References

- 10-page Probability Cheatsheet: pdf | github | website
- Introduction to Probability & Statistics

Terminology

- **random experiment**: process by which we observe something uncertain
- **outcome**: result of a random experiment.
- **sample space**: set of all possible outcomes.
- When we repeat a random experiment several times, we call each one of them a **trial**.
- **events**: a set of possible outcomes; a subset of sample space

Joint Distributions

Bivariate Normal Distribution

- X and Y are **bivariate normal** or **jointly normal**, if $aX + bY$ has a normal distribution $\forall\ a, b \in \mathbb{R}$.
    - X and Y are jointly normal => they are individually normal as well.
    - If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are independent, then they are jointly normal
        - $X + Y \sim\ N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho(X,Y)\sigma_X\sigma_Y)$
- **Standard Bivariate Normal Distribution** with correlation coefficient ρ:
    - $f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho xy + y^2]\}$
- **Bivariate Normal Distribution** with parameters μ_X, μ_Y, σ_X^2, σ_Y^2 and ρ:
    - $f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right\}$
- X and Y have bivariate normal distribution iff $\exists Z_1, Z_2 \sim N(0,1)$ s.t.
    - $X = \sigma_X Z_1 + \mu_X$
    - $Y = \sigma_Y(\rho Z_1 + \sqrt{1-\rho^2}Z_2) + \mu_Y$
    - This gives a way to generate jointly normal X and Y from standard normal Z_1 and Z_2.
- For jointly normal random variables, being independent and being uncorrelated are equivalent.

Random Vectors

- $\mathbf{X} = [X_1, \ldots, X_n]^T$
- $\mathbf{EX} = [EX_1, \ldots, EX_n]^T$
- Correlation matrix of X: $\mathbf{R_X} = \mathbf{E[XX^T]}$
- Covariance matrix of X: $\mathbf{C_X} = \mathbf{E[(X-EX)(X-EX)^T] = R_X - EXEX^T}$
    - $\mathbf{C_X}$ is psd.
    - $\mathbf{C_X}$ is +ve definite iff all its eigenvalues are larger than zero or equivalently $\det(\mathbf{C_X}) > 0$.
    - If $\mathbf{Y = AX + b}$, then $\mathbf{C_Y = AC_XA^T}$.
- **Cross Correlation matrix** of X and Y: $\mathbf{R_{XY} = E[XY^T]}$
- **Cross Covariance matrix** of X and Y: $\mathbf{C_{XY}} = E[(X-EX)(Y-EY)^T]$
- **The Method of Transformations for Random Vectors**
    - Let
        - $G : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous and invertible function with continuous partial derivatives and let $H = G^{-1}$.
        - $\mathbf{Y} = G(\mathbf{X})$ and $\mathbf{X} = G^{-1}(\mathbf{Y}) = H(\mathbf{Y}) = [H_1(\mathbf{Y}), \ldots H_n(\mathbf{Y})]^T$.
    - Then
        - PDF of $\mathbf{Y}$ is: $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(H(\mathbf{y}))|J|$
        - where J is the Jacobian of H evaluated at $\mathbf{Y}$:

$$J = \det\begin{bmatrix} \frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2} & \cdots & \frac{\partial H_1}{\partial y_n} \\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_2} & \cdots & \frac{\partial H_2}{\partial y_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial H_n}{\partial y_1} & \frac{\partial H_n}{\partial y_2} & \cdots & \frac{\partial H_n}{\partial y_n} \end{bmatrix}(y_1, \ldots, y_n),$$

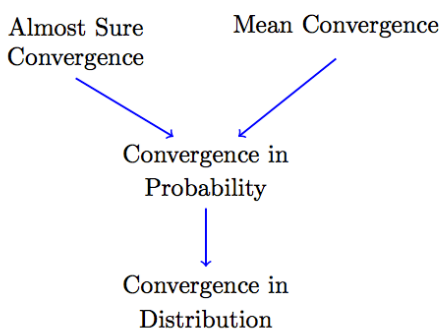
- If $\mathbf{Y = AX + b}$, then $f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det(\mathbf{A})|}f_{\mathbf{X}}\left(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})\right)$.
- **Normal (Gaussian) Random Vectors**
    - Random variables $X_1, \ldots, X_n$ are said to be **jointly normal** if $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ is normal $\forall a_i \in \mathbb{R}$

- Random vector $\mathbf{X} = [X_1, \ldots, X_n]^T$ is said to be **normal** or **Gaussian** if $X_1, \ldots, X_n$ are **jointly normal**.
- For *standard normal random vector* $f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right\}$.
- **Multivariate Normal Distribution** with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$:
  - $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \mathbf{C}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\}$
- $\mathbf{X}$ has multivariate normal distribution with mean $\mathbf{m}$ and **+ve definite** covariance matrix $\mathbf{C}$ iff there exists a standard normal $\mathbf{Z}$ s.t.
  - $\mathbf{X} = \mathbf{AZ} + \mathbf{m}$ where $\mathbf{A}$ is a matrix s.t. $\mathbf{AA}^T = \mathbf{A}^T\mathbf{A} = \mathbf{C}$
    - Note that $\mathbf{A} = \mathbf{QD}^{\frac{1}{2}}\mathbf{Q}^T$ where $\mathbf{C} = \mathbf{QDQ}^T$
  - This gives a way to generate normal $\mathbf{X} \sim N(\mathbf{m}, C)$ from standard normal $\mathbf{Z}$.
- For a normal $\mathbf{X}$, following are equivalent:
  - X_i's are independent
  - X_i's are uncorrelated
  - $\mathbf{C}_{\mathbf{X}}$ is diagonal.
- If $\mathbf{X} \sim N(\mathbf{m}, C)$ then $\mathbf{Y} = \mathbf{AX} + \mathbf{b} \sim N(\mathbf{A}E\mathbf{X} + \mathbf{b}, \mathbf{ACA}^T)$

Probability Bounds

- **Union Bound**
  - For any events $A\_1, A\_2, \ldots, A_n$, we have, $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$
- **Bonferroni Inequalities**
  - Generalization of the Union Bound using Inclusion-Exclusion Principle
- **Markov's Inequality**
  - If X is any *nonnegative* random variable, then
    * $P(X \geq a) \leq \frac{EX}{a}, \qquad \text{for any } a > 0$
    * Derive using definition of EX
- **Chebyshev's Inequality**
  - If X is any random variable, then
    - $P\left(|X - EX| \geq b\right) \leq \frac{Var(X)}{b^2}, \qquad \text{for any } b > 0$
  - Derive using Markov's Inequality
- **Chernoff Bounds** (derive using Markov Ineq)
  - $P(X \geq a) \leq e^{-sa}M_X(s), \qquad \text{for all } s > 0$
  - $P(X \leq a) \leq e^{-sa}M_X(s), \qquad \text{for all } s < 0$
  - where, $M_X(s) = E[e^{sX}]$ is the moment generating function.
  - Derive using Markov's Inequality
- **Cauchy-Schwarz**: $|EXY| \leq \sqrt{E[X^2]E[Y^2]}$

Convergence of Random Variables

- 

- Convergence of a sequence of random variables $X1, X2, X3, \cdots$ to random variable X.

Convergence in distribution

- $X_n \xrightarrow{L^r} X$ if, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$,
  - for all x at which $F_X(x)$ is continuous
- *Central Limit Theorem*

Convergence in probability

- $X_n \xrightarrow{L^r} X$ if, $\lim_{n\to\infty} P\left(|X_n - X| \geq \epsilon\right) = 0, \qquad \text{for all } \epsilon > 0$
- *Weak Law of large Numbers*

Convergence in mean:

- $X_n \xrightarrow{L^r} X$ if, $\lim_{n\to\infty} E\left(|X_n - X|^r\right) = 0$
- if $r = 2$, **mean-square convergence**: $X_n \xrightarrow{m.s.} X$

Almost sure convergence

- S is the sample space of X_i and X
- $X_n \xrightarrow{a.s.} X$ if $P\left(\{s \in S : \lim_{n\to\infty} X_n(s) = X(s)\}\right) = 1$
- $X_n \xrightarrow{a.s.} X$ if $\sum_{n=1}^{\infty} P\big(|X_n - X| > \epsilon\big) < \infty, \qquad \text{for all } \epsilon > 0$
- $X_n \xrightarrow{a.s.} X$ *iff* $\lim_{m\to\infty} P(A_m) = 1, \qquad \text{for all } \epsilon > 0$
    - where, $A_m = \{|X_n - X| < \epsilon, \qquad \text{for all } n \geq m\}$.
- *Strong Law of large Numbers*

Continuous Mapping Theorem

- For a *continuous* $h : \mathbb{R} \mapsto \mathbb{R}$
- $X_n \xrightarrow{d} X \implies h(X_n) \xrightarrow{d} h(X)$
- $X_n \xrightarrow{p} X \implies h(X_n) \xrightarrow{p} h(X)$
- $X_n \xrightarrow{a.s.} X \implies h(X_n) \xrightarrow{a.s.} h(X)$

Statistical Inference I: Frequentist Inference

Point Estimators

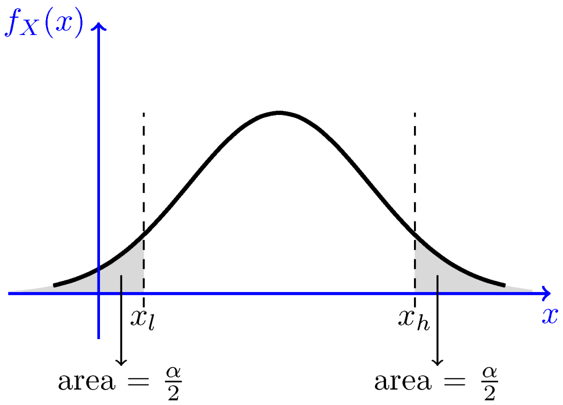
- In Frequentist Inference, θ is a deterministic (non-random) unknown/parameter that is to be estimated from the observed data.
- *Point estimator* $\hat{\Theta}$ for θ is a function of the random sample $X_1, X_2, \cdots, X_n$
    - $\hat{\Theta} = h(X_1, X_2, \cdots, X_n)$
- *Bias*: $B(\hat{\Theta}) = E[\hat{\Theta}] - \theta$
- *Unbiased Estimator*: $B(\hat{\Theta}) = 0$
- *Mean Squared Error*: $MSE(\hat{\Theta}) = E\big[(\hat{\Theta} - \theta)^2\big] = \mathrm{Var}(\hat{\Theta} - \theta) + \big(E[\hat{\Theta} - \theta]\big)^2 = \mathrm{Var}(\hat{\Theta}) + B(\hat{\Theta})^2$
- $\hat{\Theta}$ is a *consistent estimator* for θ
    - if $\lim_{n\to\infty} P\big(|\hat{\Theta}_n - \theta| \geq \epsilon\big) = 0, \text{ for all } \epsilon > 0$.
    - if $\lim_{n\to\infty} MSE(\hat{\Theta}_n) = 0$.
    - i.e. as the sample size gets larger, $\hat{\Theta}$ converges to the real value of θ.
- Let $X_1, X_2, \cdots, X_n$ be a random sample with
    - mean: $EX_i = \mu < \infty$, variance: $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$,
    - PDF: $f_X(x)$ and CDF: $F_X(X)$.
- *Order Statistics*: $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$: the random samples in ascending order.
    - $X_{(1)} = min(X_1, X_2, \cdots, X_n)$, $X_{(n)} = max(X_1, X_2, \cdots, X_n)$
    - PDF: $f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} f_X(x) \left[F_X(x)\right]^{i-1} \left[1 - F_X(x)\right]^{n-i}$
    - CDF: $F_{X_{(i)}}(x) = \sum_{k=i}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}$
    - Joint PDF:

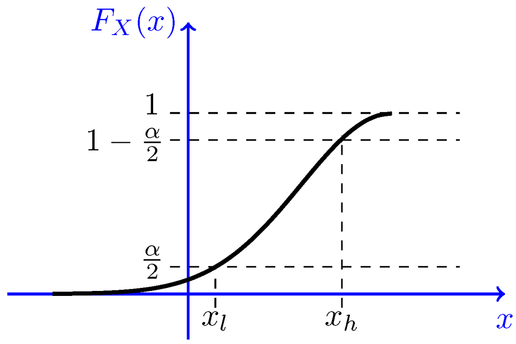
$$f_{X_{(1)},\cdots,X_{(n)}}(x_1, x_2, \cdots, x_n) = \begin{cases} n! f_X(x_1) f_X(x_2) \cdots f_X(x_n) & \text{for } x_1 \leq x_2 \leq x_2 \cdots \leq x_n \\ \\ 0 & \text{otherwise} \end{cases}$$

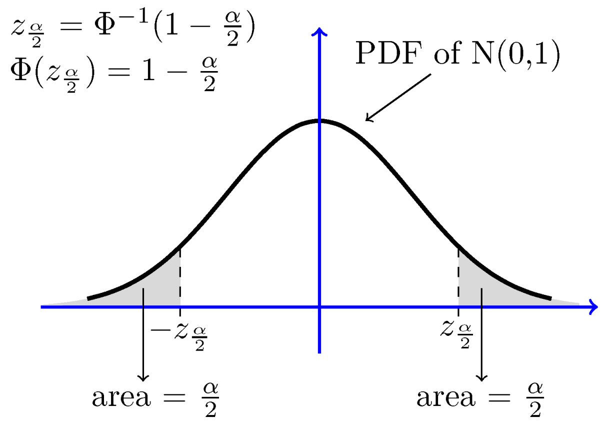
- *Sample Mean*: $\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$: unbiased and consistent estimator of μ.
    - $E[\overline{X}] = \mu, Var(\overline{X}) = \frac{\sigma^2}{n}$
- *Sample Variance*: $S^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2 = \frac{1}{n-1} \left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right)$: unbiased and consistent estimator of σ^2.
- *Sample Standard Deviation*: $S = \sqrt{S^2}$: biased estimator of σ
- *Maximal Likelihood Estimate*: $\theta_{ML} = \arg\max_{\theta} L(x_1, x_2, \cdots, x_n; \theta)$
- *Maximal Likelihood Estimator*: $\Theta_{ML} = \Theta_{ML}(X_1, X_2, \cdots, X_n)$ s.t. $\Theta_{ML}(x_1, x_2, \cdots, x_n) = \theta_{ML}$
    - MLE is asymtotically consistent: $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta| > \epsilon) = 0$
    - MLE is asymtotically unbiased: $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
    - As n becomes large, $\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{\mathrm{Var}(\hat{\Theta}_{ML})}} \xrightarrow{d} N(0, 1)$

Confidence Interval Estimation

- Let $X_1, X_2, \cdots, X_n$ be a random sample with
    - mean: $EX_i = \mu < \infty$, variance: $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$,
    - PDF: $f_X(x)$ and CDF: $F_X(X)$.
- Let θ be the distribution parameter to be estimated.

- *Interval Estimator* with *confidence level* $1-\alpha$:
  - Two estimators $\Theta_l = \Theta_l(X_1, X_2, \cdots, X_n)$ and $\Theta_h = \Theta_h(X_1, X_2, \cdots, X_n)$ s.t.
  - $P\left(\hat{\Theta}_l \leq \theta \leq \hat{\Theta}_h\right) \geq 1-\alpha$, for all possible θ
- Equivalently, $[\hat{\Theta}_l, \hat{\Theta}_h]$ is a $(1-\alpha)100\%$ *confidence interval* for θ.
- Finding Interval Estimators
  - Detour: Finding x_l and x_h s.t. $P\left(x_l \leq \theta \leq x_h\right) = 1-\alpha$.
  - One solution: $x_l = F_X^{-1}\left(\frac{\alpha}{2}\right)$ and $x_h = F_X^{-1}\left(1-\frac{\alpha}{2}\right)$
  - 



  - Example: the $(1-\alpha)$ interval for $Z \sim N(0,1)$ is given by $[-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]$
    - where, $z_p = \Phi^{-1}(1-p)$ where Φ is CDF of Z.
    - In particular, 95% confidence interval is given $[-1.96, 1.96]$
    - 

  - *Pivotal Quantity*: Q is a *pivot* if,
    - $Q = Q(X_1, X_2, \cdots, X_n, \theta)$ i.e. it doesn't depend on any unknown parameter other than θ and,
    - its distribution does not depend on θ or any other unknown parameters.
  - *Pivotal Method for finding Confidence Intervals*
    1. Find a pivotal quantity $Q = Q(X_1, X_2, \cdots, X_n, \theta)$
    * pivots already available for most important cases that appear in practice.
    3. Find an interval for Q such that, $P\left(q_l \leq Q \leq q_h\right) = 1-\alpha$
    4. Using algebraic manipulations, convert the above equation to $P\left(\hat{\Theta}_l \leq \theta \leq \hat{\Theta}_h\right) = 1-\alpha$
    - Examples of known Interval Estimators:
      - Given a random sample $X_1, X_2, \cdots, X_n$ want to estimate $\theta = EX_i$.
        - Assumption: n is large => can use CLT and sample statistics like $\overline{X}$ and S^2.
      - Example 1:
        - Assumptions:
          - **known variance** $Var(X_i) = \sigma^2 < \infty$:
        - $\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$: approximate $(1-\alpha)100\%$ confidence interval for θ.
          - approximate because CLT was used.
      - **unknown variance**:
        - Option 1: Use an upper bound for σ^2: $\sigma_{max} \geq \sigma$

      - $\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma_{max}}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma_{max}}{\sqrt{n}}\right]$: a **valid but conservative** $(1-\alpha)100\%$ confidence interval for θ.
    - Option 2: Use a point estimate for σ^2: $\hat{\sigma}^2$
      - Use of point estimate leads to additional approximation in intervals.
      - Example 2:
        - Assumptions: random sample is from $Bernoulli(\theta)$.
        - $\left[\overline{X} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \overline{X} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right]$: approximate $(1-\alpha)100\%$ confidence interval for θ.
          - Using $\hat{\sigma}^2 = \hat{\theta}(1-\hat{\theta}) = \overline{X}(1-\overline{X})$
      - **Example 3:**
        - $\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$: approximate $(1-\alpha)100\%$ confidence interval for θ.
          - Using sample variance as point estimate: $\hat{\sigma}^2 = S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\right)}$

Confidence Intervals for Normal Samples

- Let $X_1, X_2, \cdots, X_n$ be n i.i.d. $N(\mu, \sigma^2)$ r.v. with sample mean and variance $\overline{X}$ and S^2 respectively.
  - then $\overline{X}$ and S^2 are independent
- **Chi-Squared distribution**
  - $\chi^2(n) = Gamma(\frac{n}{2}, \frac{1}{2})$
  - Use to estimate variance of normal random variables
    - Let $Y = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ then $Y \sim \chi^2(n-1)$
- **Student's t-distribution**
  - $T = \frac{Z}{\sqrt{Y/n}}$ where $Z \sim N(0,1)$ and $Y \sim \chi^2(n)$ are independent r.v.
  - Use to estimate mean of normal random variables
    - Let $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$ then $T \sim T(n-1)$

Random Processes

Basics

- terminology:
  - random or stochastic process
  - **continuous-time** random process: $\{X(t), t \in J\}$, where J is an interval such as $[0,1]$, $[0,\infty]$, $[-\infty,\infty]$, etc.
  - **discrete-time** random process (or a **random sequence**): $\{X(n) = X_n, n \in J\}$ where J is a countable set such as $\mathbb{Z}$ or $\mathbb{N}$.
  - A random process is a **random function** of time.
    - **Realization** or **sample function** or **sample path** of $X(t)$.
- CDF: $F_{X(t_1)X(t_2)\cdots X(t_n)}(x_1, x_2, \cdots, x_n) = P\big(X(t_1) \le x_1, X(t_2) \le x_2, \cdots, X(t_n) \le x_n\big)$
- **Mean Function**: $\mu_X(t) : J \to \mathbb{R}$ s.t. $\mu_X(t) = E[X(t)]$
- **(Auto)correlation Function**: $R_X(t_1, t_2) = E[X(t_1)X(t_2)], \quad \text{for } t_1, t_2 \in J.$
  - $R_X(t,t) = E(X(t)^2)$
- **(Auto)covariance Function**: $C_X(t_1, t_2) = \text{Cov}\big(X(t_1), X(t_2)\big) = R_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2), \quad \text{for } t_1, t_2 \in J.$
  - $C_X(t,t) = Var(X(t))$
- **crosscorrelation Function**: $R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)], \quad \text{for } t_1, t_2 \in J.$
  - $R_X(t,t) = E(X(t)^2)$ - **expected (average) power**
- **crosscovariance Function**: $C_{XY}(t_1, t_2) = \text{Cov}\big(X(t_1), Y(t_2)\big) = R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2), \quad \text{for } t_1, t_2 \in J.$
  - $C_X(t,t) = Var(X(t))$
- **Independent Random Processes**:
  - $F_{X(t_1),X(t_2),\cdots,X(t_m),Y(t'_1),Y(t'_2),\cdots,Y(t'_n)}(x_1, x_2, \cdots, x_m, y_1, y_2, \cdots, y_n) = F_{X(t_1),X(t_2),\cdots,X(t_m)}(x_1, x_2, \cdots, x_m) \cdot F_{Y(t'_1),Y(t'_2),\cdots,Y(t'_n)}(y_1, y_2, \cdots, y_n)$
    - for all $t_i \in J, t_j \in J'$ and $x_i, y_i \in \mathbb{R}$

Stationary Processes

- The following definitions assume these two random processes:
  - continuous-time r.p. $\{X(t), t \in \mathbb{R}\}$
  - discrete-time r.p. $\{X(n) = X_n, n \in \mathbb{Z}\}$
- **Strict-sense stationary** (or simply **stationary**)

- continuous-time: $F_{X(t_1)X(t_2)\cdots X(t_r)}(x_1, x_2, \cdots, x_r) = F_{X(t_1+\Delta)X(t_2+\Delta)\cdots X(t_r+\Delta)}(x_1, x_2, \cdots, x_r)$ for all $t_i, \Delta \in \mathbb{R}$, $x_i \in \mathbb{R}$.
- discrete-time: $F_{X(n_1)X(n_2)\cdots X(n_r)}(x_1, x_2, \cdots, x_n) = F_{X(n_1+D)X(n_2+D)\cdots X(n_r+D)}(x_1, x_2, \cdots, x_r)$ for all $n_i, D \in \mathbb{Z}$, $x_i \in \mathbb{R}$.

- **Weak-sense stationary** or **Wide-sense stationary** (**WSS**)
  - continuous-time: $\mu_X(t) = \mu_X$ and $R_X(t1, t2) = R_X(t1 - t2)$
  - discrete-time: $\mu_X(n) = \mu_X$ and $R_X(n1, n2) = R_X(n1 - n2)$
  - Properties of the correlation function of WSS r.p.s $R_X(\tau)$
    - **expected (average) power** constant in time $E[X(t)^2] = R_X(0)$
    - $R_X(0) \geq 0$
    - $R_X(\tau) = R_X(-\tau), \quad \text{for all } \tau \in \mathbb{R}.$ (even function)
    - $|R_X(\tau)| \leq R_X(0), \quad \text{for all } \tau \in \mathbb{R}.$ - (prove using Cauchy-Shwartz)
- **Jointly Wide-Sense Stationary** : $X(t)$ and $Y(t)$ are JWSS if
  - both are WSS
  - $R_{XY}(t_1, t_2) = R_{XY}(t_1 - t_2)$.
- **Cyclostationary Processes**
  - *Strict-Sense*: $\exists T$ such that $X(t_1), X(t_2), \cdots, X(t_r)$ has same joint CDF as $X(t_1 + T), X(t_2 + T), \cdots, X(t_r + T)$.
  - *Weak-Sense*: $\exists T$ such that $\mu_X(t + T) = \mu_X(t)$ and $R_X(t1 + T, t2 + T) = R_X(t1, t2)$.
  - Similarly for discrete-time.
- Calculus of Random Processes
  - **Mean-square continuous**: $\lim_{\delta \to 0} E\left[|X(t + \delta) - X(t)|^2\right] = 0.$
  - Derivatives and integrals are linear operations and hence can often be interchanged with expectation.
    - $E\left[\int_0^t X(u)du\right] = \int_0^t E[X(u)]du$
    - $E\left[\frac{d}{dt}X(t)\right] = \frac{d}{dt}E[X(t)]$.

Gaussian Random Processes

- $\{X(t), t \in J\}$ is a **Gaussian (normal) random process** if, $X(t_1), \ldots, X(t_n)$ are jointly normal $\forall t_i \in J$
  - For normal random processes, wide-sense stationarity and strict-sense stationarity are equivalent.
- $\{X(t), t \in J\}$ $\{Y(t), t \in J'\}$ are **Jointly Gaussian (normal) random processes** if, $X(t_1), \ldots, X(t_n), Y(t'_1), \ldots, Y(t'_n)$ are jointly normal $\forall t_i \in J$ and $\forall t'_i \in J'$
  - For jointly Gaussian random processes being uncorrelated and being independent are equivalent.

Poisson Process

Discrete-Time Markov Chain