Theorems and Definitions

T1.1 De Morgan's law relates all three basic operations: $(A \cup B)^C = A^C \cap B^C$.

D1.1 **Outcome**: An outcome of an experiment is any possible observation of that experiment.

D1.2 **Sample space**: The sample space of an experiment is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.

D1.3 **Event**: An event is a set of outcomes of an experiment.

D1.4 **Event Space**: An event space is a collectively exhaustive, mutually exclusive set of events.

T1.2 For an event space $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are mutually exclusive and $A = C_1 \cup C_2 \cup \cdots$.

D1.5 **Axioms of Probability**: A probability measure $P[.]$ is a function that maps events in the sample space to real numbers such that Axiom 1: For any event A, $P[A] \geq 0$. Axiom 2: $P[S] = 1$. Axiom 3: For any countable collection $A_1, A_2, \ldots$ of mutually exclusive events, $P[A_1 \cup A_2 \cup \ldots] = P[A_1] + P[A_2] + \cdots$.

T1.3 For mutually exclusive events A_1 and A_2, $P[A_1 \cup A_2] = P[A_1] + P[A_2]$.

T1.4 If $A = A_1 \cup A_2 \cup \cdots \cup A_m$ and $A_i \cap A_j = \phi$ for $i \neq j$, then $P[A] = \sum_{i=1}^{m} P[A_i]$.

T1.5 The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes contained in the event: $P[B] = \sum_{i=1}^{m} P[\{s_i\}]$.

T1.6 For an experiment with sample space $S = \{s_1, \ldots, s_n\}$ in which each outcome s_i is equally likely, $P[s_i] = \frac{1}{n} \quad 1 \leq i \leq n$.

T1.7 The probability measure $P[.]$ satisfies (a) $P[\phi] = 0$. (b) $P[A^C] = 1 - P[A]$. (c) For any A and B (not necessarily disjoint), $P[A \cup B] = P[A] + P[B] - P[A \cap B]$. (d) If $A \subset B$, then $P[A] \leq P[B]$.

T1.8 For any event A, and event space $\{B_1, B_2, \ldots, B_m\}$, $P[A] = \sum_{i=1}^{m} P[A \cap B_i]$.

D1.6 **Conditional Probability**: The conditional probability of the event A given the occurrence of the event B is $P[A|B] = \frac{P[AB]}{P[B]}$.

T1.9 A conditional probability measure $P[A|B]$ has the following properties that correspond to the axioms of probability. Axiom 1: $P[A|B] \geq 0$. Axiom 2: $P[B|B] = 1$. Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ with $A_i \cap A_j = \phi$ for $i \neq j$, then $P[A|B] = P[A_1|B] + P[A_2|B] + \cdots$.

T1.10 **Law of Total Probability**: For an event space $\{B_1, B_2, \ldots, B_m\}$ with $P[B_i] > 0$ for all i, $P[A] = \sum_{i=1}^{m} P[A|B_i]P[B_i]$.

T1.11 **Bayes' Theorem**: $P[B|A] = \frac{P[AB]}{P[A]} = \frac{P[A|B]P[B]}{P[A]}$.

D1.7 **Two independent Events:** Events A and B are independent if and only if $P[AB] = P[A]P[B]$. When events A and B have nonzero probabilities, the following formulas are equivalent to the definition of independent events: $P[A|B] = P[A]$, $\quad P[B|A] = P[B]$. **Independent and disjoint are *not* synonyms.**

D1.8 **3 Independent Events**: A_1, A_2, and A_3 are independent if and only if (a) A_1 and A_2 are independent, (b) A_2 and A_3 are independent, (c) A_1 and A_3 are independent, (d) $P[A_1 \cap A_2 \cap A_3] = P[A_1]P[A_2]P[A_3]$.

D1.9 **More than Two Independent Events**: If $n \geq 3$, the sets $A_1, A_2, \ldots, A_n$ are independent if and only if (a) every set of $n - 1$ sets from $A_1, A_2, \ldots, A_n$ are independent, (b) $P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2] \cdots P[A_n]$.

D1.10 **Fundamental Principle of Counting**: If subexperiment A has n possible outcomes, and subexperiment B has k possible outcomes, then there are nk possible outcomes when you perform both experiments.

T1.12 The number of k-permutations of n distinguishable objects is $(n)_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$. E.g. sampling without replacement.

T1.13 The number of ways to choose k objects out of n distinguishable objects is $\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}$.

D1.11 ***n* choose *k***: For an integer $n \geq 0$, we define $\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & k = 0, 1, \ldots, n, \\ 0 & otherwise. \end{cases}$

T1.14 Given m distinguishable objects, there are m^n ways to choose with replacement an ordered sample of n objects. E.g. sampling with replacement.

T1.15 For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, there are m^n possible observation sequences.

T1.16 The number of observation sequences for n subexperiments with sample space $S = \{0, 1\}$ with 0 appearing n_0 times and 1 appearing $n_1 = n - n_0$ times is $\binom{n}{n_1}$.

T1.17 For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, the number of length $n = n_0 + \cdots + n_{m-1}$ observation sequences with s_i appearing n_i times is $\binom{n}{n_0, \ldots, n_{m-1}} = \frac{n!}{n_0! n_1! \ldots n_{m-1}!}$.

D1.12 **Multinomial Coefficient**: For an integer $n \geq 0$, we define $\binom{n}{n_0, \ldots, n_{m-1}} = \begin{cases} \frac{n!}{n_0! n_1! \ldots n_{m-1}!} & \begin{matrix} n_0 + \cdots + n_{m-1} = n; \\ n_i \in \{0, 1, \ldots, n\}, i = 0, 1, \ldots, m-1, \end{matrix} \\ 0 & otherwise. \end{cases}$

T1.18 The probability of n_0 failures and n_1 successes in $n = n_0 + n_1$ independent trails is $P[S_{n_0, n_1}] = \binom{n}{n_1}(1-p)^{n-n_1}p^{n_1} = \binom{n}{n_0}(1-p)^{n_0}p^{n-n_0}$.

T1.19 A subexperiment has sample space $S = \{s_0, \ldots, s_{m-1}\}$ with $P[s_i] = p_i$. For $n = n_0 + \cdots + n_{m-1}$ independent trails, the probability if n_i occurrences of s_i, $i = 0, 1, \ldots, m-1$, is $P[S_{n_0, \ldots, n_{m-1}}] = \binom{n}{n_0, \ldots, n_{m-1}} p_0^{n_0} \cdots p_{m-1}^{n_{m-1}}$.

T1.20 Reliability Problems: (a) Components in series: The probability that the operation succeeds is $P[W] = P[W_1 W_2 \ldots W_n] = p \times p \times \ldots \times p = p^n$. (b) Components in parallel: The probability that the parallel operation succeeds is $P[W] = 1 - P[W^C] = 1 - (1-p)^n$.

D2.1 **Random Variable**: A random variable consists of an experiment with a probability measure $P[.]$ defined on a sample space S and a function that assigns a real number to each outcome in the sample space of the experiment.

D2.2 **Discrete Random Variable**: X is a discrete random variable if the range of X is a countable set $S_X = \{x_1, x_2, \ldots\}$.

D2.3 **Finite Random Variable**: X is a finite random variable if the range is a finite set $\{x_1, x_2, \ldots, x_n\}$.

D2.4 **Probability Mass Function (PMF)**: The probability mass function (PMF) of the discrete random variable X is $P_X(x) = P[X = x]$.

T2.1 For a discrete random variable X with PMF $P_X(x)$ and range S_X: (a) For any x, $P_X(x) \geq 0$. (b) $\sum_{x \epsilon S_X} P_X(x) = 1$. (c) For any event $B \subset S_X$, the probability that X is in the set B is $P[B] = \sum_{x \in B} P[X = x] = \sum_{x \in B} P_X(x)$.

D2.5 **Bernoulli** (p) **Random Variable**: X is a Bernoulli (p) random variable if the PMF of X has the form $P_X(x) = \begin{cases} 1-p & x = 0, \\ p & x = 1, \\ 0 & otherwise, \end{cases}$ where the parameter p is in the range $0 < p < 1$. E.g. a coin flip.

D2.6 **Geometric** (p) **Random Variable**: X is a geometric (p) random variable if the PMF of X has the form $P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & otherwise, \end{cases}$ where the parameter p is in the range $0 < p < 1$. E.g. first success.

D2.7 **Binomial** (n, p) **Random Variable**: X is a binomial (n, p) random variable if the PMF of X has the form $P_X(x) = \binom{n}{x}p^x(1-p)^{n-x}$, where $0 < p < 1$ and n is an integer such that $n \geq 1$. E.g. number of heads from n coin flips.

D2.8 **Pascal** (k, p) **Random Variable**: X is a Pascal (k, p) random variable if the PMF of X has the form $P_X(x) = \binom{x-1}{k-1}p^k(1-p)^{x-k}$, where $0 < p < 1$ and k is an integer such that $k \geq 1$. E.g. continue until 3 failures: $P_Z(z) = \binom{z-1}{2}p^3 1 - p^{z-3} \quad z = 3, 4, 5, \ldots$.

D2.9 **Discrete Uniform** (k, l) **Random Variable**: X is a discrete uniform (k, l) random variable if the PMF of X has the form $P_X(x) = \begin{cases} \frac{1}{l-k+1} & x = k, k+1, k+2, \ldots, l, \\ 0 & otherwise, \end{cases}$ where the parameters k and l are integers such that $k < l$.

D2.10 **Poisson** (α) **Random Variable**: X is a Poisson (α) random variable if the PMF of X has the form $P_X(x) = \begin{cases} \frac{\alpha^x e^{-\alpha}}{x!} & x = 0, 1, 2, \ldots, \\ 0 & otherwise, \end{cases}$ where the parameter α is in the range $\alpha > 0$. A Poisson model often specifies an average rate, λ arrivals per second and a time interval, T seconds. In this time interval, the number of arrivals X has a Poisson PMF with $\alpha = \lambda T$. E.g. if $\lambda = \frac{1}{15} \frac{queries}{second}$ and $T = 10$ seconds, then $\alpha = \frac{10}{15} = \frac{2}{3}$ queries.

D2.11 **Cumulative Distribution Function (CDF)**: The cumulative distribution function (CDF) of random variable X is $F_X(x) = P[X \leq x]$.

T2.2 For any discrete random variable X with range $S_X = \{x_1, x_2, \ldots\}$ satisfying $x_1 \leq x_2 \leq \ldots$, (a) Going from left to right on the x-axis, $F_X(x)$ starts at zero and ends at one. (b) The CDF never decreases as it goes from left to right. (c) For a discrete random Variable X, there is a jump (discontinuity) at each value of $x \in S_X$. The height of the jump at x_i is $P_X(x_i)$. (d) Between jumps, the graph of the CDF of the discrete random variable X is a horizontal line.

T2.3 For all $b \geq a$, $F_X(b) - F_X(a) = P[a < X \leq b]$.

D2.12 **Mode**: A mode of random variable X is a number x_{mod} satisfying $P_X(x_{mod}) \geq P_X(x)$ for all x.

D2.13 **Median**: A median, x_{med}, of random variable X is a number that satisfies $P[X < x_{med}] = P[X > x_{med}]$.

D2.14 **Expected Value**: The expected value of X is $E[X] = \mu_X = \sum_{x \epsilon S_X} x P_X(x)$.

T2.4 The Bernoulli (p) random variable X has expected value $E[X] = p$.

T2.5 The geometric (p) random variable X has expected value $E[X] = \frac{1}{p}$.

T2.6 The Poisson (α) random variable in Definition 2.10 has expected value $E[X] = \alpha$.

T2.7 (a) For the binomial (n, p) random variable X of Definition 2.7, $E[X] = np$. (b) For the Pascal (k, p) random variable X of Definition 2.8, $E[X] = \frac{k}{p}$. (c) For the discrete uniform (k, l) random variable X of Definition 2.9, $E[X] = \frac{k+l}{2}$.

T2.8 Perform n Bernoulli trails. In each trail, let the probability of success be α/n, where $\alpha > 0$ is a constant and $n > \alpha$. Let the random variable K_n be the number of successes in the n trails. As $n \to \infty$, $P_{K_n}(k)$ converges to the PMF of a Poisson (α) random variable.

D2.15 **Derived Random Variable**: Each sample value y of a derived random variable Y is a mathematical function $g(x)$ of a sample value x of another random variable X. We adopt the notation $Y = g(X)$ to describe the relationship of the two random variables.

T2.9 For a discrete random variable X, the PMF of $Y = g(X)$ is $P_Y(y) = \sum_{x:g(x)=y} P_X(x)$.

T2.10 Given a random variable X with PMF $P_X(x)$ and the derived random variable $Y = g(X)$, the expected value of Y is $E[Y] = \mu_Y = \sum_{x \in S_X} g(x)P_X(x)$.

T2.11 For any random variable X, $E[X - \mu_X] = 0$.

T2.12 For any random variable X, $E[aX + b] = aE[X] + b$.

D2.16 **Variance**: The variance of random variable X is $Var[X] = E[(X - \mu_X)^2]$.

D2.17 **Standard Deviation**: The standard deviation of random variable X is $\sigma_X = \sqrt{\text{Var[X]}}$.

T2.13 $Var[X] = E[X^2] - \mu_X^2 = E[X^2] - (E[X])^2$.

D2.18 **Moments**: For random variable X: (a) The nth moment is $E[X^n]$. (b) The nth central moment is $E[(X - \mu_X)^n]$.

T2.14 $Var[aX + b] = a^2 Var[X]$.

T2.15 (a) If X is Bernoulli (p), then $Var[X] = p(1 - p)$. (b) if X is geometric, then $Var[X] = \frac{1-p}{p^2}$. (c) If X is binomial (n, p), then $Var[X] = np(1 - p)$. (d) If X is Pascal (k, p), then $Var[X] = \frac{k(1-p)}{p^2}$. (e) If X is Poisson (α), the $Var[X] = \alpha$. (f) If X is discrete uniform (k, l), then $Var[X] = \frac{(l-k)(l-k+2)}{12}$.

D2.19 **Conditional PMF**: Given the event B, with $P[B] > 0$, the conditional probability mass function of X is $P_{X|B}(x) = P[X = x|B]$.

T2.16 A random variable X resulting from an experiment with event space $B_1, \dots, B_m$ has PMF $P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x)P[B_i]$.

T2.17 $P_{X|B}(x) = \begin{cases} \frac{P_X(x)}{P[B]} & x \in B, \\ 0 & otherwise. \end{cases}$

T2.18 (a) For any $x \in B$, $P_{X|B}(x) \geq 0$. (b) $\sum_{x \in B} P_{X|B}(x) = 1$. (c) For any event $C \subset B$, $P[C|B]$, the conditional probability that X is in the set C, is $P[C|B] = \sum_{x \in C} P_{X|B}(x)$.

D2.20 **Conditional Expected Value**: The conditional expected value of random variable X given condition B is $E[X|B] = \mu_{X|B} = \sum_{x \in B} xP_{X|B}(x)$.

T2.19 For a random variable X resulting from an experiment with event space $B_1, \dots, B_m$, $E[X|B] = \sum_{i=1}^{m} E[X|B_i]P[B_i]$.

T2.20 The conditional expected value of $Y = g(X)$ given condition B is $E[Y|B] = E[g(X)|B] = \sum_{x \in B} g(x)P_{X|B}(x)$.

D3.1 **Cumulative Distribution Function (CDF)**: The cumulative distribution function (CDF) of random variable X is $F_X(x) = P[X \leq x]$.

T3.1 For any random variable X, (a) $F_X(-\infty) = 0$, (b) $F_X(\infty) = 1$, (c) $P[x_1 < X \leq x_2] = F_X(x_2) - F_X(x_1)$.

D3.2 **Continuous Random Variable**: X is a continuous random variable if the CDF $F_X(x)$ is a continuous function.

T3.2 For a continuous random variable X with PDF $f_X(x)$, (a) $f_X(x) \geq 0$ for all x, (b) $F_X(x) = \int_{-\infty}^{x} f_X(u)du$, (c) $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

T3.3 $P[x_1 < X \leq x_2] = \int_{x_1}^{x_2} f_X(x)dx$.

D3.4 **Expected Value**: The expected value of a continuous random variable X is $E[X] = \int_{-\infty}^{\infty} xf_X(x)dx$.

T3.4 The expected value of a function, $g(X)$, of random variable X is $E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$.

T3.5 For any random variable X, (a) $E[X - \mu_X] = 0$, (b) $E[aX + b] = aE[X] + b$, (c) $Var[X] = E[X^2] - \mu_X^2$, (d) $Var[aX + b] = a^2 Var[X]$.

D3.5 **Uniform Random Variable**: X is a uniform (a, b) random variable if the PDF of X is $f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x < b, \\ 0 & otherwise, \end{cases}$ where the two parameters are $a > b$.

T3.6 If X is a uniform (a, b) random variable, (a) The CDF of X is $F_X(x) = \begin{cases} 0 & x \leq a, \\ \frac{x-a}{b-a} & a < x \leq b, \\ 1 & x > b. \end{cases}$ (b) The expected value of X is $E[X] = \frac{b+a}{2}$. (c) The variance of X is $Var[X] = \frac{(b-a)^2}{12}$.

T3.7 Let X be a uniform (a, b) random variable, where a and b are both integers. Let $K = \lceil X \rceil$. Then K is a discrete uniform ($a + 1$, b) random variable.

D3.6 **Exponential Random Variable**: X is an exponential (λ) random variable if the PDF of X is $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & otherwise, \end{cases}$ where the parameter $\lambda > 0$.

T3.8 If X is an exponential (λ) random variable, (a) $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & otherwise, \end{cases}$ (b) $E[X] = \frac{1}{\lambda}$, (c) $Var[X] = \frac{1}{\lambda^2}$.

T3.9 If X is an exponential (λ) random variable, then $K = \lceil X \rceil$ is a geometric (p) random variable with $p = 1 - e^{-\lambda}$.

D3.7 **Erlang Random Variable**: X is an Erlang (n, λ) random variable if the PDF of X is $f_X(x) = \begin{cases} \lambda^n x^{n-1} e^{-\lambda x} & x \geq 0, \\ 0 & otherwise, \end{cases}$ where the parameter $\lambda > 0$, and the parameter $n \geq 1$ is an integer.

T3.10 If X is an Erlang (n, λ) random variable, then $E[X] = \frac{n}{\lambda}, Var[X] = \frac{n}{\lambda^2}$.

T3.11 Let K_α denote a Poisson (α) random variable. For any $x > 0$, the CDF of an Erlang (n, λ) random variable X satisfies $F_X(x) = 1 - F_{K_{\lambda x}}(n - 1) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}$.

D3.8 **Gaussian Random Variable**: X is $N[\mu, \sigma^2]$ or a Gaussian (μ, σ) random variable if the PDF of X is $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, where the parameter μ can be any real number and the parameter $\sigma > 0$.

T3.12 If X is $N[\mu, \sigma^2]$ or a Gaussian (μ, σ) random variable, then $E[X] = \mu, Var[X] = \sigma^2$.

T3.13 If X is $N[\mu, \sigma^2]$ or a Gaussian (μ, σ) random variable, $Y = aX + b$ is Gaussian ($a\mu + b, a\sigma$).

D3.9 **Standard Normal Random Variable**: The standard normal random variable Z is the Gaussian (0, 1) random variable.

D3.10 **Standard Normal CDF**: The CDF of the standard normal random variable Z is $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du$.

T3.14 If X is a Gaussian (μ, σ) random variable, the CDF of X is $F_X(\text{x}) = \Phi(\frac{x-\mu}{\sigma})$. The probability that X is in the interval $(a, b]$ is $[P[\text{a} < x \leq b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi(\frac{a-\mu}{\sigma})]$.

T3.15 $\Phi(-z) = 1 - \Phi(z)$.

D3.11 **Standard Normal Complementary CDF**: The standard normal complementary CDF is $Q(z) = P[Z > z] = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{\frac{-u^2}{2}} du = 1 - \Phi(z)$.

D3.12 **Unit Impulse (Delta) Function**: Let $d_\epsilon(x) = \begin{cases} 1/\epsilon & -\epsilon/2 \leq x \leq \epsilon/2, \\ 0 & otherwise. \end{cases}$ The unit impulse function is $\delta(x) = \lim_{\epsilon \to 0} d_\epsilon(x)$.

T3.16 **Sifting Property**: For any continuous functions $g(x)$, $\int_{-\infty}^{\infty} g(x)\delta(x - x_0)dx = g(x_0)$.

D3.13 **Unit Step Function**: The unit step function is $u(x) = \begin{cases} 0 & x < 0, \\ 1 & x \geq 0. \end{cases}$

T3.17 $\int_{-\infty}^{x} \delta(v)dv = u(x)$.

T3.18 For a random variable X, we have the following equivalent statements: (a) $P[X = x_0] = q$, (b) $P_X(x_0) = q$, (c) $F_X(x_0^+) - F_X(x_0^-) = q$, (d) $f_X(x_0) = q\delta(0)$.

D3.14 **Mixed Random Variable**: X is a mixed random variable if and only if $f_X(x)$ contains both impulse and nonzero, finite values.

T3.19 If $Y = aX$, where $a > 0$, then Y has CDF $F_Y(y) = F_X(\frac{y}{a})$, and PDF $f_Y(y) = \frac{1}{a} f_X(\frac{y}{a})$.

T3.20 $Y = aX$, where $a > 0$. (a) If X is uniform (b,c), then Y is uniform (ab, ac). (b) If X is exponential (λ), then Y is exponential ($\frac{\lambda}{a}$). (c) If X is Erlang (n, λ), then Y is Erlang ($n, \frac{\lambda}{a}$). (d) If X is Gaussian (μ, σ), then Y is Gaussian ($a\mu, a\sigma$).

T3.21 If $Y = X + b$, then $F_Y(y) = F_X(y - b)$, $f_Y(y) = f_X(y - b)$.

T3.22 Let U be a uniform (0,1) random variable and let $F(x)$ denote a cumulative distribution function with an inverse $F^{-1}(u)$ defined for $0 < u < 1$. The random variable $X = F^{-1}(U)$ has CDF $F_X(x) = F(x)$.

D3.15 **Conditional PDF given an Event**: For a random variable X with PDF $f_X(x)$ and an event $B \subset S_X$ with P[B] > 0, the conditional PDF of X given B is $f_{X|B}(x) = \begin{cases} \frac{f_X(x)}{P[B]} & x \in B, \\ 0 & otherwise. \end{cases}$

T3.23 Given an event space $\{B_i\}$ and the conditional PDFs $f_{X|B_i}(x)$, $f_X(x) = \sum_i f_{X|B_i}(x)P[B_i]$.

D3.16 **Conditional Expected Value Given an Event**: If $\{x \in B\}$, the conditional expected value of X is $E[X|B] = \int_{-\infty}^{\infty} x f_{X|B}(x)\,dx$.

D4.1 **Joint Cumulative Distribution Function (CDF)**: The joint cumulative distribution function of random variables X and Y is $F_{X,Y}(x,y) = P[X \le x, Y \le y]$.

T4.1 For any pair of random variables, X, Y, (a) $0 \le F_{X,Y}(x,y) \le 1$, (b) $F_X(x) = F_{X,Y}(x,\infty)$, (c) $F_Y(y) = F_{X,Y}(\infty,y)$, (d) $F_{X,Y}(-\infty,y) = F_{X,Y}(x,-\infty) = 0$, (e) If $x \le x_1$ and $y \le y_1$, then $F_{X,Y}(x,y) \le F_{X,Y}(x_1,y_1)$, (f) $F_{X,Y}(\infty,\infty) = 1$.

D4.2 **Joint Probability Mass Function (PMF)**: The joint probability mass function of discrete random variables X and Y is $P_{X,Y}(x,y) = P[X = x, Y = y]$.

T4.2 For discrete random variables X and Y and any set B in the X, Y plane, the probability of the event $\{(X,Y)\} \in B\}$ is $P[B] = \sum_{(x,y)\in B} P_{X,Y}(x,y)$.

T4.3 For discrete random variables X and Y with joint PMF $P_{X,Y}(x,y)$, $P_X(x) = \sum_{y\in S_Y} P_{X,Y}(x,y)$, $P_Y(y) = \sum_{x\in S_X} P_{X,Y}(x,y)$.

D4.3 **Joint Probability Density Function (PDF)**: The joint PDF of the continuous random variables X and Y is a function $f_{X,Y}(x,y)$ with the property $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$.

T4.4 $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$.

T4.5 $P[x_1 \le X \le x_2, y_1 < Y \le y_2] = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_2,y_1) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_1,y_1)$.

T4.6 A joint PDF $f_{X,Y}(x,y)$ has the following properties corresponding to first and second axioms of probability: (a) $f_{X,Y}(x,y) \ge 0$ for all (x, y), (b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$.

T4.7 The probability that the continuous random variables (X, Y) are in A is $P[A] = \iint_A f_{X,Y}(x,y)\,dx\,dy$.

T4.8 If X and Y are random variables with joint PDF $f_{X,Y}(x,y)$, $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$.

T4.9 For discrete random variables X and Y, the derived random variable $W = g(X,Y)$ has PMF $P_W(w) = \sum_{(x,y):g(x,y)=w} P_{X,Y}(x,y)$.

T4.10 For continuous random variables X and Y, the CDF of $W = g(X,Y)$ is $F_W(w) = P[W \le w] = \iint_{g(x,y)\le w} f_{X,Y}(x,y)\,dx\,dy$.

T4.11 For continuous random variables X and Y, the CDF of $W = max(X,Y)$ is $F_W(w) = F_{X,Y}(w,w) = \int_{-\infty}^{w}\int_{-\infty}^{w} f_{X,Y}(x,y)\,dx\,dy$.

T4.12 For random variables X and Y, the expected value of $W = g(X,Y)$ is (a) Discrete: $E[W] = \sum_{x\in S_X}\sum_{y\in S_Y} g(x,y)P_{X,Y}(x,y)$, (b) Continuous: $E[W] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$.

T4.13 $E[g_1(X,Y) + \cdots + g_n(X,Y)] = E[g_1(X,Y)] + \cdots + E[g_n(X,Y)]$.

T4.14 For any two random variables X and Y, $E[X+Y] = E[X] + E[Y]$.

T4.15 The variance of the sum of two random variables is $Var[X+Y] = Var[X] + Var[Y] + 2E[(X-\mu_X)(Y-\mu_Y)]$.

D4.4 **Covariance**: The covariance of two random variables X and Y is $Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$.

D4.5 **Correlation**: The correlation of X and Y is $r_{X,Y} = E[XY]$.

T4.16 (a) $Cov[X,Y] = r_{X,Y} - \mu_X\mu_Y$. (b) $Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$. (c) If $X = Y$, $Cov[X,Y] = Var[X] = Var[Y]$ and $r_{X,Y} = E[X^2] = E[Y^2]$.

D4.6 **Orthogonal Random Variables**: Random variables X and Y are orthogonal if $r_{X,Y} = 0$.

D4.7 **Uncorrelated Random Variables**: Random variables X and Y are uncorrelated if $Cov[X,Y] = 0$. Orthogonal means zero correlation; uncorrelated means zero covariance.

D4.8 **Correlation Coefficient**: The correlation coefficient of two random variables X and Y is $\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} = \frac{Cov[X,Y]}{\sigma_X\sigma_Y}$.

T4.17 $-1 \le \rho_{X,Y} \le 1$.

T4.18 If X and Y are random variables such that $Y = aX + b$, $\rho_{X,Y} = \begin{cases} -1 & a < 0, \\ 0 & a = 0, \\ 1 & a > 0. \end{cases}$

D4.9 **Conditional Joint PMF**: For discrete random variables X and Y and an event, B with $P[B] > 0$, the conditional joint PMF of X and Y given B is $P_{X,Y|B}(x,y) = P[X = x, Y = y|B]$.

T4.19 For any event B, a region of the X, Y plane with $P[B] > 0$, $P_{X,Y|B}(x,y) = \begin{cases} \frac{P_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & otherwise. \end{cases}$

D4.10 **Conditional Joint PDF**: Given an event B with $P[B] > 0$, the conditional joint probability density function of X and Y is $f_{X,Y|B}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & otherwise. \end{cases}$

T4.20 **Conditional Expected Value**: For random variables X and Y and an event B of nonzero probability, the conditional expected value of $W = g(X,Y)$ given B is (a) Discrete: $E[W|B] = \sum_{x\in S_X}\sum_{y\in S_Y} g(x,y)P_{X,Y|B}(x,y)$, (b) Continuous: $E[W|B] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y|B}(x,y)\,dx\,dy$.

D4.11 **Conditional variance**: The conditional variance of the random variable $W = g(X,Y)$ is $Var[W|B] = E[(W - \mu_{W|B})^2|B]$. Another notation for conditional variance is $\sigma^2_{W|B}$.

T4.21 $Var[W|B] = E[W^2|B] - (\mu_{W|B})^2$.

D4.12 **Conditional PMF**: For any event $Y = y$ such that $P_Y(y) > 0$, the conditional PMF of X given $Y = y$ is $P_{X|Y}(x|y) = P[X = x|Y = y]$.

T4.22 For random variables X and Y with joint PMF $P_{X,Y}(x,y)$, and x and y such that $P_X(x) > 0$ and $P_Y(y) > 0$, $P_{X,Y}(x,y) = P_{X|Y}(x|y)P_Y(y) = P_{Y|X}(y|x)P_X(x)$.

T4.23 **Conditional Expected Value of a Function**: X and Y are discrete random variables. For any $y \in S_Y$, the conditional expected value of $g(X,Y)$ given $Y = y$ is $E[g(X,Y)|Y = y] = \sum_{x\in S_X} g(x,y)P_{X|Y}(x|y)$.

D4.13 **Conditional PDF**: For y such that $f_Y(y) > 0$, the conditional PDF of X given $\{Y = y\}$ is $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, which implies $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.

T4.24 $f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y)$.

D4.14 **Conditional Expected Value of a Function**: For continuous random variables X and Y and any y such that $f_Y(y) > 0$, the conditional expected value of $g(X,Y)$ given $Y = y$ is $E[g(X,Y)|Y = y] = \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\,dx$.

D4.15 **Conditional Expected Value**: The conditional expected value $E[X|Y]$ is a function of random variable Y such that is $Y = y$ then $E[X|Y] = E[X|Y = y]$.

T4.25 **Iterated Expectation**: $E[E[X|Y]] = E[X]$.

T4.26 $E[E[g(X)|Y]] = E[g(X)]$.

D4.16 **Independent Random Variables**: Random variables X and Y are independent if and only if (a) Discrete: $P_{X,Y}(x,y) = P_X(x)P_Y(y)$, (b) Continuous: $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

T4.27 For independent variables X and Y, (a) $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$, (b) $r_{X,Y} = E[XY] = E[X]E[Y]$, (c) $Cov[X,Y] = \rho_{X,Y} = 0$, (d) $Var[X+Y] = Var[X] + Var[Y]$, (e) $E[X|Y = y] = E[X]$ for all $y \in S_Y$, (f) $E[Y|X = x] = E[Y]$ for all $x \in S_X$.

D4.17 **Bivariate Gaussian Random Variables**: Random variables X and Y have a bivariate Gaussian PDF with parameters $\mu_1, \sigma_1, \mu_2, \sigma_2$, and ρ if $f_{X,Y}(x,y) = \frac{\exp\left[-\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$, where μ_1 and μ_2 can be any real numbers, $\sigma_1 > 0, \sigma_2 > 0$, and $-1 < p < 1$.

T4.28 If X and Y are bivariate Gaussian random variables in Definition 4.17, X is the Gaussian (μ_1, σ_1) random variable and Y is the Gaussian (μ_2, σ_2) random variable: (a) $f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}$, (b) $f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{\frac{-(y-\mu_2)^2}{2\sigma_2^2}}$.

T4.29 If X and Y are bivariate Gaussian random variables in Definition 4.17, the conditional PDF of Y given X is $f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{\frac{-(y-\tilde{\mu}_2(x))^2}{2\tilde{\sigma}_2^2}}$, where, given $X = x$, the conditional expected value of Y is $\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$, and the variance of Y is $\tilde{\sigma}_2^2 = \sigma_2^2(1-\rho^2)$.

T4.30 If X and Y are bivariate Gaussian random variables in Definition 4.17, the conditional PDF of X given Y is $f_{X|Y}(x|y) = \frac{1}{\tilde{\sigma}_1\sqrt{2\pi}} e^{\frac{-(x-\tilde{\mu}_1(y))^2}{2\tilde{\sigma}_1^2}}$, where, given $Y = y$, the conditional expected value of X is $\tilde{\mu}_1(y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$, and the variance of X is $\tilde{\sigma}_1^2 = \sigma_1^2(1-\rho^2)$.

T4.31 Bivariate Gaussian random variables X and Y in Definition 4.17 have correlation coefficient $\rho_{X,Y} = \rho$.

T4.32 Bivariate Gaussian random variables X and Y are uncorrelated if and only if they are independent.

D5.1 **Multivariate Joint CDF**: The joint CDF of $X_1, \ldots, X_n$ is $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P[X_1 \le x_1, \ldots, X_n \le x_n]$.

D5.2 **Multivariate Joint PMF**: The joint PMF of the discrete random variables $X_1, \ldots, X_n$ is $P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P[X_1 = x_1, \ldots, X_n = x_n]$.

D5.3 **Multivariate Joint PDF**: The joint PDF of the continuous random variables $X_1, \dots, X_n$ is the function $f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \frac{\partial^n F_{X_1,\dots,X_n}(x_1,\dots,x_n)}{\partial x_1 \dots \partial x_n}$.

T5.1 If $X_1, \dots, X_n$ are discrete random variables with joint PMF $P_{X_1,\dots,X_n}(x_1,\dots,x_n)$, (a) $P_{X_1,\dots,X_n}(x_1,\dots,x_n) \geq 0$, (b) $\sum_{x_1 \in S_{X_1}} \cdots \sum_{x_n \in S_{X_n}} P_{X_1,\dots,X_n}(x_1,\dots,x_n) = 1$.

T5.2 If $X_1, \dots, X_n$ are continuous random variables with joint PDF $f_{X_1,\dots,X_n}(x_1,\dots,x_n)$, (a) $f_{X_1,\dots,X_n}(x_1,\dots,x_n) \geq 0$, (b) $F_{X_1,\dots,X_n}(x_1,\dots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,\dots,X_n}(u_1,\dots,u_n)du_1 \cdots du_n$, (c) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\dots,X_n}(x_1,\dots,x_n)dx_1 \cdots dx_n = 1$.

T5.3 The probability of an event A expressed in terms of the random variables $X_1, \dots, X_n$ is (a) Discrete:$P[A] = \sum_{(x_1,\dots,x_n)\in A} P_{X_1,\dots,X_n}(x_1,\dots,x_n)$, (b) Continuous: $P[A] = \int_A \cdots \int f_{X_1,\dots,X_n}(x_1,\dots,x_n)dx_1 dx_2 \dots dx_n$.

D5.4 **Random Vector**: A random vector is a column vector $\boldsymbol{X} = [X_1 \ \cdots X_n]'$. Each X_i is a random variable.

D5.5 **Vector Sample Value**: A sample value of a random is a column vector $\boldsymbol{x} = [x_1 \ \cdots x_n]'$. The ith component, x_i, of the vector **x** is a sample value of a random variable, X_i.

D5.6 **Random Vector Probability Functions**: (a) The **CDF of a random vector** $\boldsymbol{X}$ is $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1,\dots,X_n}(x_1,\dots,x_n)$. (b) The **PMF of a discrete random vector** $\boldsymbol{X}$ is $P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{X_1,\dots,X_n}(x_1,\dots,x_n)$. (c) The **PDF of a continuous random vector X** is $f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n)$.

D5.7 **Probability Functions of a Pair of Random Vectors**: For random vectors $\boldsymbol{X}$ with n components and $\boldsymbol{Y}$ with m components: (a) The **joint CDF** of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $F_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = F_{X_1,\dots,X_n,Y_1,\dots,Y_m}(x_1,\dots,x_n,y_1,\dots,y_m)$; (b) The **joint PMF** of discrete random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = P_{X_1,\dots,X_n,Y_1,\dots,Y_m}(x_1,\dots,x_n,y_1,\dots,y_m)$; (c) The **joint PDF** of continuous random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{X_1,\dots,X_n,Y_1,\dots,Y_m}(x_1,\dots,x_n,y_1,\dots,y_m)$.

T5.4 For a joint PMF $P_{W,X,Y,Z}(w,x,y,z)$ of discrete random variables W, X, Y, Z, some marginal PMFs are: (a) $P_{X,Y,Z}(x,y,z) = \sum_{w\in S_W} P_{W,X,Y,Z}(w,x,y,z)$, (b) $P_{W,Z}(w,z) = \sum_{x\in S_X}\sum_{y\in S_Y} P_{W,X,Y,Z}(w,x,y,z)$, (c) $P_X(x) = \sum_{w\in S_W}\sum_{y\in S_Y}\sum_{z\in S_Z} P_{W,X,Y,Z}(w,x,y,z)$.

T5.5 For a joint PDF $f_{W,X,Y,Z}(w,x,y,z)$ of continuous random varaibles W, X, Y, Z, some marginal PDFs are: (a) $f_{X,Y,Z}(x,y,z) = \int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)dw$, (b) $f_{W,Z}(w,z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)dxdy$, (c) $f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{W,X,Y,Z}(w,x,y,z)dwdydz$.

D5.8 **N Independent Random Variables**: Random Variables $X_1, \dots, X_n$ are independent if for all $x_1, \dots, x_n$, (a) Discrete: $P_{X_1,\dots,X_n}(x_1,\dots,x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_N}(x_n)$, (b) Continuous: $f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_N}(x_n)$.

D5.9 **Independent and Identically Distributed (iid)**: Random variables $X_1, \dots, X_n$ are independent and identically distributed (iid) if (a) Discrete: $P_{X_1,\dots,X_n}(x_1,\dots,x_n) = P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_N}(x_n)$, (b) Continuous: $f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_N}(x_n)$.

D5.10 **Independent Random Vectors**: Random Vectors X and Y are independent if (a) Discrete: $P_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{X}}(\boldsymbol{x})P_Y(\boldsymbol{y})$, (b) Continuous: $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})f_Y(\boldsymbol{y})$.

T5.6 For random variable $W - g(\boldsymbol{X})$, (a) Discrete: $P_W(w) = P[W = w] = \sum_{x:g(x)=w} P_{\boldsymbol{X}}(\boldsymbol{x})$, (b) Continuous: $F_W(w) = P[W = w] = \int_{g(x)\leq w} \cdots \int f_X(x)dx_1 \cdots dx_n$.

T5.7 Let $\boldsymbol{X}$ be a vector of n iid random variables each with CDF $F_X(x)$ and PDF $f_X(x)$. (a) The CDF and the PDF of $Y = \max\{X_1, \dots, X_n\}$ are $F_Y(y) = (F_X(y))^n, f_Y(y) = n(F_X(y))^{n-1} f_X(y)$. (b) The CDF and the PDF of $W = min\{X_1, \dots, X_n\}$ are $F_W(w) = 1 - (1 - F_X(w))^n, f_W(w) = n(1 - F_X(w))^{n-1} f_X(w)$.

T5.8 For a random vector $\boldsymbol{X}$, the random variables $g(\boldsymbol{X})$ has expected value (a) Discrete: $E[g(\boldsymbol{X})] = \sum_{x_1\in S_{X_1}} \cdots \sum_{x_n\in S_{X_n}} g(\boldsymbol{X})P_{\boldsymbol{X}}(\boldsymbol{x})$, (b) Continuous: $E[g(\boldsymbol{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\boldsymbol{X})f_{\boldsymbol{X}}(\boldsymbol{x})dx_1 \cdots dx_n$.

T5.9 When the components of $\boldsymbol{X}$ are independent random variables, $E[g_1(X_1)g_2(X_2)\cdots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)]\cdots E[g_n(X_n)]$.

T5.10 Given the continuous random vector $\boldsymbol{X}$, define the derived random vector $\boldsymbol{Y}$ such that $Y_k = aX_k + b$ for constants $a > 0$ and b. The CDF and PDF of $\boldsymbol{Y}$ are $F_{\boldsymbol{Y}}(\boldsymbol{y}) = F_{\boldsymbol{X}}(\frac{y_1 - b}{a}, \dots, \frac{y_n - b}{a}), f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{a^n} f_{\boldsymbol{X}}(\frac{y_1 - b}{a}, \dots, \frac{y_n - b}{a})$.

T5.11 If $\boldsymbol{X}$ is a continuous random vector and $\boldsymbol{A}$ is an invertible matrix, then $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ has PDF $f_Y(y) = \frac{1}{|det(\boldsymbol{A})|} f_{\boldsymbol{X}}(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}))$.

D5.11 **Expected Value Vector**: The expected value of a random vector X is a column vector $E[\boldsymbol{X}] = \boldsymbol{\mu_X} = [E[X_1] \ E[X_2] \ \cdots \ E[X_n]]'$.

D5.12 **Expected Value of a Random Matrix**: For a random matrix **A** with the random variable A_{ij} as its i, jth element, $E[A]$ is a matrix with i, jth element $E[A_{ij}]$.

D5.13 **Vector Correlation**: The correlation of a random vector $\boldsymbol{X}$ is an $n \times n$ matrix $\boldsymbol{R_X}$ with i, jth element $R_X(i,j) = E[X_i X_j]$. In vector notation, $\boldsymbol{R_X} = E[\boldsymbol{XX'}]$.

D5.14 **Vector Covariance**: The covariance of a random vector $\boldsymbol{X}$ is an $n \times n$ matrix $\boldsymbol{C_X}$ with components $C_X(i,j) = Cov[X_i, X_j]$. In vector notation, $\boldsymbol{C_X} = E[(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{X} - \boldsymbol{\mu_X})']$.

T5.12 For a random vector $\boldsymbol{X}$ with correlation matrix $\boldsymbol{R_X}$, covariance matrix $\boldsymbol{C_X}$, and vector expected value $\boldsymbol{\mu_X}$, $\boldsymbol{C_X} = \boldsymbol{R_X} - \boldsymbol{\mu_X}\boldsymbol{\mu_X}'$.

D5.15 **Vector Cross-Correlation**: The cross-correlation of random vectors, $\boldsymbol{X}$ with n components and $\boldsymbol{Y}$ with m components, is an $n \times m$ matrix $\boldsymbol{R_{XY}}$ with i, jth element $R_{XY}(i,j) = E[X_i Y_j]$, or, in vector notation, $\boldsymbol{R_{XY}} = E[\boldsymbol{XY'}]$.

D5.16 **Vector Cross-Covariance**: The cross-covariance of a pair of random vectors $\boldsymbol{X}$ with n components and $\boldsymbol{Y}$ with m components is an $n \times m$ matrix $\boldsymbol{C_{XY}}$ with i, jth element $C_{XY}(i,j) = Cov[X_i, Y_j]$, or, in vector notation, $\boldsymbol{C_{XY}} = E[(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{Y} - \boldsymbol{\mu_Y})']$.

T5.13 $\boldsymbol{X}$ is an n-dimensional random vector with expected value $\boldsymbol{\mu_X}$, correlation $\boldsymbol{R_X}$, and covariance $\boldsymbol{C_X}$. The m-dimensional random vector $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$, where $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{b}$ is an m-dimensional vector, has expected value $\boldsymbol{\mu_Y}$, correlation matrix $\boldsymbol{R_Y}$, and covariance matrix $\boldsymbol{C_Y}$ given by: (a) $\boldsymbol{\mu_Y} = \boldsymbol{A\mu_X} + \boldsymbol{b}$, (b) $\boldsymbol{R_Y} = \boldsymbol{AR_XA'} + (\boldsymbol{A\mu_X})\boldsymbol{b}' + \boldsymbol{b}(\boldsymbol{A\mu_X})' + \boldsymbol{bb}'$, (c) $\boldsymbol{C_Y} = \boldsymbol{AC_XA'}$.

T5.14 The vectors $\boldsymbol{X}$ and $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ have cross-correlation $\boldsymbol{R_{XY}}$ given by $\boldsymbol{R_{XY}} = \boldsymbol{R_XA'} + \boldsymbol{\mu_Xb'}$, and cross-covariance $\boldsymbol{C_{XY}}$ given by $\boldsymbol{C_{XY}} = \boldsymbol{C_XA'}$.

D5.17 **Gaussian Random Vector**: $\boldsymbol{X}$ is a Gaussian $(\boldsymbol{\mu_X}, \boldsymbol{C_X})$ random vector with expected value $\boldsymbol{\mu_X}$ and covariance $\boldsymbol{C_X}$ if and only if $f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{C_X})]^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_X})'\boldsymbol{C_X}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_X})\right)$, where $\det(\boldsymbol{C_X})$, the determinant of $\boldsymbol{C_X}$, satisfies $\det(\boldsymbol{C_X}) > 0$.

T5.15 A Gaussian random vector $\boldsymbol{X}$ has independent components if and only if $\boldsymbol{C_X}$ is a diagonal matrix.

T5.16 Given an n-dimensional Gaussian random vector $\boldsymbol{X}$ with expected value $\boldsymbol{\mu_X}$ and covariance $\boldsymbol{C_X}$, and an $m \times n$ matrix $\boldsymbol{A}$ with $rank(\boldsymbol{A}) = m$, $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ is an m-dimensional Gaussian random vector with expected value $\boldsymbol{\mu_Y} = \boldsymbol{A\mu_X} + \boldsymbol{b}$ and covariance $\boldsymbol{C_Y} = \boldsymbol{AC_XA'}$.

D5.18 **Standard Normal Random Vector**: The n-dimensional standard normal random vector $\boldsymbol{Z}$ is the n-dimensional Gaussian random vector with $E[\boldsymbol{Z}] = \boldsymbol{0}$ and $\boldsymbol{C_Z} = \boldsymbol{I}$.

T5.17 For a Gaussian $(\boldsymbol{\mu_X}, \boldsymbol{C_X})$ and random vector, let $\boldsymbol{A}$ be an $n \times n$ matrix with the property $\boldsymbol{AA'} = \boldsymbol{C_X}$. The random vector $\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{X} - \boldsymbol{\mu_X})$ is a standard normal random vector.

T5.18 Given the n-dimensional standard normal random vector $\boldsymbol{Z}$, an invertible $n \times n$ matrix $\boldsymbol{A}$, and an n-dimensional vector $\boldsymbol{b}$, $\boldsymbol{X} = \boldsymbol{AZ} + \boldsymbol{b}$ is an n-dimensional Gaussian random vector with expected value $\boldsymbol{\mu_X} = \boldsymbol{b}$ and covariance matrix $\boldsymbol{C_X} = \boldsymbol{AA'}$.

T5.19 For a Gaussian vector $\boldsymbol{X}$ with covariance $\boldsymbol{C_X}$, there always exists a matrix $\boldsymbol{A}$ such that $\boldsymbol{C_X} = \boldsymbol{AA'}$.

T9.4 Random variables X and Y have expected values μ_X and μ_Y, standard deviations σ_X and σ_Y, and correlation coefficient $\rho_{X,Y}$. The optimal linear mean square error (LMSE) estimator of X given Y is $\hat{X}_L(Y) = a^*Y + b^*$ and it has the following properties: (a) $a^* = \frac{Cov[X,Y]}{Var[Y]} = \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}, b^* = \mu_X - a^*\mu_Y$, (b) The minimum mean square estimation error for a linear estimate is $e_L^* = E[(X - \hat{X}_L(Y))^2] = \sigma_X^2(1 - \rho_{X,Y}^2)$, (c) The estimation error $X - \hat{X}_L(Y)$ is uncorrelated with Y.

Math Facts

B.0 **Devroye's Special Topics' Formulas:** (a) $\int_0^{\infty} y^n e^{-y}dy = n$; (b) $SNR = \frac{E[X^2]}{E[(X-Y)^2]} = \frac{E[X^2]}{E[N^2]} = \frac{A^2}{\sigma_N^2}, SNR_{dB} = 10log_{10}\left(\frac{E[X^2]}{E[(X-Y)^2]}\right) = 10log_{10}\left(\frac{E[X^2]}{E[N^2]}\right) = 10log_{10}\left(\frac{A^2}{\sigma_N^2}\right) = 20log_{10}\left(\frac{A}{\sigma_N}\right)$, where Y is the received signal, $Y = X + N, X$ is the transmit samples, A is the Bernoulli bit sent, and N is the Gaussian noise.

B.1 **Half Angle Formulas**: (a) $cos(A + B) = cosAcosB - sinAsinB$, (b) $sin(A + B) = sinAcosB + cosAsinB$, (c) $cos(2A) = cos2A - sin2A = 2cos^2A - 1 = 1 - 2sin^2A$, (d) $sin(2A) = 2sinAcosA$.

B.2 **Products of Sinusoids**: (a) $sinAsinB = \frac{1}{2}[cos(A - B) - cos(A + B)]$, (b) $cosAcosB = \frac{1}{2}[cos(A - B) \ cos(A + B)]$, (c) $sinAcosB = \frac{1}{2}[sin(A + B) + sin(A - B)]$.

B.3 **The Euler Formula**: The Euler formula $e^{j\theta} = cos\theta + jsin\theta$ is the source of the identities (a) $cos\theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$, (b) $sin\theta = \frac{e^{j\theta} - e^{-j\theta}}{2}$.

B.4 **Finite Geometric Series**: The finite geometric series is $\sum_{i=0}^{n} q^i = 1 + q + q^2 + \cdots + q^n = \frac{1 - q^{n+1}}{1 - q}$.

B.5 **Infinite Geometric Series**: When $|q| < 1, \sum_{i=0}^{\infty} q^i = \lim_{n\to\infty} \sum_{i=0}^{n} q^i = \frac{1}{1-q}$.

B.6 $\sum_{i=1}^{n} iq^i = \frac{q(1 - q^n[1 + n(1 - q)])}{(1 - q)^2}$.

B.7 If $|q| < 1, \sum_{i=1}^{\infty} iq^i = \frac{q}{(1-q)^2}$.

B.8 $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$.

B.9 $\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$.

B.10 **Integration by Parts**: The integration by parts formula is $\int_a^b udv = uv|_a^b - \int_a^b vdu$.