Inclusion-Exclusion Principle. Let $E_1, E_2, \cdots, E_n$ be any events, then

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) = & \sum_{i=1}^{n} P(E_i) - \sum_{1\le i_1<i_2\le n} P(E_{i_1}\cap E_{i_2}) + \cdots \\ & + (-1)^{r+1} \sum_{1\le i_1<\cdots<i_r\le n} P(E_{i_1}\cap\cdots\cap E_{i_r}) + \cdots \\ & + (-1)^{n+1} P(E_1\cap\cdots\cap E_n) \end{aligned}$$

Probability as a Continuous Set Function

- A sequence of events $\{E_n\}, n \ge 1$ is said to be an *increasing* sequence if
$$E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$$
whereas it is said be *decreasing* sequence if
$$E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$$
- If $\{E_n\}, n \ge 1$ is an *increasing* sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ as
$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$
Similarly, if $\{E_n\}, n \ge 1$ is a *decreasing* sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$ as
$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$
- If $\{E_n\}, n \ge 1$ is an *increasing* or a *decreasing* sequence of events, then
$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$$

Tail Sum Formula for Expectation. For non-negative integer-valued random variable X (that is, X takes values $0, 1, 2, \cdots$),

$$E(X) = \sum_{k=1}^{\infty} P(X \ge k) = \sum_{k=0}^{\infty} P(X \ge k)$$

Tail Sum Formula (Continuous). Suppose X is a *non-negative* continuous random variable, then

$$E(X) = \int_0^{\infty} P(X > x)\,dx = \int_0^{\infty} P(X \ge x)\,dx$$

Expectation of a Function of a Random Variable. If X is a discrete random variable that takes values x_i, $i \ge 1$, with respective probabilities $p_X(x_i)$, then for any real value function g.

$$\begin{aligned} E[g(X)] &= \sum_i g(x_i)\,p_X(x_i) \quad \text{or equivalently} \\ &= \sum_x g(x)\,p_X(x) \end{aligned}$$

Bernoulli Random Variable. A Bernoulli random variable, denoted by $Be(p)$, is defined by

$$X = \begin{cases} 1, & \text{if it is a success;} \\ 0, & \text{if it is a failure;} \end{cases}$$

It is used to model a trial in which a particular event occurs or does not. Occurrence of this event is called success and non-occurrence is called failure. Each trial has a probability of success of p and a probability of failure of $q = 1 - p$

$$E(X) = p \quad Var(X) = p(1-p)$$

Binomial Random Variable. A binomial random variable, denoted by $Bin(n, p)$, is defined by

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

where X represent the number of success in n independent Bernoulli(p) trials. Hence X takes values $0, 1, 2, \cdots, n$

$$E(X) = np \quad Var(X) = np(1-p)$$

Note: Let $X_i, i = 1, \cdots, n$ be n independent $Bernoulli(p)$ random variables. Then

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

where $X \sim Bin(n, p)$

Geometric Random Variable. A geometric random variable, denoted by $Geom(p)$, is defined as

$$P(X = k) = pq^{k-1}$$

where X represents the number of $Bernoulli(p)$ trials required to obtain the first success. Therefore, X takes values $1, 2, 3, \cdots$, and so on.

$$E(X) = \frac{1}{p} \quad Var(X) = \frac{1-p}{p^2}$$

Similarly, we could let X' represent the number of failures of $Bernoulli(p)$ trial to obtain the first success.

In that case,

$$X' = X - 1$$

$$P(X' = k) = pq^k$$

$$E(X') = \frac{1-p}{p} \quad Var(X') = \frac{1-p}{p^2}$$

Negative Binomial Random Variable. A negative binomial random variable, denoted bye $NB(r, p)$, is defined as

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

where X represents the number of $Bernoulli(p)$ trials required to obtain r successes. Therefore, X takes values $r, r+1, \cdots$, and so on.

$$E(X) = \frac{r}{p} \quad Var(X) = \frac{r(1-p)}{p^2}$$

Remark: Note that $Geom(p) = NB(1, p)$

Poisson Random Variable. A Poisson random variable, denoted by $P(\lambda)$, defined as

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

where X represents the number of occurrences of an event within a given time interval where λ is the average number of occurrences of that event within the same time interval. Therefore, X takes values $0, 1, 2, \cdots$, and so on.

$$E(X) = \lambda \quad Var(X) = \lambda$$

Hypergeometric Random Variable. A hypergeometric random variable, denoted by $H(n, N, m)$, defined as

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

Suppose that we have a set of N balls, of which m are red and $N - m$ are blue. We choose n of these balls, ***without replacement***. X represents the number of red balls in our sample.

$$E(X) = \frac{nm}{N} \quad Var(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right]$$

Expectation of a Function of a Continuous Random Variable. If X is a continuous random variable with probability density function f_X, then for any real value function $g(X)$.

(i) $E[g(X)] = \int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx$

(ii) Linearity Property

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2 Var(X)$$

$$SD(aX + b) = |a|\,SD(X)$$

(iii) Formula for variance.

$$Var(X) = E(X^2) - [E(X)]^2$$

Uniform Distribution. A random variable X is said to be uniformly distributed over the interval (a, b), denoted by $X \sim U(a, b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b; \\ 0, & \text{otherwise;} \end{cases}$$

and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy = \begin{cases} 0, & \text{if } x < a; \\ \dfrac{x-a}{b-a}, & \text{if } a \le x < b \\ 1, & \text{if } b \le x \end{cases}$$

It can be shown that

$$E(X) = \frac{a+b}{2} \quad Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution. A random variable X is said to be normally distributed with parameters μ and σ^2, denoted by $X \sim N(\mu, \sigma^2)$ if its probability density function is given by

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

Standard Normal Distribution. A normal variable is called a standard normal when $\mu = 0$ and $\sigma = 1$ and is denoted by Z, that is $Z \sim N(0, 1)$. The probability density function is denoted by ϕ and its distribution function is denoted by Φ

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

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

Let $Y \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$, then

$$\begin{aligned} P(a < Y \le b) &= P\left(\frac{a-\mu}{\sigma} < Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Exponential Distribution. A random variable X is said to be exponentially distributed with parameter $\lambda > 0$, denoted by $X \sim Exp(\lambda)$, if its probability density function is given by

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

and its distribution function is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x \le 0; \\ 1 - e^{-\lambda x}, & \text{if } x > 0; \end{cases}$$

It can be shown that

$$E(X) = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2}$$

The exponential distribution also has a memoryless property.

$$P(X > s + t \mid X > s) = P(X > t), \quad \text{for } s, t > 0$$

Exponential Distribution is usually used to model the time between events, where events occur continuously and independently at a constant average rate

Gamma Distribution. A random variable X is said to have a gamma distribution with parameters (α, λ), denoted by $Gamma(\alpha, \lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } x \ge 0; \\ 0, & \text{if } x < 0; \end{cases}$$

where $\lambda > 0, \alpha > 0$ and $\Gamma(\alpha)$, called the gamma function, is defined by

$$\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1}\,dy$$

If $X \sim Gamma(\alpha, \lambda)$, then

$$E(X) = \frac{\alpha}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2}$$

Remarks Regarding the Gamma Distribution

(a) $\Gamma(1) = \int_0^{\infty} e^{-y}\,dy = 1$

(b) It can be shown, via integration by parts, that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

(c) for integer values of α, say $\alpha = n$,

$$\Gamma(n) = (n-1)!$$

(d) $Gamma(1, \lambda) = Exp(\lambda)$

(e) If $X_i \sim Exp(\lambda)$ independently, then $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$

(f) If $X \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, then $X \sim \chi^2(n)$

(g) $\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} e^{-y} y^{-\frac{1}{2}}\,dy = \sqrt{\pi}$

(h) **Interpretation of Gamma Distribution when $\alpha = n$.** If events are occurring randomly in time then the amount of time one has to wait until a total of n events has occured is a random variable which follows a Gamma distribution with parameters (n, λ)

(i) **Weibull Distribution.** A random variable X is said to have a Weibull Distribution with parameters (ν, α, β), denoted by $W(\nu, \alpha, \beta)$, if its probability density function is given by

(j) $$f_X(x) = \begin{cases} \dfrac{\beta}{\alpha}\left(\dfrac{x-\nu}{\alpha}\right)^{\beta-1} \exp\left(-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right), & \text{if } x > \nu; \\ 0, & \text{if } x \le \nu; \end{cases}$$

(k) If $X \sim W(\mu, \alpha, \beta)$, then

(l) $E(X) = \alpha\Gamma\left(1 + \frac{1}{\beta}\right) \quad Var(X) = \alpha^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2\right]$

(m) Remark: The $Exp(\lambda)$ is a special case of a Weibull distribution with $\alpha = 1, \beta = \lambda$ and $\nu = 0$.

Cauchy Distribution. A random variable X is said to follow a Cauchy distribution with parameter θ and α, where $-\infty < \theta < \infty$ and $\alpha > 0$ if its density is given by

$$f_X(x) = \frac{1}{\pi\alpha\left[1 + \left(\frac{x-\theta}{\alpha}\right)^2\right]}, \text{ for } -\infty < x < \infty$$

Both $E(X)$ and $Var(X)$ do not exist.

Beta Distribution. A random variable X is said to have a Beta distribution with parameters (a, b), denoted by $Beta(a, b)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \dfrac{1}{\mathrm{B}(a,b)} x^{a-1}(1-x)^{b-1}, & \text{if } 0 < x < 1; \\ 0, & \text{otherwise;} \end{cases}$$

where $-\infty < a, b < \infty$ and $\mathrm{B}(a, b)$, called the Beta function, is defined by

$$\mathrm{B}(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$$

If $X \sim Beta(a, b)$, then

$$E(X) = \frac{a}{a+b} \quad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Remarks:

1. $U(0, 1)$ is a special Beta distribution. $Beta(1, 1) \equiv U(0, 1)$
2. It can be shown that

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Normal Approximation of Binomial Random Variable. Suppose that $X \sim Bin(n, p)$. Then for any $a < b$,

$$P\left(a < \frac{X - np}{\sqrt{npq}} \le b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$, where $q = 1 - p$ and $\Phi(z) = P(Z \le z)$ with $Z \sim N(0,1)$

That is,

$$Bin(n, p) \approx N(np, npq)$$

Equivalently,

$$\frac{X - np}{\sqrt{npq}} \approx Z$$

where $Z \sim N(0, 1)$

Remark: The normal approximation will be generally good for values of n satisfying $np(1-p) \ge 10$

The approximation can be further improved with continuity correction.

Continuity-correction. If $X \sim Bin(n, p)$, then

$$P(X = k) = P\left(k - \frac{1}{2} < x < k + \frac{1}{2}\right)$$

$$P(X \ge k) = P\left(X \ge k - \frac{1}{2}\right)$$

$$P(X \le k) = P\left(X \le k + \frac{1}{2}\right)$$

Poisson Approximation of Binomial Random Variable. The Poisson distribution is used as an approximation to the binomial distribution when the parameter n and p are large and small, respectively and that np is moderate.

As a working rule, use the Poisson approximation if $p < 0.1$ and put $\lambda = np$. If $p > 0.9$, put $\lambda = n(1-p)$ and work in terms of "failure".

Distribution of a Function of a Random Variable. Let X be a continuous random variable having a probability density function f_X. Suppose that $g(x)$ is a strictly monotonic, differentiable function of X. Then the random variable Y defined by $Y = g(X)$ has probability density function given by

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\left|\dfrac{d}{dy}g^{-1}(y)\right|\right), & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \ne g(x) \text{ for all } x; \end{cases}$$

where $g^{-1}(y)$ is defined to be equal that value of X such that $g(x) = y$.

Joint Probability Distribution Function of Functions of Random Variables. Let X and Y be jointly distributed random variables with joint probability density function $f_{X,Y}(x,y)$. It is sometime necessary to obtain the joint distribution of the random variables U and V, which arise as functions of X and Y.

Specifically, suppose that

$$U = g(X,Y) \quad V = h(X,Y)$$

for some functions g and h.

We want to find the joint probability function of U and V in terms of the joint probability density function $f_{X,Y}(x,y)$, g and h.

Assume the following conditons are satisfied.

1. Let X and Y be jointly continuous distributed random variables with known joint probability density function.
2. Let U and V be given functions of X and Y in the form:
$$U = g(X,Y) \quad V = h(X,Y)$$
And we can uniquely solve X and Y in terms of U and V, say $x = a(u,v)$ and $y = b(u,v)$
3. The functions g and h have continuous partial derivatives at all points (x,y) and
$$J(x,y) := \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\frac{\partial h}{\partial y} - \frac{\partial g}{\partial y}\frac{\partial h}{\partial x} \neq 0$$
at all points (x,y)

The joint probability density function of U and V is given by

$$f_{U,V}(u,v) = f_{X,Y}(x,y)\,|J(x,y)|^{-1}$$

where $x = a(u,v)$ and $y = b(u,v)$

The Bivariate Normal Distribution. We say that the random variables X, Y have a bivariate normal distribution if, for constant $\mu_x, \mu_y, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$f_{X,Y}(x,y) := \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right)$$

It can be shown that $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left(-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right)$ hence $X \sim N(\mu_X, \sigma_X^2)$

and $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y}\exp\left(-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right)$ hence $Y \sim N(\mu_Y, \sigma_Y^2)$

Expectation of Functions of Random Variables.

(a) If X and Y are jointly discrete with joint probability mass function $p_{X,Y}(x,y)$, then
$$E[g(X,Y)] = \sum_y \sum_x g(x,y)p_{X,Y}(x,y)$$

(b) If X and Y are jointly continuous with joint probability mass function $f_{X,Y}(x,y)$, then
$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\,dx\,dy$$

Boole's Inequality

Let $A_1, \cdots, A_n$ denote events and define the indicator variable $I_k, k = 1, \cdots, n$, by

$$I_k \begin{cases} 1 & \text{if } A_k \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

it can be shown that

$$P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)$$

Covariance. The covariance of jointly distributed random variable X and Y, denoted by $Cov(X,Y)$, is defined by

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Alternate Formula for Covariance

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Expectation and Independence. If X and Y are independent, then for any function $g, h: \mathbb{R} \to \mathbb{R}$, we have

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Variance of a Sum

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k) + 2\sum_{1\leq i<j\leq n} Cov(X_i, X_j)$$

If $X_i, \cdots, X_n$ are independent random variables, then

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k)$$

Hence, under independence, variance of sum is equal to the sum of variances.

Conditional Expectation.

(1) If X and Y are jointly distributed discrete random variables, then
$$E[X|Y=y] = \sum_x x p_{X|Y}(x|y), \quad \text{if } p_Y(y) > 0$$

(2) If X and Y are jointly distributed continuous random variables, then
$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx, \quad \text{if } f_Y(y) > 0$$

Important Formula Regarding Conditional Expectation.

(1)
$$E[g(X)|Y=y] = \begin{cases} \displaystyle\sum_x g(x)p_{X|Y}(x|y), & \text{for discrete case;} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y), & \text{for continuous case;} \end{cases}$$

(2)
$$E\left[\sum_{k=1}^{n} X_k | Y = y\right] = \sum_{k=1}^{n} E[X_k|Y=y]$$

Computing Probability by Conditioning. Let $X = I_A$ where A is an event, then we have

$$E(I_A) = P(A)$$
$$E(I_A|Y=y) = P(A|Y=y)$$

and hence

$$P(A) = \begin{cases} \displaystyle\sum_y P(A|Y=y)p(Y=y), & \text{for } Y \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} P(A|Y=y)\,f_Y(y)\,dy, & \text{for } Y \text{ is continuous} \end{cases}$$

Law of Total Variance.

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

Moment Generating Functions. The moment generating function of random variable X, denoted by M_X, is defined as

$$M_X(t) = E[e^{tX}] = \begin{cases} \displaystyle\sum_x e^{tx}f_X(x), & \text{if } X \text{ is discrete with pmf } p_X(x) \\ \displaystyle\int_{-\infty}^{\infty} e^{tx}f_X(x)\,dx, & \text{if } X \text{ is continous with pdf } f_X(x) \end{cases}$$

This function generates all the moments of this random variable X

For $n \geq 0$,

$$E(X^n) = M_X^{(n)}(0)$$

where

$$M_X^{(n)}(0) = \frac{d^n}{dt^n}M_X(t)\,\Big|_{t=0}$$

Multiplicative Property (Independence). If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)\,M_Y(t)$$

Uniqueness Property. Let X and Y be random variables with their moment generating functions $M_X(t)$ and $M_Y(t)$ respectively. Suppose that there exists an $h > 0$ such that

$$M_X(t) = M_Y(t), \quad \text{for all } t \in (-h, h)$$

then X and Y have the same distribution

Moment Generating Functions of Common Distribution.

1. When $X \sim Be(p)$, $M_X(t) = 1 - p + pe^t$
2. When $X \sim Bin(n,p)$, $M_X(t) = (1 - p + pe^t)^n$
3. When $X \sim Geom(p)$, $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$
4. When $X \sim Poisson(\lambda)$, $M_X(t) = \exp(\lambda(e^t - 1))$
5. When $X \sim U(\alpha,\beta)$, $M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta-\alpha)t}$
6. When $X \sim Exp(\lambda)$, $M_X(t) = \frac{\lambda}{\lambda - t}$, for $t < \lambda$
7. When $X \sim N(\mu,\sigma^2)$, $M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Markov's Inequality. Let X be a non-negative random variable. For $a > 0$, we have

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Chebyshev's Inequality. Let X be a random variable with mean μ, then for $a > 0$, we have

$$P(|X-\mu| \geq a) \leq \frac{Var(X)}{a^2}$$

Consequences of Chebyshev's Inequality. If $Var(X) = 0$, then the random variable X is a constant.

The Weak Law of Large Numbers. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, with common mean μ. Then, for any $\epsilon > 0$

$$P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \quad \text{as } n \to \infty$$

Central Limit Theorem. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}}\int_{\infty}^{x} e^{-t^2/2}\,dt$$

Normal Approximation. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables, each having mean μ and variance σ^2. Then, for large n, the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

is approximately standard normal.

In other words, for $-\infty < a < b < \infty$, we have

$$P\left(a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}$$

The Strong Law of Large Numbers. Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then with probability 1.

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \quad \text{as} \quad n \to \infty$$

One-sided Chebyshev's Inequality. If X is random variable with mean 0 and finite variance σ^2, then, for any $a > 0$

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Jensen's Inequality. If $g(x)$ is convex function, then

$$E[g(X)] \geq g(E[X])$$

provided that the expectations exist and are finite.

1. A function $g(x)$ is convex if for all $0 \leq p \leq 1$ and all $x_1, x_2 \in R_X$
$$g(px_1 + (1-p)x_2) \leq pg(x_1) + (1-p)g(x_2)$$
2. A differentiable function of one variable is convex on interval if and only if
$$g(x) \geq g(y) + g'(y)(x-y)$$
for all x and y in the interval

A twice differentiable function of one variable is convex over interval if and only if its second derivative is non-negative there.

Random Shit

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

Common Integrals

$\int k\,dx = k\,x + c$	$\int \cos u\,du = \sin u + c$	$\int \tan u\,du = \ln\lvert\sec u\rvert + c$
$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + c, n \neq -1$	$\int \sin u\,du = -\cos u + c$	$\int \sec u\,du = \ln\lvert\sec u + \tan u\rvert + c$
$\int x^{-1}\,dx = \int \frac{1}{x}\,dx = \ln\lvert x\rvert + c$	$\int \sec^2 u\,du = \tan u + c$	$\int \frac{1}{a^2+u^2}\,du = \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + c$
$\int \frac{1}{ax+b}\,dx = \frac{1}{a}\ln\lvert ax+b\rvert + c$	$\int \sec u \tan u\,du = \sec u + c$	$\int \frac{1}{\sqrt{a^2-u^2}}\,du = \sin^{-1}\left(\frac{u}{a}\right) + c$
$\int \ln u\,du = u\ln(u) - u + c$	$\int \csc u \cot u\,du = -\csc u + c$	
$\int \mathbf{e}^u\,du = \mathbf{e}^u + c$	$\int \csc^2 u\,du = -\cot u + c$	

Integration by Parts : $\int u\,dv = uv - \int v\,du$ and $\int_a^b u\,dv = uv\big|_a^b - \int_a^b v\,du$. Choose u and dv from integral and compute du by differentiating u and compute v using $v = \int dv$.

Ex. $\int x\mathbf{e}^{-x}\,dx$

$u = x \quad dv = \mathbf{e}^{-x} \quad \Rightarrow \quad du = dx \quad v = -\mathbf{e}^{-x}$

$\int x\mathbf{e}^{-x}\,dx = -x\mathbf{e}^{-x} + \int \mathbf{e}^{-x}\,dx = -x\mathbf{e}^{-x} - \mathbf{e}^{-x} + c$

Ex. $\int_3^5 \ln x\,dx$

$u = \ln x \quad dv = dx \quad \Rightarrow \quad du = \frac{1}{x}dx \quad v = x$

$\int_3^5 \ln x\,dx = x\ln x\big|_3^5 - \int_3^5 dx = (x\ln(x) - x)\big|_3^5$

$= 5\ln(5) - 3\ln(3) - 2$

***u* Substitution :** The substitution $u = g(x)$ will convert $\int_a^b f(g(x))g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du$ using $du = g'(x)\,dx$. For indefinite integrals drop the limits of integration.

Ex. $\int_1^2 5x^2\cos(x^3)\,dx$

$u = x^3 \Rightarrow du = 3x^2dx \Rightarrow x^2dx = \frac{1}{3}du$

$x = 1 \Rightarrow u = 1^3 = 1 \;::\; x = 2 \Rightarrow u = 2^3 = 8$

$\int_1^2 5x^2\cos(x^3)\,dx = \int_1^8 \frac{5}{3}\cos(u)\,du$

$= \frac{5}{3}\sin(u)\big|_1^8 = \frac{5}{3}(\sin(8) - \sin(1))$