Math 286X: Arithmetic Statistics

Spring 2020

Solutions to problem set #1

Problem 1. Fix a polynomial $f(X) \in \mathbb{Z}[X]$ of degree 1 or 2. Show that

$$\mathbb{P}(f(x) \text{ squarefree} \mid x \in \mathbb{Z}) = \prod_{p \text{ prime}} \mathbb{P}(f(x) \not\equiv 0 \mod p^2 \mid x \in \mathbb{Z}).$$

(Also think about what goes wrong in the proof for large degrees.)

Solution. If the polynomial $f(X)$ is not squarefree over $\mathbb{Q}$, both sides are clearly zero. Assume that $f(X)$ is squarefree, which implies that its discriminant is nonzero.

For any $M \geqslant 2$, we have

$$\mathbb{P}(f(x) \not\equiv 0 \mod p^2 \quad \forall p \leqslant M \mid x \in \mathbb{Z}) = \prod_{p \leqslant M} \mathbb{P}(f(x) \not\equiv 0 \mod p^2 \mid x \in \mathbb{Z})$$

by the Chinese remainder theorem. The right-hand side is decreasing as $M \to \infty$, so it must converge. (In fact, it will converge to a positive number if $\mathbb{P}(f(x) \not\equiv 0 \mod p^2 \mid x \in \mathbb{Z}) \neq 0$ for all p.)

It remains to show that the left-hand side converges to $\mathbb{P}(f(x) \text{ squarefree})$. If $f(x)$ is not squarefree, but $f(x) \not\equiv 0 \mod p^2$ for $p \leqslant M$, then $f(x) \equiv 0 \mod p^2$ for some $p > M$. It therefore suffices to show that the probability ($\mathbb{P}_{\text{sup}}$) that $f(x) \equiv 0 \mod p^2$ for some $p > M$ converges to zero as $M \to \infty$.

Note that the fact that $f(X)$ has degree 1 or 2 implies that $f(x) \ll_f T^2$ when $|x| \leqslant T$ (for large T). (The bound might depend on the coefficients of f, especially the leading coefficient.) Any prime p with $f(x) \equiv 0 \mod p^2$ must therefore satisfy $p \ll_f T$. Furthermore, if p doesn't divide the leading coefficient of f, then $f(x)$ can have at most 2 roots modulo p. If p moreover doesn't divide the discriminant of f, then Hensel's lemma shows that these roots lift to unique roots modulo p^2. We therefore have

$$\begin{aligned}
&\mathbb{P}_{\sup}(f(x) \equiv 0 \mod p^2 \text{ for some } p > M \mid x \in \mathbb{Z}) \\
&= \limsup_{T\to\infty} \mathbb{P}(f(x) \equiv 0 \mod p^2 \text{ for some } p > M \mid x \in \mathbb{Z},\ |x| \leqslant T) \\
&= \limsup_{T\to\infty} \mathbb{P}(f(x) \equiv 0 \mod p^2 \text{ for some } T \gg_n p > M \mid x \in \mathbb{Z},\ |x| \leqslant T) \\
&\ll \limsup_{T\to\infty} \sum_{M<p\ll_n T} \mathbb{P}(f(x) \equiv 0 \mod p^2 \mid x \in \mathbb{Z},\ |x| \leqslant T) \\
&\leqslant \limsup_{T\to\infty} \frac{1}{T} \sum_{M<p\ll_n T} \#\{x \in \mathbb{Z}/p^2\mathbb{Z} \mid f(x) \equiv 0 \mod p^2\} \cdot \left(\frac{2T}{p^2} + \mathcal{O}(1)\right) \\
&\ll \limsup_{T\to\infty} \frac{1}{T} \sum_{M<p\ll_n T} \left(\frac{T}{p^2} + 1\right) \\
&\ll \frac{1}{M},
\end{aligned}$$

which indeed converges to zero as $M \to \infty$. The last inequality used the fact that the number of primes $p \ll_n T$ is $o(T)$.

If the degree of $f(X)$ was 3, we would need to consider $p \ll_n T^{3/2}$. The number of such primes is far larger than T, so the above approach would fail. (Our error bound would not converge to any number less than 1 as $T \to \infty$.) $\square$

Problem 2. For each prime number p, fix a residue class $c_p \in \mathbb{F}_p$. Show that

$$\mathbb{P}(x \not\equiv c_p \mod p \quad \forall p \mid x \in \mathbb{Z}) = 0.$$

Solution. For any $M \geqslant 2$, we have

$$\begin{aligned}
&\mathbb{P}_{\sup}(x \not\equiv c_p \mod p \quad \forall p \mid x \in \mathbb{Z}) \\
&\leqslant \mathbb{P}(x \not\equiv c_p \mod p \quad \forall p \leqslant M \mid x \in \mathbb{Z}) \\
&= \prod_{p\leqslant M} \mathbb{P}(x \not\equiv c_p \mod p \mid x \in \mathbb{Z}) \\
&= \prod_{p\leqslant M} \left(1 - \frac{1}{p}\right).
\end{aligned}$$

This converges to zero as $M \to \infty$: We can rewrite it as

$$\prod_{p \leqslant M} \left(1 - \frac{1}{p}\right) = \prod_{p \leqslant M} \frac{1}{1 + p^{-1} + p^{-2} + \cdots}$$
$$= \frac{1}{\sum_{n \geqslant 1 \text{ only divisible by primes } p \leqslant M} n^{-1}},$$

which converges to zero because $\sum_{n=1}^{\infty} n^{-1} = \infty$. (We're basically computing $\frac{1}{\zeta(1)} = 0$.) $\square$

Problem 3. Fix an odd prime l. Order the quadratic number fields K by $|\operatorname{disc}(K)|$. Show that

$$\mathbb{P}(K \text{ unramified at } l \mid K \text{ quadratic number field}) = \frac{l}{l+1}.$$

Solution. If $t \neq 1$ is a squarefree integer, then $\mathbb{Q}(\sqrt{t})$ is unramified at l if and only if $t \not\equiv 0 \mod l$. The probability that t is not divisible by l is $1 - l^{-1}$. The probability that t is not divisible by l^2 is $1 - l^{-2}$. Counting quadratic number fields that are unramified at l just like we did in class (but replacing the factor $1 - l^{-2}$ by $1 - l^{-1}$) proves that

$$\mathbb{P}(K \text{ unramified at } l \mid K \text{ quadratic number field}) = \frac{1 - l^{-1}}{1 - l^{-2}} = \frac{l}{l+1}. \quad \square$$

Problem 4. Let $n \geqslant 2$. Show that the number of squarefree monic polynomials $f(X) \in \mathbb{F}_q[X]$ of degree n is $q^n - q^{n-1}$. (Hint: Every monic polynomial $a(X)$ can be written uniquely as $a(X) = f(X)g(X)^2$, where $f(X)$ is squarefree and both $f(X)$ and $g(X)$ are monic.)

Solution. Let a_n be the number of squarefree monic polynomials of degree n. Every monic polynomial $a(X)$ can be written uniquely as $a(X) = f(X)g(X)^2$, where $f(X)$ is squarefree and both $f(X)$ and $g(X)$ are monic. Write $n = \deg(a)$, $s = \deg(f)$ and $t = \deg(g)$, so that $n = s + 2t$. Since the total number of monic polynomials of degree n is q^n, we conclude that $q^n = \sum_{s,t:\ s+2t=n} a_s q^t$. We have $a_1 = q$ and the claim $a_n = q^n - q^{n-1}$ for $n \geqslant 2$ follows by induction. $\square$

Problem 5. Show that there are sets $S_p \subseteq \mathbb{F}_p$ (for prime p) such that

$$\mathbb{P}((x \bmod p) \in S_p \quad \forall p \mid x \in \mathbb{Z}) = 0,$$

but

$$\prod_p \mathbb{P}(x \in S_p \mid x \in \mathbb{F}_p) > 0.$$

Solution. Note that the proof of Problem 2 shows that we cannot pick $S_p \subsetneq \mathbb{F}_p$ for all primes p. The idea is now to imitate the nightmare scenario presented in class, but use only a sparse subset of primes.

Fix an infinite set A of prime numbers. For any $p \in A$, let $S_p \subseteq \mathbb{F}_p$ be the set of residue classes of the form $a \bmod p$, where $\lfloor \frac{1}{2}\sqrt{p} \rfloor \leqslant a \leqslant p - \lfloor \frac{1}{2}\sqrt{p} \rfloor$. For any $p \notin A$, let $S_p = \mathbb{F}_p$. For any integer $x \in \mathbb{Z}$ and any prime $p > 16x^2$ in A, we have $(x \bmod p) \notin S_p$. Therefore, there is no $x \in \mathbb{Z}$ such that $(x \bmod p) \in S_p$ for all primes p. On the other hand,

$$\prod_p \mathbb{P}(x \in S_p \mid x \in \mathbb{F}_p) = \prod_{p \in A} \frac{\#S_p}{\#\mathbb{F}_p} = \prod_{p \in A} \frac{p - 2\lfloor \frac{1}{2}\sqrt{p} \rfloor}{p} \geqslant \prod_{p \in A} \left(1 - p^{-1/2}\right)$$

Since $1 - p^{-1/2}$ converges to 1 as $p \to \infty$, one can choose the set A sufficiently sparse to make the product converge to a positive number (arbitrarily close to 1). $\square$

Problem 6. Order pairs $(x, y) \in \mathbb{N}^2$ by $\max(x, y)$. What is

$$\mathbb{P}(\gcd(x, y) = 1 \mid (x, y) \in \mathbb{N}^2)?$$

Solution. Let $M \geqslant 2$. Then,

$$\begin{aligned} &\mathbb{P}(\gcd(x, y) \not\equiv 0 \mod p \quad \forall p \leqslant M \mid (x, y) \in \mathbb{N}^2) \\ &= \prod_{p \leqslant M} (1 - \mathbb{P}(x \equiv y \equiv 0 \mod p \mid (x, y) \in \mathbb{N}^2)) \\ &= \prod_{p \leqslant M} (1 - p^{-2}), \end{aligned}$$

which converges to $\zeta(2)^{-1} = 6/\pi^2$ as $M \to \infty$.

To show that the left-hand side converges to $\mathbb{P}(\gcd(x, y) = 1 \mid (x, y) \in \mathbb{N}^2)$, we need to find an upper bound for the probability that $x \equiv y \equiv 0$ for some prime $p > M$. But

$$\begin{aligned}
&\mathbb{P}_{\text{sup}}(x \equiv y \equiv 0 \mod p \text{ for some } p > M \mid (x,y) \in \mathbb{N}^2) \\
&\ll \limsup_{T\to\infty} \mathbb{P}(x \equiv y \equiv 0 \mod p \text{ for some } p > M \mid (x,y) \in \mathbb{N},\ x, y \leqslant T) \\
&= \limsup_{T\to\infty} \mathbb{P}(x \equiv y \equiv 0 \mod p \text{ for some } p > M \mid (x,y) \in \mathbb{N}^2,\ x, y \leqslant T) \\
&\leqslant \limsup_{T\to\infty} \frac{1}{T^2} \sum_{p>M} \mathbb{P}(x \equiv y \equiv 0 \mod p \mid (x,y) \in \mathbb{N}^2,\ x, y \leqslant T) \\
&\leqslant \limsup_{T\to\infty} \sum_{p>M} \frac{1}{p^2} \\
&\ll \frac{1}{M},
\end{aligned}$$

which indeed converges to zero as $M \to \infty$. □

Problem 7 (If you know about Dirichlet series and how to make use of their complex analysis)**.** Use Dirichlet series to prove that

$$\mathbb{P}(x \text{ squarefree} \mid x \in \mathbb{N}) = \frac{1}{\zeta(2)}.$$

Solution. Consider the Dirichlet series

$$D(s) = \sum_{n \geqslant 1 \text{ squarefree}} n^{-s}.$$

Because every natural number can be written uniquely as the product of a squarefree natural number and a square, we have $\zeta(s) = D(s)\zeta(2s)$, so $D(s) = \zeta(s)/\zeta(2s)$. The right-most pole of $\zeta(s)$ is a simple pole at $s = 1$ with residue 1. The right-most zero of $\zeta(2s)$ certainly has real part less than 1/2. The right-most pole of $D(s)$ is therefore a simple pole at $s = 1$ with residue $1/\zeta(2)$. By (for example) the Wiener–Ikehara theorem, this implies that for $T \to \infty$,

$$\sum_{1 \leqslant n \leqslant T \text{ squarefree}} 1 \sim \frac{1}{\zeta(2)} \cdot T. \qquad \square$$

Problem 8. For any $t \in \mathbb{F}_q$, the discriminant of the polynomial $f_t(X) = X^3 - tX^2 + (t-3)X + 1$ is a square: $\operatorname{disc}(f_t) = (9 - 3t + t^2)^2$. Assuming the discriminant is nonzero (the polynomial $f_t(X)$ is squarefree), this implies that either $f_t(X)$ splits into linear factors, or its Galois group is the cyclic group $A_3 \subset S_3$ of degree three. Show that

$$\lim_{q\to\infty} \mathbb{P}(f_t(X) \text{ splits into linear factors} \mid t \in \mathbb{F}_q) = \mathbb{P}(g = \mathrm{id} \mid g \in A_3) = \frac{1}{3}.$$

Solution. For any $0, 1 \neq x \in \mathbb{F}_q$, there is exactly one value $r(x) \in \mathbb{F}_q$ such that $f_{r(x)}(x) = 0$. For $x = 0, 1$, there is no such value $r(x)$. The image of the map $r : \mathbb{F}_q\backslash\{0,1\} \to \mathbb{F}_q$ is the set of $t \in \mathbb{F}_q$ such that $f_t(X)$ has a root in $\mathbb{F}_q$. There are at most two values $t \in \mathbb{F}_q$ for which $\mathrm{disc}(f) = (9 - 3t + t^2)^2 = 0$. They have at most two preimages each. Any other t in the image has exactly three preimages in $\mathbb{F}_q\backslash\{0,1\}$ (because each squarefree polynomial $f_t(X)$ either splits completely or is irreducible). Therefore, the number of $t \in \mathbb{F}_q$ that split into linear factors is $\frac{q^n}{3} + \mathcal{O}(1)$. $\square$

Problem 9. Here are two ways to estimate the number $N(T)$ of pairs $(x, y) \in \mathbb{N}^2$ such that $x^2 y \leqslant T$:

a) $$N(T) = \sum_{1\leqslant x\leqslant\sqrt{T}} \#\{1 \leqslant y \leqslant \frac{T}{x^2}\} \approx \sum_{1\leqslant x\leqslant\sqrt{T}} \frac{T}{x^2} \approx T \cdot \sum_{x=1}^{\infty} \frac{1}{x^2} = \zeta(2) \cdot T.$$

b) $$N(T) = \sum_{1\leqslant y\leqslant T} \#\{1 \leqslant x \leqslant \sqrt{\frac{T}{y}}\} \approx \sum_{1\leqslant y\leqslant T} \sqrt{\frac{T}{y}} \approx \sqrt{T} \cdot \sum_{1\leqslant y\leqslant T} y^{-1/2} \approx 2 \cdot T.$$

Which is better for large T? Can you give an error bound for the better one?

Solution. To make these estimates precise, use that $\lfloor a \rfloor = a + \mathcal{O}(1)$. We also approximate sums $\sum_{a\leqslant x\leqslant b} f(x)$ by integrals $\int_a^b f(x)\mathrm{d}x$ for monotonic functions f:

$$\sum_{a\leqslant x\leqslant b} f(x) = \int_a^b f(x)\mathrm{d}x + \mathcal{O}(f(a)) + \mathcal{O}(f(b)).$$

In a), we obtain an error bound of $\mathcal{O}(T^{1/2})$, essentially because there are $\sqrt{T}$ summands and furthermore $\sum_{x>\sqrt{T}} x^{-2} = \mathcal{O}(T^{-1/2})$.

In b), we obtain an error bound of $\mathcal{O}(T)$, essentially because there are T summands and furthermore $\sum_{1\leqslant y\leqslant T} y^{-1/2} = T^{1/2} + \mathcal{O}(1)$.

Using the *hyperbola method*, one can do even better: To reduce the number of summands, and therefore the number of places where we need to round (and incur a penalty of $\mathcal{O}(1)$), note that $x^2 y \leqslant T$ implies that $x \leqslant \sqrt[3]{T}$ or $t \leqslant \sqrt[3]{T}$. We separately count the points with $x \leqslant \sqrt[3]{T}$ and the points with $y \leqslant \sqrt[3]{T}$, then subtract the points satisfying both $x \leqslant \sqrt[3]{T}$ and $y \leqslant \sqrt[3]{T}$ (which had been double-counted):

$$\begin{aligned} N(T) &= \sum_{1 \leqslant x \leqslant \sqrt[3]{T}} \left\lfloor \frac{T}{x^2} \right\rfloor + \sum_{1 \leqslant y \leqslant \sqrt[3]{T}} \left\lfloor \sqrt{\frac{T}{y}} \right\rfloor - \lfloor \sqrt[3]{T} \rfloor \cdot \lfloor \sqrt[3]{T} \rfloor \\ &= \sum_{1 \leqslant x \leqslant \sqrt[3]{T}} \frac{1}{x^2} \cdot T + 2 \cdot T^{2/3} - T^{2/3} + \mathcal{O}(T^{1/3}) \\ &= \zeta(2) \cdot T + \mathcal{O}(T^{1/3}). \end{aligned}$$

In the last step, we used that

$$\sum_{1 \leqslant x \leqslant K} x^{-2} = \zeta(2) - \sum_{x > K} x^{-2} = \zeta(2) - K^{-1} + \mathcal{O}(K^{-2})$$

for large K. □

Problem 10. Let a, b, c be a 2-cycle, an $(n-1)$-cycle, and an n-cycle in S_n (where $n \geqslant 2$). Show that they together generate the entire symmetric group S_n.

Solution. Let $H \subseteq S_n$ be a subgroup containing a 2-cycle, an $(n-1)$-cycle, and an n-cycle. Let i be the element of $\{1, \ldots, n\}$ fixed by the $(n-1)$-cycle. By conjugating the 2-cycle with an appropriate power of the n-cycle, it follows that H contains a 2-cycle of the form $(i\ j)$. By conjugating with powers of the $(n-1)$-cycle, we can show that H in fact contains all 2-cycles of this form. By conjugating with powers of the n-cycle, it follows that H contains every 2-cycle. Therefore, $H = S_n$. □