3 - Random Variables

- Estimation
- Example: radar
- Example: communication channel
- Example: force on mass
- Events corresponding to random variables
- Induced probability
- Functions of a random variable
- Cumulative distribution
- Simulation of random variables
- Expectation
- The vector space of random variables
- Variance and its interpretation
- The mean-variance decomposition

Random variables

Suppose Ω is a finite sample space, with pmf p

A function $x : \Omega \to V$ is called a *random variable*.

- The set V can be any set; it is the set of *values of x*.
- Often V is $\mathbb{R}^n$ or just $\mathbb{R}$; then x is called a *random vector*

Random variables and models

We model systems using random variables.

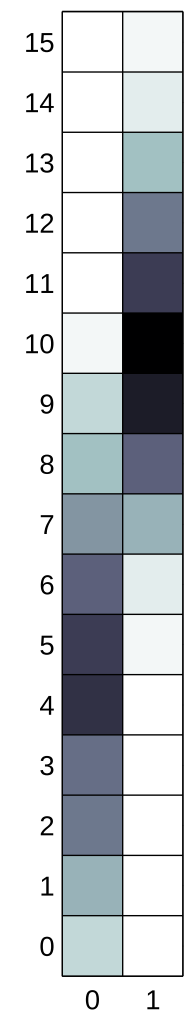
- Ω is a sample space. Exactly one outcome $\omega \in \Omega$ occurs.
- We have a *measurement random vector* $y : \Omega \to \mathbb{R}^n$.
- We have a *hypothesis random vector* $x : \Omega \to \mathbb{R}^n$

Estimation

- We measure $y(\omega)$
- We would like to estimate $x(\omega)$

Example: radar system

A radar system sends out n pulses, and receives y reflections, where $0 \leq y \leq n$.
Ideally, $y = n$ if an aircraft is present, and $y = 0$ otherwise.
In practice, reflections may be lost, or noise may be mistaken for reflections.



The set of outcomes is

$$\Omega = \Big\{ (x, y) \mid x \in \{0, 1\} \text{ and } y \in \{0, 1, \ldots, n\} \Big\}$$

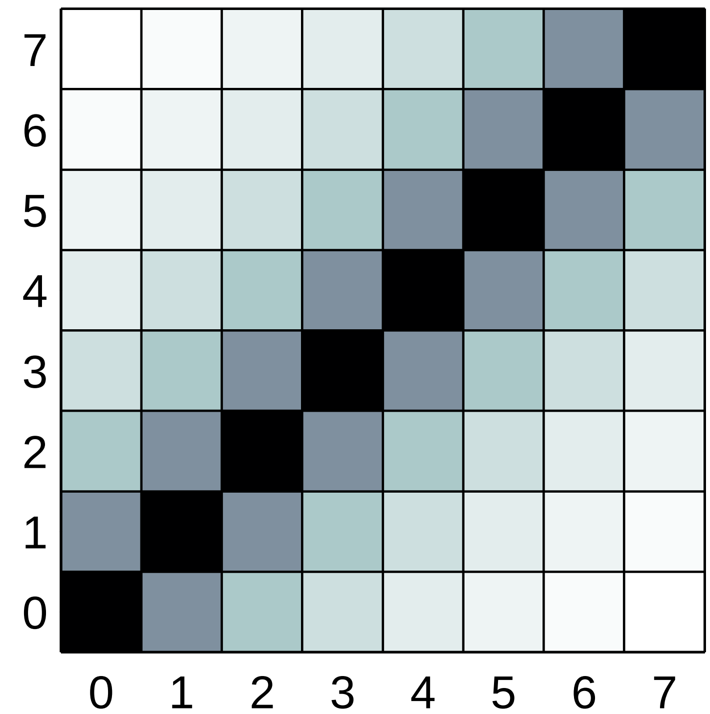
Here

- $x = 1$ if an aircraft is present, $x = 0$ otherwise
- y is the number of reflection pulses received

We measure y, and would like to determine x.

Example: communication channel

- A symbol $x \in \{0, 1, \ldots, n-1\}$ is sent.
- The channel is noisy, so the symbol received may not match what is sent.
- The symbol $y \in \{0, 1, \ldots, n-1\}$ is received.



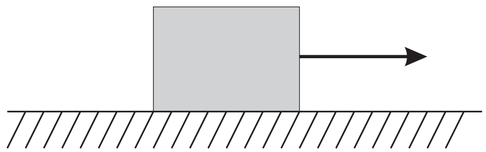
The set of outcomes is

$$\Omega = \Big\{ (x, y) \;\Big|\; x \in \{0, 1, \ldots, n-1\} \text{ and } y \in \{0, 1, \ldots, n-1\} \Big\}$$

We measure y, and would like to determine x.

Example: force on mass

Mass acted on by forces



- known sequence of forces $u_1, u_2, \ldots, u_n$
- additional random force disturbance $r_1, r_2, \ldots, r_n$
- we make a *noisy measurement* $y = v +$ position at time $n/2$ where v is random noise
- we'd like to estimate or predict the position x at time n

The set of outcomes is $\Omega = \mathbb{R}^{n+1}$, where $\omega = \begin{bmatrix} v \\ r \end{bmatrix}$

We have linear equations

$$
\begin{aligned}
y &= A(u + r) + v \\
x &= P(u + r)
\end{aligned}
$$

Estimation

We would like to *design* estimators.

Performance measures include

- the probability that the estimate is correct
- the mean size of the error, in some sense
- the bias of the estimator
  - continuous problems: are estimates on average too low or too high?
  - discrete problems: what are the probabilities of false positives or false negatives?

Random variables

We have

- sample space Ω, a finite set
- probability mass function $p : \Omega \to [0, 1]$
- a random variable $x : \Omega \to \mathbb{R}$

Suppose $a \in \mathbb{R}$. The *probability that $x = a$* is defined as

$$\mathbf{Prob}\Big(\Big\{\, \omega \in \Omega \mid x(\omega) = a \,\Big\}\Big)$$

This is equal to

$$\sum_{\omega \in \Omega \;\mid\; x(\omega) = a} p(\omega)$$

Example: two dice

We have sample space

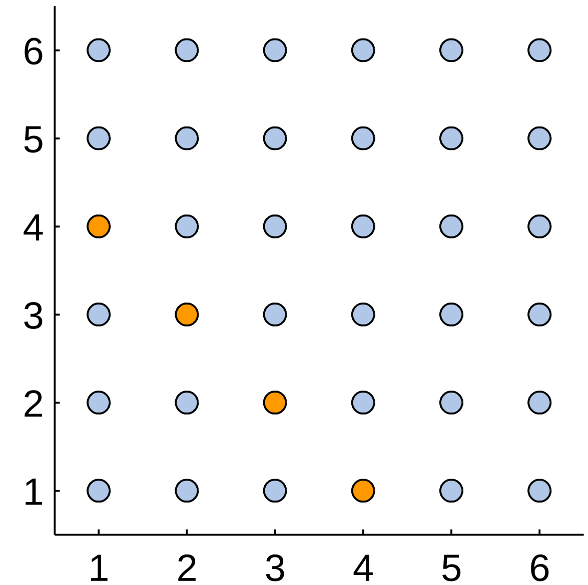
$$\Omega = \Big\{ (\omega_1, \omega_2) \in \Omega \mid \omega_i \in \{1, 2, \ldots, 6\} \Big\}$$

Define the random variable $x : \Omega \to \mathbb{R}$, the *sum of the two dice* by

$$x(\omega_1, \omega_2) = \omega_1 + \omega_2$$

Then $\mathbf{Prob}(x = 5) = \mathbf{Prob}(A)$ where the event A is

$$A = \Big\{ \omega \in \Omega \mid x(\omega) = 5 \Big\}$$

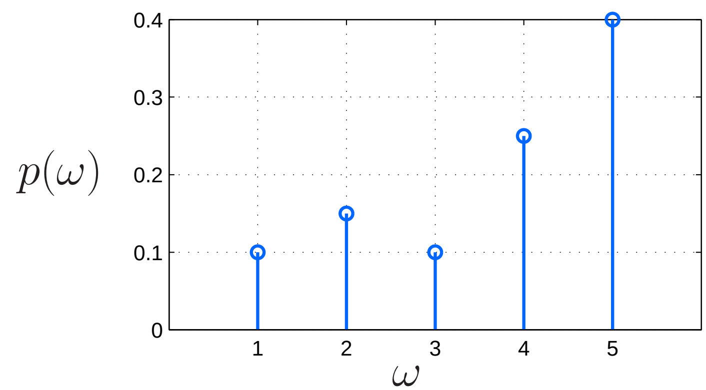


Probability and random variables

The probability of the *event* that $x = a$ is written $\mathbf{Prob}(x = a)$, i.e.,

$$\mathbf{Prob}(x = a) = \mathbf{Prob}\Big(\big\{\, \omega \in \Omega \mid x(\omega) = a \,\big\}\Big)$$

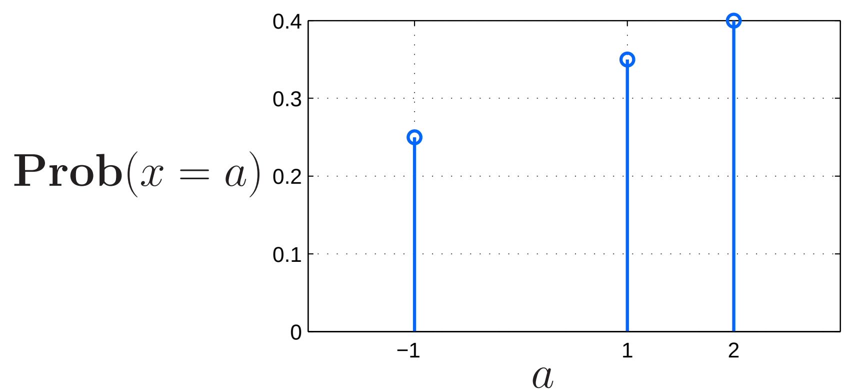
Suppose $\Omega = \{1, 2, 3, 4, 5\}$ and p is as shown



The random variable $x : \Omega \to \mathbb{R}$ is

$$x(\omega) = \begin{cases} -1 & \text{if } \omega = 1 \text{ or } \omega = 2 \\ 1 & \text{if } \omega = 3 \text{ or } \omega = 4 \\ 2 & \text{if } \omega = 5 \end{cases}$$

and $\mathbf{Prob}(x = a)$ is shown.



Notations used for random variables

- The *event* that $x = a$ is written

$$x^{-1}(a) = \left\{ \omega \in \Omega \mid x(\omega) = a \right\}$$

- The probability of this event is written as $\mathbf{Prob}(x = a)$

- This is also written $p^x(a) = \mathbf{Prob}(x = a)$

Notation for random variables

There are many similar notations used: for example, define

- $\mathbf{Prob}(x = a) = \mathbf{Prob}\big(x^{-1}(a)\big)$

- $\mathbf{Prob}(x \geq a) = \mathbf{Prob}\big(\{\, \omega \in \Omega \mid x(\omega) \geq a \,\}\big)$

- If $C \subset V$, then $\mathbf{Prob}(x \in C) = \mathbf{Prob}\big(\{\, \omega \in \Omega \mid x(\omega) \in C \,\}\big)$

Events corresponding to random variables

Suppose $x : \Omega \to V$. Each $a \in V$ defines an event

$$x^{-1}(a) = \Big\{ \omega \in \Omega \mid x(\omega) = a \Big\}$$

These events partition Ω

For example, if $\Omega = \{1, 2, 3, 4, 5\}$ and the random variable x is

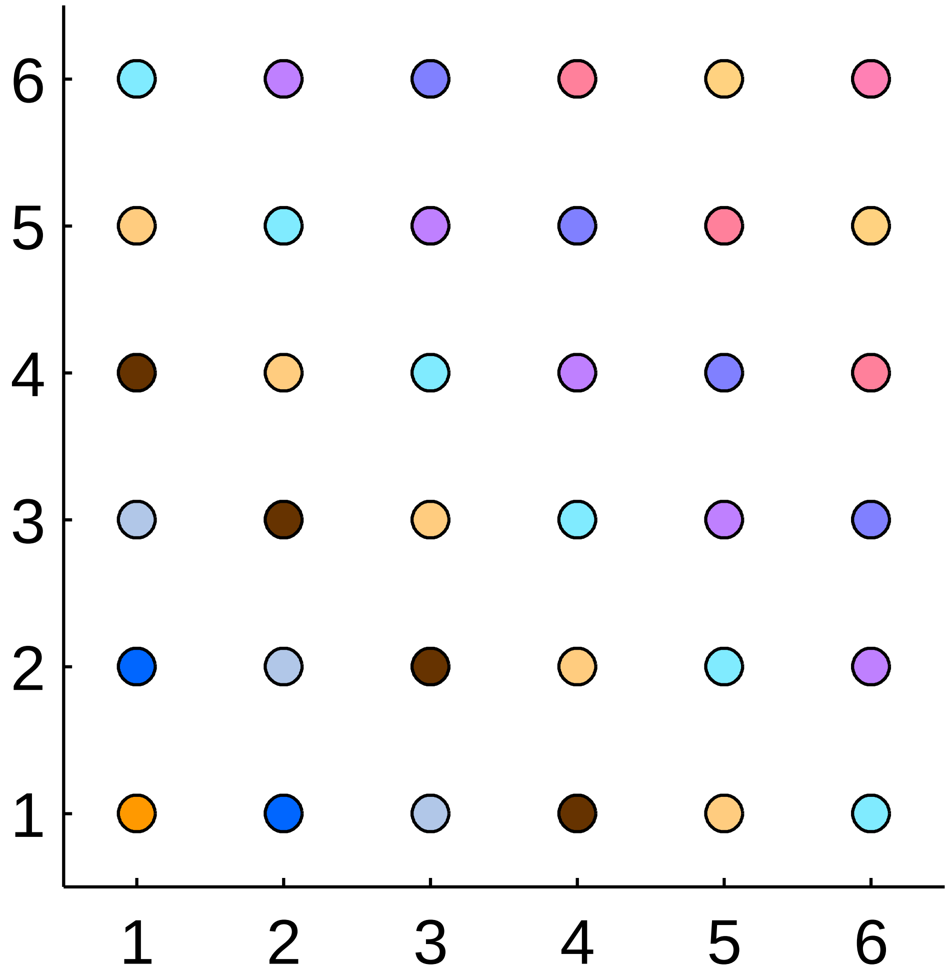
$$x(\omega) = \begin{cases} -1 & \text{if } \omega = 1 \text{ or } \omega = 2 \\ 1 & \text{if } \omega = 3 \text{ or } \omega = 4 \\ 2 & \text{if } \omega = 5 \end{cases}$$

The events associated with x are

$$x^{-1}(-1) = \{1, 2\} \qquad x^{-1}(1) = \{3, 4\} \qquad x^{-1}(2) = \{5\}$$

Example: sum of two dice

The events are



Induced probability

Suppose $x : \Omega \to V$. The *induced pmf of x* is the function $p^x : V \to [0, 1]$

$$p^x(a) = \mathbf{Prob}(x = a)$$

It satisfies the properties of a probability mass function

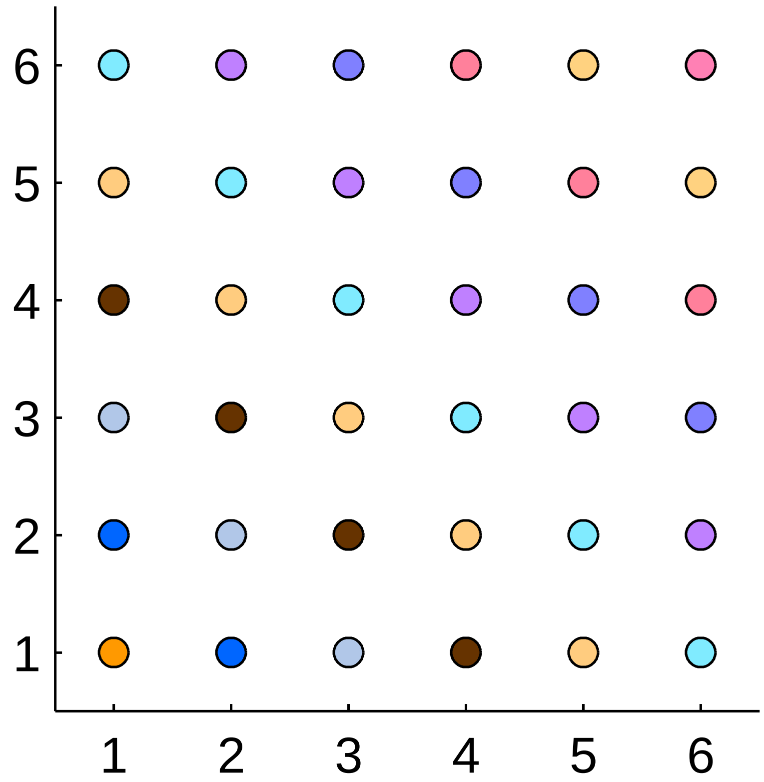
- $p^x(a) \geq 0$ for all $a \in V$
- $\displaystyle\sum_{a \in V} p^x(a) = 1$

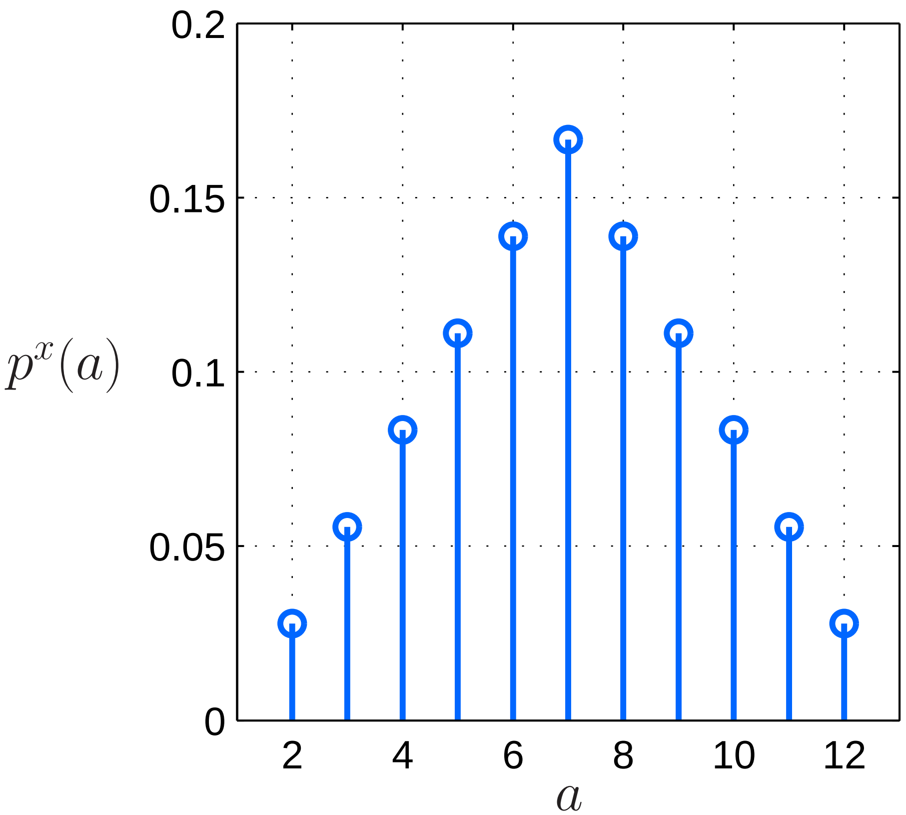
because the events $x^{-1}(a)$ partition Ω, so

$$\sum_{a \in V} p^x(a) = \sum_{a \in V} \mathbf{Prob}\big(x^{-1}(a)\big) = 1$$

Example: sum of two dice

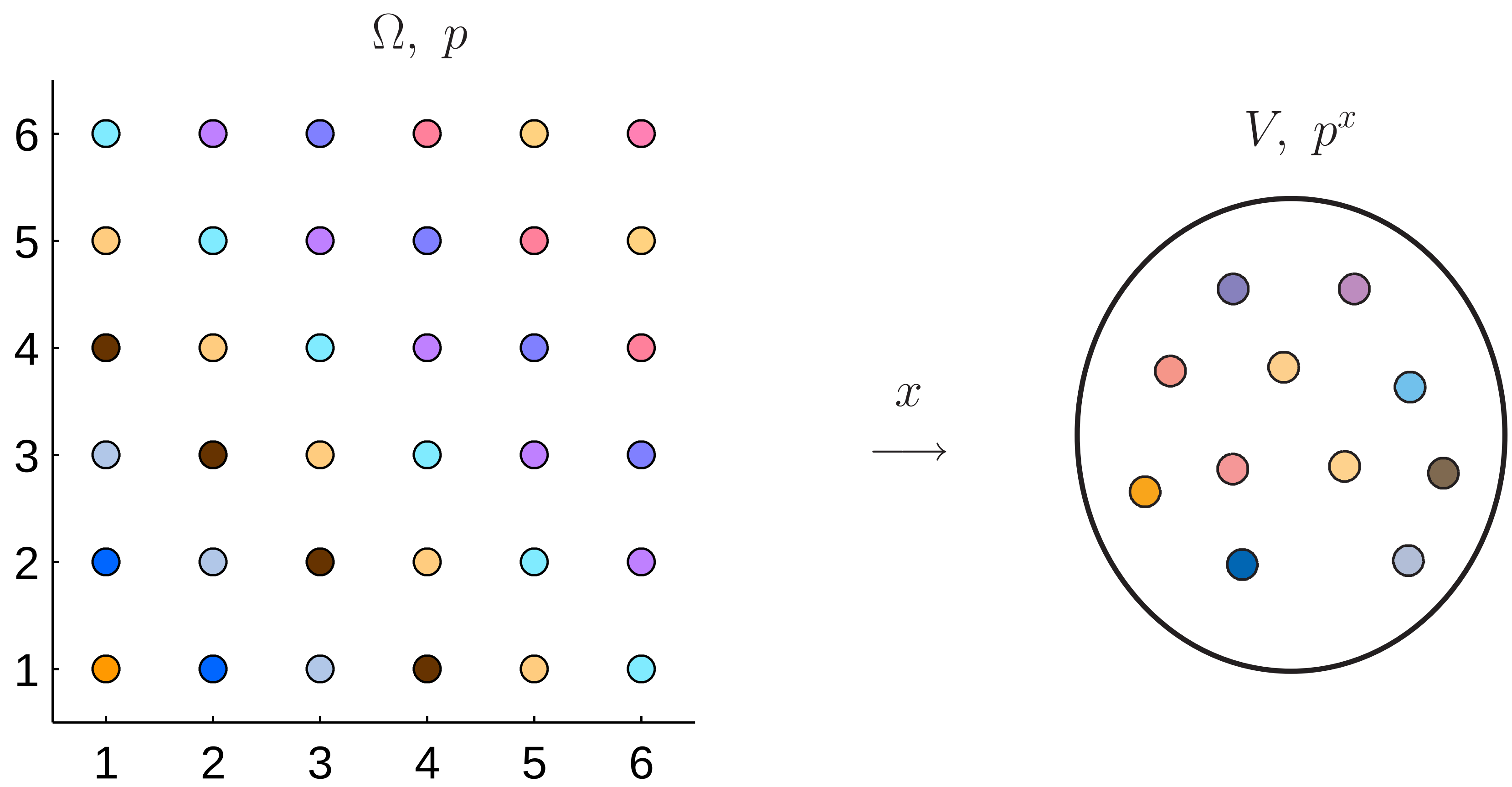
The induced pdf is below





Random variables

Another name for a random variable is a *change of variables*



The map x induces the pmf p^x on V

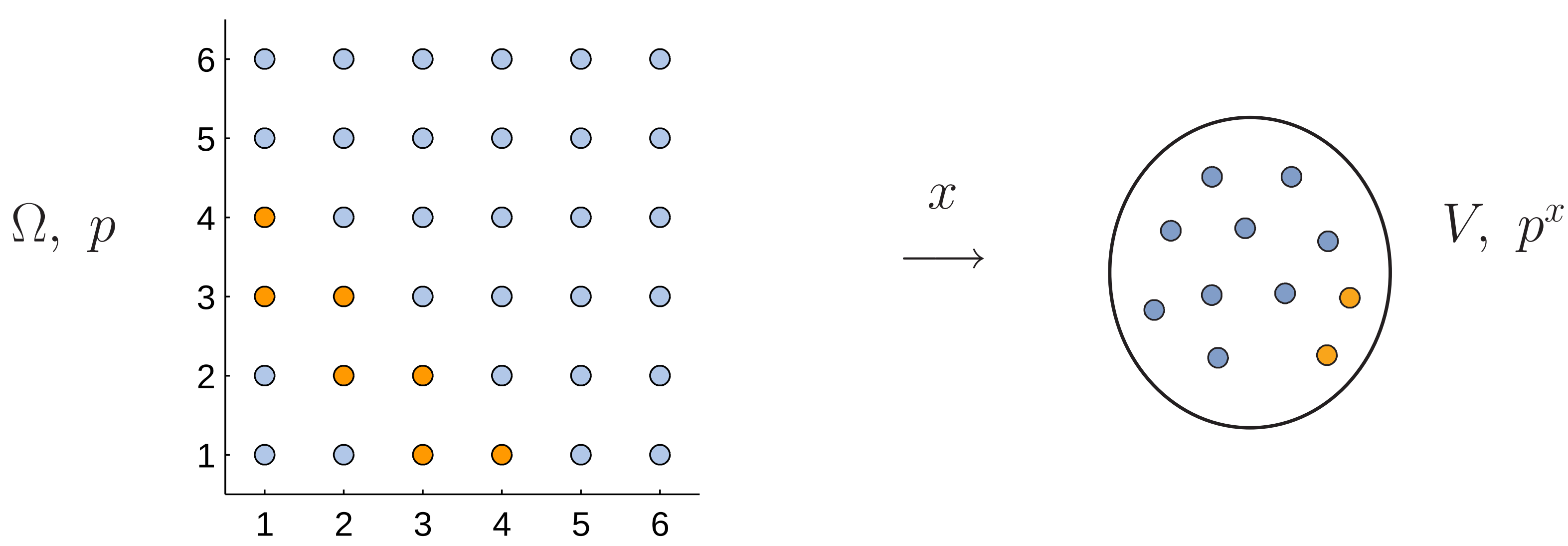
Example: sum of two dice

So there are two ways to compute, for example, $\mathbf{Prob}(x = 4 \text{ or } x = 5)$

- There are seven corresponding outcomes ω in Ω, each with probability $1/36$.

$$\mathbf{Prob}(x = 4 \text{ or } x = 5) = \mathbf{Prob}\Big(\{\, \omega \in \Omega \mid x(\omega) = 4 \text{ or } x(\omega) = 5 \,\}\Big)$$

- Or alternatively: $\mathbf{Prob}(x = 4 \text{ or } x = 5) = p^x(4) + p^x(5)$



Example: sum of two dice

Another example: suppose we want to compute

$$\mathbf{Prob}\big((x-6)^2 = 16\big)$$

- By definition

$$\mathbf{Prob}\big((x-6)^2 = 16\big) = \mathbf{Prob}\Big(\big\{\, \omega \in \Omega \mid \big(x(\omega) - 6\big)^2 = 16 \,\big\}\Big)$$

- Or using the induced pmf

$$\mathbf{Prob}\big((x-6)^2 = 16\big) = \sum_{a \in C} p^x(a)$$

where

$$C = \Big\{\, a \in V \mid (a-6)^2 = 16 \,\Big\}$$

Example: sum of two dice

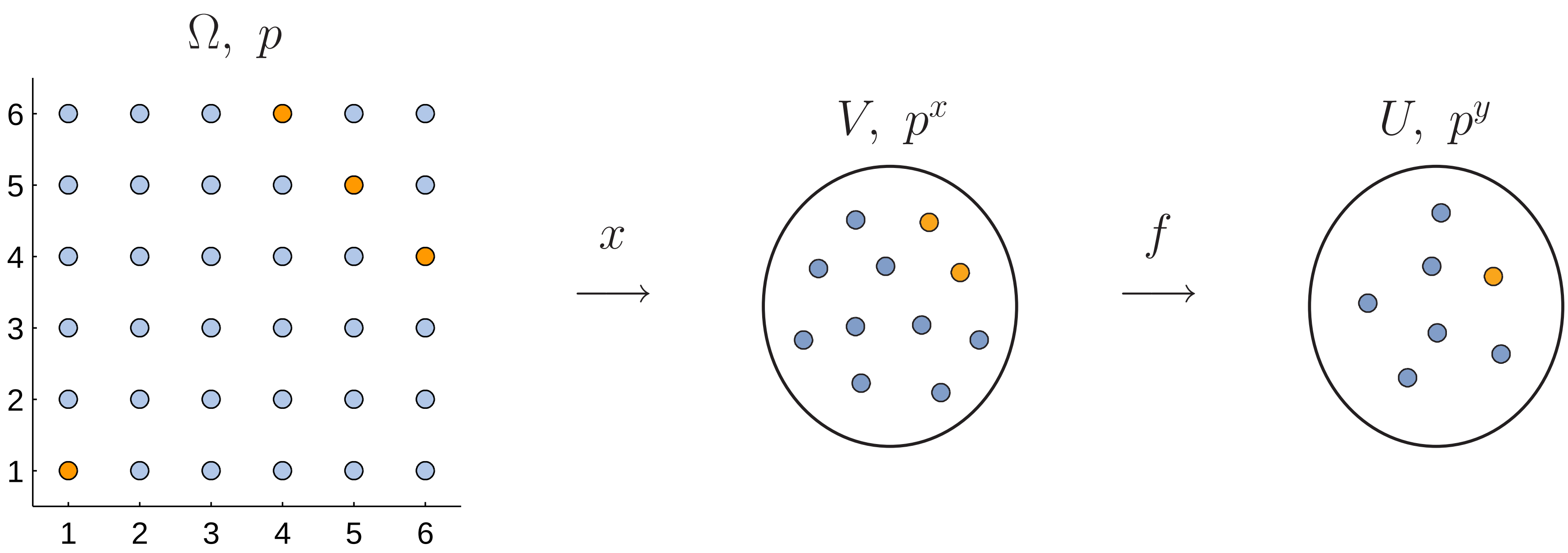
We can also do this another way: let $f : \mathbb{R} \to \mathbb{R}$ be

$$f(x) = (x-6)^2$$

and define the random variable $y = f(x)$, which means $y(\omega) = f(x(\omega))$

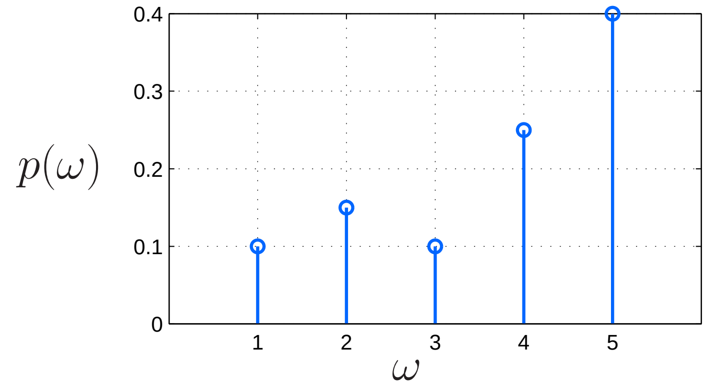
Then

$$\mathbf{Prob}(y = 16) = p^y(16)$$



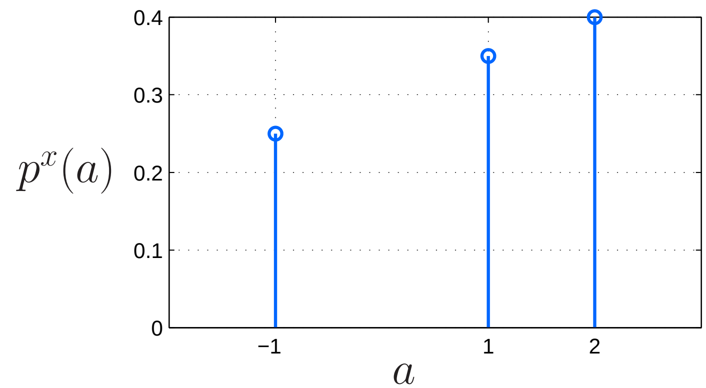
Example: functions of a random variable

- Suppose $\Omega = \{1,2,3,4,5\}$ and p is as shown



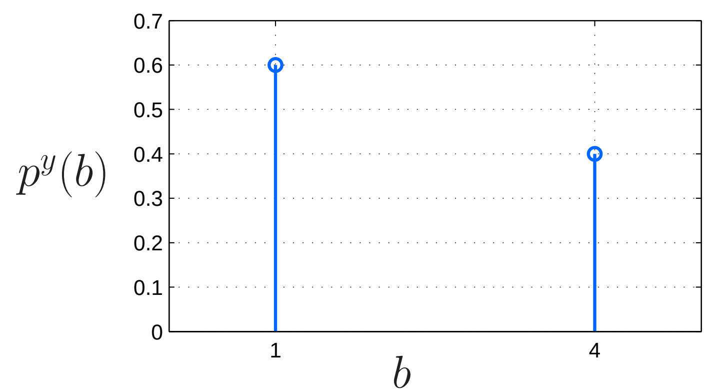
- The random variable $x : \Omega \to \mathbb{R}$ is

$$x(\omega) = \begin{cases} -1 & \text{if } \omega = 1 \text{ or } \omega = 2 \\ 1 & \text{if } \omega = 3 \text{ or } \omega = 4 \\ 2 & \text{if } \omega = 5 \end{cases}$$



- The random variable $y = x^2$, meaning

$$y(\omega) = x(\omega)^2 \quad \text{for all } \omega \in \Omega$$



Functions of a random variable

- x is a random variable $x : \Omega \to V$
- y is a function of x, that is $f : V \to U$ and $y = f(x)$

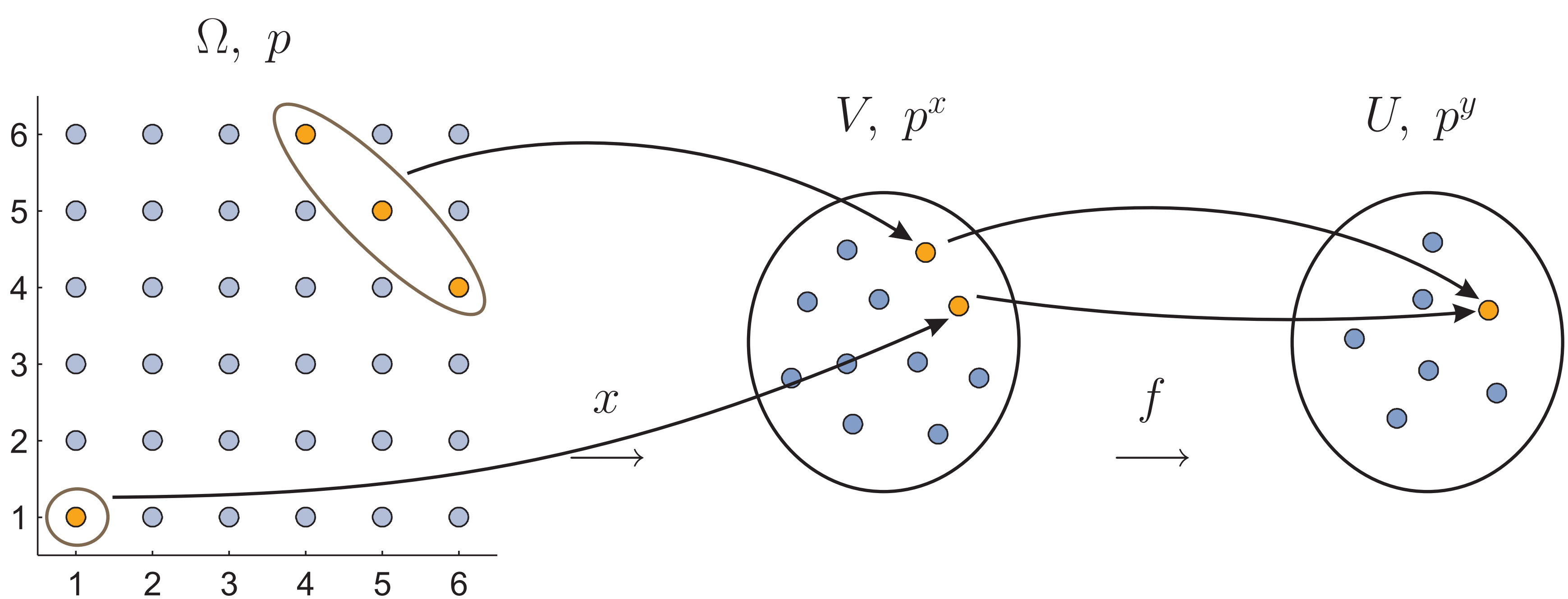
Then y defines a random variable $y(\omega) = f\big(x(\omega)\big)$. The induced pmf of y is

$$p^y(b) = \sum_{a \in V \,|\, y(a) = b} p^x(a)$$

Functions of a random variable

Because

$$
\begin{aligned}
p^y(b) &= \sum_{\omega \in \Omega \,|\, y(x(w))=b} p(\omega) \\
&= \sum_{a \in V \,|\, y(a)=b} \;\; \sum_{\omega \in \Omega \,|\, x(\omega)=a} p(\omega) \\
&= \sum_{a \in V \,|\, y(a)=b} p^x(a)
\end{aligned}
$$



Induced sample spaces

We've seen two ways to compute $\mathbf{Prob}(y = b)$

- As a sum over the sample space Ω

$$\mathbf{Prob}(y = b) = \sum_{\omega \in \Omega \,|\, y(x(w)) = b} p(\omega)$$

- As a sum over the set V

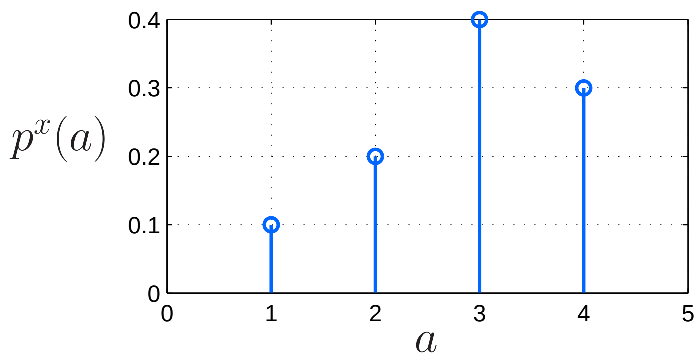
$$\mathbf{Prob}(y = b) = \sum_{a \in V \,|\, y(a) = b} p^x(a)$$

Hence we can think of V as a new sample space, called the *induced sample space*, with pmf $p^x : V \to [0, 1]$

We can compute probabilities of functions of x without knowing the original sample space Ω and the pmf p.

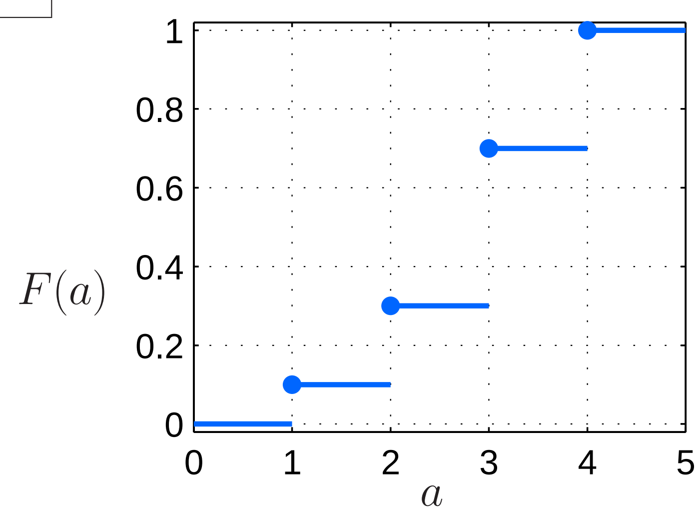
The cumulative distribution

Suppose $x : \Omega \to \mathbb{R}$ is a real-valued random variable; for example



The *cumulative distribution* (cdf) of x is a function $F : \mathbb{R} \to \mathbb{R}$ given by

$$F(a) = \mathbf{Prob}(x \le a)$$



- F is *piecewise constant*
- F is *right continuous*

The uniform random variable

In many codes, one has access to a *uniform random number generator*.

The key property is, for $0 \le a \le b \le 1$

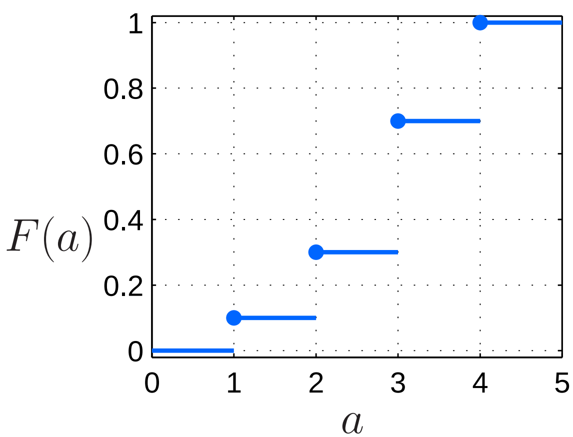
$$\mathbf{Prob}(u \in [a, b]) = b - a$$

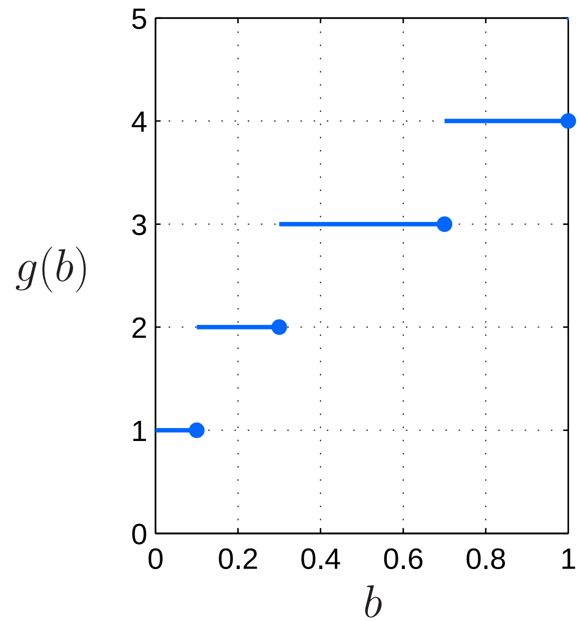
- In Matlab this is `u=rand;` *not* `randn`.
- More on continuous random variables later...

Simulation of random variables

Suppose $x : \Omega \to \mathbb{R}$ is a random variable with cdf F

Define the function $g : \mathbb{R} \to \mathbb{R}$ as below (it's almost the inverse of F)





If u is *uniform*, then

$$\mathbf{Prob}\big(g(u) = a\big) = \mathbf{Prob}(x = a)$$

and so one can simulate x by setting $x = g(u)$.

Expectation

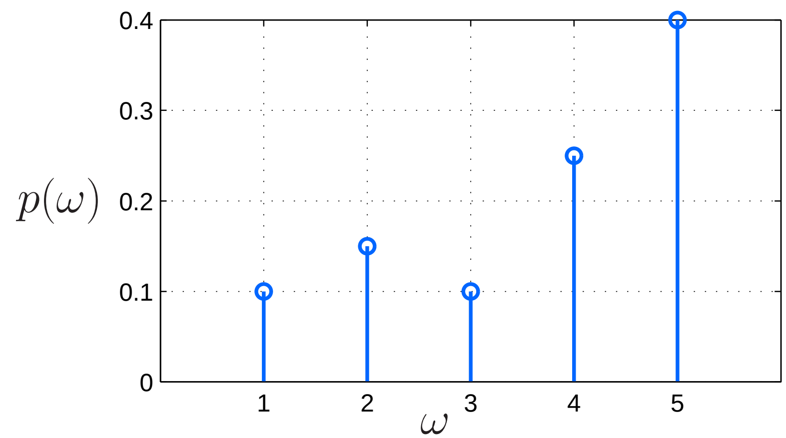
Suppose $x : \Omega \to \mathbb{R}$ is a real-valued random variable. The *expectation* of x is

$$\mathbf{E}\, x = \sum_{\omega \in \Omega} x(\omega) p(\omega)$$

- Also called the *mean* of x or the *expected value* of x

Example: expectation

Suppose $\Omega = \{1, 2, 3, 4, 5\}$, and p is plotted



Let random variable $x : \Omega \to R$ be

$$x(a) = \begin{cases} -1 & \text{if } a = 1 \text{ or } a = 2 \\ 1 & \text{if } a = 3 \text{ or } a = 4 \\ 2 & \text{if } a = 5 \end{cases}$$

The expectation is $\mathbf{E}\, x = -0.1 - 0.15 + 0.1 + 0.25 + 2(0.4) = 0.9$

Vector spaces

The set of real-valued random variables is a vector space.

Because if x and y are two random variables, so is $\lambda x + \mu y$.

- Suppose $\Omega = \{\, \omega_1, \omega_2, \dots, \omega_n \,\}$
- Suppose $x : \Omega \to \mathbb{R}$ is a random variable.
- Define the vector $r \in \mathbb{R}^n$ by

$$r_i = x(\omega_i) \qquad \text{for all } i = 1, \dots, n$$

Usually we *abuse notation* and use x to denote both the vector $r \in \mathbb{R}^n$ and the random variable $x : \Omega \to \mathbb{R}$.

Vector spaces

We can also represent the pmf $p : \Omega \to [0,1]$ by a vector.

Define the vector $p \in \mathbb{R}^n$ (again abusing notation) by

$$p_i = p(\omega_i) \qquad \text{for all } i = 1, \dots, n$$

- The vector p defines a pmf if and only if $\mathbf{1}^T p = 1$ and $p \succeq 0$, where
  - $p \succeq 0$ means $p_i \geq 0$ for all $i = 1, \dots, n$
  - $\mathbf{1}$ is the vector of all ones

- A vector p satisfying these conditions is called a *distribution* vector

Expectation and vector representations

It's easy to compute the expected value of the random variable x.

$$\boxed{\mathbf{E}\,x = p^T x}$$

Because

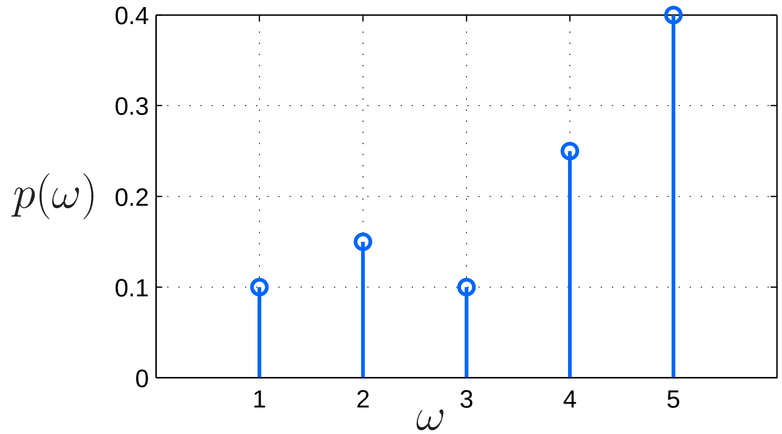
$$\begin{aligned}\mathbf{E}\,x &= \sum_{\omega\in\Omega} x(\omega)p(\omega) \\ &= \sum_{i=1}^{n} x_i p_i\end{aligned}$$

Hence expectation is *linear*

$$\boxed{\mathbf{E}(\alpha x + \beta y) = \alpha\,\mathbf{E}\,x + \beta\,\mathbf{E}\,y}$$

Example: expectation

Suppose $\Omega = \{1, 2, 3, 4, 5\}$, and p is plotted



The random variable $x = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 2 \end{bmatrix}$ and $p = \begin{bmatrix} 0.1 \\ 0.15 \\ 0.1 \\ 0.25 \\ 0.4 \end{bmatrix}$

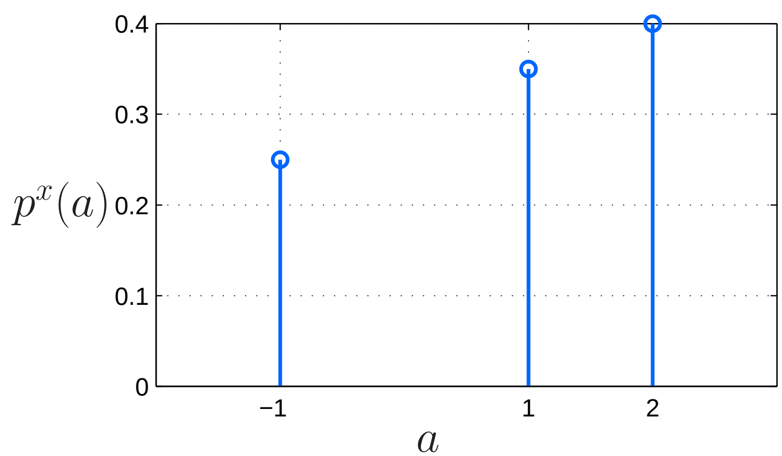
The expectation is $\mathbf{E}\, x = p^T x = -0.1 - 0.15 + 0.1 + 0.25 + 2(0.4) = 0.9$

Another way to compute expectation

Suppose $x : \Omega \to V$ and $V \subset \mathbb{R}$. The *expectation* of x is also given by

$$\mathbf{E}\, x = \sum_{a \in V} a p^x(a)$$

e.g., the random variable $x : \Omega \to \mathbb{R}$ has induced pmf as shown.



So the expectation is

$$\mathbf{E}\, x = -0.25 + 0.35 + 2(0.4) = 0.9$$

Another way to compute expectation

Because

$$\sum_{\omega\in\Omega} x(\omega)p(\omega) = \sum_{a\in V} \quad \sum_{\omega\in\Omega,\ x(\omega)=a} x(\omega)p(\omega)$$

$$= \sum_{a\in V} \ a \sum_{\omega\in\Omega,\ x(\omega)=a} p(\omega)$$

$$= \sum_{a\in V} a p^x(a)$$

Again there are two ways to compute

- summing over Ω

$$\mathbf{E}\, x = \sum_{\omega\in\Omega} x(\omega)p(\omega)$$

- summing over V

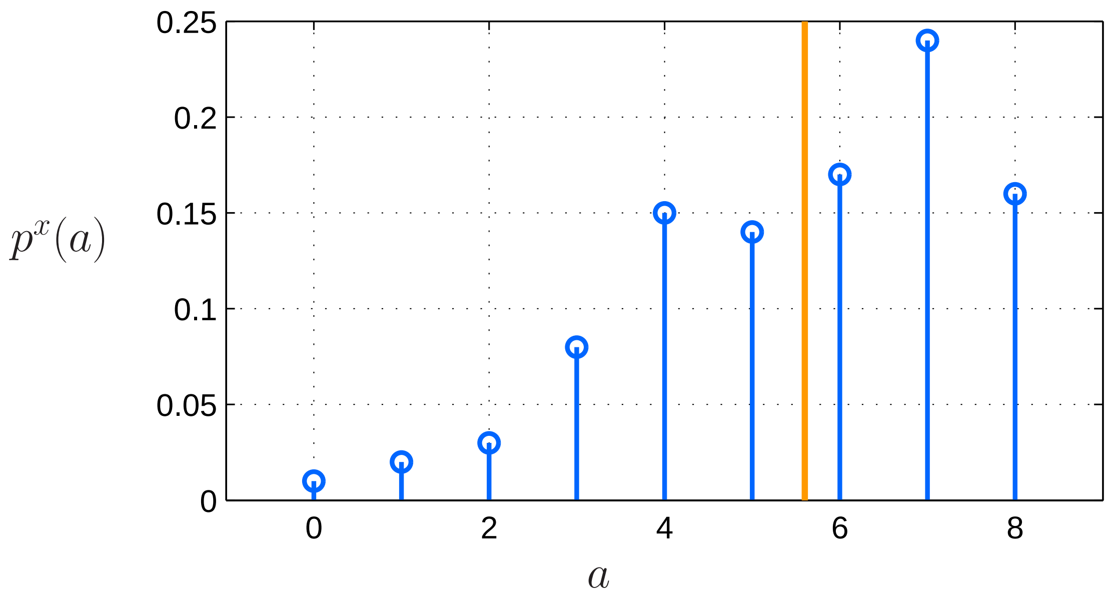
$$\mathbf{E}\, x = \sum_{a\in V} a p^x(a)$$

Interpreting the mean

The mean is

$$\mathbf{E}\,x = \sum_{a \in \mathbb{R}} a p^x(a)$$

- We interpret the mean as the *center of mass* of the distribution
- The plot below shows the *induced pmf of x*



Variance

Suppose $x : \Omega \to \mathbb{R}$ is a random variable. The *covariance* of x is

$$\boxed{\mathbf{cov}(x) = \mathbf{E}\Big((x - \mathbf{E}\,x)^2\Big)}$$

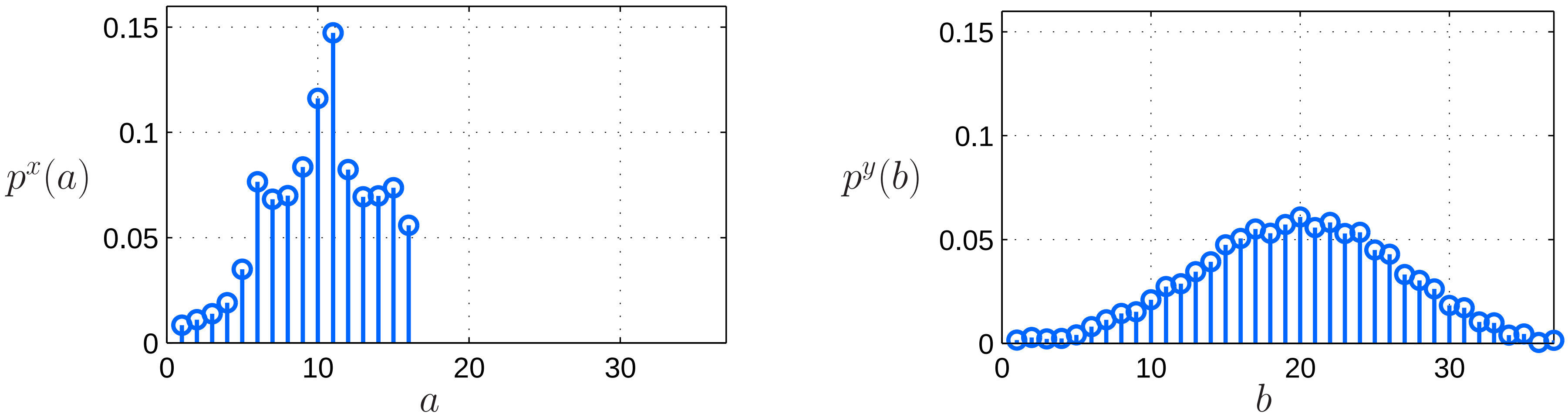
- Measures the *mean square deviation from the mean*
- Another common notation: the *standard deviation* is

$$\mathbf{std}(x) = \sqrt{\mathbf{cov}(x)}$$

- The covariance is also called the *variance*

Intepreting the covariance

The following are the induced pmfs of two random variables



Standard deviations are $\mathbf{std}(x) = 3.5$ and $\mathbf{std}(y) = 6.5$.

- The covariance gives a measure of how *wide* the range of values of a random variable extends around the mean.
- A small covariance means that the pmf is concentrated around the mean

Variance

We have the variance is

$$\mathbf{cov}(x) = \mathbf{E}\Big((x - \mathbf{E}\,x)^2\Big)$$

What this means is:

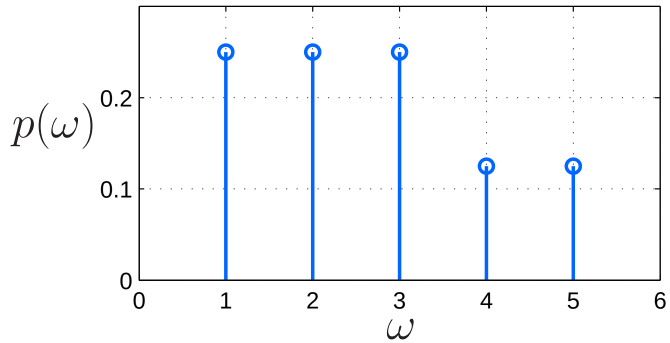
- Let $\mu \in \mathbb{R}$ be the expected value of x; i.e., $\mu = \mathbf{E}\,x$.
- Define a new random variable $y : \Omega \to \mathbb{R}$ by

$$y(\omega) = (x(\omega) - \mu)^2 \qquad \text{for all } \omega \in \Omega$$

- Then $\mathbf{cov}(x) = \mathbf{E}\,y$
- Several ways to compute this: by summing over Ω, or summing over the values of x, or summing over the values of y

Example: variance

Suppose $\Omega = \{1, 2, 3, 4, 5\}$ and p is below.



The random variable x is $x(\omega) = \begin{cases} 3 & \text{if } \omega = 1 \text{ or } \omega = 2 \\ 4 & \text{if } \omega = 3 \\ 6 & \text{if } \omega = 4 \text{ or } \omega = 5 \end{cases}$

Hence $\mathbf{E}\, x = 4$, and the random variable $y = (x - \mathbf{E}\, x)^2$ is

$$y(\omega) = \begin{cases} (3-4)^2 & \text{if } \omega = 1 \text{ or } \omega = 2 \\ (4-4)^2 & \text{if } \omega = 3 \\ (6-4)^2 & \text{if } \omega = 4 \text{ or } \omega = 5 \end{cases}$$

Hence $\mathbf{cov}(x) = \mathbf{E}(y) = 1.5$

Mean-variance decomposition

The *mean square* of x is $\mathbf{E}(x^2)$. We have

$$\boxed{\mathbf{E}(x^2) = (\mathbf{E}\, x)^2 + \mathbf{cov}(x)}$$

Called the *mean-variance decomposition.*

Easy to see; for convenience let $\mu = \mathbf{E}\, x$. Then

$$\begin{aligned}
\mathbf{cov}(x) &= \mathbf{E}\big((x-\mu)^2\big) \\
&= \mathbf{E}\big(x^2 - 2\mu x + \mu^2\big) \\
&= \mathbf{E}\big(x^2\big) - 2\mu\, \mathbf{E}\, x + \mu^2 \\
&= \mathbf{E}\big(x^2\big) - \mu^2
\end{aligned}$$

Moments of a random variable

Suppose $x : \Omega \to \mathbb{R}$ is a random variable. The *n'th moment* of x is

$$\mathbf{E}(x^n) = \sum_{\omega \in \Omega} x(\omega)^n p(\omega)$$

- The mean $\mathbf{E}\,x$ is the first moment of x
- The covariance is the second moment minus the square of the first moment.