DISCRETE LIMIT LAWS

AMIT SAHA

Abstract. Much of the analysis in generating functions is in regards to exact counting or approximations of exact parameters. However, significant interest lies in analyzing combinatorial parameters and their distributions relative to combinatorial objects. In essence, we seek data on probability distributions. Many parameters of interest follow certain laws that determine their distribution: such laws are known as *limit laws*. From bivariate and multivariate generating functions and their parameters, we can derive a wondrous collection of distributions. This paper is intended to briefly introduce the reader to some relevant methods and theorems.

1. Motivation

Example 1.1. We discuss an example of a discrete limit law before we formalize a definition – the class of binary words, $\mathcal{I}$, binary words composed of the letters $\{a, b\}$. First, we examine a parameter χ, counting the number of initial a. Suppose that

$$I^{\chi}_{n,k} := \text{number of binary words with } k \text{ initial } a \in \mathcal{I}_n.$$

It's pretty easy to examine this using elementary methods (i.e. without using generating functions). We find that

$$I^{\chi}_{n,0} = 2^{n-1}, \dots, I^{\chi}_{n,n} = 2^0 = 1.$$

The probability distribution is easy to find. If $0 \le k < n$, then our distribution is accordingly

$$\mathbb{P}_{\mathcal{I}_n}(\chi = k) = \frac{1}{2^{k+1}}$$

and if $k = n$ then we find that

$$\mathbb{P}_{\mathcal{I}_n}(\chi = k) = \frac{1}{2^n}.$$

Date: December 6, 2021.

From here, we can take the asymptotic limit of such parameters and obtain a cumulative distribution function for this class and parameter.

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{I}_n}(\chi \leq k) = 1 - \frac{1}{2^{k+1}} \quad \lim_{n\to\infty} \mathbb{P}_{\mathcal{I}_n}(\chi = k) = \frac{1}{2^{k+1}}$$

In this sense, there exists a *discrete limit law* for such a parameter. Note the similarity of such a distribution to a geometric series. Thus, this is known as a discrete limit law of the geometric type, as $n \to \infty$. Other parameters and combinatorial classes are best approached with *continuous limit laws*, but such analyses require greater analytic machinery and background. We now examine some particular machinery.

2. Preliminaries

All analysis of limit laws begins with a bivariate generating function,

$$F(z,u) = \sum_{n,k=0} f(n,k)u^k z^n,$$

where the variable n marks the size of the structure and u marks some combinatorial parameter of interest, relative to this structure in question. Recall that the goal of limit laws is to elicit information about the *probability* about some arbitrary parameter being true. Thus, it is natural to explore probability generating functions.

Ideally, we would be able to extract coefficients from our original generating function – this is the simplest case, and allows us to easily obtain a discrete limit law. However, such a task is difficult for more complex schema and generating functions, and is not particularly general.

In light of this difficulty, we use some common probabilistic machinery and use an alternative methodology.

Definition 2.1. The probability generating function of a class A over a parameter χ marked by u with bivariate generating function $A(z,u)$ is defined to be

$$\sum_{n=0}^{\infty} \mathbb{P}_{\mathcal{A}_n}(\chi = n)u^k = \frac{[z^n]A(z,u)}{[z^n]A(z,1)}$$

From here, asymptotic estimation of the coefficients of the generating function yields the desired limit law, when taken in a neighborhood around $u = 1$. Such asymptotic estimation may be developed in many ways.

Example 2.2. We again examine binary words, with the same parameter χ. We first write down the specification for such a sequence. Suppose that the variable u marks the parameter χ. Then, it follows that

$$\mathcal{I}^\chi = \textsc{Seq}(ua)\,\textsc{Seq}(\textsc{Seq}(a)b) \implies I^\chi(z,u) = \frac{1}{1-uz}\frac{1}{1-\frac{z}{1-z}} = \frac{1}{1-uz}\frac{1-z}{1-2z}.$$

We fix u and extract coefficients.

$$[z^n]W^\chi(z,u) \sim \frac{1/2}{1-u/2}2^n$$

Dividing by the coefficient of the vertical generating function (i.e. the univariate generating function), we determine that

$$\mathbb{P}_{\mathcal{I}_n}(\chi = k) = \frac{1}{2^n}[z^n]W^\chi(z,u) = \sum_{i=0}^{\infty} \frac{1}{2^{k+1}}u^k.$$

This is exactly the conclusion we derived earlier, except we have encoded it into a generating function.

Distinctions exist between various forms of limit laws that may exist for such parameters. In example [1.1], we derived two examples of discrete limit laws.

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{I}_n}(\chi \le k) = 1 - \frac{1}{2^{k+1}} \qquad \lim_{n\to\infty} \mathbb{P}_{\mathcal{I}_n}(\chi = k) = \frac{1}{2^{k+1}}$$

When we are able to obtain estimates on parameters valued *exactly* as k, we say that there exists a *local limit law*. Differences in this formulation are insignificant in the case of the convergence of discrete variables, which can be formalized by defining convergence of a limit law.

We now discuss several definitions and theorems, which we state without proof as bases off of which we develop our analytic machinery.

Definition 2.3. A discrete random variable X_n is said to converge to Y, if for all $k \ge 0$, the property that

$$\lim_{n\to\infty} \mathbb{P}(X_n \le k) = \mathbb{P}(Y \le k)$$

holds. This occurs at speed ϵ_n if the following holds.

$$\sup_k |\mathbb{P}(X_n \le k) - \mathbb{P}(Y \le k)| \le \epsilon_n \qquad \lim_{n\to\infty} \epsilon_n = 0$$

Remark 2.4. Replacing the definitions here with equivalence rather than inequalities would lead to an exactly equivalent formulation of the definition, since the variables here are discrete. Thus, local limit laws are in some sense equivalent to traditional discrete limit laws. However, in our analytic formulation using bivariate generating functions, it may be easier to deal with closed forms of functions. Thus, we use the following theorem.

Theorem 2.5. Continuity Theorem. *If ω is a set in the unit disk with at least one accumulation point, and if the probability generating functions $p_n(u) = \sum_{k\geq 0} p_{n,k}u^k$ and $q(u) = \sum_{k\geq 0} q_k u^k$ have weak convergence,*

$$\lim_{n\to+\infty} p_n(u) = q(u), \tag{2.1}$$

for all $u \in \omega$, a discrete limit law holds such that

$$\lim_{n\to+\infty} p_{n,k} = q_k \tag{2.2}$$

This theorem allows us to directly utilize the analytic properties of our bivariate generating functions. We omit the proof for brevity. A simple corollary follows.

Corollary 2.6. *If $p_n(u)$ and $q(u)$ satisfy [2.3], it similarly follows from summation of equation (2.2) that*

$$\lim_{n\to+\infty} \sum_{i\leq k} p_{n,i} = \sum_{i\leq k} q_i.$$

This theorem and its corollary give us access to a wide range of parameters that may be analyzed with such methods.

3. Applications to Permutation Statistics

Recall that the univariate specification for a permutation is

$$\mathcal{P} \cong \textsc{Set}(\textsc{Cyc}(\mathcal{Z})).$$

Thus, a permutation is essentially a set of multiple cycles. The formulation of such a law with the set construction makes it more likely that the Poisson distribution will appear in parameters of permutations. Specifically, the Poisson distribution has probabilities

$$e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{and probability generating function} \quad e^{\lambda(1-u)}.$$

Example 3.1. We first examine singletons in permutations. The bivariate generating function marking this parameter, which we'll again call χ marked by u, is

$$\mathcal{P}^{\chi} \cong \textsc{Set}(u\mathcal{Z} + \textsc{Cyc}_{\geq 2}(\mathcal{Z})) \implies P^{\chi}(z,u) = \frac{\exp(z(u-1))}{1-z}.$$

The simplest approach here is to directly extract both coefficients. We find that

$$[z^n u^k]P^\chi(z,u) = \frac{d_{n-k}}{k!},$$

and utilizing the asymptotic $d_n = 1/e$, we find that

$$\lim_{n\to+\infty} p_{n,k} = \frac{1}{e}\cdot\frac{1}{k!}.$$

As discussed earlier, this is simply a discrete law of the Poisson type.

We naturally extend this framework for the number of cycles of length m in an arbitrary permutation.

Example 3.2. The bivariate generating function marking cycles of length m with u is

$$\mathcal{P}^\chi \cong \textsc{Set}(\textsc{Cyc}_{\neq m}(\mathcal{Z}) + u\,\textsc{Cyc}_{=m}(\mathcal{Z})) \implies P^\chi(z,u) = \frac{\exp((u-1)z^m/m)}{1-z}.$$

Utilizing the framework we developed earlier, we can extract asymptotics for the coefficients. We find that

$$\lim_{n\to\infty}[z^n]P(z,u) = e^{(u-1)m},$$

which is clearly a probability generating function for a Poisson distribution. Thus, the distribution of cycles of size m in permutations follows a Poisson law.

4. Applications to Trees

We briefly discuss some complex-analytic machinery. It is well known that analysis of singularities can give significant insights on the growth rate of coefficients. Suppose that we have two classes, $\mathcal{A}$ and $\mathcal{B}$. We call their *composition schema*, $\mathcal{C}$,

$$\mathcal{C} \cong \mathcal{A}\circ(u\mathcal{B}) \implies C(z,u) = A(uB(z)).$$

This schema essentially enumerates the number of $\mathcal{B}$ components with some sort of greater structure, provided by $\mathcal{A}$. If the generating functions A and B have radii of convergence ρ_a and ρ_b respectively, then we define the quantities

$$\tau_a = \lim_{z\to\rho_a} A(z) \quad\text{and}\quad \tau_b = \lim_{z\to\rho_b} B(z). \tag{4.1}$$

We treat the case in which $\tau_b < \rho_a$. In this case, we consider the composition schema to be **subcritical**. The subcritical schema forces the composition of generating functions to have a singularity of the same type of the internal function, in this case $B(z)$, since the value the generating function approaches near its singularity is defined for $A(z)$. In some sense, it is simpler to treat and analyze this singularity. The subcritical case yields a very helpful method for treating asymptotics and obtaining limit laws.

Theorem 4.1. (Subcritical Composition Limit Laws). *Suppose that* $C(z,u) = A(uB(z))$. *If this composition is subcritical and* $B(z)$ *has a singularity at* $\rho = \rho_b$*, defined in the same way as earlier, on its disc of convergence. Then, a discrete limit law holds for the enumeration of* $\mathcal{B}$ *components. Define* $c_{n,k} = [z^n u^k]C(z,u)$, $c_n = [z^n]C(z,1)$, *and* $a_n = [z^n]A(z)$. *The discrete law is of the form*

$$\lim_{n\to\infty} \frac{c_{n,k}}{c_n} = \frac{k a_k \tau^{k-1}}{A'(\tau)}. \tag{4.2}$$

This implies that the probability generating function is equal to

$$\frac{uA'(u\tau)}{A'(\tau)}. \tag{4.3}$$

We omit the proof for brevity. This theorem allows us to treat subcritical schema and their respective limit laws in a much more methodical and direct manner. We are now fully equipped to examine some limit laws in the combinatorics of trees. Suppose we seek to count the number of components of an ordered forest of Catalan trees. Suppose that the forest is represented by class $\mathcal{F}$, and the Catalan trees by class $\mathcal{C}$. It follows directly that

$$\mathcal{F} \cong \textsc{Seq}(u\mathcal{C}) \implies F(z,u) = \frac{1}{1-uC(z)} = \frac{1}{1-\frac{u}{2}\left(1-\sqrt{(1-4z)}\right)}.$$

In this case, $\tau = 1/2$, which implies that $\tau < \rho_F$. Applying [4.2], we find that

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \frac{k}{2^{k+1}}.$$

This is a discrete limit law for this parameter and schema.

If we instead treat Cayley trees, the analysis is similar. We have the bivariate generating function

$$F(z,u) = e^{uT(z)}, \quad T(z) = ze^{T(z)}.$$

Again, the conditions of theorem [4.1] are satisfied. It follows that the limit law is

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \frac{1}{e(k-1)!}.$$

This is simply a Poisson distribution shifted by 1.

5. Conclusion

Discrete limit laws, as described in this paper, offer a clear and methodical manner in which we can derive information about the distribution of parameters in combinatorial classes. There are many other applications and specializations of limit laws, such as continuous limit laws and extensions to the critical and supercritical cases of composition schema, that were not covered in this paper.