For $S = 1 + x + x^2 + \ldots + x^n$ where $|x| < 1$

$$\lim_{n\to\infty} S = \frac{1}{1-x}$$

For $S = 1 + 2x + 3x^2 + \ldots + nx^{n-1}$

$$\lim_{n\to\infty} S = \frac{1}{(1-x)^2}$$

Definition of e^x is

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

with **Taylor Series** expansion

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

Number of ways to choose k objects from n objects (pronounced n choose k) is

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

Binomial expansion says

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

Discrete Probability

A **random variable** is noted as an uppercase letter

example: Random variable X is the number of coin tosses until we see a heads

An **outcome** of a random variable is noted by a lowercase letter

example: Outcome $x = 2$ means it took two coin tosses to see a heads

Probability

The **conditional probability** of random variable X given event A has occurred is

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

The **law of total probability** conditions on n disjoint events

$$\mathbf{P}(X) = \sum_{i=1}^{n} \mathbf{P}(X|A_i)\,\mathbf{P}(A_i)$$

Combining these, **Bayes' theorem** says

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(A_i \cap B)\,\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_i \cap B)\,\mathbf{P}(A)}{\sum_{j=1}^{n} \mathbf{P}(B|A_i)\,\mathbf{P}(A_i)}$$

Two events X and Y are independent ($X \perp Y$) if

$$\mathbf{P}(X|Y) = \mathbf{P}(X)$$

and conditionally independent on event Z if

$$\mathbf{P}(X \cap Y|Z) = \mathbf{P}(X|Z) \cdot \mathbf{P}(Y|Z)$$

Expectation

The **expectation** of a random variable X is

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} x_i \cdot \mathbf{P}(X = x_i)$$

The **linearity of expectation** says for *any* two events X and Y

$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

The **expectation of product** says if $X \perp Y$ then

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$$

The **law of total expectation** uses conditioning on n disjoint events to say

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X|A_i] \cdot \mathbf{P}(A_i)$$

The expectation of a sum S of a random number of N independently and identically distributed (*i.i.d.*) random variables X_i is given by

$$S = \sum_{i=1}^{N} X_i$$
$$\mathbf{E}[S] = \mathbf{E}[N] \cdot \mathbf{E}[X]$$

Variance

The **variance** of a random variable X is

$$\mathbf{Var}(X) = \mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2$$

The **linearity of variance** says if $X \perp Y$ then

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$$

The **squared coefficient of variance** C_X^2 is useful in quantifying *variability*

$$C_X^2 = \frac{\mathbf{Var}(X)}{\mathbf{E}[X]^2}$$

Skew

The **skew** of a random variable X is

$$\mathbf{Skew}\left(X\right) = \mathbf{E}\left[\left(X - \mathbf{E}\left[X\right]\right)^3\right]$$

Z-Transform

The **z-Transform** $\hat{X}(z)$ of a random variable X is

$$\hat{X}(z) = \mathbf{E}\left[X^z\right] = \sum_i \mathbf{P}\left(X = i\right) \cdot z^i$$

z-Transforms can give all **moments** $\mathbf{E}\left[X^i\right]$ of random variable X by saying

$$\frac{d}{dz}\hat{X}(z)\Big|_{z=1} = \mathbf{E}\left[X\right]$$
$$\frac{d^2}{dz^2}\hat{X}(z)\Big|_{z=1} = \mathbf{E}\left[X \cdot (X-1)\right]$$
$$\frac{d^3}{dz^3}\hat{X}(z)\Big|_{z=1} = \mathbf{E}\left[X \cdot (X-1) \cdot (X-2)\right]$$

The **sum in z-Transforms** of two random variables X and Y so $W = X + Y$ is

$$\hat{W}(z) = \hat{X}(z) \cdot \hat{Y}(z)$$

To **condition on z-Transforms** define random variables X, A, and B so

$$X = \begin{cases} A & \text{with probability } p \\ B & \text{with probability } 1-p \end{cases}$$

then

$$\hat{X}(z) = p \cdot \hat{A}(z) + (1-p) \cdot \hat{B}(z)$$

Discrete Distributions

	$Bernoulli(p)$	$Binomial(n,p)$	$Geometric(p)$	$Poisson(\lambda)$
event values	$k \in \{0, 1\}$	$k \in \{0, \ldots, n\}$	$k \in \{1, 2, \ldots\}$	$k \in \{0, 1, \ldots\}$
$\mathbf{P}\left(X = k\right)$	$\begin{cases} p & k=1 \\ 1-p & k=0 \end{cases}$	$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$	$p \cdot (1-p)^{k-1}$	$\frac{\lambda^k \cdot e^{-\lambda}}{k!}$
$\mathbf{E}\left[X\right]$	p	$n \cdot p$	$\frac{1}{p}$	λ
$\mathbf{Var}\left(X\right)$	$p \cdot (1-p)$	$n \cdot p \cdot (1-p)$	$\frac{1-p}{p^2}$	λ
$\mathbf{Skew}\left(X\right)$	$\frac{1-2p}{\sqrt{p \cdot (1-p)}}$	$\frac{1-2p}{\sqrt{n \cdot p \cdot (1-p)}}$	$\frac{2-p}{\sqrt{1-p}}$	$\lambda^{-\frac{1}{2}}$
$\hat{X}(z)$	$p \cdot z + 1 - p$	$(p \cdot z + 1 - p)^n$	$\frac{z \cdot p}{1 - z \cdot (1-p)}$	$e^{\lambda \cdot (z-1)}$

Continuous Probability

For continuous random variables, the probability of an exact value is zero

$$\mathbf{P}\left(X = x\right) = 0$$

Instead, a **probability density function** $f_X(x)$ describes probability over an interval

$$\mathbf{P}\left(a < X < b\right) = \int_{x=a}^{b} f_X(x) \cdot dx$$

The **cumulative density function** $F_X(x)$ gives probability of being less than a value

$$F_X(x) = \mathbf{P}\left(X < x\right) = \int_{t=-\infty}^{x} f_X(t) \cdot dt$$

The **complementary cumulative density function** $\overline{F_X}(x)$ gives probability of being greater than a value

$$\overline{F_X}(x) = \mathbf{P}\left(X > x\right) = \int_{t=x}^{\infty} f_X(t) \cdot dt$$

The **conditional probability density function** $f_{X|A}(x)$ of random variable X given A is

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathbf{P}(A)} & \text{when } x \in A \\ 0 & \text{otherwise} \end{cases}$$

Multiple Random Variables

The **joint probability** $f_{X,Y}(x, y)$ for two random variables X and Y is

$$\int_{y=c}^{d} \int_{x=a}^{b} f_{X,Y}(x, y) \cdot dx \cdot dy$$

and

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X,Y}(x, y) \cdot dy$$

If $X \perp Y$ then

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

Use **conditioning on multiple random variables** to find the probability of one random variable less than another

$$\mathbf{P}\left(X_A < X_B\right) = \int_{t=-\infty}^{\infty} \mathbf{P}\left(t < X_B\right) \cdot f_{X_A}(t) \cdot dt$$

Expectation

The **expectation** of a random variable X is

$$\mathbf{E}[X] = \int_{x=-\infty}^{\infty} x \cdot f_X(x) \cdot dx$$

The **second moment** is

$$\mathbf{E}\left[X^2\right] = \int_{x=-\infty}^{\infty} x^2 \cdot f_X(x) \cdot dx$$

The **conditional expectation** $\mathbf{E}[X|A]$ of random variable X given A is

$$\mathbf{E}[X|A] = \int_{x=-\infty}^{\infty} x \cdot f_{X|A}(x) \cdot dx$$

Rate and Memory

The **failure rate function** $r_X(t)$ of random variable X is

$$r_X(t) = \frac{f_X(t)}{\overline{F_X}(t)}$$

Increasing rate in t implies the longer it takes, the more likely an event will occur.
Decreasing rate in t implies the longer it takes, the less likely an event will occur.

A distribution is **memoryless** if the failure rate is constant. If $X \sim Exponential(\lambda)$ then

$$r_X(t) = \lambda$$

Memoryless also implies that

$$\mathbf{P}(X > t + s | X > s) = \mathbf{P}(X > t)$$

Normal Distribution

A **linear transformation** of a normal distribution is a normal distribution. If $X \sim Normal(\mu, \sigma^2)$ and $Y = a \cdot X + b$ then

$$Y \sim Normal\left(a \cdot \mu + b,\ a^2 \cdot \sigma^2\right)$$

A **transformation to standard normal** $Normal(0,1)$ can be made if $X \sim Normal(\mu, \sigma^2)$ then

$$\frac{X - \mu}{\sigma} \sim Normal(0,1)$$

The **central limit theorem** says if S_n is a sum of independently identically distributed ($i.i.d$) random variables X_i and

$$S = \sum_{i=1}^{n} X_i$$
$$Z_n = \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}}$$

then

$$\lim_{n\to\infty} \mathbf{P}\left(Z_n \leq z\right) = \frac{1}{\sqrt{2\pi}} \cdot \int_{x=-\infty}^{z} e^{-\frac{x^2}{2}} \cdot dt$$

Continuous Distributions

	$Uniform(a,b)$	$Exponential(\lambda)$	$Pareto(\alpha)$	$Normal(\mu,\sigma)$
bounds	$x \; \epsilon \; [a, \, b]$	$x \; \epsilon \; [0, \, \infty)$	$x \; \epsilon \; [0, \, \infty)$	$x \; \epsilon \; (-\infty, \, \infty)$
$f_X(x)$	$\begin{cases} \frac{1}{b-a} & x \; \epsilon \; [a,b] \\ 0 & otherwise \end{cases}$	$\lambda \cdot e^{-\lambda \cdot x}$	$\alpha \cdot x^{-(\alpha+1)}$	$\frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}$
$F_X(x)$	$\begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \; \epsilon \; [a,b] \\ 1 & x \geq b \end{cases}$	$1 - e^{-\lambda \cdot x}$	$1 - x^{-\alpha}$	
$\mathbf{E}\left[X\right]$	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	$\begin{cases} \infty & \alpha \leq 1 \\ \alpha & \alpha > 1 \end{cases}$	μ
$\mathbf{Var}\left(X\right)$	$\frac{(b-a)^2}{12}$	$\frac{1}{\lambda^2}$	$\begin{cases} \infty & \alpha \leq 2 \end{cases}$	σ^2