Week 6, Problem Set 2 - worked solutions

Wackerly 3.5-3.6 - The Geometric and Negative Binomial Distributions

> 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 telesales agent for a Baku insurance company cold-calls small businesses to sell a property-insurance policy. Past campaigns show that a single call ends in a sale with probability $p = 0.09$, and calls are to randomly chosen firms, so outcomes are independent from call to call. Let Y be the number of the call on which the agent makes her **first** sale of the day.

Give every probability to at least four significant figures.

(a) $P(Y = 8) =$ ______________

(b) The probability that the first sale comes within the first 8 calls, $P(Y \leq 8) =$ ______________

(c) The expected number of calls needed to make the first sale, $E(Y) =$ ______________

Each call is a trial with two outcomes, the trials are independent and the success probability $p = 0.09$ is the same on every call. Y, the number of the trial on which the first success occurs, therefore has the geometric distribution of Definition 3.8, $p(y) = q^{y-1}p$ for $y = 1, 2, \ldots$, with $q = 1 - p = 0.91$.

(a) The event $Y = 8$ is the single sample point "7 failures, then a success":

$$P(Y = 8) = q^7 p = (0.91)^7(0.09) = 0.0465.$$

(b) Summing the pmf is a finite geometric series:

$$P(Y \leq 8) = \sum_{y=1}^{8} q^{y-1}p = p\,\frac{1 - q^8}{1 - q} = 1 - q^8 = 1 - (0.91)^8 = 1 - 0.4703 = 0.5297.$$

The shortcut is worth seeing directly: the first sale is **not** within the first 8 calls exactly when all 8 calls fail, which has probability q^8.

(c) By Theorem 3.8, $E(Y) = 1/p = 1/0.09 = 11.1111$ calls. A sale rate of 0.09 per call means the agent should plan on roughly 11.1111 calls per sale -- the figure a sales manager uses to set daily call targets.

Problem 2

seed 20260925

The internal-audit unit of a large bank checks employee expense claims drawn at random from a very large pool. A fraction $p = 0.060$ of claims are fraudulent, and claims are drawn independently. Auditors keep checking claims until they find the first fraudulent one; let Y be the number of the audit on which that happens.

Give every probability to at least four significant figures.

(a) The probability that the first 15 audits find no fraud at all, $P(Y > 15) =$ ______________

(b) The probability that the first fraudulent claim is found on or after audit number 5, $P(Y \geq 5) =$

____

(c) The probability that the first fraudulent claim is found somewhere from audit 16 to audit 20 inclusive, $P(15 < Y \leq 20) =$ ______________

Y is geometric (Definition 3.8) with $p = 0.060$ and $q = 0.940$.

The key fact, Exercise 3.71(a), is that for a positive integer a,

$$P(Y > a) = \sum_{y=a+1}^{\infty} q^{y-1}p = p\,\frac{q^a}{1-q} = q^a,$$

which also follows at once from the sample points: $Y > a$ happens exactly when the first a trials are all failures.

(a) $P(Y > 15) = q^{15} = (0.940)^{15} = 0.3953$.

(b) $Y \geq 5$ is the same event as $Y > 4$, so

$$P(Y \geq 5) = q^4 = (0.940)^4 = 0.7807.$$

The exponent is 4, not 5: "on or after audit 5" only requires the first 4 audits to be clean.

(c) Take the difference of two tail probabilities:

$$P(15 < Y \leq 20) = P(Y > 15) - P(Y > 20) = q^{15} - q^{20} = 0.3953 - 0.2901 = 0.1052.$$

Equivalently $q^{15}(1 - q^5) = 0.3953(1 - 0.7339)$: survive the first 15 audits, then find fraud within the next 5.

Problem 3

seed 20260925

A bank's cash-logistics team models one busy ATM in Nasimi district. On any given day, independently of other days, the machine runs out of cash before the evening refill with probability $p = 0.019$. Let Y be the number of the day on which the first stock-out occurs (day 1 is tomorrow).

Give every answer to at least four significant figures.

(a) $E(Y) =$ ______________ days

(b) The standard deviation of Y, $\sigma =$ ______________ days

(c) $P(\mu - \sigma < Y < \mu + \sigma) =$ ______________

(d) The probability of at least one stock-out in the next 28 days, $P(Y \leq 28) =$ ______________

Days are independent trials with a constant "success" (stock-out) probability, so Y is geometric with $p = 0.019$, $q = 0.981$. As in the book's engine-malfunction example, "success" here is the bad outcome.

(a) Theorem 3.8: $\mu = E(Y) = 1/p = 1/0.019 = 52.6316$ days.

(b) Theorem 3.8: $\sigma^2 = (1-p)/p^2 = 0.981/0.019^2 = 2717.45$, so $\sigma = \sqrt{2717.45} = 52.1292$ days. For a geometric variable with small p the standard deviation is almost as large as the mean -- waiting times are very uncertain.

(c) The interval is $(0.5024, 104.7608)$. Its lower end is below 1 (in general $\mu - \sigma = (1-\sqrt{q})/p = 1/(1+\sqrt{q}) < 1$), so every possible value $y \geq 1$ clears it, and the event is simply $Y \leq 104$, the largest integer below 104.7608:

$$P(Y \leq 104) = 1 - q^{104} = 1 - (0.981)^{104} = 0.8640.$$

The distribution is heavily right-skewed: all the probability that is not within one standard deviation lies in the long right tail.

(d) $P(Y \leq 28) = 1 - P(Y > 28) = 1 - q^{28} = 1 - 0.5844 = 0.4156$.

Problem 4

seed 20260925

A car-leasing company's credit committee reviews applications one at a time. Each application is approved with probability $p = 0.62$, independently of the others. The branch needs 2 approved leases to fill this week's delivery of vehicles, and reviews applications until it has them. Let Y be the number of the application on which the 2nd approval occurs.

Give every answer to at least four significant figures.

(a) $P(Y = 4) =$ ______________

(b) $E(Y) =$ ______________

(c) Let $Y^* = Y - 2$ be the number of **rejected** applications reviewed before the 2nd approval. $E(Y^*) =$

____

Independent, identical two-outcome trials, stopped at the r-th success with $r = 2$: Y has the negative binomial distribution of Definition 3.9,

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r}, \qquad y = r, r+1, \ldots$$

with $p = 0.62$, $q = 0.38$.

(a) $Y = 4$ means the first 3 applications contain exactly 1 approvals (in any order) and application 4 is approved:

$$P(Y = 4) = \binom{3}{1}(0.62)^2(0.38)^2 = 3 \times 0.384400 \times 0.144400 = 0.1665.$$

(b) Theorem 3.9: $E(Y) = r/p = 2/0.62 = 3.2258$ applications.

(c) $Y^* = Y - r$ is a linear function of Y, so $E(Y^*) = E(Y) - r = 3.2258 - 2 = 1.2258$, which is also rq/p. This is the "number of failures" version of the negative binomial that R's `dnbinom` uses -- which is why R asks for $y_0 - r$ rather than y_0.

Problem 5

seed 20260925

A mobile operator's retention team phones subscribers who have asked to port their number to a competitor. Each call, independently, ends with the subscriber accepting a discounted tariff with probability $p = 0.38$. This month the team has a target of 4 retained subscribers and keeps calling until it reaches it. Each call costs the operator 3.50 AZN in agent time. Let Y be the number of calls made and $C = 3.50\,Y$ the total calling cost.

Give every answer to at least four significant figures.

(a) $E(Y) =$ ______________

(b) $V(Y) =$ ______________

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

(d) The standard deviation of C is ______________ AZN

Y is the number of the trial on which the r-th success occurs, $r = 4$, so it is negative binomial (Definition 3.9) with $p = 0.38$.

(a) Theorem 3.9: $E(Y) = r/p = 4/0.38 = 10.5263$ calls.

(b) Theorem 3.9: $V(Y) = r(1-p)/p^2 = 4(0.62)/(0.38)^2 = 17.1745$.

(c) For a constant c, $E(cY) = cE(Y)$ (Theorem 3.5; Exercise 3.33): $E(C) = 3.50 \times 10.5263 = 36.84$ AZN.

(d) $V(cY) = c^2 V(Y)$, so the standard deviation scales by c: $\sigma_C = 3.50\sqrt{17.1745} = 3.50 \times 4.1442 = 14.50$ AZN.

A budget of $E(C)$ would be overspent in a large fraction of months; a prudent manager budgets something like $E(C) + 2\sigma_C$.

Problem 6

seed 20260925

An exporter of Ganja wine files customs declarations one shipment at a time. The customs risk engine sends each declaration to physical inspection with probability $p = 0.12$, independently of past selections. Let Y be the number of the declaration that is the first one sent to inspection.

Give every answer to at least four significant figures.

(a) $P(Y > 10) =$ ______________

(b) The first 10 declarations have all cleared without inspection. Find the probability that the next 3 also clear, $P(Y > 13 \mid Y > 10) =$ ______________

(c) Given that the first 10 declarations cleared, the expected number of **additional** declarations up to and including the first inspection, $E(Y - 10 \mid Y > 10) =$ ______________

Y is geometric (Definition 3.8) with $p = 0.12$, $q = 0.88$, and $P(Y > a) = q^a$ (Exercise 3.71(a)).

(a) $P(Y > 10) = (0.88)^{10} = 0.2785$.

(b) Because $\{Y > 13\} \subset \{Y > 10\}$, the intersection is just $\{Y > 13\}$, and Definition 2.9 gives

$$P(Y > 13 \mid Y > 10) = \frac{q^{13}}{q^{10}} = q^3 = (0.88)^3 = 0.6815.$$

This is $P(Y > 3)$: exactly the probability that a **fresh** run of 3 declarations clears. That is the memoryless property, Exercise 3.71(b). The 10 clean declarations tell the exporter nothing about what comes next, because the trials are independent -- a run of luck does not make an inspection "due".

(c) Given $Y > 10$, the variable $Y - 10$ satisfies $P(Y - 10 = k \mid Y > 10) = q^{10+k-1}p/q^{10} = q^{k-1}p$, again a geometric distribution with the same p. By Theorem 3.8 its mean is $1/p = 1/0.12 = 8.3333$.

Problem 7

seed 20260925

A credit analyst models a BB-rated Caspian shipping company as defaulting in any given year with probability $p = 0.024$, provided it has not defaulted before, independently from year to year. Let Y be the number of the year in which the company first defaults (year 1 is the coming year).

(a) The company has just issued a 10-year bond. Find the probability that it survives to maturity without defaulting, $P(Y > 10) =$ ____________ (at least four significant figures)

(b) Find the largest integer y_0 such that $P(Y > y_0) \geq 0.65$. $y_0 =$ ____________

(c) Find the smallest integer n such that the probability of a default within the first n years is at least 0.45, that is $P(Y \leq n) \geq 0.45$. $n =$ ____________

Each year is a trial whose "success" is default, so Y is geometric with $p = 0.024$, $q = 0.976$, and $P(Y > y) = q^y$ (Exercise 3.71(a)).

(a) $P(Y > 10) = (0.976)^{10} = 0.7843$.

(b) $q^y \geq 0.65$. Take logarithms; $\ln q < 0$, so dividing by it reverses the inequality:

$$y \leq \frac{\ln 0.65}{\ln 0.976} = 17.7330.$$

The largest integer satisfying this is $y_0 = 17$. Check: $q^{17} = 0.6617 \geq 0.65$ but $q^{18} = 0.6458 < 0.65$.

(c) $P(Y \leq n) = 1 - q^n \geq 0.45$ means $q^n \leq 0.55$, i.e.

$$n \geq \frac{\ln 0.55}{\ln 0.976} = 24.6097,$$

so $n = 25$. Check: $1 - q^{25} = 0.4552 \geq 0.45$ while $1 - q^{24} = 0.4418 < 0.45$.

Parts (b) and (c) are the discrete analogue of a quantile: because Y takes only integer values, the inequality is solved in real numbers and then rounded in the direction the inequality allows.

Problem 8

seed 20260925

A manufacturing firm in Sumgait is moving its payroll account. Two banks, A and B, take turns presenting offers: bank A pitches in weeks 1, 3, 5, ... and bank B in weeks 2, 4, 6, ... Each pitch, whoever makes it, persuades the firm with probability $p = 0.170$, independently of earlier pitches, and the firm signs with the first bank whose pitch succeeds. Let Y be the week in which the account is won.

Give every probability to at least four significant figures.

(a) The probability that bank A wins the account, ______________

(b) Given that bank B wins the account, the probability that B won it with its **second** pitch (week 4),

____

(c) The probability that the account is still unsigned after 14 weeks, $P(Y > 14) =$ ______________

Weeks are independent trials with the same success probability, so Y is geometric (Definition 3.8) with $p = 0.170$, $q = 0.830$. Bank A wins exactly when Y is odd, bank B when Y is even.

(a) Sum the pmf over odd y; the terms form a geometric series with ratio q^2 (Exercise 3.77):

$$P(Y \text{ odd}) = p + q^2 p + q^4 p + \cdots = \frac{p}{1-q^2} = \frac{0.170}{1 - 0.688900} = 0.5464.$$

Moving first is an advantage: A wins with probability above one half, $1/(1+q)$ after simplifying, even though the two banks are equally persuasive.

(b) Bank B wins with probability $P(Y \text{ even}) = qp + q^3 p + \cdots = qp/(1-q^2) = 0.4536$, which is $1 - 0.5464$, as it must be. Since $\{Y = 4\} \subset \{Y \text{ even}\}$,

$$P(Y = 4 \mid Y \text{ even}) = \frac{q^3 p}{qp/(1-q^2)} = q^2(1-q^2) = \frac{0.09720}{0.4536} = 0.2143.$$

(c) The account is unsigned after 14 weeks exactly when all 14 pitches fail:
$P(Y > 14) = q^{14} = (0.830)^{14} = 0.0736$.

Problem 9

seed 20260925

An early-stage venture fund in Baku screens startup applications one at a time. Each screening costs the fund 800 AZN in analyst time, and a screened startup passes to the investment committee with probability $p = 0.08$, independently of the others. The fund stops at the first startup that passes and then commissions a legal and financial deep-dive on it costing 3000 AZN. Let Y be the number of startups screened and $T = 800\,Y + 3000$ the total cost of finding one investable startup.

Give every answer to at least four significant figures.

(a) $E(Y) =$ ______________

(b) $E(T) =$ ______________ AZN

(c) The standard deviation of T is ________________ AZN

(d) This quarter the fund can afford at most 14 screenings. The probability that none of them passes,

____

Y is the number of the trial on which the first success occurs: geometric with $p = 0.08$, $q = 0.92$.

(a) Theorem 3.8: $E(Y) = 1/p = 1/0.08 = 12.5000$.

(b) $T = aY + b$ with $a = 800$, $b = 3000$. By Exercise 3.33, $E(T) = aE(Y) + b$:

$$E(T) = 800(12.5000) + 3000 = 13000.00 \text{ AZN.}$$

(c) The constant b shifts T but does not spread it, so $V(T) = a^2V(Y)$. Theorem 3.8 gives $V(Y) = q/p^2 = 0.92/(0.08)^2 = 143.7500$, hence

$$\sigma_T = 800\sqrt{143.7500} = 800 \times 11.9896 = 9591.66 \text{ AZN.}$$

(d) "No pass in 14 screenings" is $Y > 14$: $P(Y > 14) = q^{14} = (0.92)^{14} = 0.3112$.

Notice how large σ_T is relative to the screening part of $E(T)$: search costs driven by a small success probability are very uncertain, which is exactly why a hard budget cap as in (d) leaves a real chance of ending the quarter with nothing.

Problem 10

seed 20260925

A microfinance lender's collections agent works through a randomly ordered list of overdue borrowers, phoning each once, and stops as soon as a borrower agrees to a restructured repayment plan. Model the number of the call on which the first agreement happens as geometric with an unknown agreement probability p. Today the first agreement came on call number $y_0 = 5$.

Give every answer to at least four significant figures.

(a) The maximum-likelihood estimate of p, $\hat{p} =$ ________________

(b) The value of the likelihood at the estimate, $L(\hat{p}) = P(Y = 5 \mid p = \hat{p}) =$ ________________

(c) The branch manager insists the agreement rate is really $p = 0.47$. Find the likelihood ratio $L(0.47)/L(\hat{p})$.

____

(d) Using $\hat{p}$ as if it were the true value, estimate the probability that tomorrow the agent needs more than 4 calls to get the first agreement. ________________

(a) The likelihood of the observation is $L(p) = P(Y = 5) = (1 - p)^4 p$. As in Example 3.13, maximize its logarithm:

$$\frac{d}{dp}[4\ln(1-p) + \ln p] = -\frac{4}{1-p} + \frac{1}{p} = 0 \quad \Longrightarrow \quad p = \frac{1}{5}.$$

The second derivative $-4/(1-p)^2 - 1/p^2$ is negative, so this is a maximum (Exercise 3.86: in general $\hat{p} = 1/y_0$). Thus $\hat{p} = 1/5 = 0.2000$.

(b) $L(\hat{p}) = (0.8000)^4(0.2000) = 0.08192$.

(c) $L(0.47) = (0.53)^4(0.47) = 0.03709$, so the ratio is $0.03709/0.08192 = 0.4527$. It can never exceed 1, because $\hat{p}$ maximizes L; the further below 1 it is, the less well the manager's figure explains what was actually observed. (One observation is thin evidence, though -- Chapter 9 makes that precise.)

(d) $P(Y > 4) = \hat{q}^4 = (0.8000)^4 = 0.4096$.

Problem 11

seed 20260925

A Baku bank is arranging a syndicated loan for a cement producer and needs 2 other banks to take a participation. It approaches banks one at a time from a long list; each bank approached agrees to join with probability $p = 0.575$, independently of the others. Let Y be the number of the approach on which the 2nd participant signs up.

Give every answer to at least four significant figures.

(a) $P(Y = 5) =$ ______________

(b) The probability that the syndicate is complete within 5 approaches, $P(Y \le 5) =$ ______________

(c) $E(Y) =$ ______________

Y is negative binomial (Definition 3.9) with $r = 2$, $p = 0.575$, $q = 0.425$: $p(y) = \binom{y-1}{1} p^2 q^{y-2}$, $y = 2, 2+1, \ldots$.

(a) $P(Y = 5) = \binom{4}{1}(0.575)^2(0.425)^3 = 4 \times 0.330625 \times 0.076766 = 0.1015$.

(b) Add the pmf from $y = 2$ to $y = 5$: p(2) = 0.3306, p(3) = 0.2810, p(4) = 0.1792, p(5) = 0.1015, which sum to $P(Y \le 5) = 0.8923$.

A useful check uses the binomial instead. The 2nd success has arrived by approach 5 exactly when the first 5 approaches contain at least 2 successes. With $X \sim \text{Bin}(5, 0.575)$ (Definition 3.7),

$$P(Y \le 5) = P(X \ge 2) = 1 - \sum_{x=0}^{1} \binom{5}{x} p^x q^{5-x} = 1 - 0.1077 = 0.8923.$$

The two framings describe the same event: the negative binomial fixes the number of successes and lets the number of trials vary; the binomial fixes the trials and counts successes.

(c) Theorem 3.9: $E(Y) = r/p = 2/0.575 = 3.4783$ approaches.

Problem 12

seed 20260925

The SME desk of a Baku bank approves each small-business loan application with probability $p = 0.26$, independently of all other applications, and reviews applications in the order they arrive. Two questions are asked about the stream of decisions starting tomorrow morning.

Give every probability to at least four significant figures.

(a) The event is: the third approval is application number 14.

Which distribution describes the random variable that this event is about?

- ○ Binomial (Definition 3.7)
- ○ Geometric (Definition 3.8)
- ○ Negative binomial (Definition 3.9)

Its probability is ____________

(b) The event is: exactly 3 of the next 14 applications are approved.

Which distribution describes the random variable that this event is about?

- ○ Binomial (Definition 3.7)
- ○ Geometric (Definition 3.8)
- ○ Negative binomial (Definition 3.9)

Its probability is ____________

All three distributions come from the same independent trials with $p = 0.26$, $q = 0.74$. What distinguishes them is **what is counted and what is fixed**.

(a) Negative binomial (Definition 3.9): the random quantity is the number of the application on which the third approval occurs, so the count of successes (3) is fixed and the number of trials varies. Probability 0.0500.

(b) Binomial (Definition 3.7): the number of applications (14) is fixed in advance and the random quantity is how many are approved. Probability 0.2331.

The three possible computations for these numbers are:

- Binomial, X = approvals among 14 applications: $P(X = 3) = \binom{14}{3} p^3 q^{11} = 364(0.26)^3(0.74)^{11} = 0.2331$.
- Geometric, Y = number of the first approved application: $P(Y = 3) = q^2 p = (0.74)^2(0.26) = 0.1424$.
- Negative binomial with $r = 3$, Y = number of the application with the third approval: $P(Y = 14) = \binom{13}{2} p^3 q^{11} = 78(0.26)^3(0.74)^{11} = 0.0500$.

The binomial and negative binomial probabilities differ only in the coefficient. “Exactly 3 approvals in 14” allows the last application to be a rejection; “the third approval is application 14” forces the last one to be an approval, and $\binom{13}{2}/\binom{14}{3} = 3/14 = 0.2143$.

Problem 13

seed 20260925

A fixed-income trading desk has a daily loss limit. The risk department models each trading day as breaching the limit with probability $p = 0.035$, independently of every other day. Let Y be the number of the trading day on which the first breach occurs, counting from the start of the quarter. The desk has now completed 19 trading days without a breach.

Give every answer to at least four significant figures.

(a) The probability that the desk also gets through the next 6 days without a breach, $P(Y > 25 \mid Y > 19) =$

____

(b) The probability that the first breach is on day 24, $P(Y = 24 \mid Y > 19) =$ ________________

(c) $E(Y \mid Y > 19) =$ ________________

(d) For comparison, the probability computed at the **start** of the quarter that the first 25 days would all be clean, $P(Y > 25) =$ ________________

Y is geometric with $p = 0.035$, $q = 0.965$, and $P(Y > a) = q^a$.

(a) $\{Y > 25\} \subset \{Y > 19\}$, so

$$P(Y > 25 \mid Y > 19) = \frac{P(Y > 25)}{P(Y > 19)} = \frac{q^{25}}{q^{19}} = q^6 = 0.8075.$$

This is the memoryless property (Exercise 3.71(b)): the answer is $P(Y > 6)$, whatever a is.

(b) For $y > 19$, $P(Y = y \mid Y > 19) = q^{y-1}p/q^{19} = q^{y-19-1}p$. At $y = 24$:

$$P(Y = 24 \mid Y > 19) = q^4 p = (0.965)^4(0.035) = 0.0304.$$

(c) Part (b) shows that, given $Y > 19$, the variable $W = Y - 19$ has pmf $q^{w-1}p$, $w = 1, 2, \ldots$ -- geometric with the same p. By Theorem 3.8 $E(W \mid Y > 19) = 1/p = 28.5714$, so

$$E(Y \mid Y > 19) = 19 + \frac{1}{p} = 19 + 28.5714 = 47.5714.$$

The expected **remaining** wait is exactly what it was on day one.

(d) $P(Y > 25) = q^{25} = 0.4104$, smaller than (a) by the factor $q^{19} = 0.5082$. Before the quarter began, 25 clean days had to include the first 19; once those are banked, only 6 remain at risk. A risk committee that treats a long clean run as evidence that a breach is “overdue” is contradicting its own model.

Problem 14

seed 20260925

A payments company is rolling out card terminals to shops in a new Baku shopping district, and a sales representative must sign up 3 merchants this week. She visits shops chosen at random; each visit signs the merchant with probability $p = 0.48$, independently. A visit that ends in a refusal takes 1.25 hours; a visit that ends in a signing takes 4.25 hours, because the contract and terminal installation are done on the spot. Let Y be the number of shops visited until the 3rd merchant signs, and T the total hours spent.

Give every answer to at least four significant figures.

(a) $E(Y) =$ ____________

(b) $V(Y) =$ ____________

(c) $E(T) =$ ____________ hours

(d) The standard deviation of T is ____________ hours

Y is negative binomial with $r = 3$ and $p = 0.48$ (Definition 3.9).

(a) Theorem 3.9: $E(Y) = r/p = 3/0.48 = 6.2500$.

(b) Theorem 3.9: $V(Y) = rq/p^2 = 3(0.52)/(0.48)^2 = 6.7708$.

(c) Of the Y visits, exactly 3 end in a signing and $Y - 3$ in a refusal. So

$$T = 1.25(Y - 3) + 4.25(3) = 1.25\,Y + (4.25 - 1.25)(3) = 1.25\,Y + 9.00.$$

As in Example 3.15, T is a linear function $aY + b$ of Y, with the extra time for the signing visits entering only as a constant. By Exercise 3.33,

$$E(T) = 1.25(6.2500) + 9.00 = 16.8125 \text{ hours.}$$

(d) $V(T) = a^2 V(Y) = (1.25)^2(6.7708) = 10.5794$, so $\sigma_T = 3.2526$ hours. The constant 9.00 does not affect the spread: every possible week contains exactly 3 signing visits, so all of the uncertainty comes from the number of refusals.

Problem 15

seed 20260925

An importer must clear two shipments through the electronic customs portal, one after the other: she submits the declaration for shipment 1 until it is accepted, then starts on shipment 2. Each submission is accepted with probability $p = 0.40$ and otherwise bounced back for a technical correction; bounces happen independently from submission to submission. Let Y_1 and Y_2 be the numbers of submissions needed for the two shipments, and $T = Y_1 + Y_2$ the total number of submissions.

Give every answer to at least four significant figures.

(a) $P(T = 8) =$ ____________

(b) $V(T) =$ ______________

(c) $P(T \le 5) =$ ______________

Y_1 and Y_2 are independent geometric variables with $p = 0.40$, $q = 0.60$.

(a) Split on the value of Y_1. For $T = 8$ we need $Y_1 = k$ and $Y_2 = 8 - k$ for some $k = 1, \ldots, 7$:

$$P(T = 8) = \sum_{k=1}^{7} q^{k-1} p \cdot q^{8-k-1} p = \sum_{k=1}^{7} p^2 q^6 = 7\,p^2 q^6 = 0.0523.$$

This is exactly Definition 3.9 with $r = 2$: $\binom{7}{1} p^2 q^6$. Of course: the submissions form one unbroken stream of independent trials, and T is the number of the trial on which the **second** acceptance occurs.

(b) Since T is negative binomial with $r = 2$, Theorem 3.9 gives $E(T) = 2/p = 5.0000$ and

$$V(T) = \frac{2q}{p^2} = \frac{2(0.60)}{(0.40)^2} = 7.5000,$$

which is twice the geometric variance of Theorem 3.8 -- as it should be for a sum of two independent copies (Chapter 5 will prove the general rule).

(c) $T \le 5$ means at least two acceptances among the first 5 submissions. The complement is zero or one acceptance in 5 trials, a binomial calculation:

$$P(T \le 5) = 1 - q^5 - 5\,p\,q^4 = 1 - 0.0778 - 0.2592 = 0.6630.$$

Problem 16

seed 20260925

A research team studying tax compliance interviews randomly selected shop owners and asks: “In the last year, have you recorded any cash sale below its true value to reduce VAT?” Suppose a fraction 0.14 of owners have done so, and that 0.80 of **those** owners deny it. The remaining 0.86 of owners have not, and all of them truthfully answer no. Let Y be the number of interviews up to and including the first “yes”.

Give every answer to at least four significant figures.

(a) The probability p that a single interview produces a “yes”, $p =$ ______________

(b) $P(Y = 3) =$ ______________

(c) $E(Y) =$ ______________

(d) The probability that 19 interviews produce no “yes” at all, ______________

(a) A “yes” requires an owner who has underreported **and** who admits it. With U = “has underreported” and A = “answers yes”, the multiplicative law gives

$$p = P(U \cap A) = P(U)P(A \mid U) = 0.14 \times 0.20 = 0.0280.$$

Owners who have not underreported contribute nothing, since $P(A \mid \bar{U}) = 0$.

The owners are randomly selected, so interviews are independent trials with this constant p, and Y is geometric (Definition 3.8) with $q = 1 - p = 0.9720$:

$$p(y) = (0.9720)^{y-1}(0.0280), \qquad y = 1, 2, \ldots$$

(b) $P(Y = 3) = q^2 p = 0.0265$.

(c) Theorem 3.8: $E(Y) = 1/p = 35.7143$ interviews.

(d) $P(Y > 19) = q^{19} = 0.5830$.

The point for survey design: denial shrinks the success probability from 0.14 to 0.0280, and the expected search length grows by the factor $1/0.20$. A naive reading of "the first yes came on interview y" through Example 3.13's estimate $1/y$ would estimate p, the **admission** rate, not the underreporting rate itself; recovering that needs the denial rate as well.

Problem 17

seed 20260925

A factoring company has bought an overdue invoice from a Baku construction supplier. Each week, independently, the debtor pays in full with probability $p = 0.16$. Let Y be the number of the week in which payment arrives. The factor's cost of carrying the invoice is

$$C = 190\,Y + 15\,Y^2 \text{ AZN},$$

where the quadratic term reflects provisioning rules that bite harder the longer an invoice stays unpaid.

Give every answer to at least four significant figures.

(a) $E[Y(Y-1)] =$ ________________

(b) $V(Y) =$ ________________ (compute it from (a); do not just quote it)

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

(d) An analyst estimates the expected cost by plugging $E(Y)$ into the formula, $190\,E(Y) + 15\,[E(Y)]^2$. By how much does this **understate** $E(C)$? ________________ AZN

Y is geometric with $p = 0.16$, $q = 0.84$.

(a) Following Exercise 3.85, factor out pq so that what remains is a second derivative:

$$E[Y(Y-1)] = \sum_{y=1}^{\infty} y(y-1)q^{y-1}p = pq\sum_{y=2}^{\infty} y(y-1)q^{y-2} = pq\,\frac{d^2}{dq^2}\left(\sum_{y=1}^{\infty} q^y\right).$$

The inner sum is $q/(1-q)$; its first derivative is $1/(1-q)^2$ and its second is $2/(1-q)^3 = 2/p^3$. Hence

$$E[Y(Y-1)] = pq \cdot \frac{2}{p^3} = \frac{2q}{p^2} = \frac{2(0.84)}{(0.16)^2} = 65.6250.$$

(b) $E(Y^2) = E[Y(Y-1)] + E(Y) = 65.6250 + 6.2500 = 71.8750$, using $E(Y) = 1/p = 6.2500$ (Theorem 3.8). Then

$$V(Y) = E(Y^2) - [E(Y)]^2 = 71.8750 - 39.0625 = 32.8125,$$

which is q/p^2, the variance in Theorem 3.8.

(c) By linearity of expectation (Theorems 3.5 and 3.6),

$$E(C) = 190E(Y) + 15E(Y^2) = 190(6.2500) + 15(71.8750) = 2265.62 \text{ AZN}.$$

(d) The plug-in figure is $190(6.2500) + 15(39.0625) = 1773.44$, so the shortfall is

$$E(C) - \text{plug-in} = 15\left(E(Y^2) - [E(Y)]^2\right) = 15\,V(Y) = 492.19 \text{ AZN}.$$

For any convex cost, replacing a random waiting time by its mean understates the expected cost, and here the error is exactly the quadratic coefficient times the variance. With a geometric wait the variance is large, so the error is not a rounding matter.

Problem 18

seed 20260925

A bank's corporate sales team pitches its payroll service to firms taken at random from among its existing loan customers. Each pitch, independently, persuades the firm to move its payroll account with probability $p = 0.72$. The team's quarterly target is 3 new payroll clients. Let Y be the number of the pitch on which the 3rd client signs, and $p(y) = P(Y = y)$.

Give probabilities and ratios to at least four significant figures.

(a) $p(6)/p(5) =$ ________________

(b) The most likely value of Y (the value y with the largest $p(y)$) is ________________

(c) $P(Y = y)$ at that most likely value is ________________

Y is negative binomial (Definition 3.9) with $r = 3$, $p = 0.72$, $q = 0.28$.

(a) For $y \geq r + 1$, the powers of p cancel and one factor of q survives (Exercise 3.98(a)):

$$\frac{p(y)}{p(y-1)} = \frac{\binom{y-1}{r-1}p^r q^{y-r}}{\binom{y-2}{r-1}p^r q^{y-1-r}} = \frac{(y-1)!\,(y-r-1)!}{(y-2)!\,(y-r)!}\,q = \frac{y-1}{y-r}\,q.$$

At $y = 6$: $\frac{5}{3}(0.28) = 0.4667$.

(b) The ratio exceeds 1 exactly when $(y-1)q > y - r$, i.e. $y(1-q) < r - q$, i.e.

$$y < \frac{r-q}{1-q} = \frac{3-0.28}{0.72} = 3.7778.$$

(Exercise 3.98(b)). So $p(y)$ increases while $y < 3.7778$ and decreases afterwards; the threshold is not an integer, so there is no tie. The largest probability is at the largest integer below it, $y = 3$.

(c) $p(3) = \binom{2}{2}(0.72)^3(0.28)^0 = 1\,(0.72)^3(0.28)^0 = 0.37325.$

Check against the neighbours: $p(2) = 0.00000$ and $p(4) = 0.31353$ are both smaller. Here the mode is r itself, and its left neighbour has probability 0 because fewer than 3 pitches cannot produce 3 clients.

Note that the mode is not the mean r/p: the negative binomial is skewed to the right, so its peak sits below its mean.

Problem 19

seed 20260925

An exploration consortium holds a licence block on the Caspian shelf. Each exploratory well costs 38 million AZN and, independently of the others, finds a commercial deposit with probability $p = 0.18$. A discovery is worth 317 million AZN (net present value of development, before exploration costs). The consortium drills wells one at a time and stops at the first discovery, but its board has approved at most 5 wells. Let N be the number of wells actually drilled, so $N = \min(Y, 5)$ where Y is geometric.

Give every answer to at least four significant figures.

(a) The probability that the programme finds a deposit, ____________

(b) $E(N) =$ ____________

(c) The expected net value of the programme (value of any discovery minus all drilling costs), ____________ million AZN. (Enter a negative number if it is negative.)

(d) Decision:

- ○ Launch the programme: its expected net value is positive
- ○ Do not launch: its expected net value is negative

Y is geometric with $p = 0.18$, $q = 0.82$: the number of the well that would make the first discovery if drilling were unlimited.

(a) A deposit is found exactly when $Y \le 5$: $P(Y \le 5) = 1 - q^5 = 1 - 0.3707 = 0.6293$.

(b) For $k = 0, 1, \ldots, 4$, $N > k$ happens exactly when the first k wells are dry, so $P(N > k) = q^k$; and N never exceeds 5. Writing $N = \sum_{k=0}^{4} I(N > k)$ (one indicator for each well that gets drilled) and taking expectations,

$$E(N) = \sum_{k=0}^{4} q^k = \frac{1 - q^5}{1 - q} = \frac{1 - q^5}{p} = \frac{0.6293}{0.18} = 3.4959.$$

(The direct sum $\sum_{y=1}^{4} y\, q^{y-1} p + 5\, q^4$ gives the same number.)

(c) The value 317 is earned with the probability from (a), and every drilled well costs 38:

$$E(\text{net}) = 317(0.6293) - 38(3.4959) = 66.63 \text{ million AZN}.$$

(d) Factor the answer to (c):

$$E(\text{net}) = (1 - q^5)\left(V - \frac{c}{p}\right) = 0.6293\,(317 - 211.11).$$

The sign is the sign of $V - c/p$, equivalently of $pV - c = 57.06 - 38 = 19.06$, the expected value of drilling one more well. Here $pV > c$, so the expected net value is positive, so the programme should be launched.

Two lessons. First, $c/p = c\,E(Y)$ is the expected exploration cost **per discovery**: the search pays if and only if a discovery is worth more than that. Second, the cap 5 changes the **size** of the expected gain or loss but not its sign -- by memorylessness every well, whenever it is drilled, is the same gamble with expected value $pV - c$.

Problem 20

seed 20260925

An investment bank is building the order book for a Baku Stock Exchange IPO and needs 3 anchor investors. It approaches institutional funds one at a time; each fund, independently, commits with probability $p = 0.48$. Let Y be the number of the approach on which the 3rd commitment arrives, and let X_n be the number of commitments among the first n approaches.

Give every answer to at least four significant figures.

(a) $P(Y = 9) =$ ______________

(b) The probability that the anchor book is complete within 9 approaches, $P(Y \leq 9) =$ ______________

(c) After 9 approaches exactly 3 funds have committed. Given only this, the probability that the 3rd commitment was from the 9th fund approached, $P(Y = 9 \mid X_9 = 3) =$ ______________

(d) Given that $Y = 12$, the probability that the very first fund approached committed, ______________

Y is negative binomial (Definition 3.9) with $r = 3$, $p = 0.48$, $q = 0.52$; X_n is binomial with parameters n and p.

(a) $P(Y = 9) = \binom{8}{2} p^3 q^6 = 28(0.48)^3(0.52)^6 = 0.0612.$

(b) $Y \leq 9$ is the same event as $X_9 \geq 3$: the 3rd commitment has arrived by approach 9 exactly when the first 9 approaches contain at least 3 commitments. So, summing only 3 binomial terms instead of 6+1 negative binomial ones,

$$P(Y \leq 9) = 1 - \sum_{x=0}^{2} \binom{9}{x} p^x q^{9-x} = 1 - 0.1111 = 0.8889.$$

(c) $\{Y = 9\} \subset \{X_9 = 3\}$, so by Definition 2.9

$$P(Y = 9 \mid X_9 = 3) = \frac{\binom{8}{2} p^3 q^6}{\binom{9}{3} p^3 q^6} = \frac{28}{84} = \frac{3}{9} = 0.3333.$$

Here $P(X_9 = 3) = 0.1837$, but the powers of p and q cancel: given that 3 of the 9 approaches succeeded, every arrangement of those successes is equally likely, and the last approach is one of them with probability r/n, whatever p is.

(d) The event "first fund commits and $Y = 12$" requires a success on approach 1, exactly 1 successes among approaches 2 to 11, and a success on approach 12:

$$P(\text{first commits} \mid Y = 12) = \frac{p \cdot \binom{10}{1} p^1 q^9 \cdot p}{\binom{11}{2} p^3 q^9} = \frac{10}{55} = \frac{2}{11} = 0.1818.$$

Again p drops out: given $Y = 12$, the other 2 successes are spread uniformly over the first 11 approaches.

Answer key

Problem	Answers
1	(a) 0.0465085, (b) 0.529747, (c) 11.1111
2	(a) 0.395292, (b) 0.780749, (c) 0.105186
3	(a) 52.6316, (b) 52.1292, (c) 0.863988, (d) 0.41557
4	(a) 0.166522, (b) 3.22581, (c) 1.22581
5	(a) 10.5263, (b) 17.1745, (c) 36.8421, (d) 14.5048
6	(a) 0.278501, (b) 0.681472, (c) 8.33333
7	(a) 0.784329, (b) 17, (c) 25
8	(a) 0.546448, (b) 0.214317, (c) 0.0736365
9	(a) 12.5, (b) 13000, (c) 9591.66, (d) 0.311193
10	(a) 0.2, (b) 0.08192, (c) 0.452701, (d) 0.4096
11	(a) 0.101523, (b) 0.892336, (c) 3.47826
12	(a) Negative binomial ... 9), (b) 0.0499532, (c) Binomial (Definition ... .7), (d) 0.233115
13	(a) 0.80754, (b) 0.0303513, (c) 47.5714, (d) 0.410377
14	(a) 6.25, (b) 6.77083, (c) 16.8125, (d) 3.2526
15	(a) 0.0522547, (b) 7.5, (c) 0.66304
16	(a) 0.028, (b) 0.026454, (c) 35.7143, (d) 0.582987
17	(a) 65.625, (b) 32.8125, (c) 2265.62, (d) 492.188
18	(a) 0.466667, (b) 3, (c) 0.373248
19	(a) 0.62926, (b) 3.49589, (c) 66.6317, (d) Launch ... positive
20	(a) 0.0612212, (b) 0.888853, (c) 0.333333, (d) 0.181818