Assignment 2 Outline Solution

A2 – 1. This question involves a k = 3 variable multinomial distribution as a probability model.

(a) $\Pr(3G, 2R, 1U) = \binom{6}{3\ 2\ 1}(\tfrac{1}{2})^3(\tfrac{1}{3})^2(\tfrac{1}{6})^1 = \tfrac{5}{36} = 0.13\dot{8}.$

(b) Pr(at least one R and exactly one U) =

$$\Pr(4G, 1R, 1U) + \Pr(3G, 2R, 1U) + \Pr(2G, 3R, 1U) + \Pr(1G, 4R, 1U) + \Pr(0G, 5R, 1U)$$
$$= \tfrac{5}{48} + \tfrac{5}{36} + \tfrac{5}{54} + \tfrac{5}{162} + \tfrac{1}{243} = \tfrac{2,882}{7,776} \simeq 0.370\,628.$$

Alternatively: Pr(at least one R and exactly one U)

$$= \Pr(1U) - \Pr(5G, 0R, 1U) = (\tfrac{5}{6})^5 - (\tfrac{1}{2})^5 = \frac{5^5 - 3^5}{6^5} = \frac{2,882}{7,776} \quad \text{(as before).}$$

(c) Pr(at least one R) = 1 – Pr(no R) = $1 - (\tfrac{2}{3})^6 = \tfrac{665}{729} \simeq 0.912\,209.$

(d) To be sure of satisfying the conditions of the event whose probability is required here, we can list all 28 of the relevant possibilities for a total of 6 items, as follows:

G	R	U	G	R	U	G	R	U	G	R	U	G	R	U	G	R	U
6	0	0	5	1	0	4	2	0	4	1	1	3	2	1	3	3	0
0	6	0	1	5	0	2	4	0	1	4	1	2	3	1	3	0	3
0	0	6	0	5	1	0	4	2	1	1	4	1	3	2	0	3	3
			0	1	5	0	2	4				1	2	3			
			5	0	1	4	0	2				3	1	2	2	2	2
			1	0	5	2	0	4				2	1	3			

Hence: Pr(at least one item which is either R or U)

$$= \Pr[(\text{at least one } R) \cup (\text{at least one } U)] = 1 - \Pr(6G) = 1 - (\tfrac{1}{2})^6 = \tfrac{63}{64} = 0.984\,375.$$

A2 – 2. This question also involves the multinomial distribution as a probability model; there is an *implied* (fourth) category of 6-9 blackflies in a tent, with a frequency of 1. The probabilities of the k = 5 outcomes of this multinomial distribution are obtained from an appropriate *Poisson* model.

With an average of 15 blackflies per cubic metre of air, we expect 9 blackflies in a tent of volume 0.6 m³.

Hence: $\Pr(y \text{ blackflies in a tent of volume } 0.6\ \text{m}^3) = \dfrac{9^y e^{-9}}{y!} \quad ; \quad y = 0, 1, 2, 3, \ldots .$

Thus, the probabilities of the five outcomes for the multinomial distribution are:

Pr(0 blackflies in a tent) = $e^{-9} \simeq 0.000\,123\,410 = \pi_1$, say;

Pr(1 or 2 blackflies in a tent) = Pr(1 blackfly) + Pr(2 blackflies)

$$= 9e^{-9}/1! + 9^2e^{-9}/2! = \tfrac{99}{2}e^{-9} \simeq 0.006\,108\,785 = \pi_2, \text{ say;}$$

Pr(3, 4 or 5 blackflies in a tent) = Pr(3 blackflies) + Pr(4 blackflies) + Pr(5 blackflies)

$$= 9^3e^{-9}/3! + 9^4e^{-9}/(4! + 9^5e^{-9}/5! = \tfrac{17,739}{20}e^{-9} \simeq 0.109\,458\,325 = \pi_3, \text{ say;}$$

Pr(6, 7, 8 or 9 blackflies in a tent) = Pr(6 blackflies + Pr(7 blackflies) + Pr(8 blackflies) + Pr(9 blackflies)

$$= 9^6e^{-9}/6! + 9^7e^{-9}/7! + 9^8e^{-9}/8! + 9^9e^{-9}/9! = \tfrac{1,712,421}{448}e^{-9} \simeq 0.471\,717\,723 = \pi_4, \text{ say;}$$

Pr(10 or more blackflies in a tent) = $1 - \pi_1 - \pi_2 - \pi_3 - \pi_4 \simeq 0.412\,591\,755 = \pi_5$, say.

(a) Then, with the categories in the same order as in the question and in the preceding Poisson calculations, and using the usual notation for the multinomial distribution:

$$\Pr(6, 4, 3, 1, 6) = \binom{20}{6\ 4\ 3\ 1\ 6}\pi_1^6\pi_2^4\pi_3^3\pi_4^1\pi_5^6 \simeq 4.892\,901 \times 10^{-28}.$$

(b) Conditioning on six tents having no blackflies as event B in the definition of $\Pr(A|B)$, we have:

$$\Pr(4, 3, 1, 6 | 6) = \frac{\Pr(6, 4, 3, 1, 6)}{\Pr(\text{no blackflies in } 6 \text{ tents})} = \frac{\Pr(6, 4, 3, 1, 6)}{\binom{20}{6}\pi_1^6(1-\pi_1)^{14}}$$
$$\simeq \frac{4.892\,900\,986 \times 10^{-28}}{1.366\,883\,032 \times 10^{-19}} \simeq 3.579\,605 \times 10^{-9}.$$

(continued overleaf)

Assignment 2 Outline Solution (continued 1)

A2 – 3. This question involves the using the Central Limit Theorem to justify an assumed approximate normality for the distribution of a sum of 200 discrete random variables.

Let the random variable Y_i represent the number of typing mistakes on the ith page; we are given in the question that the probability function of Y_i is as shown at the right.

y_i	0	1	2	3	4	5	6
$f(y_i)$	0.05	0.25	0.31	0.22	0.10	0.05	0.02

(a) Then: $E(Y_i) = 0\times 0.05 + 1\times 0.25 + \ldots\ldots + 6\times 0.02 = 2.3$ mistakes (as required);

$E(Y_i^2) = 0^2\times 0.05 + 1^2\times 0.25 + \ldots\ldots + 6^2\times 0.02 = 7.04$ mistakes^2;

so that: $s.d.(Y_i) = \sqrt{7.04 - (2.3)^2} = \sqrt{1.75} \simeq 1.3229$ mistakes (as required).

If the manuscript contains 200 typed pages, the total number of mistakes is given by: $T = \sum_{i=1}^{200} Y_i$,

so that: $E(T) = 460$ mistakes, $\text{s.d.}(T) = \sqrt{350} \simeq 18.7083$ mistakes;

also, T has approximately a normal distribution as a consequence of the Central Limit Theorem.

(b) We want to find: $\Pr(T \ge 500) \simeq \Pr[N(460, \sqrt{350}) \ge 500 - ½] \simeq \Pr[N(0,1) \ge \frac{499.5 - 460}{\sqrt{350}}]$
$\simeq \Pr[N(0,1) \ge 2.111\,364] \simeq 0.017\,371$;

i.e., the probability of at least 500 mistakes in the manuscript is about 0.017 or about 1.7%.

A2 – 4. This question involves the Poisson distribution (and its normal approximation) as a probability model.

Let the random variable Y represent the number of calls arriving at the exchange in a time interval of length t; we use the model $Y \sim \text{Po}(\lambda t)$.

i.e., $\Pr(y \text{ calls reach the exchange in time t}) = \frac{(\lambda t)^y e^{-\lambda t}}{y!}$; $y = 0, 1, 2, 3, \ldots,$

where: λ = average rate at which calls reach the exchange
= 225 per hour = 3.75 per minute.

We know that the Poisson distribution has mean $\mu = \lambda t = 225$ calls and standard deviation $\sqrt{\mu} = 15$ calls.

Also, to simplify the evaluation of the required Poisson probabilities, we use the normal approximation to the Poisson distribution; with a mean as large as 225, we expect a very good approximation.

(a) Then: $\Pr(Y \ge 250) \simeq \Pr[N(225, 15) \ge 250 - ½] \simeq \Pr[N(0,1) \ge \frac{249.5 - 225}{15}]$
$\simeq \Pr[N(0,1) \ge 1.6\dot{3}] \simeq 0.051\,199$;

i.e., the probability *at least* 250 calls arrive at the exchange in 1 hour is about 0.051 or about 5.1%.

(b) Also: $\Pr(Y = 250) \simeq \Pr[250 - ½ \le N(225, 15) \le 250 + ½] \simeq \Pr[\frac{249.5 - 225}{15} \le N(0,1) \le \frac{250.5 - 225}{15}]$
$\simeq \Pr[1.6\dot{3} \le N(0,1) \le 1.7] \simeq 0.006\,634$;

i.e., the probability *exactly* 250 calls arrive at the exchange in 1 hour is about 0.0066 or about 0.7%.

NOTE: It is of interest to compare the result obtained relatively easily by means of the normal ap proxima- tion in (b) to the more troublesome calculation of the *exact* Poisson probability:

$$\Pr(\textit{exactly } 250 \text{ calls in 1 hour}) = \frac{(225)^{250} e^{-225}}{250!} \simeq \frac{2.134\,868\,355 \times 10^{490}}{3.232\,860\,761 \times 10^{492}} \simeq 0.006\,603\,651;$$

i.e., the normal approximation is accurate to about 1 part in 200 or about 0.5% in this situation.

A2 – 5. Let the random variable Y_i represent the number of people in the ith car crossing the bridge; we are given that the probability function of Y_i is as shown at the right.

y_i	1	2	3	4	5	6
$f(y_i)$	0.05	0.43	0.27	0.12	0.09	0.04

Then: $E(Y_i) = 1 \times 0.05 + 2 \times 0.43 + \ldots.. + 6 \times 0.04 = 2.89$ people per car;

so that: $E(Y_i - 1) = 2.89 - 1 = 1.89$ passengers per car.

Hence, the following amounts are raised by the three options given:

- at \$1.00 per car, raise \$1.00 per vehicle;
- at \$0.40 per person, raise \$0.40 × 2.89 = \$1.156 per vehicle;
- at \$0.75 per car and driver and \$0.15 per passenger, raise \$0.75 + \$0.15 × 1.89 = \$1.0335 per vehicle.

Thus, the *second* option raises the most money.

(*continued*)

Assignment 2 Outline Solution (continued 2)

A2 – 6. (a) Let the random variable Y represent the number of blood tests required for a group of k people; Y has two possible values: 1 and k + 1.

Also: $\Pr(Y = 1) = \Pr(\text{pooled test is negative})$
$= \Pr(\text{none of the k people is infected}) = (1-\pi)^k$
(assuming that the results of tests on different people in the group are independent);
$\Pr(Y = k+1) = 1 - \Pr(Y=1) = 1 - (1-\pi)^k$.

Thus, the probability function of Y is as shown at the right,

y_i	1	k + 1
$f(y_i)$	$(1-\pi)^k$	$1-(1-\pi)^k$

so that: $E(Y) = \sum_y y \cdot f(y) = 1\times(1-\pi)^k + (k+1)\times[1-(1-\pi)^k]$
$= (k+1) - k(1-\pi)^k$,

which is the expression given for the average number of tests required for a group of k people.

(b) $E(Y)$ is the expected number of tests which will have to be made for a group of k people; hence, the number expected for n/k groups of k people is:

$$\frac{n}{k}\times E(Y) = \frac{n}{k}[(k+1) - k(1-\pi)^k] \qquad \text{[using (a)].}$$

(c) If π is small, $(1-\pi)^k \simeq 1 - k\pi$ (using the binomial theorem) so that, from the result in (b), the expected number of tests is given by:

$$T \simeq \frac{n}{k}[(k+1) - k(1-k\pi)] = \frac{n}{k}[k^2\pi + 1] = n(k\pi + k^{-1}), \text{ as required.}$$

Then: $\dfrac{dT}{dk} \simeq n\pi - \dfrac{n}{k^2}$, which is zero when: $k^2 = \dfrac{1}{\pi}$ or: $k \simeq 1/\sqrt{\pi}$, as required;

also: $\dfrac{d^2T}{dk^2} \simeq \dfrac{2n}{k^3}$ which is positive because n and k are positive, as required for a *minimum*.

Illustration: If $\pi = 0.01$, we take $k = 1/\sqrt{0.01} = 10$, so $T \simeq n\times 10\times 0.01 + n/10 = 0.2n$;
i.e., on average, only about *one-fifth* as many tests are needed in the grouped procedure as would be needed if n people were tested individually, when π is 0.01 or 1%.

A2 – 7. The random variable Y is *the number of prizes awarded*. To tabulate its probability function, we need to find the values Y can take on and the probabilities of these values; we can set out the relevant information as shown in the table at the right.

TYPE OF TICKET NUMBER	NUMBER OF TICKETS	PRIZES
4 digits the same	$\binom{10}{1}\frac{4!}{4!} = 10$	1
3 digits the same	$\binom{10}{1}\binom{9}{1}\frac{4!}{3!1!} = 360$	4
2 pairs of digits the same	$\binom{10}{2}\frac{4!}{2!2!} = 270$	6
2 digits the same	$\binom{10}{1}\binom{9}{2}\frac{4!}{2!1!1!} = 4{,}320$	12
all digits different	$\binom{10}{4}\frac{4!}{1!1!1!1!} = 5{,}040$	24

Using this information and remembering that there are 10,000 tickets in the lottery, we can tabulate the probability function of Y as follows:

y	1	4	6	12	24
$f(y)$	0.001	0.036	0.027	0.432	0.504

, so that: $E(Y) = 1\times 0.001 + 4\times 0.036 + \ldots\ldots + 24\times 0.504$
$= 17.587$.

With tickets costing \$1 and each prize being \$500, the lottery operator's profit is: $W = \$10{,}000 - \$500\times Y$

$\therefore$ $E(W) = \$10{,}000 - \$500\times E(Y) = \$10{,}000 - \$500\times 17.587 = \$1{,}206.50$,

which is the required lottery operator's *expected* profit.

A2 – 8. This question involves the use of the Central Limit Theorem to justify an assumed approximate normality for the distribution of a sum of 400 discrete random variables.

Let the random variable Y_i represent the number of slices of pickle on the ith hamburger; from the information given in the statement of the question, we can obtain the probability function of Y_i as shown at the right.

y_i	0	1	2	3
$f(y_i)$	0.2	0.4	0.3	0.1

Then: $E(Y_i) = 0\times 0.2 + 1\times 0.4 + \ldots\ldots + 3\times 0.1 = 1.3$ slices;
$E(Y_i^2) = 0^2\times 0.2 + 1^2\times 0.4 + \ldots\ldots + 3^2\times 0.1 = 2.5$ slices^2;

so that: $s.d.(Y_i) = \sqrt{2.5 - (1.3)^2} = 0.9$ slices;

(*continued overleaf*)

Assignment 2 Outline Solution (continued 3)

A2 – 8. (cont.) If 400 hamburgers are to be sold, the total number of slices of pickle is given by: $T = \sum_{i=1}^{400} Y_i$,

so that: $E(T) = 520$ slices, $\text{s.d.}(T) = \sqrt{324} = 18$ slices;

also, T has approximately a normal distribution as a consequence of the Central Limit Theorem.

We want to find the value of t such that: $\Pr(T \le t) = 0.99$ or: $\Pr(T > t) = 0.01$.

Hence: $\Pr(T > t) \simeq \Pr[N(520, 18) > t + ½] \simeq \Pr[N(0, 1) > \frac{t + ½ - 520}{18}] \simeq 0.01$;

$\therefore$ $\frac{t - 519.5}{18} \simeq 2.3263$ [from the $N(0, 1)$ table] so that: $t \simeq 561.37 \simeq 562$ (slices).

i.e., about 562 slices of pickle need to be on hand to be 99% sure of meeting the demand.

***A2 – 9.** This question involves finding the mean of a negative binomial (really a *geometric*) distribution, for which the parameter π is obtained from an appropriate *Poisson* model.

Let the random variable Y represent the number of mistakes on an equiprobably-selected page;

we use the model $Y \sim \text{Po}(2)$, so that: $\Pr(y \text{ mistakes on the page}) = \frac{2^y e^{-2}}{y!}$; $y = 0, 1, 2, 3, \ldots,$

Hence, for an equiprobably-selected page:

$$\begin{aligned}
\Pr(\text{page is acceptable}) &= \Pr(\text{page has 2 or fewer mistakes}) \\
&= \Pr(\text{no mistakes}) + \Pr(1 \text{ mistake}) + \Pr(2 \text{ mistakes}) \\
&= 2^0e^{-2}/0! + 2^1e^{-2}/1! + 2^2e^{-2}/2! \\
&= 5e^{-2} \simeq 0.676\,676 = \pi, \text{ say;} \\
\Pr(\text{page is } \textit{un}\text{acceptable}) &= 1 - \Pr(\text{page is acceptable}) \\
&= 1 - 5e^{-2} = 1 - \pi.
\end{aligned}$$

Then, if the random variable R represents the number of times a page must be typed before it is acceptable:

$$\Pr(\text{type a page } r \text{ times}) \equiv f(r) = (1-\pi)^{r-1} \times \pi \quad ; \quad r = 1, 2, 3, \ldots .$$

Hence:

$$\begin{aligned}
E(R) = \sum_r r \cdot f(r) = \sum_{r=1}^{\infty} r \cdot (1-\pi)^{r-1} \times \pi &= \pi \sum_{r=1}^{\infty} -\frac{d}{d\pi}(1-\pi)^r \\
&= \pi \frac{d}{d\pi} \sum_{r=1}^{\infty} -(1-\pi)^r \\
&= \pi \frac{d}{d\pi}\left(\frac{\pi - 1}{\pi}\right) \\
&= \frac{1}{\pi} = e^2/5 \simeq 1.477\,811;
\end{aligned}$$

thus, the secretary can expect to type $100 \times 1.477\,811 \simeq 148$ pages.

NOTE: Three points in interest in the final calculation of $E(R)$ are:

- recognition that $r \cdot (1-\pi)^{r-1}$ is the derivative with respect to π of $(1-\pi)^r$.
- interchange of the order of summation and differentiation, which is permitted because the two operations involve *different* variables (r and π respectively);
- the result for the sum of an infinite geometric series.