MATH 559 : BAYESIAN THEORY AND METHODS

SELECTION WITH THE NORMAL MODEL: EXPLORING DIFFERENT PRIORS

In the numerical examples, we examine the impact of changing the value of prior mean η_0 in the Normal model with known variance.

- True value: $\mu_0 = 2$
- Prior: $\pi_0(\mu) \equiv Normal(\eta_0, 1)$ (i.e. $\lambda_0 = 1$).
- η_0 varying from -5 to 5
- Sample size: $n = 20$

The following code simulates the data and computes the values of T_n, G_n, C_n and W_n across 1000 replicates.

```
set.seed(2134)
n<-20;nreps<-1000
mu0<-2;sigma0<-1
eta<-0; lambda<-1
dsq<-function(xv,ev,lv){
    dv<-xv*0
    tot<-sum(xv)
    dv<-xv-((tot-xv)+ev*lv)/length(xv-1+lv)
    return(sum(dv^2))
}
ssq<-function(xv){
    return(sum((xv-mean(xv)^2)))
}
variance.term<-function(xv,ev,lv,N=10000){

    #Monte Carlo calculation
    en<-(sum(xv)+ev*lv)/(length(xv)+lv)
    lnv<-length(xv)+lv
    mu<-rnorm(N,en,sqrt(1/lnv))
    d<-outer(xv,mu,'-')

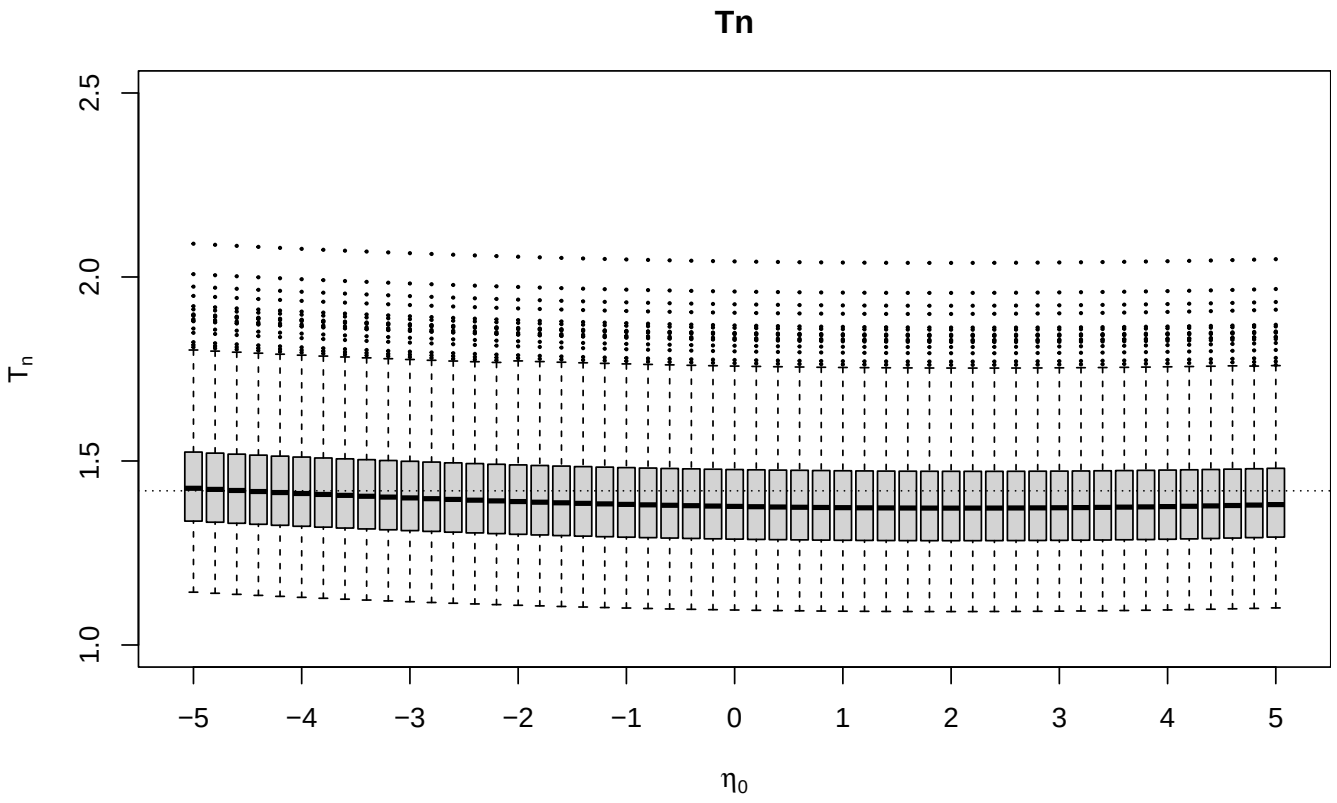
    return(mean(apply(dnorm(d,log=T),1,var)))
}
Y<-matrix(rnorm(n*nreps,2,1),ncol=n)

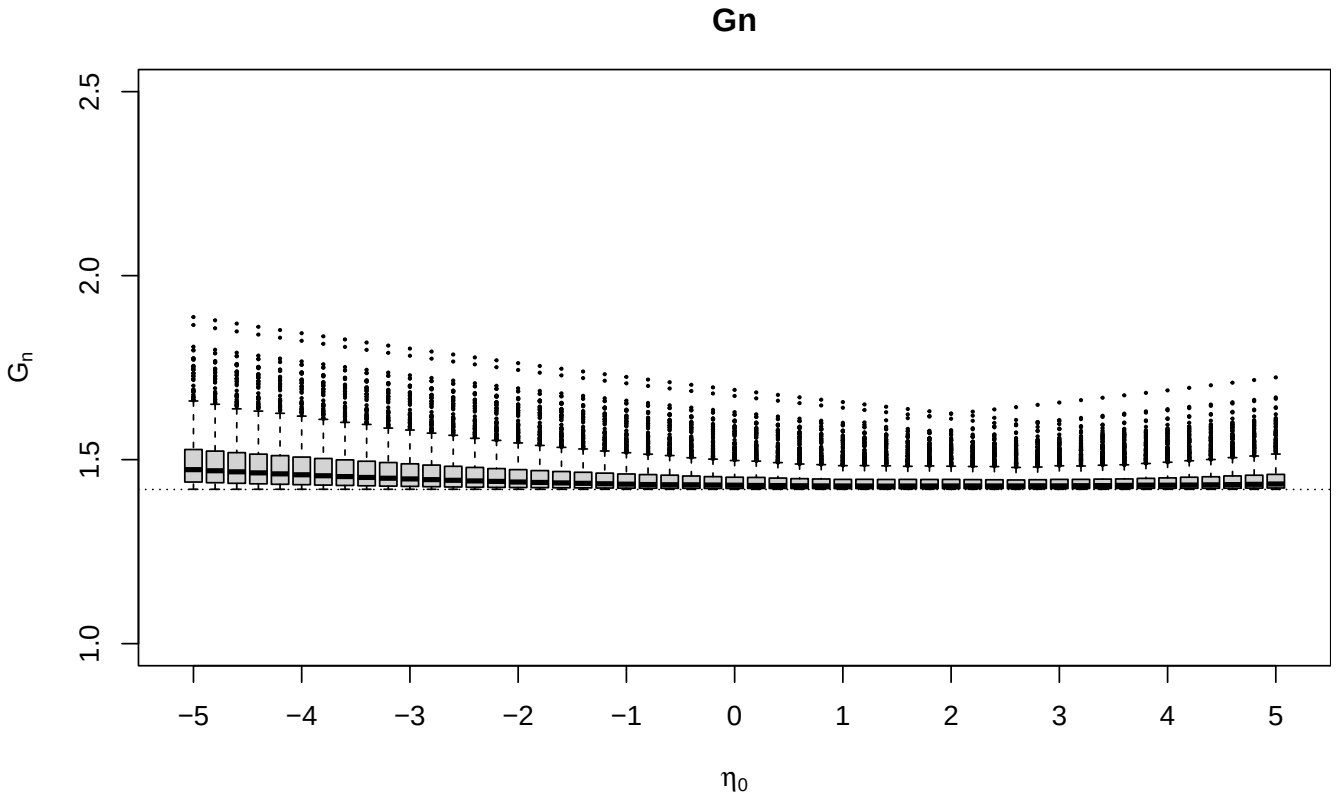
etavec<-seq(-5,5,by=0.2)
lambda.n<-n+lambda; lambda.n1<-lambda.n/(1+lambda.n)
lambda.ni<-n-1+lambda; lambda.ni1<-lambda.ni/(1+lambda.ni)
const<-0.5*log(2*pi)-0.5*log(lambda.n1)
stat.mat<-array(0,c(ncol=length(etavec),nreps,4))

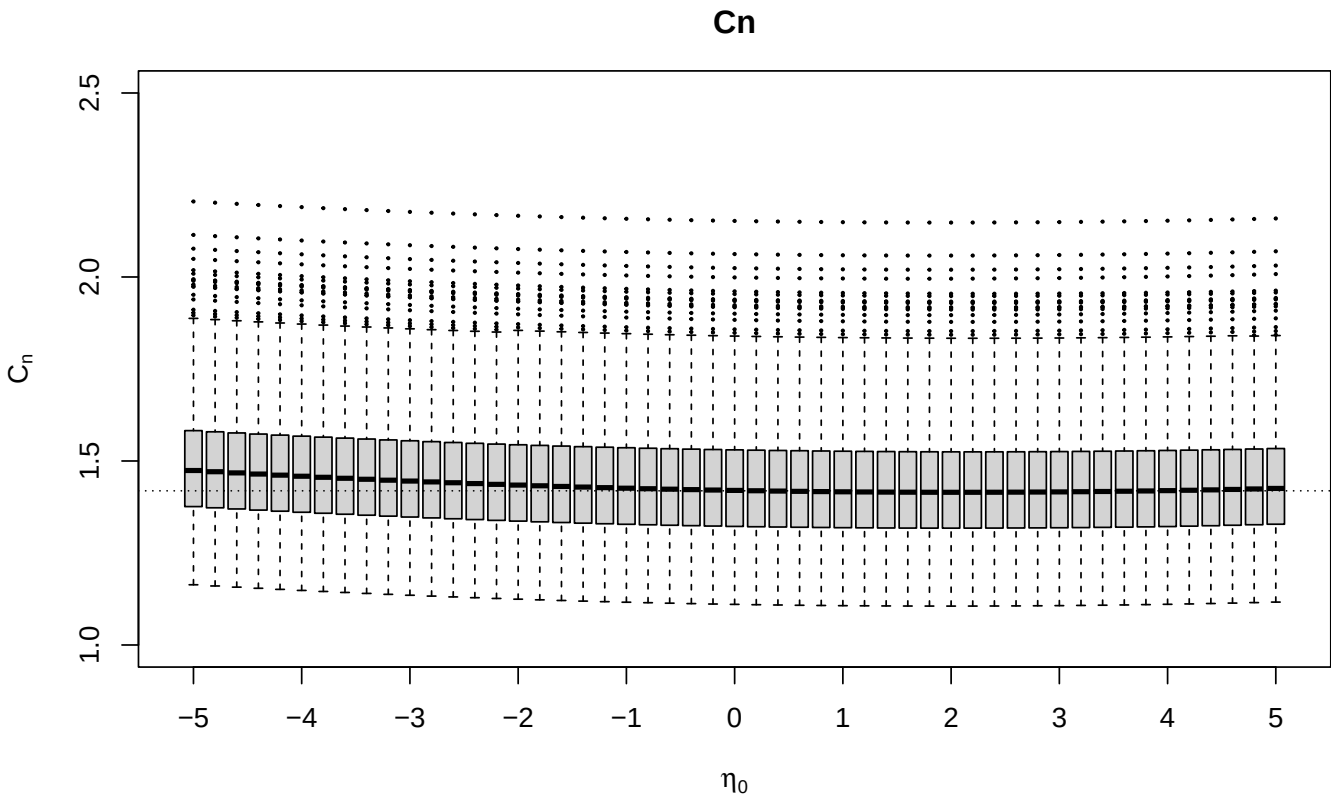
for(il in 1:length(etavec)){
    eta<-etavec[il]
    eta.n<-(n*apply(Y,1,mean)+eta*lambda)/lambda.n
    Tn<-const+0.5*lambda.n1*apply((Y-eta.n)^2,1,sum)/n
    Gn<-const+0.5*lambda.n1*(1+(eta.n-mu0)^2)
    Cn<-const+0.5*lambda.ni1*apply(Y,1,dsq,ev=eta,lv=lambda)/n
    Wn<-Tn+apply(Y,1,variance.term,ev=eta,lv=lambda)
    stat.mat[il,,]<-cbind(Tn,Gn,Cn,Wn)
}
```

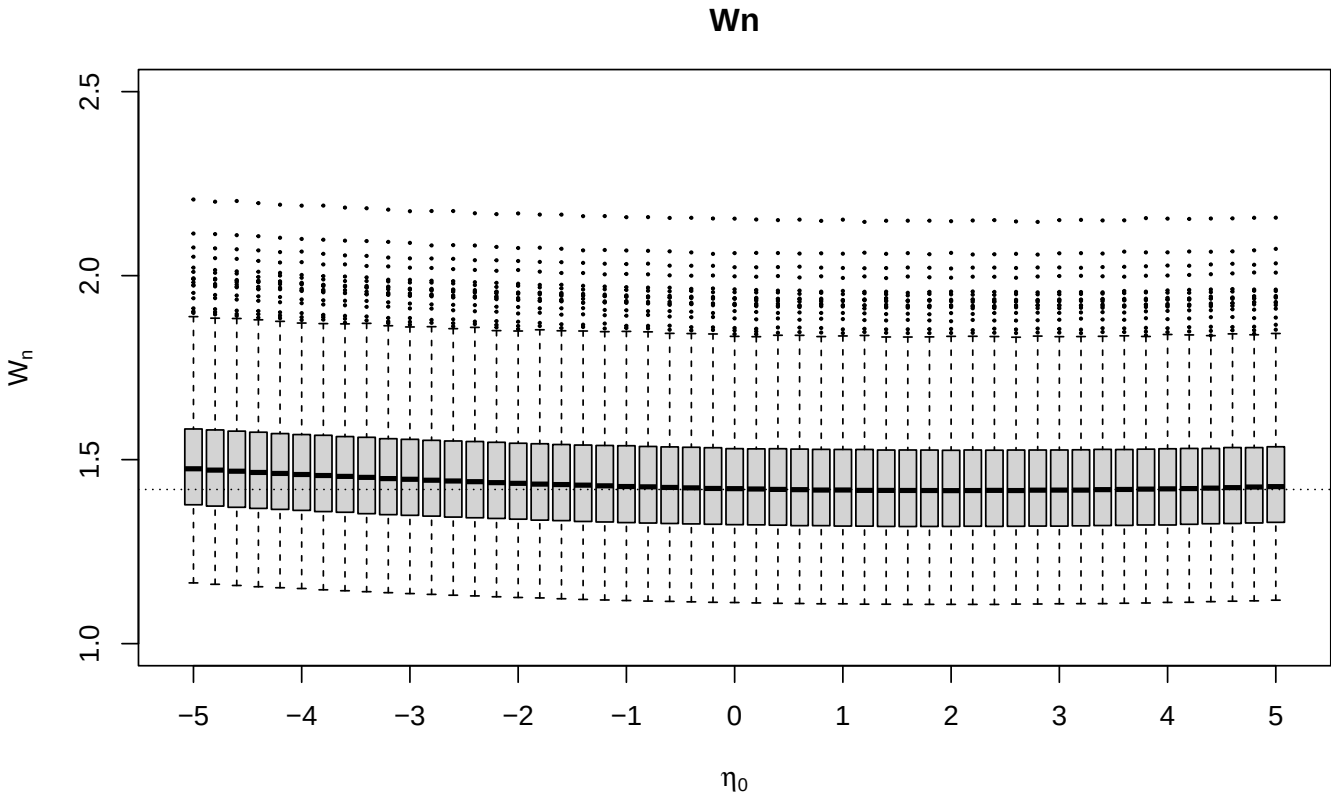
We now plot boxplots of the results for the sets of computed statistics over the selected range of η_0 values. Note that theory suggests that for large n all of the statistics should converge to the same limiting value

$$\frac{1}{2}(\log(2\pi) + 1) \doteq .$$

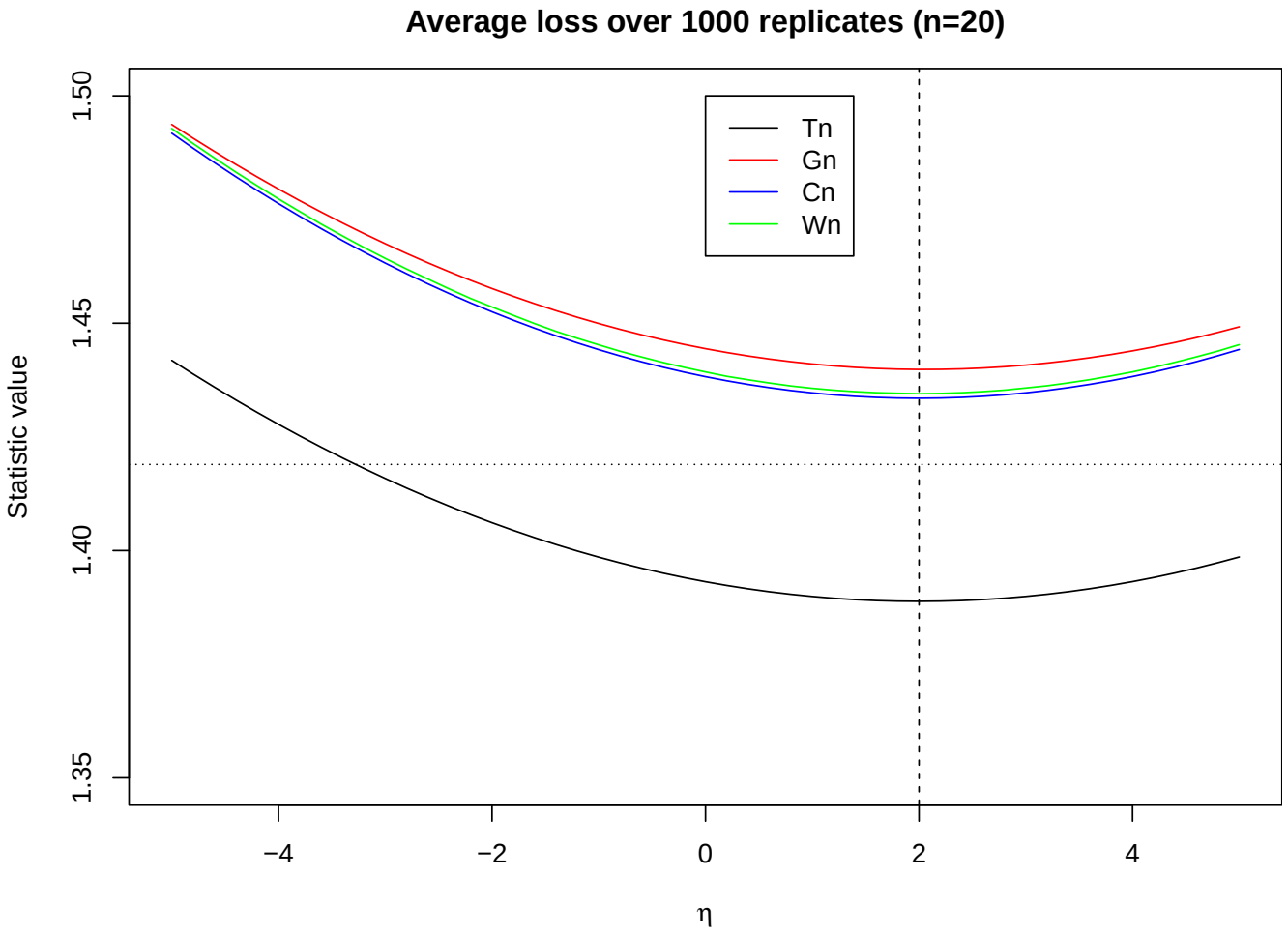

Tn
2.5
2.0
1.5
1.0
T_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Gn
2.5
2.0
1.5
1.0
G_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Cn
C_n
2.5
2.0
1.5
1.0
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Wn
W_n
2.5
2.0
1.5
1.0
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0

The replicates can be used to estimate the expected values of these quantities by taking sample averages. We see that the limiting ($n \longrightarrow \infty$) value (the horizontal dotted line) is approached in different ways by the different statistics.

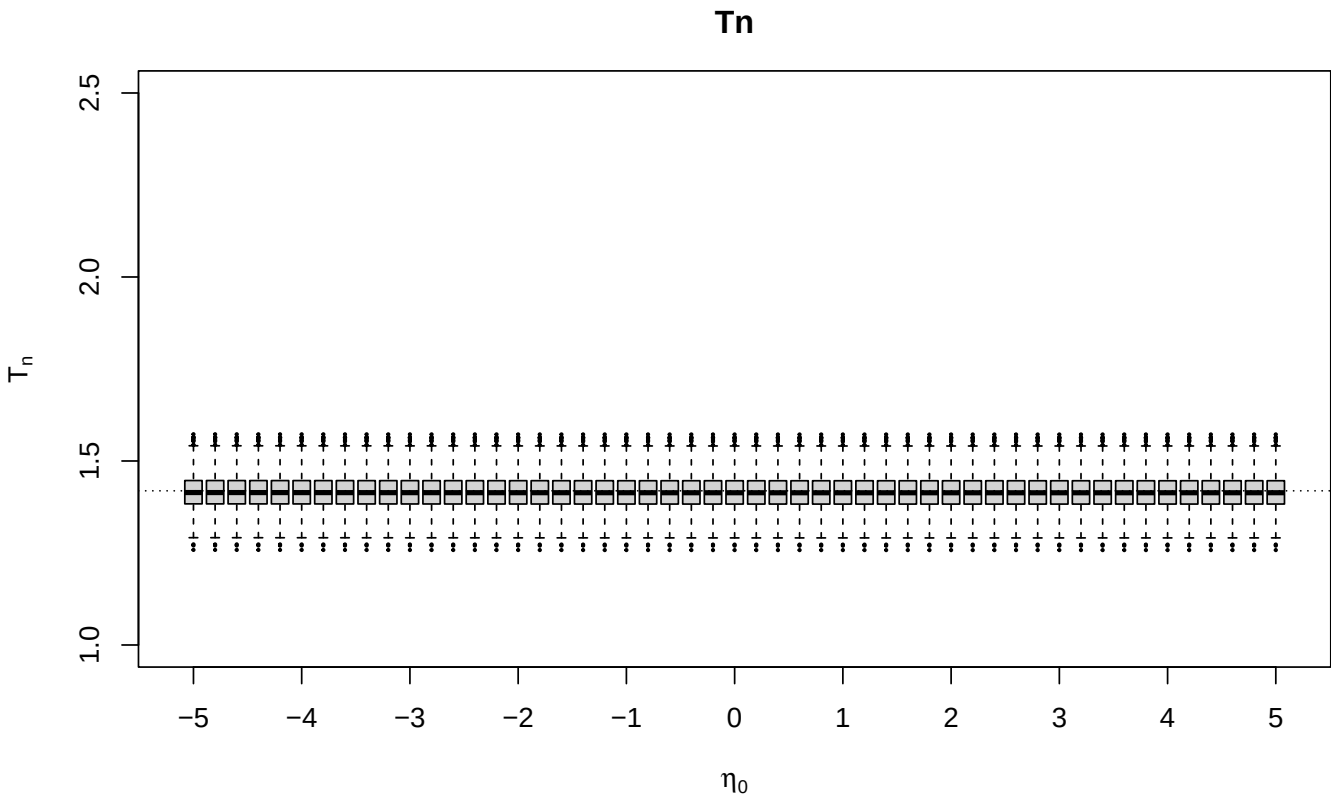


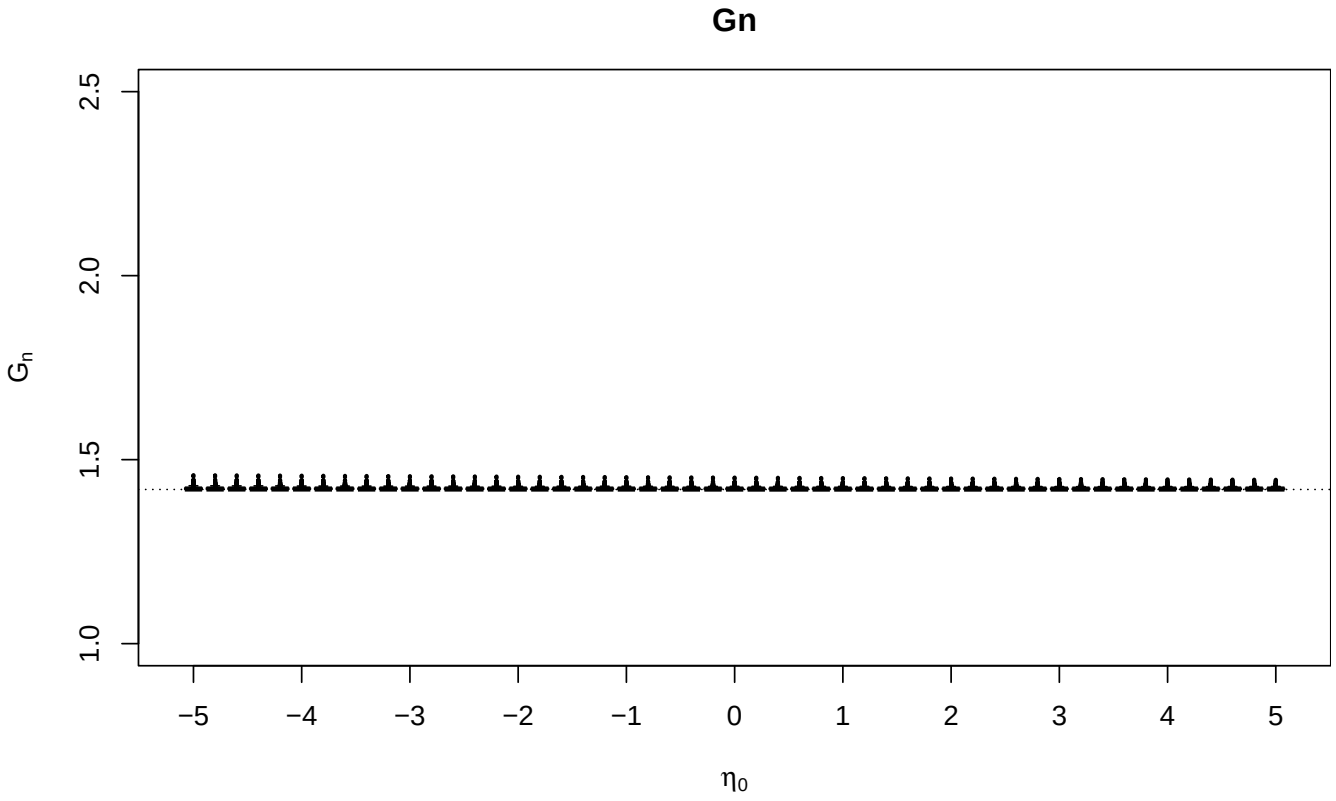
Larger sample size: $n = 200$

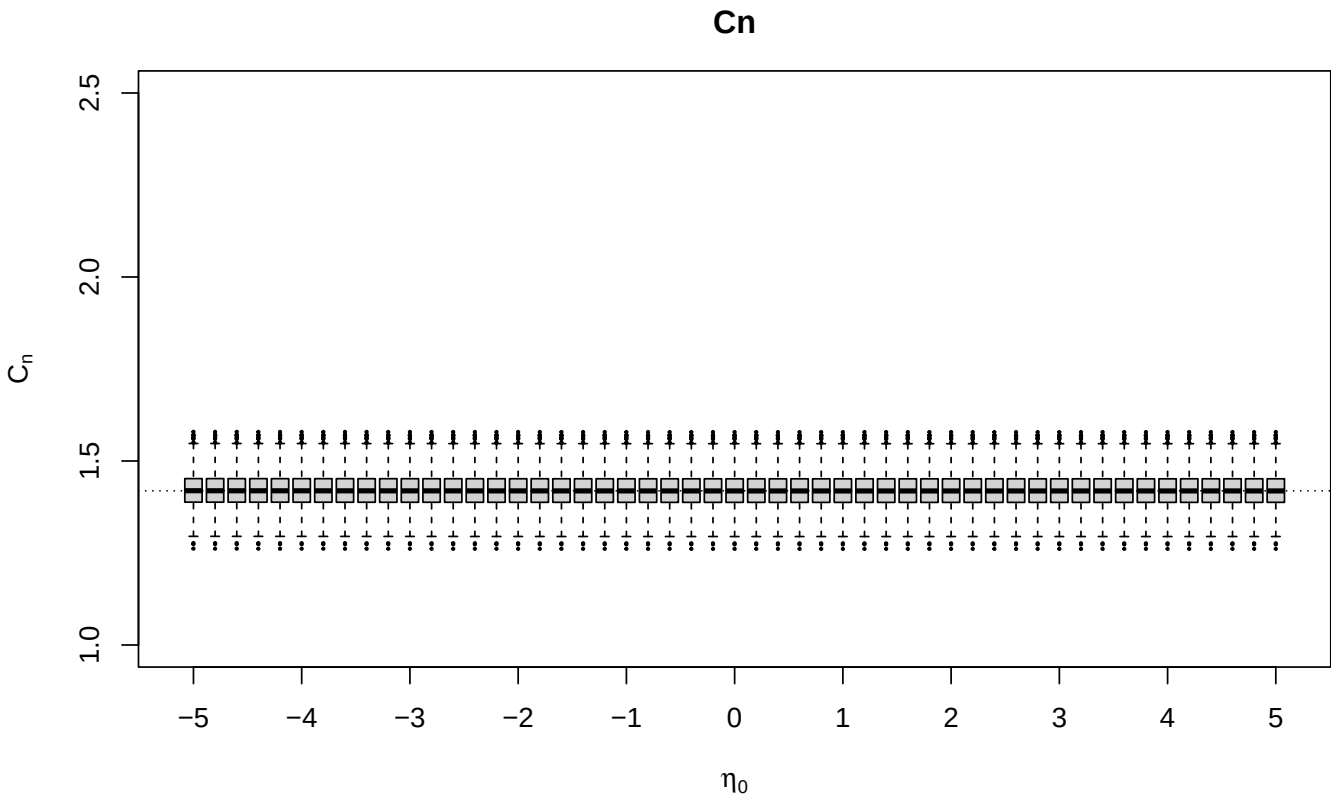
```
set.seed(2134)
n<-200;nreps<-1000
mu0<-2;sigma0<-1
eta<-0; lambda<-1
Y<-matrix(rnorm(n*nreps,2,1),ncol=n)

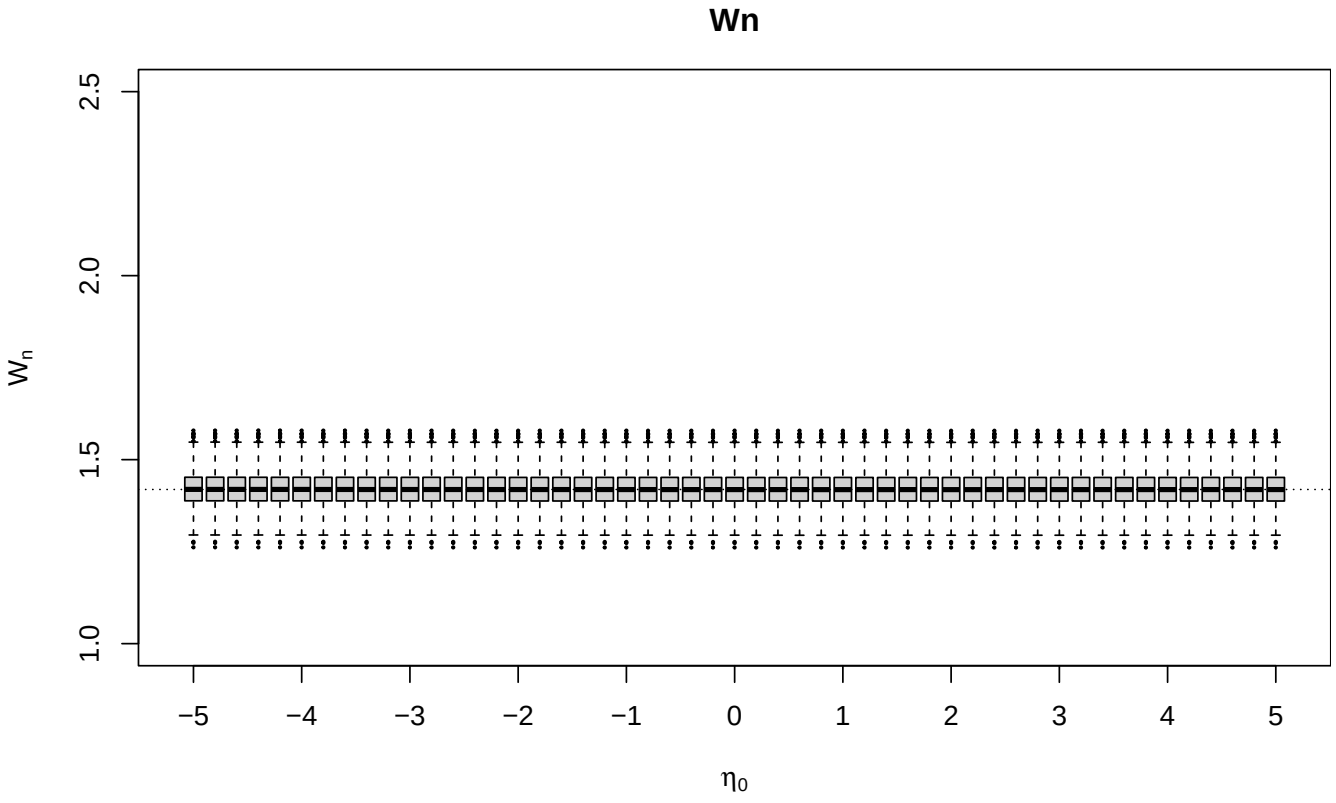
etavec<-seq(-5,5,by=0.2)
lambda.n<-n+lambda; lambda.n1<-lambda.n/(1+lambda.n)
lambda.ni<-n-1+lambda; lambda.ni1<-lambda.ni/(1+lambda.ni)
const<-0.5*log(2*pi)-0.5*log(lambda.n1)
stat.mat<-array(0,c(ncol=length(etavec),nreps,4))

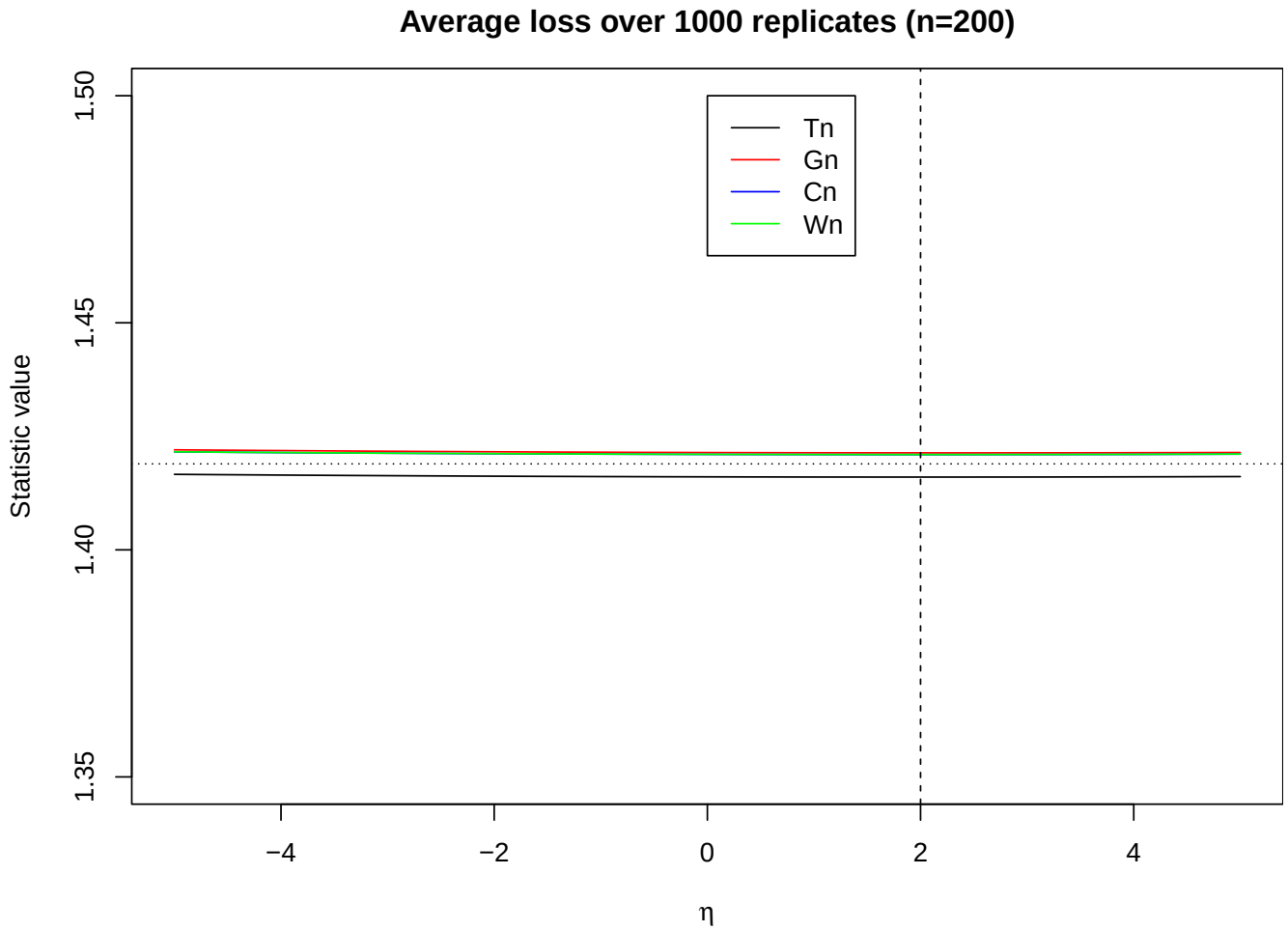
for(il in 1:length(etavec)){
    eta<-etavec[il]
    eta.n<-(rowSums(Y)+eta*lambda)/lambda.n
    Tn<-const+0.5*lambda.n1*rowMeans((Y-eta.n)^2)
    Gn<-const+0.5*lambda.n1*(1+(eta.n-mu0)^2)
    Cn<-const+0.5*lambda.ni1*apply(Y,1,dsq,ev=eta,lv=lambda)/n
    Wn<-Tn+apply(Y,1,variance.term,ev=eta,lv=lambda)
    stat.mat[il,,]<-cbind(Tn,Gn,Cn,Wn)
}
```


Tn
2.5
2.0
1.5
1.0
T_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Gn
2.5
2.0
1.5
1.0
G_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Cn
2.5
2.0
1.5
1.0
C_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0


Wn
2.5
2.0
1.5
1.0
W_n
−5
−4
−3
−2
−1
0
1
2
3
4
5
η_0



For the larger sample size, the prior choice is much less influential, and the statistic values are more or less constant.