Generation of random correlation matrices for the missed correlation problem

August 2020

1 Introduction

Suppose that we have a partial defined symmetric matrix $A(x) \in \mathbb{S}^n$ depending on the vector of m missed entries $x \in \mathbb{R}^m$. Suppose that diagonal elements are fixed to one. Denote by F_A the set $\{x \in \mathbb{R}^m : A(x) \in \mathbb{S}^n_+\}$ of all feasible solutions for the missed correlation problem. We are interested on finding a procedure to draw a random uniform sample of points from F_A when such set is non-empty.

2 Some theoretical results

Remark. Recall that the set F_A is a compact and convex subset of $\mathbb{R}^m$, since it is the feasible set of a semidefinite program. It's boundary consists of all x such that $A(x)$ is a psd matrix with determinant zero, and it's relative interior consists of all x such that $A(x)$ is pd. We know that the x that maximizes the determinant of $A(x)$ is unique when the relative interior of F_A is non-empty (there exists at least one solution that is a pd matrix).

Definition 2.1. The standard n-dimensional measure λ_n on $\mathbb{R}^n$ (Lebesgue measure, in honor of Henri Lebesgue) is given by

$$\lambda_n(A) = \int_A 1dx \tag{1}$$

for any $A \subseteq \mathbb{R}^n$.

Remark. Technically speaking, Lebesgue measure is more general than this integral, but for this discussion this simpler definition will suffice. Note that $\lambda_1(A)$ represents the length of $A \subset \mathbb{R}$, $\lambda_2(A)$ is the area of $A \subset \mathbb{R}^2$, $\lambda_3(A)$ is the volume of $A \subset \mathbb{R}^3$, and so on.

Definition 2.2. If $S \subseteq \mathbb{R}^n$ and $\lambda_n(S)$ is positive and non-infinite, then:

$$\mathbb{P}(A) = \frac{\lambda_n(A)}{\lambda_n(S)}, A \subseteq S$$

is the continuous uniform distribution on S.

A continuous uniform distribution over a set $S \subseteq \mathbb{R}^n$ is also uniquely defined by its probability density function, which takes the value $f(x) = \frac{1}{\lambda_n(S)}$ for all $x \in S$ and 0 otherwise. As well, we'll define $f(x \in S) := \int_S f(z)dz$

We can observe that this distribution cannot be defined when the Lebesgue measure of S is 0, and therefore we cannot random sample uniformly from F_A if $\lambda_m(F_A) = 0$.

3 Sampling algorithms

3.1 Rejection Sampling

One of the most efficient methods to solve the problem when n is relatively small is to use rejection sampling. This sampling method can be summarized in the following pseudo-code:

1. Draw a random point from a uniform distribution over the set $[-1, 1]^m$
2. Keep the point if it is a member of F_A (i.e. check if $A(x)$ is a correlation matrix). Otherwise discard it.
3. Repeat 1 and 2 until obtaining the desired sample.

We can now state the next theorem

Theorem 1. The distribution of a random variable obtained using the rejection sampling algorithm is continuous uniform over the feasible set F_A

Proof. Denote by Y the random variable obtained from above procedure. Let $f_X(x) = \frac{1}{2^m}$ if $x \in [-1, 1]^m$ and 0 otherwise, denote the probability density function of a random variable X with continuous uniform distribution over the set $[-1, 1]^m$. By definition of conditional probability, the density of the random variable Y is

$$f_Y(x) = f_X(x \mid x \in F_A) = \frac{f_X(x)}{f_X(x \in F_A)} = \frac{\frac{1}{2^m}}{\mathbb{P}(F_A)} = \frac{\frac{1}{2^m}}{\frac{\lambda_m(F_A)}{2^m}} = \frac{1}{\lambda_m(F_A)}$$

when $x \in F_A$ and 0 otherwise[1]. Hence, the random variable Y is continuous uniform over the set F_A. □

The disadvantage of this method is that sampling could be inefficient if $\mathbb{P}(F_A) \approx 0$, since the expected number of iterations needed to obtain a correlation matrix is $\frac{1}{\mathbb{P}(F_A)}$ which could be close to infinity. The main advantage is that uniformity of the resulting joint distribution is guaranteed.

[1]What we are doing is normalizing the density f_X such that it integrates to one over the set F_A

3.1.1 Example

Consider the partial defined matrix

$$A(x, y) = \begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \\ x & y & 1 \end{bmatrix}$$

The feasible set for this example is the unitary disk in R^2. i.e, $F_A = \{(x, y) \mid x^2 + y^2 \leq 1\}$. This set has an area of π ($\lambda_2(F_A) = \pi$), and hence the joint density is $f(x, y) = \frac{1}{\pi}$ for $(x, y) \in F_A$ and 0 otherwise. If we fix x, then $y^2 \leq 1 - x^2$ and therefore $-\sqrt{1 - x^2} \leq y \leq \sqrt{1 - x^2}$; hence, the marginal density for the random variable x is given by:

$$f_x(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1 - x^2}}{\pi}$$

for all $x \in [-1, 1]$ and 0 otherwise. This corresponds to the density of a beta distribution with shape parameters $\alpha = \beta = \frac{3}{2}$ and linear transformed to have support $[-1, 1]$. By symmetry, the marginal distribution for y is of the same form. Note that the expected value of a random vector (x, y) uniform over F_A is $(0, 0)$, which is also the point that maximizes $det(A(x, y))$ over F_A.

3.2 Rejection sampling with Sylvester's criterion

As mentioned above, rejection sampling could be unstable when dealing with a high dimensional problem. To overcome this, we can exploit the fact that $A(x)$ is pd iff every principal leading minor has a positive determinant.

We adapt the idea of rejection sampling with the following pseudocode:

1. set $k = 2$. While $k < n$:

   while the determinant of the $k \times k$ principal leading minor is less or equal 0:

   replace the missed entries in this sub-matrix with random uniform variables from $U(-1, 1)$

   $k = k + 1$

2. Repeat until obtaining the desired number of samples.

This sampling algorithm is more computationally efficient than simple rejection sampling. However, the distribution is not always uniform. Take as an example the following partially-defined matrix

$$A(x, y) = \begin{bmatrix} 1 & x & 0 \\ x & 1 & y \\ 0 & y & 1 \end{bmatrix}$$

Again, the feasible set F_A is the unitary disk. This algorithm would start by drawing x from $U(-1, 1)$ and since $1 - x^2 > 0 \forall x \in (-1, 1)$, there will be no rejections at this point. This implies that the marginal distribution for x will be $U(-1, 1)$, and from previous example, if the joint distribution is uniform over the unitary disk then the marginals should be beta distributed with shape parameters $\alpha = \beta = \frac{3}{2}$. Since the joint distribution uniquely determines the marginals distributions, the algorithm will fail to achieve a uniform distribution on F_A.

3.3 Linear Shrinking

In Statistics, the idea of perturbing one point into another is called *Linear shrinking. If we take as inputs $x_0 \in [-1, 1]^m$ and $x_1 \in F_A$ the linear shrinking method consists of the following steps:*

1. *Find the smallest $\lambda \in [0, 1]$ such that the convex combination $\lambda x_0 + (1 - \lambda)x_1$ belongs to F_A[2].*
2. *Draw a random number u from a distribution on $[0, \lambda]$*
3. *Return the point $ux_0 + (1 - u)x_1$*

The algorithm is well defined as the solution $\lambda = 0$ is always in F_A (as $x_1 \in F_A$)

We can choose the input x_0 to be a random uniform vector on $[-1, 1]^m$ and x_1 to be the solution to the maximum determinant missed correlations problem (i.e, $det(A(x_1))$ is maximum).

The above procedure is faster than rejection sampling, but the distribution of the sample is not guaranteed to be uniform on F_A.

4 When the methods fail

As outlined in definition 3.2, the algorithms based on rejection sampling will fail if the Lebesgue measure of F_A is 0 or ∞.

The linear shrinking algorithm could still work when the Lebesgue measure is 0, but it will output x_1 with probability 1[3].

Let's look at the following partially-defined matrix

$$A(x, y) = \begin{bmatrix} 1 & 1 & x \\ 1 & 1 & y \\ x & y & 1 \end{bmatrix}$$

The determinant is $-(x-y)^2$ which is non-negative if $x = y$. So F_A consists of the segment line that joins $(-1, 1)$ and $(1, 1)$. Since it is a line, it has no area

[2] This can be solved using the bisection method.

[3] A set with 0 Lebesgue measure has no interior, so the resulting line in $\mathbb{R}^m$ will intersect F_A only at x_1 for an infinite number of directions.

and hence $\lambda_2(F_A) = 0$. The probability that a random point from $[-1,1]^m$ lies in F_A is $\frac{\lambda_2(F_A)}{4} = \frac{0}{4} = 0$, so it is impossible that this random sampling method returns a valid point.