Elements of Probability and Statistics

Lecture 08: Discrete Probability Distributions

IAI, TCG-CREST September 08 & 12, 2023

8.1 (contd.) Discrete Distributions

Hypergeometric Distribution: Let there be a jar with w number of white balls and b number of black balls. n number of balls are selected from the jar, by *sampling without replacement*. The probability distribution of obtaining exactly k number of white balls is described by the Hypergeometric Distribution, written as $X \sim \text{HypGeom}(w, b, n)$, and the probabilities are given by,

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} \ .$$

The expected value of the distribution is,

$$E(X) = E(I_1 + ... + I_n) = n\frac{w}{w+b} \ .$$

Discrete Uniform Distribution: Let there be n possible outcomes, numbered from 1 to n, where all outcomes are equally likely. The distribution of any particular outcome written as $X \sim \text{DUinf}(1, n)$ is specified by the PMF:

$$P(X = k) = \frac{1}{n} \ .$$

The expected value and variance of $X \sim \text{DUinf}(n)$:

$$E(X) = \frac{1}{n}(1 + ... + n) = \frac{n+1}{2}.$$
$$Var(X) = \frac{1}{n}(1^2 + ... + n^2) - (\frac{n+1}{2})^2 = \frac{(n+1)(n-1)}{12} \ .$$

The discrete uniform distribution can also be specified more generally in terms of n items $X \sim \text{DUinf}(n)$, or more specifically in terms of n uniformly spaced outcomes between two limits a and b, with $a < b$, as $X \sim \text{DUinf}(a, b)$.

Cumulative Distribution Functions (CDF): An alternate way of specifying probability distributions is the CDF. For discrete distributions, we specify PMFs at different supports as $P(X = x)$. In comparison, for CDFs we specify cumulative probability values, e.g. as $P(X < x)$ or $P(X \leq x)$.

Defn. Properties of CDFs:

- If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
- Right continuous: $F(a) = \lim_{x \to a^+} F(x)$.

- Convergence to 0 and 1 at the limits:

$$\lim_{x\to-\infty} F(x) = 0 \text{ and } \lim_{x\to\infty} F(x) = 1.$$

Converting PMFs to CDFs: $P(X \leq a) = \sum\limits_{\substack{s\in S:\\ X(s)=k, k\leq a}} P(X = k)$.

Converting CDFs to PMFs: $P(X = a) = P(X \leq a) - P(X < a)$.

Defn. Functions of random variables: If X is a random variable defined on sample space S, $g(X) : \mathbb{R} \to \mathbb{R}$ is a function that maps $s \in S$ to $g(X(s))$, $\forall s \in S$.

If X is discrete, then the support of $g(X)$ is the set y for which $g(x) = y$ for at least one x in the support of X. Thus the PMF of $g(X)$ is,

$$P(g(X) = y) = \sum_{x:g(x)=y} P(X = x).$$

Functions can be defined on *multiple random variables* as well, e.g., $g(X,Y)$ maps $s \in S$ to $g(X(s), Y(s))$.

Defn. Independence of random variables: X and Y are independent if $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)\ \forall x, y \in \mathbb{R}$.

For discrete random variables, two variables can be said to be independent if $P(X = x, Y = y) = P(X = x)P(Y = y)\ \forall x, y \in \mathbb{R}$.

Countably infinite random variables are said to be independent if every finite subset is independent.

Thm. If X and Y are independent, then any function of X is independent of any function of Y.

i.i.d. random variables: Random variables that are independent and each follow the same distribution are called *independent and identically distributed* (i.i.d.) random variables.

Conditional PMF: For r.v.s X and Z, the conditional PMF of X given a fixed $Z = z$ is $P(X = x|Z = z)$.

Conditional Independence of random variables: X and Y are said to be conditionally independent of Z if $\forall x, y \in \mathbb{R}$ and all z in the support of Z,

$$P(X \leq x, Y \leq y|Z = z) = P(X \leq x|Z = z)P(Y \leq y|Z = z).$$

For discrete random variables, conditional independence can be defined as,

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z).$$

Q. Independence does not imply conditional independence: Two players A and B each flip a fair coin. If the coins match, A wins, otherwise B wins. Let $X = 1$ if A's coin lands as heads, otherwise $X = -1$ if it lands as tails. Similarly, $Y = 1$ if B's coin lands heads, and $Y = -1$ if it lands as tails. Let $Z = XY$, which is 1 if A wins, and -1 if B wins. Show that X and Y are independent, but not conditionally independent if $Z = 1$.

Q. Conditional independence does not imply independence: You are playing an online game against two opponents, A and B. Your chance of winning the game against A is $1/2$, and against B is $3/4$. One of the two players is randomly selected, and you play two games against that player. But you do not know who you played against. Let $Z = 0$ indicate playing against A, and $Z = 1$ indicate playing against B. Let P_1 and P_2 be indicators of winning the first and second games respectively. Then show that P_1 and P_2 are conditionally independent given $Z = 1$ or $Z = 0$, but not dependent when Z is not known.

Relation between Binomial and Hypergeometric:

Thm. If $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, and X is independent of Y, then the conditional distribution of X given $X + Y = r$ is $\text{HypGeom}(n, m, r)$.

Given $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, $X+Y = r$, and we want to find $P(X = x|X+Y = r)$. By Bayes' Rule,

$$\begin{aligned} P(X = x|X + Y = r) &= \frac{P(X + Y = r|X = x)P(X = x)}{P(X + Y = r)} \\ &= \frac{P(Y = r - x)P(X = x)}{P(X + Y = r)} \end{aligned}$$

Since X and Y are independent, $P(X + Y = r|X = x) = P(Y = r - x)$. As $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, and as X and Y are independent, $X + Y \sim \text{Bin}(n + m, p)$. Thus,

$$P(X = x|X + Y = r) = \frac{\binom{m}{r-x}p^{r-x}(1-p)^{m-r+x}\binom{n}{x}p^x(1-p)^{n-x}}{\binom{n+m}{r}p^r(1-p)^{n+m-r}} = \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}}.$$

An interesting observation is that the conditional distribution of X does not depend on p.

Thm. If $X \sim \text{HypGeom}(w, b, n)$, and $N = w + b \to \infty$ such that $p = \frac{w}{w+b}$ can be assumed to be constant, then the PMF of X converges to $\text{Bin}(n, p)$.

Proof. From the PMF of $X \sim \text{HypGeom}(w, b, n)$,

$$\begin{aligned} P(X = k) &= \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} \\ &= \binom{n}{k}\frac{\binom{w+b-n}{w-k}}{\binom{w+b}{w}} \\ &= \binom{n}{k}\frac{w!}{(w-k)!}\frac{b!}{(b-n+k)!}\frac{(w+b-n)!}{(w+b)!} \\ &= \binom{n}{k}\frac{w(w-1)...(w-k+1)b(b-1)...(b-n+k+1)}{(w+b)(w+b-1)...(w+b-n+1)} \\ &= \binom{n}{k}\frac{p(p-\frac{1}{N})...(p-\frac{k-1}{N})q(q-\frac{1}{N})...(q-\frac{n-k-1}{N})}{(1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{n-1}{N})}. \end{aligned}$$

As $N \to \infty$, $P(X = k) = \binom{n}{k}p^k q^{n-k}$.

(Review) Geometric Distribution: A sequence of independent Bernoulli trials occur that have success probability p and failure probability $q = 1 - p$. Let X be the number of failures before the first success occurs. Then $X \sim \text{Geom}(p)$, and the PMF is,

$$P(X = k) = q^k p$$

for $k = 0, 1, 2, ...$

(Defn.) First Success Distribution: For a sequence of independent Bernoulli trials with success probability p and failure probability $q = 1 - p$, let Y be the number of trials until the first success, including the successful trial. Then $Y \sim \text{FS}(p)$, and the PMF is,

$$P(Y = k) = q^{k-1} p$$

for $k = 0, 1, 2, ...$

Distributions FS and Geom are related in the following way: If $Y \sim \text{FS}(p)$, then $P(Y - 1 = k) = P(Y = k + 1) = q^k p$. Thus, $Y - 1 \sim \text{Geom}(p)$.

Conversely, if $X \sim \text{Geom}(p)$, then $X + 1 \sim \text{FS}(p)$.

(Defn.) Negative Binomial Distributions are generalizations of Geometric distributions, where the distribution of the number of failures in independent Bernoulli trials before the n-th success is achieved is described as $X \sim \text{NegBinom}$, whose PMF is,

$$P(X = k) = \binom{k + n - 1}{n - 1} p^n q^k \ ,$$

for k=0,1,2,... .

Since all trials are independent, X can be described in terms of n random variables $X_1, ..., X_n$, each describing the distribution of the number of failures before a success, i.e., each X_i is i.i.d. $\text{Geom}(p)$. Hence,

$$E(X) = \sum_{i=1}^{n} E(X_i) = n\frac{q}{p} \ .$$

(Defn.) Negative Hypergeometric: An urn contains w white balls and b black balls. The distribution of the number of black balls drawn before obtaining the n-th white ball is described by a Negative Hypergeometric distribution. If $X \sim \text{NegHypGeom}(w, b, n)$, then the PMF is,

$$P(X = k) = \frac{\binom{w}{n-1}\binom{b}{k}}{\binom{w+b}{n-1+k}} \ .$$

The relationship between four distributions are outlined in the table below:

	With Replacement	**W/O Replacement**
Fixed no. of trials	Binomial	Hypergeometric
Fixed no. of successes	Negative Binomial	Negative Hypergeometric

(Defn.) Poisson Distribution: The Poisson distribution is a discrete distribution whose mean and variance are equal. If a random variable X follows Poisson distribution with parameter $lambda > 0$, i.e., $X \sim Pois(\lambda)$, the PMF of $X = k$ is,

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},$$

for $k = 0, 1, 2, ...$

This is a valid PMF, since $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$.

The expected value of X is,

$$E(X) = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} = \lambda$$

The variance can be calculated to be $Var(X) = \lambda$.

The Poisson distribution can be interpreted as counting the number of successes that can occur over a large number of trials, when the trials each have a small probability of success. The parameter λ is interpreted the rate of occurrence of the rare events.

Defn. The Poisson Paradigm: Let $A_1, ..., A_n$ be events with probabilities $P(A_j) = p_j$, where n is large, p_j is small, and A_j are independent or weakly dependent. Let $X = \sum_{j=1}^{n} I(A_j)$ be the count of how many A_j occurs. Then X is approximately distributed as $\mathrm{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^{n} p_j$.

The Poisson paradigm is also called the 'law of rare events'. Even if p_j are small, λ is not, resulting in the possible outcomes.

The Poisson paradigm can often give good approximations to problems that are challenging to solve analytically.

Q. The birthday problem: From a group of n people, the number of pairs of people are $\binom{n}{2}$. The probability of each pair sharing a birthday can be set to $p = \frac{1}{365}$. Then by the Poisson paradigm, the distribution of the number of matching birthdays is approximately $\mathrm{Pois}(\lambda)$, with $\lambda = \binom{n}{2}\frac{1}{365}$. Then the probability of at least one match is,

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - e^{-\lambda}.$$

Q. Near-birthday problem: Let the problem be identifying pairs of people with birthdays within one day of each other. The number of pairs is still $\binom{n}{2}$. The probability of two people having birthdays within one day of each other can be $p = \frac{3}{365}$, where one birthday is fixed and the other birthday is either on that day, or on the day before, or on the next day. With the Poisson paradigm, the probability can be found with $\lambda = \binom{n}{2}\frac{3}{365}$.

Thm. If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$.

Relation between Poisson and Binomial:

Thm. If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then the conditional probability of X given $X + Y = n$ is $\mathrm{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.

Thm. If $X \sim \mathrm{Bin}(n, p)$, and $n \to \infty$ and $p \to 0$ with a constant np, then the PMF of X converges to the PMF $\mathrm{Pois}(\lambda)$.

Q. A website is estimating the daily distribution of visitors to its site. If 10^6 people can independently decide to visite the website, and the probability of visiting the website is 2×10^{-6}, what is the probability that the website will have at least three daily visitors?

Avisek Gupta
Postdoctoral Fellow
IAI, TCG CREST.