Law of Total Probability: Let A_1, A_2, A_3, $\cdots$ be events that partition Ω, i.e., disjoint ($A_i \cap A_j = \varnothing$ for $i \neq j$) and $\bigcup_{i=1}^{\infty} A_i = \Omega$. Then for event B:

$$P(B) = \sum_{i=1} P(A_i \cap B)$$

Bayes Rule: Bayes rule also applies to a countably infinite number of events.

$$P(A_j|B) = \frac{P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} P(A_j)$$

Random Variable: Random variable is actually a real-valued function $X(\omega)$ over a sample space Ω. We use upper case letters for random variables, and lower case letters for values of random variables.
We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to mean that the discrete random variable X has pmf $p_X(x)$ or $p(x)$.
Bernoulli Probability: $X \sim Bern(p)$ for $0 \leq p \leq 1$ has the pmf below. $E(X) = p$, $Var(X) = p(1-p)$
$$p_X(1) = p, \text{ and } p_X(0) = 1 - p$$

Geometric Distribution: $X \sim Geom(p)$ for $0 \leq p \leq 1$ has pmf below.
$E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$
$$p_X(k) = p(1-p)^{k-1}, \text{ for } k = 1, 2, \cdots$$

The Geometric r.v. represents, for example, the number of coin flips until the first heads shows up.
Binomial Distribution: $X \sim Binom(n, p)$ for integer $n \geq 0$ and $0 \leq p \leq 1$ has the below pmf.
$E(X) = np$, $Var(X) = np(1-p)$
$$p_X(k) = \frac{n!}{k!(n-k!)} p^k (1-p)^{n-k}$$
$$\text{for } k = 0, 1, 2, \cdots$$

with the maximum of $p_X(k)$ is obtained at $k = (n+1)p$. The binomial r.v. represents, for example, the number of heads in n independent coin flips.
Poisson Distribution: If the average number of packets arriving per unit time is λ, the probability of $k(k = 1, 2, \cdots)$ packages arrive within (0, T] can be calculated as the equation below.
$E(X) = \lambda T$, $Var(X) = \lambda T$
$$P(k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

Poisson is the limit of Binomial when $p \propto \frac{\lambda}{n}$ has $n \to \infty$.
Properties of CDF: $F_X(x)$ is right continuous, i.e., $F_X(a^+) = \lim_{x \to a^+} F_X(x) = F_X(a)$.
$P\{X = a\} = F_X(a) - F_X(a^-)$, where $F_X(a^-) = \lim_{x \to a^-} F_X(x)$
A random variable is said to be *continuous* if its cdf is a continuous function.
Properties of PDF: If $F_X(x)$ is differentiablem then X has a pdf $f(x)$ such that

$$f_X(x) = \frac{dF_X(x)}{dx}$$
$$= \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$
$$\int_a^b f(x)dx = F(b) - F(a) = P(x \in (a, b])$$

Notation: $X \sim F_X(x)$ means that X has cdf $F_X(x)$, and $X \sim f_X(x)$ means that X has pdf $f_X(x)$. PDF is a continuous version of PMF.
Uniform Distribution $X \sim U[a, b]$ where $a < b$ has pdf: $f_X(x) = \frac{1}{b-a}$ if $a \leq x \leq b$ and 0 for otherwise.
$E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$. The uniform r.v. is commonly used in modeling quantization noise and finite precision computation error.
Exponential Distribution $X \sim Exp(\lambda)$ where $\lambda > 0$ has pdf: $f_X(x) = \lambda e^{-\lambda x}$ if $x \geq 0$ and 0 for otherwise. $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$. The exponential r.v. represents interarrival time in a queue (time between two consecutive packet or customer arrivals) or service time in a queue or particle lifetime. The exponential r.v. is memoryless.
Gaussian Distribtion $X \sim \mathcal{N}(\mu, \sigma^2)$ has pdf with parameter μ (mean) and σ^2 (variance) shown as below. $E(X) = \mu$, $Var(X) = \sigma^2$.

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

Functions of a Random Variable: If we know a random variable $X \sim f_X(x)$ and the relationship between x and y as $y = g(x)$. Then $p_Y(y)$ can be calculated as:

$$f_Y(y) = \sum_{i=1}^{k} \frac{f_X(x_i)}{|g'(x_i)|}$$

This property can be used to compute probabilities for an arbitrary Gaussian r.v. from the distribution of the standard Gaussian r.v..
Generation of Random Variables
From a distribution (x) to uniform (y): 1. Find the CDF $F_X(x)$ of the distribution from which x is sampled. 2. Calculate $y = F_X(x)$ for a given sample x, which gives a uniform random variable y on the interval [0, 1].
From uniform (y) to a distribution (x): 1. Find the inverse of the CDF $F_X^{-1}(y)$ of the desired distribution. 2. Generate a uniform random variable y on the interval [0, 1]. 3. Calculate $x = F_X^{-1}(y)$ using the inverse CDF with the generated uniform y, which gives a random variable x with the desired distribution.
Joint PMF: $\sum_{x \in X} \sum_{y \in Y} p_{X,Y}(x, y) = 1$
Marginal PMF $p_X(x) = \sum_{y \in Y} p(x, y)$, $x \in \chi$
Properties of Joint and Marginal CDF:
$\lim_{x,y \to \infty} F_{X,Y}(x, y) = 1$;
$\lim_{x \to -\infty} F_{X,Y}(x, y) = 0$ and vice versa;

$\lim_{y \to \infty} F_{X,Y}(x, y) = F_X(x)$ and $\lim_{x \to \infty} F_{X,Y}(x, y) = F_Y(y)$. The $F_X(x)$ and $F_Y(y)$ are marginal cdfs of X and Y.

$$P\{a < X \leq b, c < Y \leq d\}$$
$$= F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

X and Y are jointly continuous random variables if their joint cdf is continuous in both x and y. In this case:

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)dudv, \quad x, y \in \mathbf{R}$$

If $F_{X,Y}(x, y)$ is differentiable in x and y, then:

$$f_{X,Y}(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$
$$P\{x < X \leq x + \Delta x,$$
$$= \lim_{\Delta x, \Delta y \to 0} \frac{y < Y \leq y + \Delta y\}}{\Delta x \Delta y}$$

Properties of Joint and Marginal PDF:
$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$;
$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
Conditional CDF and PDF:
CDF: Find $F_{Y|X}(y|X = x) = P\{Y \leq y|X = x\}$
$$F_{Y|X}(y|x)$$
$$= \lim_{\Delta x \to 0} P\{Y \leq y|x < X \leq x + \Delta x\}$$
$$= \lim_{\Delta x \to 0} \frac{P\{Y \leq y, x < X \leq x + \Delta x\}}{P\{x < X \leq x + \Delta x\}}$$
Assume $f_X(x)$ is constant within Δx
$$P(x \in [x, x + \Delta x]) = f_X(x)\Delta x$$
$$= \lim_{\Delta x \to 0} \frac{\int_{-\infty}^{y} f_{X,Y}(x, u)du\Delta x}{f_X(x)\Delta x}$$
$$= \int_{-\infty}^{y} \frac{f_{X,Y}(x, u)}{f_X(x)} du$$

Bayes rule of mixed random variables: Θ is a discrete random variable with $\Theta \sim p_\Theta(\theta)$ and Y is a continuous random variable and given that $f_{Y|\Theta}(y|\theta)$, then

$$p_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta)}{\sum_{\theta'} p_\Theta(\theta')f_{Y|\Theta}(y|\theta')} p_\Theta(\theta)$$

Maximum posteriori probability decoder and Maximum likelihood decoder:
MAP decoder:
$$d(y) = \begin{cases} \theta_0 & \text{if } p_{\Theta|Y}(\theta_0|y) > p_{\Theta|Y}(\theta_1|y) \\ \theta_1 & \text{otherwise} \end{cases}$$

If $p_\Theta(\theta_0) = p_\Theta(\theta_1)$, MAP can be reduces to ML decoder:
$$d(y) = \begin{cases} \theta_0 & \text{if } f_{Y|\Theta}(y|\theta_0) > p_{Y|\Theta}(y|\theta_1) \\ \theta_1 & \text{otherwise} \end{cases}$$

Minimal probability of error: Given input
$\Theta \sim p_\Theta(\theta)$, $p_\Theta(\sqrt{P}) = 0.5$, $p_\Theta(-\sqrt{P}) = 0.5$.
Noise $Z \sim N(0, N)$. Observation $Y = \Theta + Z$
$P_e = P\{d(Y) \neq \theta\}$
$$= P\{\Theta = \sqrt{P}\}P\{d(y) = -\sqrt{P}|\Theta = \sqrt{P}\} + P\{\Theta = -\sqrt{P}\}P\{d(y) = \sqrt{P}|\Theta = -\sqrt{P}\}$$
If we use the minimum distance decoder
$$= \frac{1}{2}P\{Y \leq 0|\Theta = \sqrt{P}\}$$
$$+ \frac{1}{2}P\{Y > 0|\Theta = -\sqrt{P}\}$$
$$= \frac{1}{2}P\{Z \leq -\sqrt{P}\} + \frac{1}{2}P\{Z > \sqrt{P}\}$$
$$= Q\left(\sqrt{\frac{P}{N}}\right) = Q(\sqrt{SNR})$$

Proprties of expection:
$E(g(X)) = \sum_{x \in \mathcal{X}} g(x)p_X(x) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$
$E(const) = const$;
$E[ag_1(X) + g_2(X)] = aE(g_1(X)) + E(g_2(X))$
Fundamental Theorem of Expection:
If $Y = g(X) \sim p_Y(y)$, then $E(Y) = \sum_{y \in \mathcal{Y}} yp_Y(y) = \sum_{x \in \mathcal{X}} g(x)p_X(x) = E(g(X)) = \int_{-\infty}^{\infty} yf_Y(y)dy$. Which means that $E(Y)$ can be found using either $f_X(x)$ or $f_Y(y)$.
Mean: $E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$, $E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)dx$
Variance: $Var(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2 \geq 0$; $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
Markov Inequality: $P(X \geq a) \leq \frac{1}{a}E(X)$ for $x > 0$ & $a > 1$.
$$E(X) = \int_0^{\infty} xf_X(x)dx$$
$$\geq \int_a^{\infty} xf_X(x)dx$$
$$\geq \int_a^{\infty} af_X(x)dx = aP(X \geq a)$$

Chebyshev Inequality: $P(|x - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$ for $x > 0$ & $a > 1$.
$$P(|x - \mu| \geq c) = P((x - \mu)^2 \geq c^2)$$
$$\leq \frac{1}{c^2}E[(x - \mu)^2] = \frac{\sigma^2}{c^2}$$

Expection involving two RVs $E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$. $g(X, Y)$

---

may be $X, Y, X^2, X + Y$, etc.
Correlation:
$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{X,Y}(x, y)dxdy$.
Correlation Coefficient: $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$.
$|\rho_{X,Y}| \leq 1$. ρ is basically normalizing the Cov.
Covariance:
$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$ X and Y are said to be uncorrelated if $Cov(X, Y) = 0$.
Conditioning on an event: Let $X \sim p_X(x)$ be a r.v., and A be a nonzero probability event.
$p_{X|A}(x) = \frac{p_X(x)}{P\{X \in A\}}$ or
$f_{X|A}(x) = \frac{f_X(x)}{P\{X \in A\}}$, if $x \in A$, and 0 for other wise. $\sum_{x \in A} f_{X|A}(x) = 1$. It defines a new probability distribution that only focuses on A.
Conditional Expectation on an event:
$E(g(X)|A) = \int_{-\infty}^{\infty} g(x)f_{X|A}(x)dx$
Total Expectation Theorem: Let $X \sim f_X(x)$ and $A_1, A_2, \cdots, A_n$ be disjoint nonzero probability events with $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) = 1$. Then:
$E(g(X)) = \sum_{i=1}^{n} P\{X \in A_i\}E(g(X)|A_i)$.
This theorem said: find the mean of each part in a combined distribution, and them sum them up to get the whole expectation of the total distribution.
Conditional Expectation on a random variable:
$E(g(X, Y)|Y = y) = \int_{-\infty}^{\infty} g(x, y)f_{X|Y}(x|y)dx$
Iterated Expectation:
$E_Y\{E_X(g(X, Y)|Y)\} = E_{X,Y}(g(X, Y))$
Conditional Variance:
$Var(X|Y = y) = E[(X - E(X|Y))^2|Y] = E(X^2|Y = y) - [E(X|Y = y)]^2$
Law of Conditional Variances:
$Var(X) = E_Y(Var(X|Y)) + Var(E(X|Y)) = $ bias + Variance
Mean Square Error:
$MSE = E_{X,Y}[(X - \hat{X})^2] = E_X[(X - g(Y))^2]$
Minimum MSE $MMSE = E_Y(Var(X|Y)) = E(X^2) - E[(E(X|Y))^2]$. The $\hat{x}$ that achieves the minimum MSE is called the minimum MSE estimate of x given Y. $\hat{x} = g(y) = E[X|Y = y]$. Which means that for each $Y = y$, the minimum value for $E_X[(X - g(y))^2|Y = y]$ is obtained when $g(y) = E[X|Y = y]$.
Properties:
Unbiased: $E(\hat{X}) = E(X)$.
$E[(X - \hat{X})|Y = y] = 0$.
Orthogonal:
$$E[(X - \hat{X})\hat{X}] = E_Y[E((X - \hat{X})\hat{X}|Y)]$$
$\because \hat{X}$ is a function of Y,
$\therefore$ it is a constant when condition on Y
$$= E_Y[\hat{X}E((X - \hat{X})|Y)]$$
$$= E_Y[\hat{X}(E(X|Y) - \hat{X})]$$
$\because E(X|Y) = \hat{X}$
$$= 0$$
$Var(X) = Var(\hat{X}) + E(Var(X|Y))$
MMSE Linear Estimation: Properties of **best** linear MSE estimate: $E(\hat{X}) = E(X)$; If $\rho_{X,Y} = 0$, i.e., X and Y are uncorrelated, then $\hat{X} = E(X)$, which means observation of Y is ignored and this is the disadvantage of linear estimate.; If $\rho_{X,Y} = \pm 1$, i.e., $(X - E(X))$ and $(Y - E(Y))$ are linearly dependent, then the linear estimate is perfect.
$$\hat{X} = \rho_{X,Y}\sigma_X\left(\frac{Y - E(Y)}{\sigma_Y}\right) + E(X)$$
$$MSE = (1 - \rho_{X,Y}^2)\sigma_X^2$$

Properties of Jointly Gaussian Random Variables: If X and Y are jointly Gaussian, they are individually Gaussian. The converse is not necessarily true; If X and Y are jointly Gaussian, the conditional pdf is Gaussian; If X and Y are jointly Gaussian and uncorrelated ($\rho_{X,Y} = 0$), they are also independent.

$$X|\{Y = y\} \sim \mathbb{N}\left(\rho_{X,Y}\sigma_X\frac{y - \mu_Y}{\sigma_Y} + \mu_X, (1 - \rho_{X,Y}^2)\sigma_X^2\right)$$

Random Vectors:
$CDF: F_X(\vec{x}) = P\{X_1 \leq x_1, \cdots, X_n \leq x_n\}$
$PDF: f_X(\vec{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$
$Marginals:$
$F_{X_1}(x_1) = \lim_{x_2,x_3 \to \infty} F_X(x_1, x_2, x_3)$
$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2$
$Chainrule: f_X(\vec{x}) = f_{X_1}(x_1) \cdot$
$f_{X_2|X_1}(x_1|x_1)f_{X_3|X_1,X_1}(x_3|x_1, x_2) \cdots$

Conditional Independence: X_1 and X_3 are said to be conditionally independent given X_2 if
$$f_{X_1,X_3|X_2}(x_1, x_3|x_2) = f_{X_1|X_2}(x_1, x_2) \cdots f_{X_3|X_2}(x_3|x_2)$$
$$\forall(x_1, x_2, x_3)$$

This does not mean that X_1 and X_3 are independent, or vice versa.
Mean and Covariance Matrix of random vector:
$E(X) = [E(X_1), E(X_2), \cdots, E(X_n)]^T$
$$Cov(X) = \Sigma_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$
$$= E[(X - E_X) \cdot (X - E_X)^T]$$

Σ_X is real, symmetric, and non-negative definite, i.e. $a^T \Sigma_X a \geq 0$, $\forall a$ or $|\Sigma_X| \geq 0$.
Mean and variance of sum of r.v. Let $Y = I^T X$ be the sum of X.
$$E(Y) = \sum_{i=1}^{n} E(X_i)$$
$$Var(Y) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{i \neq j}^{n} Cov(X_i, X_j)$$

Gaussian Random Vectors (GRV): A random vector X is a GRV if the joint pdf is of the form:
$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \cdot \exp{-\frac{1}{2}(x - \vec{\mu})^T \Sigma^{-1}(x - \vec{\mu})}$$

Properties of GRVs: 1. For a GRV, uncorrelation implies independence.
2. Linear transformation of a GRV yields a GRV. I.e., given any $m \times n$ matrix A, where $m \leq n$, and A has full rank m, then: $\mathbb{Y} = A\mathbb{X} \sim \mathcal{N}(A\mu, A\Sigma A^T)$
3. Marginals of a GRV are Gaussian. I.e., if X is GRV, for any subset $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ of indices, the RV: $\mathbb{Y} = [X_{i1}, X_{i2}, \ldots, X_{ik}]^T$ is a GRV.
4. Conditionals of a GRV are Gaussian, if
$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ -- \\ \mathbf{X_2} \end{bmatrix} \sim$$
$$\mathcal{N}\left(\begin{bmatrix} \mu_1 \\ -- \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ -- & -- \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

Then, $\mathbb{X}_2|\{\mathbb{X}_1 = x\} \sim$
$\mathbf{N}(\Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1) + \mu_2, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$
5. If $[\mathbb{Y}^T X]^T$ is a GRV, then the best MSE estimate of X given Y is linear.
MSE Estimation: Vector Case: When $E(X) = 0$ and $E(Y) = 0$:
$$\hat{X} = \vec{h}^T \vec{Y} = \Sigma_{i=1}^{n} h_i Y_i$$
$$\Sigma_{\vec{Y}} \vec{h} = \Sigma_{\vec{Y}X}$$
$$\vec{h} = \Sigma_{\vec{Y}}^{-1} \Sigma_{\vec{Y}X}$$
$$\hat{X} = \Sigma_{\vec{Y}X}^T \Sigma_{\vec{Y}}^{-1} \vec{Y}$$
$$MMSE = \sigma_X^2 - \Sigma_{\vec{Y}X}^T \Sigma_{\vec{Y}}^{-1} \Sigma_{\vec{Y}X}$$
$$\Sigma_{\vec{Y}X} = E[(\vec{Y} - E(\vec{Y}))(X - E(X))]$$

Almost Sure (a.s.) Convergence: We say $X_n \to X$ almost surely or with probability 1 if:
$$\mathbb{P}\left(\left\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$
Lemma:
$$\mathbb{P}\left(\max_{i \geq n} |X_i - X| > \epsilon\right) \to 0$$
$$\text{for any } \epsilon > 0 \text{ iff } X_n \xrightarrow{a.s.} X.$$

Strong Law of Large Numbers: If $X_1, X_2, \ldots$, are i.i.d (pairwise independence is suffices) with finite mean μ, then:
$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu \text{ as } n \to \infty$$

Convergence in Mean Square: We say $X_n \to X$ in mean square or L^2 if
$$\lim_{n \to \infty} \mathbb{E}(X_n - X)^2 = 0$$

Convergence in mean square does not imply convergence a.s.
Convergence in a.s. does not imply convergence in mean square.
Convergence in Probability: A sequence $X_1, X_2, \ldots$ converges to a random variable in probability if for any $\epsilon > 0$,
$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.
If $X_n \xrightarrow{a.s.} X$, then the indicator function
$\mathbf{1}_{|X_n - X| > \epsilon} \xrightarrow{a.s.} 0$. The indicator function is 1 if the absolute difference between X_n and X is greater than ϵ, and 0 otherwise.
If $X_n \to X$ in mean square, then $X_n \xrightarrow{P} X$. Converse is not ture.
Convergence in probability is weaker than both convergence a.s. and in mean square.
Weak Law of Large Numbers: Suppose $X_1, X_2, \ldots$ are such that $\mathbb{E}(X_i) = \mu$, $\mathbb{E}^2(X_i) = \sigma^2 < \infty$ and $\mathbb{E}[X_i X_j] \leq 0, \forall i \neq j$. Then,
$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu \text{ as } n \to \infty$$
$$M_n = \frac{X_1 + \cdots + X_n}{n}$$
$$\mathbb{E}[M_n] = \frac{\mathbb{E}[X_1 + \cdots + X_n]}{n} = \frac{n\mu}{n} = \mu$$
$$Var(M_n) = \frac{Var(X_1 + \cdots + X_n)}{n^2} = \frac{\sigma^2}{n}$$
$$P(|M_n - \mu| \leq \epsilon) \leq \frac{Var(M_n)}{\epsilon^2}$$
$$= \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \to \infty} 0 \text{ (Chebyshev Inequality)}$$

Convergence in distribution We say that $X_n \xrightarrow{d} X$ if for all continuous bounded functions f,
$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(x)]$$
$$OR, \ equivalently,$$
$F_{X_n}(t) \to F_X(t)$ for all continuity points of $F_X(\cdot)$

Central Limit Theorem: Consider i.i.d. random variables $X_1, X_2, \ldots$ with $\mathbb{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then
$$Z_n = \frac{S_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathbf{N}(0,1)$$

Random Process Equality: Two random process is said to be equal if:
$$x(t) = y(t) \text{ if } x(t, \zeta) = y(t, \zeta) \ \forall t, \zeta$$
$$\text{and } \mathbb{E}\{|x(t) - y(t)|^2\} = 0, \ \forall \ t$$

First Order Statistic:
  CDF: $F(x, t) = \mathbb{P}\{\mathbf{x}(t) \le x\}$
  PDF: $f(x, t) = \dfrac{\partial F(x, t)}{\partial x}$
  Mean: $\eta(t) = \mathbb{E}\{\mathbf{x}(t)\} = \displaystyle\int_{-\infty}^{\infty} x f(x, t) dx$

Second Order CDF and PDF:
  CDF: $F(x_1, x_2; t_1, t_2) = \mathbb{P}\{\mathbf{x}(t_1) \le x_1,$
  $$\mathbf{x}(t_2) \le x_2\}$$
  PDF: $f(x_1, x_2; t_1, t_2) = \dfrac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}$

Autocorrelation: It's used to measure the statistical dependence between values of the same variable at different times. In simpler terms, it quantifies how similar a signal is to itself at different points in time.
$$R_{xx}(t_1, t_2) = \mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}^*(t_2)\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2^* f(x_1, x_2; t_1, t_2) dx_1 dx_2$$
For complex processes: $R_{xx}(t_2, t_1) = R_{xx}^*(t_1, t_2)$
For real processes: $R_{xx}(t_2, t_1) = R_{xx}(t_1, t_2)$

Properties: Symmetry: $R(\tau) = R(-\tau)$. Maximum Value: $R_{xx, max} = R_{xx}(0)$. Number of max values: If $x(t)$ has one max Rxx, it is not a periodic function. If $x(t)$ has infinite max Rxx, it is a periodic function. Positive-definite function:
$$\sum_{i=0}^{m}\sum_{f=0}^{m} a_i a_j^* R_{xx}(t_i, t_j) \ge 0$$
Types: High Autocorrelation (Positively or Negatively): Indicates that the process has a strong memory, i.e., future values are highly dependent on past values. Low Autocorrelation (Close to 0): Indicates that the process is more like a random noise, with less predictability.
Autocovariance: Autocovariance measures the degree to which two variables vary together. In the context of a random process, autocovariance helps to understand how a variable at one-time point t is related to the variable at another time point $t + \tau$.
$$C_{xx}(t_1, t_2) = R_{xx}(t_1, t_2) - \eta(t_1)\eta^*(t_2).$$
Properties: Symmetry: $C(\tau) = C(-\tau)$. Maximum Value: $C_{xx, max} = C_{xx}(0)$. Relation to Autocorrelation: $R(\tau) = \dfrac{C(\tau)}{C(0)}$.
Autocovariance Coefficient:
$r(t_1, t_2) = \dfrac{C_{xx}(t_1, t_2)}{\sqrt{C_{xx}(t_1, t_1)C_{xx}(t_2, t_2)}}$. Properties: $r(t_1, t_2) \le 1$, and $r(t, t) = 1$.
Cross-correlation: Mathematically representation:
$\overline{R_{xy}(t_1, t_2)} = \mathbb{E}\{\mathbf{x}(t_1)\mathbf{y}^*(t_2)\} = R_{yx}^*(t_2, t_1)$.
$x(t)$ and $y(t)$ are orthogonal if:
$R_{xy}(t_1, t_2) = 0, \forall t_1, t_2$.
Cross-covariance: Mathematically representation:
$\overline{C_{xy}(t_1, t_2)} = R_{xy}(t_1, t_2) - \eta_x(t_1)\eta_y^*(t_2)$.
$x(t)$ and $y(t)$ are uncorrelated if:
$C_{xy}(t_1, t_2) = 0, \forall t_1, t_2$.
Power Spectrum: The power spectral density of a WSS ramdom process X(t) is defined as the Fourier transform of its autocorrelation function.
Continuous:
$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau)e^{-j\omega\tau} d\tau$$
$$R_{xx}(\tau) = \int_{-\infty}^{\infty} S_{xx}(\omega)e^{j\omega\tau}\frac{d\omega}{2\pi}$$
$$R_{xx}(\tau) = E\{\mathbf{x}(t+\tau)\mathbf{x}^*(t)\}$$
$$S_{xy}(\omega) = \int_{-\infty}^{+\infty} R_{xy}(\tau)e^{-j\omega\tau} d\tau$$
Discrete:
$$R_{xx}[m] = \int_{-\pi}^{\pi} S_{xx}(\omega)e^{jm\omega}\frac{d\omega}{2\pi}$$
$$S_{xx}(\omega) = \sum_{m=-\infty}^{\infty} R_{xx}[m]e^{-jm\omega}$$
Properties: If $x(t)$ is complex, then $S_{xx}(\omega)$ is non-negative and real. If $x(t)$ is complex, then $S_{xx}(\omega)$ is non-negative, real, and symmetric. $S_{xx}(\omega)$ is non-negative, proven by Wiener-Khinchin theorem. For the discrete process, the power spectrum is continuous and periodic with period 2pi.
Type of power spectrum:
Continuous:
  Fourier Transform:
$$h(t) \xrightarrow{FT} H(\omega)$$
$$S_{xy}(\omega) = S_{xx}(\omega)H^*(\omega)$$
$$S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$$
  Laplace Transform:
$$h(t) \xrightarrow{LT} \mathbb{H}(s), \mathbb{H}(j\omega) = H(\omega)$$
$$\mathbb{S}_{xy}(s) = \mathbb{S}_{xx}(s)\mathbb{H}(-s)$$
$$\mathbb{S}_{yy}(s) = \mathbb{S}_{xx}(s)\mathbb{H}(s)\mathbb{H}(-s)$$
Discrete:
  Discrete-time Fourier Transform:
$$h(t) \xrightarrow{DTFT} H(\omega)$$
$$S_{xy}(\omega) = S_{xx}(\omega)H^*(\omega)$$
$$S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$$

Z-Transform:
$$h[n] \xrightarrow{zT} \mathbb{H}(z), \mathbb{H}(e^{j\omega}) = H(\omega)$$
$$\mathbb{S}_{xy}(z) = \mathbb{S}_{xx}(z)\mathbb{H}(z^{-1})$$
$$\mathbb{S}_{yy}(z) = \mathbb{S}_{xx}(z)\mathbb{H}(z)\mathbb{H}(z^{-1})$$
Z-transform: $R_{xx}[m] \xrightarrow{zT} \mathbb{S}_{xx}(z) = \sum_{m=-\infty}^{\infty} R_{xx}[m]z^{-m}$. Z-transform to DTFT: $z = e^{j\omega}$

Wiener-Khinchin theorem:
$$\mathbb{E}\{|\mathbf{x}(t)|^2\} = R_{xx}(0) = \int_{-\infty}^{+\infty} S_{xx}(\omega)\frac{d\omega}{2\pi} \ge 0$$

Create a random process from Sxx: For any S(w), formulate the pdf of w like $f_w(\omega) = \dfrac{S(\omega)}{2\pi R_{xx}(0)}$. And then define w. According to the formulated pdf, find w. Then the random process can be determined like $x(t) = a\cos(\omega t + \phi)$. In which the ϕ can be an arbitrary random variable like the uniform distribution.
Ergodicity: Ergodicity refers to the property that the time average of a process equals the ensemble average.
Prerequisite: x(t) is a real stationary process, with mean $\eta = $ Ex(t).
Time average: $\eta_T = \frac{1}{2T}\int_{-T}^{T} \mathbf{x}(t)dt$. η_T is a random variable. $\mathbb{E}\{\eta_T\} = \frac{1}{2T}\int_{-T}^{T}\mathbb{E}\{x(t)\}dt = \eta$. Therefore, η_T is an estimate of η.
Mean-ergodic: x(t) is mean-ergodic process if $\eta_T \to \eta$ as $T \to \infty$ OR $\sigma_T \to 0$ as $T \to \infty$. If a process is mean-ergodic, than the ensemble average, which is hard to observe, can be replaced by time average, which is easier to observe.
Slutsky's Theorem:
Continuous: x(t) is mean-ergodic if and only if
$\frac{1}{T}\int_0^T C_{xx}(\tau)d\tau \to 0$ as $T \to \infty$.
Discrete: x(t) is mean-ergodic if and only if
$\frac{1}{M}\sum_{m=0}^{M} C_{xx}[m] \to 0$ as $M \to \infty$.
Covariance-erogdic Process: x(t) is covariance-erogdic if and only if
$$\frac{1}{T}\int_0^T C_{zz}(\tau)d\tau \to 0 \text{ as } T \to \infty$$
$$\text{Where } z(t) = x(t+\lambda)x(t)$$
$$C_{zz}(\tau) = E\{\underbrace{x(t+\lambda+\tau)x(t+\tau)}_{z(t+\tau)} \cdot$$
$$\underbrace{x(t+\lambda)x(t)}_{z(t)}\} - \underbrace{C_{xx}^2(\lambda)}_{E^2(z(t))}$$

Distribution-ergodic Process: x(t) is distribution-erogdic if and only if
$$\frac{1}{T}\int_0^T C_{yy}(\tau)d\tau \to 0 \text{ as } T \to \infty$$
$$y(t) = U[x - x(t)]$$
$$C_{yy}(\tau) = E\{y(t+\tau)y(t)\} - E^2\{y(t)\}$$
$$= F_2(x, x; \tau) - F_1^2(x)$$
White Noise Process: $C_{xx}(t_1, t_2) = 0, \forall t_1 \ne t_2$. That is to say x(t1) and x(t2) are uncorrelated for every $t_1 \ne t_2$.
Strict-Sense Stationary (SSS): A random process X(t) is said to be Strict-Sense Stationary if all its statistical properties, including mean, variance, and higher-order moments, are invariant under time translation.
$$f(x_1, \cdots, x_n; t_1, \cdots, t_n) =$$
$$f(x_1, \cdots, x_n; t_1 + c, \cdots, t_n + c), \forall c$$
Wide-Sense Stationary (WSS): A random process is said to be WSS if it satisfies the following conditions: 1.
$\mu(t) = \mu = $ constant. 2. $\delta^2(t) = \delta^2 = $ constant. 3.
$R_{xx}(t_1, t_2) = R_{xx}(t_1 + \tau, t_2 + \tau), \forall \tau$.
$$SSS \Longrightarrow WSS.$$
Memoryless System: If y(t) = g(x(t)), then the output y(t) depends only on the present input. Then such a system is called the memoryless system. $\mathbf{y}(t_n) = g(\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_n)) = g(\mathbf{x}(t_n))$. For a memoryless system, if x(t) is SSS, then y(t) is also SSS. However, if x(t) is WSS, y(t) may or may not be WSS. Find the PDf of Y: If $g_1(x_1, \ldots, x_n) = y_1'; \cdots; g_n(x_1, \ldots, x_n) = y_n'$ has one solution $x_1', \ldots, x_n'$. Then:
$$f_y(y_1', \ldots, y_n') = \frac{f_x(x_1', \ldots, x_n')}{|J(x_1', \ldots, x_n')|}$$
$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$$

Linear System: $\mathbf{y}(t) = L[\mathbf{x}(t)]$, $h(t) = L[\delta(t)]$, $\mathbf{y}(t) = \mathbf{x}(t) * h(t) = \int_{-\infty}^{+\infty}\mathbf{x}(t-\alpha)h(\alpha)d\alpha$.
Properties: 1. $L[a_1\mathbf{x}_1(t) + a_2\mathbf{x}_2(t)] = a_1 L[\mathbf{x}_1(t)] + a_2 L[\mathbf{x}_2(t)]$. 2. If x(t) is a normal process (or SSS), then y(t) is also normal process (or SSS). 3. If x(t) is WSS, then x(t) and y(t) are jointly WSS.
Fundamental Theorem: 1. $\mathbb{E}\{L[\mathbf{x}(t)]\} = L[\mathbb{E}[\mathbf{x}(t)]]$. 2. $R_{xy}(t_1, t_2) = L_2^*[R_{xx}(t_1, t_2)] = \int_{-\infty}^{\infty} R_{xx}(t_1, t_2 - \alpha)h^*(\alpha)d\alpha$ L_2 operates on t_2. 3. $R_{yy}(t_1, t_2) = L_1[R_{xy}(t_1, t_2)]$ L_1 operates on t_1.
Discrete Linear System:
$$\mathbf{y}[n] = \mathbf{x}[n] * h[n] = \sum_{k=-\infty}^{\infty} \mathbf{x}[n-k]h[k]$$
Where $h[n]$ is the impulse response of the system, $h[n] = L[\delta[n]]$.
Poisson Process: Mathematical Representation:
$$\mathbb{P}\{\mathbf{n}(t_1, t_2) = k\} = \frac{e^{-\lambda(t_1 - t_2)}(\lambda(t_1 - t_2))^k}{k!}.$$
Autocorrelation: $R_{xx}(t, t) = \lambda t + \lambda^2 t^2$,
$R_{xx}(t_1, t_2) = \lambda\min(t_1, t_2) + \lambda^2 t_1 t_2$.
Power Spectrum: $S_{xx}(\omega) = 2\pi\lambda^2\delta(\omega) + \lambda$.
Independent Property: If the intervals (t1, t2) and (t3, t4) are non-overlapping, then the random variables n(t1, t2) and n(t3, t4) are independent.
Shot Noise: Shot noise is the output of a linear system driven by a train of Poisson impulses. Shot noise:
$\mathbf{s}(t) = L\left[\sum_i \delta(t - \mathbf{t}_i)\right]$. Where L has impulse response $h(t)$, $h(t)$ is a real function, $\mathbf{t}_i$ are a set of Poisson points with average density λ.
SSS Properties: Since Poisson impulses are SSS, s(t) is also SSS.

Mean: $E\{s(t)\} = L[E\{z(t)\}] = L[\lambda] = \lambda\int_{-\infty}^{\infty} h(t)dt = \lambda H(0)$, where z(t) is the Poisson impulse. Variance: $\sigma_s^2 = \lambda\int_{-\infty}^{+\infty} h^2(t)dt$.
Autocorrelation:
$$R_{ss}(\tau) = \lambda^2 H^2(0) + \lambda h(t) * h(-\tau).$$
Power Spectrum: $S_{xx}(\omega) = 2\pi\lambda^2\delta(\omega) + \lambda$.
Modulating Process: $\mathbf{w}(t) = r(t)e^{j\phi(t)}$, where $\Re[\mathbf{w}(t)] = \mathbf{a}(t), \Im[\mathbf{w}(t)] = \mathbf{b}(t)$, $\mathbf{w}(t)$ is a complex process, $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are real and zero mean process
Autocorrelation:
$R_{ww}(\tau) = 2R_{aa}(\tau) - 2jR_{ab}(\tau)$. Power Spectrum Density: $S_{ww}(\omega) = 2S_{aa}(\omega) - 2jS_{ab}(\omega)$.
Modulated Process: $\mathbf{z}(t) = \mathbf{w}(t)e^{j\omega_0 t}$, where $e^{j\omega_0 t}$ is a complex sinusoidal carrier, ω_0 is the carrier frequency.
$\Re[\mathbf{z}(t)] = \mathbf{x}(t) = \mathbf{a}(t)\cos(\omega_0 t) - \mathbf{b}(t)\sin(\omega_0 t)$
$\qquad = \mathbf{r}(t)\cos[\omega_0 t + \phi(t)]$
$\Im[\mathbf{z}(t)] = \mathbf{y}(t) = \mathbf{b}(t)\cos(\omega_0 t) + \mathbf{a}(t)\sin(\omega_0 t)$
$\qquad = \mathbf{r}(t)\sin[\omega_0 t + \phi(t)]$
$\qquad \mathbf{r}(t)$ is for amplitude modulation
$\qquad \phi(t)$ is for phase modulation
Mean: $\mathbb{E}(\mathbf{x}(t)) = \mathbb{E}(\mathbf{y}(t)) = 0$.
Autocorrelation: $R_{zz}(\tau) = R_{ww}(\tau)e^{j\omega_0 \tau}$. If x(t) is WSS:
$$R_{xx}(\tau) = R_{yy}(\tau) = R_{aa}(\tau)\cos(\omega_0\tau)$$
$$+ R_{ab}(\tau)\sin(\omega_0\tau)$$
$$R_{xy}(\tau) = -R_{yx}(\tau) = R_{ab}(\tau)\cos(\omega_0\tau)$$
$$- R_{aa}(\tau)\sin(\omega_0\tau)$$
If x(t) is WSS, then y(t) is also WSS. x(t) is WSS if
$R_{aa}(\tau) = R_{bb}(\tau)$ and $R_{ab}(\tau) = -R_{ba}(\tau)$
Power Spectrum: $S_{zz} = S_{ww}(\omega - \omega_0)$, where S_{ww} is the base psd, and $\omega - \omega_0$ is the modulation shifting the base frequency to carrier frequency.
Demodulation: $\mathbf{w}(t) = \mathbf{z}(t)e^{-j\omega_0 t}$
Bandlimited Process: A bandlimited process refers to a signal or a stochastic process that contains no frequencies higher than a certain finite frequency.
x(t) is bandlimited if
$R_{xx}(0) < \infty$ (finite power)
$S_{xx}(\omega) = 0 \ \forall |\omega| > \sigma$ (limited spectral width)
Properties:
1. A bandlimited process can be sampled: (Shannon Sampling Theorem)
$$\mathbf{x}(t + \tau) = (L[\mathbf{x}(t)] \ with \ H(\omega) = e^{j\omega\tau})$$
$$= \sum_{n=0}^{\infty} \mathbf{x}^{(n)}(t)\frac{\tau^n}{n!}$$
Where $x^{(n)}(t)$ means the n-th order derivatives of x(t).
2. A bandlimited process is continuous and smooth.
3. The expected value of the square of the change of the signal for a value τ is upper bounded:
$$\mathbb{E}\left\{|\mathbf{x}(t+\tau) - \mathbf{x}(t)|^2\right\} \le \sigma^2\tau^2 R_{xx}(\tau). \ \tau \text{ term:}$$
For a small time difference, the signal does not change much. σ term is the bandwidth. If a signal has a narrow band, the change in the signal is also gonna be relatively small. Vice Versa. R_{xx} term is the variance of the signal.
Shannon Sampling Theorem: When a signal is bandlimited, the Shannon Sampling Theorem provides a method to perfectly reconstruct the original signal from its samples, if the sampling rate is greater than twice the bandwidth of the signal, which is also known as the Nyquist criterion. The continuous-time bandlimited signal can be represented as an infinite sum of its samples each scaled by a sinc function.
$$x(\tau) = \sum_{n=-\infty}^{\infty} x(nT) \cdot \frac{\sin[\sigma(\tau - nT)]}{\sigma(\tau - nT)}, t = 0$$
Where $x(\tau)$ is the reconstructed bandlimited signal, $x(nT)$ is the samples of the signal taken at intervals of T, $T = \frac{1}{f_s}$ is the sampling period, sin . . . represents an ideal low-pass filter impulse response, n is an integer that indexes the samples.
Matched Filter: A matched filter is a signal processing technique designed to maximize the signal-to-noise ratio (SNR) in the presence of noise by finding suitable h(t).
$$SNR = r^2 = \frac{|y_f(t_0)|^2}{\mathbb{E}\{\mathbf{y}_v^2(t_0)\}}$$
Applying to colored noise:
$$r^2 \le \int_{-\infty}^{+\infty} \frac{|F(\omega)|^2}{S_{vv}(\omega)}\frac{d\omega}{2\pi}$$
Condition that the equality holds:
$$k\left[\frac{F(\omega)}{\sqrt{S_{vv}(\omega)}}e^{j\omega t_0}\right]^* = \sqrt{S_{vv}(\omega)}H(\omega)$$
$$OR$$
$$H(\omega) = k\frac{F^*(\omega)}{S_{vv}(\omega)}e^{-j\omega t_0}$$
Markov Chain Properties: All outgoing transitions:
$\overline{\sum_j \pi_{ij}[n_1, n_2]} = 1$. All incoming transitions:
$\sum_i p_i[k]\pi_{ij}[k, n] = p_j[n]$. Where $p_i[n] = P\{\mathbf{x}_n = a_i\}$ is the state probabilities, and $\pi_{ij}[n_1, n_2] = P\{\mathbf{x}_{n_2} = a_j | \mathbf{x}_{n_j} = a_i\}$.
Chapman Kolmogoroff equation: For any $n_1 < n_2 < n_3$,
$$\pi_{ij}[n_1, n_3] = \sum_r \pi_{ir}[n_1, n_2]\pi_{rj}[n_2, n_3]$$
Homogeneous process: A process is called a homogeneous process if the transition probabilities are invariant for a shift, $\pi_{ij}[n_1, n_2] = \pi_{ij}[m]$, where $m = n_2 - n_1$
Then, CK equation is $\pi_{ij}[n + k] = \sum_r \pi_{ir}[k]\pi_{rj}[n]$
$k = n_2 - n_1, n = n_3 - n_2$
Stationary Markoff Chain: If $P[2] = P[1] = P$
$\implies P[n] = P, \forall n$. Or P is an eigenvector of the transition matrix.
Integration by Parts
$$\int_a^b u(x)v'(x)dx = $$
$$[u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$$

Trigonometry Function
$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$$
$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$$
$$\tan(\alpha \pm \beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$$
$$\sin\alpha\cos\beta = 0.5[\sin(\alpha+\beta) + \sin(\alpha-\beta)]$$
$$\cos\alpha\sin\beta = 0.5[\sin(\alpha+\beta) - \sin(\alpha-\beta)]$$
$$\cos\alpha\cos\beta = 0.5[\cos(\alpha+\beta) + \cos(\alpha-\beta)]$$
$$\sin\alpha\sin\beta = -0.5[\cos(\alpha+\beta) - \cos(\alpha-\beta)]$$
$$\sin\alpha \pm \sin\beta = 2\sin\frac{\alpha\pm\beta}{2}\cos\frac{\alpha\mp\beta}{2}$$
$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$
$$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$$
Convolution:
$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$
Taylor Expansion
$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$
$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}; \sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1}$$
$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n; \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} nx^{n-1}$$
$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n}; ln(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{-n}$$
$$\frac{1}{1-ax^{-1}} = \sum_{n=0}^{\infty} a^n x^{-n} = \sum_{n=0}^{\infty} a^{|n|}x^{-n};$$
$$\frac{ax}{1-ax} = \frac{1}{1-ax} - 1 = \sum_{n=1}^{\infty} a^n x^n$$
$$\left[\frac{1}{1-az^{-1}} + \frac{az}{1-az}\right] = \sum_{n=-\infty}^{\infty} a^{|n|}z^{-n}$$
Fourier Definition:
$$F(s) = \int_0^{\infty} f(t)e^{-(\sigma+j\omega)t}dt$$
Fourier Transformation and Inverse FT:
FT:
$$y_f(t) = \int_{-\infty}^{\infty} f(t-\alpha)h(\alpha)d\alpha$$
$$\xrightarrow{FT} Y_f(\omega) = F(\omega)H(\omega)$$
RFT:
$$Y_f(\omega) = F(\omega)H(\omega) \xrightarrow{RFT}$$
$$y_f(t) = \int_{-\infty}^{\infty} F(\omega)H(\omega)e^{j\omega t}\frac{d\omega}{2\pi}$$
P and C:
$$P_m^n = \frac{n!}{(n-m)!} = n(n-1)\ldots(n-m+1);$$
$$C_n^m = \frac{n!}{(n-m!)m!} = \frac{n(n-1)\cdots(n-m+1)}{m(m-1)\ldots 1}$$
$$C_n^m = C_n^{n-m}; C^{m-1} + C_n^m = C_{n+1}^m$$
$$C_n^0 + \cdots + C_n^m = 2^n; C_r^r + \cdots + C_r^n = C_{n+1}^{r+1}$$
$$\sum_{i=0}^{k} C_n^i C_m^{k-i} = C_{m+n}^k$$
Euler's Formula:
$$e^{i\theta} = \cos\theta + i\sin\theta, e^{-i\theta} = \cos\theta - i\sin\theta$$
$$\sin z = \frac{e^{iz} - e^{-iz}}{2i}; \cos z = \frac{e^{iz} + e^{-iz}}{2}$$
$$e^z = e^x(\cos y + i\sin y)$$
Z-Transform: $y[n] = x[n] + ay[n-1]$
$$h[n] = \begin{cases} a^n, & n \ge 0 \\ 0, & n < 0 \end{cases} \to H(z) = \frac{1}{1-az^{-1}}$$
$$H(z) = \sum_{-\infty}^{\infty} h[n]z^{-n}$$
Complex Number Tricks:
$\overline{(z_1 + z_2)*} = z_1^* + z_2^*; (z_1 \cdot z_2)^* = z_1^* \cdot z_2^*; |z| = \sqrt{z * z^*}; Re(z) = \frac{z+z^*}{2}, Im(z) = \frac{z-z^*}{2}; z^{-1} = \frac{z^*}{z \cdot z^*} = \frac{z^*}{|z|^2}$
Z-Transform Tricks: $H(z) = 1 + 0.5z^{-1}$ means current input plus half of the previous input
$s[n] = w[n] + 0.5w[n-1]$, in which z^{-1} means one unit time delay. $H(z) = \frac{1}{1-z^{-1}}$ corresponds to
$s[n] = w[n] + s[n-1]$.
Fourier Transform: $[\delta(t), 1], [u(t), \pi\delta(\omega) + \frac{1}{j\omega}], [c, c \cdot 2\pi\delta(\omega)], [\sin(\omega_0 t), j\pi(\delta(\omega - \omega_0) - \delta(\omega + \omega_0))], [\cos(\omega_0 t), \pi(\delta(\omega - \omega_0) + \delta(\omega + \omega_0))], [e^{at}, \frac{1}{j\omega - a} \forall\Re(a) \le 0], [e^{-j\omega_0 t}, \pi(\delta(\omega - \omega_0) + \delta(\omega + \omega_0))][e^{j\omega_0 t}, \pi(\delta(\omega - \omega_0) + \delta(\omega - \omega_0))], [af(t) + bh(t), aF(\omega) + bH(\omega)], [f(t-t_0), e^{-j\omega t_0}F(\omega)], [e^{j\omega_0 t}f(t), F(\omega - \omega_0)], [f(at), \frac{1}{|a|}F(\frac{\omega}{a})], [\frac{d^n f(t)}{dt^n}, (j\omega)^n F(\omega)], [\int_{-\infty}^t f(\tau)d\tau, \frac{1}{j\omega}F(\omega) + \pi F(0)\delta(\omega)]$
Z Transform: $[\delta[n], 1], [u[n], \frac{1}{1-z^{-1}}, \forall|z| > 1], [a^n u[n], \frac{1}{1-az^{-1}}, \forall|z| > |a|], [nu[n], \frac{z^{-1}}{(1-z^{-1})^2}], [n^2 u[n], \frac{z^{-1}(1+z^{-1})}{(1-z^{-1})^3}], [e^{an}u[n], \frac{1}{1-e^a z^{-1}}], [\sin(\omega_0 n)u[n], \frac{z^{-1}\sin(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}], [\cos(\omega_0 n)u[n], \frac{1-z^{-1}\cos(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}], [af[n] + bg[n], aF(z) + bG(z)], [f[n-n_0], z^{-n_0}F(z)], [f[n+n_0], z^{n_0}F(z) - \sum_{k=0}^{n_0-1} z^{n_0-k}f[k]], [a^n f[n], F(\frac{z}{a})], [nf[n], -z\frac{d}{dz}F(z)], [(f * g)[n], F(z) \cdot G(z)]$