

Homework 03

Brown University

DATA 1010

Fall 2019

Problem 1

Eytan applied gradient descent to the function $f(x) = (x - 3)(x - 1)(x + 2)(x + 3)$ starting from $x_0 = -4$, and he found that his algorithm failed to converge because $x_2 = -4$, $x_4 = -4$, and so on. In other words, the iterates alternated between -4 and another value. What learning rate did he use?

Hint: set up an equation for ϵ and use SymPy to solve it.

Solution. We set $x_0 = -4$ and determine x_2 as a function of ϵ :

```
[ ] # solution
    using SymPy
    @vars ε
    df(x) = 4x^3 + 3x^2 - 22x - 9
    x0 = -4
    x1 = x0 - ε*df(x0)
    x2 = x1 - ε*df(x1)
```

We set this expression equal to -4 and solve:

```
[ ] float.(solve(x2 + 4, ε))
```

Only one of those values makes sense as a learning rate, so the answer is approximately 0.05849.

Problem 2

Consider the following random experiment: you flip a coin, and if it turns up heads, you roll a die. If it turns up tails, then you draw a ball from an urn containing balls labeled 1 to 11.

Define a sample space Ω for this random experiment.

Solution. The set of outcomes is

$$\Omega = \{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6), ($$

Problem 3

Suppose that E , F , and G are events. Come up with events \tilde{E} , \tilde{F} , and \tilde{G} which are pairwise disjoint and which satisfy

$$E = \tilde{E}, E \cup F = \tilde{E} \cup \tilde{F}, \text{ and } E \cup F \cup G = \tilde{E} \cup \tilde{F} \cup \tilde{G}.$$

Solution. The idea is to remove from each set any elements that belong to preceding sets: we define $\tilde{E} = E$, $\tilde{F} = F \setminus E$, and $\tilde{G} = G \setminus (E \cup F)$. These sets are pairwise disjoint, since any $\omega \in E \cup F \cup G$ appears in exactly one of the sets \tilde{E} , \tilde{F} , \tilde{G} (namely, the first one it appears in, with respect to the order E, F, G).

Problem 4

Suppose that A is the event that the high temperature in Providence next Tuesday is at least 65 degrees Fahrenheit, let B be the event that the high temperature in Providence next Tuesday is at least 60 degrees Fahrenheit, and let C be the event that it rains in Providence next Tuesday. Write each of the following events using the sets A , B , and C and the operations \cap , \cup , and c .

- (a) It will be less than 60 degrees all day and rainy next Tuesday in Providence.
- (b) The high temperature in Providence next Tuesday will be at least 60 degrees but not as high as 65 degrees.
- (c) In Providence next Tuesday, it will either be dry or warm (where warm is defined to mean "daily high of at least 65 degrees").
- (d) The daily high temperature in Providence next Tuesday will be higher than 65 degrees and less than 60 degrees.

Solution.

- (a) Since B^c and C must both occur, this event is $B^c \cap C$.
- (b) Since B must occur and A must not, this event is $B \setminus A$.
- (c) We require either C^c or A to occur. Therefore, this event is $A \cup C^c$.
- (d) We require both A and B^c , so this event is $A \cap B^c$. In this instance, $A \subset B$, so $A \cap B^c = \emptyset$ (either answer is acceptable).

Problem 5

Write a Julia program to find the exact value of the probability that in eight flips of a fair coin, no two consecutive flips turn up heads.

Two hints: (i) you can write a recursive function to generate all possible outcomes, and (ii) if you store your flip sequences as strings, you can use the function `occursin` to check for the substring `HH`.

```
[ ] # solution
    """
    Returns an array of all of the length-n
    strings with characters in {H,T}
    """
    function flipsequences(n)
        if n == 1
            ["H", "T"]
        else
            [c*s for s in flipsequences(n-1) for c in ("H", "T")]
        end
    end
    # we count up the number of sequences containing HH:
    sum([occursin("HH", s) for s in
    flipsequences(8)])//length(flipsequences(8))
```

Problem 6

The 52 cards in a standard deck (13 spades, 13 clubs, 13 diamonds, 13 hearts), are dealt into four hands of 13 cards each. What is the probability that one of the hands all of the hearts?

Solution. The number of ways to deal the four hands is $\binom{52}{13} \binom{39}{13} \binom{26}{13} \binom{13}{13}$, since there are $\binom{52}{13}$ ways of dealing the first hand, then $\binom{39}{13}$ ways of dealing the second hand given any particular dealing of the first hand, and so on.

The number of dealings that have all of the hearts in the first hand is $1 \binom{39}{13} \binom{26}{13} \binom{13}{13}$, since there is one way to put all of the hearts in the first hand, then $\binom{39}{13}$ ways to put 13 of the remaining cards in the second hand, and so on. The number of ways to deal all of the hearts to the second hand is the same, and similarly for the third and fourth hands.

So the ratio of the number of dealings where someone gets all of the hearts to the total number of dealings is

$$\frac{4}{\binom{52}{13}} = 6.3 \times 10^{-12}.$$

Problem 7

Consider a random independent sequence of letters uniformly distributed in $\{a, b, \dots, z\}$. Use Monte Carlo simulation to estimate the expected number of letters that appear in the sequence up to the first appearance of aa. Repeat with ab in place of aa. Based on your findings, is the expected time to the first aa different from the expected time to the first ab?

Solution. Using the Monte Carlo results below, we estimate the mean of the number of letters till aa around 700 and the mean of the number of letters till ab around 680.

```
function rand_letters_till(word::AbstractString)
    letters = String([rand(collect('a':'z')) for i=1:length(word)])
    ctr = length(word)
    while letters[end-length(word)+1:end] ≠ word
        letters = letters[2:end] * rand(collect('a':'z'))
        ctr += 1
    end
    ctr
end

mean(rand_letters_till("aa") for i=1:10^5) # returns ~700
mean(rand_letters_till("ab") for i=1:10^5) # returns ~680
```

If we run this block repeatedly, the aa estimate is consistently 20-30 letters larger than the ab estimate. This gives us confidence that the mean is genuinely larger.

BONUS: One way to see that the mean should be larger for aa is to think about it progressively: you wait for an aa to appear, and then you hope the next letter is aa. If it isn't, you're back where you started. For ab, you're hoping the next letter after the first aa you see is ab, but if it isn't, there's a chance that it's another aa. In that case, you're in much better position hoping the next letter is aa than you would be if you were starting over from scratch.

We can actually calculate the expectations exactly. Imagine infinitely many bettors, one for each integer $n \geq 1$, placing a wager on whether the n th letter is aa. The n th bettor pays one dollar to receive 26 dollars if they win and nothing if they lose. If the bettor wins (and if the game isn't over yet), then they wager all of their winnings on the next letter being an aa. So they will end up with either 26^2 dollars or zero dollars in that case.

Once we have seen the first aa in positions $N - 1$ and N , the bettor on position $N - 1$ will have won 26^2 dollars, and the bettor on position N will have won 26 dollars. So the winnings will be exactly $26^2 + 26 = 702$ dollars. Since each bet was fair, the amount of money paid in by the bettors must be equal in expectation to the winnings. Therefore, the expected number of dollars paid in is 702.

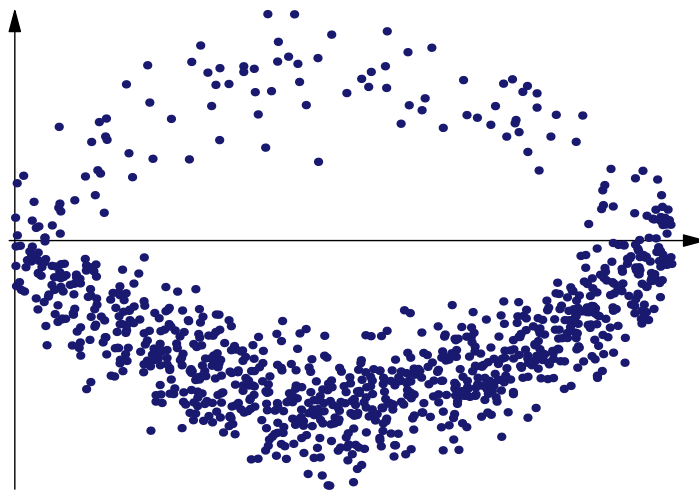
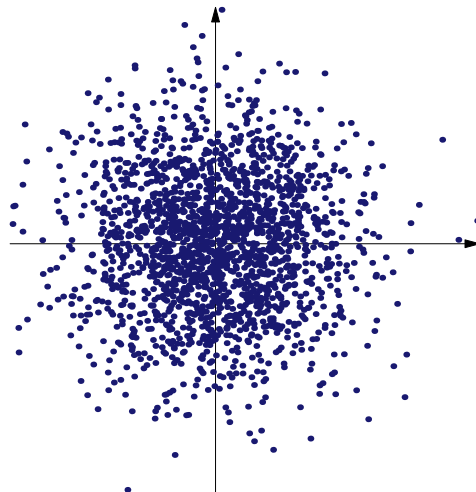
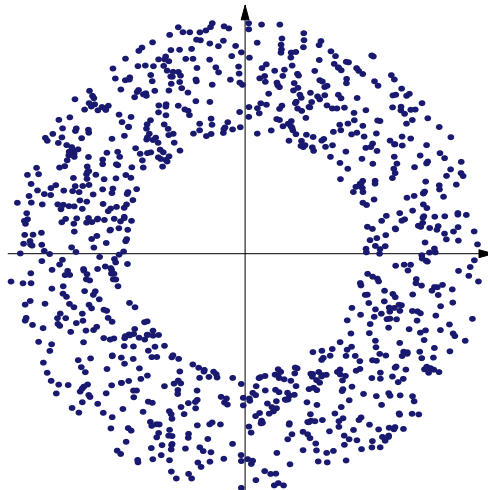
If we switch from aa to ab, the same analysis holds except that the final bettor loses. So the total winnings are $26^2 = 676$, and therefore so is the expected number of letters that have appeared when we see the first ab.

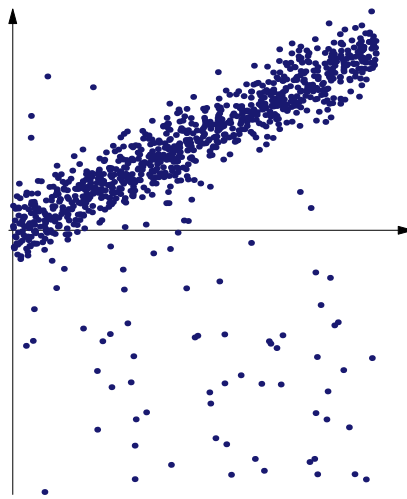
This argument uses a theorem called the *optional stopping theorem*, which says that if you're playing a progressive game which is fair at each step you decide to stop at some point, the stopped game is still fair. Those who are interested may read about this theorem

in any book which covers martingales (my favorite is David Williams' *Probability with Martingales*).

Problem 8

Suppose that many independent samples are drawn from the joint distribution of two random variables X and Y , and the results are as shown in the first figure below. Sketch, by shading, your best guess of the density function of the distribution of (X, Y) . Also sketch a graph of the function $x \mapsto \mathbb{E}[Y|X = x]$.

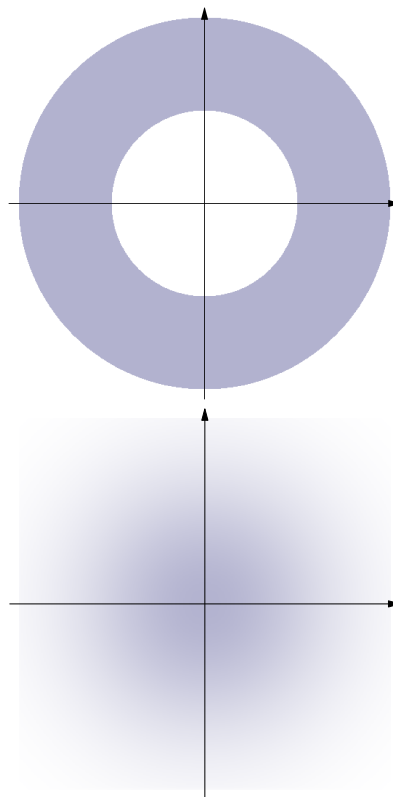


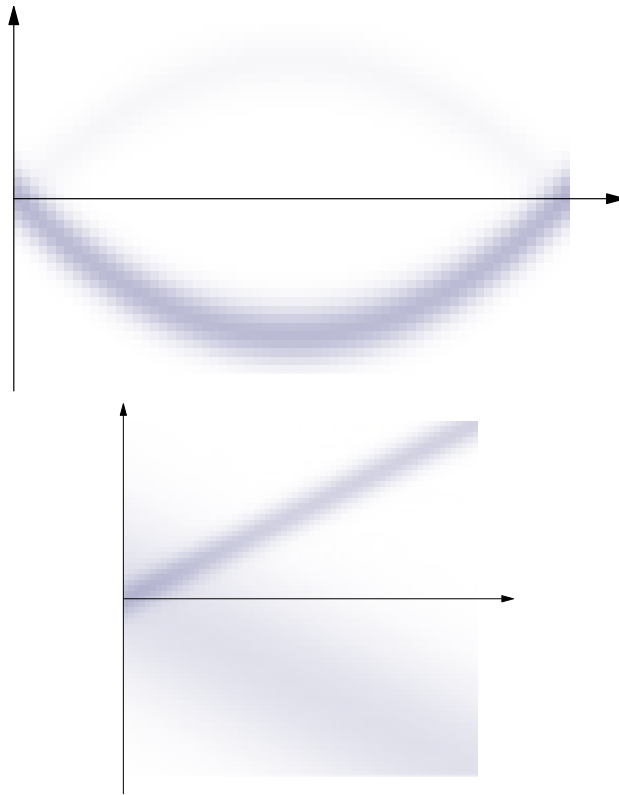


Repeat the exercise for each of the remaining figures.

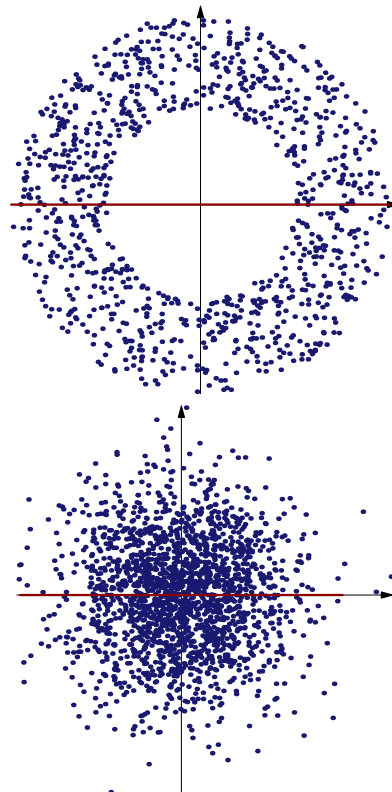
solution

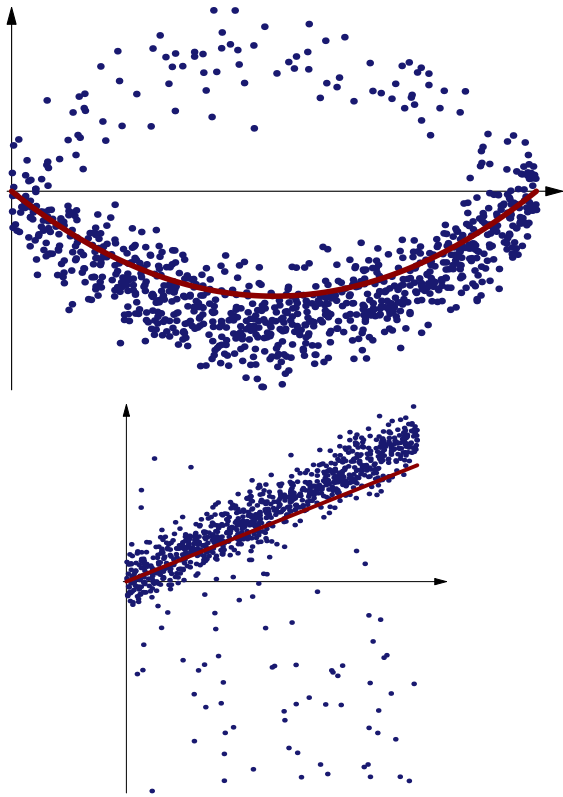
Solution. We expect that the density is large where there are lots of points, and the density is small where there are fewer points. We get something like the following:





The conditional expectation given $X = x$ is obtained by finding the center of mass of the joint density along each vertical line. We get (and these are exact, by the way):





[]

[]