Ma 322: Biostatistics
Solutions to Homework Assignment 5

Prof. Wickerhauser

Due Monday, March 1st, 2021

Read Chapter 10, "Stochastic Processes and Markov Chains," pages 160–184 of our text.

Note: Although our text has no index or table of contents, it is easy to locate words in the electronic version using the Find function of your favorite PDF reader.

1. Suppose that the list of variables $p = (p_1, \ldots, p_K)$ has a Dirichlet prior density

$$f_\alpha(p) \propto p_1^{\alpha_1 - 1} \cdots p_K^{\alpha_K - 1},$$

where $\alpha = (\alpha_1, \ldots, \alpha_K)$ is a list of shape parameters. We perform an experiment that yields counts $n = (n_1, \ldots, n_K)$ having the multinomial likelihood

$$L_n(p) \propto p_1^{n_1} \cdots p_K^{n_K}.$$

(a) For what values of α does one get a non-informative Dirichlet prior pdf?

(b) Determine the shape parameters for the posterior pdf $L_n(p)f_\alpha(p)$.

Solution: (a) Shape parameters $\alpha_1 = \cdots = \alpha_K = 1$, namely $\alpha = (1, \ldots, 1)$, yield the uniform pdf $f_\alpha(p) = 1$.

(b) The exponents combine to give the posterior pdf

$$L_n(p)f_\alpha(p) \propto p_1^{n_1 + \alpha_1 - 1} \cdots p_K^{n_K + \alpha_K - 1},$$

so the posterior Dirichlet pdf has shape parameters $\alpha + n = (\alpha_1 + n_1, \ldots, \alpha_K + n_k)$. □

2. Suppose that 100 individuals selected randomly from a population are blood-typed and the results are 49 type O, 23 type A, 18 type B, and 10 type AB.

(a) Using a non-informative prior, and assuming Hardy-Weinberg equilibrium, generate a contour plot of the posterior pdf on the proportions p_A and p_B of the A and B blood-type alleles, respectively, in the population. HINT: see `r-eg-35.txt` on the class website.

(b) Find, at least approximately, the maximum-likelihood estimator of the proportion of A,B, and O alleles in the population.

Solution: (a) Use the following code modified from `r-eg-35.txt`

```
pA <- seq(0,1, by=0.01); pB <- seq(0,1, by=0.01)
z <- matrix(0, nrow=length(pA), ncol=length(pB))
nA <- 23; nB<-18; nO<-49; nAB<-10; # experimental data
for(i in 1:length(pA)) {
for(j in 1:length(pB) ) {
a <- pA[i]; b <- pB[j];
if( a+b < 1) { # Otherwise not in the domain.
c <- 1-a-b # Shorthand for pO=1-pA[i]-pB[j].
sA <- a**2 + 2*a*c; sB <- b**2 + 2*b*c; sO <- c**2; sAB <- 2*a*b;
z[i,j] <- sA** nA * sB**nB * sO**nO * sAB**nAB
} else { z[i,j] <- 0 }
}
}
```

(Note: "**" is the same as exponentiation with a carat ˆ.) Then the requested plot is produced by

```
pdf("hw5ex2a.pdf"); contour(pA,pB,z); dev.off()
```

(b) Judging the location of the peak by eye gives $p_A \approx 0.18$, $p_B \approx 0.15$, and so $p_O = 1-p_A-p_B \approx 0.67$. □

3. Implement the function `Walk1d()` on p.167 of our text and graph three 100-step simulations starting from random seeds 123 and 4567 and 89012.

   Solution: Modify the function to make specifying the seeds and printing the plots a bit easier:

```
Walk1d<-function(n=100, seed=NULL) {
if( !is.null(seed) ) set.seed(seed);
y<-vector(length=n); y[1]<-0;
for(i in 2:n) y[i]<-y[(i-1)]+sample(c(-1,1),1);
plot(1:n,y,type='l',ylim=c(-20,20)); }
```

   Then the three requested plots are produced by

```
pdf('w1d123.pdf'); Walk1d(seed=123); dev.off();
pdf('w1d4567.pdf'); Walk1d(seed=4567); dev.off();
pdf('w1d89012.pdf'); Walk1d(seed=89012); dev.off();
```

□

4. Implement the function `Walk2d()` on p.168 of our text and graph three 500-step simulations starting from random seeds 123 and 4567 and 89012.

   Solution: Modify the function to make specifying the seeds and printing the plots a bit easier:

```
Walk2d<-function(n=500, seed=NULL) {
if( !is.null(seed) ) set.seed(seed);
x0<-0; x<-x0+cumsum(sample(c(-1,1),n,replace=TRUE));
y0<-0; y<-y0+cumsum(sample(c(-1,1),n,replace=TRUE));
plot(x,y,xlim=c(-40,40),xlab='x',ylim=c(-40,40),ylab='y',type='l');}
```

   Then the three requested plots are produced by

```
pdf('w2d123.pdf'); Walk2d(seed=123); dev.off();
pdf('w2d4567.pdf'); Walk2d(seed=4567); dev.off();
pdf('w2d89012.pdf'); Walk2d(seed=89012); dev.off();
```

□

5. A restless koala moves among three eucalyptus trees labeled 1, 2, and 3. A patient park ranger watches and makes notes every morning and evening on the koala's position, producing the following table:

Koala Tree-Change Counts

Morning Tree	Evening Tree	Count
1	1	3
1	2	12
1	3	8
2	1	10
2	2	5
2	3	20
3	1	13
3	2	4
3	3	7

(a) Treat the koala's movements as a Markov process and determine the transition matrix M from this table of counts.

(b) Starting with a uniform prior distribution on the three trees and assuming the koala's tree-change preferences remain the same, compute the posterior koala distribution, namely the stationary distribution determined by M.

Solution: (a) Input the table into a matrix of counts as follows:

```
M<-matrix(0,nrow=3,ncol=3); M[1,1]=3; M[1,2]=12; M[1,3]=8; M[2,1]=10; M[2,2]=5;
M[2,3]=20; M[3,1]=13; M[3,2]=4; M[3,3]=7;
```

The row is the starting location and the column is the ending location for each move, just as for the frog situation on p.172 of our text.

To get a matrix of transition probabilities, the matrix elements must be divided by their respective row sums:

```
for(i in 1:3) M[i,] <- M[i,]/sum(M[i,])
```

This produces the transition matrix

$$M = \begin{pmatrix} 3/23 & 12/23 & 8/23 \\ 10/35 & 5/35 & 20/35 \\ 13/24 & 4/24 & 7/24 \end{pmatrix} = \begin{pmatrix} 0.1304348 & 0.5217391 & 0.3478261 \\ 0.2857143 & 0.1428571 & 0.5714286 \\ 0.5416667 & 0.1666667 & 0.2916667 \end{pmatrix}$$

(b) Take powers of M applied (on the right) to the uniform prior $(1/3, 1/3, 1/3)$ until the result stops changing in the first few decimal places (10,000 iterations is more than we need, but the computer will not complain):

```
t0<-c(1,1,1)/3; t<-t0; for(i in 1:10000) t<-t%*%M; t
```

The result is `0.3333333, 0.2783957, 0.388271`, so the koala may be expected to spend 33% of the time in tree 1, 28% of the time in tree 2, and 39% of the time in tree 3.

Note: Taking powers of just M until its coefficients stop changing gives a matrix whose rows are identical copies of this stationary distrubution. This may be done by the following `R` commands:

```
MM<-M; for(i in 1:10000) MM<-M%*%MM; MM
```

□

6. Consider the following transition matrix for a 4-state Markov chain:

$$M = \begin{pmatrix} 0.2 & 0.1 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.3 & 0.4 \\ 0.3 & 0.4 & 0.1 & 0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{pmatrix}.$$

(a) Is M periodic or aperiodic?

(b) Is M irreducible?

(c) Is M ergodic?

(d) Does M have a stationary distribution?

(e) Is M reversible?

Solution: First use the following `R` code to test if $\lim_{n\to\infty} M^n$ exists:

```
data<-c(2,1,3,4,1,2,4,3,4,3,1,2,3,4,2,1)/10;
M<-matrix(data,4,4); MM<-M; M
for(i in 1:100) MM<-M%*%MM;
MM
```

The result is

$$M^\infty \stackrel{\text{def}}{=} \lim_{n\to\infty} M^n = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}.$$

(a) The existence of the limit of M^n as $n \to \infty$ implies that M is aperiodic.

(b) M has all nonzero elements, so M is irreducible.

(c) Since M is aperiodic and irreducible, it is ergodic.

(d) M is ergodic by part c, and from M^∞ we conclude that M has a stationary state $\pi = (.25, .25, .25, .25)$.

(e) Checking the 16 detailed balance equations we find that $\pi[1]M[1,3] = (0.25)(0.4) \neq (0.25)(0.3) = \pi[3]M[3,1]$, so M is not reversible. □

7. Consider the following transition matrix for a 5-state Markov chain:

$$F = \begin{pmatrix} 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\ 0.0 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.0 & 0.0 & 0.33 & 0.33 & 0.34 \\ 0.0 & 0.0 & 0.0 & 0.50 & 0.50 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.00 \end{pmatrix}.$$

(a) Is F periodic or aperiodic?

(b) Is F irreducible?

(c) Is F ergodic?

(d) Does F have a stationary distribution?

(e) Is F reversible?

Solution: First use the following R code to test if $\lim_{n\to\infty} M^n$ exists:

```
fdat<-c(c(1,1,1,1,1)/5,c(0,1,1,1,1)/4,c(0,0,1,1,1)/3,
c(0,0,0,1,1)/2,c(0,0,0,0,1)); F<-matrix(fdat,5,5); F
```

The result is

$$F^{\infty} \stackrel{\text{def}}{=} \lim_{n\to\infty} F^n = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

(a) The existence of the limit of F^n as $n \to \infty$ implies that F is aperiodic.

(b) Every pair of states i, j with $i > j$ has $F(i, j) = 0$, so state i can never transition to any state j with $j < i$. Thus F is reducible.

(c) Since F is not irreducible, it is not ergodic.

(d) Though F is not ergodic, the limit F^{∞} nonetheless exists, and we conclude that F has a stationary state $\pi = (0, 0, 0, 0, 1)$. This means that state 5 is *absorbing*, or equivalently that there are no transitions possible out of state 5.

(e) Checking the 25 detailed balance equations with the stationary distribution from part d shows that

$$\pi[i]F[i,j] = \pi[j]F[j,i] = \begin{cases} 0, & \text{if } (i,j) \neq (5,5) \\ 1, & \text{if } i = j = 5, \end{cases}$$

which includes all i, j, so F is reversible. $\square$

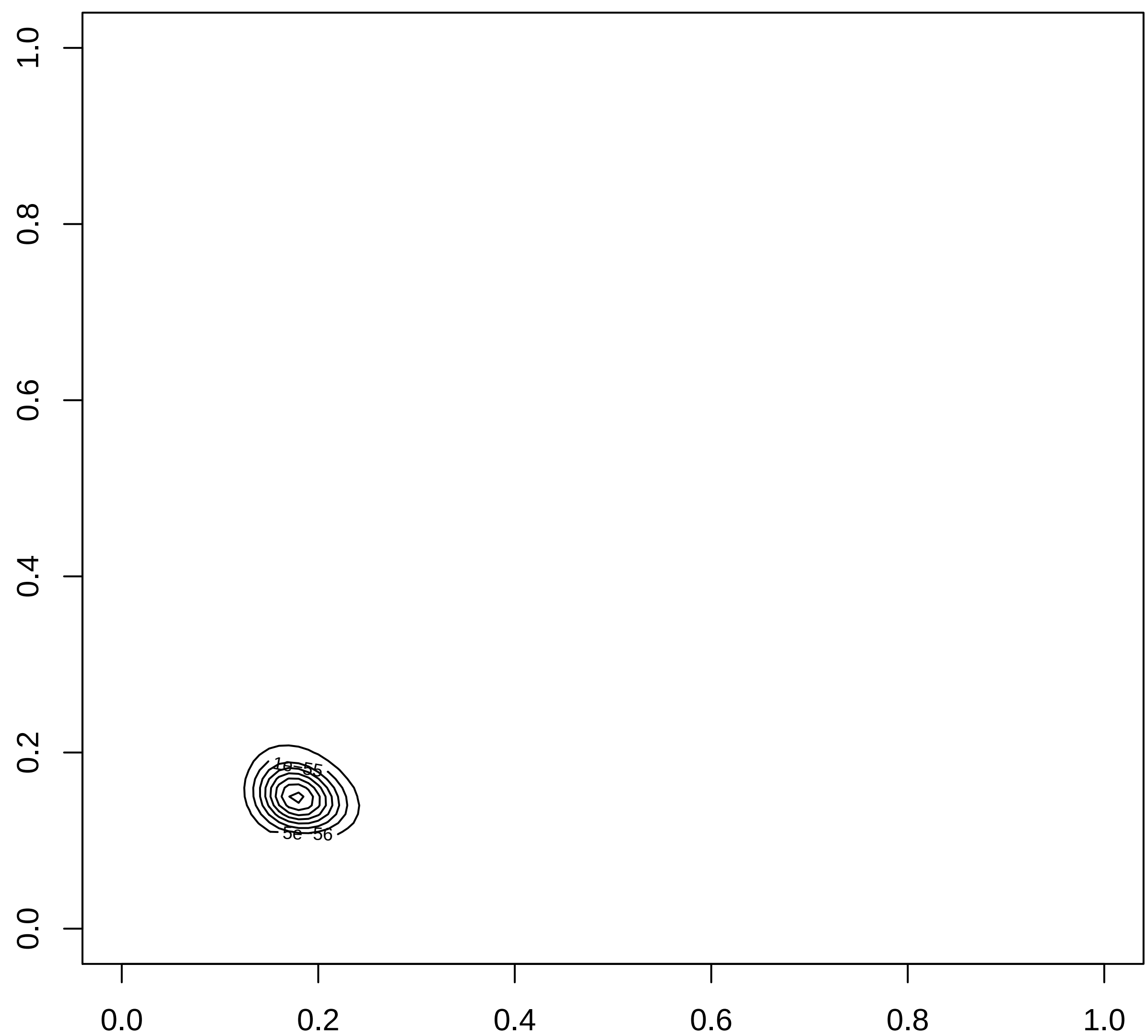

Figure 1: HW 5, Ex.2a: Contour plot of posterior pdf in (p_A, p_B)

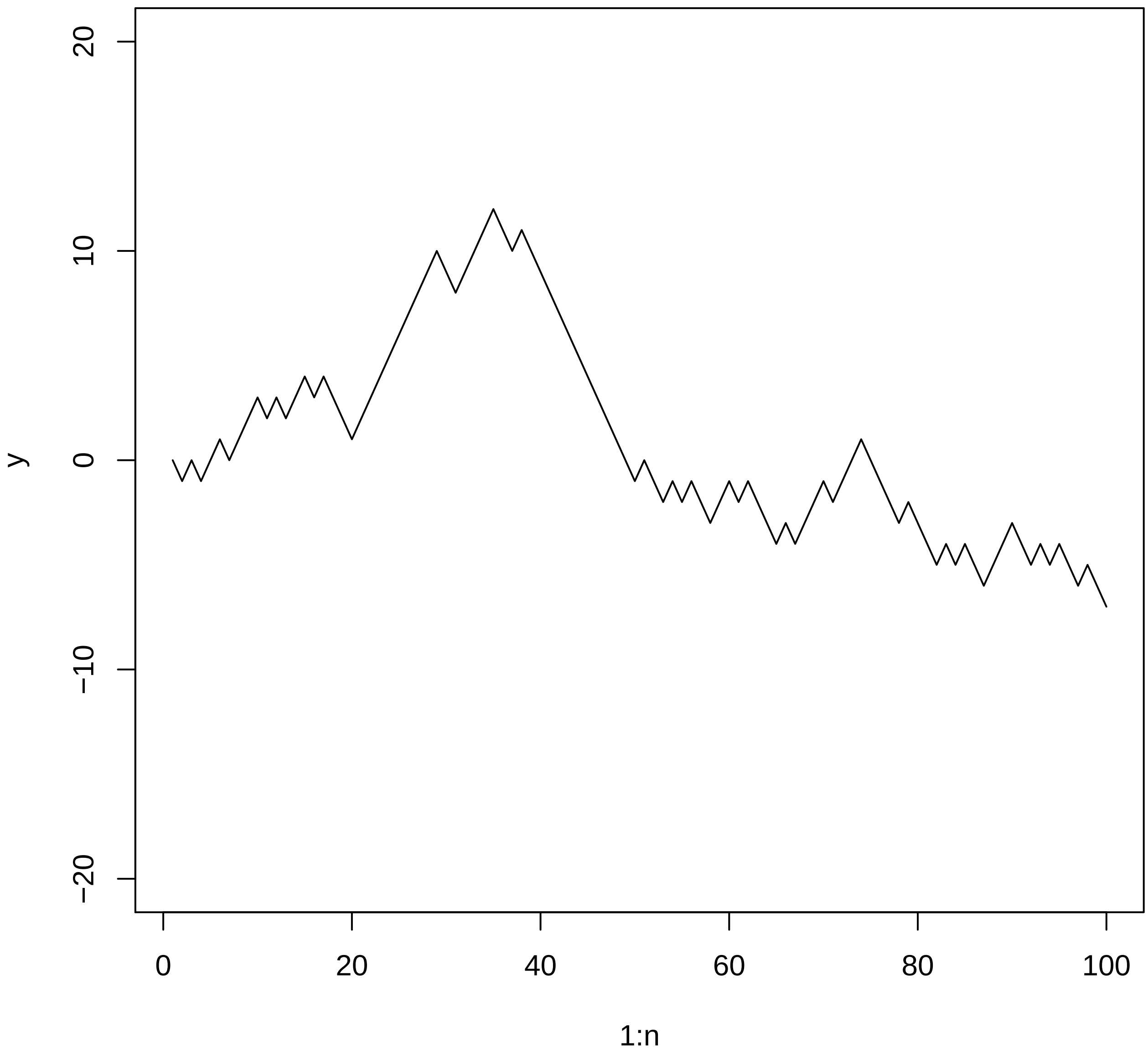

Figure 2: HW 5, Ex.2i: 1-D random walk, seed=123

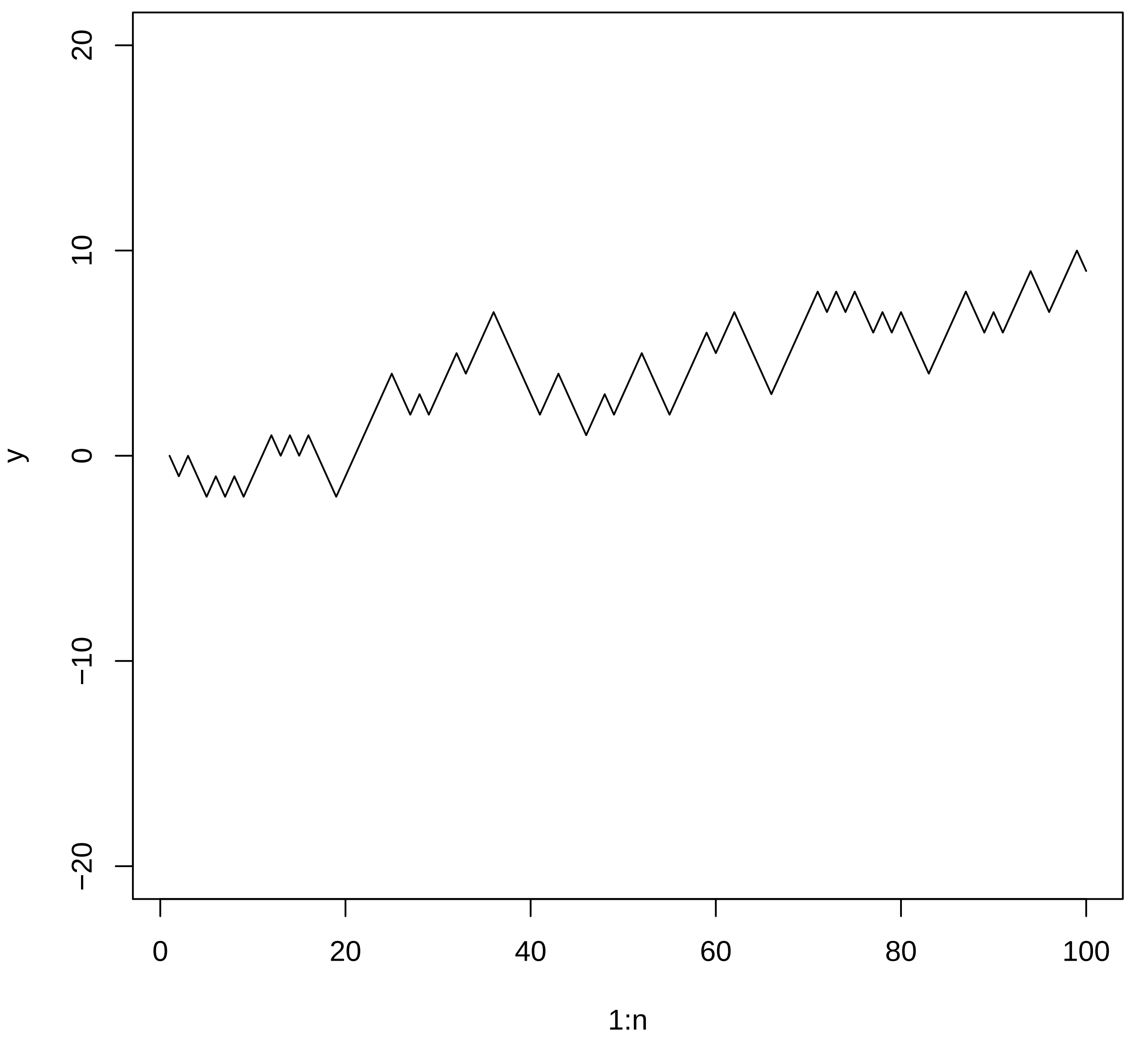

Figure 3: HW 5, Ex.2ii: 1-D random walk, seed=4567

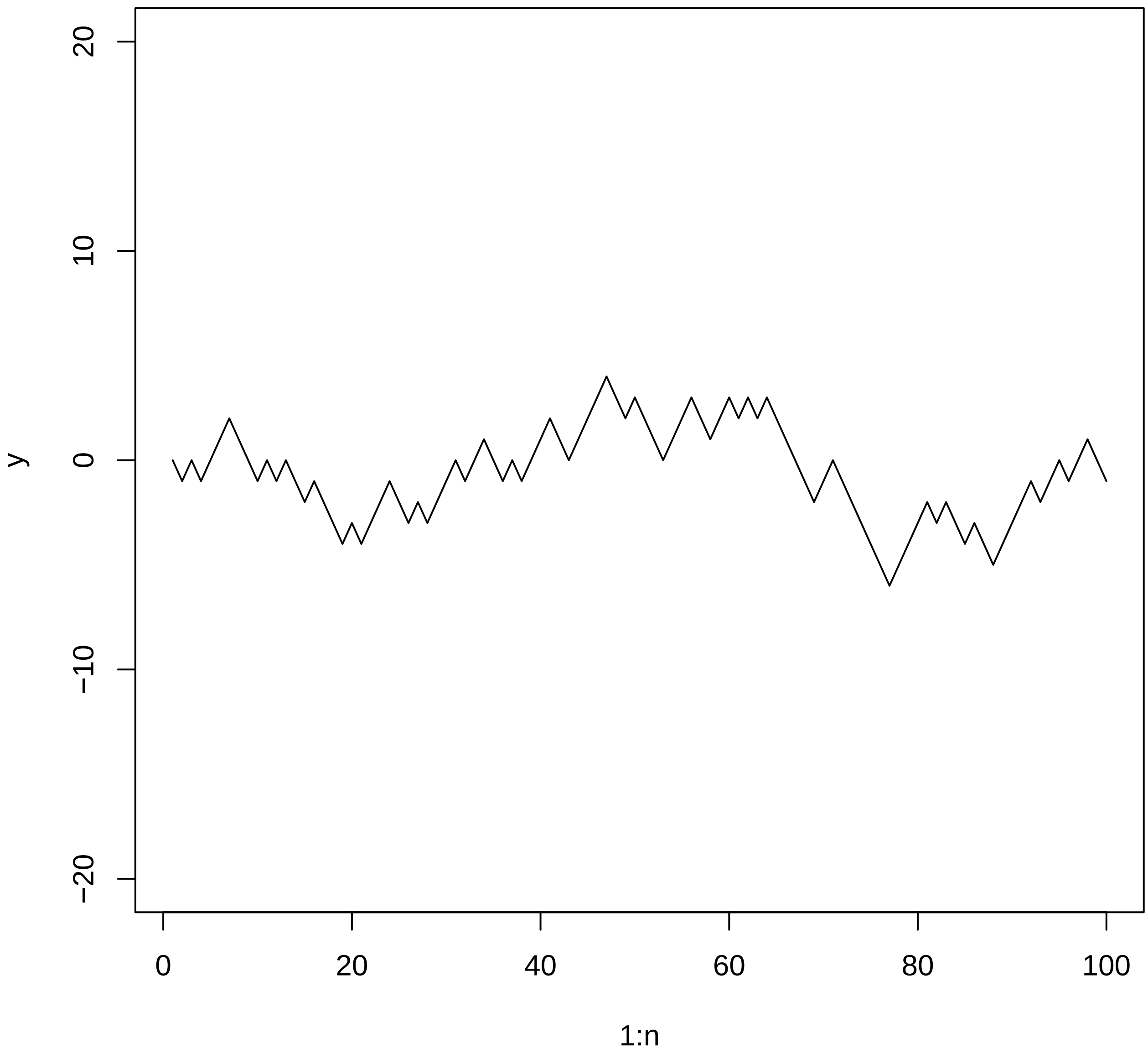

Figure 4: HW 5, Ex.2iii: 1-D random walk, seed=89012

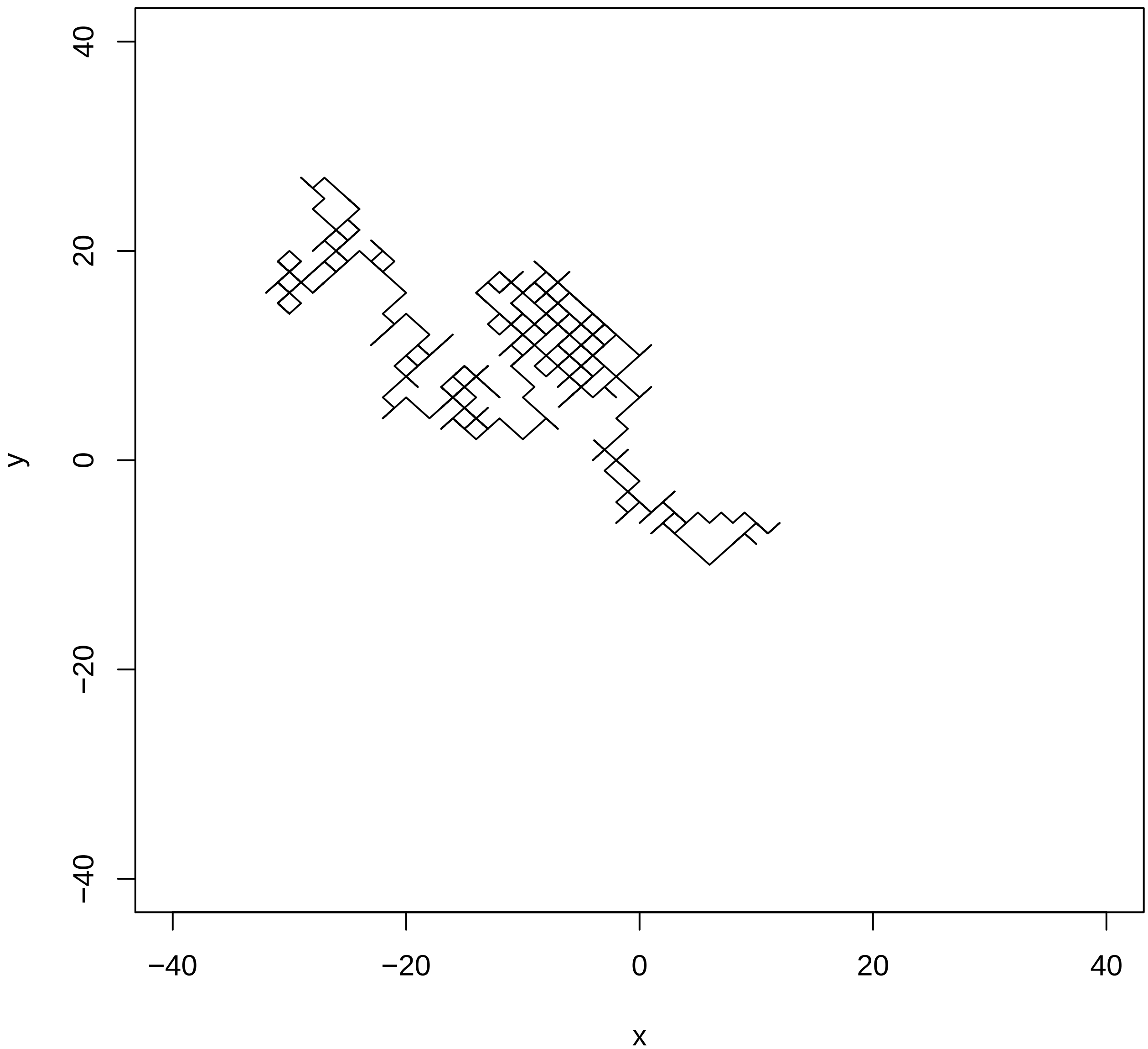

Figure 5: HW 5, Ex.3i: 2-D random walk, seed=123

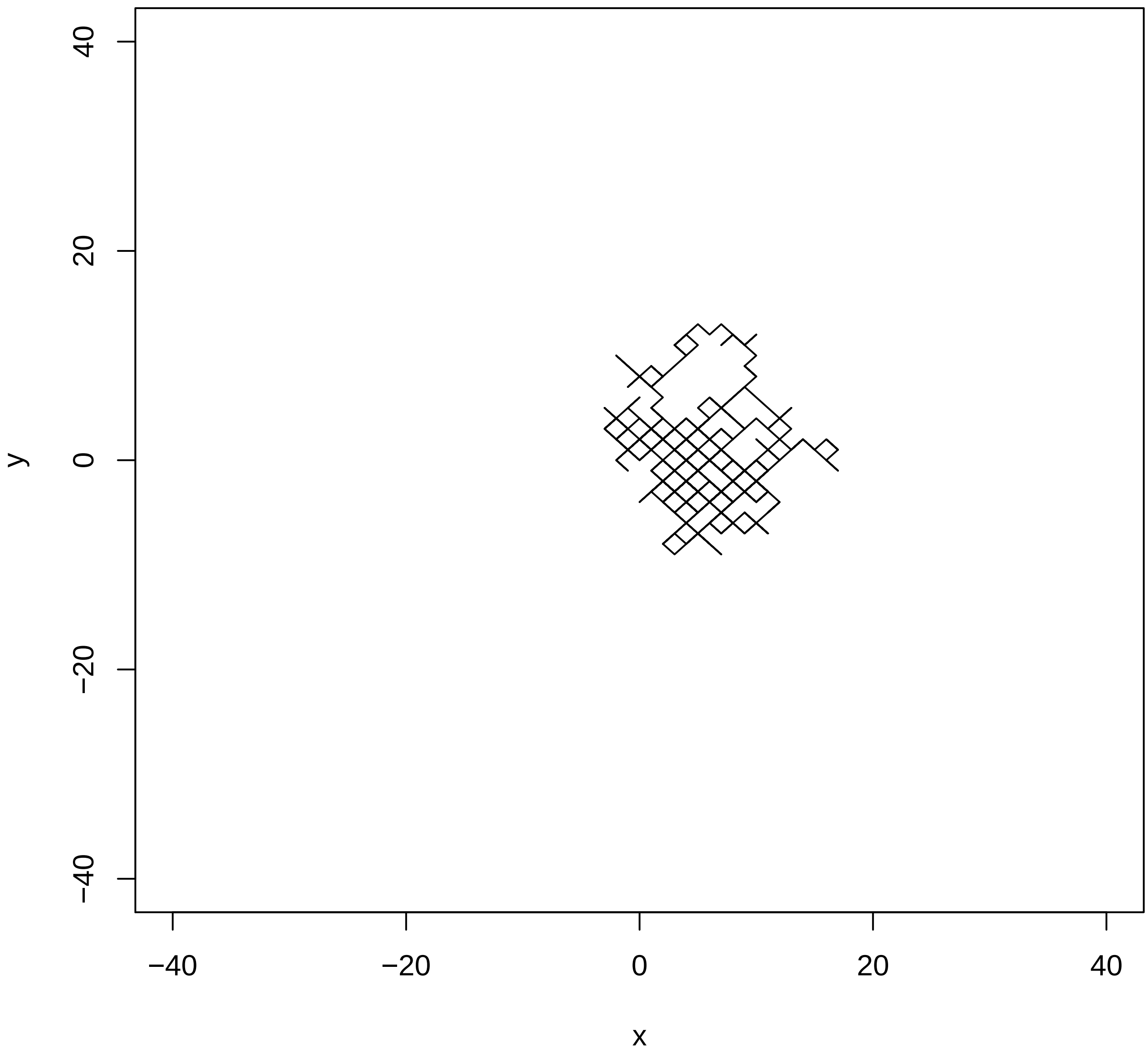

Figure 6: HW 5, Ex.3ii: 2-D random walk, seed=4567

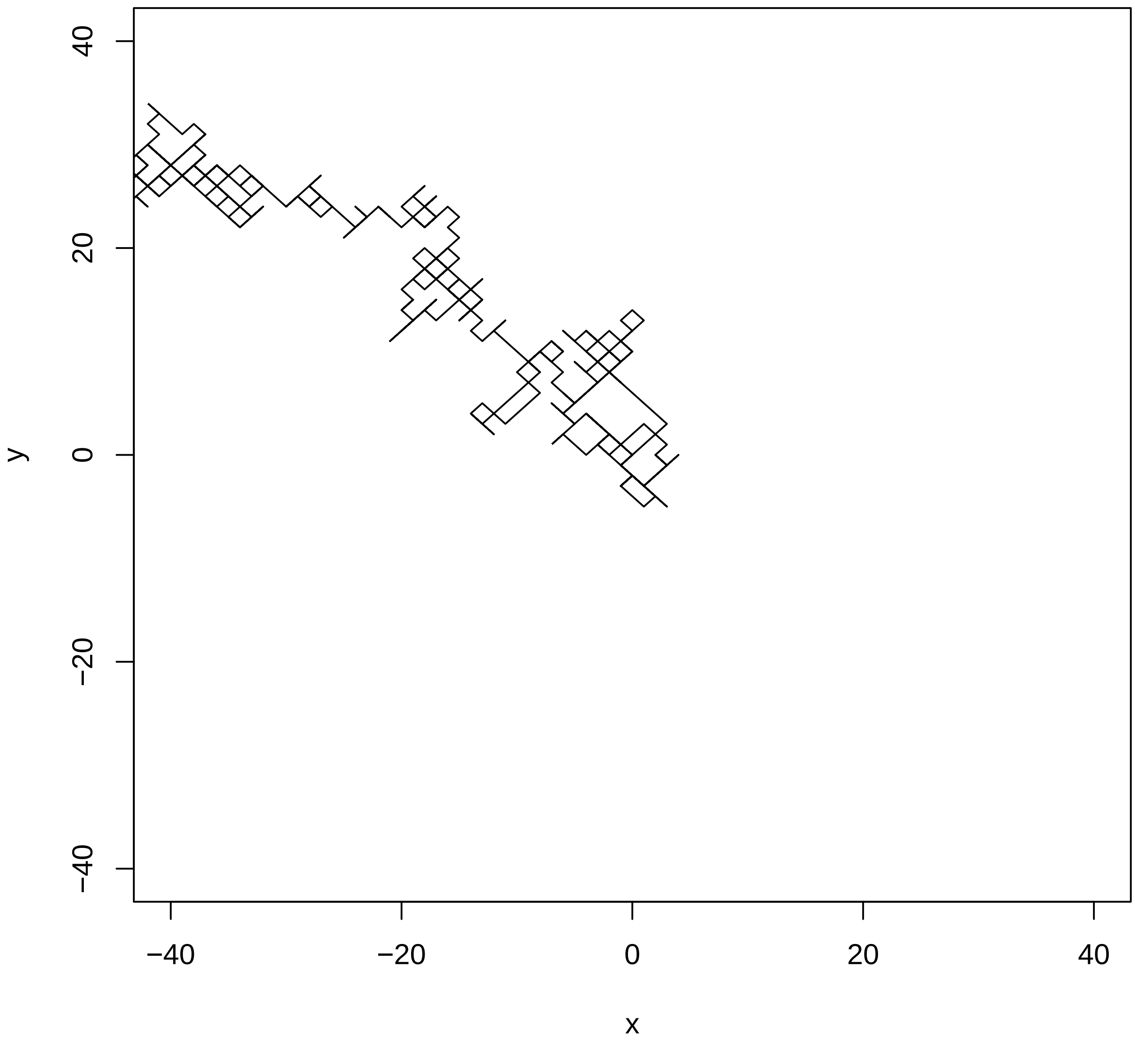

Figure 7: HW 5, Ex.3iii: 2-D random walk, seed=89012