STATISTISCHE METHODEN DER DATENANALYSE

Excersise Sheets

Physik

UNI WIEN

24. November 2023

1 Sheet

1.1 Bernoulli

1.1.1 Clopper and Pearson confidence interval

A Bernoulli experiment is repeated $n = 200$ times with $k = 121$ sucesses. Calculate the symmetric 95% interval for the parameter p. The Interval boundaries can be calculated with the inverse beta distribution.

$$G_1(k) = \beta\left(\frac{\alpha}{2}; k, n-k+1\right) \approx 0.534 \tag{1.1}$$

$$G_2(k) = \beta\left(\frac{1-\alpha}{2}; k+1, n-k\right) \approx 0.673 \tag{1.2}$$

1.1.2 Approximation by normal distribution

bootstrap

$$G_1(k) = \hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.537 \tag{1.3}$$

$$G_2(k) = \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.673 \tag{1.4}$$

robust

$$G_1(k) = \hat{p} - z_{1-\alpha/2}\frac{1}{2\sqrt{n}} \approx 0.536 \tag{1.5}$$

$$G_2(k) = \hat{p} + z_{1-\alpha/2}\frac{1}{2\sqrt{n}} \approx 0.674 \tag{1.6}$$

1.1.3 Agresti-Coull

Choose $k' = k + (z_{1-\alpha/2})^2$, $n' = n + (z_{1-\alpha/2})^2$ and $p' = k'/n'$.

$$G_1(k) = p' - z_{1-\alpha/2}\sqrt{\frac{p'(1-p')}{n}} \approx 0.546 \tag{1.7}$$

$$G_2(k) = p' + z_{1-\alpha/2}\sqrt{\frac{p'(1-p')}{n}} \approx 0.679 \tag{1.8}$$

1.2 Biased and unbiased Estimators for uniform distribution interval borders

The joint probability of the sample ist

$$g(X_1, \ldots, X_N | a, b) = \prod_n^N \frac{1}{b-a} \cdot I_{[a,b]}(X_n) \tag{1.9}$$

The ML estimator is

$$\hat{a} = \arg\max_a \prod_n^N \frac{1}{b-a} \cdot I_{[a,b]}(X_n) \tag{1.10}$$

This would not take a minimum if not for the constraint

$$\hat{a} \leq \min_n \{X_n\} \tag{1.11}$$

Therefore

$$\hat{a} = \min_n \{X_n\} \tag{1.12}$$

Similarly

$$\hat{b} \geq \max_n \{X_n\} \tag{1.13}$$

and

$$\hat{b} = \max_n \{X_n\} \tag{1.14}$$

Showing the estimators are biased To show the estimators are biased, calculate their distribution functions:

$$F_{\hat{a}}(x) = P(\hat{a} \leq x) = 1 - \prod_n P(X_n > x) \tag{1.15}$$

$$= 1 - P^N(X > x) \tag{1.16}$$

$$= 1 - \left(\frac{b-x}{b-a}\right)^N \tag{1.17}$$

The density is

$$\frac{\partial F_{\hat{a}}}{\partial x} = \frac{N}{b-a}\left(\frac{b-x}{b-a}\right)^{N-1} \tag{1.18}$$

Now the expectation value is

$$E(\hat{a}) = \int_a^b x \frac{n}{b-a} \left(\frac{b-x}{b-a}\right)^{N-1} dx \tag{1.19}$$

$$= \left[-\left(\frac{b-x}{b-a}\right)^N x\right]_a^b + \int_a^b \left(\frac{b-x}{b-a}\right)^N dx \tag{1.20}$$

Here is

$$\left[-\left(\frac{b-x}{b-a}\right)^N x\right]_a^b = \left[-\underbrace{\left(\frac{b-b}{b-a}\right)^N}_{=0} b + \underbrace{\left(\frac{b-a}{b-a}\right)^N}_{=1} a\right] = a \tag{1.21}$$

and

$$\int_a^b \left(\frac{b-x}{b-a}\right)^N dx = \left[-\frac{b-a}{N+1}\left(\frac{b-x}{b-a}\right)^{N+1}\right]_a^b \tag{1.22}$$

$$= -\frac{b-a}{N+1}\left[\underbrace{\left(\frac{b-b}{b-a}\right)^{N+1}}_{=0} - \underbrace{\left(\frac{b-a}{b-a}\right)^{N+1}}_{=1}\right] \tag{1.23}$$

$$= \frac{b-a}{N+1} \tag{1.24}$$

Together

$$E(\hat{a}) = a + \frac{b-a}{N+1} \tag{1.25}$$

Similarly for $\hat{b}$:

$$F_{\hat{b}}(x) = P\left(\hat{b} \le x\right) \tag{1.26}$$

$$= \prod_n P(X_n \le x) \tag{1.27}$$

$$= P^N (X \le x) \tag{1.28}$$

$$= \left(\frac{x-a}{b-a}\right)^N \tag{1.29}$$

The density is

$$\frac{\partial F_{\hat{b}}}{\partial x} = \frac{N}{b-a}\left(\frac{x-a}{b-a}\right)^{N-1} \tag{1.30}$$

The expectation value of $\hat{b}$ is

$$E(\hat{b}) = \int_a^b x \frac{N}{b-a}\left(\frac{x-a}{b-a}\right)^{N-1} dx \tag{1.31}$$

$$= \left[\left(\frac{x-a}{b-a}\right)^{N} x\right]_a^b - \int_a^b \left(\frac{x-a}{b-a}\right)^{N} dx \tag{1.32}$$

where

$$\left[\left(\frac{x-a}{b-a}\right)^{N} x\right]_a^b = \left[\underbrace{\left(\frac{b-a}{b-a}\right)^{N}}_{=1} b - \underbrace{\left(\frac{a-a}{b-a}\right)^{N}}_{=0} a\right] = b \tag{1.33}$$

and

$$\int_a^b \left(\frac{x-a}{b-a}\right)^{N} dx = \left[\frac{b-a}{N+1}\left(\frac{x-a}{b-a}\right)^{N+1}\right]_a^b \tag{1.34}$$

$$= \frac{b-a}{N+1}\left[\underbrace{\left(\frac{b-a}{b-a}\right)^{N+1}}_{=1} - \underbrace{\left(\frac{a-a}{b-a}\right)^{N+1}}_{=0}\right] \tag{1.35}$$

$$= \frac{b-a}{N+1} \tag{1.36}$$

Together

$$E(\hat{b}) = b - \frac{b-a}{N+1} \tag{1.37}$$

Conclusion Both estimators are biased, but asymptotically unbiased. This makes intuitivly sense. We would like to correct the estimators $\hat{a}$ and $\hat{b}$.

$$\hat{a}_c = \hat{a} - \frac{b-a}{N+1} \tag{1.38}$$

$$\hat{b}_c = \hat{b} + \frac{b-a}{N+1} \tag{1.39}$$

but a and b are not a priori known. Lets just guess an estimator.

$$\begin{aligned}
E(\hat{b}-\hat{a}) &= E(\hat{b}) - E(\hat{a}) && (1.40)\\
&= b - a - \frac{2\,(b-a)}{N+1} && (1.41)\\
&= \frac{(b-a)\,(N+1)}{N+1} - \frac{2\,(b-a)}{N+1} && (1.42)\\
&= (b-a)\,\frac{N-1}{N+1} && (1.43)\\
\Leftrightarrow E\left(\frac{N+1}{N-1}\left(\hat{b}-\hat{a}\right)\right) &= b-a && (1.44)
\end{aligned}$$

Not quite, but this means, that

$$\frac{N+1}{N-1}\left(\hat{b}-\hat{a}\right) \qquad (1.45)$$

is unbiased estimator for $b-a$. Inserting the newly found estimator yields

$$\begin{aligned}
\hat{a}_c &= \hat{a} - \frac{N+1}{N-1}\frac{\hat{b}-\hat{a}}{N+1} && (1.46)\\
&= \frac{\hat{a}\,(N-1)}{N-1} - \frac{\hat{b}-\hat{a}}{N-1} && (1.47)\\
&= \frac{\hat{a}N-\hat{b}}{N-1} && (1.48)
\end{aligned}$$

$$\begin{aligned}
\hat{b}_c &= \hat{b} + \frac{N+1}{N-1}\frac{\hat{b}-\hat{a}}{N+1} && (1.49)\\
&= \frac{\hat{b}\,(N-1)}{N-1} + \frac{\hat{b}-\hat{a}}{N-1} && (1.50)\\
&= \frac{\hat{b}N-\hat{a}}{N-1} && (1.51)
\end{aligned}$$

and we are done.