# 模擬研究與討論

## 估計式整理

整理上一章所推導之估計式，以及過去文獻中常見的估計式，並同時估計其標準差。藉由電腦模擬的方式，在不同模型的設定下，比較三種取後不放回以及兩種取後放回的估計方式之表現。以下為不同估計方法之整理：

### 取後放回

其中：

|  |
| --- |
|  |
|  |
|  |

且在上述等式中皆表示為：若 時，則；而若 時，則表示為。

### 取後不放回

其中：

並在 的估計式中代入 。

並將與 以 帶入上述式子。

其中 與 。

### 標準差、95%信賴區間、樣本涵蓋率與變異係數估計

* 標準差估計

|  |  |
| --- | --- |
|  |  |

其中

* 95%信賴區間

1. *wNew1* and *wNew2*
2. Other

|  |  |
| --- | --- |
|  |  |

其中 以此計算95%信賴區間的樣本涵蓋率 (95% confidence interval Coverage rate, 95% CI Coverage) 。

* Sample Coverage

為樣本中檢測到的物種的總發生率的百分比。

* 變異係數

為變異係數平方的估計式，即。

## 模型模擬設定

### 模擬模型假設

經由電腦模擬，可以設定不同母體並產生不同物種豐富度的結構，透過重複抽樣資料以評估估計式之估計表現。在模擬的過程中，首先需決定兩族群的共同種與特有種的數量，再選擇兩種不同種模型作為母體使用，以下為四種模擬模型之設定：

1. 同質性模型 (homogeneity model)， ， ()
2. 均勻模型 (uniform model)， ， ()，其中服從均勻分佈，為調整常數。
3. Broken-stick 模型， ， ()，其中服從指數函數分佈，為調整常數。
4. 對數常態模型 (log-normal model)，， () ，其中服從對數常態分佈，為調整常數。

### 物種與區塊數的假設

在電腦模擬時，需針對真實的群落物種與區塊數進行假設，以評估估計結果之優劣程度。故假設以下四種組合為電腦模擬的真實物種與區塊狀況做使用。

1. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)。且兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用，且。
2. 假設總物種數皆為700種物種 (700)，其中群落一與第二群落分別存在400與600種物種 (400，600)，兩群落間共有300種共同種(300)。且兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用，且。
3. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)；兩族群皆為100區塊 (100)，並從中依比例抽取與個區塊作為樣本使用，而 。
4. 假設總物種數為500種物種 (500)，其中群落一與第二群落皆存在400種物種 (400)，兩群落間共有300種共同種(300)。且兩族群分別為100與200區塊 (100，200)，並從中依比例抽取與 個區塊作為樣本使用。

而在每種群落假設下，樣本數的模擬次數為 1000次。

## 表格中名詞定義

* q：為抽樣比例，。
* Sample size：抽樣區塊數，t。
* Estimator：所採用之估計方法。
* Obs：兩樣本觀測到的共同物種。
* AVG Estimate：次模擬之平均估計值。
* Bias：次模擬之偏差 (bias)，。
* Sample SE：次模擬之樣本標準差。
* Est. SD：次模擬之標準差估計值。
* RMSE：次模擬估計量之樣本均方根誤差。
* 95% CI Coverage：次模擬中理論值於信賴區間的比例。

## 取後放回的模擬結果

首先針對樣本涵蓋率進行計算，在多種組合模擬下，在樣本大小為10的情況下，個模型的樣本涵蓋率介於0.6至0.84之間；而當樣本大小大於10時，任何模型的樣本涵蓋率皆大於0.9。

且依據模擬結果可以得知 (Table 4.1、Table 4.2、Table 4.3、Table 4.4)，在第一種物種與區塊數的假設，四種模擬模型下，偏差的表現在兩種估計方法中所估計的群落共同種，皆有低估的情況發生。尤其在小樣本時低估情況更加明顯。平均而言，在兩種估計方法的偏差結果中皆顯示，*New*所估計之群落共同種相較於原有的*Pan*方法在偏差的估計表現上更為優秀。而在樣本標準差 (Sample SE) 以及標準差估計 (Est. SD) 的結果中，可以發現無論在何種模擬情形下兩種方法皆有低估的情況發生。且*New*的樣本標準差與標準差估計結果皆比*Pan*大。

然而，單憑偏差與標準差的估計的結果無法從中判定估計式的好壞，在評估一個估計方法的優劣之時，還需要考慮該估計量的準確度以及精確度，因此將以均方根誤差 (root-mean-square error，RMSE) 以及95% CI Coverage (95% confidence interval Coverage，95% CI Coverage) 作為衡量估計式好壞之準則。在RMSE的結果方面，在兩群落皆為小樣本時，*New*的表現較優秀；但在大樣本的抽樣下，*Pan*的表現會略優於*New*。 另一方面，95% CI Coverage也是能作為評估估計式優劣的標準之一。在95% CI Coverage的結果中*New*的結果皆略優於*Pan*。

在第二種與第三種物種與區塊數的假設下所進行的模擬結果，無論是在何種估計值的結果，皆與第一種物種與區塊數的假設並無太大的差異，*New*的表現皆明顯優於*Pan*。但是，在針對第四種物種的假設下，無論是在觀測物種或是平均的估計結果上，相較於第一種假設在*New*的估計下，雖然也具有較小的偏差，容易有高估的情況發生；而在RMSE的表現上，*New*的估計也獲得較差的結果；同時，隨著樣本數的增加，在95% CI Coverage的結果反而下降。

綜上所述，在評估各項模擬與所有指標之後，本文所提出*New*估計方式的表現普遍優於*Pan*對於兩群落共同種的估計結果。且在目標區塊數以及共同物種數的真值皆為相同的情況底下，群落中的特有種數量並不會影響共同物種數的估計結果。但當目標區塊與抽樣區塊不相同時，則會影響到估計式的表現結果。

Table 4.1取後放回的抽樣方式在第一種物種與區塊假設下，群落一為同質模型，群落二為Broken-stick模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 10 | New | 125.12 | 270.79 | -29.21 | 61.1 | 46.64 | 67.69 | 0.86 |
| Pan | 246.81 | -53.19 | 40.24 | 27.93 | 66.68 | 0.82 |
| 30 | New | 240.32 | 287.18 | -12.82 | 19.64 | 18.34 | 23.44 | 0.91 |
| Pan | 277.62 | -22.38 | 13.64 | 11.92 | 26.2 | 0.89 |
| 50 | New | 267.3 | 294.68 | -5.32 | 16.04 | 14.36 | 16.89 | 0.87 |
| Pan | 288.29 | -11.71 | 11.15 | 9.45 | 16.16 | 0.86 |
| 70 | New | 277.28 | 297.93 | -2.07 | 14.03 | 12.42 | 14.18 | 0.83 |
| Pan | 292.66 | -7.34 | 9.78 | 8.19 | 12.22 | 0.82 |

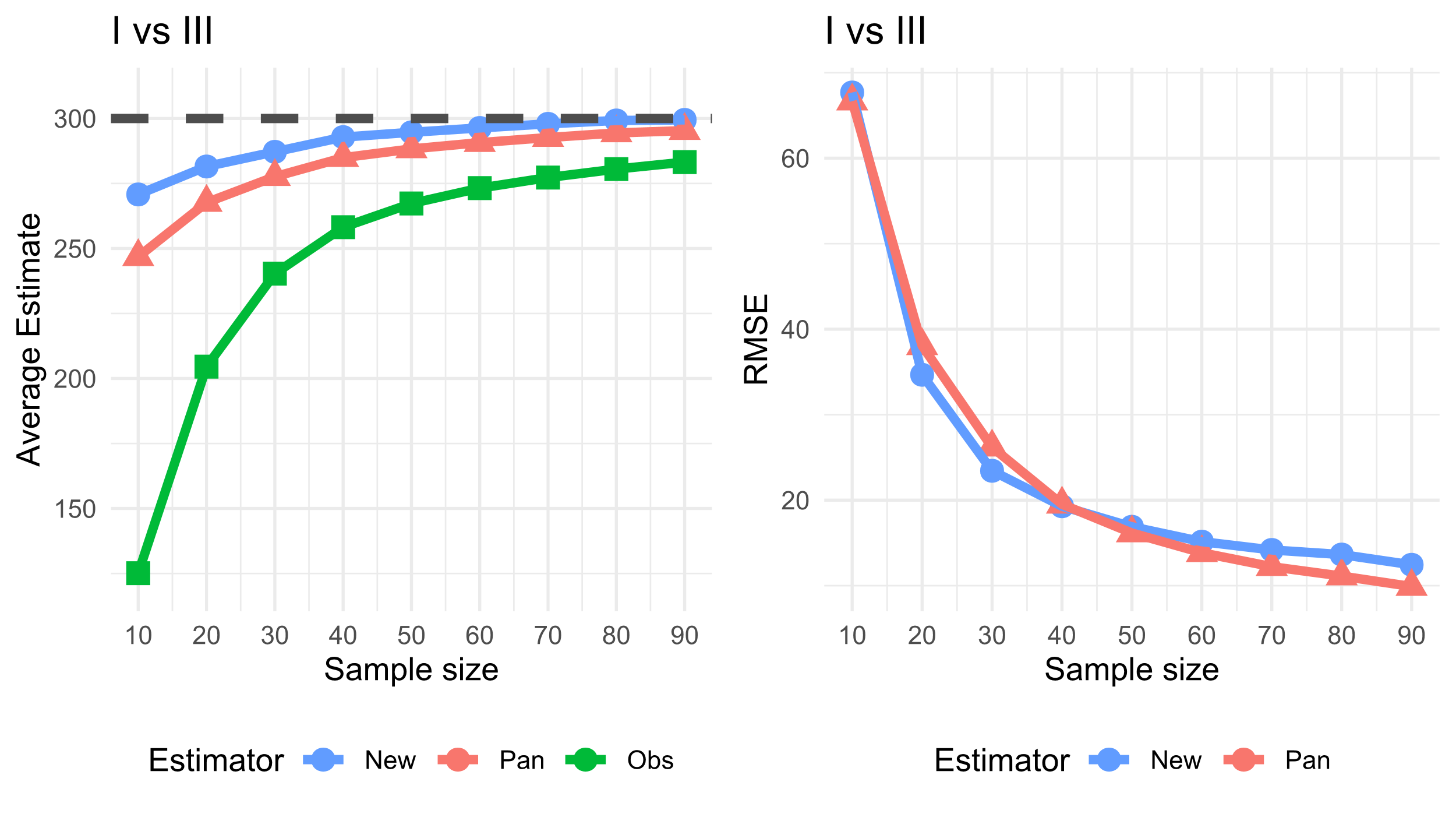


Figure 4.1取後放回的抽樣方式在第一種物種與區塊假設下， 群落一為同質模型，群落二為Broken-stick模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

Table 4.2取後放回的抽樣方式在第一種物種與區塊假設下，兩群落皆為均勻模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 10 | New | 123.34 | 244.33 | -55.67 | 52.13 | 40.13 | 76.25 | 0.86 |
| Pan | 220.02 | -79.98 | 32.53 | 22.51 | 86.34 | 0.83 |
| 30 | New | 225.85 | 277.64 | -22.36 | 22.77 | 19.67 | 31.9 | 0.89 |
| Pan | 265.65 | -34.35 | 15.31 | 12.33 | 37.6 | 0.88 |
| 50 | New | 255.45 | 289.5 | -10.5 | 16.01 | 15.42 | 19.14 | 0.9 |
| Pan | 281.66 | -18.34 | 11.37 | 10.15 | 21.58 | 0.89 |
| 70 | New | 268.74 | 295.39 | -4.61 | 13.68 | 13.72 | 14.43 | 0.91 |
| Pan | 289.14 | -10.86 | 9.76 | 9.21 | 14.6 | 0.9 |

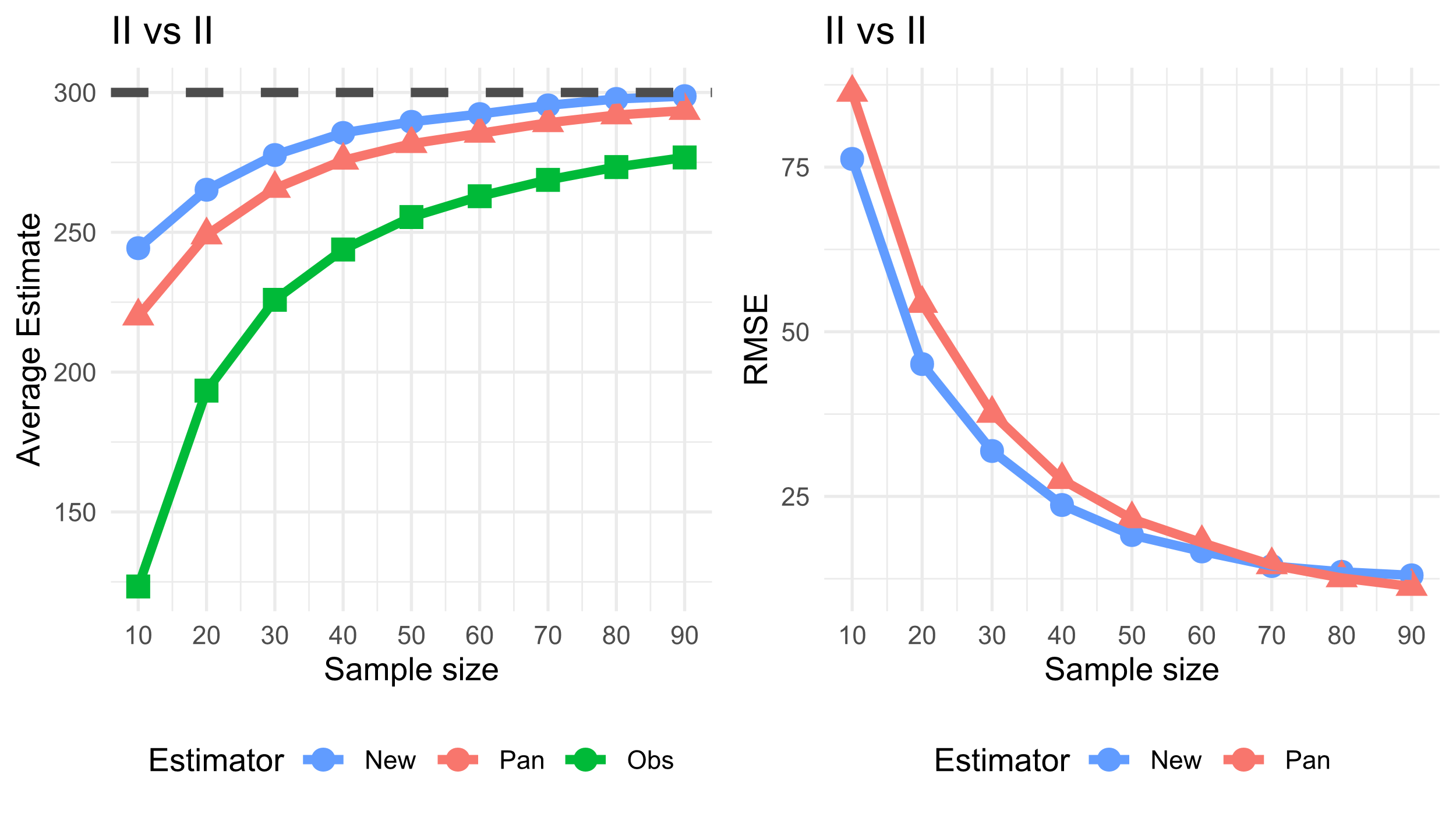


Figure 4.2取後放回的抽樣方式在第一種物種與區塊假設下，兩群落皆為均勻模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

Table 4.3取後放回的抽樣方式在第一種物種與區塊假設下，群落一為均勻模型，群落二為Broken-stick模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 10 | New | 123.27 | 232.31 | -67.69 | 47.01 | 37.17 | 82.4 | 0.85 |
| Pan | 207.78 | -92.22 | 29.49 | 20.99 | 96.82 | 0.83 |
| 30 | New | 218.54 | 276.15 | -23.85 | 24.65 | 21.91 | 34.29 | 0.89 |
| Pan | 262.17 | -37.83 | 16.27 | 13.71 | 41.18 | 0.88 |
| 50 | New | 248.57 | 290.11 | -9.89 | 19.9 | 17.98 | 22.21 | 0.88 |
| Pan | 280.02 | -19.98 | 13.76 | 11.77 | 24.26 | 0.88 |
| 70 | New | 263.41 | 295.6 | -4.4 | 17.15 | 15.73 | 17.69 | 0.88 |
| Pan | 287.96 | -12.04 | 12.17 | 10.46 | 17.11 | 0.86 |

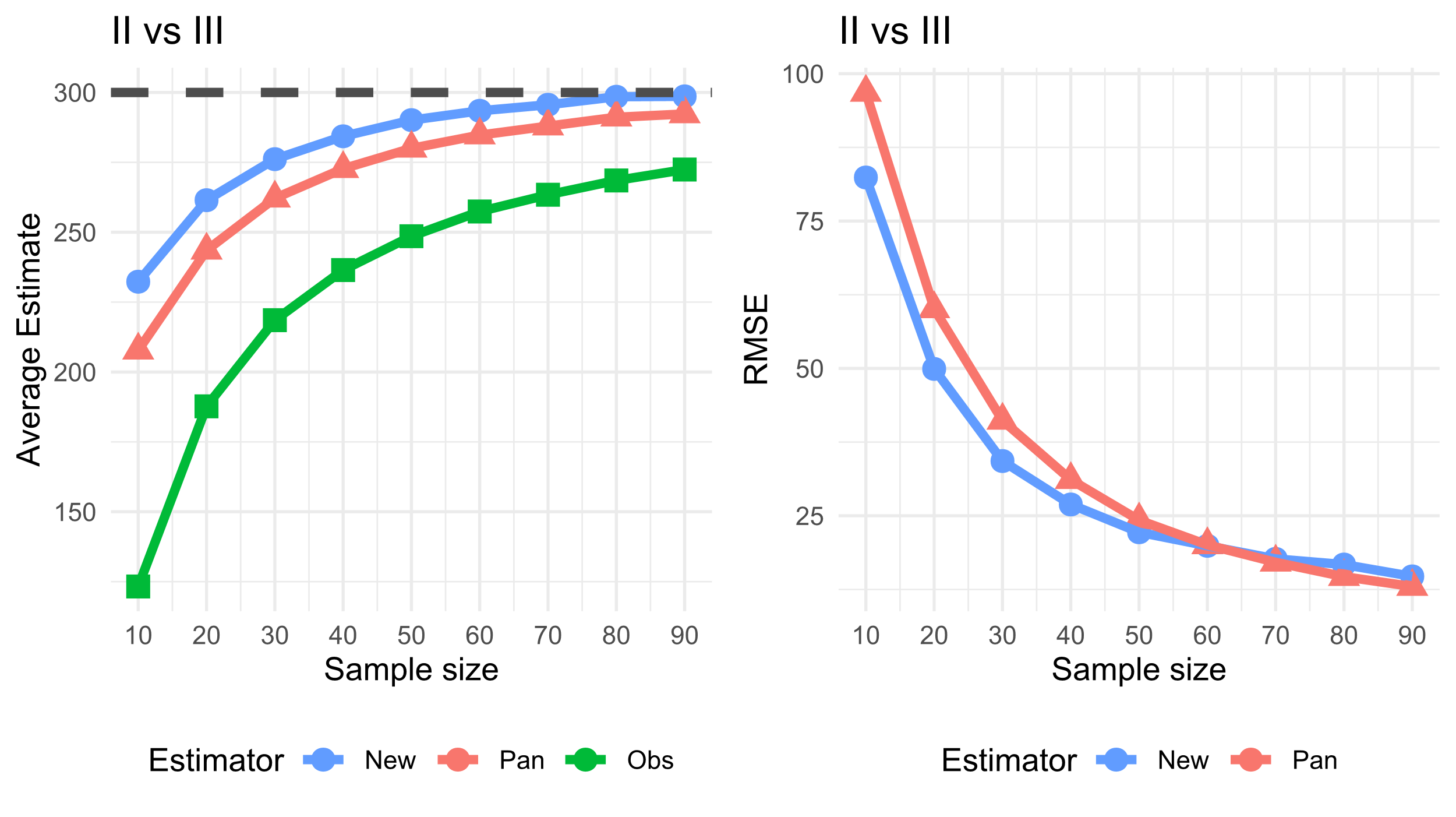


Figure 4.3取後放回的抽樣方式在第一種物種與區塊假設下，群落一為均勻模型，群落二為Broken-stick模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

Table 4.4取後放回的抽樣方式在第一種物種與區塊假設下，群落一為Broken-stick模型，群落二為對數常數模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 10 | New | 115.62 | 228.28 | -71.72 | 52.12 | 40.34 | 88.64 | 0.85 |
| Pan | 199.49 | -100.51 | 31.57 | 22.73 | 105.34 | 0.83 |
| 30 | New | 208.96 | 277.73 | -22.27 | 29.47 | 25.47 | 36.93 | 0.9 |
| Pan | 261.16 | -38.84 | 18.92 | 15.57 | 43.2 | 0.88 |
| 50 | New | 242.64 | 292.56 | -7.44 | 22.86 | 20.58 | 24.03 | 0.88 |
| Pan | 280.66 | -19.34 | 15.44 | 13.18 | 24.74 | 0.87 |
| 70 | New | 260.08 | 298.13 | -1.87 | 19.15 | 17.44 | 19.23 | 0.88 |
| Pan | 289.36 | -10.64 | 13.26 | 11.47 | 16.99 | 0.88 |

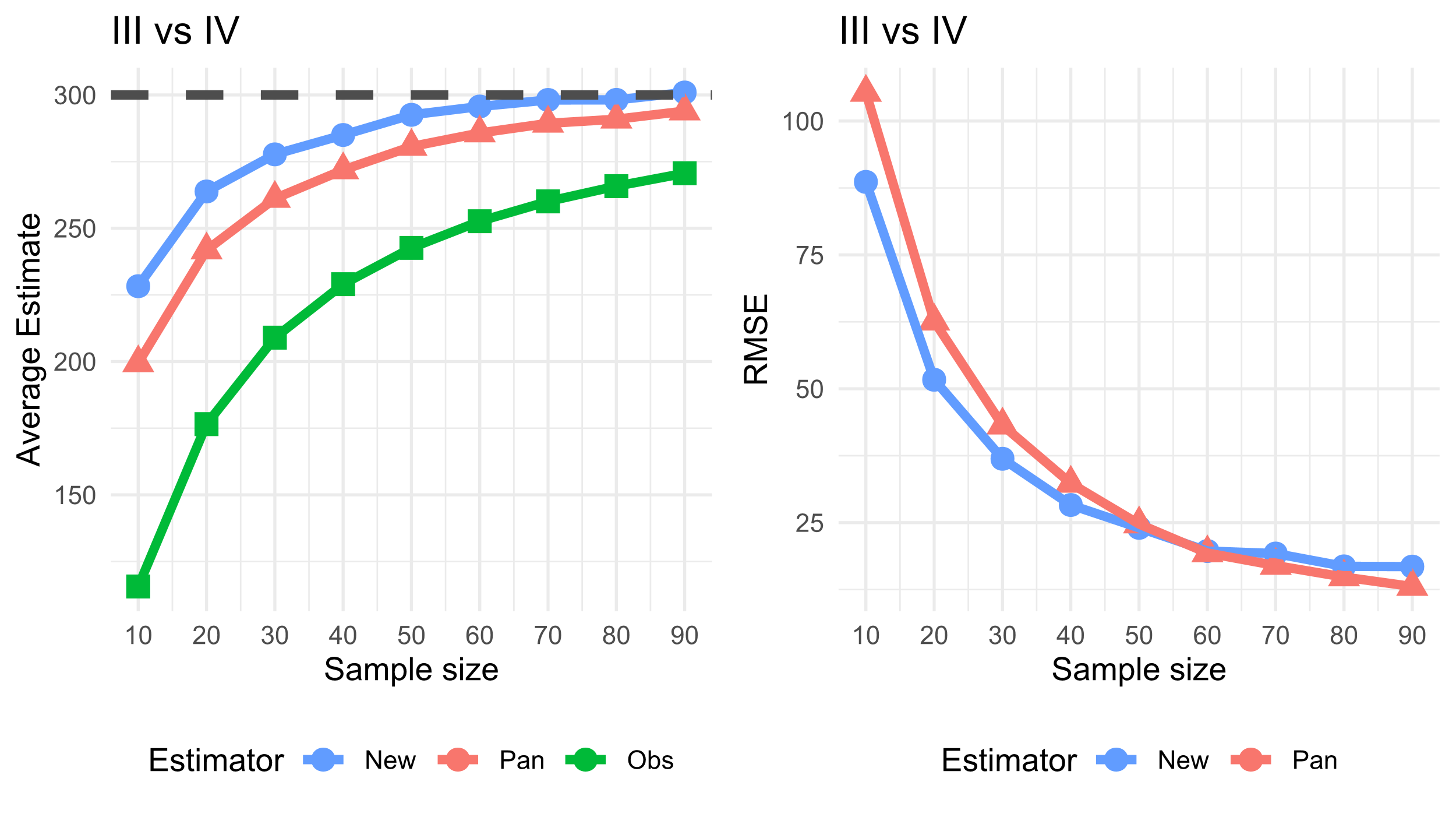


Figure 4.4 取後放回的抽樣方式在第一種物種與區塊假設下，群落一為Broken-stick模型，群落二為對數常數模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

## 取後不放回與取後放回的模擬結果比較

在取後不放回的模擬針對樣本涵蓋率型計算，同樣於在多種組合模擬下，除了在抽樣比例為0.1的情況底下，其餘樣本大小所抽樣本，無論是在何種模型或模擬組合下，所估計之樣本涵蓋率皆大於0.9。且模擬同質群落或均勻群落在抽樣比例為0.1的情況下，樣本變異係數通常小於0.4。而隨抽樣比例的增加將越接近母體變異數。

依據第一種物種假設的模擬結果可以得知 (Table 4.5、Table 4.6、Table 4.7、Table 4.8)，在偏差的表現上， 在*wNew1*與*wNew2*小樣本的估計中，同質與Broken-stick以及兩群落皆為均勻的兩種模擬組合下有高估的表現；而在均勻與Broken-stick以及Broken-stick與常對數的兩種模擬組合下則出現低估的表現。但在大樣本的情況下，所有群落組合則是三種估計方法皆呈現低估的表現。

在樣本標準差 (Sample SE) 以及平均標準差估計 (Est. SD) 的結果中，可以發現無論在何種模擬情形下三種方法皆有低估的情況發生。除此之外，在RMSE的結果中 *wNew1*與*wNew2*兩者無明顯差異，且都優於*wChao2.12*。 而在95% CI Coverage方面，由於*wChao2.12*已知為下界估計式，因此在95% CI Coverage的估計上使用log修正的方式進行估計，而在*wNew1*與*wNew2*，則使用未修正的估計方式。在該項結果下*wNew1*的表現優於其他兩者，且隨著抽樣比例的增加，*wNew1*與*wNew2*無明顯差異，且略優於*wChao2.12*。

在第二種與第三種物種與區塊數的假設底下，所進行的模擬結果，無論是在何種估計值的結果，皆與第一種物種與區塊數的假設並無太大的差異。並且，不同於第四種物種的假設在取後放回的抽樣模擬中與前三種的估計結果上有所差異，在取後不放回的模擬下，四種假設的結果皆無明顯差異。

綜合以上論述可以得知，普遍而言在取樣比例小於0.3時，取後不放回的兩估計結果較不穩定，會因群落之變異大小所致，使得高估或低估的現象皆有可能發生。而依據多次模擬結果可以得知，通常在其中一個群落的估計之樣本變異係數小於0.4時 (0.4)，*wNew1*與*wNew2*更容易發生高估的情況。反之在大樣本的條件下，採用取後放回的估計方法*New*則是會有明顯高估的情況發生。

因此在小樣本且當其中一個群落樣本所估計的變異係數小於0.4時的估計中，建議採用取後放回的抽樣方式 (*New*) 估計兩群落的共同種；而在抽樣比例達到0.3時，取後不放回的估計結果才趨近穩定，並相較於取後放回的方法有更良好的估計表現，故在抽樣比例大於0.3時，以取後不放回的兩估計是對群落的共同種進行估計會獲得較好的結果。

Table 4.5取後不放回的抽樣方式在第一種物種與區塊假設下，群落一為同質模型，群落二為Broken-stick模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95%CI Coverage |
| 0.1 | wNew1 | 130.68 | 363.12 | 63.12 | 57.16 | 51.9 | 85.14 | 0.93 |
| wNew2 | 367.93 | 67.93 | 67.03 | 56.69 | 95.41 | 0.92 |
| wChao2.12 | 246.53 | -53.47 | 30.75 | 24.33 | 61.67 | 0.89 |
| New | 282.03 | -17.97 | 57.47 | 45.49 | 60.19 | 0.89 |
| 0.3 | wNew1 | 251.5 | 311.16 | 11.16 | 12.8 | 9.77 | 16.98 | 0.88 |
| wNew2 | 313.95 | 13.95 | 16.78 | 15 | 21.81 | 0.93 |
| wChao2.12 | 281.55 | -18.45 | 9.7 | 9.15 | 20.85 | 0.92 |
| New | 300.26 | 0.26 | 19.47 | 19.11 | 19.46 | 0.91 |
| 0.5 | wNew1 | 277.66 | 299.44 | -0.56 | 6.72 | 7.5 | 6.74 | 0.98 |
| wNew2 | 297.72 | -2.28 | 6.06 | 5.89 | 6.47 | 0.94 |
| wChao2.12 | 292.87 | -7.13 | 6.15 | 5.81 | 9.41 | 0.89 |
| New | 313.19 | 13.19 | 16.35 | 16.84 | 21 | 0.91 |
| 0.7 | wNew1 | 289.19 | 299.95 | -0.05 | 3.67 | 4.06 | 3.67 | 0.97 |
| wNew2 | 298.72 | -1.28 | 3.32 | 3.51 | 3.56 | 0.97 |
| wChao2.12 | 298.19 | -1.81 | 3.62 | 3.74 | 4.05 | 0.9 |
| New | 324.51 | 24.51 | 14.65 | 17.23 | 28.55 | 0.94 |

Table 4.6取後不放回的抽樣方式在第一種物種與區塊假設下，兩群落皆為均勻模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95%CI Coverage |
| 0.1 | wNew1 | 127.72 | 306.18 | 6.18 | 44.17 | 44.69 | 44.58 | 0.95 |
| wNew2 | 310.94 | 10.94 | 51.89 | 43.22 | 53.01 | 0.92 |
| wChao2.12 | 215.98 | -84.02 | 23.97 | 19.41 | 87.37 | 0.87 |
| New | 246.21 | -53.79 | 46.24 | 38.37 | 70.92 | 0.89 |
| 0.3 | wNew1 | 234.14 | 295.51 | -4.49 | 14.16 | 14.3 | 14.85 | 0.95 |
| wNew2 | 302.6 | 2.6 | 19.56 | 17.19 | 19.72 | 0.93 |
| wChao2.12 | 269.61 | -30.39 | 11.25 | 9.92 | 32.4 | 0.89 |
| New | 294.95 | -5.05 | 24.75 | 21.59 | 25.25 | 0.9 |
| 0.5 | wNew1 | 266.2 | 297.26 | -2.74 | 8.56 | 8.55 | 8.98 | 0.94 |
| wNew2 | 300.7 | 0.7 | 9.85 | 9.02 | 9.87 | 0.92 |
| wChao2.12 | 287.64 | -12.36 | 7.82 | 6.8 | 14.62 | 0.89 |
| New | 315.69 | 15.69 | 20.48 | 19.29 | 25.79 | 0.9 |
| 0.7 | wNew1 | 282.94 | 299.07 | -0.93 | 5.17 | 4.97 | 5.25 | 0.94 |
| wNew2 | 299.6 | -0.4 | 5.23 | 4.84 | 5.24 | 0.93 |
| wChao2.12 | 296.33 | -3.67 | 5.08 | 4.55 | 6.26 | 0.89 |
| New | 337.83 | 37.83 | 20.16 | 21.22 | 42.87 | 0.94 |

Table 4.7取後不放回的抽樣方式在第一種物種與區塊假設下，群落一為均勻模型，群落二為Broken-stick模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95%CI Coverage |
| 0.1 | wNew1 | 126.47 | 290.73 | -9.27 | 43.62 | 38.75 | 44.58 | 0.92 |
| wNew2 | 299.38 | -0.62 | 55.43 | 44.92 | 55.4 | 0.92 |
| wChao2.12 | 208.8 | -91.2 | 24.3 | 19.13 | 94.38 | 0.86 |
| New | 243.7 | -56.3 | 47.86 | 38.89 | 73.87 | 0.88 |
| 0.3 | wNew1 | 226.97 | 293.64 | -6.36 | 15.81 | 14.89 | 17.03 | 0.94 |
| wNew2 | 297.21 | -2.79 | 20.82 | 17.73 | 21 | 0.93 |
| wChao2.12 | 266.41 | -33.59 | 12.31 | 10.79 | 35.77 | 0.89 |
| New | 296.35 | -3.65 | 27.04 | 24.18 | 27.27 | 0.88 |
| 0.5 | wNew1 | 261.52 | 297.11 | -2.89 | 8.34 | 9.2 | 8.82 | 0.97 |
| wNew2 | 297.61 | -2.39 | 9 | 9.31 | 9.31 | 0.96 |
| wChao2.12 | 286.59 | -13.41 | 7.55 | 7.44 | 15.38 | 0.92 |
| New | 320.96 | 20.96 | 20.55 | 21.94 | 29.35 | 0.94 |
| 0.7 | wNew1 | 281.06 | 299.6 | -0.4 | 5.32 | 5.33 | 5.34 | 0.94 |
| wNew2 | 299.04 | -0.96 | 5.3 | 5.1 | 5.38 | 0.93 |
| wChao2.12 | 296.51 | -3.49 | 5.24 | 4.9 | 6.3 | 0.9 |
| New | 342.9 | 42.9 | 20.83 | 23.26 | 47.69 | 0.95 |

Table 4.8取後不放回的抽樣方式在第一種物種與區塊假設下，群落一為Broken-stick模型，群落二為對數常數模型之情況下的估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95%CI Coverage |
| 0.1 | wNew1 | 121.47 | 282.46 | -17.54 | 44.72 | 41.29 | 48.02 | 0.94 |
| wNew2 | 290.68 | -9.32 | 56.03 | 47.67 | 56.77 | 0.93 |
| wChao2.12 | 202.82 | -97.18 | 25.04 | 19.94 | 100.36 | 0.87 |
| New | 243.48 | -56.52 | 50.92 | 41.92 | 76.06 | 0.87 |
| 0.3 | wNew1 | 223.77 | 300.8 | 0.8 | 17.43 | 19.64 | 17.44 | 0.98 |
| wNew2 | 304.2 | 4.2 | 22.7 | 19.45 | 23.08 | 0.93 |
| wChao2.12 | 269.92 | -30.08 | 13.26 | 11.78 | 32.87 | 0.91 |
| New | 304.82 | 4.82 | 31.38 | 27.43 | 31.74 | 0.91 |
| 0.5 | wNew1 | 262.89 | 301.49 | 1.49 | 8.75 | 9.85 | 8.87 | 0.97 |
| wNew2 | 303.69 | 3.69 | 10.26 | 9.95 | 10.9 | 0.94 |
| wChao2.12 | 289.26 | -10.74 | 8.02 | 7.55 | 13.4 | 0.91 |
| New | 321.77 | 21.77 | 22.08 | 21.99 | 31 | 0.94 |
| 0.7 | wNew1 | 283.16 | 300.77 | 0.77 | 4.73 | 5.18 | 4.79 | 0.96 |
| wNew2 | 301.25 | 1.25 | 4.89 | 5.12 | 5.04 | 0.96 |
| wChao2.12 | 296.82 | -3.18 | 4.69 | 4.58 | 5.66 | 0.9 |
| New | 329.86 | 29.86 | 17.54 | 19.58 | 34.62 | 0.96 |

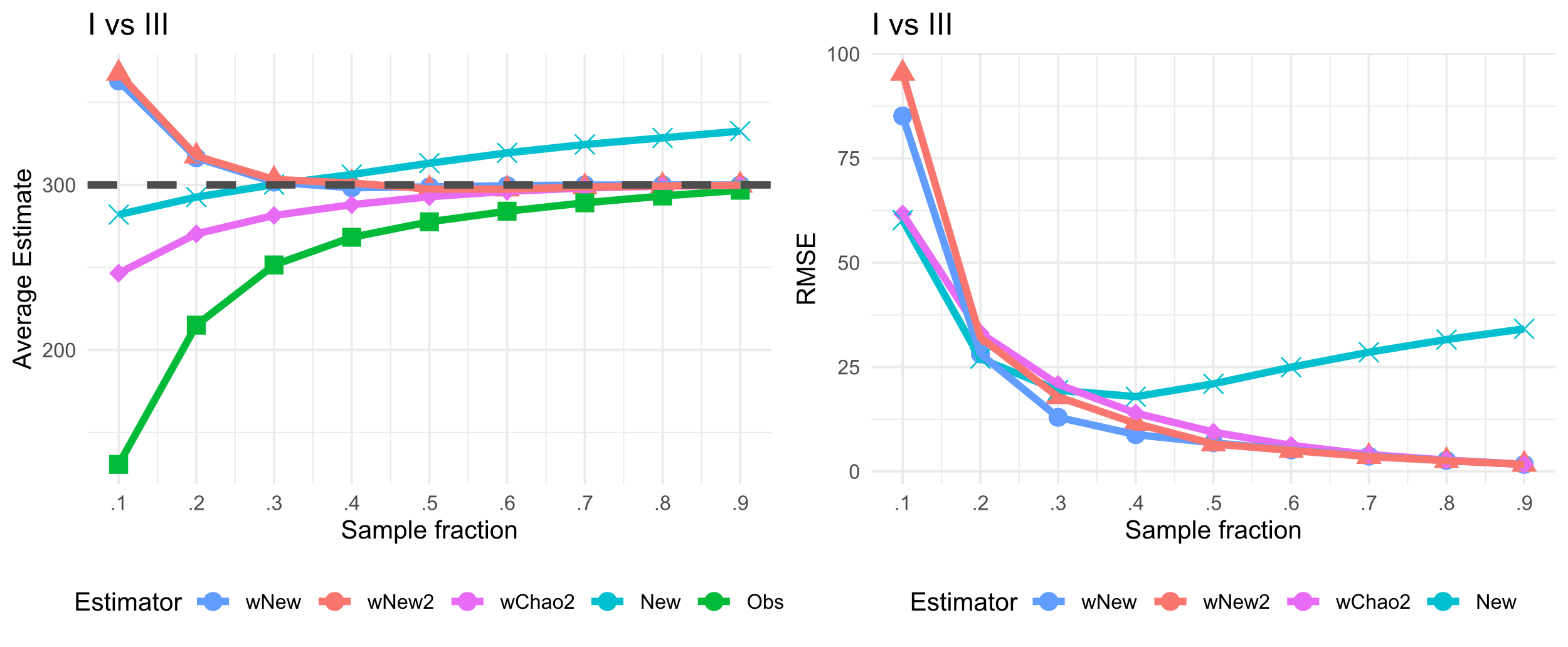


Figure 4.5取後不放回的抽樣方式在第一種物種與區塊假設下， 群落一為同質模型，群落二為Broken-stick模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

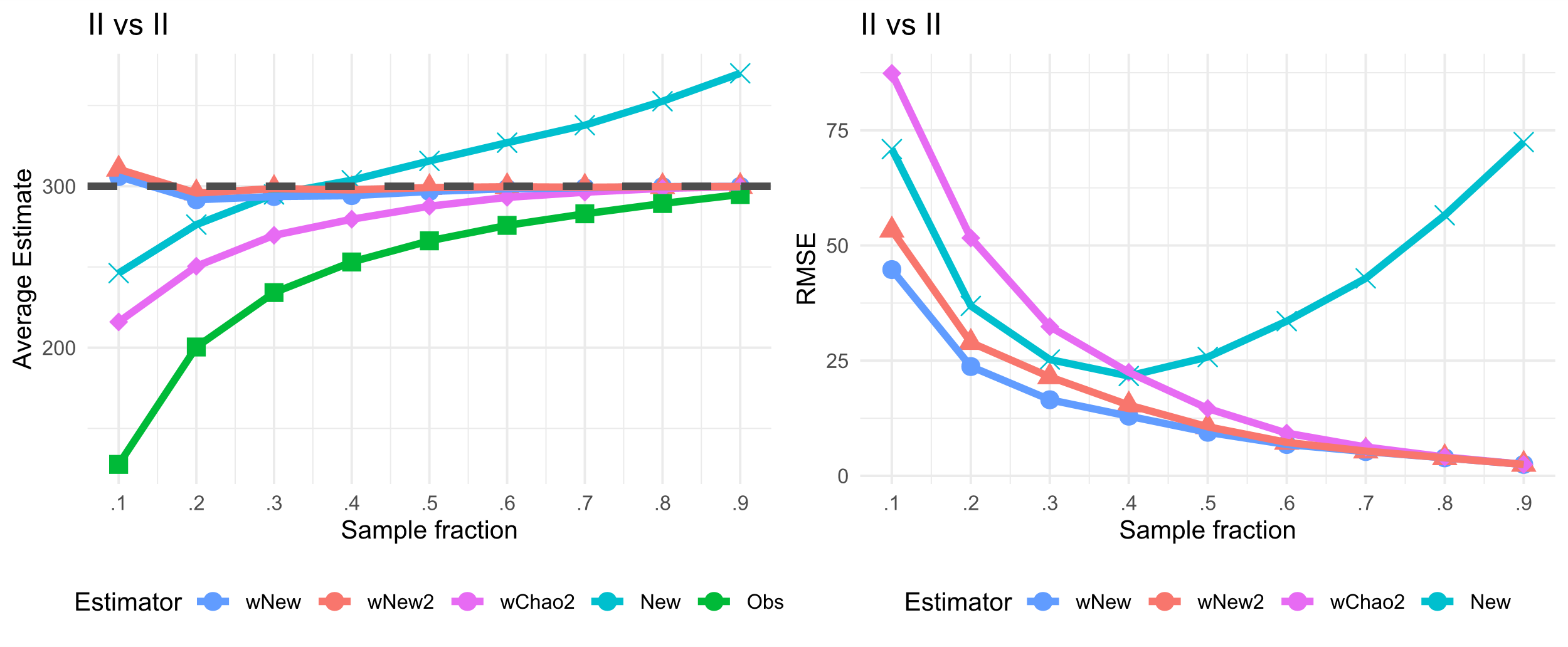


Figure 4.6取後不放回的抽樣方式在第一種物種與區塊假設下，兩群落皆為均勻模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

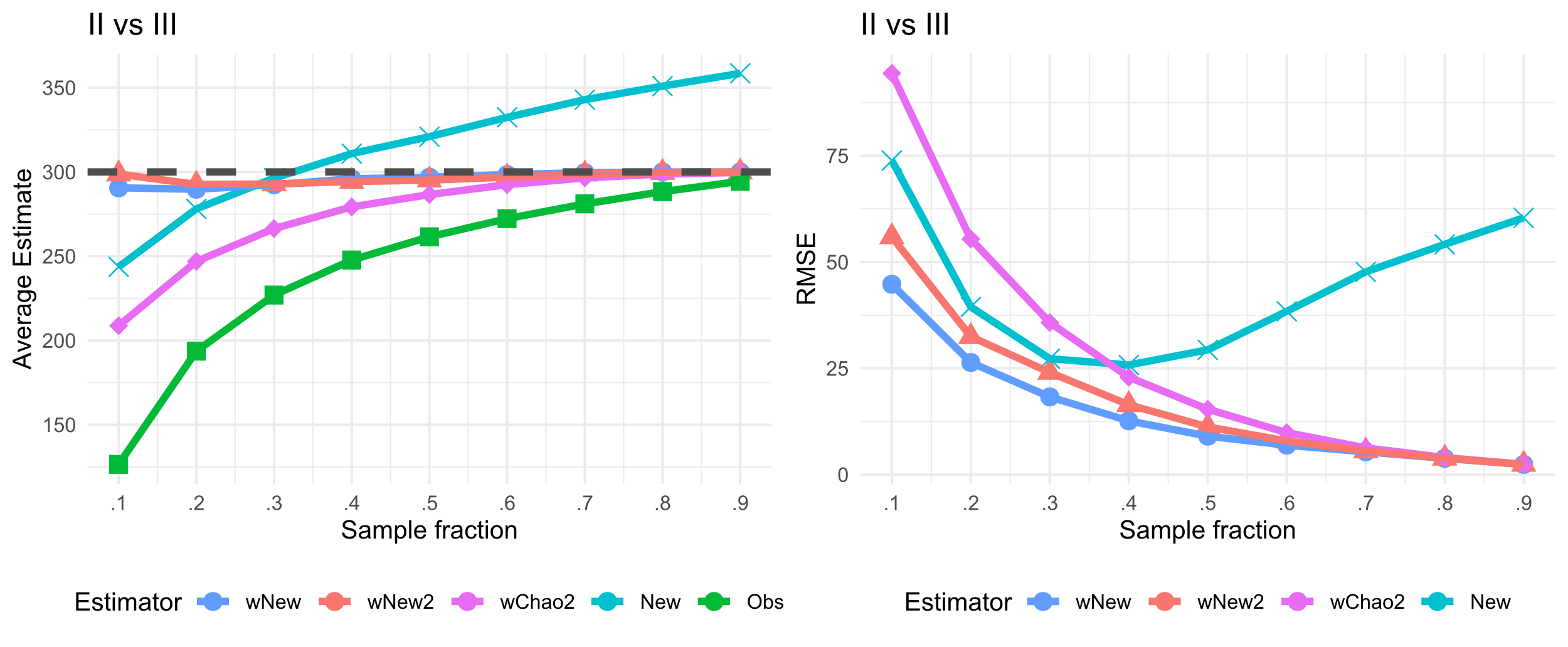


Figure 4.7取後不放回的抽樣方式在第一種物種與區塊假設下，群落一為均勻模型，群落二為Broken-stick模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

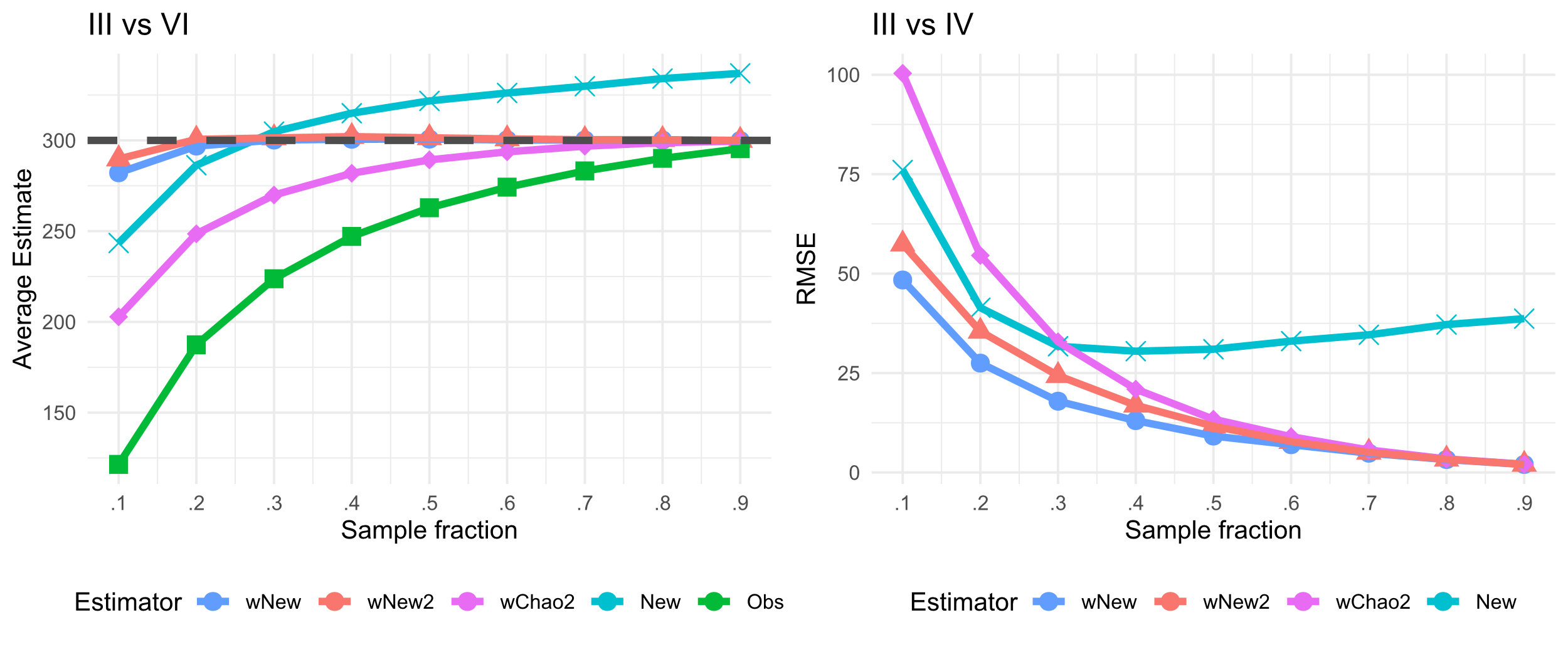


Figure 4.8取後不放回的抽樣方式在第一種物種與區塊假設下，群落一為Broken-stick模型，群落二為對數常數模型之情況下的AVG Estimate (左圖) 與RMSE (右圖) 估計結果。

## 真實資料模擬

除了使用模擬母體進行重複抽樣的模擬之外，亦使用兩筆真實數據做為母體進行1000次的重複抽樣，以評斷估計式的結果優劣。

### 澳洲三種極端氣候鳥類資料

**資料描述**

針對澳洲東南部半乾旱地區的尤加利樹林地，量化三個氣候時期178 個地點的極端氣候對鳥類物種出現 (species occurrence)、物種豐富度(species richness) 和出現型 (incidence) 的影響——聖嬰現象相關的乾旱 (Big Dry)、反聖嬰現象破壞性降雨 (Big Wet)，以及反聖嬰現象事件三年後 (Post-Big Wet)。並使用Big Dry 作為群落一使用 (Mean = 0.08, CV = 1.32)，其中共存在55種物種；Big Wet 作為群落二使用 (Mean = 0.16, CV = 1.39)，其中共存在65種物種。Post-Big Wet作為群落三使用 (Mean = 0.15, CV = 1.42)，其中共存在56種物種。

**模擬結果**

在該筆資料中使用取後放回的估計方法下，假設該筆資料為母體，並抽取其中的區塊做為樣本所使用。在偏差結果方面，使用*New*的估計結果明顯表現較*Pan*佳。在估計式評估標準上，使用RMSE與95% CI Coverage 進行評估兩種評估標準的結果表明，在多數情況下，RMSE在*New*估計方式中的評估結果劣於*Pan*；而在95% CI Coverage的模型評估上，兩者並無明顯差異，值得注意的是在本資料模擬中，隨著抽樣比例的增加，95% CI Coverage的結果會隨之下降。

Table 4.9取後放回澳洲三種極端氣候鳥類資料集中的Big Dry與Big Wet估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 20 | New | 25.54 | 40.84 | -8.16 | 15.08 | 12.58 | 17.13 | 0.8 |
| Pan | 37.54 | -11.46 | 10.53 | 8.79 | 15.56 | 0.81 |
| 60 | New | 37.39 | 46.03 | -2.97 | 9.69 | 8.18 | 10.13 | 0.75 |
| Pan | 44.16 | -4.84 | 6.97 | 5.89 | 8.48 | 0.77 |
| 100 | New | 41.48 | 47.89 | -1.11 | 8.92 | 6.78 | 8.98 | 0.7 |
| Pan | 46.44 | -2.56 | 6.49 | 4.94 | 6.97 | 0.72 |
| 140 | New | 43.66 | 48.91 | -0.09 | 7.21 | 6.13 | 7.21 | 0.69 |
| Pan | 47.72 | -1.28 | 5.35 | 4.46 | 5.49 | 0.71 |

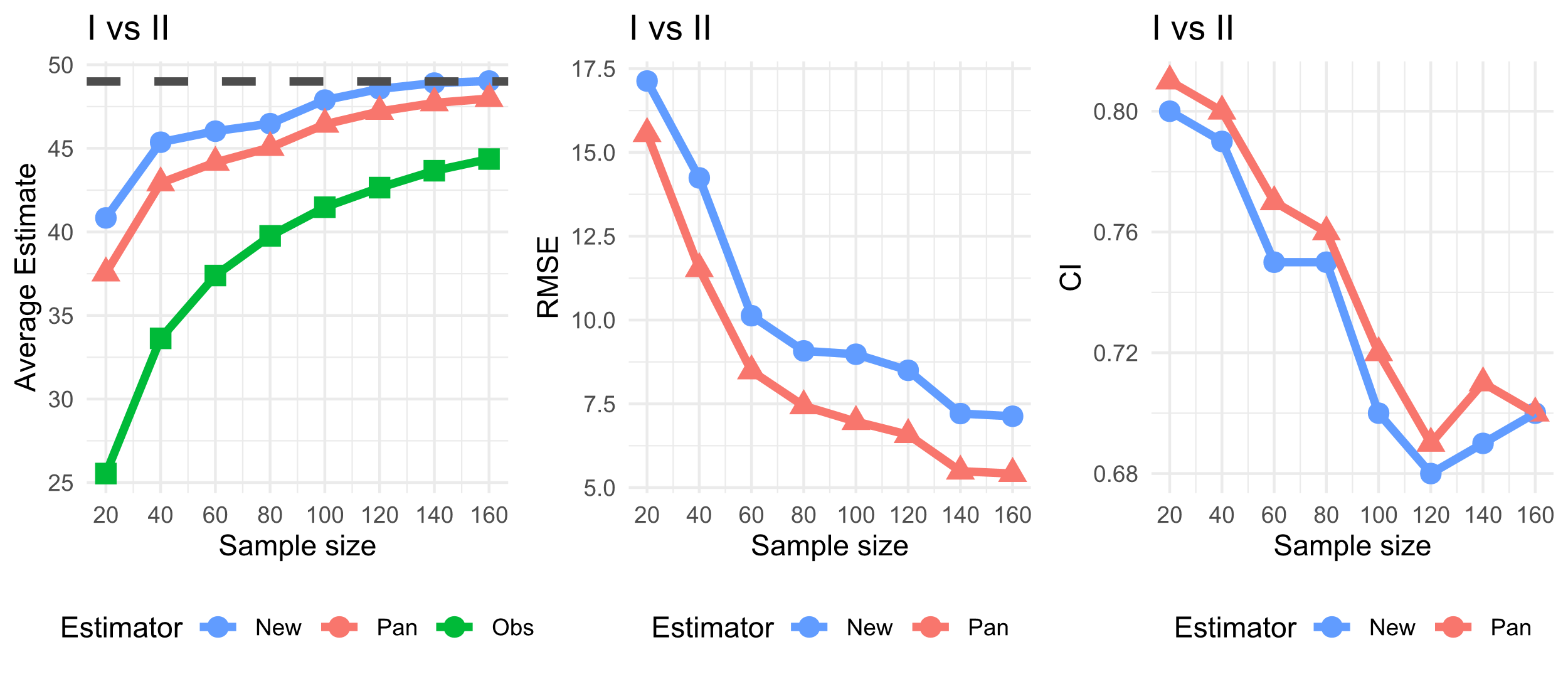


Figure 4.9取後放回澳洲三種極端氣候鳥類資料集中的Big Dry與Big Wet估計結果。

Table 4.10取後放回澳洲三種極端氣候鳥類資料集中的Big Dry與Post-Big Wet估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 20 | New | 24.27 | 40.19 | -11.81 | 14.49 | 13.1 | 18.69 | 0.82 |
| Pan | 36.69 | -15.31 | 10.14 | 9.13 | 18.36 | 0.84 |
| 60 | New | 37.16 | 47.37 | -4.63 | 10.54 | 9.4 | 11.51 | 0.8 |
| Pan | 45.27 | -6.73 | 7.85 | 6.67 | 10.33 | 0.79 |
| 100 | New | 41.93 | 49.97 | -2.03 | 9.39 | 8 | 9.6 | 0.76 |
| Pan | 48.09 | -3.91 | 6.9 | 5.76 | 7.93 | 0.75 |
| 140 | New | 44.6 | 51.59 | -0.41 | 8.53 | 7.41 | 8.53 | 0.77 |
| Pan | 49.88 | -2.12 | 6.22 | 5.31 | 6.56 | 0.76 |

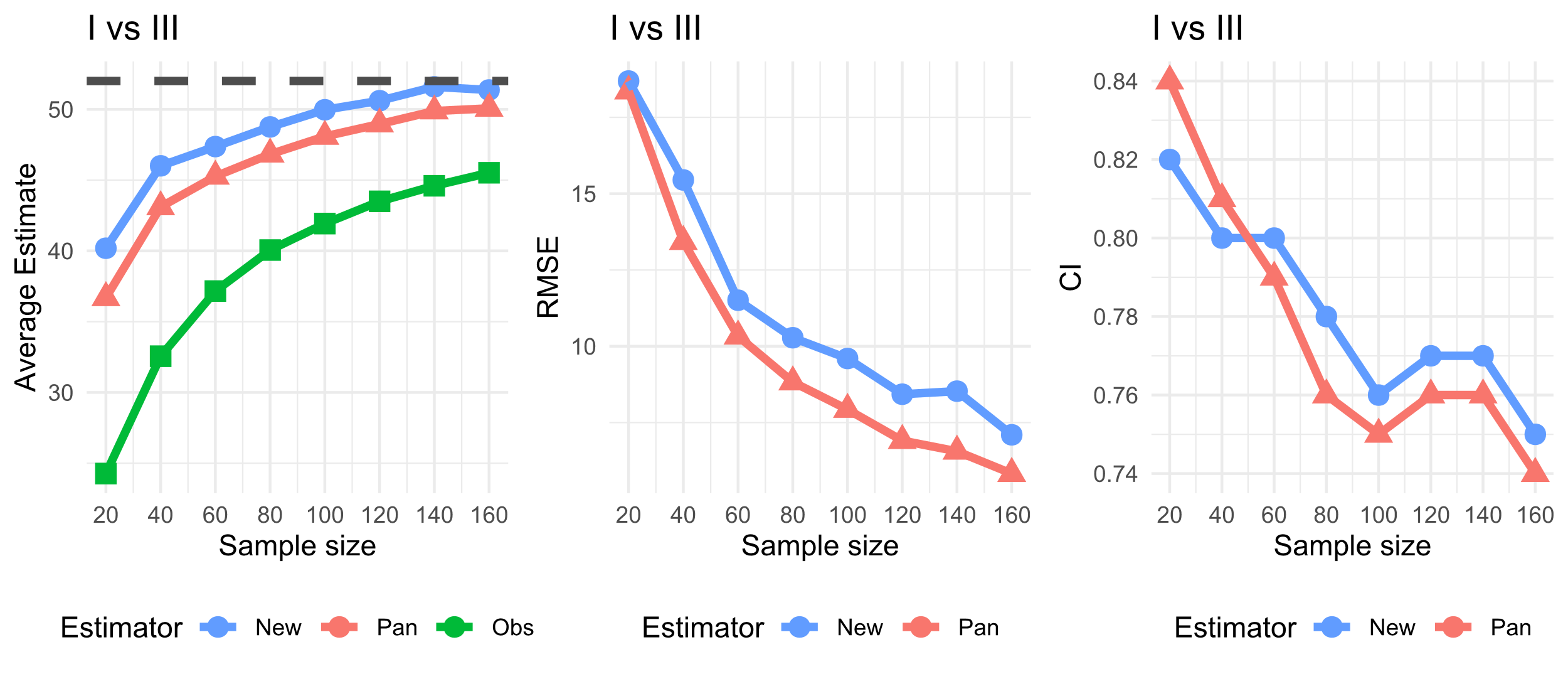


Figure 4.10取後放回澳洲三種極端氣候鳥類資料集中的Big Dry與Post-Big Wet估計結果。

Table 4.11取後放回澳洲三種極端氣候鳥類資料集中的Big Wet與Post-Big Wet估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sample size | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 20 | New | 30.82 | 45.86 | -10.14 | 14.51 | 12.74 | 17.7 | 0.8 |
| Pan | 42.88 | -13.12 | 10.5 | 8.99 | 16.8 | 0.81 |
| 60 | New | 42.91 | 51.52 | -4.48 | 9.09 | 8.42 | 10.13 | 0.8 |
| Pan | 49.88 | -6.12 | 6.76 | 6.13 | 9.12 | 0.8 |
| 100 | New | 47.31 | 54.12 | -1.88 | 8.64 | 7.23 | 8.84 | 0.76 |
| Pan | 52.73 | -3.27 | 6.3 | 5.29 | 7.1 | 0.77 |
| 140 | New | 49.53 | 55.47 | -0.53 | 7.6 | 6.77 | 7.62 | 0.74 |
| Pan | 54.15 | -1.85 | 5.59 | 4.97 | 5.88 | 0.75 |

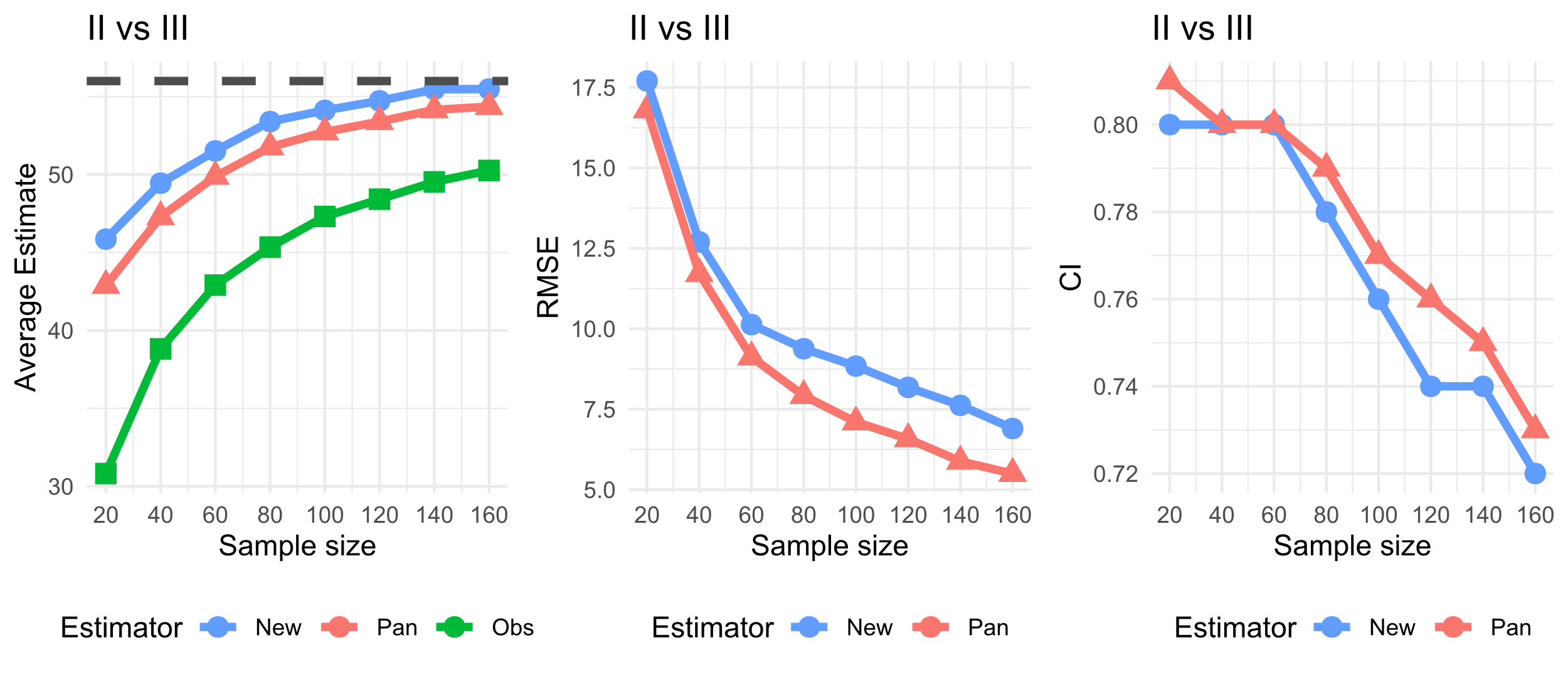


Figure 4.11取後放回澳洲三種極端氣候鳥類資料集中的Big Wet與Post-Big Wet估計結果。

### BCI

**資料描述**

本資料集中，針對1981-2016年間對Barro Colorado Island進行八次調查，紀錄該地區樹種的生長狀況。將該地區非為1250個區塊進行調查，每塊區塊的大小為400平方公尺。以第一次調查 (1981至1983年間) 與第八次調查 (2013至2016年間) 的結果作為兩群落之母體資料，並使用母體的10%、30%、50% 以及70%作為抽樣樣本大小。在兩母體中，第一次調查作為群落一 (Mean =0.18, CV = 1.34)，第八次調查作為群落二 (Mean =0.18, CV = 1.35)，在群落一中，包含了307種物種，而群落二中則擁有299種物種，兩群落的共同物種為284種。

**模擬結果**

在該筆資料中使用取後不放回的估計方法下，假設該筆資料為母體，並抽取其中的區塊做為樣本所使用。在模擬結果中，使用*wNew1*的估計方法在估計偏差方面，與真值差距最小。在抽樣比例為0.1時的估計標準差 (Est. SD) ，與樣本標準差 (Sample SE) 存在較大的差異，而此差異會隨著抽樣比例增加差異逐漸減小。在估計式評估標準上，在兩種評估標準的結果表明，在小樣本中，*wNew1*與*wNew2*的估計表現明顯較*wChao2.12*的結果較好，但在隨著樣本變大，RMSE在三種估計方式中的評估結果沒有明顯差異；而在95% CI Coverage的模型評估上，*wNew1*在小樣本中的結果明顯較其他兩者好。

Table 4.12取後不放回BCI資料集中的兩群落估計結果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| q | Estimator | Obs | AVG Estimate | Bias | Sample SE | Est. SD | RMSE | 95% CI Coverage |
| 0.1 | wNew1 | 225.24 | 273.94 | -10.06 | 23.46 | 26.09 | 25.51 | 0.97 |
| wNew2 | 298.91 | 14.91 | 46.77 | 34.39 | 49.07 | 0.92 |
| wChao2.12 | 253.31 | -30.69 | 19.56 | 13.09 | 36.38 | 0.81 |
| 0.3 | wNew1 | 256.57 | 283.48 | -0.52 | 11.3 | 10.55 | 11.31 | 0.94 |
| wNew2 | 288.77 | 4.77 | 14.42 | 11.73 | 15.18 | 0.9 |
| wChao2.12 | 272.93 | -11.07 | 9.05 | 7.26 | 14.3 | 0.84 |
| 0.5 | wNew1 | 269.18 | 284.51 | 0.51 | 6.35 | 5.87 | 6.37 | 0.93 |
| wNew2 | 285.69 | 1.69 | 6.79 | 5.86 | 6.99 | 0.92 |
| wChao2.12 | 279.89 | -4.11 | 5.77 | 4.84 | 7.08 | 0.84 |
| 0.7 | wNew1 | 276.61 | 284.09 | 0.09 | 3.37 | 3.31 | 3.37 | 0.95 |
| wNew2 | 284.32 | 0.32 | 3.41 | 3.27 | 3.43 | 0.94 |
| wChao2.12 | 282.57 | -1.43 | 3.3 | 3.01 | 3.59 | 0.87 |

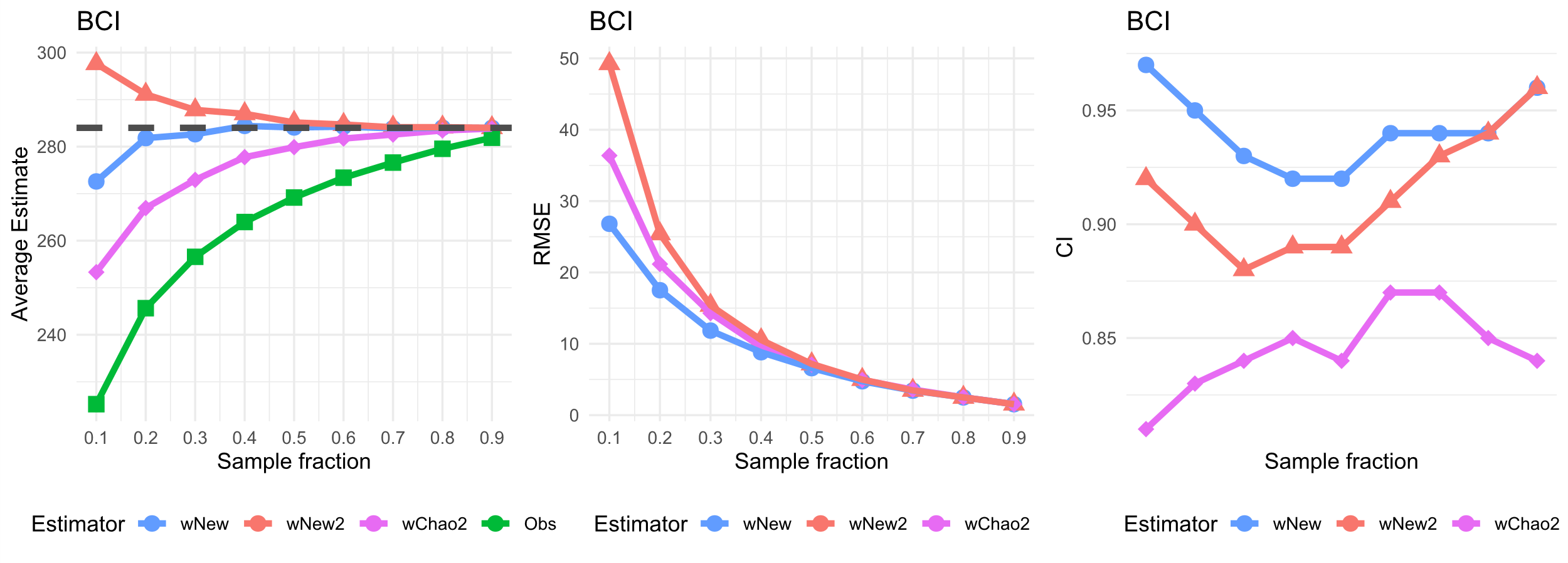


Figure 4.12取後不放回BCI資料集中的兩群落估計結果。

## 討論

總結上述結果，在第4章中使用了電腦模擬建立不同模型假設下的四種群落搭配四種不同的物種分配方式，以及分別使用兩筆真實資料作為群落母體。並從中隨機以取後放回以及取後不放回兩種不同的抽樣方式，重複抽取1000次，以計算平均的估計值、偏差、樣本標準差、估計標準差、RMSE以及95% CI Coverage的估計結果。

在取後放回的估計結果中，在第一種、第二種與第三種物種與區塊數的假設下，兩種估計方法都低估了共同種數，尤其在小樣本時更為明顯。*New*方法相較於*Pan*方法在偏差和RMSE方面表現較好。而95%信賴區間的涵蓋率方面，*New*方法也略優於*Pan*方法。但是，在第四種假設下，*New*方法偏差較小但容易發生高估的情況，而RMSE較差，且隨著樣本增加，95% CI Coverage也會隨之下降。

在取後放回的真實資料方面，以澳洲三種極端氣候鳥類資料進行模擬的結果中，*New*估計方式比*Pan*具有更好的表現，誤差較小。而在評估標準上，*New*在小樣本中的RMSE雖稍遜色於*Pan*，但隨樣本大小的增加*New*在RMSE的表現會較*Pan*好，在95% CI Coverage上兩種估計方式表現相近。

另一方面，在取後不放回的估計中，大多數情況下樣本涵蓋率都高於0.9。隨著抽樣比例增加，樣本變異係數趨於接近母體變異數。在所有的物種與區塊數假設，不同模型和組合下，*wNew1*和*wNew2*方法在小樣本中有稍微高估現象發生，在大樣本中則普遍為低估。在RMSE方面，*wNew1*和*wNew2*表現相近且皆優於*wChao2.12*。在95% CI Coverage方面，*wNew1*表現為所有中最佳者。而在真實資料的部分，BCI資料的模擬結果顯示，使用*wNew2*估計方法在小樣本中與真值的差異最小。抽樣比例為0.1時，平均估計的樣本標準差與實際樣本標準差存在較大差異，但隨著抽樣比例增加，差異逐漸減小。在評估中，RMSE和95% CI Coverage顯示，在小樣本中，*wNew1*和*wNew2*比*wChao2.12*表現更好。隨著樣本增加，三種估計方法的RMSE沒有明顯差異，但在95% CI Coverage方面，*wNew1*在小樣本中表現明顯優於其他兩者。

總結上述，本文比較了取樣後放回和不放回的估計方法對群落共同種估計的影響。兩者在樣本涵蓋率、變異性和偏差方面表現類似。兩種抽樣方法中所獲得估計結果，除了取後不放回的小樣本之外，其餘皆低估了共同物種數。且在大多數情況下，樣本涵蓋率在大多數情況下都高於0.9。總體而言，綜觀模擬出的各項指標與估計式評估結果，本文所提出的建議方法 (*New*, *wNew1*, *wNew2*) 在多數情況下表現優於原有方法 (*Pan*, *wChao2.12*)。但在取後不放回的估計中某些特定情況，特別是在小樣本和低估計方面，可能存在高估的風險，且對於不同假設條件下的估計結果表現也有所不同。