1. Estimation and testing in the Weibull distribution

   (a) The score statistic is

$$S(\theta) = \frac{n}{\theta} + \sum \log x_i - \sum x_i^\theta \log x_i$$

   and the likelihood equation can therefore be written.

$$\theta = \frac{n}{\sum x_i^\theta \log x_i - \sum \log x_i}. \tag{1}$$

   (b) The second derivative of the log-likelihood function is

$$S'(\theta) = -j(\theta) = -\frac{n}{\theta^2} - \sum x_i^\theta (\log x_i)^2$$

   and this is clearly negative, so a solution of the likelihood equation must necessarily be the MLE.

   (c) The Newton–Raphson iterative step becomes

$$\theta \leftarrow \theta + \frac{S(\theta)}{j(\theta)} = \theta + \frac{n\theta + \theta \sum \log x_i - \theta \sum x_i^\theta \log x_i}{n + \theta^2 \sum x_i^\theta (\log x_i)^2}.$$

   For the method of scoring we need to calculate the Fisher information which involves the integral

$$\mathbf{E}\{X^\theta (\log X)^2\} = \int_0^\infty x^\theta (\log x)^2 \theta x^{\theta-1} e^{-x^\theta}\, dx.$$

   Substituting $u = x^\theta$, $du = \theta x^{\theta-1}$ we get

$$\mathbf{E}\{X^\theta (\log X)^2\} = \theta^{-2} \int_0^\infty u(\log u^2) e^{-u}\, du = \theta^{-2}(\pi^2/6 + \gamma^2 - \gamma)$$

   where $\gamma = -0.5772\ldots$ is Euler's constant. This yields the Fisher information

$$i(\theta) = \frac{n}{\theta^2}(1 + \pi^2/6 + \gamma^2 - \gamma)$$

   and hence the iterative step in the method of scoring becomes

$$\theta \leftarrow \theta + \frac{S(\theta)}{i(\theta)} = \theta + \frac{n\theta + \theta \sum \log x_i - \theta \sum x_i^\theta \log x_i}{n(1 + \pi^2/6 + \gamma^2 - \gamma)}.$$

   Finally, it is tempting to use the equation (1) as a basis for an iteration

$$\theta \leftarrow \frac{n}{\sum_i x_i^\theta \log x_i - \sum \log x_i}$$

   although its convergence properties are not all that clear.

(d) The LRT rejects for large values of

$$\log \frac{f(X;2)}{f(X;1)} = n \log 2 + \sum_i (\log X_i - X_i^2 + X_i).$$

The distribution of this statistic is not explicitly available, so the critical value for the test must be calculated approximately, either referring to the Central Limit Theorem so that

$$T = t(X) = \sum_i (\log X_i - X_i^2 + X_i) \overset{\mathrm{a}}{\sim} \mathcal{N}(n\mu, n\sigma^2),$$

where

$$\mu = \mathbf{E}(\log X - X^2 + X), \quad \sigma^2 = \mathbf{V}(\log X - X^2 + X),$$

or—probably most easily and accurately, at least for moderate sample size n—by calculating a Monte–Carlo p-value as follows:
Under the null hypothesis $\theta = 1$, X is exponentially distributed. Simulate N samples of size n from the exponential distribution,

$$X_i^* = (X_{i1}^*, \ldots, X_{in}^*), i = 1, \ldots, N,$$

for example by letting $X_{ij}^* = -\log U_{ij}$, where U_{ij} are independent and uniform on $(0, 1)$.
Now calculate T above and $T_i^* = t(X_i^*)$ and define the Monte–Carlo p-value $p^* = p_N^*$ for the LRT as

$$p^* = \frac{|\{i \,|\, T_i^* > T\}|}{N}.$$

Now reject H_0 if $p^* < \alpha$.

(e) The maximized LRT for the null hypothesis $H_0 : \theta = 1$ vs. the alternative $H_A : \theta \neq 1$ rejects for large values of

$$\log L(\hat{\theta}) - \log L(1) = n \log \hat{\theta} + (\hat{\theta} - 1) \sum_i \log X_i + \sum_i (X_i^{\hat{\theta}} - X_i).$$

To determine the critical value one can either use a Monte-Carlo procedure similar to the one above or the fact that twice the above statistic has an asymptotic χ^2-distribution with one degree of freedom.

(f) The score test for the null hypothesis $H_0 : \theta = 1$ vs. the alternative $H_A : \theta \neq 1$ rejects for

$$\begin{aligned} \{S(1)\}^2 &= \left(n + \sum \log x_i - \sum x_i \log x_i\right)^2 \\ &> \chi^2(1)_{1-\alpha} i(1) = \chi^2(1)_{1-\alpha}(\pi^2/6 + \gamma^2 - \gamma). \end{aligned}$$

(g) Alternative large sample tests include the χ^2-test which rejects for

$$n(\pi^2/6 + \gamma^2 - \gamma)(\hat{\theta} - 1)^2 > \chi^2_{1-\alpha}$$

and Wald's test which rejects for

$$\frac{n}{\hat{\theta}^2}(\pi^2/6 + \gamma^2 - \gamma)(\hat{\theta} - 1)^2 > \chi^2_{1-\alpha}.$$

The latter has very small power for large alternatives.

2. Comparing Poisson rates Let $X = (X_1, \ldots, X_k)$ be independent and Poisson distributed with parameter λ_i as

$$f(x_i; \lambda_i) = \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i}, x_i = 0, 1, \ldots$$

where $\lambda_i > 0$ are unknown.

(a) The joint density is

$$f(x; \lambda) = \prod_i \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i} = \left(\prod_i \frac{1}{x_i!} \right) e^{\sum x_i \log \lambda_i - \sum \lambda_i}.$$

This is recognized as a canonical exponential family with canonical statistic $t(X) = X = (X_1, \ldots, X_k)$ so the MLE is found by equating the statistic to its expectation. Since $\mathbf{E}(X_i) = \lambda_i$ it holds that

$$\hat{\lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_k) = (X_1, \ldots, X_k).$$

This result could also have been derived easily by direct differentiation.

(b) In this case, the expression for the density reduces to

$$f(x; \lambda) = \left(\prod_i \frac{d_i^{x_i}}{x_i!} \right) e^{\log \alpha \sum x_i - \alpha \sum d_i}.$$

This is again a canonical exponential family, but this time with $t(X) = \sum X_i$ as canonical statistic. Equating the statistic to its expectation yields

$$\sum X_i = \alpha \sum d_i, \quad \hat{\alpha} = \sum X_i / \sum d_i$$

and hence

$$\hat{\hat{\lambda}} = \left\{ \frac{\sum X_i}{\sum d_i} d_1, \ldots, \frac{\sum X_i}{\sum d_i} d_k \right\}$$

as required. Again, the result could also have been obtained directly by differentiation of the log-likelihood function.

(c) The maximized LRT statistic is

$$
\begin{aligned}
-2\log\Lambda &= 2\log L(\hat{\lambda}) - 2\log L(\hat{\hat{\lambda}}) \\
&= 2\left(\sum X_i \log\hat{\lambda}_i - \sum\hat{\lambda}_i - \sum X_i \log\hat{\hat{\lambda}}_i + \sum\hat{\hat{\lambda}}_i\right) \\
&= 2\left\{\sum X_i \log\frac{X_i}{\hat{\alpha}d_i} - \sum(X_i - \hat{\alpha}d_i)\right\}.
\end{aligned}
$$

The maximized LRT follows asymptotically a χ^2-distribution with $k-1$ degrees of freedom.

Remark: Formally, there is no n which tends to infinity. The asymptotic result holds for $\alpha \to \infty$ but strictly speaking this demands another variant of the asymptotic result than the standard one given in the notes.

(d) The Fisher information matrix $i(\lambda)$ for λ is a diagonal matrix with diagonal elements $\mathbf{V}(X_i) = \lambda_i$. The Wald test statistic for H_0 is thus

$$
\begin{aligned}
W &= (\hat{\lambda} - \hat{\hat{\lambda}})^\top i(\hat{\lambda})^{-1}(\hat{\lambda} - \hat{\hat{\lambda}}) \\
&= \sum\frac{(X_i - \hat{\alpha}d_i)^2}{X_i} \\
&= \sum\frac{(\text{observed} - \text{expected})^2}{\text{observed}}.
\end{aligned}
$$

The same remark as above applies to the interpretation of the fact that W has an asymptotic distribution as a χ^2 with $k-1$ degrees of freedom.

(e) The χ^2 test statistic for H_0 is using $i(\hat{\hat{\lambda}})$ instead of $i(\hat{\lambda})$ in the Wald test. This leads to

$$
\begin{aligned}
X^2 &= (\hat{\lambda} - \hat{\hat{\lambda}})^\top i(\hat{\hat{\lambda}})^{-1}(\hat{\lambda} - \hat{\hat{\lambda}}) \\
&= \sum\frac{(X_i - \hat{\alpha}d_i)^2}{\hat{\alpha}d_i} \\
&= \sum\frac{(\text{observed} - \text{expected})^2}{\text{expected}},
\end{aligned}
$$

which is the familiar χ^2 statistic.

(f) The maximized LRT for H_1 under the assumption that H_0 is true is

$$
\begin{aligned}
-2\log\tilde{\Lambda} &= 2\log L(\hat{\hat{\lambda}}) - 2\log L(\tilde{\lambda}) \\
&= 2\left(\sum X_i \log\hat{\hat{\lambda}}_i - \sum\hat{\hat{\lambda}}_i - \sum X_i \log d_i + \sum d_i\right) \\
&= 2\left\{\sum X_i \log\frac{\hat{\alpha}d_i}{d_i} - \sum(\hat{\alpha}d_i - d_i)\right\} \\
&= 2\left\{\log\hat{\alpha}\sum X_i + (1-\hat{\alpha})\sum d_i\right\},
\end{aligned}
$$

where $\tilde{\lambda} = (d_1, \ldots, d_k)$. This LRT has an asymptotic χ^2 distribution with 1 degree of freedom.