Xinyu Tan

- **2. Model Checking**

(a) Set up posterior predictive test quantities to check the following assumptions: 1) independent Poisson distributions 2) no trend over time.
To test Poisson distributions, we use Fano factor (σ^2/μ). For Poisson distribution, it is close to 1. To test if the data are from independent distributions, I think it can be tested along with "no trend over time" using autocorreltion.

(b) Use simulations from the posterior predictive distributions to measure the discrapancies.

(i) Assume that the numbers of fatal accidents in each year are independent with a Poisson distribution.
Hence fatal accident follows:

$$p(y|\theta) = \frac{1}{y!}\theta^y \exp(-\theta)$$

The likelihood:

$$p(y|\theta) = \prod_{i=1}^{10} \frac{1}{y_i!}\theta^{y_i} \exp(-\theta) \sim \theta^{10\bar{y}} \exp(-10\theta)$$

where $\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i$.
Since I don't have any extra information, let's set prior distribution for $\theta \sim$ beta$(0,0)$.
Hence the posterior distribution

$$p(\theta|y) \sim \text{Gamma}(10\bar{y}, 10)$$

To compare the posterior predictive test quantities, we will perform the following sampling 1000 times:

(1)) $\theta^s \sim p(\theta|y_1, \cdots, y_{10})$

(2)) $\tilde{y_i^s} \sim p(y|\theta^s), \forall i \in \{1, 2, \cdots, 10\}$

For Fano factor, the p-value and graphical result are shown in 1. p-value is near 0.5, indicating that posterior predictive's fano number is similar to data.
Use lag $k=1$ autocorrelation, defined as

$$r_1 = \frac{\sum_{i=1}^{N-1}(y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=1}^{N-1}(y_i - \bar{y})^2}$$

The p-value and the graphical result are shown in 2.
We notice that $p(\text{autocorrelation}(sample) > \text{autocorrelation}(data)) \approx 0.025$, which means that the assumption that year-to-year fatal accidents are independent is inadequate.

(ii) Assumes that the numbers of fatal accidents in each year follows independent Poisson distribution with a constant rate and an exposure in each year proportional to the number of passenger miles flown.

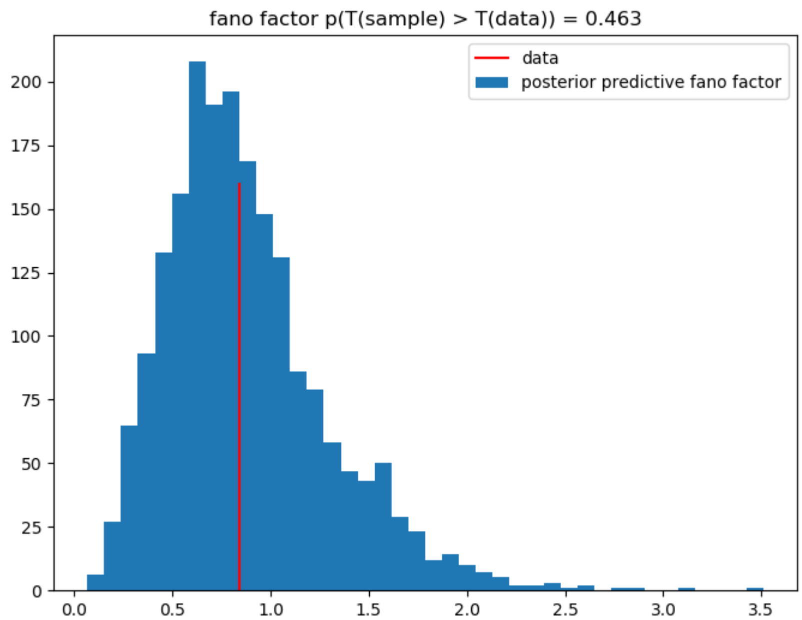

Figure 1: Posterior predictive Fano factor distribution

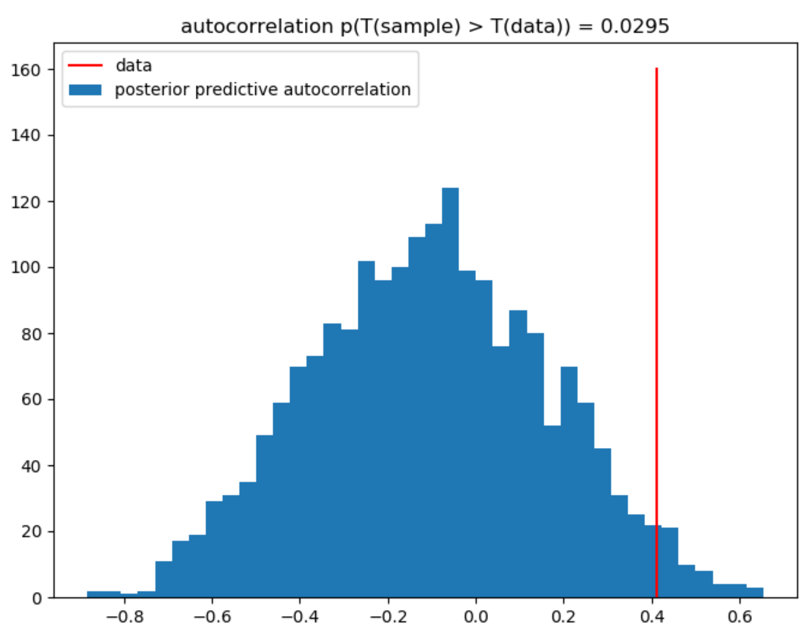

Figure 2: Posterior predictive autocorrelation distribution