Sumi Choudhury HW #3

SDGB 7844 November 2, 2016

Solutions HW 3

1) See code; based on the capture-recapture method for a population size N = 5,000 when n1 = 100 and n2 = 100 using the sample() function, in this trial m2 = 1 and N^LP = 10,000.

2) 1,000 iterations for Lincoln-Peterson estimator; N = 5,000 where n1 = 100 and n2 = 100:

> output <- capture\_recapture\_funct(N, n1, n2, iterations)

> sim <- output[[1]]

> head(sim)

m2 N\_LP

1 3 3333.333

2 3 3333.333

3 1 10000.000

4 0 Inf

5 2 5000.000

6 1 10000.000

> tail(sim)

m2 N\_LP

995 3 3333.333

996 0 Inf

997 3 3333.333

998 2 5000.000

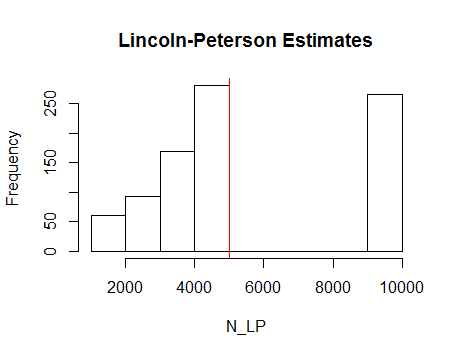
999 2 5000.000

1000 1 10000.000

> N <- output[[2]]

> N

[1] 5000



3) Percent of total infinite values in this iteration is 13.2%. This occurs whenever there is an m2 of 0, which means that the number of tagged individuals at the time of recapture is equal to 0. This can occur because the Lincoln–Peterson estimator is asymptotically unbiased as the sample size approaches infinity, but is biased at small sample sizes.

> total\_inf\_values <- length(which(sim$N\_LP == "Inf"))

> total\_inf\_values

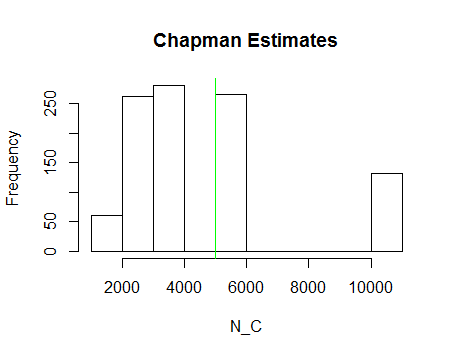
[1] 132

> P <- total\_inf\_values \* 100 / iterations

> P

[1] 13.2

4) Histogram for the Chapman estimator:



5) Bias of the Lincoln-Peterson estimator:

> bias\_N\_LP = thetacap\_N\_LP / iterations - N

> bias\_N\_LP

[1] Inf

Bias of the Chapman estimator:

> bias\_N\_C = thetacap\_N\_C / iterations - N

> bias\_N\_C

[1] -631.3409

Neither estimator is unbiased when n1, n2 = 100; on average neither of the estimators equal the true population value.

6) Based on my findings thus far, the Chapman estimator is better. The reason is that the bias in the Lincoln-Peterson population estimate (NLP) is infinite, but the bias in the Chapman population estimate (NC) is much smaller, at *E*[*θ^*]−*θ* = -631.3409.

7) Output for Chapman estimator function for N = 100,000 and 1,000 simulation runs (iterations) and multiple sample sizes:

> chapman\_estimations <- chapman\_estimator\_func(100000, 1000, seq(from=100, to=5000, by=50))

> chapman\_estimations

n bias\_NC variance

1 100 -90357.65467 46373839

2 150 -79628.30650 211648749

3 200 -66130.14830 596689262

4 250 -52802.75083 1190686341

5 300 -39094.47775 2021208538

6 350 -28890.26281 2804096804

7 400 -18197.70270 3699845990

8 450 -12088.82779 4133056614

9 500 -8690.22272 4202767649

10 550 -3019.85533 4411950964

11 600 -4336.44606 3687830137

12 650 -2974.94697 3436356760

13 700 2699.40218 4018247518

14 750 -537.85933 2522370783

15 800 -679.60820 2374949026

16 850 -1515.28543 1814533576

17 900 -2478.84728 1504415194

18 950 -728.41241 1352848536

19 1000 147.66683 1307590073

20 1050 -535.48502 1067679616

21 1100 -143.50256 964383950

22 1150 -973.79665 840796053

23 1200 513.04092 822005936

24 1250 1198.30741 732530282

25 1300 978.20593 689611896

26 1350 -28.17987 623585268

27 1400 390.50695 553519061

28 1450 261.65060 520310541

29 1500 -673.99538 465366394

30 1550 -132.29176 442751067

31 1600 256.23703 409785413

32 1650 -798.68524 371474314

33 1700 801.49627 365980909

34 1750 47.57072 340464620

35 1800 -950.22726 308782755

36 1850 -129.64815 295291644

37 1900 -38.77533 280564643

38 1950 -254.41240 263946586

39 2000 -568.21352 247715696

40 2050 -133.70079 238073530

41 2100 809.17813 233331568

42 2150 71.80907 217378499

43 2200 -128.07612 206396965

44 2250 -96.94765 195782432

45 2300 -343.25606 186171600

46 2350 383.78894 181540868

47 2400 -222.96657 169458464

48 2450 406.00029 165352131

49 2500 -264.02485 156085016

50 2550 361.69503 152671723

51 2600 335.04905 146098290

52 2650 -814.48811 135635640

53 2700 -345.29094 131961680

54 2750 471.32608 130469256

55 2800 -248.69359 122244033

56 2850 -121.97734 118651647

57 2900 60.49187 114946714

58 2950 -55.55267 110236102

59 3000 478.34557 108372019

60 3050 -213.87027 102353444

61 3100 -345.97845 98549024

62 3150 -395.79464 95269593

63 3200 -702.21045 91139018

64 3250 418.51426 91546295

65 3300 225.21591 87947339

66 3350 -91.12260 84407897

67 3400 313.04823 82991141

68 3450 -7.53502 79522562

69 3500 454.12010 78223737

70 3550 411.47590 75792406

71 3600 -710.10642 71183572

72 3650 -339.50183 69870504

73 3700 -339.22748 67979158

74 3750 -171.26402 66465570

75 3800 -180.20116 64609742

76 3850 42.71158 63199102

77 3900 -245.21195 61097951

78 3950 -119.45887 59520388

79 4000 578.62751 59466400

80 4050 297.59908 57320537

81 4100 -198.26804 54865674

82 4150 242.35399 54324853

83 4200 -47.71310 52521527

84 4250 90.45059 51424932

85 4300 -127.45483 49743523

86 4350 151.63668 49066739

87 4400 -117.94688 47460996

88 4450 -282.94385 46085639

89 4500 163.30493 45725779

90 4550 497.27073 45109190

91 4600 -10.76873 43394671

92 4650 313.52478 42809934

93 4700 50.44463 41568581

94 4750 -132.84031 40355464

95 4800 184.86562 39888869

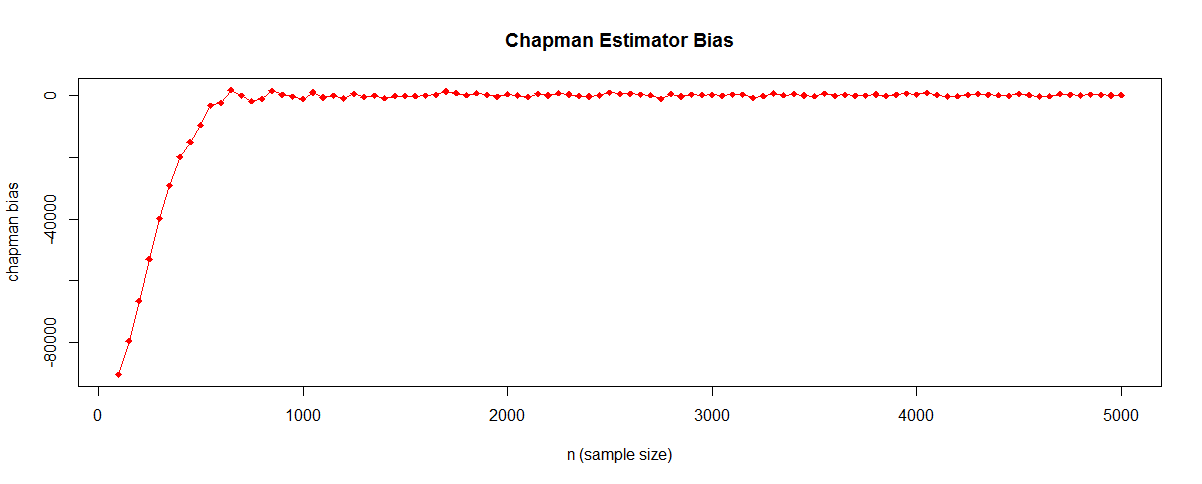
96 4850 135.68944 38976818

97 4900 -261.10334 37686580

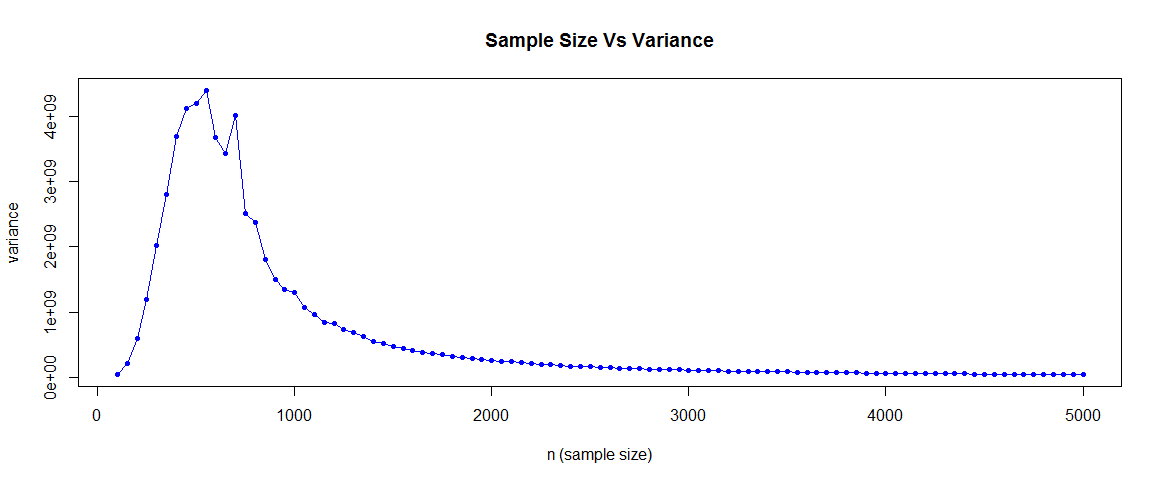
98 4950 248.71459 37418911

99 5000 166.95369 36569573

(a)



Chapman bias versus n: With a small sample number of 100 - 500, the bias towards a smaller population estimate is high (a large negative bias number), which means the estimated population will be much less than the actual count. As we increase the sample size, the bias approaches 0. Between a sample size of 500 to 1,000 the bias quickly improves and from 1,000 to 5,000 the bias is almost stable and it eventually looks closer to a flat line near zero bias.

(b) 

Chapman variance versus n: With small sample sizes of 100 to 500 the variance is very high. As we keep on increasing the sample size, the variance approaches 0. Between sample sizes 500 to 2,500, the variance is quickly reducing and from 4,500 to 5,000 the variance is almost stable and it eventually looks closer to a flat line near zero bias.

**8**. The Chapman Estimator can be a consistent estimator:

if (n1 + 1)(n2+1)/(m2+1) – 1 🡪 N (1)

i.e. (n + 1)2 / (m2 + 1) – 1 (for n1 = n2 = n),

As n 🡪 Infinity, it means all the population is captured and recaptured, hence ~ (n = N)

Since all tagged individuals will be recaptured, m2 = N

Hence, (1) will become (N + 1)2 / (N + 1) – 1

Therefore (N+1) – 1 🡪 N

We see as n approaches infinity, the Chapman estimator becomes closer to an unbiased estimate *E*[*θ^*]−*θ* = 0, thus it can be a consistent estimator.

**9**. Assumptions and reality:

a) Each individual is independently captured: Individual instincts may affect this assumption. Capture of one individual may scare / affect other individuals and those other individuals may try to escape, and so this assumption also may fail.

b) Each individual is equally likely to be captured: This is unrealistic because the population may be scattered with different densities across different geographical regions which the researcher may not be able to cover or is unaware of. He/she may be tempted to pick more samples from denser regions, and hence such samples are more likely to be captured. This can equate to non-equal probabilities of capturing samples from the population. Some individuals may also be more difficult to capture than others.

c) There are no births, deaths, immigration, or emigration of individuals: this is unrealistic. Births and deaths are not likely to be controlled. These four factors occur naturally in nature and are very difficult, if not impossible, to control.