Revision: Discrete Random Variables

In L8 we met discrete random variables and their distributions. We begin with a brisk recap; the definitions here should feel familiar.

Definition. A discrete random variable X is a quantity whose value depends on chance, taking values in some finite or countable set. Its probability mass function (pmf) is the function

$$x \mapsto \mathbb{P}(X = x),$$

and since X must take some value,

$$\sum_x \mathbb{P}(X = x) = 1$$

where the sum is over all possible values of X.

Definition. The expectation and variance of X are

$$\mu = \mathbb{E}[X] = \sum_x x\,\mathbb{P}(X = x), \qquad \sigma^2 = \mathrm{Var}[X] = \mathbb{E}[(X-\mu)^2].$$

Fact (Expectation of a function of X) — For a function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X)] = \sum_x g(x)\,\mathbb{P}(X = x).$$

In particular $\mathbb{E}[X^2] = \sum_x x^2\,\mathbb{P}(X = x)$ — note this is *not* the same as $(\mathbb{E}[X])^2$ in general.

Theorem (Computational form of the variance)

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Expanding the square inside the defining sum,

$$\begin{aligned} Var[X] = \sum_x (x-\mu)^2\,\mathbb{P}(X = x) &= \sum_x x^2\,\mathbb{P}(X = x) - 2\mu \sum_x x\,\mathbb{P}(X = x) + \mu^2 \sum_x \mathbb{P}(X = x) \\ &= \mathbb{E}[X^2] - 2\mu\cdot\mu + \mu^2\cdot 1 = \mathbb{E}[X^2] - \mu^2. \end{aligned}$$

Theorem (Linear coding)
If $a, b \in \mathbb{R}$ are constants,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \qquad \mathrm{Var}[aX + b] = a^2\,\mathrm{Var}[X].$$

Adding a constant shifts the distribution without changing its spread; scaling by a scales the standard deviation by $|a|$ and hence the variance by a^2.

For the expectation, using $\mathbb{E}[g(X)] = \sum_x g(x)\,\mathbb{P}(X = x)$ with $g(x) = ax + b$:

$$\mathbb{E}[aX + b] = \sum_x (ax + b)\,\mathbb{P}(X = x) = a \sum_x x\,\mathbb{P}(X = x) + b \sum_x \mathbb{P}(X = x) = a\mathbb{E}[X] + b.$$

For the variance, $aX+b$ has mean $a\mu+b$, so its deviations from its mean are

$$(aX+b)-(a\mu+b)=a(X-\mu),$$

and therefore

$$Var[aX+b]=\mathbb{E}\Big[a^2(X-\mu)^2\Big]=a^2\,\mathbb{E}[(X-\mu)^2]=a^2\,Var[X].$$

Example

The discrete random variable X has probability distribution

x	1	2	3	4
$\mathbb{P}(X=x)$	0.1	k	0.3	$2k$

(a) Find k.

(b) Find $\mathbb{E}[X]$ and $\text{Var}[X]$.

(c) Find $\mathbb{E}[5X-2]$ and $\text{Var}[5X-2]$.

(a) The probabilities sum to 1: $0.1+k+0.3+2k=1$, *so* $3k=0.6$ *and* $k=0.2$.

(b) With probabilities $0.1, 0.2, 0.3, 0.4$:

$$\begin{aligned}\mathbb{E}[X]&=1(0.1)+2(0.2)+3(0.3)+4(0.4)=3\\ \mathbb{E}[X^2]&=1(0.1)+4(0.2)+9(0.3)+16(0.4)=10\\ Var[X]&=10-3^2=1\end{aligned}$$

(c) $\mathbb{E}[5X-2]=5\mathbb{E}[X]-2=13$ *and* $Var[5X-2]=25\,Var[X]=25$.

Standard distributions as models

Four families of discrete distributions appear so often that they have their own names; you must know when each applies and the formulae for the mean and variance. We prove the results we can reach directly; the remaining proofs become one-liners once we have probability generating functions.

Fact (Discrete uniform distribution) — $X\sim U(n)$ takes each value in $\{1,2,\dots,n\}$ with equal probability:

$$\mathbb{P}(X=x)=\frac{1}{n},\quad x\in\{1,\dots,n\},\qquad \mathbb{E}[X]=\frac{n+1}{2},\quad \text{Var}[X]=\frac{n^2-1}{12}.$$

Using $\sum_{k=1}^n k=\frac{n(n+1)}{2}$ *and* $\sum_{k=1}^n k^2=\frac{n(n+1)(2n+1)}{6}$:

$$\mathbb{E}[X]=\frac{1}{n}\sum_{k=1}^n k=\frac{n+1}{2},\qquad \mathbb{E}[X^2]=\frac{1}{n}\sum_{k=1}^n k^2=\frac{(n+1)(2n+1)}{6},$$

so

$$Var[X]=\frac{(n+1)(2n+1)}{6}-\frac{(n+1)^2}{4}=\frac{n+1}{12}\Big(2(2n+1)-3(n+1)\Big)=\frac{(n+1)(n-1)}{12}=\frac{n^2-1}{12}.$$

Fact (Bernoulli distribution) — $X \sim \text{Bernoulli}(p)$ records a *single* trial: $X = 1$ (success) with probability p and $X = 0$ (failure) with probability $q = 1 - p$.

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = q, \qquad \mathbb{E}[X] = p, \quad \text{Var}[X] = pq.$$

Directly from the definitions:

$$\mathbb{E}[X] = 0 \cdot q + 1 \cdot p = p, \qquad \mathbb{E}[X^2] = 0^2 \cdot q + 1^2 \cdot p = p,$$

so $Var[X] = p - p^2 = p(1-p) = pq$.

Fact (Binomial distribution) — $X \sim B(n, p)$ counts the number of successes in a *fixed* number n of independent trials, each a success with the same probability p. With $q = 1 - p$:

$$\mathbb{P}(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x \in \{0, 1, \ldots, n\}, \qquad \mathbb{E}[X] = np, \quad \text{Var}[X] = npq.$$

Equivalently: X is the sum of n independent Bernoulli(p) variables — one per trial. The formulae for the mean and variance are then n copies of the Bernoulli ones; expectation algebra (later in this chapter) explains why this addition is legitimate.

Fact (Geometric distribution) — $X \sim \text{Geo}(p)$ counts the number of trials up to *and including* the first success, in repeated independent trials with constant success probability p. With $q = 1 - p$:

$$\mathbb{P}(X = x) = q^{x-1} p, \quad x \in \{1, 2, 3, \ldots\}, \qquad \mathbb{E}[X] = \frac{1}{p}, \quad \text{Var}[X] = \frac{q}{p^2}.$$

The useful tail formula $\mathbb{P}(X > x) = q^x$ ("the first x trials all fail") is quotable. The mean and variance we will prove with probability generating functions.

Example

A fair twelve-sided die is rolled repeatedly.

(a) Find the mean and variance of the score on a single roll.

(b) Find the probability that the first score of 12 occurs on the fifth roll.

(c) In 36 rolls, find the mean and variance of the number of scores of 12.

(a) Single score $X \sim U(12)$, *so* $\mathbb{E}[X] = \frac{13}{2} = 6.5$ *and* $Var[X] = \frac{144-1}{12} = \frac{143}{12} \approx 11.9$.

(b) Let $Y \sim Geo\left(\frac{1}{12}\right)$ *be the number of rolls up to and including the first 12. Then*

$$\mathbb{P}(Y = 5) = \left(\tfrac{11}{12}\right)^4 \cdot \tfrac{1}{12} = 0.0588 \text{ (3 s.f.)}.$$

(c) Let $W \sim B\left(36, \frac{1}{12}\right)$. *Then* $\mathbb{E}[W] = 36 \cdot \frac{1}{12} = 3$ *and* $Var[W] = 36 \cdot \frac{1}{12} \cdot \frac{11}{12} = \frac{11}{4} = 2.75$.

Tip

In modelling questions, always *define* your random variable ("let X be the number of ..."), *name* the distribution with its parameters, and check the modelling conditions: fixed n, independence, constant p for binomial; independent trials with constant p for geometric.

Example (OCR Further Stats, June 2024)

A discrete random variable X has the following distribution, where a, b and c are constants.

x	0	1	2	3
$\mathbb{P}(X = x)$	a	b	c	0.1

It is given that $\mathbb{E}[X] = 1.25$ and $\text{Var}[X] = 0.8875$.

(a) Determine the values of a, b and c.

(b) The random variable Y is defined by $Y = 7 - 2X$. Write down the value of $\text{Var}[Y]$.

(c) Twenty independent observations of X are obtained. The number of those observations for which $X = 3$ is denoted by T. Find the value of $\text{Var}[T]$.

(a) Three pieces of information give three equations. Probabilities sum to 1: $a + b + c = 0.9$. *The mean:* $b + 2c + 0.3 = 1.25$, *so* $b + 2c = 0.95$. *The variance:* $\mathbb{E}[X^2] = 0.8875 + 1.25^2 = 2.45$, *so* $b + 4c + 0.9 = 2.45$, *i.e.* $b + 4c = 1.55$. *Subtracting the last two:* $2c = 0.6$, *so* $c = 0.3$, $b = 0.35$, $a = 0.25$.

(b) $Var[Y] = (-2)^2\, Var[X] = 4 \times 0.8875 = 3.55$.

(c) Each observation is independently "a 3" with probability 0.1, *so* $T \sim B(20, 0.1)$ *and* $Var[T] = npq = 20 \times 0.1 \times 0.9 = 1.8$.

Textbook Exercises: [CUP.S] Ch 1 and Ch 2 (revision of discrete random variables and the standard distributions)

Joint Random Variables

So far each random variable has lived alone. Frequently we observe *two* quantities at once — a height and a weight, two dice, the scores of two players — and we care about how they vary together. This bivariate point of view underpins expectation algebra now, and correlation, regression and contingency tables later.

Definition. For discrete random variables X and Y defined on the same sample space, the joint probability distribution is the function

$$(x, y) \mapsto \mathbb{P}(X = x, Y = y),$$

usually displayed as a two-way table. As always, the probabilities over all cells sum to 1.

Definition. The marginal distribution of X is recovered by summing across the values of Y:

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y),$$

and similarly for Y. (The name comes from writing these row and column totals in the *margins* of the table.)

Definition. X and Y are independent iff

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y) \quad \text{for all } x, y.$$

Equivalently: every cell of the joint table is the product of its row and column totals.

Example

The joint distribution of X and Y is given by

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 1$	0.10	0.20	0.10
$X = 2$	0.20	0.10	0.30

(a) Write down the marginal distributions of X and of Y.

(b) Find $\mathbb{P}(X + Y = 2)$.

(c) Determine whether X and Y are independent.

(a) Row totals: $\mathbb{P}(X = 1) = 0.4$, $\mathbb{P}(X = 2) = 0.6$. *Column totals:* $\mathbb{P}(Y = 0) = 0.3$, $\mathbb{P}(Y = 1) = 0.3$, $\mathbb{P}(Y = 2) = 0.4$.

(b) $X + Y = 2$ *occurs for* $(X, Y) = (1, 1)$ *or* $(2, 0)$, *so* $\mathbb{P}(X + Y = 2) = 0.20 + 0.20 = 0.4$.

(c) Test a cell: $\mathbb{P}(X = 1)\mathbb{P}(Y = 0) = 0.4 \times 0.3 = 0.12 \neq 0.10 = \mathbb{P}(X = 1, Y = 0)$. *So X and Y are* ***not independent****.* *(One failing cell is enough; to confirm independence you must check* every *cell.)*

Expectation Algebra

Theorem (Linearity of expectation)
For any random variables X and Y and constants a, b, c:

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c.$$

This holds with *no assumption of independence whatsoever*.

The key new ingredient is $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, which we proved in L8 by summing over the joint distribution; combining it with the coding result $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$ gives the theorem. The sketch is worth seeing once:

$$\begin{aligned}\mathbb{E}[X+Y] &= \sum_x \sum_y (x+y)\,\mathbb{P}(X=x, Y=y) \\ &= \sum_x x \sum_y \mathbb{P}(X=x, Y=y) + \sum_y y \sum_x \mathbb{P}(X=x, Y=y) \\ &= \sum_x x\,\mathbb{P}(X=x) + \sum_y y\,\mathbb{P}(Y=y) = \mathbb{E}[X] + \mathbb{E}[Y],\end{aligned}$$

where we used the marginal distributions in the penultimate step. Then $\mathbb{E}[aX+bY+c] = \mathbb{E}[aX] + \mathbb{E}[bY+c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$.

Variance of Linear Combinations

Variance is not linear — but for *independent* variables something almost as good holds.

Theorem
If X and Y are independent random variables, then for constants a, b, c:

$$\mathrm{Var}[aX + bY + c] = a^2\,\mathrm{Var}[X] + b^2\,\mathrm{Var}[Y].$$

In particular, taking $a = 1, b = -1$:

$$\mathrm{Var}[X - Y] = \mathrm{Var}[X] + \mathrm{Var}[Y].$$

Write $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$. The constant c shifts the mean without affecting deviations, so it contributes nothing to the variance. The deviation of $aX+bY$ from its mean $a\mu_X + b\mu_Y$ is $a(X-\mu_X) + b(Y-\mu_Y)$, so

$$\begin{aligned}Var[aX+bY+c] &= \mathbb{E}\Big[\Big(a(X-\mu_X) + b(Y-\mu_Y)\Big)^2\Big] \\ &= a^2\,\mathbb{E}[(X-\mu_X)^2] + 2ab\,\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] + b^2\,\mathbb{E}[(Y-\mu_Y)^2].\end{aligned}$$

The outer terms are $a^2\,Var[X]$ and $b^2\,Var[Y]$. For the middle term, independence gives $\mathbb{E}[UV] = \mathbb{E}[U]\,\mathbb{E}[V]$ for the (independent) variables $U = X - \mu_X$ and $V = Y - \mu_Y$ (we will justify this product rule in the covariance section), and $\mathbb{E}[X - \mu_X] = 0$, so the cross term vanishes.

Tip

The variances always **add**, even when the variables are subtracted: $(-1)^2 = 1$. Subtracting an independent source of randomness makes the result *more* spread out, not less. Writing $\text{Var}[X-Y] = \text{Var}[X] - \text{Var}[Y]$ is one of the most heavily penalised errors on this paper (it can even give a negative variance!).

Example

X and Y are independent with $\mathbb{E}[X] = 5$, $\text{Var}[X] = 4$, $\mathbb{E}[Y] = 2$, $\text{Var}[Y] = 3$. Find the mean and variance of:

(a) $X+Y$

(b) $3X-2Y+1$

(c) $\dfrac{X+Y}{2}$

(a) $\mathbb{E}[X+Y] = 5+2 = 7$; $\mathit{Var}[X+Y] = 4+3 = 7$.

(b) $\mathbb{E}[3X-2Y+1] = 15-4+1 = 12$; $\mathit{Var}[3X-2Y+1] = 9(4)+4(3) = 48$.

(c) $\mathbb{E}\left[\frac{X+Y}{2}\right] = \frac{7}{2} = 3.5$; $\mathit{Var}\left[\frac{X+Y}{2}\right] = \frac{1}{4}(4+3) = \frac{7}{4} = 1.75$.

$2X$ is not $X_1 + X_2$

The following distinction catches out many candidates every year.

Fact — Let X_1, X_2 be *independent* random variables each with the same distribution as X. Then

$$\text{Var}[X_1+X_2] = 2\,\text{Var}[X], \qquad \text{but} \qquad \text{Var}[2X] = \text{Var}[X+X] = 4\,\text{Var}[X].$$

Why the difference? $X_1 + X_2$ is the sum of two *separate* observations: sometimes one is large while the other is small, and the fluctuations partially cancel. But $2X$ is a *single* observation doubled — there is no cancellation, every fluctuation is amplified by the factor 2, so the variance picks up the factor 2^2. Note $X+X$ means $2X$: it is the same X both times, and X is certainly not independent of itself. The means, of course, agree: $\mathbb{E}[X_1+X_2] = \mathbb{E}[2X] = 2\mathbb{E}[X]$.

Example

A machine fills bags of flour; the mass of flour in a bag has mean 1000 g and standard deviation 5 g, independently for different bags.

(a) Find the mean and variance of the total mass of flour in two randomly chosen bags.

(b) A shopkeeper instead estimates the mass of flour bought by a customer with two bags as "twice the mass of the first bag". Find the mean and variance of this estimate, and comment.

Let X_1, X_2 be the masses, independent, each with mean 1000 and variance 25.

(a) *$\mathbb{E}[X_1 + X_2] = 2000$; $Var[X_1 + X_2] = 25 + 25 = 50$ (standard deviation ≈ 7.07 g).*

(b) *$\mathbb{E}[2X_1] = 2000$ but $Var[2X_1] = 4 \times 25 = 100$ (standard deviation 10 g). The estimate is unbiased (correct mean) but twice as variable as the true total in variance terms — doubling one measurement amplifies its error rather than averaging errors out.*

Example (OCR Further Stats, June 2023 (part))

The discrete random variable W has the distribution $U(11)$. The independent discrete random variable V has the distribution $U(5)$. It is given that, for constants m and n, with $m > 0$,

$$\mathbb{E}[mW + nV] = 0 \qquad \text{and} \qquad \mathrm{Var}[mW + nV] = 1.$$

Determine the exact values of m and n.

From the uniform formulae: $\mathbb{E}[W] = 6$, $Var[W] = \frac{121-1}{12} = 10$, $\mathbb{E}[V] = 3$, $Var[V] = \frac{25-1}{12} = 2$. The two conditions become

$$6m + 3n = 0 \qquad \textit{and} \qquad 10m^2 + 2n^2 = 1.$$

The first gives $n = -2m$; substituting, $10m^2 + 8m^2 = 18m^2 = 1$, so (taking $m > 0$)

$$m = \frac{1}{\sqrt{18}} = \frac{\sqrt{2}}{6}, \qquad n = -\frac{\sqrt{2}}{3}.$$

Example (OCR S4, June 2012)

The random variables S and T are *independent* and have joint probability distribution given in the table.

	$S = 0$	$S = 1$	$S = 2$
$T = 1$	a	0.18	b
$T = 2$	0.08	0.12	0.20

(i) Show that $a = 0.12$ and find the value of b.

(ii) Find $\mathbb{P}(T - S = 1)$.

(iii) Find $\mathrm{Var}[T - S]$.

(i) *From the second row, $\mathbb{P}(T = 2) = 0.08 + 0.12 + 0.20 = 0.4$. By independence each cell is a product of marginals, so $\mathbb{P}(S = 1) = \frac{0.12}{0.4} = 0.3$, and then $0.18 = \mathbb{P}(T = 1)\mathbb{P}(S = 1)$ gives $\mathbb{P}(T = 1) = 0.6$ (as it must, since rows sum to 1). Now*

$$a + 0.08 = \mathbb{P}(S = 0) = \frac{0.08}{0.4} = 0.2 \implies a = 0.12, \qquad b + 0.20 = \mathbb{P}(S = 2) = \frac{0.20}{0.4} = 0.5 \implies b = 0.3.$$

(ii) *$T - S = 1$ occurs for $(T, S) = (1, 0)$ or $(2, 1)$: $\mathbb{P}(T - S = 1) = 0.12 + 0.12 = 0.24$.*

(iii) *Marginals: $\mathbb{P}(T = 1) = 0.6$, $\mathbb{P}(T = 2) = 0.4$; $\mathbb{P}(S = 0) = 0.2$, $\mathbb{P}(S = 1) = 0.3$, $\mathbb{P}(S = 2) = 0.5$. So*

$$\mathbb{E}[T] = 1.4, \quad \mathbb{E}[T^2] = 0.6 + 1.6 = 2.2, \quad Var[T] = 2.2 - 1.96 = 0.24,$$

$$\mathbb{E}[S] = 1.3, \quad \mathbb{E}[S^2] = 0.3 + 2.0 = 2.3, \quad Var[S] = 2.3 - 1.69 = 0.61.$$

By independence, $Var[T - S] = Var[T] + Var[S] = 0.24 + 0.61 = 0.85.$

Textbook Exercises: [CUP.S] Ch 8 §1; [S3/4] S3 Ch 2, S4 Ch 6

Covariance

When X and Y are *not* independent, how wrong is the formula $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$? Exactly by a term measuring how the two variables move together.

Definition. The covariance of X and Y is

$$\mathrm{Cov}[X,Y] = \mathbb{E}\Big[(X-\mu_X)(Y-\mu_Y)\Big] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

If large X tends to come with large Y, the product $(X-\mu_X)(Y-\mu_Y)$ is usually positive and the covariance is positive; if large X comes with small Y it is negative. Note $\mathrm{Cov}[X,X] = \mathrm{Var}[X]$: covariance generalises variance.

Theorem

For any random variables X and Y (independent or not),

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y].$$

Write $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$, so $\mathbb{E}[X+Y] = \mu_X + \mu_Y$. Then

$$\begin{aligned} Var[X+Y] &= \mathbb{E}\Big[\big((X-\mu_X)+(Y-\mu_Y)\big)^2\Big] \\ &= \mathbb{E}[(X-\mu_X)^2] + 2\,\mathbb{E}[(X-\mu_X)(Y-\mu_Y)] + \mathbb{E}[(Y-\mu_Y)^2] \\ &= Var[X] + 2\,Cov[X,Y] + Var[Y]. \end{aligned}$$

Fact — If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, so $\mathrm{Cov}[X,Y] = 0$ — which recovers our addition rule for variances. **The converse is false**: zero covariance does *not* imply independence.

Example (Uncorrelated but dependent)

Let X take the values $-1, 0, 1$ each with probability $\frac{1}{3}$, and let $Y = X^2$. Show that $\mathrm{Cov}[X,Y] = 0$, but that X and Y are not independent.

First, $\mathbb{E}[X] = \frac{1}{3}(-1+0+1) = 0$. Also $XY = X \cdot X^2 = X^3$, and X^3 takes values $-1, 0, 1$ each with probability $\frac{1}{3}$, so $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$. Hence

$$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0 - 0 \cdot \mathbb{E}[Y] = 0.$$

But X and Y are certainly not independent: Y is completely determined by X. Concretely,

$$\mathbb{P}(X=0, Y=1) = 0 \quad \textit{whereas} \quad \mathbb{P}(X=0)\,\mathbb{P}(Y=1) = \tfrac{1}{3} \cdot \tfrac{2}{3} = \tfrac{2}{9} \neq 0.$$

Covariance only detects linear *relationships; here the dependence is perfectly symmetric about 0 and the linear signal cancels.*

Textbook Exercises: [S3/4] S4 Ch 6