Probability

Saurish Chakrabarty
Department of Physics, Acharya Prafulla Chandra College
(Dated: June 11, 2025)

Syllabus (part of PHSDSC405T: Mathematical Methods II)
Definition, idea of sample space, probability addition and multiplication theorems, probability calculation by counting, probability density of random variables: Probability distribution functions

Reference
J. Mathews and R. L. Walker, *Mathematical Methods of Physics*

I. INTRODUCTION

If a fair coin is thrown N times, then we get $N_h \approx \frac{N}{2}$ heads when $N \gg 1$. This ratio, i.e., $\lim_{n\to\infty} \frac{N_h}{N}$ is said to be the *probability* of getting a head in a single toss. Such an approach gives us the *posteriori* probability.
Definition 1: If an experiment is performed N times and we get N_s successes, then $\lim_{n\to\infty} \frac{N_s}{N}$ is called the probability of a success.

Predicting probabilities before performing an experiment (*a priori* approach):
Definition 2: If an experiment has N possible *equally likely* outcomes and N_s of them result in success, then the probability of success is $\frac{N_s}{N}$.

If the two definitions give different answers, then the outcomes which were thought to be equally likely were actually not.

II. BASIC LAWS

In general, any outcome of an experiment is known as an *event*.

If A is an event, then $0 \le P(A) \le 1$. If $P(A) = 0$, A is an impossible event and if $P(A) = 1$ it is a certain event.

Consider an experiment with n equally likely outcomes. Let A and B be two events in this experiment.
n_1 = number of outcomes in which A occurs but not B, denoted by $A - B$.
n_2 = number of outcomes in which B occurs but not A.
n_3 = number of outcomes in which both A and B occur, denoted by $A \cap B$ or just AB.
n_4 = number of outcomes in which neither A nor B occur, denoted by $\overline{A \cup B}$ or just $\overline{A + B}$.
Then,

$$P(A) = \frac{n_1 + n_3}{n}$$

(and a similar expression for $P(B)$). The probability of either A or B or both occurring is denoted by $A \cup B$ or $A + B$ and,

$$P(A + B) \equiv P(A \cup B) = \frac{n_1 + n_2 + n_3}{n}$$

. The probability of both A and B occurring simultaneously is,

$$P(AB) \equiv P(A \cap B) = \frac{n_3}{n}$$

.
Conditional Probability: Probability of A occurring, given that B has occurred is denoted by $A|B$ and,

$$P(A|B) = \frac{n_3}{n_2 + n_3}$$

.
Thus,

$$P(A+B) = P(A) + P(B) - P(AB),$$

and,

$$P(AB) = P(A)P(B|A).$$

These expressions are the *laws of addition and multiplication* of two events.

Ex: Calculate the probability of getting two aces when one card is drawn from each of two decks of cards. (Ans: $\frac{25}{169}$)

Ex: Calculate the probability of getting two hearts when two cards are drawn from a deck of cards. (Ans: $\frac{1}{17}$)

Definition: Two events are *mutually exclusive* if they cannot occur simultaneously, i.e., if $P(AB) = 0$, then A and B are mutually exclusive.

Definition: Two events, A and B are *independent* if $P(AB) = P(A)P(B)$, i.e., $P(B|A) = P(B)$.

Law of Total Probability

Suppose $\{B_1, B_2, \dots\}$ is a set of mutually exclusive and exhaustive events. Then,

$$P(A) = \sum_i P(AB_i) = \sum_i P(A|B_i)P(B_i)$$

Bayes' Theorem

This is a useful theorem with diverse applications which follows from the expression of joint probability of two events, $P(AB)$, given above. It can be expressed in many forms.

Version 1: Since $P(AB) = P(BA)$

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

Version 2: For three events, A, B and C, it is easy to show that,

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

Version 3: If $\{B_1, B_2, \dots\}$ is a set of mutually exclusive and exhaustive events, then, using the first version of Bayes' theorem, we get,

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

Example: "Monty Hall problem"
In a TV game show with a host, there are three doors – two with goats behind them and one with a car. A player is asked to choose a door. After the player chooses a door, the host opens another door and a goat appears behind it. The host asks whether the player would like to change the choice made. Is changing advantageous?

Let the doors be A, B and C. Assume, without loss of generality, that the user has chosen A and the host has opened B.
We are interested in $P(Cc|Bo)$, where Cc denotes the event that door C has a car and Bo that the door B is opened. Using Bayes' theorem,

$$\begin{aligned} P(Cc|Bo) &= \frac{P(Bo|Cc)P(Cc)}{P(Bo)} \\ &= \frac{1 \times \frac{1}{3}}{P(Bo|Ac)P(Ac) + P(Bo|Bc)P(Bc) + P(Bo|Cc)P(Cc)} \\ &= \frac{\frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{2}{3} \end{aligned}$$

Example: "Clinical Test"
Consider a clinical test that gives positive result for 99% of tests performed on diseased individuals and for 3% of those performed on healthy individuals. Let the said disease affect 2% of the population. What is the probability that an individual who tests positive has the disease?
$P(+|D) = 0.99$, $P(+|H) = 0.03$, $P(D) = 0.02 \Rightarrow P(H) = 0.98$.

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+)} \\ &= \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|H)P(H)} \\ &= \frac{0.99 \times 0.02}{0.99 \times 0.02 + 0.03 \times 0.98} = \frac{198}{198 + 294} \approx 40\% \end{aligned}$$

III. DISTRIBUTIONS

Random variable: A variable which can take values from some set (discrete or continuous) with a given fixed set of probabilities.

For a discrete random variable, the *Probability Mass Function* is the function $P(x)$ that represents the probability of the random variable X, taking a value x, from the discrete set of possible values.

For a continuous random variable, the *Probability Density Function* is the function $P(x)$ for which $P(x)\delta x$ is the probability that the random variable X, takes a value between x and $x + \delta x$, where $\delta x \to 0$.

The *expectation* value of a random variable is given by, $\langle X \rangle = \sum_x x p(x)$, where $p(x)$ is the probability that the random variable X takes the value x.

The *variance* of X is $\text{var}(X) \equiv \langle (X - \langle X \rangle)^2$. By expanding the square, it is easy to show that, $\text{var}(X) = \langle X^2 \rangle - \langle X \rangle^2$. The square root of the variance is known as the *standard deviation*

In general, the n^{th} *moment* of X is defined as $\langle X^n \rangle$ and the n^{th} *central moment* is $\left\langle \left(X - \langle X \rangle\right)^n \right\rangle$. The mean is the first moment. The first central moment is zero. The second central moment is the variance.
You may also read about moment generating functions.

Binomial distribution

A *Bernoullian trial* is an experiment which has two possible outcomes – *success* and *failure*. The probability of success in each trial is denoted by p and that of failure by $q = 1 - p$.
When two trials are conducted, then the probability of no success is q^2, that of one success is $2pq$ (since we can have success first or failure first), and that of two successes is p^2.

What is the probability of n successes in N Bernoullian trials? The answer to this question is the Binomial distribution. A particular sequence of outcomes, e.g. $SSSS\dots FFFF$, where there are n successes and $N-n$ failures can occur with probability $p^n q^{N-n}$. However, any permutation of this sequence is also an outcome of the N-trial experiment which has n successes. The number of such permutations is $\binom{N}{n}$. This factor can also be understood as choosing the n trials out of the N trials where the successes occur. Therefore, the probability of n successes in N Bernoullian trials is,

$$P_N(n) = \binom{N}{n} p^n q^{N-n}.$$

This distribution of probabilities of the different possible numbers of successes is called the *binomial distribution*.

It is easy to check that $\langle n \rangle = \sum_{n=0}^{N} P_N(n) = 1$, i.e., the probabilities are normalized and adds up to one.

Expected Number of Successes.

$$\begin{aligned}\langle n \rangle &= \sum_{n=0}^{N} nP_N(n) = \sum_{n=1}^{N} n\frac{N!}{n!(N-n)!}p^n q^{N-n} = \sum_{n=1}^{N} \frac{N!}{(n-1)!(N-n)!}p^n q^{N-n} \\ &= Np\sum_{n=1}^{N} \frac{(N-1)!}{(n-1)!\big((N-1)-(n-1)\big)!}p^{n-1}q^{(N-1)-(n-1)} \\ &= Np\sum_{m=0}^{N-1} \frac{(N-1)!}{m!\big((N-1)-m\big)!}p^m q^{(N-1)-m} = Np(p+q)^{N-1} = Np \end{aligned} \quad (1)$$

Variance in the number of successes. By calculating $\langle n^2 \rangle$, it is easy to show that,

$$\text{var}(n) = \langle n^2 \rangle - \langle n \rangle^2 = Npq \quad (2)$$

Higher moments can also be calculated in the same way.

Poisson distribution

Suppose an event occurs with a constant rate μ (rate means probability per unit time). The rate of the event is small enough so that the time interval during which there is at most one event is not too small (e.g., DN-43 bus arriving in front of APC College). The probability that an event occurs in an interval of length dt is μdt. What is the probability that n events occur in an interval t? Let us denote this by $P_n(t)$. Then,

$$P_n(t+dt) = P_n(t)P_0(dt) + P_{n-1}(t)P_1(dt),$$

where dt is small enough so that not more than one event can occur during this interval.

$$P_n(t+dt) = P_n(t)(1-\mu dt) + P_{n-1}(t)\mu dt \Rightarrow \frac{dP_n(t)}{dt} = \mu P_{n-1}(t) - \mu P_n(t).$$

Plugging $n = 0$, we get,

$$\frac{dP_0(t)}{dt} = -\mu P_0(t) \Rightarrow P_0(t) = e^{-\mu t}.$$

Proceeding in the same way and using the fact that $P_n(0) = 0$ for $n \geq 1$, we get,

$$P_n(t) = \frac{e^{-\mu t}(\mu t)^n}{n!}.$$

For some fixed time interval, denoting μt by λ, and using the random variable X to denote the number of events in that interval, the probability that X takes the value x is,

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

This distribution is known as the *Poisson distribution.* It can be easily shown that both the mean and variance of a random variable following the Poisson distribution with parameter λ take the value λ.

It can be shown a binomial distribution with large N and small p such that the product $Np = \lambda$ is finite becomes a Poisson distribution with parameter λ.

Normal (Gaussian) distribution

The normal or a Gaussian distribution is simplest and most common of the distributions followed by continuous random variables. If a random variable follows a normal distribution with mean μ and standard deviation σ, then the probability density at a value x is given by,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

i.e., the probability that it takes a value between x and $x + dx$ is $P(x)dx$. It can be shown that the above distribution is normalized and has the expected mean and standard deviation. $P(\mu - k\sigma < X < \mu + k\sigma)$ takes the values 68.27%, 95.45% and 99.73% for $\in \{1, 2, 3\}$ respectively.

The *Central Limit Theorem* makes the Gaussian distribution useful and states that the mean of n independent and identically distributed (i.i.d.) random variables which follow any distribution with a specified mean μ and standard deviation σ follows a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ when $n \gg 1$. The last point is the reason why you should take many readings during experiments.

A Poisson distribution approaches a Gaussian distribution with $(\mu = \sigma^2 = \lambda)$ when we focus on large values of the random variable.

A binomial distribution approaches a Gaussian distribution with $(\mu = Np, \sigma^2 = Npq)$ when N is large and p is not too small.