Something 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

Not Real

Jonas Bruun Hubrechts

Acknowledgements

I would like to thank my....

Contents

Summary (English) i

Summary (Danish) iii

Preface v

Acknowledgements vii

1 Time to Level of Brownian motion 1

1.1 Arrivals of batches . . . 1

1.1.1 Joint distribution of Brownian and its running maximum 2

1.1.2 Joint distribution with drift and arbitrary variance . . . 4

1.1.3 Distribution of maximum of Brownian motion with drift . 6

1.1.4 Cumulative distribution of maximum . . . 6

1.1.5 Distribution of time to level . . . 7

1.1.6 MGF . . . 8

2 Problemformulering / Introduktion 13

3 Ideer til hvad der skal laves 15

4 Data 17

4.1 Adding of solids and materials . . . 18

4.2 Centrifugation . . . 18

4.3 Product transfer . . . 18

4.4 Reaction . . . 18

4.5 Post reaction . . . 18

4.6 Cooling . . . 18

4.7 Material transfer . . . 19

4.8 Cleaning operations 20

A Stuff **27**

Bibliography **29**

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

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

and 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

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

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 $Y_s = -\tilde{mu}$ 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

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

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

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

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

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

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.

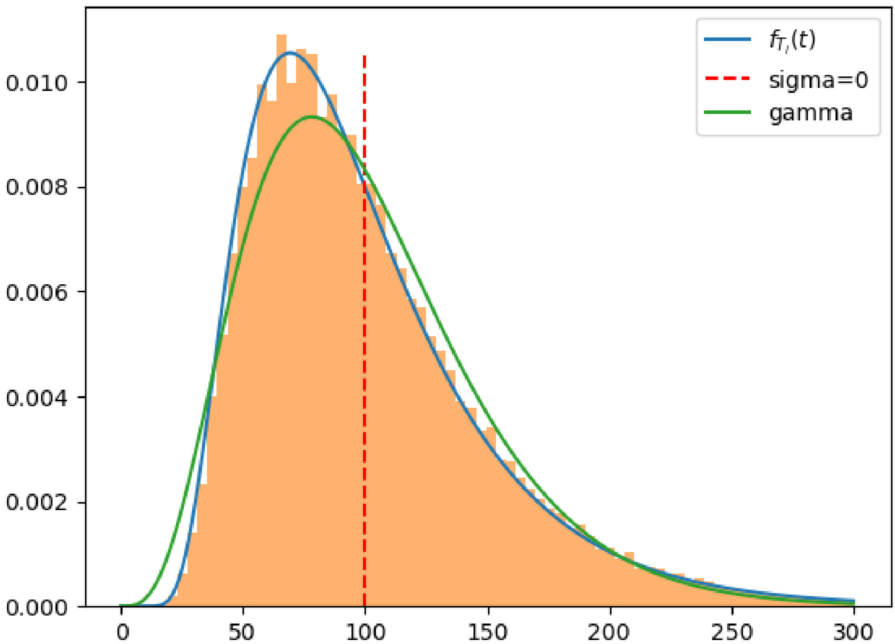

Figure 1.1: Example of simulation and actual distribution. The marked $\sigma = 0$ shows the limit as $\sigma \to 0$ corresponding to no noise on the input flow

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

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

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]

Chapter 4

Data

TIME IS IN HOURS

The data used consists of six cycles. This means that the total data set consists of six simulation runs. However, each of the runs contains many batches in sequence. The following table lists the number of batches in each of the simulations and some basic statistics

Cycle	#batch	μ_{batch}	σ^2_{batch}
A	66	14.776	3.641
B	64	15.644	3.915
C	61	17.714	2.330
D	60	18.069	6.922
E	60	18.088	9.613
F	63	17.227	7.766

Table 4.1: Per cycle batch statistics

Each batch is comprised 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).

4.1 Adding of solids and materials

This part of the process corresponds to events tagged with ID 1 through 4

4.2 Centrifugation

ID 5

4.3 Product transfer

ID 6

4.4 Reaction

ID 7

4.5 Post reaction

ID 8

4.6 Cooling

ID 9

4.7 Material transfer

ID 10

4.8 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
A	65	1.113	3.067	1.917	0.269
B	63	1.324	1.751	1.566	0.00883
C	9	1.544	3.306	2.153	0.277
D	10	1.474	2.009	1.581	0.0212
E	10	0.827	1.584	1.465	0.0462
F	10	0.748	1.610	1.466	0.0595

Table 4.2: Per cycle cleansing statistics

To verify these observations and potentially discovering more important facts of their probability distributions, histograms are plotted in the following Figure 4.1. 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.

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 4.2 below.

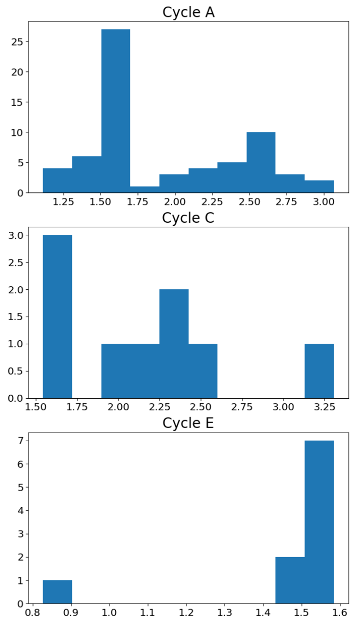

Figure 4.1: Each of the 6 cycles, cleaning operations histograms.

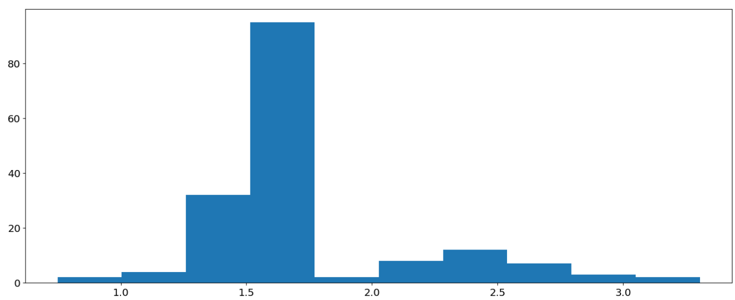

Figure 4.2: Combined cleaning operations histograms.

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 4.3. 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 4.4.

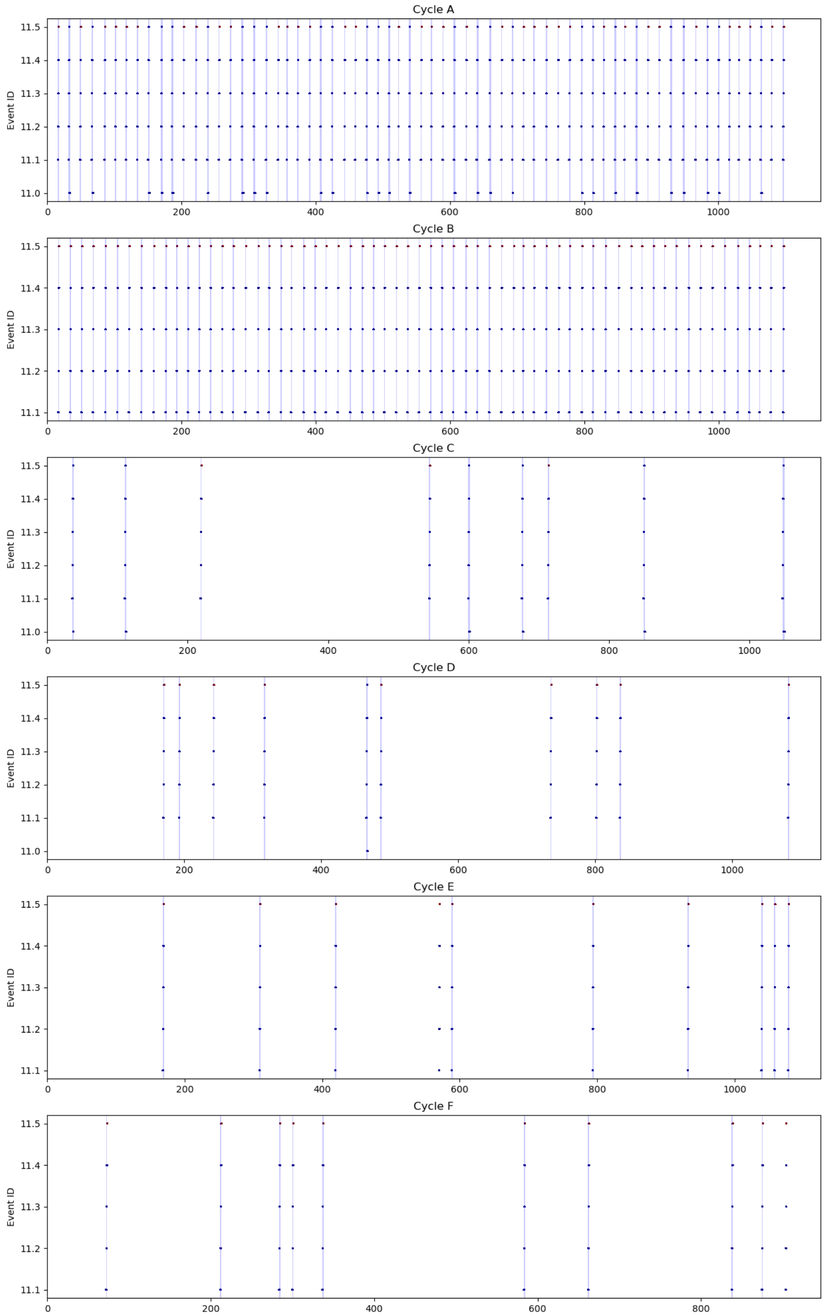

Figure 4.3: Each of the 6 cycles, cleaning (corresponding to `BatchID = 0`). Each *(*Cleaning Procedure), CIP, is highlighted with an opaque interval (the blue rectangles). The dots marked with red (only ID 11.5, but not all of these are red), is if the Cleaning ID is 0.

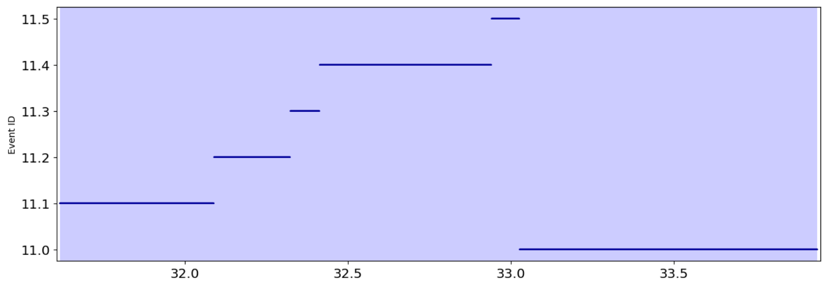

Figure 4.4: A single blue rectangle zoomed in

It is observed that the observations marked with red in figure 4.3 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 4.2, 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 4.3

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

Table 4.3: 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 4.4: 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 4.5 and note that all the ACF stay within the 95% confidence interval.

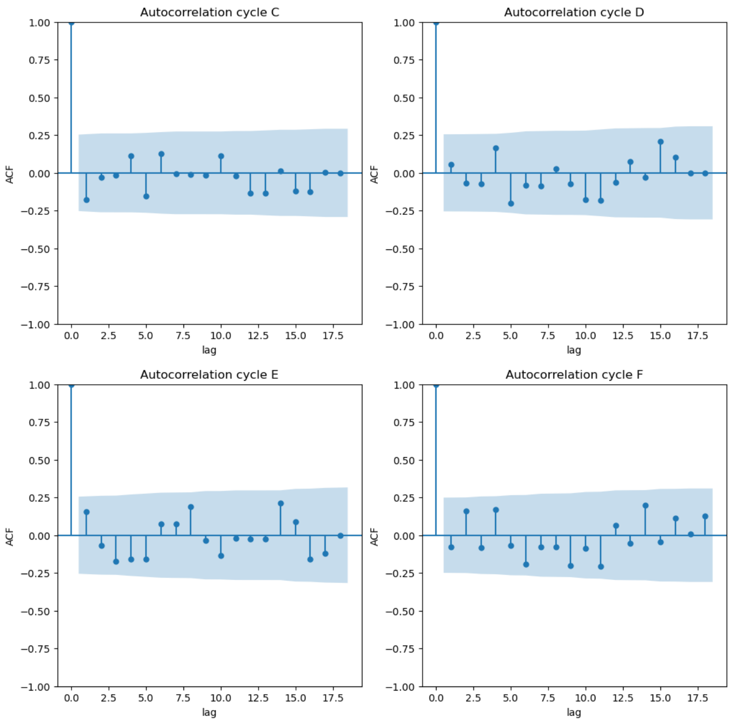

Figure 4.5: Autocorrelation function for each of the final 4 cycles. As can also be seen from this there seem to be no information to be gained of C_i from C_{i-1}.

Appendix A

Stuff

This appendix is full of stuff ...

Bibliography

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