

P.D.F./C.D.F.

- If X is a R.V. with a finite countable set of possible outcomes, $\{x_1, x_2, \dots, x_j\}$, then the **discrete probability distribution** of X

$$f(x) \text{ or } p(x_i) = \begin{cases} P\{X = x_i\} & \text{if } x = x_i, \quad i = 1, 2, \dots \\ 0 & \text{if } x \neq x_i \end{cases}$$

and **D.F. or C.D.F.**

$$P\{X \leq x_i\} = F(x_i) = \sum_{j: x_j \leq x_i} P\{X = x_j\}$$

- While, similarly, for X a R.V. taking any value **along an interval** of the **real number line**

$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(u) du$$

So if first derivative $F'(x)$ exists, then

$$F'(x) = dF(x)/dx = f(x)$$

$f(x) = F'(x)$ is the **continuous pdf**, with $\int_{-\infty}^{\infty} f(x) dx = 1$

1

EXPECTATION/VARIANCE

Clearly, $E(X) = \begin{cases} \sum_{i \in S} x_i f(x_i) & \text{discrete} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{continuous} \end{cases}$

and $Var(X) = \begin{cases} \sum_{x \in S} [x - E(X)]^2 f(x) & \text{discrete} \\ \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx & \text{continuous} \end{cases}$

2

Moments and M.G.F.'s

- For a R.V. X , and any non-negative integer k , k th moment about the origin, defined as **expected value of X^k**
- Central Moments about Mean: first = 0, second = variance
- To obtain Moments, use Moment Generating Function
- If X has a p.d.f. $f(x)$, **mgf** is expected value of e^{tX}

For a **continuous** variable, then

$$mgf(X) = E\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tX} f(x) dx$$

For a **discrete** variable, then

$$mgf(X) = E\{e^{tX}\} = \sum e^{tX} f(x)$$

- Generally:** r th moment of the R.V. is r th derivative evaluated at $t=0$

3

Examples- m.g.f.'s

- Suppose $Y = g(X) = e^{tX}$ for $X = be^{-bx}$ and $t > 0$

$$\text{Then } E[g(X)] = \int_0^{\infty} e^{tx} be^{-bx} dx = b \int_0^{\infty} e^{-(b-t)x} dx \quad \text{finite } t < b$$

$$\text{i.e. } E[g(X)] = \frac{b}{b-t} \quad \text{M.g.f. } M_X(t) \text{ for exponential } t < b$$

- Suppose p.d.f. $f(x) = \binom{n}{x} p^x q^{n-x}$ where $q = 1 - p$

$$M_X(t) = E[e^{tX}] = \sum_{x=0}^n \binom{n}{x} (pe^t)^x q^{n-x} = (pe^t + q)^n$$

4

PROPERTIES - Expectation/Variance etc. Distribution Functions

- As for R.V.'s generally. For X a **discrete** R.V. with p.d.f. $p\{X\}$, then for any real-valued function g

$$E\{g(X)\} = \sum g(x)p\{X\}$$

- e.g.

$$E\{X + Y\} = E\{X\} + E\{Y\}$$

$$E\{XY\} = E\{X\}E\{Y\}$$

Applies for more than 2 R.V.s also

- Variance - again has similar properties to previously:

$$\text{e.g. } V\{aX + b\} = a^2 V\{X\} = a^2 [E\{X^2\} - [E\{X\}]^2]$$

5

MENDEL's Example

- Let X record the no. of dominant A alleles in a randomly chosen genotype, then X is a R.V. with sample space $S = \{0, 1, 2\}$

- Outcomes in S correspond to events

$$X = \begin{cases} 0 & \text{if } aa \\ 1 & \text{if } aA, Aa \\ 2 & \text{if } AA \end{cases}$$

- Note:** Further, any **function** of X is also a R.V.

$$\text{e.g. } Z = g(X) = \begin{cases} 0 & \text{if } aa \quad (X = 0) \\ 1 & \text{if } AA, Aa, aA \quad (X > 0) \end{cases}$$

- Where Z is a variable for seed character phenotype

6

Example contd.

- So that, for Mendel's data, $Z = \begin{cases} 0 & \text{Round} \\ 1 & \text{Wrinkled} \end{cases}$
- and $f(z) = \begin{cases} P\{Z=0\} = 3/4 \\ P\{Z=1\} = 1/4 \end{cases}$ with $E(Z) = 1/4$
- and $Var(Z) = \sum_i [z_i - E(Z)]^2 f(z_i)$
 $= (0 - 1/4)^2 \times 3/4 + (1 - 1/4)^2 \times 1/4 = 9/16$

7

JOINT/MARGINAL DISTRIBUTIONS

- **Joint cumulative distribution** of X and Y , **marginal cumulative** for X , without regard to Y and **joint distribution (p.d.f.)** of X and Y then, respectively $F(x, y) = P\{X \leq x, Y \leq y\}$ (1)

$$F_x(x) = \sum_y P\{X \leq x, Y \leq y\} = F(x) \quad (2)$$

$$p(x, y) = P\{X = x, Y = y\} \quad (3)$$

$$\text{with } \sum_i \sum_j p(x_i, y_j) = 1$$

- where similarly for **continuous** case e.g. (2) becomes

$$F_1(x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, v) dv du = \int_{-\infty}^x f_1(u) du = F(x) \quad (2a)$$

8

CONDITIONAL DISTRIBUTIONS

- **Conditional distribution** of X , given that $Y=y$

$$p(x/y) = \frac{P(x, y)}{P(y)} = P\{X = x / Y = y\}$$

$$= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \quad \text{and similarly } p(y/x)$$

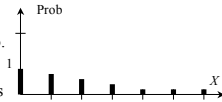
- where for X and Y **independent** $p(x/y) = p(x)$ and $p(y/x) = p(y)$

- **Example: Mendel's expt.** Probability that a round seed ($Z=1$) is a homozygote AA i.e. ($X=2$)

$$P\{X = 2 / Z = 1\} = \frac{P\{x = 2, z = 1\}}{P\{z = 1\}} = \frac{1/4}{3/4} = 1/3$$

9

Standard distributions: Geometric and Negative Binomial in brief

- **Geometric.** This arises in the "time" or No. of steps k to the first success in a series of independent Bernoulli trials. The density is 
 $Prob\{X = k\} = p(1-p)^{k-1} \quad (k = 1, 2, \dots)$
 $E[X] = 1/p$
 $VAR[X] = (1-p)/p^2$

- **Negative Binomial** This is used to model the number of failures k that occur before the r^{th} success in a series of independent Bernoulli trials. The density is

$$Prob\{X = k\} = \binom{r+k-1}{k} p^r (1-p)^k \quad (k = 0, 1, 2, \dots)$$

$$E[X] = r(1-p)/p \quad \text{Added Note: Alternative form - based directly on No. successes}$$

$$VAR[X] = r(1-p)/p^2 \quad \text{- see Tables}$$

10

Hypergeometric

- Consider a population of M items, of which W are deemed to be successes. Let X be the number of successes that occur in a sample of size n , drawn without replacement from the population. The density is

$$Prob\{X = k\} = \frac{W C_k^{M-W} C_{n-k}^W}{C_n^M} \quad (k = 0, 1, 2, \dots)$$

- Then $E[X] = n W / M$
 $VAR[X] = n W (M - W) (M - n) / \{M^2 (M - 1)\}$
- Sampling **without replacement** from a finite population

11

Standard p.d.f.'s: Gamma and Exponential in brief

- The **Gamma** distribution e.g. from queueing theory, -time to the arrival of the n^{th} customer in single-server queue, (mean arrival rate λ). P.d.f. written in terms of gamma function $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad x > 0$

$$\text{or directly } f(x) = \frac{\lambda(\lambda x)^{n-1} e^{-\lambda x}}{(n-1)!} \quad x \geq 0, n = 1, 2, \dots$$

$$= 0 \text{ otherwise} \quad \text{with } E[X] = n / \lambda \text{ and } VAR[X] = n / \lambda^2$$

- **Exponential** : special case of the Gamma distribution with $n = 1$ used e.g. to model inter-arrival time of customers, or time to arrival of first customer, in a simple queue, fragment lengths in genome mapping etc.

$$\text{The p.d.f. is } f(x) = \lambda \exp(-\lambda x), \quad x \geq 0, \lambda > 0$$

$$= 0 \text{ otherwise}$$

12

Standard p.d.f.'s : Student's t-distribution

- A random variable X has a t -distribution with n d.o.f. (t_n) if it has density

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} \quad -\infty < t < \infty$$

$$= 0 \quad \text{otherwise.}$$
- Symmetrical about the origin, with $E[X] = 0$; $VAR [X] = n / (n - 2)$.
- For small n , the t_n distribution is very flat. For $n \geq 25$, the t_n distribution \rightarrow standard normal curve.
- Suppose Z a standard Normal variable, W has a χ_n^2 distribution and Z and W independent then r.v.

$$X = \frac{Z}{\sqrt{W/n}}$$

- If x_1, x_2, \dots, x_n is a random sample from $N(\mu, \sigma^2)$, and, if define

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{then} \quad (\bar{x} - \mu) / \left(s / \sqrt{n} \right) \sim t_{n-1}$$

13

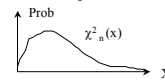
Chi-Square Distribution

- A r.v. X has a Chi-square distribution with n degrees of freedom; (n a positive integer) if it is a Gamma distribution with $\lambda = 1$, so its p.d.f. is

$$f(x) = x^{n-1} \exp(-x) / (n-1)! \quad x \geq 0$$

0 otherwise

$$E[X] = n; \quad Var [X] = 2n$$



- Two important applications:
 - If X_1, X_2, \dots, X_n a sequence of **independently** distributed Standardised Normal Random Variables, then the sum of squares $X_1^2 + X_2^2 + \dots + X_n^2$ has a χ^2 distribution (n degrees of freedom).

- If x_1, x_2, \dots, x_n is a random sample from $N(\mu, \sigma^2)$, then

$$\bar{x} = \sum_{i=1}^n x_i / n \quad \text{and} \quad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad \text{and}$$

s^2 has χ^2 distribution, $n - 1$ d.o.f., with r.v.'s \bar{x} and s^2 **independent**.

14

F-Distribution

- A r.v. X has an F distribution with m and n d.o.f. if it has a density function = ratio of gamma functions for $x > 0$ and = 0 otherwise.
- $$E[X] = n / (n - 2) \quad \text{if } n > 4$$
- $$Var[X] = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2} \quad \text{if } n > 4 \quad \text{and}$$
- For X and Y independent r.v.'s, $X \sim \chi_m^2$ and $Y \sim \chi_n^2$ then $F_{m,n} = \frac{X/m}{Y/n}$
- One consequence: if x_1, x_2, \dots, x_m ($m \geq 2$) is a random sample from $N(\mu_1, \sigma_1^2)$, and y_1, y_2, \dots, y_n ($n \geq 2$) a random sample from $N(\mu_2, \sigma_2^2)$, then

$$\frac{\sum (x_i - \bar{x})^2 / (m-1)}{\sum (y_i - \bar{y})^2 / (n-1)} \sim F_{m-1, n-1}$$

15