Linear Covariance Models - computations

Piotr Zwiernik, Caroline Uhler, Donald Richards

March 2016

In this document we present basic computations for the paper "Maximum likelihood estimation for linear Gaussian covariance models".

Code for Figure 1

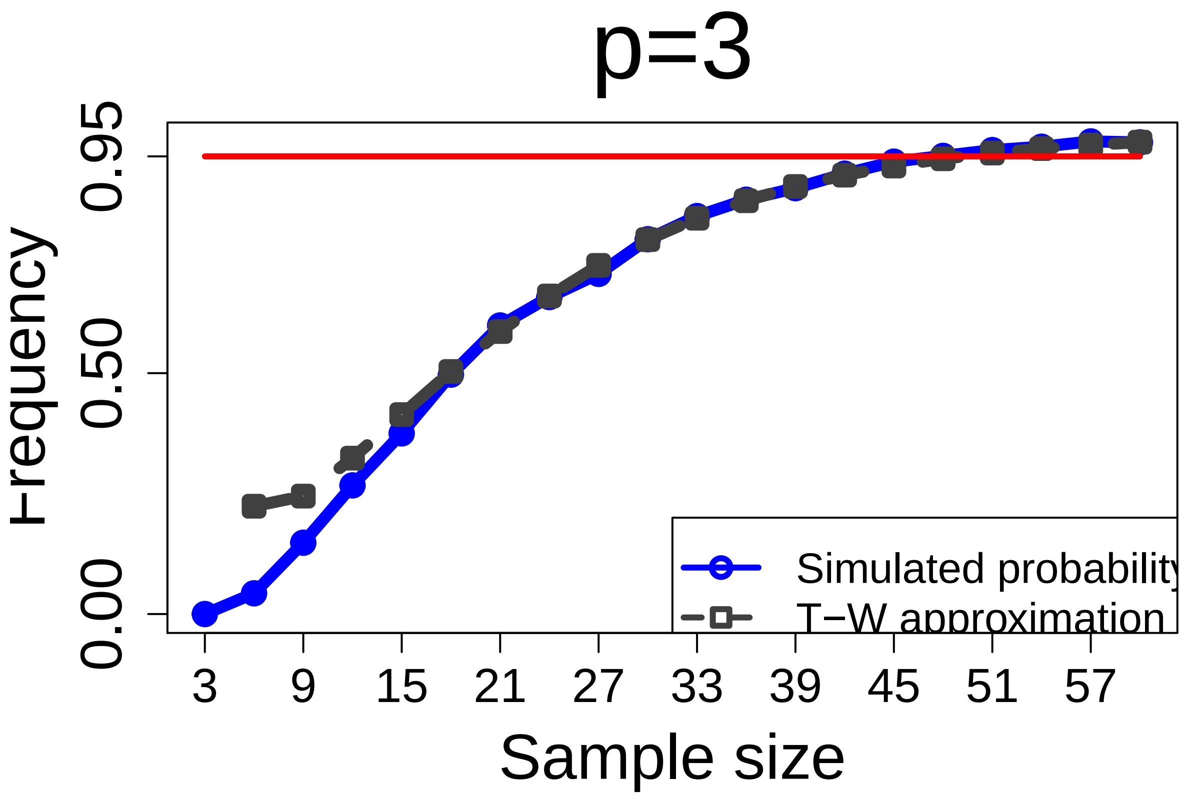
We first want to compare $\mathbb{P}(W_{n-1} > n/2)$ with its approximation given by the Tracy-Widom approximation. We are going to use R package RMTstat. We present computations for dimension $p = 3$. In this case the sample n will vary between 3 and 60. The following code produces theoretical probabilities of $\mathbb{P}(W_{n-1} > n/2)$ coming from the Tracy-Widom approximation. The code will work for any other given value of p.

```
library(RMTstat)
p <- 3
matpwt <- rep(0,20)
nn <- p*(1:20)
for (n in nn) {
    # corrected Mu's approach
    mu <- (sqrt(n-3/2)-sqrt(p-1/2))^2
    sig <- (sqrt(n-3/2)-sqrt(p-1/2))*(1/sqrt(p-1/2)-1/sqrt(n-3/2))^(1/3)
    tau <- sig/mu
    nu <- log(mu)+tau^2/8
    x <- (log(n/2)-nu)/tau
    matpwt[which(nn==n)] <- ptw(-x)
  }
```

Now we explicitly estimate these probabilities using a simple Monte Carlo approach with 1000 iterations. This is less than used to produce Figure 1 in the paper so the plot may be less smooth. To reproduce the same result set `N<-10000`. There was no need to fix a random seed because the variance of the estimators is very small.

```
library(MASS)
library(matrixcalc)
N <- 1000 # number of iterations in our simulation
samples <- p*(1:20) # sample size
# set the true covariance matrix C to be the identity matrix (white Wishart)
C <- diag(p)
# check how often 2S-I>0 (or equivalently lambda_min(W)>n/2)
in.region <- rep(0,length(samples))
for (n in samples){
  yess <- 0
  for (i in 1:N){
    dat <- mvrnorm(n, rep(0,p), C)
    S <- (n-1)*cov(dat)/n
    yess <- yess +1*(is.positive.definite(2*S-C))
  }
  # we get a simple Monte-Carlo estimate
  in.region[which(samples==n)] <- yess/N
}
```

Now we plot both together adding the `0.95` line:



Code for Figure 2

In the table in Figure 2 we check the minimal sample size that guarantees that $\mathbb{P}(W_{n-1} > n/2) > 0.95$. For large p, this minimal n will lie somewhere between $11 \cdot p$ and $12 \cdot p$. This can be checked for any fixed p using the following code.

```
p<-1000
# restrict to sample sizes in the interesting interval (for small p may not be enough)
nn <- seq(11*p,12*p,1)
resl <- rep(0,length(nn))
for (n in nn) {
    # corrected Mu's approach
    mu <- (sqrt(n-3/2)-sqrt(p-1/2))^2
    sig <- (sqrt(n-3/2)-sqrt(p-1/2))*(1/sqrt(p-1/2)-1/sqrt(n-3/2))^(1/3)
    tau <- sig/mu
    nu <- log(mu)+tau^2/8
    x <- (log(n/2)-nu)/tau
    resl[which(nn==n)] <- ptw(-x)
  }
# and this is our minimal n
(11*p+min(which(resl>0.95)))
```

```
## [1] 11759
```

Proof of Proposition 3.5

The proof of Proposition 3.5 depends on simple computations in `Mathematica`. We provide the code:

```
mu:=(Sqrt[g*p-3/2]-Sqrt[p-1/2])^2
sig:=(Sqrt[g*p-3/2]-Sqrt[p-1/2])*((1/Sqrt[p-1/2])-(1/Sqrt[g*p-3/2]))^(1/3)
tau:=sig/mu
nu:=Log[mu]+(tau^2)/8
foo:=(nu-Log[g*p/2])/tau
Assuming[g>1,Limit[foo,p->Infinity]]
```