Experiment: Underdetermined linear models

In this experiment we simulate underdetermined linear models where $p > n$. We generate n data points (X_i, y_i) where $y_i = X_i^T \beta^* + w_i$ with w_i an additive noise term. We then compute the minimum norm estimator

$$\widehat{\beta} = X^T (XX^T)^{-1} y \tag{1.1}$$

The estimated model is corrupted to

$$\eta = \widehat{\beta} + z \tag{1.2}$$

where $z_j \sim (1-\varepsilon)\delta_0 + \varepsilon Q$. The corrupted estimator is then repaired by performing median regression:

$$\widetilde{u} = \arg\min \|\eta - X^T u\|_1 \tag{1.3}$$

$$\widetilde{\beta} = X^T \widetilde{u} \tag{1.4}$$

In the experiments shown below the design is sampled as $X_{ij} \sim N(0,1)$ and we take $\beta_j^* \sim N(0,1)$ and $Q = N(1,1)$. We take $n = 50$ and $n = 100$ and vary p according to $p/n = 200/j^2$ with j ranging from 1 to 9. The plots show the empirical probability of exact repair $\widetilde{\beta} = \widehat{\beta}$ as a function of ε. Each point on the curves is the average repair success over 200 trials. The roughly equal spacing of the curves agrees with the theory, which indicates that $\sqrt{n/p}/(1-\varepsilon)$ should be sufficiently small for successful repair.

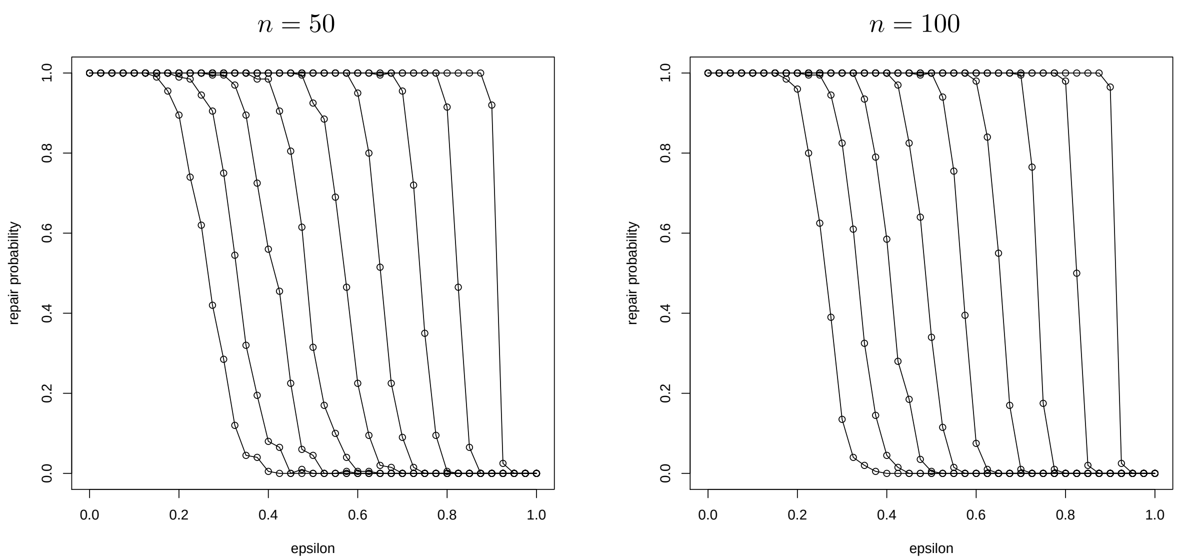

FIG 1. *Model repair for underdetermined linear models $y = X^T\beta + w$ with $p > n$. The plots show the probability of successful model repair for $n = 50$ (left) and $n = 100$ (right) with the model dimension p varying as $p/n = 200/j^2$, for $j = 1, \ldots, 9$. Each point is an average over 200 trials.*