1a. when n = 2, with the given fX(x)
P(y=2) = 0.25
P(y=3) = 0.25
P(y=4) = 0.1875
P(y=5) = 0.125
P(y=6) = 0.078125
P(y=7) = 0.0625
P(y=8) = 0.2734375
P(y=9) = 0.01171875
P(y=10) = 0.0048828125
P(y=11) = 0.0019531250
P(y=12) = 0.0009765625

1b.
Base case:

$$Y_1(y = i) = p(x = i), i = 1,2,\dots 6$$
$$Y_n(y) = 0 \; for \; all \; y \; \notin \{n, n+1, \dots 6n\}$$
$$Yn(y) = sum(Y_{n-1}(y - xi) * p(xi)), xi \in \{1,2,\dots 6\}$$

1c.
Worst running time: O(n^2)
The 7 parameters:
50 0.5 0.25 0.125 0.0625 0.03125 0.03125

Output:

```
Min probabilities:
300 5.527e-76
299 2.764e-74
298 7.323e-73
297 1.365e-71
296 2.006e-70
295 2.471e-69
294 2.651e-68
293 2.539e-67
292 2.213e-66
291 1.779e-65

Max probabilities:
97 0.04381
98 0.0438
96 0.04329
99 0.04327
100 0.04226
95 0.04225
101 0.0408
94 0.04071
102 0.03896
```

93 0.03871

Mean is 98.44
Variance is 82.76

1d.
```
plot(seq(50,300),h_Y_50, ylab='probability')
par(new=TRUE)
plot(seq(50,300),g_Y_50,type = 'l',col = 'red',xaxt ='n',yaxt = 'n',ylab='',xlab='')
par(new=TRUE)
plot(seq(50,300),f_Y_50,type = 'l',col = 'blue',xaxt ='n',yaxt = 'n',ylab='',xlab='')
legend('topright' ,legend = c('h_Y','g_Y','f_Y'),col = c('black','red','blue') )
```
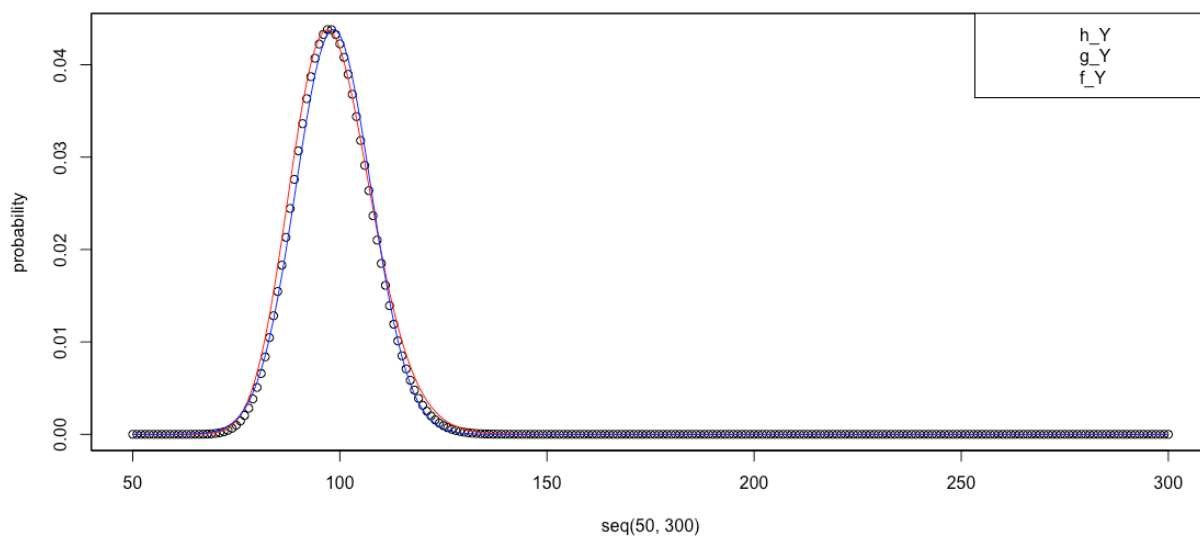
In h_Y, var = 82.76
In g_Y, (negative binomial model), var = 95.361328125
In f_Y, (normal distribution), var = 82.763671875

1e.
Exact p value:
In the dynamic modeling, as calculated above
P(y>=300) = p(y = 300) = 5.527e−76

In negative binomial model:
pnbinom(249,50,phi,lower.tail = F) = 2.485516e-35

in normal distribution:
pnorm(299,f_mu,sqrt(var_f),lower.tail = F) = 5.217155e-108
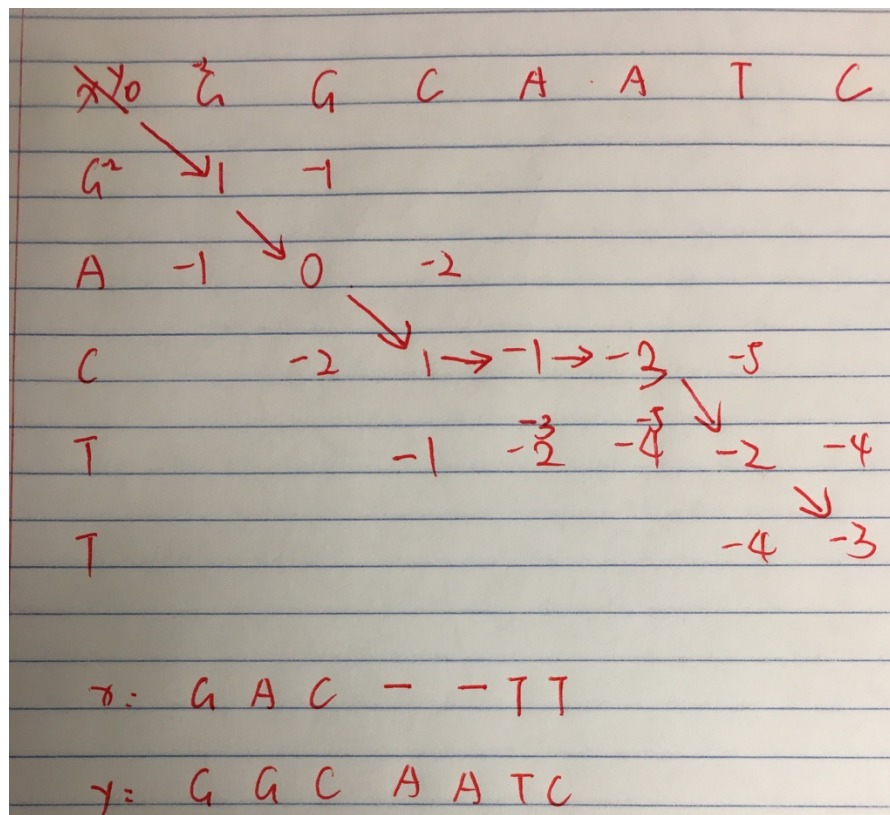
P_true – Pnbinom = -2.485516e-35
P_true – pnorm = 5.527e-76

Normal distribution decays fastest when away from mu. Can be calculated with first derivative of both negative normal distribution density function and normal distribution density function.

Pset 2
2a.

2b.
test on the sequence.fasta
`D = 100`

Alignment:
```
GGGTGGGAAA–ATAGACCAATAGG–CAGAGAGAGTCAGTGCCTATCAGAAACCCAAGAGTCTTCTCTGTCTCCACA–TGC
AAA–GGGAAACATAGA–CAG–GGGACACTCAAAGTTAGTGCCTGCTGGAAA–GC–AGA––C––CTCTGTCTCCA–AGCAC
```
```
CCAGTTTCTA–TTGGTCTCCTTAAACCTGTCTTGTAACCTTGATA
CCAACTTCTACTT–G––––––TGAG–CTGCCTTGTAACCTGGATA
```
Score: 4225

2c.
`D = 430 , e = 30`

Alignment:
```
GGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTATCAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCA
AAAGGGAAACATAGA–CAGGGGACACTCAAAGTTAGTGCCTGCTGGAAA–GC–––AGA––CCTCTGTCTCCAAGCACCCA
```
```
GTTTCTATTGGTCTCCTTAAACCTGTCTTGTAACCTTGATA
ACTTCTACTTG––––––T–GAGCTGCCTTGTAACCTGGATA
```
Score: 3077

2d.
Expensive gap:
D = 4300, e = 300
Alignment:
```
GGGTGGGAAAATAGACCAATAGGCAGAGAGAGTCAGTGCCTATCAGAAACCCAAGAGTCTTCTCTGTCTCCACATGCCCA
–––––––––––––––AAAGGGAAACATAGACAGGGGACACTCAAAGTTAGTGCCTGCTGGAAAGCAGACCTCTGTCTCCAA
```
```
GTTTCTATTGGTCTCCTTAAACCTGTCTTGTAACCTTGATA
GCACCCAACTTCTACTTGTGAGCTGCCTTGTAACCTGGATA
```
Score: –5288

Cheap gap:
D = 43 e=3
Alignment:
```
––GGGTGGGAAA–ATAGACCAATA––GG–CA–––GAGAGAGTCAGTGCCTATC–AGAAA–CCCAAGAGTCTTCTCTGTCT
AAAG–––GGAAACATAGA–C–––AGGGGACACTC–A–A–AGTTAGTGCCTG–CTGGAAAGC–––AGA––C––CTCTGTCT
```
```
CC–A–CATGCCCAGTTTCTA–TTGGTCTCCTTAAACCTGTCTTGTAACCTT–GATA
CCAAGCA––CCCAACTTCTACTT–G––––––TGAG–CTGCCTTGTAACC–TGGATA
```
Score: 6643

Evaluate the gap penalty with the true alignment by MLE.
That is to maximize the score with the given sequence based on the parameter D and e.