3rd IIA-Penn State Astrostatistics School

19–27 July, 2010

Vainu Bappu Observatory, Kavalur

Laws of Probability, Bayes' theorem, and the Central Limit Theorem

Bhamidi V Rao

Indian Statistical Institute, Kolkata

Adapted from notes prepared by

Rahul Roy and Rajeeva Karandikar

Do Random phenomena exist in Nature?

Which way a coin tossed in air will fall may be completely determined by laws of physics. The only problem in figuring out the trajectory and hence the face of the coin when it is on ground is that we have to measure too many parameters, e.g. angular momentum of rotation, force at the time of toss, wind pressure at various instants during the rotation of the coin, etc.!

Which way an electron will spin is also not known and so modelling it will require incorporating a random structure.

But we cannot exclude the possibility that sometime in future, someone will come up with a theory that will explain the spin.

Thus we often come across events whose outcome is uncertain. The uncertainty could be because of our inability to observe accurately all the inputs required to compute the outcome.

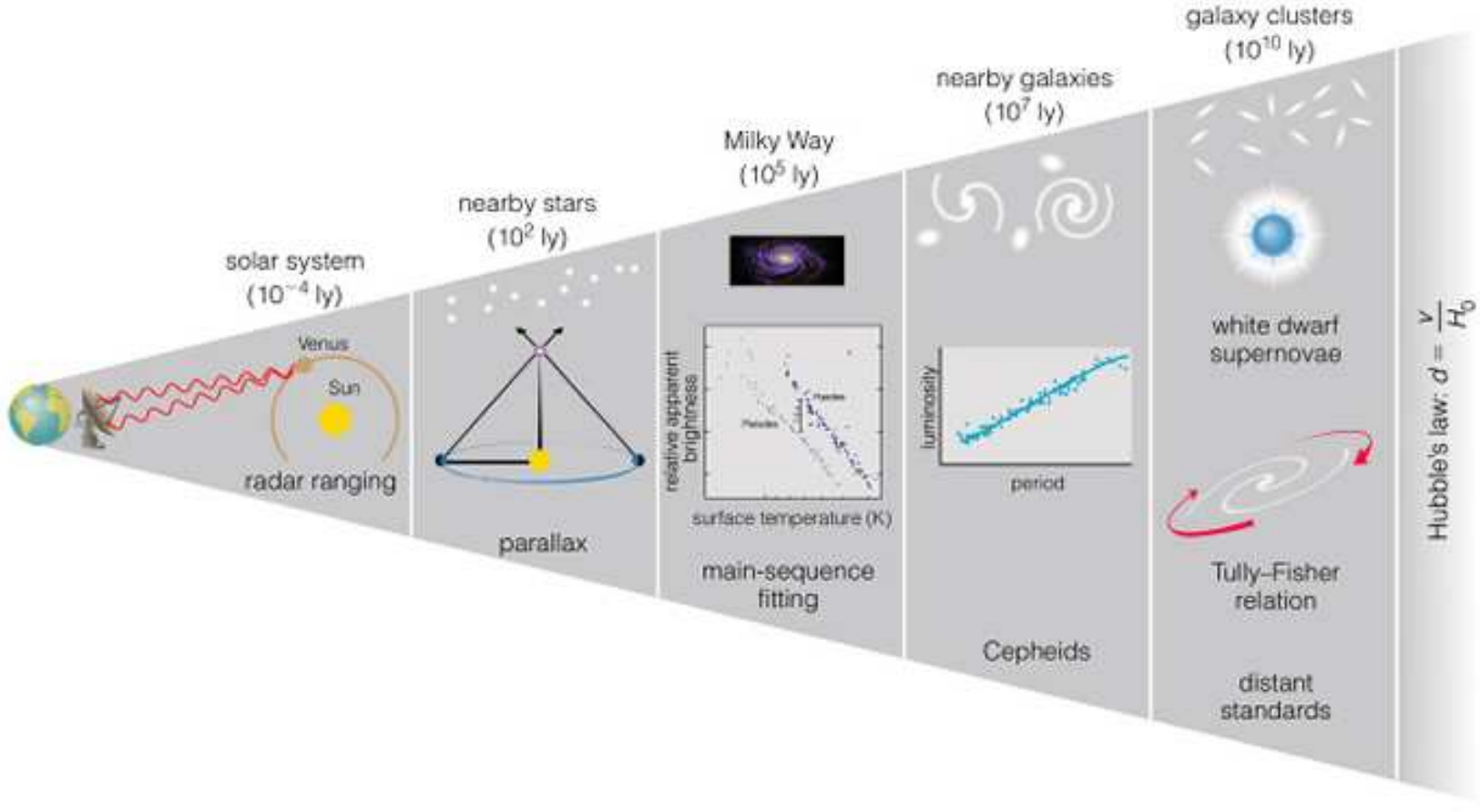
It may be too expensive or even counterproductive to observe all the inputs.

The uncertainty could be due to the current level of understanding of the phenomenon.

The uncertainty could be on account of the outcome depending on choices made by a group of people at a future time - such as outcome of an election yet to be held.

Cosmic distance ladder

The objects in the solar system were measured quite accurately by ancient Greeks and Babylonians using geometric and trigonometric methods.



see also www.math.ucla.edu/~tao/preprints/Slides/Cosmic%20Distance%20Ladder

The distance to the stars in the second lot are found by ideas of parallax, calculating the angular deviation over 6 months. First done by the mathematician Friedrich Bessel. The error here is of the order of 100 light years.

The distances of the moderately far stars are obtained by a combination of their apparent brightness and the distance to the nearby stars. This method works for stars upto 300,000 light years and the error is significantly more.

The distance to the next and final lot of stars is obtained by plotting the oscillations of their brightness. This method works for stars upto 13,000,000 light years!

At every step of the distance ladder, errors and uncertainties creep in. Each step inherits all the problems of the ones below, and also the errors intrinsic to each step tend to get larger for the more distant objects; thus the spectacular precision at the base of the ladder degenerates into much greater uncertainty at the very top.

So we need to understand UNCERTAINTY.

And the only way of understanding a notion scientifically is to provide a structure to the notion.

A structure rich enough to lend itself to quantification.

The structure needed to understand a coin toss is intuitive.

We assign a probability $\mathbf{1/2}$ to the outcome HEAD and a probability $\mathbf{1/2}$ to the outcome TAIL of appearing.

Similarly for each of the outcomes 1,2,3,4,5,6 of the throw of a dice we assign a probability $\mathbf{1/6}$ of appearing.

Similarly for each of the outcomes $\mathbf{000001}, \ldots, \mathbf{999999}$ of a lottery ticket we assign a probability $\mathbf{1/999999}$ of being the winning ticket.

Of course, we could obtain the structure of the uncertainty in a coin toss from the example of throwing a dice.

In particular if we declare as HEAD when the outcome of a throw of a dice is an even number, and if we declare as TAIL when the outcome of a throw of a dice is an odd number, then we have the same structure as that we had from a coin toss.

More generally, associated with any experiment we have an outcome space Ω consisting of outcomes $\{o_1, o_2, \ldots, o_m\}$.

Coin Toss – $\Omega = \{\text{H, T}\}$

Dice – $\Omega = \{1, 2, 3, 4, 5, 6\}$

Lottery – $\Omega = \{1, \ldots, 999999\}$

Each outcome is assigned a probability

Coin Toss – $p_H = 1/2, p_T = 1/2$

Dice – $p_i = 1/6$ for $i = 1, \ldots, 6$

Lottery – $p_i = 1/999999$ for $i = 1, \ldots, 999999$

More generally, for an experiment with an outcome space $\Omega = \{o_1, o_2, \ldots, o_m\}$. we assign a probability p_i to the outcome o_i for every i in such a way that the probabilities add up to 1.

The set $\Omega = \{o_1, o_2, \ldots, o_m\}$ is called a sample space.

A subset $E \subseteq \Omega$ is called an event.

We may be gambling with dice, so we could have a situation like

outcome	1	2	3	4	5	6
money amount	-8	2	0	4	-2	4

Our interest in the outcome is only *vis-á-vis* it's association with the monetary amount.

So we are interested in a mapping (i.e. a function) of the outcome space Ω to the reals $\mathbb{R}$

Such functions are called random variables.

The probabilistic properties of these random variables can be read out from the probabilities assigned to the outcomes of the underlying outcome space.

The probability that you win 4 rupees, i.e. $P\{X = 4\}$ means you want to find that the number 4 or the number 6 came out on the dice, i.e. $P\{4, 6\}$ Thus $P\{\omega : X(\omega) = 4\} = P\{4, 6\} = (1/6) + (1/6) = 1/3$.

Similarly the probability that you do not lose any money is the probability of the event that either 2, 3, 4 or 6 came out on the dice, and this probability is (1/6) + (1/6)+(1/6) + (1/6) = 2/3.

What are we doing?

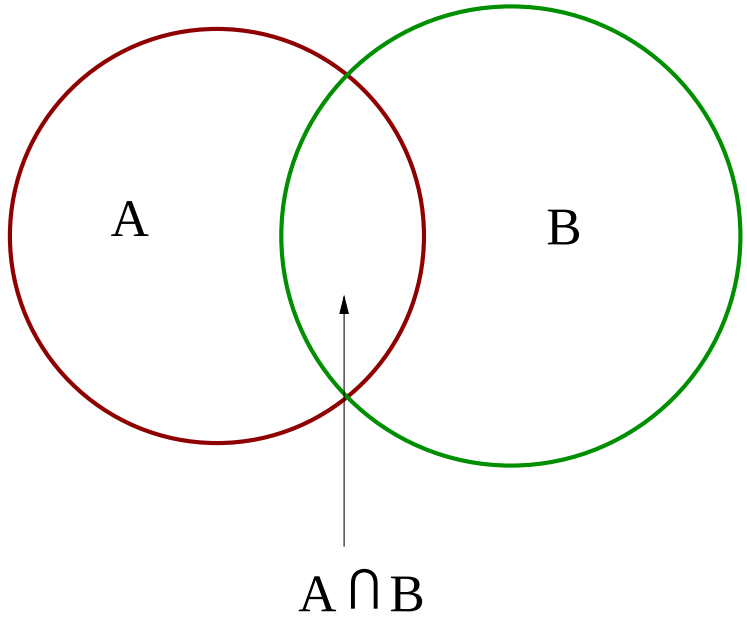
Recall our assignment of probabilities $P(o_i) = p_i$ on the outcome space $\Omega = \{o_1, o_2, \ldots, o_m\}$.

For an event $E = \{o_{i_1}, o_{i_2}, \ldots, o_{i_k}\}$, we define

$$P(E) = p_{i_1} + p_{i_2} + \cdots + p_{i_k}.$$

Easy to check that if A, B are mutually disjoint, *i.e.* $A \cap B = \phi$ then

$$P(A \cup B) = P(A) + P(B)$$



More generally, we can check that for any two events $\boldsymbol{A}$, $\boldsymbol{B}$

$$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B).}$$

Similarly, for three events A, B, C

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$
$$-P(A \cap B) - P(A \cap C) - P(B \cap C)$$
$$+P(A \cap B \cap C)$$

This has a generalization to n events.

How do we assign the probabilities p_i to the elementary outcomes?

The simplest case is when due to inherent symmetries present, we can model all the elementary events (*i.e.*outcomes) as being equally likely.

When an experiment results in m equally likely outcomes $o_1, o_2, \ldots, o_m$, the probability of an event A is

$$P(A) = \frac{\#A}{m}$$

i.e. the ratio of the number of favourable outcomes to the total number of outcomes.

Example: Toss a coin three times

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

$$p(***) = 1/8$$

Example: Toss a coin three times

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

$$p(***) = 1/8$$

No. of Heads in 3 tosses

Example: Toss a coin three times

$$\Omega = \{HHH,$$
$$HHT, HTH, THH,$$
$$HTT, THT, TTH,$$
$$TTT\}$$
$$p(***) = 1/8$$

No. of Heads in 3 tosses

$$\Omega = \{0, 1, 2, 3\}$$

Example: Toss a coin three times

$\Omega = \{HHH,$ ⟵ 3 Heads

$HHT, HTH, THH,$ ⟵ 2 Heads

$HTT, THT, TTH,$ ⟵ 1 Head

$TTT\}$ ⟵ 0 Heads

$p(***) = 1/8$

No. of Heads in 3 tosses

$\Omega = \{0, 1, 2, 3\}$

Example: Toss a coin three times

	No. of Heads in 3 tosses
	$\Omega = \{0, 1, 2, 3\}$
$\Omega = \{HHH,$	$\longleftarrow$ 3 Heads
$HHT, HTH, THH,$	$\longleftarrow$ 2 Heads
$HTT, THT, TTH,$	$\longleftarrow$ 1 Head
$TTT\}$	$\longleftarrow$ 0 Heads
$p(***) = 1/8$	$p(0) = 1/8,\ p(1) = 3/8,$
	$p(2) = 3/8,\ p(3) = 1/8$

Example: Toss a coin three times

	No. of Heads in 3 tosses
	$\Omega = \{0, 1, 2, 3\}$
$\Omega = \{HHH,$	$\longleftarrow$ 3 Heads
$HHT, HTH, THH,$	$\longleftarrow$ 2 Heads
$HTT, THT, TTH,$	$\longleftarrow$ 1 Head
$TTT\}$	$\longleftarrow$ 0 Heads
$p(***) = 1/8$	$p(0) = 1/8,\ p(1) = 3/8,$
	$p(2) = 3/8,\ p(3) = 1/8$

Note we could have done the calculations in the red part without even associating it with the blue sample space etc.

Conditional probability Let X be the number which appears on the throw of a dice.

Each of the six outcomes is equally likely, but suppose I take a peek and tell you that X is an even number.

Question: What is the probability that the outcome belongs to $\{1, 2, 3\}$?

Given the information I conveyed, the six outcomes are no longer equally likely.

Instead, the outcome is one of $\{2, 4, 6\}$ – each being equally likely.

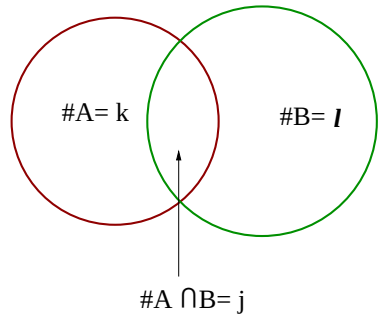
So with the information you have, the probability that the outcome belongs to $\{1, 2, 3\}$ equals 1/3.

Consider an experiment with m equally likely outcomes and let A and B be two events.

Given the information that B has happened, what is the probability that A occurs?

This probability is called the conditional probability of A given B

and written as $P(A \mid B)$.



Let $\#A = k$, $\#B = l$, $\#(A \cap B) = j$.

Given that B has happened, the new probability assignment gives a probability $1/l$ to each of the outcomes in B.

Out of these l outcomes of B,

$\#(A \cap B) = j$ outcomes also belong to A.

Hence

$$P(A \mid B) = j/l.$$

Noting that $P(A \cap B) = j/m$ and $P(B) = l/m$, it follows that

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

In general when A, B are events such that $P(B) > 0$, the conditional probability of A given that B has occurred $P(A \mid B)$ is defined by

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

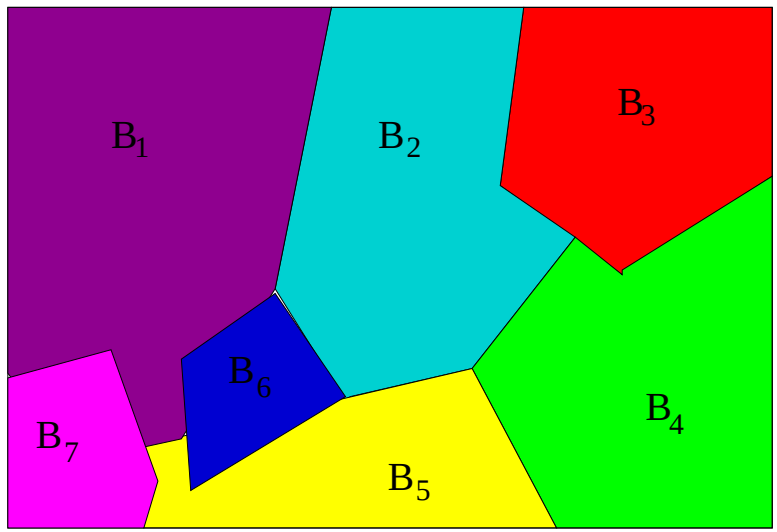
This leads to the Multiplicative law of probability

$$P(A \cap B) = P(A \mid B)P(B)$$

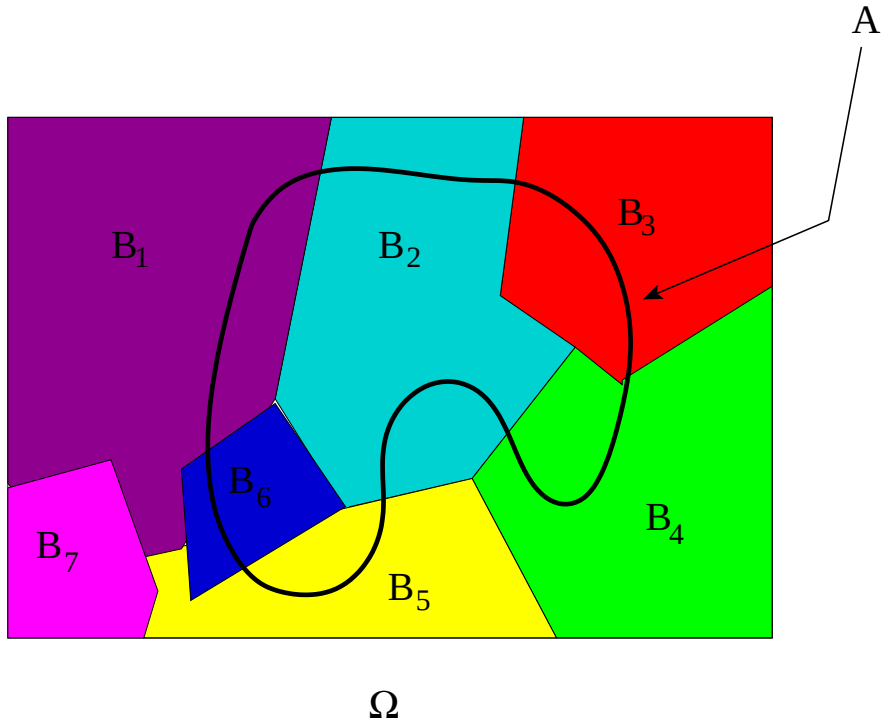
This has a generalization to n events:

$$
\begin{aligned}
&P(A_1 \cap A_2 \cap \ldots \cap A_n) \\
&= P(A_n \mid A_1, \ldots, A_{n-1}) \\
&\quad \times P(A_{n-1} \mid A_1, \ldots, A_{n-2}) \\
&\quad \times \ldots \times \\
&\quad \times P(A_2 \mid A_1) P(A_1)
\end{aligned}
$$

The Law of Total Probability Let $B_1, \ldots, B_k$ be a partition of the sample space Ω



The Law of Total Probability Let $B_1, \ldots, B_k$ be a partition of the sample space Ω and A an event



Then

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_k)$$

Also we know

$$P(A \cap B_i) = P(A|B_i)P(B_i)$$

so we get the Law of Total Probability

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k)$$

Example Suppose a bag has $\mathbf{6}$ one rupee coins, exactly one of which is a Sholay coin, i.e. both sides are HEAD. A coin is picked at random and tossed $\mathbf{4}$ times, and each toss yielded a HEAD. Two questions which may be asked here are

(i) what is the probability of the occurrence of $\boldsymbol{A} = \{$all four tosses yielded HEADS$\}$?

(ii) given that $\boldsymbol{A}$ occurred, what is the probability that the coin picked was the Sholay coin?

The first question is easily answered by the laws of total probability. Let

$B_1 =$ coin picked was a regular coin

$B_2 =$ coin picked was a Sholay coin

Then

$$\begin{aligned} P(A) &= P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) \\ &= \left(\frac{1}{2}\right)^4 \frac{5}{6} + \frac{1}{6} \\ &= \frac{21}{96} = \frac{7}{32} \end{aligned}$$

For the second question we need to find

$$P(B_2 \mid A)$$

For the second question we need to find

$$P(B_2 \mid A) = \frac{P(B_2 \cap A)}{P(A)}$$

For the second question we need to find

$$P(B_2 \mid A) = \frac{P(B_2 \cap A)}{P(A)}$$

$$= \frac{P(A \mid B_2)P(B_2)}{P(A)}$$

For the second question we need to find

$$
\begin{aligned}
P(B_2 \mid A) &= \frac{P(B_2 \cap A)}{P(A)} \\
&= \frac{P(A \mid B_2)P(B_2)}{P(A)} \\
&= \frac{1/6}{7/32} \\
&= \frac{16}{21}
\end{aligned}
$$

The previous example is atypical of the situation where we perform scientific experiments and make observations. On the basis of the observations we have to infer what was the theoretical process involved in the experiment to obtain the given observation. Occassionally we may have some (though not complete) information of the process, in which case we can use this information to help in our inference. In particular, in the example we had the prior information that there was exactly one Sholay coin among the six coins.

Suppose we have observed that A occurred. Let $B_1, \ldots, B_m$ be all possible scenarios under which A may occur, i.e. $B_1, \ldots, B_m$ is a partition of the sample space. To quantify our suspicion that B_i was the cause for the occurrence of A, we would like to obtain $P(B_i \mid A)$.

Bayes' formula or Bayes' theorem is the prescription to obtain this quantity. The theorem is very easy to establish and is the basis of Bayesian Inference.

Bayes' Theorem: If $B_1, B_2, \ldots, B_m$ is a partition of the sample space, then

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$

$$= \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{m} P(A \mid B_j)P(B_j)}$$

Suppose that A, B are events such that

$$P(A \mid B) = P(A)P(B).$$

Then we get

$$P(A \mid B) = P(A).$$

i.e. the knowledge that B has occurred has not altered the probability of A.

In this case, A and B are said to be independent events.

Let X, Y, Z be random variables each taking finitely many values. Then X, Y, Z are said to be independent if

$$P(X = i, Y = j, Z = k) = P(X = i)P(Y = j)P(Z = k)$$

for all possible values i, j, k of X, Y, Z respectively.

This can be generalized to finitely many random variables.

Expectation of a random variable Let X be a random variable taking values $x_1, x_2 \ldots, x_n$. The expected value μ of X (also called the mean of X) denoted by $E(X)$ is defined by

$$\mu = E(X) = \sum_{i=1}^{n} x_i P(X = x_i).$$

The variance of a random variable is defined by

$$\sigma^2 = \text{Var}(X) = E\{(X - \mu)^2\}.$$

Example Let X be a random variable taking values $+1$ or -1 with prob. $1/2$ each $\mu = E(X) = 0$ and $\sigma^2 = \text{Var}(X) = 1$

Example Let X be a random variable

taking values

$+1$ or -1

with prob. $1/2$ each

$\mu = E(X) = 0$

and

$\sigma^2 = \text{Var}(X) = 1$

taking values

$+10$ or -10

with prob. $1/2$ each

$\mu = E(X) = 0$

and

$\sigma^2 = \text{Var}(X) = 100$

Example Let X be a random variable

taking values	taking values
$+1$ or -1	$+10$ or -10
with prob. $1/2$ each	with prob. $1/2$ each
$\mu = E(X) = 0$	$\mu = E(X) = 0$
and	and
$\sigma^2 = \mathrm{Var}(X) = 1$	$\sigma^2 = \mathrm{Var}(X) = 100$

The variance of a random variable describes the spread of the values taken by the random variable.

Notation We will denote by CAPITAL letters the random variables, and by small letters the values taken by the random variables.

Thus X, Y, Z will stand for random variables, while x, y, z will stand for the values attained by the random variables X, Y, Z respectively.

Examples of random variables Consider n independent trials where the probability of success in each trial is p and let X denote the total number of successes, then

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots n$, $0 \leq p \leq 1$. This is known as Binomial distribution, written as $X \sim B(n, p)$. $E(X) = np$ and $Var(X) = np(1-p)$.

Consider a random variable X such that

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

for $k = 0, 1, 2, \ldots$ and $\lambda > 0$. This is known as Poisson distribution. Here
$E(X) = \lambda$ and $Var(X) = \lambda$.

If X has Binomial distribution $B(n, p)$ with large n and small p, then X can be approximated by a Poisson random variable Y with parameter $\lambda = np$, i.e.

$$P(X \leq a) \approx P(Y \leq a)$$

In order to consider random variables that may take any real number or any number in an interval as its value, we need to extend our notion of sample space and events. One difficulty is that we can no longer define probabilites for all subsets of the sample space. We will only note here that the class of events - namely the sets for which the probabilities are defined is large enough.

We also need to add an axiom called Countable additivity axiom: If $A_1, A_2, \ldots, A_k, \ldots$ are pairwise mutually exclusive events then

$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

A real valued function X on a sample space Ω is said to be a random variable if for all real numbers a, the set $\{\omega : X(\omega) \leq a\}$ is an event.

For a random variable X, the function F defined by

$$F(x) = P(X \leq x)$$

is called its distribution function. If there exists a function f such that

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

then f is called the density of X.

Examples of densities:

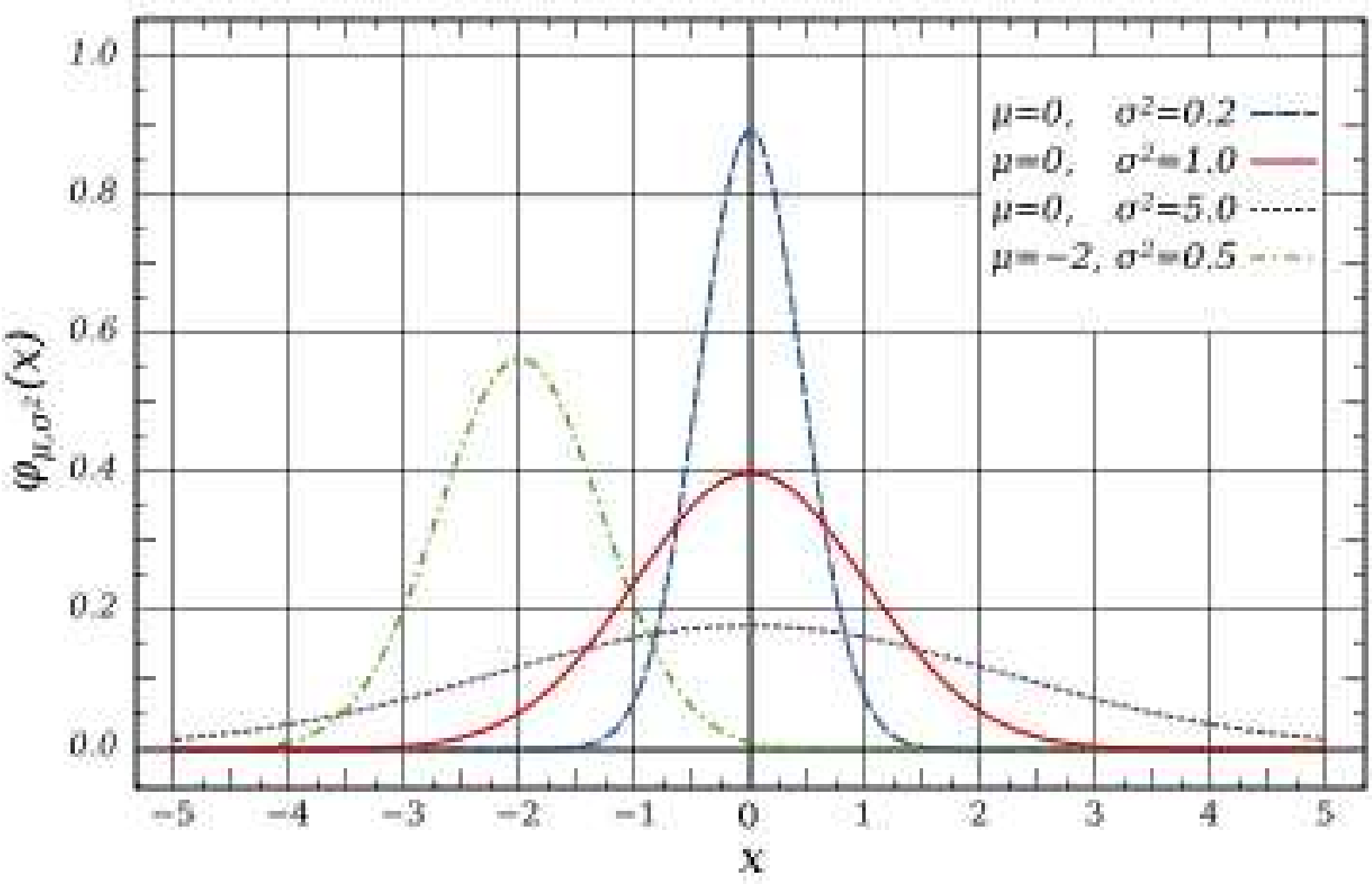
$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

This is called exponential density.

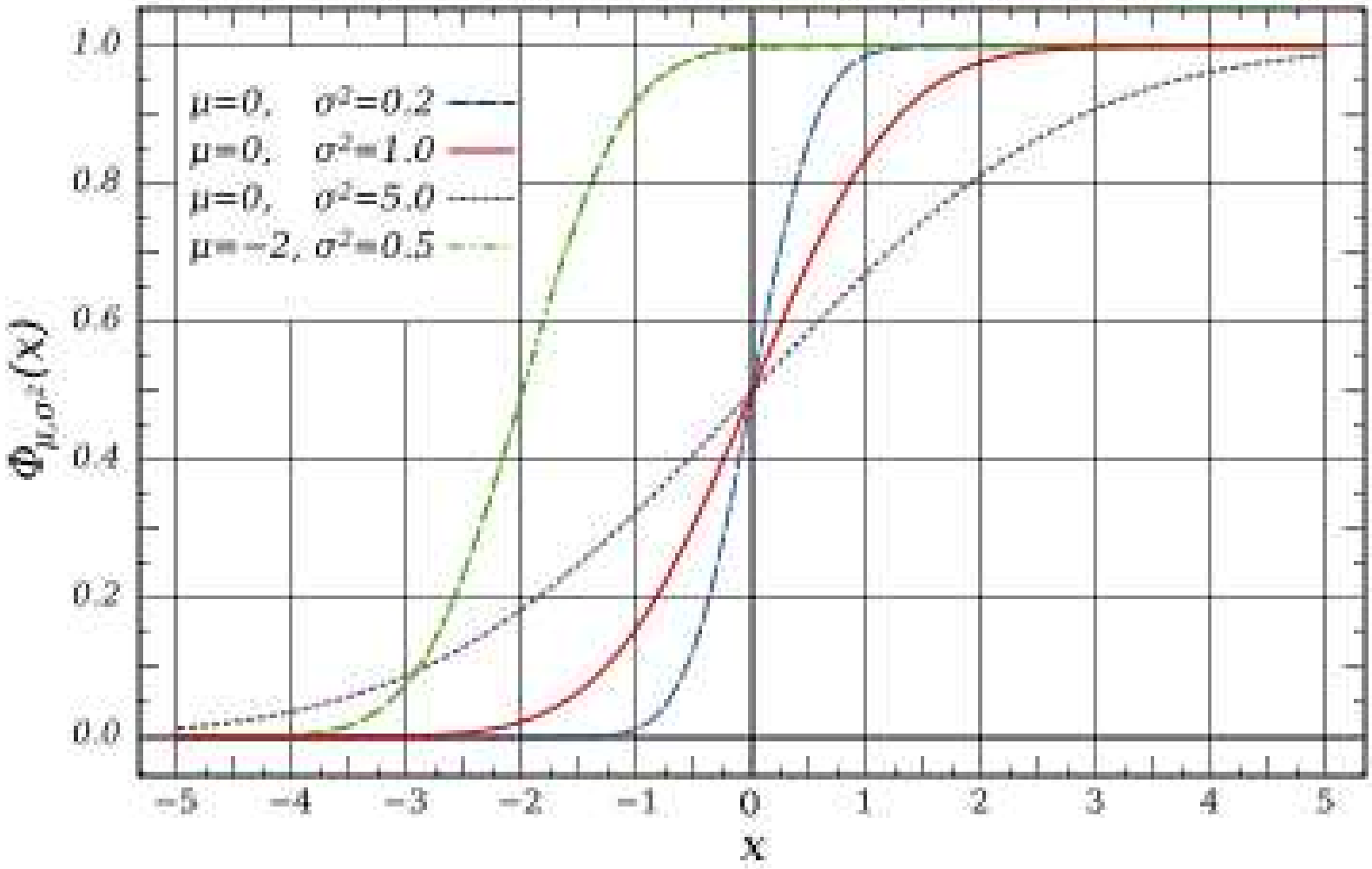
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

This is the Normal density.

Normal density function



Normal distribution function



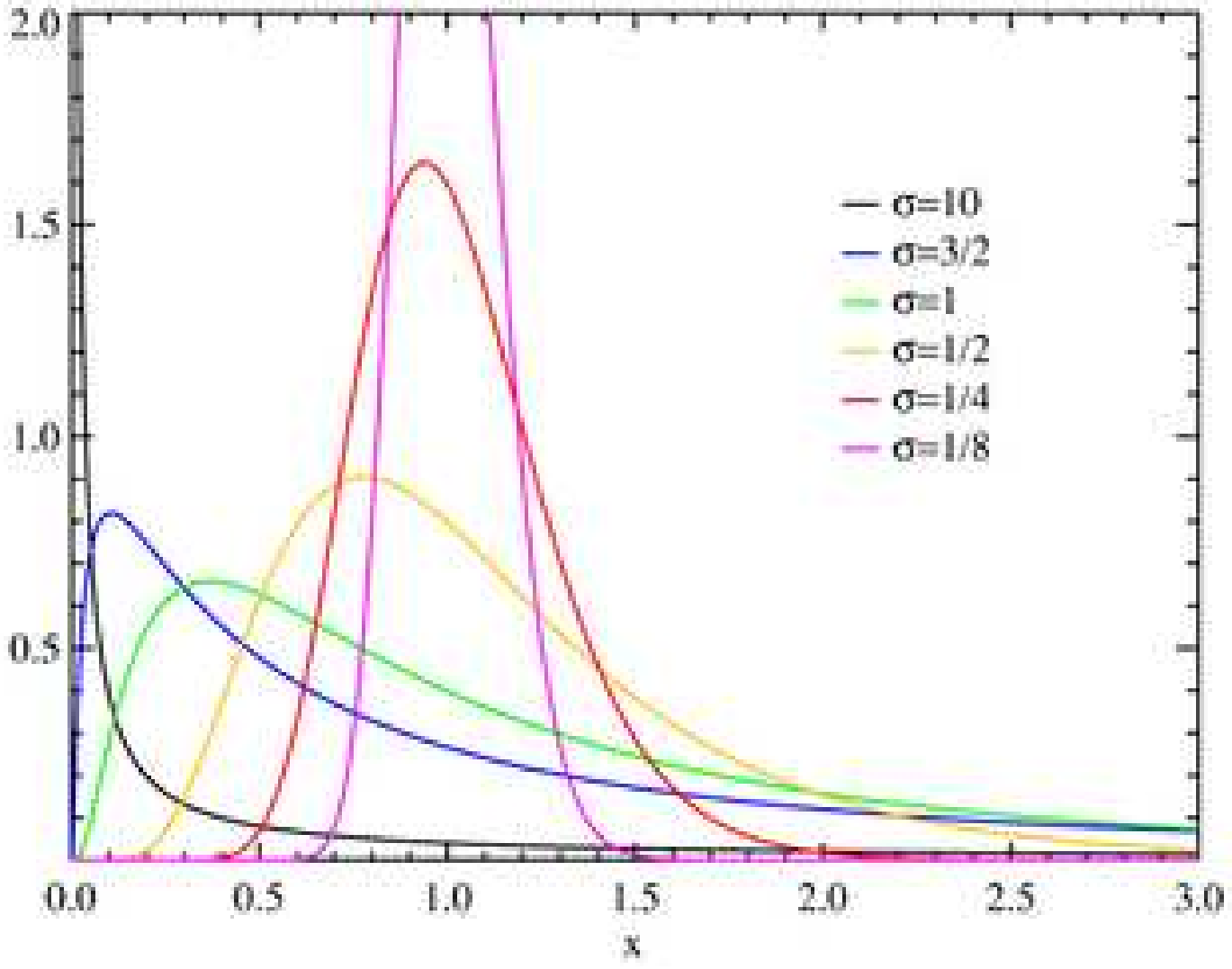
A very common density funtion encountered in astronomy is the globular cluster luminosity function GCLF.
A globular cluster (GC) is a collection of 104-106 ancient stars concentrated into a tight spherical structure structurally distinct from the field population of stars.

The distribution of GC luminosities (i.e. the collective brightness of all of its stars) is known as the globular cluster luminosity function (GCLF).

The shape of this function is said to be lognormal i.e.

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(\ln(x) - \mu)^2}{2\sigma^2} \right\} \text{ for } x > 0.$$

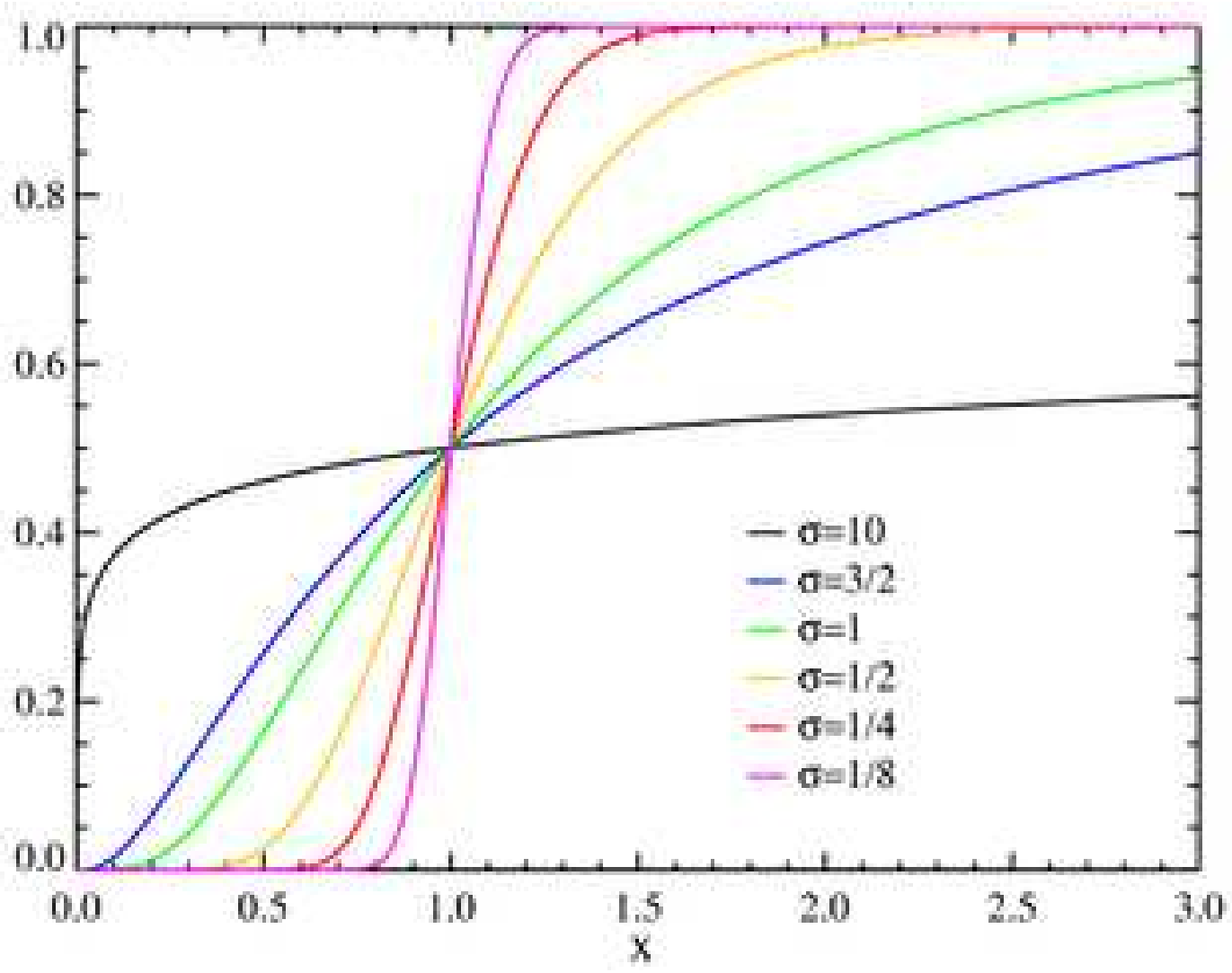
Lognormal density function



The distribution function of this random variable is difficult to compute explicitly.

It may be shown that if X is a normal random variable, then e^X has a log-normal distribution.

Lognormal distribution function



For a random variable X with density f, the expected value of X, where g is a function is defined by

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

For a random variable X with Normal density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$E(X) = \mu$$

$$Var(X) = E[(X-\mu)^2] = \sigma^2.$$

We write $X \sim N(\mu, \sigma^2)$.

If $X \sim N(0, 1)$ (i.e. the mean is 0 and variance is 1) then we call X a standard normal random variable and denote its density function and distribution function as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{(-z^2/2)}$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{(-z^2/2)} dz$$

The values of $\Phi(x)$ and those of $F(x)$ for other standard distributions are available in various computer spreadsheets.

For a random variable Y with lognormal density

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right\} \text{ for } x > 0.$$

$$E(X) = e^{\mu + (\sigma^2/2)}$$

$$Var(X) = E[(X - \mu)^2] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

When an experiment is conducted many times, e.g. a coin is tossed a hundred times, we are generating a sequence of random variables. Such a sequence is called a sequence of i.i.d. (independent identically distributed) random variables.

Suppose we gamble on the toss of a coin as follows – if HEAD appears then you give me $\mathbf{1}$ Rupee and if TAIL appears then you give me $-\mathbf{1}$ Rupee.

So if we play n round of this game we have generated i.i.d. sequence of random variables $X_1, \ldots, X_n$ where each X_i satisfies

$$X_i = \begin{cases} +1 & \text{with prob. } 1/2 \\ -1 & \text{with prob. } 1/2 \end{cases}$$

Now

$$S_n = X_1 + X_2 + \cdots + X_n$$

represents my gain after playing n rounds of this game.

Suppose we play the game n times and observe my gains/losses

OBSERVATION CHANCE

$S_{10} \leq -2$

i.e. I lost at least 6 out of 10

$S_{100} \leq -20$

i.e. I lost at least 60 out of 100

$S_{1000} \leq -200$

i.e. I lost at least 600 out of 1000

Suppose we play the game n times and observe my gains/losses

OBSERVATION	CHANCE
$S_{10} \leq -2$	
i.e. I lost at least 6 out of 10	moderate
$S_{100} \leq -20$	
i.e. I lost at least 60 out of 100	unlikely
$S_{1000} \leq -200$	
i.e. I lost at least 600 out of 1000	impossible

Suppose we play the game n times and observe my gains/losses

OBSERVATION	Probability CHANCE
$S_{10} \leq -2$	0.38
i.e. I lost at least 6 out of 10	moderate
$S_{100} \leq -20$	0.03
i.e. I lost at least 60 out of 100	unlikely
$S_{1000} \leq -200$	1.36^{-10}
i.e. I lost at least 600 out of 1000	impossible

OBSERVATION	Probability
$\lvert S_{10} \rvert \leq 1$	0.25
$\lvert S_{100} \rvert \leq 8$	0.56
$\lvert S_{1000} \rvert \leq 40$	0.8

OBSERVATION	PROPORTION	Probability
$\|S_{10}\| \leq 1$	$\frac{\|S_{10}\|}{10} \leq 0.1$	0.25
$\|S_{100}\| \leq 8$	$\frac{\|S_{100}\|}{100} \leq 0.08$	0.56
$\|S_{1000}\| \leq 40$	$\frac{\|S_{1000}\|}{1000} \leq 0.04$	0.8

Law of Large Numbers

Suppose $X_1, X_2, \ldots$ is a sequence of *i.i.d.* random variables with $E(|\ X_1\ |) < \infty$. Then

$$\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}$$

converges to $\mu = E(X_1)$: *i.e.* for all $\epsilon > 0$

$$P(|\ \bar{X}_n - \mu\ | > \epsilon) \longrightarrow 0.$$

Event	Probability	Normal
$\sqrt{1000}\frac{S_{1000}}{1000} \leq 0$	0.5126	$\Phi(0) = 0.5$
$\sqrt{1000}\frac{S_{1000}}{1000} \leq 1$	0.85	$\Phi(1) = 0.84$
$\sqrt{1000}\frac{\lvert S_{1000}\rvert}{1000} \leq 1.64$	0.95	$\Phi(1.64) = 0.95$
$\sqrt{1000}\frac{\lvert S_{1000}\rvert}{1000} \leq 1.96$	0.977	$\Phi(1.96) = 0.975$

For the sequence of *i.i.d.* random variables $X_1, X_2, \ldots$ with

$$X_i = \begin{cases} +1 & \text{with prob. } 1/2 \\ -1 & \text{with prob. } 1/2 \end{cases}$$

we have for $\bar{X}_n = S_n/n$,

$$P\left\{\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq x\right\} \longrightarrow \Phi(x)$$

Central Limit Theorem Suppose $X_1, X_2, \ldots$ is a sequence of *i.i.d.* random variables with $E(|\ X_1\ |^2) < \infty$. Let $\mu = E(X_1)$ and $\sigma^2 = E[(X_1 - \mu)^2]$. Let

$$\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}.$$

Then

$$P\left\{\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq x\right\} \longrightarrow \Phi(x)$$

$X \sim Binomial(n, p)$, n large. Then

$$P(X \leq a)$$

can be approximated by

$$\Phi\left(\sqrt{n}\left(\frac{a-np}{\sqrt{p(1-p)}}\right)\right)$$