Notes of Math 733: Probability Theory
Instructor: Timo Seppäläinen

YI WEI

Sep 2024

Contents

1 Probability Space

[**Date:** Sep 5,2024]

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 $\Omega, P(A) =$ the probability of the set (event) A

Axiom 1.1.

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

Example 1.1.

1. roll a dice: $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega) =$ power set of $\Omega =$ collection of all subset of Ω

2. # of customers to a service station in some fixed time interval

$$\Omega = \mathbb{Z}_{\geqslant 0}$$
$$\mathcal{F} = \mathcal{P}(\Omega)$$
$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k \in \Omega$$

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

3. Choose uniformly random real number from $[0,1]$
   $P(x) = 0 \quad \forall\, x \in [0,1]$
   if $0 \leqslant a < b \leqslant 1$:

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

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

   $$\begin{aligned}&\Omega = \{0,1\}^{\mathbb{Z}_{\geqslant 0}} \\ &P\{w : x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n\} = 2^{-n}\end{aligned} \tag{*}$$

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

   Exercise 1.1. 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. $\varnothing \in \mathcal{A}$
2. $A \in \mathcal{A} \Longrightarrow A^C \in \mathcal{A}$
3. $\{A_k\}_{k=1}^{\infty} \Longrightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

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

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

1. $P(\varnothing) = 0$
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

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

Definition 1.5. A probability space $(\Omega, \mathcal{F}, P)$ 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

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

P obtained from Eq. *

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

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

A random variable X is a measurable function:

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

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:

$$\begin{aligned} X_1(w) &= x_1 \quad \text{outcome of the first flip} \\ X_2(w) &= x_2 \quad \text{outcome of the second flip} \end{aligned}$$

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:

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

[**Date:** Sep 10,2024]

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

   Exercise 1.2. Check v 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 $\lambda(A) = 0 \Longrightarrow \mu(A) = 0 \quad \forall A \in \mathcal{A}$.

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

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

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

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

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

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

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

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

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

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

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

Then we know $\mu \ll$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 $(\Omega, \mathcal{F}, P)$

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.

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

$X \sim Exp(\lambda) \Longrightarrow E[g(X)] = \int_0^\infty g(x)\lambda e^{-\lambda} \mathrm{d}x$

Theorem 1.2.

Key result:

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

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

Proof.

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

   Remark. Notation: $\int_\Omega \mathbf{1}_B(X(w))P(\mathrm{d}w)$ (same as $dP(w)$)

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

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

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

3. $f \geqslant 0, \exists$ simple function $0 \leqslant f_n$

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

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

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

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

$\square$

Example 1.5. 1. $X \sim Possion(\lambda)$, $\mu =$ distribution of X. We know $\mu(B) = \sum_{k:k\in B} e^{-\lambda}\frac{\lambda^k}{k!} \Longrightarrow$ $\mu(\mathbb{R}\backslash\mathbb{Z}_{\geqslant 0}) = 0$. Then we have:

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

2. $X \sim Exp(\lambda)$

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

2 Laws of Large Numbers

[**Date:** Sep 12, 2024]

2.1 Independence

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

Definition 2.1. Let $\Omega, \mathcal{F}, P$ be a probability space.

Let $\mathcal{F}_1, \ldots, \mathcal{F}_n$ 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 $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent if $\forall A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2, \ldots, A_n \in \mathcal{F}_n$, then

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

Now r.v.'s $X_1, \ldots, X_n$ on Ω are independent if the σ-algebra $\sigma(X_1), \ldots, \sigma(X_n)$ are independent. Equivalently, $\forall$ measurable sets in the range space,

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

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

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

Claim 2.1. Fact: $X_1, \ldots, X_n$ are independent, then so are $f_1(X_1), \ldots, f_n(X_n)$

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 $P(A|Y = m) = \frac{P(A \bigcap \{Y=m\})}{P(Y=m)}$

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 2.1. 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 and independent, we have:

$$\begin{aligned} P(X_1 + X_2 = 1, X_2 + X_3 = 1) &= P(X_1 + X_2 = 1) \cdot P(X_2 + X_3 = 1) \\ \iff p(1-p) &= 4p^2(1-p)^2 \\ \iff p(1-p) &= \frac{1}{4},\ p = 0,\ \text{or } p = 1 \\ \iff p &= \frac{1}{2},\ 0,\ \text{or } 1 \end{aligned}$$

Theorem 2.2. Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be subcollection of $\mathcal{F}$, Assume that each $\mathcal{A}_i$ is closed under intersection, which means $(A, B \in \mathcal{A}_i \implies A \bigcap B \in \mathcal{A}_i)$ and $\Omega \in \mathcal{A}_i$. Assume that the probability $P(\bigcap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) \quad \forall A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$. Then the σ-algebra $\sigma(\mathcal{A}_1), \ldots, \sigma(\mathcal{A}_n)$ are independent.

Example 2.2. Collection of sets which can generate 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 theorem: Dynkin's $\pi - \lambda$ theorem.

Definition 2.2. 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) $\forall A, B \in \mathcal{A}$ and $A \subseteq B \implies B \backslash A \in \mathcal{A}$

   (c) If $A_1 \subseteq A_2 \cdots \subseteq A_n \subseteq \cdots$ and each $A_i \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Theorem 2.3. 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 2.3 to prove theorem 2.2.

Proof of theorem 2.2:

Fix $A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$, set $\mathcal{F} = A_2 \bigcap \cdots \bigcap A_n$

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

Claim 2.4. $\mathcal{A}_1 \subseteq \mathcal{L}$.

Proof of Claim 2.4.

Check that $P(F) = \prod_{i=2}^{n} P(A_i)$

Take $A_1 = \Omega$

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

Claim 2.5. $\mathcal{L}$ is a λ-system.

Proof of Claim 2.5.

1. $\Omega \in \mathcal{A}_1 \subseteq \mathcal{L}$

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

$$P\Big((B\backslash A) \bigcap F\Big) = P\Big((B \bigcap F)\backslash(A \bigcap F)\Big) = P(B \bigcap F) - P(A \bigcap F)$$

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

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

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

□

We continue the proof of theorem 2.2:

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

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

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