Binary Renyi Correlation

Information Theory Course project

Dror Drach, Co-Advised by Dr. Ofer Shayevitz and Dr. Or Ordentlich

# Background

The Hirschfeld-Gebelein-Renyi maximal correlation is a well-known measure of dependence between a pair of random variables, defined as:

This measure satisfies the tensorization property, meaning: for n i.i.d pairs each, we have . This was proven by Witsenhausen [1]

We are interested in the maximal-correlation under the additional constraint where the output size for f(x), g(y) is limited to a finite number of bits (the “1-bit-maximal-correlation” )

It is not immediately clear if the tensorization property still holds for , and if not – how does it behaves in n and with respect to different P(X,Y)

# Problem Statement

**Definition:** *for a random pair (X,Y), the Binary Maximal correlation is defined as:*

**Does the binary maximal correlation tensorize?**

* If so, how to prove it
* If not:
  + Provide a counter example
  + How does it behave with n?

# Our Contribution

We have done the following, each will be explained in detail in this chapter:

1. Implemented the general Maximal-correlation for a pair of random variables with an arbitrary discrete joint probability distribution, using SVD [4]
2. Implemented the 1-bit correlation for one and two pairs of r.v’s with discrete distributions
3. Simulated the relationships between the 1-bit-correlations of n=1 vs n=2 i.i.d pairs for different distributions of Pxy,
4. **This way, we found distributions for which the 1-bit-correlation improves with n, thus negating the tensorization conjecture for the binary case**
5. We Compared of an arbitrary pair with the gaussian case, and found a distribution P(X,Y) that can be improved by applying the central-limit-theorem (i.e. by an “educated guess” of f(x),g(y), and under )
6. We also ruled out tensorization of a “1-sided” binary correlation, which is poses compression constraints only on one of X or Y
7. During our work, we found a research note [7] on , with a supposed proof for tensorization. It is our understanding that the paper contains an error, which we will explain here.

## Implementation of the Maximal-correlation using SVD

Kumar [4] provided an easy way for computing when the r.v’s X and Y are discrete:

1. Let be the discrete joint distribution of (X,Y) with the marginals
2. Define
3. T
   * is the second largest singular value of Q
   * u,v are the resp. (left and right) singular vectors
   * With f(x), g(y) satisfying the required constraints ( )

This formulation also provides an easy and intuitive proof for the tensorization of for discrete random variables

We bring the following essence of this proof for the case of i.i.d pairs. For the complete proof under the general assumption of independence (without identity) see [4]

* Let and be 2 i.i.d random pairs. Each with (discrete) joint distribution of

*)*

* Then:

,

* Where is the Kronecker product between the distributions
* and:
* The singular values:
* Recalling that the second largest singular value for 2 i.i.d pairs satisfies

)

## Implementation of the 1-bit correlation:

**Algorithm description:**

* Select to quantize x, y to
  + The pair are jointly distributed with some .
* Find of
* Repeat for all possible quantizing functions , and select that for which is the highest.
  + **Note that are only intermediate mapping functions to a new - and legitimate - distribution:**
    1. Their norms and expectations are irrelevant with respect to the constraints for
    2. The fact we use the proper SVD solution for ensures that the norms (and means) of the total (=quantization + svd) are properly 1 (and 0) as required
  + Cost is (m,n are the cardinality of x, y resp.)
  + Calculating N pairs iruns…

**Implementation in detail:**

* Given a random pair (X,Y) with joint distribution:
* And under some quantizing functions:

,

,

* Create a quantized pair with the distribution:
* Create a function book for all possible Valid Mappings (which are excluding all-ones and all-zeros)

## :

1. We simulated the relationships between the 1-bit-correlations of n=1 vs n=2 i.i.d pairs for different distributions of Pxy
2. Recall that . **This is the computational bottleneck for this simulation**, due to which we restricted ourselves to alphabet sizes (1-sample) of 3x3 (=9x9 for 2 samples)
3. Tests were made to ensure that numerical precision errors does not generate false positives

## Counter example: 2-pairs

* **We found a 3x3 distribution for which *>***
* The numerical values are described below
* **Note: for a proper reproduction of the results, use the attached .mat file that contains the distributions with high precision: copying-pasting the printed values below might not properly reproduce the counter example**

## Counter example: Educated guess (CLT):

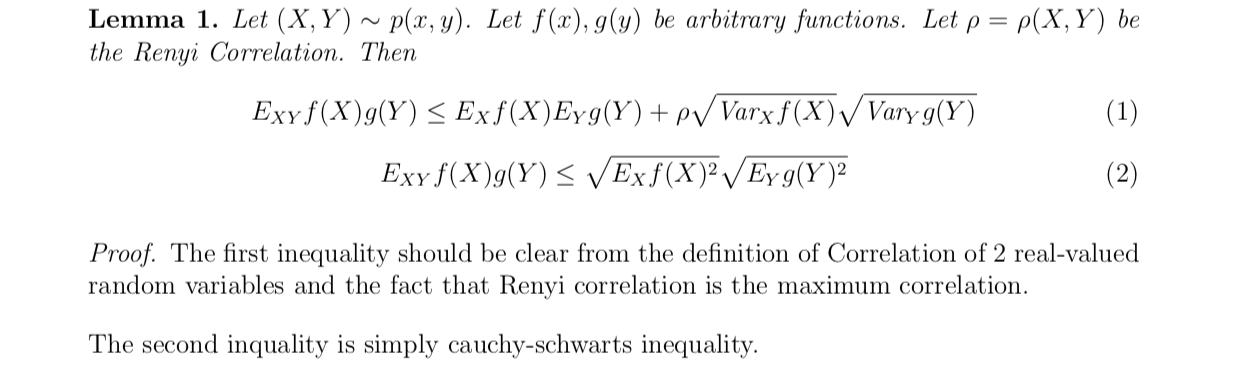
* Consider applying the central-limit theorem for each side of pairs of r.v’s (x,y) and taking the sign() from each side
* This gives us a pair of a jointly gaussian r.v under the sign() function, for which we already have [4] a closed form of the binary correlation
* This means that taking n to infinity, we can always achieve of at least
* **To summarize:**
  + consider the measure
* **This way, we compared with for various distributions, and found a distribution P(X,Y) for which , and therefore *>***

## One sided form:

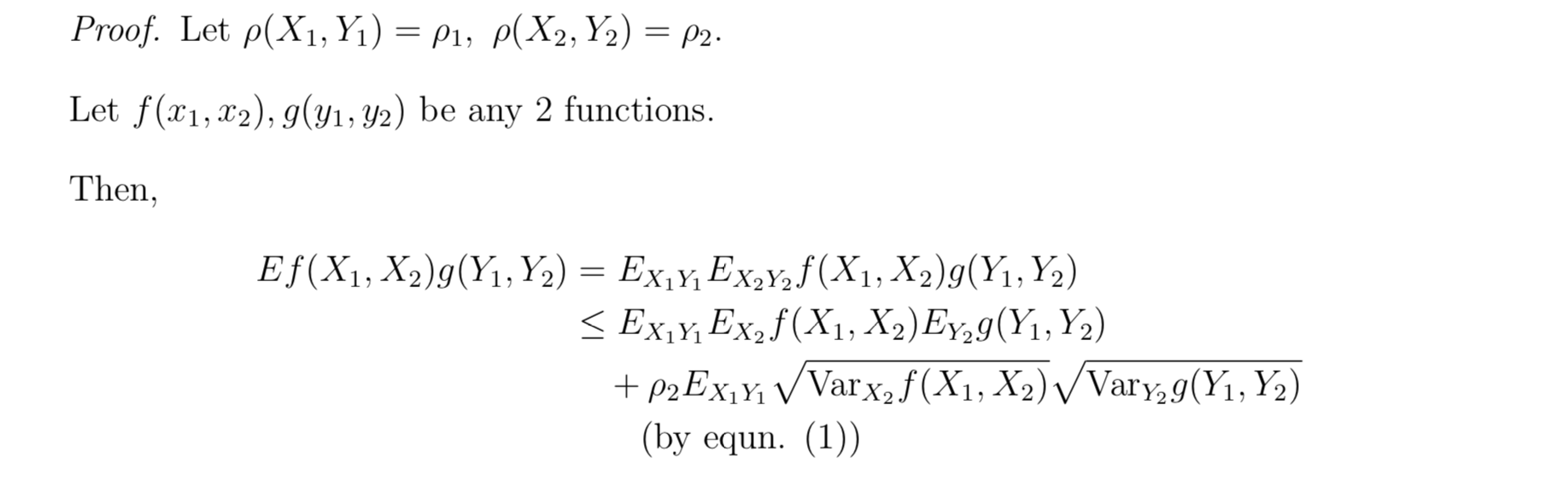
* Having ruled out the tensorization of the 1-bit correlation, we investigated a “softer” version of the measure, in which only one of the functions f,g is restricted to have binary-output
* We then found that this formulation **also doesn’t tensorize, with the same counter example for the 2-pairs** being also valid as a counter example for this case

# Kumar’s tensorization proof and its issue [6]

# The first part of the paper (see [6]) proves tensorization of for the general case. This proof is correct and builds upon a simple lemma:



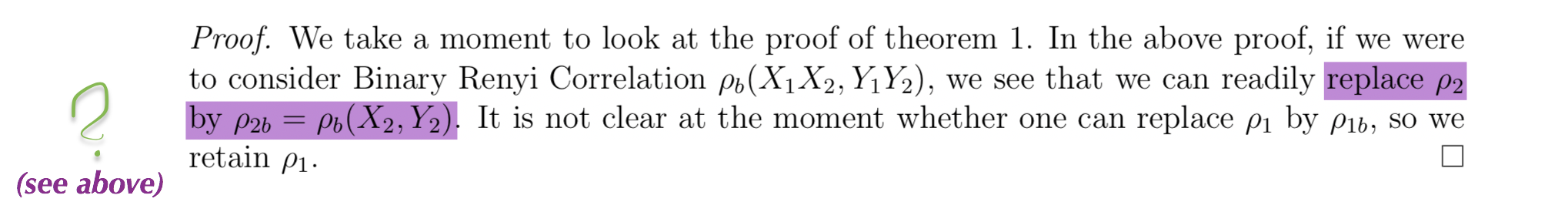
* In the proof, the lemma is used **under conditional expectation**:



# (The correct synthax should be

# The Second part of the paper deals with the binary form , and states the following lemma:

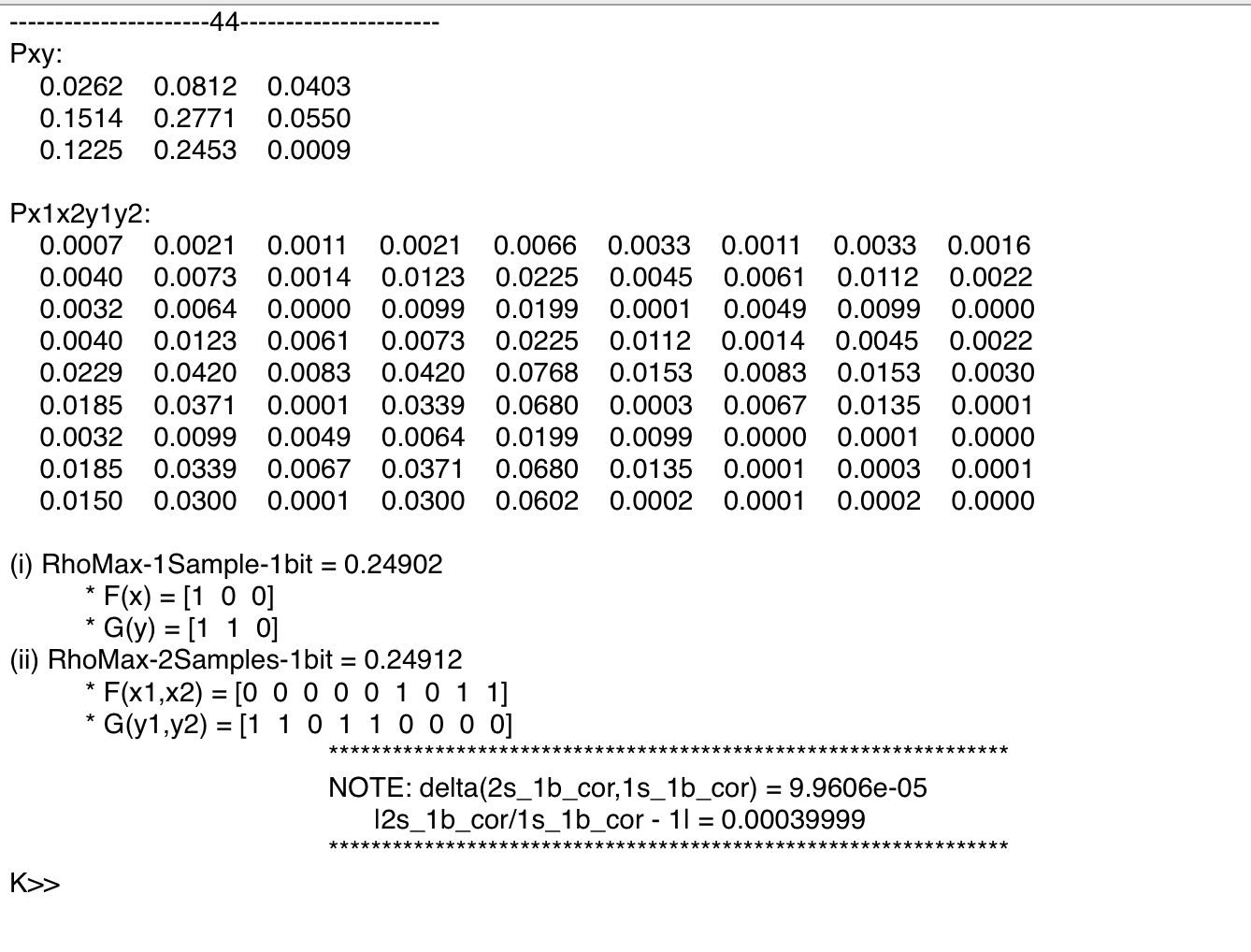
# 

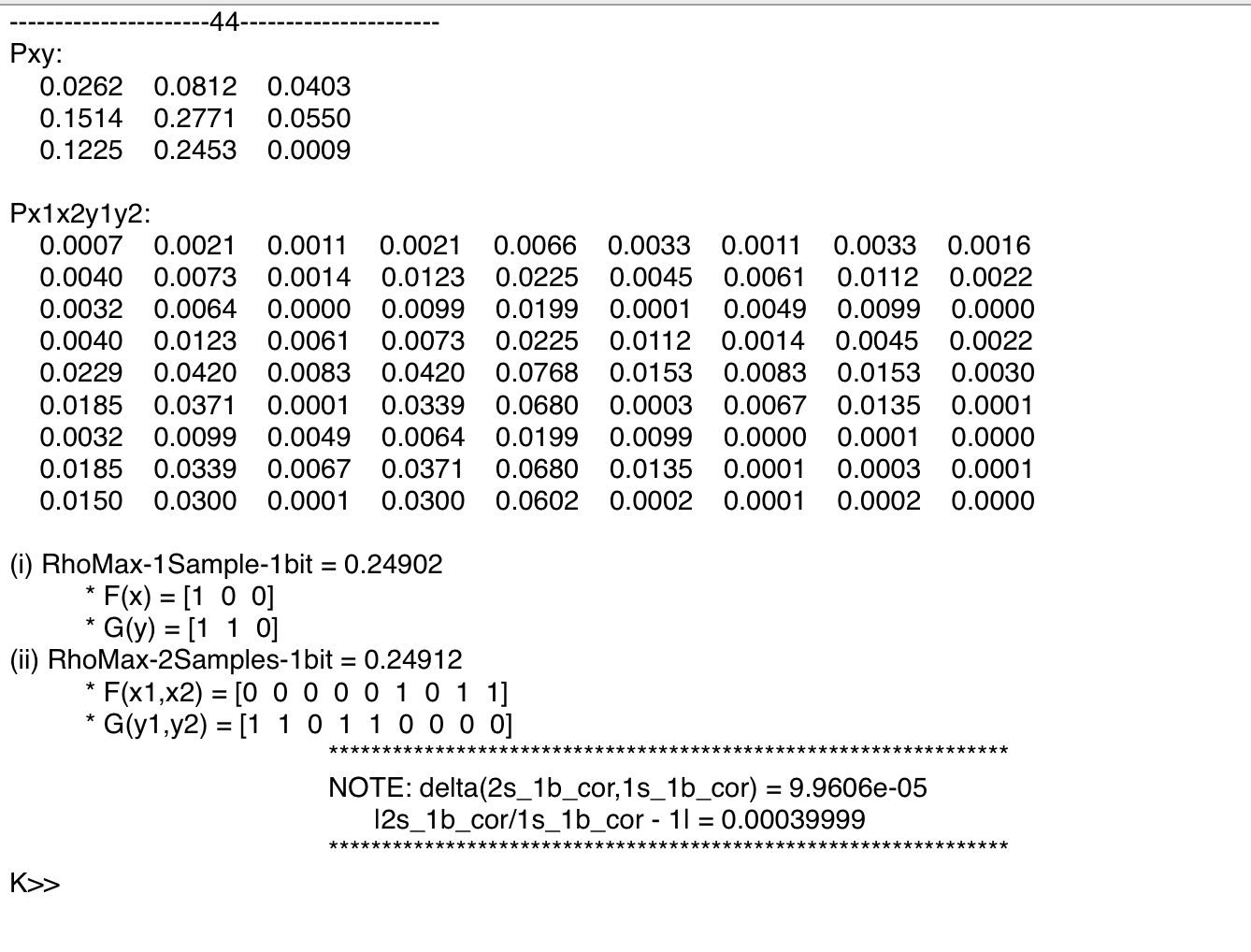


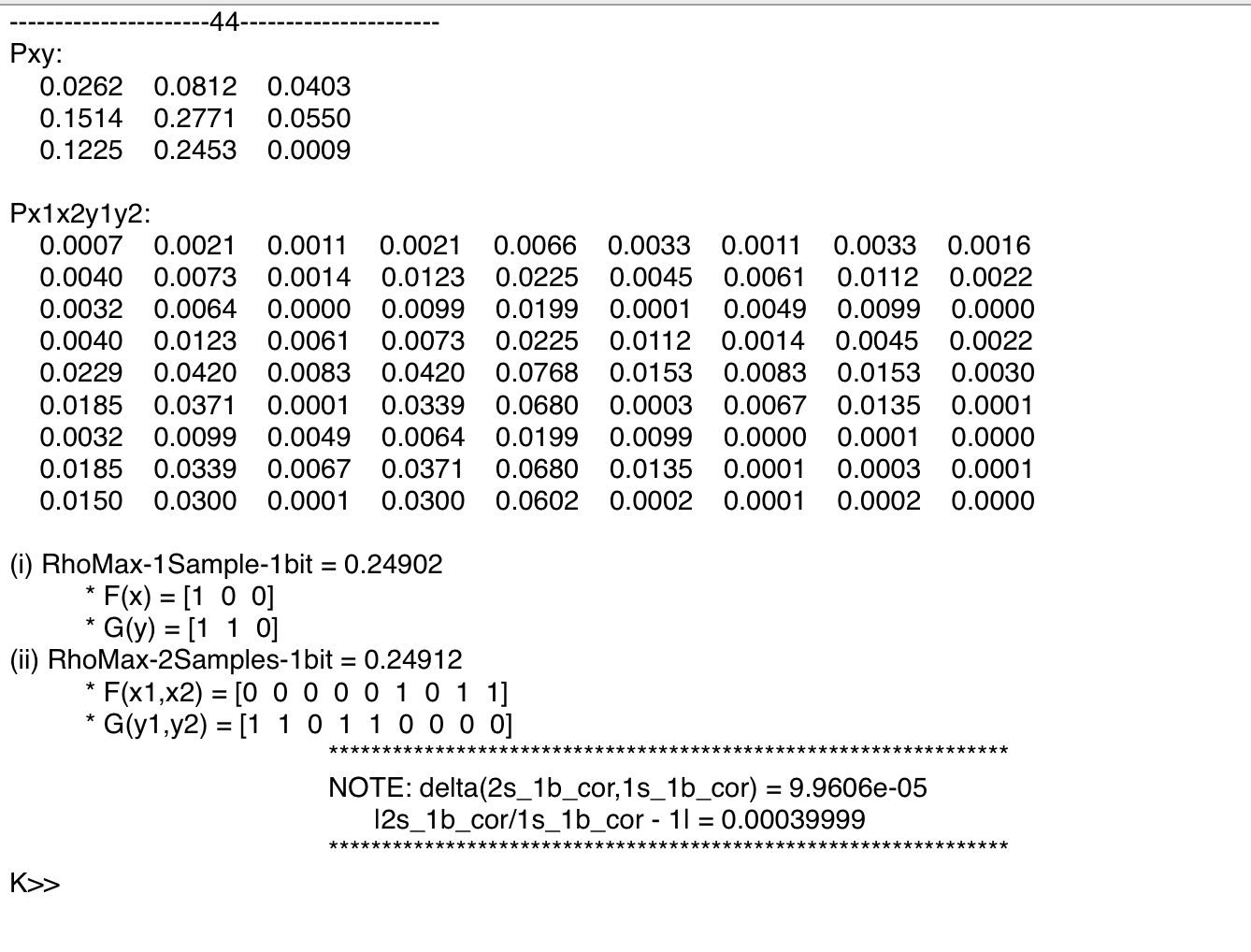
* The new claim essentially is:
* **This inequality doesn’t hold under the conditional expectation and a binary constraint.**

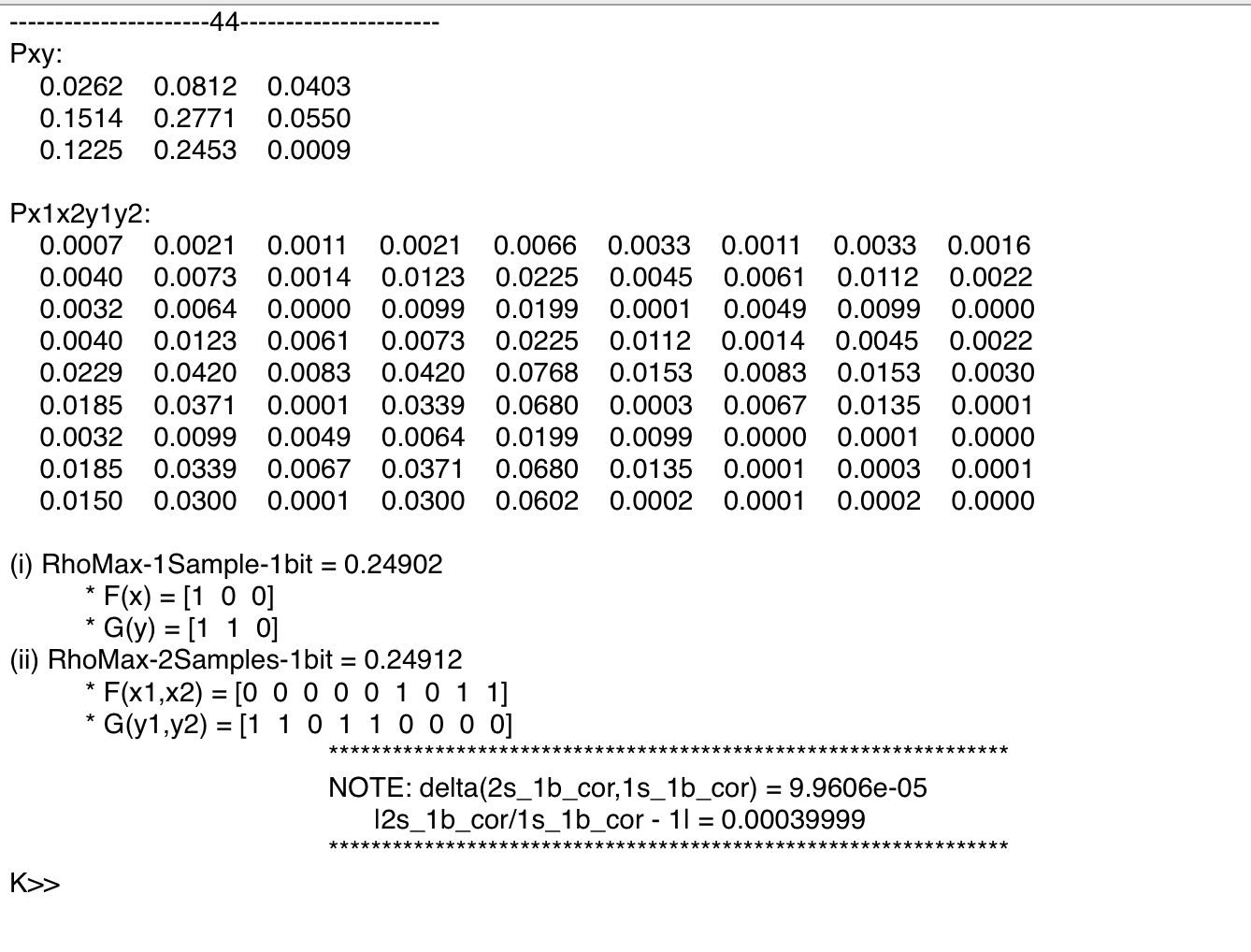
# Numerical Results

## Counter example: two pairs

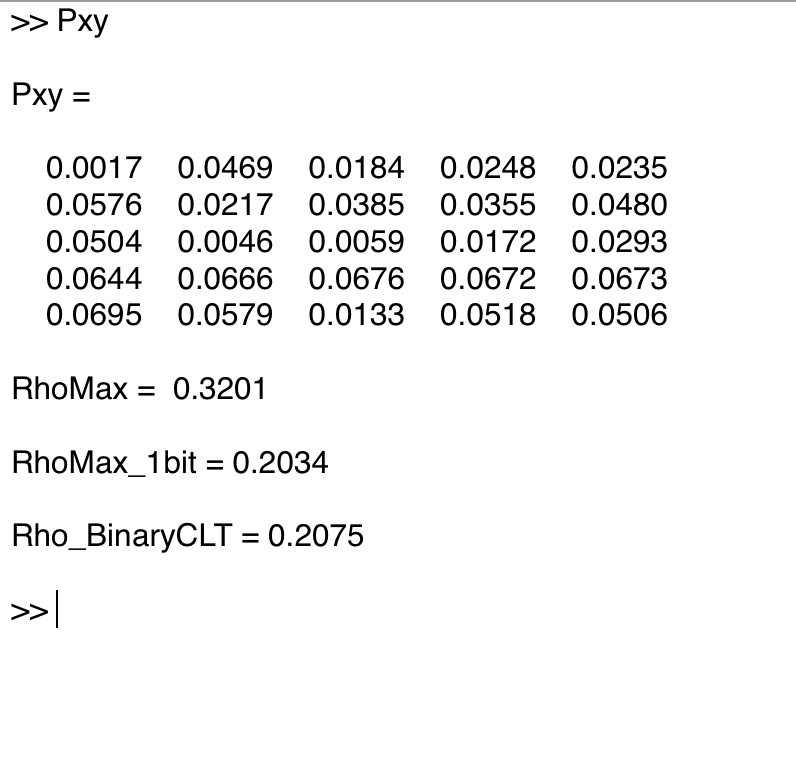








## Counter example: Educated guess (CLT):



## What’s left to find out:

* No immediate/simple connection could be found between the svd’s of the quantized optimals vs the original optimals
* No special properties were observed with the specific counter examples that we found, apart for a possible pattern in the marginals, which could not be linked to a formulation of a sufficient/necessary condition for non-tensorization (Ideas?)
* We are currently working of bounding the behavior of in high dimensions

# Summary

1. We simulated the Hirschfeld-Gebelein-Renyi maximal correlation for discrete, arbitrary distributions, using SVD
2. We simulated the maximal correlation under entropy constraint, for discrete distributions and 1 and 2 samples
3. We provided two distinct examples for which the tensorization property of the binary-maximal-correlation doesn’t hold:
   1. An arbitrary example drawn at random, for which the increase in is readily observed when considering 1 pair vs 2 pairs
   2. An “educated guess” example, where an increase in is observed comparing 1 pair vs pairs (the specific guess of the function allowed us to compute directly (as this cannot be done numerically)
4. We found a manuscript claiming the binary maximal correlation tensorizes, but the proof has an issue which we believe makes the proof invalid. We currently formalize our understanding on this issue

# References:

[1] H.S. Witsenhausen, *On sequences of pairs of dependent random variables*

[2] Anantharam et al., *On Maximal Correlation, Hypercontractivity, and the Data Processing Inequality studied by Erkip and Cover*

[3] Yin and park, *Hypercontractivity, Maximal Correlation and Non-interactive Simulation*

[4] G. Kumar, *On Sequences of Pairs of Dependent Random Variables A simpler proof of the main result using SVD*

[5] C. Borell. Tail probabilities in Gauss space. Vector Space Measures and Applications, Dublin 1977. Lecture Notes in Math. 644, 71–82 (1978). Springer-Verlag.

[6] G. Kumar, Binary Renyi Correlation A simpler proof of Witsenhausen’s result and a tighter upper bound