

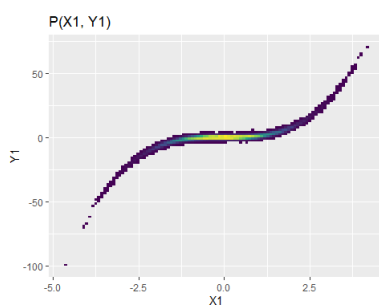
۱ تولید و تصویر داده علی

۱.۱

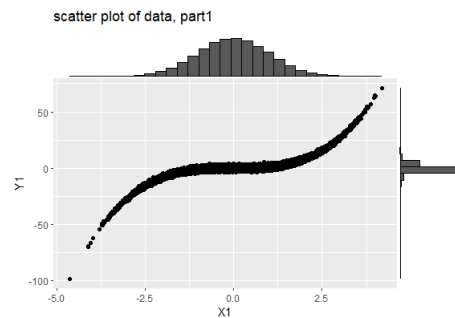
برای مدل زیر توزیع‌های خواسته شده را رسم می‌کنیم.

$$X_1 = N_X, Y_1 = X_1^3 + N_Y$$

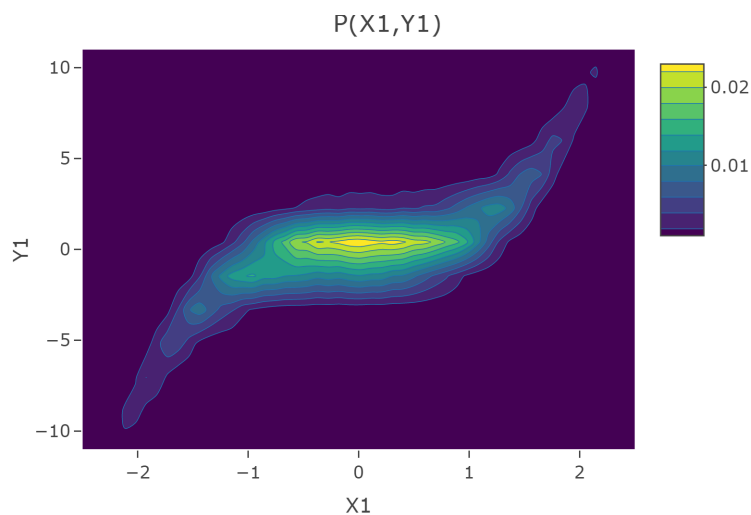
$$N_X \sim N(0, 1), \quad N_Y \sim N(0, 1), \quad N_X \perp N_Y$$



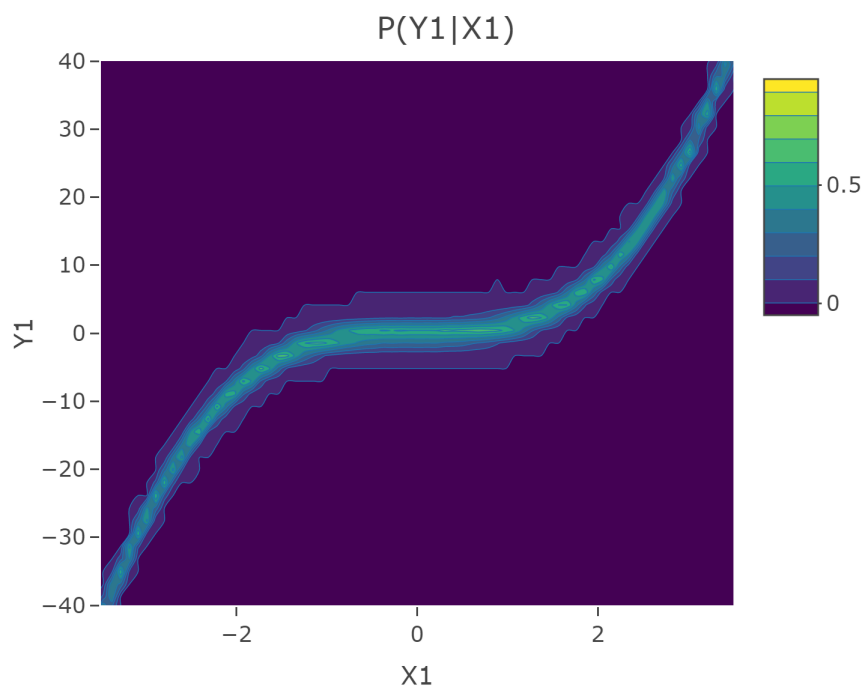
توزیع توام



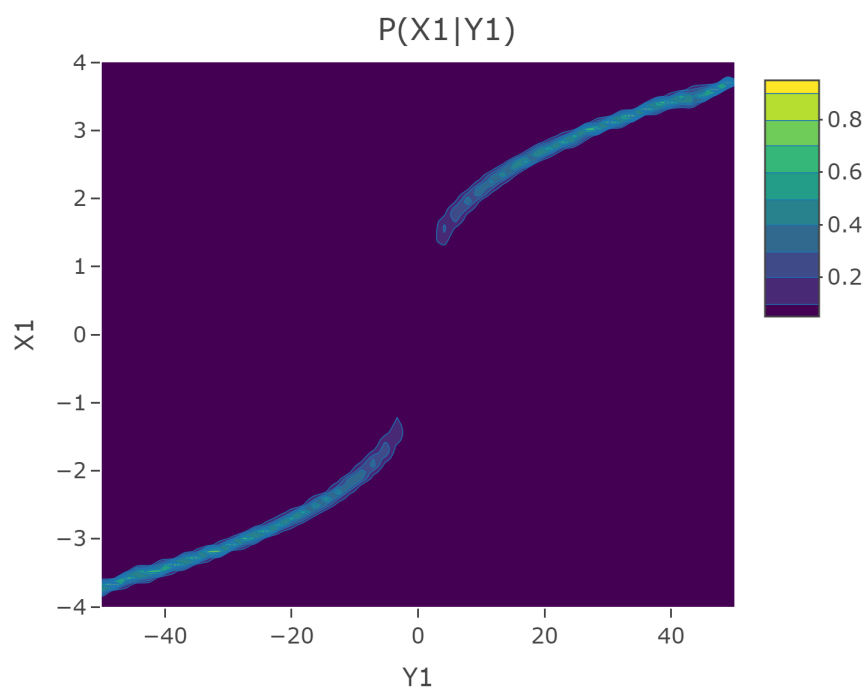
نمودار داده و توزیع‌های حاشیه‌ای



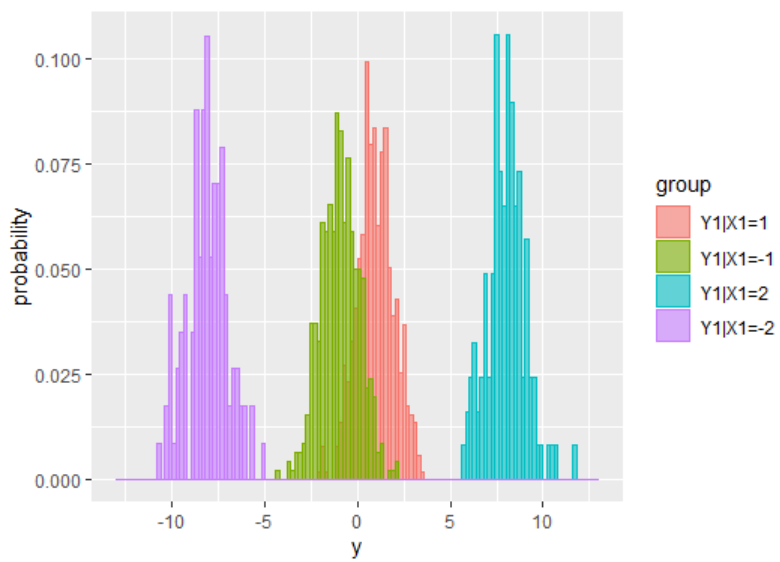
نمایش دیگر توزیع توام



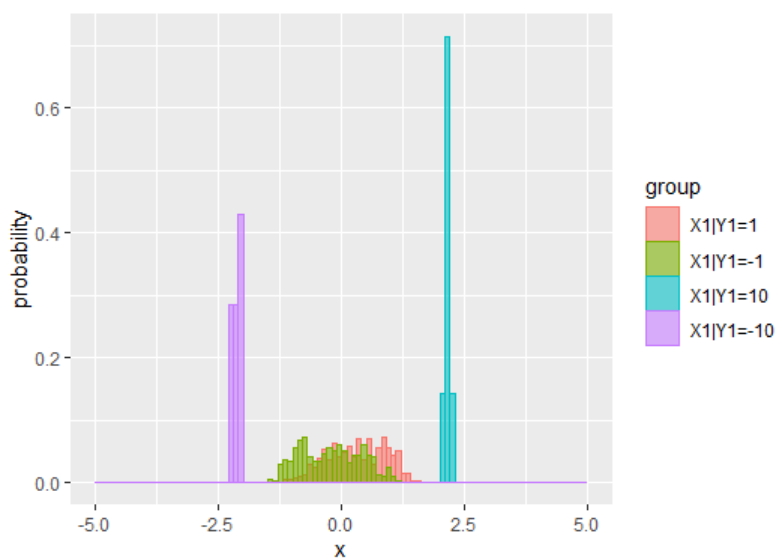
توزيع $Y_1|X_1$



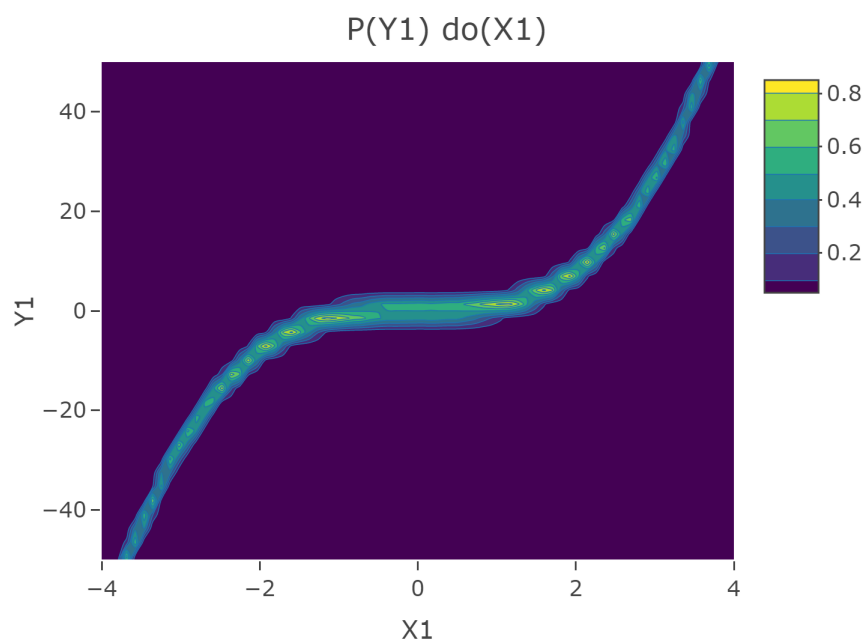
توزيع $X_1|Y_1$



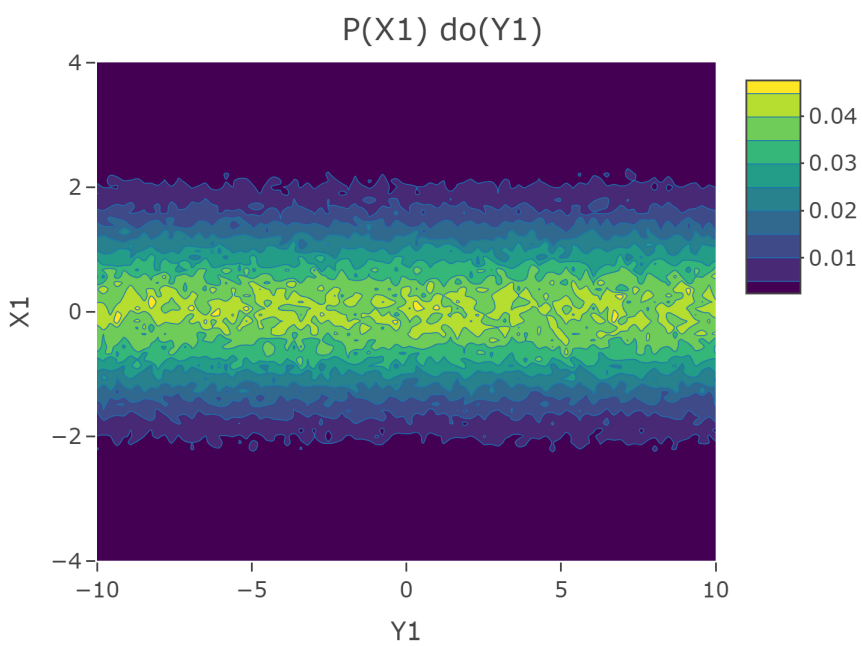
هیستوگرام مربوط به توزیع شرطی Y_1 به ازای X_1 های مختلف



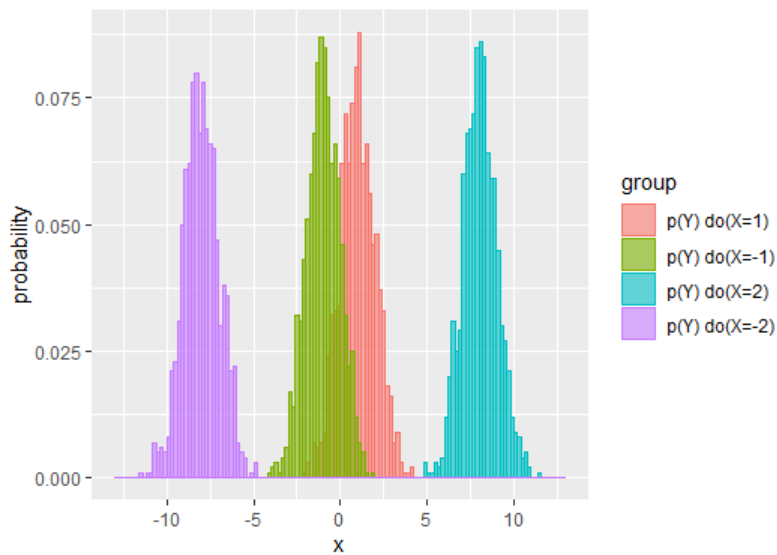
هیستوگرام مربوط به توزیع شرطی X_1 به ازای Y_1 های مختلف



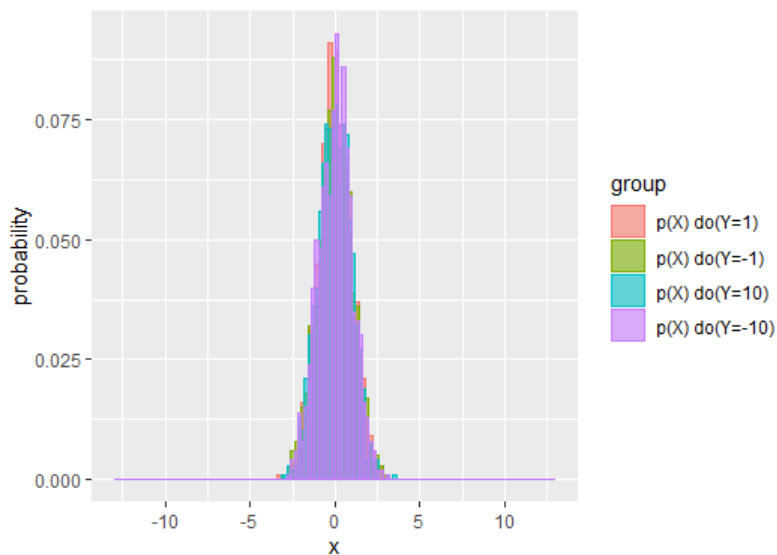
توزیع مداخله ای $P_{Y_1}^{do(X_1)}(y_1)$



توزیع مداخله ای $P_{X_1}^{do(Y_1)}(x_1)$

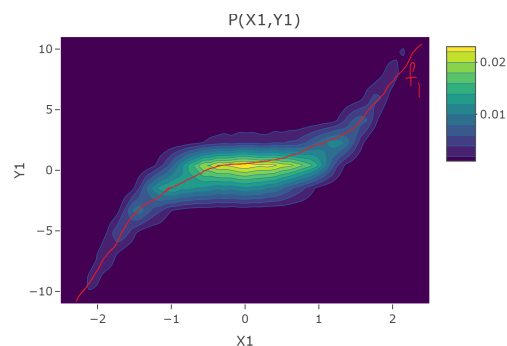


هیستوگرام مربوط به توزیع مداخله‌ای Y_1 به ازای X_1 های مختلف



هیستوگرام مربوط به توزیع مداخله‌ای X_1 به ازای Y_1 های مختلف

بررسی تشخیص پذیری: با توجه به شکل زیر تابع f_1 را می‌توان یافت به طوری که نوشت $Y = f_1(X) + N_y$ که در آن N_y و X مستقلند؛ در جهت عکس باتوجه به توزیع توام همچنین تابعی یافت نمی‌شود بنابراین طبق قضیه 4.5 کتاب می‌توان گفت X به Y جهت علی صحیح است.

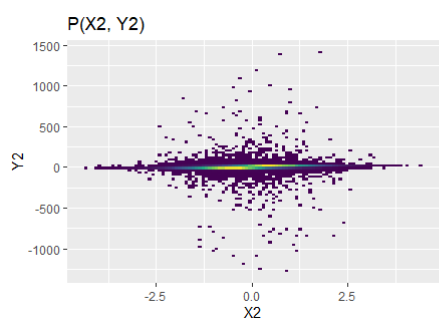


۲.۱

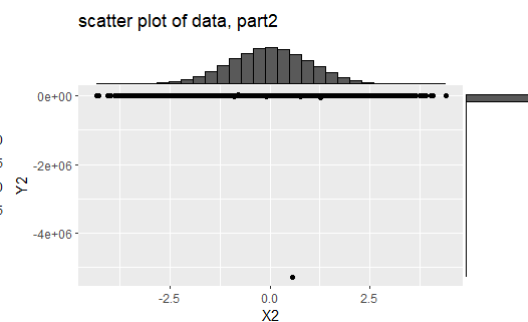
برای مدل زیر توزیع‌های خواسته شده را رسم می‌کنیم.

$$X_2 = N_X, Y_2 = 2 \times X_2 + N_Y$$

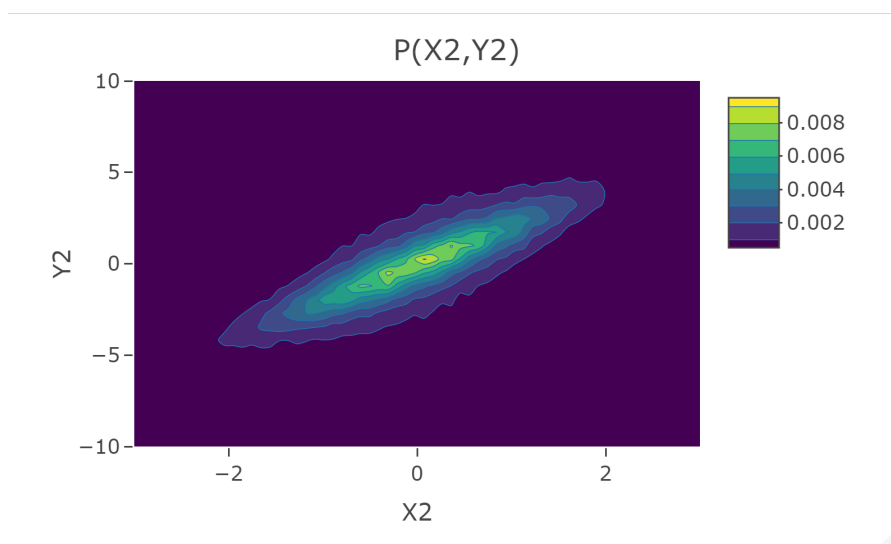
$$N_X \sim N(0, 1), \quad N_Y \sim T(1), \quad N_X \perp N_Y$$



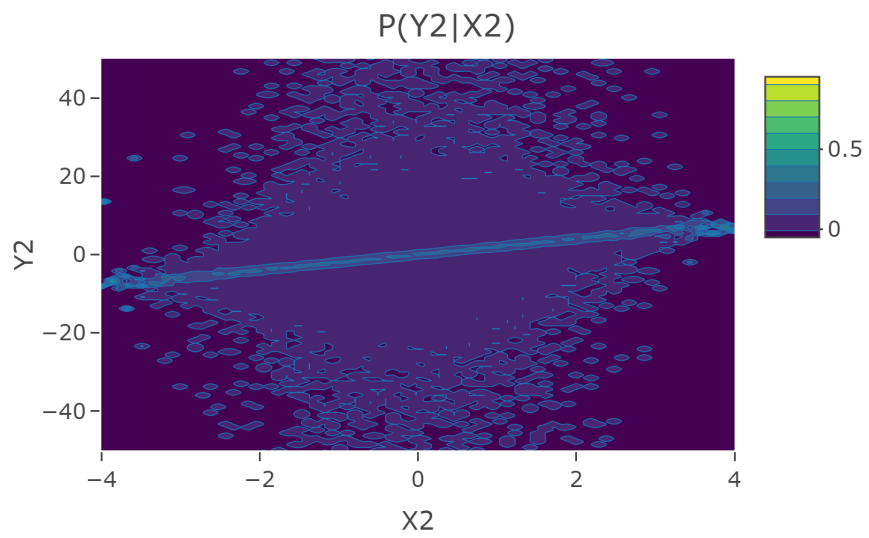
توزیع توام



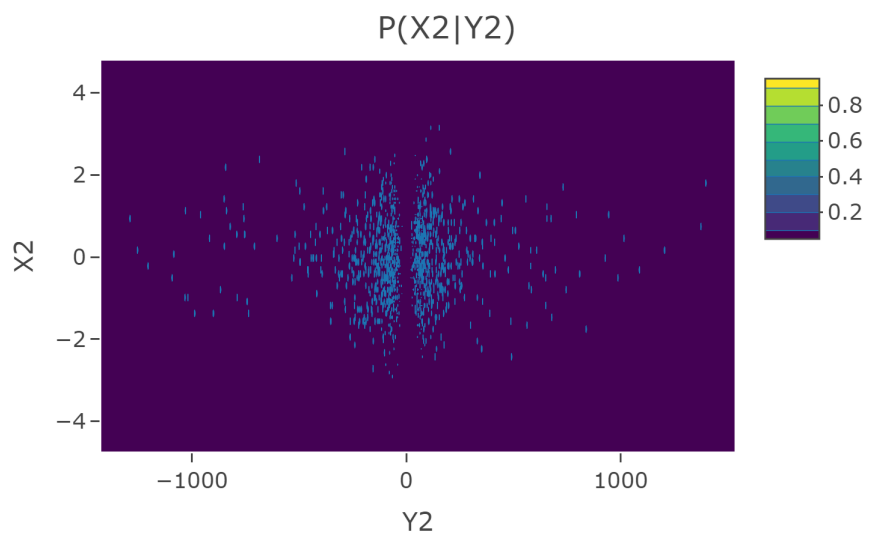
نمودار داده و توزیع‌های حاشیه‌ای



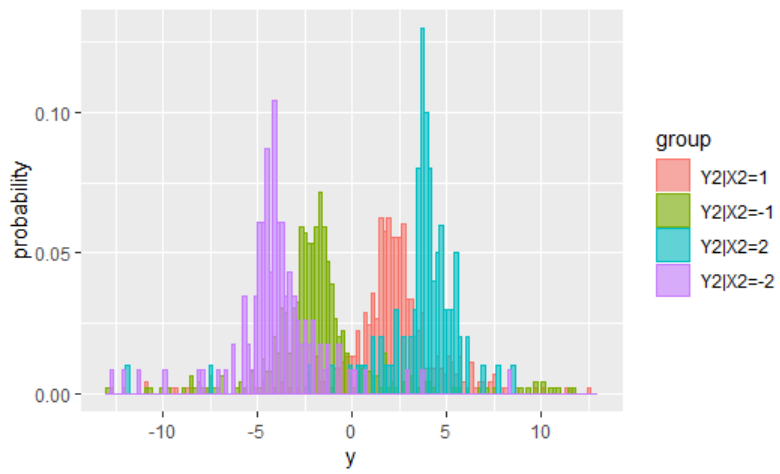
نمایش دیگر توزیع توام



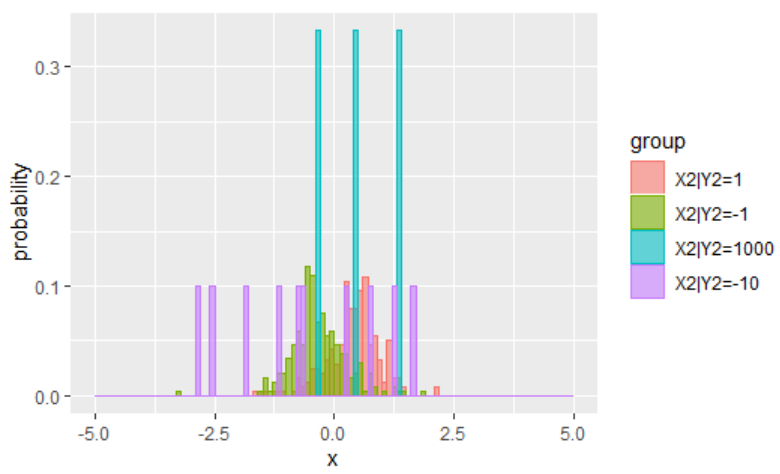
توزيع $Y_2|X_2$



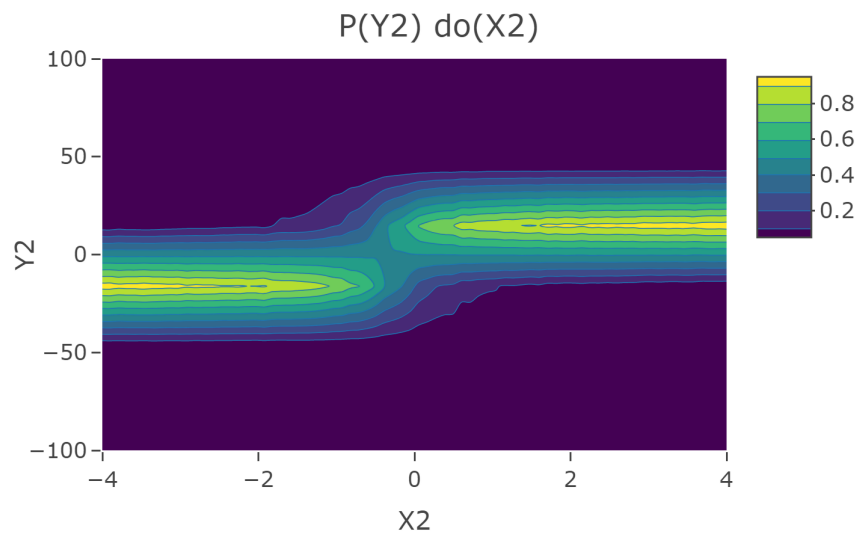
توزيع $X_2|Y_2$



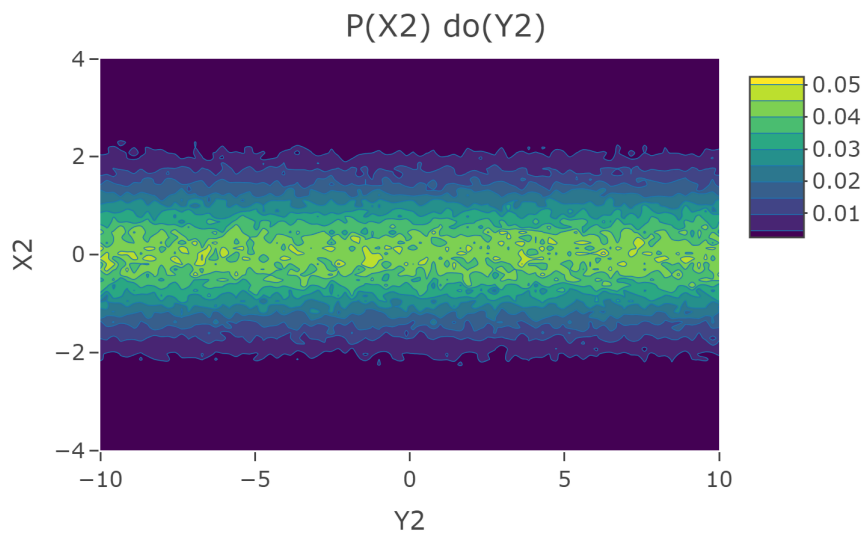
هیستوگرام مربوط به توزیع شرطی Y_2 به ازای X_2 های مختلف



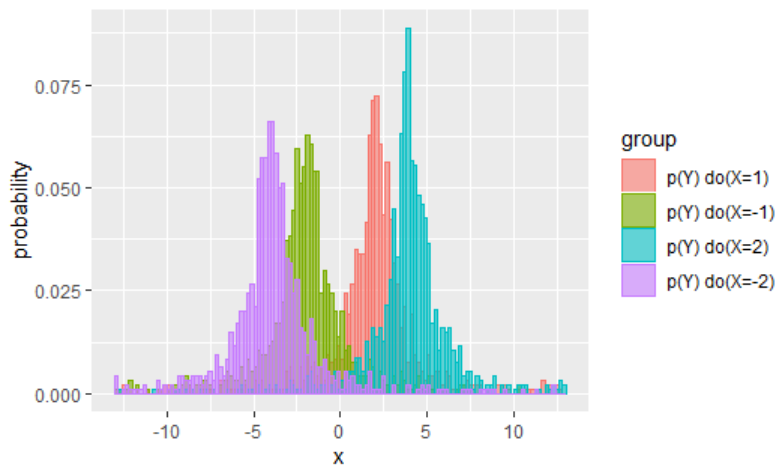
هیستوگرام مربوط به توزیع شرطی X_2 به ازای Y_2 های مختلف



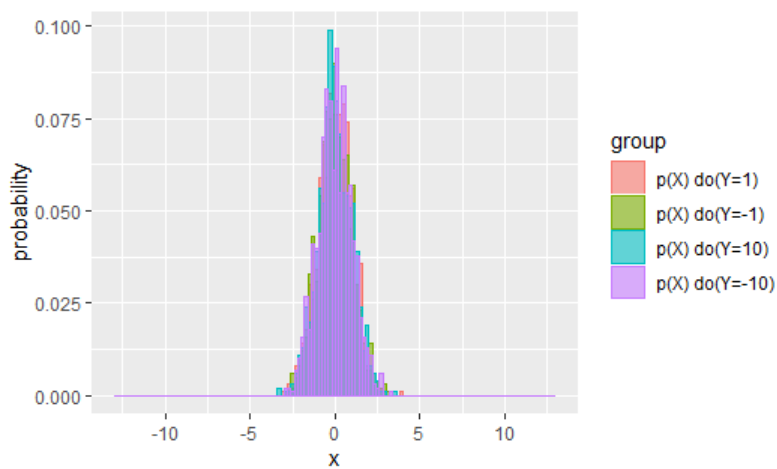
توزیع مداخله ای $P_{Y_2}^{do(X_2)}(y_2)$



توزیع مداخله ای $P_{X_2}^{do(Y_2)}(x_2)$

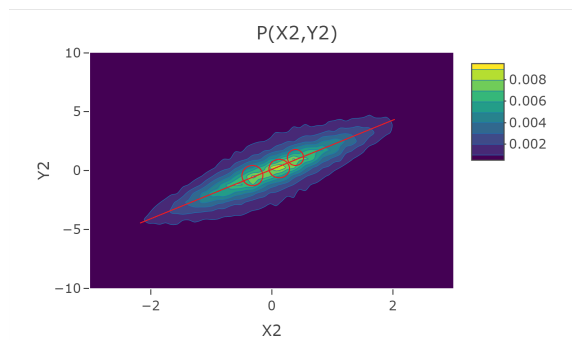


هیستوگرام مربوط به توزیع مداخله‌ای Y_2 به ازای X_2 های مختلف



هیستوگرام مربوط به توزیع مداخله‌ای X_2 به ازای Y_2 های مختلف

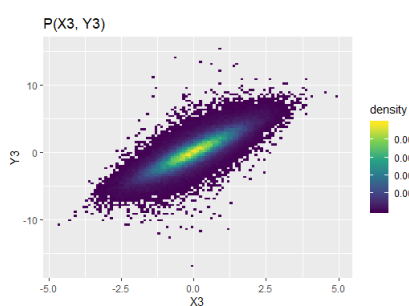
بررسی تشخیص پذیری: با توجه به شکل زیر ضریب α را می‌توان یافت به طوری که نوشت $Y = \alpha(X) + N_y$ که در آن N_y و X مستقلند؛ در جهت عکس باتوجه به توزیع توام و وجود چند قله به جای یک قله (نرمال نبودن N_y باتوجه به تراکم زیاد حول صفر و توجه به قضیه 4.2) چنین نیست، بنابراین می‌توان گفت X به Y جهت علی صحیح است. در واقع توزیع t به توزیع نرمال در حد نزدیک است اما همانگونه که در شکل می‌بینیم توزیع توام کاملاً گاوسی نیست و هنوز قابل تشخیص است.



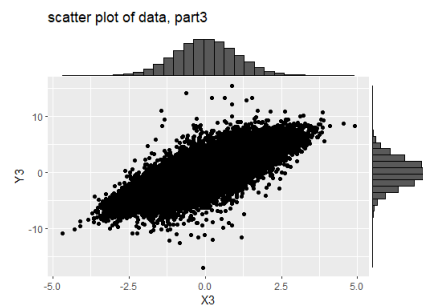
برای مدل زیر توزیع‌های خواسته شده را رسم می‌کنیم.

$$X_3 = N_X, Y_3 = 2 \times X_3 + N_Y$$

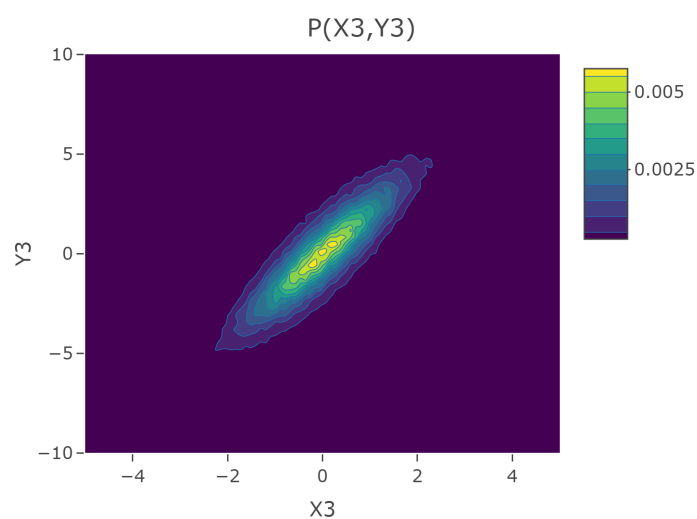
$$N_X \sim N(0, 1), \quad N_Y \sim T(5), \quad N_X \perp N_Y$$



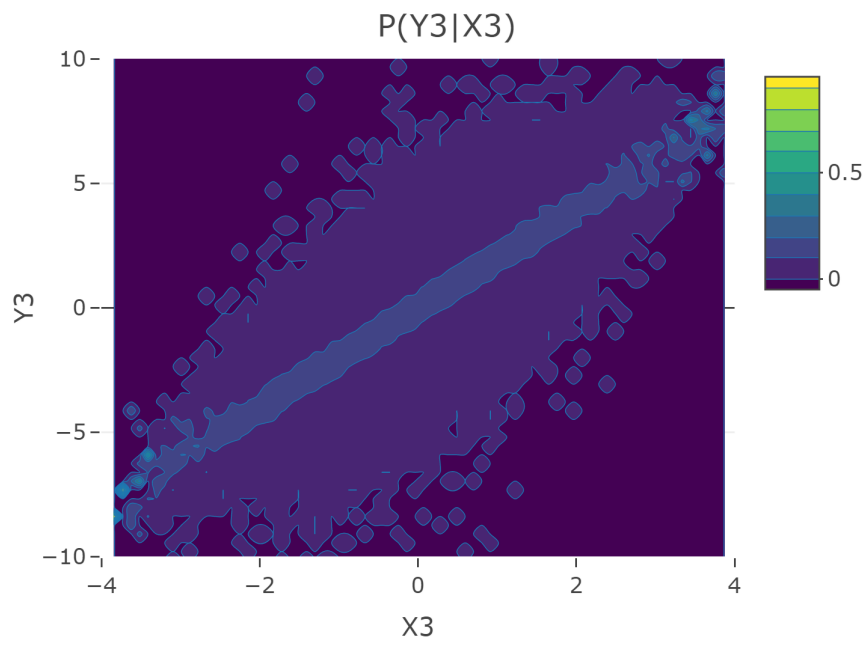
توزیع توام



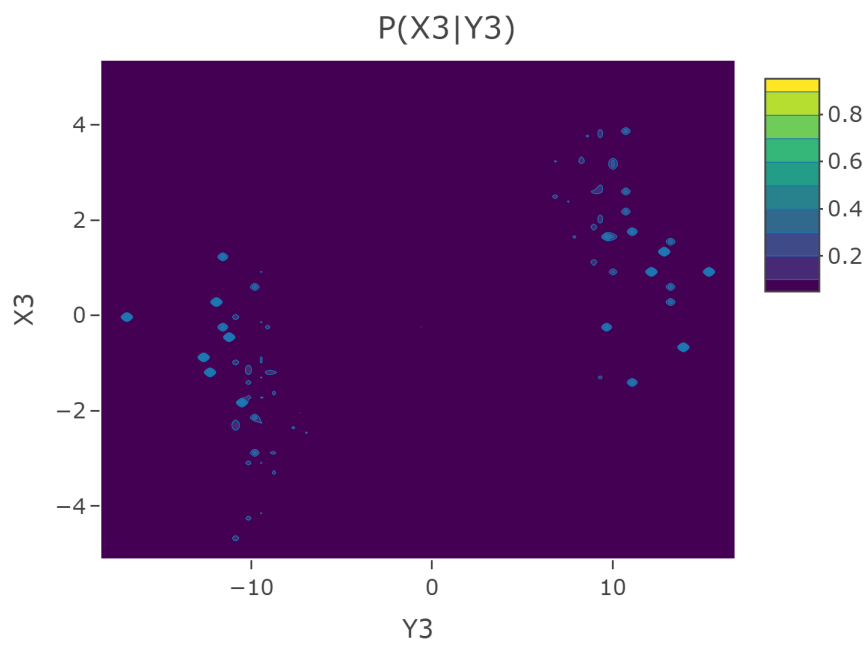
نمودار داده و توزیع‌های حاشیه‌ای



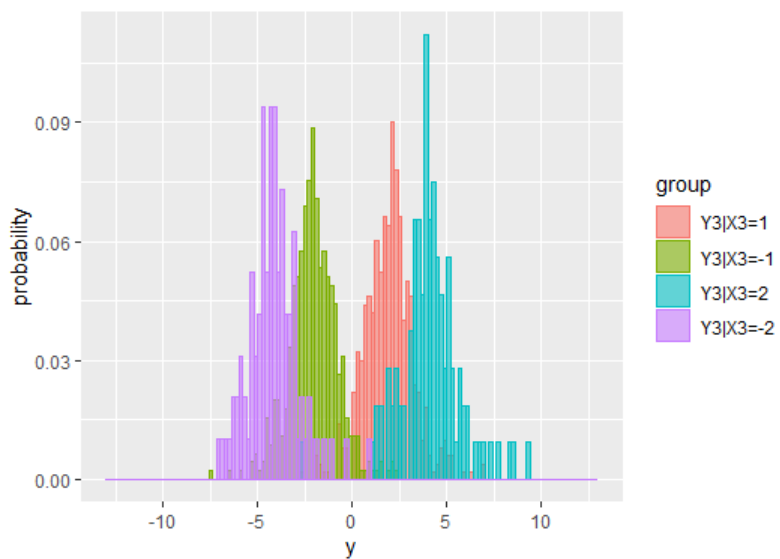
نمایش دیگر توزیع توام



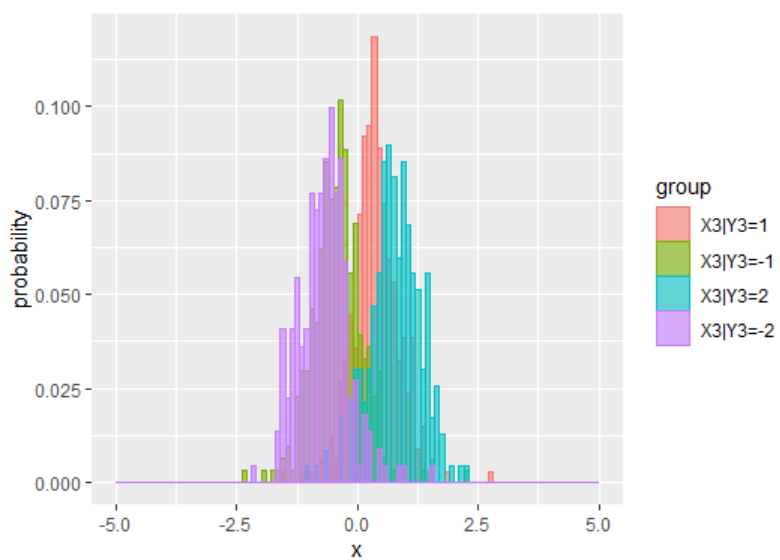
توزيع $Y_3|X_3$



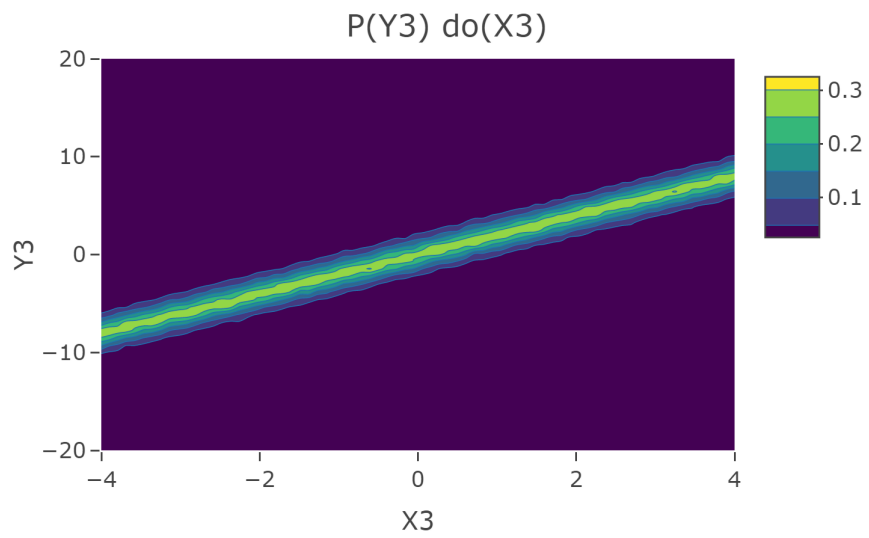
توزيع $X_3|Y_3$



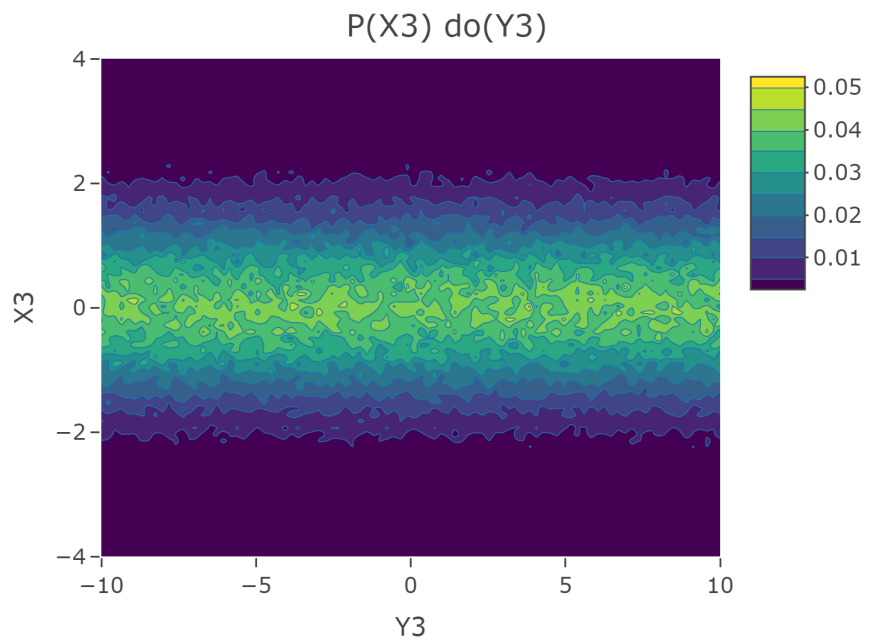
هیستوگرام مربوط به توزیع شرطی Y_3 به ازای X_3 های مختلف



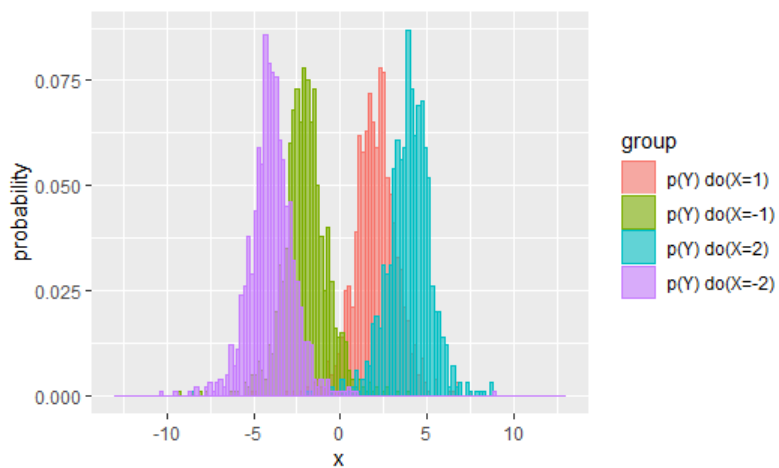
هیستوگرام مربوط به توزیع شرطی X_3 به ازای Y_3 های مختلف



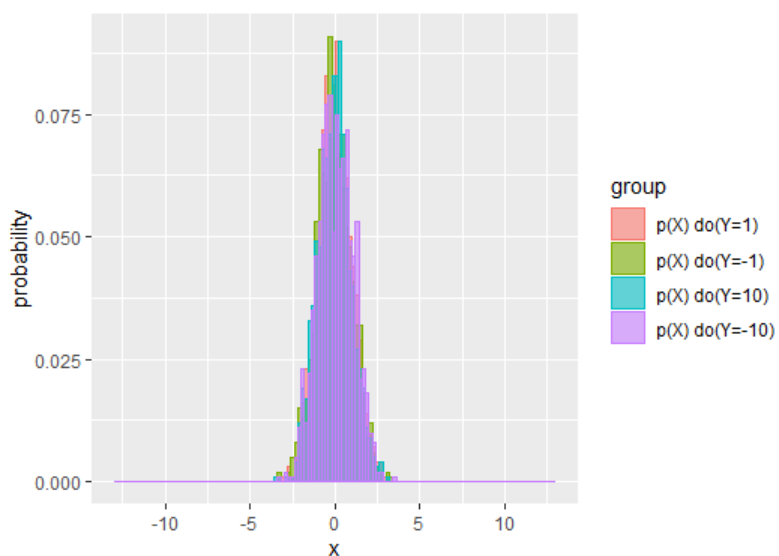
توزیع مداخله ای $P_{Y_3}^{do(X_3)}(y_3)$



توزیع مداخله ای $P_{X_3}^{do(Y_3)}(x_3)$



هیستوگرام مربوط به توزیع مداخله‌ای Y_3 به ازای X_3 های مختلف



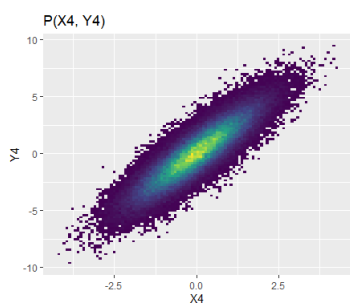
هیستوگرام مربوط به توزیع مداخله‌ای X_3 به ازای Y_3 های مختلف

بررسی تشخیص پذیری : مشابه بخش قبل است، البته در اینجا به تقارن گاوسی نزدیک تر شده ایم و با توجه به تخمینی بودن توزیع توام تشخیص پذیری میتواند سخت تر باشد.

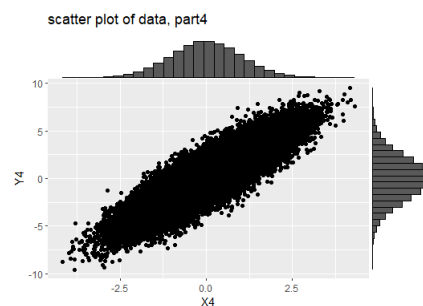
برای مدل زیر توزیع‌های خواسته شده را رسم می‌کنیم.

$$X_4 = N_X, Y_4 = 2 \times X_4 + N_Y$$

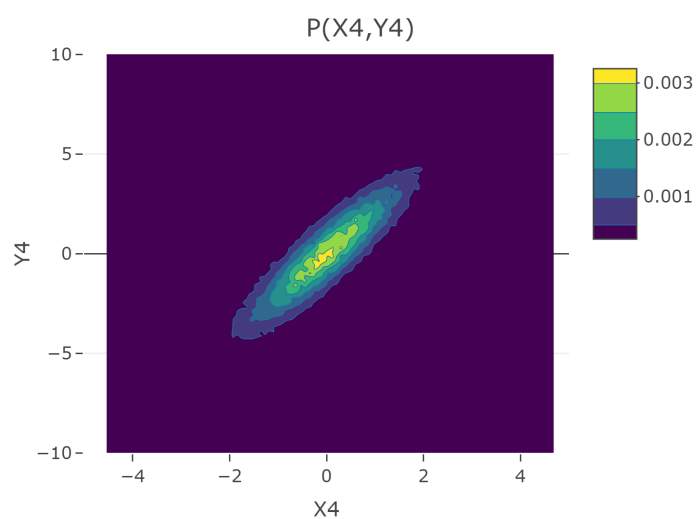
$$N_X \sim N(0, 1), \quad N_Y \sim T(20), \quad N_X \perp N_Y$$



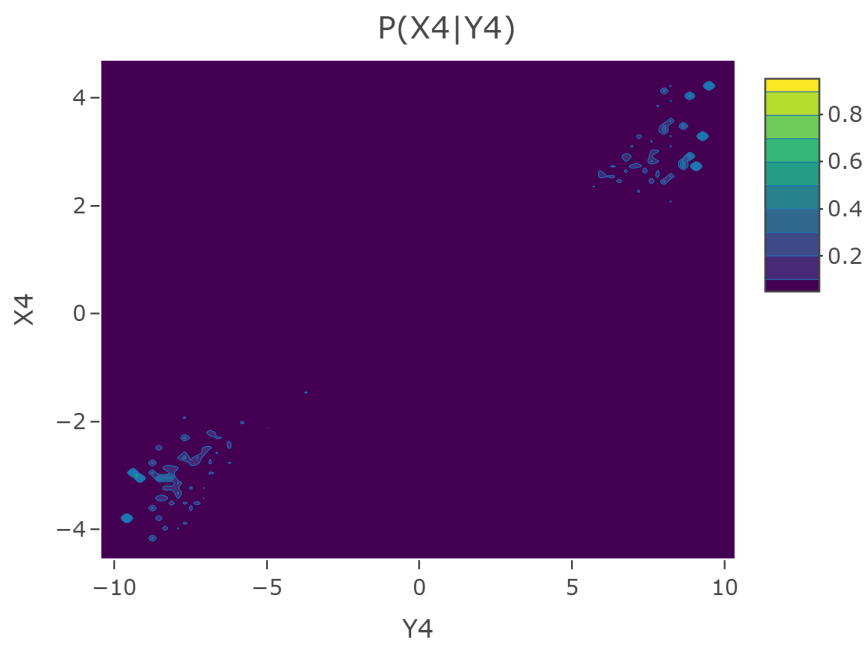
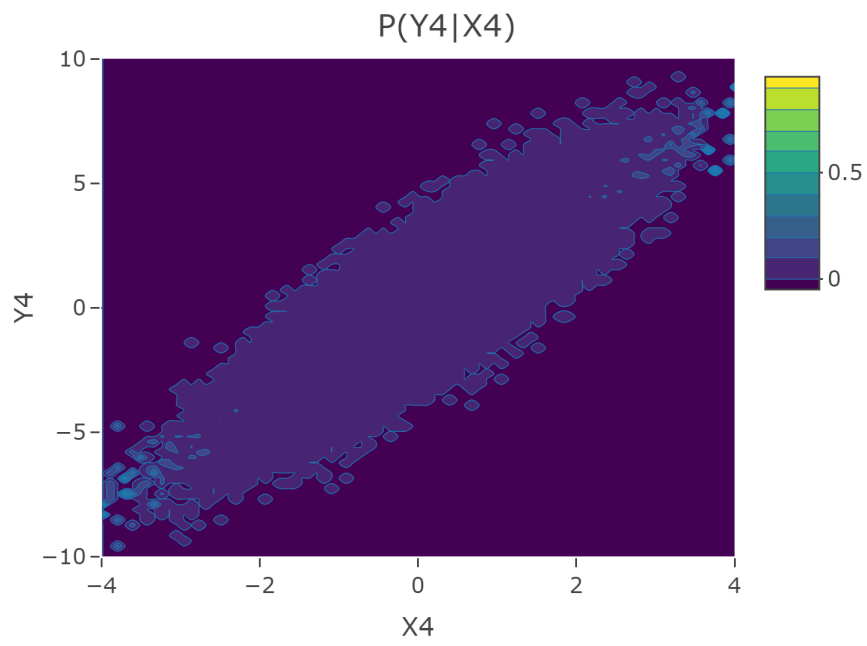
توزیع توام

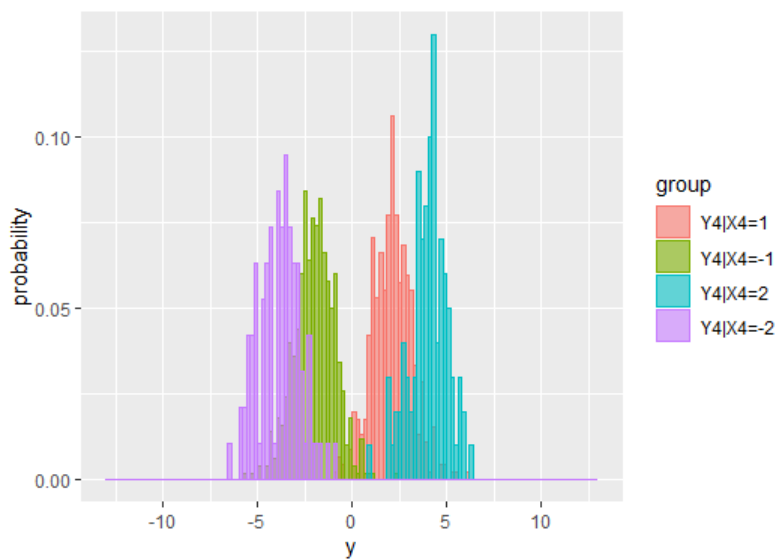


نمودار داده و توزیع‌های حاشیه‌ای

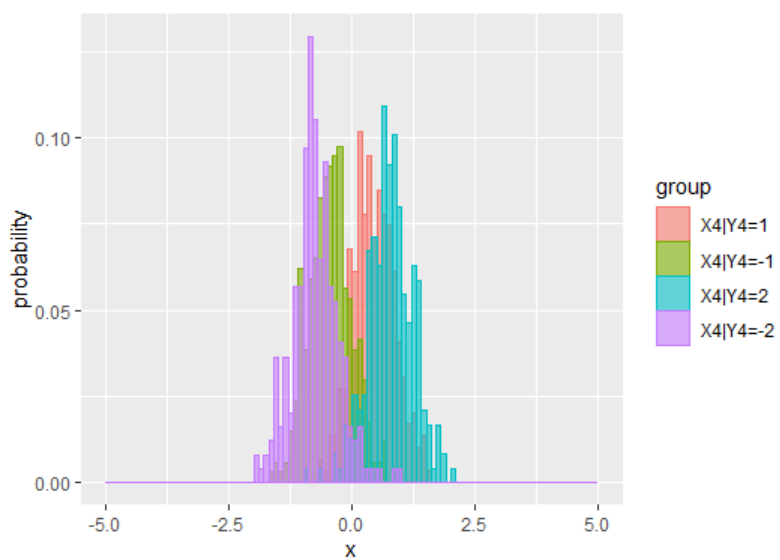


نمایش دیگر توزیع توام

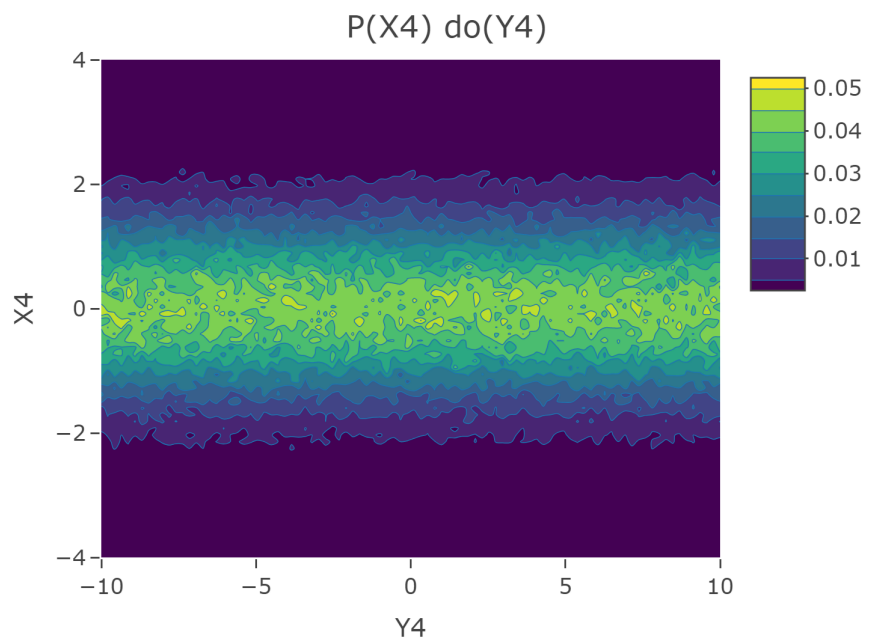
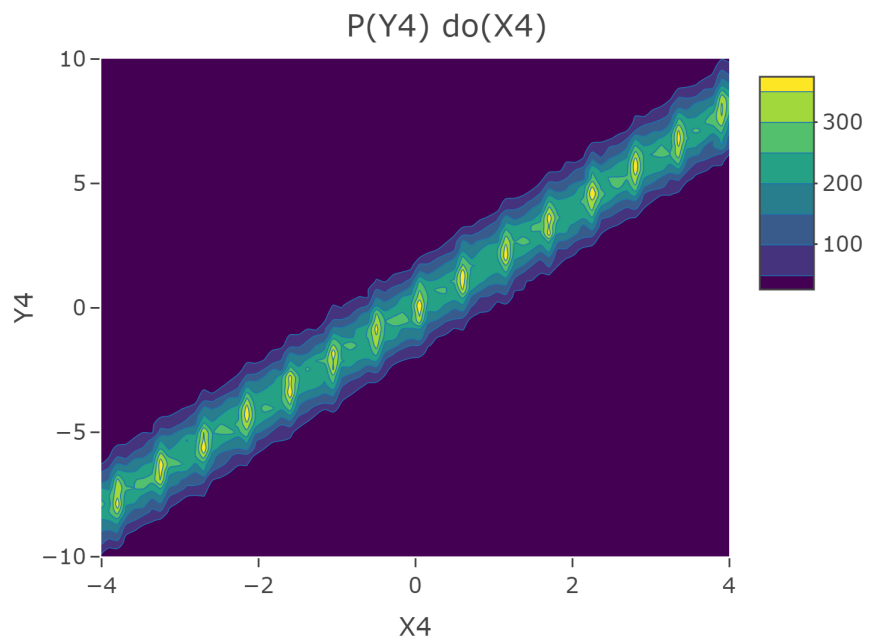


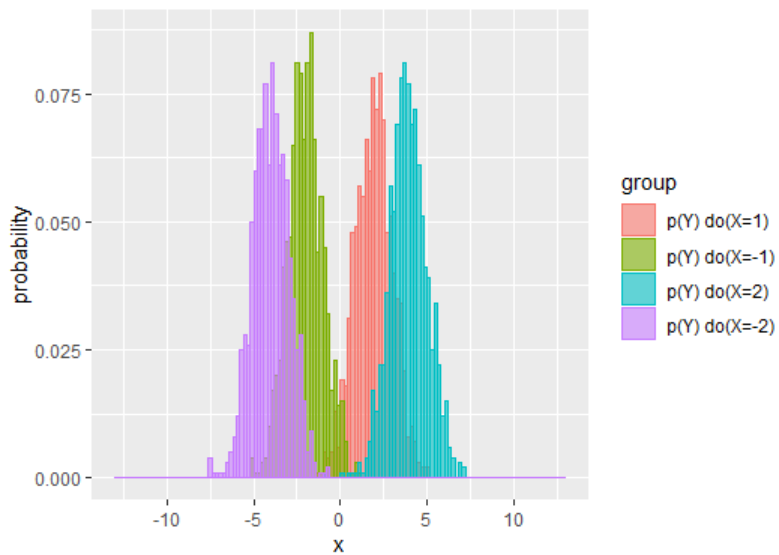


هیستوگرام مربوط به توزیع شرطی Y_4 به ازای X_4 های مختلف

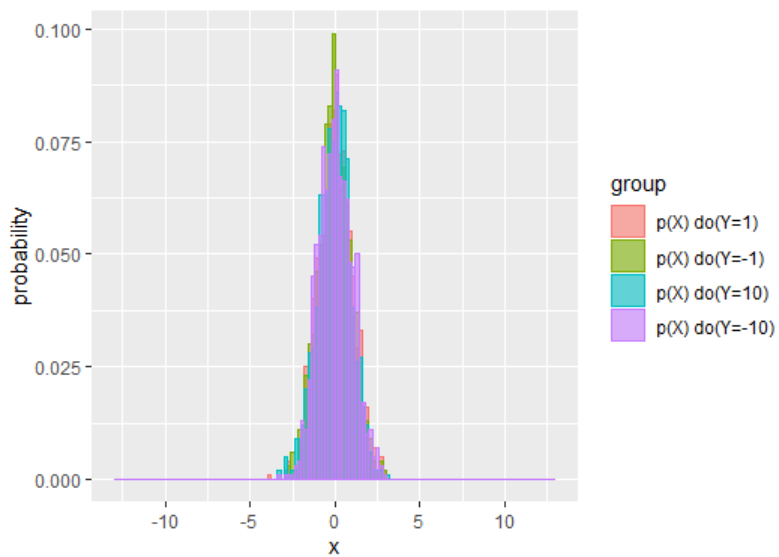


هیستوگرام مربوط به توزیع شرطی X_4 به ازای Y_4 های مختلف





هیستوگرام مربوط به توزیع مداخله‌ای Y_4 به ازای X_4 های مختلف



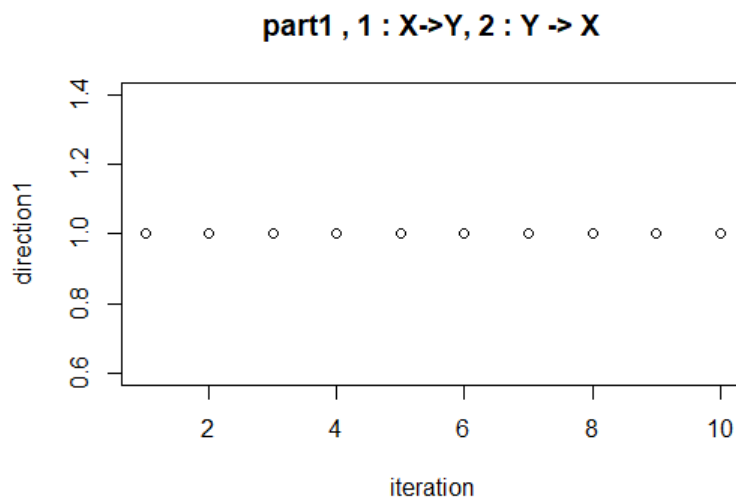
هیستوگرام مربوط به توزیع مداخله‌ای X_4 به ازای Y_4 های مختلف

بررسی تشخیص پذیری : در اینجا بسیار به تقارن گausی نزدیک شده‌ایم و با توجه به بحث های قبلی و وجود خط مناسب در هر دو جهت به طوری که علت و نویز از هم مستقل شوند می‌توان گفت این مدل تشخیص پذیر نیست.

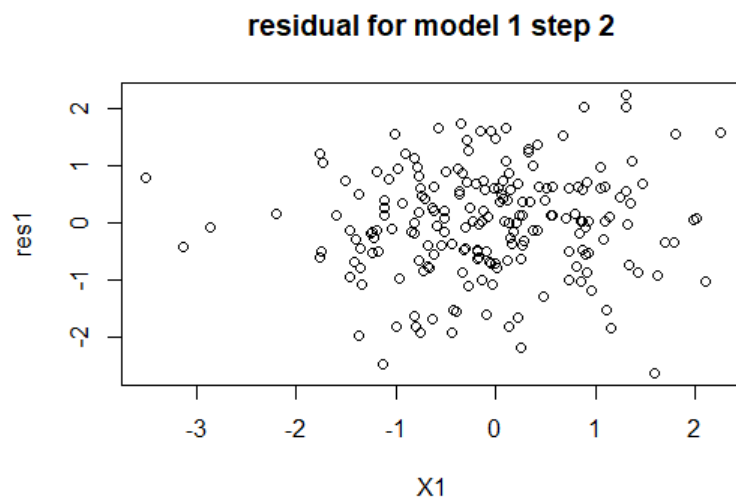
۲ اولین استنتاج علی

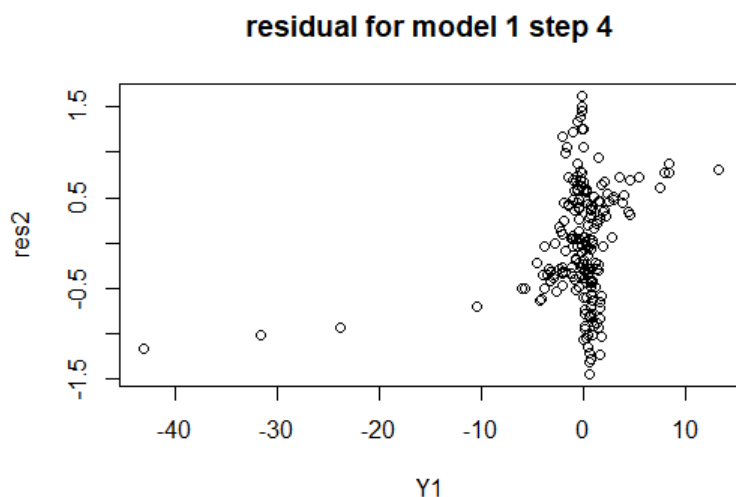
این روش در صورتی درست عمل می‌کند که در هر دو جهت نتوان رابطه‌ی علی را نوشت؛ اگر در هر دو جهت بتوان رابطه‌ی علی را نوشت که علت و نویز معلول مستقل باشند دیگر جواب این روش اعتباری ندارد (می‌توان گفت تصادفی است.) و جهت علی قابل تشخیص نیست. مراحل هم در واقع بررسی همین موضوع است و توضیح دهنده مرحله‌ای را که با سطح اطمینان بالاتری باقیمانده از آن مستقل است؛ به عنوان علت و دیگری را به عنوان معلول معرفی می‌کند. برای مدل‌های بخش قبل نتیجه به صورت زیر است:

برای مدل ۱ در هر دو تکرار جهت به درستی $X \rightarrow Y$ تشخیص داده شده است که با توجه به بخش قبل نیز قابل پیش بینی بود چون وجود رابطه غیر خطی در پیدا کردن جهت علیت کمک می‌کند.

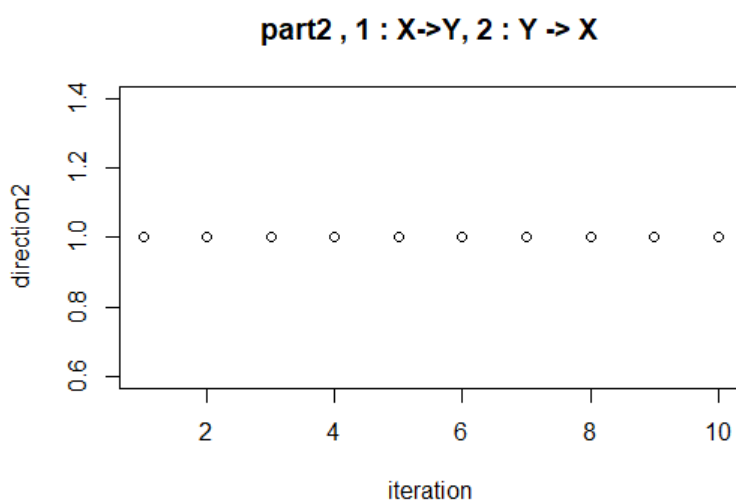


برای شهود بهتر باقی‌مانده‌ها را رسم می‌کنیم تا میزان استقلال باقی‌مانده از توضیح دهنده را در دو حالت درک کنیم.

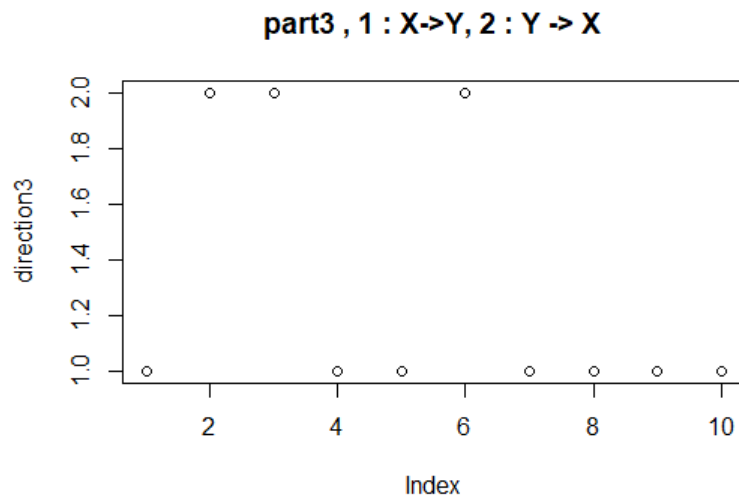




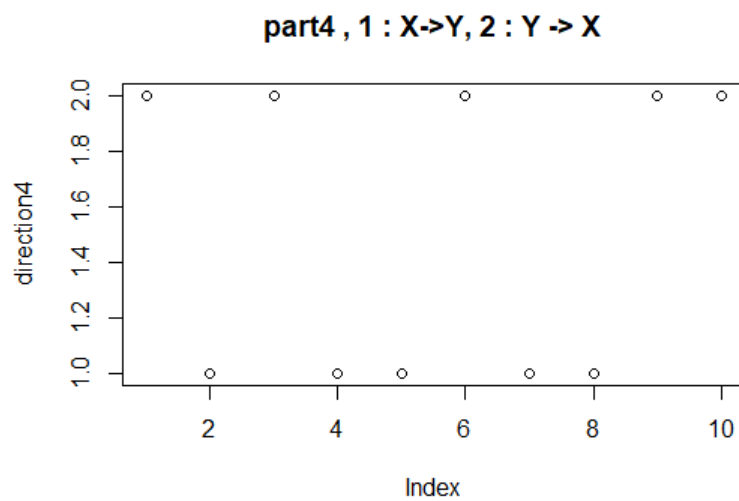
برای مدل ۲ در هر ده تکرار جهت به درستی $X \rightarrow Y$ تشخیص داده شده است که با توجه به بخش قبل و دور بودن توزیع t با پارامتر ۱ از نرمال نیز قابل پیش بینی بود.



برای مدل ۳ در ۷ از ۱۰ تکرار جهت به درستی $X \rightarrow Y$ تشخیص داده شده است که با توجه به افزایش پارامتر توزیع t و نزدیک تر شدن به حالت نرمال و به تبع آن وجود مدل علی در هر دو حالت منطقی است.

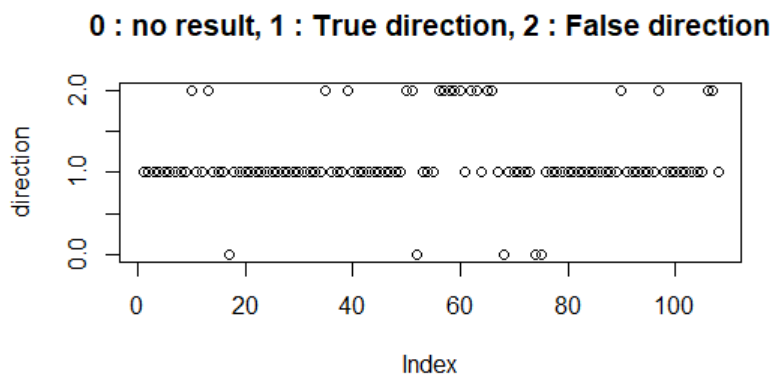


برای مدل ۴ در ۵ از ۱۰ تکرار جهت به درستی $X \rightarrow Y$ تشخیص داده شده است که در ادامه روند قبلی و میل کردن توزیع t به توزیع گاوسی قابل توجه است.

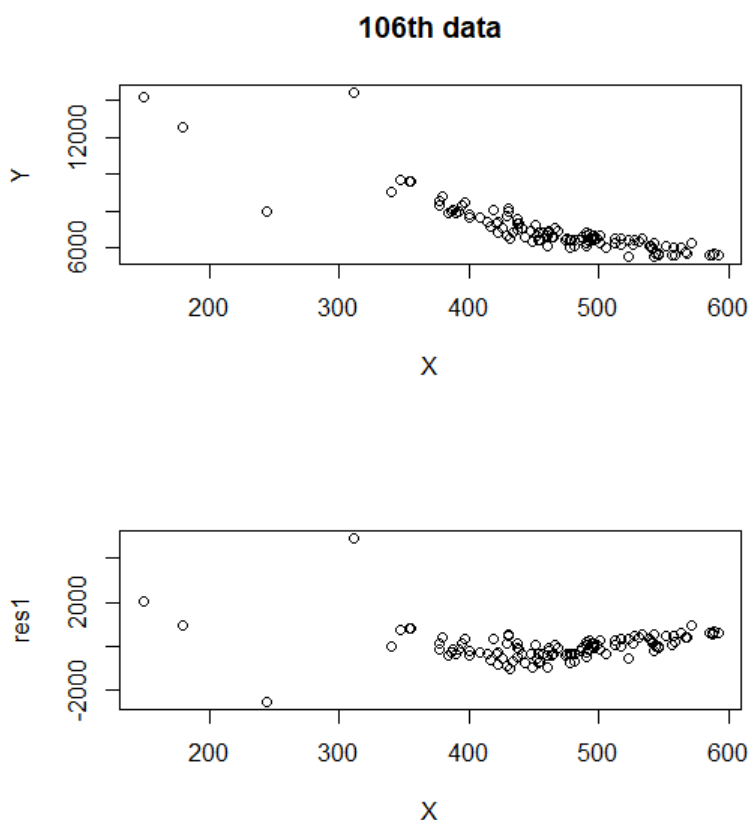


۱.۲ داده واقعی

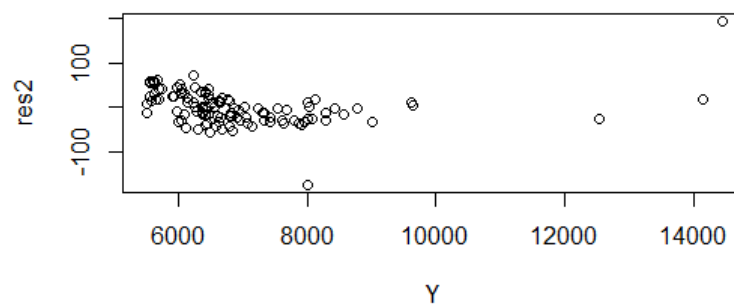
ابتدا فایل metadata را لود می‌کنیم و سپس با توجه به آن داده علت و معلول را لود می‌کنیم سپس با همان الگوریتم بخش قبل جهت علی را پیدا می‌کنیم و در برداری ذخیره می‌کنیم. عدد ۲ نشان دهنده غلط بودن جهت پیدا شده و یک نشان دهنده درستی روش است. در داده های ۱۷، ۶۸، ۷۴، ۷۵ به دلیل اورفلو کردن تابع مربوط به تست استقلال نتیجه‌ای به دست نیامد. داده های ۵۲، ۵۳، ۵۴، ۵۵، ۷۱ و ۱۰۵ نیز دوبعدی نیستند که با توجه به میانگین گرفت از اجرای الگوریتم روی دو به دوی متغیرها جهت علیت را می‌یابیم. که نتایج در نمودار زیر آمده است :



یک مورد خطا برای داده ۱۰۶ ام است که نمودار داده و باقی مانده ها در دو جهت و pvalue های آن برابر زیر است:




```
> p1  
[1] 0.00990099  
> p2  
[1] 0.03960396
```



با توجه به بالا میتوان گفت علت وجود داده های پرت و یا گاوسی بودن داده ها باشد.