CHAPTER 3

DISCRETE PROBABILITY DISTRIBUTIONS

3.1 DISCRETE UNIFORM DISTRIBUTION

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

MASS FUNCTION

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

CDF

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

MGF

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

rth MOMENT

$$\begin{aligned}
M_X^{(r)}(t) &= \sum_{x=1}^{n} x^r e^{tx}\frac{1}{n} \Longrightarrow M_X^{(r)}(0) = \frac{1}{n}\sum_{x=1}^{n} x^r \\
\Longrightarrow E_{f_X}[X] &= M_X^{(1)}(0) = \frac{1}{n}\sum_{x=1}^{n} x = \frac{(n+1)}{2} \\
E_{f_X}[X^2] &= M_X^{(2)}(0) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6} \\
\Longrightarrow Var_{f_X}[X] &= E_{f_X}[X^2] - \{E_{f_X}[X]\}^2 = \frac{(n+1)(n-1)}{12}
\end{aligned}$$

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 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

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

MGF

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

rth MOMENT

$$M_X^{(r)}(t) = \theta e^t \Longrightarrow M_X^{(r)}(0) = \theta r \Longrightarrow E_{f_X}[X] = \theta, E_{f_X}[X^2] = \theta \therefore Var_{f_X}[X] = \theta - \theta^2 = \theta(1-\theta)$$

NOTE 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.

3.3 BINOMIAL DISTRIBUTION

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

MASS FUNCTION

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

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

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

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

$$E_{f_X}[X] = n\theta Var_{f_X}[X] = n(n-1)\theta^2 + n\theta - n^2\theta^2 = n\theta(1-\theta)$$

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,

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

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, $X \sim Bin(n, \theta)$.

3.4 POISSON DISTRIBUTION

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

MASS FUNCTION

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

MGF

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

***r*th MOMENT**

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

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

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

$$E_{f_X}[X] = \lambda_{f_X}[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

NOTES

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

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

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,

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

so that $Y \sim Poisson(\lambda_1 + \lambda_2)$. Therefore, the sum of two independent Poisson random variables also has a Poisson distribution. This result can be extended easily; if $X_1, ..., X_k$ are independent random variables with $X_i \sim Poisson(\lambda_i)$ for $i = 1, ..., k$, then

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

3.4.1 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,

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

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

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

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

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,

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

or 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

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

it can be shown that

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

(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.

3.5 GEOMETRIC DISTRIBUTION

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

MASS FUNCTION

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

CDF

$$F_X(x) = 1 - (1-\theta)^x \qquad x = 1, 2, ...$$

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

$$e_{f_X}[X] = \frac{1}{\theta} \qquad Var_{f_X}[X] = \frac{2-\theta}{\theta^2} - \frac{1}{\theta^2} = \frac{1-\theta}{\theta^2}$$

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) Alternative representations:

$$f_X(x) = \phi^{x-1}(1-\phi) \qquad x = 1, 2, 3... \ \text{(that is, } \phi = 1-\theta)$$

$$f_X(x) = \phi^x(1-\phi) \qquad x = 0, 1, 2, .$$

(3) 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.

3.6 NEGATIVE BINOMIAL DISTRIBUTION

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

MASS FUNCTION

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

MGF

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

rth MOMENT

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

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

so that

$$M_X^{(1)}(0) = \frac{n}{\theta} \qquad \text{and} \qquad M_X^{(2)}(0) = \frac{n(n+(1-\theta))}{\theta^2}$$

and thus

$$E_{f_X}[X] = \frac{n}{\theta} \qquad Var_{f_X}[X] = \frac{n(n+(1-\theta))}{\theta^2} - \frac{n^2}{\theta^2} = \frac{n(1-\theta)}{\theta^2}$$

NOTES

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

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

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.

3.7 HYPERGEOMETRIC DISTRIBUTION

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

RANGE $\mathbb{X} = \{\max(0, n - N + R), ..., \min(n, R)\}$

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 $x \in X$, and zero otherwise.

NOTE

(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

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

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

$$E_{f_X}[X] = n\tfrac{R}{N}$$

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

which are the expectation and variance for a hypergeometric distribution.

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

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

so the distribution tends to a Binomial distribution.

CHAPTER 4

CONTINUOUS PROBABILITY DISTRIBUTIONS

4.1 CONTINUOUS UNIFORM DISTRIBUTION

NOTATION $X \sim Uniform(a, b)$ **RANGE** $\mathbb{X} = [a, b]$ or (a, b), for $a \leq b$

PDF

$$f_X(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

CDF

$$F_X(x) = \frac{x-a}{b-a} \qquad a \leq x \leq b$$

MGF

$$M_X(t) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{(e^{tb} - e^{ta})}{t(b-a)}$$

rth MOMENT

$$E_{f_X}[X^r] = \int_a^b x^r \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{b^{r+1}}{r+1} - \frac{a^{r+1}}{r+1} \right]$$

so therefore

$$\left. \begin{aligned} E_{f_X}[X] = & \frac{1}{b-a} \left[\frac{b^2 - a^2}{2} \right] = \frac{(a+b)}{2} \\ E_{f_X}[X^2] = & \frac{1}{b-a} \left[\frac{b^3 - a^3}{3} \right] = \frac{(a^2 + ab + b^2)}{3} \end{aligned} \right\} \Longrightarrow Var_{f_X}[X] = \frac{(b-a)^2}{12}$$

4.2 EXPONENTIAL DISTRIBUTION

NOTATION $X \sim Exp(\lambda)$ **RANGE** $\mathbb{X} = \mathbb{R}^+$

PDF

$$f_X(x) = \lambda e^{-\lambda x} \qquad x > 0 \qquad \lambda > 0.$$

CDF

$$F_X(x) = 1 - e^{-\lambda x} \qquad x > 0$$

MGF

$$M_X(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda \int_0^\infty e^{-(\lambda - t)x}dx = \frac{\lambda}{\lambda - t} \text{ for } t < \lambda$$

***r*th MOMENT**

$$M_X^{(r)}(t) = \frac{r!\lambda}{(\lambda - t)^{r+1}} \Longrightarrow M_X^{(r)}(0) = \frac{r!}{\lambda^r} \therefore E_{f_X}[X] = \frac{1}{\lambda}, E_{f_X}[X^2] = \frac{2}{\lambda^2} \Longrightarrow Var_{f_X}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

NOTES

(1) Alternative representation uses $\theta = 1/\lambda$ as the parameter of the distribution.

(2) If $X \sim Exp(\lambda)$, then, for all $x, t > 0$,

$$P[X > x + t|X > t] = \frac{P[X > x + t, X > t]}{P[X > t]} = \frac{P[X > x + t]}{P[X > t]} = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = e^{-\lambda x} = P[X > x]$$

Thus, for all $x, t > 0$, P$[X > x + t|X > t]$ =P$[X > x]$ - this is known as the Lack of Memory Property, and is unique to the exponential distribution amongst continuous distributions.

(3) Suppose that $X(t)$ is a Poisson process with rate parameter $\lambda > 0$, so that

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

.

Let $X_1, ..., X_n$ be random variables defined by $X_1 =$ "time that first event occurs", and, for $i = 2, ..., n$, $X_i =$ "time interval between occurrence of $(i-1)$st and ith events". Then $X_1, ..., X_n$ are i.i.d. $Exp(\lambda)$.

Proof : $X_1, ..., X_n$ are i.i.d. because of the assumption 2. underlying the Poisson process. So consider the distribution of X_1; in particular, consider the probability P$[X_1 > x]$ for $x > 0$. The event $[X_1 > x]$ is equivalent to the event "No events occur in the interval (0,x]", which has probability $e^{-\lambda x}$. But

$$F_{X_1}(x) = P[X_1 \leq x] = 1 - P[X_1 > x] = 1 - e^{-\lambda x} \Longrightarrow X_1 \sim Exp(\lambda)$$

.

(4) The exponential distribution is used to model failure times in continuous time. It is a continuous waiting time distribution, the continuous analogue of the geometric distribution.

(5) If $X \sim Uniform(0, 1)$, and $Y = -\log(1 - X)/\lambda$, then $Y \sim Exp(\lambda)$.

(6) If $X \sim Exp(\lambda)$, then $Y = X^{1/\alpha}$ for $\alpha > 0$ has a (two-parameter) **Weibull** distribution, and

$$f_Y(y) = \alpha\lambda y^{\alpha-1}e^{-\lambda y^\alpha} \qquad y > 0$$

4.3 GAMMA DISTRIBUTION

NOTATION $X \sim Ga(\alpha, \beta)$ **RANGE** $\mathbb{X} = \mathbb{R}^+$

PDF

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} x > 0 \qquad \alpha, \beta > 0.$$

and where, for any real number $\alpha > 0$, the **Gamma function**, $\Gamma(.)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

MGF

$$M_X(t) = \int_0^\infty e^{tx} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} = \left(\frac{\beta}{\beta-t}\right)^\alpha$$

rth MOMENT

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

$$M_X^{(1)}(t) = \frac{\alpha\beta^\alpha}{(\beta-t)^{\alpha+1}} \qquad M_X^{(2)}(t) = \frac{\alpha(\alpha+1)\beta^\alpha}{(\beta-t)^{\alpha+2}}$$

so that $M_X^{(1)}(0) = \frac{\alpha}{\beta}$ and $M_X^{(2)}(0) = \dfrac{\alpha(\alpha+1)}{\beta^2}$, and thus

$$E_{f_X}[X] = \frac{\alpha}{\beta} \qquad Var_{f_X}[X] = \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$$

NOTES

(1) If $X_1 \sim Ga(\alpha_1, \beta), X_2 \sim Ga(\alpha_2, \beta)$ are independent random variables, and $Y = X_1 + X_2$, then $Y \sim Ga(\alpha_1 + \alpha_2, \beta)$ (directly from properties of mgfs).

(2) $Ga(1, \beta) \equiv Exp(\beta)$.

(3) If $X_1, ..., X_n \sim Exp(\lambda)$ are independent random variables, and $Y = X_1 + ... + X_n$, then $Y \sim Ga(n, \lambda)$ (directly from (1) and (2)).

(4) For $\alpha > 0$,

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt = \left[-t^{\alpha-1} e^{-t}\right]_0^\infty + \int_0^\infty (\alpha-1) t^{\alpha-2} e^{-t} dt = (\alpha-1) \int_0^\infty t^{\alpha-2} e^{-t} dt = (\alpha-1)\Gamma(\alpha-1)$$

so $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$. Thus if $\alpha = 1, 2, ...$, then $\Gamma(\alpha) = (\alpha-1)!$.

(5) Special Case : If $\alpha = 1, 2, ...$ the $Ga(\alpha/2, 1/2)$ distribution is also known as the Chi-squared distribution with α degrees of freedom

(6) If $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent Chi-squared random variables with n_1 and n_2 degrees of freedom respectively, then random variable F defined as the ratio

$$F = \frac{(X_1/n_1)}{(X_2/n_2)}$$

has an **F-distribution** with (n_1, n_2) degrees of freedom.

(7) For events in a Poisson process with rate λ, then if $X(t)$ is the random variable counting the number of events that occur in the interval $[0, t)$, then

$$X(t) \sim Poisson(\lambda t) \qquad \mathrm{P}\left[X(t) = n\right] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \qquad n = 0, 1, 2, ...$$

Now consider the random variable Y_n that corresponds to the time at which the nth event occurs. To compute the distribution of Y_n consider first the cdf

$$F_{Y_n}(t) = \mathrm{P}\left[Y_n \leq t\right] = 1 - \mathrm{P}\left[Y_n > t\right]$$

But

$$Y_n > t \iff X(t) < n \iff X(t) \leq n-1$$

and so

$$F_{Y_n}(t) = 1 - \mathrm{P}\left[Y_n > t\right] = 1 - \mathrm{P}\left[X(t) \leq n-1\right] = 1 - \sum_{k=0}^{n-1} \mathrm{P}\left[X(t) = k\right] = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{1}$$

Thus, by differentiation, for $t > 0$

$$\begin{aligned} f_{Y_n}(t) &= \frac{d}{ds}\left\{1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda s}(\lambda s)^k}{k!}\right\}_{s=t} = -\sum_{k=0}^{n-1} \frac{\lambda^k}{k!}\left[-\lambda e^{-\lambda s} s^k + k e^{-\lambda s} s^{k-1}\right]_{s=t} \\ &= \sum_{k=0}^{n-1} \frac{\lambda^{k+1}}{k!} e^{-\lambda t} t^k - \sum_{k=1}^{n-1} \frac{\lambda^k}{(k-1)!} e^{-\lambda t} t^{k-1} = \lambda e^{-\lambda t}\left[\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - \sum_{k=1}^{n-1} \frac{(\lambda t)^{k-1}}{(k-1)!}\right] = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

as all other terms cancel. Hence, as $(n-1)! = \Gamma(n)$

$$f_{Y_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t} \qquad t > 0$$

and hence

$$Y_n \sim Gamma(n, \lambda)$$

Note that (1) gives a way of computing the Gamma cdf.

4.4 BETA DISTRIBUTION

NOTATION $X \sim Be(\alpha, \beta)$ **RANGE** $\mathbb{X} = (0, 1)$

PDF

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1 \qquad \alpha, \beta > 0.$$

rth MOMENT

For $r = 1, 2, ...,$

$$\begin{aligned} E_{f_X}[X^r] &= \int_0^1 x^r \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{r+\alpha-1}(1-x)^{\beta-1} dx \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(r+\alpha)\Gamma(\beta)}{\Gamma(r+\alpha+\beta)} \end{aligned}$$

$$\Longrightarrow E_{f_X}[X] = \frac{\alpha}{\alpha+\beta} \qquad E_{f_X}[X^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$\Longrightarrow Var_{f_X}[X] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

NOTES

(1) The beta distribution arises naturally in the context of order statistics; if $X_1, ..., X_k$ are i.i.d. random variables with cdf F_X, say, consider first the random variables $U_1, ..., U_k$ defined by $U_i = F_X(X_i)$ for $i = 1, ..., k$. It can be shown that $U_1, ..., U_k$ are i.i.d. $Uniform(0, 1)$ random variables. Now, consider the order statistics $Y_1, ..., Y_k$ derived from $U_1, ..., U_k$; using previous results, it can be shown that the marginal distribution of the jth order statistic is $Be(j, k - j + 1)$, for $j = 1, ..., k$.

(2) If $X_1 \sim Ga(\alpha_1, \beta)$, $X_2 \sim Ga(\alpha_2, \beta)$ are independent random variables, and $Y = X_1/(X_1 + X_2)$, then $Y \sim Be(\alpha_1, \alpha_2)$ (using standard multivariate transformation techniques).

(3) Suppose that random variables X and Y have a joint probability distribution such that the conditional distribution of X, given $Y = y$ for $0 < y < 1$, is binomial, $Bin(n, y)$, and the marginal distribution of Y is beta, $Be(\alpha, \beta)$, so that

$$f_{X|Y}(x|y) = \binom{n}{x} y^x (1-y)^{n-x} x = 0, 1, ..., n f_Y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} 0 < y < 1.$$

Then the marginal distribution of X is given by

$$f_X(x) = \int_0^1 f_{X|Y}(x|y) f_Y(y) dy = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)} x = 0, 1, 2...n$$

(3) If $\alpha = \beta = 1$, $Be(\alpha, \beta) \equiv Uniform(0, 1)$.

4.5 NORMAL DISTRIBUTION

NOTATION $X \sim N(\mu, \sigma^2)$ **RANGE** $\mathbb{X} = \mathbb{R}$

PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad x \in \mathbb{R} \qquad \mu \in \mathbb{R}, \sigma > 0.$$

MGF

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2 + tx\right\} dx \\ &= \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x - (\mu + t\sigma^2)\right)^2\right\} dx \\ &= \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \end{aligned}$$

because the integrand is a pdf, and thus the integral is equal to one.

***r*th MOMENT**

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

$$\begin{aligned} M_X^{(1)}(t) &= (\mu + t\sigma^2) \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \\ M_X^{(2)}(t) &= (\mu^2 + 2t\sigma^2\mu + t^2\sigma^4 + \sigma^2) \exp\left\{\mu t + \frac{t^2\sigma^2}{2}\right\} \end{aligned}$$

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

$$E_{f_X}[X] = \mu \qquad Var_{f_X}[X] = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

NOTES

(1) Special Case : If $\mu = 0$, $\sigma^2 = 1$, then X has a Standard or Unit normal distribution. Usually, the pdf of the unit normal is written $\phi(x)$, and the cdf is written $\Phi(x)$.

(2) If $X \sim N(0,1)$, and $Y = \sigma X + \mu$, then $Y \sim N(\mu, \sigma^2)$. Re-expressing this result, if $X \sim N(\mu, \sigma^2)$, and $Y = (X - \mu)/\sigma$, then $Y \sim N(0,1)$. (using transformation or mgf techniques)

(3) **The Central Limit Theorem** Suppose $X_1, ..., X_n$ are i.i.d. random variables with mgf M_X, with $\mathrm{E}_{f_X}[X_i] = \mu$ and $Var_{f_X}[X_i] = \sigma^2$ that is, the mgf and the expectation and variance of the X_is are specified, but the pdf is not. Let the random variable Z_n be defined by

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

and let Z_n have mgf M_{Z_n}. Then, as $n \longrightarrow \infty$,

$$M_{Z_n}(t) \longrightarrow \exp\left\{t^2/2\right\}$$

irrespective of the distribution of the X_is, that is, the distribution of Z_n tends to a **unit normal distribution** as n tends to infinity. This theorem will be proved and explained in Chapter 5, section 5.2.

This results provides a useful means of approximation. For any random variable S, say, where

$$S = \sum_{i=1}^{n} X_i$$

for independent and identically distributed random variables $X_1, ..., X_n$, the cdf of S can be approximated as follows: define

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{S - n\mu}{\sqrt{n\sigma^2}}$$

then by the theorem, and using a univariate transformation

$$F_Z(z) \approx \Phi(z) \Longrightarrow F_S(s) \approx \Phi\left(\frac{s - n\mu}{\sqrt{n\sigma^2}}\right)$$

(4) If $X \sim N(0,1)$, and $Y = X^2$, then $Y \sim \chi_1^2$, so that the square of a unit normal random variable has a chi-squared distribution with 1 degree of freedom.

(5) If $X \sim N(0,1)$, and $Y \sim N(0,1)$ are independent random variables, and Z is defined by $Z = X/Y$, the Z has a **Cauchy distribution**

$$f_Z(z) = \frac{1}{\pi}\frac{1}{1+z^2} \qquad z \in \mathbb{R}$$

(6) If $X \sim N(0,1)$, and $Y \sim Ga(n/2, 1/2)$ for $n = 1, 2, ...$ (so that $Y \sim \chi_n^2$), are independent random variables, and T is defined by

$$T = \frac{X}{\sqrt{Y/n}}$$

then T has a **Student-t distribution with n degrees of freedom**, $T \sim St(n)$,

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{n\pi}\right)^{1/2}\left\{1 + \frac{t^2}{n}\right\}^{-(n+1)/2} \qquad t \in \mathbb{R}$$

Taking limiting cases of the Student-t distribution

$$n \longrightarrow \infty : St(n) \longrightarrow N(0,1) \qquad\qquad n \longrightarrow 1 : St(n) \longrightarrow Cauchy$$

4.6 MULTIVARIATE PROBABILITY DISTRIBUTIONS

4.6.1 THE MULTINOMIAL DISTRIBUTION

The multinomial distribution is a multivariate generalization of the binomial distribution. Recall that the binomial distribution arose from an infinite Urn model with two types of objects being sampled without replacement. Suppose that the proportion of "Type 1" objects in the urn is θ (so $0 \leq \theta \leq 1$) and hence the proportion of "Type 2" objects in the urn is $1-\theta$. Suppose that n objects are sampled, and X is the random variable corresponding to the number of "Type 1" objects in the sample. Then $X \sim Bin(n, \theta)$, and

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

Now consider a generalization; suppose that the Urn contains $k+1$ types of objects ($k = 1, 2, ...$), with θ_i being the proportion of Type i objects, for $i = 1, ..., k+1$. Let X_i be the random variable corresponding to the number of type i objects in a sample of size n, for $i = 1, ..., k$. Then the joint distribution of vector $X = (X_1, ..., X_k)$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{n!}{x_1!...x_k!x_{k+1}!} \theta_1^{x_1} ....\theta_k^{x_k} \theta_{k+1}^{x_{k+1}} = \frac{n!}{x_1!...x_k!x_{k+1}!} \prod_{i=1}^{k+1} \theta_i^{x_i}$$

where $0 \leq \theta_i \leq 1$ for all i, and $\theta_1 + ... + \theta_k + \theta_{k+1} = 1$, and where x_{k+1} is defined by $x_{k+1} = n - (x_1 + ... + x_k)$. This is the mass function for the multinomial distribution which reduces to the binomial if $k = 1$. It can also be shown that the marginal distribution of X_i is $\text{Bin}(n, \theta_i)$.

EXAMPLE A dice is rolled n times; let $X_i =$"total number of i scores". Then $X = (X_1, ..., X_5)$ has a multinomial distribution with $\theta_i = 1/6$ for $i = 1, ..., 6$.

4.6.2 THE DIRICHLET DISTRIBUTION

The Dirichlet distribution is a multivariate generalization of the beta distribution. Recall that the beta distribution arose as follows; suppose that V_1 and V_2 are independent Gamma random variables with $V_1 \sim Ga(\alpha_1, \beta), V_2 \sim Ga(\alpha_2, \beta)$. Then if X is defined by

$$X = \frac{V_1}{V_1 + V_2}$$

we have that $X \sim Be(\alpha_1, \alpha_2)$, and

$$f_X(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1} (1-x)^{\alpha_2 - 1} 0 < x < 1$$

Now consider a generalization; suppose that $V_1, ..., V_{k+1}$ are independent Gamma random variables with $V_i \sim Ga(\alpha_i, \beta)$, for $i = 1, ..., k+1$. Define

$$X_i = \frac{V_i}{V_1 + ... + V_{k+1}}$$

for $i = 1, ..., k$. Then the joint distribution of vector $X = (X_1, ..., X_k)$ is given by

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_1)...\Gamma(\alpha_k)\Gamma(\alpha_{k+1})} x_1^{\alpha_1 - 1} ... x_k^{\alpha_k - 1} {x_{k+1}}^{\alpha_{k+1} - 1}$$

for $0 \leq x_i \leq 1$ for all i such that $x_1 + ... + x_k + x_{k+1} = 1$, where $\alpha = \alpha_1 + ... + \alpha_{k+1}$ and where x_{k+1} is defined by $x_{k+1} = 1 - (x_1 + ... + x_k)$. This is the density function which reduces to the beta distribution if $k = 1$. It can also be shown that the marginal distribution of X_i is $Beta(\alpha_i, \alpha)$.

EXAMPLE The composition of a mineral sample is determined in terms of percentage composition of five compounds. Let X_i be the percentage content of compound i, for $i = 1, ...4$. Then $X = (X_1, ..., X_4)$ is a vector random variable whose joint probability structure could be described using a Dirichlet distribution.

4.6.3 THE MULTIVARIATE NORMAL DISTRIBUTION

The multivariate normal distribution is a multivariate generalization of the normal distribution which can be generated in the following way. Suppose that $X_1, ..., X_k$ are i.i.d. $N(0, \sigma^2)$ random variables. Using vector notation, we can write the joint density function of $X_1, ..., X_k$ as

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \left(\frac{1}{2\pi\sigma^2}\right)^{k/2} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{x}\right\}$$

where $x = (x_1, ..., x_k)$. Now consider the multivariate transformation from $(X_1, ..., X_k)$ to $(Y_1, ..., Y_k)$ (that is, from X to Y) defined by $\mathbf{Y} = A^T\mathbf{X} + \boldsymbol{\mu}$, where A is a $k \times k$ invertible matrix of real numbers, and μ is a $k \times 1$ vector. This is a 1-1 transformation, so using the usual multivariate transformation formula, we can obtain the joint density function of $(Y_1, ..., Y_k)$ as

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) = \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right\}$$

where $\Sigma = \sigma^2 A^T A$. This is the pdf of the multivariate normal distribution. It can be shown that any marginal, joint marginal, or conditional distribution of a subset of $Y_1, ..., Y_k$ is normal or multivariate normal.