Chapter 3: Probability calculation methods

Transformations

For random variable Y, we are interested in calculating the probability distribution of the *transformed* random variable

$$U = h(Y)$$

where $h(.)$ is a real-valued function that can map a value in the set $\mathcal{Y}$ (on which Y takes values with positive probability) to another real value.

Transformations (cont.)

Discrete case: Suppose Y takes values on the (countable) set

$$\{y_1, y_2, \ldots, y_n, \ldots\}.$$

Under the transformation $h(.)$, these values map to the set

$$\{u_1, u_2, \ldots, u_n, \ldots\}$$

where $u_i = h(y_j)$ for at least one j.

To compute $P(U = u_i)$ we simply *sum* up all the values of the probabilities for the y_js that *map onto* u_i.

Transformations (cont.)

For *1-1 transformations*, this is straightforward:

- identify the set of values $\{u_1, u_2, \ldots, u_n, \ldots\}$ that the values $\{y_1, y_2, \ldots, y_n, \ldots\}$ map onto – in this case there is a one-to-one correspondence between the sets, and

  $$u_i = h(y_i)$$

  so that

  $$P(U = u_i) = P(Y = y_i).$$

- for each i, assign the probability $p_Y(y_i)$ to u_i to obtain $p_U(u_i)$

Transformations (cont.)

Example (Poisson case)

Suppose that Y has a Poisson distribution with parameter λ. Let $U = 3Y$; then only the values in the set

$$\mathcal{U} = \{0, 3, 6, 9, \ldots, 3n, \ldots\}$$

receive any probability, and

$$p_U(u) = \frac{\lambda^{u/3}}{(u/3)!} e^{-\lambda} \qquad u \in \{0, 3, 6, 9, \ldots\}$$

with $p_U(u)$ equal to zero for all other values of u.

Transformations (cont.)

Continuous case: The obstacle to using the same approach here is that for a continuous random variable Y we have

$$P(Y = y) = 0 \qquad \text{for all } y \in \mathbb{R}.$$

To get around this, we can instead consider

$$P(Y \in \mathcal{A})$$

for a subset (eg interval) of the real line. For example, we could consider $\mathcal{A} = [2, 5]$, and the probability

$$P(Y \in \mathcal{A}) \equiv P(2 \leq Y \leq 5).$$

Transformations (cont.)

The key relationship here is that

$$P(Y \in \mathcal{A}) = P(U \in \mathcal{B}).$$

We want to compute $F_U(.)$ or $f_U(.)$, so we fix $\mathcal{B}$ first, and write

$$h^{-1}(\mathcal{B})$$

for the set of y values that *map onto* $\mathcal{B}$. Then we must have

$$P(U \in \mathcal{B}) = P(Y \in h^{-1}(\mathcal{B}))$$

as all the probability gets transported by the transformation.

Transformations (cont.)

For 1–1 (or monotonic) transformations, we consider

1. $h(y)$ *increasing*: eg $U = 2Y$
2. $h(y)$ *decreasing*: eg $U = 2/Y$

Transformations (cont.)

For a 1–1 increasing transformation, we have for the cdf

$$F_U(u) = P(U \leq u) = P(h(Y) \leq u) = P(Y \leq h^{-1}(u))$$
$$= F_Y(h^{-1}(u))$$

and therefore for the pdf, by the chain rule

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y(h^{-1}(u))}{du} = f_Y(h^{-1}(u))\frac{dh^{-1}(u)}{du}.$$

If $U = 2Y$ then for fixed $u \in \mathbb{R}$

$$F_U(u) = P(U \leq u) \equiv P(2Y \leq u) = P(Y \leq u/2) = F_Y(u/2)$$

$$\therefore f_U(u) = f_Y(u/2)/2.$$

Transformations (cont.)

For a 1–1 decreasing transformation, we have instead

$$F_U(u) = P(U \leq u) = P(h(Y) \leq u) = P(Y \geq h^{-1}(u))$$
$$= 1 - F_Y(h^{-1}(u))$$

and therefore for the pdf, by the chain rule

$$f_U(u) = \frac{dF_U(u)}{du} = -\frac{dF_Y(h^{-1}(u))}{du} = -f_Y(h^{-1}(u))\frac{dh^{-1}(u)}{du}.$$

If $U = 2/Y$ then for fixed $u \in \mathbb{R}$

$$F_U(u) = P(U \leq u) \equiv P(2/Y \leq u) = P(Y \geq 2/u) = 1 - F_Y(2/u)$$

$$\therefore f_U(u) = 2f_Y(2/u)/u^2.$$

Transformations (cont.)

Therefore we can write

$$f_U(u) = f_Y(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right|$$

to cover both cases. The term

$$\left| \frac{dh^{-1}(u)}{du} \right|$$

is the *Jacobian* of the transformation.

Transformations (cont.)

Note

The identity

$$P(U \in \mathcal{B}) = P(Y \in h^{-1}(\mathcal{B}))$$

can be re-written

$$\int_{\mathcal{B}} f_U(u)\, du = \int_{h^{-1}(\mathcal{B})} f_Y(y)\, dy$$

Transformations (cont.)

Note (cont.)

Changing variables in the right-hand integral

$$u = h(y) \quad \Longleftrightarrow \quad y = h^{-1}(u)$$

we have by the rules of integration

$$\int_{\mathcal{B}} f_U(u)\, du = \int_{\mathcal{B}} f_Y(h^{-1}(u)) \left| \frac{dh^{-1}(u)}{du} \right| du$$

and the same result follows by equating the integrands.

Transformations (cont.)

Expectations

For continuous random variable Y and function $g(.)$

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y(y)\, dy.$$

However, $U = g(Y)$ is a random variable in its own right, so

$$\mathbb{E}[U] = \int_{-\infty}^{\infty} u f_U(U)\, du.$$

Therefore we must have

$$\mathbb{E}[U] \equiv \mathbb{E}[g(Y)].$$

Moment Generating Functions

Recall the definition of the mgf: for $-b \leq t \leq b$, $b > 0$,

$$m_Y(t) = \mathbb{E}[e^{tY}].$$

If $U = cY + d$, for constants $c > 0$ and d, we have that

$$m_U(t) = \mathbb{E}\left[e^{tU}\right] = \mathbb{E}\left[e^{t(cY+d)}\right] = \mathbb{E}\left[e^{ctY}e^{dt}\right] = e^{dt}\mathbb{E}\left[e^{(ct)Y}\right]$$

as e^{dt} is a constant. Therefore

$$m_U(t) = e^{dt}m_Y(ct).$$

Sums of random variables

Example (Two dice)

Two fair dice are rolled, with the outcomes of the rolls independent events.

What is the probability that the sum of the scores of the two dice exceeds 10 ?

Sums of random variables (cont.)

Example (Two dice)

Let Y_1 and Y_2 be the random variables recording the scores on the two dice:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

The outcome combinations in this table are equally likely, and occur with probability $1/36$.

Sums of random variables (cont.)

Example (Two dice)

That is

$$p(y) = \begin{cases} 1/36 & y = 2, 12 \\ 2/36 & y = 3, 11 \\ 3/36 & y = 4, 10 \\ 4/36 & y = 5, 9 \\ 5/36 & y = 6, 8 \\ 6/36 & y = 7. \end{cases}$$

Therefore

$$P[Y > 10] = P[Y = 11] + P[Y = 12] = \frac{3}{36} = \frac{1}{12}.$$

Sums of random variables (cont.)

In the example, we are computing the pmf of

$$Y = Y_1 + Y_2$$

with Y_1 and Y_2 *independent random variables*.

Independence implies we must have

$$P((Y_1 = y_1) \cap (Y_2 = y_2)) = P(Y_1 = y_1) \times P(Y_2 = y_2)$$

for all pairs of possible realizations (y_1, y_2), each value taken from the set $\{1, 2, 3, 4, 5, 6\}$.

Sums of random variables (cont.)

The crucial components of the calculation are

- the *partition* of the event $(Y = y)$ according to the possible values of Y_1;
- the *theorem of total probability*;
- the *independence* of the outcomes of the two rolls, equivalently, the independence of the random variables Y_1 and Y_2.

Using these elements, we can derive a general result.

Sums of random variables (cont.)

Suppose Y_1 and Y_2 are discrete random variables, with Y_1 and Y_2 independent. Suppose that Y_1 and Y_2 take values on the sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ respectively. Let $Y = Y_1 + Y_2$. Then

$$\begin{aligned} P(Y = y) &= P\left(\bigcup_{y_1 \in \mathcal{Y}_1} (Y = y) \cap (Y_1 = y_1) \right) \\ &= \sum_{y_1 \in \mathcal{Y}_1} P((Y = y) \cap (Y_1 = y_1)) \\ &= \sum_{y_1 \in \mathcal{Y}_1} P((Y_2 = y - y_1) \cap (Y_1 = y_1)) \\ &= \sum_{y_1 \in \mathcal{Y}_1} P(Y_2 = y - y_1) P(Y_1 = y_1) \end{aligned}$$

Sums of random variables (cont.)

That is

$$P(Y = y) = \sum_{y_1 \in \mathcal{Y}_1^*} P(Y_2 = y - y_1) P(Y_1 = y_1)$$

or

$$p_Y(y) = \sum_{y_1 \in \mathcal{Y}_1^*} p_{Y_2}(y - y_1) p_{Y_1}(y_1)$$

where

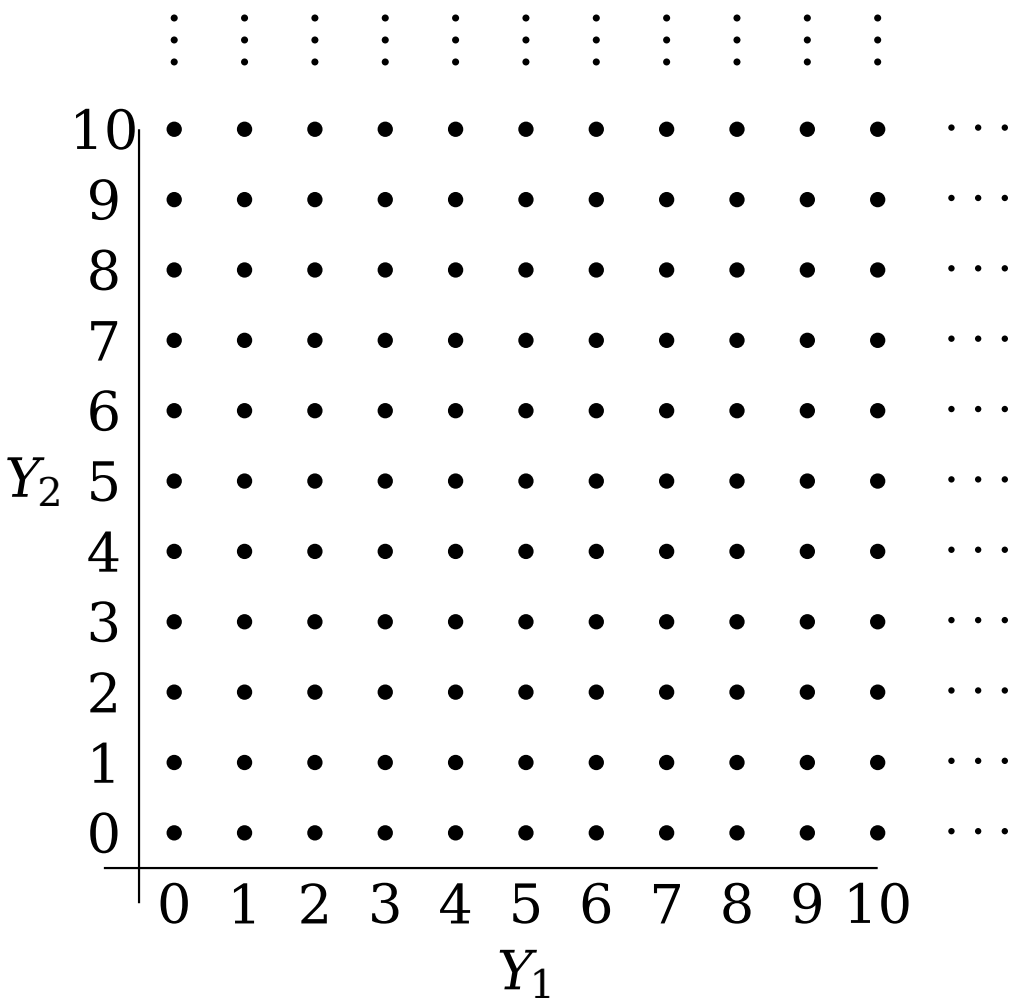
$$\mathcal{Y}_1^* = \{y_1 : \ y_1 \in \mathcal{Y}_1 \text{ and } y - y_1 \in \mathcal{Y}_2\}.$$

For example, if Y_1 and Y_2 take values $y = 0, 1, 2, \ldots$

$$P(Y = y) = \sum_{y_1 = 0}^{y} P(Y_1 = y_1) P(Y_2 = y - y_1).$$

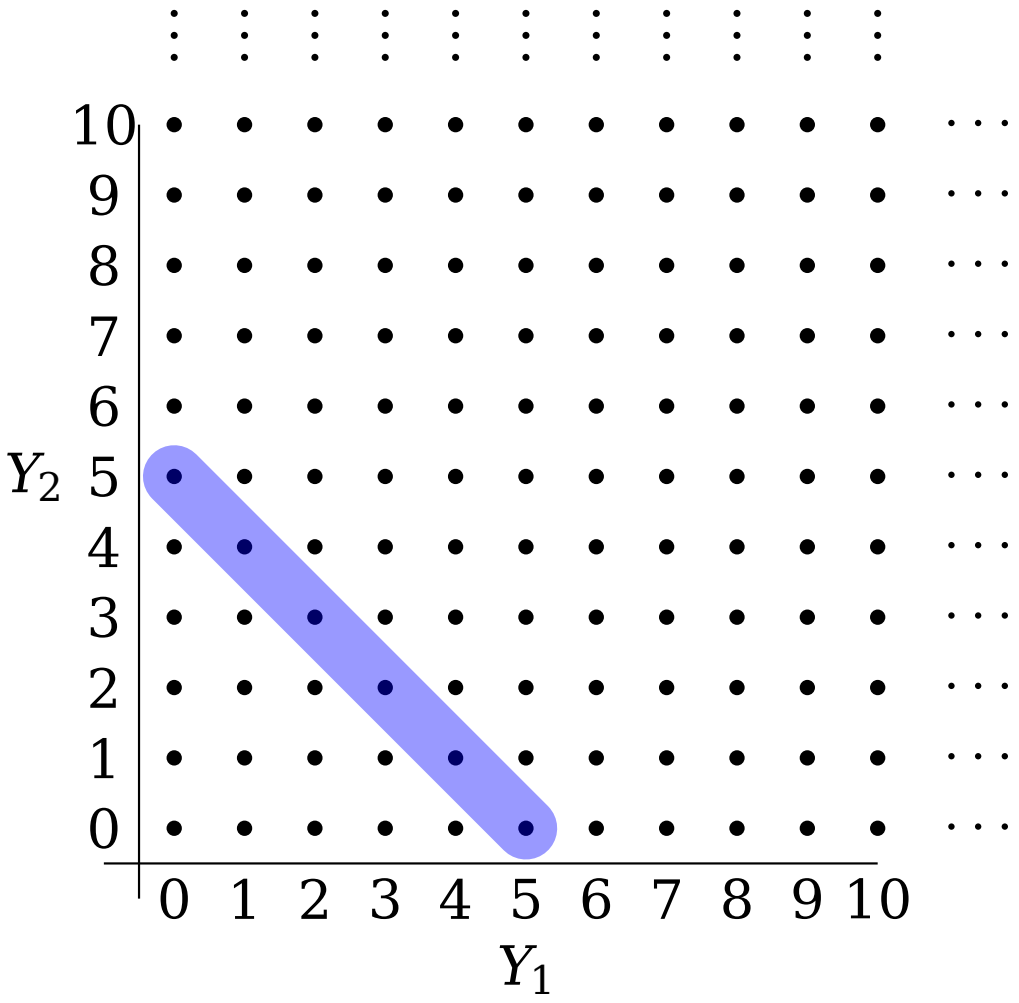
Sums of random variables (cont.)

If Y_1, Y_2 take values on $0, 1, 2, \ldots$.



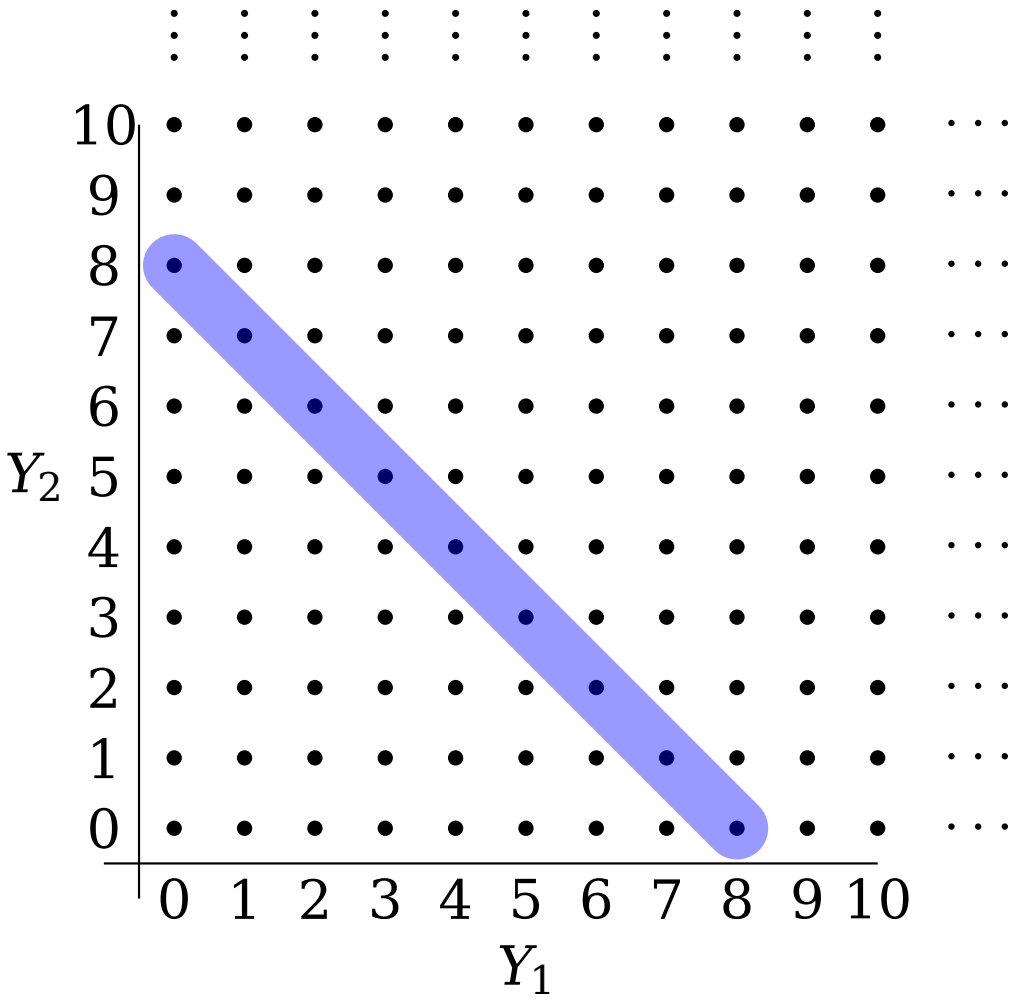
Sums of random variables (cont.)

$Y_1 + Y_2 = 5$



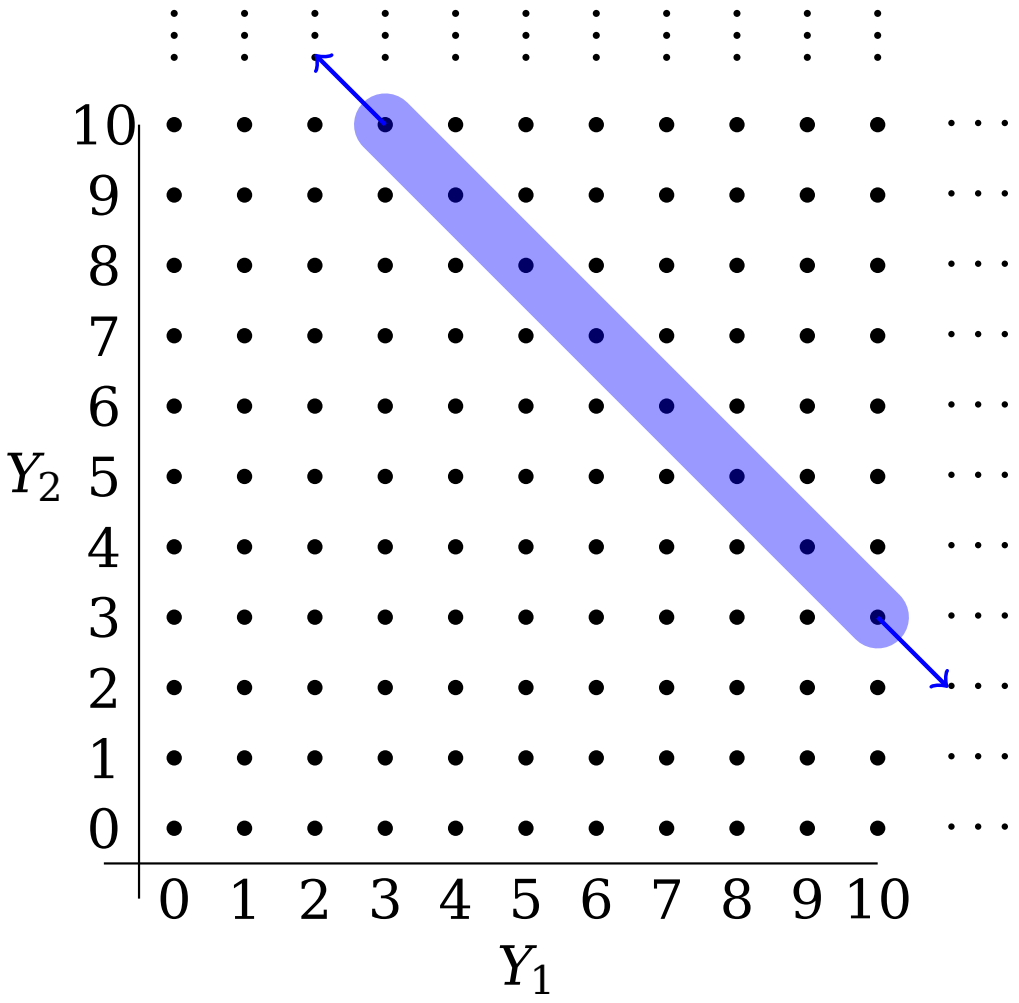
Sums of random variables (cont.)

$Y_1 + Y_2 = 8$



Sums of random variables (cont.)

$Y_1 + Y_2 = 13$



Sums of random variables (cont.)

If Y_1 and Y_2 are *continuous*, then $Y = Y_1 + Y_2$ is continuous.

For fixed $y \in \mathbb{R}$, we aim to compute

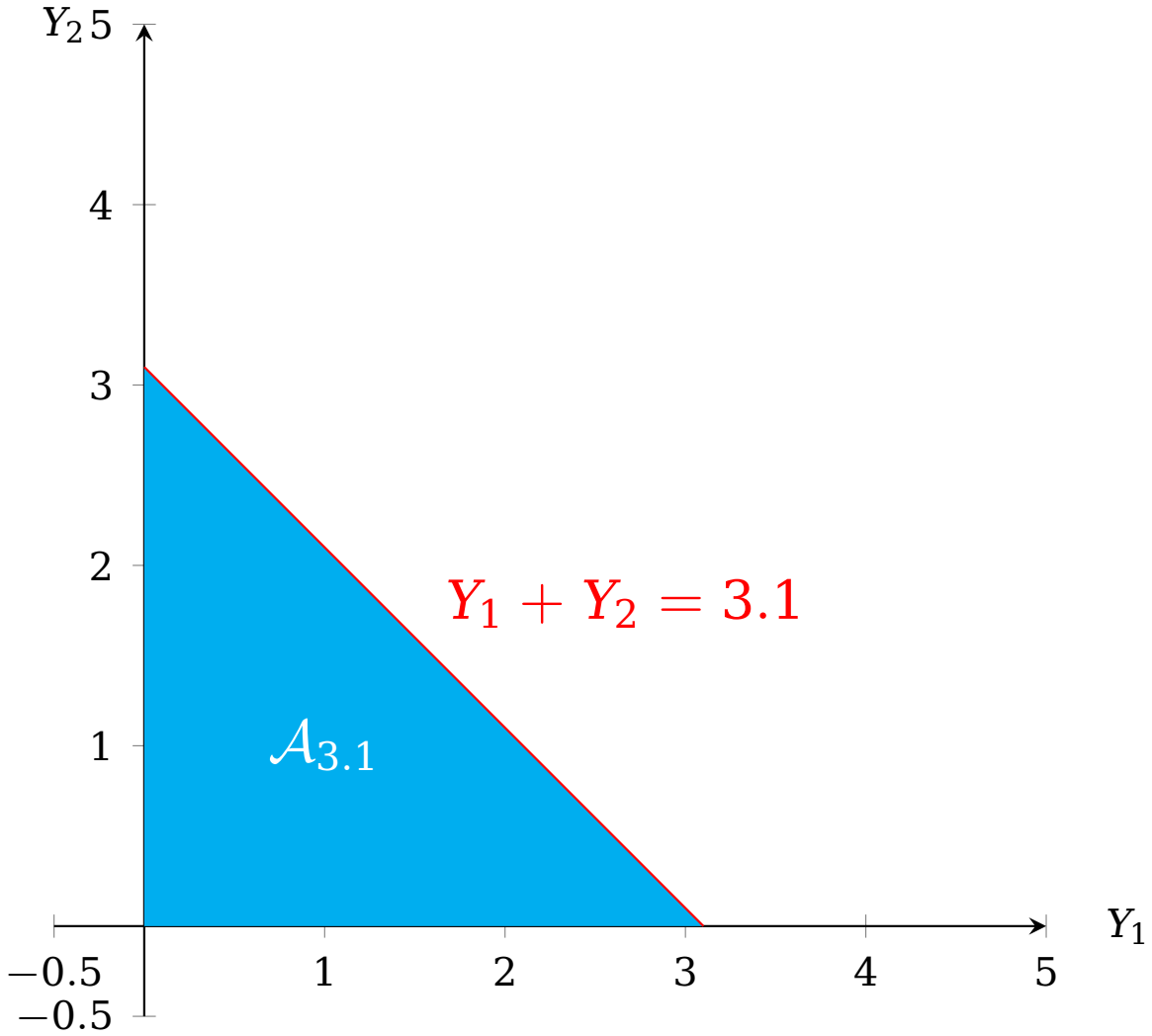
$$F_Y(y) = P(Y \leq y).$$

Let $A_y \in \mathbb{R}^2$ denote the set of values

$$A_y = \{(y_1, y_2) \in \mathbb{R}^2 : y_1 + y_2 \leq y\}.$$

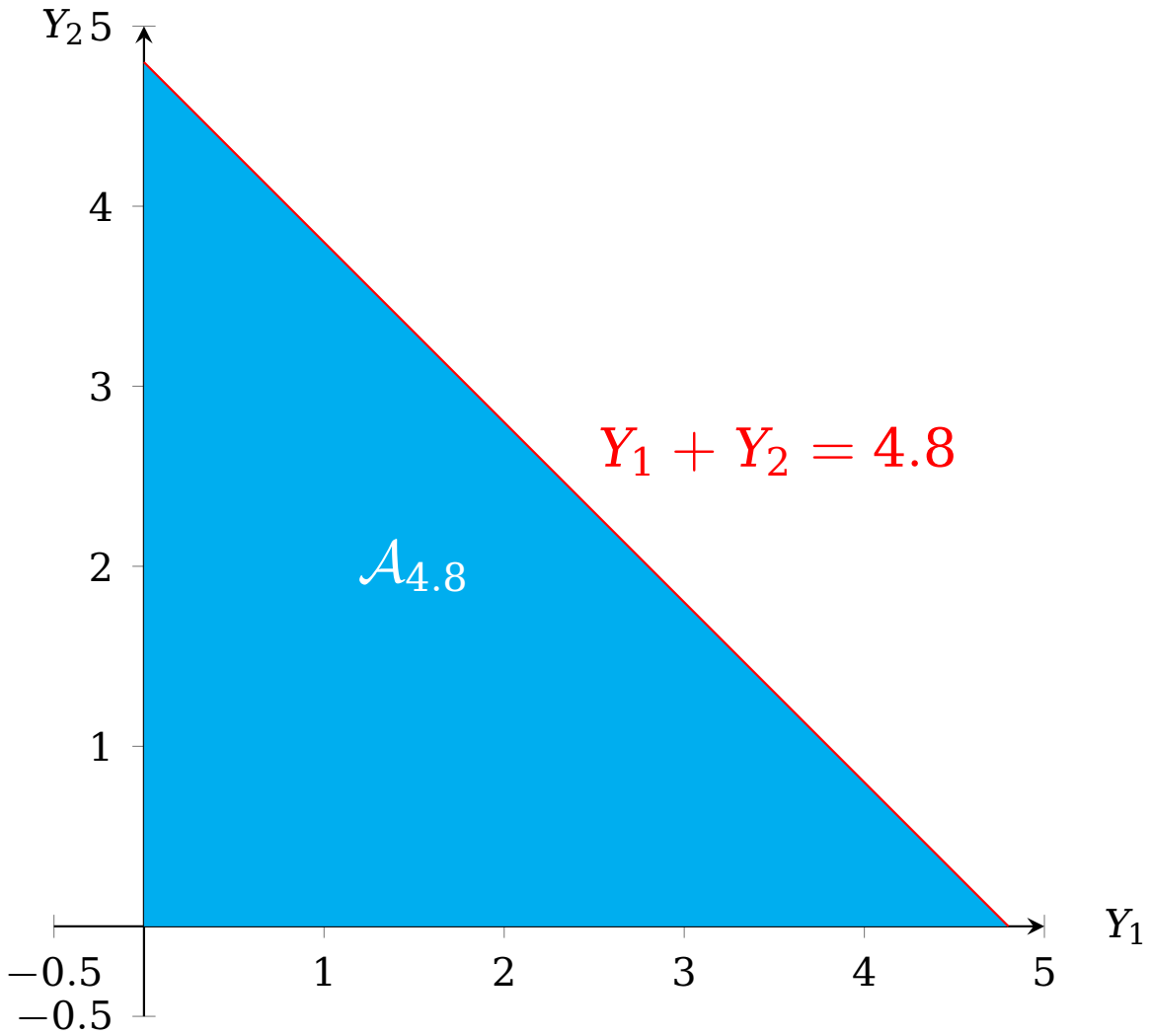
Sums of random variables (cont.)

Y_1, Y_2 taking values on $\mathbb{R}^+$; region $\mathcal{A}_{3.1}$ is where $Y_1 + Y_2 \leq 3.1$.



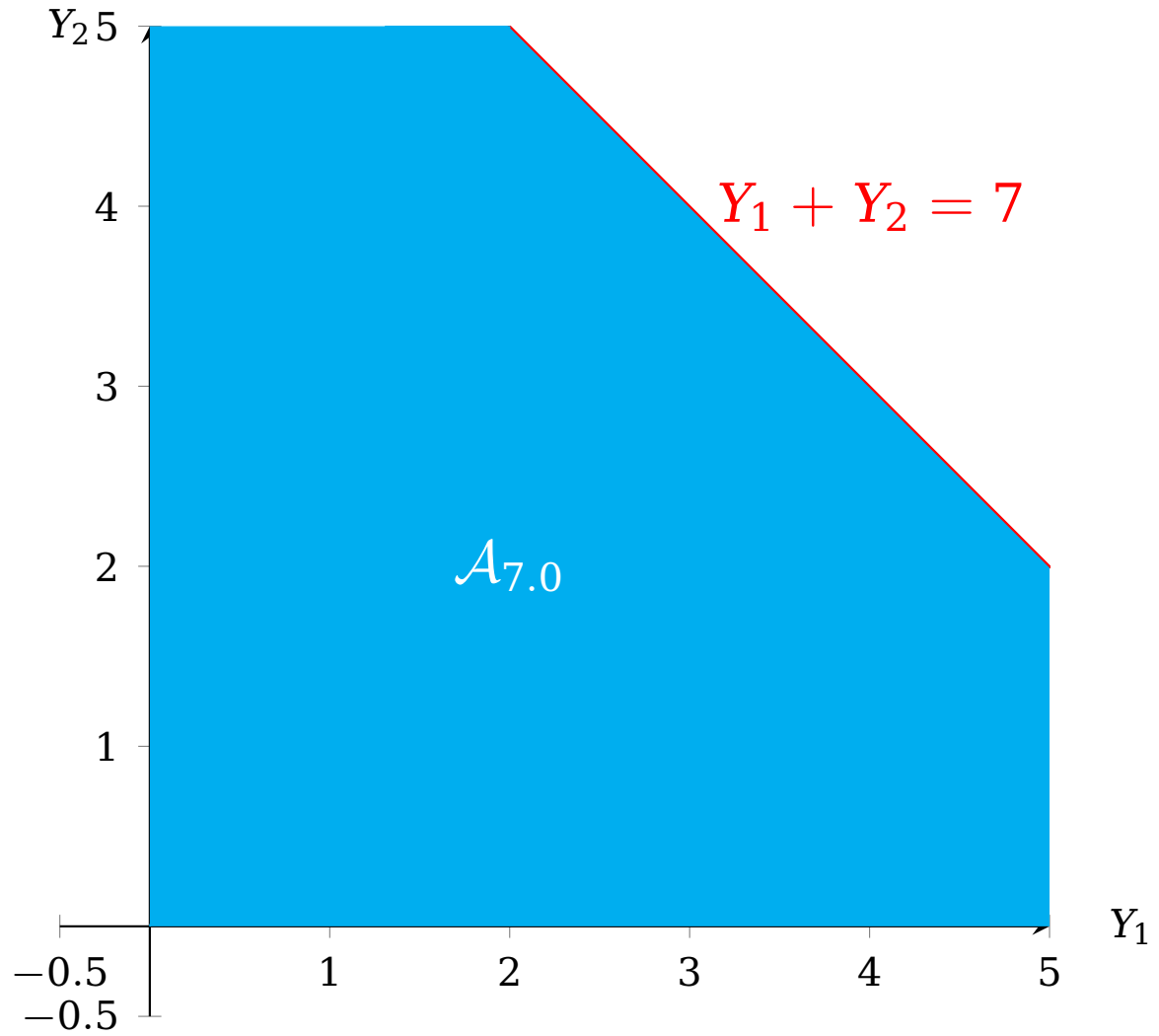
Sums of random variables (cont.)

Y_1, Y_2 taking values on $\mathbb{R}^+$; region $\mathcal{A}_{4.8}$ is where $Y_1 + Y_2 \leq 4.8$.



Sums of random variables (cont.)

Y_1, Y_2 taking values on $\mathbb{R}^+$; region $\mathcal{A}_{7.0}$ is where $Y_1 + Y_2 \leq 7$.



Sums of random variables (cont.)

Then we have by definition that

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)\, dt$$

where $f_Y(y)$ is the pdf for Y. However,

$$(Y \leq y) \equiv (Y_1 + Y_2 \leq y)$$

Therefore to compute $F_Y(y) = P(Y \leq y)$ we need to aggregate all the probability across the blue shaded region, $\mathcal{A}_y$.

Sums of random variables (cont.)

By analogy with the discrete case we may write

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y_2}(y - y_1) f_{Y_1}(y_1)\, dy_1$$

where

- $f_Y(.)$ is the pdf for Y,
- $f_{Y_1}(.)$ is the pdf for Y_1 and $f_{Y_2}(.)$ is the pdf for Y_2.

This is termed a *convolution* formula.

Sums of random variables (cont.)

Note

For *three* independent random variables

$$Y = Y_1 + Y_2 + Y_3$$

we use the convolution formula *recursively*; we have

$$Y = (Y_1 + Y_2) + Y_3 = Y_{12} + Y_3$$

say, and so we would compute the pmf or pdf of $Y_{12} = Y_1 + Y_2$ first, and then convolve it with the pmf or pdf for Y_3.