Topics in Physical Chemistry and Biophysics

1 Review of probability

18.04.23 lec 1

Definition 1

Probability. If N is the total number of outcomes, and n_A fall in category A, then

$$P_A = \frac{n_A}{N} = \frac{\text{outcomes cat.} A}{\text{all outcomes}}$$

Rules of composite events:

1. Mutually exclusive: outcomes $(A_1, A_{2,\ldots})$ are *mutually exclusive* if one outcome precludes another outcomes. (Event A_1 prevents even A_2 from happening simultaneously.)
2. Collectively exhaustive: if all known outcomes are also all possible outcomes. $\sum P_i = 1$.
3. Independence: outcomes do not depend on each other.
4. Multiplicity: total number of ways in which outcomes occur.

Rules of calculation:

1. Let there be 3 outcomes A, B, C with probability P_A, P_B, P_C. What is the probability that either one occurs (A or B or C)?

$$P(A \cup B \cup C) = P_A + P_B + P_C$$

   That's the addition rule.

2. Probability that all outcomes occur? (Assuming independence)

$$P(A \cap B \cap C) = P_A\, P_B\, P_C$$

3. Probability that an event A is not happening? $P = 1 - P_A$

Example 1. We roll a die twice. What is the probability of rolling a 1 first **or** a 4 second?

Split the problem to parts. Note that the events are not mutually exclusive. Condition applies if:

- 1 first and not a 4 second: $\frac{1}{6} \cdot \frac{5}{6}$
- not a 1 first and a 4 second: $\frac{5}{6} \cdot \frac{1}{6}$
- 1 first and 4 second: $\frac{1}{6} \cdot \frac{1}{6}$

Now sum up all of the options to get result.

Definition 2

Correlated events. $P(B|A)$ is the probability that B occurs given A has occured.

Joint probability. $P(AB)$ that both A and B occur.

Definition 3

General multiplication rule.

$$P(AB) = P(B|A)\, P(A)$$

$P(A)$ is called the a priori probability and $P(B|A)$ is called the a posterior probability

Theorem 1

Bayes theorem.

$$P(B|A)\, P(A) = P(A|B)\, P(B)$$

Example 2. 1% of population has breast cancer. We use mammography to detect cancer.

Event A: breast cancer. $P(A) = 0.01$. $P(\bar{A}) = 1 - P(A) = 0.99$.

Event B: diagnosis. $P(B|A) = 0.8$. $P(B|\bar{A}) = 0.096$. (i.e. false positive0

What is the chance that a doctor has diagnosed someone with cancer? i.e. $P(A|B)$

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

$P(B)$ is the diagnosis of breast cancer irrespective wheter it's there or not there.

$$P(B) = P(BA) + P(B\bar{A}) = P(B|A)\, P(A) + P(B|\bar{A})\, P(\bar{A}) = 0.8 \cdot 0.01 + 0.096 \cdot 0.99 = 0.103$$

$$P(A|B) = \frac{0.8 \cdot 0.01}{0.103} = 0.078 = 7.8\%$$

The reason that $P(A|B)$ is so small is that the rate of false positive is really low and the rate of having breast cancer is really low.

Combinatorics: concerned with composition of events, and not their order.

Example 3. How many combinations of there are of N amino acids?

$$W = N! = N\,(N-1)\,(N-2)\cdots$$

Example 4. Distinguish or not Distinguish: What are the possible number of ways to arrange N amino acids? Divide all permutations (assuming objects are distinguishable) by the number of permutations of objects that are indistinguishable.

$$W = \frac{N!}{N_A}$$

In general, for N objects consisting of t categories in which the objects are indistinguishable:

$$W = \frac{N!}{(n_1!)\,(n_2!)\cdots(n_t!)}$$

So, if $t = 2$, (e.g. possible number of ways to arrange three acids A,A,H)

$$W = \frac{N!}{n_1! \cdot n_2!} = \frac{N!}{n_1!\,(N - n_1)!} = \binom{N}{n}$$

Definition 4

Distribution functions. Describe collections of probabilities. Relevant for continuous variables.

$$\sum_i p_i \rightarrow \int_a^b p(x)\,\mathrm{d}x$$

Popular distributions:

1. Binomial Distribution. Relevent when there are only two outcomes.

   Example 5. What is the probability that a series of N trials has n_H heads and n_T tails in any order?

   P_H, P_T are mutually exclusive, so the probability of one sequence is

   $$P_H^{n_H} \cdot P_T^{n_T} = P_H^{n_H}\,(1 - P_H)^{N - n_H}; \quad N = n_H + n_T$$

   and the number of ways to arrange the coins is

   $$W = \frac{N!}{n_H!\,(N - n_H)!}$$

   therefore, the possibility for the outcome (getting n_H and n_T) in any order is

   $$\boxed{P(n_H, N) = \binom{N}{n_H} p_H^{n_H}\,(1 - p_H)^{N - N_H}}$$

   that's the binomial distribution.

   Example 6. Given the molecule $C_{27}H_{44}O$ such that 1.1% is ^{13}C and the rest are ^{12}C, the fraction of molecules without a single ^{13}C is given by the binomial distribution.

2. Multinomial distribution. Basically the extension of the binomial distribution.

   $$\boxed{P(n_1, n_2, \ldots, n_t, N) = \left(\frac{N!}{n_1!\,n_2! \cdots n_t!}\right) p_1^{n_1}\,p_2^{n_2} \cdots p_t^{n_t}}$$

Definition 5

Moments of distributions. Averages and Variances of distribution functions.

Given $p(i)$ s.t. $\sum_i p_{(i)} = 1$, the **Average** is defined as

$$\langle i \rangle = \sum_i i\, p(i) \rightarrow \langle x \rangle = \int x\, p(x)\, \mathrm{d}x$$

Given $f(x)$,

$$\langle f(x) \rangle = \int f(x)\, p(x)\, \mathrm{d}x$$

Given $a \in \mathbb{R}$

$$\langle a\, f(x) \rangle = \int a\, f(x)\, p(x)\, \mathrm{d}x = a\, \langle f(x) \rangle$$

Given 2 functions $f(x), g(x)$,

$$\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle$$

$$\langle f(x) \cdot g(x) \rangle \neq \langle f(x) \rangle\, \langle g(x) \rangle$$

The 2nd and 3nd **Moments** of the distributions $p(x)$ are

$$\langle x^2 \rangle = \int x^2\, p(x)\, \mathrm{d}x$$

$$\langle x^3 \rangle = \int x^3\, p(x)\, \mathrm{d}x$$

The **Variance** of the distribution, σ^2 is defined as

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle$$