CMPUT 655

A Short Review on Probability and Statistics

Fall 2022

Outline

- Basic Probability
  - Probability Space
  - Random Variables
  - Expectation and Conditional Expectation
- Basic Statistical Inference
  - I.I.D. Random Sample
  - Bias and variance of an estimator

Useful Resources

- Chapter 2 of Martha’s ML notes
  - https://marthawhite.github.io/mlbasics/notes.pdf
- For a more rigorous introduction, see Chapter 2 of Csaba’s bandit book
  - https://tor-lattimore.com/downloads/book/book.pdf
- For more measure theory, see STAT 571 lecture notes
  - https://sites.ualberta.ca/~kashlak/data/stat571.pdf

Probability Space

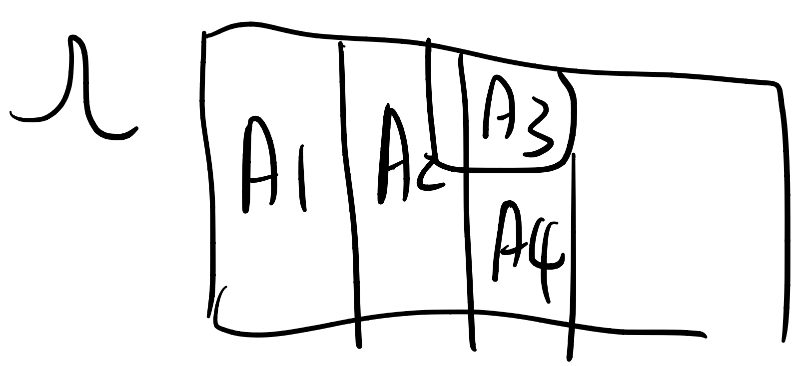
- A probability space $(\Omega, \mathscr{F}, P)$ consists of:
  - a sample space Ω, which is a set of possible outcomes
    - Example: Flip a coin $\Omega = \{H, T\}$
  - an event space $\mathscr{F}$, which is a set of subsets of Ω

    $\mathcal{F} = \{\phi, \Omega, \{H\}, \{T\}\}$
  - a probability measure P that assigns a probability for each element in $\mathscr{F}$

    $P(\phi) = 0, \; P(\Omega) = 1, \; P(\{H\}) = 0.6$

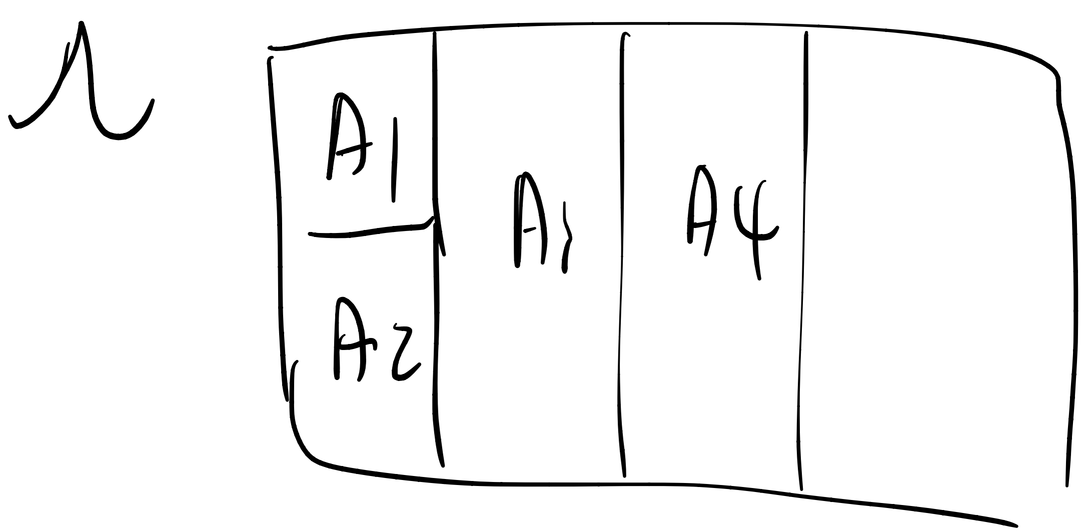
Event Space (σ-field)

- $\mathcal{F}$ is a set of subsets of Ω such that
    - $\emptyset, \Omega \in \mathcal{F}$
    - if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
    - For a countable collection of set $A_1, A_2, \ldots \in \mathcal{F}$, then $\cup_i A_i \in \mathcal{F}$



Probability Measure

- P is a function $P : \mathscr{F} \to [0,1]$ such that
  - $P(\emptyset) = 0$ and $P(\Omega) = 1$
  - For a countable pairwise disjoint collection of sets $A_1, A_2, \ldots \in \mathscr{F}$, then $P(\cup_i A_i) = \sum_i P(A_i)$

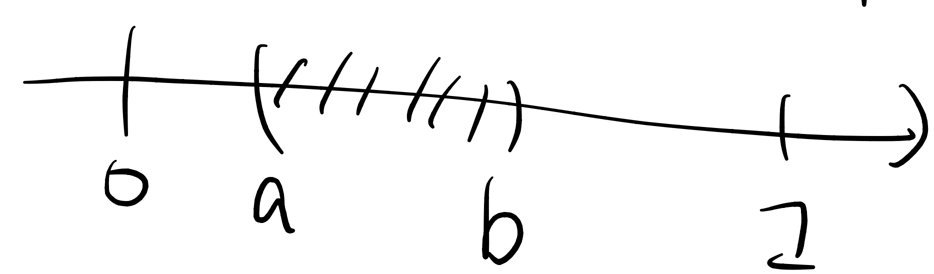


Examples of probability spaces

- Discrete sample space
    - $\Omega = \{1, \ldots, n\}$ $\quad \mathcal{F} = \mathcal{P}(\Omega)$
    - $\mathscr{F} = \{\emptyset, \{1\}, \{2\}, \ldots, \{n\}, \{1,2\}, \ldots, \{1, \ldots, n\}\}$
    - $P(A) = |A|/n$ $\quad P(\{1\}) = 1/n$ $\quad (0,1) \in \mathcal{F}$

$(0,1)^c = \{0,1\} \in \mathcal{F}$

- Continuous sample space
    - $\Omega = [0,1]$ $\quad \mathcal{F} \neq \mathcal{P}(\Omega)$



    - $\mathscr{F} = B(\Omega)$, the smallest σ-field generated by the set of all open intervals (a, b) with $0 \leq a < b \leq 1$ $\quad |\mathcal{F}_1|, |\mathcal{F}_2|, |\mathcal{F}_3|$
    - $P((a, b)) = b - a$ $\quad P(\{x\}) = 0$

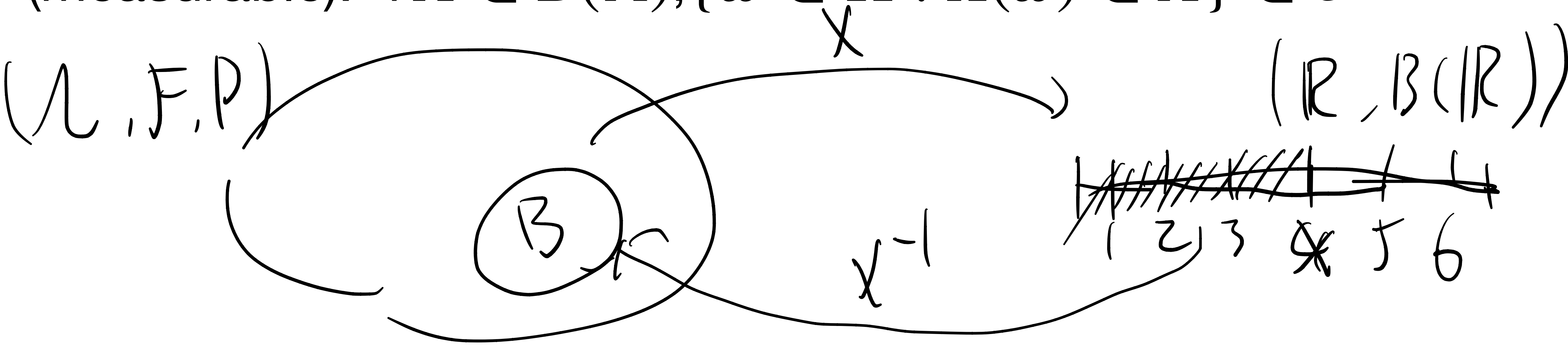
Random Variables

$\frac{1}{2} \circ X + \frac{1}{2} \circ X_{normal}$

- A (real-valued) random variable X is a function $X : \Omega \to \mathbb{R}$ such that

$P(X \leq 1) = 1/6 \quad P(X \leq 2) = 2/6$

  - (measurable): $\forall A \in B(\mathbb{R}), \{\omega \in \Omega : X(\omega) \in A\} \in \mathscr{F}$



- The law of X is $P_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\})$

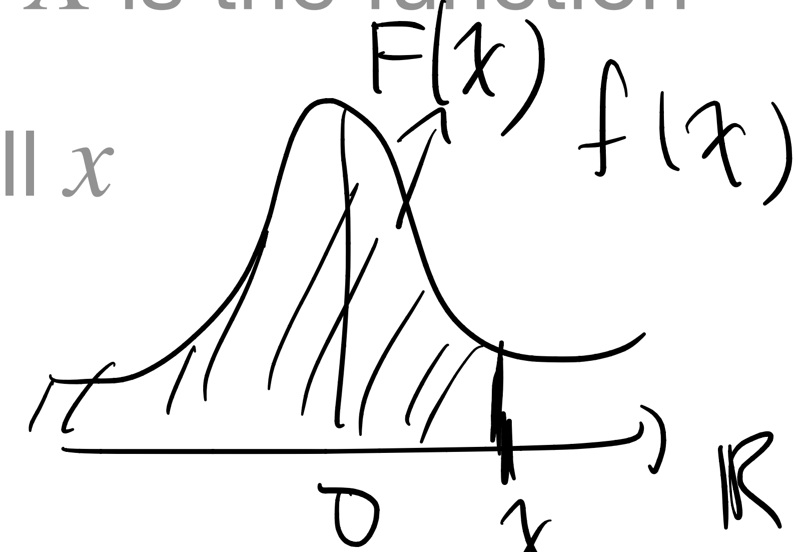
  - Any probabilistic question about X can be answered from P_x. We don't need to know the details of the underlying probability space!!!

CDF, PMF, and PDF

- $F_X(x) = P_X(X \leq x)$ is called the cdf of X

- A discrete random variable X only takes a countable number of different values $x_1, x_2, \ldots$.
  The pmf of X is $f_X(x_i) = P_X(X = x_i)$

- The pdf of a continuous random variable X is the function f_X that satisfies $F(x) = \int_{-\infty}^{x} f_X(t)dt$ for all x

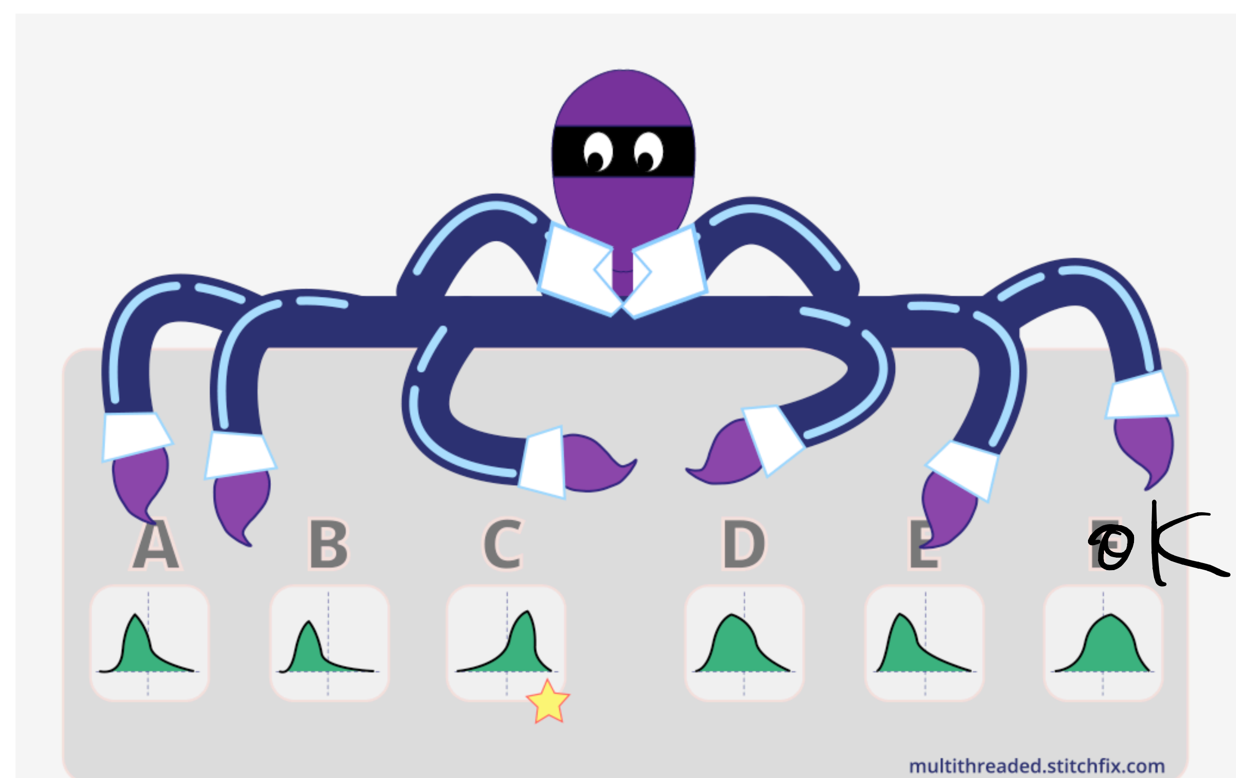


Multi-arm Bandit Example

← a dist. over $\{1, \dots, k\}$

- We choose an arm A_1 from π_1, and receive a reward R_1
- Repeat again, choose A_2 from π_2 and receive R_2
- A_1, R_1, A_2, R_2 are all random variables

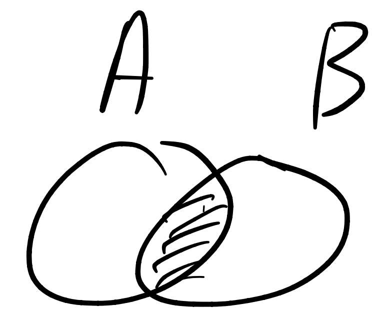
$R_t \sim P_{\underline{\underline{A_t}}}$



K

Independence

- Two events $A, B \in \mathscr{F}$ are independent if $P(A \cap B) = P(A)P(B)$



- Back to our MAB example, R_1, R_2 are independent if
  - $P(R_1 = r_1, R_2 = r_2) = P(R_1 = r_1)P(R_2 = r_2)$ for all r_1, r_2
  - More formally, for any $A \in \sigma(R_1)$ and $B \in \sigma(R_2)$, A and B are independent
  - Question: is R_1, R_2 independent in our example?

Conditional Probability

- Let $A, B \in \mathcal{F}$, the conditional probability is defined as
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- Back to our MAB example,
$$P(R_1 = r_1 \mid A_1 = a_1) = \frac{P(R_1 = r_1, A_1 = a_1)}{P(A_1 = a_1)}$$

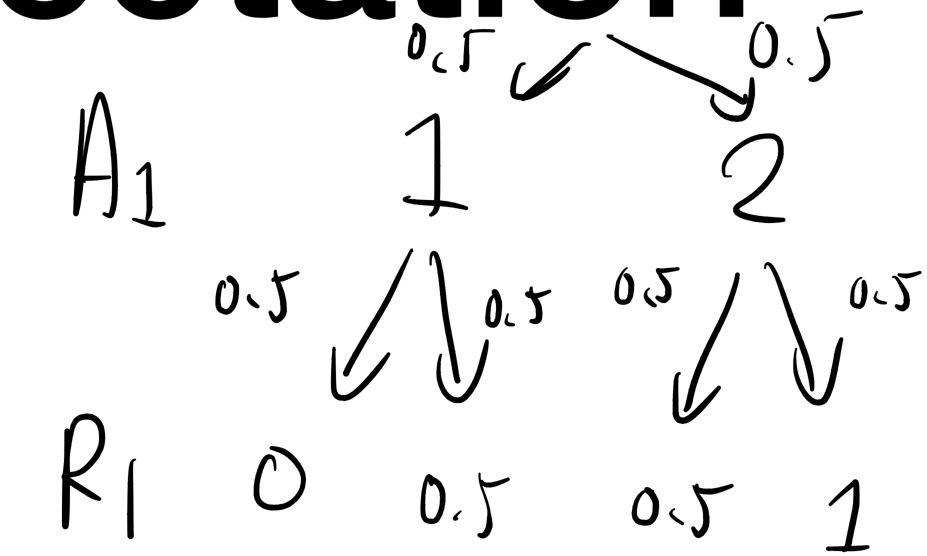
Expectation

- For a discrete random variable
  $$\mathbb{E}[X] = \sum_i x_i \underbrace{P(X = x_i)}_{\text{pmf}}$$

- For a continuous random variable
  $$\mathbb{E}[X] = \int x \underbrace{f_X(x)}_{\text{pdf}} dx$$

- Back to our MAB example, $\mathbb{E}[R_1] = \sum_i r_i P(R_1 = r_i)$

Conditional Expectation



- For discrete random variables X, Y, $\mathbb{E}[X \mid Y = y] = \sum_i x_i P(X = x_i \mid Y = y)$

  - Conditional expectation $\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}$ is a random variable given by $\mathbb{E}[X \mid Y](\omega) = \mathbb{E}[X \mid Y = Y(\omega)]$

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

- For continuous random variables, $\mathbb{E}[X \mid Y = y] = \int x f(x \mid y) dx$

$$\frac{P(R_1 = 0, A_1 = 1)}{P(A_1 = 1)} = \frac{0.5 \cdot \cancel{0.5}}{\cancel{0.5}}$$

- Back to our MAB example, $\mathbb{E}[R_1 \mid A_1 = a_1] = \sum_i r_i P(R_1 = r_i \mid A_1 = a_1)$

$$\|$$

$$q(a_1)$$

$$= 0 \, P(R_1 = 0 \mid A_1 = 1) + 0.5 \, P(R_1 = 0.5 \mid A_1 = 1)$$

$$= 0\left(\frac{1}{2}\right) + 0.5\left(\frac{1}{2}\right) = 0.25$$

Basic Statistical Inference

Random Sample from a distribution

- Let $X_1, \ldots, X_n$ be i.i.d. *(independent and identically distributed)* random variables sampled from a distribution F
    - What does i.i.d. sample mean?

Random Sample from a distribution

- Let $X_1, \ldots, X_n$ be i.i.d. *(independent and identically distributed)* random variables sampled from a distribution F
  - What does i.i.d. sample mean?

    $P(X_1 = x_1 \ldots X_n = x_n)$

    $= \prod_i P(X_i = x_i)$
    - $X_1, \ldots, X_n$ are mutually independent
    - Each X_i has the same distribution F

Estimating the mean of the distribution

- Suppose the distribution has mean μ and variance σ^2. Both are unknown. We want to estimate μ from the random sample $X_1, \ldots, X_n$.

  $\{ f(x;\theta) : \theta \in \Theta \}$

  - What can we do?

Estimating the mean of the distribution

- Suppose the distribution has mean μ and variance σ^2. Both are unknown. We want to estimate μ from the random sample $X_1, \ldots, X_n$.
    - What can we do?
- The sample average estimate is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
    - How good is the estimator?

Bias and Variance

X_i r.v.s a_i constant $\mathbb{E}(\sum a_i X_i) = \sum a_i \mathbb{E}(X_i)$

- The bias of the estimator is defined as $\mathbb{E}[\bar{X}] - \mu$

$$\mathbb{E}[\bar{X}] - \mu = \mathbb{E}[\frac{1}{n}\sum_i X_i] - \mu = \frac{1}{n}\sum_i \mathbb{E}[X_i] - \mu = \frac{1}{n}\sum_i \mu - \mu = 0$$

$\rightarrow \bar{X}$ is an unbiased estimator of μ

- The variance is defined as $\mathbb{V}(\bar{X}) = \mathbb{E}[(\bar{X} - \mathbb{E}\bar{X})^2]$

$$\mathbb{V}(\bar{X}) = \frac{1}{n^2}\sum_i \mathbb{V}(X_i) = \frac{\sigma^2}{n}$$

$V(\sum a_i X_i) = \sum a_i^2 V(X_i)$

$MSE(\bar{X})$

$= bias(\bar{X})^2 + V(\bar{X})$

Questions?