CHARTA LEARNING · LESSON NOTES

Random Variables

FRM Part I

1 A story: how many loans will default?

- A lender holds four loans: how many default next year?
- The count could be 0, 1, 2, 3 or 4, each with a chance
- A random variable puts a number on an uncertain outcome
- We need: a full description, a summary, and the tail

IN THE LESSON

- Picture a small lender holding four loans.
- Nobody knows how many will go bad next year.
- It could be none, one, two, three or all four.
- That count is uncertain, but it is not a mystery, because we can attach a chance to each possible count.
- A quantity whose value is settled by chance, and that we describe by listing its possible values with their chances, is called a random variable.
- Risk management is full of them: the number of defaults, tomorrow's return, next year's loss.
- To manage risk we need three things.
- A full description of the chances, a short summary of the centre and the spread, and a way to read off the bad outcomes in the tail.

2 Discrete and continuous variables

- Discrete: separate values you can list (defaults, notches)
- Continuous: any value in a range (returns, losses, rates)
- Discrete: the chance of an exact value makes sense
- Continuous: needs a different tool, the density

IN THE LESSON

- There are two families of random variable, and almost every rule in this lesson depends on which one you have.
- A discrete variable takes separate values that you could list, such as the number of defaults, or the number of rating notches a bond falls.
- A continuous variable can take any value in a range, such as a return, a loss measured in dollars, or an interest rate.
- For a discrete variable we can talk about the chance of an exact value.
- For a continuous variable we cannot, and we will see why later.
- First we work with the discrete case, using our four loans, because the arithmetic is small enough to see everything.

3 The probability mass function

Illustration: defaults among four loans, called X

Defaults x	Probability p(x)
0	0.40
1	0.30
2	0.15
3	0.10
4	0.05

IN THE LESSON

- Here is the probability mass function for our four loans.
- It is just a table.
- For each possible number of defaults, it gives the probability of exactly that number.
- The chance of no defaults is forty per cent, of one default thirty per cent, of two default fifteen per cent, of three ten per cent, and of all four five per cent.
- These numbers are an illustration, not real data.
- Two properties always hold.
- Every probability lies between zero and one, and across all possible values they add up to exactly one, because something must happen.
- The mass function is written p of x, and it means the probability that the variable equals exactly x.

4 Mass function and cumulative function

p(x) = P(X = x)
F(x) = P(X ≤ x) = sum of p(t), t ≤ x

Mass function, p(x)	Chance of exactly x
Cumulative, F(x)	Running total through x
Jump size	p(x) = F(x) − F(just below x)
Total is one	All the masses add to 1

IN THE LESSON

- Now meet the second description.
- The cumulative distribution function, written capital F of x, is the probability that the variable is at most x.
- It is a running total of the mass, adding every value up to and including x.
- So the two are linked in both directions.
- Going forward, F of x is the sum of the masses up to x.
- Going back, the mass at x is the size of the jump in F at x: F at x minus F just below x.
- The masses add to one, so F climbs from zero all the way to one.
- The mass function gives the probability of an exact value.
- The cumulative function gives the probability of landing at or below a value.

5 Example 1: build the CDF from the PMF

Illustration: F(x) is the running total of p(x)

F(0) = p(0)	**0.40**
F(1) = 0.40 + 0.30	**0.70**
F(2) = 0.70 + 0.15	**0.85**
F(3) = 0.85 + 0.10	**0.95**
F(4) = 0.95 + 0.05	**1.00**

IN THE LESSON

- Let's build the cumulative function for our four loans, one running total at a time.
- At zero, F is just the mass at zero: point four zero.
- At one, add point three zero to get point seven zero.
- At two, add point one five to get point eight five.
- At three, add point one zero to get point nine five.
- At four, add the last point zero five, and we reach one point zero zero.
- Notice what this says in words.
- There is a seventy per cent chance of one default or fewer, and an eighty five per cent chance of two or fewer.
- The running total never falls and ends at one, which is a useful check on any cumulative function you build.

6 The CDF is a staircase



IN THE LESSON

- Here is that running total drawn as a picture.
- Along the bottom is the number of defaults, and up the side is the cumulative probability.
- For a discrete variable the cumulative function is a step function.
- It is flat between the possible values, and it jumps up at each possible value.
- The height of each jump is the mass there: at zero it jumps by point four, at one by point three, at two by point one five.
- The steps never go down, and they climb from zero to one.
- Because it is flat between values, the cumulative function at two and a half equals its value at two, point eight five.
- The dots show the value at each whole number, because the function includes the upper value.

7 Example 2: probabilities of intervals

Illustration: using F(x) for intervals and exact values

P(1 < X ≤ 3) = F(3) − F(1)	0.95 − 0.70	**0.25**
P(1 ≤ X ≤ 3): add p(1)	0.25 + 0.30	**0.55**
P(1 ≤ X < 3): less p(3)	0.55 − 0.10	**0.45**
p(2) = F(2) − F(1)	0.85 − 0.70	**0.15**
F(2.5) = F(2)	flat between values	**0.85**

IN THE LESSON

- Now use the cumulative function to answer interval questions.
- First, the chance that the count is more than one and at most three.
- Take F of three, point nine five, and subtract F of one, point seven zero: point two five.
- The cumulative function includes its upper value and excludes the lower.
- Next, one up to three, both included.
- We must add back the mass at one, point three zero, giving point five five.
- Then one up to but not including three, so we subtract the mass at three, point one zero, giving point four five.
- Fourth, the mass at two is F of two minus F of one: point eight five minus point seven zero, point one five.
- Finally F of two and a half is F of two, point eight five.

8 Watch the endpoints

- F includes its upper value, not the value below it
- $P(a < X \le b) = F(b) - F(a)$
- $P(a \le X \le b) = F(b) - F(a) + p(a)$
- $P(a \le X < b) = F(b) - F(a) + p(a) - p(b)$
- Write the inequality signs first, then check each end

IN THE LESSON

- The most common slip with the cumulative function is the wrong endpoint, so let's fix the rules in words.
- The function includes the upper value but not the lower one.
- So F of b minus F of a gives the probability of being above a and at most b.
- If you want a included as well, add the mass at a.
- If you want b excluded, subtract the mass at b.
- For a continuous variable none of this matters, because single points carry no probability, but for a discrete variable every endpoint counts.
- A good habit is to write the inequality signs out before you touch any numbers, and then check each endpoint against the rule.

9 The expected value

$E[X] = \Sigma\, x \times p(x)$
$E[aX + bY + c] = a \cdot E[X] + b \cdot E[Y] + c$

Expected value	Probability-weighted average
Value, x	Each possible value
Probability, p(x)	Its probability
Linearity	True even if X and Y are dependent

IN THE LESSON

- The second big idea is expectation.
- The expected value of a random variable is its probability weighted average.
- For a discrete variable, multiply each possible value by its probability, and add everything up.
- For a continuous variable the sum becomes an integral of the value times the density.
- Think of it as the long run average, the balance point of the distribution.
- It does not have to be a value the variable can actually take: the average roll of a die is three and a half.
- And it is not in general the median or the mode.
- Its most useful property is linearity.
- The expected value of a times one variable plus b times another plus a constant is a times the first mean, plus b times the second mean, plus the constant, whether or not the two variables are independent.

10 Example 3: the expected number of defaults

Illustration: $E[X] = \Sigma\, x \times p(x)$

$0 \times 0.40 + 1 \times 0.30$		**0.30**
2×0.15		**0.30**
3×0.10		**0.30**
4×0.05		**0.20**
Expected defaults	0 + 0.30 + 0.30 + 0.30 + 0.20	**1.10**

IN THE LESSON

- Let's compute the expected number of defaults for our four loans.
- Multiply each value by its probability.
- Zero times point four zero is zero, and one times point three zero is point three zero, so the first row shows point three zero.
- Two times point one five is point three zero.
- Three times point one zero is point three zero.
- Four times point zero five is point two zero.
- Add them: zero plus point three, plus point three, plus point three, plus point two, gives one point one zero.
- So the expected number of defaults is one point one.
- Notice that no year can actually produce one point one defaults.
- The expected value is an average over many possible years, not a prediction of any single one.

11 The rules of expectation

- Linear always: E[aX + bY + c] = a·E[X] + b·E[Y] + c
- Not multiplicative: E[XY] = E[X]·E[Y] + Cov(X, Y)
- Nonlinear g: E[g(X)] is usually not g(E[X])
- Total expectation: weight the conditional means

IN THE LESSON

- Four rules are worth memorizing, and the mean need not be a value the variable can take.
- First, expectation is linear.
- You can take the expected value of a combination by combining the expected values, and it works with or without independence.
- Second, it is not multiplicative.
- The expected value of a product is the product of the expected values, plus the covariance of the two variables.
- Only when the covariance is zero do the two agree.
- Third, and most important for risk, if you apply a nonlinear function to a variable, the expected value of the result is usually not the function of the expected value.
- Option payoffs are the classic case.
- Fourth, the law of total expectation: the overall mean is the probability weighted average of the conditional means.

12 Example 4: a nonlinear payoff

Illustration: price 80, 100, 120; call strike 100

Average price E[S]	80×0.25+100×0.50+120×0.25	**100**
Payoff at the average price	max(100 − 100, 0)	**0**
Average payoff E[g(S)]	0×0.25+0×0.50+20×0.25	**5**
The two differ	5 is not 0	**5 vs 0**

IN THE LESSON

- Here is the trap in numbers.
- A stock ends the year at eighty, one hundred or one hundred twenty, with probabilities one quarter, one half and one quarter.
- A call option, a contract giving the right to buy at one hundred, pays the stock price minus one hundred if that is positive, and zero otherwise.
- The expected stock price is twenty plus fifty plus thirty: one hundred.
- Plug that into the payoff and you get zero.
- But the payoffs are zero, zero and twenty, so the average payoff is twenty times one quarter: five.
- The payoff at the average price is zero, the average payoff is five.
- They differ, because the payoff is nonlinear.
- Average the function, never apply it to the average.

13 Example 5: total expectation and linearity

Illustration: regimes, then a linear gain

Calm: probability 0.80	conditional mean	**2%**
Stress: probability 0.20	conditional mean	**−5%**
Overall mean return	0.80×2% + 0.20×(−5%)	**0.6%**
Gain = 5 + 1,000 × return	5 + 1,000 × 0.006	**11**

IN THE LESSON

- Next, the law of total expectation.
- Suppose next year is either calm, with probability eighty per cent, or stressed, with probability twenty per cent.
- In a calm year the expected return is two per cent.
- In a stressed year it is minus five per cent.
- The overall mean is the weighted average of the two conditional means: point eight times two, plus point two times minus five, which is one point six minus one, so point six per cent.
- Now linearity.
- A fund earns a fixed fee of five, plus one thousand times the return.
- Its expected gain is five plus one thousand times point zero zero six, which is eleven.
- We only needed the mean return, not the whole distribution.

14 The four population moments

Four numbers describe most of a distribution

Moment	What it describes
Mean (first)	Location, the centre
Variance (second central)	Spread around the mean
Skewness (third, scaled)	Asymmetry, longer tail
Kurtosis (fourth, scaled)	Tail weight, extremes

IN THE LESSON

- A distribution has many details, but four numbers capture most of its character, and they are called the four commonly used population moments.
- The first is the mean, which describes location, the centre.
- The second central moment is the variance, which describes spread; its square root is the standard deviation.
- The third is skewness, which describes asymmetry, which side has the longer tail.
- The fourth is kurtosis, which describes the weight of the tails, how often extreme outcomes occur.
- Central means we measure deviations from the mean.
- For skewness and kurtosis we also standardize, by dividing by sigma to the third and sigma to the fourth, so that the numbers do not depend on the units.

15 The moment formulas

Mean = E(X) Variance = E[(X − μ)²]
Skew = E[(X − μ)³] ÷ σ³ Kurt = E[(X − μ)⁴] ÷ σ⁴

Mean	Centre of the distribution
Variance	Average squared deviation
Std deviation	Square root of variance
Skewness	Zero if symmetric
Kurtosis	Three for a normal

IN THE LESSON

- Here are the four formulas.
- The mean is the expected value of the variable.
- The variance is the expected squared deviation from the mean, mu.
- The standard deviation, sigma, is its square root, and it is in the same units as the variable.
- Skewness is the expected cubed deviation, divided by sigma cubed.
- It is zero for a symmetric distribution, and negative when the left tail is longer.
- Kurtosis is the expected fourth power deviation, divided by sigma to the fourth.
- For a normal distribution it equals three, and excess kurtosis is kurtosis minus three.
- A common slip is forgetting the division by sigma cubed or sigma to the fourth, or forgetting to subtract three when excess kurtosis is asked.

16 Example 6: variance of the defaults

Illustration: $Var(X) = E[X^2] - (E[X])^2$

Mean E[X]	from Example 3	**1.10**
E[X²]	0 + 0.30 + 0.60 + 0.90 + 0.80	**2.60**
Variance	$2.60 - 1.10^2$	**1.39**
Standard deviation	$\sqrt{1.39}$	**1.18**

IN THE LESSON

- Let's find the spread of our default count.
- We already know the mean is one point one.
- The easiest route is the shortcut: variance equals the expected square minus the square of the mean.
- First, the expected square.
- Square each value, multiply by its probability, and add: zero, plus point three, plus point six, plus point nine, plus point eight.
- That is two point six zero.
- Square the mean: one point one squared is one point two one.
- The variance is two point six zero minus one point two one: one point three nine.
- Taking the square root, the standard deviation is about one point one eight defaults.
- If you compute it the long way, with squared deviations from the mean, you get exactly the same answer.

17 Example 7: skewness and kurtosis

Illustration: standardized third and fourth moments

Third central moment	$\Sigma\ p \times (x - 1.1)^3$	**1.48**
Skewness	$1.48 \div \sigma^3$ (1.64)	**0.90**
Fourth central moment	$\Sigma\ p \times (x - 1.1)^4$	**5.52**
Kurtosis	$5.52 \div \sigma^4$ (1.93)	**2.86**
Excess kurtosis	$2.86 - 3$	**−0.14**

IN THE LESSON

- Now the third and fourth moments for the same variable.
- Take each deviation from the mean, raise it to the third power, weight by its probability and add.
- The third central moment comes to one point four eight.
- The standard deviation cubed is about one point six four, so the skewness is about point nine zero.
- It is positive: the long tail is on the right, where many defaults sit.
- For kurtosis, use the fourth power instead.
- The fourth central moment is about five point five two.
- Divide by sigma to the fourth, which is the variance squared, one point nine three, and kurtosis is about two point eight six.
- Minus three gives excess kurtosis of about minus point one four, slightly thinner tails than a normal.

18 Reading skewness and kurtosis

- Skewness 0 if symmetric; negative means a longer left tail
- Kurtosis is 3 for the normal; excess = kurtosis − 3
- High kurtosis: fatter tails, more extreme outcomes
- Same mean and sigma can hide very different tails
- Our defaults: skew +0.90, kurtosis 2.86, excess −0.14

IN THE LESSON

- How do you read these two numbers?
- Skewness tells you which side the extremes lie on.
- A symmetric distribution has skewness zero.
- Negative skewness means a longer left tail, which for returns means the occasional large loss; positive skewness means a longer right tail.
- Kurtosis tells you how heavy the tails are.
- The normal distribution has kurtosis three, so excess kurtosis is measured against that benchmark.
- Higher kurtosis means fatter tails and more extreme outcomes than a normal model implies.
- This is why risk managers care: two return series can share the same mean and the same standard deviation, and still differ greatly in how often they produce a disaster.
- Our defaults had slight positive skew and kurtosis just below three.

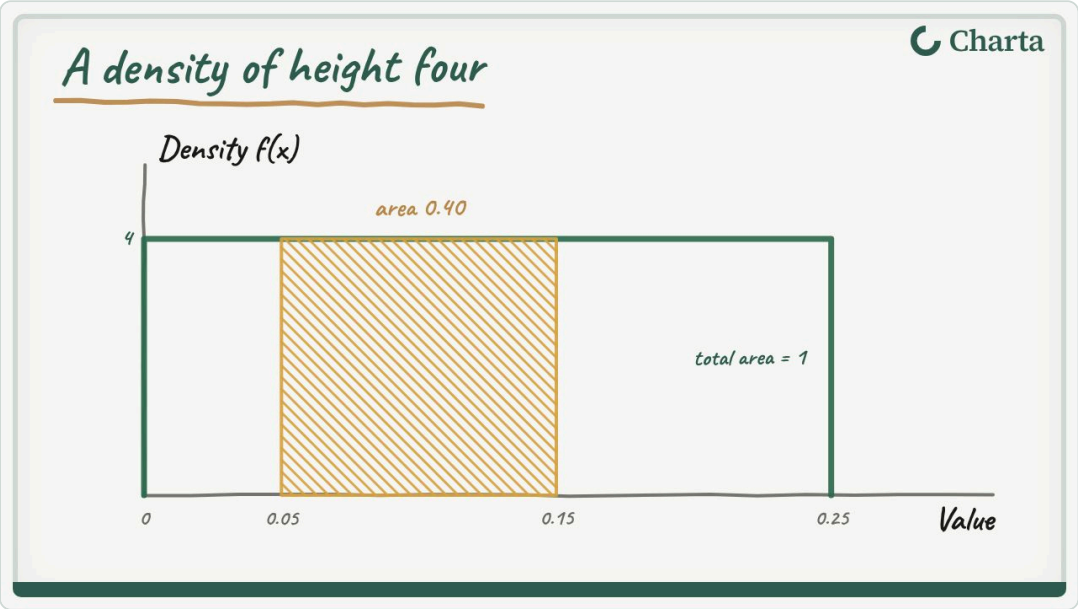
19 Mass versus density

- PMF: discrete; each value carries a real probability
- PDF: continuous; the height is not a probability
- P(X = any single value) = 0 for a continuous variable
- Probability = area under the curve; total area = 1
- A density can exceed 1; a probability never can

IN THE LESSON

- Now the continuous case.
- A probability mass function belongs to a discrete variable, and gives a genuine probability for each possible value, so its entries lie between zero and one and add to one.
- A probability density function belongs to a continuous variable, and it is a different kind of object.
- The chance that a continuous variable equals any single exact value is zero, because a single point has zero width, so it has zero area under the density.
- Instead, probability is the area under the density curve between two points, and the total area is one.
- A density value is a height, not a probability, so it can exceed one.
- The cumulative function is the area up to a point, and the density is the slope of the cumulative function.

20 A density of height four



IN THE LESSON

- Here is the simplest density, a uniform one.
- Suppose a quantity is equally likely to be anywhere between zero and point two five, for example a return somewhere in that range.
- The curve is a flat line, a rectangle.
- The width is point two five, and the area must be one, so the height has to be one divided by point two five, which is four.
- A height of four is not a probability of four, which would be nonsense.
- It is a density.
- Probability is area.
- The shaded slice between point zero five and point one five has width point one, so its area is four times point one, point four zero.
- The chance of any exact point, such as point one, has zero width, so zero area.

21 Example 8: from density to probability

Illustration: uniform between 0 and 0.25

Height of the density	1 ÷ 0.25	**4**
P(0.05 < X < 0.15)	4 × 0.10	**0.40**
P(X = 0.10)	width of zero	**0**
F(0.20) = P(X ≤ 0.20)	4 × 0.20	**0.80**
Density = slope of F	0.80 ÷ 0.20	**4**

IN THE LESSON

- Let's turn this density into answers.
- The height is one divided by the width, one over point two five, which is four.
- The probability of landing between point zero five and point one five is height times width: four times point one, point four zero.
- The probability of exactly point one is zero, because the width of a single point is zero.
- The cumulative function at point two zero is the area up to there: four times point two zero, point eight zero.
- And going the other way, the density is the slope of the cumulative function.
- It rises point eight zero over a width of point two zero, so the slope is four, matching the height.
- Remember: read the area, never the height, as the probability.

22 PMF and PDF side by side

Same questions, two kinds of variable

Feature	PMF | PDF
Variable type	Discrete | Continuous
Value at a point	Probability | Height
Probability of a range	Add | Area
Total	Sums to 1 | Area is 1
Link to F	Running sum | Integral

IN THE LESSON

- Let's put the two objects side by side.
- The mass function is for discrete variables, the density for continuous ones.
- At a point, the mass function gives a probability, while the density gives only a height.
- To find the probability of a range, you add the masses for the discrete variable, but you find an area, an integral, for the continuous one.
- Both total one: the masses sum to one, and the area under the density is one.
- And both connect to the cumulative function.
- For the discrete variable F is a running sum of the masses.
- For the continuous variable F is the integral of the density.
- The usual mistake is reading a density value as a probability.

23 The quantile function

q(p) = F^{−1}(p), so F(q(p)) = p
Normal: q(p) = μ + z × σ

Quantile, q(p)	Value with probability p below it
Probability, p	A probability between 0 and 1
Normal z-value	1.645 at 95%, 2.326 at 99%
Mean, mu	Centre of the normal
Sigma	Scale of the normal

IN THE LESSON

- The last way to describe a distribution is the quantile function.
- It is the inverse of the cumulative function.
- You give it a probability, p, and it returns the value below which a fraction p of the probability lies.
- So the median is the quantile at one half, and a ninety nine per cent value at risk is the ninety ninth percentile of the loss distribution.
- For a normal variable with mean mu and standard deviation sigma, the quantile equals mu plus z times sigma, where z is the standard normal quantile.
- One tailed, z is one point six four five at ninety five per cent and two point three two six at ninety nine.
- For a discrete variable, the quantile is the smallest value whose cumulative probability is at least p.

24 Example 9: a discrete quantile

Illustration: F = 0.40, 0.70, 0.85, 0.95, 1.00

Median q(0.50)	F(0)=0.40, F(1)=0.70	**1**
q(0.70)	F(1) = 0.70 reaches 0.70	**1**
q(0.90)	F(2)=0.85, F(3)=0.95	**3**
q(0.99)	only F(4) = 1.00 reaches it	**4**

IN THE LESSON

- Let's find quantiles for our default count, where the cumulative values are point four, point seven, point eight five, point nine five and one.
- The rule is the smallest value whose cumulative probability reaches at least p.
- For the median, p is point five.
- At zero, F is only point four, too small.
- At one, F is point seven, which clears it.
- So the median is one.
- For p equal to point seven, F at one is exactly point seven, which reaches it, so the quantile is again one.
- For p equal to point nine, F at two is point eight five, too small, but at three it is point nine five, so the answer is three.
- For p equal to point nine nine, only four reaches it.
- The answer is four.

25 Example 10: normal quantiles

Illustration: loss ~ normal, mean 10, sigma 4

95th percentile, z = 1.645	10 + 1.645 × 4	**16.58**
99th percentile, z = 2.326	10 + 2.326 × 4	**19.30**
Median, z = 0	10 + 0 × 4	**10**
5th percentile, z = −1.645	10 − 1.645 × 4	**3.42**

IN THE LESSON

- Now a normal case, the one behind many value at risk calculations.
- Suppose a yearly loss is normal with mean ten and standard deviation four, in millions.
- These are illustrative numbers.
- The ninety fifth percentile uses z of one point six four five: ten plus one point six four five times four, which is sixteen point five eight.
- The ninety ninth uses two point three two six: ten plus nine point three zero four is nineteen point three zero.
- The median has z equal to zero, so it is just the mean, ten.
- And watch the direction.
- For lower tail quantiles, such as the fifth percentile, use the negative z: ten minus one point six four five times four, which is three point four two.

26 Example 11: quantile-based estimators

Illustration: quartiles −4, 3 and 5 for a return

IQR = q(0.75) − q(0.25)	5 − (−4)	**9**
Lower gap: median − q(0.25)	3 − (−4)	**7**
Upper gap: q(0.75) − median	5 − 3	**2**
Robust skewness	(5 + (−4) − 2×3) ÷ 9	**−0.56**

IN THE LESSON

- Quantiles also give estimators that are robust to outliers, because one extreme value does not move the quartiles much.
- The interquartile range, or I Q R, measures spread: the upper quartile minus the lower quartile.
- Suppose a return series has a lower quartile of minus four, a median of three and an upper quartile of five.
- The range is five minus minus four, nine.
- For robust skewness, compare the gaps either side of the median.
- The lower gap is three minus minus four, seven.
- The upper gap is five minus three, two.
- The measure is upper quartile plus lower quartile minus twice the median, over the range: minus five over nine, minus point five six.
- Negative, because the lower quartile sits further from the median.

27 A linear change of a variable

$Y = a + b \times X$ Mean: $a + b \cdot E(X)$
Variance: $b^2 \cdot Var(X)$ Std dev, IQR: $|b| \times$ old

Shift, a	Moves the centre, not the spread
Multiplier, b	Stretches spread by size of b
Location	Mean and median: a + b × centre
Spread	Spread: $|b| \times \sigma$, $b^2 \times$ variance
Shape	Skewness and kurtosis

IN THE LESSON

- The final topic is what happens when you transform a random variable linearly: new variable Y equals a plus b times X.
- Here a is a shift and b is a multiplier.
- This is how you convert units, rescale a loss, or flip gains into losses.
- The rules are simple.
- The mean becomes a plus b times the old mean, and the median follows the same rule.
- The variance becomes b squared times the old variance, so the shift a has no effect on dispersion.
- The standard deviation becomes the absolute value of b times sigma.
- The interquartile range scales the same way, by the absolute value of b.
- Skewness and kurtosis are standardized moments, so they ignore the shift and the size of the multiplier.

28 What a linear change does

For $Y = a + bX$, where a is the shift and b the multiplier

Measure	Becomes
Mean	a + b × E(X)
Median	a + b × median(X)
Standard deviation	$|b| \times \sigma$
Variance	$b^2 \times Var(X)$
Interquartile range	$|b| \times IQR$

IN THE LESSON

- Here is the summary for the location and scale measures.
- The mean becomes a plus b times the mean.
- The median becomes a plus b times the median, for either sign of b.
- The standard deviation becomes the absolute value of b times sigma, never negative, because spread cannot be negative.
- The variance becomes b squared times the variance.
- And the interquartile range becomes the absolute value of b times the original.
- Two points to remember.
- The shift a affects only location, so it never enters the variance, the standard deviation or the range.
- And if b is negative, the sign is dropped for spread.
- The usual errors are carrying the sign of b into a standard deviation, and adding a to a variance.

29 Shape under a linear change

- Skewness and kurtosis ignore a and the size of b
- Kurtosis never changes
- Skewness: unchanged if $b > 0$, flips sign if $b < 0$
- Negative b mirrors the distribution: quantiles swap places
- The median still maps to a + b × median, either sign

IN THE LESSON

- What about shape?
- Kurtosis and skewness are standardized, so the shift and the size of the multiplier cancel out.
- Kurtosis is therefore unchanged by any linear transformation.
- Skewness is unchanged when b is positive.
- When b is negative, the distribution is mirrored, the long tail swaps sides, and skewness changes sign.
- The same mirroring explains one more point about quantiles.
- The median still maps to a plus b times the median for either sign of b, but other quantiles change places when b is negative, because the order of the values reverses.
- The old upper quartile becomes the new lower quartile.
- Check the sign of b before you decide which quantile you need.

30 Example 12: Y = 5 − 2X

Illustration: Y = 5 − 2X, with X the default count

Mean	5 − 2 × 1.10	**2.80**
Variance	(−2)² × 1.39	**5.56**
Standard deviation	|−2| × 1.18	**2.36**
Skewness (b < 0: flips)	was +0.90	**−0.90**
Kurtosis (unchanged)	stays 2.86	**2.86**

IN THE LESSON

- Let's apply a linear change to our default count.
- Say a lender books a gain of five, minus two for every default, so Y equals five minus two times X.
- The old mean was one point one, so the new mean is five minus two point two, which is two point eight zero.
- The old variance was one point three nine.
- The new variance is minus two squared, four, times one point three nine, which is five point five six.
- The standard deviation uses the absolute value of minus two, two times one point one eight, which is two point three six.
- Skewness was positive point nine zero.
- The multiplier is negative, so it flips to minus point nine zero.
- And kurtosis stays at two point eight six.

31 Four mistakes to avoid

- Wrong endpoint: F includes b, so add p(a) when a counts
- Function of the mean: average the function, g(X), instead
- Density read as probability: use the area under it
- Linear change: |b| for σ, b² for variance, never add a

IN THE LESSON

- Four mistakes cause most errors in this topic.
- One, the wrong endpoint.
- The cumulative function includes the upper value, so add the mass at a when a is included, and subtract the mass at b when b is excluded.
- Two, applying a nonlinear function to the mean.
- Average the function over the distribution instead, as with the option payoff.
- Three, reading a density value as a probability.
- The probability of a range is the area, and the height can be greater than one.
- Four, mishandling a linear change.
- Spread uses the absolute value of b and the square of b for variance, and the shift a never goes into the variance.
- Keep these four checks in mind.

32 Recap

- PMF: chance of each value; CDF: running total
- E[X] = Σ x·p(x); linear, but E[g(X)] is not g(E[X])
- Mean, variance, skewness, kurtosis (excess = kurtosis − 3)
- PDF height is not a probability; area is
- Quantile q(p) inverts F; Y = a + bX: |b|σ, b²·Var

IN THE LESSON

- Let's recap the module.
- The probability mass function gives the chance of each exact value of a discrete variable, the cumulative function is its running total, and each can be rebuilt from the other.
- The expected value is the probability weighted average.
- It is linear, but not multiplicative, and the mean of a nonlinear function is not the function of the mean.
- The four moments are the mean, variance, skewness and kurtosis, with the last two standardized.
- A density is a height, and probability is area.
- The quantile function inverts the cumulative function, giving value at risk.
- And a linear change alters the measures in simple, predictable ways.

33 Try three quick questions

1. X is 1, 2 or 3; F(1)=0.2, F(2)=0.6. Find p(2), P(1<X≤2)
   Answer: p(2) = F(2) − F(1) = 0.6 − 0.2 = 0.4. Same as P(1 < X ≤ 2).
2. X has mean 4 and σ = 3. Find mean and σ of 2 − 5X.
   Answer: Mean = 2 − 5×4 = −18. σ = |−5| × 3 = 15.
3. Return ~ normal, mean 2, σ 5. What is the 1st percentile?
   Answer: 2 − 2.326 × 5 = −9.63. Lower tail uses the negative z.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- A discrete variable has cumulative values of point two at one, point six at two and one at three.
- What is the mass at two, and the chance of more than one and at most two?
- Both are point six minus point two, point four.
- Question two.
- A variable has a mean of four and a standard deviation of three.
- What are the mean and standard deviation of two minus five times the variable?
- The mean is two minus twenty, minus eighteen, and the standard deviation is fifteen.
- Question three.
- A normal return has a mean of two and sigma of five.
- What is the first percentile?
- Two minus two point three two six times five, minus nine point six three.