Brief Theory of Probability: Notes from MATH 431

Compiled by Harry Luo

Contents

1 Sample Spaces, collection of events, probability measure

- Sample space Ω: set of all possible outcomes of an experiment. Comes in n-tuples where n represents number of repeated trials.
- Collection of events $\mathcal{F}$: subset of state space to which we assign a probability.
- Probability measure: function that assigns a probability to each event.$P : F \to \mathbb{R}$.
  - Range is $[0, 1]$.
  - $P(\Omega) = 1$ and $P(\emptyset) = 0$
  - For pairwise disjoint events $A_1, A_2, \ldots$, $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

---

2 Sampling: Uniform, Replacement, Order

- uniform sampling: each outcome is equally likely
- Binomial coeff

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{1}$$

2.1 Replacement

- ex: sample K distinct marked balls from N balls in a box, **with** Replacement

$$\begin{gathered} \Omega = \{1, 2, 3, \ldots, N\}^K \\ \|\Omega\| = N^K \\ P(\text{none of the balls is marked 1}) = \frac{(N-1)^K}{N^K} \end{gathered} \tag{2}$$

- ex: sample K distinct marked balls from N balls in a box, **without** Replacement

$$\begin{gathered} \Omega = \{(i_1, i_2, \ldots, i_K) \mid i_1, \ldots, i_K \in \{1, 2, \ldots, N\}, \text{distinct} \\ \|\Omega\| = \binom{N-1}{K} \\ P(\text{none of the balls is marked 1}) = \frac{\binom{N-1}{K}}{\binom{N}{K}} = \frac{N-K}{N} \end{gathered} \tag{3}$$

2.2 Order

- order matters: $A_n^k = \frac{n!}{(n-k)!}$
- order doesn't matter: $\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$

---

3 Infinite Sample Spaces

3.1 discrete

$$\Omega = \{\infty, 1, 2, \ldots\} \tag{4}$$

3.2 continuous

$$\begin{gathered} P([a', b']) = \frac{\text{length of } [a', b']}{\text{length of} [a, b]} \\ \text{single point, or sets of points: } P(\{x\}) = P(\cup_{i=1}^{\infty} \{x_i\}) = 0 \end{gathered} \tag{5}$$

- Complements: $P(A) = 1 - P(A^C)$

---

4 Conditioinal Probability, Law of Total Prob., Bayes' Theorem, Independence

4.1 Conditional prob.

$$P(A|B) = \frac{|A \cap B|}{|B|} \Rightarrow P(AB) = P(B)P(A|B) \tag{6}$$

(new sample space is B, total number of outcomes is $A \cap B$)

4.2 Law of total probability:

Given partitions $B_1, B_2, \ldots$ of Ω,

$$P(A) = \sum_i P(A|B_i)P(B_i) \tag{7}$$

4.3 Bayes' Theorem:

Given events A, B, P(A) and P(B) >0,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} \tag{8}$$

Considering the law of total prob., the generalized form, when B_i are partitions, is given as:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)} \tag{9}$$

4.4 Independence:

$$P(AB) = P(A)P(B) \Leftrightarrow P(B|A) = P(B) \tag{10}$$

Note: By virtue of conventions, we write $A \cap B$ as AB in Probability.
If A,B,C,D are independent, it follows that $P(ABCD) = P(A)P(B)P(C)P(D)$; however, the inverse is not always true.

- Independence of Random Variables (messy as hell...)

Given 2 random variables

$$\begin{gathered} X_1 \in \{x_{11}, x_{12}, x_{13}, \ldots, x_{1m}\} \\ X_2 \in \{x_{21}, x_{22}, x_{23}, \ldots, x_{2n}\} \\ \text{Random variables X\_1 and X\_2 are independent} \Leftrightarrow \\ P(X_1 = x_{1i}, X_2 = x_{2j}) = P(X_1 = x_{1i})P(X_2 = x_{2j}) \end{gathered} \tag{11}$$

Need to check n*m equations to verify independence.

4.5 Conditional Independence:

For events $A_1, A_2, \ldots, A_n, B$, any set of events in A: A_{i1}, A_{i2}, A_{i3}, they are conditionally independent given B if

$$P(A_{i1}A_{i2}A_{i3}|B) = P(A_{i1}|B) * P(A_{i2}|B) * P(A_{i3}|B) \tag{12}$$

5 Independent Trials, Distributions

5.1 Bernoulli dirtribution:

a single trial, with success probability p, and failure probability 1-p. Prameter being the success probability.

$$X \sim \text{Ber}(p) \Rightarrow P(X = x) = p^x * (1 - p)^{1-x}, x \in \{0, 1\} \tag{13}$$

5.2 Binomial Distribution:

multiple independent Bernoulli trials, with success probability p, and failure probability 1-p. Parameters being the number of trials n and the success probability p.

$$X \sim \text{Bin}(n, p) \Rightarrow P(X = k) = \binom{n}{k} p^k * (1 - p)^{n-k}, k \in \{0, 1, ..., n\} \tag{14}$$

5.3 Geometric distribution:

multiple independent Bernoulli trials with success probability p, while stoping the experiment at the first success.

$$X \sim \text{Geom}(p) = p * (1 - p)^{k-1}, k \in \{1, 2, ...\} \tag{15}$$

5.4 Hypergeometric distribution:

There are N objects of type A, and $N_A - N$ objects of type B. Pick n objects without replacement. Denote number of A objects we picked as k. Parameters are N, N_A, n.

$$P(X = k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}} \tag{16}$$

choose k from N_A, choose n-k from N-N_A, divide by total number of ways to choose n from N

6 Random Variables

6.1 Discrete random variable

Discrete random variables are random variables that can take on a countable number of values. It comes naturally from discrete, finite or infinitly countable sample spaces. (As briefly discussed in Section 3.1)

For $A = \{k_1, k_2, ..., \}$ s.t. random variable$X \in A$, or $P(X \in A) = 1$, X is a random variable, with possible values $k_1, k_2, ...$ and $P(X = k_n) > 0$

6.1.1 Probability Mass Function (pmf)

The PMF is a function that defines the probability distribution for a discrete random variable. It gives the probability of the random variable taking on each possible value. The PMF, denoted as

$$p_X(k) = P(X = k) \text{ , where } k \text{ are possible values of X} \tag{17}$$

It is a function of k, and

$$p_X : S \to [0, 1], \tag{18}$$

where:

S is the support set, i.e., the set of all possible values that the discrete random variable X can take. [0, 1] represents the range of the function, as probabilities are always between 0 and 1. For each value k in the support set S, the PMF assigns a probability p_X(k), which represents the likelihood of the random variable X taking the value k.

The PMF satisfies the following properties:

Non-negativity: $p_{X(k)} \geq 0$ for all k in S.
Total probability: $\sum_k p_{X(k)} = 1$ where the sum is taken over all k in S.

Example: For a fair six-sided die, the PMF would be $P(X = x) = \frac{1}{6}$ for $x = 1, 2, 3, 4, 5, 6$. Or more elegantly,

$$p_X(k) = \frac{1}{6}, \text{for every } k \in \{1, 2, 3, 4, 5, 6\} \tag{19}$$

6.2 continuous Random Variables

Not rigorously defined in this class, but a continuous random variable is one that can take on any value in a range. The probability of a continuous random variable taking on a specific value is 0. It came natually from continuous sample spaces.The probability is assigned to intervals of values, and they are assigned by the **probability density function**.

6.2.1 Probability Density Function (pdf)

continuous r.v are defined in this class by having a probability density function. A random variable X is continuous if there exists a function f(x) such that

$$\begin{gathered} \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1, f(x) > 0 \text{ everywhere} \\ \text{and } P(X \leq b) = \int_{-\infty}^{b} f(x)\,\mathrm{d}x \Leftrightarrow P(a \leq X \leq b) = \int_{a}^{b} f(x)\,\mathrm{d}x \end{gathered} \tag{20}$$

6.2.2 Cumulative Distribution Function (cdf)

cdf of a r.v. is defined as

$$F(x) = P(X \leq x) \tag{21}$$

and it follows that

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) \tag{22}$$

- Continuous r.v.

  it looks suspiciously like an indefinite integral, and when we are dealing with continuous r.v., it is.

  $$\boxed{F(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)\,\mathrm{d}x}$$

  Recall the fundamental theorm of calculus,

  $$F'(x) = f(x), \tag{23}$$

  so the pdf is the derivative of the cdf.

- Discrete r.v.

pmf and cdf is connected by

$$F(x) = P(X \leq s) = \sum_{k \leq x} p_{X(k)} \tag{24}$$

where the sum is taken over all k such that $k \leq x$.
In english, the cdf is the sum of the pmf up to the value x, or "compound probability thus far"

If the cdf graph is stepped (piecewise constant), it is a discrete r.v. If it is continuous except at several points, it is a continuous r.v.

6.3 continuous Distribution

Based on different pdf, we have different behaviors of random variables. We call them distributions.

6.3.1 Uniform Distribution

r.v. X has the uniform distribution on the interval [a,b] if its pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ for } a \leq x \leq b \\ 0 \text{ otherwise} \end{cases} \tag{25}$$

TODO(MORE TO COME, ORDER OF BOOK WEIRD)