Probability

Modeling uncertainty

Michael Psenka

Random experiment (random variable)

 $\longrightarrow \{H, T\}$ (70%, 30%)

 $\longrightarrow \{1, 2, 3, 4, 5, 6\}$

Elements of probability theory

Sample space

Definition. *For a "random experiment", the sample space* Ω *is the set of possible outcomes of the random experiment.*

Example

1. Coin: sample space is $\{H, T\}$
2. Die: sample space is $\{1, 2, 3, 4, 5, 6\}$
3. Pair of dice: $\{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$

   e.g. $(2, 4)$

Events

Definition. *An event of a random variable with sample space* Ω *is a subset* $E \subset \Omega$.

2 Die :  $\Omega = \{1,2,\ldots,6\} \times \{1,2,\ldots,6\}$

$\mathbb{P}(i+j \text{ is even})$.

$\downarrow$

$E = \{(1,1), (1,3), (1,5), (3,1), (2,2), \ldots\}$

Distribution

(collection)
"set of events"

Definition. *A distribution* $\pi : \mathcal{P}(\Omega) \to \mathbb{R}$ *over a sample space* Ω *is a function of subsets of* Ω *that satisfies the following properties:*

"π" $:= \mathbb{P}$

1. "*Normalization*": $\pi(\emptyset) = 0, \pi(\Omega) = 1$.
2. "*Monotonicity*": $A \subset B \Rightarrow \pi(A) \leq \pi(B)$.
3. "*Additivity*": $A \cap B = \emptyset \Rightarrow \pi(A \cup B) = \pi(A) + \pi(B)$.

2 die

e.g. $A =$ "both die even" ; $A = \{(2,2), (2,4), (2,6), \dots\}$

$B =$ "1st die even" ; $B = \{(2,2), (2,3), (2,4), \dots\}$

$A \subseteq B$

Principle of Inclusion/Exclusion

$$(\star)\quad \mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{S \subset \{1,\dots,n\} : |S| = k} \mathbb{P}\left(\bigcap_{i \in S} A_i\right) \right)$$

Proof (sketch)

1. Use addition, to show subtraction formula
2. Use add + subtract to show $\mathbb{P}(A \cup B)$ formula
3. Use $\mathbb{P}(A \cup B)$, add + subtract to get $(\star)$

$\pi(A_1 \cup (A_2 \cup \dots \cup A_n)) = \pi(A_1) + \pi(A_2 \cup \dots \cup A_n)$

Measure-theoretic definition

$= \sum_{i=1}^{N} \pi(A_i)$

1 Probability Spaces and Random Variables

Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space. The set Ω is called the *sample space*; its elements are called *outcomes*. The σ-algebra $\mathcal{H}$ may be called the grand *history*; its elements are called *events*. We repeat the properties of the probability measure $\mathbb{P}$; all sets here are events:

1.1 *Norming:* $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1.$

Monotonicity: $H \subset K \quad \Rightarrow \quad \mathbb{P}(H) \leq \mathbb{P}(K).$

Finite additivity: $H \cap K = \emptyset \quad \Rightarrow \quad \mathbb{P}(H \cup K) = \mathbb{P}(H) + \mathbb{P}(K).$

Countable additivity: (H_n) disjointed $\quad \Rightarrow \quad \mathbb{P}(\bigcup_n H_n) = \sum_n \mathbb{P}(H_n).$

Sequential continuity: $H_n \nearrow H \quad \Rightarrow \quad \mathbb{P}(H_n) \nearrow \mathbb{P}(H),$
$H_n \searrow H \quad \Rightarrow \quad \mathbb{P}(H_n) \searrow \mathbb{P}(H).$

Boole's inequality: $\mathbb{P}(\bigcup_n H_n) \quad \leq \quad \sum_n \mathbb{P}(H_n).$

$\mathbb{Z}$

$\mathbb{Z}^{\times 2}$

$\mathbb{Z}^{\times 3}$ ...

$\mathbb{Z}^{\times \mathbb{N}}$ ("$\mathbb{Z}^{\infty}$")

Erhan Çinlar, *Probability and Stochastics* (you will not be tested on this)

Distribution (notes)

"codomain"

Definition. *A distribution* $\mathbb{P}: \mathcal{P}(\Omega) \to \mathbb{R}$ *over a sample space* Ω *is a function of subsets of* Ω, *first defined over the singletons* $\{a\}, a \in \Omega$ *such that:*

$\mathbb{P}(\text{"dice even"}) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6)$

1. $\mathbb{P}(A) = \sum_{a \in A} \mathbb{P}(a)$, "$\sum_{\emptyset} := 0$"
2. "range" $0 \leq \mathbb{P}(a) \leq 1$ *for all* $a \in \Omega$,
3. $\mathbb{P}(\Omega) = \sum_{a \in \Omega} \mathbb{P}(a) = 1$.

Equivalence of definitions

(lecture) $\Longrightarrow$ (notes)

Suppose P satisfies (lecture)

1. Since Ω is finite, any event $E = \bigcup_{i=1}^{N} \{a_i\}$, all disjoint. (is a finite union)
   Thus, $\mathbb{P}(E) = \sum_{i=1}^{N} \mathbb{P}(a_i)$

2. For all E, $\phi \subset E \subset \Omega$. By monotonicity, we get
   $\mathbb{P}(\phi) \leq \mathbb{P}(E) \leq \mathbb{P}(\Omega)$. By normalization $0 \leq \mathbb{P}(E) \leq 1$

3. Normalization.

Equivalence of definitions

(notes) $\Rightarrow$ (lecture) Assume $\mathbb{P}$ has (notes) properties,

1. Normalization assumed

2. $A \subset B$ $A = \bigcup_{i=1}^{N} \{x_i\}$, $B = \bigcup_{i=1}^{N+M} \{x_i\}$.

$\mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(a_i)$. Since $\mathbb{P}(a_i) \geq 0$, $\sum_{i=N+1}^{N+M} \mathbb{P}(a_i) \geq 0$.

Thus, $\mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(a_i) \leq \sum_{i=1}^{N+M} \mathbb{P}(a_i) = \mathbb{P}(B)$

3. $A = \bigcup_{i=1}^{N} \{x_i\}$, $B = \bigcup_{i=N+1}^{M} \{x_i\}$. If disjoint, all x_i's distinct,

and $\mathbb{P}(A \cup B) = \sum_{i=1}^{N+M} \mathbb{P}(x_i) = \sum_{i=1}^{N} \mathbb{P}(x_i) + \sum_{i=N+1}^{N+M} \mathbb{P}(x_i) = \mathbb{P}(A) + \mathbb{P}(B)$

Random variables

(or $(\Omega, \mathbb{P})$)

Definition. *A random variable is a double* $X = (\Omega, \pi)$*, where* Ω *is the sample space of* X*, and* π *is the distribution of* X.

Normal coin: $\Omega = \{H, T\}$, $\mathbb{P}(H) = 0.5$, $\mathbb{P}(T) = 0.5$

Weighted coin: $\Omega = \{H, T\}$, $\mathbb{P}(H) = 0.65$, $\mathbb{P}(T) = 0.35$

2 coin tosses: $\Omega = \{H, T\} \times \{H, T\}$ $\mathbb{P}((H, H)) = 0.25$

Uniform distribution

<u>Def</u> For sample space Ω, the uniform distribution is the unique distribution $\mathbb{P}$ such that $\mathbb{P}(a) = \mathbb{P}(b)$, for all $a, b \in \Omega$

$$\sum_{a \in \Omega} \mathbb{P}(a) = 1 \implies \sum_{a \in \Omega} p = 1$$

$$\implies p \left(\sum_{a \in \Omega} 1 \right) = 1$$

$$\implies p|\Omega| = 1 \implies p = \frac{1}{|\Omega|}$$

$$\mathbb{P}(A) = \sum_{a \in A} \left(\frac{1}{|\Omega|} \right) = \boxed{\frac{|A|}{|\Omega|}}$$

Examples of random variables: balls and bins

10 bins, 20 balls

Assume "uniformly distributed" what is $\mathbb{P}$(no balls in 1^{st} bin)?

Approach

$\Omega = \{1,2,3,\dots,10\} \times \{1,2,3,\dots,10\} \times \dots \times \{1,2,\dots,10\}$ (20 copies)

A: {no balls in bin 1} $= \{(i_1,\dots,i_{20}) \in [10]^{20} : i_k > 1 \ \forall \ k \in [10]\}$

Example: balls and bins

$|A| = 9^{20}$ (9 remaining choices for each) $|\Omega| = 10^{20}$

$\Rightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{9^{20}}{10^{20}} = \boxed{\left(\frac{9}{10}\right)^{20} \simeq 0.122}$

Approach 2: $\Omega = \{$stars & bars reps$\}$, e.g. $||{*}{*}||{*}{*}{*}|{*}| \ldots \in \Omega$

$A = \{$no balls in bin 1$\} = \{$stars & bars w/ no stars to left of bar 1$\}$

→ equivalent to 1 less bar, $|A| = \binom{28}{20} \Rightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \boxed{\frac{\binom{28}{20}}{\binom{29}{20}} \simeq 0.3103}$

$\boxed{\text{Conclusion}}$: Choosing sample space is important, and induces different distributions. Here, if each ball has uniform distribution, Approach 1 is correct.

Cont. on next page

Example: “birthday paradox”

$\boxed{\text{Approach 1}}$ $\mathbb{P}(A) < \mathbb{P}(B)$

$\boxed{A}$ 20 balls $\rightarrow$

$(1,1,1,1,\dots,1)$

$\boxed{B}$ 2 balls $\rightarrow$

$(1,1,2,2,\dots)$

$(2,2,1,1,\dots)$

$\boxed{\text{Approach 2}}$ $\mathbb{P}(A) = \mathbb{P}(B)$

n students, 365 possible birthdays

$\mathbb{P}(\text{no one has the same birthday})$?

$[n] := \{1,\dots,n\}$

$\Omega = [365] \times [365] \times \dots \times [365]$ (e.g. $(2, 365, 101, \dots)$)

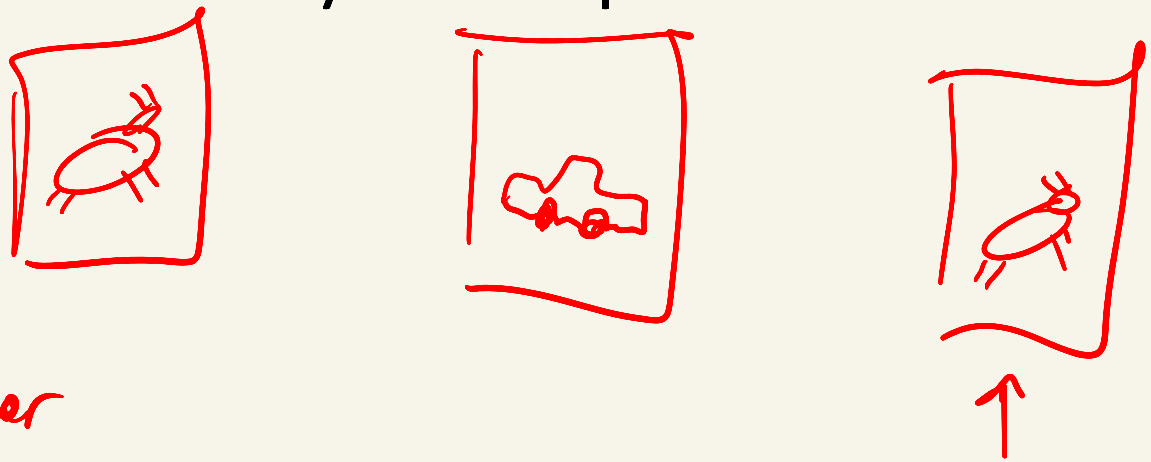
$A = \{a \in \Omega \mid a_i \neq a_j\}$

$|\Omega| = 365^n$, $|A| = \frac{365!}{(365-n)!} = 365 \cdot 364 \cdot \dots \cdot (365-(n-1))$

$\Longrightarrow \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \boxed{\frac{365 \cdot \dots \cdot (365-(n-1))}{365^n}}$

Example: Monty Hall problem

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



1. Choose door

2. Host reveals non-chosen goat door

3. Host asks if you want to switch?

(a, b, c)

which door has car

which did you choose?

which does host reveal?

No switch, $\mathbb{P}(\text{win}) = 1/3$

Switch, $\mathbb{P}(\text{win}) = 2/3$