Discrete random variables: expectation, variance and the binomial, geometric and Poisson

Ten problems on discrete random variables: expectation and variance from a probability table, a fair die and Var(aX + b), the binomial probabilities and its mean and variance from indicators, linearity of expectation without independence in the matching problem, the geometric distribution and memorylessness, Poisson counts and the Poisson limit of the binomial, the laws of total expectation and variance, and the expected number of distinct examples in a bootstrap sample, with worked solutions and the mistakes of squaring the mean, dropping the binomial coefficient and adding probabilities that overlap.

BEFORE YOU START

A discrete random variable is a number that comes out of a random experiment: the count of correct predictions in a batch, the number of retries before a request succeeds, the number of events in a minute. Its distribution is a list of values and their probabilities, and almost everything asked about it is a weighted sum over that list. Two facts do most of the work. Expectation is linear, $E[X+Y]=E[X]+E[Y]$, whether or not X and Y are independent, which turns hard counts into sums of easy indicators. Variance is not linear, and the place it goes wrong is the place the squares are. These ten problems build the binomial, geometric and Poisson distributions from first principles, and end with the expected number of distinct examples in a bootstrap sample, the 63.2% behind bagging's out-of-bag estimate.

- A discrete random variable X has a probability mass function $p(x)=P(X=x)$ with $p(x)\ge 0$ and $\sum_x p(x)=1$.
- **Expectation.** $E[X]=\sum_x x\,p(x)$, and for any function g, $E[g(X)]=\sum_x g(x)\,p(x)$. In general $E[g(X)]\neq g(E[X])$; in particular $E[X^2]\neq (E[X])^2$.
- **Linearity.** $E[aX+b]=aE[X]+b$, and $E[X+Y]=E[X]+E[Y]$ for **any** X and Y, independent or not.
- **Variance.** $\operatorname{Var}(X)=E[(X-E[X])^2]=E[X^2]-(E[X])^2$, and the standard deviation is its square root. $\operatorname{Var}(aX+b)=a^2\operatorname{Var}(X)$. For independent X and Y, $\operatorname{Var}(X+Y)=\operatorname{Var}X+\operatorname{Var}Y$; otherwise add $2\operatorname{Cov}(X,Y)$.
- **Indicators.** For an event A, $I_A=1$ if A happens and 0 otherwise. $E[I_A]=P(A)$ and $I_A^2=I_A$.
- **Binomial**(n,p): the number of successes in n independent trials with success probability p. $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ for $k=0,\dots,n$, where $\binom{n}{k}=\frac{n!}{k!\,(n-k)!}$ counts the orders in which the k successes can occur.

- **Geometric**(p): the number of the trial on which the first success happens, $P(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, \ldots$ Some books count the failures before it instead, which shifts everything by 1.
- **Poisson**(λ): $P(X = k) = e^{-\lambda}\lambda^k / k!$ for $k = 0, 1, 2, \ldots$, a count of events that occur independently at an average rate λ per interval. The series $\sum_{k \ge 0} \lambda^k / k! = e^{\lambda}$ makes the probabilities sum to 1.
- **Conditioning.** $E[X] = E\big[E[X \mid Y]\big]$ and $\operatorname{Var}(X) = E\big[\operatorname{Var}(X \mid Y)\big] + \operatorname{Var}\big(E[X \mid Y]\big)$.

Problems

1. X takes the values $0, 1, 2, 3$ with probabilities 0.1, 0.3, 0.4 and 0.2. Find $E[X]$, $E[X^2]$, $\operatorname{Var}(X)$ and the standard deviation. `difficulty ·`

2. X is the roll of a fair six-sided die. Find $E[X]$, $E[X^2]$ and $\operatorname{Var}(X)$, and then $E[2X + 3]$ and $\operatorname{Var}(2X + 3)$. `difficulty ·`

3. A classifier labels 10 independent examples, each correctly with probability 0.8. Find the probability that exactly 8 are correct, the probability that at least 9 are, and the mean and variance of the number correct. `difficulty ··`

4. Show that a Binomial(n, p) variable has mean np and variance $np(1-p)$ by writing it as a sum of indicators. `difficulty ··`

5. $n \ge 2$ letters are put into n addressed envelopes in a uniformly random order. Let X be the number that land in the right envelope. Find $E[X]$ and $\operatorname{Var}(X)$. `difficulty ··`

6. A request succeeds on each attempt with probability 0.2, independently, and is retried until it succeeds. Let X be the number of attempts. Find $E[X]$, $P(X > 3)$, the probability of success within 3 attempts, and $P(X > 5 \mid X > 2)$. `difficulty ··`

7. Errors arrive in a log at an average of 3 per minute, as a Poisson count. Find the probability of no errors in a minute, of at most 2, and of at least 3, and show from the series that $E[X] = \lambda$. `difficulty ··`

8. A rare label occurs with probability 0.002 per example, independently, in a batch of 1000. Compare the exact binomial probabilities of 0 and of 3 occurrences with the Poisson approximation, and say why the approximation works. difficulty ···

9. A fair die is rolled, and if it shows K, then K fair coins are flipped. Let X be the number of heads. Find $E[X]$ and $\mathrm{Var}(X)$. difficulty ···

10. A bootstrap sample draws n examples uniformly at random, with replacement, from a training set of n. Find the expected number of distinct examples in it, evaluate it for $n = 10$, and find the fraction it approaches as n grows. difficulty ···

Answers

1. $E[X] = 1.7$, $E[X^2] = 3.7$, $\operatorname{Var}(X) = 0.81$, standard deviation 0.9
2. $E[X] = 3.5$, $E[X^2] = \frac{91}{6}$, $\operatorname{Var}(X) = \frac{35}{12}$; $E[2X + 3] = 10$, $\operatorname{Var}(2X + 3) = \frac{35}{3} \approx 11.67$
3. $P(X = 8) \approx 0.302$, $P(X \geq 9) \approx 0.376$, $E[X] = 8$, $\operatorname{Var}(X) = 1.6$
4. $E[X] = np$, $\operatorname{Var}(X) = np(1 - p)$
5. $E[X] = 1$ and $\operatorname{Var}(X) = 2 - 1^2 = 1$, for every $n \geq 2$
6. $E[X] = 5$, $P(X > 3) = 0.512$, $P(X \leq 3) = 0.488$, $P(X > 5 \mid X > 2) = 0.512$
7. $P(0) \approx 0.0498$, $P(X \leq 2) \approx 0.4232$, $P(X \geq 3) \approx 0.5768$, $E[X] = 3$
8. Exact 0.13506 and 0.18063; Poisson 0.13534 and 0.18045
9. $E[X] = 1.75$ and $\operatorname{Var}(X) = \frac{7}{8} + \frac{35}{48} = \frac{77}{48} \approx 1.604$
10. $E[D] = n\big(1 - (1 - \frac{1}{n})^n\big)$; 6.513 for $n = 10$; the fraction tends to $1 - e^{-1} \approx 0.632$

Worked solutions

Problem 1

> X takes the values $0, 1, 2, 3$ with probabilities 0.1, 0.3, 0.4 and 0.2. Find $E[X]$, $E[X^2]$, $\mathrm{Var}(X)$ and the standard deviation.

$0.1 + 0.3 + 0.4 + 0.2 = 1$	Check the table is a distribution before using it.
$E[X] = 0(0.1) + 1(0.3) + 2(0.4) + 3(0.2) = 1.7$	Each value weighted by its probability.
$E[X^2] = 0(0.1) + 1(0.3) + 4(0.4) + 9(0.2) = 3.7$	Square the values, not the probabilities, and keep the same weights.
$\mathrm{Var}(X) = 3.7 - 1.7^2 = 3.7 - 2.89 = 0.81$	The shortcut $E[X^2] - (E[X])^2$.
$E[X] = 1.7$, $E[X^2] = 3.7$, $\mathrm{Var}(X) = 0.81$, **standard deviation** 0.9	Check from the definition: $0.1(1.7)^2 + 0.3(0.7)^2 + 0.4(0.3)^2 + 0.2(1.3)^2 = 0.289 + 0.147 + 0.036 + 0.338 = 0.81$.

Problem 2

> X is the roll of a fair six-sided die. Find $E[X]$, $E[X^2]$ and $\mathrm{Var}(X)$, and then $E[2X + 3]$ and $\mathrm{Var}(2X + 3)$.

$E[X] = \dfrac{1 + 2 + \cdots + 6}{6} = \dfrac{21}{6} = 3.5$	Each face has probability $\frac{1}{6}$.
$E[X^2] = \dfrac{1 + 4 + 9 + 16 + 25 + 36}{6} = \dfrac{91}{6}$	The same weights on the squares.
$\mathrm{Var}(X) = \dfrac{91}{6} - \dfrac{49}{4} = \dfrac{182 - 147}{12} = \dfrac{35}{12}$	About 2.917.

$E[2X+3] = 2(3.5)+3 = 10$ and $\mathrm{Var}(2X+3) = 4 \cdot \dfrac{35}{12} = \dfrac{35}{3}$	Shifting by 3 moves every value and the mean together, so the spread is unchanged; scaling by 2 doubles every deviation and so multiplies their squares by 4.
$E[X] = 3.5$, $E[X^2] = \frac{91}{6}$, $\mathrm{Var}(X) = \frac{35}{12}$; $E[2X+3] = 10$, $\mathrm{Var}(2X+3) = \frac{35}{3} \approx 11.67$	Check: $2X+3$ takes the values $5, 7, \dots, 15$, and the mean of their squares minus 10^2 is $\frac{25+49+81+121+169+225}{6} - 100 = \frac{670}{6} - 100 = \frac{35}{3}$.

Problem 3

> A classifier labels 10 independent examples, each correctly with probability 0.8. Find the probability that exactly 8 are correct, the probability that at least 9 are, and the mean and variance of the number correct.

$X \sim \mathrm{Binomial}(10, 0.8)$	A fixed number of independent trials with the same success probability.
$P(X=8) = \binom{10}{8}(0.8)^8(0.2)^2 = 45 \cdot 0.16777 \cdot 0.04 \approx 0.3020$	$\binom{10}{8} = \binom{10}{2} = 45$: the two errors can be any two of the ten examples.
$P(X \ge 9) = \binom{10}{9}(0.8)^9(0.2) + (0.8)^{10} \approx 0.2684 + 0.1074 = 0.3758$	The events $X=9$ and $X=10$ are disjoint, so their probabilities add.
$E[X] = np = 8$ and $\mathrm{Var}(X) = np(1-p) = 10 \cdot 0.8 \cdot 0.2 = 1.6$	Problem 4 derives both.
$P(X=8) \approx 0.302$, $P(X \ge 9) \approx 0.376$, $E[X] = 8$, $\mathrm{Var}(X) = 1.6$	Even at the mean, 8 correct happens less than a third of the time; the standard deviation is $\sqrt{1.6} \approx 1.26$ examples.

Problem 4

> Show that a Binomial(n, p) variable has mean np and variance $np(1-p)$ by writing it as a sum of indicators.

$X = I_1 + I_2 + \cdots + I_n$, where $I_j = 1$ if trial j succeeds	Counting successes is adding a 1 for each one.
$E[I_j] = 1 \cdot p + 0 \cdot (1-p) = p$, so $E[X] = np$	Linearity of expectation; no independence needed for this step.
$E[I_j^2] = E[I_j] = p$, so $\mathrm{Var}(I_j) = p - p^2 = p(1-p)$	An indicator is 0 or 1, so it equals its square.
$\mathrm{Var}(X) = \sum_j \mathrm{Var}(I_j) = np(1-p)$	The trials are independent, so the covariances are 0 and the variances add. This step does use independence.
$E[X] = np$, $\mathrm{Var}(X) = np(1-p)$	The variance is largest at $p = \frac{1}{2}$ and is 0 at $p = 0$ or 1, when the outcome is certain.

Problem 5

> $n \geq 2$ letters are put into n addressed envelopes in a uniformly random order. Let X be the number that land in the right envelope. Find $E[X]$ and $\mathrm{Var}(X)$.

$X = I_1 + \cdots + I_n$ with $I_j = 1$ if letter j is in its own envelope, and $P(I_j = 1) = \frac{1}{n}$	Each letter is equally likely to be in any of the n envelopes.
$E[X] = n \cdot \frac{1}{n} = 1$	Linearity holds even though the I_j are dependent: if $n-1$ letters are right, the last one must be.
$E[I_i I_j] = P(\text{both right}) = \dfrac{1}{n} \cdot \dfrac{1}{n-1}$ for $i \neq j$	Given letter i is right, letter j has $n-1$ envelopes left.

$E[X^2] = \sum_j E[I_j^2] + \sum_{i \neq j} E[I_i I_j] = n \cdot \frac{1}{n} + n(n-1) \cdot \frac{1}{n(n-1)} = 2$	Expand the square of the sum: n diagonal terms and $n(n-1)$ ordered pairs.
$E[X] = 1$ **and** $\mathrm{Var}(X) = 2 - 1^2 = 1$, **for every** $n \geq 2$	Each pair has $\mathrm{Cov}(I_i, I_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} > 0$, and the $n(n-1)$ of them add exactly the $\frac{1}{n}$ that the independent formula $n \cdot \frac{1}{n}(1 - \frac{1}{n})$ misses (Mistake 5).

Problem 6

A request succeeds on each attempt with probability 0.2, independently, and is retried until it succeeds. Let X be the number of attempts. Find $E[X]$, $P(X > 3)$, the probability of success within 3 attempts, and $P(X > 5 \mid X > 2)$.

$P(X = k) = 0.8^{k-1}(0.2)$, so $E[X] = \sum_{k \geq 1} k\,(0.8)^{k-1}(0.2) = \dfrac{0.2}{(1-0.8)^2} = 5$	$\sum_{k \geq 1} k q^{k-1} = \frac{1}{(1-q)^2}$ is the derivative of the geometric series $\sum q^k = \frac{1}{1-q}$. In general $E[X] = \frac{1}{p}$.
$P(X > 3) = 0.8^3 = 0.512$	$X > 3$ means the first three attempts all failed.
$P(X \leq 3) = 1 - 0.512 = 0.488$	The complement: at least one success in three attempts.
$P(X > 5 \mid X > 2) = \dfrac{P(X > 5)}{P(X > 2)} = \dfrac{0.8^5}{0.8^2} = 0.8^3 = 0.512$	$X > 5$ is contained in $X > 2$, so the intersection is $X > 5$.
$E[X] = 5$, $P(X > 3) = 0.512$, $P(X \leq 3) = 0.488$, $P(X > 5 \mid X > 2) = 0.512$	Memorylessness: after two failures, the wait for success starts afresh, so the past attempts are no evidence that success is due.

Problem 7

Errors arrive in a log at an average of 3 per minute, as a Poisson count. Find the probability of no errors in a minute, of at most 2, and of at least 3, and show from the series that $E[X] = \lambda$.

$P(X = 0) = e^{-3} \approx 0.0498$	$\lambda^0 / 0! = 1$.

$P(X \le 2) = e^{-3}\left(1 + 3 + \frac{9}{2}\right) = 8.5e^{-3} \approx 0.4232$	The three terms $k = 0, 1, 2$.
$P(X \ge 3) = 1 - 8.5e^{-3} \approx 0.5768$	The complement, instead of an infinite sum.
$E[X] = \sum_{k \ge 1} k \dfrac{e^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k \ge 1} \dfrac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$	The $k = 0$ term is 0; cancel k against $k!$ and re-index with $j = k - 1$.
$P(0) \approx 0.0498$, $P(X \le 2) \approx 0.4232$, $P(X \ge 3) \approx 0.5768$, $E[X] = 3$	The same trick on $E[X(X-1)] = \lambda^2$ gives $\mathrm{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$: for a Poisson count, mean and variance are equal.

Problem 8

> A rare label occurs with probability 0.002 per example, independently, in a batch of 1000. Compare the exact binomial probabilities of 0 and of 3 occurrences with the Poisson approximation, and say why the approximation works.

$P(X = 0) = 0.998^{1000} \approx 0.13506$	Binomial with $n = 1000$, $p = 0.002$.
Poisson with $\lambda = np = 2$: $P(0) = e^{-2} \approx 0.13534$	Same mean.
$P(X = 3) = \binom{1000}{3}(0.002)^3(0.998)^{997} \approx 0.18063$	$\binom{1000}{3} = 166{,}167{,}000$.
Poisson: $\dfrac{e^{-2}2^3}{3!} = \dfrac{4}{3}e^{-2} \approx 0.18045$	Agreement to about 0.1%.
Exact 0.13506 **and** 0.18063; **Poisson** 0.13534 **and** 0.18045	With $p = \lambda/n$, $(1 - \frac{\lambda}{n})^n \to e^{-\lambda}$ and $\binom{n}{k}p^k \to \frac{\lambda^k}{k!}$ as $n \to \infty$ with λ fixed: many trials, each rarely a success, give a Poisson count.

Problem 9

> A fair die is rolled, and if it shows K, then K fair coins are flipped. Let X be the number of heads. Find $E[X]$ and $\mathrm{Var}(X)$.

Given $K = k$, $X \sim \mathrm{Binomial}(k, \frac{1}{2})$, so $E[X \mid K] = \frac{K}{2}$ and $\mathrm{Var}(X \mid K) = \frac{K}{4}$	Problem 4 with $p = \frac{1}{2}$.
$E[X] = E\left[\frac{K}{2}\right] = \frac{3.5}{2} = 1.75$	The law of total expectation, with $E[K] = 3.5$ from Problem 2.
$E\left[\mathrm{Var}(X \mid K)\right] = \frac{3.5}{4} = \frac{7}{8}$	The average spread within each value of K.
$\mathrm{Var}\left(E[X \mid K]\right) = \mathrm{Var}\left(\frac{K}{2}\right) = \frac{1}{4} \cdot \frac{35}{12} = \frac{35}{48}$	The spread of the conditional means, using $\mathrm{Var}(K) = \frac{35}{12}$.
$E[X] = 1.75$ **and** $\mathrm{Var}(X) = \frac{7}{8} + \frac{35}{48} = \frac{77}{48} \approx 1.604$	Using only the first term, $\frac{7}{8}$, would ignore the uncertainty in how many coins are flipped.

Problem 10

> A bootstrap sample draws n examples uniformly at random, with replacement, from a training set of n. Find the expected number of distinct examples in it, evaluate it for $n = 10$, and find the fraction it approaches as n grows.

$D = \sum_{j=1}^{n} I_j$ with $I_j = 1$ if example j is drawn at least once	Count distinct examples one example at a time.
$P(I_j = 0) = \left(1 - \frac{1}{n}\right)^n$	Each of the n independent draws misses example j with probability $1 - \frac{1}{n}$.
$E[D] = n\left(1 - \left(1 - \frac{1}{n}\right)^n\right)$	Linearity; the I_j are dependent, which does not matter for the mean.
$n = 10$: $E[D] = 10(1 - 0.9^{10}) \approx 10(1 - 0.3487) = 6.513$	About 6.5 distinct examples out of 10 draws.

$E[D] = n\big(1 - (1 - \frac{1}{n})^n\big)$; 6.513 **for** $n =$ 10; **the fraction tends to** $1 - e^{-1} \approx$ 0.632	$(1 - \frac{1}{n})^n \to e^{-1}$. So each bootstrap sample leaves out about 36.8% of the training set, which bagging uses as a free validation set: the out-of-bag examples.

Where this goes wrong

1. E[X²] taken as (E[X])²

The expectation of a function is computed by putting the mean into the function.

$E[X] = 1.7$ for the table of Problem 1	Right so far.
"$E[X^2]$ is the square of the mean."	The step that causes the mistake: expectation does not pass through a square.
$E[X^2] = 1.7^2 = 2.89$, so $\mathrm{Var}(X) = 2.89 - 2.89 = 0$ ✗	A variance of 0 would mean X is always 1.7, which it never is. $E[X^2] = \sum x^2 p(x) = 3.7$, and $\mathrm{Var}(X) = 0.81$. The gap between $E[X^2]$ and $(E[X])^2$ is exactly the variance, so it is never negative.

2. Var(2X + 3) taken as 2 Var(X)

$E[2X + 3] = 2E[X] + 3$ is linear, and the variance is expected to behave the same way.

$\mathrm{Var}(X) = \frac{35}{12}$ for a fair die	Right so far: Problem 2.
"Scaling by 2 doubles the variance, and the $+3$ drops out."	The $+3$ does drop out; the factor is what goes wrong.
$\mathrm{Var}(2X + 3) = 2 \cdot \frac{35}{12} = \frac{35}{6} \approx 5.83$ ✗	Variance is in squared units: doubling every deviation from the mean multiplies every squared deviation by 4. $\mathrm{Var}(2X + 3) = \frac{35}{3} \approx 11.67$, and the standard deviation, not the variance, doubles.

3. Binomial probability without the coefficient

The probability of one particular sequence of results is written down, and it looks like the answer.

10 independent examples, each correct with probability 0.8	Right so far: Problem 3.
“Eight correct and two wrong has probability $0.8^8 \cdot 0.2^2$.”	True for one specified pattern, such as the first eight right and the last two wrong.
$P(X = 8) = 0.8^8(0.2)^2 \approx 0.00671$ ✗	There are $\binom{10}{8} = 45$ patterns with exactly eight correct, each with that probability, so $P(X = 8) = 45 \times 0.00671 \approx 0.302$. Without the coefficient, the eleven probabilities $P(X = 0), \ldots, P(X = 10)$ would not sum to 1.

4. Adding probabilities that overlap

Three tries at a 20% chance sounds like a 60% chance.

Each attempt succeeds with probability 0.2, independently	Right so far: Problem 6.
“Success within three attempts is success on the first, the second or the third.”	Also right; the events are not disjoint, though.
$P(X \le 3) = 0.2 + 0.2 + 0.2 = 0.6$ ✗	Adding counts the outcomes where two or three attempts would succeed more than once; with six attempts the same rule would give 1.2. Go through the complement: $P(X \le 3) = 1 - 0.8^3 = 0.488$.

5. Variances of dependent indicators added as if independent

Problem 4 adds the variances of indicators, and the same move looks fine for the matching problem.

$X = I_1 + \cdots + I_n$ with $\mathrm{Var}(I_j) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$	Right so far: Problem 5.
"So $\mathrm{Var}(X)$ is n times that."	The step that causes the mistake: it needs the I_j to be uncorrelated, and they are not.
$\mathrm{Var}(X) = 1 - \frac{1}{n}$, which is 0.75 for $n = 4$ ✗	Every pair has a positive covariance $\frac{1}{n(n-1)} - \frac{1}{n^2}$, and the $n(n-1)$ ordered pairs add $\frac{1}{n}$, so $\mathrm{Var}(X) = 1$ for every $n \geq 2$. Linearity of expectation needs no independence; adding variances does.

More at eulerpad.app. This is one of 82 free practice sets with worked solutions at eulerpad.app/practice, alongside guides to the topics behind them.

EulerPad, the app, reads your handwritten math on iPad, iPhone and Android and checks it line by line. It's free to start.

This set is also Chapter 3 of *Probability & Statistics for Machine Learning*, a 250-problem workbook with worked solutions: eulerpad.app/workbooks

eulerpad.app/practice/discrete-random-variables-and-expectation