Stationary Poisson Process Project Ideas

During the R exercises, you ran a simulation to predict the number of gas escapes expected during the next 10 years. The three options below are extensions of that idea.

1. Extend the current code and estimate T_{10}, T_{100}, and T_{200}, i.e. the times of the 10-th, 100-th and 200-th gas escapes. Remember that T_n is the sum of the first n interfailure times and its distribution, in this case, is a Gamma $\mathcal{G}(n, \lambda)$. In addition to point estimates for each time, show some histograms and compare.

2. A Bayesian approach. If we assume that our prior distribution is a Gamma $\mathcal{G}(\alpha, \beta)$, then our posterior distribution is a Gamma $\mathcal{G}(\alpha + n, \beta + sT)$.

   Think about how to choose α and β. What is the relationship between them and what an expert would know?

   Here is some guidance: We know that the interfailure times have an exponential distribution $\mathcal{E}(\lambda)$ so that $E(X_1) = \frac{1}{\lambda}$, so $\lambda = \frac{1}{E(X_1)}$. We also know that choosing a Gamma $\mathcal{G}(\alpha, \beta)$ prior implies $E(\lambda) = \frac{\alpha}{\beta}$ and that $\text{var}(\lambda) = \frac{\alpha}{\beta^2}$.

   If you were discussing this with an engineer, she would have an opinion about X_1. You can use that opinion to guess a value $\tilde{\lambda}$ about λ and use it to specify the prior distribution.

   The main question here is: What does the prior variance mean in a Bayesian framework?

   Use different pairs of α and β to get different Bayesian estimators (posterior mean of gamma) and compare histograms.

   Choose a pair (α_1, β_1) and plot both prior and posterior distributions when using the data from the entire gas network.

   Optional: compare triples of (prior mean, posterior mean, MLE) when changing parameters. Try for very small variance and very large variance as well.

3. Use the methodology we learned on 2/24 to analyze a new dataset about accidents in construction sites in Spain. These data go from January 1988 to November 1990.

   In this example, the number of workers corresponds to the kilometers of gas pipes in the Milano example and the number of accidents corresponds to the number of gas escapes.

   Given the accidents data, calculate an individual propensity to have an accident. Repeat more or less what we did and compute an MLE for each month, an MLE for each year, and, if you'd like, try the Bayesian approach.

For all of these options, please provide numerical and graphical results, but also **explanations**!