STA355 Homework 1

student number: 1003942326

Yulin WANG

2020-01-31

Question1

(a)(i)

Since $Z \sim N(0, \sigma^2)$, so pdf of Z is:

$$f_Z(z;\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{z^2}{2\sigma^2})$$

Then the pdf of $|Z|$ is:

$$f_{|Z|}(z;\sigma) = 2f_Z(z;\sigma) = \frac{2}{\sqrt{2\pi\sigma^2}}\exp(-\frac{z^2}{2\sigma^2}) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\exp(-\frac{z^2}{2\sigma^2}) \quad ;\text{ where } \ z \geq 0$$

So the cdf of half-normal variable $|Z|$ is:

$$F_{|Z|}(z;\sigma) = \int_0^z \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\exp(-\frac{x^2}{2\sigma^2})dx$$

By using change of variables, set $t = \frac{x}{\sqrt{2}\sigma}$, then $x = \sqrt{2}\sigma \cdot t$ and $dx = \sqrt{2}\sigma dt$
Then we have:

$$F_{|Z|}(z;\sigma) = \int_0^{\frac{z}{\sqrt{2}\sigma}} \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\exp(-\frac{2\sigma^2 t^2}{2\sigma^2})\sqrt{2}\sigma dt = \frac{2}{\sqrt{\pi}}\int_0^{\frac{z}{\sqrt{2}\sigma}} \exp(-t^2)dt$$

Thus the cdf of $|Z|$:

$$G(x) = \frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{\sqrt{2}\sigma}} \exp(-t^2)dt$$

Since we know that cdf of a $N(0,1)$ random variable is:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp(-\frac{t^2}{2})dt$$

Then we have:

$$2\Phi(\frac{x}{\sigma}) - 1 = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x}{\sigma}} \exp(-\frac{t^2}{2})dt - 1 = \frac{2}{\sqrt{\pi}}\int_0^{\frac{x}{\sqrt{2}\sigma}} \exp(-t^2)dt = G(x)$$

Therefore, we've proved that $G(x) = 2\Phi(\frac{x}{\sigma}) - 1$

(a)(ii)

Want to show: the τ quantile of the distribution of $|Z|$ is $G^{-1}(\tau) = \sigma\Phi^{-1}((\tau+1)/2)$

Note that based on the definition of cdf of $|Z|$, $G(x) = P(|Z| \le x) = \tau$, $G^{-1}(\tau)$ is the value of $|Z|_\tau$, which is called the τ quantile of the distribution of $|Z|$.
So it is equivalent to show that $G(\sigma\Phi^{-1}(\frac{\tau+1}{2})) = \tau$

Since from part(i) we know that $G(x) = 2\Phi(\frac{x}{\sigma}) - 1$, so we have:

$$
\begin{aligned}
G(\sigma\Phi^{-1}(\frac{\tau+1}{2})) &= 2\Phi[\frac{\sigma\Phi^{-1}(\frac{\tau+1}{2})}{\sigma}] - 1 \\
&= 2\Phi(\Phi^{-1}(\frac{\tau+1}{2})) - 1 \\
&= 2(\frac{\tau+1}{2}) - 1 \\
&= \tau + 1 - 1 \\
&= \tau
\end{aligned}
$$

Therefore, we've proved that the τ quantile of the distribution of $|Z|$ is $G^{-1}(\tau) = \sigma\Phi^{-1}((\tau+1)/2)$

(b)

$W(k)$ is called a central order statistic. Intuitively, $W(k)$ (where $k \approx \tau n$) is an estimator of the τ-quantile $G^{-1}(\tau)$. Therefore, we might expect that

$$W(k_n) = W(k) \xrightarrow{p} G^{-1}(\tau)$$

as $n \to \infty$ if $k_n/n \to \tau$ (i.e. $k \approx \tau n$)
Since $\{k_n\}$ is a sequence of integers with $\sqrt{n}(\frac{k_n}{n} - \tau) \to 0$ for some $\tau \in (0,1)$ and $f(G^{-1}(\tau)) > 0$
So

$$\sqrt{n}(W(k_n) - G^{-1}(\tau)) \xrightarrow{d} N(0, \frac{\tau(1-\tau)}{f^2(G^{-1}(\tau))})$$

then

$$\sqrt{n}(\frac{W(k_n)}{\Phi^{-1}(\frac{\tau_k+1}{2})} - \frac{G^{-1}(\tau)}{\Phi^{-1}(\frac{\tau_k+1}{2})}) \xrightarrow{d} N(0, \frac{\tau(1-\tau)}{f^2(G^{-1}(\tau))} \cdot \frac{1}{[\Phi^{-1}(\frac{\tau_k+1}{2})]^2})$$

Since from part(a)(ii) we have $G^{-1}(\tau) = \sigma\Phi^{-1}((\tau+1)/2)$, so that $\frac{G^{-1}(\tau)}{\Phi^{-1}((\tau+1)/2)} = \sigma$
And we have $\tau_k = \frac{k}{n+1} \to \tau \in (0,1)$ as $k, n \to \infty$, and $\hat{\sigma}_k = \frac{W(k)}{\Phi^{-1}(\frac{\tau_k+1}{2})}$
Thus

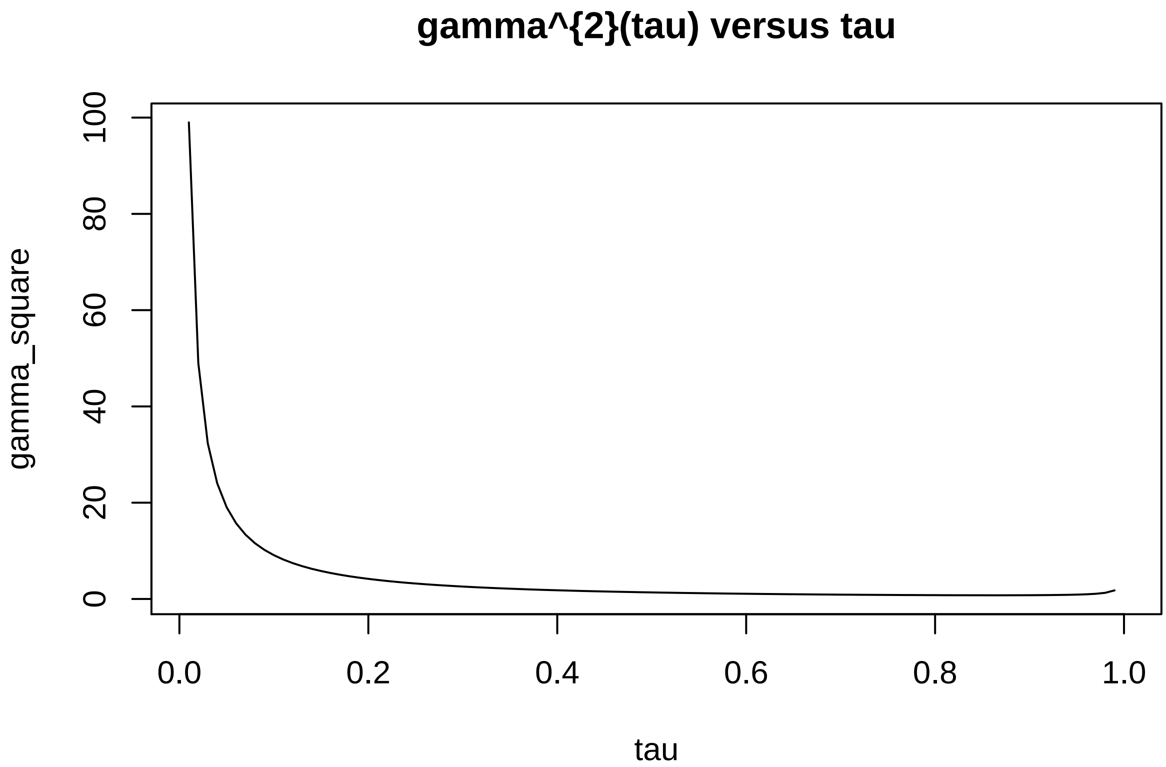
$$\sqrt{n}(\sigma_k - \sigma) \xrightarrow{d} N(0, \frac{\tau(1-\tau)}{f^2(G^{-1}(\tau))} \cdot \frac{1}{[\Phi^{-1}(\frac{\tau+1}{2})]^2})$$

Then we can give an expression for $\gamma^2(\tau)$:

$$\gamma^2(\tau) = \frac{\tau(1-\tau)}{f^2(G^{-1}(\tau))} \cdot \frac{1}{[\Phi^{-1}(\frac{\tau+1}{2})]^2}$$

We can use R to implement this function of $\gamma^2(\tau)$ about τ and create a plot about the value of $\gamma^2(\tau)$ versus the value of $\tau \in (0,1)$ as follows:

```r
library(LaplacesDemon) # the package need to use halfnorm
tau <- 1:100/100 # grid search sequence 0.01,0.02,...,0.99,1
gamma_square <- ((tau * (1-tau))/((dhalfnorm(qhalfnorm(tau)))**2)) / (qnorm((tau+1)/2))**2
plot(tau, gamma_square, type = "l", main = "gamma^{2}(tau) versus tau")
```



As it shows in the plot, the value of $\gamma^2(\tau)$ appears to be decreasing as τ increases from $(0, 1)$
Thus, when $\tau \to 1$, the value of $\gamma^2(\tau)$ minimized.

(c)

Since we have the cdf of $|U|$:

$$\begin{aligned} F(|U|) = P(|U| \leq u) &= P(-u \leq U \leq u) \\ &= P(U \leq u) - P(U \leq -u) \\ &= \Phi(\frac{u - \mu_1}{\sigma}) - \Phi(\frac{-u - \mu_1}{\sigma}) \end{aligned}$$

Then the distribution of $|U|$ depends on μ_1 so that we can assume that $\mu_1 > \mu_2 \geq 0$.

Since the graph of the standard normal cdf Φ has 2-fold rotational symmetry around the point $(0, 1/2)$, that is, $\Phi(-x) = 1 - \Phi(x)$, so $\Phi(-(\frac{x+\mu}{\sigma})) = 1 - \Phi(\frac{x+\mu}{\sigma})$
Then we have the cdf of variable X:

$$\begin{aligned} F(|X|) = P(|X| \leq x) &= P(-x \leq X \leq x) \\ &= P(X \leq x) - P(X \leq -x) \\ &= \Phi(\frac{x - \mu}{\sigma}) - \Phi(\frac{-x - \mu}{\sigma}) \\ &= \Phi(\frac{x - \mu}{\sigma}) - \Phi(-(\frac{x + \mu}{\sigma})) \\ &= \Phi(\frac{x - \mu}{\sigma}) + \Phi(\frac{x + \mu}{\sigma}) - 1 \end{aligned}$$

Then we can take derivative of $P(|X| \le x)$ in terms of μ:

$$\begin{aligned}\frac{\partial P(|X| \le x)}{\partial \mu} &= \frac{\partial \Phi(\frac{x-\mu}{\sigma})}{\partial \mu} + \frac{\partial \Phi(\frac{x+\mu}{\sigma})}{\partial \mu} \\ &= -\Phi^{'}(\frac{x-\mu}{\sigma}) + \Phi^{'}(\frac{x+\mu}{\sigma}) \le 0\end{aligned}$$

Thus, $P(|X| \le x)$ is decreasing but not strictly decreasing as μ increases, that is $P(|U| \le x) \le P(|V| \le x)$ for all x with $P(|U| \le x) < P(|V| \le x)$ for some x.
Therefore, we've proved that U is stochastically greater than V.

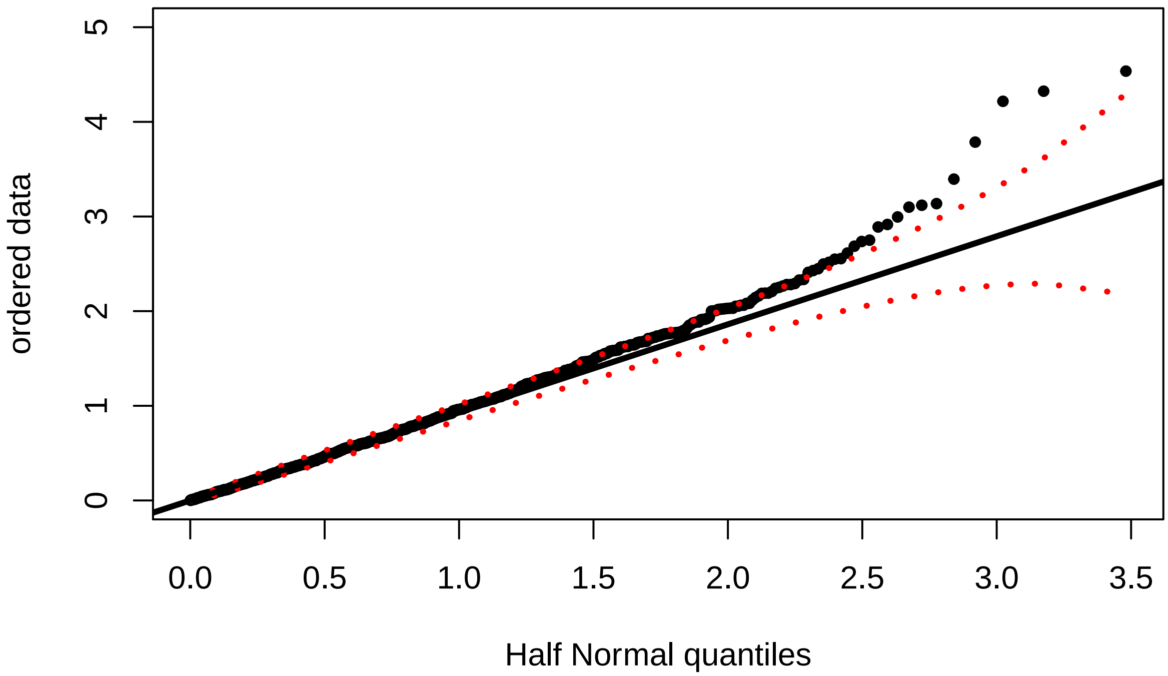
(d)

Firstly, we use the halfnormal function to create a half-normal plot.
And then add a few rows of code to estimate how many of the 1000 means are non-zero.
Note: if the ordered data lies beyond the upper envelope, then we can consider it has a non-zero mean.

```r
# modified the function halfnormal
halfnormal <- function(x,tau=0.5,ylim) {
       sigma <- quantile(abs(x),probs=tau)/sqrt(qchisq(tau,1))
       n <- length(x)
       pp <- ppoints(n)
       qq <- sqrt(qchisq(pp,df=1))
# upper envelope
       upper <- sigma*(qq + 3*sqrt(pp*(1-pp))/(2*sqrt(n)*dnorm(qq)))
# lower envelope
       lower <- sigma*(qq - 3*sqrt(pp*(1-pp))/(2*sqrt(n)*dnorm(qq)))
# add upper and lower envelopes to plot
       if (missing(ylim)) ylim <- c(0,max(c(upper,abs(x))))
       plot(qq,sort(abs(x)),
          xlab="Half Normal quantiles",ylab="ordered data",pch=20,
          ylim=ylim)
       lines(qq,lower,lty=3,lwd=3,col="red")
       lines(qq,upper,lty=3,lwd=3,col="red")
       abline(a=0,b=sigma,lwd=3)

       # add these code to estimate how many observations have non-zero mean.
       num_non_zero <- 0
       for (i in 1:n){
         if (sort(abs(x))[i] > upper[i]){
           num_non_zero <- num_non_zero + 1
         }
       }
       num_non_zero
}

# load the data and use the halfnormal function
x <- scan("data.txt")
halfnormal(x, ylim = c(0, 5))
```




```
## [1] 43
```

Thus, approximately 43 of the 1000 means are non-zero.

Question2

(a)

Since we have:

$$h(x) = \frac{f(x)}{1-F(x)} \quad \text{for } x \geq 0$$

So:

$$h(F^{-1}(\tau)) = \frac{f(F^{-1}(\tau))}{1-F(F^{-1}(\tau))} = \frac{f(F^{-1}(\tau))}{1-\tau}$$

then:

$$\frac{1}{h(F^{-1}(\tau))} = \frac{1-\tau}{f(F^{-1}(\tau))}$$

thus:

$$\int_0^1 \frac{1}{h(F^{-1}(\tau))} d\tau = \int_0^1 \frac{1-\tau}{f(F^{-1}(\tau))} d\tau$$

By using change of variables, set $x = F^{-1}(\tau)$, then $\tau = F(x)$ and $d\tau = f(x)dx$
Then we have:

$$\int_0^1 \frac{1-\tau}{f(F^{-1}(\tau))} d\tau = \int_0^\infty \frac{1-F(x)}{f(x)} f(x)dx = \int_0^\infty (1-F(x))dx = E(X)$$

Therefore, we've proved that $E(X) = \int_0^1 \frac{1}{h(F^{-1}(x))} d\tau$

(b)

Since $k \approx \tau n$, then $\tau \approx \frac{k}{n} \approx \frac{k-1}{n}$, so we have:

$$(n-k+1)D_k = (1 - \frac{k-1}{n})nD_k \approx (1-\tau)nD_k$$

thus:

$$(n-k+1)D_k \xrightarrow{d} (1-\tau)nD_k$$

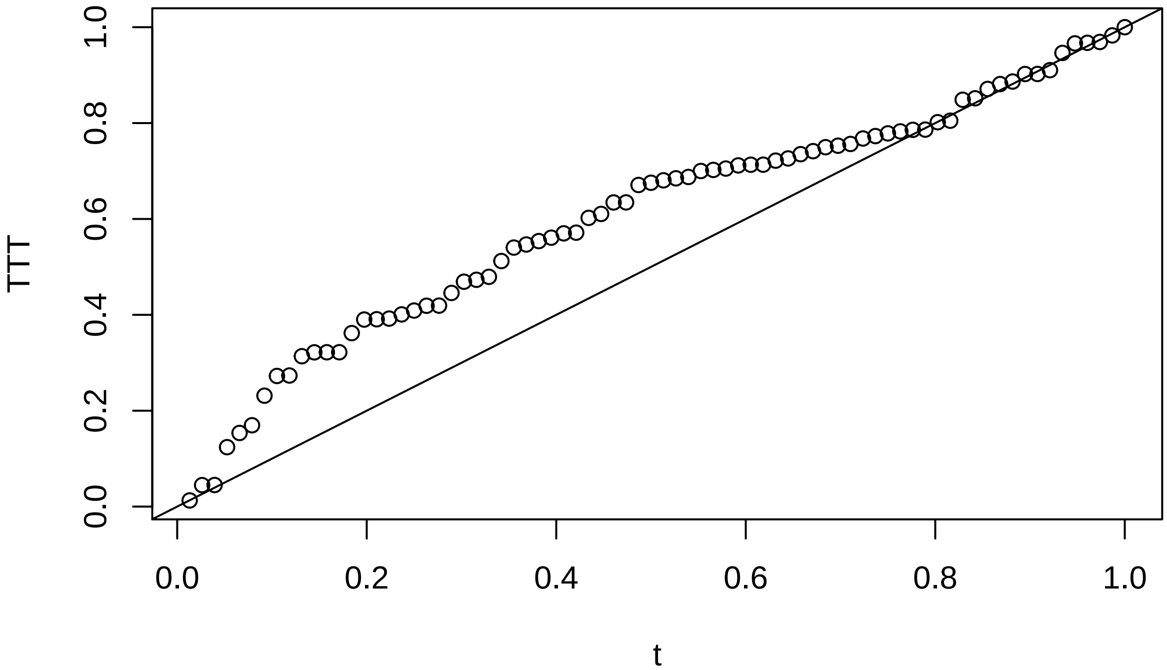
Note that from part(a) we have $\frac{1}{h(F^{-1}(\tau))} = \frac{1-\tau}{f(F^{-1}(\tau))}$

Since we know that the distribution of nDk is approximately Exponential with mean $\frac{1}{f(F^{-1}(\tau))}$,
so that the distribution of $(1-\tau)nD_k$ is approximately Exponential with mean $\frac{1-\tau}{f(F^{-1}(\tau))} = \frac{1}{h(F^{-1}(\tau))}$

Therefore, we've proved that the distribution of $(n-k+1)D_k$ is approximately Exponential with mean $\frac{1}{h(F^{-1}(\tau))}$

(c)

```
# construct a TTT plot for "kevlar.txt".
kevlar <- scan("kevlar.txt")
x <- sort(kevlar) # order elements from smallest to largest
n <- length(x) # find length of x
d <- c(n:1)*c(x[1],diff(x))
plot(c(1:n)/n, cumsum(d)/sum(x), xlab="t", ylab="TTT")
abline(0,1) # add 45 degree line to plot
```



Since the points mostly lie above the 45° line, so it suggests that the hazard function appears to be increasing with time.