STATISTISCHE METHODEN DER DATENANALYSE

Excersise Sheets

Physik

UNI WIEN

17. November 2023

1 Sheet

1.1 ML Estimator

1.1.1 Calculating the Estimator

The Joint probability density of an exponentially distributed sample is

$$g\left(x_1, \ldots, x_m | \tau\right) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-\frac{x_i}{\tau}} \tag{1.1}$$

The logarithm of this density interpreted as a likelihood function is

$$\begin{aligned} l\left(\tau\right) = \ln g\left(x_1, \ldots, x_n | \tau\right) &= \sum_{i=1}^{n} \ln\left(\frac{1}{\tau} e^{-\frac{x_i}{\tau}}\right) && (1.2) \\ &= \sum_{i=1}^{n} - \ln\left(\tau\right) - \frac{1}{\tau} x_i && (1.3) \end{aligned}$$

Maximizing the chance to draw this particular sample:

$$\begin{aligned} \frac{\partial l}{\partial \tau} &= \sum_{i=1}^{n} -\frac{1}{\tau} + \frac{1}{\tau^2} x_i \overset{!}{=} 0 && (1.4) \\ \Leftrightarrow - n + \sum_{i=1}^{n} \frac{1}{\tau} x_i &= 0 && (1.5) \\ \Leftrightarrow \hat{\tau} &= \frac{1}{n} \sum_{i=1}^{n} x_i && (1.6) \end{aligned}$$

Inserting $s = \sum x_i = 395.25$ and $n = 250$ yields

$$\hat{\tau} = 1.581 \tag{1.7}$$

1.1.2 Showing that the Estimator is efficient

Showing $\hat{\tau}$ is unbiased

$$\begin{aligned} E_\tau\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] &= \frac{1}{n}\sum_{i=1}^{n} E\left[x_i\right] && (1.8) \\ &= \frac{1}{n}\sum_{i=1}^{n} \tau && (1.9) \\ &= \tau && (1.10) \end{aligned}$$

Showing $\hat{\tau}$ is efficient Calculate the Fisher Information

$$
\begin{aligned}
I_\tau &= E\left[-\frac{\partial^2 \ln g\left(x_1,\ldots,x_2|\tau\right)}{\partial\tau^2}\right] && (1.11)\\
&= E\left[-\sum_{i=1}^{n}\frac{1}{\tau^2}-\frac{2}{\tau^3}x_i\right] && (1.12)\\
&= -\sum_{i=1}^{n}\frac{1}{\tau^2}-\frac{2}{\tau^3}E\left[x_i\right] && (1.13)\\
&= -\sum_{i=1}^{n}\frac{1}{\tau^2}-\frac{2}{\tau^2} && (1.14)\\
&= \sum_{i=1}^{n}\frac{1}{\tau^2} && (1.15)\\
&= \frac{n}{\tau^2} && (1.16)
\end{aligned}
$$

Comparing with estimator variance

$$
\begin{aligned}
V\left[\hat{\tau}\right] &= V\left[\frac{1}{n^2}\sum_{i=1}^{n}V\left[x_i\right]\right] && (1.17)\\
&= \frac{\tau^2}{n} && (1.18)
\end{aligned}
$$

Conclusion, estimator is as efficient as it gets.

1.2 Laplace distribution

1.2.1 Expectation Value and Variance

The Laplace distribution density is

$$
f\left(x;m,s\right) = \frac{1}{2s}\exp\left[-\frac{|x-m|}{s}\right] \quad (1.19)
$$

Expectation Value The expecation value of laplace distributed variable X is

$$
\begin{aligned}
E\left[X\right] &= \int_{-\infty}^{\infty}\frac{x}{2s}\exp\left[-\frac{|x-m|}{s}\right]\,dx && (1.20)\\
&= \frac{1}{2s}\int_{-\infty}^{\infty}\left(u+m\right)\exp\left[-\frac{|u|}{s}\right]\,du && (1.21)\\
&= \frac{1}{2s}\int_{-\infty}^{\infty}u\exp\left[-\frac{|u|}{s}\right]\,du+m && (1.22)\\
&= m && (1.23)
\end{aligned}
$$

Variance

$$\begin{aligned}
V[X] &= E\left[(X-m)^2\right] && (1.24)\\
&= \frac{1}{2s}\int_{-\infty}^{\infty}(x-m)^2\exp\left\{-\frac{|x-m|}{s}\right\}dx && (1.25)\\
&= \frac{1}{2s}\int_{-\infty}^{\infty}u^2\exp\left\{-\frac{|u|}{s}\right\}du && (1.26)\\
&= \frac{1}{s}\int_{0}^{\infty}u^2\exp\left\{-\frac{u}{s}\right\}du && (1.27)\\
&= s^3\int_{0}^{\infty}v^2e^{-v}\,dv && (1.28)\\
&= 2s^3 && (1.29)
\end{aligned}$$

1.2.2 Estimators for m and s

For a given sample the joint density is

$$g(x_1,\ldots,x_n|s,m) = \prod_{i=1}^{n}\frac{1}{2s}e^{-\frac{|x_i-m|}{s}} \tag{1.30}$$

The log likelihood function is

$$\begin{aligned}
\ln g &= \sum_{i=1}^{n}\ln\left(\frac{1}{2s}\exp\left\{-\frac{|x_i-m|}{s}\right\}\right) && (1.31)\\
&= \sum_{i=1}^{n}\ln\left(\frac{1}{2s}\right)-\frac{|x_i-m|}{s} && (1.32)\\
&= n\ln\left(\frac{1}{2s}\right)-\frac{1}{s}\sum_{i=1}^{n}|x_i-m| && (1.33)
\end{aligned}$$

The maximum likelihood estimator for m can now be found

$$\begin{aligned}
\frac{\partial\ln g}{\partial s} &= \frac{\partial}{\partial s}\left[n\ln\left(\frac{1}{2s}\right)-\frac{1}{s}\sum_{i=1}^{n}|x_i-m|\right] && (1.34)\\
&= -\frac{n}{s}+\frac{1}{s^2}\sum_{i=1}^{n}|x_i-m| \stackrel{!}{=} 0 && (1.35)\\
\Rightarrow \hat{s} &= \frac{1}{n}\sum_{i=1}^{n}|x_i-m| && (1.36)
\end{aligned}$$

The maximum likelihood estimator for s can also be found

$$\frac{\partial \ln g}{\partial m} = \frac{\partial}{\partial m}\left[n \ln\left(\frac{1}{2s}\right) - \frac{1}{s}\sum_{i=1}^{n}|x_i - m|\right] \tag{1.37}$$

$$= -\frac{1}{s}\sum_{i=1}^{n}\frac{\partial}{\partial m}|x_i - m| \tag{1.38}$$

$$= -\frac{1}{s}\sum_{i=1}^{n}\operatorname{sgn}(x - m) \tag{1.39}$$

1.3 Survey

The multimodal distribution is defined as

$$f(n_A, n_B, n_C, n_D | p_A, p_B, p_C, p_D) = n! \prod_{i=A,B,C,D}\frac{1}{n_i!}p_i^{n_i} \tag{1.40}$$

with the log density

$$\ln f = \ln\left(n! \prod_{i=A,B,C,D}\frac{1}{n_i!}p_i^{n_i}\right) \tag{1.41}$$

$$= \ln(n!) + \sum_{i=A,B,C,D}\ln\left(\frac{1}{n_i!}\right) + \ln\left(p_i^{n_i}\right) \tag{1.42}$$

$$= \ln(n!) + \sum_{i=A,B,C,D}\ln\left(\frac{1}{n_i!}\right) + n_i \ln(p_i) \tag{1.43}$$

Now the estimators for the voter shares can be found with

$$\frac{\partial \ln f}{\partial p_i} = \frac{n_i}{p_i} \tag{1.44}$$