CHAPTER 3

3. DISCRETE PROBABILITY DISTRIBUTIONS

3.1 DISCRETE UNIFORM DISTRIBUTION

NOTATION $X \sim Uniform(n)$ **RANGE** $\mathbb{X} = \{1, 2, ..., n\}$

MASS FUNCTION

$$f_X(x) = \frac{1}{n} \qquad \text{for } x \in \{1, 2, ..., n\}$$

CDF

$$F_X(x) = \frac{x}{n} \qquad \text{for } x \in \{1, 2, ..., n\}$$

MGF

$$M_X(t) = \sum_{x=1}^{n} \mathrm{e}^{tx} \frac{1}{n} = \frac{\mathrm{e}^t}{n}\left[1 + \mathrm{e}^t + ... + \mathrm{e}^{(n-1)t}\right] = \frac{\mathrm{e}^t}{n}\frac{\left[1 - \mathrm{e}^{nt}\right]}{\left[1 - \mathrm{e}^t\right]}$$

rth MOMENT

$$M_X^{(r)}(t) = \sum_{x=1}^{n} x^r \, \mathrm{e}^{tx} \frac{1}{n} \qquad \Longrightarrow M_X^{(r)}(0) = \frac{1}{n}\sum_{x=1}^{n} x^r$$

$$\Longrightarrow \mathrm{E}_{f_X}[X] = M_X^{(1)}(0) = \frac{1}{n}\sum_{x=1}^{n} x = \frac{(n+1)}{2}$$

$$\mathrm{E}_{f_X}[X^2] = M_X^{(2)}(0) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6}$$

$$\Longrightarrow \mathrm{Var}_{f_X}[X] = \mathrm{E}_{f_X}[X^2] - \{\mathrm{E}_{f_X}[X]\}^2 = \frac{(n+1)(n-1)}{12}$$

Note

We can define a discrete uniform distribution over any finite set of values rather that merely the integers $\{1, 2, ..., n\}$; in this case, the moments of the distribution will depend on the nature of the range $\mathbb{X}$, but in the same techniques for calculation of moments, mgf etc. can be used.

3.2 BERNOULLI DISTRIBUTION

NOTATION $X \sim Bernoulli(\theta)$ **RANGE** $\mathbb{X} = \{0, 1\}$

MASS FUNCTION

$$f_X(x) = \theta^x (1-\theta)^{1-x} \quad \text{for } x \in \{0,1\} \quad \text{where } 0 \leq \theta \leq 1.$$

CDF

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-\theta & \text{for } x = 0 \\ 1 & x \geq 1 \end{cases}$$

MGF

$$M_X(t) = \sum_{x=0}^{1} \mathrm{e}^{tx} \theta^x (1-\theta)^{1-x} = 1 - \theta + \theta \mathrm{e}^t$$

rth MOMENT

$$\begin{aligned} M_X^{(r)}(t) &= \theta \mathrm{e}^t \\ \Longrightarrow M_X^{(r)}(0) &= \theta \quad r = 1, 2, ... \\ \Longrightarrow \mathrm{E}_{f_X}[X] &= \theta \\ \mathrm{E}_{f_X}[X^2] &= \theta \\ \Longrightarrow \mathrm{Var}_{f_X}[X] &= \theta - \theta^2 = \theta(1-\theta) \end{aligned}$$

Notes

(1) The Bernoulli distribution is used for modelling when the outcome of an experiment is either a "success" or a 'failure", where the probability of getting a success is equal to θ.

(2) The mass function of X can be re-written

$$\mathrm{P}[X = x] = \exp\{t(x)q(\theta) + \log h(x) + \log c(\theta)\}$$

where $t(x) = x$, $q(\theta) = \log(\theta/(1-\theta))$, $h(x) = 1$, $c(\theta) = 1 - \theta$.

3.3 BINOMIAL DISTRIBUTION

NOTATION $X \sim Bin(n, \theta)$ **RANGE** $\mathbb{X} = \{0, 1, 2, ..., n\}$

MASS FUNCTION

$$f_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad \text{for } x \in \{0, 1, 2, ..., n\} \qquad \text{where } n \geq 0,\ 0 \leq \theta \leq 1.$$

MGF

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} \theta^x (1-\theta)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(\theta e^t\right)^x (1-\theta)^{n-x} = \left(1-\theta+\theta e^t\right)^n$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = n\theta \mathrm{e}^t \left(1-\theta+\theta \mathrm{e}^t\right)^{n-1} \qquad M_X^{(2)}(t) = n(n-1)\left\{\theta \mathrm{e}^t\right\}^2 \left(1-\theta+\theta \mathrm{e}^t\right)^{n-2} + n\theta \mathrm{e}^t \left(1-\theta+\theta \mathrm{e}^t\right)^{n-1}$$

so that $M_X^{(1)}(0) = n\theta$ and $M_X^{(2)}(0) = n(n-1)\theta^2 + n\theta$, and thus

$$\mathrm{E}_{f_X}[X] = n\theta \qquad \mathrm{Var}_{f_X}[X] = n(n-1)\theta^2 + n\theta - n^2\theta^2 = n\theta(1-\theta)$$

Notes

(1) If $X_1, ..., X_k$ are independent and identically distributed (i.i.d.) $Bernoulli(\theta)$ random variables, and $Y = X_1 + ..., X_k$, then by the standard result for mgfs,

$$M_Y(t) = \{M_X(t)\}^k = \left(1-\theta+\theta \mathrm{e}^t\right)^k$$

so therefore $Y \sim Bin(k, \theta)$ because of the uniqueness of mgfs. Thus the binomial distribution is used to model the **total** number of successes in a series of independent and identical experiments.

(2) Alternatively, consider **sampling without replacement** from **infinite** collection, or **sampling with replacement** from a **finite** collection of objects, a proportion θ of which are of Type I, and the remainder are of Type II. If X is the number of Type I objects in a sample of n, $X \sim Bin(n, \theta)$.

(3) The mass function of X can be re-written

$$\mathrm{P}[X = x] = \exp\left\{t(x)q(\theta) + \log h(x) + \log c(\theta)\right\}$$

where $t(x) = x$, $q(\theta) = \log(\theta/(1-\theta))$, $h(x) = \binom{n}{x}$, $c(\theta) = (1-\theta)^n$.

3.4 POISSON DISTRIBUTION

NOTATION $X \sim Poisson(\lambda)$ **RANGE** $\mathbb{X} = \{0, 1, 2, ..\}$

MASS FUNCTION

$$f_X(x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} \qquad \text{for } x \in \{0, 1, 2, ...\} \qquad \text{where } \lambda > 0.$$

MGF

$$M_X(t) = \sum_{x=0}^{\infty} \mathrm{e}^{tx} \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!} = \mathrm{e}^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda \mathrm{e}^t\right)^x}{x!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda \mathrm{e}^t} = \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\}$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \lambda \mathrm{e}^t \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\} \qquad M_X^{(2)}(t) = \left(\lambda \mathrm{e}^t\right)^2 \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\} + \lambda \mathrm{e}^t \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\}$$

so that $M_X^{(1)}(0) = \lambda$ and $M_X^{(2)}(0) = \lambda^2 + \lambda$, and thus

$$\mathrm{E}_{f_X}[X] = \lambda \qquad \mathrm{Var}_{f_X}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Notes

(1) If $X \sim Bin(n, \theta)$, let $\lambda = n\theta$. Then

$$M_X(t) = \left(1 - \theta + \theta \mathrm{e}^t\right)^n = \left(1 + \lambda(e^t - 1))/n\right)^n \longrightarrow \exp\left\{\lambda\left(\mathrm{e}^t - 1\right)\right\}$$

as $n \longrightarrow \infty$, which is the mgf of a Poisson random variable. Therefore, the Poisson distribution arises as the limiting case of the binomial distribution, when $n \longrightarrow \infty, \theta \longrightarrow 0$ with $n\theta = \lambda$ constant (that is, for "large" n and "small" θ).

(2) Suppose that X_1 and X_2 are independent, with $X_1 \sim Poisson(\lambda_1)$, $X_2 \sim Poisson(\lambda_2)$, then if $Y = X_1 + X_2$, using the general mgf result for independent random variables,

$$M_Y(t) = M_{X_1}(t) M_{X_2}(t) = \exp\left\{\lambda_1\left(\mathrm{e}^t - 1\right)\right\} \exp\left\{\lambda_2\left(\mathrm{e}^t - 1\right)\right\} = \exp\left\{(\lambda_1 + \lambda_2)\left(\mathrm{e}^t - 1\right)\right\}$$

so that $Y \sim Poisson(\lambda_1 + \lambda_2)$. Therefore, the sum of two independent Poisson random variables also has a Poisson distribution.

By induction, this result can be extended to give the distribution of the sum of k independent random Poisson random variables, that is, if $X_1, ..., X_k$ are independent random variables with $X_i \sim Poisson(\lambda_i)$ for $i = 1, ..., k$, then

$$Y = \sum_{i=1}^{k} X_i \Longrightarrow Y \sim Poisson\left(\sum_{i=1}^{k} \lambda_i\right).$$

(3) **The Poisson Process**

Consider an experiment involving events that occur repeatedly in time. Let $X(t)$ be the random variable representing the number of events that occur in the interval $(0, t]$, so that $X(t)$ takes values $0, 1, 2, ...$ Suppose that

1. $X(0) = 0$

2. For all $0 < s \leq t$, $h > 0$, and non-negative integers n and m,

$$\mathrm{P}[X(t+h) - X(t) = n | X(s) = m] = \mathrm{P}[X(t+h) - X(t) = n]$$

that is, the numbers of events occurring in disjoint intervals are probabilistically **independent**.

3. For $\delta t > 0$ small,

$$\mathrm{P}[X(t+\delta t) - X(t) = 1] = \lambda \delta t + O(\delta t)$$

for some $\lambda > 0$, where

$$\lim_{\delta t \longrightarrow 0} \frac{O(\delta t)}{\delta t} = 0$$

that is, the probability of **exactly one** event occurring in the small interval $(t, t + \delta t]$ is, for small δt, proportional to the length of the interval, δt.

4. For $\delta t > 0$ small,

$$\mathrm{P}[X(t+\delta t) - X(t) \geq 2] = O(\delta t)$$

for some $\lambda > 0$, that is, the probability of **more than one** event occurring in a small interval $(t, t + \delta t]$ is essentially zero.

Then, if

$$P_n(t) = \mathrm{P}[n \text{ events occur in } (0, t]]$$

it can be shown that

$$P_n(t) = \mathrm{P}[X(t) = n] = \frac{\mathrm{e}^{-\lambda t}(\lambda t)^n}{n!}$$

(that is, the random variable corresponding to the number of events that occurs in the interval $(0, t]$ has a Poisson distribution with parameter λt.)

Examples : Failures/breakdowns of mechanical components, occurrence of accidents, emission of particles from radioactive sources etc.

(4) The mass function of X can be re-written

$$\mathrm{P}[X = x] = \exp\{t(x)q(\lambda) + \log h(x) + \log c(\lambda)\}$$

where $t(x) = x$, $q(\lambda) = \log\lambda$, $h(x) = \dfrac{1}{x!}$, $c(\lambda) = \mathrm{e}^{-\lambda}$.

3.5 GEOMETRIC DISTRIBUTION

NOTATION $X \sim Geometric(\theta)$ **RANGE** $\mathbb{X} = \{1, 2, ...\}$

MASS FUNCTION

$$f_X(x) = (1-\theta)^{x-1}\theta \qquad \text{for } x \in \{1,2,...\} \qquad \text{where } 0 \leq \theta \leq 1.$$

CDF

$$F_X(x) = 1 - (1-\theta)^x \qquad \text{for } x = 1, 2, ...$$

MGF

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx}(1-\theta)^{x-1}\theta = \theta e^t \sum_{x=1}^{\infty} e^{t(x-1)}(1-\theta)^{x-1} = \theta e^t \sum_{x=0}^{\infty} \left(e^t(1-\theta)\right)^x = \frac{\theta e^t}{1-e^t(1-\theta)}$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \frac{\theta e^t}{\left[1-e^t(1-\theta)\right]^2} \qquad M_X^{(2)}(t) = \frac{\theta e^t\left[1-e^t(1-\theta)\right]\left[1+e^t(1-\theta)\right]}{\left[1-e^t(1-\theta)\right]^4}$$

so that $M_X^{(1)}(0) = \dfrac{1}{\theta}$ and $M_X^{(2)}(0) = \dfrac{2-\theta}{\theta^2}$, and thus

$$E_{f_X}[X] = \frac{1}{\theta} \qquad \text{Var}_{f_X}[X] = \frac{2-\theta}{\theta^2} - \frac{1}{\theta^2} = \frac{1-\theta}{\theta^2}$$

Notes

(1) If $X \sim Geometric(\theta)$, then for $x, j \geq 1$,

$$P[X = x+j | X > j] = \frac{P[X = x+j, X > j]}{P[X > j]} = \frac{P[X = x+j]}{P[X > j]} = \frac{(1-\theta)^{x+j-1}\theta}{(1-\theta)^j} = (1-\theta)^{x-1}\theta = P[X = x]$$

So $P[X = x+j | X > j] = P[X = x]$. This property is unique (among **discrete** distributions) to the geometric distribution, and is called the **lack of memory** property.

(2) The geometric distribution is used to model the number, X, of independent, identical Bernoulli trials until the first success is obtained. It is a discrete **waiting time** distribution.

(3) The mass function of X can be re-written

$$P[X = x] = \exp\left\{t(x)q(\theta) + \log h(x) + \log c(\theta)\right\}$$

where $t(x) = (x-1)$, $q(\theta) = \log(1-\theta)$, $h(x) = 1$, $c(\theta) = \theta$.

3.6 NEGATIVE BINOMIAL DISTRIBUTION

NOTATION $X \sim NeBi(n,\theta)$ **RANGE** $\mathbb{X} = \{n, n+1, n+2, ...\}$

MASS FUNCTION

$$f_X(x) = \binom{x-1}{n-1} \theta^n (1-\theta)^{x-n} \qquad \text{where } n \in \{1,2,3,...\}\,,\ 0 \le \theta \le 1.$$

MGF

$$M_X(t) = \sum_{x=n}^{\infty} \mathrm{e}^{tx} \binom{x-1}{n-1} \theta^n (1-\theta)^{x-n} = \left(\theta \mathrm{e}^t\right)^n \sum_{x=n}^{\infty} \binom{x-1}{n-1} \left(\mathrm{e}^t(1-\theta)\right)^{x-n} = \left\{\frac{\theta \mathrm{e}^t}{1-\mathrm{e}^t(1-\theta)}\right\}^n$$

rth MOMENT

No simple general expression for $M_X^{(r)}(t)$, but

$$M_X^{(1)}(t) = \frac{n(\theta \mathrm{e}^t)^n}{\left[1-\mathrm{e}^t(1-\theta)\right]^{n+1}} \qquad M_X^{(2)}(t) = \frac{n(\theta \mathrm{e}^t)^n \left[n + \mathrm{e}^t(1-\theta)\right]}{\left[1-\mathrm{e}^t(1-\theta)\right]^{n+2}}$$

so that $M_X^{(1)}(0) = \dfrac{n}{\theta}$ and $M_X^{(2)}(0) = \dfrac{n(n+(1-\theta))}{\theta^2}$, and thus

$$\mathrm{E}_{f_X}[X] = \frac{n}{\theta} \qquad \mathrm{Var}_{f_X}[X] = \frac{n(n+(1-\theta))}{\theta^2} - \frac{n^2}{\theta^2} = \frac{n(1-\theta)}{\theta^2}$$

Notes

(1) If $X \sim Bin(n,\theta)$, $Y \sim NeBi(r,\theta)$, then for $r \le n$, $\mathrm{P}[X \ge r] = \mathrm{P}[Y \le n]$.

(2) The Negative Binomial distribution is used to model the number, X, of independent, identical Bernoulli trials needed to obtain **exactly** n successes.

(3) Alternative representation: let Y be the number of **failures** in a sequence of independent, identical Bernoulli trials that contains **exactly** n successes. Then $Y = X - n$, and hence

$$f_Y(y) = \binom{n+y-1}{n-1} \theta^n (1-\theta)^y \qquad y \in \{0,1,...\}$$

(4) If $X_i \sim Geometric(\theta)$, for $i = 1,...n$, are i.i.d. random variables, and $Y = X_1 + ... + X_n$, then $Y \sim NeBi(n,\theta)$ (result immediately follows using mgfs).

(5) If $X \sim NeBi(n,\theta)$, let $n(1-\theta) = \lambda$ and $Y = X - n$. Then

$$M_Y(t) = e^{-nt} M_X(t) = \left\{\frac{\theta}{1-\mathrm{e}^t(1-\theta)}\right\}^n = \left\{1 + \lambda \frac{(\mathrm{e}^t-1)}{n - \lambda \mathrm{e}^t}\right\}^n \longrightarrow \exp\left\{\lambda(\mathrm{e}^t-1)\right\}$$

as $n \longrightarrow \infty$, hence the alternate form of the negative binomial distribution tends to the Poisson distribution as $n \longrightarrow \infty$ with $n(1-\theta) = \lambda$ constant.

(6) The mass function of X can be re-written

$$\mathrm{P}[X = x] = \exp\left\{t(x)q(\theta) + \log h(x) + \log c(\theta)\right\}$$

where $t(x) = x$, $q(\theta) = \log(\theta/(1-\theta))$, $h(x) = \binom{x-1}{n-1}$, $c(\theta) = (1-\theta)^n$.

3.7 HYPERGEOMETRIC DISTRIBUTION

NOTATION $X \sim \text{HypGeom}(N, R, n)$ for $N \geq R \geq n$

RANGE $\mathbb{X} = \{\text{Max}(0, n-N+R), ..., \text{Min}(n, R)\}$

MASS FUNCTION

$$f_X(x) = \frac{\binom{N-n}{R-x}\binom{n}{x}}{\binom{N}{R}} = \frac{\binom{N-R}{n-x}\binom{R}{x}}{\binom{N}{n}}$$

for $x \in \mathbb{X}$, and zero otherwise.

Notes

(1) The hypergeometric distribution is used as a model for experiments involving sampling *without* replacement from a *finite* population. The mass function for the hypergeometric distribution can be obtained by using combinatorics/counting techniques. However the form of the mass function does not lend itself readily to calculation of moments etc..

Consider obtaining the sample of size n by drawing sequentially, and let X_i for $i = 1, ..., n$ represent the number of Type I objects obtained on the ith draw (so that $X_i = 0$ or 1). Then $X_1, ..., X_n$ are *dependent* Bernoulli random variables, and

$$X_1 \sim \text{Bernoulli}(R/N), \quad X_2|X_1 = x_1 \sim \text{Bernoulli}((R-x_1)/(N-1)), ...$$

Using the successive conditioning, and general results for the expectation and variance, it can be shown that

$$\text{E}_{f_X}[X] = n\frac{R}{N}$$

$$\text{Var}_{f_X}[X] = n\frac{R}{N}\left(1-\frac{R}{N}\right)\left(\frac{N-n}{N-1}\right)$$

which are the expectation and variance for a hypergeometric distribution.

(2) As $N, R \longrightarrow \infty$ with $R/N = \theta$(constant), then

$$\text{P}[X = x] \longrightarrow \binom{n}{x}\theta^x(1-\theta)^{n-x},$$

so the distribution tends to a Binomial distribution.