S2 Revision Notes

Binomial distribution

Key Words: fixed trials, large n small p, expectation, variance, standard deviation

- Use Binomials when: exact probability of an event happening in a fixed no. of trials, success or failure, \geq independent, constant probability between trials
- \geq $X \sim N(n, p)$

$$\circ \quad P(X=x) = \binom{n}{x} p^{x} (1-p)^{n-x}$$

- $E(X) = \mu = np$ 0
- $Var(X) = \sigma^2 = np(1-p)$ 0
- \geq

Poisson distribution

```
Key Words: time, expectation, variance, standard deviation, E(X) = Var(X)
```

- ۶
- ≻



Continuous random variables

Key Words: p.d.f., c.d.f, quartiles, percentiles

- \geq Probability density function (f(x) is the relative frequency of x in the given sample)
 - $\int_{-\infty}^{\infty} f(x) dx = 1$ 0
 - $P(a < x < b) = \int_{a}^{b} f(x) dx$ 0
 - $f(x) \ge 0$ for all x 0
- Cumulative distribution function (F(x) is the cumulative frequency of x for the same f(x)) \geq
 - $\circ \quad P(x \le a) = F(a) = \int_{-\infty}^{a} f(x) \, dx$

$$\circ \quad E(X) = \mu = \int_{-\infty}^{\infty} xf(x) \, dx$$

•
$$Var(X) = \sigma^2 = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \left(\int_{-\infty}^{\infty} x f(x) \, dx\right)^2$$

f (x)

1 b-a

b

а

x

- Lower Quartile (Q_1) : $F(Q_1) = 0.25$ 0
- *Median* (Q_2) : $F(Q_2) = 0.5$ 0
- Upper Quartile (Q_3) : $F(Q_3) = 0.75$ 0
- Percentile (P_n) : $F(P_n) = \frac{n}{100}$ 0
- Mode is the highest point on a p.d.f 0

Continuous uniform distribution

Key Words: p.d.f, c.d.f, expectation, variance



The same continuous random variable X has uniform c.d.f \geq



Normal Approximations

Key Words: discrete, continuous, continuity correction

When crossing the black line (going from discrete to continuous data) a continuity correction needs to be applied

Since: $P(X = 5) \rightarrow P(4.5 \le X \le 5.5)$

Key Words: population, samples, finite, infinite, sampling unit, frame, statistic

- Population: total collection of all individual items
- Sample: selection of individual items from a population
- Finite population: each items can be counted (given a number)
- Infinite population: impossible to give each item a number
- Sampling unit: individual items of a population
- Sampling frame: list of sampling units used in practice to represent a population
- Statistic: random variable calculated solely from the observations in a sample (no parameters used)

Hypothesis testing

Key Words:

 \geq A hypothesis test is a mathematical procedure to compare a null hypothesis to an alternative hypothesis within a given degree of accuracy

• *H*₀(null): "unchanged" hypothesis

- H₁(alternative): "alternative" / "new" 0 hypothesis
- The Critical region is the range of values that would ≻ lead vou to reject H₀



- One-tailed test: increase or decrease (one end of the \geq distribution)
- Two-tailed test: both increase and decrease (both ends of the distribution therefore half for each tail) ≻
- The significance level of a test is the actual probability of rejecting H_0 \triangleright

