CHARTA LEARNING · LESSON NOTES

Multivariate Random Variables

FRM Part I

1 A story: two loans, one risk

- Two loans: each can fail, maybe together, maybe apart
- Single-loan chances cannot show how the pair behaves
- The joint behavior of the pair is what drives bank risk
- Tools: joint table, covariance, correlation, weights

IN THE LESSON

- Imagine a small bank with two business loans.
- For each loan we can ask whether it defaults.
- Looked at alone, each loan has its own chance of trouble.
- But the bank cares about something more.
- If the two borrowers sell to the same customers, they tend to fail in the same bad year, and the bank could lose both at once.
- If they work in unrelated trades, one failing tells us little about the other.
- The single loan distributions cannot tell these two banks apart.
- What separates them is the joint behavior of the pair.
- This module gives you the tools to describe that joint behavior with numbers, and to use it to measure risk.

2 The joint pmf and the probability matrix

- Rows: values of X. Columns: values of Y
- Each cell f(x, y) = P(X = x and Y = y)
- Rule 1: every cell lies between 0 and 1
- Rule 2: ALL cells together sum to 1
- Rows or columns on their own need not sum to 1

IN THE LESSON

- Start with two discrete random variables, which take a list of separate values.
- The joint probability mass function, or P M F, gives the chance that the first variable takes one value and the second takes another at the same time.
- We lay it out as a probability matrix.
- The values of the first variable run down the rows, the values of the second run across the columns, and each cell holds the chance of that pair.
- Two rules make a matrix valid.
- Every cell lies between zero and one.
- And all the cells together add up to exactly one, because some pair must occur.
- Notice what is not required: individual rows and columns do not have to add to one.

3 Our running example: the probability matrix

Illustration: X down the rows, Y across the columns

X \ Y	Y=1	Y=2	Y=3	total
X = 1	0.20	0.15	0.05	0.40
X = 2	0.10	0.15	0.35	0.60
Column total	0.30	0.30	0.40	1.00

IN THE LESSON

- Here is the matrix we will use all lesson.
- It is an illustration with made up numbers.
- The row variable, X, takes the values one and two.
- The column variable, Y, takes one, two and three.
- In the first row, the cells are point two zero, point one five and point zero five, and the row adds to point four zero.
- In the second row, the cells are point one zero, point one five and point three five, and that row adds to point six zero.
- At the bottom the column totals are point three zero, point three zero and point four zero.
- Everything together adds to one point zero zero, so the matrix is valid.
- Keep these six numbers in mind.

4 Example 1: events and a missing cell

Add the cells that satisfy the event; the whole matrix sums to 1

Sum of the five known cells	0.20+0.15+0.05+0.10+0.15	**0.65**
Missing cell f(2, 3)	1 − 0.65	**0.35**
P(X > Y): cell (2, 1) only	f(2, 1)	**0.10**
P(X + Y = 3)	f(1, 2) + f(2, 1)	**0.25**

IN THE LESSON

- Two skills first.
- Suppose one cell is hidden, say the cell for X equals two and Y equals three.
- We add the five known cells: point two, point one five, point zero five, point one, point one five, which total point six five.
- The missing cell is one minus that, point three five, matching the table.
- Next, events are found by adding the cells that satisfy them.
- The event that X is greater than Y happens only in the cell where X is two and Y is one, so its probability is point one zero.
- The event that the total X plus Y equals three happens in two cells: the cell for one and two, point one five, and the cell for two and one, point one zero.
- The sum is point two five.

5 Marginal and conditional distributions

$f_X(x) = \Sigma_y f(x, y)$
$f(x \mid y) = f(x, y) \div f_Y(y)$
Independent: $f(x, y) = f_X(x) \times f_Y(y)$

Marginal: a row or column total	Add across the other variable
Conditional: cell over given total	Divide by the GIVEN variable total
Joint probability of a cell	The cell itself
Independence test	Must hold in every cell

IN THE LESSON

- From the joint matrix we can recover two kinds of single variable distribution.
- The marginal of the row variable is its row total: add the joint probabilities over all values of the other variable.
- The marginal of the column variable is its column total.
- A marginal describes one variable while ignoring the other.
- The conditional distribution is different.
- To find the row variable given a value of the column variable, take that column and divide every cell by the column total, so the new numbers add to one.
- And the two variables are independent only if the joint probability equals the product of the marginals, in every cell, without exception.

6 Example 2: marginals and a conditional

Illustration: marginals are totals; a conditional divides by the given total

Marginal of X	row totals	**0.4, 0.6**
Marginal of Y	column totals	**0.3, 0.3, 0.4**
f(X = 2 | Y = 3)	0.35 ÷ 0.40	**0.875**
f(X = 1 | Y = 3)	0.05 ÷ 0.40	**0.125**
Check: conditional sums	0.875 + 0.125	**1**

IN THE LESSON

- Back to the running matrix.
- The marginal of the row variable is the pair of row totals: point four and point six.
- The marginal of the column variable is the column totals: point three, point three and point four.
- Now condition on the column variable equaling three.
- That is the third column, with cells point zero five and point three five, and the column total is point four.
- The chance that the row variable is two, given that the column variable is three, is point three five divided by point four, which is point eight seven five.
- The chance that it is one is point zero five over point four, point one two five.
- Check that the conditional probabilities add to one: they do.
- Given a large column outcome, the row variable is very likely two.

7 Reading marginals, conditionals, independence

- Marginal = add over the variable you want to ignore
- Conditional = cell ÷ total of the variable that is GIVEN
- Wrong divisor reverses the conditioning
- Independent: EVERY cell = row total × column total
- One failing cell proves dependence

IN THE LESSON

- Let's put the rules in words.
- A marginal is found by adding over the variable you want to ignore.
- A conditional is found by dividing a cell by the total of the variable you are given.
- The usual error is dividing by the wrong total, which reverses the conditioning, so always ask which variable is given.
- Independence is a strong claim.
- If the two variables are independent, learning the value of one tells you nothing about the other, and every single cell must equal its row total times its column total.
- If even one cell fails the test, the variables are dependent.
- One failing cell is enough to settle the matter.

8 Example 3: testing independence

Independent only if f(x, y) = f_X(x) × f_Y(y) in every cell

Actual cell f(1, 1)	from the matrix	**0.20**
Product of marginals	0.40 × 0.30	**0.12**
Verdict	0.20 ≠ 0.12	**Dependent**
Another cell f(2, 3)	0.35 vs 0.6 × 0.4 = 0.24	**Differs**

IN THE LESSON

- Are our two variables independent?
- Check the first cell, the one for X equals one and Y equals one.
- Its actual probability is point two zero.
- The product of the marginals is the row total, point four, times the column total, point three, which is point one two.
- Point two does not equal point one two.
- That alone says the variables are dependent.
- Let's confirm with another cell.
- The cell for two and three has probability point three five.
- The row total is point six, the column total is point four, and their product is point two four.
- Again they differ.
- The verdict is dependence: the variables move together in a way the marginals alone cannot explain.

9 Expected value of a function of two variables

$E[g(X, Y)] = \Sigma\, g(x, y) \times f(x, y)$
$E(X) = \Sigma\, x \times f_X(x)$ $E(XY) = \Sigma\, x \cdot y \cdot f(x, y)$

Function g of both variables	Evaluated in each cell
Probability of the cell	Weight each cell by its chance
Sum over every cell	Add over all cells
Nonlinear payoff	e.g. max(Y−X, 0)

IN THE LESSON

- Now expected values.
- For any function of the two variables, written g, the expected value is the sum over all cells of the function evaluated in that cell times the probability of that cell.
- You compute the function cell by cell, weight it by the cell's probability, and add.
- This one rule covers everything we need.
- If the function is just the row variable, you get its mean.
- If it is the product of the two variables, you get the quantity needed for covariance.
- If it is a nonlinear payoff, such as the larger of a difference and zero, you get the average payoff.
- For the plain mean of one variable, you can also use its marginal probabilities, which gives the same answer.

10 Example 4: E(X), E(Y) and E(XY)

$E[g] = \Sigma\, g(x, y)\, f(x, y)$. Illustration with our matrix

E(X)	1×0.40 + 2×0.60	**1.6**
E(Y)	1×0.3 + 2×0.3 + 3×0.4	**2.1**
E(XY), row X = 1	1 × (0.20+0.30+0.15)	**0.65**
E(XY), row X = 2	2 × (0.10+0.30+1.05)	**2.90**
E(XY), total	0.65 + 2.90	**3.55**

IN THE LESSON

- First the two means, using the marginals.
- The row variable has mean one times point four plus two times point six, which is one point six.
- The column variable has mean one times point three, plus two times point three, plus three times point four, which is two point one.
- Now the expected product.
- Go row by row.
- In the first row, the products times probabilities are point two, point three and point one five, which add to point six five, and the row variable is one, so the row contributes point six five.
- In the second row, the values add to one point four five, times two, giving two point nine zero.
- The expected product is the total: three point five five.

11 Example 5: a nonlinear payoff

Payoff $g = \max(Y - X, 0)$. Other cells pay 0

Cell (1, 2): payoff 1	1 × 0.15	**0.15**
Cell (1, 3): payoff 2	2 × 0.05	**0.10**
Cell (2, 3): payoff 1	1 × 0.35	**0.35**
True expected payoff	0.15 + 0.10 + 0.35	**0.60**
Wrong: plug in the means	max(2.1 − 1.6, 0)	**0.50**

IN THE LESSON

- Here is why we weight cell by cell.
- Suppose a payoff equals the larger of Y minus X, and zero.
- Only three cells pay anything.
- In the cell for one and two, the payoff is one, with probability point one five, contributing point one five.
- In the cell for one and three, the payoff is two, with probability point zero five, contributing point one zero.
- In the cell for two and three, the payoff is one, with probability point three five, contributing point three five.
- The expected payoff is point six zero.
- Now the tempting shortcut: plug the means into the function.
- The mean of Y minus the mean of X is point five, so that gives point five.
- The answers differ.
- Plugging in means works only for linear functions.

12 Covariance: moving together in a line

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

Covariance	Co-movement in units of X × Y
Shortcut form	E(XY) less E(X)E(Y)
Sign of the co-movement	Sign: together or opposite
Covariance with itself	Cov(X, X) = Var(X)

IN THE LESSON

- Covariance measures how two variables move together in a straight line.
- It is the average, weighted by probability, of the product of the two deviations from the means.
- If the variables tend to be above their means together, or below together, the products are mostly positive and the covariance is positive.
- If one tends to be high when the other is low, it is negative.
- In practice we use a shortcut: the expected product minus the product of the expected values.
- The covariance of a variable with itself is its variance.
- One drawback: covariance carries the units of the first variable times the units of the second, so its size is hard to judge.
- The common slip is forgetting to subtract the product of the means.

13 Example 6: covariance of our matrix

Cov = E(XY) − E(X)E(Y). Means 1.6 and 2.1, E(XY) = 3.55

Covariance	3.55 − 1.6 × 2.1	**0.19**
Sign of the covariance	0.19 > 0	**Together**
Var(X)	$2.80 - 1.6^2$	**0.24**
Var(Y)	$5.10 - 2.1^2$	**0.69**

IN THE LESSON

- Now compute it.
- The expected product is three point five five.
- The product of the means is one point six times two point one, which is three point three six.
- So the covariance is three point five five minus three point three six: point one nine.
- It is positive, so the two variables tend to move up and down together.
- We will also need the variances.
- For the row variable, the expected square is one times point four plus four times point six, which is two point eight.
- Subtract the mean squared, two point five six, and the variance is point two four.
- For the column variable, the expected square is five point one, the mean squared is four point four one, and the variance is point six nine.

14 Correlation: covariance without units

$\rho = \mathrm{Cov}(X, Y) \div (\sigma_X \times \sigma_Y)$
$\mathrm{Cov}(X, Y) = \rho \times \sigma_X \times \sigma_Y$

Correlation	Unit-free co-movement
Covariance	Numerator: the covariance
Standard deviations	Square roots of the variances
Range of correlation	Between −1 and +1

IN THE LESSON

- To make covariance easy to read, we rescale it.
- Correlation is the covariance divided by the product of the two standard deviations.
- The units cancel, and the result always lies between minus one and plus one.
- A correlation of plus one means an exact straight line sloping up, and minus one an exact straight line sloping down.
- Zero means no linear co-movement.
- This is why correlations can be compared across different pairs, while raw covariances cannot.
- The same fact says the size of a covariance can never exceed the product of the two standard deviations.
- Whenever you are given a correlation and two standard deviations, you can turn them back into a covariance by multiplying the three.

15 Example 7: correlation of our matrix

Illustration: Cov = 0.19, Var(X) = 0.24, Var(Y) = 0.69

σ_X	√0.24	**0.490**
σ_Y	√0.69	**0.831**
σ_X × σ_Y	0.490 × 0.831	**0.407**
Correlation	0.19 ÷ 0.407	**0.467**

IN THE LESSON

- Let's turn the covariance of point one nine into a correlation.
- The variance of the row variable is point two four, so its standard deviation is the square root, about point four nine zero.
- The variance of the column variable is point six nine, so its standard deviation is about point eight three one.
- Their product is about point four zero seven.
- The correlation is point one nine divided by point four zero seven, which is about point four six seven.
- So the two variables have a moderate positive linear relationship.
- It is well above zero but far below one, so knowing one of them helps only partly in guessing the other.
- Always check your answer lands between minus one and plus one.

16 Zero covariance does not mean independent

- Independent ⇒ Cov = 0 and correlation = 0
- The reverse is FALSE: zero only rules out a straight line
- Example: X symmetric about 0, and $Y = X^2$
- Correlation of +1 or −1 means an exact straight line

IN THE LESSON

- Now the key link between independence and covariance, and it is worth knowing cold.
- If two variables are independent, the expected product equals the product of the expected values, so the covariance is zero and so is the correlation.
- Independence implies zero covariance.
- But the reverse is false.
- Zero covariance only rules out a straight line relationship.
- Two variables can be strongly dependent in a curved way, and still have zero covariance and zero correlation.
- The classic case is a variable that is symmetric around zero, paired with its own square.
- Knowing the first tells you the second exactly, yet no straight line connects them.
- And a correlation of plus or minus one means an exact straight line, nothing less.

17 Example 8: dependent, yet covariance zero

Illustration: X = −1, 0, +1 with probability 1/3 each; $Y = X^2$

E(X)	(−1 + 0 + 1) ÷ 3	**0**
E(XY) = E(X³)	(−1 + 0 + 1) ÷ 3	**0**
Covariance	0 − 0 × E(Y)	**0**
Yet knowing X fixes Y	$Y = X^2$	**Dependent**

IN THE LESSON

- Let's check that claim with numbers.
- Let the first variable take the values minus one, zero and plus one, each with probability one third.
- Let the second variable be its square, so it takes one, zero and one.
- The mean of the first variable is minus one plus zero plus one, over three, which is zero.
- The expected product of the two is the expected value of the first variable cubed: minus one, zero and one again, averaging to zero.
- The covariance is zero minus zero times the mean of the second, which is zero.
- Yet the second variable is completely determined by the first.
- They are perfectly dependent.
- Zero covariance did not mean independence.

18 Linear transformations

Cov(aX + b, cY + d) = a × c × Cov(X, Y)
Corr unchanged if a×c > 0; sign flips if a×c < 0

Multipliers a and c	Covariance scales by a × c
Added constants	Constants b and d: no effect
Correlation after the change	Same size; sign of a × c

IN THE LESSON

- What happens if we change the units, or shift the variables?
- Take the first variable, multiply by a, and add b.
- Take the second, multiply by c, and add d.
- The new covariance equals a times c times the old covariance.
- Adding a constant has no effect, because the mean moves with the variable and the deviations stay the same.
- Multiplying scales the covariance by the product of the multipliers, signs included.
- Correlation is unit free, so it is unchanged when a times c is positive.
- When a times c is negative, the sign flips but the size stays the same.
- Common mistakes are scaling the correlation by the multipliers, forgetting the sign, or letting the constants matter.

19 Example 9: changing units and signs

Illustration: Cov(X, Y) = 0.19, correlation 0.467

Cov(2X + 5, −3Y + 1)	2 × (−3) × 0.19	**−1.14**
Correlation after	a × c < 0: sign flips	**−0.467**
Decimals to percent: Cov	0.0019 × 100 × 100	**19**
Decimals to percent: corr	unchanged	**0.467**

IN THE LESSON

- Take our covariance of point one nine, and correlation of point four six seven.
- Now form two new variables: two times the first plus five, and minus three times the second plus one.
- The multipliers are two and minus three, and the constants five and one do nothing.
- The new covariance is two times minus three times point one nine, which is minus one point one four.
- The correlation keeps its size, but the product of the multipliers is negative, so the sign flips: minus point four six seven.
- Now a unit change.
- Convert two returns from decimals to percent: each is multiplied by one hundred, so a covariance of point zero zero one nine becomes nineteen, while the correlation stays at point four six seven.

20 Variance of a weighted combination

Var(aX + bY) = a²Var(X) + b²Var(Y) + 2ab·Cov(X, Y)
Cov(X, Y) = ρ × σ_X × σ_Y
SD(aX + bY) = √Var(aX + bY)

Squared weights	Weights squared on variances
Cross term	2ab: doubled, appears once
Signs of the weights	Signs matter: X − Y has b = −1
Covariance	Or use ρ × σ_X × σ_Y

IN THE LESSON

- Now the formula behind portfolio risk.
- Combine two variables with weights a and b.
- The variance of the combination is a squared times the first variance, plus b squared times the second variance, plus two times a times b times the covariance.
- If you are given a correlation and standard deviations instead, replace the covariance by the correlation times the two standard deviations.
- Notice the pattern.
- The weights are squared on the variances, but appear once, doubled, on the covariance.
- The signs of the weights matter, so for a difference, with weights one and minus one, the covariance term is subtracted.
- The standard deviation is simply the square root of the variance.
- The common errors are omitting the two, not squaring, and ignoring a sign.

21 Example 10: a two-asset portfolio

Illustration: σ_X 20%, σ_Y 30%, ρ = 0.5, weights 0.6 and 0.4

Term 1: 0.6² × 0.04	0.36 × 0.04	**0.0144**
Term 2: 0.4² × 0.09	0.16 × 0.09	**0.0144**
Cross: 2ab × Cov	0.48 × (0.5 × 0.2 × 0.3)	**0.0144**
Variance	0.0144 × 3	**0.0432**
Volatility	√0.0432	**20.78%**

IN THE LESSON

- Let X and Y be returns, with standard deviations of twenty per cent and thirty per cent, and a correlation of point five.
- Put sixty per cent in the first and forty per cent in the second.
- The first term is point six squared, point three six, times the variance point zero four, which is point zero one four four.
- The second term is point four squared, point one six, times point zero nine, also point zero one four four.
- The cross term is two times point six times point four, point four eight, times the covariance, point five times point two times point three, point zero three, giving point zero one four four again.
- The variance is point zero four three two.
- The square root is a volatility of twenty point seven eight per cent.

22 Example 11: the spread X minus Y

Same σ_X = 20%, σ_Y = 30%. Spread X − Y has weights 1 and −1

Var(X) + Var(Y)	0.04 + 0.09	**0.13**
Minus 2 × Cov, ρ = 0.5	2 × 0.5 × 0.2 × 0.3	**0.06**
Spread variance	0.13 − 0.06	**0.07**
Spread volatility	√0.07	**26.46%**
If uncorrelated: volatility	√0.13	**36.06%**

IN THE LESSON

- Now a long short position, which is the first variable minus the second.
- The weights are one and minus one.
- The variance is the sum of the two variances, point zero four plus point zero nine, which is point one three, minus two times the covariance.
- With a correlation of point five, the covariance is point zero three, so we subtract point zero six.
- The spread variance is point zero seven, and the volatility is the square root, twenty six point four six per cent.
- Compare the uncorrelated case.
- With no correlation, the cross term vanishes and the variances add, even though we subtract.
- The variance is point one three, and the volatility thirty six point zero six per cent.
- Positive correlation makes the spread less volatile.

23 Conditional expected value

- Take the row for x; divide each cell by the row total
- Average the y values with those new weights
- Condition on an event: divide by the event's probability
- Independent: E(Y | x) = E(Y) for every x
- Slip to avoid: forgetting to divide by the row total

IN THE LESSON

- Next, the conditional mean.
- The expected value of the column variable, given that the row variable equals a particular value, is the mean of the column variable under the conditional distribution.
- In the matrix, take the row for that value, divide each cell by the row total so the row sums to one, and then average the column values using those new weights.
- You can also condition on an event, such as the row variable being at least one, and the method is the same, with the probability of the event as the divisor.
- If the variables are independent, the conditional mean equals the ordinary mean for every value.
- The frequent slip is forgetting to divide by the row total.

24 Example 12: E(Y | X = 2) and E(Y | X = 1)

Divide the row by its total, then average the y values

Row total, X = 2	0.10 + 0.15 + 0.35	**0.60**
Weights f(y | X = 2)	1/6, 1/4, 7/12	**0.17,0.25,0.58**
E(Y | X = 2)	1.45 ÷ 0.60	**2.417**
E(Y | X = 1)	0.65 ÷ 0.40	**1.625**
Check: weighted average	0.4×1.625 + 0.6×2.4167	**2.1**

IN THE LESSON

- Condition on the row variable equaling two.
- That row has cells point one zero, point one five and point three five, with row total point six zero.
- Divide each by point six: the conditional probabilities are about point one seven, point two five and point five eight.
- They add to one.
- The conditional mean of Y is one times point one, plus two times point one five, plus three times point three five, which is one point four five, divided by point six: two point four one seven.
- For the first row, the weighted sum is point six five, divided by the row total point four, giving one point six two five.
- As a check, weight these by the row totals: point four times one point six two five, plus point six times two point four one seven, equals the overall mean, two point one.

25 Independent and identically distributed

- Independent: one outcome says nothing about the others
- Identically distributed: same mean, variance and shape
- iid: fair dice, draws with replacement
- Not iid: volatility clustering, regime shifts
- Uncorrelated is not the same as independent

IN THE LESSON

- A sequence of random variables is independent and identically distributed, or I I D, when two conditions hold.
- First, the variables are mutually independent: the outcome of one carries no information about the others.
- Second, each has exactly the same distribution, with the same mean, variance and shape.
- Rolls of a fair die, or repeated draws with replacement from one population, are I I D.
- Many real series are not.
- Volatility clustering and autocorrelation make observations dependent, and draws without replacement from a small population are dependent too.
- Shifts between calm and stressed regimes break the identical part.
- And being uncorrelated is not the same as being independent.
- Risk models often assume I I D returns for convenience, so spotting where this fails is a key skill.

26 Why iid makes sums easy

$S = X_1 + \ldots + X_n$
$E(S) = n\mu$ $Var(S) = n\sigma^2$ $SD(S) = \sigma\sqrt{n}$
Var(sample mean) = $\sigma^2 \div n$

Mean of the sum	n times the single mean
Variance of the sum	n times the single variance
Square root of time	Grows as √n, not n
Sample mean	Variance falls to $\sigma^2 \div n$

IN THE LESSON

- Why do we like this assumption?
- Let a sum of n variables each have mean mu and variance sigma squared.
- The mean of the sum is n times mu, and that holds whatever the dependence between them.
- With independence, every covariance term is zero, and with identical distributions all n variances are the same, so the variance of the sum is n times sigma squared.
- The standard deviation is therefore sigma times the square root of n.
- This is why risk scales with the square root of time, not with time itself.
- Dividing the sum by n gives the sample mean, and its variance is sigma squared over n.
- Do not scale a standard deviation by n, and do not add standard deviations.

27 Example 13: sixteen iid daily returns

Illustration: mean 0.1% and σ = 1% per day, 16 days

Mean of the sum	16 × 0.1%	**1.6%**
Variance of the sum	16 × 0.0001	**0.0016**
SD of the sum	√0.0016	**4%**
Variance of sample mean	0.0001 ÷ 16	**0.00000625**
SD of sample mean	√0.00000625	**0.25%**

IN THE LESSON

- Suppose daily returns are I I D, each with a mean of point one per cent and a standard deviation of one per cent, so a variance of point zero zero zero one.
- Add sixteen days.
- The mean of the sum is sixteen times point one per cent, which is one point six per cent.
- The variance is sixteen times point zero zero zero one, or point zero zero one six.
- The standard deviation is the square root, four per cent.
- Four, not sixteen: it grows with the square root of the number of days.
- Now the sample mean, the sum divided by sixteen.
- Its variance is point zero zero zero one divided by sixteen, which is point zero zero zero zero zero six two five, and its standard deviation is point two five per cent.

28 Example 14: when iid fails

Illustration: same 16 days, but every pair has correlation 0.2

Extra term $n(n-1)\rho\sigma^2$	240 × 0.2 × 0.0001	**0.0048**
Variance with correlation	0.0016 + 0.0048	**0.0064**
SD of the sum	√0.0064	**8%**
iid shortcut said	√0.0016	**4%**
Shortcut understates risk	4% against 8%	**by half**

IN THE LESSON

- Now see how badly the shortcut can mislead.
- Suppose the sixteen daily returns share a correlation of point two.
- The variance of the sum gains an extra term: n times n minus one, times the correlation, times the variance.
- That is sixteen times fifteen, two hundred and forty, times point two, times point zero zero zero one, which is point zero zero four eight.
- Add the I I D variance of point zero zero one six, and the total is point zero zero six four.
- The standard deviation is the square root, eight per cent.
- The I I D shortcut said four per cent.
- So a small correlation of point two doubled the risk of the sum, and the shortcut understated it by half.

29 Four mistakes to avoid

- Joint sums to 1 overall; divide by the GIVEN total
- Cov = E(XY) − E(X)E(Y); zero Cov is not independence
- Square the weights, keep 2ab·Cov, respect the signs
- SD of a sum grows as √n, not n; check the iid claim

IN THE LESSON

- Mistake one: forcing each row or column of the joint matrix to sum to one, or dividing by the wrong total when conditioning.
- The whole matrix sums to one, and you divide by the total of the variable that is given.
- Mistake two: forgetting to subtract the product of the means in the covariance, or believing that zero covariance proves independence.
- It does not.
- Mistake three: errors in the variance of a weighted sum.
- Square the weights, keep the doubled cross term, and respect the signs.
- Mistake four: scaling a standard deviation by n, or assuming I I D when the data have clustering or correlation.
- For a sum, the standard deviation grows with the square root of n.

30 Recap: the main ideas

- Cells sum to 1; marginal = total; conditional = ÷ given
- E[g(X, Y)] = Σ g × f; Cov = E(XY) − E(X)E(Y)
- ρ = Cov ÷ (σ_X σ_Y); independent ⇒ 0, not reverse
- Var(aX + bY) = a²Var(X) + b²Var(Y) + 2ab·Cov
- iid sum: mean nμ, variance nσ², SD σ√n

IN THE LESSON

- Let's pull it together.
- A joint P M F is a matrix whose cells sum to one.
- Marginals are row and column totals, and conditionals divide a cell by the total of the given variable.
- Expected values of functions are sums over all cells, weighted by probability.
- Covariance is the expected product minus the product of the means, and correlation rescales it into the range from minus one to plus one.
- Independence gives zero covariance, but zero covariance does not give independence.
- Linear changes scale the covariance and leave correlation alone, up to a sign.
- The variance of a weighted sum includes a doubled covariance term.
- And for I I D variables, the mean of a sum is n mu, and the standard deviation is sigma root n.

31 Try three quick questions

1. f(1,1)=0.1, f(1,2)=0.3, f(2,1)=0.2. f(2,2)? P(Y=2 | X=2)?
   Answer: f(2,2) = 0.4. P = 0.4 ÷ 0.6 = 0.67.
2. Var(X) = 4, Var(Y) = 9, Cov = −3. Var(X − Y)?
   Answer: 4 + 9 − 2 × (−3) = 19.
3. iid daily returns, σ = 2%. SD of a 25-day sum?
   Answer: 2% × √25 = 10%, not 50%.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- A two by two matrix has f one one equal to point one, f one two equal to point three and f two one equal to point two.
- What is the last cell, and what is the chance that Y equals two, given that X equals two?
- The last cell is point four, and the conditional chance is point four over point six, about point six seven.
- Question two.
- Variances are four and nine and the covariance is minus three.
- What is the variance of the difference?
- Thirteen plus six is nineteen.
- Question three.
- Daily returns are I I D with a standard deviation of two per cent.
- What is the standard deviation over twenty five days?
- Two times five, ten per cent.