarrow \mathbb{R} = f^+ - f^-$ Borel function where $f^+, f^- \geq 0$

$$\begin{aligned} \int_{\Omega} f(X) dP &= \int_{\Omega} f^+(X) dP - \int_{\Omega} f^-(X) dP \\ &= \int_{\mathbb{R}} f^+ d\mu - \int_{\mathbb{R}} f^- d\mu \\ &= \int_{\mathbb{R}} f d\mu \end{aligned}$$

□

Example 1.5. 1. $X \sim \text{Poisson}(\lambda)$, μ = distribution of X . We know $\mu(B) = \sum_{k:k \in B} e^{-\lambda} \frac{\lambda^k}{k!} \implies \mu(\mathbb{R} \setminus \mathbb{Z}_{\geq 0}) = 0$. Then we have:

$$\begin{aligned}
 E[e^{-tX}] &= \int_{\mathbb{R}} e^{-tx} \mu(dx) \\
 &= \int_{\mathbb{Z}_{\geq 0}} e^{-tx} \mu(dx) \\
 &= \sum_{k \in \mathbb{Z}_{\geq 0}} \int_{\{k\}} e^{-tx} \mu(dx) \\
 &= \sum_{k \geq 0} e^{-tk} e^{-\lambda} \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} e^{\lambda e^{-t}} \\
 &= e^{\lambda(e^{-t} - 1)}
 \end{aligned}$$

2. $X \sim \text{Exp}(\lambda)$

$$\begin{aligned}
 E[e^{-tX}] &= \int_{\mathbb{R}} e^{-tx} \mu(dx) = \int_{[0, \infty)} e^{-tx} \lambda e^{-\lambda x} dx \\
 &= \lim_{M \rightarrow \infty} \int_{[0, M]} \lambda e^{-(t+\lambda)x} dx = \lim_{M \rightarrow \infty} R \int_0^M \lambda e^{-(t+\lambda)x} dx \\
 &= \lim_{M \rightarrow \infty} \left(-\frac{\lambda}{t+\lambda} \right) e^{-(t+\lambda)x} \Big|_0^M \\
 &= \lim_{M \rightarrow \infty} \left(\left(-\frac{\lambda}{t+\lambda} \right) e^{-(t+\lambda)M} + \frac{\lambda}{t+\lambda} \right) \\
 &= \frac{\lambda}{t+\lambda}
 \end{aligned}$$

[Date: Sep 12, 2024]

1.2 Independence

Perhaps you recall this: events A and B are independent if $P(A \cap B) = P(A)P(B)$

Definition 1.12. Let Ω, \mathcal{F}, P be a probability space.

Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be sub- σ -algebra of \mathcal{F} . (Means each \mathcal{F}_i is a σ -algebra and $\mathcal{F}_i \subseteq \mathcal{F}$)

Then we say $\mathcal{F}_1, \dots, \mathcal{F}_n$ are independent if $\forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \dots, A_n \in \mathcal{F}_n$, then

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

Now r.v.'s X_1, \dots, X_n on Ω are independent if the σ -algebra $\sigma(X_1), \dots, \sigma(X_n)$ are independent.

Equivalently, \forall measurable sets in the range space,

$$P\left(\bigcap_{i=1}^n \{X_i \in B_i\}\right) = \prod_{i=1}^n P\{X_i \in B_i\}$$

Events A_1, \dots, A_n are independent if the r.v.'s $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ are independent.

And arbitrary collection $\{\mathcal{F}_\beta : \beta \in \mathcal{J}\}$ of sub- σ -algebra is independent if \forall distinct $\beta_1, \dots, \beta_n \in \mathcal{J}$, $\mathcal{F}_{\beta_1}, \dots, \mathcal{F}_{\beta_n}$ are independent.

Remark. Fact: X_1, \dots, X_n are independent, then so are $f_1(X_1), \dots, f_n(X_n)$

Remark. Why product?

X, Y discrete r.v.'s. We're interested in the event $\{X = k\}$. Suppose we learn that $Y = m$. We replace P with $P(\cdot, Y = m)$ defined by $P(A|Y = m) = \frac{P(A \cap \{Y=m\})}{P(Y=m)}$

When is $P(X = k) = P(X = k|Y = m)$?

$$\begin{aligned} P(X = k) &= P(X = k|Y = m) \\ \iff P(X = k)P(Y = m) &= P(X = k, Y = m) \end{aligned}$$

We need some notions/tool to check easily if two r.v.'s are independent.

1. Develop a simpler criterion for checking independence of a given collection of r.v.'s.
2. To construct a probability space with desired independent r.v.'s.

Example 1.6. Let X_1, X_2, X_3 be independent *Bernolli*(p) r.v.'s.

$$P(X_i = 1) = p = 1 - P(X_i = 0)$$

Consider the following events:

$$\begin{cases} \{X_1 + X_2 = 1\} \\ \{X_2 + X_3 = 1\} \end{cases}$$

Firstly we have:

$$P(X_1 + X_2 = 1) = P(01) + P(10) = 2p(1 - p) = P(X_2 + X_3 = 1)$$

And we have:

$$P(X_1 + X_2 = 1, X_2 + X_3 = 1) = P(101) + P(010) = p^2(1 - p) + p(1 - p)^2 = p(1 - p)$$

If the two events are independent, we have:

$$P(X_1 + X_2 = 1, X_2 + X_3 = 1) = P(X_1 + X_2 = 1) \cdot P(X_2 + X_3 = 1)$$

$$\begin{aligned}
&\iff p(1-p) = 4p^2(1-p)^2 \\
&\iff p(1-p) = \frac{1}{4}, p = 0, \text{ or } p = 1 \\
&\iff p = \frac{1}{2}, 0, \text{ or } 1
\end{aligned}$$

Theorem 1.3. Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be subcollection of \mathcal{F} , Assume that each \mathcal{A}_i is closed under intersection, which means $(A, B \in \mathcal{A}_i \implies A \cap B \in \mathcal{A}_i)$ and $\Omega \in \mathcal{A}_i$. Assume that the probability $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$. Then the σ -algebra $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$ are independent.

Example 1.7.

The basis of Borel-algebra. $A_i = \{(a, b) : -\infty < a < b < \infty\}$, then $\sigma(A_i) = \mathcal{B}_{\mathbb{R}}$.

Or you can take $(-\infty, b]$

The tool for proving the thm: Dynkin's $\pi - \lambda$ theorem.

Definition 1.13. Let \mathcal{A} be a collection of subset of Ω

1. \mathcal{A} is a π -system if it is closed under intersections.
2. \mathcal{A} is a λ -system if it has the following three properties:

- (a) $\Omega \in \mathcal{A}$
- (b) $\forall A, B \in \mathcal{A} \text{ and } A \subseteq B \implies B \setminus A \in \mathcal{A}$
- (c) If $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq \dots$ and each $A_i \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Theorem 1.4. Suppose \mathcal{P} is a π -system, \mathcal{L} is a λ -system and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$

We use theorem 1.4 to prove theorem 1.3.

Proof of theorem 1.3: Fix $A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$, set $\mathcal{F} = A_2 \cap \dots \cap A_n$

$$\mathcal{L} = \{A \in \mathcal{F} : P(A \cap \mathcal{F}) = P(A)P(\mathcal{F})\}$$

Claim 1.5. $\mathcal{A}_1 \subseteq \mathcal{L}$

proof of claim 1.1. Check that $P(\mathcal{F}) = \prod_{i=2}^n P(A_i)$

Take $A_1 = \Omega$

Let $A_1 \in \mathcal{A}_1$. $P(A_1 \cap \mathcal{F}) = P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) = P(A_1)P(\mathcal{F})$ □

Claim 1.6. \mathcal{L} is a λ -system.

proof of claim 1.2. 1. $\Omega \in \mathcal{A}_1 \subseteq \mathcal{L}$

2. Let $A, B \in \mathcal{L}, A \subseteq B$. We want $B \setminus A \in \mathcal{L}$.

$$P((B \setminus A) \cap \mathcal{F}) = P((B \cap \mathcal{F}) \setminus (A \cap \mathcal{F})) = P(B \cap \mathcal{F}) - P(A \cap \mathcal{F})$$

3. Let $\mathcal{L} \ni A_i \nearrow A$. We want: $A \in \mathcal{L}$

$$P(A \bigcap F) = \lim_{n \rightarrow \infty} P(A_n \bigcap F) \quad \text{because } A_n \bigcap F \nearrow A \bigcap F$$

We've checked that \mathcal{L} is a λ -system. So $\sigma(A_1) \subseteq \mathcal{L}$

□

Thus $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \sigma(\mathcal{A}_1), A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$

We can use the same argument to upgrade each \mathcal{A}_i in turn to $\sigma(\mathcal{A}_i)$. At the end we have the product properties for all members of $\sigma(\mathcal{A}_1), \dots, \sigma(\mathcal{A}_n)$

□

Corollary 1.7. \mathbb{R} -valued r.v.'s X_1, \dots, X_n are independent iff

$$P\left(\bigcap_{i=1}^n \{X_i \leq s_i\}\right) = \prod_{i=1}^n P\{X_i \leq s_i\}$$