Probability rules and conditional expectations

Quantitative Methods

1 A story: should you take an umbrella?

- Forecast: 30% chance of rain = 0.30
- Probability: a score from 0 (impossible) to 1 (certain)
- New news can change the chance
- This topic: combine, average, update

IN THE LESSON

- Picture a cloudy morning.
- The forecast says there is a thirty per cent chance of rain.
- That number is a probability.
- It is a score between zero and one, or between zero and a hundred per cent, for how likely something is.
- Zero means it cannot happen.
- One means it is certain.
- Thirty per cent is zero point three.
- Now suppose you hear that rain is much more likely if the wind comes from the west.
- Your estimate changes when you learn something new.
- This topic is about doing that carefully.

2 Events, and the word given

- Event: something that might happen (A, B)
- P(A): the chance of event A
- P(A | B): chance of A GIVEN that B happened
- Different from P(A) on its own

IN THE LESSON

- Two words first.
- An event is just something that might happen, like it rains, or a share price falls.
- We label events with capital letters, A and B.
- The probability of A is written P of A.
- Next, the word given.
- Written as a vertical bar, it means given that something else has already happened.
- So P of A given B is the chance of A once you know B has happened.
- For example, the chance of rain given a west wind.
- A chance with a given is a different number from the chance on its own.
- Mixing them up is the most common error here.

3 The addition rule: either one happens

P(A or B) = P(A) + P(B) − P(AB)

P(A)	Chance of the first event
P(B)	Chance of the second event
P(AB), both	Chance both happen: counted twice

IN THE LESSON

- Here is the first rule, the addition rule.
- It answers: what is the chance that A happens, or B happens, or both?
- The probability of A or B equals P of A, plus P of B, minus P of A and B.
- Let's take each term.
- P of A is the chance of the first event.
- P of B is the chance of the second.
- Now why subtract?
- Because an outcome where both happen has been counted twice, once inside P of A and once inside P of B.
- So we take one copy away.
- Think of two college clubs.
- A student in both must only be counted once.

4 Example 1: the addition rule (set-up)

Pick one company at random from the list

A: pays a dividend		**0.40**
B: is a large company		**0.35**
Both A and B	the overlap	**0.15**

IN THE LESSON

- Example one.
- A fund looks at a list of companies.
- Forty per cent of them pay a dividend, which means they hand out some profit to shareholders.
- Thirty five per cent are large companies.
- And fifteen per cent are both: large and paying a dividend.
- The question is: if you pick one company at random, what is the chance it is large, or pays a dividend, or both?
- Call A paying a dividend, and B being large.
- The overlap, P of A and B, is zero point one five.
- Next, we put them together.

5 Example 1: the answer

P(A or B) = P(A) + P(B) − P(AB)

P(A) + P(B)	0.40 + 0.35	**0.75**
Subtract the overlap	− 0.15	**0.60**
P(A or B)	dividend, large, or both	**0.60**

Forgetting the subtraction gives 0.75: counted twice

IN THE LESSON

- Now apply the rule.
- P of A or B equals zero point four zero, plus zero point three five, minus zero point one five.
- Forty plus thirty five is seventy five.
- Take away the overlap of fifteen, and we get sixty.
- So the answer is zero point six zero, or sixty per cent.
- Had we forgotten to subtract, we would have said seventy five per cent, counting the large dividend payers twice.
- Sense check: the answer can never be smaller than either single chance, or more than one.
- If two events can never happen together, we call them mutually exclusive, the overlap is zero, and we simply add.

6 The multiplication rule: both happen

P(AB) = P(A ¦ B) × P(B)

Joint P(AB)	Chance both happen
Conditional P(A | B)	Chance of A once B has happened
P(B)	Chance of B happening first

IN THE LESSON

- Second rule, the multiplication rule.
- It answers: what is the chance that A and B both happen?
- The probability of A and B, called the joint probability, equals P of A given B, times P of B.
- In words: first B has to happen, which has chance P of B.
- Then, given B, A has to happen, which has chance P of A given B.
- You multiply the two steps.
- Think of getting into a concert.
- First you need a ticket.
- Then, given the ticket, you need the doors to let you in.
- Multiply the two chances, and the joint chance is smaller than either part.

7 Independent events, and a trap

- Independent: P(A | B) = P(A), so P(AB) = P(A) × P(B)
- Two coin flips: 0.5 × 0.5 = 0.25 for two heads
- Mutually exclusive: cannot both happen
- Exclusive is NOT independent

IN THE LESSON

- A special case.
- Two events are independent if knowing one happened tells you nothing about the other.
- Then P of A given B is just P of A, and the rule becomes the simple version: multiply the two ordinary chances.
- Two coin flips are independent.
- The chance of two heads is a half times a half, a quarter.
- Now the trap.
- Independent does not mean mutually exclusive.
- Mutually exclusive means they cannot both happen, so if one happens, the other cannot.
- That is the opposite of knowing nothing.
- Exclusive events with real chances are in fact dependent.
- Never assume independence unless the question says so.

8 Example 2: a recession tree (set-up)

The stock either falls or holds, depending on the economy

P(recession)	**0.30**
P(no recession)	**0.70**
P(falls | recession)	**0.80**
P(falls | no recession)	**0.10**

IN THE LESSON

- Example two.
- A recession has a thirty per cent chance.
- So no recession has a seventy per cent chance.
- If there is a recession, a certain stock falls eighty per cent of the time.
- If there is no recession, the stock falls only ten per cent of the time.
- These last two are conditional chances, given what the economy does.
- Over all years, how often does the stock fall?
- There are two ways to reach a fall: through a recession, or through no recession.
- We multiply along each path, then add the two paths.

9 Example 2: follow each path

Multiply along each path, then add the paths

Recession and falls	0.30 × 0.80	**0.24**
No recession and falls	0.70 × 0.10	**0.07**
P(stock falls)	0.24 + 0.07	**0.31**

0.80 is not 0.31, so a fall is not independent of the economy

IN THE LESSON

- Path one: recession, then a fall.
- Multiply along the path.
- Zero point three zero times zero point eight zero is zero point two four.
- Path two: no recession, then a fall.
- Zero point seven zero times zero point one zero is zero point zero seven.
- The paths never overlap, so we just add.
- Zero point two four plus zero point zero seven is zero point three one.
- So the stock falls thirty one per cent of the time.
- Not eighty.
- The eighty only applies in a recession.
- Also, if a fall were independent of the economy, the chance given a recession would equal thirty one.
- It is eighty, so they are dependent.

10 Random variables and the expected value

- Random variable: an uncertain quantity
- Expected value: probability-weighted average
- The long-run average, not a promise
- It may be a value that never occurs

IN THE LESSON

- Next idea.
- A random variable is a quantity whose value is uncertain, such as next year's return on a share.
- Think of a game where you roll a die for prize money.
- The expected value is the probability weighted average of all the possible outcomes.
- It is what you would average per play if you could repeat the game many times.
- It need not be a value that can actually happen.
- A fair die has an expected roll of three and a half, and you can never roll that.
- And it tells you the centre, not how risky things are.

11 Expected value and variance

$E(X) = \Sigma\, p \times x$ Variance $= \Sigma\, p \times (x - E(X))^2$

Probability p	Chance of each outcome
Outcome x	Value of each outcome
Expected value E(X)	The probability-weighted average
Variance	Squared distance, weighted

IN THE LESSON

- Here are the formulas.
- The expected value of X is the sum of each outcome times its probability.
- We multiply every possible outcome, x, by its probability, p, and add them up.
- The probabilities must add to one.
- The variance measures spread.
- For each outcome, take the distance from the expected value, square it, multiply by that outcome's probability, and add them up.
- Squaring makes every distance positive and big misses count more.
- The standard deviation is the square root of the variance, and is easier to read, because it has the same units as the outcomes.

12 Example 3: expected return

Recession −10% (0.25), normal 8% (0.50), boom 22% (0.25)

0.25 × (−10%)	recession	**−2.5**
0.50 × 8%	normal	**4.0**
0.25 × 22%	boom	**5.5**
Expected return	−2.5 + 4.0 + 5.5	**7.0%**

IN THE LESSON

- Example three.
- Next year's return on a share depends on the economy.
- In a recession, with probability zero point two five, it returns minus ten per cent.
- In a normal year, probability zero point five zero, it returns eight per cent.
- In a boom, probability zero point two five, it returns twenty two per cent.
- Now weight each return.
- A quarter of minus ten is minus two point five.
- A half of eight is four.
- A quarter of twenty two is five point five.
- Add them.
- Minus two point five plus four plus five point five is seven.
- The expected return is seven per cent, a figure that none of the three years will actually deliver.

13 Example 3: variance and standard deviation

Distances are measured from the expected return, 7%

Distances from 7%		**−17, +1, +15**
Squared		**289, 1, 225**
Weighted squares	72.25 + 0.5 + 56.25	**129**
Variance		**129**
Standard deviation	√129	**11.36%**

IN THE LESSON

- Now the spread.
- Measure each return's distance from the expected return of seven, not from zero.
- Minus ten is seventeen below.
- Eight is one above.
- Twenty two is fifteen above.
- Square them: two eighty nine, one, and two twenty five.
- Now weight each square by its probability.
- A quarter of two eighty nine is seventy two point two five.
- A half of one is zero point five.
- A quarter of two twenty five is fifty six point two five.
- Add them up and the variance is one hundred twenty nine.
- The standard deviation is the square root, eleven point three six per cent.
- So: expected seven, typical miss eleven point three six.
- The spread is bigger than the average, so this share is risky.

14 Bayes: changing your mind with evidence

- Test: catches 90% of fraud, wrongly flags 5% of honest
- Only 1% of accounts are fraud (the prior)
- Flagged. Fraud chance: 90%? No, about 15%
- Bayes: update a belief with evidence

IN THE LESSON

- A bank runs a test to find fraud.
- One account in a hundred is fraudulent.
- The test catches ninety per cent of fraud, but it also wrongly flags five per cent of honest accounts.
- An account has just been flagged.
- How likely is it to be fraud?
- Most people say ninety per cent.
- The honest answer is about fifteen.
- Here is why.
- The prior, our starting belief before the evidence, is only one per cent.
- Honest accounts are so common that even a five per cent false alarm rate flags more of them than there are real frauds.
- Bayes' formula updates a starting belief using new evidence, and does the counting properly.

15 Bayes' formula

P(A ¦ B) = P(B ¦ A) × P(A) ÷ P(B)

Prior P(A)	Belief before the evidence
Likelihood P(B | A)	Chance of evidence if A is true
Total P(B)	Chance of the evidence overall
Posterior P(A | B)	Updated belief after the evidence

IN THE LESSON

- Here is Bayes' formula.
- The updated chance of A given B equals P of B given A, times P of A, all divided by P of B.
- Take each term.
- P of A is the prior, what you believed before the evidence.
- P of B given A is the likelihood: how probable the evidence is if A is true.
- P of B is the total chance of seeing that evidence at all, from any cause.
- You usually build it by adding up the ways it can happen.
- The answer, P of A given B, is the posterior, your updated belief.
- Strong evidence moves a belief a long way, but the starting point always matters.

16 Example 4: count the flags (set-up)

Imagine 10,000 accounts

Fraudulent accounts	1% of 10,000	**100**
Frauds flagged	90% of 100	**90**
Honest accounts	99% of 10,000	**9,900**
Honest ones flagged	5% of 9,900	**495**

IN THE LESSON

- Example four, with the bank.
- Imagine ten thousand accounts to make the counting easy.
- One per cent are fraudulent, which is one hundred accounts.
- The test catches ninety per cent of those, so ninety frauds are flagged.
- The other ninety nine per cent are honest, which is nine thousand nine hundred accounts.
- The test wrongly flags five per cent of them, which is four hundred ninety five honest accounts.
- Ninety real frauds are flagged, but four hundred ninety five honest accounts too.
- False alarms outnumber real catches by more than five to one.

17 Example 4: apply Bayes

P(fraud | flag) = P(flag | fraud) × P(fraud) ÷ P(flag)

Fraud and flagged	0.01 × 0.90	**0.0090**
Honest and flagged	0.99 × 0.05	**0.0495**
Any flag	0.0090 + 0.0495	**0.0585**
P(fraud | flag)	0.0090 ÷ 0.0585	**15.4%**

IN THE LESSON

- Now the formula, using probabilities.
- The chance of fraud and a flag is zero point zero one times zero point nine zero, which is zero point zero zero nine zero.
- The chance of honest and a flag is zero point nine nine times zero point zero five, which is zero point zero four nine five.
- The total chance of a flag is the two added together, zero point zero five eight five.
- That is P of B.
- Bayes then says: zero point zero zero nine zero divided by zero point zero five eight five, which is about fifteen point four per cent.
- The prior grew fifteen fold, yet the account is still more likely honest than fraudulent.

18 Covariance and correlation

Cov(A,B) = Σ (a − ā)(b − b̄) ÷ (n − 1)
ρ = Cov(A,B) ÷ (σA × σB)

Covariance	Do they move together?
Std dev of A	Spread of A
Std dev of B	Spread of B
Correlation	Unit-free, from −1 to +1

IN THE LESSON

- Now, do two things move together?
- Take two share returns, A and B.
- Covariance measures whether they tend to rise and fall together.
- For each period, take A's distance from its average and B's distance from its average, multiply the pair, add them up, and divide by the number of periods minus one.
- Positive means they move together, negative means they offset.
- But its size depends on the units, so you cannot compare it between pairs.
- Correlation fixes that.
- Divide covariance by the standard deviation of A times the standard deviation of B.
- The result has no units and always lies between minus one and plus one.

19 Example 5: two share returns (set-up)

Returns in %. Both averages are 5%.

Period	A and B returns
1	2 and 3
2	4 and 2
3	6 and 7
4	8 and 8

IN THE LESSON

- Example five.
- Two shares, A and B, over four periods.
- Returns for A are two, four, six and eight per cent.
- Returns for B are three, two, seven and eight per cent.
- Both tend to go up, but not perfectly, because B dipped in the second period while A rose.
- Both averages are exactly five per cent.
- For A, twenty divided by four is five.
- For B, three plus two plus seven plus eight is twenty, and divided by four is five.
- Next we find how far each return sits from its average.

20 Example 5: the covariance

Cov = Σ (a − 5)(b − 5) ÷ (4 − 1)

Distances for A		**−3, −1, +1, +3**
Distances for B		**−2, −3, +2, +3**
Multiply the pairs		**6, 3, 2, 9**
Add, divide by 3	20 ÷ 3	**6.67**

IN THE LESSON

- Step one, the distances from average.
- For A they are minus three, minus one, plus one and plus three.
- For B they are minus two, minus three, plus two and plus three.
- Step two, multiply each pair.
- Minus three times minus two is six.
- Minus one times minus three is three.
- Plus one times plus two is two.
- Plus three times plus three is nine.
- Step three, add them: six plus three plus two plus nine is twenty.
- Divide by the number of periods minus one, which is three.
- Twenty over three is six point six seven.
- The covariance is positive, so the shares tend to move together.
- But is six point six seven big or small?
- On its own, we cannot tell, because it is in per cent squared.

21 Example 5: the correlation

ρ = Cov ÷ (σA × σB)

Std dev of A	√(20 ÷ 3)	**2.58**
Std dev of B	√(26 ÷ 3)	**2.94**
Covariance	from the last slide	**6.67**
Correlation	6.67 ÷ (2.58 × 2.94)	**0.88**

IN THE LESSON

- To judge the size, convert to correlation.
- First the standard deviations.
- For A, square the distances: nine, one, one, nine, which add up to twenty.
- Divide by three and take the square root: two point five eight.
- For B, square: four, nine, four, nine, which add to twenty six.
- Divide by three and take the square root: two point nine four.
- Now correlation equals the covariance, six point six seven, divided by two point five eight times two point nine four.
- The bottom is about seven point five nine, so correlation is about zero point eight eight.
- That is close to plus one, so a strong positive straight line relationship.
- It has no units, so we can compare it with any other pair.

22 What correlation does and does not say

- +1: perfect straight line up. −1: straight line down
- 0: no STRAIGHT-LINE link (a curve can hide)
- Says nothing about how big the moves are
- Never proves one thing causes the other

IN THE LESSON

- Let's be careful about what these numbers tell you.
- A correlation of plus one means a perfect upward straight line.
- Minus one means a perfect downward straight line.
- Zero means there is no straight line pattern.
- But three limits matter in practice.
- First, correlation only captures straight line links.
- Two variables can be strongly related along a curve and still show a correlation of zero.
- Second, it says nothing about size of moves, a correlation of point nine nine could involve tiny or huge moves.
- Third, it never proves one thing causes another.
- Two shares can rise together because of a third factor, such as the economy.

23 Conditional expectations: it depends

- Forecast = scenarios + chance of each
- E(X | S): average outcome GIVEN scenario S
- Variance given S: spread inside that scenario
- Overall forecast: weight the what-ifs

IN THE LESSON

- Now let's bring the ideas together.
- Forecasts usually hide a story.
- Say a company's earnings depend on the economy.
- If the economy expands, earnings are likely to be one thing.
- If it falls into recession, they are likely to be another.
- A conditional expected value is the average outcome given that a particular scenario has happened.
- Written E of X given S.
- The same goes for conditional variance, the spread given the scenario.
- Each answers a what if question.
- The overall forecast weights each answer by the chance of its scenario, and you can update it properly when those chances change.

24 The total expectation rule

$$E(X) = \Sigma\, P(S) \times E(X \mid S)$$
$$Var(X \mid S) = \Sigma\, p \times (x - E(X \mid S))^2$$

Scenario chance	How likely each scenario is
Average given S	Average outcome in that scenario
Conditional variance	Spread inside that scenario
Overall E(X)	Weighted overall forecast

IN THE LESSON

- The rule is called the total probability rule for expected values.
- The overall expected value of X equals the sum, over every scenario, of the chance of that scenario times the expected value of X given that scenario.
- Take each term.
- The scenario chance is how likely the scenario is, and these chances must add up to one.
- The conditional expected value is the average outcome inside that scenario.
- Multiply and add.
- For conditional variance, measure each outcome's distance from its own scenario's average, not from the overall average.

25 Example 6: earnings by scenario (set-up)

EPS forecast. Chances inside each scenario are in brackets.

Expansion: chance 0.60	$3.00 (0.7), $2.50 (0.3)
Recession: chance 0.40	$2.00 (0.5), $1.40 (0.5)

IN THE LESSON

- Example six.
- An analyst forecasts earnings per share, called E P S, which is profit divided by the number of shares.
- If the economy expands, which has probability zero point six zero, earnings are three dollars with probability zero point seven, or two dollars fifty with probability zero point three.
- If the economy falls into recession, probability zero point four zero, earnings are two dollars with probability zero point five, or one dollar forty with probability zero point five.
- The chances inside the scenarios are conditional on the economy.
- We want the average and spread inside each, then the overall forecast.

26 Example 6: inside each scenario

Conditional expected values and variances

E(EPS | expansion)	0.7×3.00 + 0.3×2.50	**$2.85**
Variance | expansion	SD about $0.23	**0.0525**
E(EPS | recession)	0.5×2.00 + 0.5×1.40	**$1.70**
Variance | recession	SD $0.30	**0.09**

IN THE LESSON

- Expansion first.
- The conditional expected value is zero point seven times three dollars, plus zero point three times two fifty.
- That is two ten plus seventy five cents, so two dollars eighty five.
- The distances from that are fifteen cents and thirty five cents.
- Squared and weighted, zero point seven times zero point zero two two five, plus zero point three times zero point one two two five, gives a variance of zero point zero five two five, and a standard deviation of about twenty three cents.
- Recession next.
- A half of two dollars plus a half of one forty is one dollar seventy.
- Both outcomes sit thirty cents away, so the variance is zero point zero nine.

27 Example 6: the forecast, then the news

$E(EPS) = P(S_1) \times E(X \mid S_1) + P(S_2) \times E(X \mid S_2)$

Before: weights 0.6, 0.4	expansion, recession	
0.6 × 2.85 + 0.4 × 1.70		**$2.39**
After: weights 0.4, 0.6	expansion, recession	
0.4 × 2.85 + 0.6 × 1.70		**$2.16**

IN THE LESSON

- Now the overall forecast.
- Weight each scenario's average by its chance.
- Zero point six times two eighty five is one seventy one.
- Zero point four times one seventy is sixty eight cents.
- Add them: two dollars thirty nine.
- Then weak economic data arrives.
- The analyst now thinks recession has a sixty per cent chance, and expansion forty.
- Only the weights moved.
- Zero point four times two eighty five is one fourteen.
- Zero point six times one seventy is one oh two.
- Add: two dollars sixteen.
- The forecast fell twenty three cents purely because the odds of the bad scenario rose.

28 Common mistakes to avoid

- Addition rule: forget the overlap
- Assume independence without being told
- Variance from zero, or without weights
- Bayes: ignore the prior. Correlation ≠ cause

IN THE LESSON

- Here are the mistakes that lose the most marks.
- One.
- Forgetting to subtract the overlap in the addition rule, so shared outcomes are counted twice.
- Two.
- Assuming independence when nobody told you so, and then multiplying the plain chances.
- Three.
- Measuring variance from zero instead of from the expected value, or forgetting to weight by probability.
- Four.
- In Bayes, ignoring the prior, which makes a rare event look common after a positive test.
- Five.
- Reading correlation as proof of cause, or as proof of no relationship when it is zero.
- And six.
- Using the overall average instead of the scenario's own average in a conditional variance.

29 Recap: the toolkit

P(A or B) = P(A) + P(B) − P(AB) P(AB) = P(A given B) × P(B)
E(X) = Σ p x Posterior = likelihood × prior ÷ total
ρ = Cov ÷ (σA σB) E(X) = Σ P(S) E(X given S)

Addition rule	Subtract the overlap
Multiplication rule	Given, then multiply
Expected value	Weight by probability
Bayes' formula	Prior, likelihood, total
Correlation	Unit-free, −1 to +1

IN THE LESSON

- Let's pull it together.
- The addition rule: P of A or B is P of A plus P of B minus the overlap.
- The multiplication rule: P of A and B is P of A given B times P of B.
- Expected value is each outcome times its probability, added up, and variance is the weighted squared distance from it.
- Bayes' formula updates a prior: posterior equals likelihood times prior, divided by the total chance of the evidence.
- Correlation is covariance divided by the two standard deviations, always between minus one and one.
- And a conditional expected value is the average inside one scenario; weight those by scenario chances to get the overall forecast.

30 Try three quick questions

1. P(A) = 0.30, P(B) = 0.50, P(AB) = 0.10. What is P(A or B)?
   Answer: 0.30 + 0.50 − 0.10 = 0.70
2. EPS is $5 if the economy grows (chance 0.6) and $3 otherwise. Expected EPS?
   Answer: 0.6 × 5 + 0.4 × 3 = $4.20
3. Cov(A,B) = 12, σA = 4, σB = 5. What is the correlation?
   Answer: 12 ÷ (4 × 5) = 0.6

IN THE LESSON

- Now try three questions yourself.
- Pause after each one.
- Question one.
- The chance of event A is zero point three zero, event B is zero point five zero, and both together is zero point one zero.
- What is the chance of A or B?
- Thirty plus fifty minus ten gives zero point seven zero.
- Question two.
- Earnings are five dollars if the economy grows, with chance zero point six, and three dollars otherwise.
- What is expected earnings?
- Zero point six times five plus zero point four times three is four dollars twenty.
- Question three.
- The covariance is twelve, and the standard deviations are four and five.
- What is the correlation?
- Twelve divided by twenty is zero point six.
- Well done.