Simulation10

Mengqi Liu

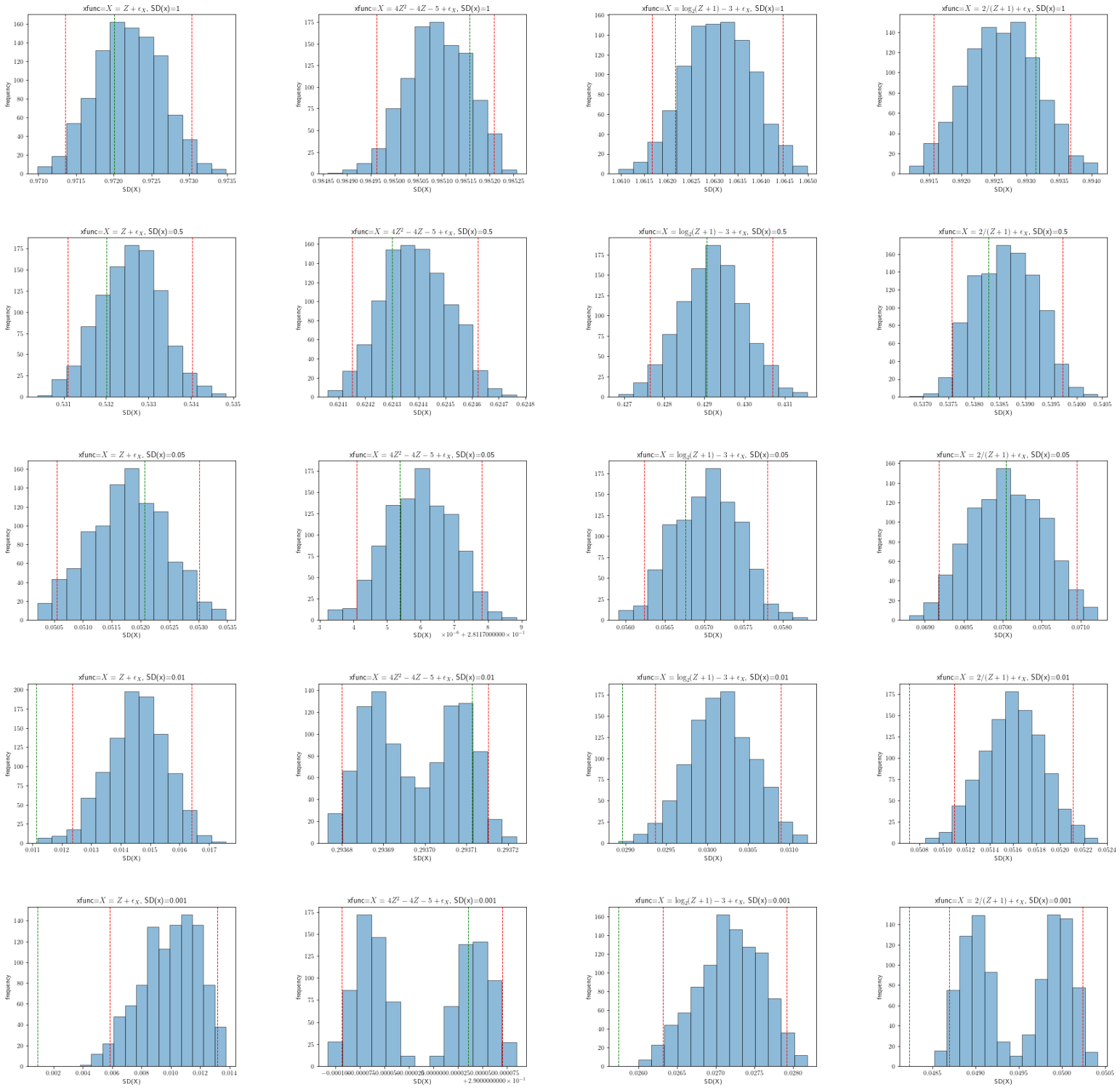
Sep 12, 2023

- Methods: ($\tilde{Z}$ is the discretized Z, and the data belonging to the same group share the same $\tilde{Z}$.)
  - "double_Z/cor_Z": permute X within each bin. At each time, regress Y on $1, Z$ and regress X on $1, Z$ separately, and take the *absolute correlation* between residuals from two linear regressions as the test statistic.
  - "XY_meanZ": permute X within each bin. At each time, regress Y on $1, \tilde{Z}$ and regress X on $1, \tilde{Z}$ separately, and take the *absolute product* between residuals from two linear regressions as the test statistic.
  - "cor_noZ": use $cor(X, Y)$ as test statistic with local permutation in X with respect to Z.
  - "XY_Z": regress Y on $1, Z$ and regress X on $1, Z$ separately. Permute residuals from regression on for X and take the *absolute product* between residuals as the test statistic.
  - "XY_noZ": use $X^\top Y$ as test statistic with local permutation in X with respect to Z.

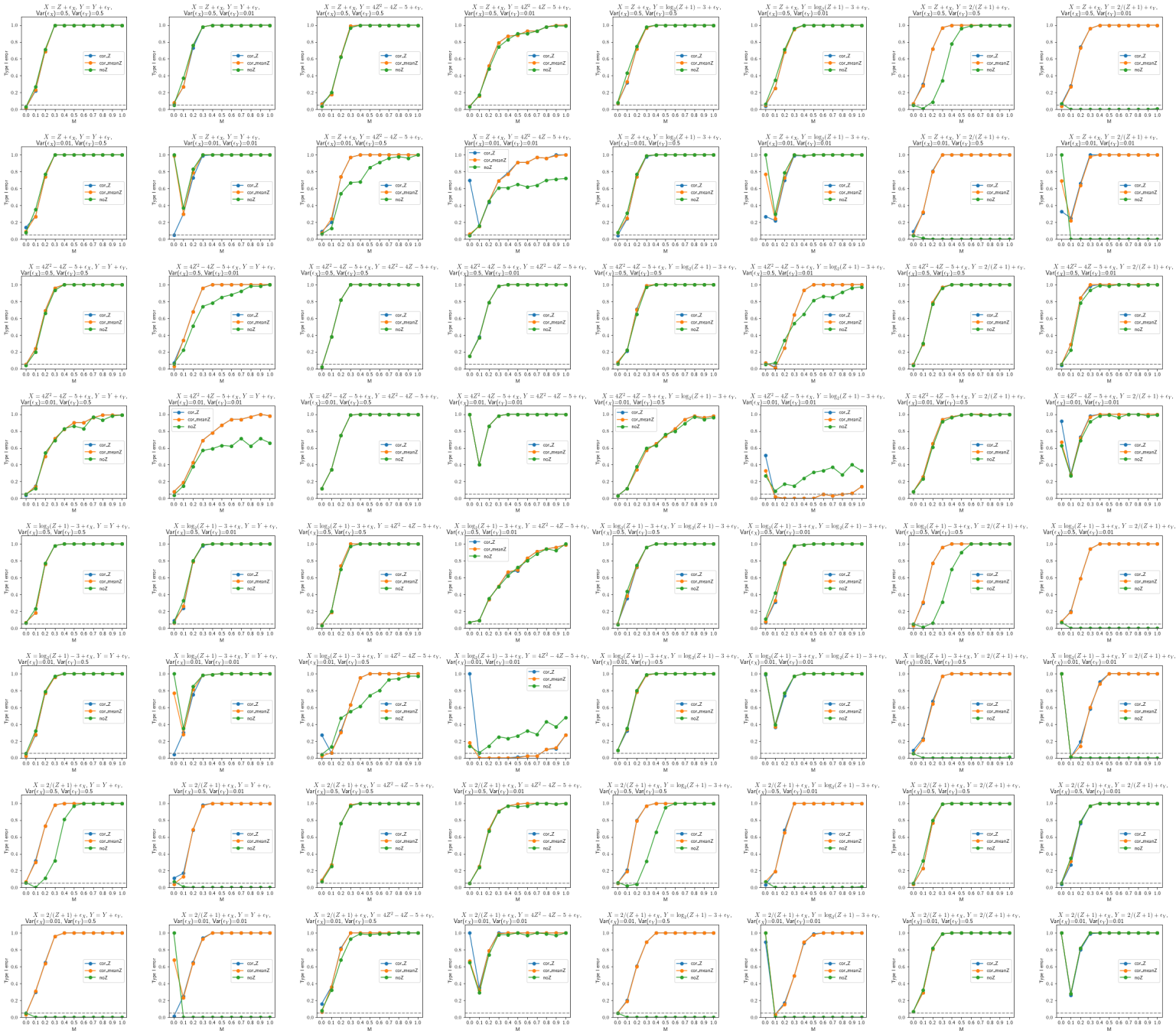
Summary

- cor_meanZ = XY_meanZ (same denominator)
- Only using nominator (XY_Z) will improve the ability of cor_Z to control for type-1 error
- cor_noZ != XY_noZ due to non-zero $E(X)$ & $E(Y)$
- noZ has lower power under situtation that at least one of X and Y is smooth. If both are non-smooth, sometimes noZ can have better type-1 error control and power at the same time. noZ & XY_noZ sometimes have catastrophic problem in zero power.
- why XY_Z doesn't gain more power than XY_meanZ?
  - too small sample? (N=1000)
  - too large M? (M$\leq$25 cannot guarantee type-1 error control)
  - good property of gaussian noise? (skewed normal distribution)
  - linear in Z? (both non-linear)

Deviation of SD($P_{Z^\perp}X$) to SD($P_{Z^\perp}X_\sigma$)

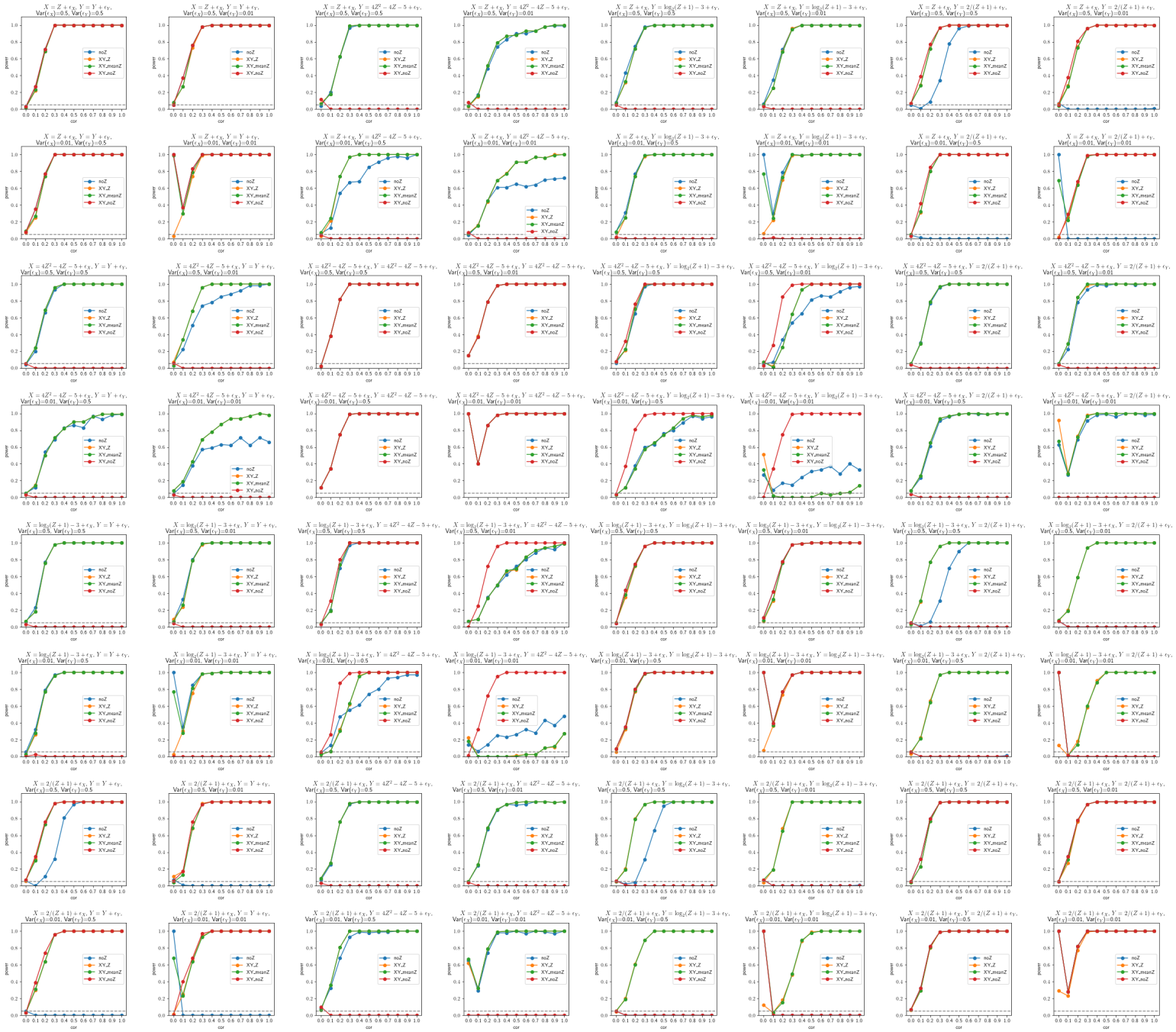


Original results (transformed to cor-power plot)



Comparing cor_Z and XY_Z

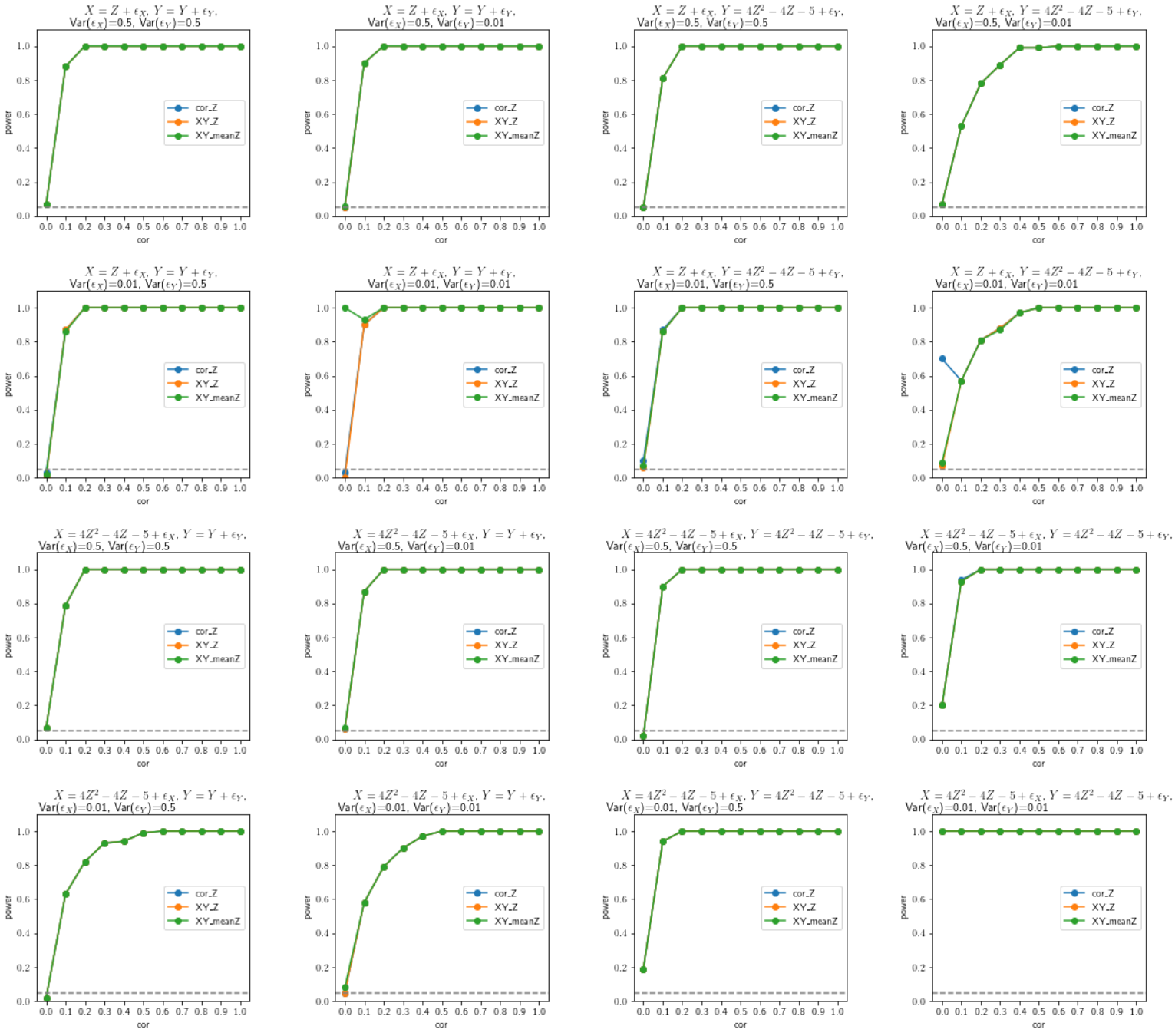
Comparing type-1 error and power between using information of Z or not



Comparing XY_Z and XY_meanZ

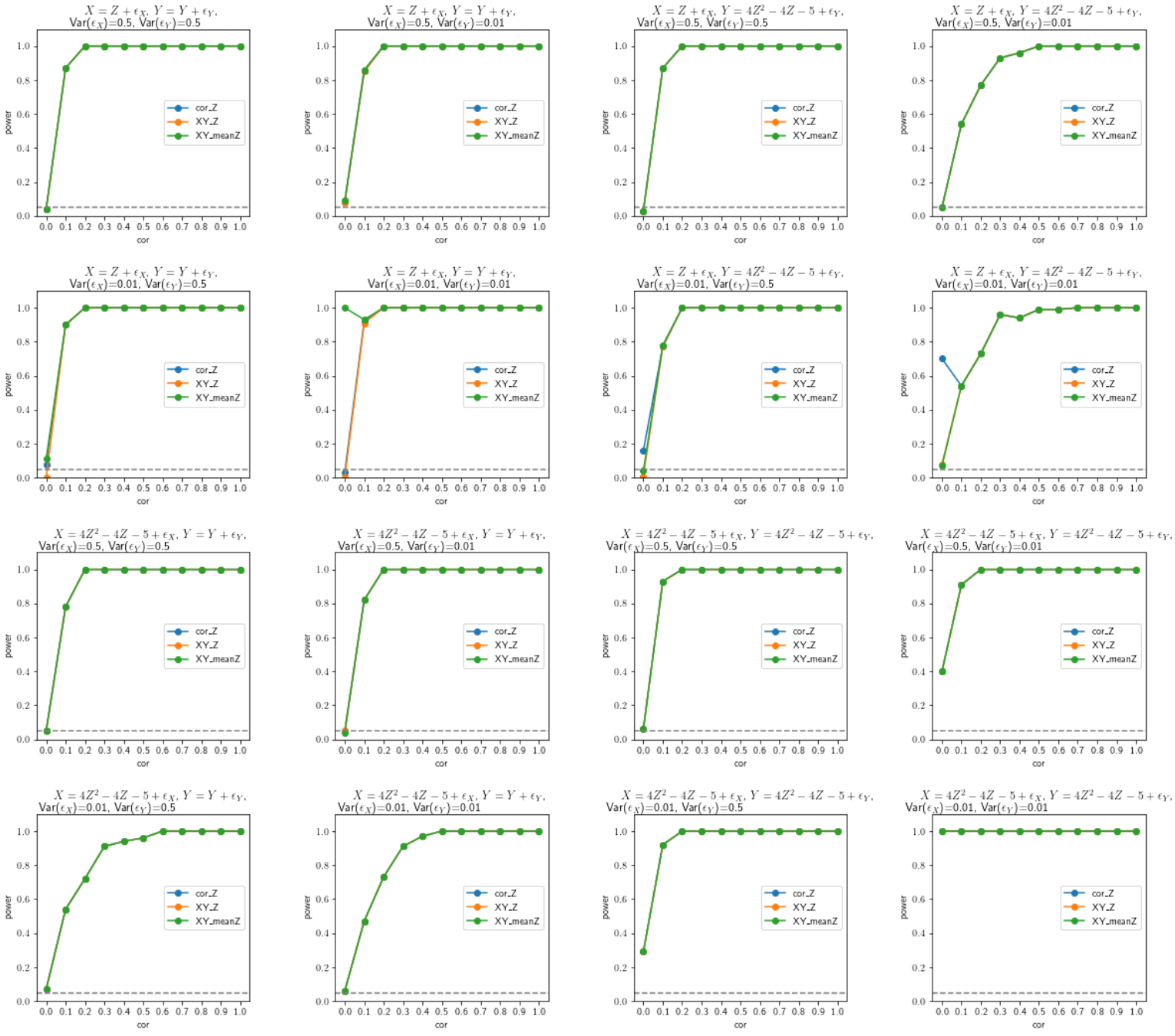
- large N=1000

Normal Distribution



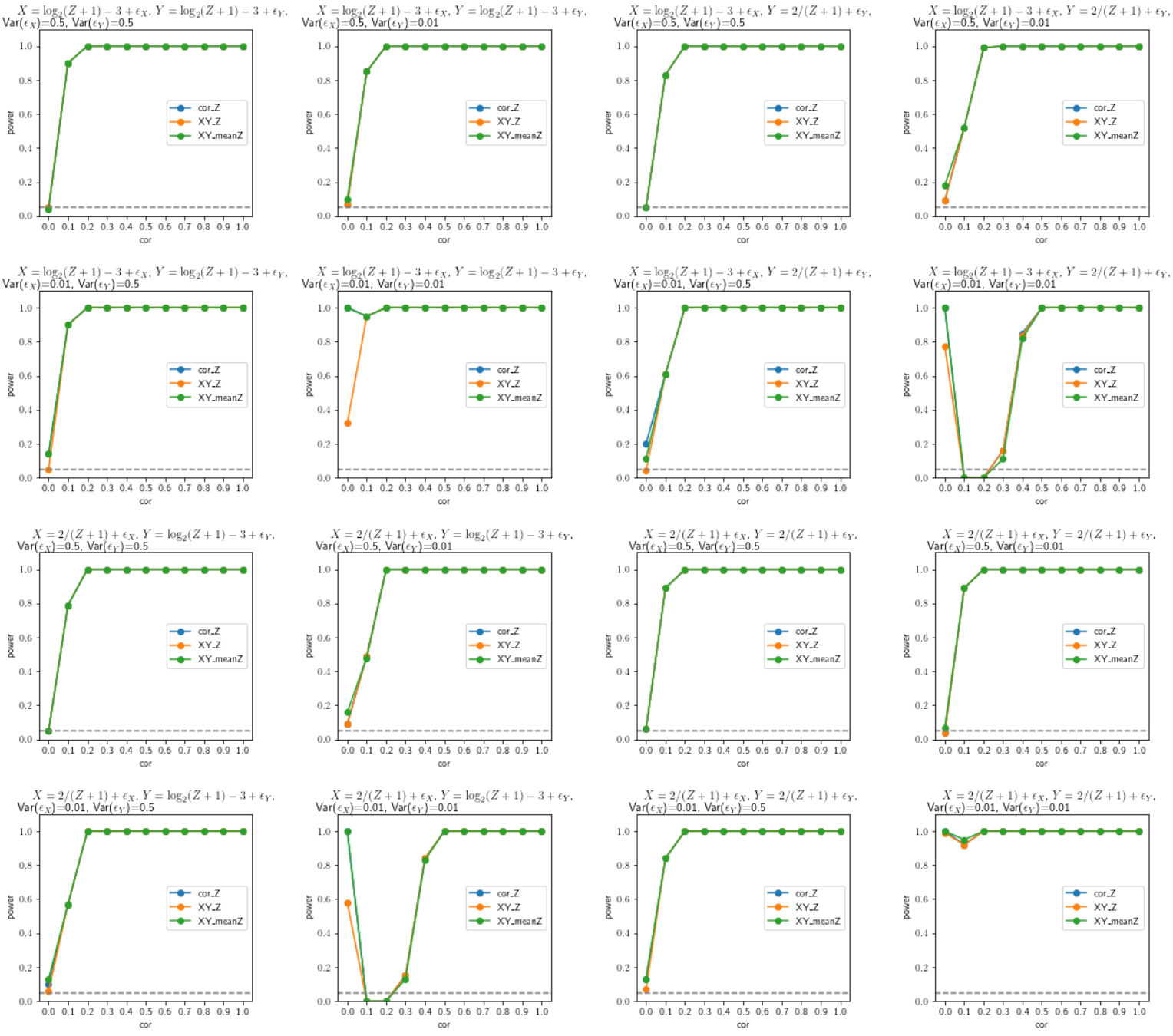
- large N=1000 + distribution="skewed_normal"

Normal Distribution



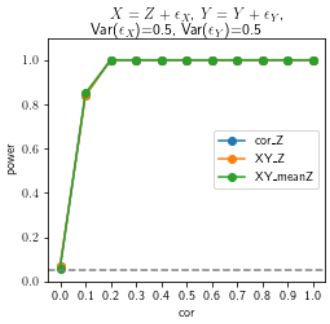
- large N=1000 + distribution="skewed_normal" + nonlinear function in Z

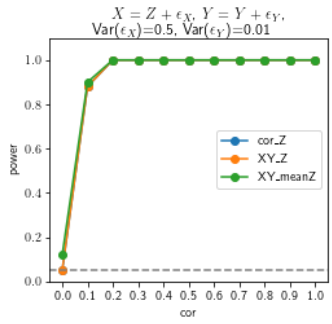
Skewed Normal Distribution

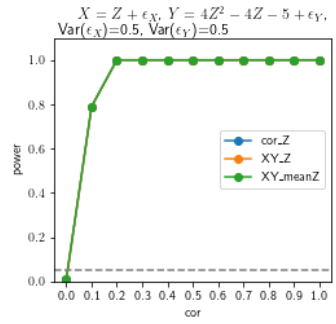


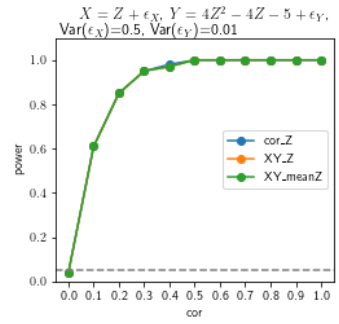
- M=25

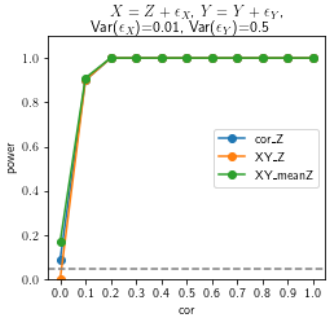
Normal Distribution

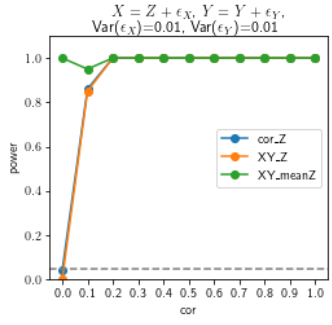


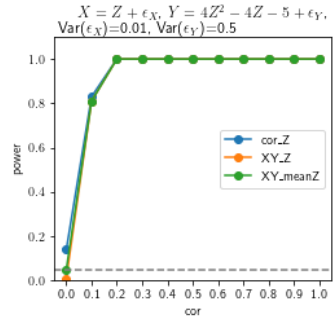


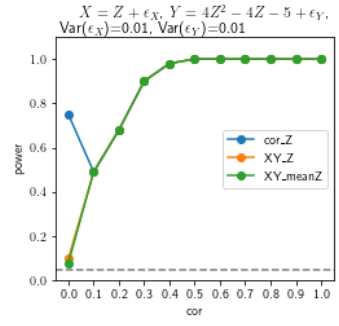


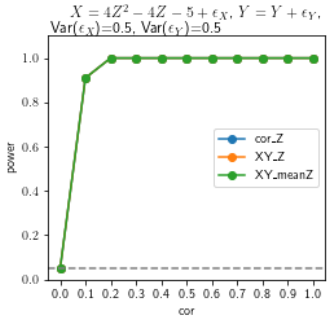


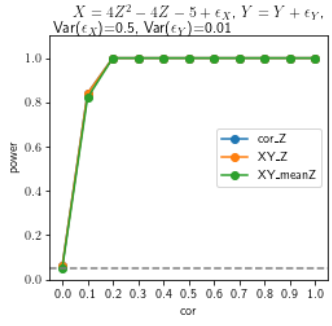


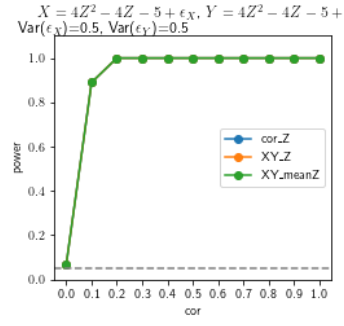


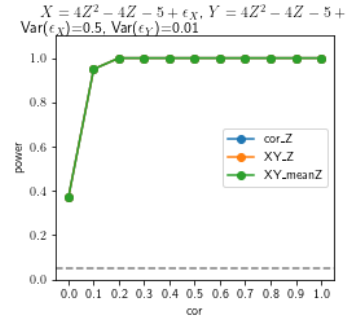


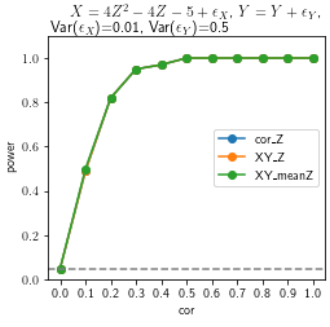


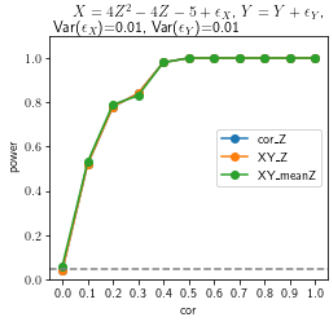


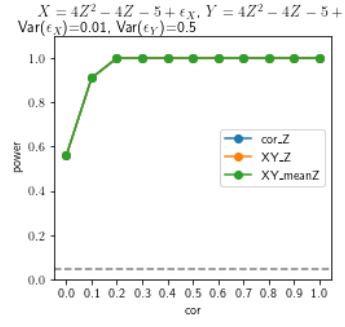


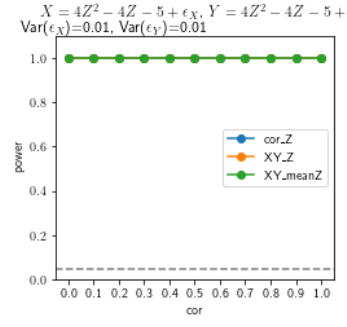




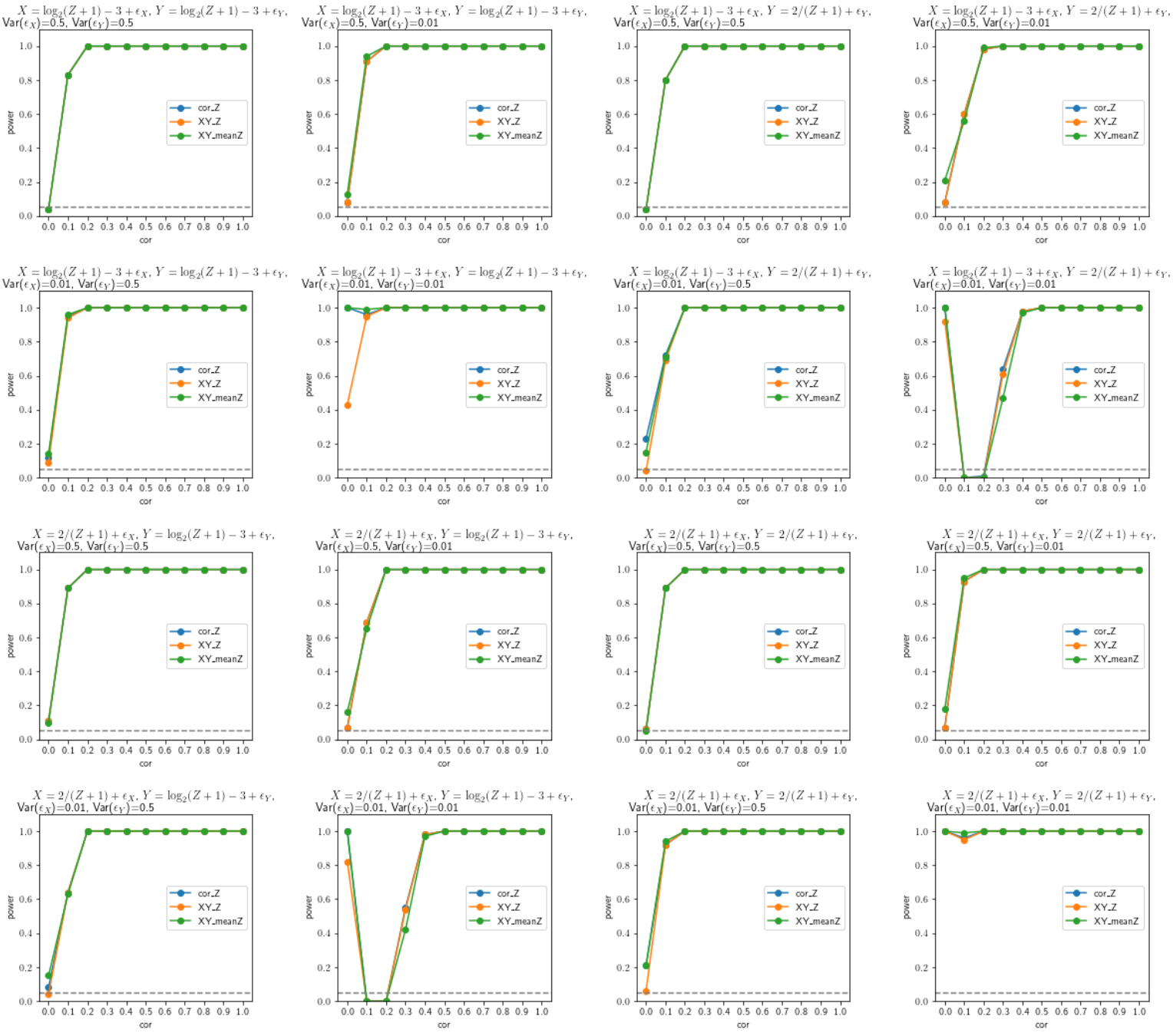






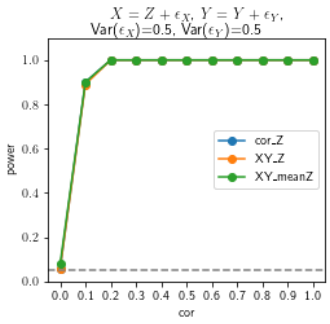


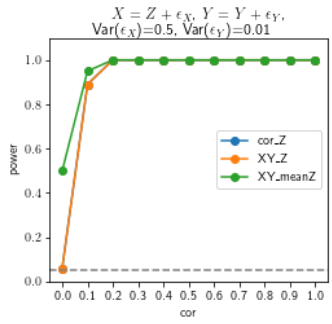
Normal Distribution

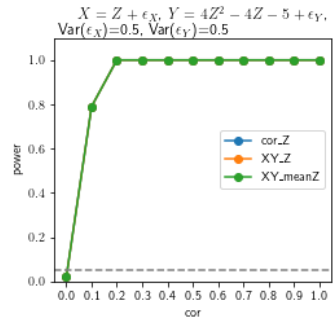


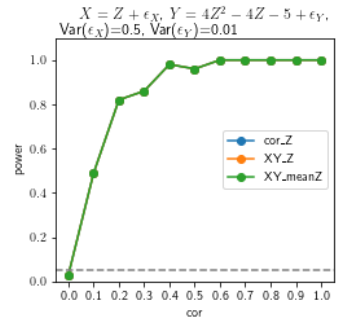
- M = 10

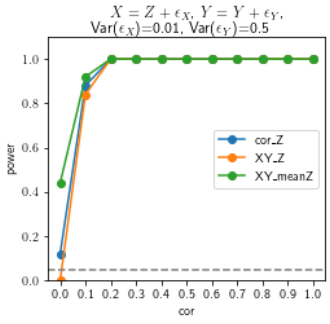
Normal Distribution

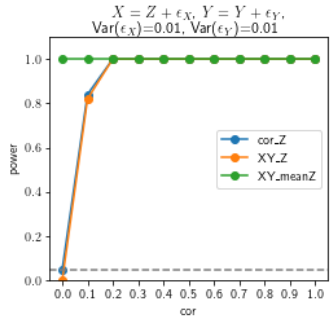


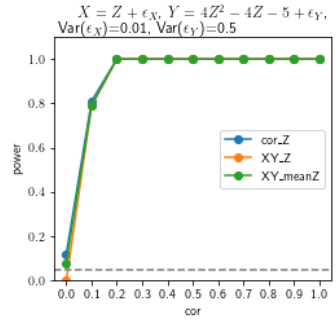


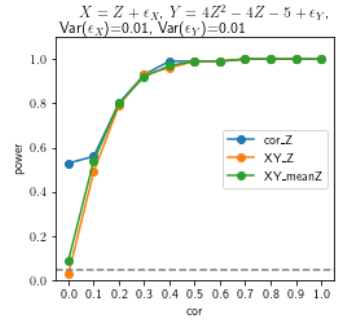


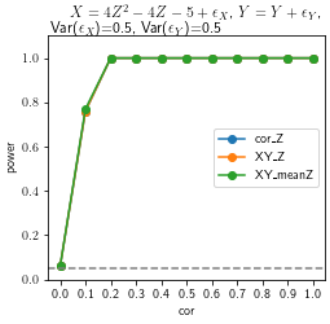


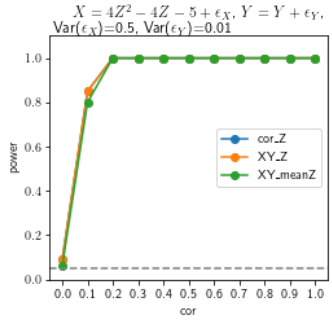


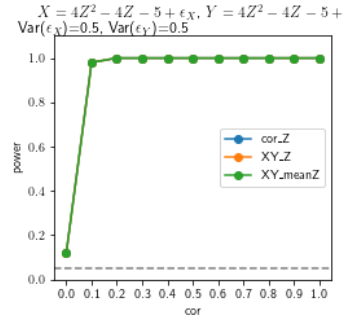


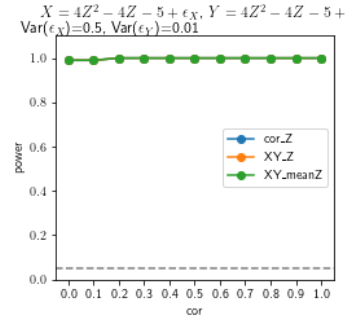


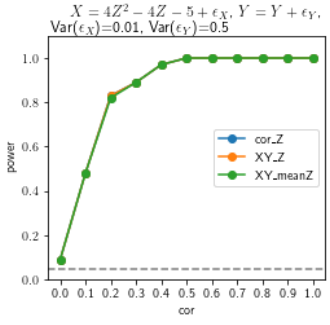


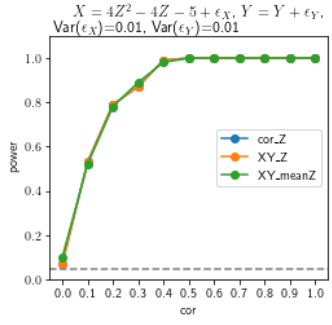


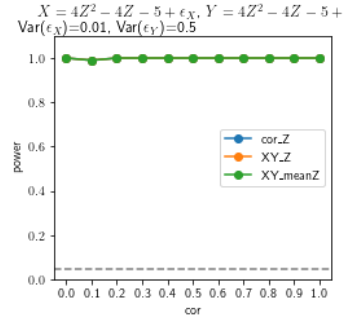


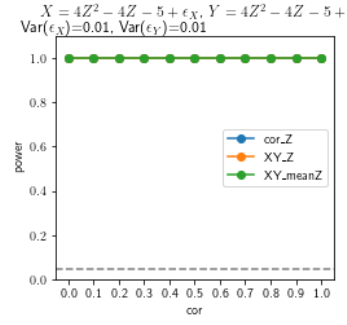




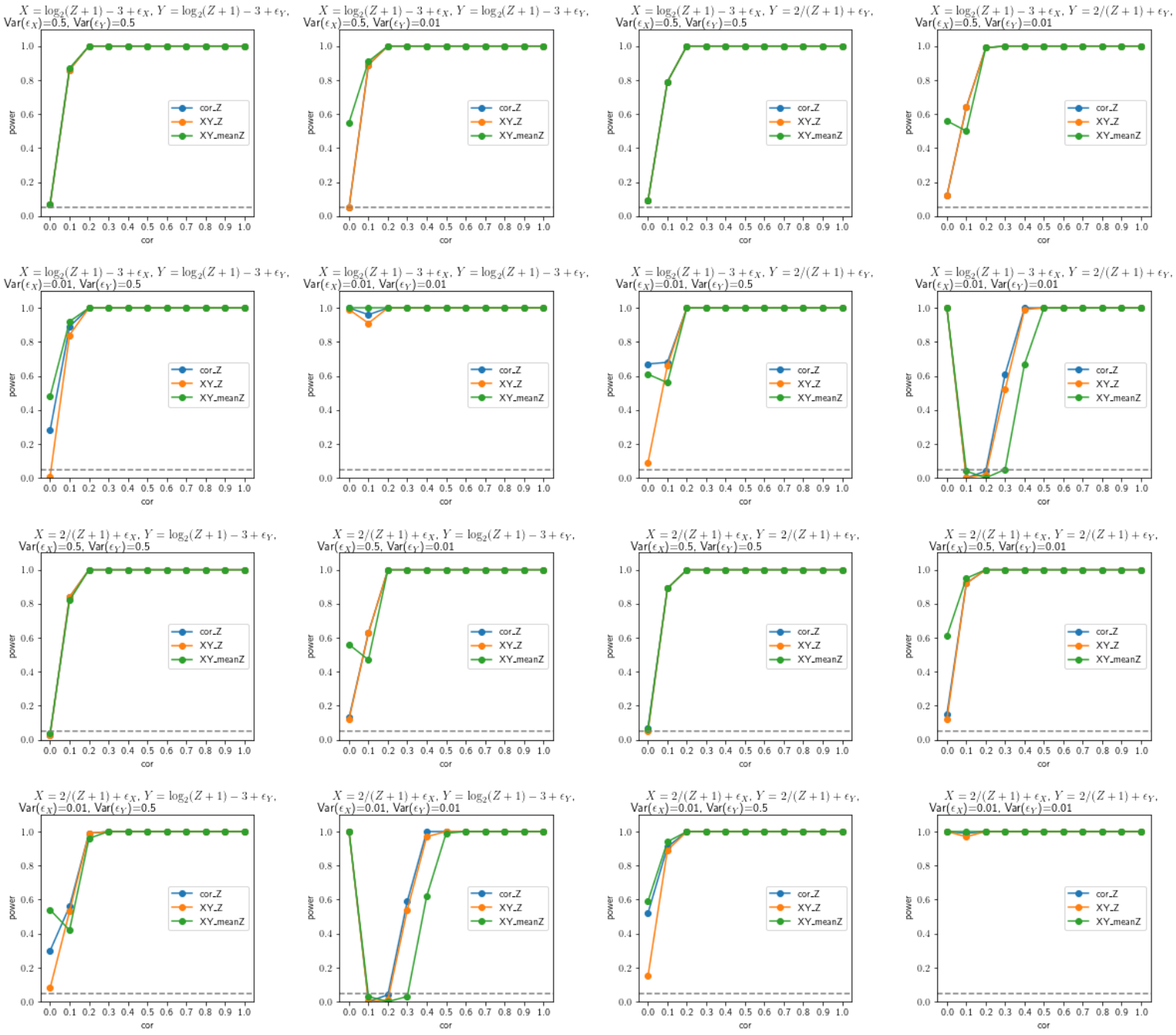








Normal Distribution



More statistics

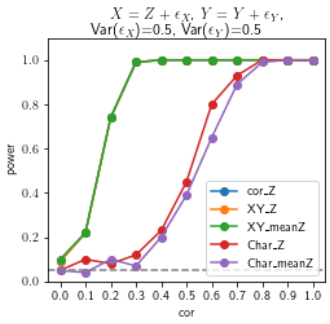
- new conditional correlation coefficient by Mona Azadkia & Sourav Chatterjee (2021)

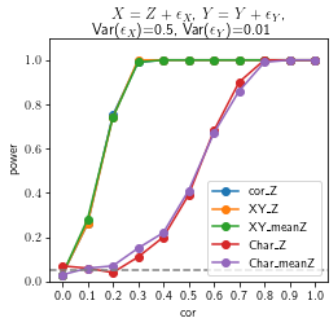
$$T = T(X, Y|Z) = \frac{\int \mathbb{E}(\mathrm{Var}(\mathbb{P}(X \geq t|Y, Z)|Z))d\mu(t)}{\int \mathbb{E}(\mathrm{Var}(1_{\{X \geq t\}}|Z))d\mu(t)}$$

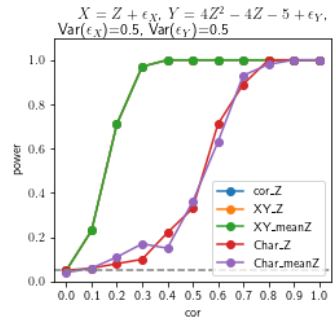
$$T_n = T_n(X, Y|Z) = \frac{\sum_{i=1}^{n}(\min\{R_i, R_{M(i)}\} - \min\{R_i, R_{N(i)}\})}{\sum_{i=1}^{n}(R_i - \min\{R_i, R_{N(i)}\})}$$

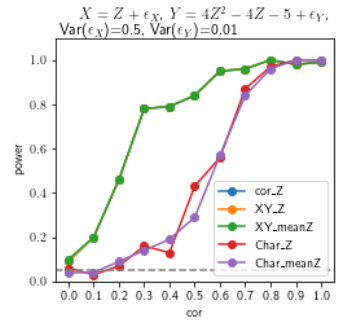
where $N(i)$ is the index of nearest to Z_i, $M(i)$ is the index of nearest to (Z_i, Y_i) and R_i is the rank of X_i.

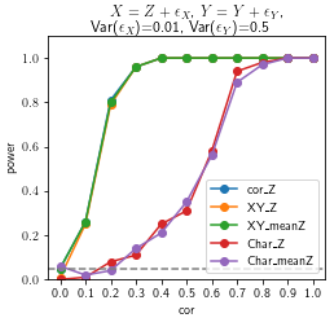
Normal Distribution

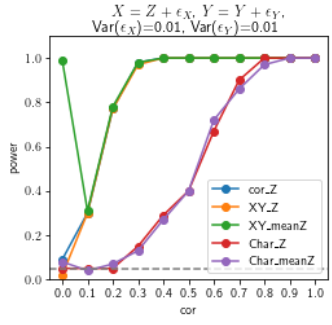


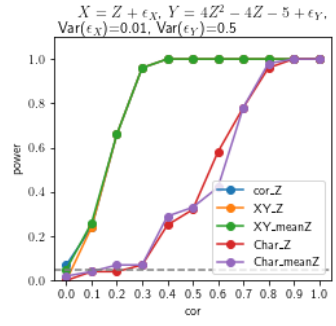


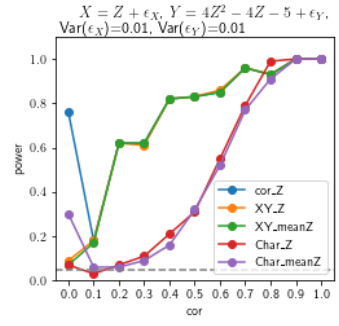


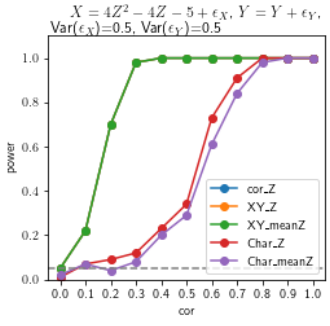


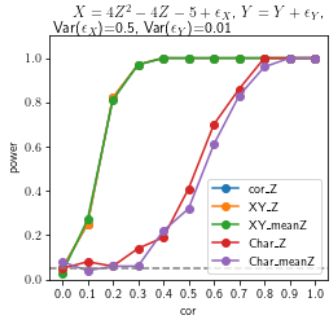


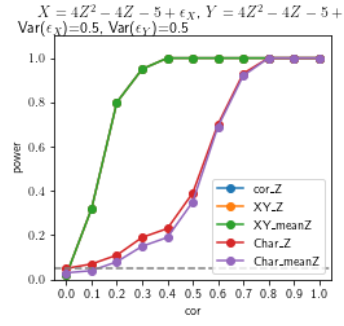


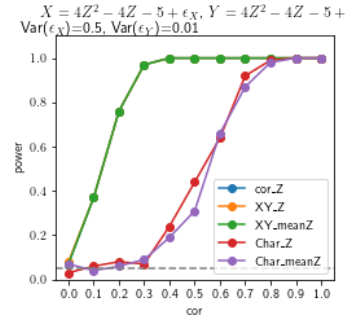


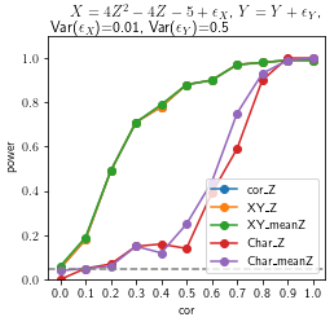


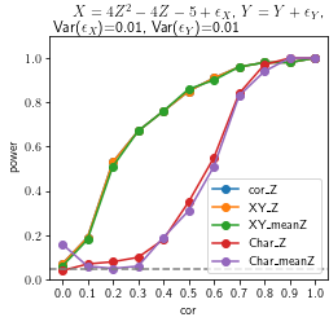


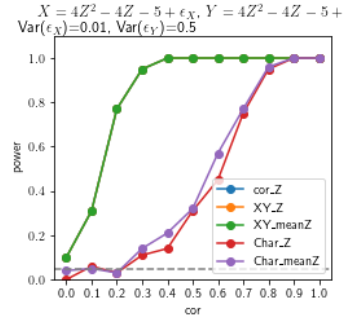


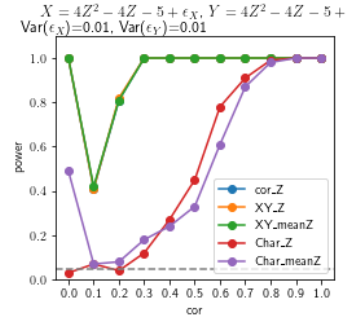












Normal Distribution