Discrete Distributions and Random-Variable Calculations

Lectures 11–13

QBA201 – Quantitative Business Analysis

Shi Qiu – Fall 2026

A random variable turns an outcome into a number

For one coin flip, define $X = 1$ for heads and $X = 0$ for tails.

X is a rule. Its realized value depends on the outcome.

Define what is counted or measured before writing a probability.

Discrete and continuous variables

Discrete

Countably many possible values, such as the number of correct answers.

Continuous

A model for measurements varying over an interval, such as elapsed time.

A two-dice *sum* takes values 2 through 12. One die takes values 1 through 6.

Uppercase variable, lowercase realized value

Let X be daily sales. A particular value is x.

$$p(x) = \mathrm{P}(X = x).$$

For example, $p(3)$ means $\mathrm{P}(X = 3)$.

The random variable and the value it takes are different objects.

A discrete distribution must satisfy two checks

$$0 \le p(x) \le 1 \quad \text{for every possible } x,$$
$$\sum_{x} p(x) = 1.$$

Discuss: Can five nonnegative numbers be probabilities if their sum is 1.20?

Divide each frequency by the total

Let X be the number of residents in a randomly selected household. The historical textbook frequencies below are in millions.

Persons	Relative frequency
1	$8.1/26.5 = 0.3057$
2	$9.1/26.5 = 0.3434$
3	$4.1/26.5 = 0.1547$
4	$3.4/26.5 = 0.1283$
5 or more	$1.8/26.5 = 0.0679$

Use the unrounded ratios in calculations.

An event can contain several values

Four or more persons includes both final categories.

$$\mathrm{P}(X \geq 4) = \frac{3.4 + 1.8}{26.5} = 0.1962.$$

Discuss: Why can we answer this event without knowing each household size in the final category?

Expected value is a probability-weighted mean

$$\mu = \mathrm{E}(X) = \sum_x x\, p(x).$$

Each possible value receives its probability as a weight.

An expected value need not be a possible realized value.

Discuss: Can an average household size of 2.3 describe any one household?

Variance measures spread under the distribution

$$\text{Var}(X) = \sum_x (x - \mu)^2 p(x).$$

Equivalent formula:

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

$$SD(X) = \sqrt{\text{Var}(X)}.$$

An open-ended category needs an extra assumption

For this calculation only, replace every five-or-more household size by five.

$$\mathrm{E}(X^*) = \frac{1(8.1) + 2(9.1) + 3(4.1) + 4(3.4) + 5(1.8)}{26.5} = 2.3094.$$

X^* is the capped household-size variable.

This is a lower bound for the true mean size, not an exact mean of the open-ended original data.

Calculate spread for the capped variable

For the second moment, square each value before weighting: $\mathrm{E}[(X^*)^2] = \sum_x x^2 p(x)$.

$$\mathrm{E}[(X^*)^2] = \frac{180.8}{26.5} = 6.8226\ldots$$

$$\mathrm{Var}(X^*) = \frac{180.8}{26.5} - \left(\frac{61.2}{26.5}\right)^2 = 1.4892.$$

$$SD(X^*) = \sqrt{1.489156\ldots} = 1.2203.$$

The binomial model has four conditions

1. A fixed number n of trials.
2. Two defined outcomes on each trial.
3. The same success probability p on each trial.
4. Independent trials.

The binomial variable counts successes among those n trials.

The binomial probability formula

For $X \sim \text{Bin}(n, p)$:

$$\text{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

The power terms give one specified sequence's probability. The combination counts sequences with x successes.

Binomial mean and variance

$$\mathrm{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p).$$
$$SD(X) = \sqrt{np(1-p)}.$$

Discuss: For $n = 10$ and $p = 0.2$, calculate the mean and variance.

A guessing model for a multiple-choice quiz

A hypothetical quiz contains ten questions, each with five options and one correct answer.

A student guesses independently and uniformly on every question.

Discuss: What is the probability of no correct answers? Exactly two correct answers?

Translate the quiz scenario into parameters

Let X be the number of correct answers.

$$n = 10, \qquad p = \frac{1}{5} = 0.2.$$

$$X \sim \text{Bin}(10, 0.2)$$

under the stated guessing assumptions.

The number of answer options sets the trial probability, not the number of trials.

No correct answers

$$P(X = 0) = \binom{10}{0}(0.2)^0(0.8)^{10}.$$

$$= 1(1)(0.8)^{10} = 0.1074.$$

There is about a 10.74% chance of zero correct answers in this model.

Exactly two correct answers

$$\mathrm{P}(X = 2) = \binom{10}{2}(0.2)^2(0.8)^8.$$

$$= 45(0.04)(0.16777216) = 0.3020.$$

Discuss: What would be missing if you calculated only $(0.2)^2(0.8)^8$?

Translate a cutoff into integer values

Suppose five or more correct answers are needed to pass.

Failure means fewer than five:

$$X < 5 \quad \Longleftrightarrow \quad X \leq 4.$$

For an integer count, less than five includes exactly four.

A cumulative probability adds the included values

$$\mathrm{P}(X \leq 4) = \sum_{x=0}^{4} \binom{10}{x} (0.2)^x (0.8)^{10-x} = 0.9672.$$

In Excel: `BINOM.DIST(4,10,0.2,TRUE)`.

Discuss: How would you obtain the probability of passing?

Expected value respects linear transformations

For constants a, b, c:

$$\mathrm{E}(c) = c, \qquad \mathrm{E}(X + c) = \mathrm{E}(X) + c,$$

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b.$$

Linearity of expectation does not require independence.

Translate sales into profit

Let S be monthly sales in euros and G monthly profit.

$$\mathrm{E}(S) = 25,000, \quad SD(S) = 4,000.$$

A 30% contribution margin and fixed costs of 6,000 give:

$$G = 0.30S - 6,000.$$

Discuss: Which number affects the mean but not the variance?

Calculate expected profit

$$\mathrm{E}(G) = \mathrm{E}(0.30S - 6,000).$$

$$= 0.30\,\mathrm{E}(S) - 6,000 = 0.30(25,000) - 6,000 = 1,500.$$

Expected profit is 1,500 euros per month under the stated model.

Variance squares the scale factor

$$\mathrm{Var}(c) = 0, \qquad \mathrm{Var}(X + b) = \mathrm{Var}(X).$$
$$\mathrm{Var}(aX + b) = a^2 \,\mathrm{Var}(X).$$
$$SD(aX + b) = |a| SD(X).$$

Adding a fixed amount shifts every observation equally, without changing spread.

Calculate profit variability

First square the supplied SD: $\text{Var}(S) = 4,000^2 = 16,000,000$.

$$\text{Var}(G) = (0.30)^2(16,000,000) = 1,440,000.$$

$$SD(G) = \sqrt{1,440,000} = 1,200.$$

Check directly: $SD(G) = 0.30(4,000) = 1,200$ euros.

Variance has squared units. Standard deviation returns to euros.

A bivariate distribution preserves pairs

$$p(x, y) = \mathrm{P}(X = x \text{ and } Y = y).$$

A joint table assigns probability to every possible pair.

Marginal distributions describe X and Y separately.

Discuss: Why cannot two marginal tables alone tell us how the variables move together?

Monthly house sales for two agents

Let X and Y be the monthly numbers sold by Xavier and Yvette.

	$X=0$	$X=1$	$X=2$	$\mathrm{P}(Y=y)$
$Y=0$	.12	.42	.06	.60
$Y=1$	.21	.06	.03	.30
$Y=2$	.07	.02	.01	.10
$\mathrm{P}(X=x)$	.40	.50	.10	1.00

Use marginals to find each variable's moments

$$\mathrm{E}(X) = 0(.4) + 1(.5) + 2(.1) = 0.7,$$

$$\mathrm{Var}(X) = [0^2(.4) + 1^2(.5) + 2^2(.1)] - 0.7^2 = 0.41.$$

Similarly, $\mathrm{E}(Y) = 0.5$ and $\mathrm{Var}(Y) = 0.45$.

Thus $SD(X) = 0.6403$ and $SD(Y) = 0.6708$.

Covariance requires joint probabilities

$$\mathrm{Cov}(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x, y).$$

Equivalent calculation:

$$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y).$$

Multiply each paired product by the probability of that pair.

Correlation standardizes the joint variation

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

It is defined when both variances are positive.

Discuss: Which pieces can be found from marginals, and which need the joint table?

Compute the expected product

Only pairs with both values nonzero contribute:

$$\mathrm{E}(XY) = 1(1)(.06) + 2(1)(.03) + 1(2)(.02) + 2(2)(.01) = 0.20.$$

$$\mathrm{Cov}(X, Y) = 0.20 - (0.7)(0.5) = -0.15.$$

Zero-valued products can be omitted only after checking their values.

Interpret the agents' correlation

$$\rho = \frac{-0.15}{\sqrt{0.41(0.45)}} = -0.3492.$$

The monthly counts have a negative linear association under this model.

The sign does not establish why the agents' sales move differently.

Find the distribution of total sales

Let $T = X + Y$.

t	$\mathrm{P}(T = t)$
0	.12
1	.63
2	.19
3	.05
4	.01

For $T = 2$, add $p(0, 2) + p(1, 1) + p(2, 0) = .07 + .06 + .06$.

Expectations add; variances need covariance

$$\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y).$$

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

Independence implies zero covariance when moments exist.

Zero covariance alone does not generally imply independence.

Check the total using the sum rules

$$\mathrm{E}(X + Y) = 0.7 + 0.5 = 1.2.$$
$$\mathrm{Var}(X + Y) = 0.41 + 0.45 + 2(-0.15) = 0.56.$$

Discuss: Would omitting the covariance overstate or understate the variance here?

Mean and variance of a weighted sum

Let R_1, R_2 be asset returns. For fixed weights $w_1 + w_2 = 1$, define $R_p = w_1 R_1 + w_2 R_2$.

$$\mathrm{E}(R_p) = w_1 \,\mathrm{E}(R_1) + w_2 \,\mathrm{E}(R_2).$$

$$\mathrm{Var}(R_p) = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \rho \sigma_1 \sigma_2.$$

Here σ_j is asset j's return SD and ρ their correlation.

Keep returns consistently in decimal form.

The source's two-asset assumptions

$$w_1 = 0.25, \quad w_2 = 0.75.$$

$$\mathrm{E}(R_1) = 0.08, \quad \mathrm{E}(R_2) = 0.15, \quad \sigma_1 = 0.12, \quad \sigma_2 = 0.22.$$

Find the mean and compare SD for $\rho = 1$, 0.5 and 0.

Calculate the portfolio mean

$$\mathrm{E}(R_p) = 0.25(0.08) + 0.75(0.15) = 0.1325.$$

The expected return is 13.25% in this model.

Discuss: Does changing only the correlation change this expected return?

Correlation changes the variance

$$\text{Var}(R_p) = 0.028125 + 0.0099\rho.$$

ρ	Variance	Standard deviation
1	0.038025	19.5000%
0.5	0.033075	18.1865%
0	0.028125	16.7705%

More than two variables: include every covariance pair

For $R_p = \sum_{i=1}^{k} w_i R_i$:

$$\mathrm{E}(R_p) = \sum_{i=1}^{k} w_i \, \mathrm{E}(R_i),$$

$$\mathrm{Var}(R_p) = \sum_{i=1}^{k} w_i^2 \, \mathrm{Var}(R_i) + 2 \sum_{i<j} w_i w_j \, \mathrm{Cov}(R_i, R_j).$$

Discuss: Why does each covariance pair appear only once inside the final sum?