4 Further Probability & Statistics – Part 2

Name: _____ Class: _____ Date: _____

Vocabulary

Write each word three times.

English	**Write it 3 times** (English)		
chi-squared test			
probability generating function			

Recall

In **Part 1** you met continuous random variables (the density $f(x)$, the cumulative $F(x)$, and the median). Part 2 adds a **test** that compares observed and expected counts, and a **function** that packs a whole distribution into one.

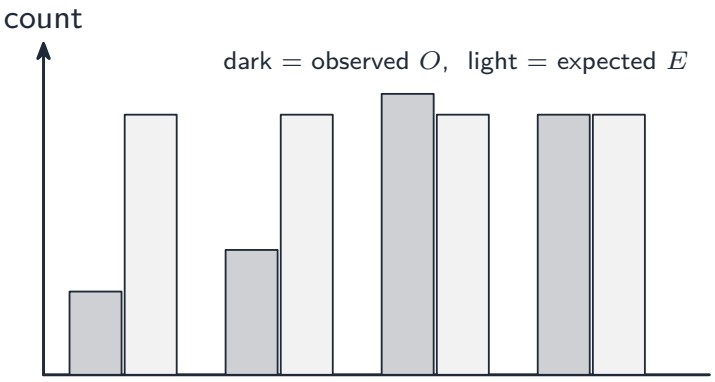
New ideas

■ 1 The chi-squared test

A **chi-squared test** compares **observed** counts O with the **expected** counts E from a model:

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

A **small** χ^2 means the data fits the model well; a **large** χ^2 (above the table value) means it fits badly.



Worked example. Observed $20, 30, 25, 25$ with each expected $= 25$: $\chi^2 = \dfrac{(20-25)^2 + (30-25)^2 + 0 + 0}{25} = \dfrac{50}{25} = 2.$

Micro-check. In a chi-squared test, does a **large** value mean the model fits well or badly? [1]

■ 2 Probability generating functions

The **probability generating function** of a discrete variable X is

$$G(t) = E(t^X) = \sum_x P(X = x)\, t^x.$$

It packs the whole distribution into one function. The mean comes from its derivative at $t = 1$: $E(X) = G'(1)$.

Worked example. If $P(X{=}0) = 0.5$, $P(X{=}1) = 0.3$, $P(X{=}2) = 0.2$, then $G(t) = 0.5 + 0.3t + 0.2t^2$, so $G'(t) = 0.3 + 0.4t$ and $E(X) = G'(1) = 0.7$.

Micro-check. The PGF gives the mean by $E(X) = G'(?)$ – state the value. [1]

Watch out: in χ^2, divide each $(O - E)^2$ by its **own** E before adding – not by the total. A **large** χ^2 means a **bad** fit. The PGF gives the mean from $G'(1)$, not $G(1)$ (which is always 1).

Practice

4.1 Write the formula for the chi-squared statistic χ^2. [1]

4.2 Observed counts are $10, 20, 30$ with expected counts $20, 20, 20$. Find χ^2. [2]

4.3 State what a **large** value of χ^2 tells you about the model. [1]

4.4 A variable has $G(t) = 0.4 + 0.4t + 0.2t^2$. Find $G'(t)$ and hence $E(X) = G'(1)$. [2]

4.5 State what the probability generating function $G(t)$ "packs" into one function. [1]

4.6 Challenge. A variable has $G(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$. Find $E(X)$. [3]