

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

CS723

Probability and Stochastic Processes

CS723 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

COMPREHENSIVE STUDY HANDOUT

CS723

Probability and Stochastic Processes

Lecture-wise study notes prepared for Virtual University of Pakistan students.

Covers all major topics with objectives, key definitions, summaries and exam pointers.



13 study units

Table of Contents

Placeholder for table of contents

0

Unit 1: Probability Foundations: Sample Spaces and Axioms

Learning Objectives

- Objective 1:** Define sample spaces, events and probability axioms
- Objective 2:** Apply counting, conditional probability and Bayes' theorem
- Objective 3:** Distinguish independence from mutual exclusivity

Probability theory begins with a sample space S , the set of all possible outcomes of an experiment, and events, which are subsets of S . Kolmogorov's three axioms ground everything: probabilities are non-negative, the probability of the whole sample space is 1, and the probability of a countable union of disjoint events is the sum of their probabilities. Every rule you will ever use, from the addition rule to Bayes' theorem, is a consequence of these axioms.

Counting matters because many probabilities are ratios of favorable to total outcomes. Permutations count ordered selections, combinations unordered ones; the binomial coefficient nCr appears everywhere from coin tosses to packet arrivals. Conditional probability $P(A|B) = P(A \text{ and } B)/P(B)$ captures how information updates belief, and Bayes' theorem inverts conditioning: $P(B|A)$ lets us reason from evidence back to causes, which is the engine of medical testing and spam filtering alike.

Two confusions must never occur. Mutually exclusive events cannot happen together ($P(A \text{ and } B) = 0$); independent events do not influence each other ($P(A \text{ and } B) = P(A)P(B)$). A single coin toss's heads and tails are mutually exclusive but not independent; two separate tosses are independent but not mutually exclusive. Very important: exams deliberately probe this distinction, so state definitions precisely before computing.

Key Definitions

Sample space is defined as the set of all possible outcomes of a random experiment.

Conditional probability is defined as $P(A|B)$, the probability of A given that B has occurred, equal to $P(A \cap B)/P(B)$.

Bayes' theorem is defined as $P(B|A) = P(A|B)P(B)/P(A)$, which reverses the direction of conditioning.

Key Points to Remember

- Kolmogorov axioms: non-negativity, $P(S)=1$, countable additivity for disjoint events.
- Counting: permutations (ordered), combinations (unordered); nCr is ubiquitous.
- Bayes' theorem reasons from evidence to causes; powers testing and filtering.
- Mutually exclusive = cannot co-occur; independent = no influence. Never confuse them.

Unit Summary

Probability rests on sample spaces and Kolmogorov's axioms; counting, conditioning and Bayes' theorem build on them, with independence carefully distinguished from mutual exclusivity.

Exam Pointer

Very important: Bayes' theorem numericals (e.g. disease testing with false positives) are classic. Also expect a question forcing you to distinguish independence from mutual exclusivity with examples.

Unit 2: Random Variables and Distributions

Learning Objectives

Objective 1: Define discrete and continuous random variables

Objective 2: Work with PMF, PDF and CDF

Objective 3: Apply Bernoulli, binomial, geometric and Poisson models

A random variable X is a function mapping outcomes to real numbers, letting us do arithmetic with randomness. Discrete random variables take countable values and are described by a probability mass function $p(x) = P(X = x)$; continuous ones take uncountably many values and are described by a probability density function $f(x)$, where probabilities are areas under the curve. The cumulative distribution function $F(x) = P(X \leq x)$ unifies both, and is non-decreasing from 0 to 1.

Four discrete distributions model most counting situations. Bernoulli(p) is a single trial with success probability p . Binomial(n, p) counts successes in n independent trials. Geometric(p) counts trials until the first success, and is memoryless: past failures do not change future odds. Poisson(λ) counts events in a fixed interval when events occur at constant rate independently; it is the workhorse for arrivals in queueing and networks.

Choosing the right model is a modeling skill: fixed trials with success counts suggests binomial; waiting time for the first success suggests geometric; events per unit time suggests Poisson. Very important: the Poisson arises as the limit of the binomial when n is large and p is small with np fixed, a fact exams love to test both as derivation and as approximation guidance.

Key Definitions

Random variable is defined as a function assigning a real number to each outcome of a sample space.

Probability mass function is defined as $p(x) = P(X = x)$ for a discrete random variable.

Probability density function is defined as $f(x)$ for a continuous variable, with $P(a < X < b)$ as the area under f between a and b .

Key Points to Remember

- Discrete: PMF; continuous: PDF (probabilities are areas); CDF $F(x) = P(X \leq x)$ unifies both.
- Bernoulli: one trial; binomial: successes in n trials; geometric: trials until first success (memoryless); Poisson: events per interval.
- Model choice follows the story: fixed trials, waiting time, or rate over time.
- Poisson is the large- n , small- p limit of the binomial.

Unit Summary

Random variables are described by PMF/PDF/CDF; Bernoulli, binomial, geometric and Poisson models cover most discrete counting scenarios, chosen by the problem's story.

Exam Pointer

Very important: be fluent in identifying which distribution fits a word problem, and know the Poisson-as-binomial-limit result. CDF properties are short-question staples.

Unit 3: Continuous Distributions: Uniform, Exponential, Normal

Learning Objectives

Objective 1: Use uniform, exponential and normal distributions

Objective 2: Exploit the memoryless property of the exponential

Objective 3: Standardize normal variables and use Z-tables

Three continuous distributions dominate applications. The uniform(a, b) spreads probability evenly, modeling complete ignorance within bounds and serving as the source for simulation via inverse-transform sampling. The exponential(λ) models waiting times between Poisson events with density $\lambda e^{-\lambda x}$; its hallmark is memorylessness, $P(X > s+t \mid X > s) = P(X > t)$, meaning the wait so far tells you nothing about the remaining wait. Only the exponential (continuous) and geometric (discrete) have this property.

The normal (Gaussian) distribution is the most important in all of statistics. Its bell curve is parameterized by mean μ and variance σ^2 , and the central limit theorem explains its ubiquity: sums of many small independent effects are approximately normal. Standardization $Z = (X - \mu)/\sigma$ converts any normal to the standard normal, whose probabilities come from Z-tables or software. The empirical rule (68-95-99.7%) gives quick estimates.

In stochastic processes these distributions are the building blocks: exponential inter-arrival times define the Poisson process, and normal increments define Brownian motion. Very important: when an exam says 'memoryless continuous waiting time', it is telling you to use the exponential; recognize the cue rather than deriving it each time.

Key Definitions

Memoryless property is defined as $P(X > s+t \mid X > s) = P(X > t)$: the remaining wait does not depend on time already waited.

Standard normal is defined as the normal distribution with mean 0 and variance 1, denoted Z .

Central limit theorem is defined as the sum (or mean) of many independent variables is approximately normal.

Key Points to Remember

- Uniform: equal likelihood; basis of random-number generation.
- Exponential: waiting times; uniquely memoryless among continuous distributions.
- Normal: parameterized by mean and variance; CLT explains its ubiquity.
- Standardize with $Z = (X - \mu)/\sigma$; use the 68-95-99.7 rule for estimates.

Unit Summary

Uniform, exponential (memoryless waiting times) and normal (CLT-backed) are the core continuous models, feeding directly into Poisson processes and Brownian motion.

Exam Pointer

Very important: memoryless-property problems and normal standardization numericals are exam certainties. Memorize the exponential density and the Z transformation.

Unit 4: Expectation, Variance and Joint Distributions

Learning Objectives

Objective 1: Compute expectation and variance for discrete and continuous variables

Objective 2: Use joint, marginal and conditional distributions

Objective 3: Apply covariance, correlation and conditional expectation

Expectation $E[X]$ is the probability-weighted average, the long-run center of a random variable. For discrete X it is the sum of $x \cdot p(x)$; for continuous, the integral of $x \cdot f(x)$. Variance $\text{Var}(X) = E[(X - \mu)^2] = E[X^2] - \mu^2$ measures spread. Linearity of expectation, $E[aX + bY] = aE[X] + bE[Y]$, holds always, even without independence, which makes it the most powerful computational shortcut in probability.

Multiple random variables are described jointly: $p(x,y)$ or $f(x,y)$ gives probabilities of pairs, marginals recover single-variable distributions by summing or integrating out the other variable, and conditionals describe one variable given the other. Independence of X and Y means the joint factors: $p(x,y) = p(x)p(y)$. Covariance $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$ measures linear association, and correlation normalizes it to $[-1,1]$.

Conditional expectation $E[Y|X]$ is itself a random variable (a function of X), and the tower property $E[E[Y|X]] = E[Y]$ lets us compute expectations by conditioning on useful information. This technique, conditioning on the first step, solves many stochastic-process problems elegantly. Very important: correlation measures linear relationship only; uncorrelated does not imply independent (though independent always implies uncorrelated).

Key Definitions

Expectation is defined as $E[X]$, the probability-weighted average value of a random variable.

Variance is defined as $\text{Var}(X) = E[(X - \mu)^2]$, measuring spread around the mean.

Covariance is defined as $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$, measuring linear co-movement of two variables.

Key Points to Remember

- Linearity of expectation needs no independence: $E[aX + bY] = aE[X] + bE[Y]$.
- Joint $\rightarrow$ marginals by summing/integrating out; independence means the joint factors.
- Correlation in $[-1,1]$ measures linear association only.
- Tower property $E[E[Y|X]] = E[Y]$; condition on the first step to simplify.

Unit Summary

Expectation and variance summarize location and spread; joint distributions, covariance and conditional expectation handle multiple variables, with linearity and the tower property as key tools.

Exam Pointer

Very important: linearity-of-expectation tricks and 'uncorrelated does not imply independent' (with a counterexample) are favourite exam items.

Unit 5: Moment Generating Functions and Limit Theorems

Learning Objectives

Objective 1: Use MGFs to identify distributions and compute moments

Objective 2: State and apply the weak law of large numbers

Objective 3: State and apply the central limit theorem

The moment generating function $M_X(t) = E[e^{tX}]$ encodes all moments: the k -th derivative at zero gives $E[X^k]$. Its real power is identification: if two variables share an MGF in a neighborhood of zero, they share the distribution. MGFs also trivialize sums of independent variables, since $M_{\{X+Y\}}(t) = M_X(t)M_Y(t)$, which is how we prove that sums of independent normals or Poissons stay in the family.

The weak law of large numbers says the sample mean converges in probability to the true mean as n grows: averages stabilize. This is the mathematical license for Monte Carlo simulation and for trusting empirical frequencies. Chebyshev's inequality, $P(|X - \mu| \geq k\sigma) \leq 1/k^2$, gives a universal (if loose) bound behind the proof.

The central limit theorem says that the standardized sum of iid variables with finite variance converges to the standard normal, regardless of the original distribution. It justifies normal approximations for binomial and Poisson probabilities, confidence intervals, and the normal noise models of engineering. Very important: CLT needs independence and finite variance; it describes the sum's distribution, not the individual variables becoming normal.

Key Definitions

Moment generating function is defined as $M_X(t) = E[e^{tX}]$; derivatives at zero yield moments and it identifies distributions.

Weak law of large numbers is defined as the sample mean converges in probability to the population mean.

Central limit theorem is defined as standardized sums of iid finite-variance variables converge in distribution to $N(0,1)$.

Key Points to Remember

- MGF: differentiate at 0 for moments; MGF of independent sum = product of MGFs.
- WLLN: averages stabilize; proved via Chebyshev's inequality.
- CLT: sums become normal; needs iid + finite variance.

- CLT justifies normal approximations and confidence intervals.

Unit Summary

MGFs encode distributions and simplify sums; the law of large numbers stabilizes averages; the CLT makes sums approximately normal.

Exam Pointer

Very important: use MGFs to prove sums of Poissons are Poisson, and apply CLT to approximate binomial probabilities with continuity correction.

Unit 6: Introduction to Stochastic Processes

Learning Objectives

Objective 1: Define stochastic processes and classify them

Objective 2: Explain state spaces, index sets and sample paths

Objective 3: Distinguish stationary and independent increments

A stochastic process is a collection of random variables $\{X_t\}$ indexed by time t , modeling systems that evolve randomly: stock prices, queue lengths, network traffic. Classification runs along two axes: the index set (discrete-time $t = 0, 1, 2, \dots$ vs continuous-time $t \geq 0$) and the state space (discrete states like counts vs continuous values like prices). A Bernoulli process and a Poisson process are discrete-state; Brownian motion is continuous-state.

Each fixed outcome ω yields a sample path, one possible trajectory of the system; the process is the ensemble of all paths weighted by probability. Finite-dimensional distributions (the joint law of $X_{\{t_1\}}, \dots, X_{\{t_n\}}$) characterize the process. Two structural properties recur: stationary increments (the distribution of $X_{\{t+s\}} - X_t$ depends only on s) and independent increments (non-overlapping increments are independent).

These properties define the great families: the Poisson process has stationary independent increments with Poisson counts; Brownian motion has stationary independent normal increments.

Very important: when classifying a new process on an exam, always state both axes (time and state) and check the increment properties; that disciplined checklist prevents most errors.

Key Definitions

Stochastic process is defined as a family of random variables $\{X_t\}$ indexed by time, modeling random evolution.

Sample path is defined as the trajectory $t \rightarrow X_t(\omega)$ for a single fixed outcome.

Stationary increments is defined as the law of $X_{\{t+s\}} - X_t$ depends only on the lag s , not on t .

Key Points to Remember

- Classify by index set (discrete/continuous time) x state space (discrete/continuous).
- Sample path = one realized trajectory; process = ensemble of paths.
- Key properties: stationary increments, independent increments.

- Poisson process and Brownian motion are defined by their increment laws.

Unit Summary

Stochastic processes are time-indexed random variables classified by time and state type; sample paths, finite-dimensional distributions and increment properties are the core vocabulary.

Exam Pointer

Exam pointer: practice classifying processes (Bernoulli, Poisson, Brownian, random walk) on both axes and stating their increment properties.

Unit 7: Markov Chains: Concepts and Classification

Learning Objectives

- Objective 1:** Define Markov chains and the Markov property
- Objective 2:** Classify states: recurrent, transient, absorbing, periodic
- Objective 3:** Work with transition matrices and n-step probabilities

A Markov chain is a discrete-time, discrete-state process with the Markov property: the future depends on the past only through the present, $P(X_{n+1}|X_n, \dots, X_0) = P(X_{n+1}|X_n)$. This memorylessness makes chains tractable and ubiquitous, from web page ranking (PageRank) to genetics to queueing models. The one-step transition probabilities p_{ij} form the transition matrix P , whose rows sum to 1.

The Chapman-Kolmogorov equations give n-step probabilities: $P^{(n+m)} = P^{(n)}P^{(m)}$, so $P^{(n)} = P^n$ by matrix powering. States are classified by return behavior: recurrent states are returned to with probability 1, transient states may be abandoned forever, and absorbing states ($p_{ii} = 1$) trap the chain. Periodicity matters too: a state with period $d > 1$ is visited only at multiples of d steps, which affects convergence.

First-step analysis computes hitting probabilities and expected hitting times by conditioning on the first transition, producing linear equations. Very important: this 'condition on the first step' technique reappears in gambler's ruin, birth-death processes and queueing; mastering it once pays off across the course.

Key Definitions

Markov property is defined as the future depends on history only through the present state.

Transition matrix is defined as P with entries $p_{ij} = P(X_{n+1}=j|X_n=i)$; rows sum to 1.

Recurrent state is defined as a state the chain returns to with probability 1.

Key Points to Remember

- Markov property = memoryless evolution; transition matrix P captures one-step moves.
- Chapman-Kolmogorov: $P^{(n+m)} = P^{(n)}P^{(m)}$; n-step = P^n .
- States: recurrent vs transient; absorbing ($p_{ii}=1$); periodic vs aperiodic.
- First-step analysis: condition on the first transition to get linear equations.

Unit Summary

Markov chains evolve memorylessly via a transition matrix; Chapman-Kolmogorov gives multi-step probabilities; states are classified by recurrence and periodicity; first-step analysis solves hitting problems.

Exam Pointer

Very important: classify states of a given chain and compute n-step probabilities by matrix powering. First-step analysis problems (gambler's ruin style) are standard.

Unit 8: Stationary Distributions and Long-Run Behavior

Learning Objectives

- Objective 1:** Define and compute stationary distributions
- Objective 2:** State conditions for existence and uniqueness
- Objective 3:** Apply Markov chains to PageRank-style problems

A stationary distribution π satisfies $\pi = \pi P$ with entries summing to 1: if the chain starts in π , it stays in π forever. For finite irreducible aperiodic chains, a unique stationary distribution exists and the chain converges to it regardless of starting state, with $\pi_j = 1/(\text{expected return time to } j)$. This is the workhorse result: long-run fraction of time in each state is given by π .

Computation is linear algebra: solve $\pi(P - I) = 0$ with the normalization constraint. For two-state chains this is hand-solvable; larger chains need systematic elimination. Irreducibility (all states communicate) plus aperiodicity guarantees convergence; periodic chains oscillate and need Cesaro averages. Positive recurrence is the infinite-state analogue ensuring a stationary distribution exists.

The flagship application is PageRank: the web is a Markov chain where pages are states and links are transitions, and the stationary distribution ranks pages by long-run visit frequency. Similar ideas rank sports teams, model population genetics and analyze queueing networks. Very important: exams often ask you to interpret π physically as long-run proportions, not just solve equations.

Key Definitions

Stationary distribution is defined as π with $\pi = \pi P$ and $\sum \pi_i = 1$; invariant under transitions.

Irreducible chain is defined as every state can reach every other state.

Aperiodicity is defined as the chain does not cycle deterministically through state subsets.

Key Points to Remember

- Stationary: $\pi = \pi P$, normalized; long-run fraction of time in each state.
- Finite irreducible aperiodic $\Rightarrow$ unique stationary distribution and convergence.
- $\pi_j = 1 / (\text{mean return time to } j)$.
- PageRank = stationary distribution of the web's link chain.

Unit Summary

Stationary distributions describe long-run behavior via $\pi = \pi P$; irreducibility and aperiodicity guarantee uniqueness and convergence, powering applications like PageRank.

Exam Pointer

Very important: solve $\pi_i = \pi_i P$ for 2- and 3-state chains and interpret the result as long-run proportions. State the convergence conditions precisely.

Unit 9: The Poisson Process

Learning Objectives

- Objective 1:** Define the Poisson process via its postulates
- Objective 2:** Relate counts to exponential inter-arrival times
- Objective 3:** Apply splitting, merging and conditioning properties

The Poisson process $N(t)$ counts events up to time t at rate λ . It is defined by: $N(0) = 0$, independent increments, and $P(\text{one event in } (t, t+h]) = \lambda h + o(h)$ with negligible chance of two. These postulates imply $N(t) \sim \text{Poisson}(\lambda t)$: the count in any interval is Poisson with mean proportional to length. It models call arrivals, radioactive decays, and network packets.

The deep structure lies in the inter-arrival times: they are iid $\text{exponential}(\lambda)$, and memorylessness makes the process 'restart' at every event. Equivalently, given $N(t) = n$, the n arrival times are distributed as order statistics of n iid $\text{uniform}(0, t)$ variables, a beautiful result that simplifies many computations.

Two operations preserve the family: splitting each event independently with probability p yields a Poisson process of rate λp (thinning), and merging independent Poisson processes sums their rates (superposition). Conditioning tricks, like the uniform order-statistics result, turn hard problems into easy ones. Very important: whenever a problem gives a rate and asks about counts in intervals, reach for $\text{Poisson}(\lambda t)$; when it asks about waiting times, reach for exponential.

Key Definitions

Poisson process is defined as a counting process with stationary independent increments and $\text{Poisson}(\lambda t)$ counts.

Thinning (splitting) is defined as independently keeping each event with probability p gives $\text{Poisson}(\lambda p)$.

Superposition (merging) is defined as the sum of independent Poisson processes is Poisson with summed rates.

Key Points to Remember

- Postulates: start at 0, independent increments, $P(\text{event in } h) = \lambda h + o(h)$.
- $N(t) \sim \text{Poisson}(\lambda t)$; inter-arrival times iid $\text{exponential}(\lambda)$.
- Given $N(t)=n$, arrival times look like sorted uniforms.
- Thinning splits rates; superposition adds them.

Unit Summary

The Poisson process counts events at rate λ with Poisson counts and exponential inter-arrivals, closed under thinning and superposition.

Very important: numericals on counts in intervals, waiting-time questions via exponential, and thinning/superposition problems are exam mainstays.

Unit 10: Continuous-Time Markov Chains and Birth-Death Processes

Learning Objectives

Objective 1: Define CTMCs via transition rates

Objective 2: Analyze birth-death processes and their stationary distributions

Objective 3: Connect CTMCs to queueing models

A continuous-time Markov chain stays in state i for an exponential holding time with rate ν_i , then jumps according to transition probabilities. The memoryless holding times preserve the Markov property in continuous time. The generator matrix Q has off-diagonal $q_{ij} = \nu_i \cdot p_{ij}$ and rows summing to zero; it plays the role $P - I$ played discretely, with Kolmogorov forward/backward equations governing $P(t)$.

Birth-death processes are CTMCs where transitions only go to neighboring states: births at rate λ_n , deaths at rate μ_n . They model populations, queues and inventories. Detailed balance $\pi_n \cdot \lambda_n = \pi_{n+1} \cdot \mu_{n+1}$ often yields the stationary distribution in closed form via a simple recursion, avoiding large linear systems. The M/M/1 queue is the famous special case with constant rates.

Stationary distributions exist when the chain is positive recurrent, which for birth-death processes reduces to a summability condition on the rate ratios. Very important: detailed balance is both a computational shortcut and a conceptual check; if you can guess a distribution satisfying it, you have the answer. Exams frequently ask for the stationary distribution of a birth-death process via this route.

Key Definitions

Continuous-time Markov chain is defined as a Markov process with exponential holding times and jump probabilities.

Generator matrix Q is defined as rates q_{ij} with zero row sums, governing continuous-time evolution.

Birth-death process is defined as a CTMC with transitions only between neighboring states.

Key Points to Remember

- CTMC: exponential holding times + jump chain; Q matrix has zero row sums.
- Birth-death: only neighbor transitions; detailed balance gives stationary distribution by recursion.
- Positive recurrence (summability of rate ratios) ensures stationarity.
- M/M/1 queue is a birth-death process with constant birth/death rates.

Unit Summary

CTMCs combine exponential holding times with jump probabilities; birth-death processes yield stationary distributions through detailed balance, underpinning queueing models.

Exam Pointer

Very important: derive the stationary distribution of a birth-death process using detailed balance. Know the generator matrix structure.

Unit 11: Queueing Theory: M/M/1 and Beyond

Learning Objectives

Objective 1: Explain Kendall notation and queueing terminology

Objective 2: Analyze the M/M/1 queue: stability, stationary distribution, performance measures

Objective 3: Apply Little's law

Queueing theory studies waiting lines with Kendall notation A/B/c/K: arrival process / service process / servers / capacity. M/M/1 means Poisson arrivals, exponential service, one server, infinite capacity; M/M/c adds servers; M/G/1 allows general service. Key random quantities: the arrival rate λ , service rate μ , traffic intensity $\rho = \lambda/\mu$, and the number in system L with its queue and waiting-time relatives.

The M/M/1 queue is a birth-death process with birth rate λ and death rate μ . Stability requires $\rho < 1$; otherwise the queue grows without bound. The stationary distribution is geometric: $\pi_n = (1-\rho)\rho^n$. Performance measures follow: mean number in system $L = \rho/(1-\rho)$, mean number waiting $L_q = \rho^2/(1-\rho)$, and mean sojourn time $W = 1/(\mu - \lambda)$. Note how everything blows up as ρ approaches 1: a server at 95% utilization has 19 customers on average.

Little's law, $L = \lambda W$, relates time-averages to customer-averages for any stable system, no distributional assumptions needed. It lets you get waiting times from counts or vice versa, and generalizes to $L_q = \lambda W_q$. Very important: Little's law is the single most-used queueing result; exams test it both directly and as a shortcut inside larger problems.

Key Definitions

Kendall notation is defined as A/B/c/K describing arrival process, service process, servers and capacity.

Traffic intensity is defined as $\rho = \lambda/\mu$; stability of M/M/1 needs $\rho < 1$.

Little's law is defined as $L = \lambda W$: mean number in system equals arrival rate times mean sojourn time.

Key Points to Remember

- M/M/1: Poisson arrivals, exponential service, single server; birth-death with rates λ, μ .
- Stability: $\rho < 1$; stationary $\pi_n = (1-\rho)\rho^n$; $L = \rho/(1-\rho)$; $W = 1/(\mu - \lambda)$.
- Utilization near 1 explodes queue lengths: $\rho=0.95 \rightarrow L=19$.
- Little's law $L=\lambda W$ holds for any stable system; use it as a shortcut.

Unit Summary

M/M/1 queues are birth-death processes stable when $\rho < 1$, with geometric stationary distribution and closed-form performance measures linked by Little's law.

Exam Pointer

Very important: M/M/1 numericals (find L, W, P(wait)) and Little's law applications are guaranteed exam material. Memorize the formulas.

Unit 12: Brownian Motion and Gaussian Processes

Learning Objectives

Objective 1: Define Brownian motion (Wiener process)

Objective 2: Explain its properties: increments, scaling, reflection

Objective 3: Introduce geometric Brownian motion for finance

Brownian motion (Wiener process) $W(t)$ is the continuous-state counterpart of the random walk: $W(0) = 0$, independent increments, $W(t) - W(s) \sim \text{Normal}(0, t-s)$, and continuous sample paths. It models particle diffusion, noise in circuits, and asset prices. The variance grows linearly with time, so uncertainty fans out as the square root of time, a scaling with profound consequences.

Key properties reward study. Scaling: $cW(t/c^2)$ is again Brownian motion. Reflection principle: $P(\max_{s \leq t} W(s) > a) = 2P(W(t) > a)$, which prices barrier options. Nowhere-differentiable paths mean velocity is undefined, so stochastic calculus (Ito) had to be invented to integrate against it. These are not curiosities; they are the working tools of mathematical finance.

Geometric Brownian motion $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$ keeps prices positive and is the engine of the Black-Scholes model. Very important: understand why the $-\sigma^2/2$ correction appears (it comes from Ito's lemma via Jensen's effect on the exponential); exams in applied tracks probe exactly this.

Key Definitions

Brownian motion (Wiener process) is defined as $W(0)=0$, independent normal increments with variance equal to time lag, continuous paths.

Reflection principle is defined as $P(\max_{s \leq t} W(s) \geq a) = 2P(W(t) \geq a)$ for $a > 0$.

Geometric Brownian motion is defined as exp of Brownian motion with drift; the standard stock-price model.

Key Points to Remember

- Brownian motion: start 0, independent $N(0, t-s)$ increments, continuous paths.
- Variance grows linearly; paths are continuous but nowhere differentiable.
- Reflection principle prices barrier events; scaling is a useful symmetry.
- Geometric BM models prices; the $-\sigma^2/2$ drift correction comes from Ito calculus.

Unit Summary

Brownian motion is the continuous Gaussian process with independent normal increments, underlying diffusion models and, via geometric BM, mathematical finance.

Exam Pointer

Exam pointer: state the defining properties precisely and use the reflection principle in a barrier-probability problem.

Unit 13: Estimation, Simulation and Applications

Learning Objectives

Objective 1: Estimate parameters: MLE and method of moments

Objective 2: Simulate stochastic processes via Monte Carlo

Objective 3: Survey applications across engineering and computing

Parameters like arrival rates must be estimated from data. Maximum likelihood estimation picks the parameter maximizing the probability of the observed data; for Poisson counts the MLE of λ is the sample mean. The method of moments equates sample moments to theoretical ones. Estimators are judged by unbiasedness, consistency and efficiency, and large-sample theory leans on the law of large numbers and CLT from earlier units.

When analysis is intractable, simulate. Monte Carlo methods generate sample paths: exponential inter-arrivals for Poisson processes, normal increments for Brownian motion, uniform draws for event times. Repeating many runs estimates probabilities and expectations with quantifiable error shrinking as $1/\sqrt{n}$. Discrete-event simulation extends this to queues and networks, driving real capacity planning.

Applications close the loop: Erlang's queueing formulas dimension telephone networks; Markov models compress and predict in speech and bioinformatics (hidden Markov models); Poisson processes model internet traffic and reliability failures; Brownian motion prices derivatives. Very important: a strong exam answer always connects the mathematics to one concrete application, showing you know what the theory is for.

Key Definitions

Maximum likelihood estimation is defined as choosing the parameter value that maximizes the likelihood of the observed data.

Monte Carlo simulation is defined as estimating quantities by repeated random sampling of a model.

Hidden Markov model is defined as a Markov chain observed only through noisy emissions, used in speech and bioinformatics.

Key Points to Remember

- MLE: maximize likelihood; Poisson rate MLE = sample mean. Method of moments: match moments.
- Monte Carlo: simulate paths (exponential/normal/uniform draws); error $\sim 1/\sqrt{n}$.
- Applications: Erlang/phones, HMMs/speech-bioinformatics, Poisson/traffic-reliability, BM/finance.

- Always tie theory to a concrete application in exam answers. 

Unit Summary

Parameters are estimated by MLE/moments, intractable models are simulated by Monte Carlo, and the course's processes power telephony, computing, biology and finance.

Exam Pointer

Very important: compute an MLE for a Poisson or exponential sample, and describe how to simulate a Poisson process. One application-linked question is typical.