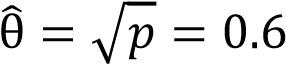
Stat 427/627 Statistical Machine Learning Quiz 6 Jun Lu. American University

Quiz 6. Jack knife and Bootstrap

Name: Lisa Liubovich Attempt (circle one): BEFORE AFTER

Given 25 samples of water, 16 samples are polluted with arsenic (0) and 9 samples are clean (1). Thus, the proportion of clean samples, p, is estimated by 𝑝̂ = 9/25 = 0.36.

However, the standard approach for water sampling is to estimate . An obvious estimator of 𝜃 is

̂ . But this θ̂ is biased, i.e., E ̂ ̂ .

Starting with θ̂ = √𝑝̂ , compute the jackknife estimator θ̂𝐽𝐾.

1. Delete one at a time and compute 𝜃̂(−𝑖),   𝑖 = 1, 2, … , 25, of the obtained data. Note that, given the sample, 16 times the deleted sample is arsenic (0), 9 times the deleted sample is clean (1).

When a clean sample is removed (xi = 1)

𝑝̂(-i) = 8/24 = 1/3 🡪 θ̂(-i) = square root of 1/3 = 0.57735026919

When a polluted sample is removed (xi = 0)

𝑝̂(-i) = 9/24 = 3/8 🡪 θ̂(-i) = square root of 3/8 = 0.61237243569

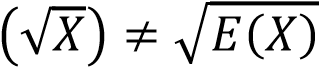
θ̂(.)= (9\*0.57735026919 + 16 \* 0.61237243569)/25 = 14.9941113937/25 = 0.59976445575

1. Compute the Jackknife estimate 𝜃̂𝐽𝐾.

𝜃̂𝐽𝐾 = n\*θ̂ - (n-1) θ̂(.) = 25 \* 0.60 – 24 \* 0.59976445575 = 0.605653062

1. Compute the estimated bias of 𝜃̂.

Bias = θ̂ - 𝜃̂𝐽𝐾 = 0.60 - 0.605653062 = -0.005653062

1. **Stat-627 only**. Use the following distribution to illustrate that 𝐸. Recall that: 𝐸(𝑋) = ∑(𝑥 ⋅ 𝑃(𝑋 = 𝑥)).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | P(X = x) | x ⋅ P(X = x) | √𝑋 | √𝑥 ⋅ P(X = x) |
| 0 | 0.5 | 0 \* 0.5 = 0 | 0 | 0 |
| 9 | 0.5 | 9\* 0.5 = 4.5 | 3 | 3\*0.5 = 1.5 |
|  |  | 𝐸(𝑋) = 0 + 4.5 = 4.5  √ 2.12 |  | 𝐸(√𝑋) = 0 + 1.5 = 1.5 |