STA 623 – Fall 2009 – Dr. Charnigo

Section 3.2: Discrete Distributions

Discrete uniform distribution. A random variable X has the discrete uniform distribution on $\{1, 2, \ldots, N\}$, N a positive integer, if

$$P(X = x) = (1/N)1_{\{x \in \{1,2,\ldots,N\}\}}.$$

Using the formulas $\sum_{x=1}^{N} x = N(N+1)/2$ and $\sum_{x=1}^{N} x^2 = N(N+1)(2N+1)/6$, your textbook authors show that

$$E[X] = (N+1)/2 \quad \text{and} \quad Var[X] = (N+1)(N-1)/12.$$

We can also define a discrete uniform distribution on any set of N distinct points $\{a_1, a_2, \ldots, a_N\} \subset \mathbb{R}$. If these points are equispaced, in the sense that $a_2 - a_1 = a_3 - a_2 = \cdots = a_N - a_{N-1}$, then we can exploit the textbook authors' computations to find the mean and variance.

Example (discrete uniform distribution). Let Y be a random variable with probability mass function $P(Y = y) = (1/9)1_{\{y \in \{1.0,1.5,2.0,2.5,3.0,3.5,4.0,4.5,5.0\}\}}$. If we put $X :=\quad Y + \quad$, then X has the discrete uniform distribution on $\{1, 2, \ldots, 9\}$. We have $E[X] = 5$ and $Var[X] = 20/3$. As such,

$E[Y] =$ and $Var[Y] =$

Hypergeometric distribution. A random variable X has the hypergeometric distribution with parameters M, N, and K (positive integers with $K+M-N \leq \min\{K, M\}$) if

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \quad \text{for integers } x \in [\max\{0, K+M-N\}, \min\{K, M\}].$$

The usual interpretation of the hypergeometric distribution with parameters M, N, and K is that we randomly select K objects without replacement from among N objects, of which M belong to "Group 1" and $N - M$ belong to "Group 2". Then X is the number of the K objects that belong to "Group 1".

Example (hypergeometric distribution). To calculate $E[X]$, we note that

$$
\begin{aligned}
& \sum_{x=\max\{0,K+M-N\}}^{\min\{K,M\}} x\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\
= & \sum_{x=\max\{1,K+M-N\}}^{\min\{K,M\}} \frac{x\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\
= & \sum_{x=\max\{1,K+M-N\}}^{\min\{K,M\}} \frac{M\binom{M-1}{x-1}\binom{N-M}{K-x}}{\binom{N-1}{K-1}N/K} \\
= & \sum_{y=\max\{1,K+M-N\}-1}^{\min\{K,M\}-1} \frac{MK}{N}\frac{\binom{M-1}{y}\binom{N-M}{K-(y+1)}}{\binom{N-1}{K-1}} \\
= & \sum_{y=\max\{0,(K-1)+(M-1)-(N-1)\}}^{\min\{(K-1),(M-1)\}} \frac{MK}{N}\frac{\binom{(M-1)}{y}\binom{(N-1)-(M-1)}{(K-1)-y}}{\binom{(N-1)}{(K-1)}} \\
=
\end{aligned}
$$

One can also establish that

$$Var[X] = \frac{KM(N-M)(N-K)}{N^2(N-1)}.$$

If $K+M-N = \min\{K, M\}$, we have $Var[X] = 0$. To what does the probability mass function simplify in this case, and how we can interpret this in terms of randomly selecting objects without replacement?

Binomial distribution. A random variable X has the binomial distribution with parameters $p \in [0, 1]$ and $n \in \{1, 2, \ldots\}$ if

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} 1_{\{x \in \{0,1,\ldots,n\}\}}.$$

The usual interpretation of the binomial distribution with parameters p and n is that we conduct n independent trials, each of which results in "success" with probability p and "failure" with probability $1 - p$, and then let X denote the total number of successes. These are sometimes called Bernoulli trials. You saw this in Written Assignment 2, in which the n independent trials were the interviews of grocery shoppers, selection of Brand A was a success, and selection of Brand B was a failure.

We have

$$E[X] = np \quad \text{and} \quad Var[X] = np(1-p).$$

Note that we can write

$$X = T_1 + T_2 + \cdots + T_n,$$

where $T_i := 1_{\{\text{“success” on trial i}\}}$ for $i \in \{1, 2, \ldots, n\}$. Simple calculations show that $E[T_i] = p$ and $Var[T_i] = p(1-p)$. Thus, we see that

$$E[T_1 + T_2 + \cdots + T_n] = E[T_1] + E[T_2] + \cdots + E[T_n],$$

which is no surprise, but we also discover that

$$Var[T_1 + T_2 + \cdots + T_n] = Var[T_1] + Var[T_2] + \cdots + Var[T_n].$$

We will see later this semester that the latter result is quite general (i.e., not specific to problems involving the binomial distribution): variances are additive for independent random variables.

Example (binomial distribution). Can you provide an intuitive explanation for the manner in which $Var[X]$ depends on p? Can you think of any good reason that the "parameter space" for p should be $[0, 1]$ rather than $(0, 1)$?

Poisson distribution. A random variable X has the Poisson distribution with parameter $\lambda \in (0, \infty)$ if

$$P(X = x) = \exp[-\lambda]\lambda^x/x! \quad \text{for } x \in \{0, 1, \ldots\}.$$

We have $E[X] = Var[X] = \lambda$.

The usual interpretation of the Poisson distribution with parameter λ is that events are occurring over time (space) such that: (i) the number of events occurring in one time interval (spatial region) is independent of the number of events occurring in another nonoverlapping time interval (spatial region); (ii) as the length of a time interval (area of a spatial region) shrinks to zero, the probability of there being exactly one event divided by the length of the time interval (area of the spatial region) converges to λ; and, (iii) as the length of a time interval (area of a spatial region) shrinks to zero, the probability of there being more than one event divided by the length of the time interval (area of the spatial region) converges to 0. Then X is the number of events occurring over a time interval (spatial region) of unit length (area).

Although less relevant with modern computational resources, the Poisson distribution with parameter np has historically been used to approximate the binomial distribution with parameters p and n when p is small and n is large.

Example (Poisson distribution). To justify the Poisson approximation to the binomial distribution, put $p_n := \min\{1, \lambda/n\}$ and let X_n have the binomial distribution with parameters p_n and n for $n \in \{1, 2, \ldots\}$. Then, as your textbook authors have documented,

$$M_{X_n}(t) = [p_n \exp[t] + (1 - p_n)]^n.$$

For large enough n, we have

$$[p_n \exp[t] + (1 - p_n)]^n = [1 + \lambda(\exp[t] - 1)/n]^n,$$

so that $M_{X_n}(t)$ converges to which is the moment generating function of the Poisson distribution with parameter λ. Hence, for any non-integral $x \in (0, \infty)$, we have $\lim_{n\to\infty} P(X_n \le x) =$

Negative binomial distribution. A random variable X has the negative binomial distribution with parameters $p \in (0, 1]$ and $r \in \{1, 2, \ldots\}$ if

$$P(X = x) = \binom{r + x - 1}{x} p^r (1 - p)^x \quad \text{for } x \in \{0, 1, \ldots\}.$$

We have $E[X] = r(1 - p)/p$ and $Var[X] = r(1 - p)/p^2$.

The usual interpretation of the negative binomial distribution with parameters p and r is that we are conducting Bernoulli trials. However, instead of fixing the number of trials at n, we continue until we attain r successes. Thus, the number of trials is random. We let X denote the number of failures required to attain r successes.

An alternative definition for the negative binomial distribution entails letting Y denote the number of trials required to attain r successes. Obviously $Y = r + X$, so we have

$$P(Y = y) =$$

$$E[Y] = \qquad\qquad \text{and} \quad Var[Y] =$$

Where necessary to distinguish these two definitions, I will refer to the latter as the "offset" negative binomial distribution, as the support of Y is offset from 0.

Example (negative binomial distribution). Take another look at $E[X]$ and $Var[X]$. Can you formulate a conjecture about the circumstances under which there exists a good Poisson approximation to the negative binomial distribution?

Or, going in the other direction, suppose that you originally planned to model the number of events occurring over a time interval (spatial region) of unit length (area) as a Poisson random variable but that your data led you to believe that $E[X] = 5$ and $Var[X] = 6$. With what parameters p and r might you consider describing your data using the negative binomial distribution?

Geometric distribution. A random variable X has the geometric distribution with parameter $p \in (0, 1]$ if

$$P(X = x) = p(1 - p)^x \quad \text{for } x \in \{0, 1, \ldots\}.$$

This is just a special case of the negative binomial distribution with parameters p and $r = 1$. As such, we have $E[X] = (1 - p)/p$ and $Var[X] = (1 - p)/p^2$.

An alternative definition for the geometric distribution entails putting $Y := 1 + X$, so we have

$$P(Y = y) =$$

$$E[Y] = \qquad\qquad\qquad \text{and} \quad Var[Y] =$$

Where necessary to distinguish these two definitions, I will refer to the latter as the "offset" geometric distribution, as the support of Y is offset from 0.

Example (geometric distribution). Why have we excluded 0 from the parameter space for p? Does the reason given for including 0 in the case of the binomial distribution apply here?

Suppose that $X_1, X_2, \ldots, X_r$ are independent geometric random variables with parameter $p \in (0, 1]$. Formulate a conjecture about what will be the distribution of $X_1 + X_2 + \cdots + X_r$.

Can you formulate similar conjectures about any of the other types of discrete random variables that we have studied this week? (The theoretical machinery to prove these conjectures will come later in the semester.)