Meeting 29

Shows results, were very interesting!

Behavior makes sense as dropping for *a* but increasing with *T*, compare 0.99^51 to 0.999^501 for example.

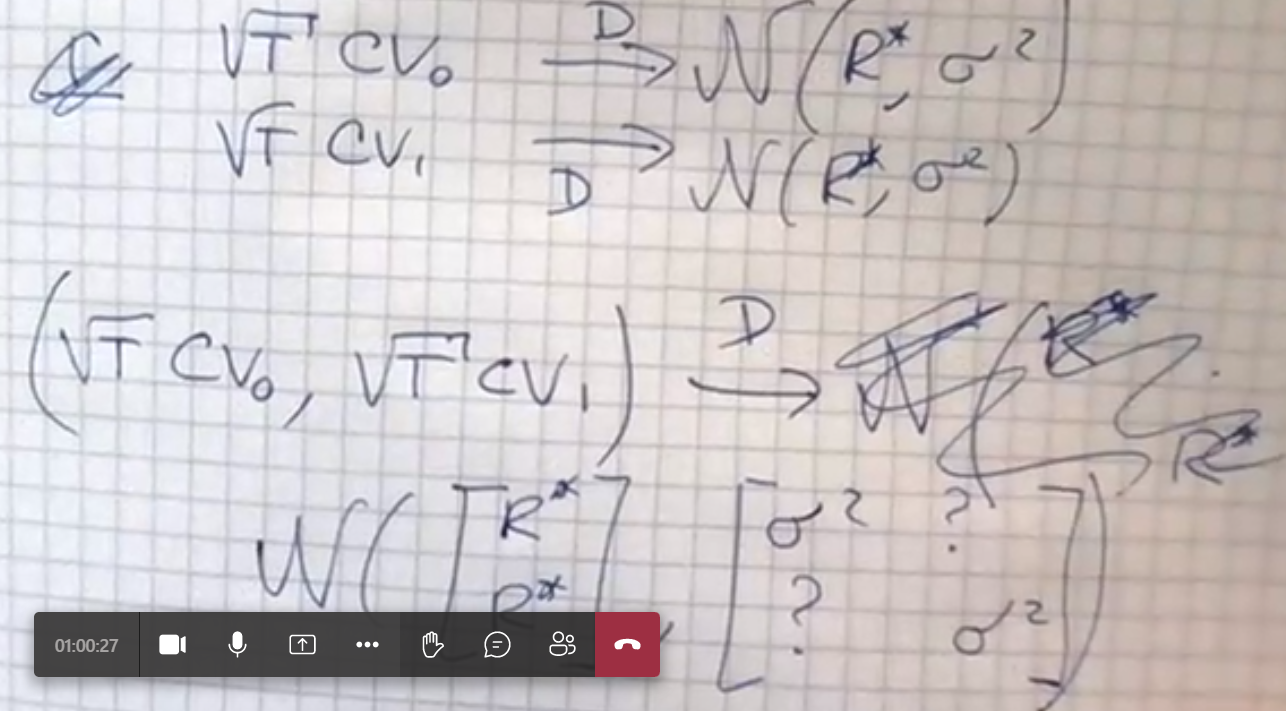
Check for T = 5001

Hypothesis: as 0.99^5001 = 0.9^501, we expect to see the success probability of 0.9 at 501 to be equal to 0.99 at 5001. Indeed, both have ~0.85.

Now, 0.9999^5001 is about 0.999^501, which is approximately 0.63 for a = 0.999 and T = 501, **but also** for a = 0.9999 and T = 5001, interesting!

Compare Risks

Understand Plots



Both scores for T infinitely large converge in distribution to a = 0, and covariance

R\* is 1 (risk of a).

Sigma, fischer information.

Next two weeks out.

27th 6th may

Next week 10:30 AM next teusdat.

Start writing, introduction.

Thesis = Presentation

Six pieces of paper

Onion

CV\_0 will have distribution exactly chisq\_T-1

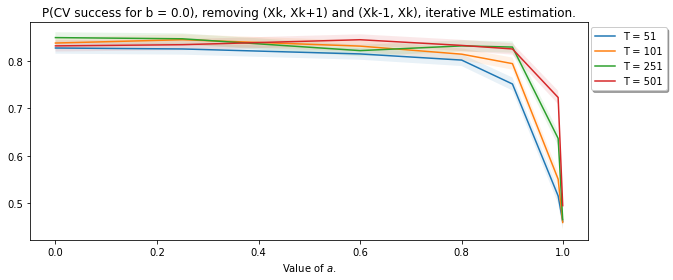
CV\_0 will have distribution approximately N(T – 1, 2 (T – 1)) => divide by T – 1

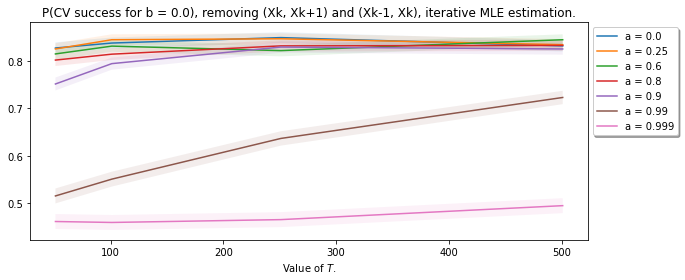
CV\_1 will have distribution approximately

CV\_diff will have a mean of – 1 / (T – 1).

# Reinvestigating the estimates

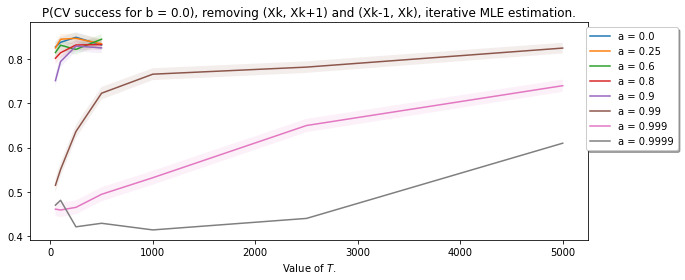
Small mistake, I used the reiterated ML estimate for CV(0) when using a = a0. However, this should be the exact b estimate based on a0. Luckily, this did not change much. Some excerpts:





## Investing behavior for much larger *T*.

Hypothesis, for e.g. T = 5001, the P(CV success) will be much larger for a = 0.999 and a = 0.99.

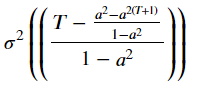


We see that for larger *T*, the values indeed “grow”. We also see an interesting trend. If we add another 9 to the end (so 0.999 compared to 0.9999), to get approx. the same probability of success, we need 10 times as much timesteps. As 0.999^T is approximately 0.9999^(10T) = 0.9999^10^T = 0.999^T, which is quite interesting.

# Deriving

We will take a look at an – a.

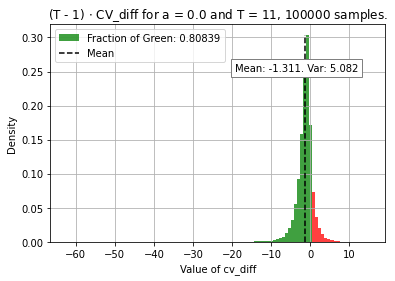
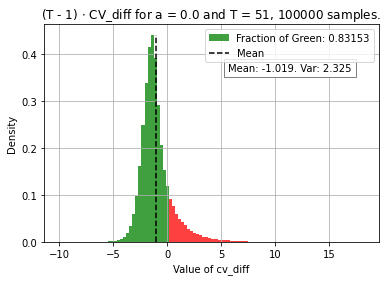
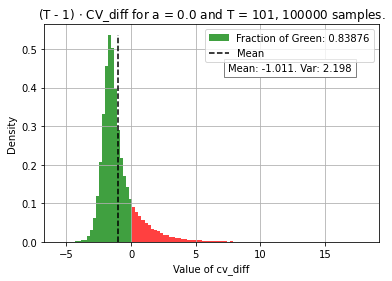
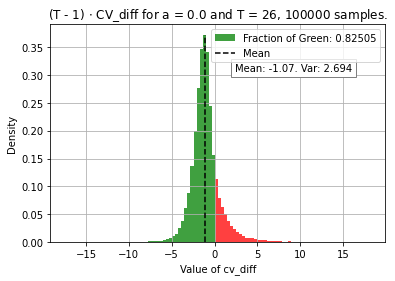
The Fisher Information is E[sum(X\_i^2)], which has a closed form expression, but is asymptotically equal to (T – 1) / (1 – a^2)

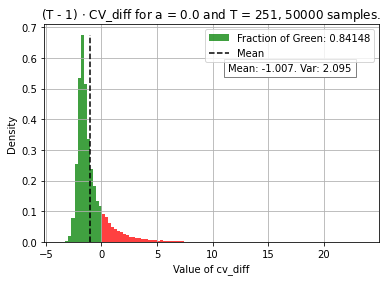
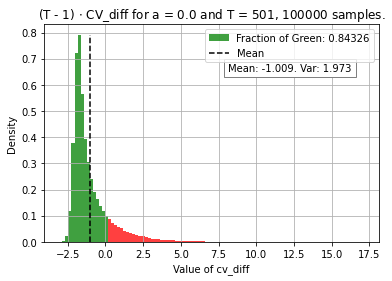


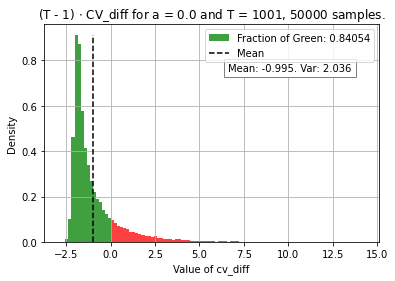
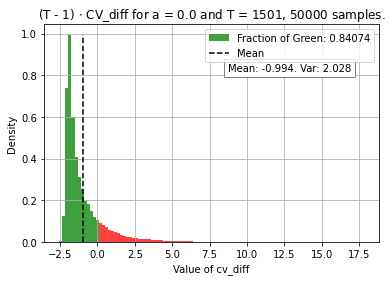
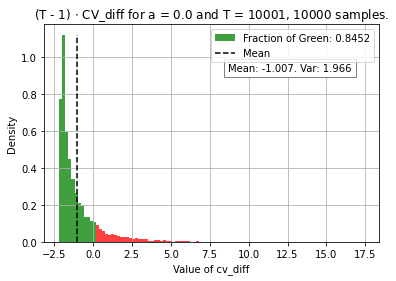
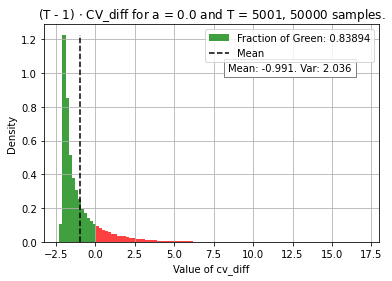
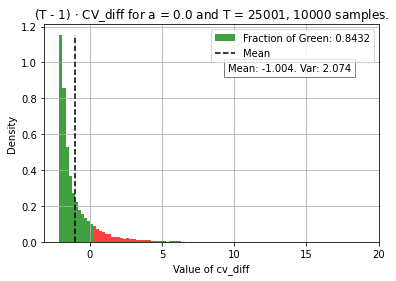
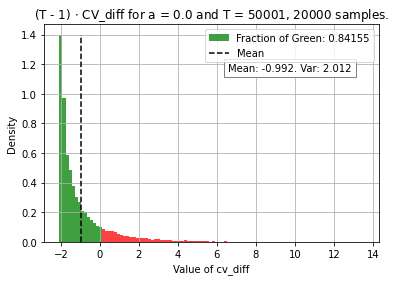
Asymptotically 1 / (1 – a^2), so 1 / Fisher Is 1 – a^2

# Asymptotic Behavior of CV\_diff

**Histograms as a function of *T*, start with *a = 0.0*.**







We see that the value of CV\_diff has on average a value of – 1 / (T – 1), which is due to the covariance of the LOOCV estimate of a. Nevertheless, the distribution is the difference of two heavily correlated random variables who are both normally distributed in the limit. Nevertheless, their difference is far from normally distributed, with seemingly a minimum of -2, which is interestingly also the median.

# Investigate behavior for *a* close to one.

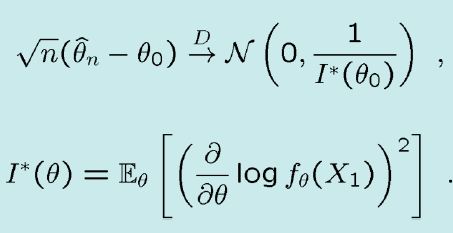
Idea, perhaps it is not too bad to have smaller probability when *a* is large, as both estimates will be poor. Is this the case though? We inspect the empirical risk of both models, and see nothing strange unfortuntaly, we do see that the expected risk deviates a lot, which makes sense as for CV(0) we use the true *a*, so that one has an expected error close to perfection.

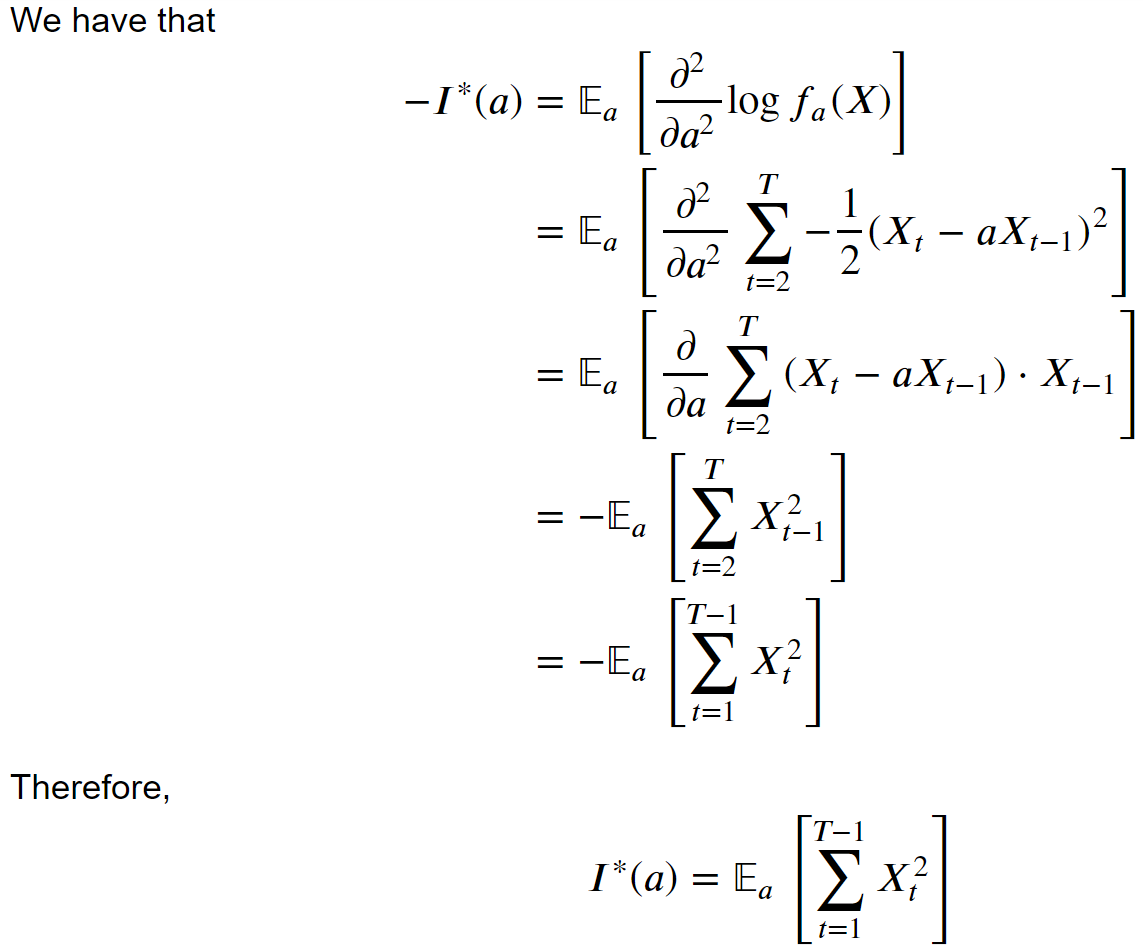
*What I think is the reason*.

For *a* close to one, we see that CV(1) can achieve a much lower empirical risk, as we can tweak *a* and this free parameter is especially more useful for values of *a* close to one. This difference in empirical risk translates to a larger difference in cross validation risk, meaning that CV(1) < CV(0) more often, resulting in a smaller probability of success.

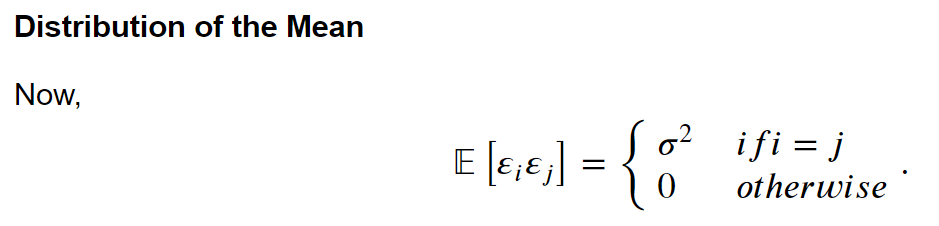
# Deriving

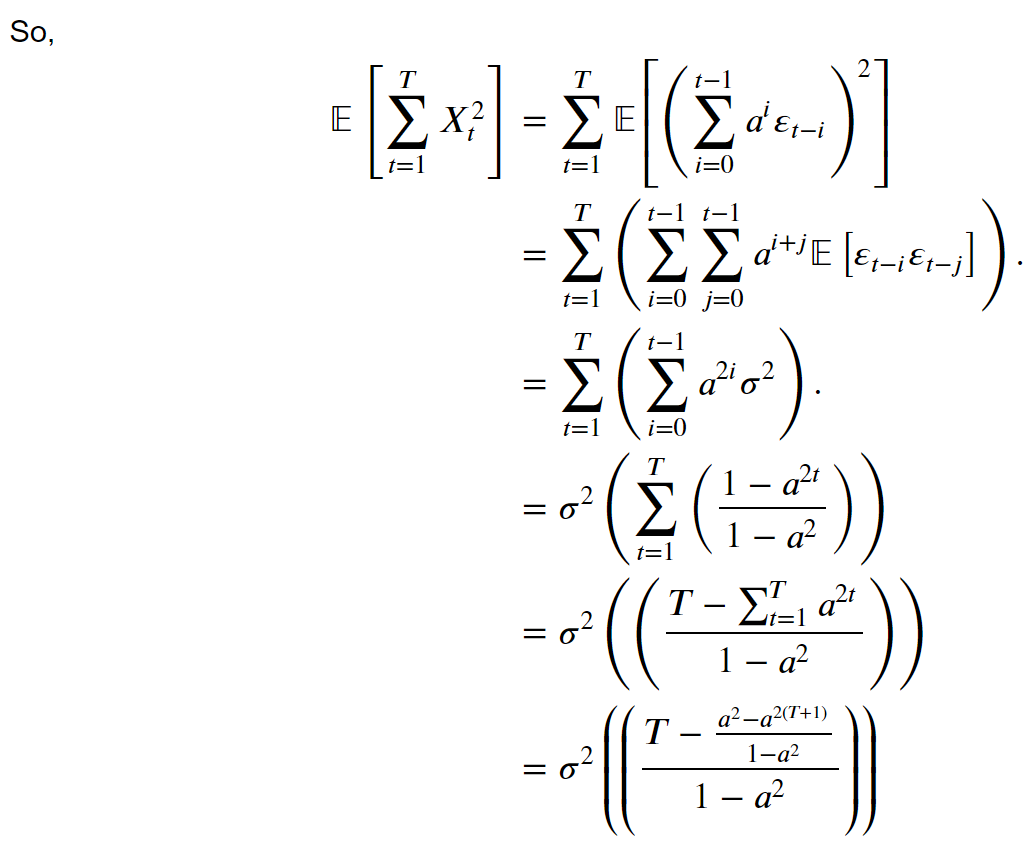
Used the lecture notes:





Now, for any value of *a*, we have that





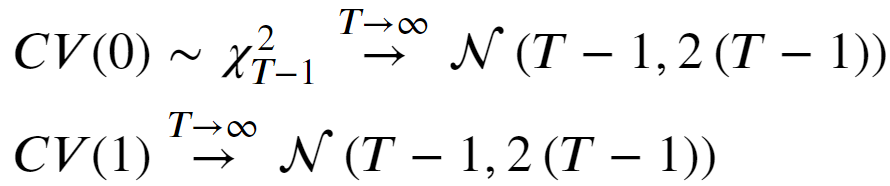
For -1 < a < 1, we know by its tailorexpansion that this value over T converges linearly to **sigma^2 / (1 – a^2).**

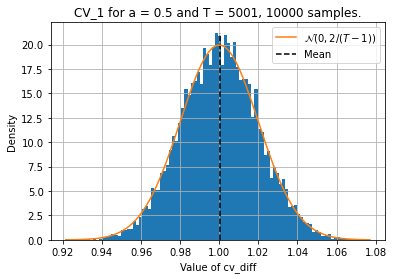
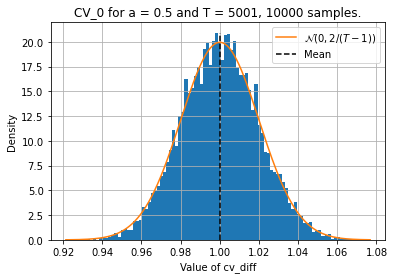
So, we know the distribution of the estimates. Furthermore, the mean of the differences is also known. Indeed, we have that

# 

However, this only holds for **large enough *T***. Especially for *a* close to one, the value for *T* needs to be quite large:

Now, we know that CV(0).





## Plot of difference in estimators.

# Write about bootstrapping

## Bootstrapping to create 99% OMP CI.

Not really CI for LS estimates, but quite similar. Source says that the ARB works very well, gives nice procedure on how to do it.

## Bootstrapping to verify recoverability of OMP coefs.

Does not work super well on few samples, which is unfortunate, but idea seems to work well for moderately large data.

# Write about L1-Regularization

Use Lasso on dense recovered solution. This yields a certain sparse matrix. Re-estimate recovered coefficients using OLS again to get the best estimates.