DATE

Probability theory

Probability theory is nothing but common sense reduced to calculation.

Ex. Marbles in a jar

A jar has 2 orange, 5 blue, 3 red, and 4 yellow marbles. A marble is drawn at random from the jar and is replaced back again. Each marble has a probability of 1/14

Probability the marble is orange $2/14 = 1/7$
Probability that the marble is red $= 3/14$
Probability that the marble is orange or blue $= 7/14 = 1/2$

Ex Coins in your purse.

You have 5 ₹1
3 ₹2
2 ₹5

A coin is taken at random. If x equals the value of coin, what is the average value x?

$$P(₹1) = \frac{5}{10} = 1/2$$

$$P(₹2) = \frac{3}{10} \quad \text{and} \quad P(₹5) = \frac{2}{10} = 1/5$$

$$\langle x \rangle = 1 \times P(₹1) + 2\,P(₹2) + 5\,P(₹5)$$
$$= \frac{1}{2} + \frac{3 \times 2}{10} + 5 \times \frac{2}{10} = \frac{5}{10} + \frac{6}{10} + \frac{10}{10}$$
$$= \frac{21}{10} = 2.1\ ₹$$

On the other hand the naive expectation would be ₹4.

Rules of probability

Suppose that there is an operation or a process that has several distinct possible outcomes(ω_i). The process might be the flip of a coin or the roll of a six-sided dice. We call each flip a trial. The list of all possible events or outcomes is called the sample space

For example if you are tossing a coin then the sample space $S = \{H, T\}$. If you are rolling a dice then the sample space $S = \{1, 2, 3, 4, 5, 6\}$.
If you are observing a Brownian particle then the sample space is the set of all points accessible to the particle.

The choice of sample space is dictated by the particular system one is asking, and type of probabilistic question one is interested. As a general rule the elements (outcomes) ω_i and ω_j must be mutually exclusive - that is two events ω_i and ω_j do not occur simultaneously.
Furthermore the set of elements should be complete, that is union of all ω_i should give S $\cup_i \omega_i = S$.
A subset of A of S, $A \subset S$ is then called a event.

If there are N such mutually exclusive events then we assign probabilities $P(i)$ that satisfy

$$1 \geq P(i) \geq 0 \quad \rightarrow \quad \text{Positivity}$$

$$\sum_i P(i) = 1 \quad \rightarrow \quad \text{Normalization}$$

Additivity → if A and B are of two disjoint events that is $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

The additivity axiom can be further generalized. Let A_i $(i = 1, \ldots N)$ be pairwise disjoint, that is $A_i \cap A_j = \emptyset$, then $P(\cup_i A_i) = \sum_i P(A_i)$

Ex Now, for a fair dice the probability of rolling any face is 1/6. The probability of throwing a 3 or a 6 with one throw is then

$$P(3 \text{ or } 6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Interpretation of probability - there are many different interpretations of probability, because any interpretation that satisfies the above criteria can serve as a probability.

Interpretation based on symmetry
Examples include tossing a coin. There are two faces and both of them are equally likely to occur.

Six faced dice. Possible outcomes are 1, 2, 3, 4, 5, 6. and therefore probability of getting any face is 1/6.

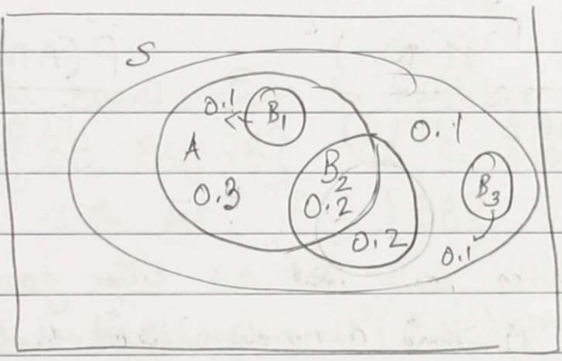
Frequency interpretation is also an alternative interpretation. Suppose an experiment is performed N times, and N_A times event A occurs. If N is sufficiently large then $P(A) = N_A/N$.

Conditional probability

We now want to consider probability for an event A given that an event B has occured. This is called a Conditional probability defined as $P(A|B)$

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

This may be visualized as restricting the sample space to B. The implication of the above equation is that if the outcomes are restricted to B, then this set serves as the new sample space.



$P(A|B_2) = \frac{0.2}{0.4} = \frac{1}{2} = 0.5$

$P(A) = 0.6$

$P(A|B_1) = \frac{0.1}{0.1} = 1$

$P(A|B_3) = 0.$

Formal Derivation → Let S be a sample space with elemantary event $\{\omega\}$. Suppose that an even $B \subseteq S$ has occured. Now, a new probability is to be assigned on $\{\omega\}$ to reflect this.

$\omega \in B \quad P(\omega|B) = \alpha P(\omega)$

$\omega \notin B \quad P(\omega|B) = 0$

$$\sum_{\omega \in S} P(\omega|B) = 1$$

DATE

$$\sum_{\omega \in S} P(\omega|B) = \sum_{\omega \in B} P(\omega|B) + \sum_{\omega \notin B} P(\omega|B)$$

$$= \alpha \sum_{\omega \in B} P(\omega) = 1$$

$$\Rightarrow \alpha P(B) = 1$$

$$\alpha = 1/P(B)$$

$$\Rightarrow P(\omega|B) = \frac{P(\omega)}{P(B)} \quad \text{for } \omega \in B$$

$$= 0 \quad \text{for } \omega \notin B.$$

For a general event A,

$$P(A|B) = \sum_{\omega \in A \cap B} P(\omega|B) + \sum_{\omega \in A \cap B^c} \cancelto{0}{P(\omega|B)}$$

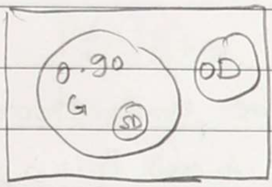
$$P(A|B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

Ex A machine produces parts that are either good (90%) slightly defective (2%) and obviously defective (8%). Produced parts get through an automatic inspection machine which is able to detect any part that is obviously defective. What is the quality of the parts that make it through the machine?

$$P(G) = 0.90 \qquad P(SD) = 0.02 \quad \text{and} \quad P(OD) = 0.08$$

We want to know the probability $P(G|OD^c)$, that is the probability that it is good given that it passed inspection (that is it not obviously defective)

$$P(G|OD^c) = \frac{P(G \cap OD^c)}{P(OD^c)} = \frac{0.90}{1 - P(OD)} = \frac{0.90}{0.92} \sim 0.98$$



Ex Your neighbor has 2 children. You learn that he has a son Joe. What is the probability that Joe's sibling is also a brother?

Now, the sample space here is given by $S = \{BB, BG, GB, GG\}$.

Assuming boys and girls are equally likely, then the 4 elements of S are also equally likely.

Now, the event that the neighbor has a son is the set $E = \{BB, BG, GB\}$ and the event that the neighbor has two boys is the set $F = \{BB\}$

We want to know $P(F|E) = \dfrac{P(F \cap E)}{P(E)} = \dfrac{P(\{BB\})}{P(\{BB, BG, GB\})}$

$$= \frac{1/4}{3/4} = \frac{1}{3}$$

The expression for the conditional probability can be given an interpretation in terms of frequencies. Let an experiment be performed N times, and N_A denote the number of times A appears. If N is large enough $P(A) = \dfrac{N_A}{N}$ and $P_B = \dfrac{N_B}{N}$ and $P(A \cap B) = N_{A \cap B}/N$.

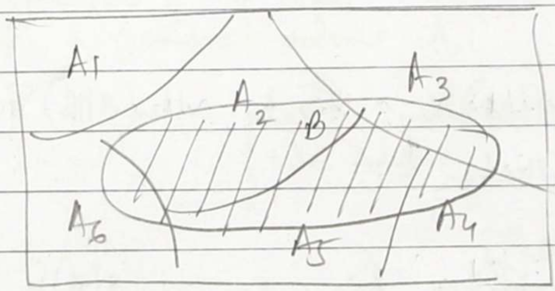
DATE

Therefore, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{N_{A \cap B}}{N_B}$

So the conditional probability $P(A|B)$ can be understood as the relative number of times both A and B appear to the number of times B appears.

Partition theorem

Let $S = \bigcup_i A_i$. Then $\{A_i\}$ is called a partition of S.



For a subset $B \subset S$ we can write

$$B = \bigcup_i (B \cap A_i)$$

Therefore

$$P(B) = P\left(\bigcup_i (B \cap A_i)\right) = \sum_i P(B \cap A_i)$$

$$= \sum_i P(A_i)\, P(B|A_i)$$

Now we put is in the conditional probability expression.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i \cap B)}{\sum_i P(A_i)\, P(B|A_i)}$$

Bayes theorem $\rightarrow$ It provides us with a way of understanding how the probability that a hypothesis is true is affected by new evidence.

DATE

$$P(A) = P(A|B) + P(A|-B)$$

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

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

$$\Rightarrow \quad P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

$$\Rightarrow \quad P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad \text{Baye's theorem.}$$

If I now look at multiple possible outcomes A_i which are mutually exclusive then

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

Ex A computer chess program has two modes Expert (E) and novice (N). The expert mode beats you 75% and the novice mode wins 50% of the time. You close your eyes and randomly choose one mode. The computer wins (W) both times. What is the probability that you chose the novice mode?

The probability that is of interest is $P(N|WW)$, which is difficult to calculate. We therefore use Baye's theorem.

$$P(N|WW) = \frac{P(WW|N)\,P(N)}{P(WW)}$$

Now, $P(N) = 1/2$. Therefore $P(WW|N) = \left(\frac{1}{2}\right)^2 = 1/4$.

We now calculate $P(WW)$. There are two ways the program won the two games —

1) You chose the novice mode and it won twice
2) You chose the expert mode and it won twice

Since N and E are mutually exclusive

$$P(WW) = P(N \text{ and } WW) + P(E \text{ and } WW)$$

$$= P(WW|N)\,P(N) + P(WW|E)\,P(E)$$

$$= \left(\tfrac{1}{4}\right)\left(\tfrac{1}{2}\right) + \left(\tfrac{3}{4}\right)^2\left(\tfrac{1}{2}\right) = \frac{13}{32}$$

$$P(N|WW) = \frac{(1/4)(1/2)}{13/32} = \frac{4}{13} \simeq 0.31$$

Note that the probability of winning by choosing a novice mode has decreased from 50% to 31%.

Discrete distributions

Consider 3 spins which can be in +1 state with probability p
-1 state with probability q.

$\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow$	$\uparrow\uparrow\downarrow$	$\downarrow\uparrow\uparrow$
p^3	p^2q	p^2q	p^2q
$\uparrow\downarrow\downarrow$	$\downarrow\uparrow\downarrow$	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\downarrow$
q^2p	q^2p	q^2p	q^3

Let $P_N(n, n')$ denote the probability that n spins are up and $n' = N - n$ spins are down.

$$P_3(3,0) = p^3, \quad p_3(2,1) = 3p^2 q, \quad p_3(1,2) = 3q^2 p, \quad p_3(0,3) = q^3$$

For N spins, therefore the probability distribution can be written as

$$P_N(n, n') = W_N(n, n') \, p^n q^{n'}$$

The factor $W(n, n')$ can be written down as $W_N(n, n') = {}^N C_n = \dfrac{N!}{n!\, n'!}$

So that $$p_N(n, n') = \frac{N!}{n!\, n'!} p^n q^{n'} = \frac{N!}{n!\,(N-n)!} p^n q^{N-n}$$

↳ Binomial Distribution.

Consider now a sequence of random events and let $t_1, t_2, \ldots$ be the time at each successive event occurs. Examples include when a phone call is received and the times when a Gueiger Counter registers a decay process.

Suppose that we register the sequence over a very long time T that is much greater than any of the intervals $t_i - t_{i-1}$. We also assume that the average number of events is λ per unit time so that in the time interval t, the mean number of events is λt.

Assume, that the events are random and are independent of each other. Given λ, the mean number of events per unit time, we want to find the probability distribution $w(t)$ of the interval between two events.

DATE

Now, if an event has occured at $t=0$, the probability that another will happen within $[0,t]$ is given by

$$\int_0^t w(t)\,dt'$$

$\Rightarrow$ The probability that no event occurs in the interval t is $1 - \int_0^t w(t')\,dt'$

Therefore, the probability that duration of the interval between the two events is between t and $t+dt$ is given $w(t)\,dt$

$w(t)\,dt$ = probability that no event occurred in $[0,t]$ × probability that one event occurs in $[t, t+dt]$

$$= \left(1 - \int_0^t w(t')\,dt'\right) \lambda\,dt$$

$$\Rightarrow \frac{dw}{dt} = -\lambda w(t)$$

$$w(t) = A\,e^{-\lambda t} \qquad \int_0^\infty w(t')\,dt' = 1$$

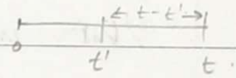
$\Rightarrow A = \lambda$ so that $w(t) = \lambda e^{-\lambda t}$.

Now we ask a different question —
Let us divide a long time interval T into smaller time intervals T/n. What is the probability n events occur in the time interval, given λ?

For $n=0$, $$P_{n=0}(t) = 1 - \lambda \int_0^t e^{-\lambda t'}\,dt' = e^{-\lambda t}$$

DATE

For $n=1$. $\rightarrow$ Let the event happen at t'. Since there can be no further events in $(t-t')$.



$$P_{n=1}(t) = \int_0^t \lambda e^{-\lambda t'} e^{-\lambda(t-t')} \, dt' \qquad \lambda e^{-\lambda t} t$$

$$= \lambda \int_0^t dt' \, e^{-\lambda t} = \lambda t \, e^{-\lambda t'}$$

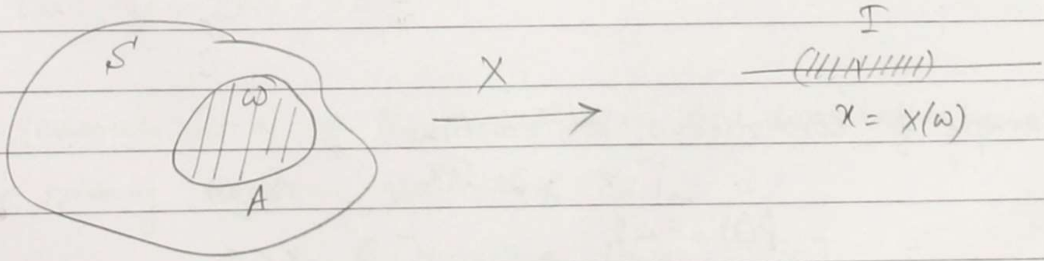
This can be now generalized. If there n events in the interval $[0,t]$, then the first event occurs at t' and the rest $(n-1)$ events must occur within the time interval $(t-t')$

$$\Rightarrow P_n(t) = \int_0^t \lambda e^{-\lambda t'} P_{n-1}(t-t') \, dt'$$

$$\boxed{P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}} \rightarrow \text{Poisson Distribution.}$$

Continuous probability distribution

Random variables $\rightarrow$ So far we have formulated the probability theory in an abstract space S. Now, we would like to assign numbers to the outcomes of an experiment.
Formally this is acheived by a map X from Sample Space S into $\mathbb{R}$ with the property that the inverse mapping X^{-1} of every interval $I \subset \mathbb{R}$ corresponds to an event $A \subset S$.



PAGE

Ex: we may define binary random variable for tossing a coin with sample space $S = \{H, T\}$

$$X : \begin{cases} H \to 0 \\ T \to 1 \end{cases}$$

Starting from the probability measure P defined on S we can now define an induced probability measure

$$P(\{x \in I\}) := P(A = X^{-1}(I))$$

$$= P(\{\omega \in S : X(\omega) \in I\})$$

the probability for the random variable X to take values in the interval $I \subset \mathbb{R}$ is defined as the probability of the event A in sample space S.

For a continuous random variable which can take on a continuous set of values, it is difficult to build on the concept of a probability for the random variable to take a particular value x; since such a probability would be strictly zero. Instead we define the probability density function $p(x)$

$$p(x) \geq 0 \qquad \int_{\mathbb{R}} p(x)\, dx = 1$$

$$p(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

Example of continuous distributions-

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

DATE

Gaussian Distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Characterization of probability distributions.

Moments

Mean $\langle x \rangle \equiv E[x] = \int_{\mathbb{R}} x\, p(x)\, dx$

Mean of any function of x can then be written as

$$E[h(x)] = \int h(x)\, p(x)\, dx$$

Variance

$$\langle x^2 \rangle \equiv Var[x] = \sigma^2 := E\left[(x - E[x])^2\right] = E[x^2] - (E[x])^2$$

Covariance and Correlation function

If we have two random variable X, Y.

$$Cov[X,Y] = E\left[(X - E[X])(Y - E[Y])\right]$$

$$= E[XY] - E[X]E[Y]$$

If X and Y are independent then $E[XY] = E[X]E[Y]$ and $Cov[XY] = 0$.

A generalization of the covariances are correlation functions for many random variables X_i

$$C_{ij} = E[X_i X_j] - E[X_i]E[X_j]$$

If X is now a function of time — $X(t)$ such $\langle X \rangle = 0$
Now, the value of X at two times t_1 and t_2 are random variables. We can therefore write

$$C(t_1, t_2) = E[X(t_1) X(t_2)] \equiv \langle X(t_1) X(t_2) \rangle$$

if $C(t_1, t_2) \equiv C(|t_1 - t_2|)$ then the process is a stationary process; otherwise it is called non-stationary process.

Generating Functions

Generating functions are a very convinient way of storing information. Consider a sequence $\{a_0, a_1, a_2, \dots a_n\}$.
One defines the ordinary generating function of the sequence as

$$G(s) = \sum_{n=0}^{\infty} a_n s^n$$

The sequence can be reconstructed by using derivatives

$$a_n = \frac{1}{n!} \frac{d^n G(s)}{ds^n}\bigg|_{s=0}.$$

Given a discrete random variable X_1, which take values $0, 1, 2, \dots$ and the set of probabilities $p_n = Prob(X=n)$.

$$G_X(s) := E[s^X] = \sum_n s^n p_n \qquad G_X(0) = 1$$

Ex $p_n(\mu) = \frac{\mu^n}{n!} e^{-\mu}$

$$G_p(s) = \sum_n s^n \frac{\mu^n}{n!} e^{-\mu} = e^{-\mu} \sum_n \frac{(\mu n)^s}{n!}$$

$$= e^{-\mu} e^{\mu s} = e^{+\mu(s-1)}$$

Binomial distribution $p_B(n, N, p) = \frac{N!}{n!\,(N-n)!} p^n (1-p)^{N-n}$

$$G_B(s, N, p) = \sum_{n=0}^{\infty} s^n \frac{N!}{n!\,(N-n)!} p^n (1-p)^{N-n}$$

$$G_B(s, N, p) = \left[ps + (1-p) \right]^N$$

Now consider the limit where the average $\mu = Np$ of the binomial distribution remains finite but $N \to \infty$

$$\lim_{N\to\infty} G_B(s, N, p) = \lim_{N\to\infty} \left[\frac{s\mu}{N} + \left(1 - \frac{\mu}{N}\right) \right]^N = \left[\frac{\mu}{N}(s-1) + 1 \right]^N$$

$$\lim_{x\to\infty} \left(1 + \frac{\kappa}{x}\right)^{mx} = e^{m\kappa}$$

$$\lim_{N\to\infty} G_B(s, N, p) = e^{\mu(s-1)} = G_p(s, \mu)$$

For any random variable X

$Y = X_1 + X_2$

$P_Y(y) = \int dx_1\, p(x_1)\, p(x_2)\, \delta(y - x_1 - x_2)$

$$G_X(1) = \sum_n p_n = 1$$

$$G_X' = \sum n p_n = E[X]$$

$$G''(X) = \sum n(n-1) p_n = \sum n^2 p_n - \sum n p_n$$
$$= E[X^2] - E[X].$$

If we have $X_1, X_2, \dots X_r$ independent random variables then define $Y = X_1 + X_2 + \dots + X_r$

$$G_Y(s) := E[s^Y] = E\left[s^{\sum X_i}\right]$$

$\because X_i$ are independent, then $G_Y(s) = \prod_i E[s^{X_i}] = \prod G_{X_i}(s)$

Conditional probabilities in continuous case

If (x, y) are random variables and have a joint probability density function $p(x, y)$ then the conditional probability is given by

$$p_{X|Y}(x|y) = \frac{p_{XY}(x\ y)}{p_Y(y)}$$

to motivate this multiply L.H.S by dx and R.H.S by $\frac{dx\,dy}{dy}$

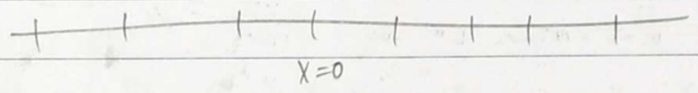
$$p_{X|Y}(x|y)\,dx = \frac{p_{XY}(x, y)\,dx\,dy}{p_Y(y)\,dy}$$

$$= \frac{P\{x \le X \le x+dx,\ y \le Y \le y+dy\}}{P\{y \le Y \le y+dy\}}$$

Further, $p_Y(y) = \int p_{XY}(x, y)\,dx$

One dimensional random walk

Consider a particle located initially at position $X_0 = 0$, on a one dimensional lattice



At every time it can move forward a distance $+a$ with probability p or move backward $-a$ with probability p'

$t_n = n\tau$

DATE

If X_i is the position of the particle at $t_i = i\tau$ then

$$X_i = X_{i-1} + Z_i$$

Where Z_i is a random variable which can take values $\pm a$.

We want to calculate the probability distribution for X_n after n steps. To get to a distance $X_n = ka$ after n steps means taking n_+ steps forward & n_- steps backward.

$$n_+ + n_- = n \qquad n_+ = \frac{(n+k)}{2}$$

$$n_+ a - n_- a = ka \qquad n_+ - n_- = k \qquad n_- = \left(\frac{n-k}{2}\right)$$

This occurs $\frac{n!}{n_+!\, n_-!}$ ways so that distribution is given by

$$P(n, k, p) = \frac{n!}{n_+!\, n_-!} p^{n_+} p^{n_-} = \frac{n!}{n_+!\, n_-!} p^n$$

We are now interested in a continuum limit such $n \to \infty$, $a, \tau \to 0$.
If you calculate the mean $E[X_n]$ and variance we have

$(Z - \langle Z \rangle)^2$ $\quad (p-q)a = \mu$

$$\langle X_n \rangle = \sum_i \langle Z_i \rangle$$

$\langle Z^2 \rangle - \langle Z \rangle^2$

$pa^2 + qa^2 - a^2(p-q)^2$

$$\text{and} \quad \langle X_n^2 \rangle = \sum \langle Z_i^2 \rangle$$

$(p^2 + q) - p^2 + q^2 + 2pq$

$1 - p^2 - (1-p)^2$
$+ 2p(1-p)$

Now, Z are Bernoulli trials and therefore

$1 - p^2 - (1 - 2p + p^2)$
$+ 2p^2 - 2p^2$

$$\langle Z \rangle = \sum z_i\, p(z_i) = p(+a) + p(-a) = 0.$$

$1 - p^2 - 1 + 2p - p^2$
$+ 2p - 2p^2$

$$\langle Z^2 \rangle = \sum z_i^2\, p(z_i) = a^2 p + a^2 p = 2a^2 p$$

$1 - 4p^2 + 4p$
$1 - 4p(1-p)$
$1 - 4pq$

$4pq$

DATE

$\langle x_n \rangle = 0$ $\qquad t_n = n\tau$

and $\langle x_n^2 \rangle = 2pa^2 n = 2pa^2 \frac{t_n}{\tau} = 4p\frac{a^2}{2\tau}t_n$

$D = \frac{a^2}{2\tau}$ → Diffusivity one of the macroscopic quantity that has emerged from microscopic dynamics

In the continuum limit with $n \to \infty$, the binomial distribution

$$p(n,k,p) = \frac{1}{\sqrt{2\pi n}} e^{-k^2/2n}$$

we now define the probability density as

$$\rho(x, t_n) = \frac{1}{a} p(n,k,p)$$

$$= \frac{1}{\sqrt{2\pi a^2 n}} e^{-k^2/2n}$$

$$= \frac{1}{\sqrt{4\pi \frac{a^2 t}{2\tau}}} e^{-x^2/4\frac{a^2 t}{2\tau}}$$

$$= \frac{1}{\sqrt{4\pi D t}} e^{-x^2/4Dt}$$

We may now interpret this result "macroscopically" as a Brownian walker undergoing a diffusion process. The probability density $\rho(x,t)$ is actually a solution of the equation $\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2}$

To derive this result, we note that the Brownian particle found at x at a time $t+\tau$ can only come from $x-a$ or $x+a$ at time t

$$\rho(x, t+\tau) = p\rho(x-a, t) + p\rho(x+a, t)$$

$$\rho(x, t+\tau) = \rho(x,t) + \tau \frac{\partial \rho(x,t)}{\partial t}$$

Further, $\rho(x \pm a, t) = \rho(x,t) \pm a\frac{\partial \rho(x,t)}{\partial x} + \frac{1}{2}a^2 \frac{\partial^2 \rho(x,t)}{\partial x^2} + O(a^3)$

$$\rho(x,t) + \tau\frac{\partial \rho}{\partial t} = \rho(x,t) + \frac{1}{2}a^2\frac{\partial^2 \rho}{\partial x^2} + O(a^3)$$

$$\frac{\partial \rho}{\partial t} = \frac{a^2}{2\tau}\frac{\partial^2 \rho}{\partial x^2} \Rightarrow \frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2}$$

Central limit theorem

Let there be mutually independent random variables $X_1, X_2, \dots X_N$ which are characterized by common but independent probability distribution $W(x_1), W(x_2) \dots, W(x_N)$. Suppose that the mean value and variance exists for $X_1, X_2, X_3 \dots X_N$.
We require the probability distribution for the sum

$$Y = X_1 + X_2 + \dots X_N = \sum_i X_i \quad \text{in the limit of } N \to \infty$$

Now consider the random variable $Z = \sum_i (X_i - \langle x \rangle)/\sqrt{N}$

$$= \frac{Y}{\sqrt{N}} - \sqrt{N}\langle x \rangle$$

Then by definition $W_Z(z) = \int dx_1\, dx_2 \dots dx_N\, \delta\left(z - \frac{Y}{\sqrt{N}} + \sqrt{N}\langle x \rangle\right) W(x_1) W(x_2) \dots W(x_N)$

DATE

$$W_Z(z) = \int dx_1\, dx_2 \ldots dx_N\; W(x_1)\, W(x_2) \ldots W(x_N)\; \delta\left(z - \frac{x_1 + x_2 + \cdots + x_N}{\sqrt{N}} + \sqrt{N}\langle x\rangle\right)$$

$$= \int \frac{dk}{2\pi} \int \prod dx_i \; e^{ik\left(z - \frac{x_1 + x_2 + \cdots + x_N}{\sqrt{N}} + \sqrt{N}\langle x\rangle\right)}$$

$$= \int \frac{dk}{2\pi} e^{ikz} \int \prod_i dx_i \; e^{-ik\sum_i x_i/\sqrt{N}} \; e^{ik\sqrt{N}\langle x\rangle} \; W(x_1)\, W(x_2) \ldots W(x_N)$$

$$= \int \frac{dk}{2\pi} e^{ikz} \prod_i \int dx_i \; e^{-i\frac{k}{\sqrt{N}} x_i} \; W(x_i) \; e^{ik\sqrt{N}\langle x\rangle}$$

$$= \int \frac{dk}{2\pi} \; e^{ikz + ik\sqrt{N}\langle x\rangle} \left[\Phi\left(\frac{k}{\sqrt{N}}\right)\right]^N$$

By definition

$$\Phi(k) = \langle e^{-ikx} \rangle = \int dx \; e^{-ikx} \; W(x)$$

$\Phi(k)$ is called the characteristic function of probability distribution.

$$\ln \varphi(k) = \ln \langle e^{-ikx} \rangle$$

$$= \ln \left\langle 1 - ikx + \frac{(ikx)^2}{2!} + \cdots \right\rangle$$

$$= \ln \left[1 - ik\langle x\rangle + \frac{(ik)^2}{2!}\langle x^2\rangle + \cdots \right]$$

$$= \sum_{m=1}^{\infty} \frac{(-ik)^m}{m!} C_m$$

$C_1 = \langle x\rangle \qquad C_2 = \langle x^2\rangle - \langle x\rangle^2 = \sigma^2 \qquad C_3 = \langle x^3\rangle - 3\langle x^2\rangle\langle x\rangle + 2\langle x\rangle^3$

DATE

The characteristic function therefore generates all the cumulants.

Coming back to the problem at hand

$$W_z(z) = \int \frac{dk}{2\pi} e^{ikz + ik\sqrt{N}\langle x\rangle} \left[\phi\left(\frac{k}{\sqrt{N}}\right)\right]^N$$

Now $\ln \varphi(k) = -ik\langle x\rangle + \frac{(ik)^2}{2!}\sigma^2 + \cdots +$

$$\varphi(k) = e^{-ik\langle x\rangle + \frac{(ik)^2}{2!}\sigma^2 - \frac{(ik)^3}{3!}C_3 \cdots}$$

$$W_z(z) = \int \frac{dk}{2\pi} e^{ikz + ik\sqrt{N}\langle x\rangle} e^{-ik\sqrt{N}\langle x\rangle + \frac{(ik)^2}{2!}\sigma^2 - \frac{(ik)^3}{3!\,N^{1/2}}C_3 + \cdots}$$

$$= \int \frac{dk}{2\pi} e^{ikz - \frac{k^2\sigma^2}{2}}$$

In the limit $N \to \infty$ all other terms vanish.

$$= \int \frac{dk}{2\pi} e^{-\left[\frac{k^2\sigma^2}{2} - ikz\right]}$$

$\left(k^2 - 2i\frac{z}{\sigma^2}k - \frac{z^2}{\sigma^4} + \frac{z^2}{\sigma^4}\right.$

$$= \int \frac{dk}{2\pi} e^{-\left(k^2 - \frac{2ikz}{\sigma^2}\right)\frac{\sigma^2}{2}}$$

$$W_z(z) = \int \frac{dk}{2\pi} e^{-\frac{\sigma^2}{2}\left[\left(k - \frac{2ikz}{\sigma^2}\right)^2 + \frac{z^2}{\sigma^4}\right]} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}$$

$$W_y(y)\,dy = W_z(z)\,dz$$

$$W_y(y) = W_z(z)\frac{dz}{dy} = \frac{1}{\sqrt{2\pi N\sigma^2}} e^{-\frac{(y - N\langle x\rangle)^2}{2N\sigma^2}}$$

This is the central limit theorem: $W_Y(y)$ is a Gaussian distribution, although we did not in any way assume that $w(x)$ was such a distribution.

$$\text{Mean value} \quad \langle Y \rangle = N\langle x \rangle$$

$$\sigma^2 : \quad \sigma_y^{2} = \sqrt{N}\sigma_x$$

$$\text{relative deviation:} \quad \frac{\sigma_y}{\langle y \rangle} = \frac{\sqrt{N}\,\sigma_x}{N\langle x \rangle} = \frac{\sigma_x}{\langle x \rangle}\frac{1}{\sqrt{N}}$$

Information and uncertainity

Consider an experiment which has two outcomes, E_1 and E_2 with probabilities P_1 and P_2. For example the experiment could be a coin tossing experiment. If the coin is a fair coin then we have $P_1 = P_2 = 1/2$. On the other hand, say we have a bent coin and $P_1 = 1/5$ $P_2 = 4/5$. Intuitively, we would say that the results of the first experiment is more uncertain.

Consider now two additional experiments. In the third experiment there are four outcomes with $P_1 = P_2 = P_3 = P_4 = 1/4$ and in the fourth there are 6 outcomes with $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6$. Intuitively, the 4^{th} experiment is the most uncertain and the 2^{nd} the least. We do not know how to rank the 1^{st} and the 3^{rd} experiment.

We therefore want to define a measure that is consistent with our intuitive sense. Let us define the uncertainity function $S(P_1, P_2, \ldots P_i, \ldots)$, where P_i is the probability of event i. Consider now the case where all the probabilities are equal $P_i = \frac{1}{\Omega}$: Ω is the total number of outcomes.

DATE

In this case we have $S(\Omega)$. It is easy to see that $S(\Omega)$ has to satisfy certain conditions:

* For only one outcome $\Omega = 1$ and $S(\Omega=1) = 0$
* For two values of Ω_1 and Ω_2 with $\Omega_1 > \Omega_2$ $S(\Omega_1) > S(\Omega_2)$ that is $S(\Omega)$ is an montonically increasing function of Ω.

Next consider multiple events — we throw a die with Ω_1 outcomes and flip a coin with Ω_2 outcomes. The total number of outcomes is $\Omega_1 \Omega_2$. Now, say we know the outcome of one event - the result of the die is known so that the uncertainity associated with it becomes known. But still the uncertainity associated with the tossing of the coin remains. Similarly we can also reduce the uncertainity in the reverse order, but the total uncertainity remains non-zero. We have therefore

$$S(\Omega_1 \Omega_2) = S(\Omega_1) + S(\Omega_2)$$

There is a unique functional form that satisfies the above form

$$S(xy) = S(x) + S(y)$$

Take $z = xy$

$$\frac{\partial S(z)}{\partial x} = \frac{\partial z}{\partial x}\frac{dS}{dz} = y\frac{dS}{dz}$$

$$\frac{\partial S(z)}{\partial y} = \frac{\partial z}{\partial y}\frac{dS}{dz} = x\frac{dS}{dz}$$

$$\frac{\partial S(z)}{\partial x} = \frac{dS(x)}{dx} \quad \text{and} \quad \frac{\partial S(z)}{\partial y} = \frac{dS(y)}{dy}$$

$$\frac{dS(x)}{dx} = y\frac{dS(z)}{dz} \qquad \frac{dS(y)}{dy} = x\frac{dS(z)}{dz}$$

DATE

$$x \frac{dS}{dx} = y \frac{dS}{dy}$$ Multiplying the previous expression by x and y respectively.

Since the L.H.S depends only on x and the R.H.S depends only on y, we have

$$x \frac{dS}{dx} = y \frac{dS}{dy} = A$$

$$S(x) = A \ln x + B$$

Now, if $x = 1$, $S(x) = 0 \Rightarrow B = 0$

The constant A is now arbitrary, which we set to 1.

$$S(x) = \ln x$$

and therefore, for equal probabilities we have

$$S(\Omega) = \ln \Omega$$

Now, what if the probabilities for the outcomes are different?

Consider a loaded die where the probabilities P_j are not equal. Imagine that we roll the dice a large number of times N. Then each outcome would occur $N_j = NP_j$ times. However, these outcomes can occur in different order. Thus, the original uncertainity about the outcome of one roll of a die is converted into an uncertainity about order. However, all the possible orders are equally likely in an experiment of N rolls. We can count that number

$$\Omega_R = \frac{N!}{N_1!\,\cancel{(N-N_1)!}} \; \frac{\cancel{(N-N_1)!}}{(N-N_1-N_2)!\; N_2!} \cdots \frac{(N-N_1-N_2-\cdots N_{n-1})!}{(N-N_1-N_2-\cdots N_n)!\; N_n!}$$

But $N = N_1 + N_2 + \cdots N_N$.

$$\Omega_R = \frac{N!}{N_1!\, N_2!\, N_3! \cdots N_N!} = \frac{N!}{\prod_j N_j!}$$

and then I can define the uncertainity S_N as

$$S_N = \ln \Omega_R = \ln\left[\frac{N!}{\prod_j N_j!}\right]$$

$$S_N = \ln N! - \sum_j \ln N_j!$$

The uncertainity associated with one roll is then given by

$$S = \mathrm{Lt}_{N\to\infty} \frac{1}{N} S_N = \mathrm{Lt}_{N\to\infty} \frac{1}{N}\left[\ln N! - \sum_j \ln N_j!\right]$$

$\ln N! = N \ln N - N \longrightarrow$ Sterlings Formula

$$S = \mathrm{Lt}_{N\to\infty} \frac{1}{N}\left[N \ln N - N - \sum_j (N_j \ln N_j - N_j)\right]$$

$$= \mathrm{Lt}_{N\to\infty} \frac{1}{N}\left[N \ln N - N - \sum_j (NP_j \ln NP_j - NP_j)\right]$$

$$= \mathrm{Lt}_{N\to\infty} \frac{1}{N}\left[N \ln N - N - \sum N \ln N\, P_j - \sum_j N P_j \ln P_j - N \sum_j P_j\right]$$

$$= \mathrm{Lt}_{N\to\infty} \frac{1}{N}\left[N \ln N - N - N \ln N - N - N \sum_j P_j \ln P_j\right]$$

$$= -\sum P_j \ln P_j$$

DATE

Properties of Statistical Entropy

1) Maximal Entropy - For a fixed number of events/outcomes, S reaches a maximum equal to $\sigma(N) = \ln N$, where all the events are equiprobable.

Let $\{p_i\}$ an $\{q_i\}$ be two probability laws. Then

$$\sum p_i \ln q_i - \sum p_i \ln p_i \leq 0$$

$$\ln x \leq x - 1$$

$$\sum_i p_i \ln q_i - \sum_i p_i \ln p_i = \sum_i p_i \ln \frac{q_i}{p_i} \leq \sum_i p_i \left(\frac{q_i}{p_i} - 1\right)$$

$$\leq \sum q_i - \sum p_i \leq 0.$$

Now choose $q_i = 1/N$ $\longrightarrow$ where all the events are equiprobable

$$\sum p_i \ln \frac{1}{N} - \sum p_i \ln p_i \leq 0$$

$$-\ln N + S(\{p_i\}) \leq 0$$

$$S(\{p_i\}) \leq \ln N = \sigma(N)$$

2) Minimum property: The statistical entropy reaches a minimum, equal to zero when one of the events is a certainity.

3) Extensivity: In case of N equiprobable events the statistical entropy increases with N.

4) Impossibility of events → If some of the events have zero probability then $S(p_1, p_2, \cdots p_N, 0, 0, \ldots 0) = S(p_1, p_2, \cdots p_N)$.

5) Additivity : For multiple events, the statistical entropy is given by

$$S = \sum_j S_j$$