**LECTURE 7**
**EXAMPLES OF DISCRETE DISTRIBUTIONS**

In this lecture, we consider some common examples of *families* of discrete distributions: collections of PMFs described by one or more *parameters*.

# 1 Bernoulli trials

This is probably the simplest example of a distribution. It arises when the outcome of an experiment is classified in one of two mutually exclusive categories: success or failure. A Bernoulli random variable is an indicator function for success in an experiment.

**Definition 1. (Bernoulli distribution)** We say that $X$ has a Bernoulli distribution *with parameter* $p \in [0,1]$, denoted

$$X \sim Bernoulli(p),$$

if the support of $X$ is $S_X = \{0,1\}$, and the PMF is

$$\begin{aligned} p_X(1) &= p, \\ p_X(0) &= 1-p. \end{aligned}$$

*Remark.*

1. An alternative and more compact representation for the PMF of a Bernoulli random variable is
$$p_X(x) = p^x(1-p)^{1-x}, \quad x \in \{0,1\}.$$
2. Definition 1 introduces a *family (or collection) of distributions*: we have a different distribution for every value of the parameter $p \in [0,1]$.
3. Once we say that some random variable follows Bernoulli distribution, its behavior is completely described by a single parameter $p$.
4. Sometimes when we need to emphasize the dependence of a distribution on the value of a parameter, we would write the PMF as $p_X(x;p)$.
5. When a distribution of $X$ is determined by a parameter $p$, its mean, variance and other moments are also functions of $p$:
$$\begin{aligned} \mu_X(p) &= EX = \sum_{x \in S_X} x p_X(x;p), \\ \sigma_X^2 &= Var(X) = \sum_{x \in S_X} (x - \mu_X(p))^2 p_X(x;p). \end{aligned}$$

Lets find the mean and variance of $X \sim Bernoulli(p)$. The mean of $X$ is

$$EX = 1 \times p + 0 \times (1-p) = p.$$

Thus, in the case of Bernoulli distributions the mean provides a complete information on the distribution of the random variable. To find the variance of $X$, first note that

$$EX^2 = 1^2 \times p = p,$$

and, therefore, by Theorem 2(d) in Lecture 6,

$$\begin{aligned} Var(X) &= EX^2 - (EX)^2 \\ &= p - p^2 \\ &= p(1-p). \end{aligned}$$

Lastly, the MGF of $X$ is

$$\begin{aligned} M_X(t) &= Ee^{tX} \\ &= e^t p + e^0(1-p) \\ &= pe^t + (1-p). \end{aligned}$$

## 2 Geometric distribution

Consider a sequence of independent Bernoulli trials each having the same probability of success $p$, and let $X$ denote the number of trials until the first success. The distribution of $X$ is called geometric.

**Definition 2. (Geometric distribution)** We say that $X$ has a geometric distribution with parameter $p \in [0,1]$, denoted

$$X \sim Geometric(p),$$

if the support of $X$ is $S_X = \{1, 2, \ldots\}$ and the PMF of $X$ is

$$p_X(x) = p(1-p)^{x-1}. \tag{1}$$

*Remark.*

1. The family of Geometric distributions is described by one parameter $p$ (the probability of success in a single trial).

2. To have a Geometric distribution, the trials must be independent and have the same probability of success in each trial.

The function $p_X(\cdot)$ in (1) is a PMF since $p_X(x) > 0$ for all $x \in S_X$ and, for $q = 1 - p$,

$$\begin{aligned}\sum_{x=1}^{\infty} p_X(x) &= \sum_{x=1}^{\infty} pq^{x-1} \\ &= p(1 + q + q^2 + \ldots) \\ Q &= p\frac{1}{1-q} \\ &= p\frac{1}{p} \\ &= 1.\end{aligned}$$

To find the MGF of the Geometric distribution with parameter $p$, consider

$$\begin{aligned}\sum_{x=1}^{\infty} e^{tx} p(1-p)^{x-1} &= pe^t \sum_{x=1}^{\infty} e^{t(x-1)}(1-p)^{x-1} \\ &= pe^t \sum_{x=1}^{\infty} (e^t(1-p))^{x-1}.\end{aligned}$$

For this sum of infinitely many terms to be finite, we need that

$$e^t(1-p) < 1,$$

or

$$t < -\log(1-p). \tag{2}$$

Hence, the MGF of a geometric distribution is defined only for the values of $t$ satisfying (2). For such values of $t$,

$$\begin{aligned}M_X(t) &= pe^t \sum_{x=1}^{\infty} ((1-p)e^t)^{x-1} \\ &= \frac{pe^t}{1-(1-p)e^t} \\ &= p\left(e^{-t} - (1-p)\right)^{-1}.\end{aligned}$$

To compute the mean and variance of a Geometric distribution,

$$\begin{aligned}\frac{dM_X(t)}{dt} &= p\left(e^{-t} - (1-p)\right)^{-2} e^{-t} \\ \frac{d^2M_X(t)}{dt^2} &= 2e^{-2t} p\left(e^{-t} - (1-p)\right)^{-3} - p\left(e^{-t} - (1-p)\right)^{-2} e^{-t}.\end{aligned}$$

Hence,

$$\begin{aligned} EX &= \frac{dM_X(0)}{dt} = \frac{1}{p} \\ EX^2 &= \frac{d^2M_X(0)}{dt^2} = \frac{2}{p^2} - \frac{1}{p}. \end{aligned}$$

Therefore,

$$\begin{aligned} Var(X) &= EX^2 - (EX)^2 \\ &= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{1}{p^2} - \frac{1}{p} \\ &= \frac{1-p}{p^2}. \end{aligned}$$

## 3 Binomial distribution

Consider a sequence of $n$ independent, identical Bernoulli trials with a probability of success in a single trial equal to $p$, and let $X$ denote the number of successes. In this case, $X$ can take values in $\{0, 1, \ldots, n\}$. For example, consider the following sequence of $n$ trials with $x$ successes and $n - x$ failures:

$$\underbrace{1, 1, \ldots, 1}_{x \text{ times}}, \underbrace{0, 0, \ldots, 0}_{n-x \text{ times}}.$$

Since the trials are independent, the probability of observing this sequence is

$$\underbrace{p \times p \times \ldots \times p \times}_{x \text{ times}} \underbrace{(1-p) \times (1-p) \times \ldots \times (1-p)}_{n-x \text{ times}} = p^x(1-p)^{n-x}.$$

However, since the order of successes and failures does not matter, we can have

$$C(n, x) = \frac{n!}{(n-x)!x!} = \binom{n}{x}$$

distinct sequences of trials with $x$ successes and $n - x$ failures. Hence,

$$P(X = x) = \left( \begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{n-x}.$$

**Definition 3. (Binomial distribution)** We say that $X$ has a Binomial distribution with parameters $n \in \{1, 2, \ldots\}$ and $p \in [0, 1]$, denoted

$$X \sim Binomial(n, p),$$

if the support of $X$ is $S_X = \{0, 1, \ldots\}$, and the PMF of $X$ is

$$p_X(x) = \left( \begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}. \tag{3}$$

*Remark.*

1. The family of Binomial distributions is described by two parameters: $n$ (the number of trials) and $p$ (the probability of success in a single trial).

2. To have a Binomial distribution, the trials must be independent and have the same probability of success in each trial.

3. In MS Excel, binomial probabilities can be computed using a function BINOM.DIST($x$,$n$,$p$,CUMULATIVE), where CUMULATIVE=FALSE for the PMF at $x$, and CUMULATIVE=TRUE for the cumulative probability $P(X \leq x)$ (the CDF at $x$).

4. In Stata, cumulative binomial probabilities can be computed using a command `display binomial(`$n$`,`$x$`,`$p$`)`. For example, when $n = 10$ and $p = 0.4$, to compute all cumulative probabilities for $x = 0, 1, \ldots, 10$ type:
   ```
   for num 0/10:  display binomial(10,X,0.4)
   ```

To verify that (3) is a PMF, we need to check that $\sum_{x=0}^{n} p_X(x) = 1$. However,

$$\begin{aligned} \sum_{x=0}^{n} \left( \begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} &= (p + (1-p))^n \\ &= 1^n \\ &= 1, \end{aligned}$$

where the first equality follows from the properties of binomial coefficients and binomial expansions (see equation (1.2-1) on page 15 in Hogg, Tanis, and Zimmerman). Next, we will derive the MGF of the Binomial distribution:

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} \\ &= (e^t p + (1-p))^n, \end{aligned}$$

where the third equality is again by the properties of binomial coefficients and binomial expansions.

Given the MGF, we can compute the mean and variance of a Binomial distribution given parameter values $n$ and $p$.

$$\begin{aligned} \frac{dM_X(t)}{dt} &= n(e^t p + (1-p))^{n-1} p e^t, \\ \frac{d^2 M_X(t)}{dt^2} &= n(n-1)(e^t p + (1-p))^{n-2} p^2 e^t + n(e^t p + (1-p))^{n-1} p e^t. \end{aligned}$$

Hence,

$$\begin{aligned} EX &= \frac{dM_X(0)}{dt} \\ &= np. \end{aligned}$$

The result is quite intuitive: since there are $n$ independent trials with probability $p$ of success, we should expect on average $np$ successful trials.

Next,

$$\begin{aligned} EX^2 &= \frac{d^2 M_X(0)}{dt^2} \\ &= n(n-1)p^2 + np. \end{aligned}$$

Hence,

$$\begin{aligned} Var(X) &= EX^2 - (EX)^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= np - np^2 \\ &= np(1-p). \end{aligned}$$

Note that the variance of a Binomial random variable is $n$ times the variance of a Bernoulli random variable. We will come back to this connection later.

## 4 Poisson distribution

The last distribution we consider in this lecture can be derived as a limiting case of the Binomial distribution with parameters $n$ and $p$. Suppose that $n$ Bernoulli trials must be performed in a fixed time interval. Suppose further that the number of trials is very large ($n \to \infty$), and the probability of success in each trial is very small ($p \to 0$). However, assume that the expected number of successes is fixed: $np = \lambda$ for some $\lambda > 0$. Thus, we have many trials with very small probability of success performed every $1/n$ of the time interval. The distribution arising in the limit as $n \to \infty$ and $p \to 0$ while $np = \lambda$ is called Poisson with parameter $\lambda \in (0, \infty)$. It is useful for modeling the number of arrivals (taxis, phone calls, web-page hits) in a given time interval, assuming that arrivals are independent.

The PMF of the Poisson distribution can be derived by taking the limit of the Binomial PMF. The Binomial PMF for $x$ successes is

$$\begin{aligned} \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} &= \frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\ &= \frac{\lambda^x}{x!} \frac{n(n-1)\dots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{\lambda}{n}\right)^{n}, \end{aligned} \tag{4}$$

where in the first line we used $p = \lambda/n$. To find the limit of this expression, we will consider it term-by-term. Fix $x$. First,

$$\begin{aligned} \frac{n(n-1)\dots(n-x+1)}{n^x} &= \frac{n}{n}\frac{(n-1)}{n}\dots\frac{(n-x+1)}{n} \\ &\to 1 \end{aligned} \tag{5}$$

as $n \to \infty$ since $x$ is fixed. Next, as $n \to \infty$,

$$\left(1 - \frac{\lambda}{n}\right)^{-x} \to 1. \tag{6}$$

Lastly, we will show that

$$\left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda} \tag{7}$$

as $n \to \infty$.

**Lemma 4.** *For $\lambda \in (0, \infty)$, $\lim_{n\to\infty}(1 - \lambda/n)^n = e^{-\lambda}$.*

*Proof.* The result is equivalent to

$$\begin{aligned} \log\left(\left(1 - \frac{\lambda}{n}\right)^{n}\right) &= n \log\left(1 - \frac{\lambda}{n}\right) \\ &\to -\lambda, \end{aligned} \tag{8}$$

since

$$a = e^{\log a}.$$

Consider the second-order Taylor expansion of $\log(1 + u)$ around zero:

$$\log(1 - u) = 0 - u - \frac{u^2}{2} + R_n,$$

where $R_n$ is the remainder term satisfying

$$|R_n| \leq Cu^2$$

for some constant $C > 0$. Using this with $u = \lambda/n$, we obtain:

$$\begin{aligned} n \log\left(1 - \frac{\lambda}{n}\right) &= n\left(-\frac{\lambda}{n} - \frac{1}{2}\frac{\lambda^2}{n^2} + R_n\right) \\ &= -\lambda - \frac{1}{2}\frac{\lambda^2}{n} + nR_n. \end{aligned}$$

As $n \to \infty$,

$$\frac{\lambda^2}{n} \to 0$$

since $\lambda$ is a fixed number. Similarly,

$$\begin{aligned} n|R_n| &\leq nC\frac{\lambda^2}{n^2} \\ &= C\frac{\lambda^2}{n} \\ &\to 0. \end{aligned}$$

Hence, (8) holds and the result follows. □

Combining the results in (4)-(7), we obtain that

$$\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \to \frac{\lambda^x}{x!}e^{-\lambda}.$$

**Definition 5. (Poisson distribution)** We say that $X$ has a Poisson distribution with parameter $\lambda \in (0, \infty)$, denoted

$$X \sim Poisson(\lambda),$$

if the support of $X$ is $S_X = \{0, 1, \ldots\}$, and the PMF of $X$ is

$$p_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}. \tag{9}$$

*Remark.* The family of Poisson distributions can be used to approximate Binomial distributions, provided that the number of Bernoulli trials is large and successes are rare. Note that calculations using the Poisson distribution distribution are simpler than those with Binomial probabilities.

To verify that $p_X(\cdot)$ in (9) is indeed a PMF, note first that $e^\lambda$ has an infinite-order Taylor expansion around $e^0 = 1$:

$$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} \cdots.$$

Therefore,

$$\begin{aligned}
\sum_{x=0}^{\infty} p_X(x) &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} \\
&= e^\lambda e^{-\lambda} \\
&= 1.
\end{aligned}$$

To find the MGF of the Poisson distribution,

$$\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} \\
&= \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} e^{-\lambda}. \qquad (10)
\end{aligned}$$

Define

$$\theta = e^t \lambda,$$

so that

$$\lambda = \theta e^{-t},$$

and re-write (10) as

$$\begin{aligned}
M_X(t) &= \sum_{x=0}^{\infty} \frac{\theta^x}{x!} e^{-\theta e^{-t}} \\
&= e^{-\theta e^{-t}} e^{\theta} \sum_{x=0}^{\infty} \frac{\theta^x}{x!} e^{-\theta} \\
&= e^{\theta(1-e^{-t})} \\
&= e^{e^t \lambda (1-e^{-t})} \\
&= e^{\lambda(e^t - 1)}.
\end{aligned}$$

Hence, the MGF of the Poisson distribution is

$$M_X(t) = e^{\lambda(e^t-1)}.$$

To find the mean and variance of the Poisson distribution,

$$\begin{aligned}\frac{dM_X(t)}{dt} &= \lambda e^{\lambda(e^t-1)+t},\\ \frac{d^2M_X(t)}{dt^2} &= \lambda e^{\lambda(e^t-1)+t}(\lambda e^t + 1).\end{aligned}$$

Hence,

$$\begin{aligned}EX &= \frac{dM_X(0)}{dt} = \lambda,\\ EX^2 &= \frac{d^2M_X(0)}{dt^2} = \lambda^2 + \lambda,\end{aligned}$$

and therefore,

$$\begin{aligned}Var(X) &= (\lambda^2 + \lambda) - \lambda^2\\ &= \lambda.\end{aligned}$$