CmpE 343 Lecture Notes
5: Discrete Distributions

Ethem Alpaydın

October 28, 2014

1 Introduction

Frequently, certain distributions arise in different applications, and they are given names and parameterized. We start by discussing distributions where the random variable is discrete.

2 Uniform Distribution

The random variable X takes one of $K \geq 2$ different values, $X \in \{x_1, x_2, \ldots, x_K\}$, and in the uniform distribution all K values are equally likely, so

$$P\{X = x_i\} = f(x_i) = \frac{1}{K}, \forall i = 1, \ldots, K \tag{1}$$

For example a coin has two faces and if it is a fair coin, both faces are equally likely and have a probability of $1/2$. A die has six faces and if it is a fair die, all faces are equally likely and have a probability of $1/6$.

Let us calculate the expected value and variance of a uniform random variable:

$$\begin{aligned} E[X] &= \sum_{i=1}^{K} x_i f(x_i) = \sum_{i=1}^{K} \frac{x_i}{K} = \frac{\sum_{i=1}^{K} x_i}{K} = \overline{x} \\ \text{Var}(X) &= \sum_{i=1}^{K} (x_i - E[X])^2 f(x_i) = \frac{\sum_{i=1}^{K} (x_i - \overline{x})^2}{K} \end{aligned} \tag{2}$$

where $\overline{x} \equiv \sum_{i=1}^{K} x_i / K$ is the average value for X.

3 Bernoulli Distribution

When we toss a coin there are two outcomes and if the coin is fair, the two probabilities are equal. Consider the general case where the probability of heads is p_0 and hence the probability of tails is $1 - p_0$. This is the Bernoulli distribution where there are two outcomes $X \in \{0, 1\}$, and

$$P\{X = 1\} = f(1) = p_0 \text{ and } P\{X = 0\} = f(0) = 1 - p_0 \tag{3}$$

When $p_0 = 1/2$ we get the uniform distribution but with Bernoulli, the parameter p_0 can take any value between 0 and 1. The two outcomes of 0 and 1 are generally named "failure" and "success" but Bernoulli distribution can denote any experiment with two outcomes, for example, a patient may live or die, it may rain tomorrow or not, a product may be defective or not, an email may be spam or not, and so on.

The two cases of equation (3) can be written as

$$P\{X = x\} = f(x; p_0) = p_0^x (1 - p_0)^{1-x}, x \in \{0, 1\} \tag{4}$$

$f(x;p_0)$ denotes that the probability that the random variable takes the value x, and what follows the semicolon is the parameter of the distribution (which should be set to a particular value for us to be able to calculate the probability).

Let us calculate the expected value and variance of a Bernoulli random variable:

$$\begin{aligned} E[X] &= 0 \cdot f(0) + 1 \cdot f(1) = p_0 & (5) \\ \mathrm{Var}(X) &= (0-p_0)^2 f(0) + (1-p_0)^2 f(1) = (0-p_0)^2 \cdot (1-p_0) + (1-p_0)^2 p_0 = p_0(1-p_0) \end{aligned}$$

4 Multinoulli Distribution

Let us generalize to Bernoulli from two outcomes to arbitrary K, and we get the Multinoulli distribution. There are K distinct states $X \in \{x_1, x_2, \dots, x_K\}$ with probabilities $p_1, p_2, \dots, p_K\}$ satisfying $p_i \geq 0, \forall i$ and $\sum_{i=1}^{K} p_i = 1$. The probability distribution is defined as

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_K = x_K\} = f(x_1, x_2, \dots, x_K; p_1, p_2, \dots, p_K) = \prod_{i=1}^{K} p_i^{x_i} \tag{6}$$

When $p_i = 1/K$ we get the uniform distribution, which is a special case. It is used to represent random experiments with $K \geq 3$ outcomes, for example, a patient may suffer from one of K different diseases, a customer may buy one of K different products, and so on.

5 Binomial Distribution

Let us say we have a coin (not necessarily fair) and we toss it ten times and we are interested in the number of heads. We have n independent repetitions of the same Bernoulli experiment with outcomes of 0 ("failure") and 1 ("success") with probabilities $1 - p_0$ and p_0 respectively and we are interested in the total number of "successes." The probability distribution of such a Binomial random variable is written as

$$P\{X = x\} = f(x; n, p) = \binom{n}{x} p_0^x (1-p_0)^{n-x}, x = 0, 1, \dots, n \tag{7}$$

Each experiment is a success with p_0, so the probability that there are x successes is p_0^x; similarly the probability that there are $n-x$ failures is $(1-p_0)^{n-x}$, and hence the probability that there are both is $p_0^x(1-p_0)^{n-x}$. But this is one possible case of x successes and $n-x$ failures, for example, when the first x experiments are all successes and all the remaining experiments are failures. But there are many possible such cases, for example, the first can be a success, and the next failure and then two successes and so on. Actually there are $\binom{n}{x}$ different ways of having x successes and $n-x$ failures, so we sum over all them. For example there are $\binom{3}{1}$ ways of seeing one heads in three tosses, namely $\{HTT, THT, TTH\}$.

Let us calculate the expected value and variance of a Binomial random variable and in doing that, we see that we can write the Binomial random variable X as the sum of n independent 0/1 Bernoulli random variables (each for example representing the outcome of one coin toss):

$$P\{X = x\} = \sum_{j=0}^{n} I_j \tag{8}$$

and hence

$$\begin{aligned} E[X] &= E\left[\sum_{j=0}^{n} I_j\right] = \sum_{j=0}^{n} E[I_j] = np_0 & (9) \\ \mathrm{Var}(X) &= \mathrm{Var}\left(\sum_{j=0}^{n} I_j\right) = \sum_{j=0}^{n} \mathrm{Var}(I_j) = np_0(1-p_0) \end{aligned}$$

Note that in calculating the variance we used the fact that I_j are independent.

6 Multinomial Distribution

Just like the Binomial distribution is the number of occurrences of Bernoulli events, the Multinomial is the number of occurrences of Multinoulli events. Bernoulli has two outcomes and hence Binomial keeps track of two counts, Multinoulli has $K \geq 3$ outcomes each with probability $p_i, i = 1, \dots, K$ and hence the Multinomial keeps track of K counts, namely, $X_1, X_2, \dots, X_K$ where X_i is the number of occurrence of outcome i in n independent repetitions. Generalizing equation [7) from 2 to K, the probability distribution of the Multinomial random variables $X_1, X_2, \dots, X_K$ is written as

$$P\{X_1 = x_1, X_2 = x_2, \dots, X_K = x_K\} = f(x_1, x_2, \dots, x_K; n, p_1, p_2, \dots, p_K) = \frac{n!}{x_1! x_2! \cdots x_K!} \prod_{i=1}^{K} p_i^{x_i} \tag{10}$$

7 Hypergeometric Distribution

The Binomial distribution assumes that repetitions are independent, and so that it is always the same Bernoulli event that is repeated independently. If the repetitions are not independent, Binomial distribution cannot be used because it is not the same Bernoulli event that is repeated. If we are tossing a coin or rolling a die, the repeated events are independent and we can use the Binomial or the Multinomial because the outcome of the previous experiment has no effect on the coin or the die and hence does not affect the probabilities. But if for example we are drawing balls from a bag or drawing cards from a deck, if we do it *with replacement*, that is, if we return the ball to the bag or the card to the deck then successive events are independent and we can use the Bernoulli or the Multinoulli for each event, but if we do it *without replacement* because the composition of the bag or the deck changes after each event, it is not the same event that is repeated, and the probabilities change depending on the outcome of the previous event—we need to start using conditional probabilities.

For example, let us say we have a bag that contains four balls of which two are red and two are black. Let us say we draw two balls at random and want to calculate the probability that of the two, one is red. For there to be one red ball, the other should be black and the red ball can be seen in the first draw ('R_1B_2') or the second draw ('B_1R_2') ; we need to calculate the probability of each and sum them: $P(X = 1) = P(\text{'}R_1B_2') + P(\text{'}B_1R_2')$.

Drawing with replacement: The probability of drawing a red ball is $2/4 = 1/2$ and it does not change after the first draw. The probability of drawing a black ball is also $1/2$. So $P(\text{'}R_1B_2') = 1/2 \cdot 1/2$ (the two draws are independent) and $P(\text{'}BR') = 1/2 \cdot 1/2$, and so $P(X = 1) = P(\text{'}RB') + P(\text{'}BR') = 1/2$. Or we can say that each draw is Bernoulli where the probability of red ball is $1/2$ and so the probability that we see one red in two repetitions is

$$P(X = 1) = \binom{2}{1} (1/2)^1 (1/2)^1 = 1/2$$

Drawing without replacement: In this case, the probabilities in the second draw depend on the first draw. The number of balls that remain in the bag is one less and also either the number of reds or blacks change—the two draws are dependent and we need to use conditional probabilities.

$$P(\text{'}R_1B_2') = P(R_1)P(B_2|R_1) = (2/4) \cdot (2/3) = 1/3$$

Similarly $P(\text{'}B_1R_2) = 1/3$ and $P(X = 1) = 2/3$. Or we can calculate this as

$$P(X = 1) = \frac{\binom{2}{1}\binom{2}{1}}{\binom{4}{2}} = 2/3$$

You can show that $P(X = 0) = P(X = 2) = 1/6$ and that $\sum_{x=0}^{2} P(X = x) = 1$.

This last formula is the Hypergeometric distribution. Let us say we have a set of N objects where k of them are of one kind (and $N - k$ are of the second kind)and we draw n instances at random without replacement, the probability that x of these n are of the first kind (and $n - x$ are of the second kind) is given by

$$f(x; N, n, k) = \binom{k}{x}\binom{N-k}{n-x} \Big/ \binom{N}{n} \tag{11}$$

Hypergeometric distribution is used frequently in settings for testing and quality control where we choose a small subset from a large batch of products and try to estimate the quality of the large batch from the proportion of defectives in the small subset. In such a case, it does not make sense to return a tested item back into the subset, and most of the time we cannot—testing may make the item unusable. This is called *acceptance testing*.

The expected value and variance of a Hypergeometric distributed random variable X is

$$\begin{aligned} E[X] &= n\frac{k}{N} \qquad (12)\\ \mathrm{Var}(X) &= \left(\frac{N-n}{N-1}\right) n\frac{k}{N}\left(1 - \frac{k}{N}\right) \end{aligned}$$

Let us compare these with those of Binomial given in equation (**??**). We see that k/N corresponds to p_0 of the Binomial; we also see that $(N - n)/(N - 1)$ gets closer to 1 as n is small with respect to N, and the formulas for variance become the same too. This make sense if we are drawing a small subset from a very large set, whether we do it with replacement or not does not make much difference.

8 Multivariate Hypergeometric Distribution

Just as the Multinoulli is the generalization of Bernoulli, and the Multinomial is the generalization of Binomial to Multinomial from two to $K \geq 3$, Multivariate Hypergeometric is the generalization of the Hypergeometric distribution. We have N items where N_1 are of the first kind, N_2 of the second kind, until N_K of the Kth kind. We draw n items at random without replacement. The probability that in these n, we see x_1 of the first kind, and so on () is

$$\begin{aligned} P\{X_1 = x_1, X_2 = x_2, \ldots, X_K = x_K\} &= \\ f(x_1, x_2, \ldots, x_K; N, n, N_1, N_2, \ldots, N_K) &= \binom{N_1}{x_1}\binom{N_2}{x_2}\cdots\binom{N_K}{x_K} \Big/ \binom{N}{n} \qquad (13) \end{aligned}$$

where $\sum_{i=1}^{N} N_i = N$ and $\sum_{i=1}^{N} x_i = n$.

9 Negative Binomial and Geometric Distributions

In the Bionomial distribution, we repeat the Bernoulli experiment n times and we are interested in the probability of seeing X successes. In the Negative Binomial, we fix the number of successes as k and we are interested in the probability that it takes X repetitions to see that many successes.

Let us think as follows: We are interested in the probability that it takes X repetitions to see k successes. So the kth success should occur in the Xth repetition, which means that we need to see $k - 1$ successes in the preceding $X - 1$ repetitions, the probability of which can be calculated using the Binomial distribution. Following this independently, we see another success, with probability p_0. Hence we have

$$\begin{aligned} P\{X = x\} = f(x; k, p_0) &= \binom{x-1}{k-1} p_0^{k-1}(1 - p_0)^{x-k} \cdot p \\ &= \binom{x-1}{k-1} p_0^{k}(1 - p_0)^{x-k}, x = 1, 2, 3, \ldots \qquad (14) \end{aligned}$$

Geometric distribution is a special case where we are interested in the probability that we see the *first* success in the Xth repetition. The previous (independent) $X - 1$ repetitions should all be failures and then we see a success:

$$P\{X = x\} = f(x; p_0) = (1 - p_0)^{x-1} p_0, x = 1, 2, 3, \ldots \tag{15}$$

The expected value and variance of a Geometric random variable X is given as

$$E[X] = \frac{1}{p_0} \tag{16}$$

$$\text{Var}(X) = \frac{1-p_0}{p_0^2} \tag{17}$$

So for example if the probability of a spam email is 0.01, we will see on the average one after 100 emails. We use Binomial or the Negative Binomial depending on whether we fix the number of repetitions and count the successes, or fix the successses and count the repetitions.