Worksheet 17 (Solutions)

1. A fridge manufacturer tests how long it takes before their devices need to be repaired. Looking at 4 randomly selected units, it took 450, 1200, 1500, and 2400 days until replacement. (a) describe the distribution in terms of one of our known probability densities and (b) estimate the parameters of the model.

The distribution to use is the Geometric. Let's use the MME to find the parameter p:

$$\frac{1}{p} = \bar{X}$$
$$\widehat{p} = \frac{1}{\bar{X}}$$

Which gives that p is estimated as: 0.0007207207, for a 0.07% chance of failure in a given day.

2. You want to estimate what proportion of cities in the US have over 1 million people in them. All you know is that the average city has an average 20k people. Construct a probability model for this question and make an estimate of the proportion of cities with more than 1 million people. Hints: Do this question in 'thousands of people'. Also, remember that the Pareto distribution has a mean $\frac{\alpha}{\alpha-1}$ and density:

$$f(x) = \frac{\alpha}{x^{\alpha+1}}, x > 1.$$

The correct distribution is the Pareto. We can again use the method of moments to find $\widehat{\alpha}$:

$$\begin{aligned}\frac{\widehat{\alpha}}{\widehat{\alpha}-1} &= \bar{X} \\ &= 20.\end{aligned}$$

Solving for α:

$$\begin{aligned}\widehat{\alpha} &= 20\widehat{\alpha} - 20 \\ \widehat{\alpha} &= \frac{20}{19}.\end{aligned}$$

So, we can now integrate to find the answer the question:

$$\begin{aligned}\int_{1000}^{\infty} \frac{\alpha}{x^{\alpha+1}} &= \left[x^{-\alpha}\right]_{1000}^{\infty} \\ &= (1000)^{-20/19} \\ &= 0.0006951928\end{aligned}$$

So, about 0.06% of cities will be that large.

3. There are only 10 US cities with more than 1 million people. Ways of counting the number of cities differ, but there are somewhere between 3000 and 20000 cities in the US. How does your estimate from question 2 compare to the actual data?

If there are 3000 cities, I would expect about 2 using the model and if there are 20000 I would expect about 14. So yes, the model is quite reasonable.

4. During my last four office hours I had: 1, 2, 2, and 4 students. Build a probability model and estimate the parameter(s) that describe it given the data. Predict the probability that nobody shows up to my next office hours.

This looks like a Poisson distribution. Let's use MME to estimate λ:

$$\begin{aligned} \widehat{\lambda} &= \bar{X} \\ &= 2.25 \end{aligned}$$

So the density function gives:

$$\begin{aligned} \mathbb{P}(X = 0) &= \frac{\lambda^0 e^{-\lambda}}{k!} \\ &= e^{-2.25} \approx 0.1053992 \end{aligned}$$

So about a 10% chance.

5. A school group is giving out T-Shirts in the dining hall. Each shirt has a tag with an number on it; the first shirt was numbered 1, the second was 2, the third 3, and so forth. Shirts were handed out at random, regardless of the number. When you arrive (late) for lunch, only five shirts remain. They have the numbers 10, 45, 102, 210, and 230. Build a probability model the describes the situation (approximately). Estimate the total number of T-Shirts that were handed out using both an MLE and the MME.

This can be approximately modeled as a $U(1, \theta)$ distribution, for an unknown θ (the only difference is that we are sampling without replacement rather than with it). The MLE estimator just yields that the best guess is 230. The MME gives:

$$\begin{aligned} \bar{X} &= \frac{1}{2}(\widehat{\theta} + 1) \\ \widehat{\theta} &= 2\bar{X} - 1 \end{aligned}$$

And so it gives and estimate of:

$$\widehat{\theta} = 2 \cdot 119.4 - 1 = 237.8$$