Performance Evaluation

Homework 2
Random Waypoint Simulation
With Solutions

Jean-Yves Le Boudec
Eleni Stai

Due date: March 20, 2017, 23:55PM

Important: (1) In some problems we ask you to give the assumptions, comments or explanations. All these questions are accounted for grading. (2) write your answers in LATEX. (3) Submit a single zip file which includes the pdf file for your answers and your Matlab files. Don’t clutter the pdf document with Matlab code as this makes it difficult to read.

1 Statistics Warmup

Problem 1. Write a program in Matlab that generates a sample of \( n \) iid standard normal variables, and display the corresponding histogram. Repeat the operation 9 times, for \( n = 10, 30, 90, 270, \ldots \) and display the results on a \( 3 \times 3 \) panel. Solution. (Figure 1)

![Histograms of normal distribution samples. The numbers in the titles are the numbers of samples.](image)

Figure 1: Histograms of normal distribution samples. The numbers in the titles are the numbers of samples.

Problem 2. Plots and Distributions
1. Plot the densities of the following distributions: Uniform(0, 50), Normal(0, 0.5), Student(3), Student(10), Exponential(0.5), Exponential(2), Beta(10, 90), Beta(100, 10). Solution. (Figure 2)
2. Write a program which generates a sample of \( n = 1000 \) RVs having a distribution in one of the above. Do it for all the distributions given above. Display the corresponding standard normal QQ-plots. **Solution.** (Figure 3)

![QQ-plots](image)

**Figure 3:** Standard normal QQ-plots of the distributions.

3. How do you interpret an S-shape in a normal QQ-plot? A U-shape? Justify. **Solution.** Adopted from the S-PLUS online help: “QQ-plots are used to assess whether data have a particular distribution, or whether two data sets have the same distribution. If the distributions are the same, then the plot will be approximately a straight line. The extreme points have more variability than points toward the center. A plot with a “U” shape means that one distribution is skewed relative to the other. In the case of the normal QQ-plot “U” shape means that the distribution is not symmetrical. An “S” shape implies that one distribution has longer or shorter tails than the other. In the case of the normal QQ-plot “S” shape that is bent down on the left and bent up on the right means that the data have longer tails than the Gaussian.”

2 **SIMULATE RANDOM WAYPOINT AND LOOK AT WHAT YOU HAVE DONE**

**Problem 3.** Simulate the random waypoint model defined in the chapter “Simulation” (Example 6.5). Use the following parameters.
• Number of users: \( N = 100 \),

• The area is a rectangle of dimensions \( l \times L \) with \( l = L = 1500m \),

• Let \( S \) be one of the two group members’ sciper number. \( v_{\text{min}} = (0.7 + 0.01 \times S \% 21) \text{m/s}, v_{\text{max}} = (1.9 + 0.1 \times S \% 11) \text{m/s} \), where \( \% \) is the modulo operator.

• The simulation terminates at (simulated) time \( T_s = 86400 \text{ s} = 1 \text{ day} \).

Write a simulation program in Matlab that

1. computes and displays the mean, the minimum and the maximum number of waypoints reached by the different mobiles for only one simulation run \textbf{Solution.} \ Number of reached waypoints:
   \[ nbWP_{\text{mean}} = 137, nbWP_{\text{max}} = 159.23, nbWP_{\text{mean}} = 177. \]

2. displays the trajectory and waypoints of one user; of 5 users. \textbf{Solution.} (Figure 4 and Figure 5)

![Figure 4: Trajectory and waypoints of one user.](image)

\[ \text{Trajectory of one user} \]

\[ \text{Waypoints of one user} \]

\[ \text{0 500 1000 1500} \]
\[ \text{0 500 1000 1500} \]

\[ \text{0 500 1000 1500} \]
\[ \text{0 500 1000 1500} \]

How much real time did your program take for 1 day of simulated time? \textbf{Solution.} Real time depends on your machine – a fraction of a second. We obtained: 0.170 \text{ s}.

3 \ \textbf{DIFFERENT VIEWPOINTS}

\textbf{EVENT AVERAGE (PALM) VIEWPOINT}

a) Display the histogram of speeds sampled at transition epochs \( T_n \), based on the samples for one mobile and for all mobiles.
Figure 5: Trajectory and waypoints of 5 users.

b) Show a histogram of mobile positions based on the samples for one mobile and for all mobiles. To this end, discretize the area into square bins; display a grey shade diagram where the intensity of the grey is proportional to the frequency.

You have obtained the view experienced by an observer who comes at an arbitrary transition time $T_n$.

Solution. (Figure 6 and Figure 7) Both speed and positions have uniform distributions, as specified by the model.

**TIME AVERAGE VIEWPOINT**

a) Now sample the position and speed of mobiles every 12 seconds. Display the histogram of speeds sampled every 12 seconds (based on the samples for one mobile and for all mobiles).

b) Similarly, display a histogram of mobile positions sampled every 12 seconds (based on the samples for one mobile and for all mobiles).

c) You have obtained the view experienced by an observer who comes at an arbitrary point in time. Comment on what you see and compare the obtained results with those from the Event Average Viewpoint.

Solution. (Figure 8 and Figure 9) Speed and positions are not uniform. Speed is more likely to be small than large; this is because a trip at low speed takes more time, therefore a mobile spends more time at low speed than at high speed. Positions are more likely to be around the center than at the edge because in a trip the time is spent between the waypoints.
Figure 6: Palm distribution of the speed (histogram).

Figure 7: Palm distribution of the position.
Figure 8: Time average based distribution of the speed (histogram).

Figure 9: Time average based distribution of the position.
4 CONFIDENCE INTERVALS

4.1 CONFIDENCE INTERVALS FOR MEDIANS AND MEANS

For each of the $N$ mobiles compute the average of the speed values sampled at transition epochs $T_n$ (event average speed for each user). This gives you a data sequence $X_1, X_2, \ldots, X_N$. Similarly, for each of the $N$ mobiles compute the average of the instant speed values sampled at arbitrary instants of time (time average speed for each user). This gives you another data sequence $Y_1, Y_2, \ldots, Y_N$.

a) First, take $N = 100$. For the sequences $X$ and $Y$, compute and display on the same plot: median, confidence interval for the median at level 0.95, mean and confidence interval for the mean at level 0.95. Write explicitly and verify all the assumptions you use for the above computations.

b) Compare the time average speed sequence with the event average speed sequence in terms of their mean/median values and the corresponding confidence intervals. How do you interpret the mean value of the event average speed sequence?

c) Do the same for $N = 25$.

Solution. (Figure 10)

![Figure 10: Median (solid line in the center of the box), confidence interval for the median (notched area of the box), mean (dashed line), confidence interval for the mean (dotted line)](image)

The numerical values are:

**Time average speed**:
- median, $N=100$: 1.43777
- median, $N=25$: 1.42951
- mean, $N=100$: 1.43954
- mean, $N=25$: 1.43432
- confidence interval for the mean, $N=100$: (1.434, 1.445)
- confidence interval for the mean, $N=25$: (1.423, 1.445)

**Event average speed**:
- median, $N=100$: 1.49973
- median, $N=25$: 1.49914
- mean, $N=100$: 1.49795
- mean, $N=25$: 1.49794
- confidence interval for the mean, $N=100$: (1.493, 1.502)
- confidence interval for the mean, $N=25$: (1.489, 1.507)

c) Compare the results obtained for $N = 100$ against the corresponding results obtained for $N = 25$, and give the conclusion on how the number of samples affect the confidence intervals.
Solution.

![Normal qq-plots](#)

One assumption is that the average values obtained for different users are iid random variables, which is true by the simulation design. To compute confidence intervals for the means, we assume that the data is approximately normally distributed (a data point is a corresponding average value obtained for one user). From the qq-plots (Figure 11) we find visually that the data points can be considered as approximately normal. We show qq-plots only for \( N = 100 \), as both sets of samples (\( N = 100 \) and \( N = 25 \)) come from the same distribution and more samples give more evidence about the distribution. To compute means, medians and confidence intervals for the medians, we do not need any assumption about the data distribution.

The mean value for the event average speed sequence (around 1.5) can be easily guessed as the speed values at the transition epochs are sampled from the uniform distribution between 1 and 2 in the simulation performed for these solutions. The mean and median values of sequences \( X \) and \( Y \) differ significantly and there is no overlapping between the corresponding confidence intervals for \( X \) and \( Y \). So the sequences \( X \) and \( Y \) differ significantly statistically. Specifically, the time average speed sequence, \( Y \), has a lower mean value as a trip at low speed takes more time, therefore a mobile spends more time at low speed than at high speed (same reason as in the previous section).

Comparing the results for \( N = 100 \) and \( N = 25 \), the estimate uncertainty decreases with the number of samples, so confidence intervals decrease with the number of samples.

### 4.2 Prediction Intervals For Samples

For each of the \( N \) mobiles compute the average of the instant speed values sampled at arbitrary instants of time (time average speed for each user). This gives you a data sequence \( Y_1, Y_2, \ldots, Y_N \). First take \( N = 100 \). For this sequence compute and display on the same plot:

1. prediction interval for a sample at level 0.95, computed assuming the data are normal and using the
estimates of the mean and variance, and
b) prediction interval for a sample at level 0.95, computed using order statistic.

Do another plot for \( N = 50 \), and one for \( N = 25 \).

**Solution.** Solution is given at Figure 12. For \( N = 25 \), prediction intervals at confidence level 0.95 cannot be computed using order statistic (at least 39 samples is needed).

![Figure 12: Prediction intervals for samples: using normal approximation (solid lines) and using order statistic (dotted lines).](image)

The numerical values are:

**Time average speed, prediction intervals:**

- \( N=100 \), order statistic: (1.383 1.511)
- \( N=100 \), normal approximation: (1.384 1.495)
- \( N=50 \), order statistic: (1.383 1.518)
- \( N=50 \), normal approximation: (1.382 1.500)
- \( N=25 \), order statistic: there are no enough samples to compute 0.95 prediction interval
- \( N=25 \), normal approximation: (1.378 1.491)

Write explicitly and verify all the assumptions you used for the above computations.

c) Compare the results obtained in (a) against the corresponding results obtained in (b): are they similar or very different and why?

d) Compare the results obtained for \( N = 100 \) against the results obtained for \( N = 50 \) and \( N = 25 \): how does the value of \( N \) affect the prediction intervals in (a) and how in (b)? Explain.

**Solution.** One assumption is that the average values obtained for different users are iid random variables, which is true by the simulation design. To apply normal-approximation method for computing prediction intervals, we assume the data is approximately normally distributed. From the qq-plots (Figure 11) we find visually that the data points can be considered as approximately normal. To apply order-statistic method for computing prediction intervals we do not need any assumption about the data distribution.

"(a) versus (b)"; for \( N = 100 \) and \( N = 50 \), the results in (a) are similar to those in (b), because the data is approximately normal. But for \( N=25 \), the method used in (b) can not be applied. Use of the additional information that the data points are approximately normal allows to compute the confidence interval using the method from (a) even if the number of the samples is very small and the method from (b) cannot be used. If the data is not approximately normal, only (b) can be used and (a) usually gives wrong results.

"Different N values"; decreasing \( N \) does not affect generally the size of prediction interval. Very small \( N \) (\( N=25 \)) does not allow to compute the interval by the method used in (b).