Week 7, Problem Set 1 - worked solutions

Wackerly 3.7-3.8 - The Hypergeometric and Poisson Distributions; Midterm I Review

> This is one randomly generated version of the set. Every student receives different numbers, so the values here will not match your own copy — follow the method, not the arithmetic.

Problem 1

seed 20260925

A Central Bank examiner visits a branch that holds 16 mortgage files. Unknown to the examiner, 7 of them have a documentation error. The examiner pulls 4 files at random, all at once, so that every set of 4 files is equally likely to be chosen. Let Y be the number of files with an error in the examiner's sample.

Give every probability to at least four significant figures (enter a decimal number).

(a) $P(Y = 0) =$ ____________

(b) $P(Y = 1) =$ ____________

(c) $P(Y \geq 2) =$ ____________

The population has $N = 16$ files, $r = 7$ with the attribute "has an error" and $N - r = 9$ without; a sample of $n = 4$ is taken without replacement. By Definition 3.10, Y is hypergeometric:

$$p(y) = \frac{\binom{7}{y}\binom{9}{4-y}}{\binom{16}{4}}, \qquad \binom{16}{4} = 1820.$$

(a) No error means all 4 files come from the 9 sound ones:

$$p(0) = \frac{\binom{7}{0}\binom{9}{4}}{\binom{16}{4}} = \frac{126}{1820} = 0.069231.$$

(b) One file from the 7 with errors and 3 from the 9 sound ones (the mn rule):

$$p(1) = \frac{\binom{7}{1}\binom{9}{3}}{\binom{16}{4}} = \frac{7 \times 84}{1820} = 0.323077.$$

(c) By the complement, $P(Y \geq 2) = 1 - p(0) - p(1) = 1 - 0.069231 - 0.323077 = 0.607692$.

The draws are not independent trials: once an erroneous file is pulled, fewer remain, so the binomial model of Section 3.4 does not apply exactly. That is precisely the situation the hypergeometric distribution is built for.

Problem 2

seed 20260925

A regional equity index lists 24 stocks, 12 of which are banks. A new fund picks 6 different index stocks at random, every set of 6 being equally likely. Let Y be the number of bank stocks the fund picks.

Give every answer to at least four significant figures.

(a) $E(Y) =$ ____________

(b) $V(Y) =$ ____________

(c) The standard deviation of Y is ____________

(d) The factor by which $V(Y)$ is smaller than the variance of a binomial random variable with $n = 6$ and $p = r/N$: $V(Y)/[np(1-p)] =$ ____________

The fund samples $n = 6$ of $N = 24$ stocks without replacement, and $r = 12$ of the N are banks, so Y has a hypergeometric distribution (Definition 3.10). Theorem 3.10 gives its mean and variance.

(a) $E(Y) = \dfrac{nr}{N} = \dfrac{6 \times 12}{24} = 3.0000.$

(b)

$$V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 6\left(\frac{12}{24}\right)\left(\frac{12}{24}\right)\left(\frac{18}{23}\right) = 1.500000 \times 0.782609 = 1.173913.$$

(c) $\sigma = \sqrt{1.173913} = 1.0835.$

(d) Writing $p = r/N$, the first three factors of $V(Y)$ are exactly $np(1-p) = 1.500000$, the variance a binomial count would have if the fund drew with replacement. What is left over is

$$\frac{N-n}{N-1} = \frac{18}{23} = 0.782609.$$

Sampling without replacement removes some variability: every bank stock picked makes the next pick less likely to be a bank. The correction is noticeable here because n is not small relative to N; for fixed n it tends to 1 as $N \to \infty$.

Problem 3

seed 20260925

Calls to the card-services line of a Baku bank arrive at random, and the number Y of calls in a five-minute interval has a Poisson distribution with mean $\lambda = 2.20$.

Give every probability to at least four significant figures.

(a) $P(Y = 0) =$ ____________

(b) $P(Y = 6) =$ ____________

(c) $P(Y \leq 2) =$ ____________

(d) $P(Y > 6) =$ ____________

By Definition 3.11, $p(y) = \dfrac{\lambda^y}{y!}e^{-\lambda}$ for $y = 0, 1, 2, \ldots$, here with $\lambda = 2.20$.

(a) $p(0) = e^{-2.20} = 0.110803$.

(b) $p(6) = \dfrac{2.20^6}{6!} e^{-2.20} = 0.017448$.

(c) Add the first three probabilities: $p(1) = 2.20\, e^{-2.20} = 0.243767$ and $p(2) = \dfrac{2.20^2}{2} e^{-2.20} = 0.268144$, so

$$P(Y \le 2) = 0.110803 + 0.243767 + 0.268144 = 0.622714.$$

(d) "More than 6" has infinitely many values, so use the complement:

$$P(Y > 6) = 1 - \sum_{y=0}^{6} p(y) = 1 - 0.992539 = 0.007461.$$

In R the same numbers are `dpois(6, 2.20)` and `1 - ppois(6, 2.20)`.

Problem 4

seed 20260925

A bank's card-fraud system raises alerts at random; the number Y of alerts on a given day has a Poisson distribution with mean $\lambda = 3.0$. Monitoring costs 380 AZN a day plus 30 AZN for every alert investigated, so the daily cost is $C = 380 + 30Y$ AZN.

Give every answer to at least four significant figures.

(a) $V(Y) =$ ____________

(b) $E(Y^2) =$ ____________

(c) $E(C) =$ ____________ AZN

(d) The standard deviation of C is ____________ AZN

(a) By Theorem 3.11 a Poisson random variable has $E(Y) = V(Y) = \lambda$, so $V(Y) = 3.0$ (and the standard deviation of Y is $\sqrt{3.0} = 1.7321$).

(b) Theorem 3.6 says $V(Y) = E(Y^2) - \mu^2$, so

$$E(Y^2) = V(Y) + [E(Y)]^2 = \lambda + \lambda^2 = 3.0 + 3.0^2 = 12.00.$$

(c) By Theorems 3.3-3.5 (expectation of a constant, of a constant times a function, and of a sum),

$$E(C) = 380 + 30\, E(Y) = 380 + 30 \times 3.0 = 470.00 \text{ AZN}.$$

(d) Adding a constant does not change spread, and multiplying by 30 multiplies the variance by 30^2:

$$V(C) = 30^2\, V(Y) = 900 \times 3.0 = 2700.00, \qquad \sigma_C = \sqrt{2700.00} = 51.9615 \text{ AZN}.$$

Because the Poisson mean and variance are the same number, a busier day is also a less predictable one: the standard deviation of the alert count grows like $\sqrt{\lambda}$.

Problem 5

seed 20260925

A microfinance lender holds 140 small loans to unrelated market traders. Each loan defaults this year with probability $p = 0.019$, independently of the others, so the number Y of defaults is binomial with $n = 140$ and $p = 0.019$.

Give every answer to at least four significant figures (enter decimal numbers).

(a) The parameter of the approximating Poisson distribution: $\lambda =$ _______________

(b) The Poisson approximation to $P(Y = 0)$: _______________

(c) The Poisson approximation to $P(Y \le 2)$: _______________

(d) The exact binomial value of $P(Y = 0)$: _______________

Section 3.8 obtains the Poisson probability function as the limit of the binomial one when $n \to \infty$ with $np = \lambda$ held fixed. So for large n, small p and $\lambda = np$ below roughly 7, binomial probabilities are well approximated by Poisson probabilities with the same mean.

(a) $\lambda = np = 140 \times 0.019 = 2.66$.

(b) $P(Y = 0) \approx e^{-\lambda} = e^{-2.66} = 0.069948$.

(c) Summing Poisson probabilities with $\lambda = 2.66$: $p(0) = 0.069948$, $p(1) = 2.66e^{-2.66} = 0.186062$, $p(2) = 0.247463$, $p(3) = 0.219417$, so taking $y = 0, \ldots, 2$,

$$P(Y \le 2) \approx \sum_{y=0}^{2} \frac{2.66^y}{y!} e^{-2.66} = 0.503473.$$

(d) Exactly, no default means every loan survives: $P(Y = 0) = (1 - p)^n = 0.981^{140} = 0.068181$.

The approximation in (b) is off by only 0.001768. It works because $(1 - \lambda/n)^n \to e^{-\lambda}$, which is exactly the step in the derivation of Definition 3.11.

Problem 6

seed 20260925

Customers walk into a small branch of a bank in Ganja according to a Poisson process with an average of 10 customers per hour.

Give every answer to at least four significant figures.

(a) The teller steps away for 8 minutes. The expected number of customers who arrive in that time is

____

(b) The probability that at least one customer arrives while the teller is away is _______________

(c) The probability that exactly two customers arrive in a 15-minute window is _______________

For a Poisson process with mean λ occurrences per unit, the number of occurrences in a units has a Poisson distribution with mean $a\lambda$ (Section 3.8). Here the unit is one hour and $\lambda = 10$.

(a) 8 minutes is $a = 8/60$ hours, so the mean is $a\lambda = 10 \times 8/60 = 1.3333$.

(b) With $Y \sim$ Poisson(1.3333), $P(Y \geq 1) = 1 - p(0) = 1 - e^{-1.3333} = 1 - 0.263597 = 0.736403$.

(c) In 15 minutes the mean is $10 \times 15/60 = 2.5000$, so

$$P(Y = 2) = \frac{2.5000^2}{2!} e^{-2.5000} = 0.256516.$$

The rate must always be converted to the length of the interval being asked about before Definition 3.11 is applied; using $\lambda = 10$ in (b) would describe a whole hour, not 8 minutes.

Problem 7

seed 20260925

A payments company buys card terminals in lots of 18. Testing a terminal is slow, so the company tests 5 terminals chosen at random from each lot (without replacement) and rejects the lot if more than one tested terminal is faulty; a rejected lot goes back to the supplier.

A particular lot contains 6 faulty terminals. Let Y be the number of faulty terminals among the 5 tested.

Give every answer to at least four significant figures.

(a) The probability that this lot is rejected is ______________

(b) $E(Y) =$ ______________

(c) $V(Y) =$ ______________

Y counts faulty items in a sample of $n = 5$ drawn without replacement from $N = 18$ items of which $r = 6$ are faulty, so Y is hypergeometric (Definition 3.10) with $p(y) = \binom{6}{y}\binom{12}{5-y} \Big/ \binom{18}{5}$ and $\binom{18}{5} = 8568$.

(a) The lot is rejected when $Y \geq 2$. The complement is quicker:

$$p(0) = \frac{\binom{12}{5}}{\binom{18}{5}} = \frac{792}{8568} = 0.092437, \qquad p(1) = \frac{\binom{6}{1}\binom{12}{4}}{\binom{18}{5}} = \frac{6 \times 495}{8568} = 0.346639.$$

The lot is accepted with probability $p(0) + p(1) = 0.439076$, so $P(Y \geq 2) = 1 - 0.439076 = 0.560924$.

(b) By Theorem 3.10, $E(Y) = nr/N = 5 \times 6/18 = 1.6667$.

(c) Again by Theorem 3.10,

$$V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = 5\left(\frac{6}{18}\right)\left(\frac{12}{18}\right)\left(\frac{13}{17}\right) = 0.849673.$$

The rejection probability in (a) is the power of the sampling plan against a lot this bad: the larger it is, the fewer faulty lots slip through.

Problem 8

seed 20260925

A district tax office has a register of 15 small firms, 6 of which under-report their turnover. It has capacity for 4 audits this year.

Each week the office draws one firm at random from the whole register, and a firm that has already been audited stays in the register and can be drawn again. This is repeated for 4 weeks.

Let Y be the number of audits that land on an under-reporting firm.

(a) Which model describes Y exactly?

- ◯ Hypergeometric with N = 15, r = 6, n = 4
- ◯ Binomial with n = 4, p = 6/15
- ◯ Poisson with mean 4 x 6/15
- ◯ Geometric with p = 6/15

Give the remaining answers to at least four significant figures.

(b) $P(Y = 1) =$ ______________

(c) $V(Y) =$ ______________

(a) Because a firm is put back before the next draw, the 4 draws are independent trials with the same success probability 6/15: Y is binomial (Definition 3.7).

The two candidate answers, with $p = r/N = 0.400000$:

Binomial (with replacement). By Definition 3.7 and Theorem 3.7,

$$P(Y = 1) = \binom{4}{1} p\, q^3 = 4(0.400000)(0.600000)^3 = 0.345600, \qquad V(Y) = npq = 0.960000.$$

Hypergeometric (without replacement). By Definition 3.10 and Theorem 3.10,

$$P(Y = 1) = \frac{\binom{6}{1}\binom{9}{3}}{\binom{15}{4}} = \frac{6 \times 84}{1365} = 0.369231, \qquad V(Y) = npq\,\frac{N-n}{N-1} = 0.960000 \times \frac{11}{14} = 0.754286.$$

(b)-(c) The model chosen in (a) gives $P(Y = 1) = 0.345600$ and $V(Y) = 0.960000$.

Had the firms been drawn without replacement, the hypergeometric answers would have been 0.369231 and 0.754286. With a register this small relative to n, the choice of model changes the answers substantially; the binomial is only a good approximation to the hypergeometric when N is large relative to n.

Problem 9

seed 20260925

A bank receives a batch of 19 newly personalised payment cards. Exactly 5 of them have a faulty chip, but nobody knows which. Before the batch is mailed to customers, the card centre tests n cards chosen at random, without replacement. Let Y be the number of faulty cards among those tested.

The centre wants the probability of catching at least one faulty card to be at least 0.70.

(a) The smallest sample size n that achieves this is ____________

Give the probabilities to at least four significant figures.

(b) With that n, $P(Y \geq 1) =$ ____________

(c) With one card fewer, $P(Y \geq 1) =$ ____________

Y is hypergeometric (Definition 3.10) with $N = 19$, $r = 5$ and sample size n. The event "at least one faulty card" is the complement of $Y = 0$, and

$$P(Y = 0) = \frac{\binom{5}{0}\binom{14}{n}}{\binom{19}{n}} = \frac{14}{19} \cdot \frac{14-1}{19-1} \cdots \frac{14-n+1}{19-n+1},$$

the probability that each tested card in turn is sound. This product only falls as n grows, so $P(Y \geq 1) = 1 - P(Y = 0)$ only rises, and we look for the first n at which it reaches 0.70.

(c) At $n = 3$: $P(Y = 0) = 0.375645$, so $P(Y \geq 1) = 0.624355 < 0.70$, which is not enough.

(b) Multiplying in one more factor, $11/16$, at $n = 4$: $P(Y = 0) = 0.258256$, so $P(Y \geq 1) = 0.741744 \geq 0.70$.

(a) Hence the smallest sample size is $n = 4$.

The same question under a binomial model would never need a smaller n, and often needs a larger one: drawing without replacement exhausts the sound cards, so each extra test is worth a little more than it would be with replacement.

Problem 10

seed 20260925

A bank's data centre has 14 identical-looking servers in stock; 5 of them shipped with outdated firmware. Tonight the team takes 4 servers at random from stock and installs them. A server with current firmware takes 40 minutes to install; one with outdated firmware must be re-flashed first and takes 130 minutes. Let Y be the number of outdated servers among the 4, and T the total installation time in minutes.

Give every answer to at least four significant figures.

(a) The probability that none of the 4 servers needs re-flashing is ____________

(b) $E(T) =$ ____________ minutes

(c) $V(T) =$ ____________ squared minutes

(d) The standard deviation of T is ______________ minutes

Y is hypergeometric (Definition 3.10) with $N = 14$, $r = 5$, $n = 4$. Each of the 4 servers takes 40 minutes, and each outdated one takes $130 - 40 = 90$ minutes more, so

$$T = 40 \times 4 + 90\,Y = 160 + 90\,Y.$$

(a) $P(Y = 0) = \binom{9}{4} \Big/ \binom{14}{4} = 0.125874$, the chance that all 4 servers come from the 9 good ones.

(b) By Theorem 3.10, $E(Y) = nr/N = 1.428571$. By Theorems 3.3-3.5,

$$E(T) = 160 + 90\,E(Y) = 160 + 90 \times 1.428571 = 288.5714.$$

(c) By Theorem 3.10,

$$V(Y) = 4\left(\frac{5}{14}\right)\left(\frac{9}{14}\right)\left(\frac{10}{13}\right) = 0.706436.$$

The constant 160 does not affect the variance, and the factor 90 enters squared:

$$V(T) = 90^2\,V(Y) = 8100 \times 0.706436 = 5722.1350.$$

(d) $\sigma_T = \sqrt{5722.1350} = 75.6448$ minutes.

A common slip is to write $V(T) = 130^2 V(Y)$: only the **extra** time per outdated server is random, so it is the difference 90 that is squared.

Problem 11

seed 20260925

The number Y of claims per week on a small portfolio of cargo-insurance policies at a Baku insurer is modelled as Poisson with an unknown mean λ. Over many years, a fraction 0.120 of weeks have had no claims at all, and the actuary sets $P(Y = 0) = 0.120$. Weeks are treated as disjoint intervals of a Poisson process.

Give every answer to at least four significant figures.

(a) $\lambda =$ ______________

(b) In a week in which at least one claim arrives, the probability that at least 4 arrive: $P(Y \geq 4 \mid Y \geq 1) =$

____

(c) The probability of exactly two claims in total over the next 3 weeks is ______________

(a) By Definition 3.11, $P(Y = 0) = e^{-\lambda}$. Setting $e^{-\lambda} = 0.120$ gives $\lambda = -\ln(0.120) = 2.120264$.

(b) By Definition 2.9,

$$P(Y \geq 4 \mid Y \geq 1) = \frac{P(\{Y \geq 4\} \cap \{Y \geq 1\})}{P(Y \geq 1)} = \frac{P(Y \geq 4)}{1 - P(Y = 0)},$$

because $\{Y \geq 4\}$ is contained in $\{Y \geq 1\}$. With $\lambda = 2.120264$: $p(0) = 0.120$, $p(1) = 0.254432$, $p(2) = 0.269731$, $p(3) = 0.190634$, so

$$P(Y \geq 4) = 1 - \sum_{y=0}^{3} p(y) = 1 - 0.834796 = 0.165204, \qquad P(Y \geq 4 \mid Y \geq 1) = \frac{0.165204}{0.880} = 0.187731.$$

(c) For a Poisson process the count over $a = 3$ weeks is Poisson with mean $a\lambda = 3 \times 2.120264 = 6.360791$, so

$$P(\text{2 claims in 3 weeks}) = \frac{6.360791^2}{2!} e^{-6.360791} = 0.034957.$$

Note that (a) needed only one number from the data: because the Poisson family has a single parameter, pinning down any one probability fixes all of them.

Problem 12

seed 20260925

A payment switch in Baku receives card payments through two independent channels. In a given 10-millisecond interval, the number Y_1 of payments arriving through the mobile app is Poisson with mean 1.6, and the number Y_2 arriving through shop terminals is Poisson with mean 2.8. Treat the two counts as independent, and let $W = Y_1 + Y_2$.

Give every answer to at least four significant figures.

(a) $P(W = 3) =$ ______________

(b) $V(W) =$ ______________

(c) Given that $W = 3$, the probability that every one of the 3 payments came through the app: $P(Y_1 = 3 \mid W = 3) =$ ______________

(a) The event $W = 3$ is the union of the mutually exclusive events $\{Y_1 = j, Y_2 = 3 - j\}$, $j = 0, \ldots, 3$, and by independence each has probability $p_1(j)\, p_2(3 - j)$:

$$P(W = 3) = \sum_{j=0}^{3} \frac{1.6^j e^{-1.6}}{j!} \cdot \frac{2.8^{3-j} e^{-2.8}}{(3-j)!}.$$

The terms are $0.044919, 0.077003, \ldots, 0.008381$. Rather than add them one by one, factor out $e^{-(1.6+2.8)}/3!$:

$$P(W = 3) = \frac{e^{-4.4}}{3!} \sum_{j=0}^{3} \binom{3}{j} 1.6^j 2.8^{3-j} = \frac{4.4^3}{3!} e^{-4.4} = 0.174305,$$

using the binomial theorem. The same step works for every total, so W is itself Poisson with mean $1.6 + 2.8 = 4.4$.

(b) By Theorem 3.11 applied to W, $V(W) = 4.4$.

(c) By Definition 2.9,

$$P(Y_1 = 3 \mid W = 3) = \frac{P(Y_1 = 3, Y_2 = 0)}{P(W = 3)} = \frac{\frac{1.6^3}{3!} e^{-1.6} \cdot e^{-2.8}}{\frac{4.4^3}{3!} e^{-4.4}} = \left(\frac{1.6}{4.4}\right)^3 = 0.363636^3 = 0.048084.$$

Given the total, each payment behaves as if it independently chose the app with probability $1.6/4.4$: a binomial count hides inside the pair of Poissons.

Problem 13

seed 20260925

A catalytic cracking unit at a refinery near Baku breaks down at random, and the number Y of breakdowns in a day is Poisson with mean $\lambda = 1.4$. Each breakdown costs more than the one before (the unit must be restarted from a colder state), and the day's revenue in thousand AZN is modelled as

$$R = 19000 - 250\,Y^2.$$

Give every answer to at least four significant figures.

(a) $E(Y^2) =$ ____________

(b) $E(R) =$ ____________ thousand AZN

(c) The probability that the day's revenue is below 18000 thousand AZN is ____________

(a) By Theorem 3.11, $E(Y) = V(Y) = \lambda$. By Theorem 3.6, $V(Y) = E(Y^2) - [E(Y)]^2$, so

$$E(Y^2) = V(Y) + [E(Y)]^2 = \lambda + \lambda^2 = 1.4 + 1.4^2 = 3.36.$$

(b) By Theorems 3.3-3.5,

$$E(R) = 19000 - 250\,E(Y^2) = 19000 - 250 \times 3.36 = 18160.00.$$

Note that $E(R) \neq 19000 - 250\,[E(Y)]^2$: the expected value of a function is not the function of the expected value, and the difference here is $250\,V(Y)$.

(c) $R < 18000$ means $19000 - 250\,Y^2 < 19000 - 1000$, that is $Y^2 > 4$, and because Y is a non-negative integer, $Y \geq 3$:

$$P(Y \geq 3) = 1 - p(0) - p(1) - p(2) = 1 - 0.246597 - 0.345236 - 0.241665 = 0.166502.$$

Problem 14

seed 20260925

The security team of a bank logs cyber incidents, which occur according to a Poisson process with a long-run average of 2.8 incidents per month. Over the last 4 months, 15 incidents were logged. Let Y be the number of incidents in a 4-month period if the long-run rate still holds.

Give numerical answers to at least four significant figures.

(a) $E(Y) =$ ____________

(b) The upper end of the interval $\mu \pm 2\sigma$: $\mu + 2\sigma =$ ______

(c) Is the observed count of 15 beyond $\mu + 2\sigma$?

- ◯ Yes: it lies more than two standard deviations above the mean
- ◯ No: it lies within two standard deviations of the mean

(d) The exact probability $P(Y \geq 15) =$ ______

(a) For a Poisson process with $\lambda = 2.8$ per month, the count in $a = 4$ months is Poisson with mean $\lambda^{\star} = a\lambda = 4 \times 2.8 = 11.20$.

(b) By Theorem 3.11, $\sigma^2 = \lambda^{\star} = 11.20$, so $\sigma = \sqrt{11.20} = 3.3466$ and $\mu + 2\sigma = 11.20 + 2(3.3466) = 17.8933$.

(c) The observed 15 is below 17.8933, so by the empirical-rule yardstick it is not unusual, even though it is above the mean.

(d) The tail has infinitely many terms, so use the complement:

$$P(Y \geq 15) = 1 - \sum_{y=0}^{14} \frac{11.20^y}{y!} e^{-11.20} = 1 - 0.839101 = 0.160899.$$

(In R: `1 - ppois(14, 11.20)`.) As in Example 3.22, the two-sigma check is a quick screen; the exact tail probability says how surprising the count really is if the rate has not changed.

Problem 15

seed 20260925

An agent selling life-insurance policies by phone makes calls one after another. Each call ends in a sale with probability $p = 0.14$, independently of every other call.

Give every answer to at least four significant figures.

(a) The probability that the agent's first sale comes on call number 6: ______

(b) The variance of the number of the call on which the first sale comes: ______

The agent's daily target is 2 sales.

(c) The probability that the second sale comes on call number 8: ______

(d) The expected number of calls needed to reach 2 sales: ______

(e) The standard deviation of the number of calls needed to reach 2 sales: ______

The calls are independent trials with the same success probability, so this is the setting of Sections 3.5 and 3.6. The random variable is the **number of the trial** on which a given success occurs, not the number of successes in a fixed number of trials; the binomial is the wrong model here.

(a) Let Y be the call of the first sale. By Definition 3.8, Y is geometric: the first 5 calls fail and call 6 succeeds,

$$p(6) = q^5 p = (0.86)^5(0.14) = 0.065860.$$

(b) By Theorem 3.8, $V(Y) = \dfrac{1-p}{p^2} = \dfrac{0.86}{0.14^2} = 43.8776$.

(c) Let $Y^\star$ be the call of the second sale. By Definition 3.9 it is negative binomial with $r = 2$: the first 7 calls must contain exactly 1 sales (in any order) and call 8 must be a sale,

$$p^\star(8) = \binom{7}{1} p^2 q^6 = 7\,(0.14)^2(0.86)^6 = 0.055507.$$

(d) By Theorem 3.9, $E(Y^\star) = r/p = 2/0.14 = 14.2857$.

(e) By Theorem 3.9, $V(Y^\star) = \dfrac{r(1-p)}{p^2} = 87.7551$, so the standard deviation is $\sqrt{87.7551} = 9.3678$.

The negative binomial count is the sum of 2 independent geometric waiting times, which is why its mean and variance are 2 times those of (b).

Problem 16

seed 20260925

A leasing company has 38 car leases outstanding to unrelated clients. Each lease goes into arrears this quarter with probability 0.08, independently of the others. Y is the number of leases that go into arrears this quarter.

(a) Which distribution does Y have?

- ◯ Binomial
- ◯ Geometric
- ◯ Negative binomial
- ◯ Hypergeometric
- ◯ Poisson

Give the remaining answers to at least four significant figures.

(b) $E(Y) =$ ______________

(c) $V(Y) =$ ______________

(a) Binomial. A fixed number 38 of independent trials, each a success (arrears) with the same probability 0.08, and Y counts successes: Definition 3.7. By Theorem 3.7, E(Y) = np and V(Y) = np(1 - p).

(b) Substituting the numbers in the story, $E(Y) = 38 \times 0.08 = 3.0400$.

(c) Likewise $V(Y) = 38 \times 0.08 \times 0.92 = 2.7968$.

To tell the five models apart, ask two questions. Is the number of trials fixed in advance (binomial, hypergeometric), or is it the thing being counted (geometric, negative binomial)? And for a fixed number of draws, are the draws independent with a constant probability (binomial) or made without replacement from a small finite group (hypergeometric)? A count of events in an interval of time or space, with no fixed number of trials at all, is Poisson.

Problem 17

seed 20260925

A leasing company in Sumgait holds 26 equipment leases to unrelated small firms. Each lease defaults this year with probability $p = 0.075$, independently of the others. Each lease has an exposure of 11 thousand AZN, and the company recovers part of it on default, losing a fraction 0.65 of the exposure. Let Y be the number of defaults and L the total credit loss in thousand AZN.

Give every answer to at least four significant figures (enter decimal numbers).

(a) $E(L) =$ ______________ thousand AZN

(b) The standard deviation of L is ______________ thousand AZN

(c) The exact probability that the loss exceeds 14.30 thousand AZN: ______________

(d) The Poisson approximation to the probability in (c): ______________

Each default loses $11 \times 0.65 = 7.15$ thousand AZN, so $L = 7.15\,Y$. The leases are 26 independent trials with the same default probability, so Y is binomial with $n = 26$, $p = 0.075$ (Definition 3.7).

(a) By Theorem 3.7, $E(Y) = np = 1.9500$, and by Theorem 3.4, $E(L) = 7.15\,E(Y) = 7.15 \times 1.9500 = 13.9425$.

(b) By Theorem 3.7, $V(Y) = npq = 26(0.075)(0.925) = 1.803750$, so $\sigma_Y = 1.343038$. Multiplying Y by the constant 7.15 multiplies its standard deviation by 7.15: $\sigma_L = 7.15 \times 1.343038 = 9.6027$.

(c) $L > 14.30$ means $7.15\,Y > 2 \times 7.15$, i.e. $Y \geq 3$:

$$p(0) = q^{26} = 0.131730, \quad p(1) = 26\,p\,q^{25} = 0.277701, \quad p(2) = \binom{26}{2} p^2 q^{24} = 325\,p^2 q^{24} = 0.281454,$$

$$P(Y \geq 3) = 1 - 0.131730 - 0.277701 - 0.281454 = 0.309115.$$

(d) With $\lambda = np = 1.9500$ (Section 3.8), $p(0) \approx e^{-\lambda} = 0.142274$, $p(1) \approx \lambda e^{-\lambda} = 0.277434$, $p(2) \approx \frac{\lambda^2}{2} e^{-\lambda} = 0.270499$, so

$$P(Y \geq 3) \approx 1 - 0.142274 - 0.277434 - 0.270499 = 0.309793.$$

The two answers differ by 0.000678. The approximation is usable because p is small and λ is well below 7, but n is only 26, so it is not exact; the error shrinks as n grows with np held fixed, which is the limit that defines the Poisson distribution in Section 3.8.

Problem 18

seed 20260925

An electronics shop in Baku is clearing an old tablet model with a list price of 300 AZN. For every unit sold during the day, the shop multiplies the price of the next unit by 0.90. So if Y units are sold during the day, the price displayed at closing time is $300\,(0.90)^{Y}$ AZN.

Give every answer to at least four significant figures.

(a) Suppose the number of buyers Y is Poisson with mean 2.0. The expected closing price is ____________ AZN

(b) The closing price computed at the mean number of buyers, $300\,(0.90)^{2.0}$, is ____________ AZN

(c) Suppose instead that exactly 10 customers will see the offer and each buys one unit with probability 0.30, independently of the others. The expected closing price is then ____________ AZN

In both cases the closing price is $g(Y) = 300\,d^{Y}$ with $d = 0.90$, and Theorem 3.2 gives $E[g(Y)] = \sum_y g(y)p(y)$.

(a) With $p(y) = \lambda^y e^{-\lambda}/y!$ (Definition 3.11),

$$E\left(d^Y\right) = \sum_{y=0}^{\infty} d^y \frac{\lambda^y e^{-\lambda}}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda d)^y}{y!} = e^{-\lambda} e^{\lambda d} = e^{-\lambda(1-d)},$$

using the series for e^x as in Example 3.18. With $\lambda = 2.0$, $\lambda d = 1.8000$ and $\lambda(1-d) = 0.2000$:

$$E(\text{price}) = 300\,e^{-0.2000} = 245.6192 \text{ AZN}.$$

(b) $300 \times 0.90^{2.0} = 243.0000$ AZN. This is smaller than (a): d^y is a convex function of y, so $E(d^Y) > d^{E(Y)}$. The expected value of a function is not the function of the expected value.

(c) Now Y is binomial with $n = 10$, $p = 0.30$ (Definition 3.7), and

$$E\left(d^Y\right) = \sum_{y=0}^{10} d^y \binom{10}{y} p^y q^{10-y} = \sum_{y=0}^{10} \binom{10}{y} (pd)^y q^{10-y} = (q + pd)^{10}$$

by the binomial theorem. Here $q + pd = 0.70 + 0.30 \times 0.90 = 0.9700$, so

$$E(\text{price}) = 300 \times 0.9700^{10} = 221.2272 \text{ AZN}.$$

Problem 19

seed 20260925

A customs officer at the Baku port has a batch of 14 import declarations, of which 5 understate the value of the goods. She checks the declarations one at a time in a random order (every ordering equally likely), never checking the same one twice.

Give every answer to at least four significant figures.

(a) The probability that the **second** understated declaration she finds is the one checked in position 5: ______________

(b) The probability that at least two of the first 5 declarations she checks are understated: ______________

(c) The expected number of understated declarations among the first 5 checked: ______________

The first m declarations checked in a random order are a random sample of size m drawn without replacement, so the number of understated ones among them is hypergeometric (Definition 3.10) with $N = 14$, $r = 5$, $n = m$.

(a) "The second understated declaration is at position 5" is the intersection of two events:

- A: exactly one understated declaration among the first 4;
- B: the declaration at position 5 is understated.

By the hypergeometric distribution with $n = 4$,

$$P(A) = \frac{\binom{5}{1}\binom{9}{3}}{\binom{14}{4}} = \frac{5 \times 84}{1001} = 0.419580.$$

Given A, 10 declarations are left, and 4 of them are understated, so $P(B \mid A) = 4/10 = 0.400000$. By the multiplicative law (Theorem 2.5),

$$P(A \cap B) = P(A)\,P(B \mid A) = 0.419580 \times 0.400000 = 0.167832.$$

(b) Now $n = 5$, and $P(Y \geq 2) = 1 - p(0) - p(1)$ with

$$p(0) = \frac{\binom{9}{5}}{\binom{14}{5}} = \frac{126}{2002} = 0.062937, \qquad p(1) = \frac{\binom{5}{1}\binom{9}{4}}{\binom{14}{5}} = \frac{5 \times 126}{2002} = 0.314685,$$

so $P(Y \geq 2) = 1 - 0.062937 - 0.314685 = 0.622378$.

The answers to (a) and (b) are different questions: (b) is "the second understated declaration turns up **by** position 5", and summing the answer to (a) over positions $2, \ldots, 5$ gives (b).

(c) By Theorem 3.10 with $n = 5$, $E(Y) = nr/N = 5 \times 5/14 = 1.7857$.

Problem 20

seed 20260925

The compliance desk of a bank flags large-value transfers for review. The number Y of transfers flagged in a day is modelled as Poisson with an unknown mean λ. From its records the desk estimates

$$E[Y(Y-1)] = 15.3664.$$

Give numerical answers to at least four significant figures.

(a) $\lambda =$ ______________

(b) $E(Y^2) =$ ______________

(c) The most likely number of flagged transfers in a day (the value y with the largest $p(y)$) is ______________

(d) The probability of that most likely number: ______________

(a) By Theorem 3.2, and dropping the $y = 0$ and $y = 1$ terms, which are zero,

$$E[Y(Y-1)] = \sum_{y=2}^{\infty} y(y-1)\frac{\lambda^y e^{-\lambda}}{y!} = \lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2}e^{-\lambda}}{(y-2)!} = \lambda^2 \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \lambda^2,$$

the last sum being the total probability of a Poisson distribution. So $\lambda^2 = 15.3664$ and $\lambda = \sqrt{15.3664} = 3.92$.

(b) $E(Y^2) = E[Y(Y-1)] + E(Y) = \lambda^2 + \lambda = 15.3664 + 3.92 = 19.2864$. (Subtracting λ^2 then gives $V(Y) = \lambda$, Theorem 3.11.)

(c) Consecutive probabilities satisfy

$$\frac{p(y)}{p(y-1)} = \frac{\lambda^y e^{-\lambda}/y!}{\lambda^{y-1}e^{-\lambda}/(y-1)!} = \frac{\lambda}{y}.$$

So $p(y) > p(y-1)$ exactly when $y < \lambda = 3.92$: the probabilities rise up to $y = 3$ and fall afterwards. Check the two neighbours: $p(3)/p(3-1) = 3.92/3 = 1.306667 > 1$ and $p(4)/p(3) = 3.92/4 = 0.980000 < 1$. The most likely value is $y = 3$, the integer part of λ.

(d) $p(3) = \dfrac{3.92^3}{3!}e^{-3.92} = 0.199192.$

Because λ is not an integer, there is a single mode; when λ is an integer, $p(\lambda - 1) = p(\lambda)$ and there are two.

Answer key

Problem	Answers
1	(a) 0.0692308, (b) 0.323077, (c) 0.607692
2	(a) 3, (b) 1.17391, (c) 1.08347, (d) 0.782609
3	(a) 0.110803, (b) 0.0174484, (c) 0.622714, (d) 0.00746135
4	(a) 3, (b) 12, (c) 470, (d) 51.9615
5	(a) 2.66, (b) 0.0699482, (c) 0.503473, (d) 0.0681806
6	(a) 1.33333, (b) 0.736403, (c) 0.256516
7	(a) 0.560924, (b) 1.66667, (c) 0.849673
8	(a) Choice 2, (b) 0.3456, (c) 0.96
9	(a) 4, (b) 0.741744, (c) 0.624355
10	(a) 0.125874, (b) 288.571, (c) 5722.14, (d) 75.6448
11	(a) 2.12026, (b) 0.187731, (c) 0.0349571
12	(a) 0.174305, (b) 4.4, (c) 0.0480841
13	(a) 3.36, (b) 18160, (c) 0.166502

14	(a) 11.2, (b) 17.8933, (c) No: it ... of the mean, (d) 0.160899
15	(a) 0.0658598, (b) 43.8776, (c) 0.0555066, (d) 14.2857, (e) 9.36777
16	(a) Binomial, (b) 3.04, (c) 2.7968
17	(a) 13.9425, (b) 9.60272, (c) 0.309115, (d) 0.309793
18	(a) 245.619, (b) 243, (c) 221.227
19	(a) 0.167832, (b) 0.622378, (c) 1.78571
20	(a) 3.92, (b) 19.2864, (c) 3, (d) 0.199192