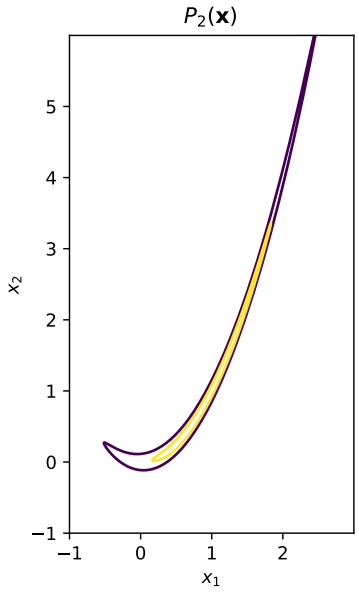
SOLUTIONS: Example Sheet 2

Christopher J. Moore

Jan 2024

1. The function $f_n(\mathbf{x}) \geq 0$, this follows from the fact that it is defined as the sum of squared quantities. From the definition $f_n(1, 1, \ldots, 1) = 0$, so this is a minimum. That this is the unique global minimum follows from that the function is a sum of squared quantities and so at any zero all of these squares must individually equal zero. In particular, this implies that $(1 - x_i)^2$ must equal zero for each i. But $x_i = 1$ is just the minimum we have already found and so this is unique. *Depending on the value of n there can be other local minima. The function f_n is a polynomial order ≤ 4 in all of the x_i, and therefore the gradient vector is a polynomial order ≤ 3 which can have as many as three real roots in each coordinate giving possibly as many as 3^n stationary points altogether. To say anything further, it is necessary to proceed on a case-by-case basis for different n. In fact, it can be shown that f_n has exactly 1 minimum (the global minimum) when $n = 2$ or 3 and exactly two minima (one global and one local) for $4 \leq n \leq 7$.*

2. See figure below, contours enclose 50% and 90% of the probability.



3. Deterministic methods are doomed to fail.

4. and 5. I find $\mathrm{E}_{\mathbf{x}}[|\mathbf{x}|] = 1.69 \pm 0.01$. But the answer is much less important than the method used to obtain it; see accompanying notebook.

This example class is intended to take the form of a discussion. There are (deliberately) many ways that this question could be tackled. What approach did everyone take? which ways worked well? how easy were they to implement? what choices were you forced to make in your implementation? and how did you test for convergence? Some possible discussion points:

- Why don't deterministic methods work? *Deterministic methods "waste" nearly all function evaluations in regions where the probability is very small. The number of points needed scales exponentially with the dimension; so even if, somehow, you manage to do it when $n = 3$, the case of $n = 4$ is much worse.*

- Whats' the purpose of "thinning" the Markov chain? To what extent is this needed to answer (i), (ii) and (iii)? *Thinning, done properly, gives us independent samples which is vital for estimating the error in*

the Monte Carlo sum. It is less important for other applications, such as making the corner plot.

- How many samples do you need for a given uncertainty on your integral? *I found I needed $N \approx 4 \times 10^4$ i.i.d. samples to get the uncertainty ± 0.01 stated above. The uncertainty scales as $N^{-1/2}$, if and only if the samples are independent. The $N^{-1/2}$ scaling is a pain; getting the error down by another order of magnitude requires 100 times as many samples! But what other methods are there? Monte Carlo may be a bad option, but it's the only option.*

- What other diagnostics of convergence are possible? *Besides the autocorrelation length, you could run multiple chains and test for similarity. E.g. Gelman-Rubin which will be discussed later in this course.*

- Once we have the $n = 3$ case working, the case of $n = 4$ is easy. *This hardly involves any extra work. Maybe the autocorrelation length of our Markov chains is a little bit longer, but the $N^{-1/2}$ scaling of our uncertainty doesn't depend on the dimension.*