1. Weierstrass' approximation theorem and Fejér's theorem

Unless we say otherwise, all our functions are allowed to be complex-valued. For eg., $C[0,1]$ means the set of complex-valued continuous functions on $[0,1]$.

Theorem 1 (Weierstrass). *If $f \in C[0,1]$ and $\varepsilon > 0$ then there exists a polynomial P such that $\|f-P\|_{\sup} < \varepsilon$. If f is real-valued, we may choose P to be real-valued.*

Proof. Define $B_n f(x) := \sum_{k=0}^{n} f(k/n)\binom{n}{k}x^k(1-x)^{n-k}$, a polynomial of degree n. Make the following observations about the coefficients $p_{n,x}(k) = \binom{n}{k}x^k(1-x)^{n-k}$.

$$\sum_{k=0}^{n} p_{n,x}(k) = 1, \quad \sum_{k=0}^{n} k p_{n,x}(k) = nx, \quad \sum_{k=0}^{n} (k-nx)^2 p_{n,x}(k) = nx(1-x),$$

all of which can be easily checked using the binomial theorem. In probabilistic language, $p_{n,x}$ is a probability distribution on $0,1,\ldots,n$ whose mean is nx and standard deviation is $nx(1-x)$. From these observations we immediately get

$$\sum_{k:|\frac{k}{n}-x|\geq\delta} p_{n,x}(k) \leq \frac{1}{\delta^2 n^2}\sum_{k=0}^{n}(k-nx)^2 p_{n,x}(k) \leq \frac{x(1-x)}{n\delta^2}.$$

Thus, denoting $\omega_f(\delta) = \sup_{|x-y|\leq\delta}|f(x)-f(y)|$, we get

$$\begin{aligned}
|B_n f(x) - f(x)| &\leq \sum_{k\,:\,|\frac{k}{n}-x|<\delta} |f(x)-f(k/n)|p_{n,x}(k) + \sum_{k\,:\,|\frac{k}{n}-x|\geq\delta} |f(x)-f(k/n)|p_{n,x}(k) \\
&\leq \omega_f(\delta)\sum_{k\,:\,|\frac{k}{n}-x|<\delta} p_{n,x}(k) + 2\|f\|_{\sup}\frac{x(1-x)}{n\delta^2} \\
&\leq \omega_f(\delta) + \frac{1}{2n\delta^2}\|f\|_{\sup}.
\end{aligned}$$

First pick $\delta > 0$ so that $\omega_f(\delta) < \varepsilon/2$ and then pick $n > \frac{\|f\|_{\sup}}{\varepsilon\delta^2}$ to get $\|B_n f - f\|_{\sup} < \varepsilon$. ■

Let S^1 denote the unit circle which we may identify with $[-\pi,\pi)$. Continuous functions on S^1 may be identified with continuous functions on $I = [-\pi,\pi]$ such that $f(-\pi) = f(\pi)$. Let $e_k(t) = e^{ikt}$ for $t \in [-\pi,\pi)$. A supremely important fact is that e_k are orthonormal in $L^2(I, dt/2\pi)$, i.e., $\int_I e_k(t)\overline{e_\ell}(t)\frac{dt}{2\pi} = \delta_{k,\ell}$. The question of whether this is a complete orthonormal basis is answered to be "yes" by the following theorem.

Theorem 2 (Fejér). *Given any $f \in C(S^1)$ and $\varepsilon > 0$, there exists a trigonometric polynomial $P(e^{it}) = \sum_{k=-N}^{N} c_k e^{ikt}$ such that $\|f-P\|_{\sup} < \varepsilon$.*

Proof. Define $\hat{f}(k) = \int_I f(t)e^{-ikt}\frac{dt}{2\pi}$ and set

$$\begin{aligned}
\sigma_N f(t) &= \sum_{k=-N}^{N}\left(1-\frac{|k|}{N+1}\right)\hat{f}(k)e^{ikt} \\
&= \sum_{k=-N}^{N}\left(1-\frac{|k|}{N+1}\right)e^{ikt}\int_I f(s)e^{-iks}\frac{ds}{2\pi} \\
&= \int_I f(s)K_N(t-s)ds
\end{aligned}$$

where the *Fejér kernel* K_N is defined as

$$K_N(u) = \sum_{k=-N}^{N}\left(1-\frac{|k|}{N+1}\right)e^{iku} = \frac{1}{N+1}\frac{\sin^2\left(\frac{N+1}{2}u\right)}{\sin^2\left(\frac{u}{2}\right)}$$

The key observations about K_N (use the two forms of K_N whichever is convenient)

$$K_N(u) \geq 0 \text{ for all } u, \qquad \int\limits_I K_N(u)\frac{du}{2\pi} = 1, \qquad \int\limits_{I\setminus[-\delta,\delta]} K_N(u)\frac{du}{2\pi} \leq \frac{1}{N+1}\frac{1}{\sin^2(\delta/2)}.$$

In probabilistic language, $K_N(\cdot)$ is a probability density on I which puts most of its mass near 0 (for large N). Therefore,

$$\begin{aligned}|\sigma_N f(t) - f(t)| &\leq \int_{-\delta}^{\delta} |f(t)-f(s)|K_N(t-s)ds + \int_{I\setminus[-\delta,\delta]} |f(t)-f(s)|K_N(t-s)ds \\ &\leq \omega_f(\delta) + 2\|f\|_{\sup}\frac{1}{N+1}\frac{1}{\sin^2(\delta/2)}.\end{aligned}$$

Pick δ so that $\omega_f(\delta) < \varepsilon/2$ and then pick $N+1 > \frac{4\|f\|_{\sup}}{\varepsilon\sin^2(\delta/2)}$ to get $\|\sigma_N f - f\|_{\sup} < \varepsilon$. ■

Extensions and alternate proofs: In the following exercise, derive Weierstrass' theorem from Fejér's theorem!

Exercise 3. Let $f \in C_{\mathbb{R}}[0,1]$.

(1) Construct a function $g : [-\pi,\pi] \to \mathbb{R}$ such that (a) g is even, (b) $g = f$ on $[0,1]$ and (c) g vanishes outside $[-2,2]$.
(2) Invoke Fejér's theorem to get a trigonometric polynomials T such that $\|T - g\|_{\sup} < \varepsilon$.
(3) Use the series $e^z = \sum_{k=0}^{\infty}\frac{1}{k!}z^k$ to replace the exponentials that appear in T by polynomials. Be clear about the uniform convergence issues.
(4) Tie everything together to get a polynomial P with *real* coefficients such that $\|f - P\|_{\sup} < 2\varepsilon$.

Natural questions are whether such theorems extend to multivariable setting, for example, are polynomials in two variables dense in $C([0,1]\times[0,1])$? While one can retrace the proofs above, it is most clearly captured by the very general Stone-Weierstrass theorem.

Theorem 4 (Stone Weierstrass theorem)**.** *Let X be a compact Hausdorff space and let $\mathcal{A} \subseteq C_{\mathbb{R}}(X)$ be (a) a real vector space, (b) closed under multiplication, (c) contain constant functions and (d) separate points of X. The last condition means that for any distinct points $x, y \in X$, there is some $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. Then, $\mathcal{A}$ is dense in $C(X)$ in sup-norm.*

We skip the proof of Stone-Weierstrass theorem here (see references later). As a corollary deduce the answer to the questions raised earlier about extensions to multivariable setting.

Exercise 5.

(1) If $K \subseteq \mathbb{R}^k$ is compact, then show that polynomials in k variables are dense in the space $C(K)$. This works for both real and complex valued functions.
(2) Show that the previous statement fails in $\mathbb{C}$. More precisely, show that there exists a continuous function $f(z)$ on $\overline{\mathbb{D}} := \{z : |z| \leq 1\}$ that cannot be uniformly approximated by polynomials of the form $\sum_{k=0}^{n} c_k z^k$.
(3) State the analogue of Fejér's theorem in multi-variable setting and derive it from Stone-Weierstrass theorem.

Here is an alternate proof of Weierstrass' theorem that was suggested by Subhroshekhar Ghosh. It is in some way more natural and removes the magical Bernstein polynomials from the picture.

Exercise 6. Formulate precisely and prove the following steps. For definiteness work with real-valued function on $\mathbb{R}$.

(1) A real-analytic function can be uniformly approximated on compact sets by polynomials.

(2) Convolution[1] of a real-analytic function with a compactly-supported continuous function is also a real-analytic function.

(3) If φ is a probability density that is real-analytic, then $\varphi_\sigma(x) := \frac{1}{\sigma}\varphi(x/\sigma)$ is also a real-analytic probability density.

(4) If f is a compactly supported continuous function, then it can be uniformly approximated by real-analytic functions.

(5) Deduce Weierstrass' theorem.

There are many examples of real-analytic probability densities. We mention two (if the facts we say below are not familiar to you, just take them as facts till you see proofs in some other course).

(1) The Cauchy density $\varphi(x) = \frac{1}{\pi(1+x^2)}$. In this case, $(f * \varphi_y)(x) = u(x,y)$ where $u : \overline{\mathbb{H}} \to \mathbb{R}$ is the unique function that solves the Dirichlet problem on the upper-half plane $\mathbb{H} := \{(x,y) : y > 0\}$ with boundary condition f. What this means is that
   (a) u is continuous on $\overline{\mathbb{H}}$,
   (b) $u(\cdot,0) = f(\cdot)$,
   (c) $\Delta u = 0$ on $\mathbb{H}$.
   The point is that $(f * \varphi_y)$ is just u restricted to the line with y-co-ordinate equal to y and approaches f (at least pointwise) when $y \to 0$.

(2) The normal density $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. In this case $(f * \varphi_t)(x) = u(x,t)$ where u solves the heat equation with initial condition f. What this means is that
   (a) u is continuous on $\mathbb{R} \times \overline{\mathbb{R}_+}$,
   (b) $u(\cdot,0) = f(\cdot)$,
   (c) $\frac{\partial}{\partial t}u(x,t) = \frac{1}{2}\frac{\partial^2}{\partial x^2}u(x,t)$ on $\mathbb{R} \times \mathbb{R}_+$.
   Again, $(f * \varphi_{\sqrt{t}}) = u(\cdot,t)$ is the function ("temperature") at time t, and approaches the initial condition f (at least pointwise) as t approaches 0.

Brief history of Fejér's theorem: This is a cut-and-dried history, possibly inaccurate, but only meant to put things in perspective!

(1) The vibrating string problem is an important PDE that arose in mathematical physics, and asks for a function $u : [a,b] \times \overline{\mathbb{R}_+} \to \mathbb{R}$ satisfying $\frac{\partial^2}{\partial t^2}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t)$ for $(x,t) \in (a,b) \times \mathbb{R}_+$ and satisfying the initial conditions $u(x,0) = f(x)$ and $\frac{\partial}{\partial t}u(x,t)\mid_{t=0} = g(x)$, where f and g are specified initial conditions.

(2) Taking $[a,b] = [-\pi,\pi]$ (without loss of generality), it was observed that if $f(x) = e^{ikx}$ and $g(x) = e^{i\ell x}$, then $u(x,t) = \cos(kt)e^{ikx} + \frac{1}{\ell}\sin(\ell t)e^{i\ell x}$ solves the problem.

(3) Linearity of the system meant that if f and g are trigonometric polynomials, then by taking linear combinations of the above solution, one could obtain the solution to the vibrating string problem.

(4) Thus, the question arises, whether given f and g we can approximate them by trigonometric polynomials (and hopefully the corresponding solutions will be approximate solutions).

[1]Convolution of f and g is defined by $(f*g)(x) := \int f(u)g(x-u)du = \int f(x-u)g(u)$. It is well-defined when f is bounded and g is integrable (absolutely). One can give many other conditions on f and g, but this will suffice for us.

(5) Fourier made the fundamental observation that $e_k(\cdot)$ are orthonormal on $[-\pi,\pi]$ and deduced that if the notion of approximation is in mean-square sense (i.e., L^2 distance $\sqrt{\int |f-g|^2}$, then the best degree-n trigonometric polynomial approximation to f is $S_n f(x) \sum_{k=-n}^{n} \hat{f}(k) e^{ikx}$.

(6) But it was an open question whether $\|S_n f - f\|_{L^2} \to 0$ as $n \to \infty$. In other words, is $\{e_k\}_{k\in\mathbb{Z}}$ a complete orthonormal set for $L^2([-\pi,\pi])$?

(7) Since continuous functions are dense is $L^2[-\pi,\pi]$, it suffices to show that continuous functions can be uniformly approximated by trigonometric polynomials.

(8) It is no longer the case that $S_n f$ is the best approximation. Fejér's innovative idea was to consider averages of $S_n f$, i.e., $T_n f := \frac{1}{2n+1}\sum_{k=0}^{2n} S_k f$ (the same trigonometric polynomials that appeared in the proof!) and show that they do converge to f uniformly.

References and further reading:

(1) B.Sury, *Weierstrass' theorem - leaving no stone unturned*, a nice expository article on Weierstrass' theorem available at `http://www.isibang.ac.in/~sury/hyderstone.pdf`.

(2) Rudin, *Principles of mathematical analysis* or Simmon's *Topology and modern analysis* for a proof of Stone-Weierstrass' theorem.

(3) Katznelson, *Harmonic analysis* or many other book on Fourier series for basics of Dirichlet and Fejeér kernels.