LECTURE 15 – PROBABILITY, MOMENTS, AND CONCENTRATION

JAY PANTONE

Let $\mathcal{A}$ be a combinatorial class with counting sequence A_n. Let χ be a function $\chi : \mathcal{A} \to \mathbb{Z}_{\geq 0}$ that measures some parameter of $\mathcal{A}$ (e.g., number of cycles in a permutation, number of parts in a set partition). Let $a_{n,k}$ be the number of objects α in $\mathcal{A}$ of size n such that $\chi(\alpha) = k$.

The probability that a randomly selected object α of size n has $\chi(\alpha) = k$ is

$$\mathbb{P}_{\mathcal{A}_n}\{\chi = k\} = \frac{A_{n,k}}{A_n}.$$

Functions like χ are examples of *discrete random variables*. We omit a formal definition, but one can just imagine that any discrete random variable X is, for our purposes, a function like χ.

Definition: Given a discrete random variable X, define the *probability generating function* $p(u)$ of X to be

$$p(u) = \sum_{k \geq 0} \mathbb{P}\{X = k\} u^k.$$

Theorem 15.1. *Let χ be a parameter function, and let $A(z, u)$ be a bivariate generating function such that tracks χ over a class $\mathcal{A}$ (i.e., the coefficient of $z^n u^k$ is the number of objects of size n in $\mathcal{A}$ with χ-value k). Then, the probability generating function of χ over $\mathcal{A}_n$ is*

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

The probability generating function encodes a tremendous about information regarding the distribution of χ. First we introduce some notation and probabilistic terms.

Definitions: Let X be a discrete random variable. The *expectation of $f(X)$* is

$$\mathbb{E}[f(X)] = \sum_{k \geq 0} \mathbb{P}\{X = k\} \cdot f(k).$$

In particular, the *rth power moment* is

$$\mathbb{E}[X^r] = \sum_{k \geq 0} \mathbb{P}\{X = k\} \cdot k^r.$$

The *expected value* (or, *mean*) of a discrete random variable X is

$$\mathbb{E}[X] = \sum_{k \geq 0} \mathbb{P}\{X = k\} \cdot k.$$

The variance is

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

and the standard deviation is

$$\sigma(X) = \sqrt{\mathbb{V}[X]}.$$

It turns out that, given a probability generating function, it is more convenient to work with factorial moments than with power moments. The *rth factorial moment* of X is

$$\mathbb{E}[X(X-1)\cdots(X-(r-1))].$$

Theorem 15.2. *The rth factorial moment can be calculated by r-fold differentiation of the probability generating function:*

$$\mathbb{E}[X(X-1)\cdots(X-(r-1))] = \frac{[z^n]\,[(D_u^r)A(z,u)]_{u=1}}{[z^n]A(z,1)} = [(D_u^r)p_{\mathcal{A}_n}(u)]_{u=1}\,.$$

This permits direct calculation of the expected value and variance of X from $p(u)$.

Notation: Henceforth, we use notation f_z to mean the derivative of f with respect to z and f_u to mean the derivative of f with respect to u.

Example: Let $A(z,u)$ be the BGF for binary words of length n with k 0's. (As usual, z tracks length and u tracks number of 0's.) We determined in the previous section that

$$A(z,u) = \frac{1}{1-z(1+u)}.$$

Let X be the discrete random variable tracking the number of 0's. From the BGF we determine the first factorial moment to be

$$\begin{aligned}
\mathbb{E}[X] &= \frac{[z^n]W_u(z,1)}{[z^n]W(z,1)} \\
&= \frac{[z^n]\left[\frac{\partial}{\partial u}\left(\frac{1}{1-z(1+u)}\right)\right]_{u=1}}{2^n} \\
&= 2^{-n}[z^n]\left[\frac{z}{(1-z(1+u))^2}\right]_{u=1} \\
&= 2^{-n}[z^n]\frac{z}{(1-2z)^2} \\
&= 2^{-n}(n2^{n-1} \\
&= \frac{n}{2}.
\end{aligned}$$

Therefore, the expected number of 0's in a binary word of length n is $n/2$ (as we... well... expected). Let's calculate the standard deviation now. To do so, we first need to calculate the variance, and to do that we will first calculate the 2nd factorial moment.

$$
\begin{aligned}
\mathbb{E}[X(X-1)] &= \frac{[z^n]W_{uu}(z,1)}{[z^n]W(z,1)} \\
&= 2^{-n}[z^n]\left[\frac{2z^2}{(1-z(1+u))^3}\right]_{u=1} \\
&= 2^{-n}[z^n]\frac{2z^2}{(1-2z)^3} \\
&= 2^{-n}(n(n-1)2^{n-2}) \\
&= \frac{n(n-1)}{4}.
\end{aligned}
$$

It follows now that

$$
\begin{aligned}
\mathbb{V}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= (\mathbb{E}[X^2] - \mathbb{E}[X]) + \mathbb{E}[x] - \mathbb{E}[X^2] \\
&= \mathbb{E}[X^2 - X] + \mathbb{E}[X] - \mathbb{E}[X^2] \\
&= \frac{n(n-1)}{4} + \frac{n}{2} - \frac{n^2}{4} \\
&= \frac{n}{4}.
\end{aligned}
$$

Therefore, the standard deviation is

$$\sigma(X) = \sqrt{\mathbb{V}[X]} = \frac{\sqrt{n}}{2}.$$

Further moments are easily calculated as well, but the expected value and standard deviation tell us a lot of information about this so-called *binomial distribution*[1]. In particular, the property that the standard deviation is asymptotically negligible when compared to the expected value as n tends to infinity implies that the distribution is *concentrated around the mean*.

To formally state what concentration means, we first need two famous probability results: Markov's inequality and Chebyshev's inequality.

Theorem 15.3. *Let X be a non-negative discrete random variable and Y an arbitrary discrete random variable. Then,*

(1) *(Markov's inequality)* $\mathbb{P}\{X \geq t\mathbb{E}[X]\} \leq \frac{1}{t}$,

(2) *(Chebyshev's inequality)* $\mathbb{P}\{|Y - \mathbb{E}[Y]|\} \geq t\sigma(Y)\} \leq \frac{1}{t^2}$.

[1]This is commonly introduced in a beginner statistics course as the distribution of the number of heads when flipping a fair coin n times.

Proof. To prove Markov's inequality, we simply see that

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k\geq 0} \mathbb{P}\{X=k\}\cdot k \\ &\geq \sum_{k\geq t\mathbb{E}[X]} \mathbb{P}\{X=k\}\cdot k \\ &\geq \sum_{k\geq t\mathbb{E}[X]} \mathbb{P}\{X=k\}\cdot t\mathbb{E}[X] \\ &= t\mathbb{E}[X]\sum_{k\geq t\mathbb{E}[X]} \mathbb{P}\{X=k\} \\ &= t\mathbb{E}[X]\mathbb{P}\{X\geq t\mathbb{E}[X]\},\end{aligned}$$

and we divide both sides by $\mathbb{E}[X]$ to get the result. To derive Chebyshev's inequality, apply Markov's inequality to the random variable $T = (Y - \mathbb{E}[Y])^2$. ☐

We can now give a more concrete statement about concentration.

Theorem 15.4. *Consider a family $\{X_n\}$ of discrete random variables. Let $\mu_n = \mathbb{E}[X]$ and $\sigma_n = \sigma(X_n)$. Suppose the condition*

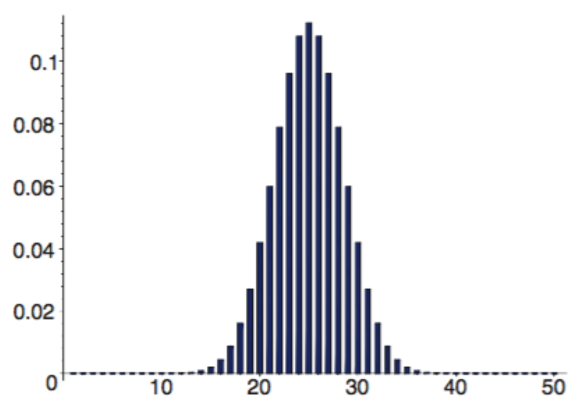
$$\lim_{n\to\infty} \frac{\sigma_n}{\mu_n} = 0.$$

Then we say that the distribution of the family $\{X_n\}$ is concentrated in the sense that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{1-\epsilon \leq \frac{X_n}{\mu_n} \leq 1+\epsilon\right\} = 1.$$

Proof. Hint: Use Chebyshev's inequality. ☐

Let us illustrate what concentration of distribution around the mean tells us in a practical sense. Below is a plot of the distribution of the number of 0's in binary words of length 50. Note how a large proportion of words are "close" to the mean of 25.



Furthermore, here are the number of 0's found in 10 randomly selected binary words of length 100,000:

$$49798, 79873, 49968, 49980, 49999, 50017, 50029, 50080, 50101, 50284.$$

As n increase, we see tighter and tighter clusters around the mean. This is concentration.

Example: For the next example, let us consider permutations counted according the number of cycles. We previously showed that the BGF is

$$P(z,u) = (1-z)^{-u}.$$

The expected number of cycles in a permutation of length n is

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}_n}[\chi] &= \frac{[z^n]P_u(z,1)}{[z^n]P(z,1)} \\
&= [z^n]\left[-(1-z)^{-u}\log(1-z)\right]_{u=1} \\
&= [z^n]\frac{1}{1-z}\log\left(\frac{1}{1-z}\right) \\
&= 1 + \frac{1}{2} + \cdots + \frac{1}{n}.
\end{aligned}$$

Asymptotically, $\mathbb{E}_{\mathcal{P}_n}[\chi] \sum \log(n)$. One can also show that the second factorial moment is

$$\mathbb{E}_{\mathcal{P}_n}[X(X-1)] = \frac{1}{1-z}\left(\log\left(\frac{1}{1-z}\right)\right)^2,$$

which yields

$$\mathbb{V}_{\mathcal{P}_n}[X] = \log(n) + \gamma - \frac{\pi^2}{6} + O\left(\frac{1}{n}\right),$$

and so

$$\sigma_n \sim \sqrt{\log(n)}.$$

Therefore, this distribution is also concentrated around the mean.

Example: For an example of a distribution that is not concentrated about the mean, let's exaqmine the class of permutations counted according to the number of cycles of size r. We showed earlier that the expected number of cycles of length r in a permutation of length $n \geq r$ is $1/r$. The numerical calculations are a bit more complicated, but it can be shown that the full distribution is a Poisson distribution with mean $1/r$ and therefore standard deviation $1/\sqrt{r}$. Since there are both constant, their quotient does not go to zero as $n \to \infty$, and therefore the distribution is *not* concentrated around the mean.

Example: For our last example, we revisit the class of rooted unlabeled planar trees counted by total path length. Previously, we showed that the BGF $G(z,u)$ satisfies the functional equation

$$G(z,u) = \frac{z}{1-G(uz,z)}.$$

To calculate the mean and variance, we must take the derivative with respect to u of *each side* of the functional equation, then substitute $u=1$ into the resulting functional equation. Using the multivariate chain rule on the right-hand side, we see

$$G_u(z,u) = \frac{z}{(1-G(uz,u))^2}(zG_z(uz,u) + G_u(uz,u)).$$

Substituting $u=1$:

$$G_u(z,1) = \frac{z}{(1-G(z,1))^2}(zG_z(z,1) - G_u(z,1))$$

and therefore

$$G_u(z,1) = \frac{z^2 G_z(z,1)}{(1-G(z,1))^2 - z}.$$

Noting that $G(z,1) = G(z)$ (the original univariate GF for these trees, i.e., the Catalan GF), we have

$$G_u(z,1) = \frac{z^2 G'(z)}{(1-G(z))^2 - z} = \frac{z}{2(1-4z)} - \frac{z}{2\sqrt{1-4z}}.$$

Coefficient extraction yields

$$[z^n]G_u(z,1) = 2^{2n-3} - \frac{1}{2}\binom{2n-2}{n-1}.$$

Some quick asymptotic analysis (which we have not yet learned) shows that

$$\frac{[z^n]G_u(z,1)}{[z^n]G(z,1)} \sim \frac{\sqrt{\pi}}{2} n^{3/2}.$$

An interesting corollary is that in a randomly selected tree on n nodes, the expected distance from the root to a randomly selected node is on the order of $\sqrt{n}$. In a completely balanced binary tree, this quantity if on the order of $\log(n)$, so this shows that the "average" planar tree is quite unbalanced.