Edexcel Further Maths A-level

Further Statistics 1

Formula Sheet

Provided in formula book	Not provided in formula book

Discrete Distributions

Discrete Random Variables

For a discrete random variable X taking values x_i with probabilities $\mathrm{P}(X = x_i)$:	
Expectation (mean)	$\mathrm{E}(X) = \mu = \Sigma\, x_i\, \mathrm{P}(X = x_i)$
Variance	$\mathrm{Var}(X) = \sigma^2 = \Sigma\,(x_i - \mu)^2 \mathrm{P}(X = x_i) = \mathrm{E}(X^2) - \left(\mathrm{E}(X)\right)^2$
For a function $\mathrm{g}(X)$:	$\mathrm{E}(\mathrm{g}(X)) = \Sigma\, \mathrm{g}(x_i)\, \mathrm{P}(X = x_i)$

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

Standard Discrete Distributions

Distribution	Binomial $B(n, p)$	Poisson $Po(\lambda)$	Geometric $Geo(p)$ on $1, 2, \ldots$	Negative binomial on $r, r+1, \ldots$
$\mathrm{P}(X = x)$	$\binom{n}{x} p^x (1-p)^x$	$e^{-\lambda} \frac{\lambda^x}{x!}$	$p(1-p)^{x-1}$	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$
$Mean$	np	λ	$\frac{1}{p}$	$\frac{r}{p}$
$Variance$	$np(1-p)$	λ	$\frac{1-p}{p^2}$	$\frac{r(1-p)}{p^2}$
$P.G.F$	$(1 - p + pt)^n$	$e^{\lambda(t-1)}$	$\frac{pt}{1-(1-p)t}$	$\left(\frac{pt}{1-(1-p)t}\right)^r$

Poisson Distribution

If two Poisson distributions X, Y are independent:	$X + Y \sim Po(\lambda_x + \lambda_y)$
If each observation of X is independent and $X \sim Po(\lambda)$:	$aX \sim Po(a\lambda)$

Binomial approximation	*If* $X \sim B(n, p)$ *and* n *is large and* p *close to* 0 *then* $X \approx \sim Po(np)$ *where* $\lambda = np$.

Geometric Distribution

Cumulative Distribution	$P(X \leq x) = 1 - (1-p)^x$
	$P(X \geq x) = (1-p)^{x-1}$

Probability Generating Functions

For a discrete random variable X:	$\mathrm{G}_X(t) = \mathrm{E}(t^x) = \Sigma \mathrm{P}(X = x)t^x$

For $Z = X + Y$, where X and Y are independent:	$\mathrm{G}_Z(\mathrm{t}) = \mathrm{G}_X(\mathrm{t}) \times \mathrm{G}_Y(\mathrm{t})$

$$\mathrm{E}(X) = \mathrm{G}'_X(1)$$

$$\mathrm{Var}(X) = \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - [\mathrm{G}'_X(1)]^2$$

$$\mathrm{G}_{aX+b}(t) = t^b \mathrm{G}_X(t^a)$$

Hypothesis Testing

Null hypothesis	$\boldsymbol{H_0 : \theta = m}$
One tailed test	$H_1 : \theta > m$ *or* $\theta < m$
Two-tailed test	$H_1 : \theta \neq m$

Central Limit Theorem

If $X_1, X_2, \dots, X_n$ is a random sample of size n from a population with mean μ and variance σ^2, then

$$X \approx \sim N(\mu, \frac{\sigma^2}{n}).$$

Chi Squared Tests

Measure of Goodness of Fit

$O_i = observed\ frequency$
$E_i = expected\ frequenct$
$N = number\ of\ trials$

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \sum \frac{O_i^2}{E_i} - N$$

The greater the value of X^2, the less good the fit.

Degrees of Freedom

$No. of\ degrees\ of\ freedom = No. of\ cells\ (after\ necessary\ combining) - No. of\ parameters$

Contingency Tables

$$Expected\ frequency = \frac{Row\ total\ \times Column\ total}{Grand\ total}$$

Number of degrees of freedom $\nu = (h-1)(k-1)$ for an $h \times k$ table

Quality of Tests

$Power = 1 - \mathrm{P}(Type\ II\ error) = \mathrm{P}(being\ in\ the\ critical\ region\ when\ H_0\ is\ false)$