ISyE 6404 Enrichment Project #1 (EP-2): K-M Estimation, Kernel Regression and Spline

All enrichment and computer projects should follow the guidelines stated in EP-1 for preparing and submitting your reports.

1. **K-M Estimation (25%):**

   Locate *a data set* with right-censoring (in Type-I Censoring) in the field of your interest, e.g., eCommerce, medical study, drug development, supply-chain/logistics operations, for applying the K-M Estimator to estimate the survival function with pointwise confidence intervals.

2. **Kernel and Related Regression with One Explanatory Variable (40%):**

   Locate *a data set* suitable for nonparametric regression (usually has nonlinear y-x relationship) in the field of your interest, e.g., eCommerce, medical study, drug development, supply-chain/logistics operations. Apply all of the procedures below

   (i)   Kernel Regression,
   (ii)   Local Polynomial Regression,
   (iii) LOESS,
   (iv)   Smoothing Spline,

   to the y-x data-fit. Compare fits from the four methods.

3. **Cross-Validation With the "Leave-One-Out" Procedure (10%):**

   Compare the above four methods with a leave-one-out cross-validation procedure.

4. **Resampling Procedures: Bootstrap and Jackknife (25%):**

   1) Select an input x_0 in the [min(x-data), max(x-data)].
   2) Use all four regression models built in Task #2 to make point-predictions of Y at x_0.
   3) Use both bootstrap ($B = 1000$) and jackknife resampling procedures to find a 90% pointwise confidence interval (CI) for the point-prediction. If the resampled distribution of the point-prediction is symmetric, use 5% in each tail to find the CI-bounds. If the distribution is not symmetric, use the HPD-interval idea to find the CI-bound. Compare the results from four regression methods.