MATH 466: ADVANCED APPLIED ANALYSIS: FALL 2017
COMMENTS ON HW PROBLEMS

STEVEN J. MILLER (SJM1@WILLIAMS.EDU): MATH 466, FALL 2017

ABSTRACT. A key part of any math course is doing the homework. This ranges from reading the material in the book so that you can do the problems to thinking about the problem statement, how you might go about solving it, and why some approaches work and others don't. Another important part, which is often forgotten, is how the problem fits into math. Is this a cookbook problem with made up numbers and functions to test whether or not you've mastered the basic material, or does it have important applications throughout math and industry? Below I'll try and provide some comments to place the problems and their solutions in context.

Date: January 14, 2020.

1. HW #2: DUE SEPTEMBER 15, 2017

1.1. **Assignment.** (1) Consider the solution to assigning Chapter 70 aid by Linear Regression. For each explain why or why not. (a) Will a solution exist for all data sets? (b) Will the percentages always be between 0 and 1? (c) Are the percentages stable under small changes? (d) Will the three percentages sum to 1? (2) Was \$70,000 a good price for the Cleveland Indians winning 15 or more games in a row?

1.2. **Solutions. Problem #1:** (a) A solution exists so long as the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is invertible. (b) No, the percentages will not always be in $[0,1]$. (c) The percentages are stable under small changes by linearity. (d) They will sum to 1, which is interesting. For $i \in \{1,2,3\}$ we have the model

$$\overrightarrow{y}_i \ = \ \mathbf{X}\overrightarrow{\beta}_i.$$

The theory of linear regression, obtained by doing a least squares analysis, gives us that the best-fit values of β are

$$\overrightarrow{\beta}_i \ = \ (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\overrightarrow{y}_i;$$

notice that for all three schools it is the same combination of matrices, and all that changes are the vectors $\overrightarrow{y}_i$. We are assuming we have T years of data, and for each year we have a six-vector:

$$\overrightarrow{x}_t^T \ = \ (x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5}, x_{t6}) \ = \ (1, \mathrm{LES}_{\mathrm{pop;t}}, \mathrm{WES}_{\mathrm{pop;t}}, \mathrm{MtG}_{\mathrm{pop;t}}, \mathrm{L}_{\mathrm{EQV;t}}, \mathrm{W}_{\mathrm{EQV;t}}).$$

We assume the data is such that the six columns are linearly independent; this can be checked (note that it would be very unlikely for there to be a dependency). Thus the matrix X^TX is invertible. If it were not, then as it is square it must have determinant zero, which means it has a zero eigenvector $\overrightarrow{v} \neq \overrightarrow{0}$, which means $X^TX\overrightarrow{v} = \overrightarrow{0}$. Multiplying on the left by $\overrightarrow{v}^T$, we find $||X\overrightarrow{v}||^2 = 0$, and thus $X\overrightarrow{v} = \overrightarrow{0}$; this implies that the columns of X are linearly dependent, contradiction.

Thus the estimated best values are

$$\overrightarrow{y}_i^{\mathrm{est}} \ = \ \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\overrightarrow{y}_i.$$

If we sum the above over $i \in \{1,2,3\}$ then, since the three choices of i correspond to percentages that add to 1, we get $\sum_i \overrightarrow{y}_i = \overrightarrow{1}$. Thus

$$\sum_{i=1}^{3} \overrightarrow{y}_i^{\mathrm{est}} \ = \ \mathbf{M}\overrightarrow{1}, \quad \text{where} \quad \mathbf{M} \ = \ \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}.$$

Notice that

$$\mathbf{MX} \ = \ \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X} \ = \ \mathbf{X};$$

thus each column of $\mathbf{X}$ is an eigenvector of M with eigenvalue 1 (*we are assuming no column of* $\boldsymbol{X}$ *is identically zero; if that happened then the corresponding variable would not be in play and the matrix* $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ *would not be invertible*).

Thus, if $\mathbf{X}$ is an $r \times c$ matrix (r rows and c columns) then $\mathbf{X}$ has c eigenvectors with eigenvalue 1. What is the size of $\mathbf{M}$? As

$$\mathbf{M} \ = \ \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}},$$

we see $\mathbf{M}$ is the product of an $r \times c$ matrix with a $c \times c$ matrix with a $c \times r$ matrix. Thus $\mathbf{M}$ is an $r \times r$ matrix with c eigenvalues of 1.

What is the relationship between c and r? Recall the rank of a matrix is the minimum of the column rank and the row rank. In practice we have more rows than columns, and thus we expect $r > c$. It all comes down to whether or not the vector of all 1's is in the column space of $\mathbf{X}$. Fortunately this is easily determined, as the first column of $\mathbf{X}$ is all 1's! Thus $\overrightarrow{1}$ is an eigenvector of $\mathbf{M}$ with eigenvalue 1, and the sum of the estimations is always 1.

Well, that's what I wrote a few years ago; turns out the result is correct but the proof is lacking. We need to show that if we have any data then the percentages sum to 1. In year t imagine we have the vector

$$\overrightarrow{x}_t^T \ = \ (x_{t1}, x_{t2}, x_{t3}, x_{t4}, x_{t5}, x_{t6}) \ = \ (1, \mathrm{LES}_{\mathrm{pop;t}}, \mathrm{WES}_{\mathrm{pop;t}}, \mathrm{MtG}_{\mathrm{pop;t}}, \mathrm{L}_{\mathrm{EQV;t}}, \mathrm{W}_{\mathrm{EQV;t}}).$$

We know if t is drawn from one of the years in our data set that

$$\overrightarrow{x}_t^T\beta \ = \ 1,$$

where β is the vector that is the sum of the three individual β_i's. We also know the row rank of X equals the column rank of X; as the column rank of X is 6 the row rank is also 6. We can therefore write our vector $\overrightarrow{x}_t$ as a linear combination of six of the rows of X, say

$$\overrightarrow{x}_t = \alpha_1 \overrightarrow{x}_{t_{k_1}} + \cdots + \alpha_6 \overrightarrow{x}_{t_{k_6}}.$$

As each vector has a 1 in its first entry, we know $\alpha_1 + \cdots + \alpha_6 = 1$; this is the key fact that was not written explicitly before. As each year from the X matrix, when dotted with β gives 1, we now have this holds for any year and any data!

Note how crucial it was in our analysis that the first entry is always a 1.

Problem #2: We can write code to simulate the answer, though of course it is possible to derive it theoretically as well.

```
winsinarow[p_, numwins_, numseasons_] := Module[{},
  (* p is the probability of a win *)
  (* numwins is the target number of wins in a row *)
  (* numseasons is the number of seasons we simulate *)
  successes = 0;
  For[n = 1, n <= numseasons, n++,
   {
    streak  = 0;
    For[g = 1, g <= 162, g++,
     {
      If[Random[] <=  p, streak = streak + 1, streak = 0];
      If[streak >= numwins,
       {
        g = 2000;
        successes = successes + 1;
        }];
      }]; (* end of g loop *)
    }]; (* end of n loop *)
  Print["Percent of time have winning streak of at least ", numwins,
   " games with probability ", 100.0 p, "% of winning a game is ",
   100.0 successes / numseasons, "%."];
  Print["Number of successful seasons is ", successes, " out of ",
   numseasons, "."];
  ]

winsinarow[.5, 15, 1000000]
Percent of time have winning streak of at least 15 games with
    probability 50.% of winning a game is 0.2275%.
Number of successful seasons is 2275 out of 1000000.

winsinarow[.6, 15, 1000000]
Percent of time have winning streak of at least 15 games with
    probability 60.% of winning a game is 2.7924%.
Number of successful seasons is 27924 out of 1000000.
```

Can also solve theoretically and exactly.

```
streakresult[k_, p_, games_] := Module[{},
   For[ell = 0, ell <= k, ell++, y[ell] = 1];
```

```
   For[n = k, n <= games, n++,
    y[n] = Sum[p^(ell - 1) (1 - p) y[n - ell], {ell, 1, k}]
    ];
   Print["Probability win at least ", k, " in a row in ", games,
    " games is ", 100. - 100*y[games], "%."];
   ];

streakresult[15, .5, 162]
streakresult[15, .6, 162]

Probability win at least 15 in a row in 162 games is 0.227149%.
Probability win at least 15 in a row in 162 games is 2.78014%.
```

Here is a plot of the probability of having a 15 game or more winning streak as a function of the probability of winning a given game.

Of course, our work above is under the assumption that all opponents are equally likely to lose to the Indians. In a real season you will have stretches against stronger or weaker teams (or stronger or weaker starters). Handling that theoretically is quite difficult, and one should probably resort to simulations.

Homework #3: Due September 22, 2017: (1) Problem 11.2.5. (2) Problem A.2.7. (3) Problem 11.2.10. (4) Problem 11.3.6.

2. HW #3: Due September 22, 2017

2.1. **Assignment:** (1) Problem 11.2.5. (2) Problem A.2.7. (3) Problem 11.2.10. (4) Problem 11.3.6.

2.2. **Solutions:** From Professor Morrison (when he took Math 406 with me in Spring '18).

(1) Problem 11.2.5. If $\langle f, f\rangle, \langle g, g\rangle < \infty$, we know that $\langle f, g\rangle < \infty$. However, there do exist $f, g : [0,1] \to \mathbb{C}$ such that $\int_0^1 |f(x)|dx, \int_0^1 |g(x)|dx < \infty$ but $\int_0^1 f(x)\overline{g}(x)dx = \infty$. Consider, for instance, $f(x) = g(x) = \left(\frac{1-x^2}{x}\right)^{.5}$. Although this does not integrate to a nice closed form, Mathematica gives us that $\int_0^1 |f(x)|dx = \int_0^1 f(x)dx \approx 1.748$. However, $\int_0^1 f(x)\overline{g}(x)dx = \int_0^1 \frac{1-x^2}{x}dx$ diverges. As $f \in L^1([0,1])$ but $f \notin L^2([0,1])$, we have that $f \in L^2([0,1])$ is a stronger assumption than $f \in L^1([0,1])$.

(Note: I'm not positive that this example works as it is defined on (0,1] instead of [0,1]; however, it seems to get at the idea of functions with some blow-up built into them that ends up being finite.)

(2) Problem A.2.7. The Taylor coefficients for

$$f(x) \;=\; \begin{cases} e^{-1/x^2} \text{ when } x \neq 0 \\ 0 \text{ when } x = 0 \end{cases}$$

expanded around the origin vanish.

We will show inductively that $f^{(n)}(0) = 0$, thereby implying that all Taylor coefficients vanish at the origin. We already know that it holds for $n = 0$, so we must simply show that if it holds for $n = k$, it will hold for $k + 1$. By our hypothesis $f^{(k)}(0) = 0$, and the chain rule along with various other rules of differentiation gives us that $f^{(k)}(x) = e^{-1/x^2}\left(\sum_{n=0}^{N} \frac{b_n}{x^{c_n}}\right)$ for $x \neq 0$, where $N, c_n \in \mathbb{N}_0$ and $b_n \in \mathbb{R}$. Let us now calculate $f^{(k+1)}(0)$.

$$\begin{aligned} f^{(k+1)}(0) &= \lim_{h\to 0} \frac{f^{(k)}(0+h) + f(0)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} e^{-1/h^2} \left(\sum_{n=0}^{N} \frac{b_n}{h^{c_n}}\right). \end{aligned}$$

We will now calculate $\lim_{h\to 0^+}$ and $\lim_{h\to 0^-}$ and show that they are equal.

$$\begin{aligned} \lim_{h\to 0^+} \frac{1}{h} e^{-1/h^2} \left(\sum_{n=0}^{N} \frac{b_n}{h^{c_n}}\right) &= \lim_{t\to\infty} te^{-t^2} \left(\sum_{n=0}^{N} b_n t^{c_n}\right) \\ &=0 \end{aligned}$$

since we have a polynomial function multiplied by a negative exponential. Similarly,

$$\begin{aligned} \lim_{h\to 0^-} \frac{1}{h} e^{-1/h^2} \left(\sum_{n=0}^{N} \frac{b_n}{h^{c_n}}\right) &= \lim_{t\to -\infty} te^{-t^2} \left(\sum_{n=0}^{N} b_n t^{c_n}\right) \\ &=0 \end{aligned}$$

since e^{-t^2} is unchanged by the fact that t is now approaching $-\infty$. As the limits from the left and right are equal, we conclude that

$$f^{(k+1)}(0) = \lim_{h\to 0} \frac{1}{h} e^{-1/h^2} \left(\sum_{n=0}^{N} \frac{b_n}{h^{c_n}}\right) = 0$$

and that all Taylor coefficients of $f(x)$ vanish at the origin.x

This implies that a Taylor series expansion does not correspond to a unique function. Consider, for instance, the Taylor series expansion $\sum_{n=0}^{\infty} \frac{x^n}{n!}$, which is of course about the origin. Both e^x and $e^x + f(x)$ have this same expansion about 0, and are not the same function.

(3) Problem 11.2.10. We define the Dirichlet Kernel as follows, and give an equivalent formula:

$$D_N(x) := \sum_{n=-N}^{N} e_n(x) = \frac{\sin((2N+1)\pi x)}{\sin \pi x}.$$

Recall the geometric series formula

$$\sum_{n=N}^{M} r^n = \frac{r^N - r^{M+1}}{1-r},$$

and note that $e_n(x) = (e^{2\pi i x})^n$. This allows us to write

$$\sum_{n=-N}^{N} e_n(x) = \sum_{n=-N}^{N} (e^{2\pi i x})^n = \frac{e^{-2\pi i x N} - e^{2\pi i x(N+1)}}{1 - e^{2\pi i x}}.$$

Consider the numerator of this expression. Factoring out $e^{\pi i x}$, we may write

$$\begin{aligned} e^{-2\pi i x N} - e^{2\pi i x(N+1)} =& e^{\pi i x}\left(e^{-\pi i x(2N+1)} - e^{\pi i x(2N+1)}\right) \\ =& -2ie^{\pi i x}\sin((2N+1)\pi x). \end{aligned}$$

Considering the denominator of our expression and factoring out $e^{\pi i x}$, we may write

$$\begin{aligned} 1 - e^{2\pi i x} =& e^{\pi i x}(e^{-\pi i x} - e^{\pi i x}) \\ =& -2ie^{\pi i x}\sin(\pi x). \end{aligned}$$

These two expressions allow us to write

$$D_N(x) = \frac{-2ie^{\pi i x}\sin((2N+1)\pi x)}{-2ie^{\pi i x}\sin(\pi x)} = \frac{\sin((2N+1)\pi x)}{\sin \pi x}$$

as claimed.

We define the Fejer Kernel as follows, and give an equivalent formula:

$$F_N(x) = \frac{1}{N}\sum_{n=0}^{N-1} D_n(x) = \frac{\sin^2(N\pi x)}{N\sin^2 \pi x}.$$

We will prove this formula by induction. Note that it will suffice to show that $\sum_{n=0}^{N-1} D_n(x) = \frac{\sin^2(N\pi x)}{\sin^2 \pi x}$ for all $N \geq 1$. In the case of $N = 1$, we have $D_0 = 1$ (the only term in our sum), and $\frac{\sin^2(\pi x)}{\sin^2 \pi x} = 1$, so our claim holds for the base case.

Let us assume that $\sum_{n=0}^{k-1} D_n(x) = \frac{\sin^2(k\pi x)}{\sin^2 \pi x}$; we will show that the formula holds for $N = k+1$. Note that

$$\begin{aligned} \sum_{n=0}^{(k+1)-1} D_n(x) =& \frac{\sin^2(k\pi x)}{\sin^2 \pi x} + D_k(x) \\ =& \frac{\sin^2(k\pi x)}{\sin^2 \pi x} + \frac{\sin((2k+1)\pi x)}{\sin \pi x} \\ =& \frac{\sin^2(k\pi x) + \sin((2k+1)\pi x)\sin(\pi x)}{\sin^2 \pi x}. \end{aligned}$$

Rewriting the numerator of this expression in exponential form, we have

$$\begin{aligned}\sin^2(k\pi x)+&\sin((2k+1)\pi x)\sin(\pi x) =\\ &(e^{ik\pi x}-e^{-ik\pi x})^2/(2i)^2+(e^{i(2k+1)\pi x}-e^{-i(2k+1)\pi x})(e^{i\pi x}-e^{i\pi x})/(2i)^2\\ &=(e^{i2k\pi x}+e^{-i2k\pi x}-2+e^{i2(k+1)\pi x}-e^{i2k\pi x}-e^{-i2k\pi x}+e^{-i2(k+1)\pi x})(-1/4)\\ &=(e^{i2(k+1)\pi x}+e^{-i2(k+1)\pi x}-2)(-1/4)\\ &=(2\cos(4(k+1)\pi x)-2)(-1/4)\\ &=(2(1-2\sin^2(2(k+1)\pi x))-2)(-1/4)\\ &=-4\sin^2(2(k+1)\pi x)(-1/4)\\ &=\sin^2(2(k+1)\pi x).\end{aligned}$$

We may now write

$$\begin{aligned}\sum_{n=0}^{(k+1)-1} D_n(x) =&\frac{\sin^2(k\pi x)+\sin((2k+1)\pi x)\sin(\pi x)}{\sin^2\pi x}\\ =&\frac{\sin^2(2(k+1)\pi x)}{\sin^2\pi x}.\end{aligned}$$

Thus we have shown that if our claim holds for k, it also holds for $k+1$. Combined with the base case, this completes the proof.

(4) Problem 11.3.6. The Weierstrass Approximation Theorem, which states that any continuous periodic function can be uniformly approximated by trigonometric polynomials, implies the original version of Weierstrass' Theorem, which states that if f is continuous on $[a,b]$, then for any $\varepsilon > 0$ there is a polynomial $p(x)$ such that $|f(x)-p(x)| < \varepsilon$ for all $x \in [a,b]$.

Let f be continuous on $[a,b]$, and let $\varepsilon > 0$. By the Weierstrass Approximation Theorem, we know there exists a finite trigonometric polynomial $T(x)$ such that $|f(x)-T(x)| < \varepsilon/2$ for all $x \in [a,b]$ (as we may construct a continuous, periodic function that is equal to f at all points in $[a,b]$). Note that by the definition of trigonometric polynomials, we may write $T(x) = \sum_{n=1}^{N} c_n e_n(x)$. Also note that we may write $c_n e_n(x) = c_n \sum_{n=0}^{\infty} \frac{1}{n!(2\pi i)^n} x^n$. As this is a convergent sum, we may choose k such that $|c_n e_n(x) - c_n e_n'(x)| < \varepsilon/(2N)$ for all $x \in [a,b]$, where $c_n e_n'(x) = c_n \sum_{n=0}^{k} \frac{1}{n!(2\pi i)^n} x^n$. Let us consider $p(x) = \sum_{n=1}^{N} c_n e_n'(x)$, a finite polynomial. Note that

$$\begin{aligned}|T(x)-p(x)| =&\left|\sum_{n=1}^{N} c_n e_n(x) - \sum_{n=1}^{N} c_n e_n'(x)\right|\\ =&\left|\sum_{n=1}^{N}(c_n e_n(x) - c_n e_n'(x))\right|\\ \leq&\sum_{n=1}^{N}|c_n e_n(x) - c_n e_n'(x)|\\ <&N(\varepsilon/2N)\\ =&\varepsilon/2.\end{aligned}$$

Thus we have that $|f(x)-T(x)| < \varepsilon/2$ and $|T(x)-p(x)| < \varepsilon/2$ for all $x \in [a,b]$. By the triangle inequality, we know that $|f(x)-p(x)| < \varepsilon/2+\varepsilon/2 = \varepsilon$ for all $x \in [a,b]$. Thus if f is continuous on $[a,b]$, then for any $\varepsilon > 0$ there is a polynomial $p(x)$ such that $|f(x)-p(x)| < \varepsilon$ for all $x \in [a,b]$.

3. HW #4: DUE OCTOBER 18, 2017

Assignment: Due Wednesday, Oct 18: Exercise 12.3.8, 12.3.16, 12.3.19, and 5.5.6 (just think about the sine part of the question, no need to write anything).

12.3.8: Prove for k a positive integer, if $\alpha \in \mathbb{Q}$ then $\{n^k\alpha\}$ is periodic while if $\alpha \notin \mathbb{Q}$ then no two $\{n^k\alpha\}$ are equal.

Solution: Let $\alpha = p/q$ be a rational in lowest terms. Note that we can write any n as $mq + \ell$ with m an integer and $0 \le \ell < q$. As $(mq+\ell)^k \equiv \ell^k \bmod q$, we have $n^k\alpha \bmod 1$ cycles through $\ell^k\alpha \bmod 1$, and thus is periodic with period at most q.

If α is irrational, imagine $n_1^k\alpha = n_2^k\alpha \bmod 1$. Then there is an integer m such that $n_1^k\alpha = n_2^k\alpha + m$, or after simplifying $\alpha = m/(n_1^k - n_2^k)$. If $n_1 \neq n_2$ then α is an integer; contradiction.

12.3.16: Show that the sequence $\{n!e\}$ is not equidistributed. In fact, the only limit point of this sequence is 0.

Solution: Since $e = \sum_{k=0}^{\infty} 1/k!$, we have $n!e = \text{integer} + n!\sum_{k=n+1}^{\infty} 1/k!$. For n large, $n!e \bmod 1$ is very close to 0:

$$n!\sum_{k=n+1}^{\infty}\frac{1}{k!} \;<\; \frac{1}{n+1}\sum_{m=0}^{\infty}\left(\frac{1}{n+1}\right)^m \;=\; \frac{1}{n+1}\left(1-\frac{1}{n+1}\right)^{-1} \;=\; \frac{1}{n}.$$

Thus the sequence cannot be equidistributed as it is converging to zero.

12.3.19: Triangle Inequality: Prove for $a, b \in \mathbb{C}$ that $|a+b| \le |a| + |b|$.

Solution: There are lots of proofs; here's an algebra intense but a follow-your-nose. Say $a = a_x + ia_y$ and $b = b_x + ib_y$. Then $a + b = (a_x + b_x) + i(a_y + b_y)$, and

$$\begin{aligned} |a+b| &= \sqrt{(a_x+b_x)^2 + (a_y+b_y)^2} \\ |a| &= \sqrt{a_x^2 + a_y^2} \\ |b| &= \sqrt{b_x^2 + b_y^2}. \end{aligned}$$

Now, if we have an inequality between two non-negative numbers then squaring preserves the inequality, and thus it suffices to show $|a+b|^2 \le (|a|+|b|)^2 = |a|^2 + 2|a||b| + |b|^2$. We have

$$(a_x+b_x)^2 + (a_y+b_y)^2 \;\text{ versus }\; (a_x^2+a_y^2) + 2\sqrt{a_x^2+a_y^2}\sqrt{b_x^2+b_y^2} + (b_x^2+b_y^2).$$

Expanding and subtracting common terms gives

$$2(a_xb_x + a_yb_y) \;\text{ versus }\; 2\sqrt{a_x^2+a_y^2}\sqrt{b_x^2+b_y^2}.$$

Without loss of generality we can assume a_x, b_x, a_y and b_y are all non-negative, as that leads to the left side being as large as possible. In that case we cancel the 2's and then square, finding

$$a_x^2b_x^2 + 2a_xb_xa_yb_y + a_y^2b_y^2 \;\text{ versus }\; (a_x^2+a_y^2)(b_x^2+b_y^2) \;=\; a_x^2b_x^2 + a_x^2b_y^2 + a_y^2b_x^2 + a_y^2b_y^2.$$

Canceling again and then subtracting what's left on the LHS yields

$$0 \;\text{ versus }\; a_x^2b_y^2 - 2a_xb_xa_yb_y + a_y^2b_x^2 \;=\; (a_xb_y - a_yb_x)^2;$$

as the final expression is non-negative we obtain the Triangle Inequality.

FIGURE 1. Plot of sum of $\cos(n)^n$.

5.5.6: Use Exercises 5.5.5 and 5.4.19 (where we prove π is irrational) to show that $\sum_{n=1}^{\infty}(\cos n)^n$ diverges; the argument of the cosine function is in radians. Harder: what about $\sum_{n=1}^{\infty}(\sin n)^n$?

Solution: Let's plot! See Figure 1.

```
list = {};
sum = 0;
For[m = 1, m <= 100000, m++,
 {
  sum = sum + 1.0 Cos[m]^m;
   list = AppendTo[list, {m, sum}]
  }]
ListPlot[list]
```

Thus it looks clear that it diverges. To show it diverges it suffices to show the individual terms do not converge to zero (note of course that if a sum converges the terms must converge to zero, but the terms converging to zero is not sufficient to ensure the sum converges). Thus we are done if we can show $\cos(n)^n$ is near 1 infinitely often.

From the referenced exercises, we can find infinitely many pairs of integers (p_m, q_m) such that $|2\pi - p_m/q_m| < 1/q_m^2$. Thus

$$\cos(p_m) \;=\; \cos\left(2\pi p_m + \theta_m/q_m\right),$$

where $|\theta_m| \le 1$. As cosine is periodic with period 2π, we have

$$\cos(p_m) \;=\; \cos(\theta_m/q_m) \;=\; 1 - \frac{\theta_m^2}{2q_m^2} + \frac{1}{4!}\frac{\theta_m^4}{q_m^4} - \cdots$$

(just use the Taylor series expansion of cosine). As $|\theta_m| \le 1$ and all the powers of θ_m are even, we obtain a lower bound for m large if we replace the infinite sum with just the first term (we have an alternating series that is strictly decreasing in absolute value), and then get another lower bound by replacing θ_m^2 with 1. Thus

$$\cos(p_m) \;=\; 1 - \frac{1}{2q_m^2} \;>\; 1 - \frac{1}{q_m}.$$

(Part of the goal here is to show that we can often be VERY crude in approximating, if we have sufficient convergence!). Raising to the p_m-power gives

$$\cos(p_m)^{p_m} \ \geq \ \left(1 - \left(\frac{1}{2q_m^2}\right)\right)^{p_m}.$$

As $6q_m \leq p_m \leq 7q_m$ (since $2\pi \approx p_m/q_m$) we see that if we replace the exponent of p_m with $7q_m$ we make the right hand side smaller (note it is close to 1 so is positive), hence

$$\cos(p_m)^{p_m} \ \geq \ \left(1 - \left(\frac{1}{2q_m^2}\right)\right)^{7q_m}.$$

We want to say this expression is close to 1. The best way to handle this, as we have such a high power, is to take a logarithm and then exponentiate. We find, using $\log(1-u) \geq -u$, that

$$\begin{aligned} p_m \log\cos(p_m) &\geq 7q_m \log\left(1 - \left(\frac{1}{2q_m^2}\right)\right) \\ &= -7q_m\left(\frac{1}{2q_m^2} + \frac{1}{2(2q_m^2)^2} + \frac{1}{3(2q_m^2)^3} + \cdots\right) \\ &> -7q_m/q_m \\ \cos(p_m)^{p_m} &\geq e^{-7/q_m}. \end{aligned}$$

As $7/q_m \to 0$, its exponential goes to 1 and thus we do find, infinitely often, that $\cos(n)^n$ is close to 1.

NOTE: as we are just trying to show that $\cos(p_n)^{p_n}$ *doesn't converge to 0, we can argue more crudely. The discussion above shows that it is arbitrarily close to 1 infinitely often; all we need is that infinitely often it is bounded away from 0.*