School of Engineering and Applied Science (SEAS), Ahmedabad University

CSE 400: Fundamentals of Probability in Computing

Group-6 Intelligent Transportation System Scribe

Scribe 2 (Part 1) : Q1 to Q3

---

1 Project System and Objective

This project addresses a *probabilistic route selection problem* under specific uncertainties. Unlike deterministic shortest-path approaches that assume fixed travel times, this project models travel time as a random variable to capture the inherent unpredictability of real-world traffic conditions and identify the reliable route.

System Objective

The primary objective of the system is to **maximize the probability of arriving at a destination before calculated time or a specified deadline.** Instead of minimizing average or expected travel time, the system focuses on identifying the *most reliable route* in terms of on-time arrival
Given a starting point s, a destination point d, and a maximum allowable travel time $T_{\max}$, the objective of the system is defined as

$$\max_{P \in \mathcal{P}(s,d)} \Pr(T_{\text{route}}(P) \leq T_{\max}),$$

where P denotes a feasible route from s to d, $\mathcal{P}(s,d)$ is the set of all such routes, and $T_{\text{route}}(P)$ is the random variable representing the total travel time along route P.

1.1 Project System Overview

The end-to-end system follows a probabilistic reasoning pipeline that integrates traffic data, uncertainty modeling, and reliability-based optimization:

- **Historical Congestion Data:** Past traffic patterns and delay statistics are utilized to estimate probability distributions of travel times.
- **Probabilistic Travel-Time Model:** Travel times on individual road segments are modeled as random variables and their distributions are estimated from SUMO-generated traffic data.
- **Optimal Route Selection:** At each intersection, On-time arrival probability is evaluated using empirical CDF and it is bounded analytically using Markov's Inequality.

1.2 Sources of Uncertainty

The probabilistic nature of the problem arises from multiple sources of uncertainty inherent in traffic systems:

- **Traffic Congestion Variability:** Fluctuating traffic demand and congestion levels cause random delays on road segments.
- **Uncertain Waiting Times:** Signal timings and queue formation at intersections lead to stochastic waiting times.
- **Simulation Stochasticity:** Random seed variation in SUMO simulations introduces variability in vehicle interactions, routing dynamics, and delayed propagation.
- **Data Uncertainty and Model Approximation:** Travel-time estimation relies on historical and sensor data that may be noisy or incomplete, requiring simplifying modeling assumptions.

1.3 Problem Interpretation

This project addresses travel time uncertainty in road networks where each segment travel time is a random variable derived from stochastic SUMO simulations and it is done by integrating historical traffic data into a probabilistic framework. The total route time taken is calculated as the sum of the time taken at different segment of the routes. The PDF and CDF graphs are constructed from from SUMO-generated trip-time data using kernel density estimation to obtain the required distributions. Moreover, to reduce unnecessary high delay in reaching the destination, an upper bound is introduced using Markov's inequality, to bound the probability of delay in reaching destination.

2 Key Random Variables and Uncertainty Modeling

This section describes the probabilistic formulation of the routing problem by identifying the key random variables and explaining how uncertainty due to traffic congestion is modeled. The formulation captures variability at both the road-segment level and the route level, and directly supports reliability- based optimization.

2.1 Travel Time on Each Road Segment

Consider a route consisting of n road segments. The travel time required to traverse the i^{th} road segment is modeled as a random variable

$$T_i, \quad i = 1, 2, \ldots, n.$$

Each random variable T_i represents uncertainty caused by traffic congestion, signal delays, road conditions, and stochastic traffic flow. Instead of assigning a deterministic weight, each segment is associated with a probability distribution

$$P(T_i = t),$$

which may be estimated from historical traffic data or simulated traffic conditions.

2.2 Total Route Travel Time

The total travel time for a selected route is defined as the random variable

$$T_{\text{route}} = \sum_{i=1}^{n} T_i.$$

Since each segment travel time is uncertain, the total route travel time is also stochastic. Variability in T_{route} increases with route length and congestion intensity, making deterministic shortest-path approaches insufficient for reliable routing.

2.3 Arrival Time Relative to Deadline

To evaluate whether a route satisfies time constraints, an arrival indicator random variable is defined as

$$A = \begin{cases} 1, & \text{if } T_{\text{route}} \leq T_{\max}, \\ 0, & \text{otherwise,} \end{cases}$$

where $T_{\max}$ denotes the desired arrival deadline.

This variable indicates whether a selected route achieves on-time arrival. The probability $\Pr(A = 1)$ is used as the primary performance metric of the system.

2.4 Uncertainty Modeling and Probabilistic Assumptions

The following assumptions are made to model uncertainty in the routing problem:

- **Initial Independence Assumption:** For the base model, travel times on individual road segments, represented as T_i are assumed to be independent random variables.
- **Traffic Congestion Variability:** The probability distributions of segment travel times can change from one time interval to another to account for peak and off-peak traffic conditions.
- **Ignoring External Factors:** Throughout the project, we assume that there are no other external factors such as accidents, roadworks and change in weather conditions (like high humidity or fog) affecting the traffic congestion on any road networks.

2.5 Reliability-Based Objective

Based on the defined random variables, the routing problem is formulated as the maximization of on-time arrival probability:

$$\max \ \Pr(T_{\text{route}} \leq T_{\max}).$$

This formulation emphasizes reliability under uncertainty rather than minimization of expected travel time, making it suitable for congestion-aware and deadline-sensitive routing scenarios.

2.6 Modeling Assumptions

In order to model the uncertainty in travel times while keeping the model tractable, the following assumptions are made:

(a) **Stochastic Travel Times:** The travel time on each route segment is modeled as a random variable to represent the uncertainty introduced by traffic congestion and signal delays.

(b) **Baseline Independence Assumption:** To keep the model simple at this stage, the travel times on individual route segments T_i are modeled as independent random variables. This allows for easy aggregation of the uncertainty on individual segments to estimate the total route travel time.

(c) **Limited Dependency Modeling:** There could be some dependencies among the travel times on different route segments because of common traffic conditions but these are not modeled at this stage.

(d) **Excluding External Factors:** Throughout the project, we assume that there are no other external factors such as accidents, roadworks and change in weather conditions (like high humidity or fog) affecting the traffic congestion on any road networks.

2.7 Delay Bounding using Markov's Inequality

Here, assuming the non-negative random variable T_{route}, Markov's Inequality provides us an upper bound on delay probability as given below:

$$\Pr(T_{\text{route}} \geq D) \leq \frac{\mathbb{E}[T_{\text{route}}]}{D},$$

where D denotes the arrival deadline.

This bound is distribution-free and it is introduced to validate empirical delay probabilities derived from simulation-based CDF graphs.

3 Probabilistic Reasoning and Dependencies section

This section is a overview how the probabilistic relationships are used in the project to support reasoning and decision-making under uncertain travel conditions.

3.1 Probabilistic Representation of Travel Time

- Travel time on each route segment is modeled as a random variable. This representation reflects uncertainty arising from varying traffic conditions.
- The probabilistic model enables the system to reason about a range of possible travel outcomes rather than assuming a single fixed travel time.

3.2 Route-Level Inference

- The total travel time of a route is obtained by aggregating the random travel times of its individual segments.
- This aggregation is used to compute the probability of arriving within a given time limit, which supports comparison and selection of candidate routes during decision-making.

3.3 Independence and Dependence Assumptions

- In the initial model, travel times of different route segments are assumed as independent in order to simplify probabilistic reasoning and implementation.
- Possible dependence between adjacent segments due to shared traffic conditions is acknowledged but not explicitly modeled at the current stage.

3.4 Simulation-Based Reasoning Implementation

The probabilistic reasoning framework implemented in this project is based on Monte Carlo simulation in SUMO, where multiple traffic situations generate travel-time samples. These samples drive empirical inference, distribution fitting, and inequality-based delay analysis.

4 Probabilistic Route Evaluation and CDF Construction

This section explains how the probability of on-time arrival is computed using simulation data and empirical distribution estimation.

4.1 Monte Carlo Simulation Framework

To estimate travel-time distributions, multiple stochastic simulations are performed in SUMO under varying traffic seeds and demand patterns. For each candidate route P, we collect N independent samples:

$$T_{\text{route}}^{(1)}, T_{\text{route}}^{(2)}, \ldots, T_{\text{route}}^{(N)}.$$

Each sample represents the total travel time observed in one simulation run.

4.2 Kernel Density Estimation (KDE)

To visualize the underlying probability density function (PDF), kernel density estimation is applied to the collected samples. This provides a smooth approximation of the travel-time distribution without assuming a specific parametric model.

The combination of empirical CDF and KDE allows both quantitative probability estimation and qualitative distribution analysis.

5 Comparative Route Selection and Reliability Analysis

This section describes how routes are compared and chosen based on reliability.

5.1 Comparison Criterion

Assume there are two routes to compare, $P1$ and $P2$. The system chooses the route with the highest probability of reaching before the deadline $Tmax$. In other words, the best route is the one that provides a better opportunity to arrive at the destination on time.

5.2 Expected Time vs Reliability

The route with the shortest expected travel time is not necessarily the most reliable route. A route may be the shortest on average but also very unreliable. Another route may be slightly longer on average but with more predictable travel times. When there is a deadline to meet, predictability is more valuable than average travel time.

5.3 Variance and Risk Consideration

Variance of travel time is also a significant factor. A route with highly variable travel times is more likely to encounter extreme delays. This further reduces the chances of arriving on time. Routes with smaller variance are more reliable, even if they are slightly longer on average.

5.4 Decision Interpretation

Thus, the best route is the most reliable route, not necessarily the shortest route on average. This is more realistic since arriving on time is more valuable than minimizing average travel time.

6 Analytical Bounds and Theoretical Guarantees

This section provides a theoretical validation of the empirical simulation results using Markov's inequality. The goal is to establish a distribution-free probabilistic guarantee on delay.

6.1 Markov's Inequality

Let T_{route} denote the total route travel time, which is a non-negative random variable.

For any deadline $D > 0$, Markov's inequality states:

$$\Pr(T_{\text{route}} \geq D) \leq \frac{\mathbb{E}[T_{\text{route}}]}{D}.$$

This provides an upper bound on the probability of exceeding the deadline.

6.2 Lower Bound on On-Time Arrival Probability

Rewriting the inequality in terms of on-time arrival:

$$\Pr(T_{\text{route}} \leq D) = 1 - \Pr(T_{\text{route}} \geq D) \geq 1 - \frac{\mathbb{E}[T_{\text{route}}]}{D}.$$

Thus, even without knowing the full distribution of travel time, we obtain a guaranteed lower bound on reliability.

6.3 Interpretation in the Routing Context

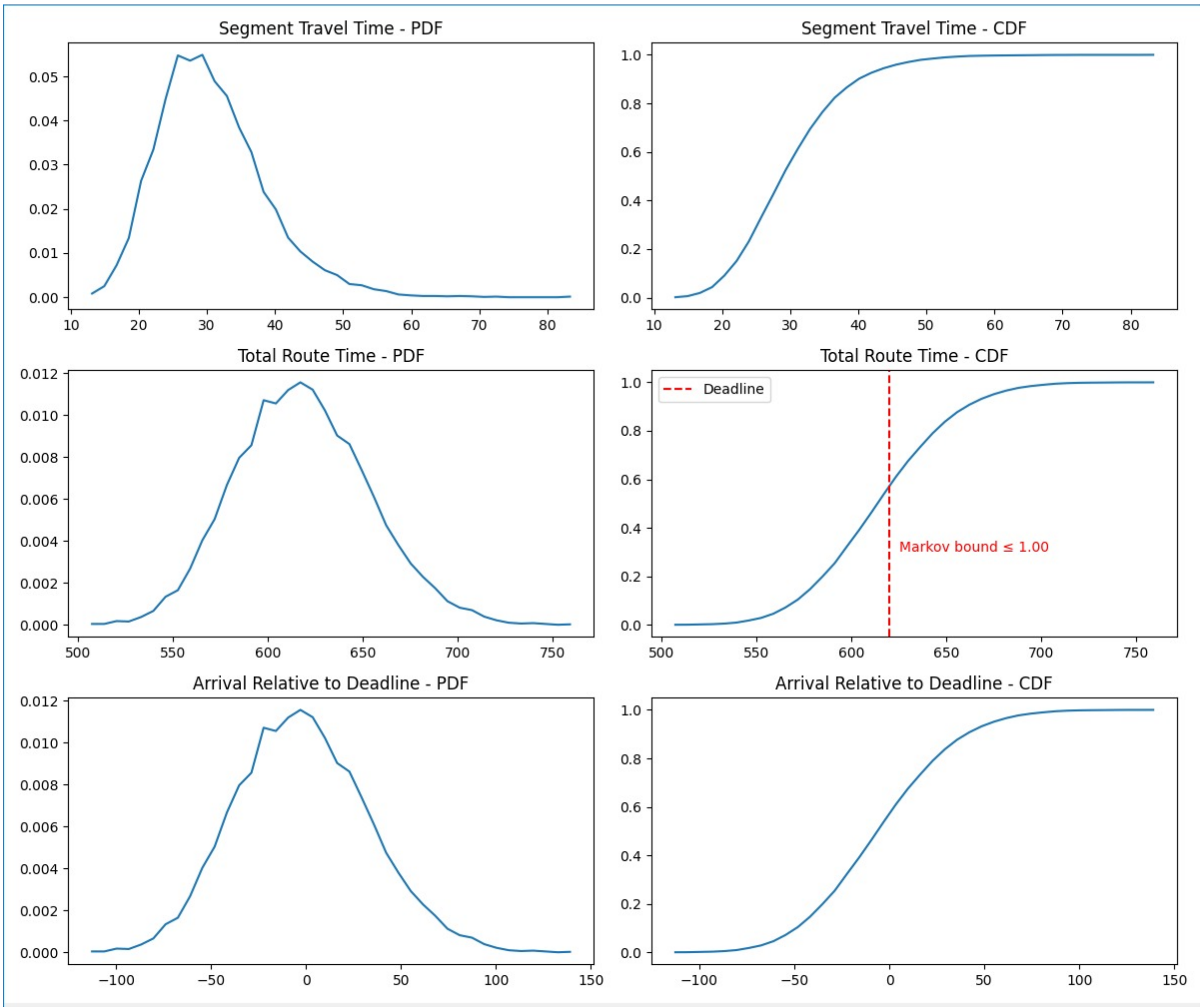
- The empirical CDF obtained from Monte Carlo simulation provides an estimate of $\Pr(T_{\text{route}} \leq D)$.
- Markov's inequality provides a conservative theoretical lower bound based solely on the mean travel time.
- If the empirical probability is significantly higher than the Markov bound, it indicates that the inequality is loose but still valid.

6.4 Comparison with Simulation Results

Figure 6.4 illustrates:

- The empirical PDF and CDF of segment travel time.
- The empirical PDF and CDF of total route time.
- The CDF evaluated at the deadline D.

The vertical red dashed line represents the deadline. The CDF value at this point corresponds to the empirical on-time arrival probability.



6.5 Discussion

Markov's inequality depends only on the expected travel time and does not account for variance or distribution shape. Therefore, it typically produces a loose bound. However, it is valuable because:

- It is distribution-free.
- It provides a guaranteed reliability bound.
- It theoretically validates simulation-based inference.

6.6 Conclusion

The analytical bound confirms that the empirical on-time arrival probability observed in simulations satisfies

$$\Pr(T_{\text{route}} \le D) \ge 1 - \frac{\mathbb{E}[T_{\text{route}}]}{D}.$$

Hence, the routing framework is supported both by simulation-based empirical reasoning and by theoretical probabilistic guarantees.

End of Submission