Beräkningar inlämningsuppgift 1

Erik Ödmann, David Carlsson

Exponentialfördelningen

Täthetsfunktionen för exponentialfunktionen är

$$p_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Likelihood funktionen för en observation blir då

$$L(\lambda) = \lambda e^{-\lambda x}$$

Vi applicerar den naturliga logaritmen på likelihood funktionen vilket ger oss log-likelihood funktionen för en observation

$$\begin{aligned} l(\lambda) &= \ln(\lambda e^{-\lambda x}) \\ &= \ln(\lambda) + \ln(e^{-\lambda x}) \\ &= \ln(\lambda) - \lambda x \ln(e) \\ &= \ln(\lambda) - \lambda x \end{aligned}$$

Log-likelihood funktionen för hela urvalet får vi genom att ta summan över log-likehood funktionen för en observation

$$\begin{aligned} l_n(\lambda) &= \sum_{i=1}^{n} (\ln(\lambda) - \lambda x_i) \\ &= n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \end{aligned}$$

Nästa steg är att derivera likelihood funktionen för urvalet

$$\begin{aligned} l'_n(\lambda) &= \frac{d}{d\lambda}(n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i) \\ &= n\lambda^{-1} - \sum_{i=1}^{n} x_i \\ &= \frac{n}{\lambda} - n\bar{x} \\ &= n(\lambda^{-1} - \bar{x}) \end{aligned}$$

ML skattningen ges utav att lösa ekvationen

$$
\begin{aligned}
l_n'(\lambda) &= n\lambda^{-1} - \sum_{i=1}^{n} x_i = 0 \\
\Leftrightarrow \sum_{i=1}^{n} x_i &= \frac{n}{\lambda} \\
\Leftrightarrow \hat{\lambda} &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}
$$

För att beräkna fisherinformationen behöver vi andraderivatan utav log-likelihood funktionen som ges utav

$$
\begin{aligned}
l_n''(\lambda) &= \frac{d}{d\lambda}(n\lambda^{-1} - \sum_{i=1}^{n} x_i) \\
&= -n\lambda^{-2}
\end{aligned}
$$

Fisherinformationen för urvalet blir då

$$
\begin{aligned}
I_n(\lambda) &= -E[l_n''(\lambda)] \\
&= -E[-n\lambda^{-2}] \\
&= n\lambda^{-2}
\end{aligned}
$$

Vi kan nu beräkna medelfelet för ML-skattningen som

$$
\begin{aligned}
Sd(\hat{\lambda}) &= I_n(\hat{\lambda})^{-1/2} \\
&= (n\hat{\lambda}^{-2})^{-1/2} \\
&= \hat{\lambda} n^{-1/2}
\end{aligned}
$$

Score-testet

Vi börjar med att härleda konfidensintervallet baserat på score-testet

$$
T_{\text{score}} = \frac{l_n'(\lambda_0)}{\sqrt{I_n(\lambda_o)}} \approx N(0,1)
$$

Från tidigare beräkningar vet vi att detta är lika med

$$
\begin{aligned}
T_{\text{score}} &= \frac{n(\lambda_0^{-1} - \bar{x})}{\sqrt{n\lambda_0^{-2}}} \\
&= \frac{\lambda_0}{\sqrt{n}} \times n(\frac{1}{\lambda_0} - \bar{x}) \\
&= \sqrt{n}(1 - \bar{x}\lambda_0)
\end{aligned}
$$

Vi kan nu härleda $100(1-\alpha)\%$-igt konfidensintervall för λ_0

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < T < z_{\alpha/2}) \\
&= P(-z_{\alpha/2} < \sqrt{n}(1-\bar{x}\lambda_0) < z_{\alpha/2}) \\
&= P(-z_{\alpha/2}\frac{1}{\sqrt{n}} < 1-\bar{x}\lambda_0 < z_{\alpha/2}\frac{1}{\sqrt{n}}) \\
&= P(-1-z_{\alpha/2}\frac{1}{\sqrt{n}} < -\bar{x}\lambda_0 < -1+z_{\alpha/2}\frac{1}{\sqrt{n}}) \\
&= P(-\frac{1}{\bar{x}}-z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}} < -\lambda_0 < -\frac{1}{\bar{x}}+z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}}) \\
&= P(\frac{1}{\bar{x}}-z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}} < \lambda_0 < \frac{1}{\bar{x}}+z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}})
\end{aligned}
$$

Vilket ger oss ett konfidensintervall för λ_0 baserat på score-testet

Wald-testet

Här ges tesstatistikan utav

$$T_{\text{wald}} = \frac{\hat{\lambda}-\lambda_0}{Sd(\hat{\lambda})} \approx N(0,1)$$

Från tidigare beräkningar vet vi att vi kan skriva detta som

$$T_{\text{wald}} = \frac{\hat{\lambda}-\lambda_0}{\sqrt{\hat{\lambda}n^{-1/2}}} \approx N(0,1)$$

Vi kan nu härleda $100(1-\alpha)$%-igt konfidensintervall för λ_0

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < T < z_{\alpha/2}) \\
&= P(-z_{\alpha/2} < \frac{\hat{\lambda}-\lambda_0}{\sqrt{\hat{\lambda}n^{-1/2}}} < z_{\alpha/2}) \\
&= P(-z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}} < \hat{\lambda}-\lambda_0 < z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}}) \\
&= P(-\hat{\lambda}-z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}} < -\lambda_0 < -\hat{\lambda}+z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}}) \\
&= P(\hat{\lambda}-z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}} < \lambda_0 < \hat{\lambda}+z_{\alpha/2}\sqrt{\hat{\lambda}n^{-1/2}})
\end{aligned}
$$

Eftersom att $\hat{\lambda} = \frac{1}{\bar{x}}$ får vi

$$1-\alpha = P(\frac{1}{\bar{x}}-z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}} < \lambda_0 < \frac{1}{\bar{x}}+z_{\alpha/2}\frac{1}{\bar{x}\sqrt{n}})$$

Vilket ger oss ett konfidensintervall för λ_0 baserat på wald-testet

Resultat

I figuren nedan ser vi resultatet för exponentialfördelningen.

Table 1: Resultatet av testerna på exponentialfördelningen

n	p	Score-testet	Wald-testet
10	0.1	0.0457	0.0457
50	0.1	0.0467	0.0467
100	0.1	0.0490	0.0490
10	1.0	0.0461	0.0461
50	1.0	0.0451	0.0451
100	1.0	0.0505	0.0505
10	10.0	0.0445	0.0445
50	10.0	0.0511	0.0511
100	10.0	0.0499	0.0499

Binomialfördelningen

Täthetsfunktionen för binomialfördelninge är

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = \text{Antal utfall}$$

Likelihood funktionen för hela urvalet ges utav

$$L_n(p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Log-likelihood för hela urvalet kan vi då beräkna till

$$\begin{aligned} l_n(p) &= \ln(\binom{n}{k} p^k (1-p)^{n-k}) \\ &= \ln(\binom{n}{k}) + \ln(p^k) + \ln((1-p)^{n-k}) \\ &= \ln(\binom{n}{k}) + k\ln(p) + (n-k)\ln(1-p) \end{aligned}$$

Första derivatan utav log-likelihood funktionen för hela urvalet. Vi vet att $k = \sum_{i=1}^{n} x_i = n\bar{x}$ vilket ger oss

$$\begin{aligned} l_n'(p) &= \frac{d}{dp}(\ln(\binom{n}{k}) + k\ln(p) + (n-k)\ln(1-p)) \\ &= kp^{-1} - (n-k)(1-p)^{-1} \\ &= \frac{k(1-p) - p(n-k)}{p(1-p)} \\ &= \frac{k - pn}{p(1-p)} \\ &= \frac{n}{p(1-p)}(\bar{x} - p) \end{aligned}$$

ML-skattninges ges utav att lösa ekvationen för p

$$\begin{aligned} l_n'(p) &= kp^{-1} - (n-k)(1-p)^{-1} = 0 \\ &\Leftrightarrow kp^{-1} = (n-k)(1-p)^{-1} \\ &\Leftrightarrow (1-p)k = p(n-k) \\ &\Leftrightarrow k - pk = pn - pk \\ &\Leftrightarrow k = pn \\ &\Leftrightarrow \hat{p} = \frac{k}{n} = \bar{x} \end{aligned}$$

För att beräkna fisherinformationen behöver vi andraderivatan utav log-likelihood funktionen

$$\begin{aligned} l_n''(p) &= \frac{d}{dp}(kp^{-1} - (n-k)(1-p)^{-1}) \\ &= -kp^{-2} - (n-k)(1-p)^{-2} \end{aligned}$$

Fisherinformationen för hela urvalet blir då

$$
\begin{aligned}
I_n(p) &= -E[l_n''(p)] \\
&= -E[-kp^{-2} - (n-k)(1-p)^{-2}] \\
&= -E[\frac{-k+2kp-np^2}{p^2(1-p)^2}]
\end{aligned}
$$

Eftersom $E[k] = np$ får vi

$$
\begin{aligned}
I_n(p) &= \frac{np + np^2}{p^2(1-p)^2} \\
&= \frac{n}{p(1-p)}
\end{aligned}
$$

Medelfelet för ML-skattningen ges då utav

$$
\begin{aligned}
Sd(\hat{p}) &= I_n(\hat{p})^{-1/2} \\
&= \sqrt{\hat{p}(1-\hat{p})/n}
\end{aligned}
$$

Score-testet

Vi börjar härleda konfidensintervallet baserat på score-testet. Teststatistikan ges utav

$$
T_{\text{score}} = \frac{l_n'(p_0)}{\sqrt{I_n(p_o)}} \approx N(0,1)
$$

Första derivatan av likelihood funktionen samt fisherinformationen beräknade vi tidigare vilket ger oss

$$
\begin{aligned}
T_{\text{score}} &= \frac{l_n'(p_0)}{\sqrt{I_n(p_o)}} \\
&= \frac{\frac{n}{p_0(1-p_0)}(\bar{x}-p_0)}{\sqrt{\frac{n}{p_0(1-p_0)}}} \\
&= \sqrt{\frac{p_0(1-p_0)}{n}} \times \frac{n}{p_0(1-p_0)}(\bar{x}-p_0) \\
&= \sqrt{\frac{n}{p_0(1-p_0)}}(\bar{x}-p_0) \\
&= \frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}} \approx N(0,1)
\end{aligned}
$$

Vi kan nu härleda $100(1-\alpha)\%$-igt konfidensintervall för p_0

$$
\begin{aligned}
1-\alpha &= P(-z_{\alpha/2} < T < z_{\alpha/2}) \\
&= P(-z_{\alpha/2} < \frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}} < z_{\alpha/2}) \\
&= P(-z_{\alpha/2}\sqrt{p_0(1-p_0)/n} < \bar{x}-p_0 < z_{\alpha/2}\sqrt{p_0(1-p_0)/n}) \\
&= P(-\bar{x}-z_{\alpha/2}\sqrt{p_0(1-p_0)/n} < -p_0 < -\bar{x}+z_{\alpha/2}\sqrt{p_0(1-p_0)/n}) \\
&= P(\bar{x}-z_{\alpha/2}\sqrt{p_0(1-p_0)/n} < p_0 < \bar{x}+z_{\alpha/2}\sqrt{p_0(1-p_0)/n})
\end{aligned}
$$

Vilket ger oss ett konfidensintervall för p_0 baserat på score-testet

Wald-testet

Här ges tesstatistikan utav

$$T_{\text{wald}} = \frac{\hat{p} - p_0}{Sd(\hat{p})} \approx N(0,1)$$

Från tidigare beräkningar vet vi att vi kan skriva detta som

$$T_{\text{wald}} = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx N(0,1)$$

Vi kan nu härleda $100(1-\alpha)\%$-igt konfidensintervall för p_0

$$\begin{aligned}
1 - \alpha &= P(-z_{\alpha/2} < T < z_{\alpha/2}) \\
&= P(-z_{\alpha/2} < \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} < z_{\alpha/2}) \\
&= P(-z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < \hat{p} - p_0 < z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}) \\
&= P(-\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < -p_0 < -\hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}) \\
&= P(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < p_0 < \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n})
\end{aligned}$$

Eftersom $\hat{p} = \bar{x}$ blir detta

$$1 - \alpha = P(\bar{x} - z_{\alpha/2}\sqrt{\bar{x}(1-\bar{x})/n} < p_0 < \bar{x} + z_{\alpha/2}\sqrt{\bar{x}(1-\bar{x})/n})$$

Vilket ger oss ett konfidensintervall för p_0 baserat på wald-testet

Resultat

I figuren nedan ser vi resultatet för binomialfördelningen.

Table 2: Resultatet av testerna på binomialfördelningen

n	Lambda	Score-testet	Wald-testet
10	0.1	0.0718	0.3563
50	0.1	0.0282	0.1221
100	0.1	0.0625	0.0634
10	0.3	0.0755	0.1652
50	0.3	0.0423	0.0671
100	0.3	0.0632	0.0499
10	0.5	0.0209	0.1054
50	0.5	0.0643	0.0643
100	0.5	0.0545	0.0545

Vi kan se att för score-testet så ligger förkastningsandelen kring 5% vilket var vår signifikansnivå. Vi kan också notera att vi ser en tendens av att andelen närmar sig 5% då stickprovsstorleken blir större.

Slutsats