

Solution of assignment 11, ST2304

Problem 1 1. R code:

```
recomb<-function(r,n){
  n.draw=0
  for(i in 1:n)
  {
    X<-rpois(1,r)
    if(X%%2){
      n.draw=n.draw+1
    }
  }
  n.draw/n
}

recomb(r=0.01,n=10000)
```

Alt.

```
n<-10000
r<-5
X<-rpois(n,r)
sumX<-sum(X%%2)##summarized over all the
##odd numbers because these are 1's and even
sumX/n##probability of odd number/recombinations
```

r	0.01	0.1	0.5	5
P(recombination)	0.0111	0.0855	0.3146	0.5067

This indicates that larger physical distance between the loci increases the probability of recombinations.

2. To make the graph we include r in a function as a variables.

The limiting values of the probability of recombinations seems to be 0.5 when r goes to infinity. This is resonable as only an odd number of crossover events (a 50-50 chance between even and odd number of crossover events) would result in a recombination.

R code:

```
recomb2<-function(n){
  r<-seq(0.01,6,by=0.1)
  prob<-rep(NA,length(r))
  for(i in 1:length(r))
  {
    n.draw=0
    for(j in 1:n)
    {
      X<-rpois(1,r[i])
      if(X%%2)
      {
        n.draw=n.draw+1
      }
    }
  }
}
```

```

}
prob[i]=n.draw/n
}
prob
plot(x=r,y=prob,xlab="Physical distance, r",ylab="Probability of recombination",typ

}

recomb2(n=10000)
Alt.
  r<-seq(0.01,6,by=0.1)
  prob<-rep(NA,length(r))
  for(i in 1:length(r))
  {
X<-rpois(n,r[i])
sumX<-sum(X%%2)
prob[i]=sumX/n
}

```

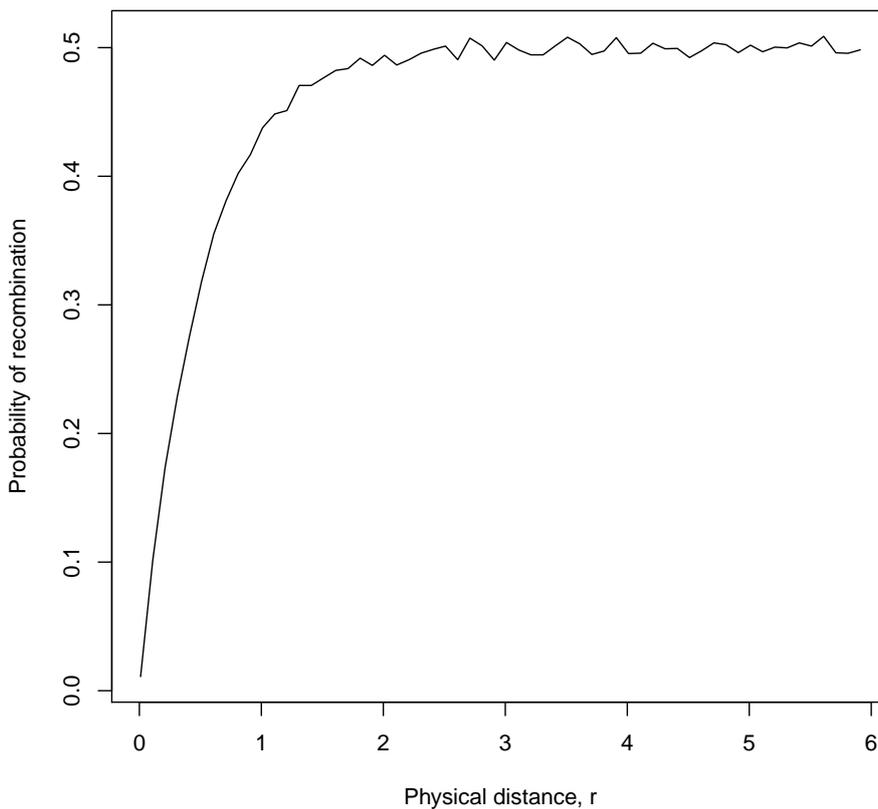


Figure 1: Probability of recombination against physical distance, r , on the chromosome.

Problem 2

S^2 is the unbiased estimate of the variance, σ^2 , based on the sample X_1, X_2, \dots, X_n , from the unknown/true population. The distribution of S^2 is chisquare distributed with $n - 1$ degrees of freedom.

$$Y = X_1^2 + X_2^2 + \dots + X_n^2 = \frac{(n-1)S^2}{\sigma^2} \quad (1)$$

where $X_1 + X_2 + \dots + X_n$ is independent normally distributed variables. The chisquare distribution is not symmetrical, but the quantiles, that is dependent on the degrees of freedom, is defined as

$$P(\chi_{1-\alpha/2} < Y < \chi_{\alpha/2}) = 1 - \alpha, \quad (2)$$

setting in for Y and rearrange the equalities, we get the confidence interval for S^2 . Thus we can find the confidence interval for the true variance σ^2 in the distribution from a sample. Here we see that the confidence interval (0.91,1.08) contains σ^2 , chosen to be 1^2 (sd=1).

```
n=1000
Xnorm=rnorm(n, mean=0 ,sd=1)

S=(1/(n-1))*sum((Xnorm-mean(Xnorm))^2)

confS=list(lower=(S*(n-1))/qchisq(0.025,df=n-1,lower.tail=F),
           upper=(S*(n-1))/qchisq(0.975,df=n-1,lower.tail=F))
> confS
$lower
[1] 0.9074894

$upper
[1] 1.081561
```

Problem 3 1. Under the full model, all n p_i 's are free parameters (no relationship $p_i = q\phi(\beta_0 + \beta_1 \text{time}_i)$ is imposed) and the MLEs are $\hat{p}_i = x_i/n$ which can be computed as follows in R.

```
> phat <- x/n
> phat
[1] 0.0000000 0.0000000 0.0000000 0.0000000 0.1875000 0.1190476 0.2000000
[8] 0.1851852 0.4000000 0.3181818 0.2857143 0.4615385 0.0000000 0.5000000
[15] 0.6250000 0.8055556 0.7272727 0.6666667 0.6551724 0.6969697 0.8214286
[22] 0.8571429 0.9333333 0.8000000 0.9166667 0.7826087 0.7857143 0.7826087
[29] 0.8461538 1.0000000 0.8000000 0.9285714 0.6666667 1.0000000 0.7500000
[36] 0.9000000 0.9000000 0.7777778 0.7500000 1.0000000 0.8571429 1.0000000
[43] 1.0000000 1.0000000 0.5000000 1.0000000 0.0000000 1.0000000 1.0000000
```

2. The maximum log likelihood under the full model is the log likelihood at the point $(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_n)$ in the parameter space. At this point the log likelihood $\ln L(p_1, p_2, \dots, p_n) = \sum \ln f(x_i)$ is

```
> sum(dbinom(x,size=n,prob=phat,log=T))
[1] -47.56002
```

3. From the solution to assignment 10, the maximum log likelihood of the model $p_i = q\phi(\beta_0 + \beta_1 \text{time}_i)$ is -68.21 (the maximum negative log likelihood is in the \$value component of the list returned by `optim`).

4. The observed deviance is two times the difference between the maximum log likelihoods, that is,

```
> 2*((-47.56)-(-68.21))  
[1] 41.3
```

5. Under the null hypothesis that the fitted model is correct the deviance D is chi-square distributed with $n - p = 49 - 3 = 46$ degrees of freedom. We reject this null hypothesis if D is larger than the upper 0.05-quantile of the chi-square distribution,

```
> qchisq(.05,df=46,lower=F)  
[1] 62.82962
```

that is, $\chi_{46}^2 = 62.83$ so we can not reject the hypothesis that the model is correct. The P -value becomes

```
> pchisq(41.3,df=46,lower=F)  
[1] 0.6691562
```

6. The expected value of a chi-square distributed variable is equal to its degrees of freedom, that is, in our case 46. The fact that the observed value of D is slightly smaller than this indicates that there is some (statistically non-significant) under-dispersion in the data.