Supplemental Materials to
"Model-based Differentially Private Data Synthesis"

Fang Liu[*]

Department of Applied and Computational Mathematics and Statistics

University of Notre Dame, Notre Dame, IN 46556

fang.liu.131@nd.edu

January 9, 2017
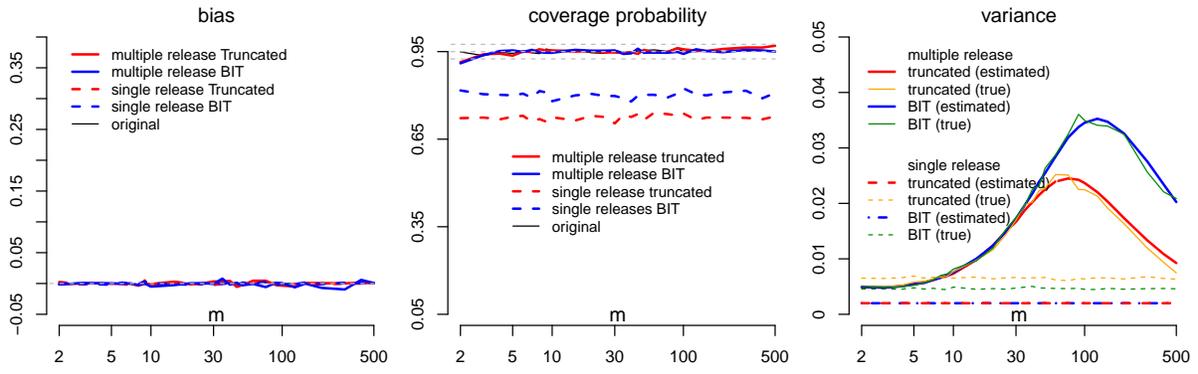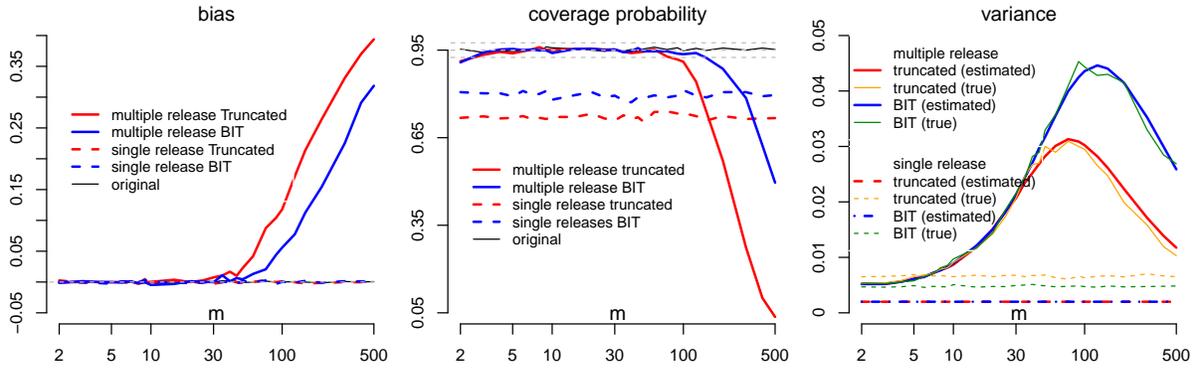
(a) $n = 1000, (c_0, c_1) = (-4, 4)$

(b) $n = 1000, (c_0, c_1) = (-4, 5)$
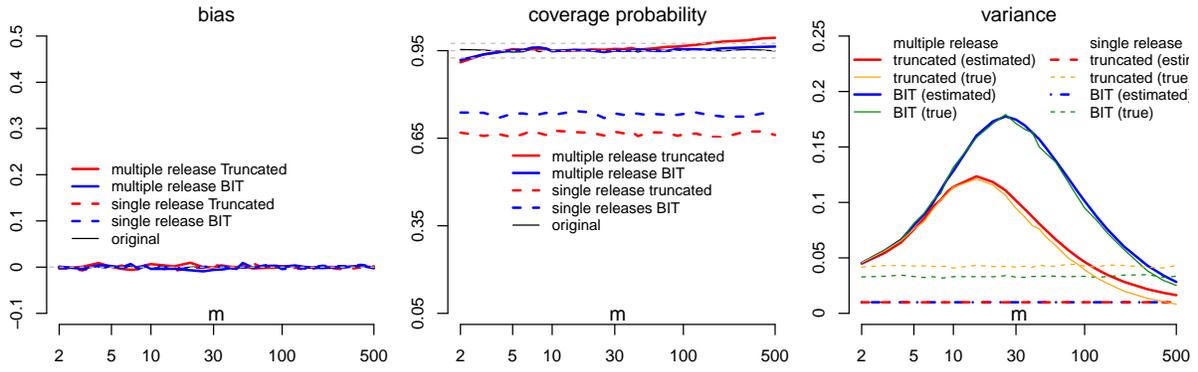
(c) $n = 100, (c_0, c_1) = (-4, 4)$
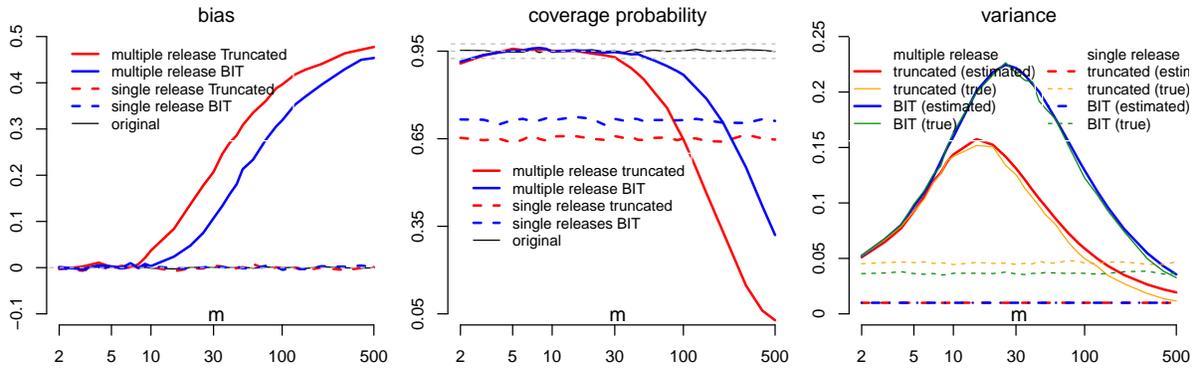
(d) $n = 100, (c_0, c_1) = (-4, 5)$

Figure S1: inferences based on synthetic data via modips ($m \in [2, 500]$)

(a) $s = -0.03$

(b) $s = 0.09$
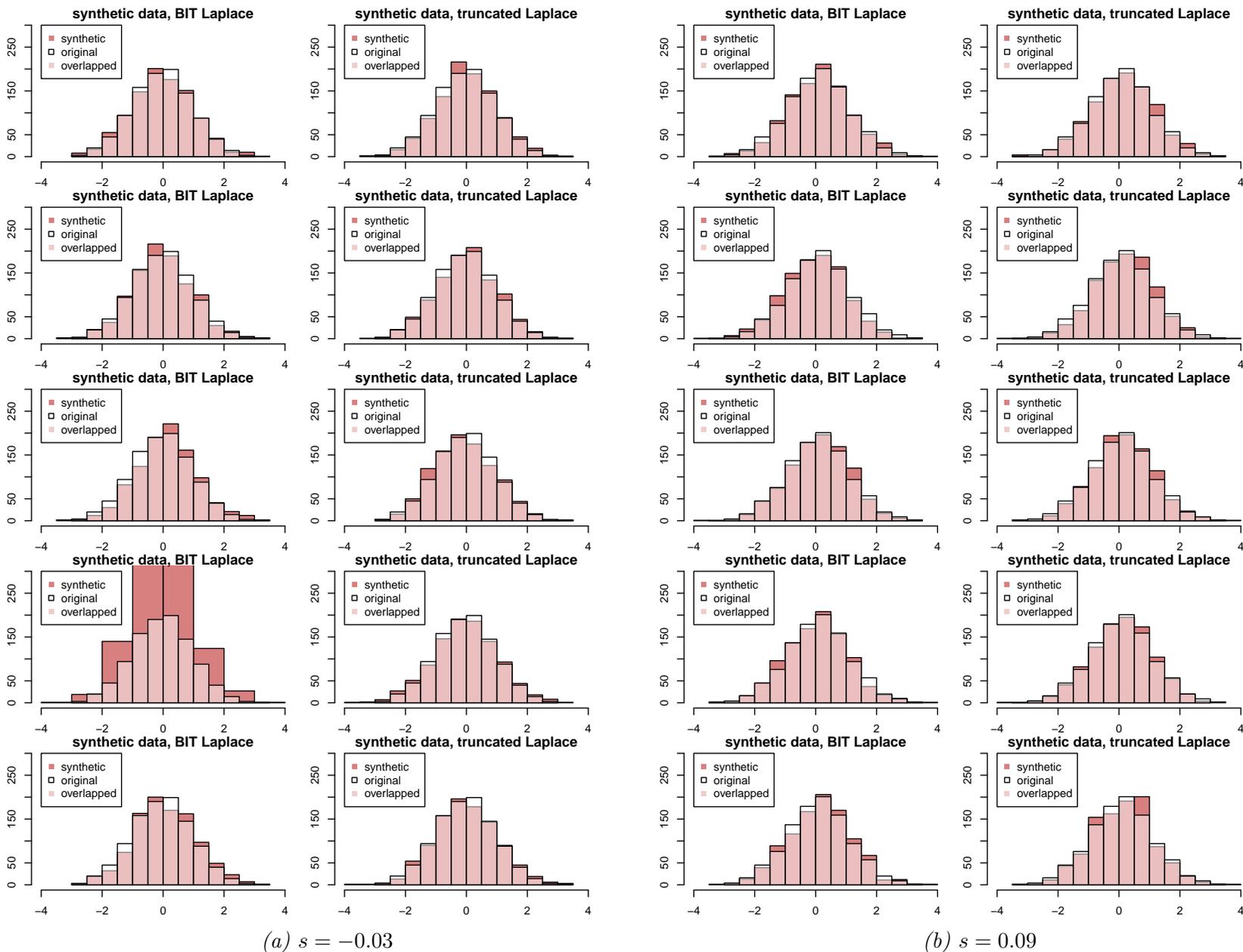
Figure S2: examples of synthetic data via modips in the simulation study ($n = 1000, m = 5, [c_0, c_1] = [-4, 4]$)
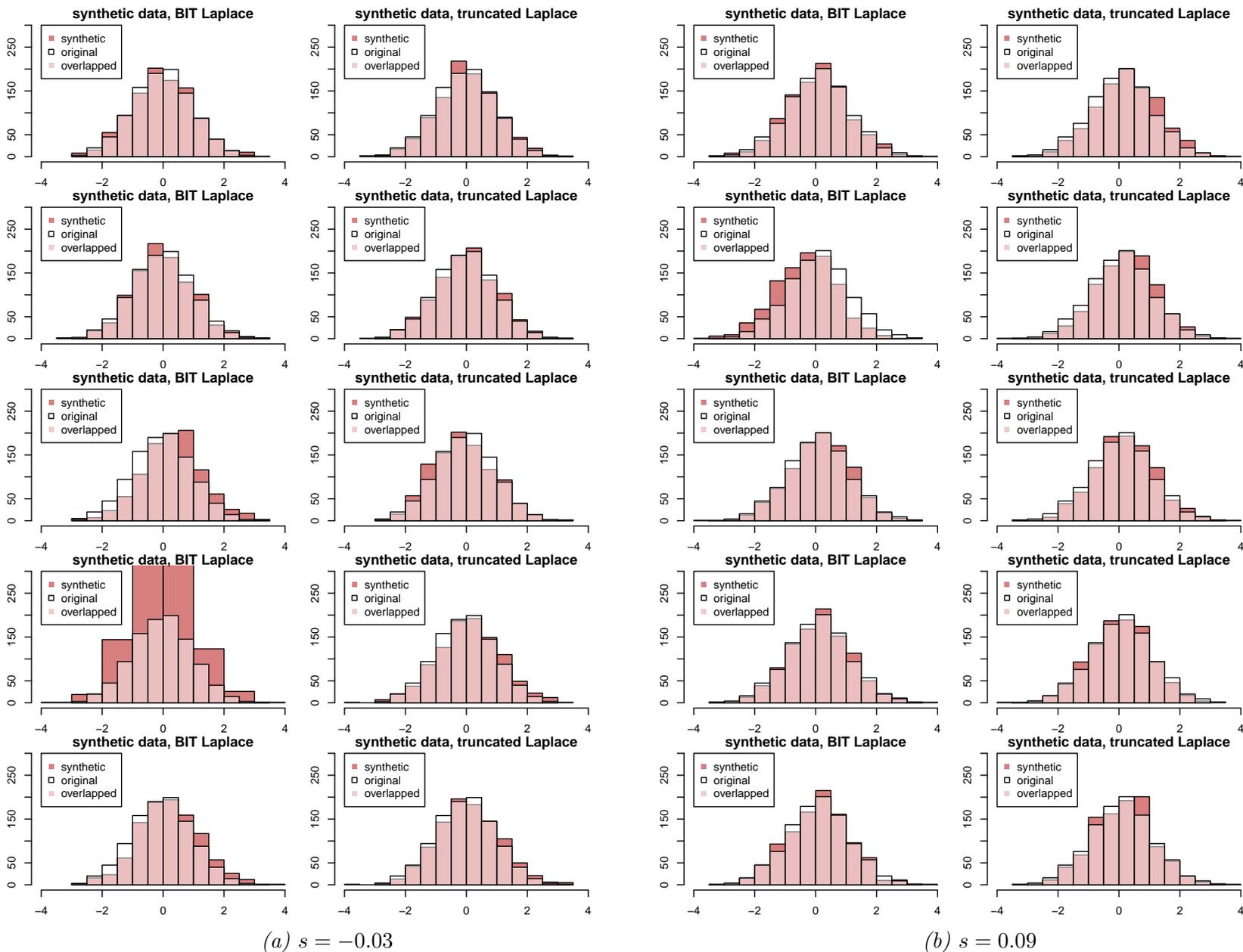
Figure S3: examples of synthetic data via modips in the simulation study ($n = 1000, m = 15, [c_0, c_1] = [-4, 4]$)

(a) $s = -0.03$

(b) $s = 0.09$
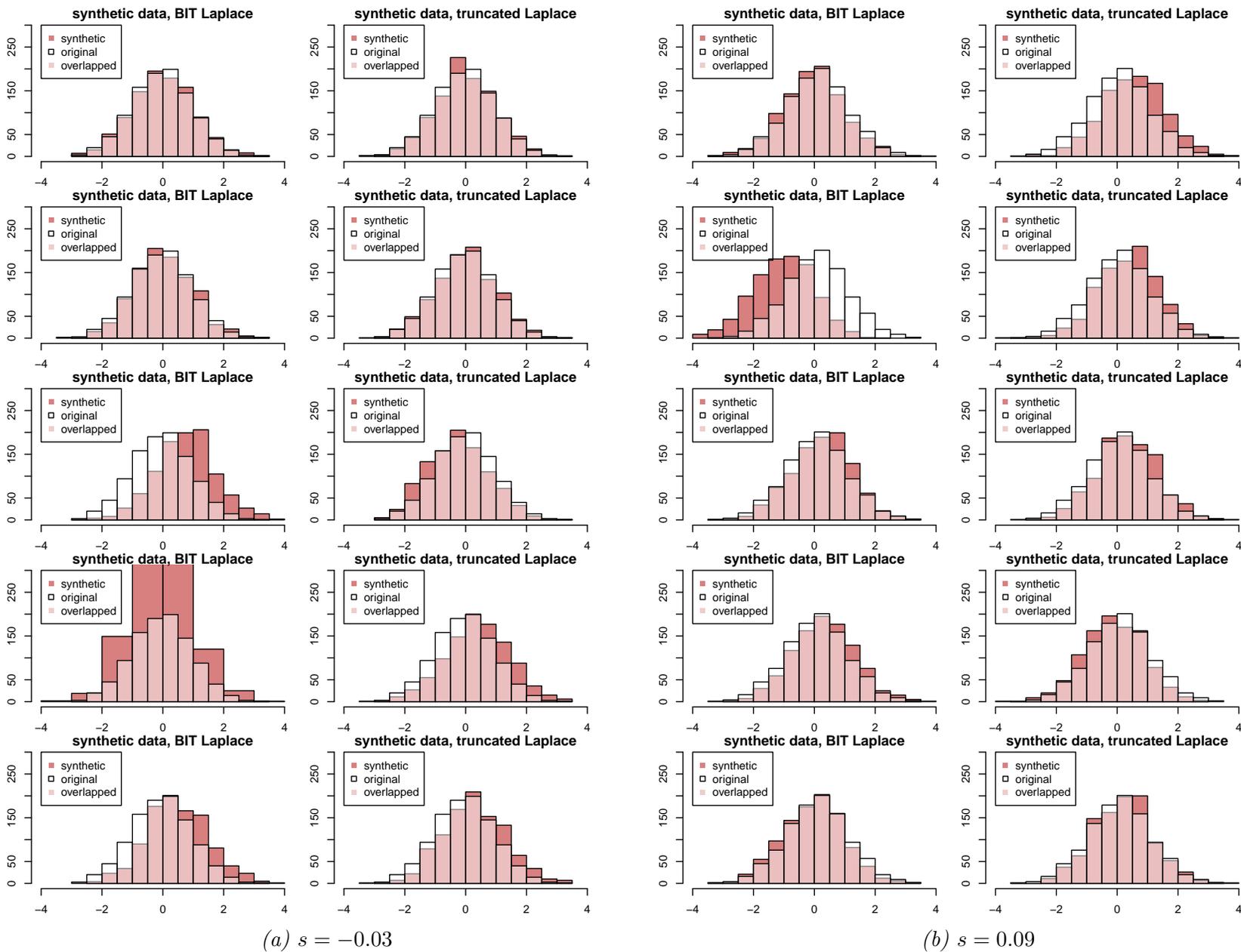
Figure S4: examples of synthetic data via modips in the simulation study $(n = 1000, m = 40, [c_0, c_1] = [-4, 4])$

6
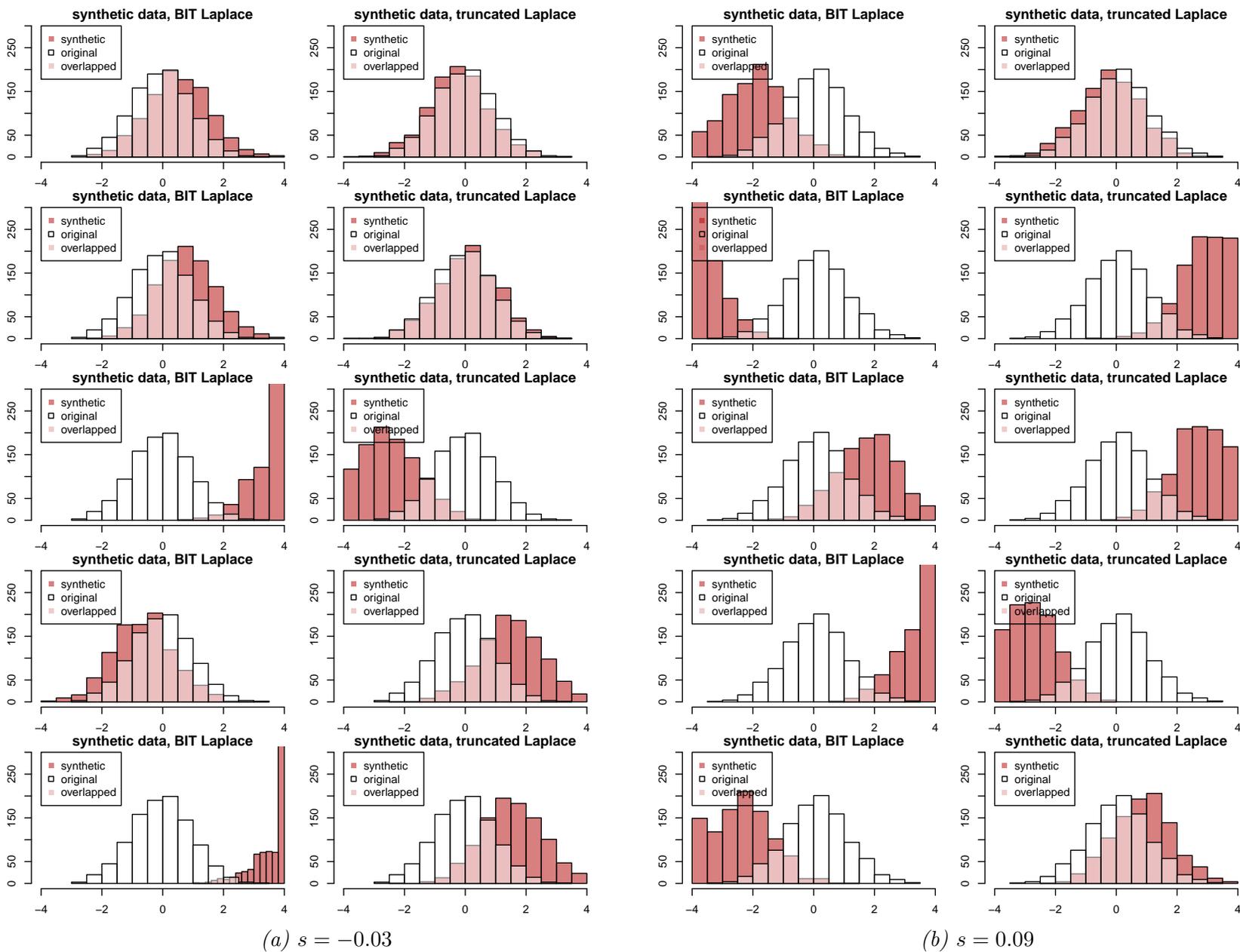
Figure S5: *examples of synthetic data via modips in the simulation study* ($n = 1000, m = 500, [c_0, c_1] = [-4, 4]$)
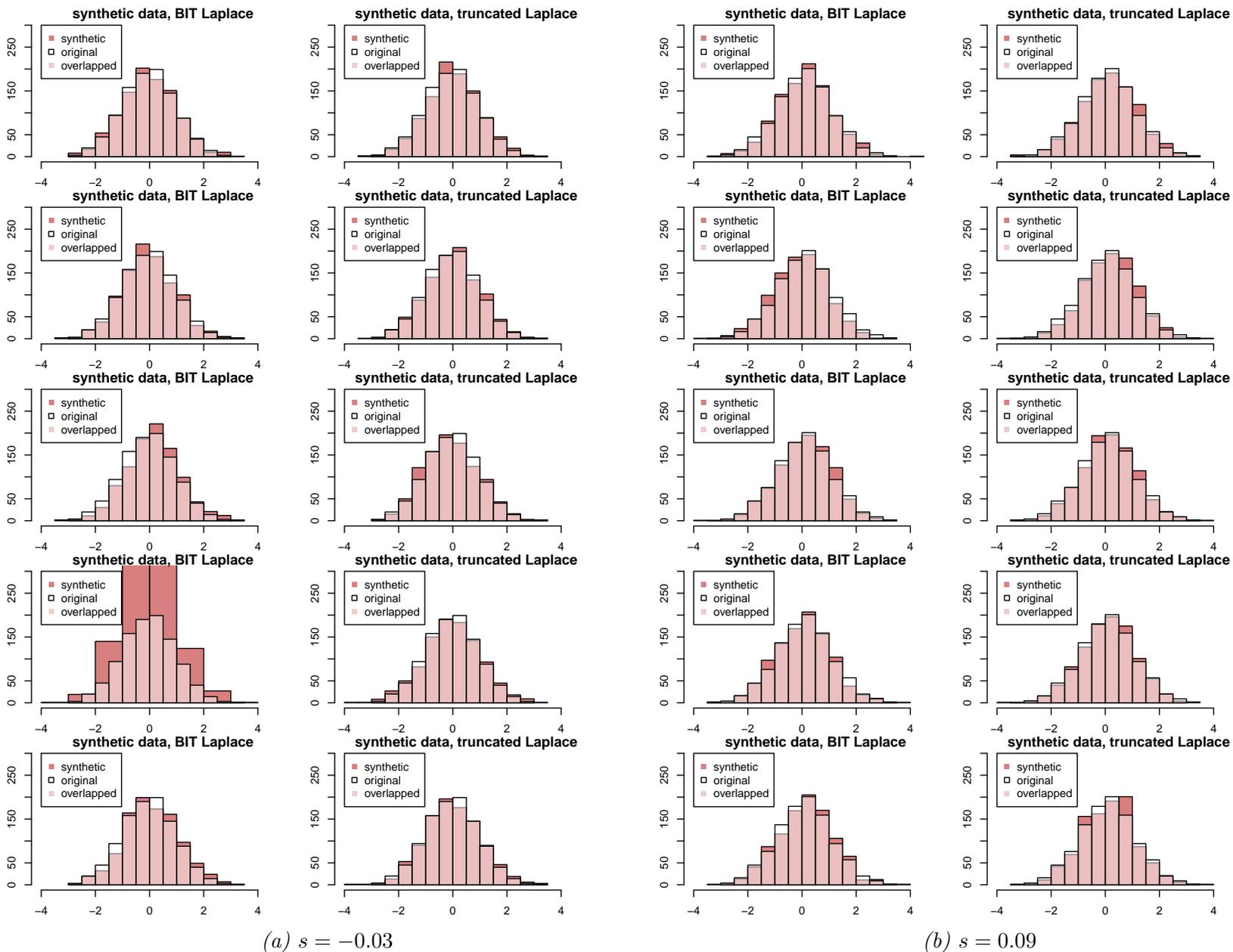
Figure S6: examples of synthetic data via modips in the simulation study ($n = 1000, m = 5, [c_0, c_1] = [-4, 5]$)
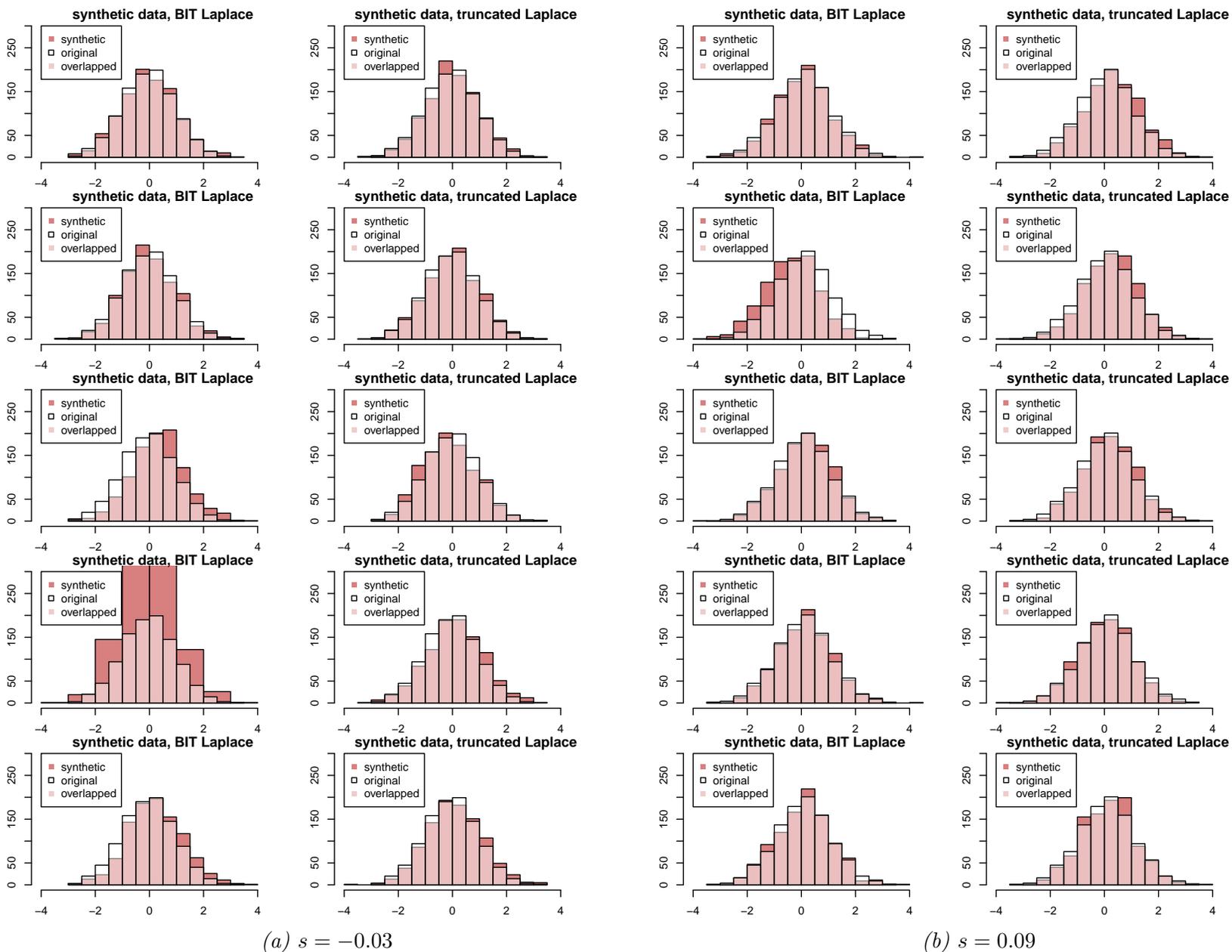
Figure S7: examples of synthetic data via modips in the simulation study $(n = 1000, m = 15, [c_0, c_1] = [-4, 5])$
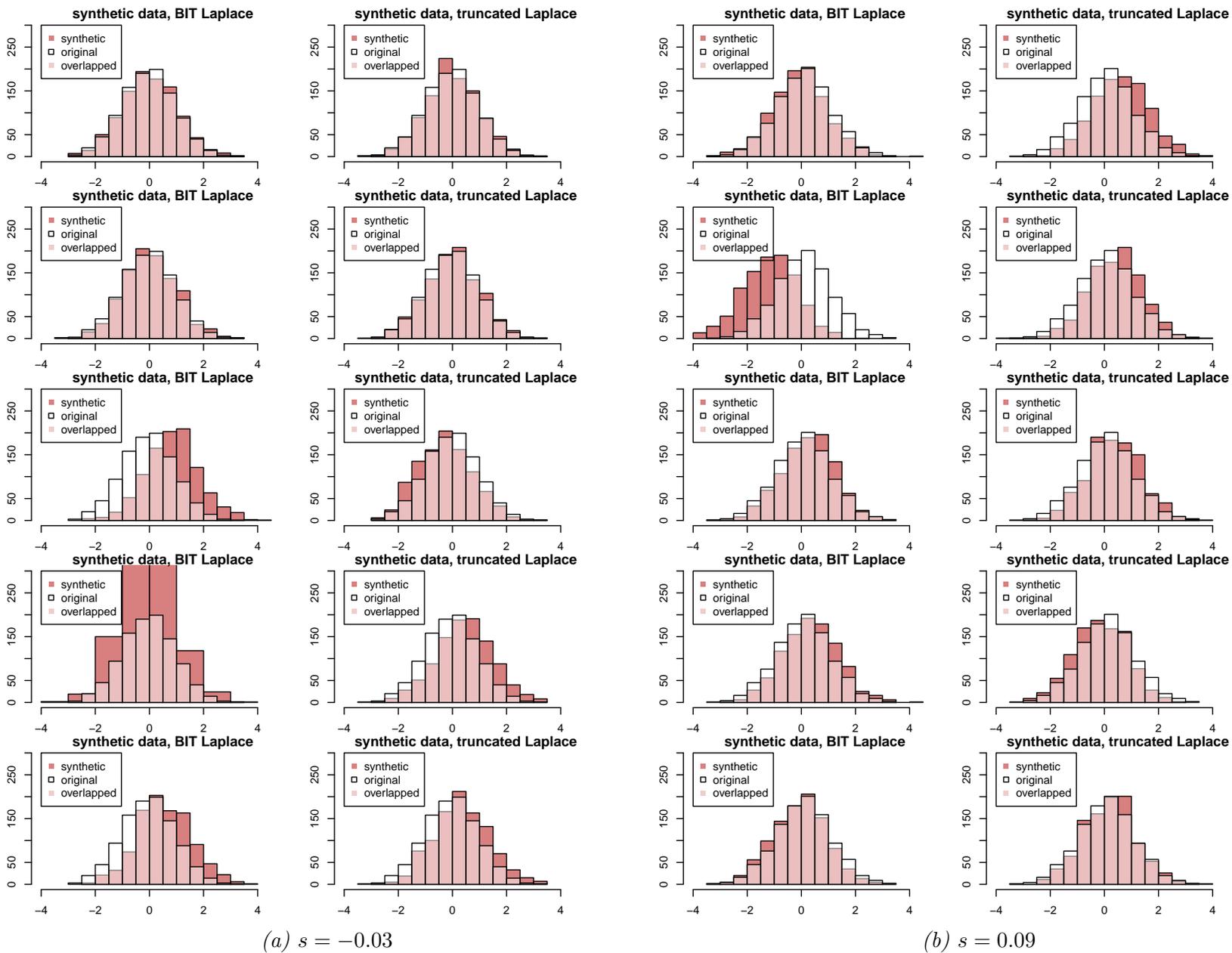
Figure S8: examples of synthetic data via modips in the simulation study ($n = 1000, m = 40, [c_0, c_1] = [-4, 5]$)
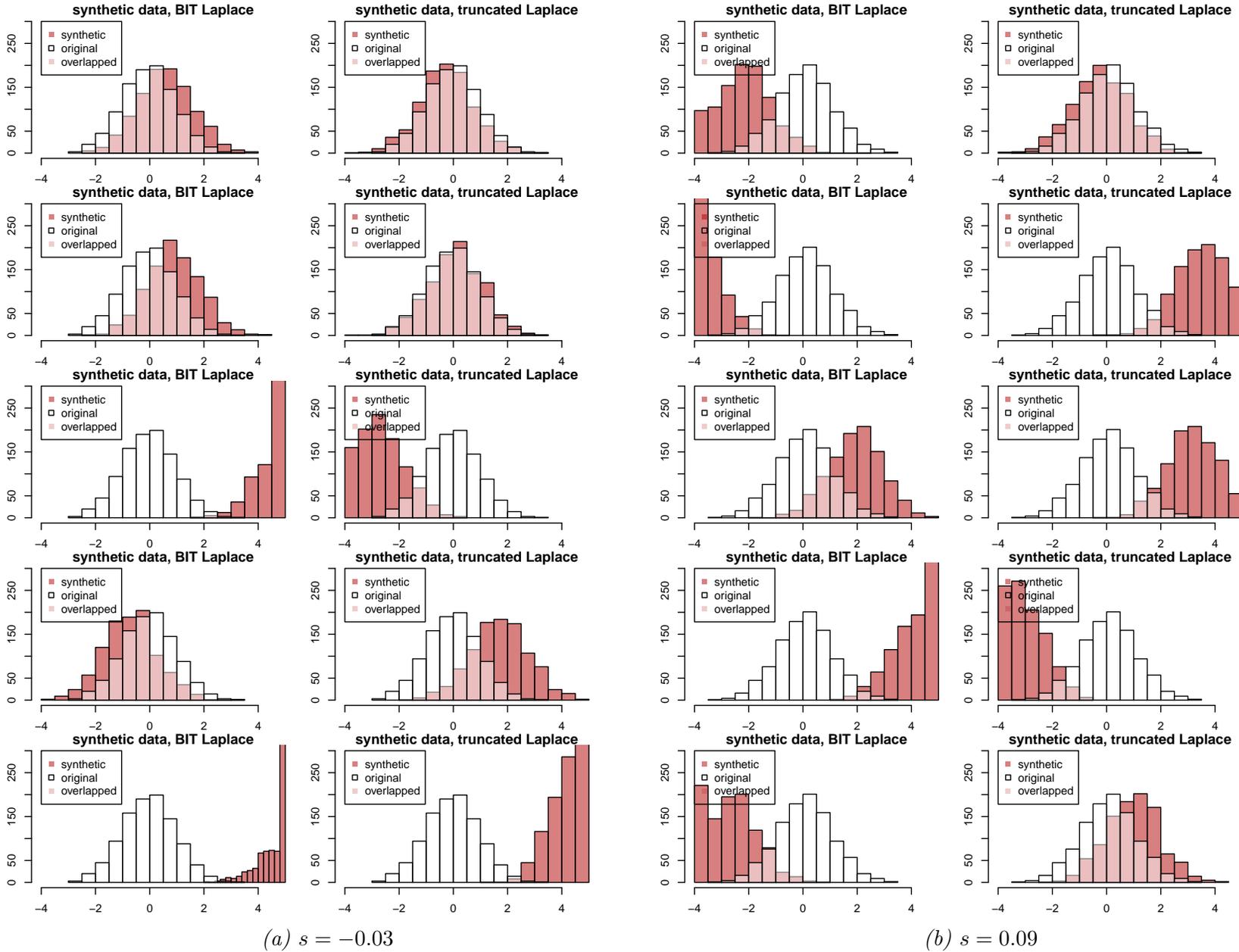
Figure S9: examples of synthetic data via modips in the simulation study ($n = 1000, m = 500, [c_0, c_1] = [-4, 5]$)
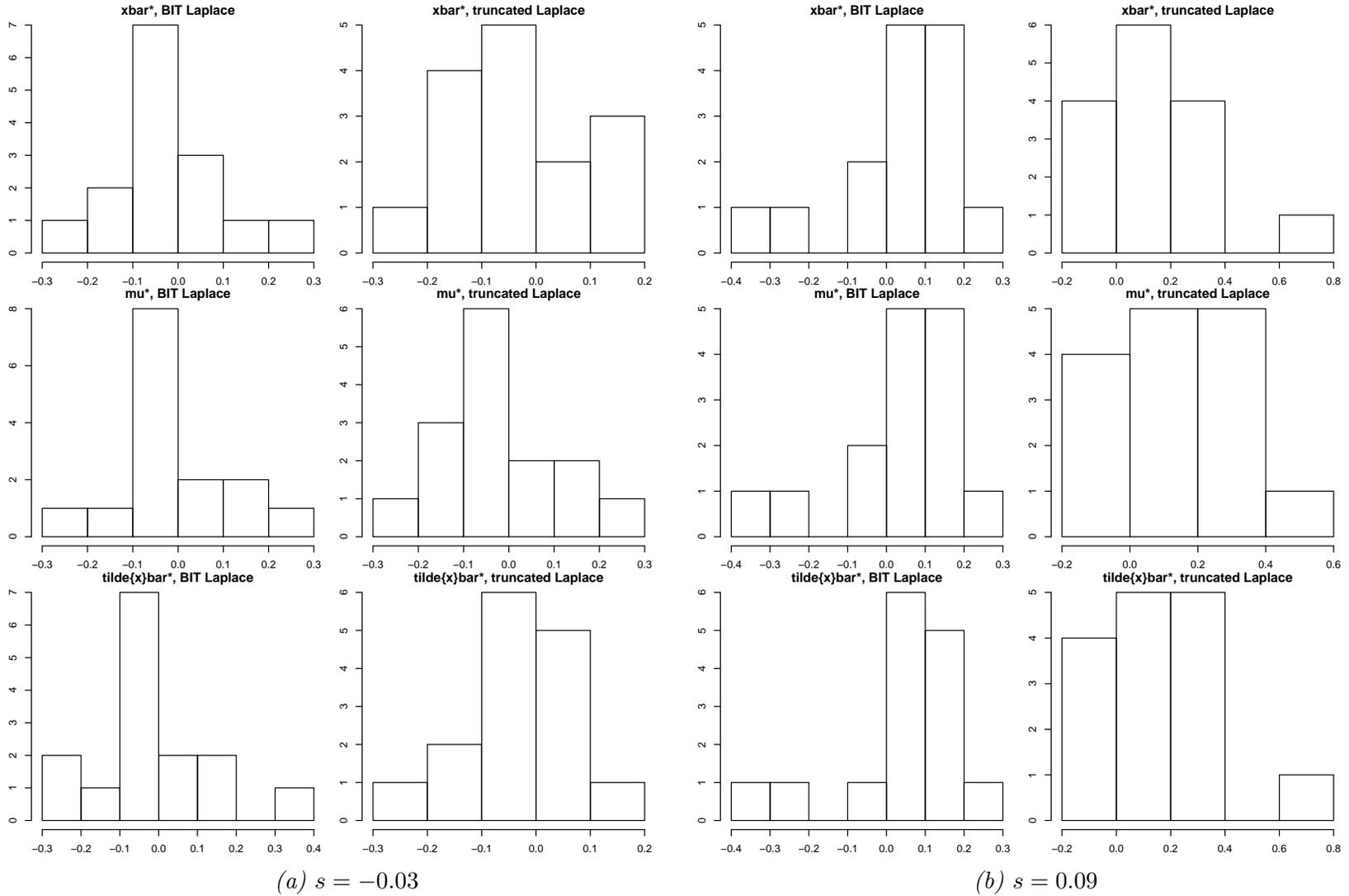
Figure S10: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 1000, m = 15, [c_0, c_1] = [-4, 4]$)

11
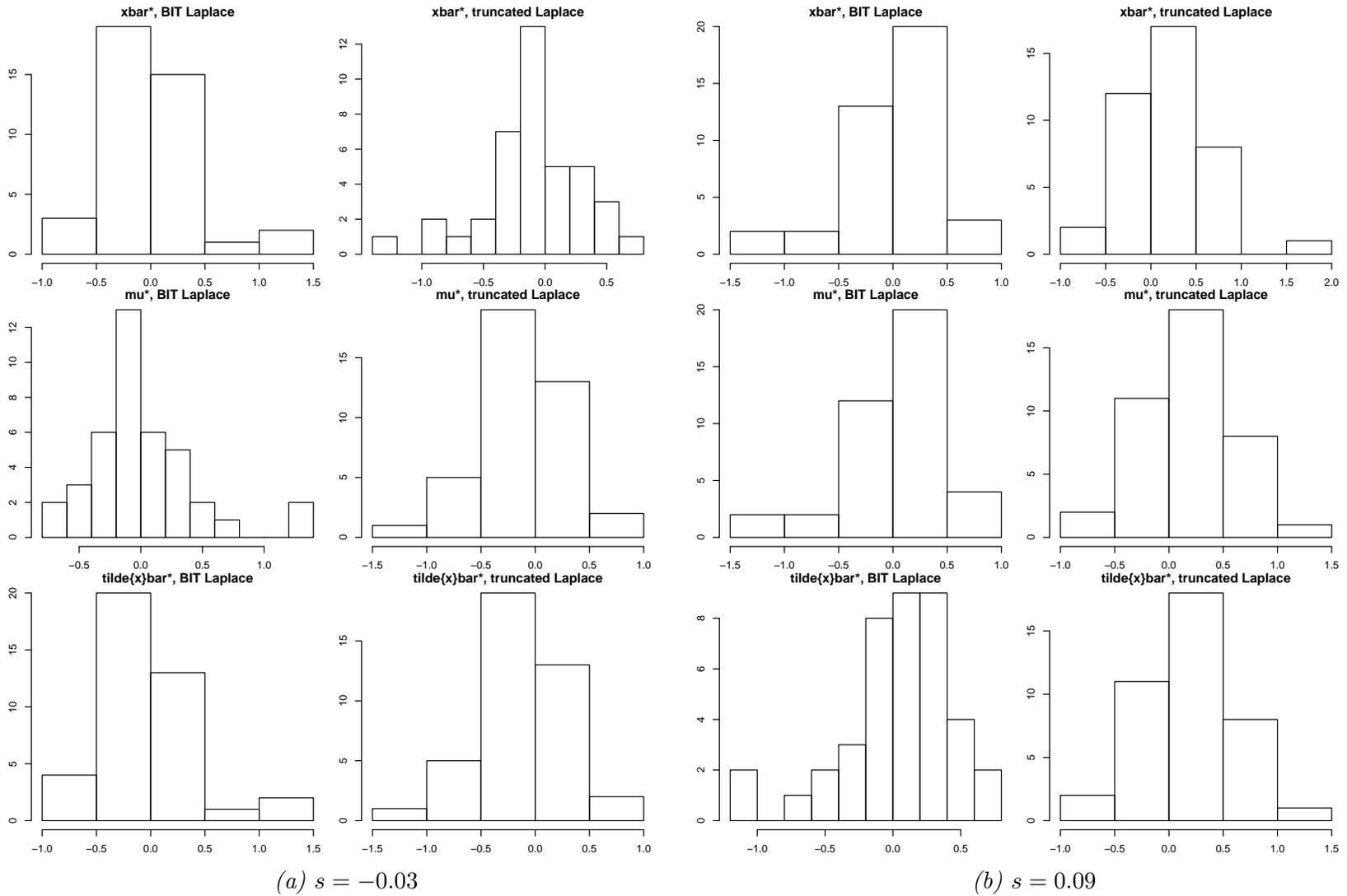
Figure S11: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 1000, m = 40, [c_0, c_1] = [-4, 4]$)

(a) $s = -0.03$

(b) $s = 0.09$
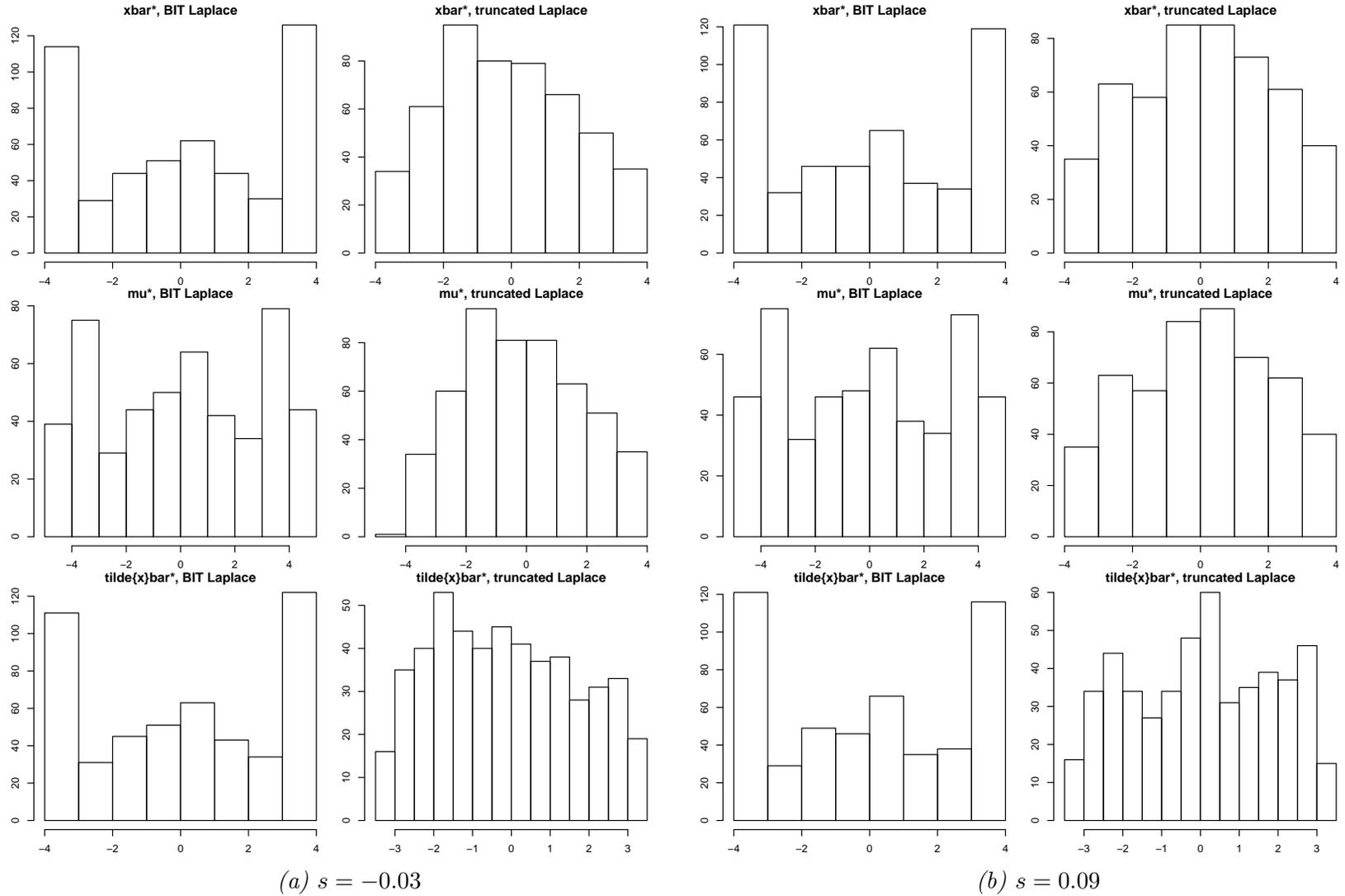
Figure S12: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 1000, m = 500, [c_0, c_1] = [-4, 4]$)

(a) $s = -0.03$ (b) $s = 0.09$
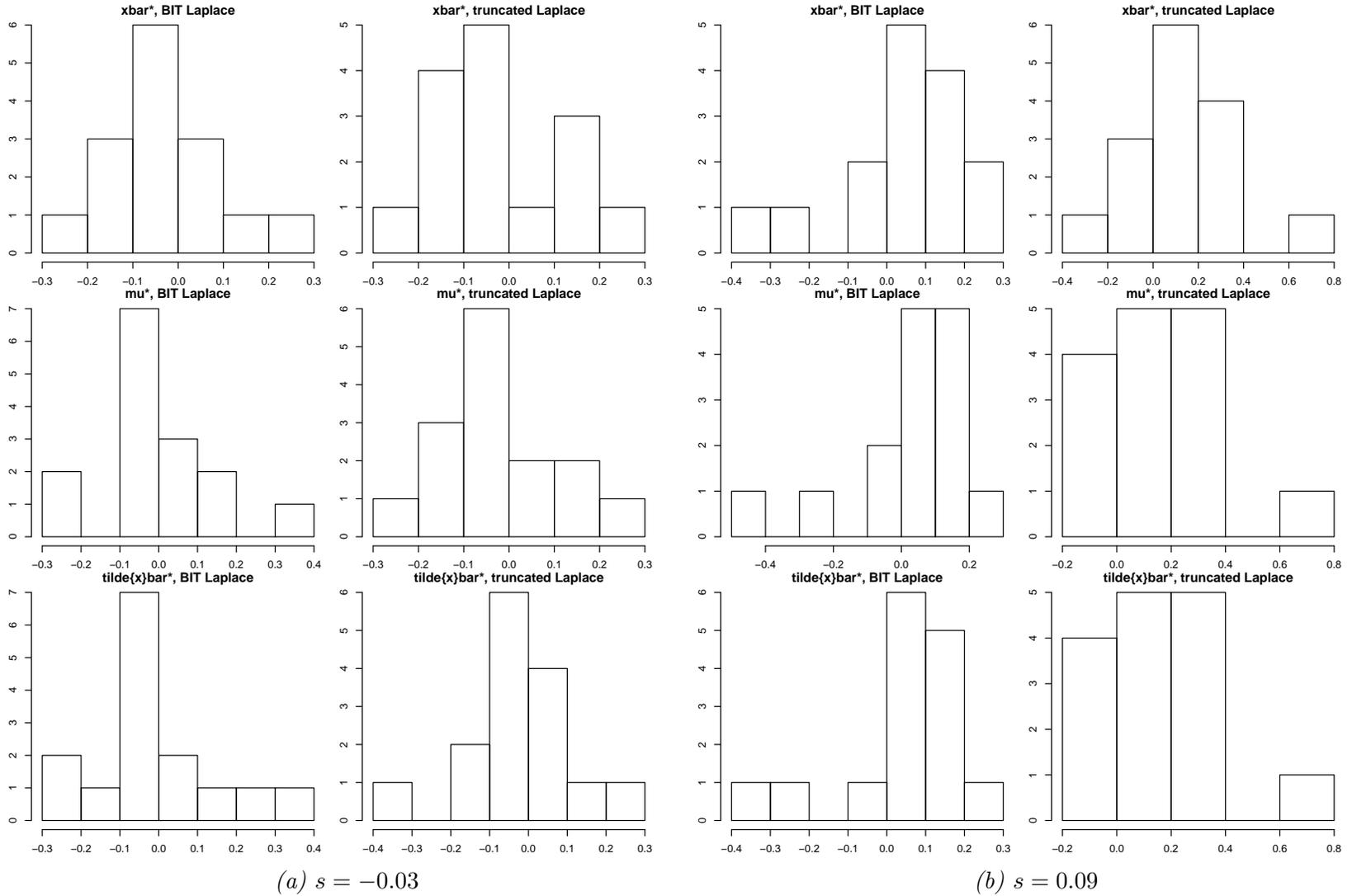
Figure S13: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study $(n = 1000, m = 15, [c_0, c_1] = [-4, 5])$

(a) $s = -0.03$

(b) $s = 0.09$
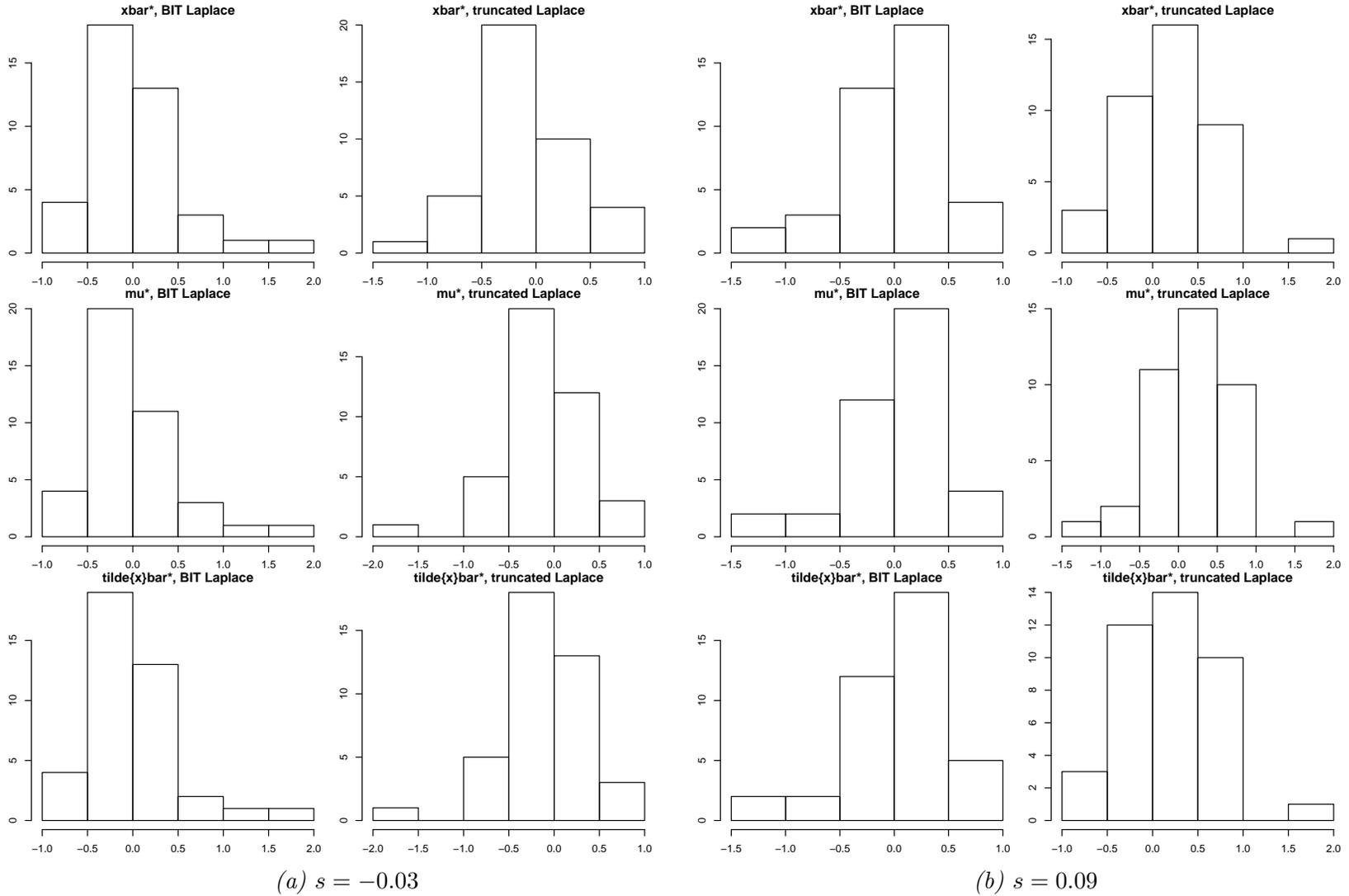
Figure S14: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across the m sets in the simulation study $(n = 1000, m = 40, [c_0, c_1] = [-4, 5])$
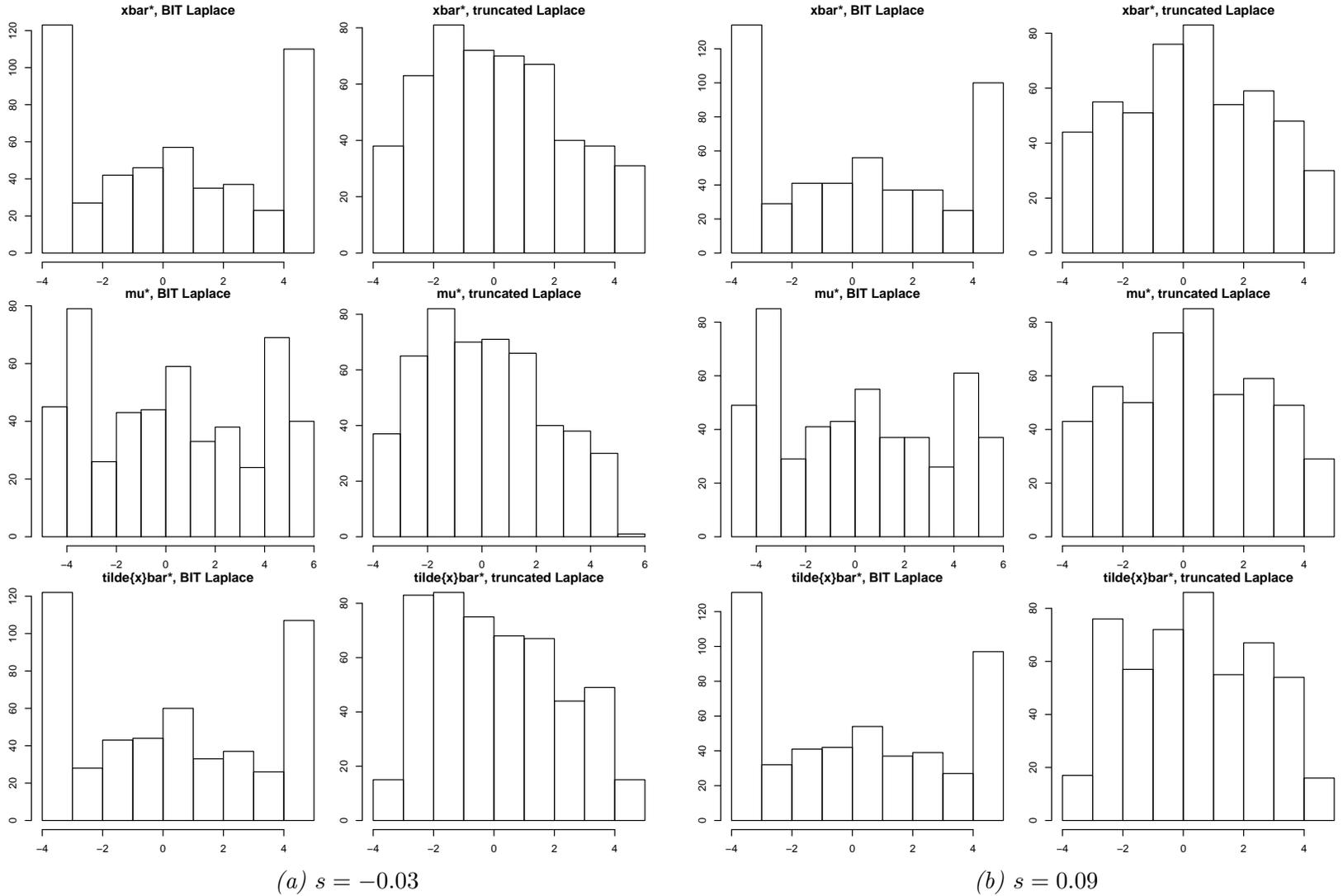
Figure S15: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 1000, m = 500, [c_0, c_1] = [-4, 5]$)
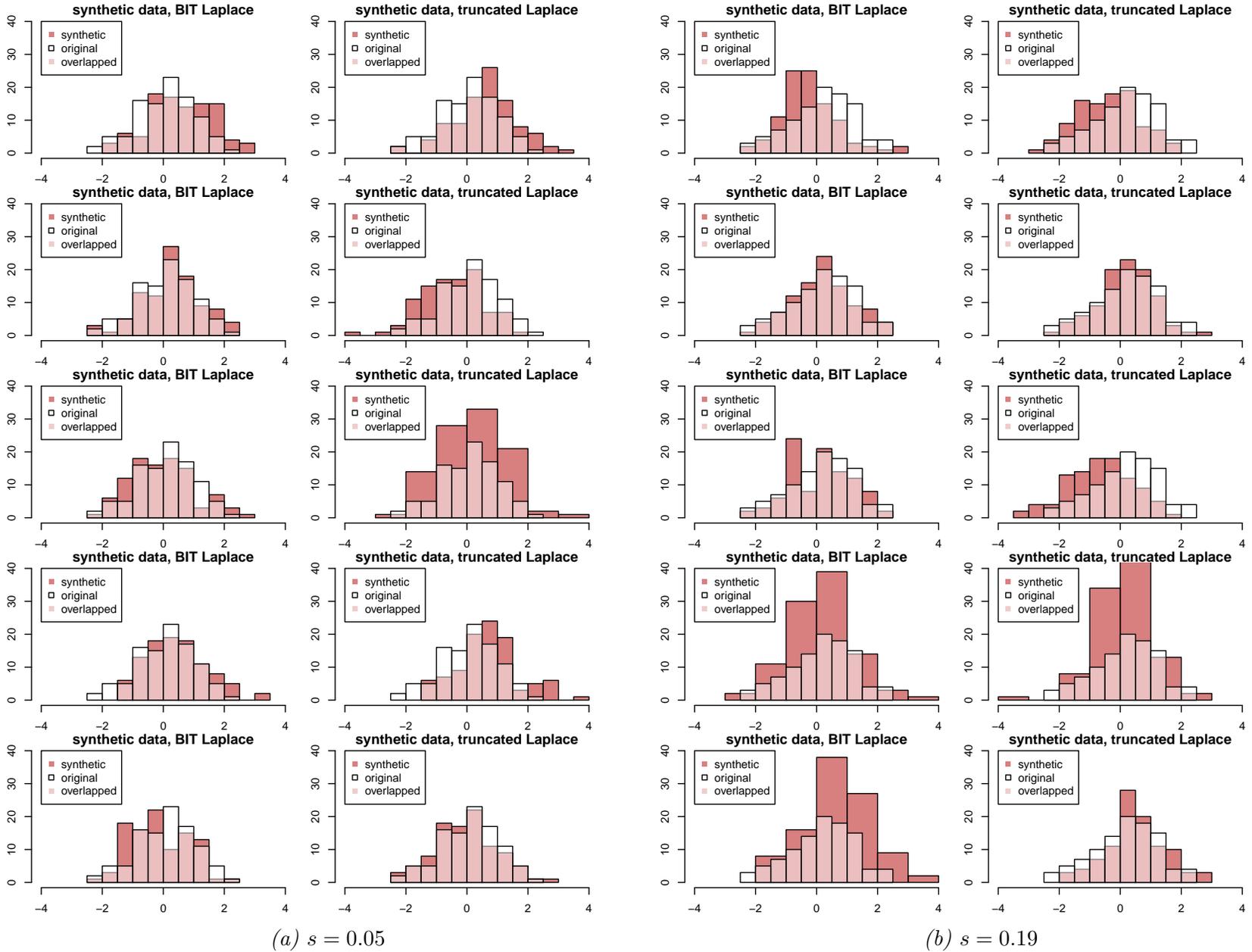
Figure S16: examples of synthetic data via modips in the simulation study ($n = 100, m = 5, [c_0, c_1] = [-4, 4]$)
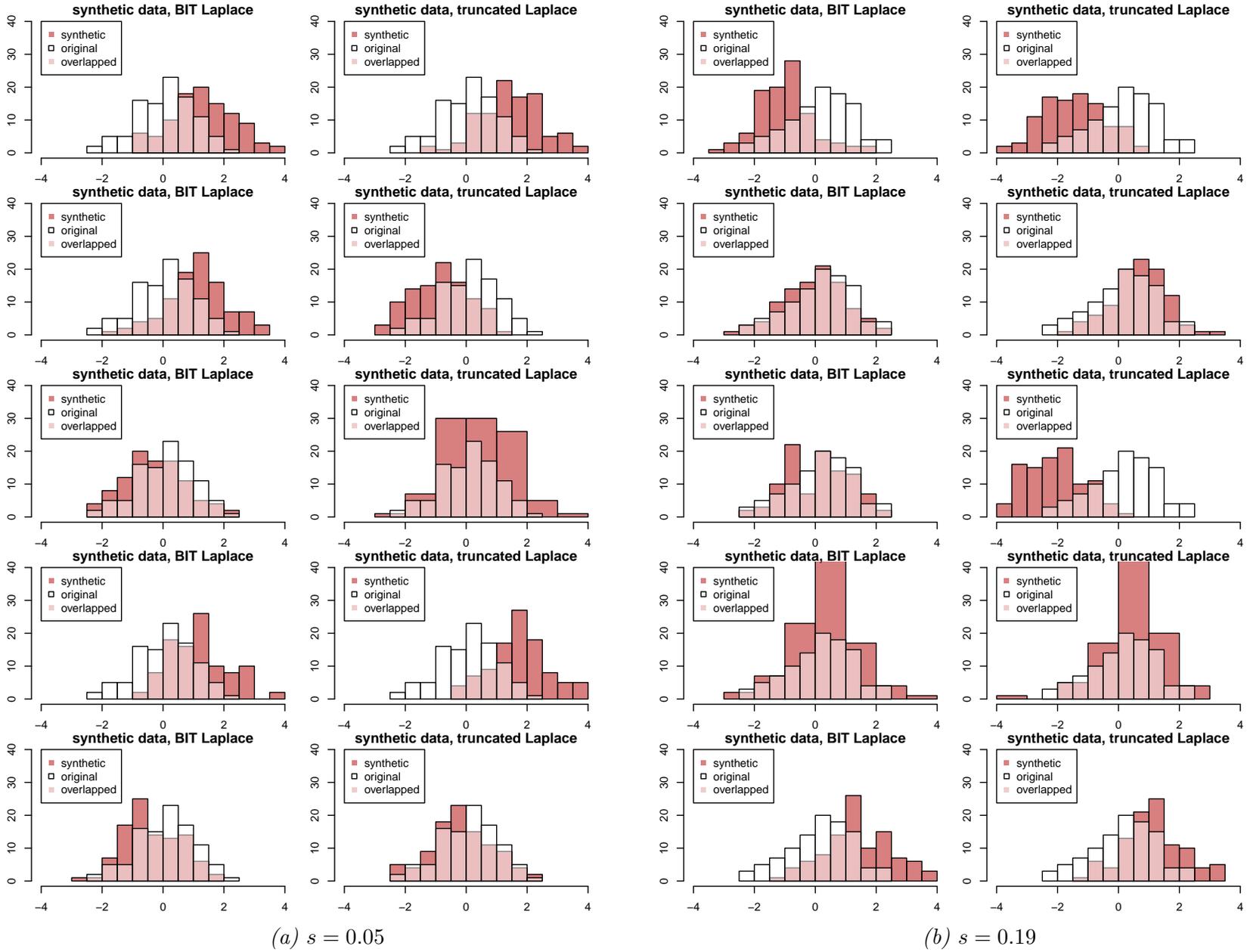
Figure S17: examples of synthetic data via modips in the simulation study ($n = 100, m = 15, [c_0, c_1] = [-4, 4]$)
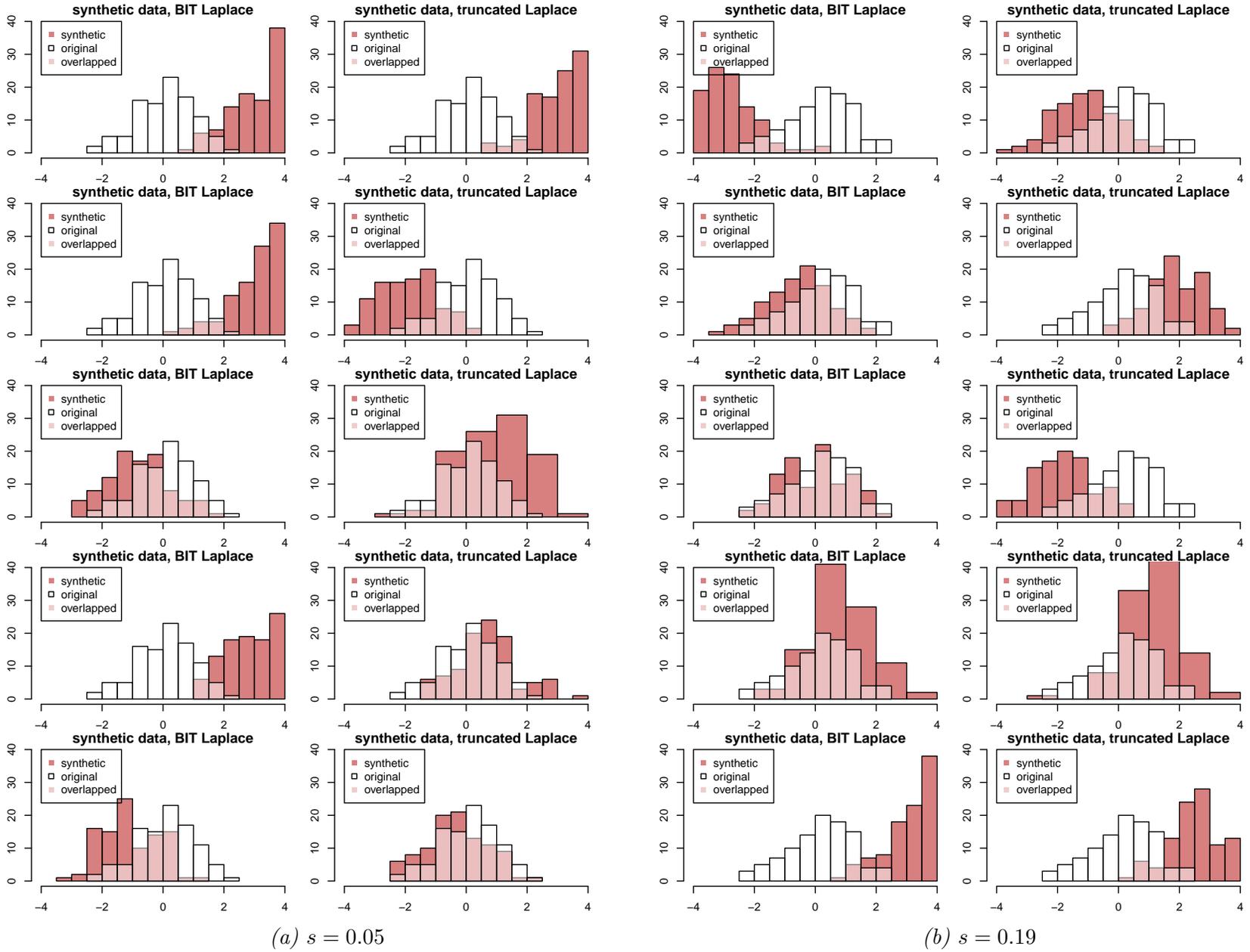
Figure S18: examples of synthetic data via modips in the simulation study $(n = 100, m = 40, [c_0, c_1] = [-4, 4])$
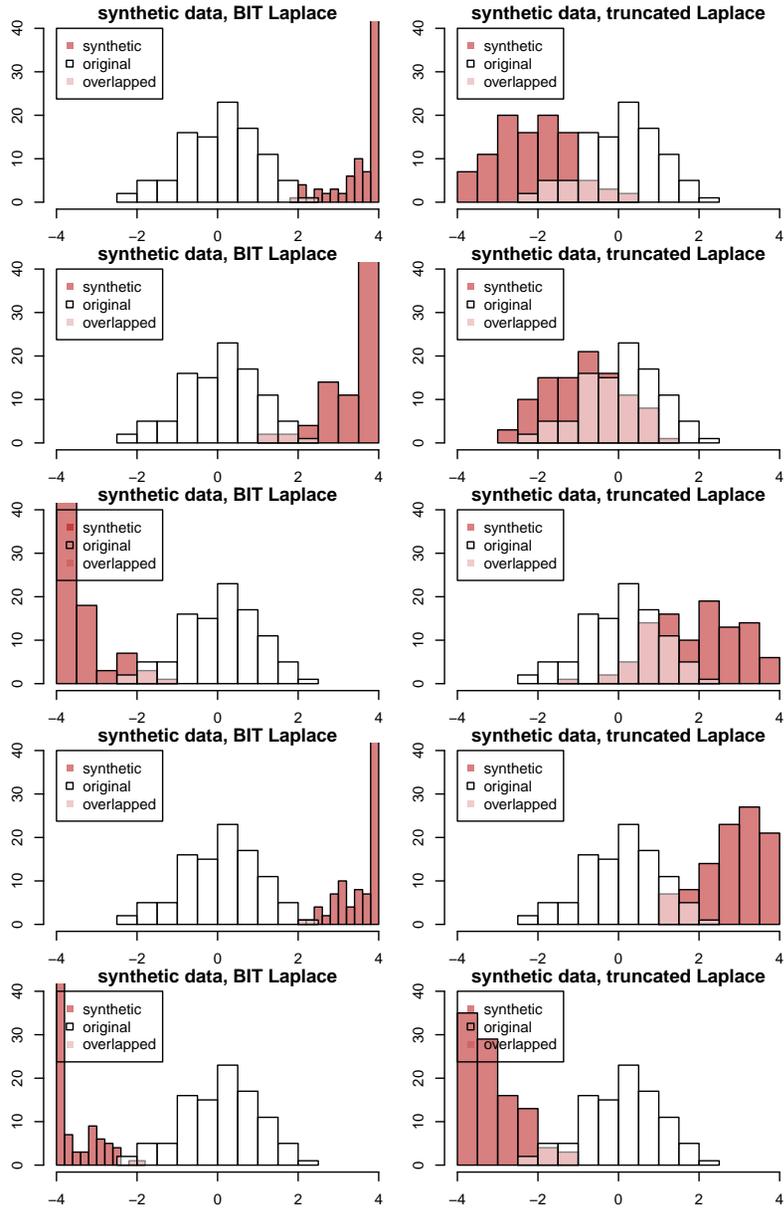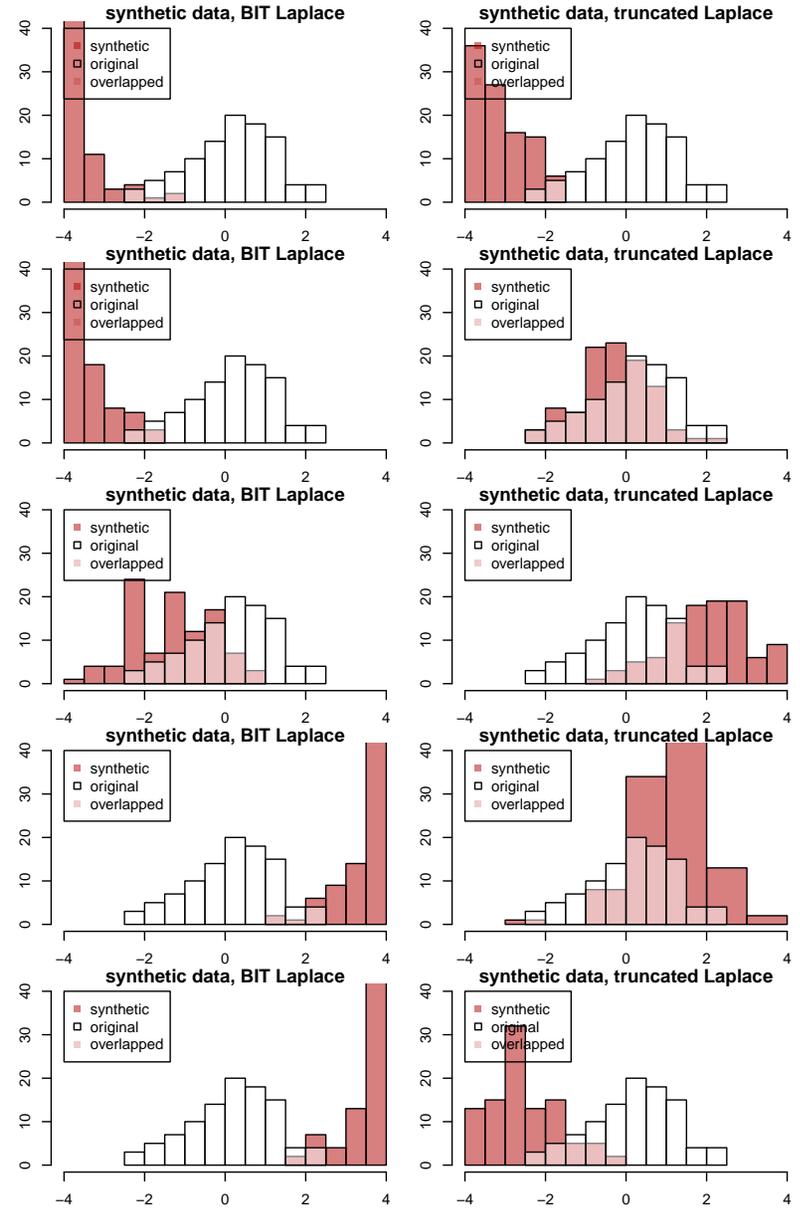
19

Figure S19: examples of synthetic data via modips in the simulation study ($n = 100, m = 500, [c_0, c_1] = [-4, 4]$)

(a) s = 0.05

(b) s = 0.19
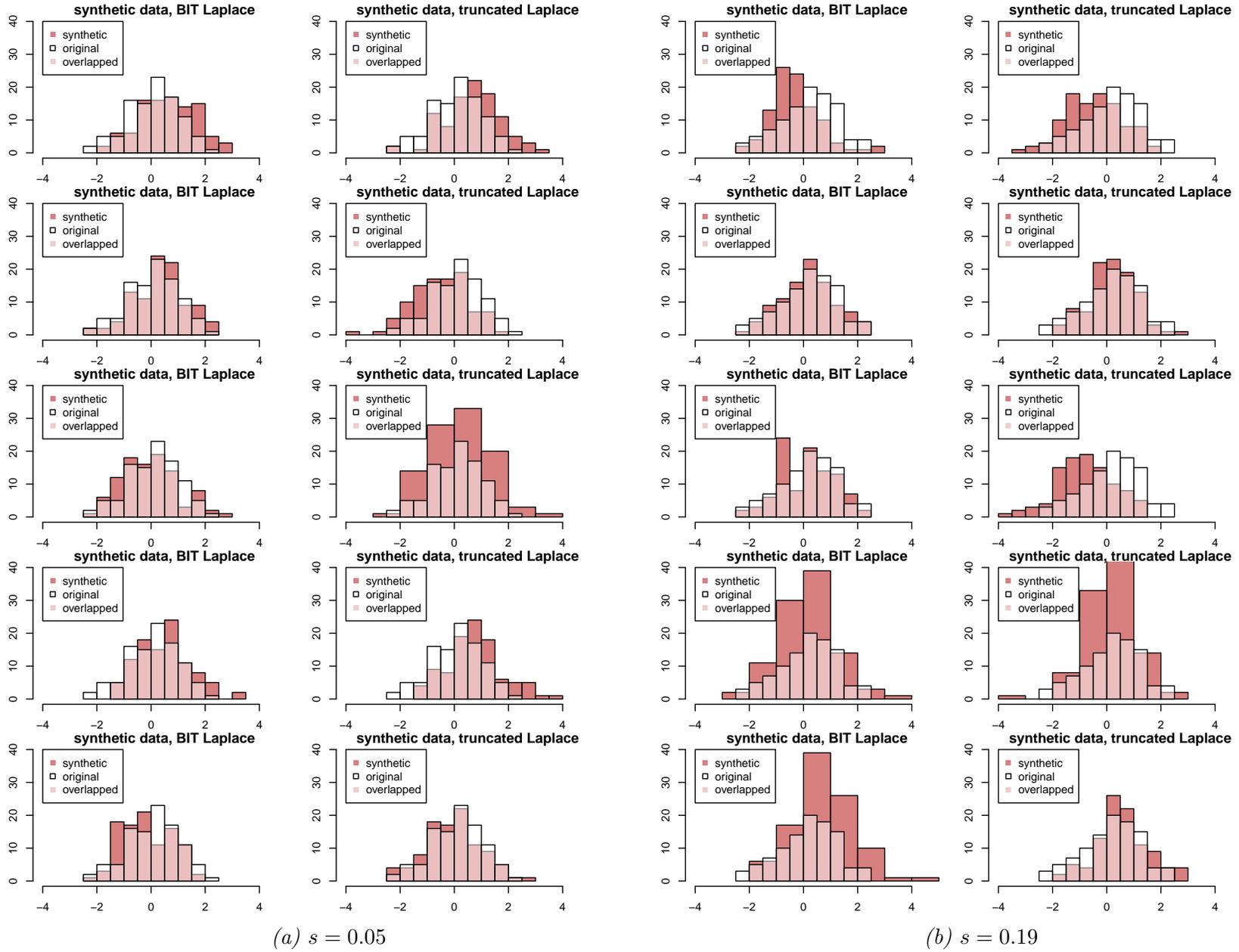
Figure S20: examples of synthetic data via modips in the simulation study ($n = 100, m = 5, [c_0, c_1] = [-4, 5]$)
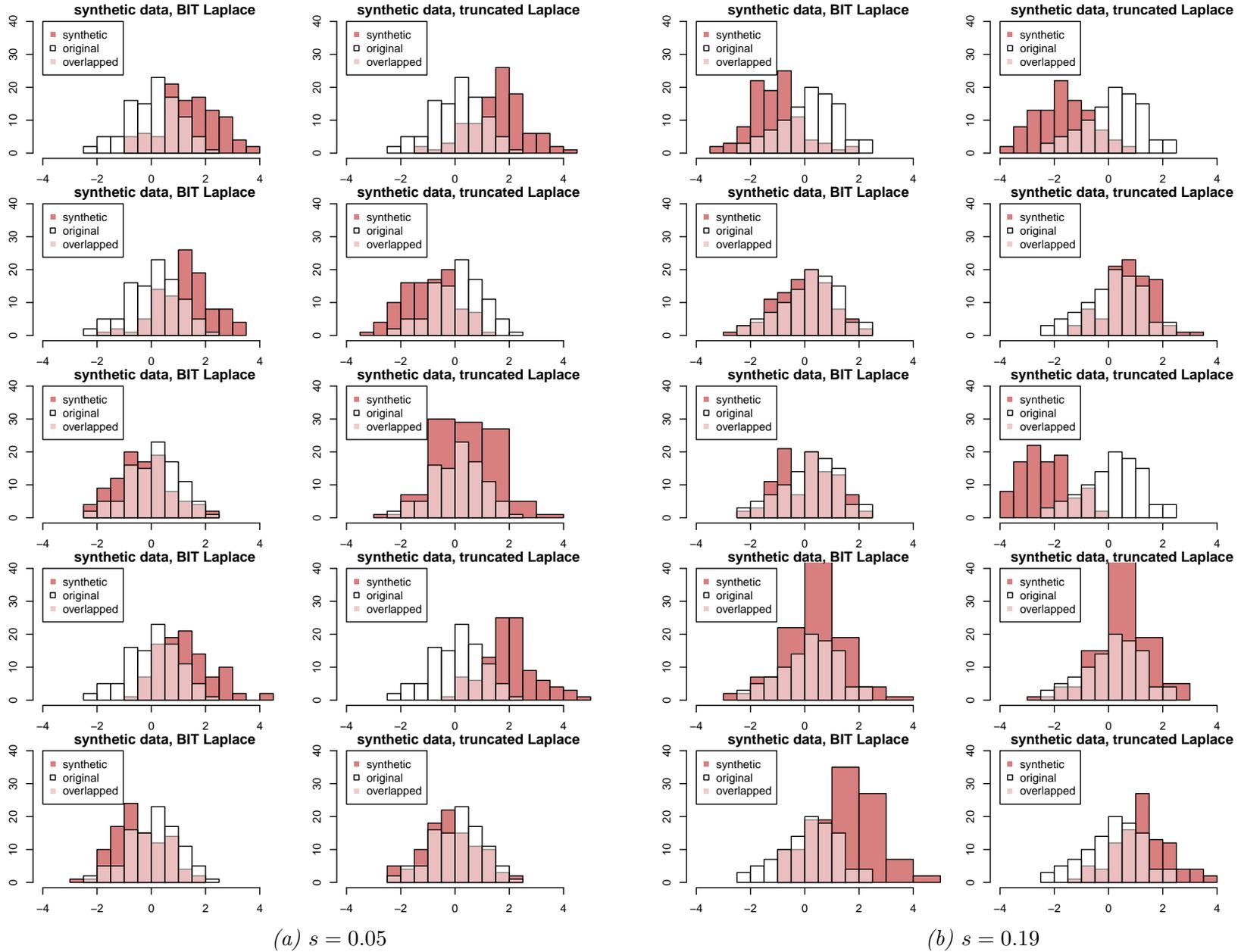
Figure S21: examples of synthetic data via modips in the simulation study $(n = 100, m = 15, [c_0, c_1] = [-4, 5])$
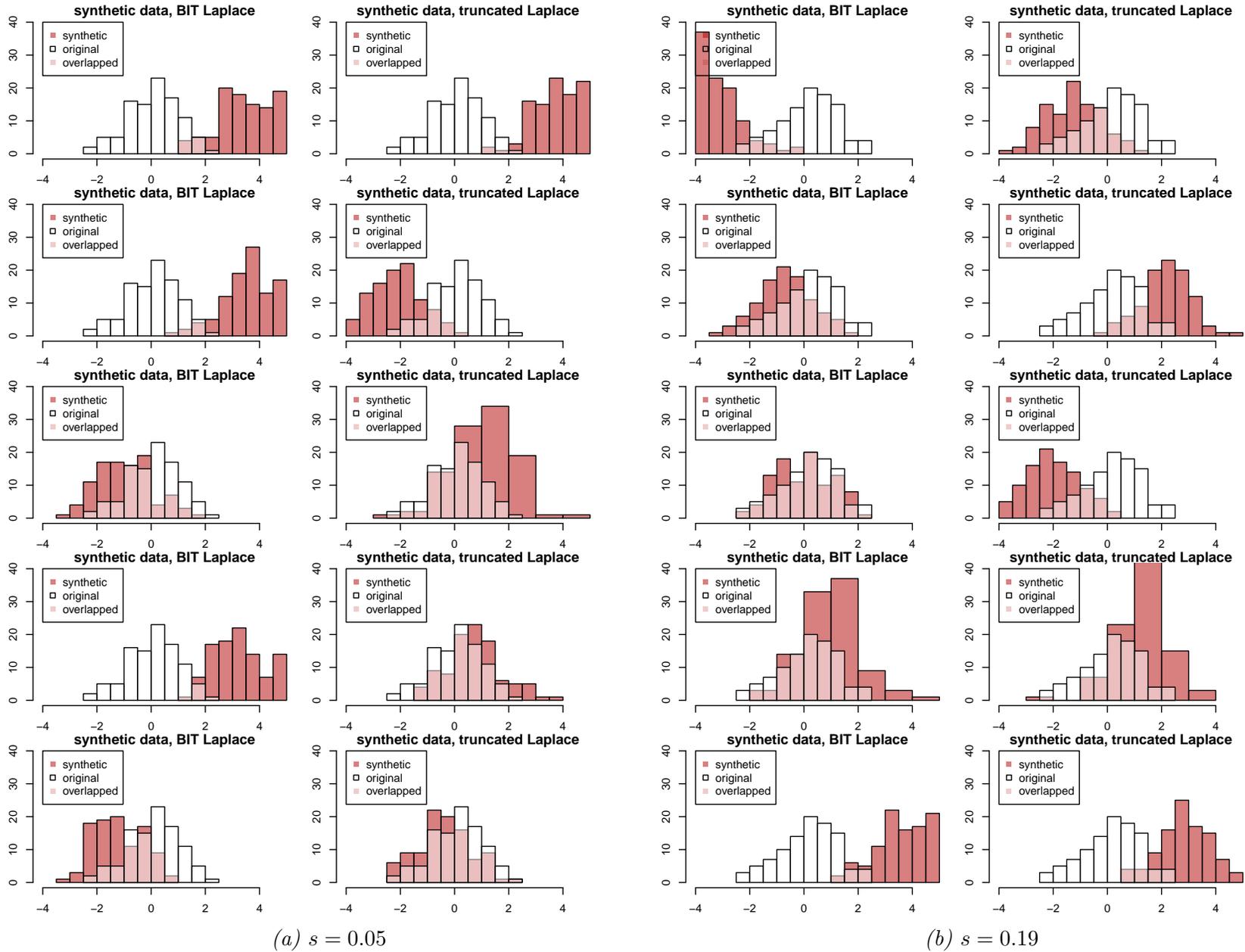
Figure S22: examples of synthetic data via modips in the simulation study ($n = 100, m = 40, [c_0, c_1] = [-4, 5]$)

23
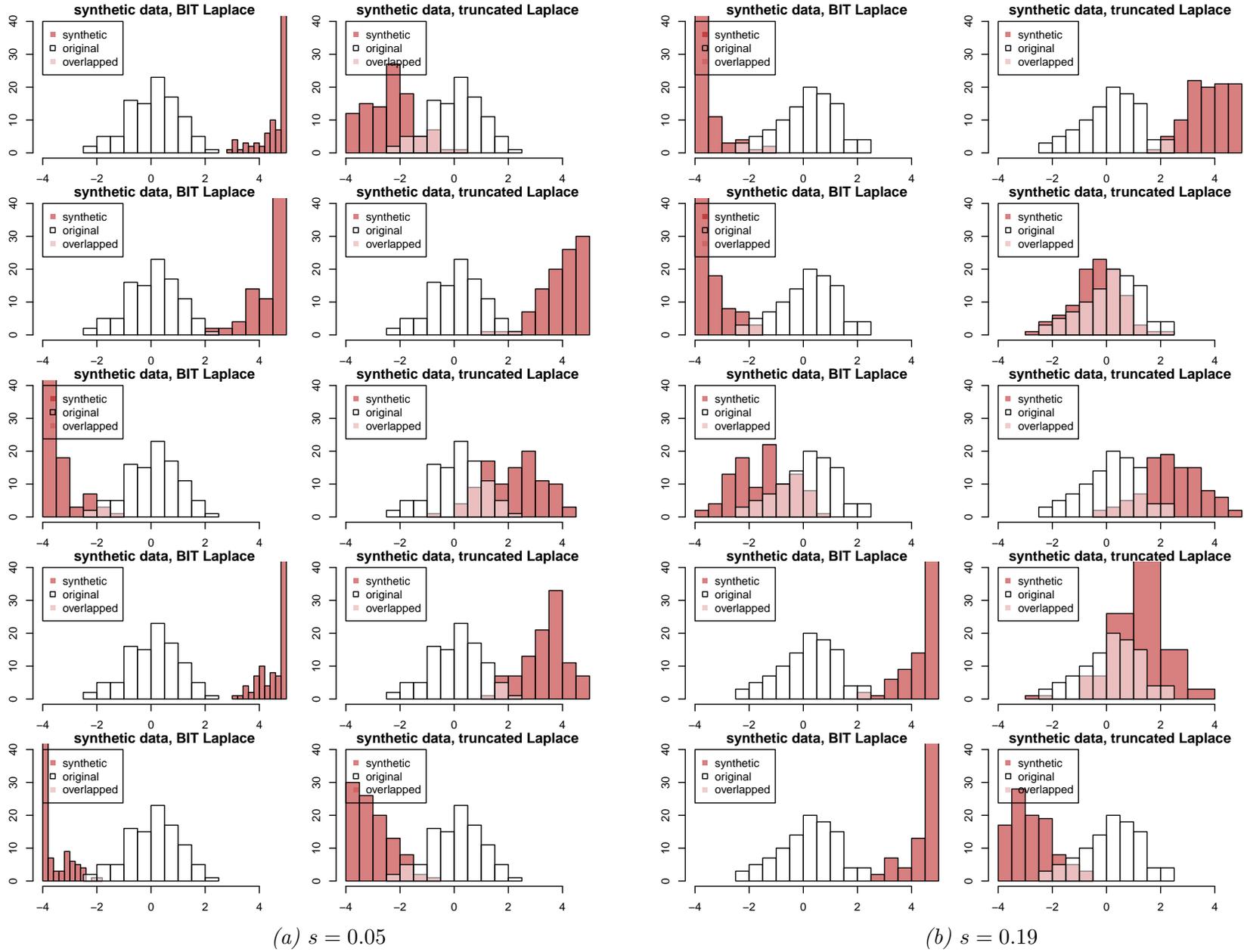
Figure S23: examples of synthetic data via modips in the simulation study ($n = 100, m = 500, [c_0, c_1] = [-4, 5]$)
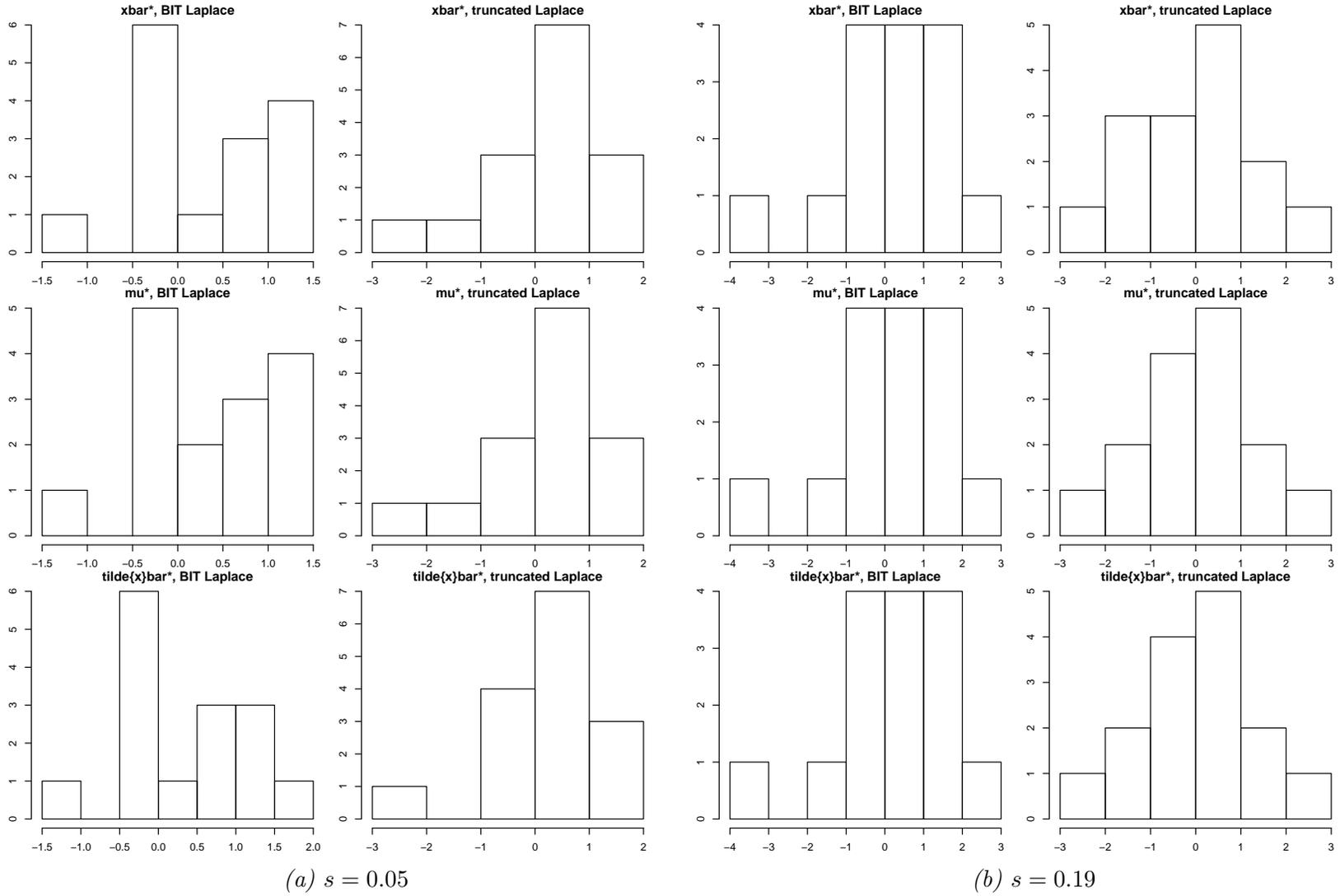
Figure S24: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 15, [c_0, c_1] = [-4, 4]$)

(a) $s = 0.05$

(b) $s = 0.19$
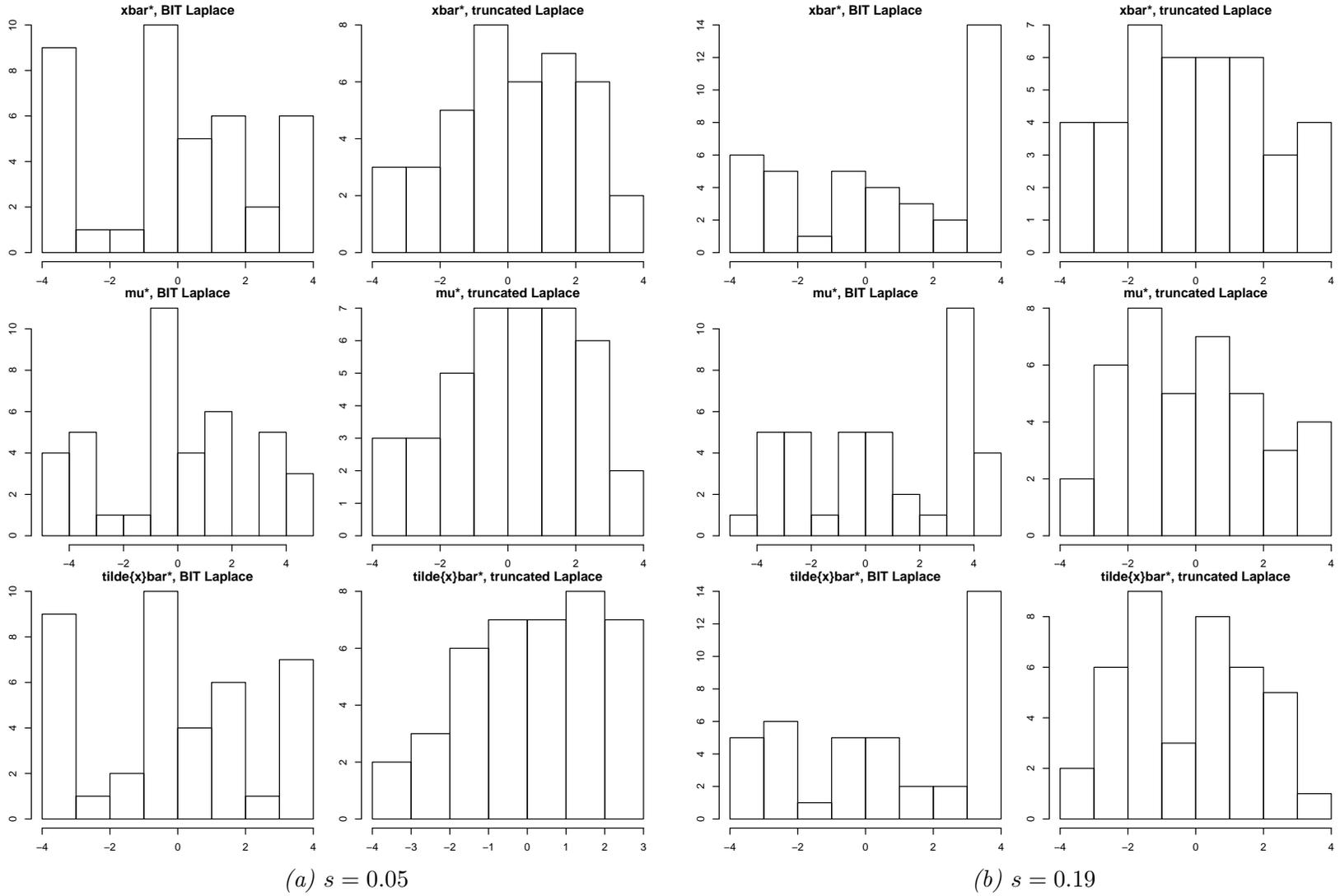
Figure S25: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 40, [c_0, c_1] = [-4, 4]$)

(a) $s = 0.05$

(b) $s = 0.19$
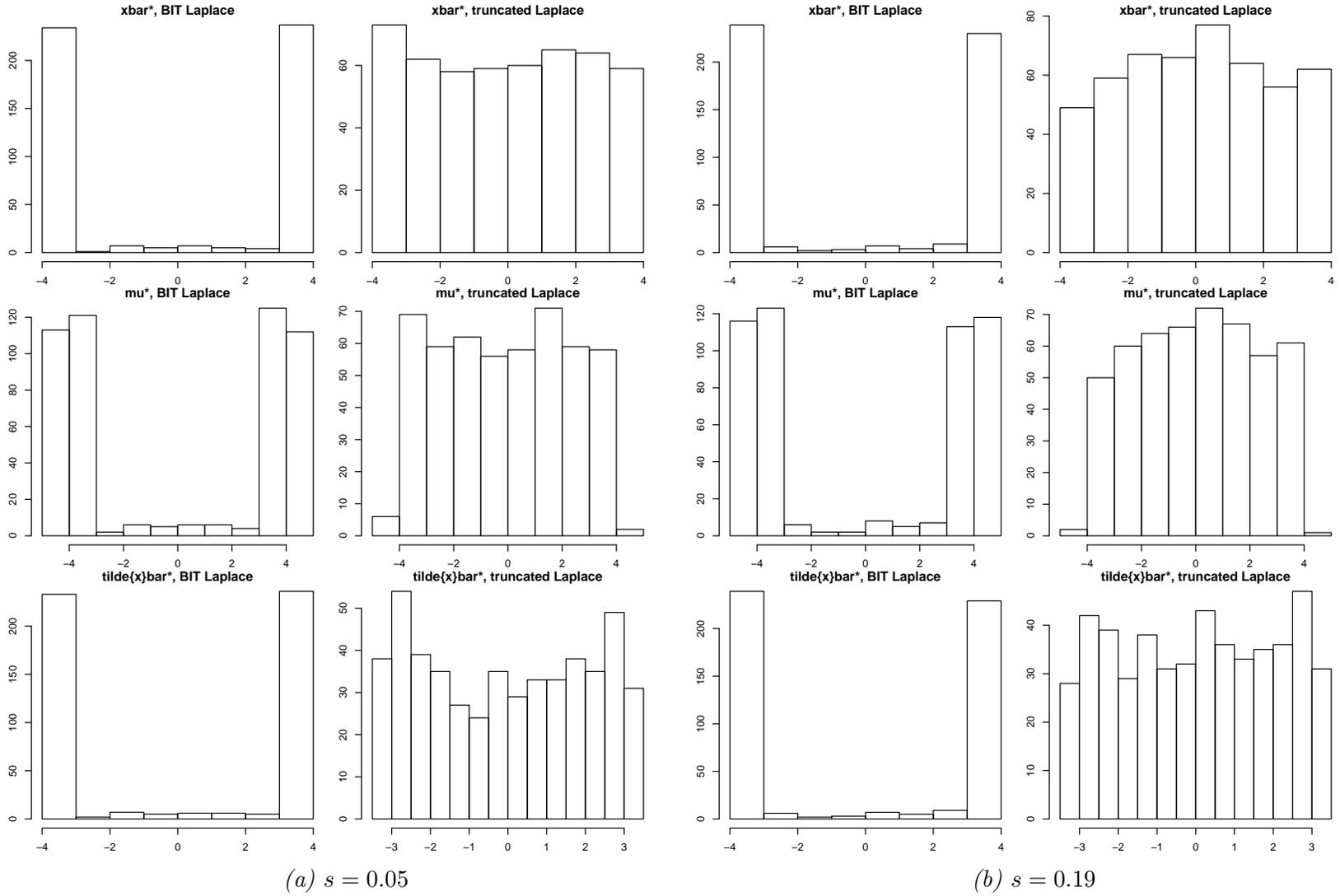
Figure S26: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\bar{\tilde{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 500, [c_0, c_1] = [-4, 4]$)
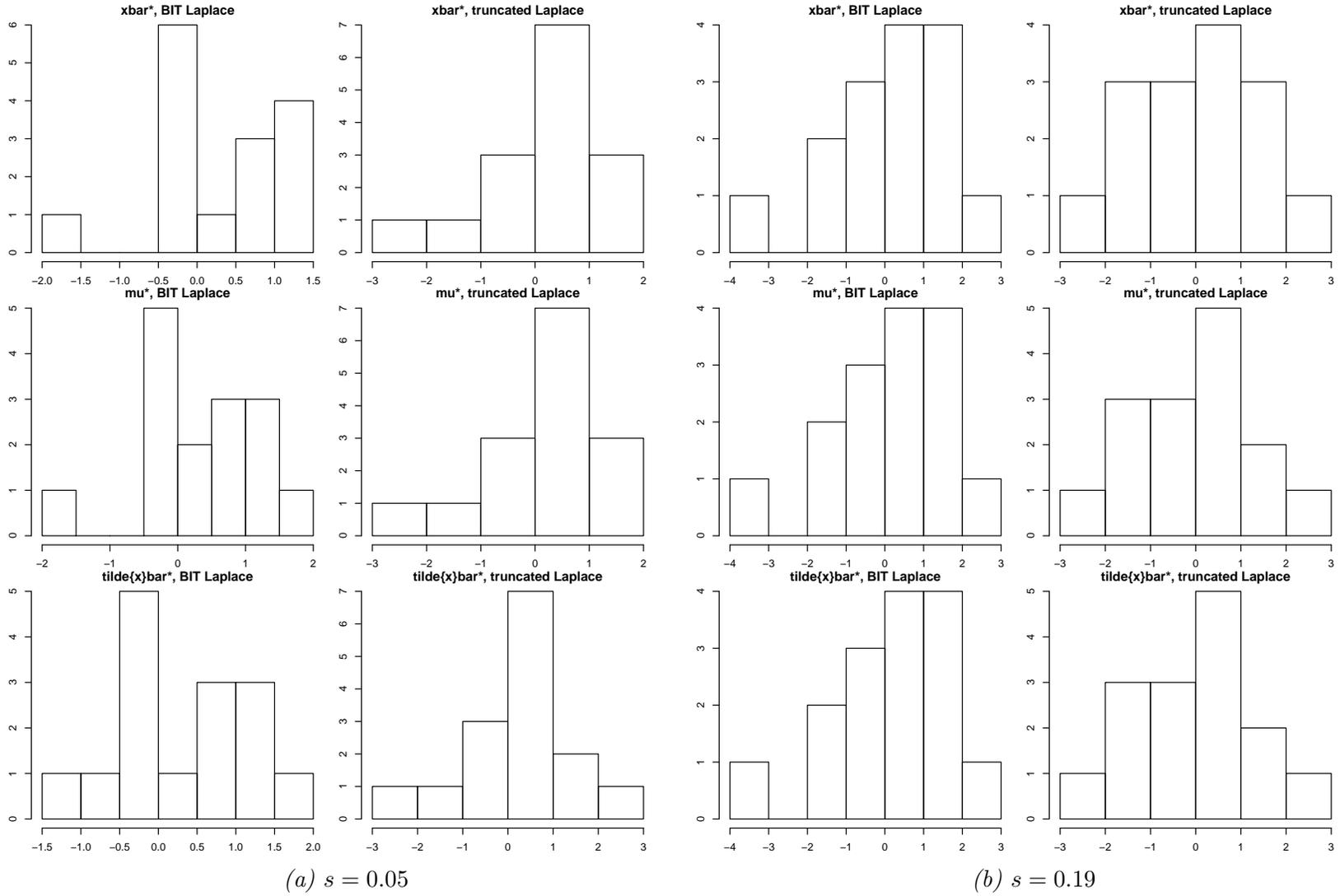
Figure S27: distribution of $\bar{x}^{*(k)}$, $\mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 15, [c_0, c_1] = [-4, 5]$)
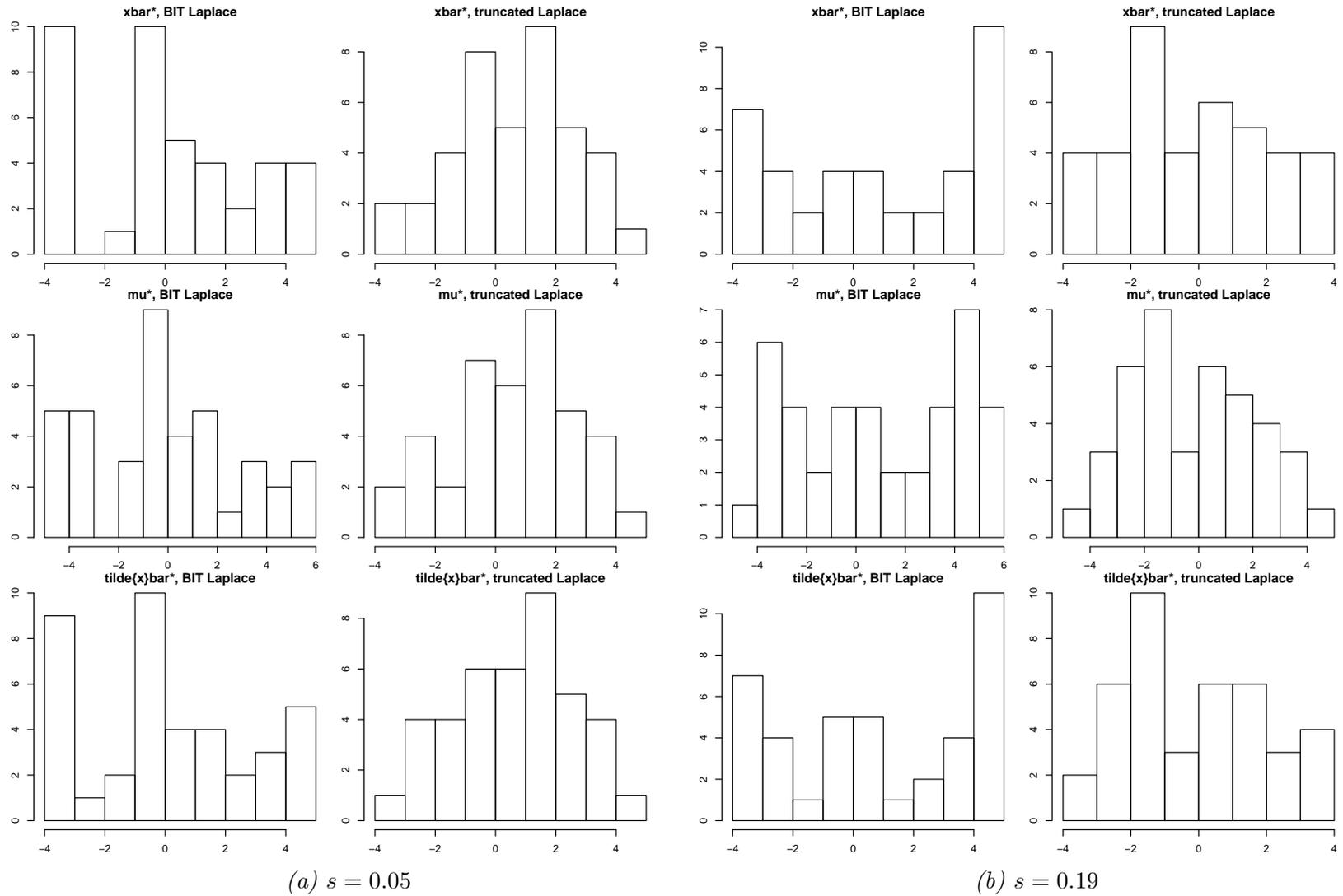
Figure S28: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 40, [c_0, c_1] = [-4, 5]$)

(a) $s = 0.05$

(b) $s = 0.19$
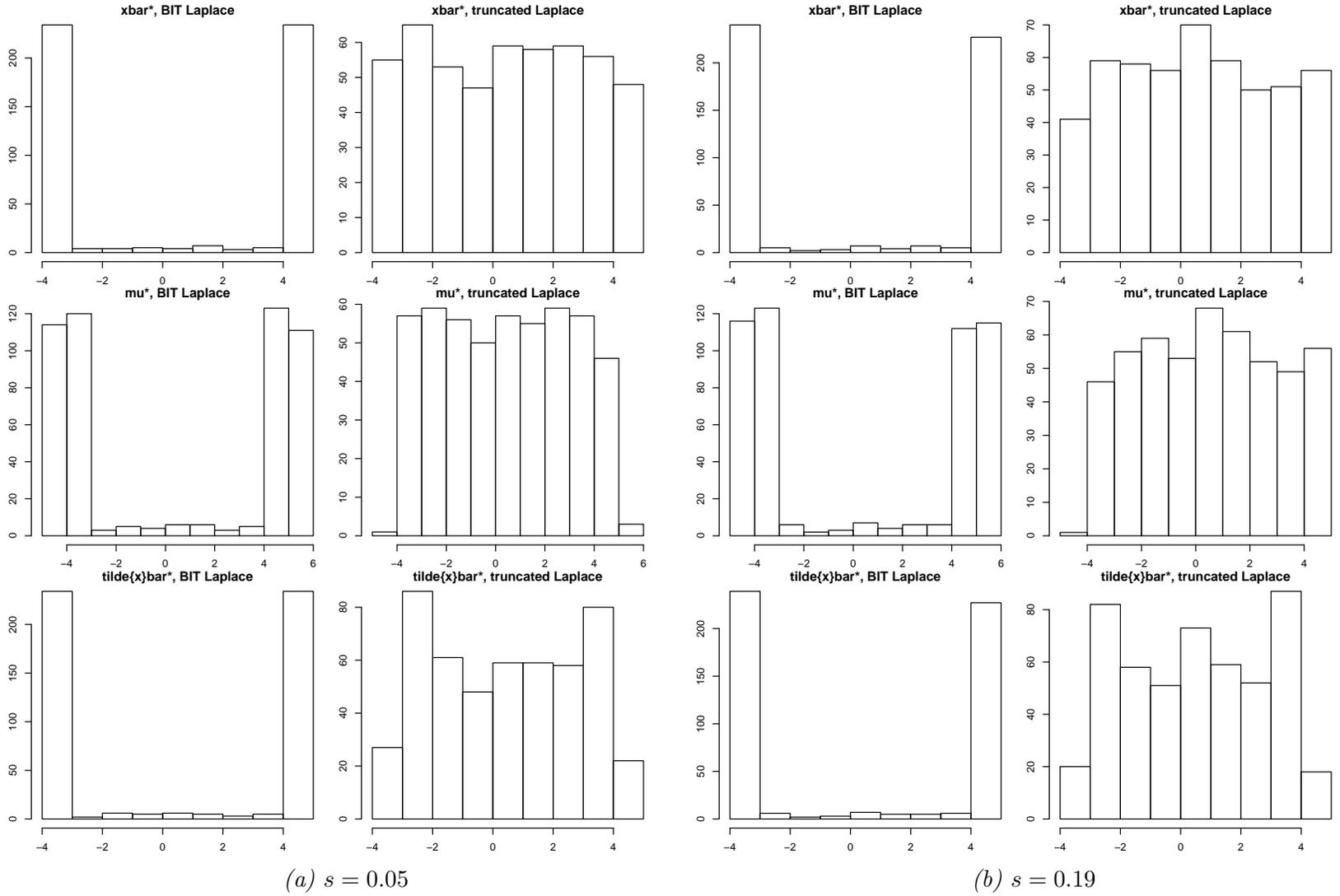
Figure S29: distribution of $\bar{x}^{*(k)}, \mu^{*(k)}$ and $\tilde{\bar{x}}^{*(k)}$ across synthetic sets in the simulation study ($n = 100, m = 500, [c_0, c_1] = [-4, 5]$)