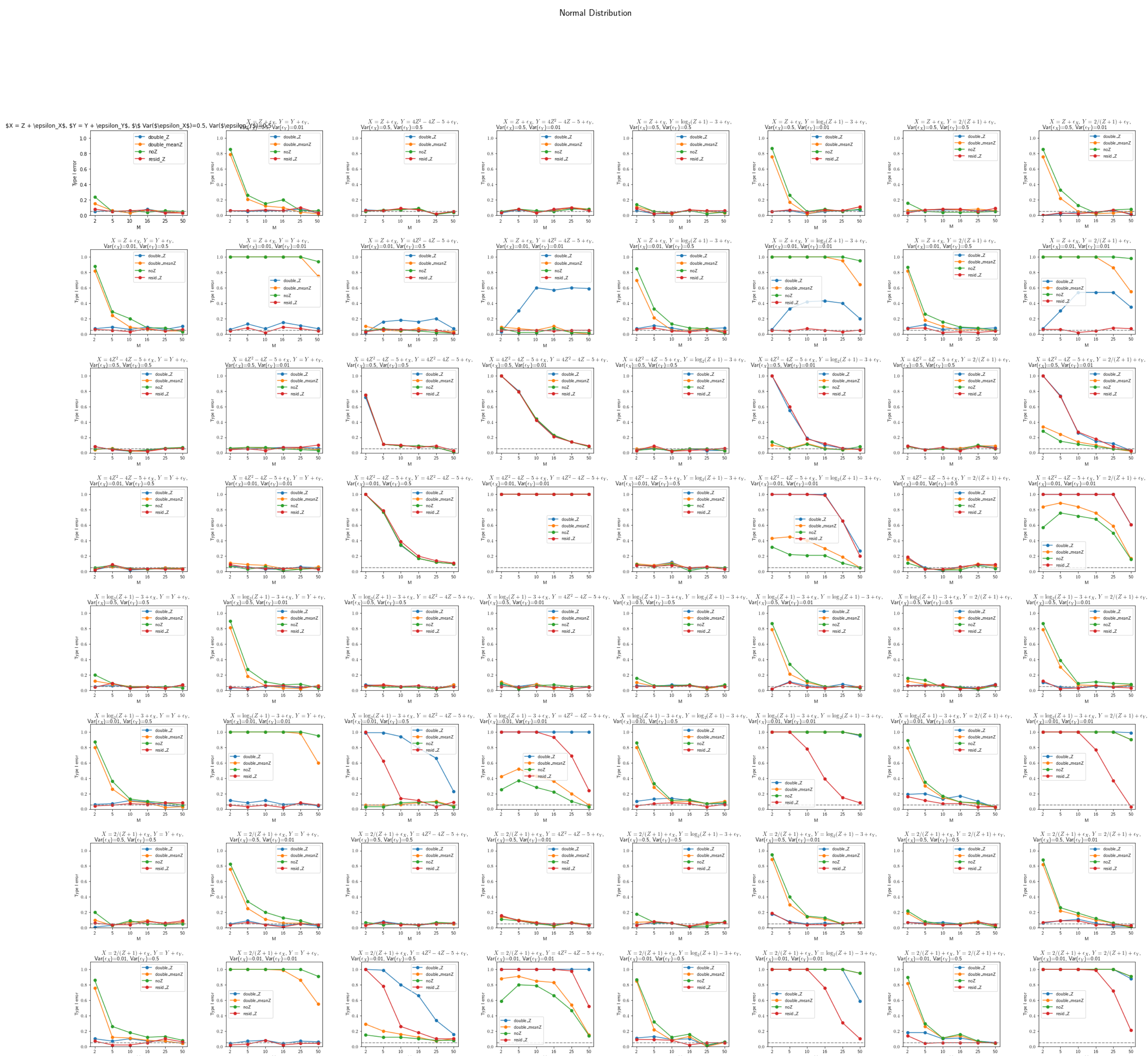
Simulation9

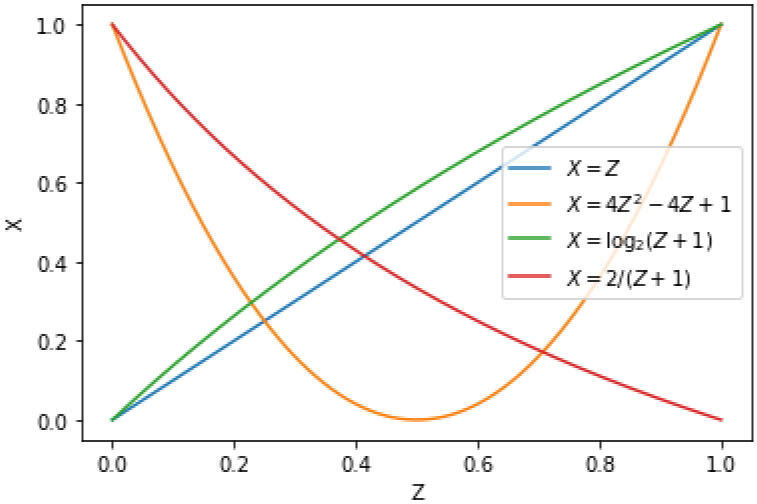
Mengqi Liu

Aug 29, 2023

Simulation result correction



Different generating functions



Analysis

- Methods: ($\tilde{Z}$ is the discretized Z, and the data belonging to the same group share the same $\tilde{Z}$.)
  - "double_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.

$$\begin{aligned} \operatorname{cor}(P_{Z^\perp}X_\sigma, P_{Z^\perp}Y) &= \frac{\mathbb{E}(X_\sigma^\top P_{Z^\perp}P_{Z^\perp}Y) - \mathbb{E}(P_{Z^\perp}X_\sigma)\mathbb{E}(P_{Z^\perp}Y)}{\sqrt{\operatorname{Var}(P_{Z^\perp}X_\sigma)\operatorname{Var}(P_{Z^\perp}Y)}} && (1) \\ &\propto \frac{\mathbb{E}(X_\sigma^\top P_{Z^\perp}Y)}{\operatorname{SD}(X_\sigma)} && (2) \end{aligned}$$

- "double_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* correlation between residuals from two linear regressions as the test statistic.

$$\begin{aligned} \operatorname{cor}(P_{\tilde{Z}^\perp}X_\sigma, P_{\tilde{Z}^\perp}Y) &= \frac{\mathbb{E}(X_\sigma^\top P_{\tilde{Z}^\perp}P_{\tilde{Z}^\perp}Y) - \mathbb{E}(P_{\tilde{Z}^\perp}X_\sigma)\mathbb{E}(P_{\tilde{Z}^\perp}Y)}{\sqrt{\operatorname{Var}(P_{\tilde{Z}^\perp}X_\sigma)\operatorname{Var}(P_{\tilde{Z}^\perp}Y)}} && (3) \\ &\propto \mathbb{E}(X_\sigma^\top P_{\tilde{Z}^\perp}Y) && (4) \end{aligned}$$

- "noZ": use *cor*(X, Y) as test statistic with local permutation in X with respect to Z.
- "resid_Z": regress Y on $1, Z$ and regress X on $1, Z$ separately. Permute residuals from regression on for X and take the *absolute* correlation between permuted residuals as the test statistic.

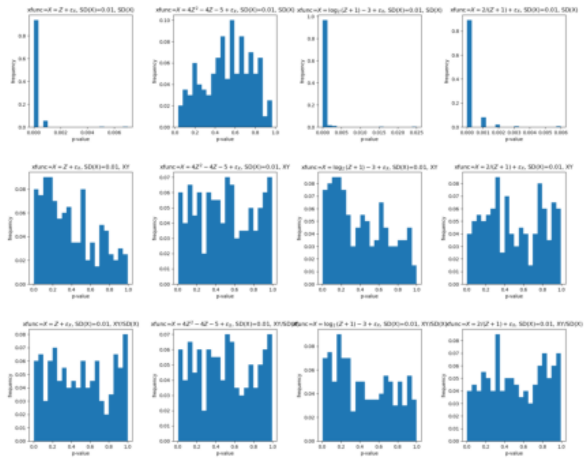
$$\begin{aligned} \operatorname{cor}((P_{Z^\perp}X)_\sigma, P_{Z^\perp}Y) &= \frac{\mathbb{E}((X^\top P_{Z^\perp})_\sigma P_{Z^\perp}Y) - \mathbb{E}((P_{Z^\perp}X)_\sigma)\mathbb{E}(P_{Z^\perp}Y)}{\sqrt{\operatorname{Var}((P_{Z^\perp}X)_\sigma)\operatorname{Var}(P_{Z^\perp}Y)}} && (5) \\ &\propto \mathbb{E}((X^\top P_{Z^\perp})_\sigma \epsilon_Y) && (6) \end{aligned}$$

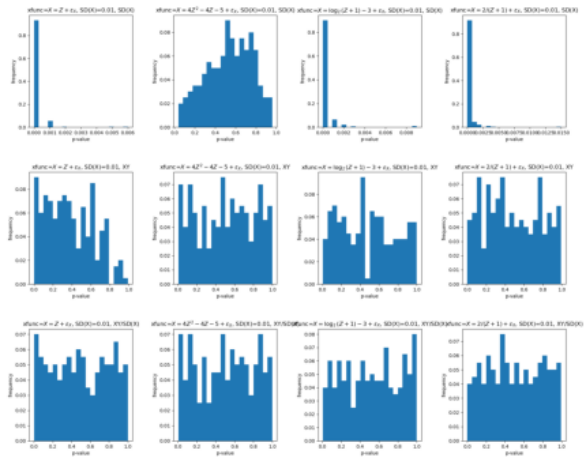
- Basically "double_meanZ" and "resid_Z" are very similar - both permuting residuals after regressing $X \sim 1 + \tilde{Z}$ and $X \sim 1 + Z$ respectively.

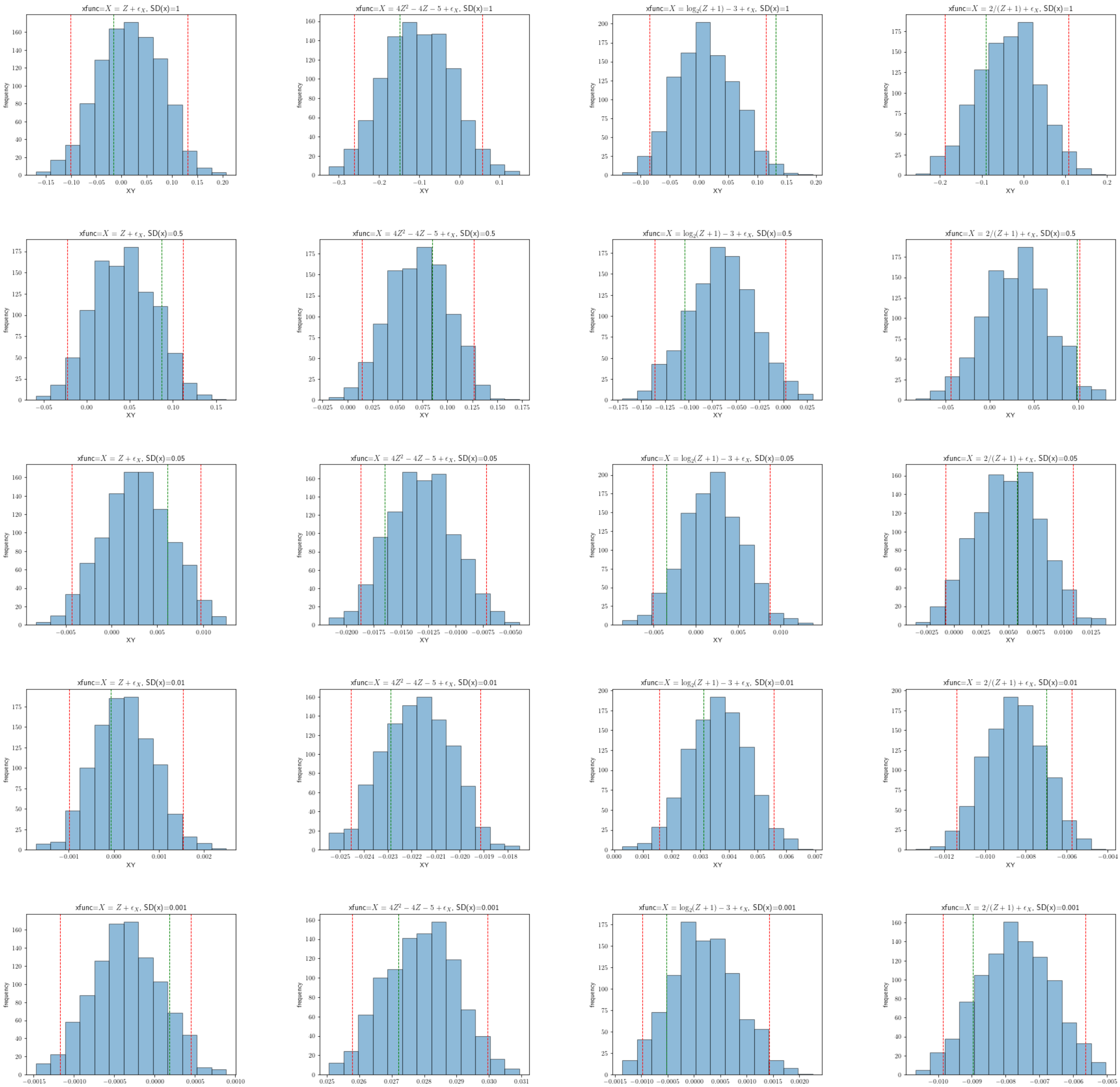
- More generally, "no_Z" is also similar - permuting residuals after regressing $X \sim 1$ but without using information of Z.

double_Z

- when Y in linear in Z:
  - sometimes with uniform distribution on $X_\sigma^\top P_{Z^\perp} Y$ and $\mathrm{SD}(X_\sigma)$ (column2)
  - sometimes with extremely small $\mathrm{SD}(X_\sigma)$ but $X_\sigma^\top P_{Z^\perp} Y$ with mean 0 (column1,3,4) -$\longrightarrow$ uniform distribution on $\frac{X_\sigma^\top P_{Z^\perp} Y}{\mathrm{SD}(X_\sigma)}$

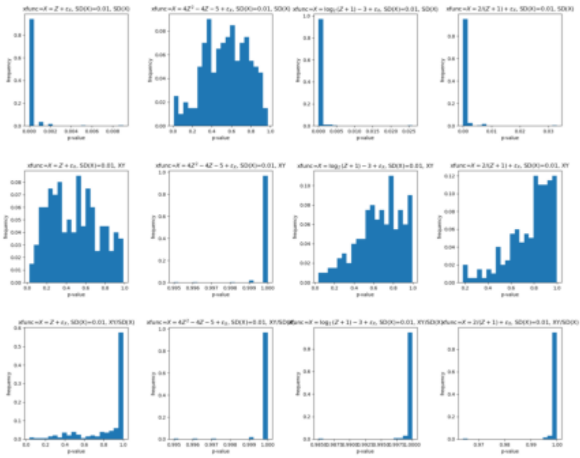


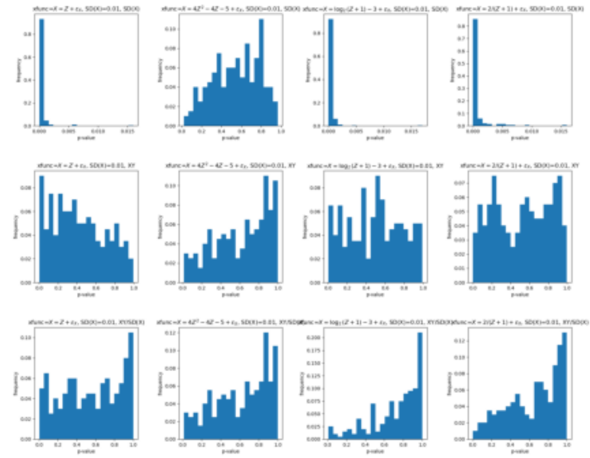


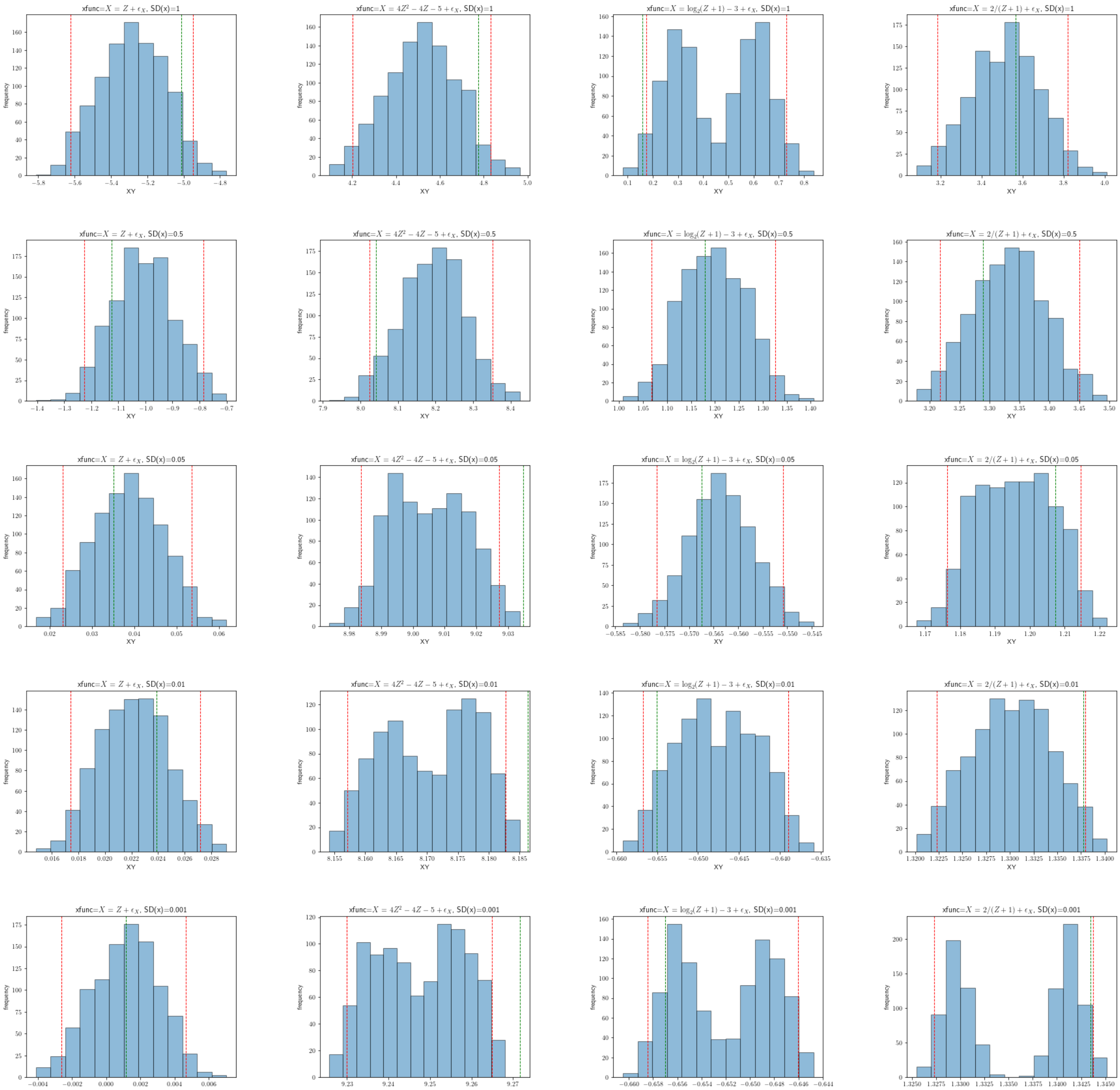


double_Z

- when Y in linear in Z:
  - sometimes with uniform distribution on $\mathrm{SD}(X_\sigma)$ but extremely high $X_\sigma^\top P_{Z^\perp} Y$ (column2) $\rightarrow$ maybe due to the same generating function?
  - sometimes with extremely small $\mathrm{SD}(X_\sigma)$ and $X_\sigma^\top P_{Z^\perp} Y$ with generally non-zero mean (column1,3,4) -$\rightarrow$ heavy density on large percentile(>0.95) $\frac{X_\sigma^\top P_{Z^\perp} Y}{\mathrm{SD}(X_\sigma)}$



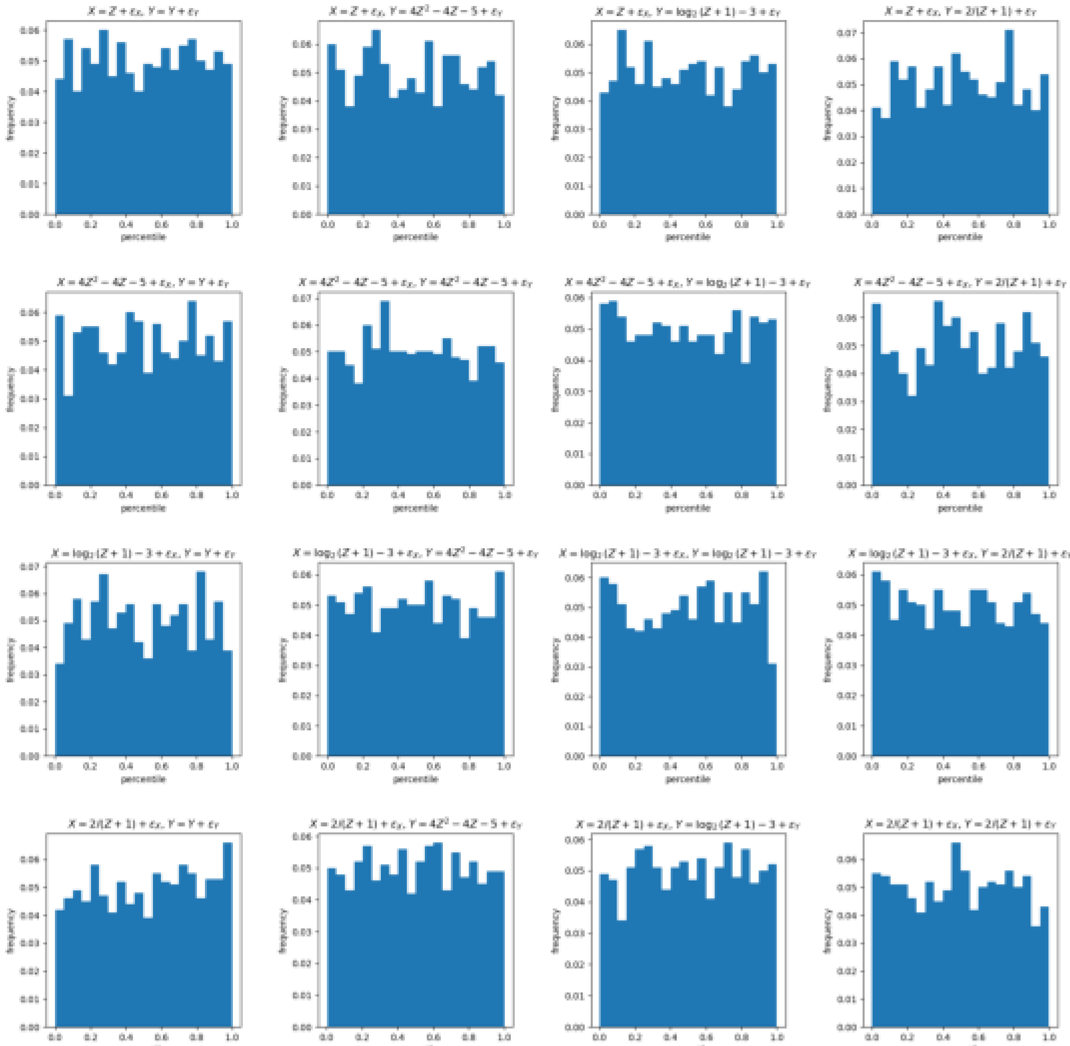




double_meanZ

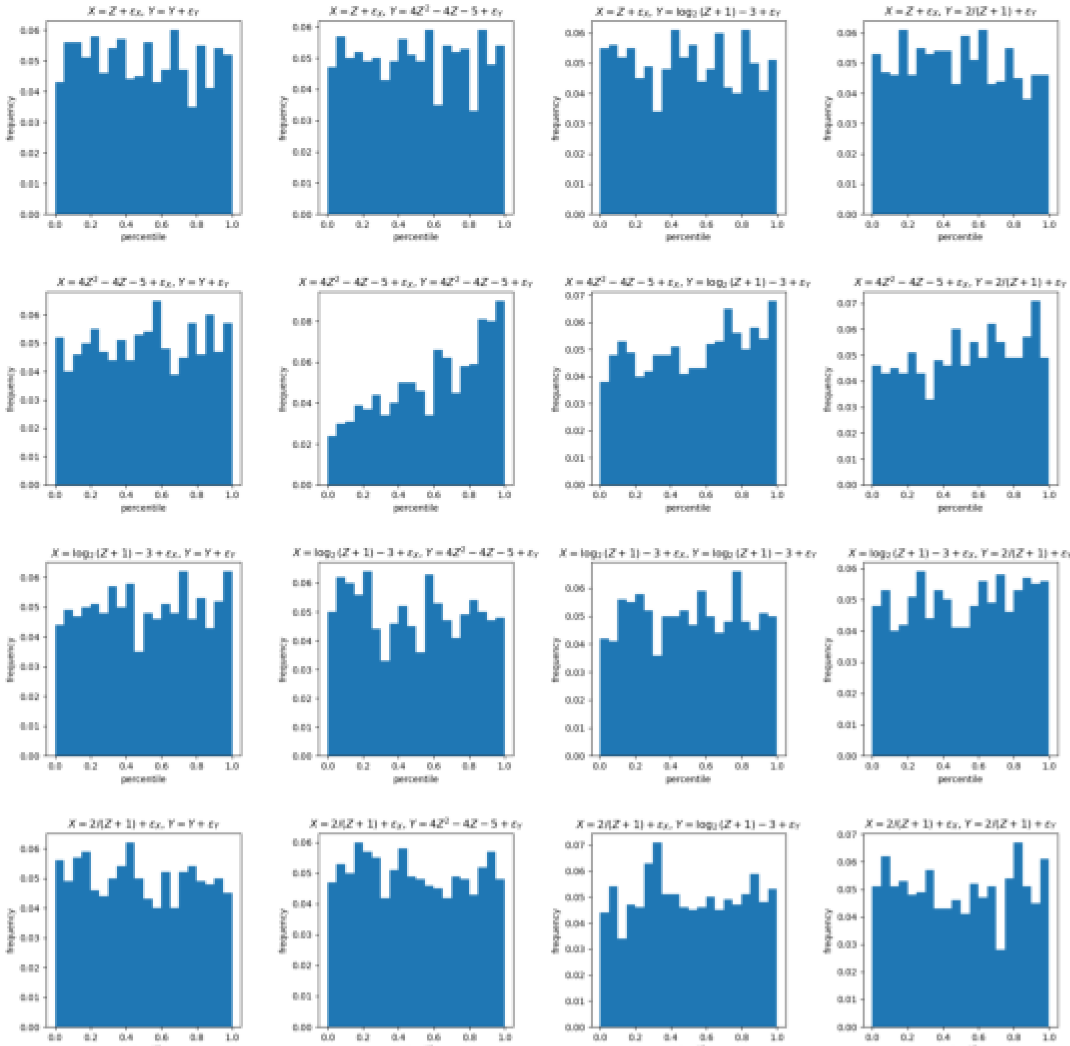
- $\mathrm{SD}(X) = 0.5, \mathrm{SD}(Y) = 0.5$

```
<matplotlib.image.AxesImage at 0x7fe3f3498580>
```



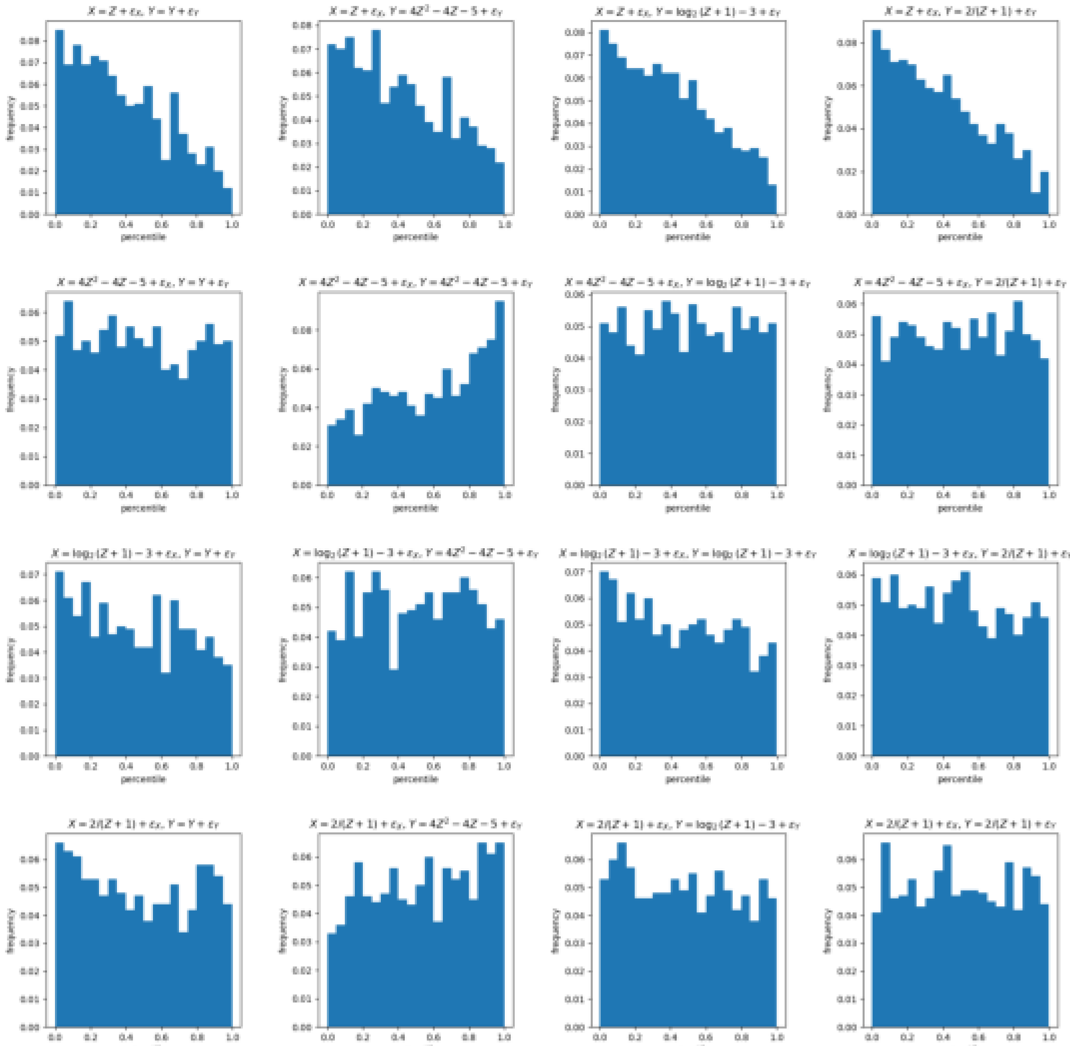
- $\mathrm{SD}(X) = 0.5, \mathrm{SD}(Y) = 0.01$

```
<matplotlib.image.AxesImage at 0x7fe3f33d3a30>
```



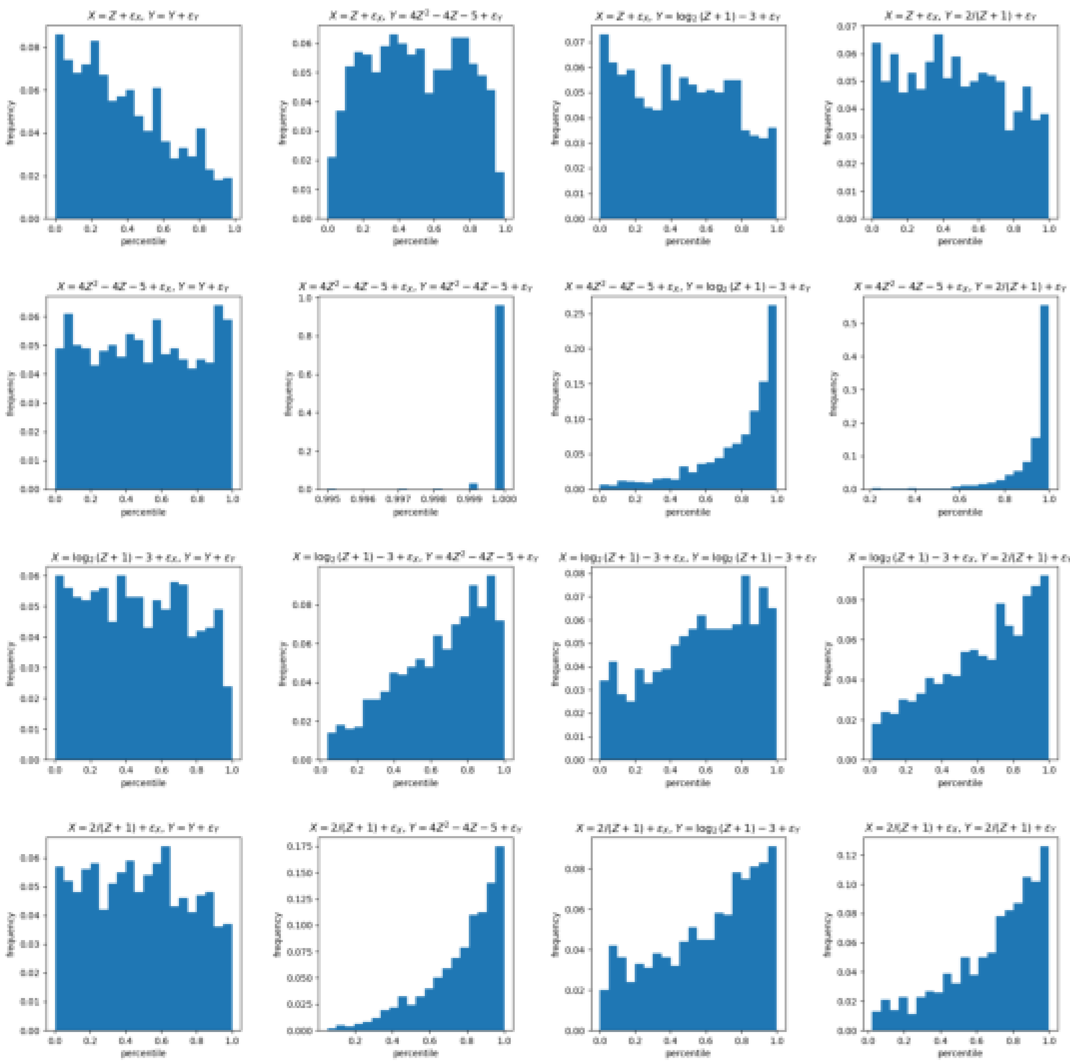
- $\mathrm{SD}(X) = 0.01, \mathrm{SD}(Y) = 0.5$

```
<matplotlib.image.AxesImage at 0x7fe3f3a7f160>
```



- $\mathrm{SD}(X) = 0.01, \mathrm{SD}(Y) = 0.01$

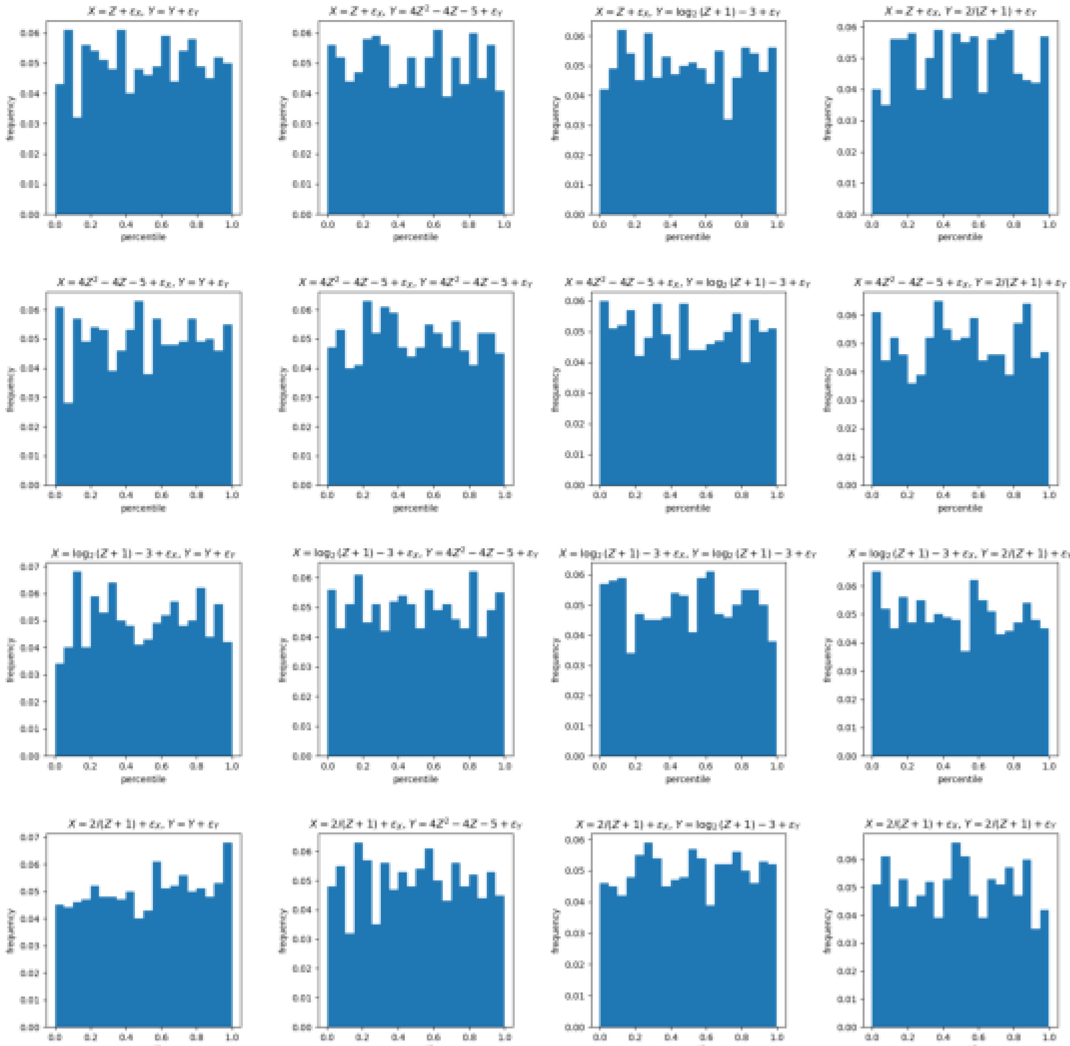
```
<matplotlib.image.AxesImage at 0x7fe3e3954490>
```



resid_Z

- $\mathrm{SD}(X) = 0.5, \mathrm{SD}(Y) = 0.5$

```
<matplotlib.image.AxesImage at 0x7fe3e3215f70>
```

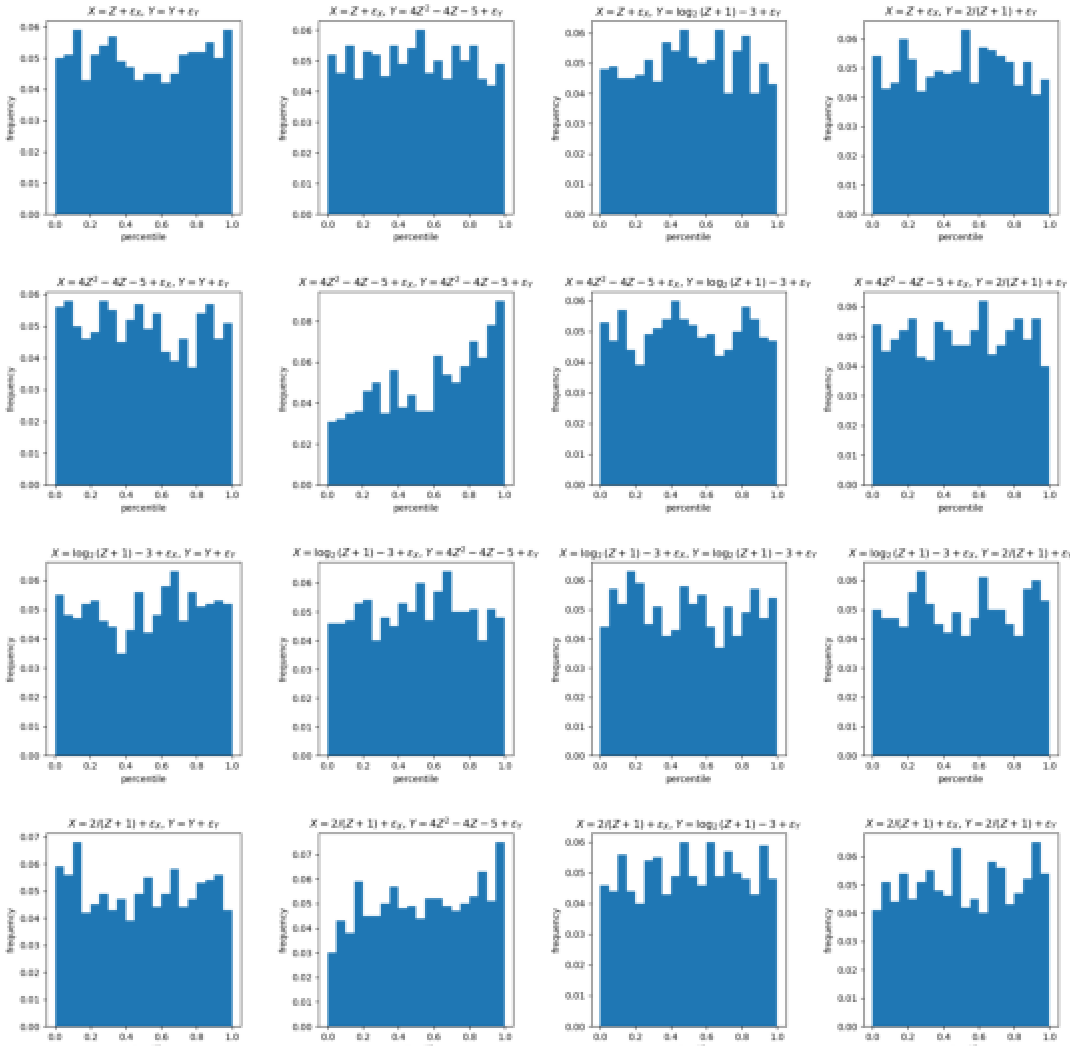


- $\mathrm{SD}(X) = 0.5, \mathrm{SD}(Y) = 0.01$

```
<matplotlib.image.AxesImage at 0x7fe3d3c791c0>
```

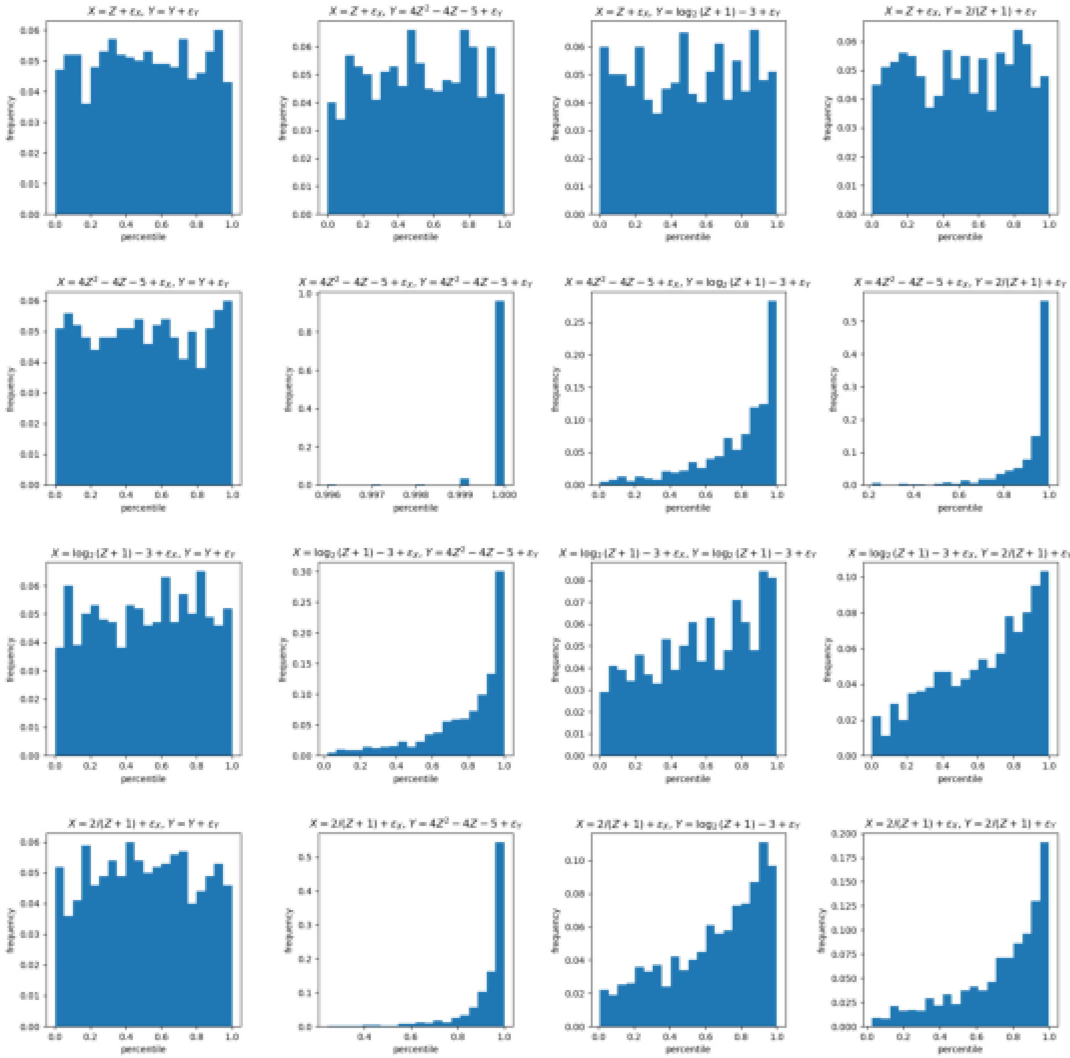
- $\mathrm{SD}(X) = 0.01, \mathrm{SD}(Y) = 0.5$

```
<matplotlib.image.AxesImage at 0x7fe3d41ebf10>
```



- $\mathrm{SD}(X) = 0.01, \mathrm{SD}(Y) = 0.01$

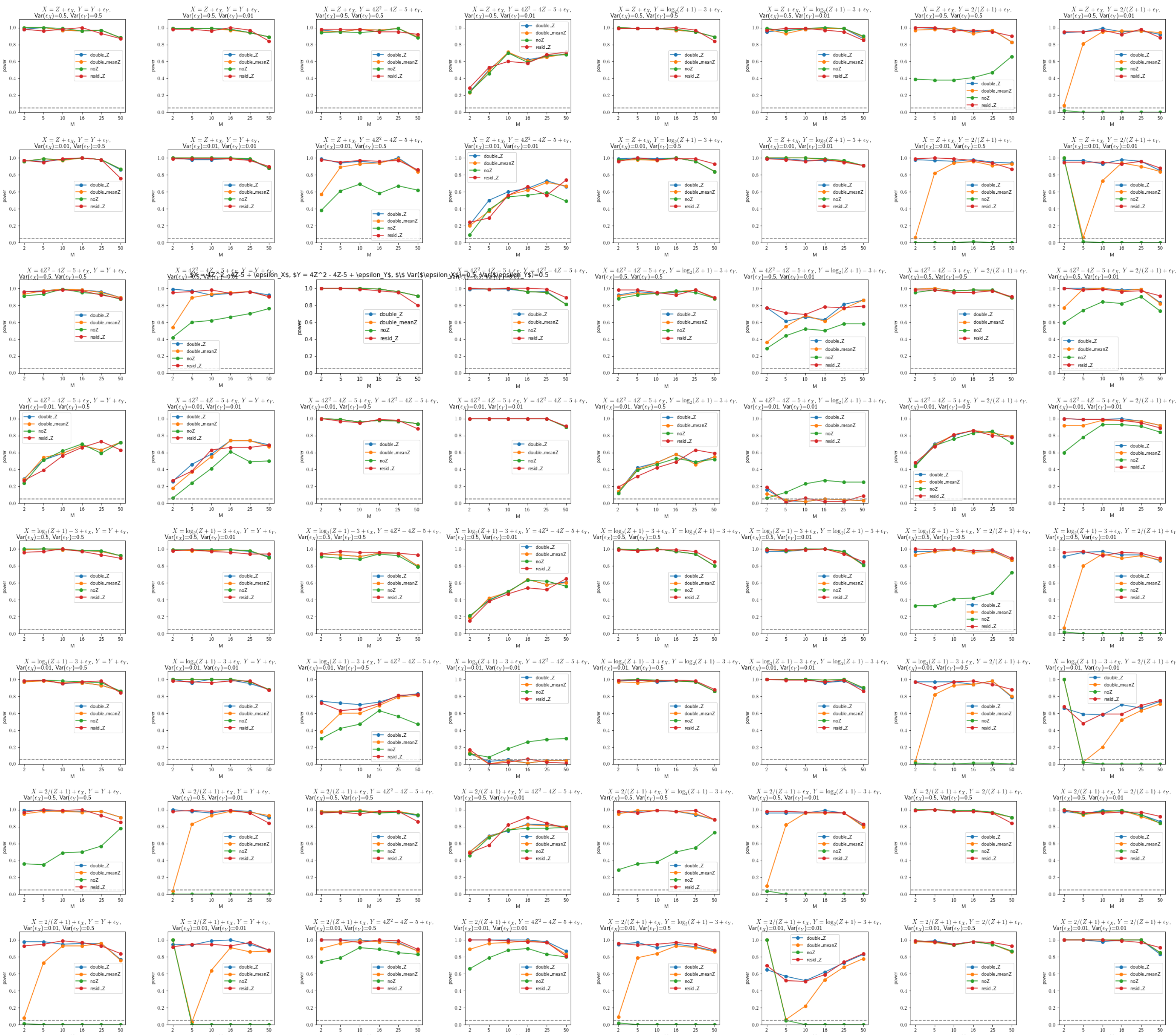
```
<matplotlib.image.AxesImage at 0x7fe3d506aa00>
```



Power

- cor = 0.4

Normal Distribution



- cor = 0.2

Normal Distribution