Ma 322: Biostatistics
Homework Assignment 3

Prof. Wickerhauser

Read Chapter 8, pages 108–134 of our text.

1. Every morning, Rosencrantz flips a coin. If it turns up heads, he rises out of bed and rolls two dice to decide what he will have for breakfast. If the sum of the dice is more than 9 he has eggs, otherwise he has cereal.

   If the coin turns up tails, Rosencrantz sleeps for another hour and then has cereal for breakfast.

   (a) Assuming a fair coin and fair dice, what is the conditional probability that Rosencrantz eats cereal for breakfast given that the coin flip turned up heads?

   (b) Assuming a fair coin and fair dice, what is the probability that Rosencrantz eats eggs for breakfast? What is the probability that he eats cereal for breakfast?

   Solution: First build the conditional probability tree, with the probabilities here written in parentheses:

   - Coin flip Heads (1/2):
     - Dice sum more than 9 (6/36): Eggs
     - Dice sum 8 or less: (30/36): Cereal
   - Coin flip Tails (1/2): Cereal

   (a) The part of the tree below "Coin flip Heads" has the desired conditional probability: 6/36 that he has eggs, 30/36 that he has cereal, given that the coin flip was Heads.

   (b) Multiplying and adding to get the totals, the probability that Rosencrantz eats eggs is $(1/2)(6/36)+(1/2)(0) = 3/36$ and the probability that he eats cereal is $(1/2)(30/36) + (1/2)(1) = 33/36$. □

2. Use the joint and marginal probability table 8-2 on page 114 of our text to answer the following questions:

   (a) What is P(G at P1)?

   (b) What is P(G at P2)?

   (c) What is P(G at both P1 and P2)?

   (d) What is P(T at P1 | G at P2)?

   Solution: (a) This is the marginal probability value for column G: 0.2

   (b) This is the marginal probability value for row G: 0.1

   (c) This is the joint probability value for row G, column G: 0.0

   (d) This is the P(T at P1 and G at P2)/ P(G at P2) = 0.1/0.1 = 1. □

3. Suppose that X, Y are continuous random variables, each taking values in $[0,1]$, with joint probability density function

$$f(x,y) = c(1 - x^2 y),$$

where c is a constant.

(a) Find c.

(b) Find the marginal density function $f_X(x)$.

(c) Find the marginal density function $f_Y(y)$.

(d) Are X and1 Y independent? Supply a proof or a counterexample to justify your answer.

(e) Compute $P(x < \frac{1}{2} \mid y = \frac{1}{2})$.

Solution: (a) Since $\iint f(x,y)\,dxdy = 1$, we must have

$$1 = c\int_0^1 dx \int_0^1 dy\,(1 - x^2 y) = c\int_0^1 dx\,(1 - \frac{x^2}{2}) = c(\frac{5}{6}),$$

so $c = 6/5$.

(b) Integrate out the y variable:

$$f_X(x) = \int_0^1 f(x,y)\,dy = \int_0^1 \frac{6}{5}(1 - x^2 y)\,dy = \frac{6}{5}(1 - \frac{x^2}{2}).$$

(c) Integrate out the x variable:

$$f_Y(y) = \int_0^1 f(x,y)\,dx = \int_0^1 \frac{6}{5}(1 - x^2 y)\,dx = \frac{6}{5}(1 - \frac{y}{3}).$$

(d) If X and Y were independent then we would have $f(x,y) = f_X(x)f_Y(y)$ for all $x, y \in [0,1]$. But this is not true: the counterexample $x = 1, y = 1$ gives $f(1,1) = 0$ while $f_X(1)f_Y(1) = 6/13$.

(e) First compute the conditional probability density from the definition:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{6}{5}(1 - x^2 y)}{\frac{6}{5}(1 - \frac{y}{3})}.$$

Thus $f(x|y = 1/2) = (1 - x^2/2)/(1 - 1/6) = \frac{6}{5}(1 - \frac{x^2}{2})$. Then

$$P(x < 1/2|y = 1/2) = \int_0^{1/2} f(x|y = 1/2)\,dx = \int_0^{1/2} \frac{6}{5}(1 - \frac{x^2}{2})\,dx = \frac{6}{5}(x - \frac{x^3}{6})|_0^{1/2} = \frac{6}{5}(\frac{1}{2} - \frac{1}{48}) = \frac{23}{48}.$$

□

4. Suppose that genotypes AA, Aa, and aa have respective occurence probabilities $P_{AA} = 0.1$, $P_{Aa} = 0.2$, and $P_{aa} = 0.7$ in a population of diploid organisms.

(a) What is the probability of getting 1 AA, 2 Aa's, and 7 aa's in a random sample of 10 individuals from this population?

(b) What is the probability of getting no AA's in a random sample of 10 individuals from this population?

(c) Simulate taking 8 independent random samples of 10 individuals from this population and use the simulation to estimate P_{AA}, P_{Aa}, and P_{aa}.

Solution: (a) Use the following R commands:

```
pAA<-0.1; pAa<-0.2; paa<-0.7; p<-c(pAA,pAa,paa);
dmultinom(c(1,2,7), size=10, prob=p)       #  0.1185902
```

(b) We must sum over all samples of size 10 that have no AAs. Use the following R commands:

```
pAA<-0.1; pAa<-0.2; paa<-0.7; p<-c(pAA,pAa,paa);  s<-0;
for (Aa in 0:10) s<-s+dmultinom(c(0,Aa,10-Aa), size=10, prob=p);
s     #  0.3486784
```

(c) Use the following R commands:

```
pAA<-0.1; pAa<-0.2; paa<-0.7; p<-c(pAA,pAa,paa);
sim<-rmultinom(8, size=10, prob=p); sim
#  2    0    2    1    1    0    0    0
#  1    1    2    3    1    1    1    3
#  7    9    6    6    8    9    9    7
rowMeans(sim)/10         # 0.075 0.1625 0.7625
```

□

5. (a) Let $B = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$. Compute $\Sigma = B^T B$ and find Σ^{-1}.

(b) Use `persp()` and `contour()` to display the bivariate normal density with mean $(0, 1)$ and covariance matrix Σ (from part a) over the range $[-9, 9] \times [-9, 9]$

(c) Use `mvrnorm()` to simulate 1000 samples from the bivariate normal density of part b. Display the resulting scatterplot.

(d) Display the histograms of the X and Y marginal densities of the simulation in part c.

(e) Calculate the covariance matrix of the samples in part c. Hint: check your work by comparing the sample covariance matrix with Σ.

Solution: (a) Compute

$$\Sigma = B^T B = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}.$$

Then

$$\Sigma^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}^{-1} = \frac{1}{9}\begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix}.$$

R commands:

```
B<-matrix(c(1,1,1,-2),2,2); S <- t(B)\%*\% B; S; solve(S)
```

(b) Use the function `bvnpdf()` defined on the class website:

```
source("bvnpdf.R");
x<-seq(-9,9,by=0.25); y<-x; z<-bvnpdf(x,y,mu=c(0,1),Sigma=S );
pdf(file="persp.pdf"); persp(x,y,z);  dev.off();
pdf(file="contour.pdf"); contour(x,y,z);  dev.off();
```

(c) R commands:

```
require(MASS);
set.seed(10293847); sim<-mvrnorm(1000, mu=c(0,1), Sigma=S );
pdf(file="scatter.pdf"); plot(sim); dev.off();
```

(d) R commands:

```
pdf(file="xmargin.pdf"); hist(sim[,1]); dev.off();
pdf(file="ymargin.pdf"); hist(sim[,2]); dev.off();
```

(e) R commands:

```
cov(sim);  # or var(sim)
  2.1323976 -0.9797013
 -0.9797013  5.1052048
```

□

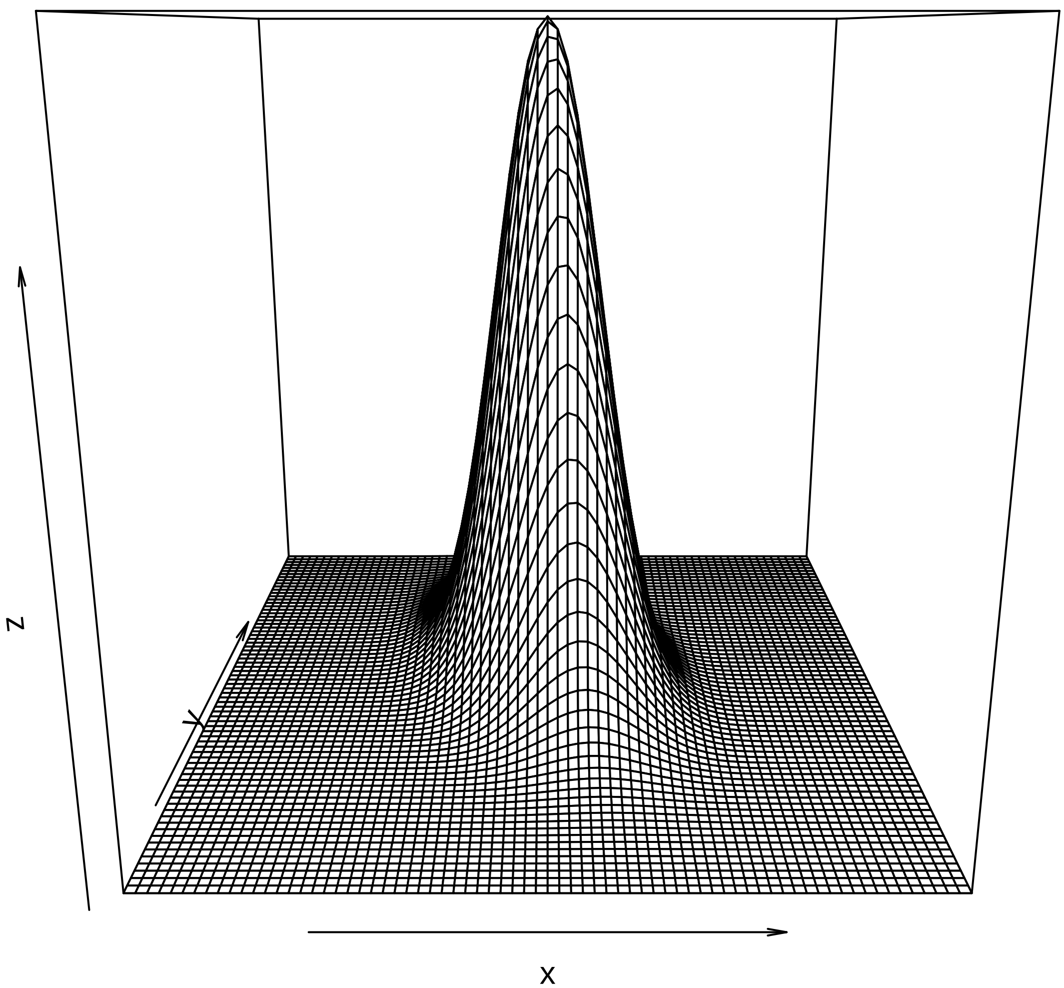

Figure 1: HW 3-5b1: Perspective Plot of a Bivariate Normal PDF

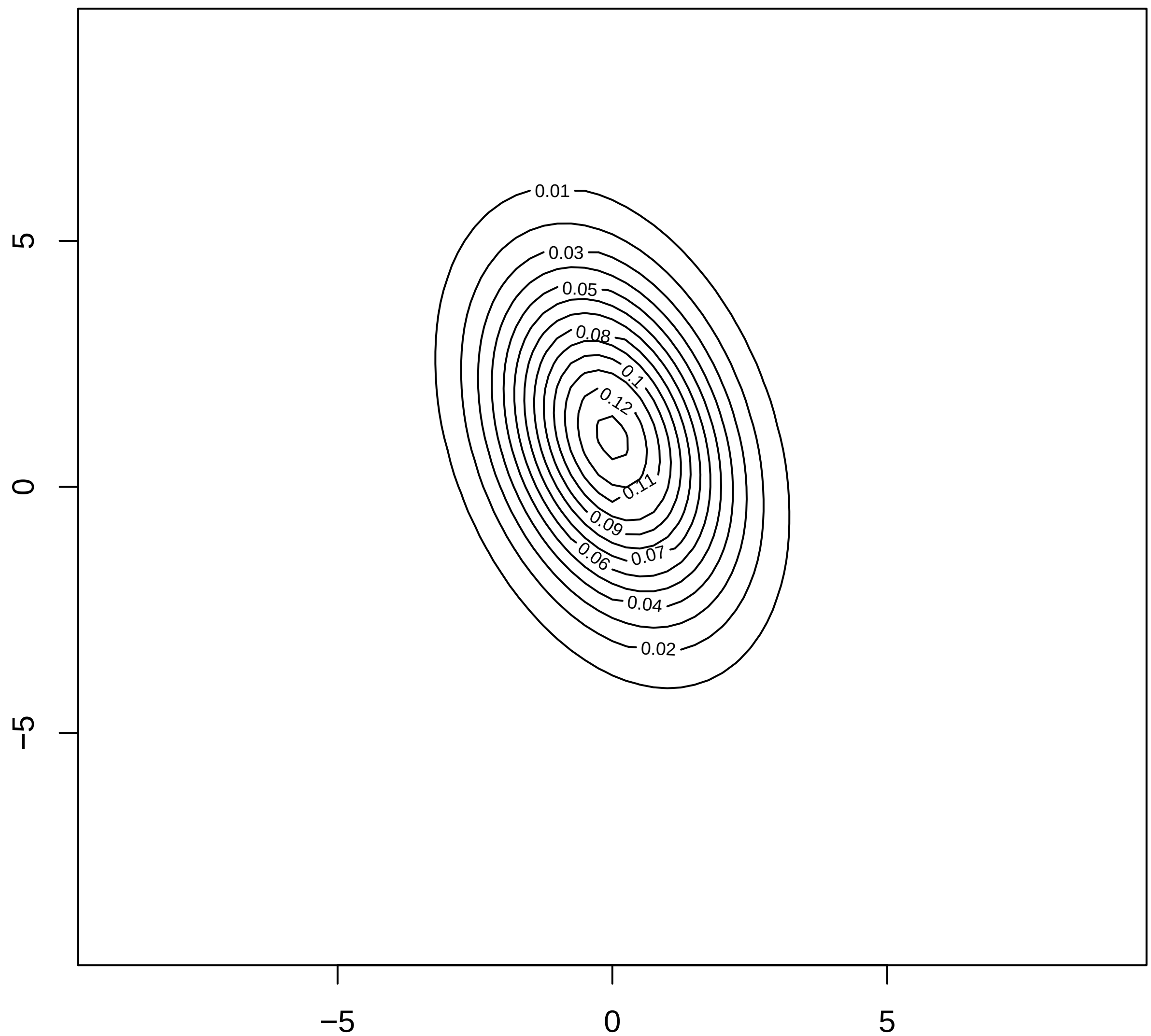

Figure 2: HW 3-5b2: Contour Plot of a Bivariate Normal PDF

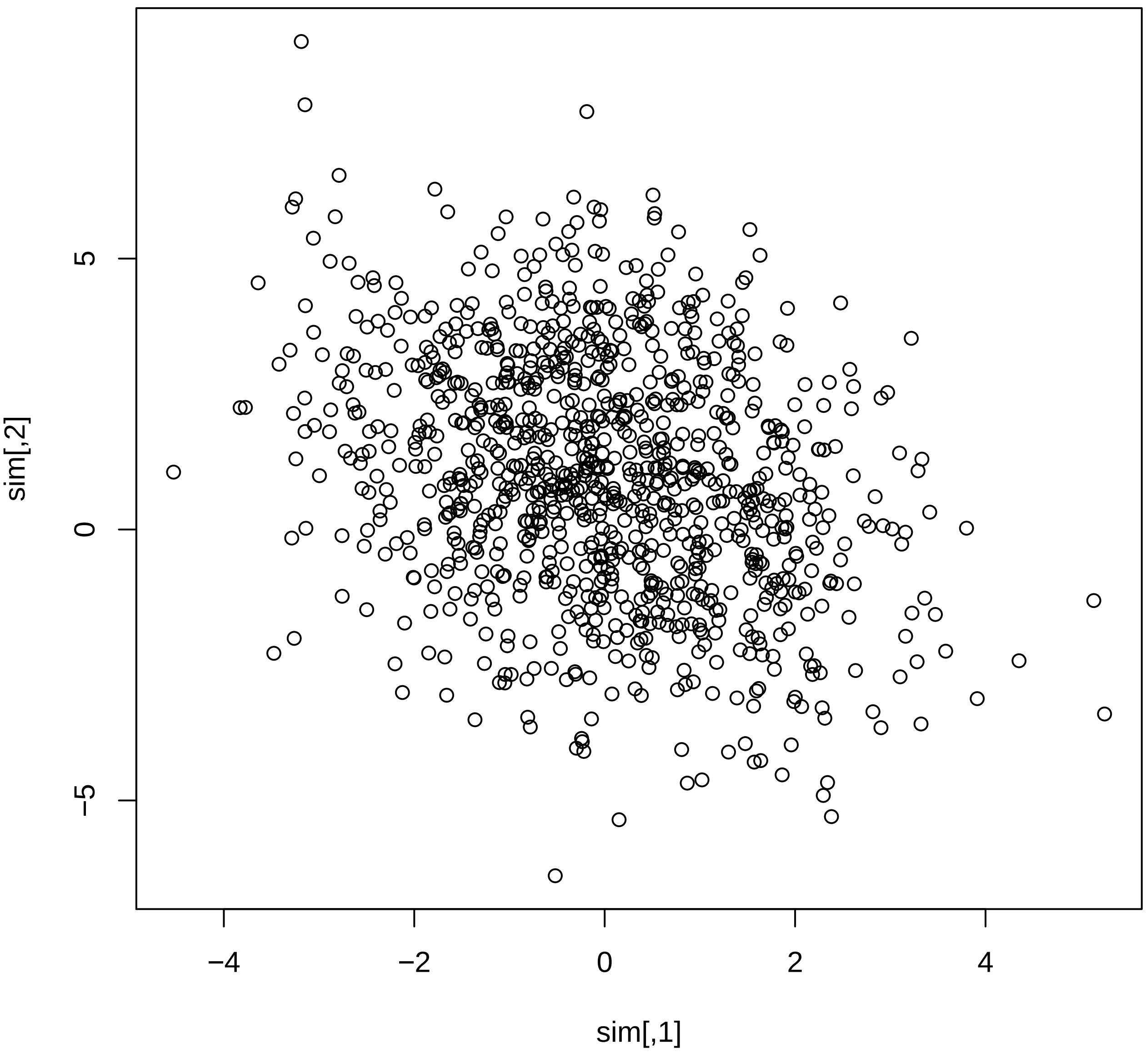

Figure 3: HW 3-5c: Scatterplot of Samples from a Bivariate Normal PDF

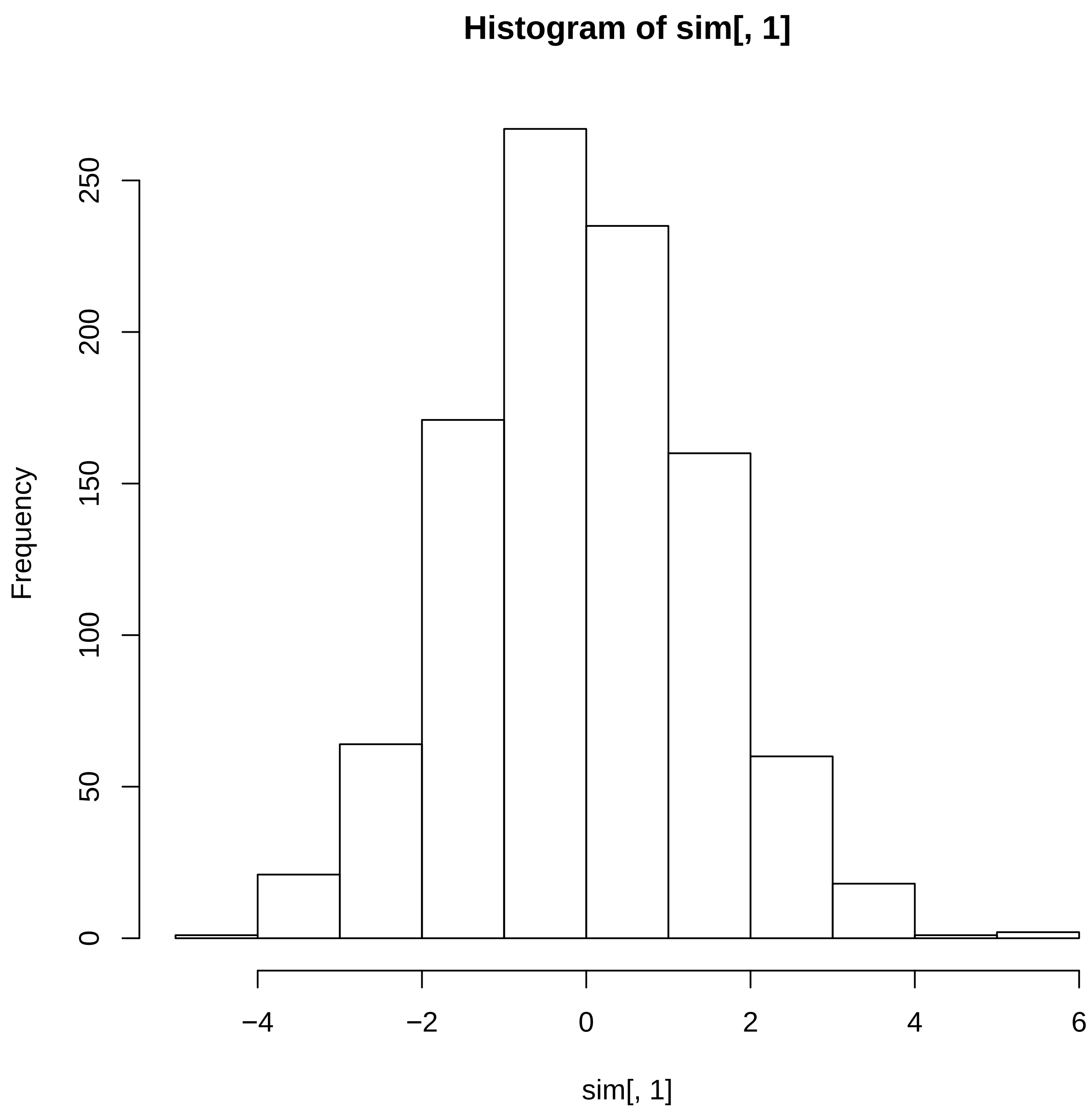

Figure 4: HW 3-5d1: X Marginal PDF of a Bivariate Normal PDF

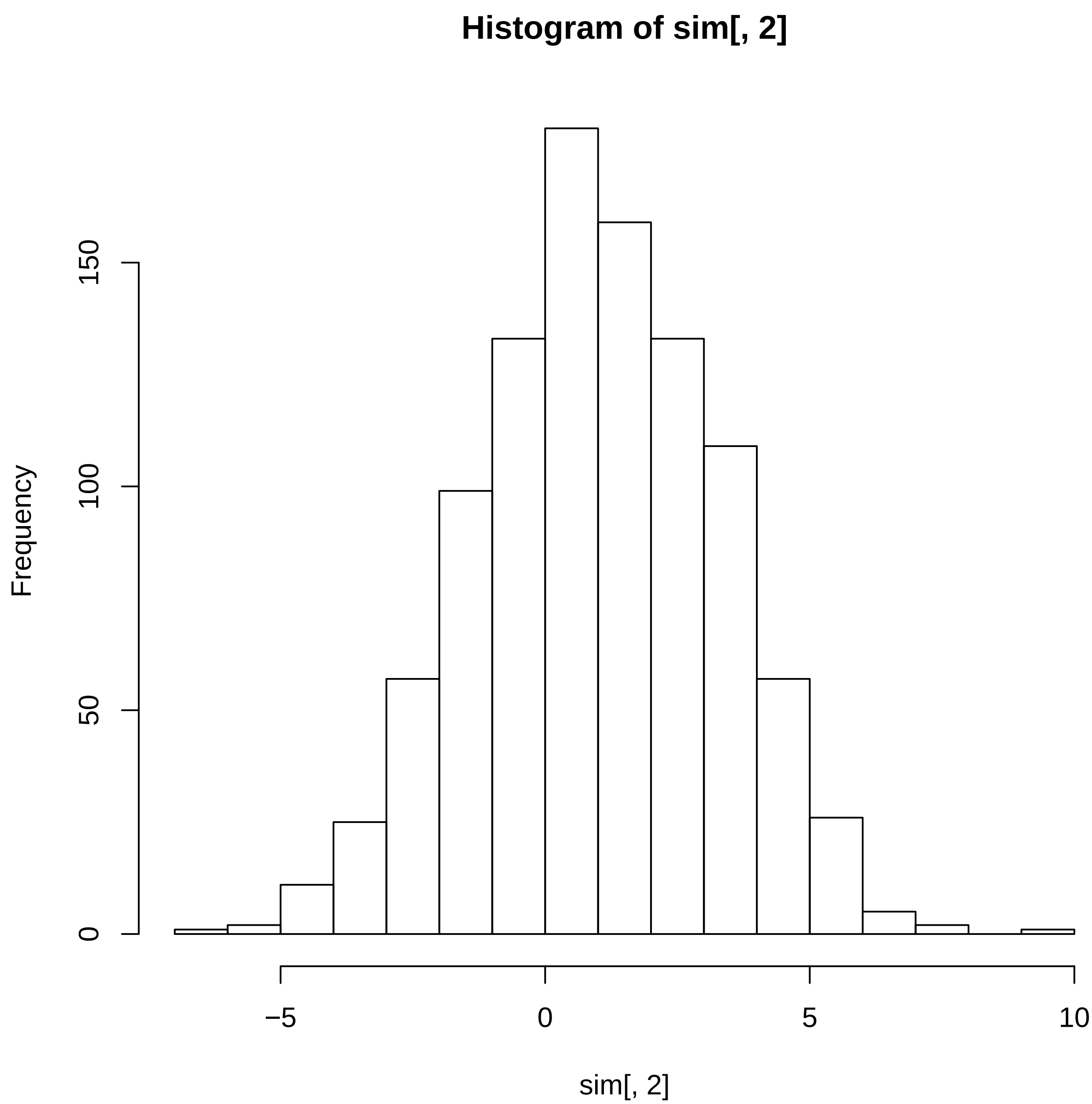

Figure 5: HW 3-5d2: Y Marginal PDF of a Bivariate Normal PDF