# Probabilistic Travel Time Modeling
# Data Source Description

Milestone 2

## 1. Introduction

This document describes the synthetic data generation and probabilistic assumptions used in the travel time reliability model.

## 2. Segment Travel Time Model

Each road segment travel time is modeled as a Lognormal random variable:

$$T_i \sim \text{Lognormal}(\mu = 3.4, \sigma = 0.25)$$

Which implies:

$$\ln(T_i) \sim \mathcal{N}(3.4, 0.25^2)$$

**Justification:**

- Travel time is strictly positive
- Traffic delays produce right-skewed distributions
- Lognormal captures multiplicative traffic effects

## 3. Monte Carlo Data Generation

Synthetic data was generated using 8000 samples:

```
segment_time = np.random.lognormal(mean=3.4, sigma=0.25, size=8000)
```

# 4. Route Travel Time

Each route consists of 20 independent segments:

$$T_{route} = \sum_{i=1}^{20} T_i$$

Simulated as:

```
total_route_time = np.sum(
np.random.lognormal(3.4, 0.25, (8000, 20)), axis=1)
```

# 5. Deadline and Reliability

A fixed deadline was defined:

$$D = 620$$

Reliability objective:

$$P(T_{route} \leq D)$$

# 6. Markov Inequality Bound

To obtain a distribution-free bound:

$$P(T_{route} \geq D) \leq \frac{E[T_{route}]}{D}$$

This provides a conservative upper bound on delay probability.

# 7. Libraries Used

- NumPy – data generation and computation
- Matplotlib – plotting PDF and CDF

# 8. Summary

The model uses synthetic lognormal data and Monte Carlo simulation to estimate travel time uncertainty and evaluate deadline reliability using probabilistic bounds.