1 Review of Probability

Random variables are denoted by X, Y, Z, etc. The *cumulative distribution function (c.d.f.)* of a random variable X is denoted by $F(x) = P(X \leq x),\ -\infty < x < \infty$, and if the random variable is continuous then its probability density function is denoted by $f(x)$ which is related to $F(x)$ via

$$\begin{aligned} f(x) &= F'(x) = \frac{d}{dx}F(x) \\ F(x) &= \int_{-\infty}^{x} f(y)dy. \end{aligned}$$

The *probability mass function (p.m.f.)* of a discrete random variable is given by

$$p(k) = P(X = k),\ -\infty < k < \infty,$$

for integers k.

$1 - F(x) = P(X > x)$ is called the *tail* of X and is denoted by $\overline{F}(x) = 1 - F(x)$. Whereas $F(x)$ increases to 1 as $x \to \infty$, and decreases to 0 as $x \to -\infty$, the tail $\overline{F}(x)$ decreases to 0 as $x \to \infty$ and increases to 1 as $x \to -\infty$.

If a r.v. X has a certain distribution with c.d.f. $F(x) = P(X \leq x)$, then we write, for simplicity of expression,

$$X \sim F. \tag{1}$$

1.1 Moments and variance

The expected value of a r.v. is denote by $E(X)$ and defined by

$$\begin{aligned} E(X) &= \sum_{k=-\infty}^{\infty} kp(k),\ \text{discrete case}, \\ E(X) &= \int_{-\infty}^{\infty} xf(x)dx,\ \text{continuous case}. \end{aligned}$$

$E(X)$ is also referred to as the *first moment* or mean of X (or of its distribution).

Higher moments $E(X^n),\ n \geq 1$ can be computed via

$$\begin{aligned} E(X^n) &= \sum_{k=-\infty}^{\infty} k^n p(k),\ \text{discrete case}, \\ E(X^n) &= \int_{-\infty}^{\infty} x^n f(x)dx,\ \text{continuous case}, \end{aligned}$$

and more generally $E(g(X))$ for a function $g = g(x)$ can be computed via

$$\begin{aligned} E(g(X)) &= \sum_{k=-\infty}^{\infty} g(k)p(k),\ \text{discrete case}, \\ E(g(X)) &= \int_{-\infty}^{\infty} g(x)f(x)dx,\ \text{continuous case}. \end{aligned}$$

(Leting $g(x) = x^n$ yields moments for example.)

Finally, the variance of X is denoted by $Var(X)$, defined by $E\{|X - E(X)|^2\}$, and can be computed via

$$Var(X) = E(X^2) - E^2(X), \tag{2}$$

the second moment minus the square of the first moment.

For any r.v. X and any number a

$$E(aX) = aE(X), \text{ and } Var(aX) = a^2 Var(X). \tag{3}$$

For any two r.v.s. X and Y

$$E(X + Y) = E(X) + E(Y). \tag{4}$$

If X and Y are independent, then

$$Var(X + Y) = Var(X) + Var(Y). \tag{5}$$

The above properties generalize in the obvious fashion to to any finite number of r.v.s.

In general (independent or not)

$$Var(X + Y) = Var(X) + V(Y) + 2Cov(X, Y),$$

where

$$Cov(X, Y) \stackrel{\text{def}}{=} E(XY) - E(X)E(Y).$$

When $Cov(X, Y) > 0$, X and Y are said to be *positively correlated*, whereas when $Cov(X, Y) < 0$, X and Y are said to be *negatively correlated*. When $Cov(X, Y) = 0$, X and Y are said to be *uncorrelated*, and in general this is weaker than independence of X and Y: *there are examples of uncorrelated r.v.s. that are not independent.*

1.2 Moment generating functions

The moment generating function (mgf) of a r.v. X (or its distribution) is defined for all $s \in (-\infty, \infty)$ by

$$\begin{aligned} M(s) &\stackrel{\text{def}}{=} E(e^{sX}) \\ &= \int_{-\infty}^{\infty} e^{sx} f(x) dx \quad \Big(= \sum_{-\infty}^{\infty} e^{sk} p(k) \text{ in the discrete r.v. case}\Big) \end{aligned} \tag{6}$$

It is so called because it generates the moments of X by differentiation at $s = 0$:

$$M'(0) = E(X), \tag{7}$$

and more generally

$$M^{(n)}(0) = E(X^n), \ n \geq 1. \tag{8}$$

The mgf uniquely determines a distribution in that no two distinct distributions can have the same mgf. So knowing a mgf characterizes the distribution in question.

If X and Y are independent, then $E(e^{s(X+Y)}) = E(e^{sX} e^{sY}) = E(e^{sX})E(e^{sY})$, and we conclude that *the mgf of an independent sum is the product of the individual mgf's.*

Sometimes to stress the particular r.v. X, we write $M_X(s)$. Then the above independence property can be concisely expressed as

$$M_{X+Y}(s) = M_X(s) M_Y(s), \text{ when } X \text{ and } Y \text{ are independent.}$$

Remark 1.1 For a given distribution, $M(s) = \infty$ is possible for some values of s, but there is a large useful class of distributions for which $M(s) < \infty$ for all s in *a neighborhood of the origin*, that is, for $s \in (-\epsilon, \epsilon)$ with $\epsilon > 0$ suffiently small. Such distributions are referred to as *light-tailed* because their tails can be shown to tend to zero quickly. There also exists distributions of non-negative r.v.s. for which $M(s) = \infty,\ s > 0$, and this can be so even if the distribution has finite moments of all orders (see the lognormal distribution for example). A large class of such distributions are referred to as *heavy-tailed* because their tails tend to zero slowly. An example of a very heavy-tailed distribution is the Pareto distribution; for $a > 0$ a constant its tail is of the form by $\overline{F}(x) = x^{-a},\ x \geq 1,\ \overline{F}(x) = 1,\ x \in [0, 1)$.

Remark 1.2 For non-negative r.v.s. X, it is sometimes more common to use the *Laplace transform*, $\mathcal{L}(s) = E(e^{-sX}),\ s \geq 0$, which is always finite, and then $(-1)^n \mathcal{L}^{(n)}(0) = E(X^n),\ n \geq 1$.

For discrete r.v.s. X, it is sometimes more common to use

$$M(z) = E(z^X) = \sum_{k=-\infty}^{\infty} z^k p(k),\ |z| \leq 1$$

for the mgf in which case moments can be generated via $M'(1) = E(X)$, $M''(1) = E((X)(X-1))$, $M^{(n)}(1) = E(X(X-1)\cdots(X-(n-1))),\ n \geq 1$.

1.3 Examples of well-known distributions

Discrete case

1. (Bernoulli distribution with success probability p) With $0 < p < 1$ a constant, X has p.m.f. $p(k) = P(X = k)$ given by

$$\begin{aligned} p(1) &= p, \\ p(0) &= 1 - p, \\ p(k) &= 0, \text{ otherwise.} \end{aligned}$$

Thus X only takes on the values 1 (success) or 0 (failure).

A simple computation yields

$$\begin{aligned} E(X) &= p \\ Var(X) &= p(1-p) \\ M(s) &= pe^s + 1 - p. \end{aligned}$$

Bernoulli r.v.s. arise naturally as the *indicator function*, $X = I\{A\}$, of an event A, where

$$I\{A\} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if the event } A \text{ occurs;} \\ 0, & \text{otherwise.} \end{cases}$$

Then $p = P(X = 1) = P(A)$ is the probability that the event A occurs. For example, if you flip a coin once and let $A = \{\text{coin lands heads}\}$, then for $X = I\{A\}$, $X = 1$ if the coin lands heads, and $X = 0$ if it lands tails. Because of this elementary and intuitive coin-flipping example, a Bernoulli r.v. is sometimes referred to as a coin flip, where p is the probability of landing heads.

Observing the outcome of a Bernoulli r.v. is sometimes called *performing a Bernoulli trial*, or experiment.

Keeping in the spirit of (1) we denote a Bernoulli p r.v. by

$$X \sim Bern(p).$$

2. (Binomial distribution with success probability p and n trials) If we consecutively perform n independent Bernoulli p trials, $X_1, \ldots, X_n$, then the total number of successes $X = X_1 + \cdots + X_n$ yields the Binomial r.v. with p.m.f.

$$p(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } 0 \le k \le n; \\ 0, & \text{otherwise.} \end{cases}$$

In our coin-flipping context, when consecutively flipping the coin exactly n times, $p(k)$ denotes the probability that exactly k of the n flips land heads (and hence exactly $n-k$ land tails).

A simple computation (utilizing $X = X_1 + \cdots + X_n$ and independence) yields

$$\begin{aligned} E(X) &= np \\ Var(X) &= np(1-p) \\ M(s) &= (pe^s + 1 - p)^n. \end{aligned}$$

Keeping in the spirit of (1) we denote a binomial n, p r.v. by

$$X \sim bin(n, p).$$

3. (geometric distribution with success probability p) The number of independent Bernoulli p trials required until the first success yields the geometric r.v. with p.m.f.

$$p(k) = \begin{cases} p(1-p)^{k-1}, & \text{if } k \ge 1; \\ 0, & \text{otherwise.} \end{cases}$$

In our coin-flipping context, when consecutively flipping the coin, $p(k)$ denotes the probability that the k^{th} flip is the first flip to land heads (all previous $k-1$ flips land tails). The tail of X has the nice form $\overline{F}(k) = P(X > k) = (1-p)^k,\ k \ge 0$.

It can be shown that

$$\begin{aligned} E(X) &= \frac{1}{p} \\ Var(X) &= \frac{(1-p)}{p^2} \\ M(s) &= \frac{pe^s}{1-(1-p)e^s}. \end{aligned}$$

(In fact, computing $M(s)$ is straightforward and can be used to generate the mean and variance.)

Keeping in the spirit of (1) we denote a geometric p r.v. by

$$X \sim geom(p).$$

Remark 1.3 As a variation on the geometric, if we change X to denote the number of failures before the first success, then (since the first flip might be a success yielding no failures at all), the p.m.f. becomes

$$p(k) = \begin{cases} p(1-p)^k, & \text{if } k \geq 0; \\ 0, & \text{otherwise,} \end{cases}$$

and $p(0) = p$. Then $E(X) = (1-p)p^{-1}$ and $Var(X) = (1-p)p^{-2}$. *Both of the above are called the geometric distribution.*

4. (Poisson distribution with mean (and variance) λ) With $\lambda > 0$ a constant, X has p.m.f.

$$p(k) = \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!}, & \text{if } k \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The Poisson distrubution has the interesting property that both its mean and variance are identical $E(X) = Var(X) = \lambda$. Its mgf is given by

$$M(s) = e^{\lambda(e^s - 1)}.$$

The Poisson distribution arises as an approximation to the binomial (n, p) distribution when n is large and p is small: Letting $\lambda = np$,

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda}\frac{\lambda^k}{k!}, \ 0 \leq k \leq n.$$

Keeping in the spirit of (1) we denote a Poisson λ r.v. by

$$X \sim Poiss(\lambda).$$

Continuous case

1. (uniform distribution on (a, b)) With a and b constants, X has density function

$$f(x) = \begin{cases} \frac{1}{b-a}; & \text{if } x \in (a, b) \\ 0, & \text{otherwise,} \end{cases}$$

c.d.f.

$$F(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } x \in (a, b); \\ 1, & \text{if } x \geq b; \\ 0, & \text{if } x \leq a, \end{cases}$$

and tail

$$\overline{F}(x) = \begin{cases} \frac{b-x}{b-a}, & \text{if } x \in (a, b); \\ 0, & \text{if } x \geq b; \\ 1, & \text{if } x \leq a. \end{cases}$$

A simple computation yields

$$\begin{aligned} E(X) &= \frac{a+b}{2} \\ Var(X) &= \frac{(b-a)^2}{12} \\ M(s) &= \frac{e^{sb} - e^{sa}}{s(b-a)}. \end{aligned}$$

When $a = 0$ and $b = 1$, this is known as the *uniform distribution over the unit interval*, and has density $f(x) = 1, x \in (0,1)$, $E(X) = 0.5$, $Var(X) = 1/12$, $M(s) = s^{-1}(e^s - 1)$.

Keeping in the spirit of (1) we denote a uniform (a,b) r.v. by

$$X \sim unif(a,b).$$

2. (exponential distribution) With $\lambda > 0$ a constant, X has density function

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

c.d.f.

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0, \end{cases}$$

and tail

$$\overline{F}(x) = \begin{cases} e^{-\lambda x}, & \text{if } x \geq 0; \\ 1, & \text{if } x < 0, \end{cases}$$

A simple computation yields

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ Var(X) &= \frac{1}{\lambda^2} \\ M(s) &= \frac{\lambda}{\lambda - s}. \end{aligned}$$

The exponential is famous for having the unique *memoryless property*,

$$P(X - y > x | X > y) = P(X > x), \ x \geq 0, \ y \geq 0,$$

in the sense that it is the unique distribution with this property.

(The geometric distribution satisfies a discrete version of this.)

Keeping in the spirit of (1) we denote an exponential λ r.v. by

$$X \sim exp(\lambda).$$

3. (normal distribution with mean μ and variance σ^2: $N(\mu, \sigma^2)$) The normal distribution is extremely important in applications because of the Central Limit Theorem (CLT). With $-\infty < \mu < \infty$ (the mean) and $\sigma^2 > 0$ (the variance) constants, X has density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \ -\infty < x < \infty.$$

This is also called the *Gaussian* distribution. We denote it by $N(\mu, \sigma^2)$. When $\mu = 0$ and $\sigma^2 = 1$ it is called the *standard* or unit normal, denoted by $N(0,1)$. If Z is $N(0,1)$, then $X = \sigma Z + \mu$ is $N(\mu, \sigma^2)$. Similarly, if X is $N(\mu, \sigma^2)$, then $Z = (x - \mu)/\sigma$ is $N(0,1)$. Thus the c.d.f. $F(x)$ can be expressed in terms of the c.d.f. of a unit normal Z. We therefore give the $N(0,1)$ c.d.f. the special notation $\Theta(x)$;

$$\Theta(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy,$$

and we see that

$$\begin{aligned} F(x) &= P(X \le x) \\ &= P(\sigma Z + \mu \le x) \\ &= P(Z \le (x-\mu)/\sigma) \\ &= \Theta((x-\mu)/\sigma). \end{aligned}$$

$\Theta(x)$ does not have a closed form (e.g., a nice formula that we can write down and plug into); hence the importance of good numerical recipes for computing it, and tables of its values.

The moment generating function of $N(\mu, \sigma^2)$ can be shown to be

$$M(s) = e^{s\mu + s^2\sigma^2/2}.$$

Keeping in the spirit of (1) we denote a $N(\mu, \sigma^2)$ r.v. by

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

4. (lognormal distribution) If Y is $N(\mu, \sigma^2)$, then $X = e^Y$ is a non-negative r.v. having the *lognormal distribution*; called so because its natural logarithm $Y = \ln(X)$ yields a normal r.v.

   X has density

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}}, & \text{if } x \ge 0; \\ 0, & \text{if } x < 0. \end{cases}$$

   Observing that $E(X)$ and $E(X^2)$ are simply the moment generating function of $N(\mu, \sigma^2)$ evaulated at $s = 1$ and $s = 2$ respectively yields

$$\begin{aligned} E(X) &= e^{\mu + \frac{\sigma^2}{2}} \\ Var(X) &= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \end{aligned}$$

   (It can be shown that $M(s) = \infty$ for any $s > 0$.)

   The lognormal distribution plays an important role in financial engineering since it is frequently used to model stock prices. As with the normal distribution, the c.d.f. does not have a closed form, but it can be computed from that of the normal via $P(X \le x) = P(Y \le \ln(x))$ due to the relation $X = e^Y$, and we conclude that $F(x) = \Theta((\ln(x)-\mu)/\sigma)$. Thus computations for $F(x)$ are reduced to dealing with $\Theta(x)$, the c.d.f. of $N(0, 1)$.

   Keeping in the spirit of (1) we denote a lognormal μ, σ^2 r.v. by

$$X \sim lognorm(\mu, \sigma^2).$$

5. (Pareto distribution) With constant $a > 0$, X has density

$$f(x) = \begin{cases} ax^{-(1+a)}, & \text{if } x \ge 1; \\ 0, & \text{if } x < 1, \end{cases},$$

c.d.f.

$$F(x) = \begin{cases} 1 - x^{-a}, & \text{if } x \geq 1; \\ 0, & \text{if } x \leq 1, \end{cases}$$

and tail

$$\overline{F}(x) = \begin{cases} x^{-a}, & \text{if } x \geq 1; \\ 1, & \text{if } x \leq 1. \end{cases}$$

(In many applications, a is an integer.) A simple computation yields

$$\begin{aligned} E(X) &= \frac{a}{a-1},\ a > 1;\ (= \infty \text{ otherwise}) \\ Var(X) &= \frac{a}{a-2} - \Big(\frac{a}{a-1}\Big)^2,\ a > 2;\ (= \infty \text{ otherwise}). \end{aligned}$$

(It can be shown that $M(s) = \infty$ for any $s > 0$.)

It is easily seen that $E(X^n) = \infty$ for all $n \geq a$: *The Pareto distribution has infinite moments for high enough n.* The Pareto distribution has the important feature that its tail $\overline{F}(x) = x^{-a}$ tends to 0, as $x \to \infty$, slower than does any exponential tail $e^{-\lambda x}$ or any lognormal tail. It is an example of a distribution with a very *heavy* or *fat* tail. Data suggests that the distribution of stock prices resembles the Pareto more than it does the widely used lognormal.

Keeping in the spirit of (1) we denote a Pareto a r.v. by

$$X \sim Pareto(a).$$

Remark 1.4 Variations on the Pareto distribution exist which allow the mass to start at different locations; $\overline{F}(x) = (c/(c+x))^a,\ x \geq 0$ with $c > 0$ and $a > 0$ constants for example.

1.4 Calculating expected vaues by integrating the tail

Given a continuous non-negative random variable X, we typically calculate, by definition, its *expected value* (also called its *mean*) via

$$E(X) \stackrel{\text{def}}{=} \int_0^\infty x f(x) dx,$$

where $f(x)$ is the density function of X. However, it is usually easier to calculate $E(X)$ by integrating the tail $\overline{F}(x)$:

Proposition 1.1 (Computing $E(X)$ via Integrating the Tail Method) *If X is a non-negative random variable, then $E(X)$ can be computed via*

$$E(X) = \int_0^\infty \overline{F}(x) dx. \tag{9}$$

Proof : Letting

$$I(x) = I\{X > x\} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } X > x; \\ 0, & \text{if } X \leq x, \end{cases}$$

denote the indicator function for the event $\{X > x\}$,

$$X = \int_0^X dx = \int_0^\infty I(x)dx,$$

which is easily seen by graphing the function $I(x)$ as a function of x (which yields a rectangle with length X and height 1, thus with area X). Taking expectations we conclude that

$$E(X) = E\{\int_0^\infty I(x)dx\}.$$

Finally, interchanging the order of integral and expected value (allowed since everything here is non-negative; formally this is an application of *Fubini's Theorem*) and recalling that $E(I\{B\}) = P(B)$ for any event B, where here $B = \{X > x\}$, yields

$$E(X) = \int_0^\infty E(I(x))dx = \int_0^\infty P(X > x)dx,$$

as was to be shown.

■

1.4.1 Examples

1. (*Exponential distribition:*) $\overline{F}(x) = e^{-\lambda x}$ which when integrated yields

$$E(X) = \int_0^\infty e^{-\lambda x}dx = \frac{1}{\lambda}.$$

Note that computing this expected value by integrating $xf(x) = x\lambda e^{-\lambda x}$ would require integration by parts.

2. (*Computing* $E(\min\{X, Y\})$ *for independent* X *and* Y *:*)

Consider two r.v.s. X and Y. Let $Z \stackrel{\text{def}}{=} \min\{X, Y\}$ (the minimum value of X and Y; $Z = X$ if $X \leq Y$, $Z = Y$ if $Y < X$). Then $P(Z > z) = P(X > z, Y > z)$ because the minimum of X and Y is greater than z if and only if both X and Y are greater than z. If we also assume that X and Y are independent, then $P(X > z, Y > z) = P(X > z)P(Y > z)$, and so (when $\min\{X, Y\}$ is non-negative) we can compute $E(Z)$ via

$$E(Z) = \int_0^\infty P(Z > z)dz = \int_0^\infty P(X > z)P(Y > z)dz,$$

a very useful result. For example, suppose $X \sim exp(1)$ and $Y \sim exp(2)$ are independent. Then for $Z = \min\{X, Y\}$, $P(Z > z) = e^{-z}e^{-2z} = e^{-3z}$ and we conclude that $E(Z) = \int_0^\infty e^{-3z}dz = 1/3$.

(In fact what we have shown here, more generally, is that the minimum of two independent expontially distributed r.v.s. is itself expontially distributed with rate as the sum of the individual rates.)

3. (*Computing* $E\{(X - k)^+\}$ *:*) For any number a, the *positive part* of a is defined by $a^+ \stackrel{\text{def}}{=} \max\{0, a\}$ (the maximum value of 0 and a; $a^+ = a$ if $a > 0$; 0 otherwise). For fixed $k \geq 0$ and any random variable X, with c.d.f. $F(x)$, let $Y = (X - k)^+$, the positive part of $X - k$. Then since Y is non-negative we can compute $E(Y)$ by integrating its tail.

But for any $x \geq 0$, it holds that $a^+ > x$ if and only if $a > x$, yielding $(X-k)^+ > x$ if and only if $(X-k) > x$; equivalently if and only if $X > x+k$. We thus conclude that $P(Y > x) = P(X > x+k) = \overline{F}(x+k)$ yielding

$$E(Y) = \int_0^\infty \overline{F}(x+k)dx = \int_k^\infty \overline{F}(x)dx,$$

where we changed variables, $u = x+k$, to obtain the last integral.

For example suppose $X \sim unif(0,1)$ and $k = 0.5$. Then

$$E(X-0.5)^+ = \int_{0.5}^\infty (1-x)dx = 0.125$$

Application of $E\{(X-k)^+\}$: Suppose that you own an option to buy a stock at price $K = 2$, at time $T = 6$ (months from now). The stock price at time $T = 6$ will have value X (random). You will exercise the option if and only if $X > 2$; and do nothing otherwise. Then $(X-2)^+$ is your profit at time T, and $E\{(X-2)^+\}$ your expected profit.

1.4.2 Computing higher moments

It can be shown more generally that the n^{th} moment of X, $E(X^n) = \int_0^\infty x^n f(x)dx$, can be computed as

$$E(X^n) = \int_0^\infty nx^{n-1}P(X > x)dx,\ n \geq 1, \tag{10}$$

yielding the second moment of X when $n = 2$:

$$E(X^2) = \int_0^\infty 2xP(X > x)dx. \tag{11}$$

1.4.3 Discrete-time case

Integrating the tail method also is valid for non-negative discrete-time random variables; the integral is replaced by a summation:

$$E(X) = \sum_{k=0}^\infty P(X > k).$$

For example, if X has a geometric distribution, then
$\overline{F}(k) = (1-p)^k,\ k \geq 0$ yielding

$$E(X) = \sum_{k=0}^\infty (1-p)^k = \frac{1}{p},$$

via the sum of a geometric series.

1.5 Strong Law of Large Numbers and the Central limit theorem (CLT)

A *stochastic process* is a collection of r.v.s. $\{X_t : t \in T\}$ with index set T. If $T = \{0, 1, 2, \ldots\}$ is the discrete set of integers, then we obtain a sequence of random variables $X_0, X_1, X_2, \ldots$ denoted by $\{X_n : n \geq 0\}$ (or just $\{X_n\}$). In this case we refer to the value X_n as the *state* of the process at time n. For example X_n might denote the stock price of a given stock at the end of the n^{th} day. If time n starts at $n = 1$, then we write $\{X_n : n \geq 1\}$ and so on. If time is continuous (meaning that the index set $T = [0, \infty)$) then we have a continuous-time stochastic process denoted by $\{X_t : t \geq 0\}$.

A very special (but important) case of a discrete-time stochastic process is when the r.v.s. are independent and identically distributed (i.i.d.). In this case there are two classical and fundamental results:

Theorem 1.1 (SLLN) *If $\{X_n : n \geq 1\}$ are i.i.d. with finite mean $E(X) = \mu$, then w.p.1.,*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mu, \ n \to \infty.$$

One of the practical consequences of the SLLN is that we can, for n large enough, use the approximation

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx E(X),$$

when trying to determine an apriori unknown mean. The SLLN is fundamental in *Monte Carlo Simulation.*

Theorem 1.2 (CLT) *If $\{X_n : n \geq 1\}$ are i.i.d. with finite mean $E(X) = \mu$ and finite non-zero variance $\sigma^2 = Var(X)$, then*

$$Z_n \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{n}}\Big(\sum_{i=1}^{n} X_i - n\mu\Big) \Longrightarrow N(0,1), \ n \to \infty, \ \textit{in distribution};$$

$\lim_{n\to\infty} P(Z_n \leq x) = \Theta(x), \ -\infty < x < \infty.$

If $\mu = 0$ and $\sigma^2 = 1$, then the CLT becomes

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \Longrightarrow N(0,1).$$

The CLT allows us to approximate sums of i.i.d. r.v.s. endowed with any c.d.f. F (even if unknown) by the c.d.f. of a normal, as long as the variance of F is finite. The famous normal approximation to the binomial distribution is but one example, for the binomial is the sum of i.i.d. Bernoulli r.v.s., and thus the CLT applies.