CHARTA LEARNING · LESSON NOTES

Statistical distributions

CFA Level I · Quantitative Methods

You will learn: Match the right distribution to the question: uniform, binomial, normal, lognormal, Student's t, chi-square or F, and compute with each.

Running example: Examples use made-up figures, such as a manager who beats the market in 4 of 5 quarters (60% chance each quarter if skilled) and a portfolio with an expected return of 8% and a standard deviation of 12%.

1. Distributions and the discrete uniform

- A distribution is a list of everything that could happen, with how likely each outcome is. Discrete outcomes are counted (a die roll); continuous outcomes are measured (a return).
- Two rules always hold: every probability is between 0 and 1, and all the probabilities add up to exactly 1.
- The discrete uniform distribution makes every outcome equally likely: n outcomes, each with probability 1 ÷ n. A loaded die is not uniform.

Discrete uniform: each outcome has probability 1 ÷ n

ONE ROLL OF A FAIR SIX-SIDED DIE

Any single face	1 out of 6	**0.1667**
Four or less	4 faces out of 6	**0.6667**
An even number	2, 4 or 6: 3 faces	**0.5**

2. The binomial distribution

- It counts successes in a fixed number of tries. It needs: a fixed number of tries n; two results per try (success or failure); the same chance of success p every time; independent tries.
- C(n, k) counts the different orders in which the successes can happen. The mean is n × p; the variance is n × p × (1 − p); the standard deviation is its square root.
- In finance both assumptions are shaky: price moves are not truly independent and default chances change with the economy. Treat it as a model. It still underlies binomial option pricing, where a price moves up or down by a factor each period.

$P(X = k) = C(n, k) \times p^k \times (1 - p)^{n-k}$. Mean = n × p. Variance = n × p × (1 − p)

FIVE QUARTERS; WIN EACH INDEPENDENTLY

Skilled (p = 0.6): mean, variance, sd	5 × 0.6; 5 × 0.6 × 0.4; √1.2	**3.0; 1.2; 1.10**
Exactly four wins	C(5, 4) = 5; 5 × 0.6^4 × 0.4	**0.2592**
All five wins	0.6^5	**0.0778**
Four or more	0.2592 + 0.0778	**0.3370**
No skill (p = 0.5): four or more	5 ÷ 32 + 1 ÷ 32 = 6 ÷ 32	**0.1875**

Watch out: Using the binomial when tries are not independent, and writing the variance as n × p instead of n × p × (1 − p).

3. The normal distribution and the z-score

- The normal (bell curve) is continuous and symmetric. The mean is its middle and the standard deviation sets its width; these two numbers describe it completely. About 68% of outcomes fall within 1 standard deviation of the mean, about 95% within 2, about 99% within 3.
- The z-score counts the distance from the mean in standard deviations. Converting to z gives the standard normal (mean 0, standard deviation 1), so one table works for every normal.
- Real returns have fatter tails than the normal curve, so normal-based probabilities understate how often extreme events happen.

$z = (x - \mu) \div \sigma$

PORTFOLIO: MEAN 8%, STANDARD DEVIATION 12%, RETURNS NORMAL		
Chance of −10% or worse	z = (−10 − 8) ÷ 12 = −1.50	**6.68%**
Chance of beating +20%	z = +1.00; 1 − 0.8413	**15.87%**
Between −4% and +20%	z from −1 to +1	**68.27%**
Within 2 / 3 standard deviations	−16% to +32% / −28% to +44%	**95.45% / 99.73%**
A year of −28% or worse	z = −3; about 1 year in 740	**0.135%**

Watch out: Treating the normal curve as the truth.

4. Continuous compounding, log returns and the lognormal

- Continuous compounding adds interest at every instant. £1 for one year at a continuously compounded rate r grows to e^r, where e is about 2.718. The effective annual rate is e^r − 1.
- A continuously compounded (log) return is the natural logarithm of the price relative, ending price ÷ starting price. Log returns add over time, simple returns do not.
- A variable is lognormal when its natural log is normal. If log returns are normal, prices are lognormal: a floor at zero and a long tail to the right. This underlies the Black-Scholes option pricing model.

$r = \ln(S_1 \div S_0)$. $S_1 = S_0 \times e^{r}$

£100 AT 6% FOR ONE YEAR; AND A ROUND TRIP 100 → 120 → 100		
Yearly / quarterly / monthly	100×1.06; 1.015^4; 1.005^{12}	**£106.00 / £106.14 / £106.17**
Continuously	100 × e^0.06	**£106.18**
Round trip, simple returns added	+20.00% and −16.67%	**+3.33% (wrong)**
Round trip, log returns added	ln(1.20) + ln(100 ÷ 120) = 0.1823 − 0.1823	**0.0000**
+10% then +5%: log returns	ln(1.10) + ln(1.05) = 0.0953 + 0.0488	**0.1441**
Final price	100 × e^0.1441	**£115.50**

Watch out: Adding simple returns over several periods. Add log returns, or multiply growth factors.

5. Student's t distribution

- Use it for a mean when the true standard deviation is unknown and estimated from a small sample. It is a symmetric bell like the normal, but with fatter tails.
- Degrees of freedom for a mean are sample size − 1. With few degrees of freedom the tails are very fat; as they grow, t approaches the normal. With a small sample, the t gives a wider interval.

Degrees of freedom = n − 1

95% TWO-SIDED CRITICAL VALUES, AND A T INTERVAL (16 YEARS, MEAN 7%, SAMPLE SD 8%)		
Normal		**1.960**
t with 5 / 30 / 120 degrees of freedom		**2.571 / 2.042 / 1.980**
Standard error	8 ÷ √16	**2**
t critical value, 15 degrees		**2.131**
Margin of error	2.131 × 2	**4.26**
Interval for the average	7 ± 4.26	**2.74 to 11.26**

Watch out: Using the normal where the t belongs.

6. Chi-square and F

- Chi-square is for questions about one variance. It is never negative and skewed to the right. The test number is (n − 1) × sample variance ÷ the variance tested against.
- F compares two variances. It is the ratio of two independent chi-square variables, each divided by its own degrees of freedom. It is never negative and is skewed right. Put the larger variance on top. F is also behind the overall significance test of a regression.

Chi-square statistic = (n − 1) × s^2 ÷ σ_0^2. F = larger variance ÷ smaller variance

TESTING VARIANCES AT THE 5% LEVEL (UPPER TAIL)		
Chi-square: claim 16, n = 21, s^2 = 24	20 × 24 ÷ 16	**30.0**
Critical value, 20 degrees of freedom		**31.41**
F: variances 36 and 25, 11 observations each	36 ÷ 25	**1.44**
Critical value, 10 and 10 degrees of freedom		**2.98**
Verdict	both below their critical values	**do not reject**

Watch out: Forgetting that chi-square and F can never be negative.

Key takeaways

- Probabilities lie between 0 and 1 and add up to 1; a fair die is discrete uniform, each outcome 1 ÷ n.
- Binomial: mean = n × p and variance = n × p × (1 − p); it needs independent tries and a constant p.
- Normal: z = (x − μ) ÷ σ; about 68%, 95% and 99% lie within 1, 2 and 3 standard deviations, but real tails are fatter.
- Log return = $\ln(S_1 \div S_0)$; log returns add, and if they are normal, prices are lognormal.
- Student's t is the fat-tailed cousin of the normal, used for a mean when the standard deviation is estimated.
- Chi-square is for one variance, F compares two variances and tests a regression as a whole.

Quick check

1. Binomial with n = 20 and p = 0.25. What are the mean and the variance?
   Answer: Mean 5; variance 20 × 0.25 × 0.75 = 3.75.
2. A return is 20%, the mean is 10% and the standard deviation is 5%. z-score?
   Answer: (20 − 10) ÷ 5 = +2.0.
3. Which distribution: a mean (sd estimated), one variance, two variances?
   Answer: Student's t; chi-square; F.