Something

Jonas Bruun Hubrechts

Kongens Lyngby 2024

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

The goal of the thesis is to ...

Summary (Danish)

Målet for denne afhandling er at ...

Preface

This thesis was prepared at DTU Compute in fulfilment of the requirements for acquiring an M.Sc. in Engineering.

The thesis deals with ...

The thesis consists of ...

Lyngby, 01-July-2024

Jonas Bruun Hubrechts

Acknowledgements

I would like to thank my....

Contents

Chapter 1

Time to Level of Brownian motion

1.1 Arrivals of batches

Assuming that the in-flow from the previous section in the production obeys the following SDE

$$dS_t = rdt + \sigma dB_t$$

I.e. Brownian motion with drift. And assuming that every time the accumulated mass hits a level l, the batch is ready to be processed by the next step, we wish to first find the distribution for these times. Note that the above model allows for negative flow and thus also negative accumulated mass. However, for $\sigma << r$ this becomes very unlikely as

$$\mathbb{P}\left(S_t \leq 0\right) = \Phi\left(\frac{-r\sqrt{t}}{\sigma}\right)$$

Thus, only for small t this is probable, as otherwise it is dominated by $\frac{r}{\sigma}$ which

is large and thus the probability very low.

Furthermore, if one allows periods without inflow, the running maximum could be a good model. Either way, the probability distribution for between batch times is the same.

To derive the distribution for the between batch times, T, we shall use the Girsanov Theorem as well as the joint distribution of the maximum of a standard Brownian motion and is running maximum. Thus, let B_t be a standard Brownian motion, and M_t the running maximum defined as

$$M_t := \sup_{s \in [0,t]} \{B_s\}$$

- Udledning af joint fordeling mellem M og B
- Change of measure for at opnå med drift og sigma
- Marginal fordeling for M

1.1.1 Joint distribution of Brownian and its running maximum

To derive the joint density of a standard Brownian motion and its running maximum, consider the following probability

$$\mathbb{P}\left(M_t \geq m, B_t \leq w\right)$$

Let T_m be defined as the first time B_t hits the level m, i.e. $T_m := \inf_t \left(B_t = m\right)$. Then $M_t \geq m \iff T_m \leq t$. Thus, the above probability is reexpressed as

$$\mathbb{P}\left(M_t \geq m, B_t \leq w\right) = \mathbb{P}\left(T_m \leq t, B_t \leq w\right)$$

To proceed, we use the principle of reflection which is admissible due to B_t being a martingale. In particular, we define $\tilde{B}_t$ as follows

$$\tilde{B}_t := \begin{cases} B_t & t \leq T_m \\ 2m - B_t & t > T_m \end{cases}$$

It follows that $\tilde{B}_t$ is also a standard Brownian motion. By the definition of $\tilde{B}_t$, we then have that

$$\mathbb{P}\left(T_m \leq t, B_t \leq w\right) = \mathbb{P}\left(T_m \leq t, 2m - w \leq \tilde{B}_t\right)$$

Notice that the original expression is only sensible for $m \geq w$ as $w > m$ is a contradiction to the definition of M_t. Thus, $2m - w \geq m$ hence $\tilde{B}_t \geq 2m - w$ implies that the original Brownian motion B_t has hit the level m and thus $T_m \leq t$. This means that

$$\mathbb{P}\left(T_m \leq t, 2m - w \leq \tilde{B}_t\right) = \mathbb{P}\left(2m - w \leq \tilde{B}_t\right) = 1 - \Phi\left(\frac{2m - w}{\sqrt{t}}\right)$$

Thus, in total we have found that

$$\mathbb{P}\left(M_t \geq m, B_t \leq w\right) = 1 - \Phi\left(\frac{2m - w}{\sqrt{t}}\right)$$

And thus, the joint distribution is obtained by differentiation

$$\begin{aligned} f_{M_t, B_t}(m, w) &= \frac{\partial^2}{\partial m \, \partial w} \mathbb{P}\left(M_t \leq m, B_t \leq w\right) \\ &= \frac{\partial^2}{\partial m \, \partial w} \left(\mathbb{P}\left(B_t \leq w\right) - \mathbb{P}\left(M_t \geq m, B_t \leq w\right)\right) \\ &= \frac{\partial^2}{\partial m \, \partial w} \Phi\left(\frac{2m - w}{\sqrt{t}}\right) \\ &= \frac{2(2m - w)}{t^{3/2}} \phi\left(\frac{2m - w}{\sqrt{t}}\right) \quad , m \leq w, \ m \geq 0 \end{aligned}$$

Note:

Now, define instead $\tilde{B}_t = \sigma B_t$. We then find a similar expression for the joint density of ... and its running maximum. Namely, as

$$\mathbb{P}\left(\tilde{M}_t \geq m, \tilde{B}_t \leq w\right) = \mathbb{P}\left(\sigma M_t \geq m, \sigma B_t \leq w\right)$$

Same formula, but with m and w divided by σ

1.1.2 Joint distribution with drift and arbitrary variance

Let B_t be a standard Brownian motion defined on the probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, define $\tilde{B}_t$ to be a Brownian motion with drift as follows

$$\tilde{B}_t := \tilde{\mu} t + B_t$$

To derive the joint density $f_{\tilde{M}_t, \tilde{B}_t}(m, w)$ on measure $\mathbb{P}$, we use a corollary of the Girsanov theorem. Namely, suppose B_t is Brownian motion under measure $\mathbb{P}$, then there exists a measure $\mathbb{Q}$ such that $\tilde{B}_t = B_t - \langle B, X \rangle_t$ is a Brownian motion (without drift) under this new measure given that X_t is an adapted process. Furthermore, as $\tilde{B}_t$ is a martingale, the Radon-Nikodym derivative is equal to the stochastic exponential $Z_t = \exp\left(X_t - \frac{1}{2}\langle X \rangle_t\right)$.

Now, if X_t is of the form $\int_0^t Y_s \, dB_s$ where $\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{1}{2}\int_0^T Y_s^2 \, ds\right)\right] < \infty$, a special case, the Cameron-Martin-Girsanov implies that $\tilde{B}_t = B_t - \int_0^t Y_s \, ds$ is then a $\mathbb{Q}$ Brownian motion. This can easily be shown when Y_s fulfills Noviko's condition, then Z_t is a martingale and the Girsanov theorem applies as clearly X_t is also adapted to B_t. Then, from the above corollary,

$$\begin{aligned}
\tilde{B}_t &= B_t - \langle B, X \rangle_t \\
&= B_t - \lim_{||P|| \to 0} \sum_i \left(B_{t_{i+1}} - B_{t_i}\right) \left(\int_{t_i}^{t_{i+1}} Y_s \, dB_s\right) \\
&= B_t - \lim_{||P|| \to 0} \sum_i \left(B_{t_{i+1}} - B_{t_i}\right)^2 Y_{t_i^*} \\
&= B_t - \int_0^t Y_s \, ds
\end{aligned}$$

As it has now been shown that there exists as measure $\mathbb{Q}$ under which $\tilde{B}_t$ is a Brownian motion as choosing $Y_s = -\tilde{m}u$ we reproduce the initial definition of $\tilde{B}_t$. To then derive the joint distribution of $\tilde{B}_t$ and its running maximum $\tilde{M}_t$,

we compute the Radon-Nikodym derivative, Z_t, hence given by

$$\begin{aligned}\left.\frac{d\mathbb{Q}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t &= \exp\left(\int_0^t Y_s\, dB_s - \frac{1}{2}\int_0^t Y_s^2\, ds\right)\\ &= \exp\left(-\tilde{\mu}\int_0^t dB_s - \frac{1}{2}\tilde{\mu}^2\int_0^t ds\right)\\ &= \exp\left(-\tilde{\mu}B_t - \frac{1}{2}\tilde{\mu}^2 t\right)\\ &= \exp\left(-\tilde{\mu}\tilde{B}_t + \frac{1}{2}\tilde{\mu}^2 t\right)\end{aligned}$$

With the above derivative, we have that

$$\mathbb{Q}(A) = \int_A Z_t\, d\mathbb{P}$$

And thus also

$$\mathbb{P}(A) = \int_A Z_t^{-1}\, d\mathbb{Q}$$

as $Z_t : X \to (0, \infty)$. It then simply follows that

$$f_{\tilde{M}_t,\tilde{B}_t}(m, w) = \tilde{f}_{\tilde{M}_t,\tilde{B}_t}(m, w)e^{\tilde{\mu}w - \frac{1}{2}\tilde{\mu}^2 t}$$

where $\tilde{f}$ is the probability distribution under measure $\mathbb{Q}$. Hence,

$$f_{\tilde{M}_t,\tilde{B}_t}(m, w) = \frac{2(2m - w)}{t^{3/2}} e^{\tilde{\mu}w - \frac{1}{2}\tilde{\mu}^2 t}\phi\left(\frac{2m - w}{\sqrt{t}}\right)$$

To introduce the standard deviation σ, first define $\tilde{\mu} = \mu/\sigma$ and $\hat{B}_t = \sigma\tilde{B}_t$. Then, $\hat{B}_t$ is also a Brownian with drift, μ, but with variance $\sigma^2 t$. Furthermore, the joint distribution is

$$f_{\hat{M}_t,\hat{B}_t}(m, w) = \frac{2(2m - w)}{\sigma^3 t^{3/2}} e^{\frac{1}{\sigma^2}(\mu w - \frac{1}{2}\mu^2 t)}\phi\left(\frac{2m - w}{\sigma\sqrt{t}}\right)$$

1.1.3 Distribution of maximum of Brownian motion with drift

The distribution of the running maximum $\hat{M}_t$ is given by the marginal of the above, namely

$$f_{\hat{M}_t}(m) = \int_{-\infty}^{m} f_{\hat{M}_t,\hat{B}_t}(m,w)\,dw$$

Integration by parts admits

$$f_{\hat{M}_t}(m) = \frac{2}{\sigma\sqrt{t}}\phi\left(\frac{m-\mu t}{\sigma\sqrt{t}}\right) - \frac{2\mu}{\sigma^2}e^{\frac{2m\mu}{\sigma^2}}\Phi\left(-\frac{m+\mu t}{\sigma\sqrt{t}}\right)$$

1.1.4 Cumulative distribution of maximum

As we shall later need the survival function of $\hat{M}_t$, we first compute the cumulative distribution. Namely

$$\mathbb{P}\left(\hat{M}_t \le m\right) = \int_0^m \int_{-\infty}^{\eta} f_{\hat{M}_t,\hat{B}_t}(\eta,w)\,dw\,d\eta$$

To compute the above, we split the inner integral over the line $w = 0$ in the η, w plane and reformulate

$$\mathbb{P}\left(\hat{M}_t \le m\right) = \underbrace{\int_0^m \int_w^m f_{\hat{M}_t,\hat{B}_t}(\eta,w)\,d\eta\,dw}_{I_1} + \underbrace{\int_{-\infty}^0 \int_0^m f_{\hat{M}_t,\hat{B}_t}(\eta,w)\,d\eta\,dw}_{I_2}$$

The antiderivative of $f_{\hat{M}_t,\hat{B}_t}(m,w)$ w.r.t. m is simple and calculated to be

$$\int f_{\hat{M}_t,\hat{B}_t}(m,w)\,dm = -\frac{1}{\sigma\sqrt{2\pi t}}e^{\frac{1}{\sigma^2}\left(\mu w - \frac{1}{2}\mu^2 t\right)}e^{-\frac{1}{2}\left(\frac{2m-w}{\sigma\sqrt{t}}\right)^2}$$

The first of the above integrals, I_1, is then

$$I_1 = -\frac{1}{\sigma\sqrt{2\pi t}}e^{-\frac{1}{2\sigma^2}\mu^2 t}\int_0^m e^{\frac{\mu w}{\sigma^2}-\frac{1}{2}\left(\frac{2m-w}{\sigma\sqrt{t}}\right)^2} - e^{\frac{\mu w}{\sigma^2}-\frac{1}{2}\left(\frac{w}{\sigma\sqrt{t}}\right)^2}\,dw$$

And similar for the second integral I_2

$$I_2 = -\frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2}\mu^2 t} \int_{-\infty}^{0} e^{\frac{\mu w}{\sigma^2} - \frac{1}{2}\left(\frac{2m-w}{\sigma\sqrt{t}}\right)^2} - e^{\frac{\mu w}{\sigma^2} - \frac{1}{2}\left(\frac{w}{\sigma\sqrt{t}}\right)^2} \, dw$$

It is observed that the integrands are the same, thus

$$\mathbb{P}\left(\hat{M}_t \leq m\right) = -\frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{1}{2\sigma^2}\mu^2 t} \int_{-\infty}^{m} e^{\frac{\mu w}{\sigma^2} - \frac{1}{2}\left(\frac{2m-w}{\sigma\sqrt{t}}\right)^2} - e^{\frac{\mu w}{\sigma^2} - \frac{1}{2}\left(\frac{w}{\sigma\sqrt{t}}\right)^2} \, dw$$

From simple substitution, and a few calculations, on gets that

$$\mathbb{P}\left(\hat{M}_t \leq m\right) = \Phi\left(\frac{m-\mu t}{\sigma\sqrt{t}}\right) - e^{\frac{2m\mu}{\sigma^2}} \Phi\left(-\frac{m+\mu t}{\sigma\sqrt{t}}\right)$$

1.1.5 Distribution of time to level

As $\mathbb{P}(M_t \geq l) = \mathbb{P}(T_l \leq t)$. It thus follows that $f_{T_l}(t) = \frac{d}{dt}\mathbb{P}(M_t \geq l)$ which is easily calculated from the above. Namely

$$\begin{aligned} f_{T_l}(t) &= \frac{d}{dt}\left(1 - \mathbb{P}(M_t \leq l)\right) \\ &= -\frac{d}{dt}\left(\Phi\left(\frac{l-\mu t}{\sigma\sqrt{t}}\right) - e^{\frac{2l\mu}{\sigma^2}} \Phi\left(-\frac{l+\mu t}{\sigma\sqrt{t}}\right)\right) \\ &= \frac{\mu t + l}{2\sigma t^{3/2}} \phi\left(\frac{l-\mu t}{\sigma\sqrt{t}}\right) + \frac{l-\mu t}{2\sigma t^{3/2}} e^{\frac{2\mu l}{\sigma^2}} \phi\left(-\frac{\mu t + l}{\sigma\sqrt{t}}\right) \end{aligned}$$

Note that although the distribution above is parameterized by 3 parameters, it can be completely specified by $\tilde{\mu} = \mu/\sigma$ and $\tilde{l} = l/\sigma$ which is clear also from the following

Let $Z_t = \mu t + \sigma B_t$ and similarly $\tilde{Z}_t = Z_t/\sigma = \tilde{\mu} + B_t$. Then $\mathbb{P}(T_l \leq t) = \mathbb{P}(M_t \geq l) = \mathbb{P}\left(\tilde{M}_t \geq \tilde{l}\right) = \mathbb{P}\left(\tilde{T}_{\tilde{l}} \leq t\right)$ where $\tilde{M}_t$ and $\tilde{T}_{\tilde{l}}$ are the running maximum and time to level of $\tilde{Z}_t$. Thus, equivalent to a probability of non-scaled Brownian motion with drift.

To verify the above probability distribution, a Monte-Carlo simulation is carried out for 100.000 simulations with parameters $l = 10$, $\mu = 0.1$, $\sigma = 0.5$. As the shape resembles a gamma distribution, a simple fit, matching the mean and variance is also plotted. Although the gamma family of probability distributions is also a two-parameter family, they do not quite overlap as can be seen in the following plot.

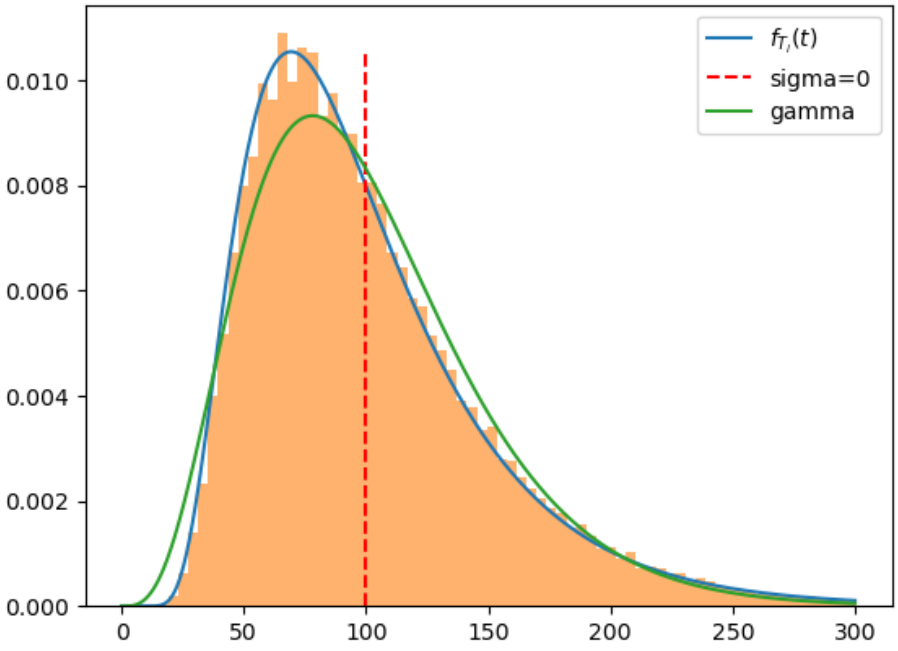


Figure 1.1: Example of simulation and actual distribution. The marked $\sigma = 0$ shows the limit as $\sigma \to 0$ corresponding to no noise on the input flow

1.1.6 MGF

$$\begin{aligned}\mathbb{E}\left[e^{\theta T_l}\right] &= \int_0^\infty e^{\theta t} f_{T_l}(t)\, dt \\ &= \underbrace{\int_0^\infty e^{\theta t} \frac{\mu t + l}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{l-\mu t}{\sigma\sqrt{t}}\right)^2} dt}_{I_1} + e^{\frac{2\mu l}{\sigma^2}} \underbrace{\int_0^\infty e^{\theta t} \frac{l - \mu t}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(-\frac{\mu t + l}{\sigma\sqrt{t}}\right)^2} dt}_{I_2}\end{aligned}$$

We shall only consider the first integral I_1 as the second follows directly from

the result of the first by substituting μ with $-\mu$

$$\begin{aligned}
I_1 &= \int_0^\infty \frac{\mu t + l}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(l-\mu t)^2 - 2\theta\sigma^2 t^2}{\sigma^2 t}}\, dt \\
&= \int_0^\infty \frac{\mu t + l}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(\mu^2 - 2\theta\sigma^2)t^2 - 2l\mu t + l^2}{\sigma^2 t}}\, dt \\
&= e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \int_0^\infty \frac{\mu t + l}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sqrt{\mu^2 - 2\theta\sigma^2}t - l}{\sigma\sqrt{t}}\right)^2}\, dt \\
&= e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \int_0^\infty \left(\frac{\sqrt{\mu^2 - 2\theta\sigma^2}t + l}{2\sigma t^{3/2}} + \frac{\mu - \sqrt{\mu^2 - 2\theta\sigma^2}}{2\sigma\sqrt{t}} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sqrt{\mu^2 - 2\theta\sigma^2}t - l}{\sigma\sqrt{t}}\right)^2}\, dt
\end{aligned}$$

Once again, we split the integral, now as follows

$$I_1 = e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \left(\underbrace{\int_0^\infty \frac{\sqrt{\mu^2 - 2\theta\sigma^2}t + l}{2\sigma t^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sqrt{\mu^2 - 2\theta\sigma^2}t - l}{\sigma\sqrt{t}}\right)^2} dt}_{I_{11}} + \frac{\mu - \sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma} \underbrace{\int_0^\infty \frac{1}{2\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\sqrt{\mu^2 - 2\theta\sigma^2}t - l}{\sigma\sqrt{t}}\right)^2} dt}_{I_{12}} \right)$$

For the first integral, the substitution $u = \frac{\sqrt{\mu^2 - 2\theta\sigma^2}t - l}{\sigma\sqrt{t}}$ reveals that

$$I_{11} = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du = 1$$

As for the second integral I_{12} it can be rewritten as

$$I_{12} = e^{\frac{l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \int_0^\infty \frac{1}{2\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{\mu^2}{2\sigma^2} - \theta\right)t - \frac{l^2}{2\sigma^2}t^{-1}}\, dt$$

Then, substituting $u = \sqrt{\frac{\mu^2}{2\sigma^2} - \theta}\sqrt{t}$

$$I_{12} = e^{\frac{l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{\mu^2}{2\sigma^2} - \theta}} \int_0^\infty e^{-u^2 - \frac{l^2}{2\sigma^2}\left(\frac{\mu^2}{2\sigma^2} - \theta\right)u^{-2}}\, du$$

To solve the above integral, we consider the following family of integrals, parameterized by s

$$I(s) = \int_0^\infty e^{-u^2 - s^2 u^{-2}}\, du$$

It follows that

$$I'(s) = -2\int_0^\infty e^{-u^2 - s^2 u^{-2}} s u^{-2}\, du$$

letting $z = s/u$, it follows that $dz = -s/u^2 du$ and (assuming $s > 0$)

$$I'(s) = -2\int_0^\infty e^{-s^2 z^{-2} - z^2}\, dz = -2I(s)$$

Also,

$$I(0) = \int_0^\infty e^{-u^2}\, du = \frac{\sqrt{\pi}}{2}$$

Hence

$$I(s) = \frac{\sqrt{\pi}}{2} e^{-2s} \quad \text{, for } s \geq 0$$

Note that due to symmetry, $I(s) = I(-s)$, and hence

$$I(s) = \frac{\sqrt{\pi}}{2} e^{-2|s|}$$

Thus, letting $s = \frac{l}{2\sigma}\sqrt{\frac{\mu^2}{\sigma^2} - 2\theta}$ i.e. resulting in the integral from I_{12}, the integral I_{12} is simply

$$\begin{aligned} I_{12} &= e^{\frac{l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \frac{1}{\sqrt{2\pi}\sqrt{\frac{\mu^2}{2\sigma^2} - \theta}} \frac{\sqrt{\pi}}{2} e^{-2\frac{l}{2\sigma}\sqrt{\frac{\mu^2}{\sigma^2} - 2\theta}} \\ &= \frac{\sigma}{2\sqrt{\mu^2 - 2\theta\sigma^2}} \end{aligned}$$

Combining the above results, we finally have I_1

$$\begin{aligned} I_1 &= e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \left(1 + \frac{\mu - \sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma} \frac{\sigma}{2\sqrt{\mu^2 - 2\theta\sigma^2}}\right) \\ &= e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \left(\frac{1}{2} + \frac{\mu}{2\sqrt{\mu^2 - 2\theta\sigma^2}}\right) \end{aligned}$$

Similar calculations results in the integral I_2 or simply by letting $\mu = -\mu$ as discussed before. In total, we find the moment generating function to be

$$\begin{aligned} \mathbb{E}\left[e^{\theta T_l}\right] &= e^{\frac{l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \left(\frac{1}{2} + \frac{\mu}{2\sqrt{\mu^2 - 2\theta\sigma^2}}\right) \\ &\quad + e^{\frac{2\mu l}{\sigma^2}} e^{\frac{-l\mu - l\sqrt{\mu^2 - 2\theta\sigma^2}}{\sigma^2}} \left(\frac{1}{2} - \frac{\mu}{2\sqrt{\mu^2 - 2\theta\sigma^2}}\right) \\ &= e^{\frac{l}{\sigma^2}\left(\mu - \sqrt{\mu^2 - 2\theta\sigma^2}\right)} \end{aligned}$$

From the above calculations, this is clearly defined for θ in some neighborhood of 0, thus the above is indeed a proper moment generating function. Furthermore, all derivatives exists at $\theta = 0$.

This also shows that T_l does not belong to neither the Gamma family nor Phase-Type

The first 3 moments are given by

$$\mathbb{E}\left[T_l\right] = \frac{l}{\mu}, \quad \mathbb{E}\left[T_l^2\right] = \frac{l\sigma^2}{\mu^3} + \frac{l^2}{\mu^2}, \quad \mathbb{E}\left[T_l^3\right] = \frac{3l\sigma^4}{\mu^5} + \frac{3l^2\sigma^2}{\mu^4} + \frac{l^3}{\mu^3}$$

Interestingly, the average is exactly what one would expect if no stochasticity was present. Furthermore, the variance has a simple nice form, namely $\frac{l\sigma^2}{\mu^3}$.

Continuing the simulation from above, the theoretical mean evaluates to 100 whereas the simulated mean evaluated to 100.343. The theoretical variance is 2500 whereas the variance from simulation was 2468.224.

Chapter 2

Problemformulering / Introduktion

In many production facilities, planning is a big part of maximizing some index. Whether this is production throughput over some time period and thus often also the economic surplus or some other key index, it is of great importance to have an underlying model to describe the observed variation. In particular in operational research, the schedules may drift in suboptimal ways if the variation is not considered.

Furthermore, from a salesman point of view, expected production and time intervals can be of great use when planning and also building production facilities. Namely, one might find that increasing the volume or efficiency of some part of the facility would increase the production throughput and profitability. This is also known as bottleneck analysis and require some understanding of the underlying mechanics and a stochastic model of this could improve the strength of such results.

Therefore, the primary objective of this paper/thesis is to investigate and model the yield and time of a production flow with focus on the pharmaceutical and chemical production industry. More precisely, we will be building a statistical model for a single process, with the purpose of being able to describe the variation in the yield of the production cycle and production times. This will then be used to analyze potential bottlenecks.

Furthermore, it will be interesting to construct a network of such processes as is typically the case in industry. We shall see how much can be said about such a network and what obstacles one may encounter when trying to analyze such networks which is this thesis will initially be treated as networks of queues.

Chapter 3

Ideer til hvad der skal laves

Overall model for throughput of system. I.e. model the system as e.g. a system of queues and how much is produced at each step and this propagate. The important aspect is breakdown (extra processing time) and possibility of having to trowing out some production along the way, either due to error or some other (unforeseen) causes.

Need to investigate different ways of modelling this (starting with a simple system with no queuing, i.e. a single batch; this is what is done above). Discuss the pros and cons and how much information they preserve (aggregation models etc. may need to model so'me part of the system by throwing away)

- Petri Net
- ODE Stochastic Chemical Reaction (first order)
- Database of pharmacokinetic time-series data
- Chemical Manufacturing Process Data'
- [fCM, sample reference]
- [Bal07]

Chapter 4

Data

In this section, we will describe and analyze the data used in this thesis. The data stems from a simulation study on a chemical batch production system and comprises six time series from which the time stamps (in hours) and level in a tank is of the most interest. The goal of this section is to describe how different phases of the production covary. The data is chosen as it resembles an actual production data set but is also implemented with what at first glance seem to be fairly realistic variation and noise in measurements. Overall, the point of this example data is to exemplify what one may encounter in a real batch production system and from this try to build a model in order to predict or quantify the behavior of the system or learn hidden (causal) structure important for optimization etc.

We define a set of phases/units $\mathcal{U}$ that each batch comprises. In this case, units are identified with IDs 1 through 10 (with subunits such as 3.1) described further in Table 4.2. Then, for each unit, we define the stochastic variables X_u and X_u^D to be the duration and delay after a unit respectively. It is also important to keep track of the level in the tank after each unit is finished, and we thus define variables M_u and M_u^D to be the level in the tank after unit u and its associated delay respectively. As the units are executed in sequence, as is the case in this data, a simple representation of the variables can easily be visualized and is shown in Figure 4.1 below.

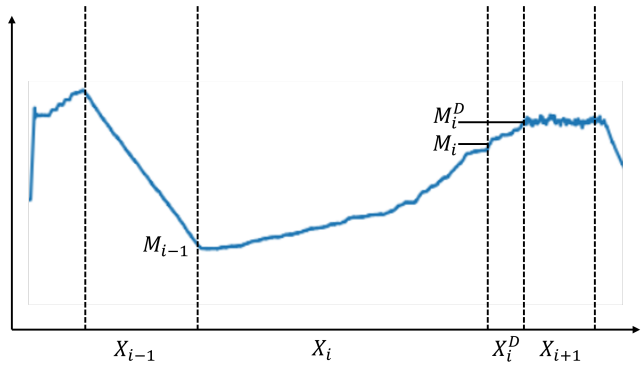


Figure 4.1: Exemplification of the variables X_i, X_i^D, M_i, M_i^D.

4.1 Basic statistics

Initially, we present some basic statistics in Table 4.1 for batches and proceed to discuss the nature of the batches, removing outliers and faulty observations etc.

Cycle	#batch	μ	σ^2	σ	σ/μ
A	66	14.776	3.641	1.908	0.1291
B	64	15.644	3.915	1.979	0.1265
C	61	17.714	2.330	1.526	0.08617
D	60	18.069	6.922	2.631	0.1456
E	60	18.088	9.613	3.100	0.1714
F	63	17.227	7.766	2.787	0.1618

Table 4.1: Per cycle batch duration statistics

Each batch comprises several states. These include adding materials (IDs 1 through 4), centrifugation (ID 5), product transfer (the precipitate generated from the centrifugation, ID 6), chemical reaction (ID 7), a post operation state (Probably to let it cool down to a point where it is ready for further processing, ID 8), Cooling of the product (ID 9), material transfer (transfer the gained product before cleaning of the reaction vessel and/or prepare for the next reaction batch, ID 10). Notice that there is a total of 374 batches throughout the 6 observed cycles.

4.2 Incompleteness on trailing batches

As it may be of interest to investigate the correlation structure of different metrics and variables later on, it is important to understand how each of the batches across the cycles behave. Initially, when looking through the dataset, we observe a few negative phase IDs which will need investigation. However, before we do so, we check that each of the batches actually go through all the states mentioned in [Vic21]. Thus, we take the absolute value of the negative phase IDs to ease the analysis prior to the analysis of the negative phase IDs. After this is done, we observe that not all batches go through all the phases and that some seem to have extra phases not described by [Vic21]. Namely, from Table 4.2, we see that IDs 3 and 4 (which are not described in [Vic21]) have significantly fewer batches going through this phase. But perhaps even more interesting is the final 4 phases where almost all batches goes through these phases.

ID	Count	Description
1.0	374	Addition of liquid raw material `Educt1`
2.0	374	Addition of liquid raw material `Educt2`
3.0	181	
3.1	374	Addition of liquid raw material `Educt3`
3.2	374	Agitation
4.0	163	
4.1	374	Waiting for field operation
4.2	374	Addition of solids
4.3	374	Waiting for control operator
5.0	374	centrifugation
6.0	374	Product transfer
7.0	370	Reaction
8.0	369	Post reaction
9.0	369	Cooling
10.0	368	Material transfer

Table 4.2: The number of batches across all cycles that contains at least one observation for each different absolute phase ID.

Investigating when these inadequacies occur reveals that they are the last batch from each of the cycles. For example, the final batch from cycle A only goes to phase 6 (the product transfer). This can however be explained from the fact that simulation only last for 1100 hours for each cycle and is thus simply cut-off here. As we do not know if these final operations were done at the time the simulations were cut off (which is likely not the case), the final phase for each of the final batches should be disregarded. The cut-off can also be observed in

Figure 4.2. Furthermore, throwing away 6 incomplete batches out of the total 374 will likely not harm the analysis and is thus thrown away as this will make the analysis much simpler later on.

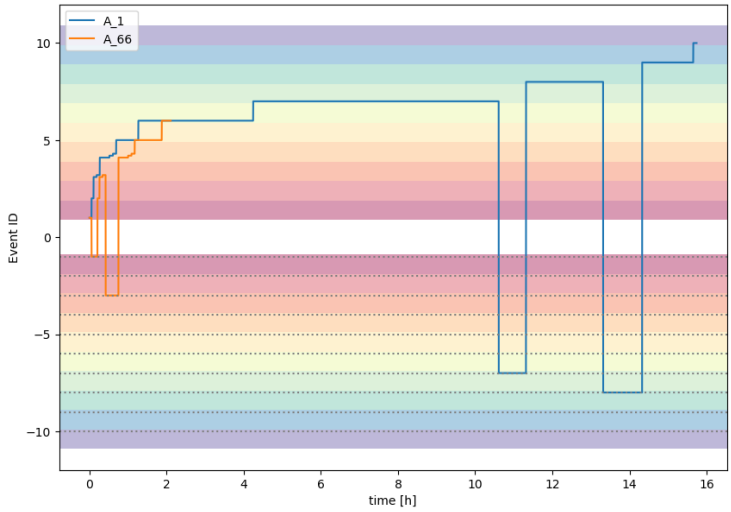


Figure 4.2: The first and last batch from cycle A. It is clear that the final batch is cut-off even before the current phase it finished.

After cutting of the final 6 batches, we have a total of 368 batches of which each goes through all the phases. We thus proceed to discuss the negative phase IDs in the following section, where we also discuss the first four phases.

4.3 Production phases

This part of the process corresponds to events tagged with ID 1 through 10 but will initially concern itself with ID 1 through 4 as much can be learned from the data set here. In Figure 4.3 an example of how the process evolves over time through the different phases is shown. Immediately, we observe something weird, namely the negative event IDs.

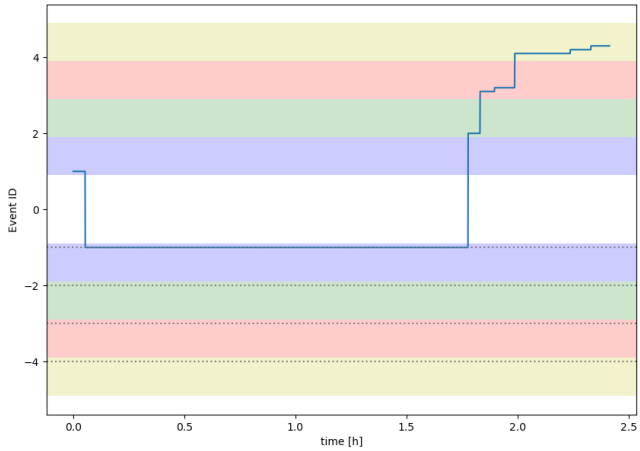


Figure 4.3

To see what is going on here, from data we can see that negative values occur throughout all the six cycles. More specifically, for each negative phase observed, we log in which cycles this occurs. The result is shown in Table 4.3. Notice that -4.1, -4.2 and -4.3 only show up in cycle F, which from [Vic21] is known to be the only one with wrongly labelled phases. We thus suspect that this is indeed the case for these labels and might just have supposed to be the original 4.1, 4.2 and 4.3. To see if nothing funny goes on with these values, these batches are plotted as in Figure 4.3 in Figure 4.4.

Figure 4.4 shows that nothing weird is going on except for the negation of the sub phase's ID. The same can be said for the remaining of the cases where phase ID -4.1, -4.2 and/or -4.3 is used. We thus conclude that these may simply be wrongly labelled thus we convert every such instance to its absolute value and continue with this modified data set from this point on.

Event \ Cycle	A	B	C	D	E	F
-1	■	■	■	■		
-2				■	■	■
-3	■		■	■	■	■
-4		■	■	■	■	
-4.1						■
-4.2						■
-4.3						■
-5	■	■	■	■	■	■
-6		■	■	■	■	■
-7	■		■	■	■	■
-8	■		■	■	■	■
-9				■	■	■
-10		■		■	■	■

Table 4.3: Occurrences of negative phases IDs. It is observed that sub phases 4.1, 4.2, 4.3 only occur in cycle F which is known to be the only cycle with wrongly labelled phases.

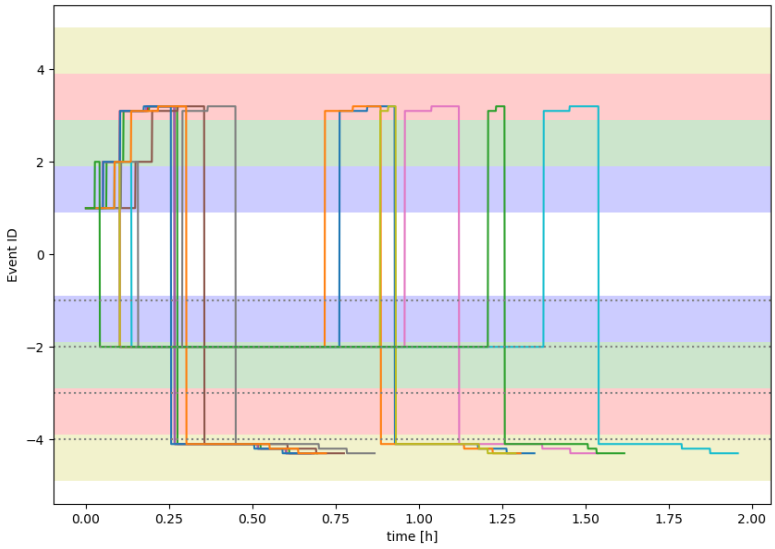


Figure 4.4: 13 of the 48 batches with at least one of the sub phases 4.1, 4.2 4.3 negative.

Having converted the above sub phase IDs we summarize the current situation in regard to negative phase IDs in the following table, Table 4.4. Now all the remaining occurrences of negative phase IDS does not seem to exhibit any structure from looking at Table 4.4. We thus proceed to understand what is going on with the remaining negative phase IDs.

Event \ Cycle	A	B	C	D	E	F
-1	■	■	■	■		
-2				■	■	■
-3	■		■	■	■	■
-4		■	■	■	■	
-5	■	■	■	■	■	■
-6		■	■	■	■	■
-7	■		■	■	■	■
-8	■		■	■	■	■
-9				■	■	■
-10		■		■	■	■

Table 4.4: Occurrences of negative phases IDs. It is observed that sub phases 4.1, 4.2, 4.3 only occur in cycle F which is known to be the only cycle with wrongly labelled phases.

When plotting different batches, it is clear that the negative phase IDs only occur at the end of a phase e.g. -1 only happens after 1 and so on. This together with the fact that -3 and -4 also only happen after 3.1, 3.2 and 4.1, 4.2, 4.3 respectively (and we never see a phase labelled 3 or 4) indicate that the negative phase IDs could very well correspond to delays at the end of a phase, which both makes sense from a production point of view but also from [Vic21] where they note that all simulated cycles have been implemented with delays.

At this point, it would seem that the labels of the processes are understood for phases 1 through 10 corresponding to the actual production in each batch. Thus, we proceed by searching relationships and otherwise quantifying the durations of each phase, both delays and duration for each of the phases. As a beginning, histograms for each of the phases and delays are plotted in Figure 4.5. Notice that phases 4.1, 4.3 and 8 are not shown, this is because they always last 15 min, 5 min and 2 hours respectively with the only derivation being in machine precision either when loaded or during calculations. Furthermore, notice that for the negative IDs i.e. the delays, the orange bar. This bar represents the cases where no delay was observed which is thus modelled as an atom at 1.

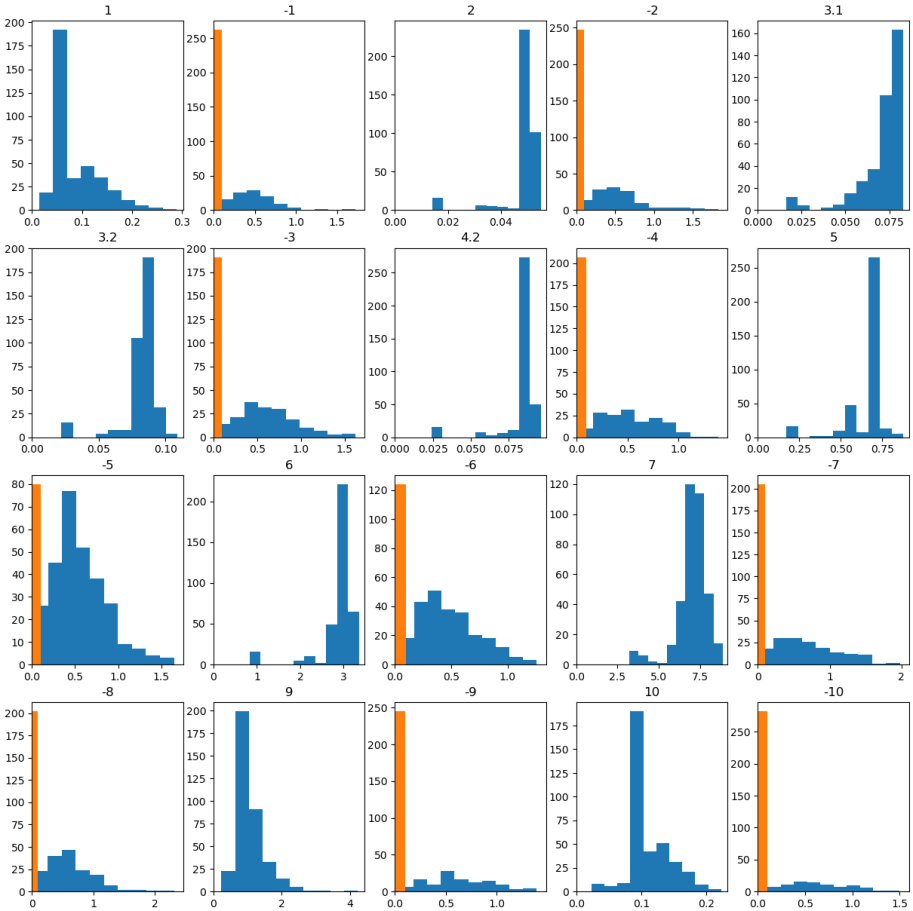


Figure 4.5: Histograms of all phases and delays which are non-constant.

Apart from the above comments, not much catches the eye when looking at Figure 4.5, and we thus proceed by checking if any correlation is immediately present.

4.3.1 Correlations

Lige en korrelationsmatrix

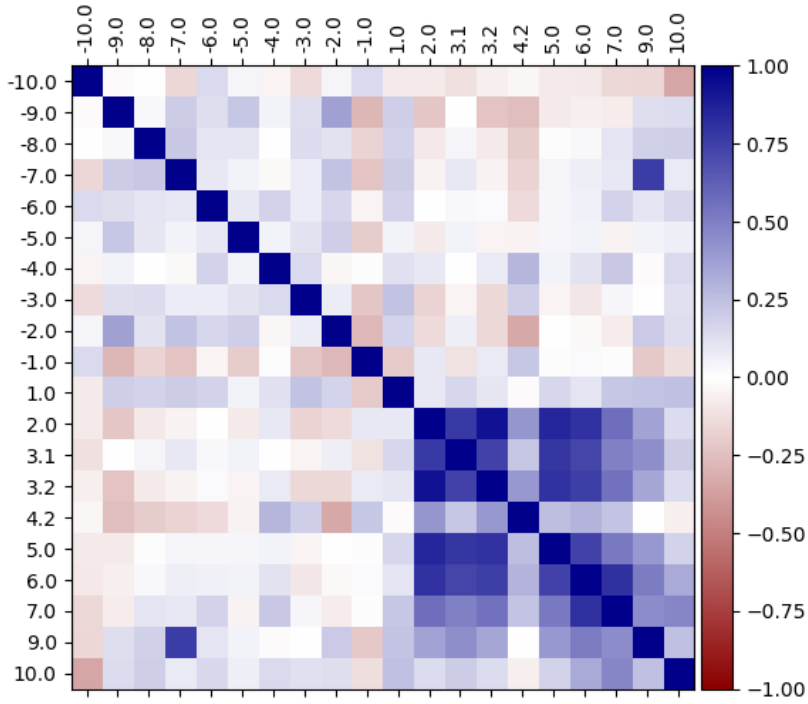


Figure 4.6: Correlation matrix for all phases with non-constant duration.

9 og 10 er ikke specielt korreleret med noget (afkøling og materiale overførsel). Ellers er 2 fremt il og med reaktionen alle korrelerede med hinanden. Eftersom rent fysisk det udvikles i tid, må handlingen i 2 påvirke de næste osv.

Umiddelbart lidt spøjst hvis delay på 7 (reaktion) skulle have noget med tiden for afkøling at gøre, især at den skulle være positiv (ville man ikke tro delay efter produktion ville afkøle mere og dermed reducere behov for afkøling, medmindre varmt steam bliver tilføjet også under delay på 7)

Herunder er samme korrelationsmatrix, dog hvor delay og phasens varighed lagt sammen (også med sub phases såsom 3.1, 3.2 og -3 tilsammen bliver 3)

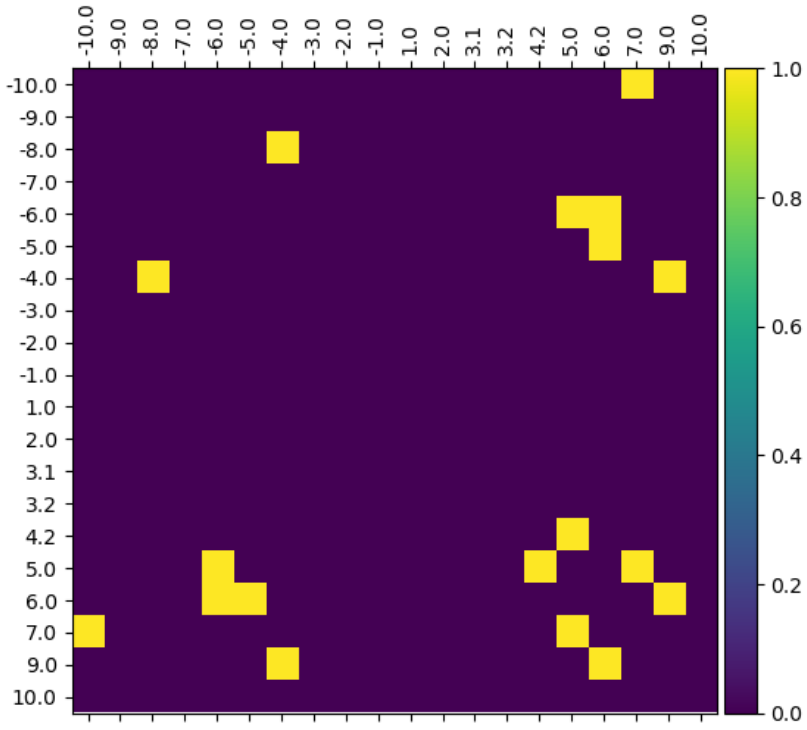


Figure 4.7: Permuteringstest med $\alpha = 0.05$. Also run with less simulations but same result at 1 mil and 10k sims. The Benjamin-Hochberg procedure on the upper (or lower) triangle reveals that none of the correlations are significant.

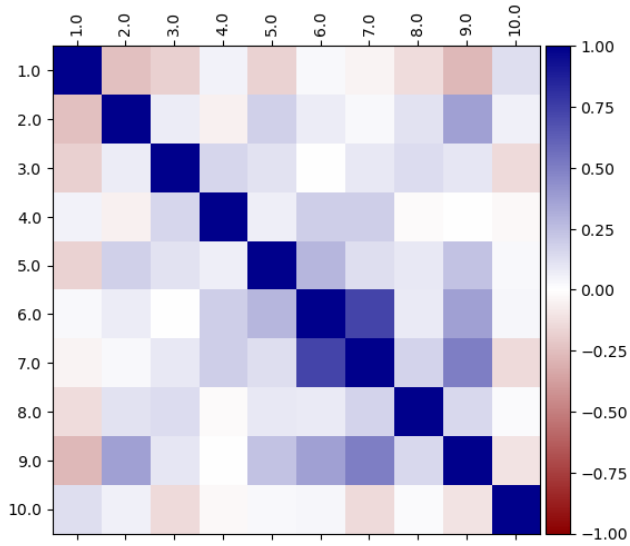


Figure 4.8: Correlation matrix for all phases collapsed

Ligeledes scatter plots for superdiagonalen i ovenstående matrix. Altså phase 1 overfor phase 2, phase 2 overfor phase 3 osv. Er farvelagt efter hvilken cycle de kommer fra. Table 4.4 forklarer hvorfor nogle af de horisontale fremkommer sammen med Figure 4.5 (selve produktionstiden er ret kort sammenlignet med delay.)

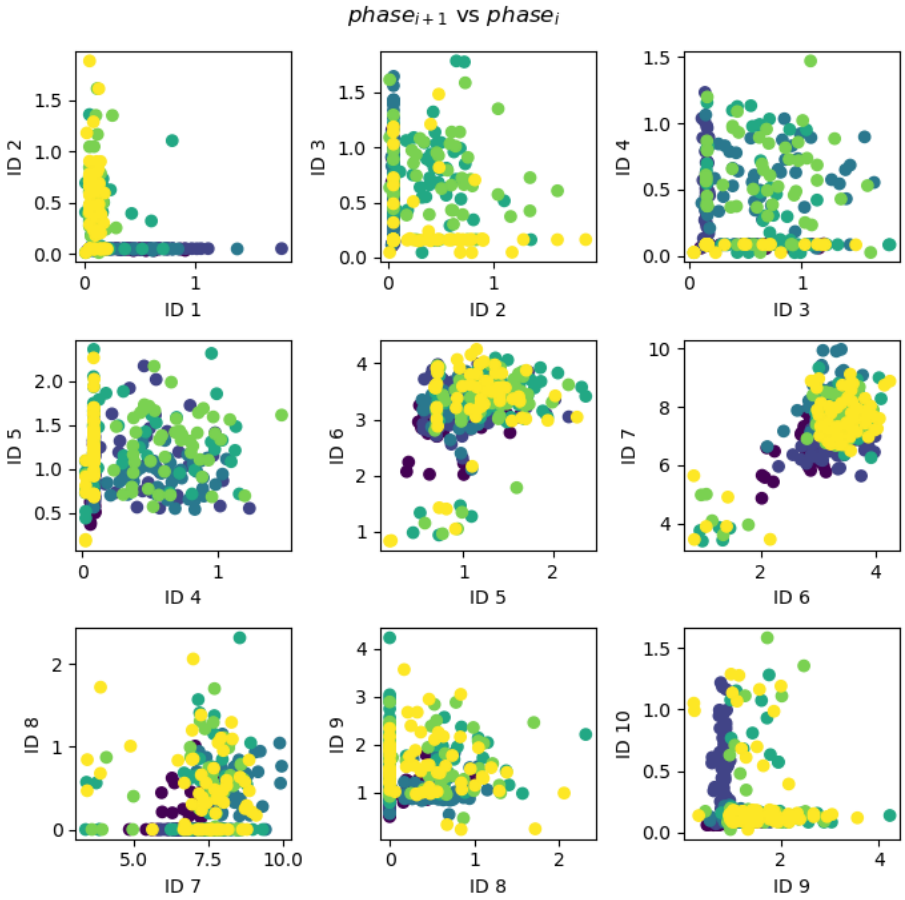


Figure 4.9: Phases vs their next phase when collecting everything regarding a single phase into a total time duration

4.4 Cleaning operations

Sometimes, the vessel is cleansed. This is however not every time after a batch so might be interesting to investigate further. Initially, per cycle, the cleanings are summarized in the following table with basic statistics. As can be seen, there is quite some differences.

The most notifiable differences per batch are the number of cleanses especially when comparing to Table 4.1. For the first two cycles, the cleanses seem to be in between every batch, which is indeed also the while the later four are only sometimes. Furthermore, although the cleanses are between every batch for cycles A and B, the variances are extremely different. For the last four cycles, they seem to be grouped further, E and F are very alike while cleanses in C and D are generally longer although D has a substantially smaller variance than C.

Cycle	#ops	min	max	μ	σ^2	σ	σ/μ
A	65	1.113	3.067	1.917	0.269	0.518	0.270
B	63	1.324	1.751	1.566	0.00883	0.0939	0.0600
C	9	1.544	3.306	2.153	0.277	0.526	0.245
D	10	1.474	2.009	1.581	0.0212	0.146	0.0922
E	10	0.827	1.584	1.465	0.0462	0.215	0.147
F	10	0.748	1.610	1.466	0.0595	0.244	0.166

Table 4.5: Per cycle cleansing statistics

Cycle	A	B	C	D	E	F
$\widehat{\mathbb{E}[\sum X_u]}$	13.993	13.898	15.343	14.471	14.589	14.418
$\widehat{\mathrm{Var}(\sum X_u)}$	0.95636	0.46587	0.76111	4.9589	4.2678	5.3545
$\widehat{\sum \mathrm{Var}(X_u)}$	0.50590	0.31182	0.36667	1.8322	1.5788	1.9696
$\widehat{\mathbb{E}[\sum X_u^D]}$	0.96398	1.9402	2.4503	3.6050	3.7390	3.0041
$\widehat{\mathrm{Var}(\sum X_u^D)}$	0.31843	0.39117	0.90187	1.2468	1.2787	1.0462
$\widehat{\sum \mathrm{Var}(X_u^D)}$	0.34921	0.53198	0.74914	1.4357	1.2454	1.3099
$\widehat{\mathbb{E}[\sum X_u X_u^D]}$	1.9321	1.5001	4.8191	6.4225	6.0405	6.3343
$\widehat{\mathrm{Var}(\sum X_u X_u^D)}$	3.7798	0.89920	16.870	22.133	12.660	16.194

Table 4.6: Each of the time related variables X_i and D_i and variance description.

To verify these observations and potentially discovering more important facts of their probability distributions, histograms are plotted in the following Figure 4.10. We indeed again observe the likeliness between the cycles A and B, C

and D, E and F respectively. Also, for the first two cycles and more so cycle B, the cleaning times are somewhat normally distributed although cycle A has a very heavy right tail in that case. The later four cycles only have 10 observations but the mode (i.e. peak) seem to be about the same.

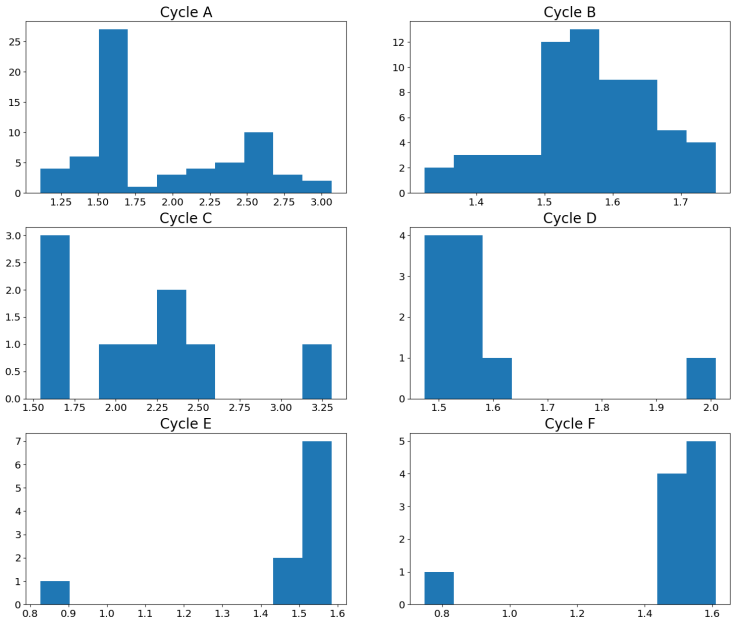


Figure 4.10: Each of the 6 cycles, cleaning operations histograms.

From the above observation of like modes one may want to observe the histogram of the combined set of cleaning times. In particular, under the hypothesis that the durations are actually from the same probability distributions and realized independently within each cycle a histogram of all the observations are of interest and is shown in Figure 4.11 below.

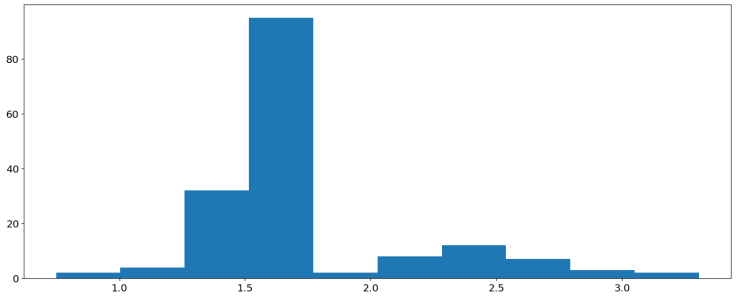


Figure 4.11: Combined cleaning operations histograms.

Finally, to get a better overview of the irregularities is the number of cleaning periods (mostly concerning cycles C through F), each cleaning operation is shown in the following Figure 4.12. The vertical shaded rectangles signify the period in which a cleaning operation is taking place. Furthermore, the event IDs are shown but to get a clearer view on what is going on, a single rectangle (zoomed in) is shown in Figure 4.13.

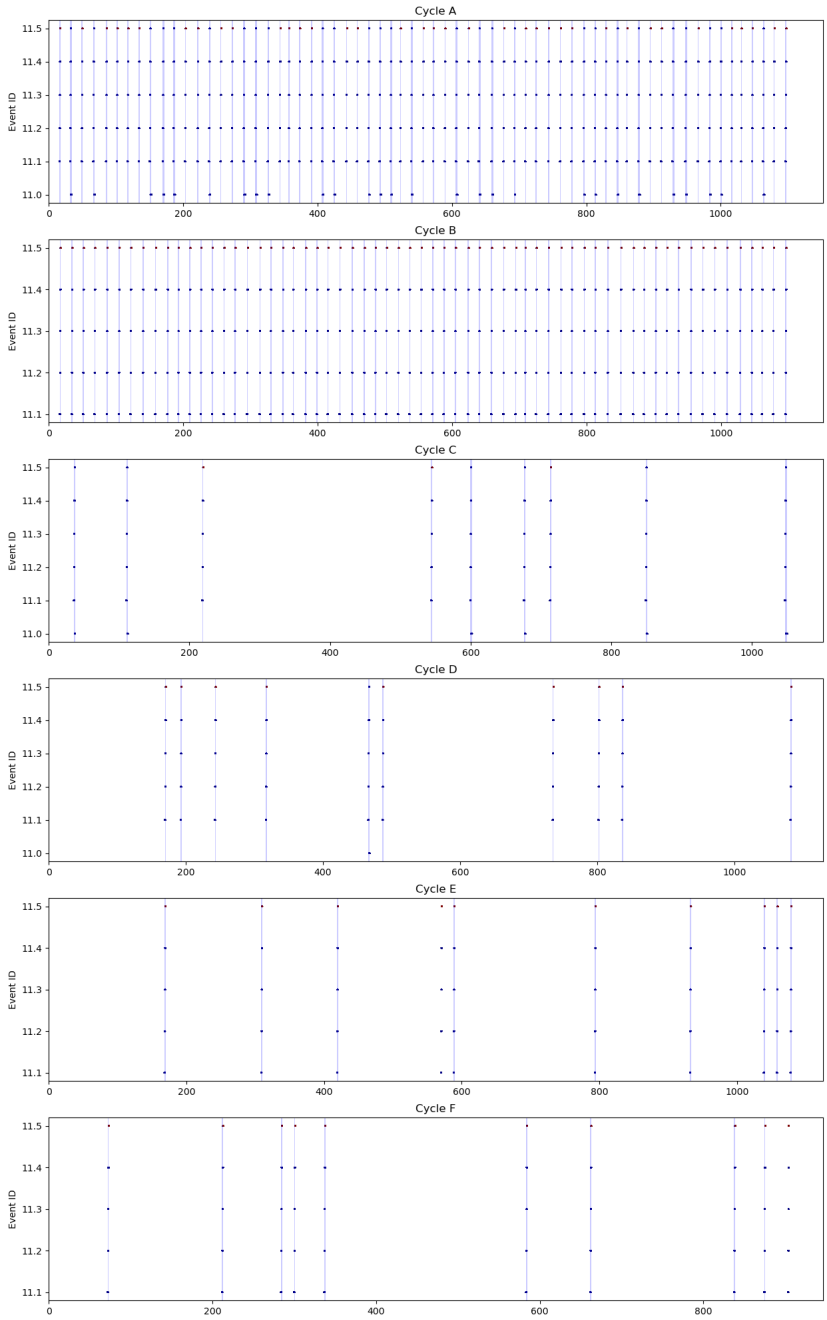


Figure 4.12: Each of the 6 cycles, cleaning (corresponding to `BatchID = 0`). Each *(*Cleaning Procedure), CIP, is highlighted with an opaque interval (the blue rectangles). The dots marked with red (only ID 11.5, but not all of these are red), is if the Cleaning ID is 0.

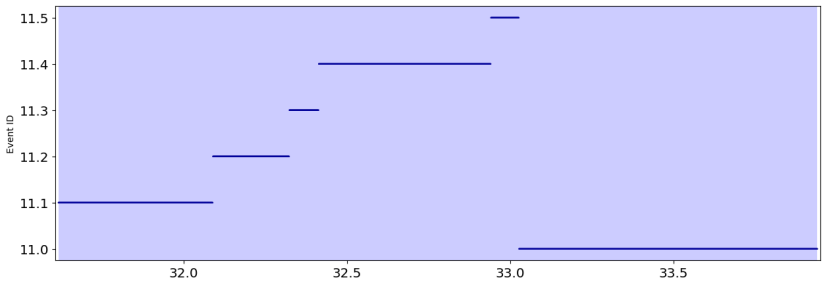


Figure 4.13: A single blue rectangle zoomed in

It is observed that the observations marked with red in figure 4.12 occur exactly when that specific cleaning operation does not go to the state 11.0 after the flush of the tank (event ID 11.5) and vice versa. It is hard to conclude what this may mean, but the cleaning being in state 11.0 may indicate that the system is idle before continuing the next batch like what is observed from the other steps of the process flow. Also, it is noted that while the red dots occur nothing else is happening according to the dataset.

From a modelling point of view, the cycles C through F can be thought of as the cleansing operation having a probability of not happening or equivalently as having a duration of 0. It is thus of interest to observe what the probability of cleaning after an operation is. From Table 4.1 and Table 4.5, we that indeed for cycles A and B, the probability is 100 % when disregarding the possibility of cleaning after the final batch. Hence, we see that for the remaining cycles, the probabilities of cleaning the tank after an operation are as in Table 4.7

Cycle	% cleaning
A	100.00
B	100.00
C	15.00
D	16.95
E	16.95
F	16.13

Table 4.7: Per cycle probability of cleaning

Furthermore, let C_i denote whether the ith batch is followed by a cleaning of the tank or not. It is then of interest if the next batch is followed by a cleaning given whether the current batch is followed by a cleaning. In particular, we count for each of the cycles the transitions which are shown in the following tables. Notice that the number of observations is two less than the total number of batches within each specific cycle. This is due to the last batch is never followed by

a cleaning (nor is the first batch superseded by a cleaning procedure) which results in one less observation and also due to the fact that we are logging transitions and hence lose another observation. To test for randomness, a Chi-squared test is carried out on each of the cycles to check for independence. It is observed all the cycles exhibit independence between the groups i.e. there is no statistical evidence for information is gained about if the next batch is followed by a cleaning operation given whether the current batch is followed by a cleaning operation.

C_i \ C_{i+1}	No	Yes
No	41	9
Yes	9	0

(a) C, $p = 0.3293$

C_i \ C_{i+1}	No	Yes
No	41	8
Yes	7	2

(b) D, $p = 0.6456$

C_i \ C_{i+1}	No	Yes
No	41	7
Yes	7	3

(c) E, $p = 0.3532$

C_i \ C_{i+1}	No	Yes
No	41	9
Yes	9	1

(d) F, $p = 1.0000$

Table 4.8: Contingency table for Cycle C-F

Thus collecting the observations from all the last four cycles, we may want to model the atom of the cleaning procedure independently of the previous batch and with a probability of 0.8375 corresponding to the cleaning procedure only being carried out $16,25\%$ of cases.

Finally, we show the autocorrelation function for each the four cycles C-F in Figure 4.14 and note that all the ACF stay within the 95% confidence interval.

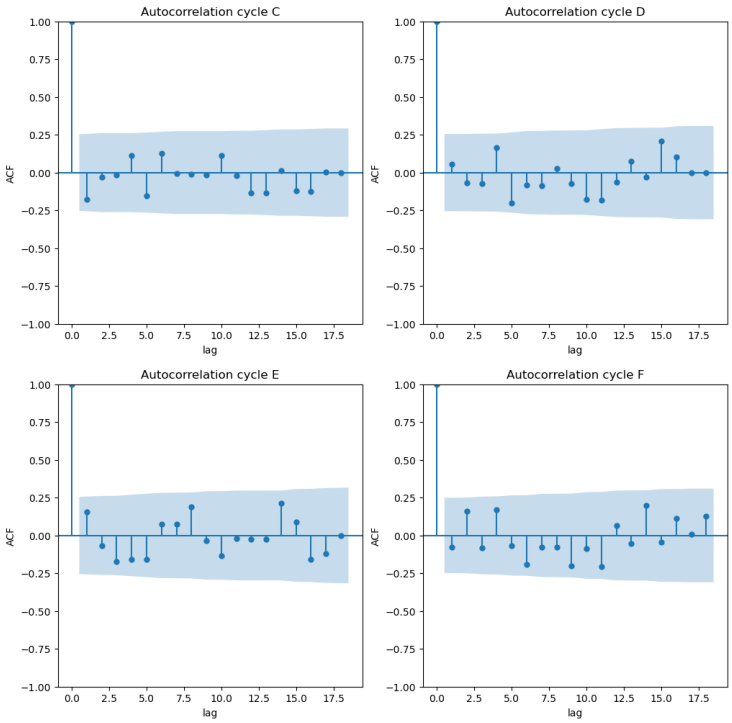


Figure 4.14: Autocorrelation function for each of the final 4 cycles. As can also be seen from this there seem to be no information to be gained of C_i from C_{i-1}.

Chapter 5

Method

Definition 5.1. *Induced Norm.*

A matrix norm $|||\cdot|||$ *is said to be induced by the vector norm* $||\cdot||$ *when*

$$|||A||| = \sup_{||x||=1} ||Ax||$$

See e.g. https://www.anandinstitute.org/pdf/Roger_A.Horn.%20_Matrix_Analysis_2nd_editio

Definition 5.2. *Sub-multiplicative Matrix norm*

A matrix norm $|||\cdot|||$ *is said to be sub-multiplicative, if for every* $A, B \in \mathbb{F}^{n\times n}$ *where* $\mathbb{F}$ *is either the real or complex field:*

$$|||AB||| \leq |||A||| \cdot |||B|||$$

5.1 Copula based network discovery

Suppose a set of random variables (X_i) is given. The method presented in this section aims to discover direct relationships between pairs X_i and X_j for

$i \neq j$. These relationships can be presented using an undirected graph. We will later discuss methods of directing edges such that a causal network may be discovered.

One way of discovering such a network from a set of observations is through mutual information. Namely, suppose that we have obtained a matrix G_{obs} with similarities between each pair of random variables. We will use mutual information as similarity. Let G_{dir} be the information on each edge of the graph i.e. only the direct effects between pairs of variables.

$$G_{obs} = G_{dir} + G_{indir} = G_{dir} + G_{dir}^2 + G_{dir}^3 + \cdots = (I - G_{dir})^{-1} - I = G_{dir} (I - G_{dir})^{-1}$$

where it is assumed that the above infinite sum converges. A necessary and sufficient condition for this is that $\rho(G_d ir) < 1$ where $\rho(\cdot)$ denotes the spectral radius.

It follows that G_{dir} can be calculated as

$$G_{dir} = G_{obs} (I + G_{obs})^{-1}$$

Furthermore, as described by ... . As G_{obs} is symmetric it is diagonalizable with a matrix U such that $\Lambda_{obs} = U^T G_{obs} U$ where Λ_{obs} is a diagonal matrix containing the spectral values with multiplicity in any ordering. It follows that since $\rho(G_{obs}) < 1$, $\lambda_{obs} + I$ is also invertible and

$$G_{dir} = U \Lambda_{dir} U^T$$

Where $\Lambda_{dir} = \Lambda_{obs} (I + \Lambda_{obs})^{-1}$ i.e. $[\Lambda_{dir}]_{ii} = \frac{\lambda_{obs}}{1+\lambda_{obs}}$ and 0 elsewhere. Whether this is better than a normal inversion of $I + G_{obs}$ is at this point unknown.

At this point, we only need G_{obs} and as mentioned earlier we will use mutual information. However, to make the calculations more robust and efficient we will use a closely related measure of dependence, namely CE which is defined as follows

$$CE(X_1, \ldots, X_n) = -\int \cdots \int_{[0,1]^n} c(u_1, \ldots, u_n) \log_b c(u_1, \ldots, u_n)\, du_1 \ldots du_n$$

where $c(\cdot)$ is the uniquely defined copula of the joint distribution $f_{\mathbf{X}}(\cdot)$. We show that the above is indeed equal to the negative mutual information. First, we realize that from definition,

$$\begin{aligned} c(u_1, \ldots, u_n) &= \partial_{\mathbf{u}} C(u_1, \ldots, u_n) \\ &= \partial_{\mathbf{u}} F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right) \\ &= f(x_1, \ldots, x_n) \frac{1}{f_1(x_1) \ldots f_n(x_n)} \end{aligned}$$

Thus

$$\begin{aligned}
-CE &= \int \dots \int_{[0,1]^n} c\left(u_1, \dots, u_n\right) \log c\left(u_1 \dots, u_n\right)\, du_1, \dots, du_n \\
&= \int \dots \int_{[0,1]^n} c\left(u_1, \dots, u_n\right) \log c\left(u_1 \dots, u_n\right)\, dF_1(x_1), \dots, dF_n(x_n) \\
&= \int \dots \int_{\mathbb{R}^n} \frac{f(x_1, \dots, x_n)}{f_1(x_1) \dots f_n(x_n)} \log \left(\frac{f(x_1, \dots, x_n)}{f_1(x_1) \dots f_n(x_n)} \right) f_1(x_1) \dots f_n(x_n)\, dx_1 \dots dx_n \\
&= \int \dots \int_{\mathbb{R}^n} f(x_1, \dots, x_n) \log \left(\frac{f(x_1, \dots, x_n)}{f_1(x_1) \dots f_n(x_n)} \right) dx_1 \dots dx_n \\
&= MI
\end{aligned}$$

Hence, we obtain G_{obs} from the pairwise copula entropy (CE)

$$G_{obs} = \begin{bmatrix} 0 & -CE_{12} & \dots & -CE_{1n} \\ -CE_{21} & 0 & \dots & -CE_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -CE_{n1} & -CE_{n2} & \dots & 0 \end{bmatrix}$$

Although theoretically $-NCE_{ii} = \infty$ for all i, we put 0 in the diagonal because we do not want self explanation as these are trivial. The argument for calculating CE instead of MI are due to the finite volume integral and simpler integrand. In particular, using the copulas, we avoid the fraction $\frac{f(x_1,\dots,x_n)}{f_1(x_1)\dots f_n(x_n)}$ which could easily result in numerical instability e.g. when both f and f_is are close to 0.

Finally, from the deconvoluted information matrix D_{dir} we may choose a threshold t for choosing which edges are significant. The choice of t

Theorem 5.3 (Sklar's theorem)**.** *For a random vector* $\boldsymbol{X}$ *with CDF* F *and univariate marginal CDFs* $F_1, \ldots, F_d$*. There exists a copula* C *such that*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

If X *is continuous,* C *is unique.*

The following corollary follows immediately

Corollary 5.3.1. *Coordinate transformation*

Under the assumptions of Theorem 5.3, given any set $(T_1, \ldots, T_d)$ *of strictly increasing functions, if* C *is a copula of* $(X_1, \ldots, X_d)$ *then it is also a copula of* $(T_1(X_1), \ldots, T_d(X_d))$*.*

Proof. Suppose $(X_1, \ldots, X_d)$ permits a copula C and let T_i be given as stated. Consider coordinate wise the result of the transformation $Y_i = T_i(X_i)$ and consider the CDF $F_{Y_i}(y_i)$

$$F_{Y_i}(y_i) = \mathbb{P}(Y_i \leq y_i) = \mathbb{P}\left(T_i^{-1}(Y_i) \leq T_i^{-1}(y_i)\right) = \mathbb{P}(X_i \leq x_i) = F_{X_i}(x_i)$$

Thus

$$\begin{aligned} F_{\boldsymbol{X}}(x_1, \ldots, x_d) &= C\left(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right) \\ &= C\left(F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)\right) \\ &= F_{\boldsymbol{Y}}(y_1, \ldots, y_d) \end{aligned}$$

where Sklar's theorem have been used for the final equality. □

The above corollary is actually equivalent with a seemingly stronger statement and follows easily

Proposition 5.4. *Since* T_i *is strictly increasing, the inverse* T_i^{-1} *exists and is also strictly increasing. Thus, the above implication is bidirectional and hence for strictly increasing functions* T_i*,* C *is a copula of* $(X_1, \ldots, X_d)$ *if and only if it is a copula of* $(T_1(X_1), \ldots, T_d(X_d))$*.*

5.2 Algorithms

Algorithm 1 G_{obs} computation

Require: $N > 0$ ▷ Number of variables
  for $1 \leq i < j \leq N$ **do**
    Estimate F_i and F_j from $x_i^{\mathcal{D}}$ and $x_j^{\mathcal{D}}$
    $u_i^{\mathcal{D}} \leftarrow F_i(x_i^{\mathcal{D}})$
    $u_j^{\mathcal{D}} \leftarrow F_j(x_j^{\mathcal{D}})$
    Estimate C_{ij} from $u_i^{\mathcal{D}}$ and $u_j^{\mathcal{D}}$
    Compute NCE_{ij}
    $G_{ij}, G_{ji} \leftarrow -NCE_{ij}$
  end for

Algorithm 2 (ND) Network Deconvolution

Require: G_{obs} ▷ Input observational matrix
  $[G_{obs}]_{ii} \leftarrow 0,\ \forall 1 \leq i \leq N$ ▷ zero-diagonal
  $Q_p \leftarrow G_{[1-\alpha]}$
  Set G_{obs} 0, where $G_{obs} < Q_p$
  Compute eigendecomposition Q, Λ of G_{obs}
  $\lambda^+ \leftarrow \max(\lambda^{\max}, 0)$
  $\lambda^- \leftarrow -\min(\lambda^{\min}, 0)$
  $m^+ \leftarrow \frac{1-\beta}{\beta}\lambda^+$
  $m^- \leftarrow \frac{1+\beta}{\beta}\lambda^-$
  $m \leftarrow \max(m^+, m^-)$
  $\hat{\Lambda} \leftarrow \Lambda(mI + \Lambda)^{-1}$
  return $Q\hat{\Lambda}Q^T$

Remark. The $1-\alpha$ quantile, denoted by $G_{[1-alpha]}$ (of the upper triangular matrix), can be computed in many ways. As it is only used to filter the G_{obs} matrix, its precise value does not matter. Only the property α part of the observations are below $Q_{1-\alpha}$ and $1-\alpha$ are above (or equal to) $Q_{1-\alpha}$. This property is for example fulfilled by the quantile function `quantile` from `NumPy` (v. 1.26.4). Thus setting $\alpha = 1$ will result in G_{obs} retaining all entries (except for the diagonal entries).

Remark. The $\beta \in (0, 1)$ parameter serves as a sort of regularization. The algorithm above also maps the maximum absolute value of the eigenvalues (the spectral radius) to β as is also pointed out in the implementation of Source code from Nature paper at https://compbio.mit.edu/nd/index.html. The proof of this is shown below

Proof. To show that eigenvalues i.e. the diagonal elements of Λ resulting from Algorithm 2 all fall in the interval $[-\beta, \beta]$ (i.e. $\sigma\left(Q\Lambda Q^T\right) \subseteq [-\beta, \beta]$) where at least one λ is mapped to either $-\beta$ or β, first notice that clearly the resulting eigenvalues of $G_{dir} = Q\hat{\Lambda}Q^T$ are clearly given by $\frac{\lambda_i}{m+\lambda_i}$ where $(\lambda_i)_{\{1,\dots,N\}}$ are the (real) eigenvalues of G_{obs} from the definition of $\hat{\Lambda}$. We will show the above by first considering $\lambda \geq 0$ and $\lambda < 0$.

For $\lambda \geq 0$, clearly $m \geq \frac{1-\beta}{\beta}\lambda^+$, thus

$$\frac{\lambda}{m+\lambda} = \frac{1}{1+m/\lambda} \leq \frac{1}{1+\frac{\lambda^+}{\lambda}\frac{1-\beta}{\beta}} \leq \frac{1}{1+\frac{1-\beta}{\beta}} = \beta$$

where the final inequality follows from $\lambda \leq \lambda^+$. Hence $[0, \lambda^+] \to [0, \beta]$.

Furthermore, for $0 > \lambda \geq -\lambda^-$, note that also $m \geq \frac{1+\beta}{\beta}\lambda^-$. Since $\beta \in (0, 1]$, $m + \lambda \geq \frac{1+\beta}{\beta}\lambda^- + \lambda > 0$ and thus $\frac{\lambda}{m+\lambda} < 0$ which implies

$$-\frac{\lambda}{m+\lambda} \leq \frac{-\lambda}{\frac{1+\beta}{\beta}\lambda^- + \lambda} = \frac{1}{\frac{1+\beta}{\beta}\frac{\lambda^-}{-\lambda} - 1} \leq \frac{1}{\frac{1+\beta}{\beta} - 1} = \beta$$

i.e. $[-\lambda^-, 0) \to [-\beta, 0)$. This shows that indeed all the eigenvalues of G_{dir} is numerically less that or equal to β. Finally, assuming $m \neq 0$ or equivalently that $G_{obs} \neq \mathbf{0}$, either $m = \frac{1-\beta}{\beta}\lambda^+$ (and thus $\lambda^+ \neq 0$ is an eigenvalue of G_{obs}) for which the above shows that indeed λ^+ is mapped to β or $m = \frac{1+\beta}{\beta}\lambda^-$ (and hence $\lambda^- \neq 0$ and thus $-\lambda^-$ is an eigenvalue of G_{obs}) for which $-\lambda^-$ is mapped to $-\beta$. This shows that G_dir indeed has an eigenvalue which numerical value is β. □

5.3 Examples

In this section, we will investigate how the algorithms Algorithm 1 and Algorithm 2 works in junction and, if so, observe how the algorithm can fail and what may be done to correct such cases. Initially, a few simple examples involving exponentiated multivariate Gaussians $\boldsymbol{Y}$.

Example 5.1. *Exponentiated multivariate Gaussian*

Let us consider a simple case with $\mathbf{Y} = e^{\mathbf{X}}$ *(element wise exponentiation) where* $X \sim \mathcal{N}(\mathbf{0}, \Sigma)$ *where*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0.9\sigma_1\sigma_2 & 0 \\ 0.9\sigma_1\sigma_2 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix}$$

It is clear that to Algorithm 1, the mean is non-important as simply corresponds to a scaling of the Y_i *variables. Furthermore, because of Corollary 5.3.1, theoretically, due to the uniqueness of the Copula* C *(as* $\boldsymbol{Y}$ *is continuous) we should expect near equal or very similar results for* $\boldsymbol{Y}$ *and* $\boldsymbol{X}$ *from Algorithm 1. Additionally, different* σ *corresponds to different scaling of* $\boldsymbol{X}$*, and thus we should observe equal or near equal* G_{dir} *for all* $\boldsymbol{Y}$*. Initially, we shall see how this hypothesis holds up to the following three examples*

$$\boldsymbol{\sigma} = (0.07, 0.3, 0.9), \quad \boldsymbol{\sigma} = (1, 1, 1), \quad \boldsymbol{\sigma} = (1, 2, 3)$$

In order for the sample size to not influence the results, we simulate a generous number of samples, namely, for the following results we have used $n = 10{,}000$ *samples. For* $\boldsymbol{\sigma} = (1, 1, 1)$*, Algorithm 1 and Algorithm 2 returns the following (using* $\alpha = 1$ *and* $\beta = 0.99$*)*

$$G_{dir} = \begin{bmatrix} -0.33396 & 0.6660 & 0.02512 \\ 0.6660 & -0.3341 & 0.02730 \\ 0.02512 & 0.02730 & -0.0020583 \end{bmatrix} \tag{5.1}$$

Similarly, for $\boldsymbol{\sigma} = (0.07, 0.3, 0.9)$*:*

$$G_{dir} = \begin{bmatrix} -0.3335 & 0.6665 & 0.01414 \\ 0.6665 & -0.3335 & 0.01418 \\ 0.01414 & 0.01418 & -0.00060124 \end{bmatrix} \tag{5.2}$$

Finally, for $\boldsymbol{\sigma} = (1, 2, 3)$*:*

$$G_{dir} = \begin{bmatrix} -0.1490 & 0.09535 & 0.3599 \\ 0.09535 & -0.2989 & 0.5831 \\ 0.3599 & 0.5831 & -0.4037 \end{bmatrix}$$

For $\boldsymbol{\sigma} = (1, 1, 1)$ and $\boldsymbol{\sigma} = (0.07, 0.3, 0.9)$ we observe the most resemblance to the Σ, although the resulting G_{dir} deviate in the final column. The difference is likely produced by Algorithm 1 as if the resulting G_{obs} was the same, then so would G_{dir} and from the above argument, we know that theoretically this should be the case. For the final example, $\boldsymbol{\sigma} = (1, 2, 3)$, we see a completely different result and immediately suspect that there must be some numerical errors. Investigating the partial results of Algorithm 1 we immediately see a flaw in the supposedly uniform variables U_i as shown in figure Figure 5.1

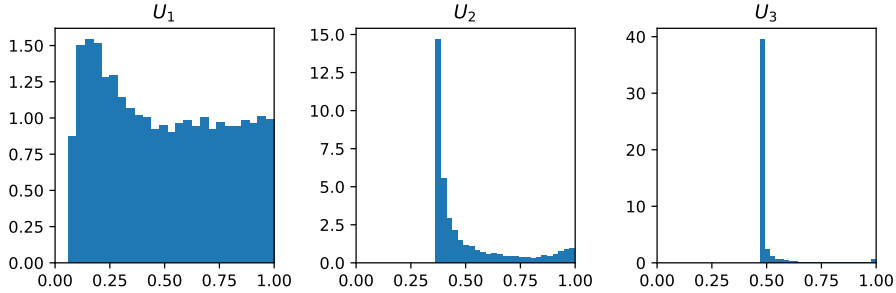


Figure 5.1: The samples transformed using $U_i = F_i(X_i)$ for $\boldsymbol{\sigma} = (1, 2, 3)$. These should be uniformly distributed, but clearly this is not the case for U_2 and U_3. Even U_1 does not quite resemble 10,000 samples from a uniform distribution.

Before handling this, the non-uniformity of U_1 in Figure 5.1 is likely also present in the case when $\boldsymbol{\sigma} = (1, 1, 1)$. Indeed, Figure 5.2 shows that this is indeed the case.

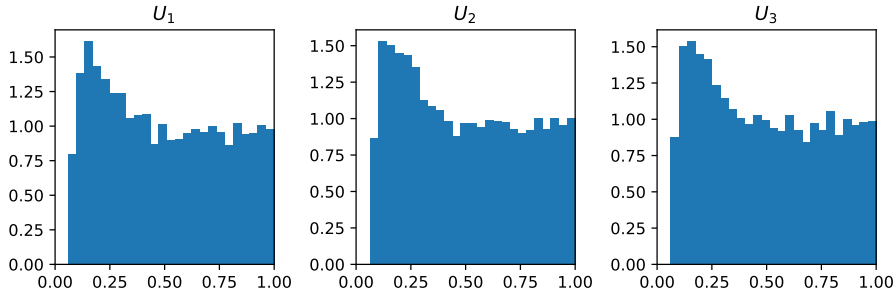


Figure 5.2: The samples transformed using $U_i = F_i(X_i)$ for $\boldsymbol{\sigma} = (1, 1, 1)$.

Finally, just to be sure, $\boldsymbol{\sigma} = (0.07, 0.3, 0.9)$ is also shown in Figure 5.3 and

seems very reasonable, except for U_3.

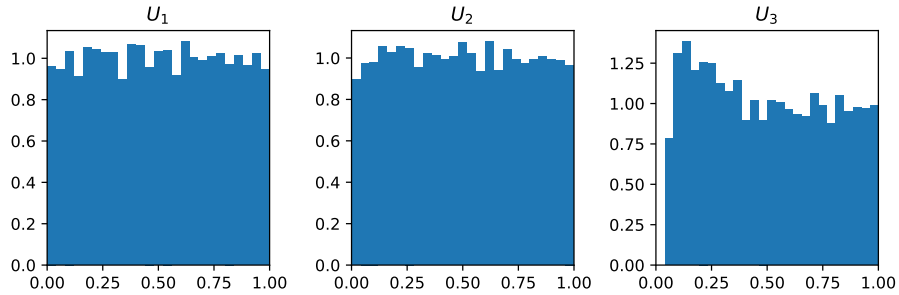


Figure 5.3: The samples transformed using $U_i = F_i(X_i)$ for $\boldsymbol{\sigma} = (0.07, 0.3, 0.9)$.

From the above examples, it seems that the larger the variance, the worse the uniforms turn out. Reasons for this could include numerical issues when trying to calculate $u_i^{(j)}$ *form* $y_i^{(j)}$ *by* $u_i^{(j)} = \int_{-\infty}^{y_i^{(j)}} f_i(y)\,dy$ *and bad fitting of the kernel density estimate from observations. In particular, for values similar, which happens in the case for large* σ *such that we observe large negative realizations of* X_i, $y_i^{(j)}$ *are almost 0, and when computing the integral could result in identical values. Furthermore, from Figure 5.4 we see that indeed the fit is quite poor. Note that we have zoomed in on the interval* $[-200, 200]$ *which contains* 96.2% *of observations. The poor fit is primarily due to the use of Scott's Rule as discussed above which in this case overshoots the optimal bandwidth by a lot.*

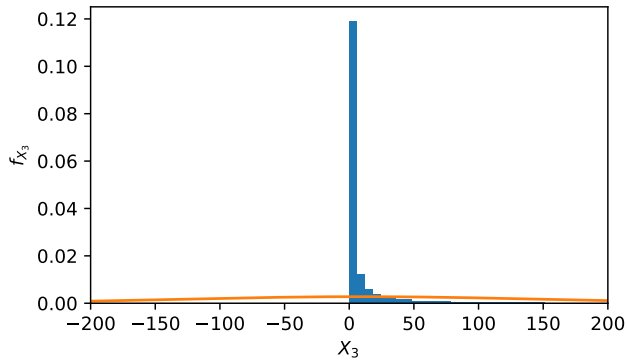


Figure 5.4

The poor fit also explains the high concentration of U_3 *around* 0.5 *in Figure 5.1 as only* 54.5% *of the probability mass lies above* 0.

However, also here Corollary 5.3.1 proves to be useful. Namely, we can get rid of the numerical issues by transforming Y_i *using e.g.* $\log(\cdot)$ *or* $(\cdot)^p$ *for* $p > 0$ *to get even out the observations more. As the first simply inverts the initial transformation of* X_i*, we choose the latter as a more interesting case. In particular, choosing* $p < 1$ *will result in a more even distribution. In the following,* $p = 1/10$ *has been used to transform* $\boldsymbol{Y}$ *prior to running Algorithm 1 and the resulting* $u_i^{(j)}$ *is shown in Figure 5.5.*

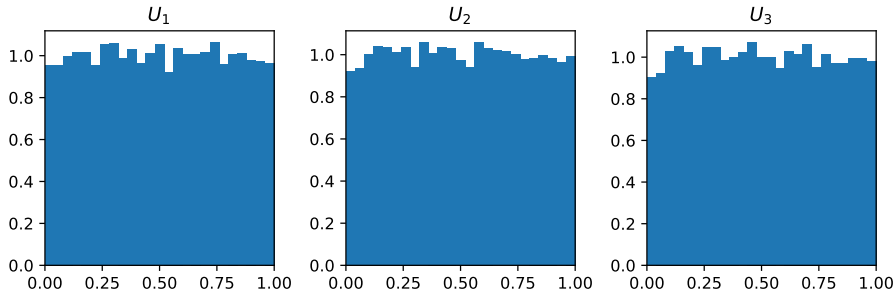


Figure 5.5

The resulting $u_i^{(j)}$ *now seem to follow a uniform distribution and indeed the KDE fits much better as seen in Figure 5.6.*

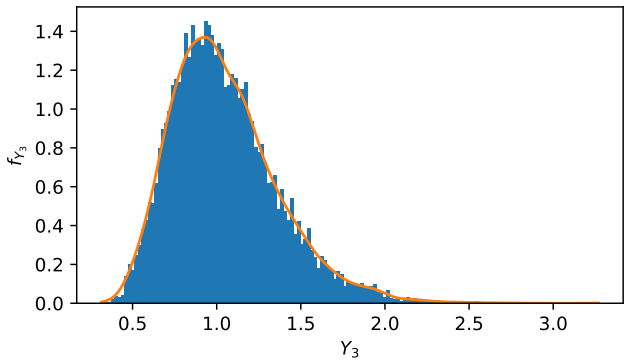
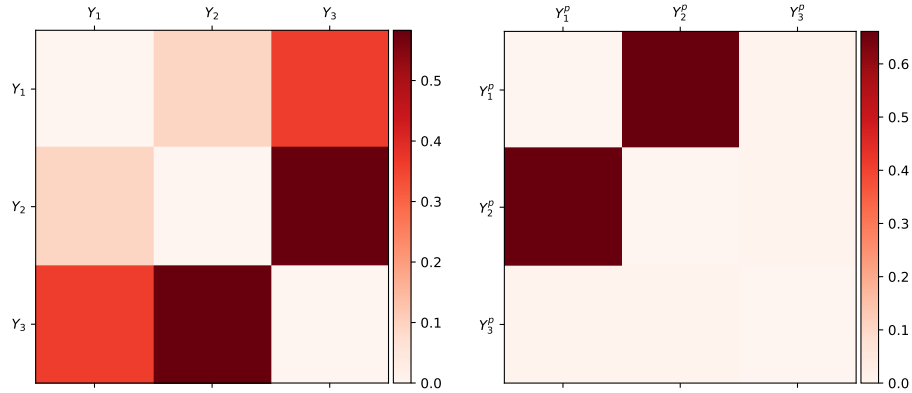


Figure 5.6

Turning to Algorithm 1 and Algorithm 2 we now find that G_{dir} is given by

$$G_{dir} = \begin{bmatrix} -0.3290 & 0.6610 & 0.008440 \\ 0.6610 & -0.3290 & 0.008150 \\ 0.008440 & 0.008150 & -0.0002061 \end{bmatrix}$$

Which is indeed much more comparable with the result from before in Equation 5.1 and Equation 5.2. The difference between G_{dir} from $\boldsymbol{Y}$ and $\boldsymbol{Y}^p$ is clearly visible in Figure 5.7 and also Figure 5.7b resembles the original correlation structure.



(a) **(b)**

Figure 5.7: G_{dir} resulting from 10,000 samples from multi variate Gaussian with $\boldsymbol{\sigma} = (1, 2, 3)$ in **(a)** with raw samples from $\boldsymbol{Y}$ and in **(b)** the transformed data corresponding to $\boldsymbol{Y}^p$.

Finally, to end this example we shall compare with some theoretical results. Namely, the output G_{obs} of Algorithm 1 can also be calculated theoretically. For this, we shall use Proposition 5.5 which permits a theoretical result, namely

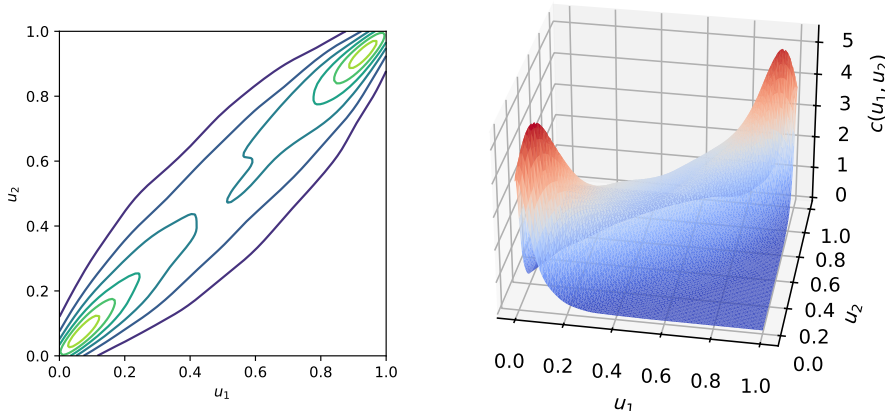
$$\begin{aligned} G_{obs} &= \begin{bmatrix} 0 & -\frac{1}{2}\ln\left(1-\rho_{12}^2\right) & -\frac{1}{2}\ln\left(1-\rho_{13}^2\right) \\ -\frac{1}{2}\ln\left(1-\rho_{21}^2\right) & 0 & -\frac{1}{2}\ln\left(1-\rho_{23}^2\right) \\ -\frac{1}{2}\ln\left(1-\rho_{31}^2\right) & -\frac{1}{2}\ln\left(1-\rho_{32}^2\right) & 0 \end{bmatrix} \\ &\cong \begin{bmatrix} 0 & 0.83037 & 0 \\ 0.83037 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Similarly, prior to deconvolution, using just the sampled $\boldsymbol{X}$ (i.e. no exponential

transform), Algorithm 1 returns

$$G_{obs} = \begin{bmatrix} 0. & 0.71841756 & 0.01781815 \\ 0.71841756 & 0. & 0.01769672 \\ 0.01781815 & 0.01769672 & 0. \end{bmatrix}$$

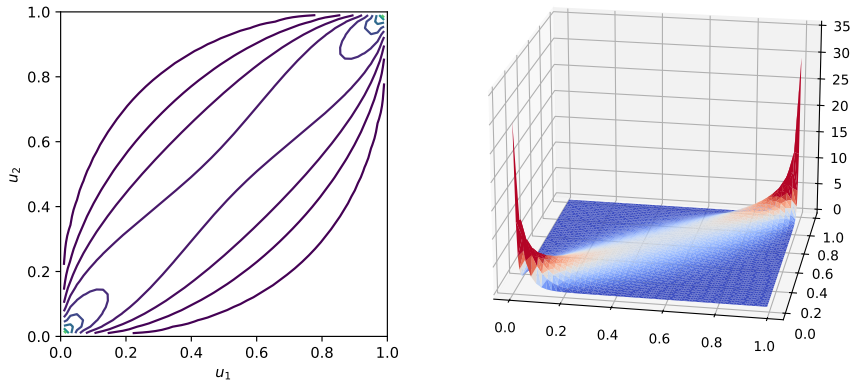
Clearly these are not equal, but in this case, the error is suspected to originate from the estimated joint density. For example, considering X_1 and X_2, we compare the estimated joint copula density and compare to the theoretical reference til et sted hvor gausisk copula står shown in Figure 5.8 and Figure 5.9 respectively.



(a) **(b)**

Figure 5.8: Estimated copula density c with $\rho = 0.9$ corresponding to X_1 and X_2.

The noticeable difference is in the corners $(0, 0)$ and $(1, 1)$ where the theoretical copula density tends to infinity whereas the estimated density has modes at $(0.1, 0.1)$ and $(0.9, 0.9)$. In particular, simply rescaling the copula density in Algorithm 1 does not resemble the theoretical boundary which is a known issue reference til artikel om undershoot peaks og boundary conditions for KDE. A better approach may be to use jackknifing link til afsnit of jackknifing, som også indeholder reference til artikel hvor dette gøres.



(a) **(b)**

Figure 5.9: Theoretical copula density c with $\rho = 0.9$ corresponding to X_1 and X_2.

We note however, that the underlying structure is still captured i.e. that Y_1 and Y_2 covary while Y_3 does not inform Y_1 or Y_2 and vice versa.

We continue with a similar example to the previous one. The key difference is the number of variables and a more complicated correlation structure to test the algorithms further.

Example 5.2. *From Example 5.1 we saw how one could handle some numerical issues. Thus, in this example we shall not bother ourselves with such computations and merely focus on the correlation structure. In particular, we shall sample $\boldsymbol{X}$ from a 10 dimensional*

Proposition 5.5. *Given a bivariate normal distribution* $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *where*

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma^2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Then the mutual information $I(X_1, X_2) = -\frac{1}{2}\ln\left(1-\rho^2\right)$.

Proof. □

$$G_{obs} = \sum_{k\geq 1} G_{dir}^k \tag{5.3}$$

5.3.1 Ensuring convergence and the effect of β

From Network Deconvolution - A General Method to Distinguish Direct Dependencies over Networks - Supplementary Notes and their implementation found at their webpage at https://compbio.mit.edu/nd/ there seem to be an inconsistency between code and theory. In this section, we shall thus investigate from where the discrepancy arises. Initially, from the formulation in Equation 5.3, for the right-hand side to converge, it must have spectral radius at most 1 and to ensure convergence, less than 1. In the latter case,

5.3.2 Robustness to noise

They show that the procedure is robust noise by considering that the observed information matrix is influenced by some noise $N \in \mathbb{R}^{n\times n}$ and characterize the noise by its Euclidean norm $||N||_2 := \sup_{||x||_2=1}||Nx||_2$. They show that

$$||G_{dir} - \hat{G}_{dir}||_2 \leq \gamma + \mathcal{O}\left(\delta^2 + \gamma^2 + \delta\gamma\right)$$

where γ is the spectral norm of N and δ is the spectral norm of $\hat{G}_{obs}$. However, this upper bound can actually be computed when $\delta + \gamma < 1$ and diverges when $\delta + \gamma > 1$. Furthermore, the result can be generalized to other norms than the spectral norm. In particular, the Frobenius norm admits a similar upper bound on the difference. Consider any matrix norm $|||\cdot|||$ for which $|||AB||| \leq$

$|||A||| \cdot |||B|||$. It then follows that

$$\begin{aligned}
\left|\left|\left|G_{dir} - \hat{G}_{dir}\right|\right|\right| &= \left|\left|\left|G_{obs}\left(I + G_{obs}\right)^{-1} - \hat{G}_{obs}\left(I + \hat{G}_{obs}\right)^{-1}\right|\right|\right| \\
&= \left|\left|\left|-\sum_{k\geq 1}\left(-G_{obs}\right)^{k} + \sum_{k\geq 1}\left(-\hat{G}_{obs}\right)^{k}\right|\right|\right| \\
&\leq \sum_{k\geq 1}\left|\left|\left|G_{obs}^{k} - \left(G_{obs} + N\right)^{k}\right|\right|\right| \\
&\leq \sum_{k\geq 1}\sum_{i=1}^{k}\binom{k}{i}|||N|||^{i}\,|||G_{obs}|||^{k-i} \\
&= \sum_{k\geq 1}\left(\left(\,|||N||| + |||G_{obs}|||\,\right)^{k} - |||G_{obs}|||^{k}\right) \\
&= \frac{|||N||| + |||G_{obs}|||}{1 - \left(|||N||| + |||G_{obs}|||\right)} - \frac{|||G_{obs}|||}{1 - |||G_{obs}|||}
\end{aligned}$$

Where the final equality assumes that $|||N||| + |||G_{obs}||| < 1$ and by splitting up the sum into 2 geometric series. However, we also show that the series diverge when $|||N||| + |||G_{obs}||| > 1$ which is not directly apparent as it is a difference geometric series. Namely, by the ratio test, letting $\gamma = |||N|||$ and $\delta = |||G_{obs}|||$ as by supp. notes NetworkDeconvolution-AGeneralMethodtoD...

$$\begin{aligned}
\lim_{n\to\infty}\left|\frac{(\gamma+\delta)^{n+1} - \delta^{n+1}}{(\gamma+\delta)^{n} - \delta^{n}}\right| &= \lim_{n\to\infty}\left|\frac{(\gamma+\delta)\left(1+\frac{\gamma}{\delta}\right)^{n} - \delta}{\left(1+\frac{\gamma}{\delta}\right)^{n} - 1}\right| \\
&= \lim_{n\to\infty}\left|\delta + \gamma\frac{\left(1+\frac{\gamma}{\delta}\right)^{n}}{\left(1+\frac{\gamma}{\delta}\right)^{n} - 1}\right| \\
&= \lim_{n\to\infty}\left|\delta + \gamma\frac{1}{1 - \left(1+\frac{\gamma}{\delta}\right)^{-n}}\right| \\
&= |\gamma + \delta| = \gamma + \delta
\end{aligned}$$

as $\gamma, \delta > 0$ unless they are the zero-matrix in which case the above is nonsensical from a perspective of interest. The above shows that indeed the above result diverges when $\gamma + \delta > 1$.

Thus, for the spectral norm, denoted by $|||\cdot|||_2$, one simply needs the fact that it is sub-multiplicative (Definition 5.2) which follows from the fact that the spectral norm is induced by the l_2 norm (see Definition 5.1) on $\mathbb{R}$, and hence

$$\sup_{||x||_2=1}||ABx|| \leq ||A||_2 \cdot \sup_{||x||_2=1}||Bx|| = ||A||_2\,||B||_2$$

which by definition means that $||AB||_2 \leq ||A||_2 ||B||_2$. The Frobenius norm $|||\cdot|||_F$ is also sub-multiplicative and depending on the use case may be very useful.

5.3.3 Quick made up example

As a thought up experiment for the underlying

Appendix A

Stuff

This appendix is full of stuff ...

Bibliography

[Bal07] Gianfranco Balbo. *Introduction to Generalized Stochastic Petri Nets*, pages 83–131. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[fCM] The Association for Computing Machinery. Acm turing award honors founders of automatic verification technology.

[Vic21] Margarida L. C. Vicente. A benchmark model to generate batch process data for machine learning testing and comparison. 2021.