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 $\Omega$ 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.

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

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

\[
\Omega = \left\{ (x, y) \mid x \in \{0, 1\} \text{ and } y \in \{0, 1, \ldots, n\} \right\}
\]

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

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

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* $y = v + \text{position at time } n/2$
  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

\[
y = A(u + r) + v \\
x = P(u + r)
\]
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 $\Omega$, a finite set
- probability mass function $p : \Omega \rightarrow [0, 1]$
- a random variable $x : \Omega \rightarrow \mathbb{R}$

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

$$\text{Prob}\left(\left\{ \omega \in \Omega \mid x(\omega) = a \right\}\right)$$

This is equal to

$$\sum_{\omega \in \Omega \mid x(\omega) = a} p(\omega)$$
Example: two dice

We have sample space

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

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

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

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

$$A = \left\{ \omega \in \Omega \mid x(\omega) = 5 \right\}$$
Probability and random variables

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

$$\text{Prob}(x = a) = \text{Prob}\left(\left\{ \omega \in \Omega \mid x(\omega) = a \right\}\right)$$

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

The random variable $x : \Omega \rightarrow \mathbb{R}$ 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 $\text{Prob}(x = a)$ 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 $\text{Prob}(x = a)$

- This is also written $p^x(a) = \text{Prob}(x = a)$
Notation for random variables

There are many similar notations used: for example, define

- \( \text{Prob}(x = a) = \text{Prob}(x^{-1}(a)) \)

- \( \text{Prob}(x \geq a) = \text{Prob}(\{ \omega \in \Omega \mid x(\omega) \geq a \}) \)

- If \( C \subset V \), then \( \text{Prob}(x \in C) = \text{Prob}(\{ \omega \in \Omega \mid x(\omega) \in C \}) \)
Events corresponding to random variables

Suppose \( x : \Omega \rightarrow V \). Each \( a \in V \) defines an event

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

These events partition \( \Omega \)

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

\[
x^{-1}(-1) = \{1, 2\} \quad x^{-1}(1) = \{3, 4\} \quad x^{-1}(2) = \{5\}
\]
Example: sum of two dice

The events are

\[ \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 2 & 3 & 4 & 5 & 6 \\
\end{array} \]
### Induced probability

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

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

It satisfies the properties of a probability mass function

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

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

$$\sum_{a \in V} p^x(a) = \sum_{a \in V} \text{Prob}(x^{-1}(a)) = 1$$
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, $\text{Prob}(x = 4 \text{ or } x = 5)$

- There are seven corresponding outcomes $\omega$ in $\Omega$, each with probability $1/36$.

$$\text{Prob}(x = 4 \text{ or } x = 5) = \text{Prob}\left(\{\omega \in \Omega \mid x(\omega) = 4 \text{ or } x(\omega) = 5\}\right)$$

- Or alternatively:

$$\text{Prob}(x = 4 \text{ or } x = 5) = p_x(4) + p_x(5)$$
Example: sum of two dice

Another example: suppose we want to compute

\[ \text{Prob}\left((x - 6)^2 = 16\right) \]

• By definition

\[ \text{Prob}\left((x - 6)^2 = 16\right) = \text{Prob}\left(\left\{ \omega \in \Omega \mid (x(\omega) - 6)^2 = 16 \right\}\right) \]

• Or using the induced pmf

\[ \text{Prob}\left((x - 6)^2 = 16\right) = \sum_{a \in C} p^x(a) \]

where

\[ C = \left\{ a \in V \mid (a - 6)^2 = 16 \right\} \]
Example: sum of two dice

We can also do this another way: let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be

\[
f(x) = (x - 6)^2
\]

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

Then

\[
\text{Prob}(y = 16) = p^y(16)
\]
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

  $$y(\omega) = x(\omega)^2 \text{ for all } \omega \in \Omega$$
Functions of a random variable

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

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

$$p^y(b) = \sum_{a \in V \mid y(a) = b} p^x(a)$$
Functions of a random variable

Because

\[ p^y(b) = \sum_{\omega \in \Omega \mid y(x(\omega)) = b} p(\omega) \]

\[ = \sum_{a \in V \mid y(a) = b} \sum_{\omega \in \Omega \mid x(\omega) = a} p(\omega) \]

\[ = \sum_{a \in V \mid y(a) = b} p^x(a) \]
Induced sample spaces

We’ve seen two ways to compute \( \text{Prob}(y = b) \)

- As a sum over the sample space \( \Omega \)
  \[
  \text{Prob}(y = b) = \sum_{\omega \in \Omega \mid y(x(w)) = b} p(\omega)
  \]

- As a sum over the set \( V \)
  \[
  \text{Prob}(y = b) = \sum_{a \in V \mid y(a) = b} p^x(a)
  \]

Hence we can think of \( V \) as a new sample space, called the \textit{induced sample space}, with pmf \( p^x : V \rightarrow [0, 1] \)

We can compute probabilities of functions of \( x \) without knowing the original sample space \( \Omega \) and the pmf \( p \).
The cumulative distribution

Suppose \( x : \Omega \rightarrow \mathbb{R} \) is a real-valued random variable; for example

\[
P_x(a) = \begin{cases} 
0 & a < 0 \\
0.1 & 0 \leq a < 1 \\
0.4 & 1 \leq a < 2 \\
0.3 & 2 \leq a < 3 \\
0.2 & 3 \leq a < 4 \\
0.2 & 4 \leq a < 5 \\
1 & a \geq 5
\end{cases}
\]

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

\[
F(a) = \text{Prob}(x \leq a)
\]

- \( F \) 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 $0 \leq a \leq b \leq 1$

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

- In Matlab this is \texttt{u=rand}; *not* \texttt{randn}.

- More on continuous random variables later…
Simulation of random variables

Suppose $x : \Omega \rightarrow \mathbb{R}$ is a random variable with cdf $F$
Define the function $g : \mathbb{R} \rightarrow \mathbb{R}$ as below (it’s almost the inverse of $F$)

If $u$ is uniform, then

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

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

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

\[
E_x = \sum_{\omega \in \Omega} x(\omega)p(\omega)
\]

- 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

$$E x = -0.1 - 0.15 + 0.1 + 0.25 + 2(0.4) = 0.9$$
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 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$
- Suppose $x : \Omega \to \mathbb{R}$ is a random variable.
- Define the vector $r \in \mathbb{R}^n$ by
  \[ r_i = x(\omega_i) \quad \text{for all } i = 1, \ldots, n \]

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 $p : \Omega \rightarrow [0, 1]$ by a vector.

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

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

- The vector $p$ defines a pmf if and only if $1^T p = 1$ and $p \succeq 0$, where
  - $p \succeq 0$ means $p_i \geq 0$ for all $i = 1, \ldots, n$
  - $1$ 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 $x$.

$$E x = p^T x$$

Because

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

$$= \sum_{i=1}^{n} x_i p_i$$

Hence expectation is *linear*

$$E(\alpha x + \beta y) = \alpha E x + \beta E y$$
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 \textit{expectation} of \( x \) is also given by

\[
E_x = \sum_{a \in V} a p^x(a)
\]

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

So the expectation is

\[
E_x = -0.25 + 0.35 + 2(0.4) = 0.9
\]
Another way to compute expectation

Because

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

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

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

Again there are two ways to compute

- summing over $\Omega$

$$E_x = \sum_{\omega \in \Omega} x(\omega)p(\omega)$$

- summing over $V$

$$E_x = \sum_{a \in V} ap^x(a)$$
Interpreting the mean

The mean is

\[ E x = \sum_{a \in \mathbb{R}} ap^x(a) \]

- We interpret the mean as the \textit{center of mass} of the distribution
- The plot below shows the \textit{induced pmf of} \( x \)
**Variance**

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

$$\text{cov}(x) = \mathbb{E}\left((x - \mathbb{E}x)^2\right)$$

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

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

- The covariance is also called the *variance*
Interpreting the covariance

The following are the induced pmfs of two random variables

\[ p^x(a) \]
\[ p^y(b) \]

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

- The covariance gives a measure of how \textit{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

\[ \text{cov}(x) = E\left((x - E(x))^2\right) \]

What this means is:

- Let \( \mu \in \mathbb{R} \) be the expected value of \( x \); i.e., \( \mu = E(x) \).
- Define a new random variable \( y : \Omega \rightarrow \mathbb{R} \) by
  \[ y(\omega) = (x(\omega) - \mu)^2 \quad \text{for all } \omega \in \Omega \]
- Then \( \text{cov}(x) = E(y) \)
- Several ways to compute this: by summing over \( \Omega \), 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 $E(x) = 4$, and the random variable $y = (x - E(x))^2$ 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 $\text{cov}(x) = E(y) = 1.5$
Mean-variance decomposition

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

\[
\mathbb{E}(x^2) = (\mathbb{E}x)^2 + \text{cov}(x)
\]

Called the *mean-variance decomposition*.

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

\[
\text{cov}(x) = \mathbb{E}((x - \mu)^2)
\]

\[
= \mathbb{E}(x^2 - 2\mu x + \mu^2)
\]

\[
= \mathbb{E}(x^2) - 2\mu \mathbb{E}x + \mu^2
\]

\[
= \mathbb{E}(x^2) - \mu^2
\]
Moments of a random variable

Suppose \( x : \Omega \rightarrow \mathbb{R} \) is a random variable. The \( n \)'th moment of \( x \) is

\[
E(x^n) = \sum_{\omega \in \Omega} x(\omega)^n p(\omega)
\]

- The mean \( E x \) is the first moment of \( x \)
- The covariance is the second moment minus the square of the first moment.