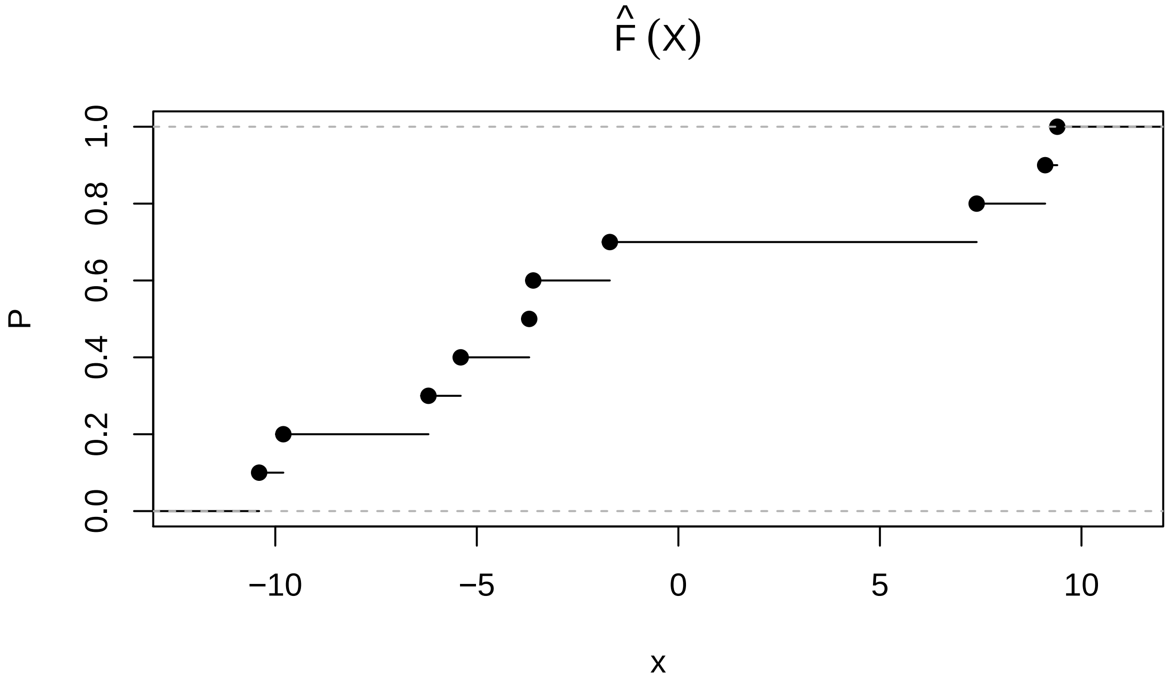
SMA 2018, Ex #4 Solution

Guy Ashiri-Prossner

Q1

$$X = [-9.8, -6.2, -1.7, 7.4, 9.4, -3.6, -10.4, -5.4, -3.7]$$

a



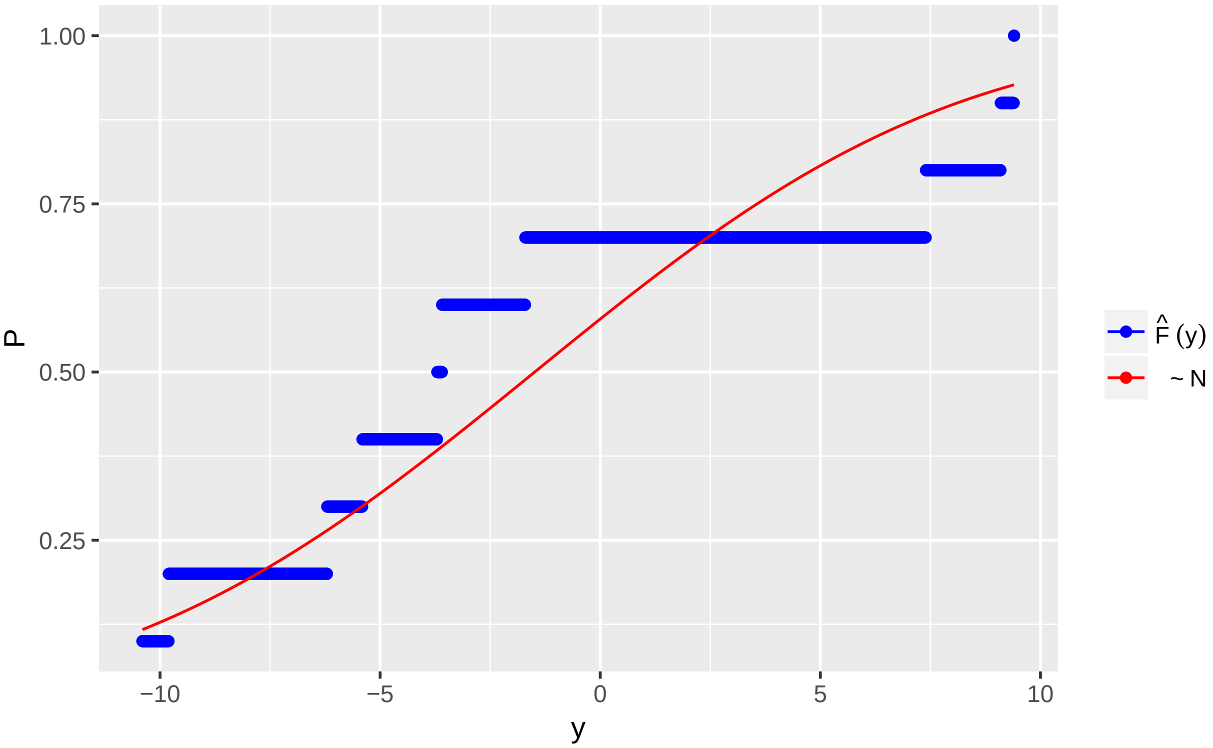
b

$$D = max_{t\in[-\infty,\infty]}\{|\hat{F}_n(t) - \Phi(\frac{t-\bar{X}}{s})|\}$$

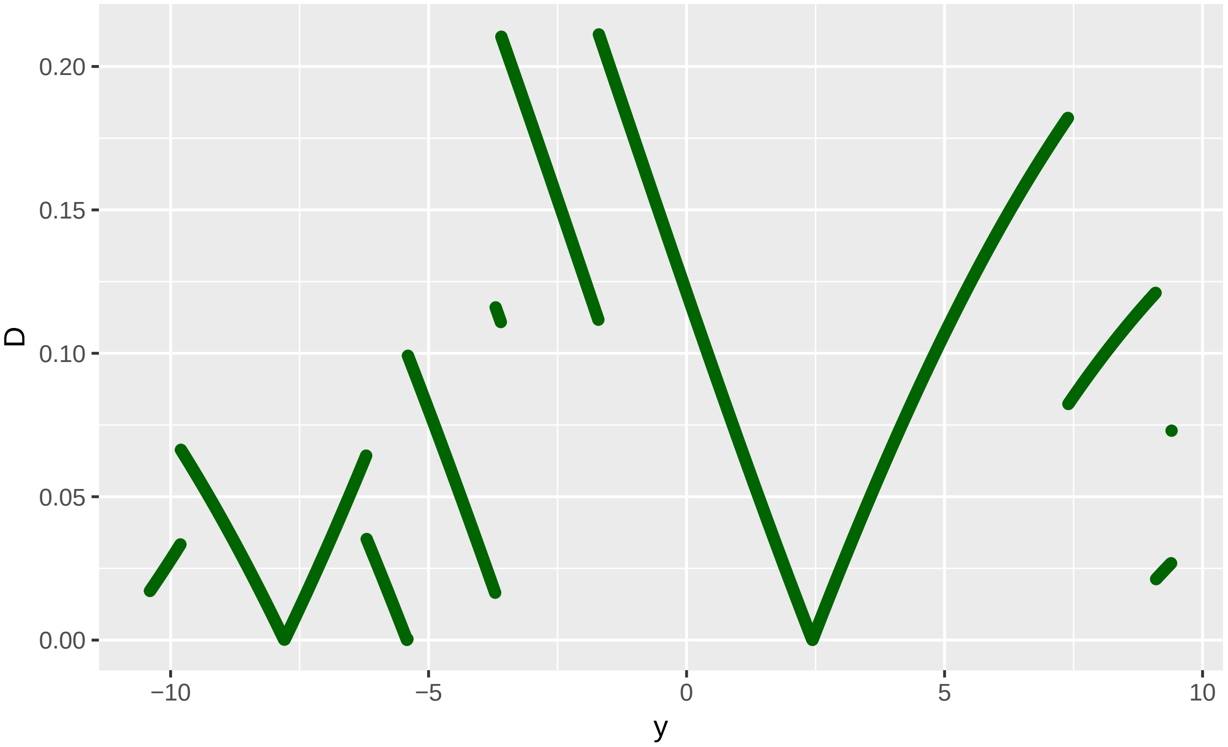
$\bar{X}$ = -1.49, $s = \sqrt{s^2}$ = 7.491

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
t	-10.4000	-9.8000	-6.2000	-5.4000	-3.700	-3.6000	-1.7000	7.4000	9.1000	9.400
d	0.0171	0.0663	0.0352	0.0991	0.116	0.2109	0.2112	0.0823	0.0213	0.073

We can see that the K-S statistic is maximized for t_7. If we take a look at all these t_i points and immediately before them, we still get that $D = 0.2112$. To illustrate the correctness of this answer, let's take this a step further and look at all values in range $[min(X), max(X)]$, step size = 0.01:



If we look at d values for these points, we get:



With the maximum $D = 0.2112$ attained at $t =$ -1.7, which is exactly our previous t_7.

Q2

$$Drug : [2.59, 3.75, 2.08, 2.23, 2.19], Placebo : [2.65, 0.78, 2.46, 2.45, 1.19]$$

a

$$R^D = [8, 10, 3, 5, 4], R^P = [9, 1, 7, 6, 2]$$

b

As we have $\binom{10}{5} = 252$ combinations, let's look at all of the distinct possible values and their frequency:

w	Freq	$P(W_s = w)$	$P(W_s \leq w)$
15	1	0.004	0.004
16	1	0.004	0.008
17	2	0.008	0.016
18	3	0.012	0.028
19	5	0.020	0.048
20	7	0.028	0.075
21	9	0.036	0.111
22	11	0.044	0.155
23	14	0.056	0.210
24	16	0.063	0.274
25	18	0.071	0.345
26	19	0.075	0.421
27	20	0.079	0.500
28	20	0.079	0.579
29	19	0.075	0.655
30	18	0.071	0.726
31	16	0.063	0.790
32	14	0.056	0.845
33	11	0.044	0.889
34	9	0.036	0.925
35	7	0.028	0.952
36	5	0.020	0.972
37	3	0.012	0.984
38	2	0.008	0.992
39	1	0.004	0.996
40	1	0.004	1.000

Our Wilcoxon rank-sum statistic is $W_S = 8 + 10 + 3 + 5 + 4 = 30$.

c

$$E[W_S] = \frac{n_2}{2}(N + 1) = \frac{5}{2} \cdot (10 + 1) = 2.5 \cdot 11 = 27.5$$

$$Var(W_S) = \frac{n_1 n_2 (N + 1)}{12} = \frac{5 \cdot 5 \cdot (10 + 1)}{12} = \frac{25 \cdot 11}{12} = 22.91667$$

d

$$Z = \frac{W_S - E[W_S]}{\sqrt{Var(W_S)}} = \frac{30 - 27.5}{\sqrt{22.91667}} = \frac{2.5}{4.787} = 0.522 \rightarrow P(z > Z) = 0.3$$

Q3

a

```
q3a <- function(n1, n2){
  N <- n1 + n2
  combs <- combn(1:N, n2)
  w <- colSums(combs)
  return(table(w))
}
print(q3a(5, 5))
```

```
## w
## 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
##  1  1  2  3  5  7  9 11 14 16 18 19 20 20 19 18 16 14 11  9  7  5  3  2  1
## 40
##  1
```

b

```
h <- function(q, N, m){
  if((m > 0) && (N > 0) && (q > 0) && (N > m)){
    return(h(q - N, N - 1, m - 1) + h(q, N - 1, m))
  }
  else{
    if(((m == 1) && (q <= N)) || ((m == N) && (q == sum(1:N)))){
      return(1)
    }
    else{
      return(0)
    }
  }
}
print(h(30, 10, 5))
```

```
## [1] 18
```

c

```
n1 <- 5
n2 <- 5
N <- n1 + n2
print(1 - (cumsum(q3a(n1, n2))  / choose(N, n2)))
```

```
##          15          16          17          18          19          20
## 0.996031746 0.992063492 0.984126984 0.972222222 0.952380952 0.924603175
##          21          22          23          24          25          26
## 0.888888889 0.845238095 0.789682540 0.726190476 0.654761905 0.579365079
##          27          28          29          30          31          32
## 0.500000000 0.420634921 0.345238095 0.273809524 0.210317460 0.154761905
##          33          34          35          36          37          38
## 0.111111111 0.075396825 0.047619048 0.027777778 0.015873016 0.007936508
##          39          40
## 0.003968254 0.000000000
```

We get p-value of 0.2738, while the normal approximation gave us 0.3.

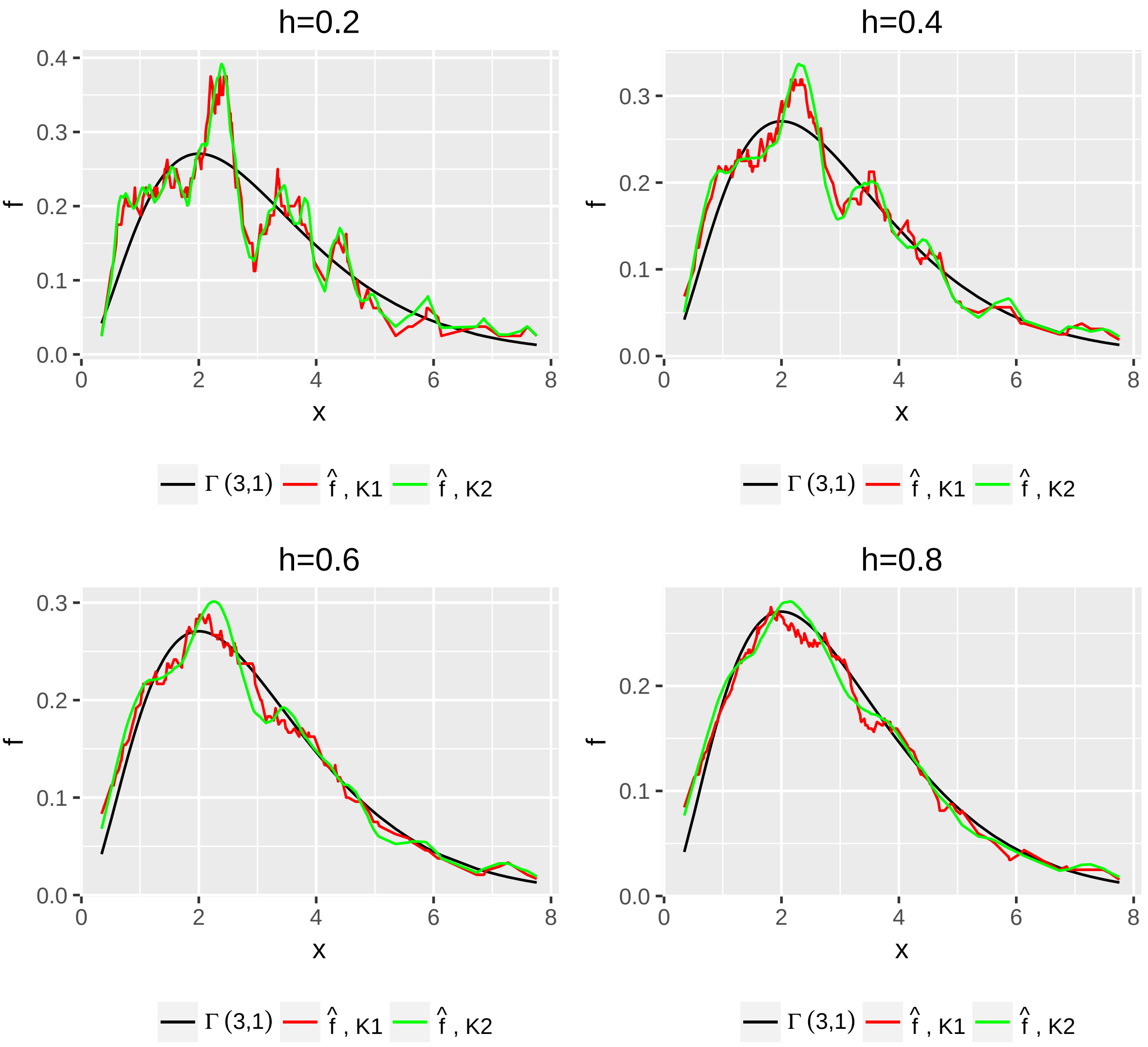
Q4

a

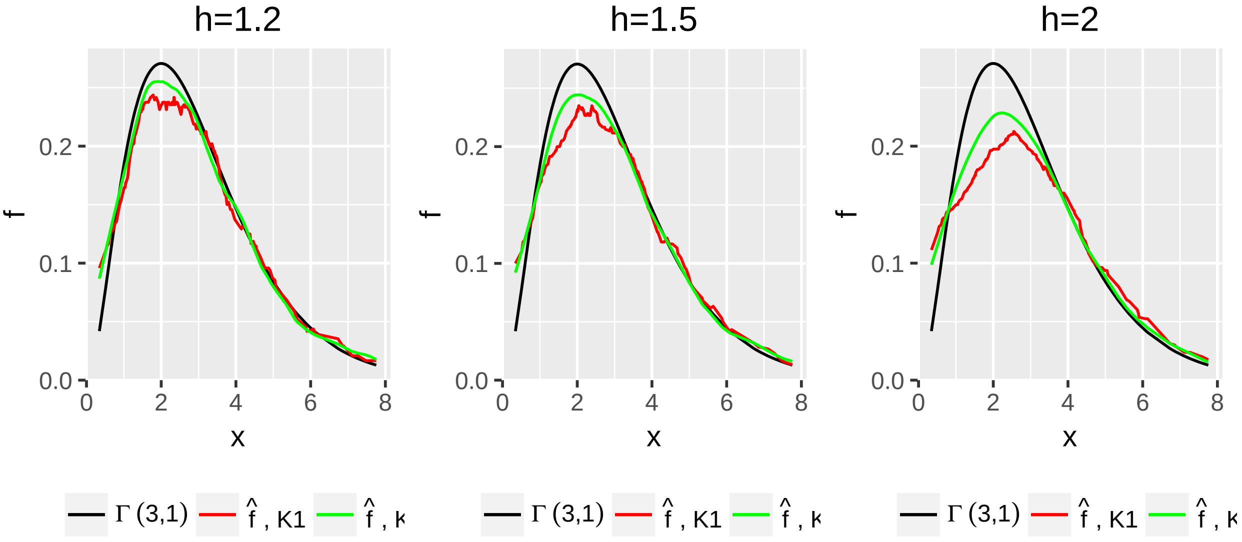
```r
q4a <- function(h, K, x, X){
  f.hat <- 0
  for(i in 1: length(X)){
    f.hat <- f.hat + K((X[i] - x) / h)
  }
  return(f.hat / (length(X) * h))
}
```

b

```r
set.seed(1)
x <- rgamma(200, 3, 1)

K1 <- function(u){
  return(0.5 * ifelse(abs(u) <= 1, 1, 0))
}

K2 <- function(u){
  return(0.75 * (1 - u^2) * ifelse(abs(u) <= 1, 1, 0))
}

h <- c(0.2, 0.4, 0.6, 0.8)
p <- list()
for(i in 1:length(h)){
  df <- data.frame(x = x)
  df$f <- dgamma(x, 3, 1)
  df$f.hat_1 <- q4a(h[i], K1, x, x)
  df$f.hat_2 <- q4a(h[i], K2, x, x)
  p[[i]] <- ggplot(data = df, aes(x = x))+
    geom_line(aes(y = f, colour = "t"))+
    geom_line(aes(y = f.hat_1, colour = "k1"))+
    geom_line(aes(y = f.hat_2, colour = "k2"))+
    scale_colour_manual("", labels = c(expression(Gamma ~ ("3,1")),
                                       expression(hat(f) ~ ", K1"),
                                       expression(hat(f) ~ ", K2")),
                        breaks = c("t", "k1", "k2"),
                        values = c("t"="black", "k1"="red", "k2"="green"))+
    labs(title = paste("h=",h[i], sep="")) +
    theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom")
}
library(gridExtra)
grid.arrange(p[[1]], p[[2]], p[[3]], p[[4]], ncol=2)
```



We can see that the closer h gets to 1, the better fit we get. Note that for higher values, this doesn't hold and the fit gets worse:


h=1.2
h=1.5
h=2
f
x
Γ(3,1)
$\hat{f}$, K1
$\hat{f}$, K