S1 p1

Cumulative frequency step polygon
Stem and leaf
Grouped frequency

class

5 - 9   9.5 ← boundary
7 ← mid
limit

Histograms + relative frequency
Median by interpolation p33
Quartiles p34

$$\mu = \bar{x} = \frac{\sum fx}{\sum f}$$

$$\sigma^2 = \frac{\sum (x-\mu)^2}{n} \quad \text{or} \quad \frac{\sum f(x-\mu)^2}{\sum f} = \frac{\sum fx^2}{\sum f} - \mu^2$$

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1} = \frac{1}{\sum f - 1}\left\{\sum fx^2 - \frac{\left(\sum fx\right)^2}{\sum f}\right\}$$

p 60

CODING $y = \frac{x-a}{b} \Rightarrow \bar{x} = b\bar{y} + a$

$s_x = b\,s_y$

Skew

+ve   -ve   symmetrical

mode < median < mean   mean < median < mode

"where the tail is"

Probability $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(M|B) = \frac{P(M \cap B)}{P(B)} \quad \text{or} \quad P(M \cap B) = P(M|B) \times P(B)$$

Independent $\Rightarrow P(A \cap B) = P(A) \times P(B)$

Mutually exclusive $\Rightarrow P(A \cap B) = 0$

Product-moment correlation coefficient (PMCC) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum (x_i - \bar{x})^2 \qquad S_{yy} = \sum (y_i - \bar{y})^2$$

OR

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$S_{xx} = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

CODING EFFECT !

Regression $\quad y = a + bx \quad b = \dfrac{S_{xy}}{S_{xx}} \quad a = \bar{y} - b\bar{x}$

+Residuals

Probability function of a discrete random variable $\quad P(X=x) \text{ or } p(x)$

$$\sum_{\forall x} P(X=x) = 1$$

Cumulative distribution function $F(x_0) = P(X \leq x_0)$

Discrete random variable $\Rightarrow F(x_0) = \sum_{x \leq x_0} P(X=x)$

Expected value $\mu = E(X) = \sum_{\forall x} x P(X=x)$

$$\sigma^2 = \mathrm{Var}(X) = E(X^2) - \mu^2$$

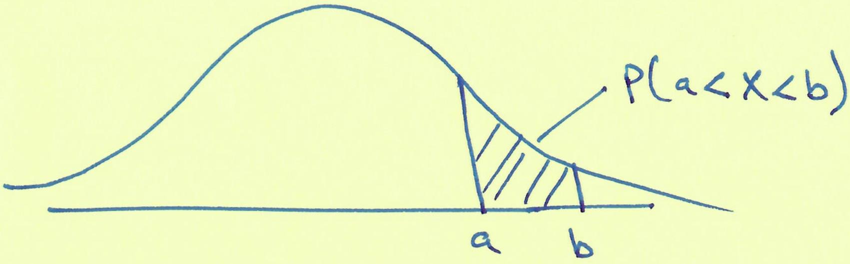
$$E(aX) = aE(X) \qquad E(aX+b) = aE(X) + b$$

$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X) \qquad \mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)$$

Discrete Uniform distribution p163

51 p3

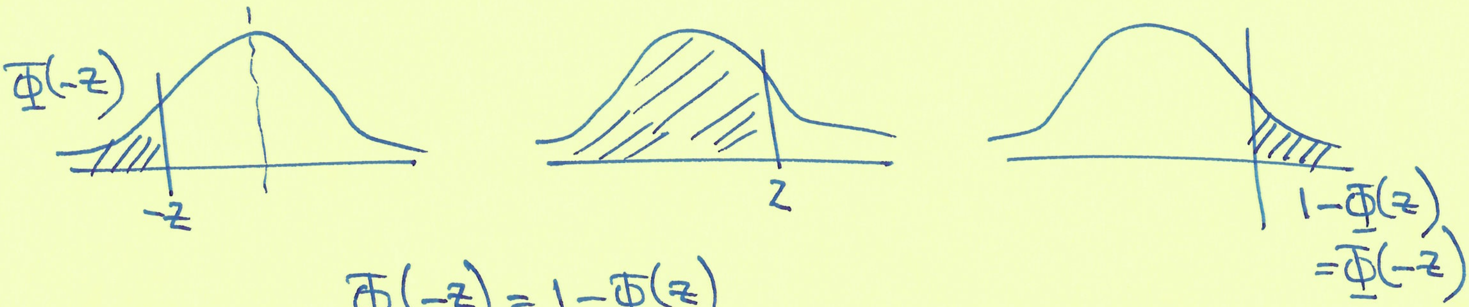
$X \sim N(\mu, \sigma^2)$



CONTINUOUS RANDOM VARIABLE

If $X \sim N(\mu, \sigma^2)$, and $Z = \dfrac{X-\mu}{\sigma}$ then $Z \sim N(0, 1^2)$

$\Phi(z_1) = P(Z < z_1)$



$\Phi(-z) = 1 - \Phi(z)$

Binomial Distribution

- fixed Number of trials
- "success" or "failure"
- independent

$P(X=r) = \binom{n}{r} p^r q^{n-r}$ $\quad q = 1-p$ $\qquad$ Careful using table p7

$X \sim B(n,p)$, then $E(X) = np$, $Var(X) = \sigma^2 = npq = np(1-p)$

Derivation p9

Poisson Distribution

events @ random
independently from each other
The average rate remains constant
zero probability of simultaneous occurrences

$X \sim Po(\lambda)$ $\qquad P(X=r) = \dfrac{e^{-\lambda}\lambda^r}{r!}$ $\quad$ where λ = mean number of occurrences in the "period", "length", "area"

$P(X=k+1) = \dfrac{\lambda}{k+1} \times P(X=k)$

e.g.
If $X \sim Po(\lambda)$ in one week, then $X \sim Po(2\lambda)$ in two weeks....

Careful using tables

$X \sim Po(\lambda)$, then $E(X) = \lambda$; $Var(X) = \lambda$, $\sigma = \sqrt{\lambda}$

Poisson likely if mean and variance very close

Poisson approximation to the Binomial distribution

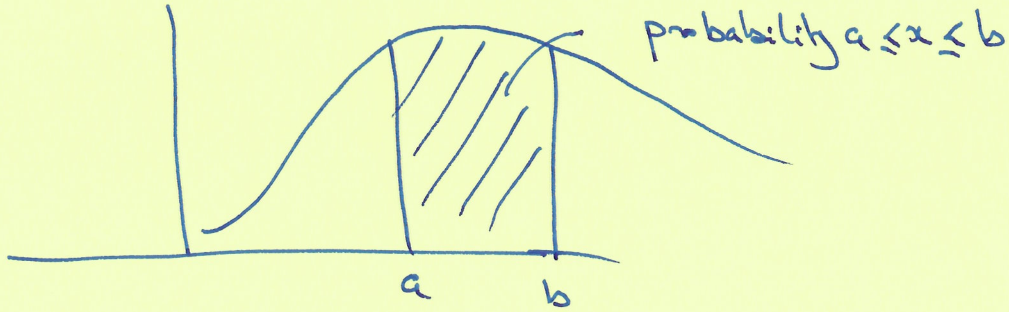
$X \sim B(n,p)$ large number of trials n
probability of "success" small $p <<$

then $X \sim$ approx $Po(\lambda)$ with $\lambda = np$

$\Longrightarrow$ np close to npq, a property of the Poisson distribution

CONTINUOUS RANDOM Variables

probability density function pdf $\quad f(x) \geqslant 0 \ \forall x \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$



Cumulative distribution function $F(x_0) = \int_{-\infty}^{x_0} f(x)\,dx$

c/f discrete random variable where $\mu = E(X) = \sum p x$

CONTINUOUS " " $\quad \mu = \int_{-\infty}^{\infty} x f(x)\,dx$ over definition of $f(x)$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx \qquad$ Variance $= \sigma^2 = E(X^2) - \mu^2$

MODE of a continuous random variable is where max value of $f(x)$

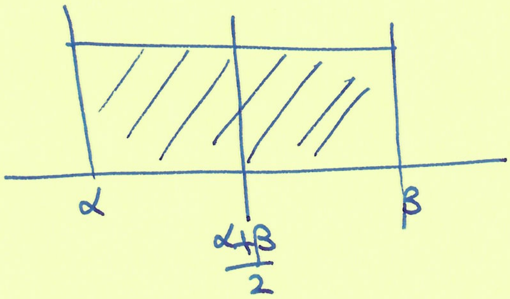
Median is when $F(m) = 0.5$

Q_1 lower quartile $F(Q_1) = 0.25$

Q_3 upper " $F(Q_3) = 0.75$

Sampling Statistics depend only on the sample..... *

CONTINUOUS Distributions – UNIFORM (rectangular) distribution



median = mean = $E(X) = \frac{\alpha+\beta}{2}$ p99!

$Var(X) = \frac{(\beta-\alpha)^2}{12}$

Approximating a discrete random variable, Y continuous random variable used to approximate X

* $P(X=x) \overset{approx}{\approx} P(x-0.5 < Y < x+0.5)$ *

$P(X \leq x) \approx^{approx} P(y \leq x+.5)$

e.g. if $X \sim B(n,p)$ $E(X) = np$ and $Var(X) = npq = np(1-p)$

Large n, p close to 0.5 (np and $n(1-p) > 5$)

approx by $N(np, npq)$

---

if $X \sim Po(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$

for large λ (>10) can use $N(\lambda, \lambda)$ to approx X.

Hypothesis Testing

identify null hypothesis H_0 $p =$ $\lambda =$

Alternative hypothesis H_1 $p <$ or $p >$ one-sided test

$p \neq$ two-sided test

Significance level

"true"

CONCLUSION, At the {significance level} not sufficient evidence to reject

there is " " to reject in favour of alternative

"false"

Critical Region

Reject H_0 if the observed value lies in the region