Jackknife: introduction (1/2)

The Jackknife, like the Bootstrap, is a ressampling method to estimate the bias and the standard error of an estimator.

To estimate the bias for example, the idea is to compute estimates on subsamples of size $n-1$ and then average them. This average is then compared to the estimate computed on the whole sample.

The Jackknife method bears ressemblance to leave-one-out Cross-validation.

Jackknife: introduction (2/2)

Suppose that we observe some data $x_1, ..., x_n$ and an estimator $\hat{\theta}$ is computed on the data, for example the mean. The Jackknife is based on estimates computed on *Jackknife subsamples* of size $n-1$.

Let $\mathbf{x}_{(i)} = (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ be the sample for which the i^{th} observation is omitted and let the estimator $\hat{\theta}_{(i)}$ be the i^{th} jackknife replication of $\hat{\theta}$, the estimator computed on $\mathbf{x}_{(i)}$. Then the Jackknife estimate of the bias of $\hat{\theta}$ is given by

$$\widehat{bias}_J = (n-1)(\hat{\theta}_{(.)} - \hat{\theta})$$

where $\hat{\theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{(i)}$ The Jackknife estimate of the standard error of $\hat{\theta}$ is given by

$$\widehat{se}_J = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(.)}\right)^2}$$

Working example (1/8)

Example: Compute the Jackknife and Bootstrap bias of the Maximum Likelihood Estimator for θ in an Exponential model, given by $\hat{\theta}_{MLE} = 1/\bar{x}$. To do so, generate a sample of size $n = 20$ from an Exponential distribution with parameter $\theta = 3$. Compare the result with the theoretical bias which is $bias_{\hat{\theta}_{MLE}} = \frac{\theta}{n-1}$. Then compute the standard error of $\hat{\theta}_{MLE}$.

	x
1	0.27
2	0.43
3	1.18
4	0.25
5	0.27
6	0.61
7	0.03
8	0.14
9	0.09
10	0.31
11	0.14
12	0.09
13	0.16
14	0.72
15	0.17
16	0.11
17	0.15
18	0.39
19	0.21
20	0.13

Working example (2/8)

Maximum likelihood estimation of an exponential model:

$$\mathcal{L}(\lambda \mid \mathbf{x}) = \prod_{i=1}^{n} f_\lambda(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\left(\lambda \sum_{i=1}^{n} x_i\right)}$$

$$l(\lambda \mid \mathbf{x}) = ln(\mathcal{L}(\lambda \mid \mathbf{x})) = n\ ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial ln\mathcal{L}(\lambda \mid \mathbf{x})}{\partial \lambda} = n\frac{1}{\lambda} - \sum_{i=1}^{n} x_i$$

Setting the derivative equal to 0 and solving for λ then yields

$$n\frac{1}{\lambda} - \sum_{i=1}^{n} x_i = 0 \quad \Leftrightarrow \quad n\frac{1}{\lambda} = \sum_{i=1}^{n} x_i \quad \Leftrightarrow \hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

Working example (3/8)

We know that $1/\overline{x}$ is the point where the log-likelihood function reaches its maximum value since the second derivative of the function is negative, i.e.

$$\frac{\partial^2 ln\mathcal{L}(\lambda \mid \mathbf{x})}{\partial \lambda^2} = -\frac{n}{\lambda^2} \qquad (< 0)$$

An estimator $\hat{\theta}$ is said to be unbiased for θ if the following equality holds

$$E[\hat{\theta}] = \theta$$

.

Working example (4/8)

We note that if $x_1, x_2, ..., x_n$ are iid $Exp(\lambda)$ r.v., then the sum $\sum_{i=1}^n x_i \sim Gamma(n, \lambda)$. Therefore we define $y = \sum_{i=1}^n x_i$ and use the pdf of y.

$$E[\hat{\lambda}_{MLE}] = E\left[\frac{1}{\bar{x}}\right] = E\left[\frac{n}{\sum_{i=1}^n x_i}\right] = E\left[\frac{n}{y}\right] = n\ E\left[\frac{1}{y}\right]$$

$$= n \int_0^\infty \frac{1}{y}\ f_\lambda(y)\ dy$$

$$= n \int_0^\infty \frac{1}{y}\ \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}\ dy$$

$$= n\ \frac{\lambda^n}{\Gamma(n)} \int_0^\infty y^{n-2}\ e^{-\lambda y}\ dy \qquad (u = \lambda y \Leftrightarrow y = u/\lambda, \quad du = u' dy = \lambda dy \Leftrightarrow dy = du/\lambda)$$

$$= n\ \frac{\lambda^n}{\Gamma(n)} \int_0^\infty \left(\frac{u}{\lambda}\right)^{n-2} e^{-u}\ \frac{du}{\lambda}$$

Working example (5/8)

$$
\begin{aligned}
&= n \; \frac{\lambda^n}{\Gamma(n)} \; \frac{1}{\lambda} \; \frac{1}{\lambda^{n-2}} \int\limits_0^\infty u^{n-2} \; e^{-u} \; du \\
&= n \; \frac{\lambda^n}{\Gamma(n)} \; \frac{1}{\lambda} \; \frac{1}{\lambda^{n-2}} \int\limits_0^\infty u^{n-2} \; e^{-u} \; du \qquad\qquad \text{where} \int\limits_0^\infty u^{n-2} \; e^{-u} \; du = \Gamma(n-1) \\
&= n \; \frac{\lambda^n}{(n-1)\;\Gamma(n-1)} \; \frac{1}{\lambda} \; \frac{1}{\lambda^{n-2}} \; \Gamma(n-1) \\
&= \frac{n}{n-1} \; \lambda
\end{aligned}
$$

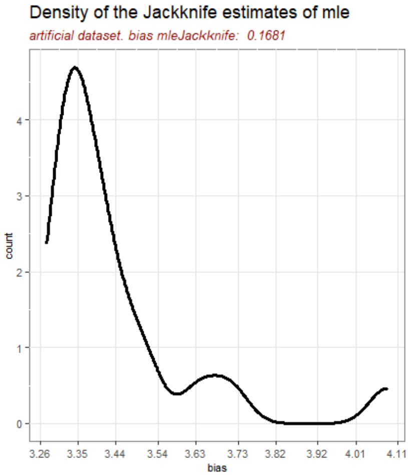
We therefore conclude that $\hat{\lambda}_{MLE}$ is a **biased estimator** for λ. The bias is then

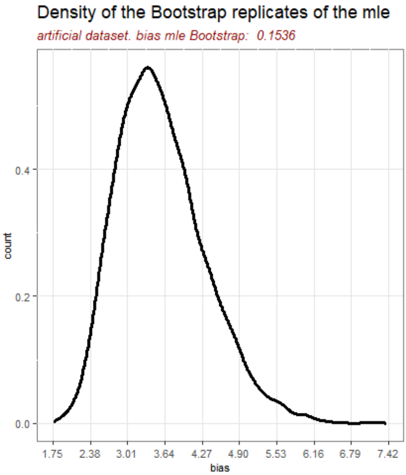
$$
b_\lambda(\hat{\lambda}_{MLE}) = E[\hat{\lambda}_{MLE}] - \lambda = \frac{n}{n-1} \; \lambda - \lambda = \frac{\lambda}{n-1}
$$

Working example (6/8)

```
# 2.1 Jackknife estimate of the bias
theta.jack = numeric(n)
for(i in 1:n) {
  theta.jack[i] <- 1 / mean(sample1$x[-i])
}

# estimate of the bias
bias.J <- (n-1) * (mean(theta.jack) - lambda.hat)
bias.J # [1] 0.1681113

# 3.2 Bootstrap estimate of the bias
set.seed(2023)
B = 10000                                    # B: set the number of
     bootstrap replicates
bootstrap_object <- matrix(rep(0, B*length(sample1$x)),  nrow = B)
theta.boot <- numeric(B)

# perform the bootstrap using the function sample() with replacement
for(i in 1:B) {

  bootstrap_object[i,] <- sample(sample1$x, size = length(sample1$x), replace =
     TRUE)
  theta.boot[i] <- 1 / mean(bootstrap_object[i, ])
}

# estimate of the bias
bias.B <- mean(theta.boot - lambda.hat)
bias.B # [1] 0.1535831

# 4. theoretical bias
lambda / (n-1) # [1] 0.1578947
```

Working example (7/8)





Working example (8/8)

The bootstrap estimate of the bias is closer to the theoretical bias compared to the Jackknife estimate. For more details, see the remarks in the next slide.

	bias.Jackknife	bias.Bootstrap	bias.theoretical
1	0.1681	0.1539	0.1579

Remarks

(i) In general, the Boostrap performs better than the Jackknife since it uses more Bootstrap samples (than Jackknife samples).

(ii) The Bootstrap is more computationally intensive.

(iii) The Jackknife can fail if the statistic $\hat{\theta}$ is not smooth, like for example the median or quantiles in general. In that case, we will prefer the Bootstrap to estimate the bias.

(iv) Jackknife can also produce estimates that are quite different from the Bootstrap estimates on small samples.

References

Rizzo, M.L. (2019). Statistical Computing with R, Second Edition (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429192760

Efron, B. and Tibshirani, R. J. (1993), An introduction to the Bootstrap, Chapman Hall.

The R Project for Statistical Computing: https://www.r-project.org/

Python: https://www.python.org/

course notes