MATH 559 - ASSIGNMENT 2 - SOLUTIONS

A total of $n = 24$ times are recorded, each measuring the operating life (time from installation to failure) of a different electrical component. Suppose that the times are presumed to be realizations from an infinitely exchangeable sequence $\{Y_i\}$ such that in the de Finetti representation.

$$f_Y(y;\theta) = \theta e^{-\theta y} \qquad y > 0$$

where $\theta \in \Theta = \mathbb{R}^+ \equiv \{t \in \mathbb{R} : t > 0\}$; that is, given θ, $Y_1, \ldots, Y_n$ are presumed independent and identically distributed, with $Y_i \sim Exponential(\theta)$. It is recorded that

$$s_n \equiv \sum_{i=1}^{n} y_i = 21.17.$$

```
n<-24
s.n<-21.17
t.n<-30.37
a.0<-2
b.0<-0.1
a.n<-n+a.0
b.n<-s.n+b.0
```

(a) *Find the posterior distribution for θ if the prior is chosen to be $Gamma(2, 0.1)$.*

Solution: Setting $\alpha_0 = 2$, $\beta_0 = 0.1$, we have that the posterior is given by

$$\pi_n(\theta) \propto \left\{\prod_{i=1}^{n} \theta e^{-\theta y_i}\right\} \times \theta^{\alpha_0 - 1} e^{-\beta_0 \theta}$$

$$= \theta^{n+\alpha_0-1} \exp\{-\theta(s_n + \beta_0)\}\theta\}$$

so that

$$\pi_n(\theta) \equiv Gamma(n + \alpha_0, s_n + \beta_0) \equiv Gamma(26.00, 21.27)$$

3 MARKS

(b) *Find the Jeffreys prior for θ, and find the posterior under this prior.*

Solution: We have

$$\ell(y, \theta) = \log f_Y(y;\theta) = \log\theta - y\theta$$

$$\dot{\ell}(y, \theta) = \frac{d \log f_Y(y;\theta)}{d\theta} = \theta^{-1} - y$$

$$\ddot{\ell}(y, \theta) = \frac{d^2 \log f_Y(y;\theta)}{d\theta^2} = -\theta^{-2}$$

so therefore Jeffreys's prior is

$$\pi_0(\theta) \propto \{|\mathbb{E}_Y[\ddot{\ell}(Y, \theta); \theta]|\}^{1/2} = \theta^{-1}$$

which is equivalent to choosing $\alpha_0 = 0, \beta_0 = 0$ in the above. It is an improper prior on $\Theta \equiv \mathbb{R}^+$. 3 MARKS

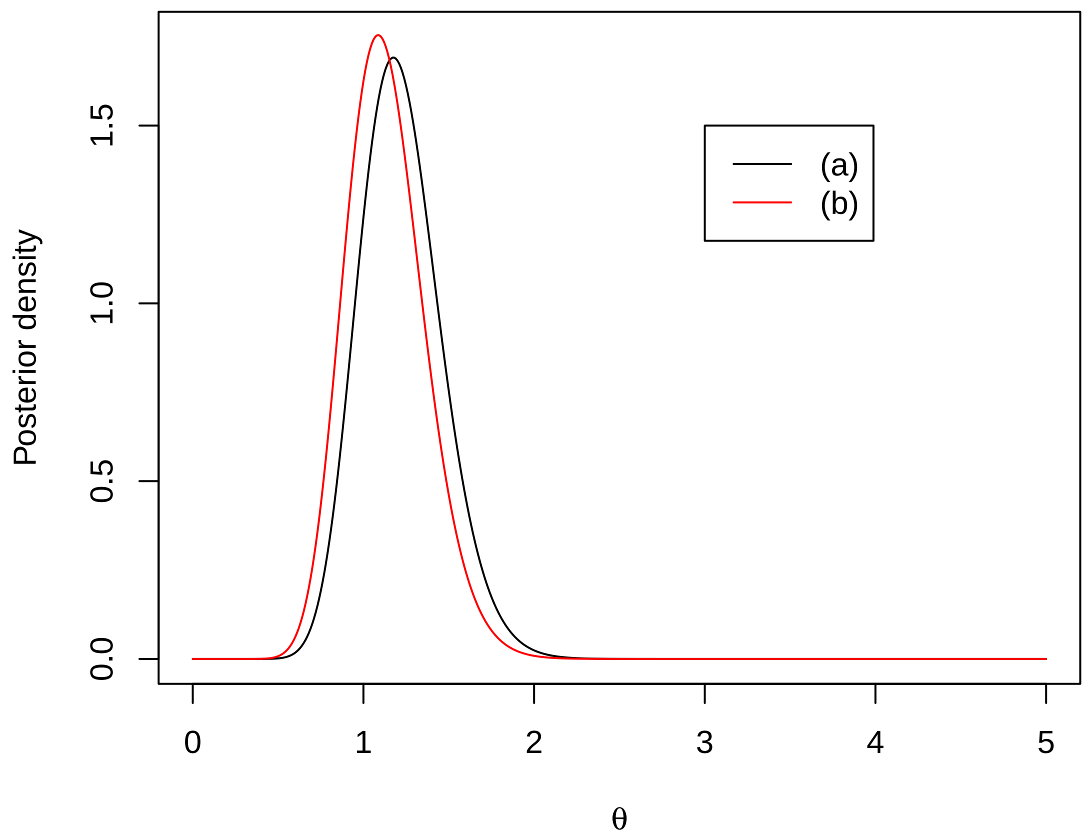
Therefore

$$\pi_n(\theta) \equiv Gamma(n, s_n) \equiv Gamma(24.00, 21.17)$$

2 MARKS

Here are the two posteriors

```
thv<-seq(0,5,by=0.01)
pi.n<-dgamma(thv,a.n,b.n)
pi.nJ<-dgamma(thv,n,s.n)
par(mar=c(4,4,1,2))
plot(thv,pi.n,type='l',ylim=range(0,1.75),xlab=expression(theta),ylab='Posterior density')
lines(thv,pi.nJ,col='red')
legend(3,1.5,c('(a)','(b)'),col=c('black','red'),lty=1)
```



(c) *For a quadratic loss $L(t, \theta) = (t - \theta)^2$, compute the Bayesian estimates (ie the numerical values) from the posteriors in (a) and (b).*

Solution: From lectures, the estimator under this loss is the **posterior mean** which for the Gamma distribution is α_n/β_n, so we have for the two estimates

$$\text{(a)} \quad \frac{26.00}{21.27} = 1.22 \qquad\qquad \text{(b)} \quad \frac{24.00}{21.17} = 1.13$$

2 MARKS

(d) *For a quadratic loss $L(t, \theta) = (t - \theta)^2$, plot (using any suitable software) the minimum expected posterior loss*

$$\min_t \int_0^\infty L(t, \theta)\pi_n(\theta)\, d\theta$$

if the posterior is computed under the prior $\pi_0(\theta) \equiv Gamma(2, \beta_0)$ for $0 < \beta_0 \leq 100$.

Solution: Here we have that the minimized loss is equal to

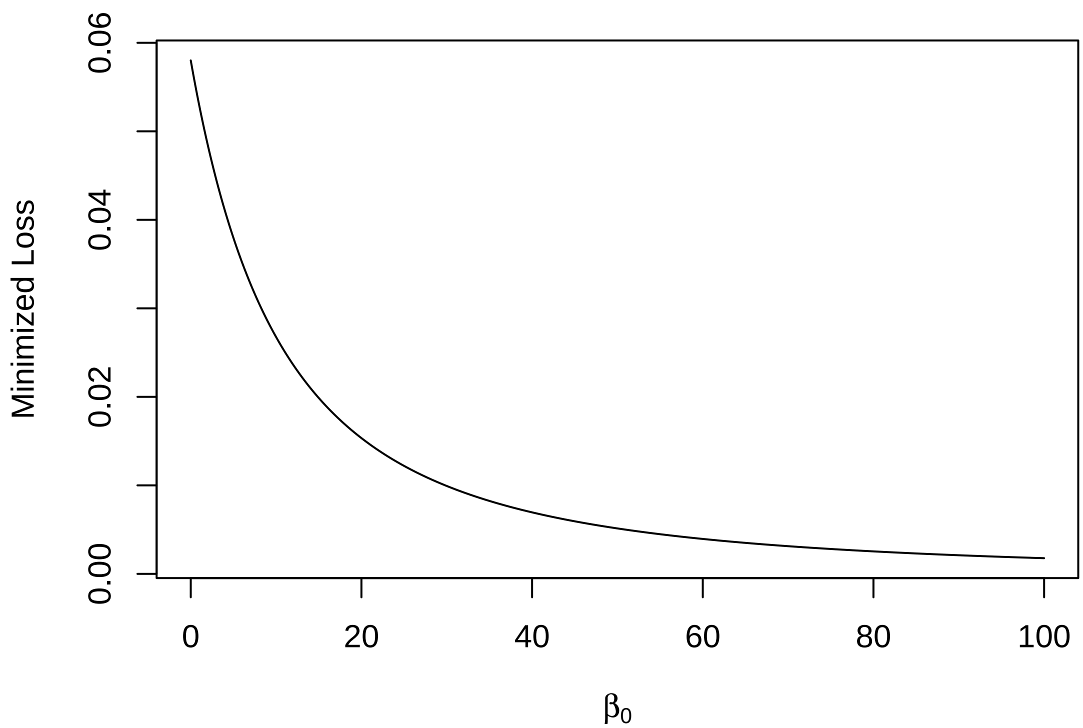
$$\int_0^\infty (\widehat{\theta}_{nB} - \theta)^2 \pi_n(\theta)\, d\theta$$

but this is simply the posterior variance, as $\widehat{\theta}_{nB}$ is the posterior mean. The posterior variance is equal to

$$\frac{\alpha_n}{\beta_n^2} = \frac{n+2}{(s_n+\beta_0)^2} = \frac{26.00}{(21.17+\beta_0)^2}$$

Here is the plot in R:

```
b0v<-seq(0,100,by=0.1)
min.loss<-(n+2)/(s.n+b0v)^2
par(mar=c(4,4,2,2))
plot(b0v,min.loss,type='l',xlab=expression(beta[0]),ylab='Minimized Loss')
```



4 MARKS

(e) Find (using any suitable software) the *one-sided* Bayesian credible interval $\mathcal{C}_{0.95} \equiv (0, c)$ such that

$$\int_0^c \pi_n(\theta)\, d\theta = 0.95.$$

for the posterior in (a).

Solution: In R:

```
cval<-qgamma(0.95,a.n,b.n)
cval

+ [1] 1.641565
```

so the interval is $(0, 1.64)$.

4 MARKS

(f) It is also recorded that

$$t_n \equiv \sum_{i=1}^{n} y_i^2 = 30.37.$$

Comment on the plausibility of the presumed Exponential model.

Solution: There are many ways to approach this question, but I think all of them need to compare the moments. The key feature of the Exponential model is that

$$\mathbb{E}_Y[Y;\theta] = \frac{1}{\theta} \qquad \text{Var}_Y[Y;\theta] = \frac{1}{\theta^2}$$

so that

$$\mathbb{E}_Y[Y^2;\theta] = \frac{2}{\theta^2} = 2\{\mathbb{E}_Y[Y;\theta]\}^2$$

and therefore

$$\frac{\mathbb{E}_Y[Y^2;\theta]}{\{\mathbb{E}_Y[Y;\theta]\}^2} = 2.$$

Now, here for this finite sample, and based on the summary statistics

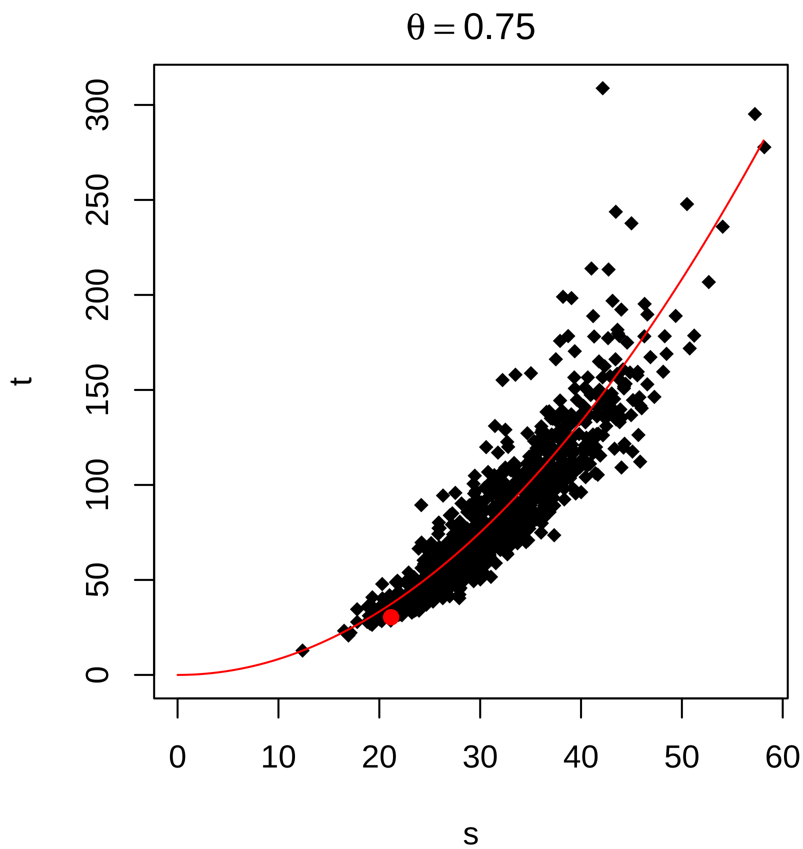
$$\frac{t_n/n}{\{s_n/n\}^2} = 1.63$$

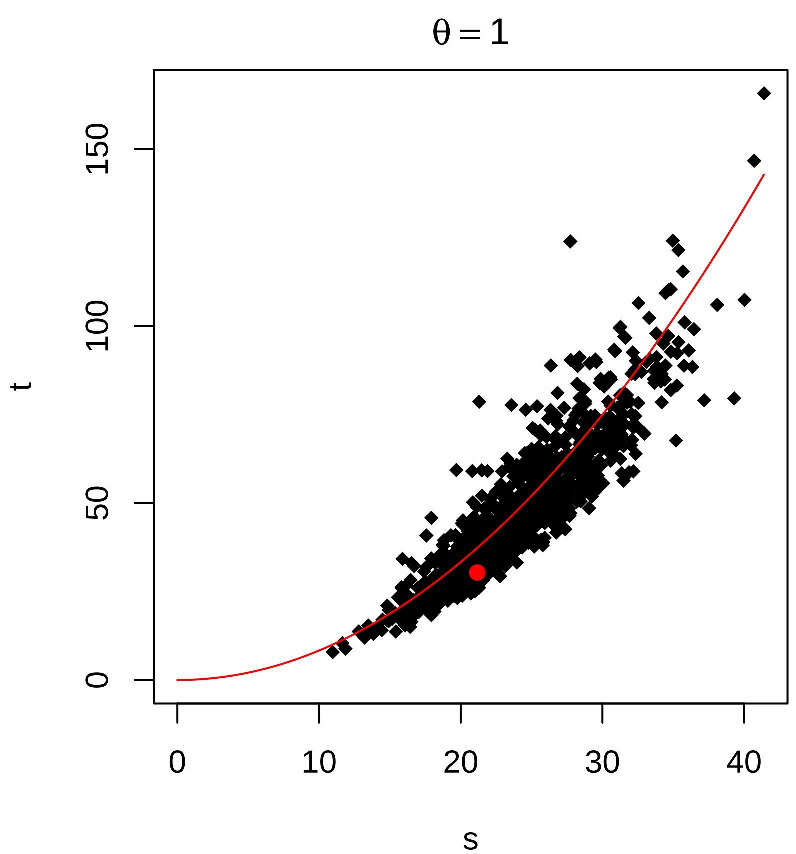
suggesting that the Exponential assumption is not implausible.

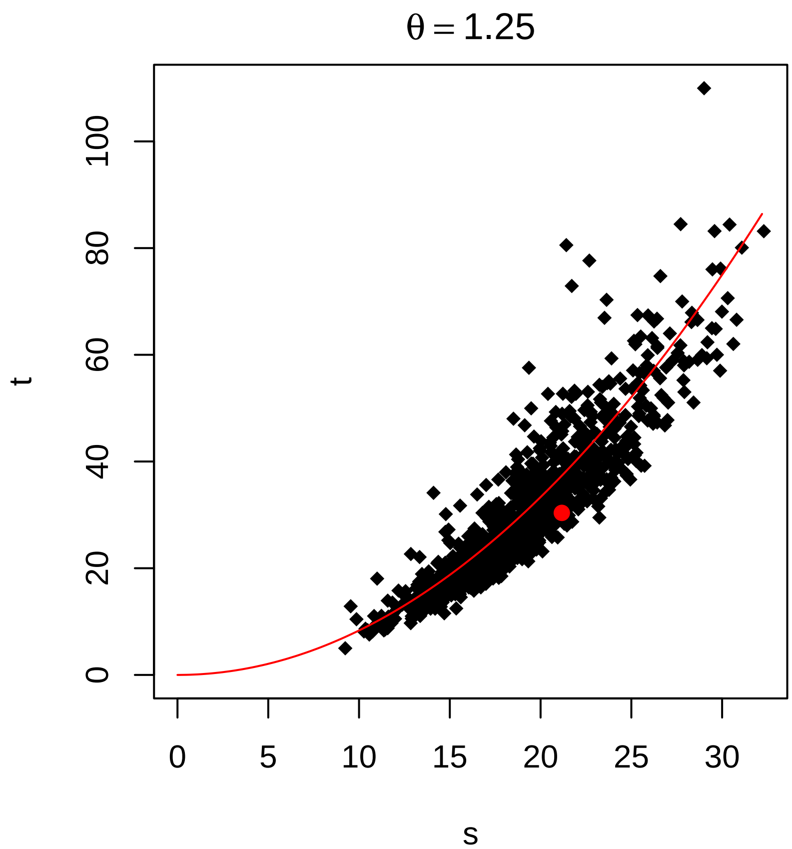
2 MARKS

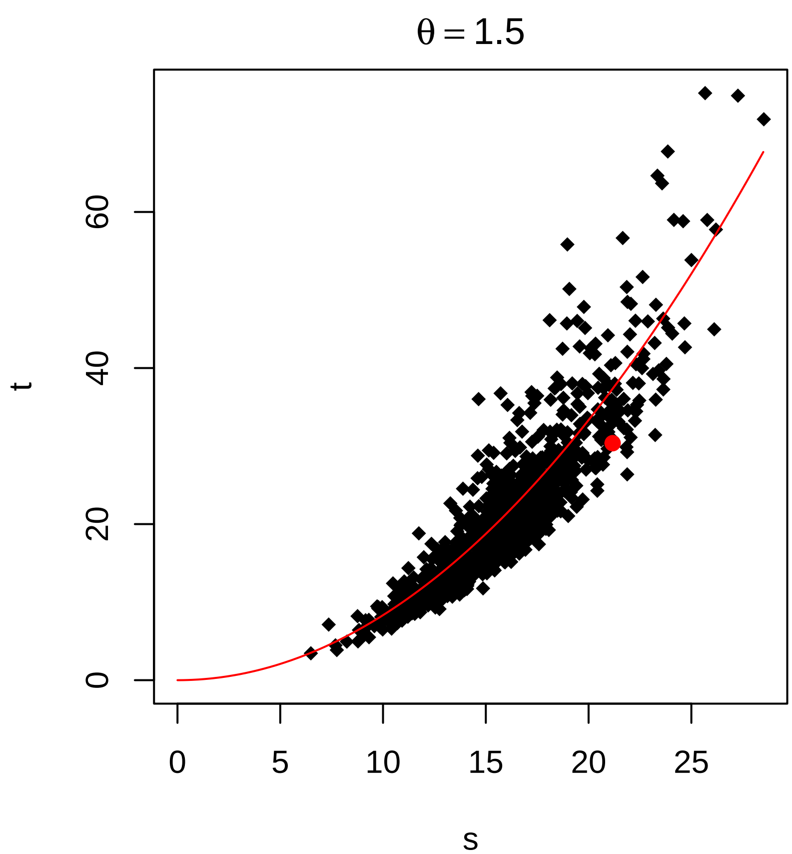
In simulation, we can examine the empirical behaviour by actively varying θ over a suitable range. The red line in each plot is the line $y = (2/n)x^2$, which is the relationship implied by the Exponential model if we plot T_n vs S_n; the red dot is the observed values of (s_n, t_n). The observed statistics seem reasonable for an Exponential model with θ in the range 0.75 to 1.5

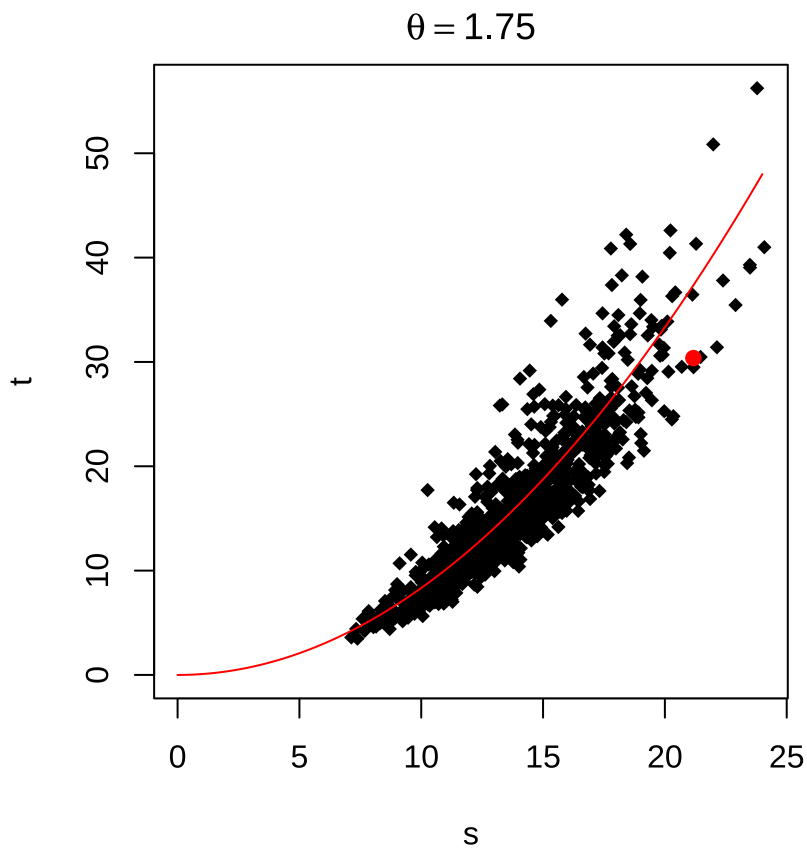
```
thv<-seq(0.75,2.0,by=0.25)
nreps<-1000
s.mat<-matrix(0,nrow=length(thv),ncol=nreps)
t.mat<-matrix(0,nrow=length(thv),ncol=nreps)
exp.sim<-function(nv,thval){
    yv<-rexp(nv,thval)
    return(c(sum(yv),sum(yv^2)))
}
for(i in 1:length(thv)){
    r.mat<-replicate(nreps,exp.sim(n,thv[i]))
    s.mat[i,]<-r.mat[1,]
    t.mat[i,]<-r.mat[2,]
}
par(mar=c(4,4,2,2))
for(i in 1:length(thv)){
    xm<-max(s.mat[i,])
    ym<-max(t.mat[i,])
    plot(s.mat[i,],t.mat[i,],xlab='s',ylab='t',xlim=range(0,xm),ylim=range(0,ym),pch=18)
    xv<-seq(0,xm,by=0.1)
    yv<-2*xv^2/n
    lines(xv,yv,col='red')
    title(substitute(theta==tv,list(tv=thv[i])))
    points(s.n,t.n,col='red',pch=19)
}
```

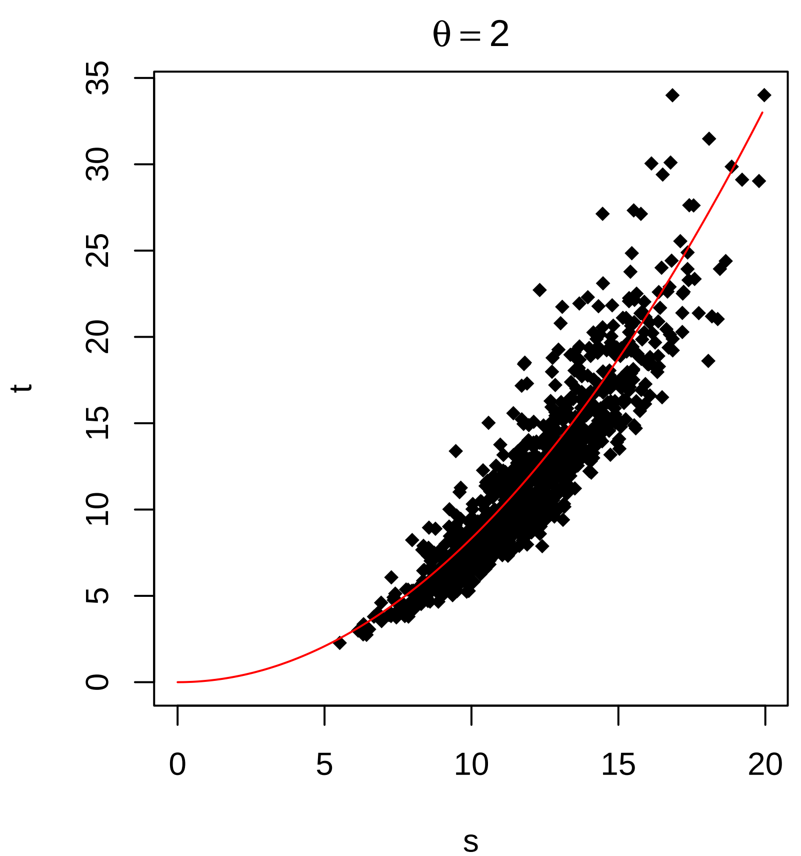

θ = 0.75
t
s


θ = 1
t
s


θ = 1.25
t
0
20
40
60
80
100
0
5
10
15
20
25
30
s


θ = 1.5
t
0
20
40
60
0
5
10
15
20
25
s


θ = 1.75
t
s
0
10
20
30
40
50
0
5
10
15
20
25


θ = 2
t
s
0
5
10
15
20
25
30
35
0
5
10
15
20