Opgave 2

Som bekendt anvendes en logaritmisk transformation ofte for at opnå normalitet af stokastiske variable. I denne opgave anvendes et konkret datamateriale fra sundhedsøkonomi, hvori der betragtes det antal dage 1469 patienter, der er blevet opereret for “større ledoperationer og påsætning af lemmer” har været indlagt. De fleste patienter har været indlagt under to uger heraf 6 kun én dag, mens enkelte har været indlagt meget længere - den længste tid er 158 dage. Figuren viser et histogram over disse data med den indtegnede tæthedsfunktion for den tilpassede logaritmiske normalfordeling.

Den statistiske model er, at y-erne er uafhængige, logaritmisk normalfordelte med parametre μ og σ^2, hvilket betyder, at z = log(y) er normalfordelt med middelværdi μ og varians σ^2.

Det kan vises, at middelværdien af y er E[y] = $\exp\left(\hat{\mu}+\frac{1}{2}\hat{\sigma}^2\right)$og at variansen for y er

$$\mathrm{var}(y) = \exp(2\mu+\sigma^2)(\exp(\sigma^2)-1) = E[y]^2\left(\exp\left(\sigma^2\right) -1\right).$$

Da z = log(y) ~ N(μ, σ^2) er det oplagt, at maksimum likelihood estimatoren for parametrene er givet ved

$$\hat{\mu} = \bar{z} = \frac{1}{n}\sum_{i=1}^{n}\log(y_i) \quad \text{og} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\log\left(y_i\right)-\hat{\mu}\right)^2$$

mens matricen **H** er givet ved $\mathbf{H} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$

For opgavens data er gennemsnittet $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 13.49$, mens parametrene estimeres ved

$\hat{\mu} = 2.44$ og $\hat{\sigma}^2 = 0.32$.

Middelværdien opfattes som en funktion af parametrene μ og σ^2, $g\begin{pmatrix}\mu \\ \sigma^2\end{pmatrix} = \exp\left(\mu+\frac{1}{2}\sigma^2\right)$,

som estimeres ved at indsætte maksimum likelihoodestimatorerne for μ og σ^2.

1) Find den asymptotiske fordeling af middelværdi estimatoren $\exp\left(\hat{\mu}+\frac{1}{2}\hat{\sigma}^2\right)$. Beregn dette estimats i talværdi i eksemplet og angiv talværdien af den approksimative varians.

2) Opstil Wald testet for hypotesen om, at middelværdien er lig med 14, altså $H_0 : \exp\left(\mu + \frac{1}{2}\sigma^2\right) = 14$ og gennemfør testet i eksemplet.

Antag at man anvender gennemsnittet af y-erne som estimator $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ = 13.49 dage.

3) Angiv den asymptotiske fordeling af denne estimator for middelværdien af y og beregn variansen på denne estimator i taleksemplet. Diskuter fordele og ulemper ved de to estimatorer. Er estimatorerne middelrette? Er der forskel på deres varianser?

Under hypotesen $\exp\left(\mu + \frac{1}{2}\sigma^2\right) = 14$ er σ^2 = 2(log(14) - μ) og dermed bliver tæthedsfunktionen for log(y) ~ N(μ, 2(log(14) - μ)) altså :

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2\left(\log(14) - \mu\right)}} \exp\left\{-\frac{1}{2} \frac{\left(\log(y) - \mu\right)^2}{2\left(\log(14) - \mu\right)}\right\}$$

4) Opstil log-likelihoodfunktionen for denne ene parameter μ og vis, at likelihoodligningen bliver en anden gradsligning i den ubekendte μ.