Golomb's distribution and related

Ümit Işlak

August 20, 2020

1 Uniform random variables

Definition 1.1 *A random variable X is said to be* ***uniform*** *over the set S, $|S| < \infty$, if its pmf is*

$$f(x) = \begin{cases} \frac{1}{|S|}, & \text{if } x \in S, \\ 0, & \text{otherwise.} \end{cases}$$

In this case we write $X \sim U(S)$.

Example 1.1 *Let $p_1 \neq p_2$ be prime numbers in $\{1, 2, \ldots, n\}$. Also let $N \sim U(\{1, 2, \ldots, n\})$.*

(a) Find $\mathbb{P}(p_1 \mid N)$.

(b) Find $\mathbb{P}(p_2 \mid N \big| p_1 \mid N)$.

(c) Let $n = 12$, $p_1 = 3$, $p_2 = 5$ in previous part to conclude that the events $p_1 \mid N$ and $p_2 \mid N$ are not independent.

(d) Show that

$$\lim_{n\to\infty} \frac{\mathbb{P}(p_2 \mid N \big| p_1 \mid N)}{\mathbb{P}(p_2 \mid N)} = 1.$$

Solution: (a) We have

$$\mathbb{P}(p_1 \mid N) = \frac{\left\lfloor \frac{n}{p_1} \right\rfloor}{n}.$$

(b) We have

$$\mathbb{P}(p_2 \mid N \big| p_1 \mid N) = \frac{\mathbb{P}(p_1 \mid N, p_2 \mid N)}{\mathbb{P}(p_1 \mid N)} = \frac{\frac{\left\lfloor \frac{n}{p_1 p_2} \right\rfloor}{n}}{\frac{\left\lfloor \frac{n}{p_1} \right\rfloor}{n}} = \frac{\left\lfloor \frac{n}{p_1 p_2} \right\rfloor}{\left\lfloor \frac{n}{p_1} \right\rfloor}.$$

(c) When $n = 12$, $p_1 = 3$, $p_2 = 5$, we have

$$\mathbb{P}(5 \mid N \big| 3 \mid N) = \frac{\left\lfloor \frac{12}{15} \right\rfloor}{\left\lfloor \frac{12}{3} \right\rfloor} = 0,$$

but

$$\mathbb{P}(5 \mid N) = \frac{1}{6}.$$

Since these two are not equal we conclude that these events are really independent.

(d) Left for you. $\square$

Note that for large n, the probabilities in previous example will be close to each other. Next problem is asking you to verify this.

Remark 1.1 *Note that in setting of previous example we have*

$$\lim_{n\to\infty} \mathbb{P}(p \mid N) = \lim_{n\to\infty} \frac{\left\lfloor \frac{n}{p} \right\rfloor}{n} = \frac{1}{p}.$$

Remark 1.2 *In next section we will that according to a certain probability measure $\mathbb{P}_s$ on $\mathbb{N}$, the events of Example 1.1 are independent. This will lead to several interesting results.*

Remark 1.3 *The set of all prime numbers is not finite as we know for a long while, and we can not have a uniform distribution in it.*

Problem 1.1 *Show that we can not have a uniform distribution over any infinite set. (Hint: Use contradiction)*

2 Golomb's distribution

Let N_1 and N_2 be independent uniform random variables over the set $\{1, 2, \ldots, n\}$. Also let p_1, p_2 be distinct prime numbers. Our interest in this section is on divisibility relations such as $\mathbb{P}(p_1 \mid N_1)$, $\mathbb{P}(p_1 \mid N_1, p_2 \nmid N_1)$ $\mathbb{P}(\gcd(N_1, N_2) = 1)$ and $\lim_{n\to\infty} \mathbb{P}(\gcd(N_1, N_2) = 1)$. The main difficulty arising in such questions is that the divisibility relations such as $p_1 \mid N_1$ and $p_2 \mid N_1$ are not statistically independent which we have already seen Example XX above.

Discuss asymptotic independence in standard asymptotic density case.

2.1 Golomb distribution

Recall that the Riemann zeta function is defined by $\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s}$. In this section we take the domain of ζ to be $(1, \infty) \subset \mathbb{R}$. We define Golomb's distribution following the work of Solomon Golomb. In the literature, you may also see this distribution called as the Dirichlet distribution, the zeta distribution and various others.[1]

Definition 2.1 *A random variable X is said to have the* ***Golomb distribution*** *with parameter $s \in (1, \infty)$ if its pmf is given by*

$$f(n) = \frac{1}{\zeta(s) n^s}, \qquad n \geq 1.$$

We write $X \sim G_s$.

That this defines a pmf is clear from the definition of ζ. Below we write $\mathbb{P}_s$ for the underlying probability measure, that is:

$$\mathbb{P}_s(X = n) = f(n), \qquad n \geq 1.$$

Also for given $p \geq 1$, D_p represents the event that $p \mid X$ in this section.

[1]As an admirer of Golomb's work on various fields, I preferred to call it this way following his paper.

Proposition 2.1 *Let X be a random variable with distribution G_s and p be a prime number. We have*

$$\mathbb{P}_s(D_p) = \frac{1}{p^s}$$

Proof: We have

$$\mathbb{P}_s(D_p) = \sum_{k=1}^{\infty} \mathbb{P}_s(X = kp) = \sum_{k=1}^{\infty} \frac{1}{\zeta(s)(kp)^s} = \frac{1}{\zeta(s)p^s} \sum_{k=1}^{\infty} \frac{1}{k^s} = \frac{1}{\zeta(s)p^s}\zeta(s) = \frac{1}{p^s}.$$

$\square$

Corollary 2.1 *Consider the setting in Proposition 2.1.*
(i) $\mathbb{P}_s(p \nmid X) = 1 - p^{-s}$.
(ii) $\mathbb{P}_s(2 \mid X) = \frac{1}{2^s}$.

Proposition 2.2 *Let D_p be defined as in Proposition 2.1. Then the events D_p are independent for prime p with respect to the Golomb distribution.*

Proof: We show that they are pairwise-independent, and leave the independence to the reader. Proof of Proposition 2.1 can be slightly modified to get $\mathbb{P}_s(D_m) = \frac{1}{m^s}$ for any $m \geq 2$. Let now p, q be distinct primes. Then we have

$$\mathbb{P}_s(D_p \cap D_q) = \mathbb{P}_s(D_{pq}) = \frac{1}{(pq)^s} = \frac{1}{p^s}\frac{1}{q^s} = \mathbb{P}_s(D_p)\mathbb{P}_s(D_q).$$

Hence, we conclude that the events $\{D_p : p \text{ prime}\}$ are pairwise-independent. $\square$

Problem 2.1 *Modify the proof to extend the pair-wise independence to independence.*

2.2 Asymptotic density

There are various ways to define the asymptotic density of a subset A of $\mathbb{N}$ in $\mathbb{N}$. We begin with the natural asymptotic density. For a subset A of $\mathbb{N}$, we define the upper and lower natural density of A to be

$$\overline{d}(A) = \limsup_{n\to\infty} \frac{\#(a \in A : a \leq n)}{n}$$

and

$$\underline{d}(A) = \liminf_{n\to\infty} \frac{\#(a \in A : a \leq n)}{n},$$

respectively.

Problem 2.2 *Find a subset A of $\mathbb{N}$ where $\overline{d}(A)$ and $\underline{d}(A)$ are distinct.*

Definition 2.2 *For a subset A of $\mathbb{N}$, if $\overline{d}(A)$ and $\underline{d}(A)$ are equal, then the natural asymptotic density of A is defined to be*

$$d(A) := \overline{d}(A) = \underline{d}(A).$$

For example, the natural asymptotic density of even numbers is 1/2 (How do you verify this rigorously? How can you generalize this argument?).

Next we relate the natural asymptotic density of a set A to the uniform discrete distribution. Let $A \subset \mathbb{Z}^+$, $n \in \mathbb{N}$. Also, let X_n be a uniformly distributed random variable over the set $\{1, 2, \dots, n\}$. Then

$$d(A) = \lim_{n \to \infty} \frac{\#(a \in A : a \leq n)}{n}$$

can be considered as

$$\lim_{n \to \infty} \mathbb{P}(X_n \in A).$$

Note that our definition above for asymptotic density can be rewritten as

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} \mathbf{1}(i \in A)}{n},$$

where $\mathbf{1}(i \in A)$ equals 1 when $i \in A$, and equals 0 otherwise.

We define two more asymptotic densities that will be useful in the sequel. First one is the logarithmic asymptotic density.

Definition 2.3 *A subset A of $\mathbb{N}$ is said to have* ***logarithmic asymptotic density*** *(LAD) if we have*

$$\lim_{n \to \infty} \sum_{i=1}^{n} \frac{\mathbf{1}(i \in A)}{i \ln n} = 1.$$

Logarithmic asymptotic density will be of interest when we analyze the Benford law and related matters. The second type of asymptotic density of interest will be the zeta density.

Definition 2.4 *For a subset A of $\mathbb{N}$, the zeta asymptotic density of A is defined to be*

$$\delta(A) := \lim_{s \to 1^+} \sum_{a \in A} \frac{a^{-s}}{\zeta(s)},$$

whenever the limit exists.

Problem 2.3 *These three definitions of asymptotic density give you the chance to define a much more generalized class of asymptotic densities. How?*

We continue here with the discussion of Golomb distribution and the zeta density. Assuming $Z \sim G_s$, observe that we may write

$$\delta(A) = \lim_{s \to 1^+} \mathbb{P}_s(Z \in A).$$

Recall that if $Z \sim G_s$, $n \in \mathbb{N}$, then

$$\mathbb{P}_s(n \mid Z) = \frac{1}{n^s}$$

and so when $gcd(m,n)=1$

$$\mathbb{P}_s(nm \mid Z) = \frac{1}{(nm)^s} = \frac{1}{n^s}\frac{1}{m^s} = \mathbb{P}_s(n \mid Z)\mathbb{P}_s(m \mid Z).$$

In other words, with respect to $\mathbb{P}_s$, the events $\{n \mid Z\}$ and $\{m \mid Z\}$ are independent in this setting.

A final note on divisibility before concluding this section: If $X_n \sim U\{1, 2 \ldots, n\}$ and p is a prime number, then

$$\lim_{n\to\infty} \mathbb{P}(p \mid X_n) = \frac{1}{p} = \lim_{s\to 1^+} \mathbb{P}_s(p \mid Z).$$

2.3 Euler's prime number identity revisited

Next we give a probabilistic proof of the well known result in multiplication function theory (which we have seen earlier in a different setting) known as Euler's prime number identity. This accepts various other proofs including analytic ones.

Proposition 2.3 *(Euler's prime number identity) We have*

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}.$$

Proof: Letting X have G_s distribution, recall that we have $\mathbb{P}_s(X = k) = \frac{k^{-s}}{\zeta(s)}$. This in particular says $\mathbb{P}_s(X = 1) = \frac{1}{\zeta(s)}$. We now have

$$\begin{aligned}
\frac{1}{\zeta(s)} = \mathbb{P}_s(X = 1) &= \mathbb{P}_s(X \text{ is not divisible by any prime}) \\
&= \mathbb{P}_s\left(\bigcap_{p \text{ prime}} D_p^c\right) \\
&= \prod_{p \text{ prime}} \mathbb{P}_s(D_p^c) \\
&= \prod_{p \text{ prime}} (1 - p^{-s}).
\end{aligned}$$

We conclude that $\zeta(s) = \prod_p (1 - p^{-s})^{-1}$ as required. $\square$

2.4 Square free

Proposition 2.4 *Let X have G_s distribution. The probability that no square greater than one divides X is $1/\zeta(2s)$.*

Proof: A similar argument to proof of Proposition 2.2 shows that the events $\{D_{p^2}\}$ are independent since $gcd(p^2, q^2) = 1$ for distinct primes p, q. Then

$$\begin{aligned}
\mathbb{P}_s(\text{no square divides } X) &= \mathbb{P}_s(\text{no prime square divides } X) \\
&= \mathbb{P}_s\left(\bigcap_{p \text{ prime}} D^c_{p^2}\right) \\
&= \prod_{p \text{ prime}} \mathbb{P}_s(D^c_{p^2}) \\
&= \prod_{p \text{ prime}} (1 - p^{-2s}) \\
&= \frac{1}{\zeta(2s)}.
\end{aligned}$$

□

Show that the previous statement is equivalent to

Problem 2.4 *(i) Let X and Y be independent and each have G_s distribution. Let H be the greatest common divisor of X and Y. Then*

$$\mathbb{P}_s(H = n) = \frac{1}{n^{2s}\zeta(2s)}.$$

(ii) Conclude from the first part that the probability of X and Y being coprime is $\frac{1}{\zeta(2s)}$.

2.5 Greatest common divisor

Let X_1^n, X_2^n be independent uniformly distributed random variables over the set $\{1, 2, \ldots, n\}$. We are interested in finding

$$\lim_{n\to\infty} \mathbb{P}(gcd(X_1^n, X_2^n) = 1).$$

To analyze this, let Z_1, Z_2 be independent with G_s distribution. Then we have

$$\begin{aligned}
\mathbb{P}_s(gcd(Z_1, Z_2) = 1) &= \mathbb{P}_s\left(\bigcap_{p \text{ prime}} \{p \nmid Z_1 \text{ or } p \nmid Z_2\}\right) \\
&= \prod_{p \text{ prime}} \mathbb{P}_s(p \nmid Z_1 \text{ or } p \nmid Z_2) \\
&= \prod_{p \text{ prime}} (1 - \mathbb{P}_s(p \mid Z_1, p \mid Z_2)) \\
&= \prod_{p \text{ prime}} (1 - p^{-2s}) \\
&= \frac{1}{\zeta(2s)},
\end{aligned}$$

where the last step follows from Euler's identity. Then we obtain

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(gcd(X_1^n, X_2^n) = 1) &= \lim_{n\to\infty} \prod_{p\le n} (1 - \mathbb{P}(p \mid X_1^n, p \mid X_2^n)) \\
&= \lim_{n\to\infty} \prod_{p\le n} (1 - \mathbb{P}(p \mid X_1^n)\mathbb{P}(p \mid X_2^n)) \\
&= \prod_{p \text{ prime}} (1 - p^{-2}) \\
&= \frac{1}{\zeta(2)} \\
&= \lim_{s\to 1^+} \mathbb{P}_s(gcd(Z_1, Z_2) = 1).
\end{aligned}$$

Note that this in particular says that

$$\lim_{n\to\infty} \mathbb{P}(gcd(X_1^n, X_2^n) = 1) = \frac{6}{\pi^2}.$$

Problem 2.5 *How can you generalize this discussion to k many independent uniform random numbers?*

3 Distributions with prime divisibility condition; Khintchine distributions

Definition 3.1 *Let X be a random variable having distribution $\mathbb{P}$ on natural numbers. $\mathbb{P}$ is said to have the* ***factorization property*** *if*

$$\mathbb{P}(p \mid X, q \mid X) = \mathbb{P}(p \mid X)\mathbb{P}(q \mid X)$$

for any distinct prime numbers p, q. A distribution $\mathbb{P}$ with the factorization property is called a ***Khintchine distribution****.*

We have already seen that the Golomb distribution has the factorization property. Are there any others? Next theorem gives two characterization of all such distributions.

Theorem 3.1 *Let X be a random variable from some distribution $\mathbb{P}$ on natural numbers whose (random) prime factorization is given by*

$$X = \prod_{i=1}^{\infty} p_i^{N_i}$$

where p_i is the i-th prime number.

(a) $\mathbb{P}$ has the factorization property if and only if the prime powers $\{N_i : i \ge 1\}$ are independent.

(b) $\mathbb{P}$ has the factorization property if and only if

$$\mathbb{P}(X = n) = \frac{f(n)}{n^s F(s)}$$

where $F(s)$ is the Dirichlet series of some non-negative arithmetic function, i.e. $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$.

The proof of the theorem will be given in the Appendix.