Statistical Inference
R: Assessment 1
Solution

I. Kiss and F. Di Lauro

The solution is given in the form of a PDF document as well as an example of R code *R_assessment1_solution.R*.

1 Identify distributions from data.

The required estimations are done in the R code and table 1 gives the adjusted values.

Column	Distribution	Estimated parameters
1	Binomial	$(n, p) = (5, 0.504)$
2	Log-normal	$(\mu = 0.55, \sigma = 1.89)$
3	Beta	$(\alpha, \beta) = (3.25, 1.69)$
4	Uniform	$(a, b) = (0.01, 9.93)$
5	Exponential	$\lambda = 2.69$
6	Poisson	$\lambda = 9.17$

Table 1: Adjusted distributions for each column of the dataset.

2 Likelihood ratio test.

1. It is sufficient to replace the likelihood function in the definition by the particular case of the exponential distribution:

$$\begin{aligned}\Lambda(x) &= \frac{\mathcal{L}(\lambda_0; x)}{\mathcal{L}(\lambda_1; x)} \\ &= \prod_{i=1}^{n} \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left((X_i - \mu_0)^2 - (x_i - \mu_1)^2\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2}\left(n(\mu_0^2 - \mu_1^2) - 2\sum_{i=1}^{n} X_i(\mu_0 - \mu_1)\right)\right]\end{aligned} \tag{1}$$

2. We can manipulate the expression (1) to

$$
\begin{aligned}
\log \Lambda(x) &= -\frac{1}{2\sigma^2}\left(n(\mu_0^2-\mu_1^2) - 2\sum_{i=1}^{n} X_i(\mu_0-\mu_1)\right) \\
\rightarrow \frac{\sum_{i=1}^{n} x_i}{n} &= \frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)
\end{aligned}
\tag{2}
$$

Now, since all X_i are i.i.d. with distribution $\mathcal{N}(\mu_0, \sigma^2)$, we know that $\frac{\sum_{i=1}^{n} x_i}{n} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$.

3. From the previous point, we can standardize the random variable $\bar{X}$ to get

$$
\mathbb{P}\left(\bar{X} > \kappa | H_0\right) = \mathbb{P}\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \kappa' | H_0\right), \tag{3}
$$

where $\kappa' = \frac{\kappa-\mu_0}{\sigma/\sqrt{n}}$, is the well-known $q_{0.95}$ quantile of a standard Normal variable $q_{0.95} \sim 1.645$. By inverting the last expression, we get

$$
\kappa = \frac{\sigma}{\sqrt{n}} q_{0.95} + \mu_0 \simeq 11.645 \tag{4}
$$

4. We start from the previous expression and manipulate it to get an expression for Λ. Note that $\mu_0 - \mu_1 < 0$ and this is important in determining the sign of the inequality:

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} \frac{x_i}{n} \geq k\right) &= q_{0.95}, \\
\Leftrightarrow \mathbb{P}\left(\frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) \geq k\right) &= q_{0.95}, \\
\Leftrightarrow \mathbb{P}\left(\ln(\Lambda(x)) < \frac{n(\mu_0-\mu_1)}{\sigma^2}\kappa - \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) &= q_{0.95}, \\
\Leftrightarrow \mathbb{P}\left(\Lambda(x) < \exp\left[\frac{n(\mu_0-\mu_1)}{\sigma^2}\kappa - \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right]\right) &= q_{0.95}.
\end{aligned}
\tag{5}
$$

In other words, one must choose $k_\Lambda = \exp\left[\frac{n(\mu_0-\mu_1)}{\sigma^2}\kappa - \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right]$.

5. Using R one has $\Lambda(x) = 5.2 \cdot 10^{11}$ and $k_\Lambda = 71.93$ thus $\Lambda(x) \geq k_\Lambda$ and H_0 is not rejected.

6. In this test, the p-value is defined as follows:

$$
\begin{aligned}
p &= \mathbb{P}(\Lambda(x') \leq \Lambda(x) | H_0) \\
&= \mathbb{P}\left(\frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x')) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) > \frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) | H_0\right) \\
&= \mathbb{P}\left(\bar{x}'_i > \frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) | H_0\right) \\
&= 1 - F\left(\frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)\right),
\end{aligned}
\tag{6}
$$

where $F(x)$ is the cumulative distribution function of the $\mathcal{N}(\mu_0, \frac{\sigma^2}{n})$ distribution. The numerical application gives $p = 0.99$.

7. By definition, the power of this test is defined as:

$$\begin{aligned}
\Pi(\mu_1) &= \mathbb{P}(\text{Probability of rejecting } H_0|H_1) = \mathbb{P}\left(\Lambda(x) \leq k_\Lambda|\mu_1\right) \\
&= \mathbb{P}\left(\frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(\Lambda(x)) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right) > \frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(k_\Lambda) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)|H_1\right) \\
&= \mathbb{P}\left(\bar{x} > \frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(k_\Lambda) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)|H_1\right) \\
&= 1 - F\left(\frac{\sigma^2}{n(\mu_0-\mu_1)}\left(\ln(k_\Lambda) + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma^2}\right)\right),
\end{aligned} \tag{7}$$

where F is the cdf of the $\mathcal{N}(\mu_1, \frac{\sigma^2}{N})$. The numerical application gives $\Pi(\mu_1) = 1 - \beta = 0.999$.

8. As proven during the course, this test is the uniformly most powerful test.