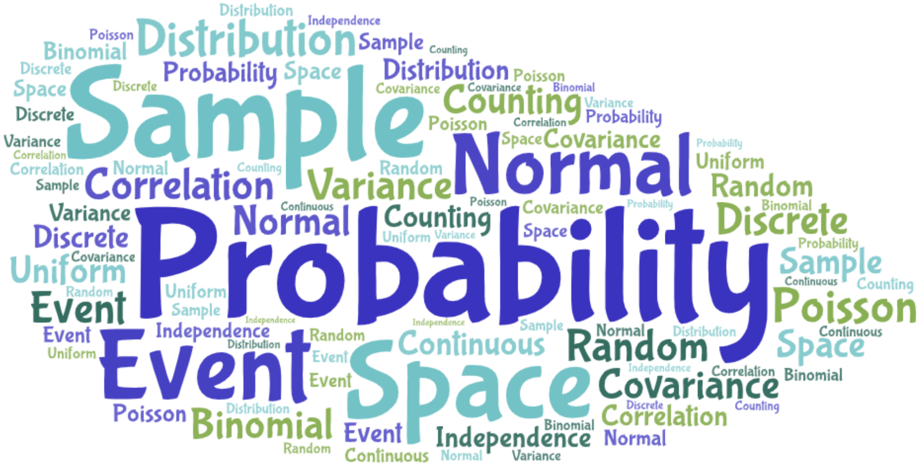
Discrete Mathematics and Probability

Week 8



Chris Heunen

Topics

- Independence: what information changes probability
- Random variables: when variables depend on chance
- Expectation: most likely outcomes of experiment
- Variance: how much the experiment can deviate

Independence

Independence

Sometimes partial information on an experiment
does not change the likelihood of an event.

Definition

Events E and F are *independent* if $\mathbf{P}(E \mid F) = \mathbf{P}(E)$.
Equivalently: $\mathbf{P}(E \cap F) = \mathbf{P}(E) \cdot \mathbf{P}(F)$.
Equivalently: $\mathbf{P}(F \mid E) = \mathbf{P}(F)$.

Proposition

If E and F are independent events,
then E and F^c are also independent.

independent events $\neq$ mutually exclusive events !

independence usually either trivial or tricky...

Examples

$\Omega =$ two dice

$E = \{\text{sum is } 6\}$

$F = \{\text{first is } 3\}$

$\Big\}$ not independent: $\frac{1}{36} = P(E \cap F) \neq P(E) \cdot P(F) = \frac{5}{36} \cdot \frac{1}{6}$

$E' = \{\text{sum is } 7\}$

$F = \{\text{first is } 3\}$

$\Big\}$ independent! $\frac{1}{36} = P(E' \cap F) = P(E) \cdot P(F) = \frac{6}{36} \cdot \frac{1}{6} = \frac{1}{36}$

$G = \{\text{second is } 4\}$

$\Big\}$ pairwise independent: $\frac{1}{36} = P(E' \cap G) = P(E') \cdot P(G) = \frac{6}{36} \cdot \frac{1}{6} = \frac{1}{36}$

$\frac{1}{36} = P(F \cap G) = P(F) \cdot P(G) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$

but $\quad 1 = P(E' \mid F \cap G) \neq P(E) = \frac{1}{6}$ ..

equivalently $\quad \frac{1}{36} = P(E' \cap F \cap G) \neq P(E') \cdot P(F \cap G) = P(E') \cdot P(F) \cdot P(G) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{6}{36} = \frac{1}{6}$

Independence

Definition

Three events E, F, G are *(mutually) independent* if:

$$\begin{aligned}
\mathbf{P}\{E \cap F\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{F\}, \\
\mathbf{P}\{E \cap G\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{G\}, \\
\mathbf{P}\{F \cap G\} &= \mathbf{P}\{F\} \cdot \mathbf{P}\{G\}, \\
\mathbf{P}\{E \cap F \cap G\} &= \mathbf{P}\{E\} \cdot \mathbf{P}\{F\} \cdot \mathbf{P}\{G\}.
\end{aligned}$$

For more events the definition is that any (finite) subset of them have this factorisation property.

Examples

Fix parameter $0<p<1$

n independent experiments, each succeeds with probability p

- chance that every one succeeds?

$$p^n \xrightarrow[n\to\infty]{} 0$$

- chance that at least one succeeds? Murphy's Law

$$1 - P(\text{each one fails}) = 1-(1-p)^n \xrightarrow[n\to\infty]{} 1$$

- chance that exactly k succeed?

$$\binom{n}{k}\cdot p^k \cdot (1-p)^{n-k}$$

Random variables

Random variables

"Random variables $\approx$ random numbers". But *random* means that there must be some kind of experiment behind these numbers.

Definition

A *random variable* is a function from the sample space Ω to the real numbers $\mathbb{R}$.

flip 3 coins, X counts Heads:

$$X(T,T,T) = 0$$
$$X(H,T,T) = X(T,H,T) = X(T,T,H) = 1$$
$$X(H,H,T) = X(H,T,H) = X(T,H,H) = 2$$
$$X(H,H,H) = 3$$

$$P(X=1) = P(\{(T,T,H),(T,H,T),(H,T,T)\}) = \frac{3}{8}$$

Discrete random variables

Definition

A random variable X that can take on countably many possible values is called *discrete*.

e.g.:
- nr Heads in 3 coin flips
- nr flips needed to first see a Head

Probability mass function

Definition

The *probability mass function (pmf)*, or *distribution* of a discrete random variable X gives the probabilities of its possible values:

$$\mathfrak{p}_X(x_i) = \mathbf{P}(X = x_i),$$

Proposition

$$\mathfrak{p}(x_i) \geq 0 \qquad \textit{and} \qquad \sum_i \mathfrak{p}(x_i) = 1$$

Vice versa: any function with these properties that is nonzero on only countably many values x_i, is a p.m.f.

Examples

- nr Heads in 3 coin flips:

$$p(0) = p(3) = \frac{1}{8} \qquad p(1) = p(2) = \frac{3}{8} \qquad \text{indeed} \quad \sum_{i=0}^{3} p(i) = 1$$

- fix parameter $\lambda > 0$. define $p(i) = c \cdot \frac{\lambda^i}{i!}$.

  which c make this a p.m.f.?

$$p(i) \geq 0 \quad \text{iff} \quad c \geq 0.$$

$$\sum_{i=0}^{\infty} p(i) = c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = c \cdot e^{\lambda} = 1 \implies c = e^{-\lambda}$$

so e.g. $P(X=0) = p(0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda}$

$$\begin{aligned} P(X>2) &= 1 - P(X \leq 2) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - e^{-\lambda} - e^{-\lambda} \cdot \lambda - e^{-\lambda} \frac{\lambda^2}{2} \end{aligned}$$

Cumulative distribution function

Definition

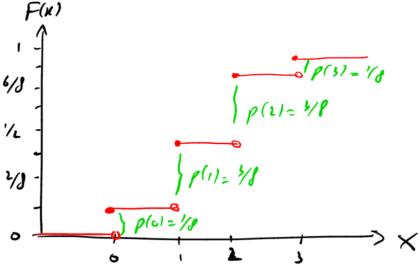
The *cumulative distribution function (cdf)* of a random variable X:

$$F\colon \mathbb{R} \to [0, 1], \qquad x \mapsto F(x) = \mathbf{P}(X \leq x).$$

contains all relevant info about X

e.g. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

Flip 3 coins, X counts Heads



Cumulative distribution function

Proposition

A cumulative distribution function F:

- *is non-decreasing: if $x \leq y$ then $F(x) \leq F(y)$*
- *has limit $\lim_{x \to -\infty} F(x) = 0$ on the left*
- *has limit $\lim_{x \to \infty} F(x) = 1$ on the right*

Expectation

Expectation

Once we have a random variable, we want to quantify its *typical* behaviour in some sense. Two of the most often used quantities for this are the *expectation* and the *variance*.

Definition

The *expectation* of a discrete random variable X is:

$$\mathbf{E}X = \sum_i x_i \cdot \mathfrak{p}(x_i)$$

provided the sum exists. Also called *mean*, or *expected value*.

weighted average / center of mass

Examples

$$X = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E^c \text{ occurs} \end{cases}$$ "indicator variable"

$p(1) = P(E)$

$p(0) = 1 - P(E)$

$\mathbb{E}X = 0 \cdot p(0) + 1 \cdot p(1) = P(E)$

X = nr on fair die after roll

$$\mathbb{E}X = \sum_{i=1}^{6} i \cdot p(i) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \dots + \frac{6}{6} = \frac{7}{2}$$

expectation need not be possible value!

Properties of expectation

Proposition (expectation of a function of a random variable)

If X is a discrete random variable, and $g : \mathbb{R} \to \mathbb{R}$ a function, then:

$$\mathbf{E}g(X) = \sum_i g(x_i) \cdot \mathfrak{p}(x_i) \qquad \textit{(if it exists)}$$

Corollary (expectation is linear)

If X is a discrete random variable, and a, b fixed real numbers:

$$\mathbf{E}(aX + b) = a \cdot \mathbf{E}X + b.$$

Proof: $\mathbb{E}(aX+b) = \sum_i (ax_i + b) \cdot p(x_i) = a \cdot \sum_i x_i p(x_i) + b \sum_i p(x_i)$

$= a \cdot \mathbb{E}X + b \cdot 1$

Moments

Definition (moments)

Let $n \in \mathbb{N}$. The n^{th} *moment* of a random variable X is:

$$\mathbf{E}X^n$$

The n^{th} *absolute moment* of X is:

$$\mathbf{E}|X|^n$$

$$\mathbb{E}X^n = \mathbb{E}(X^n) \neq (\mathbb{E}X)^n \quad !$$

Variance

Example

$X = 0$

$Y = \begin{cases} 1 & \text{w/prob } 1/2 \\ -1 & \text{— } 1/2 \end{cases}$

$Z = \begin{cases} 2 & \text{— } 1/5 \\ -1/2 & \text{— } 4/5 \end{cases}$

$U = \begin{cases} 10 & \text{— } 1/2 \\ -10 & \text{— } 1/2 \end{cases}$

$\mathbb{E}X = \mathbb{E}Y = \mathbb{E}Z = \mathbb{E}U = 0$

expectation cannot distinguish X, Y, Z, U

but clearly different.

3. Variance

$\mathrm{Var}\, X = \mathbb{E}(X-0)^2 = 0^2 = 0$

$SD\, X = \sqrt{0} = 0$

$\mathrm{Var}\, Y = \mathbb{E}(Y-0)^2 = 1^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} = 1$

$SD\, Y = \sqrt{1} = 1$

Definition (variance, standard deviation)

The *variance* and the *standard deviation* of a random variable are:

- $\mathbf{Var}\, X = \mathbf{E}(X - \mathbf{E}X)^2$.
- $\mathrm{SD}\, X = \sqrt{\mathbf{Var}\, X}$.

$\mathrm{Var}\, Z = \mathbb{E}(Z-0)^2 = 2^2 \cdot \frac{1}{5} + (-\frac{1}{2})^2 \cdot \frac{4}{5} = 1$

$SD\, Z = \sqrt{1} = 1$

$\mathrm{Var}\, U = \mathbb{E}(U-0)^2 = 10^2 \cdot \frac{1}{2} + (-10)^2 \cdot \frac{1}{2} = 100$

$SD\, U = \sqrt{100} = 10$

variance gives finer information but still cannot distinguish X and Z

Example

Properties of the variance

Proposition (equivalent form of the variance)

$\mathbf{Var}\, X = \mathbf{E}X^2 - (\mathbf{E}X)^2$ *for any random variable* X.

Corollary

$\mathbf{E}X^2 \geq (\mathbf{E}X)^2$ *for any random variable* X,
with equality only if X *is constant.*

(Pf: book)

Examples

X = roll of fair die

$\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = (1^2+2^2+\cdots+6^2)\cdot\frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

$\mathrm{SD}\, X = \sqrt{35/12} \simeq 1.71$

two most important numbers of fair die: average 3.5
typical deviation 1.71

X = indicator of event E

$\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1^2 \cdot P(E) - (P(E))^2 = P(E)\cdot(1-P(E))$

$\mathrm{SD}\, X = \sqrt{P(E)\cdot(1-P(E))}$

Properties of the variance

Proposition (variance is not linear)

Let X be a random variable, a and b fixed real numbers. Then:

$$\mathbf{Var}(aX + b) = a^2 \cdot \mathbf{Var}\, X$$

Proof: book.

$\mathbb{V}\mathrm{ar}\,(X+b) = \mathbb{V}\mathrm{ar}\,(X) = \mathbb{V}\mathrm{ar}\,(-X)$:

variance is invariant under reflection
and shifting by a constant.

Properties of the variance

Proposition (variance is not linear)

Let X be a random variable, a and b fixed real numbers. Then:

$$\mathbf{Var}(aX + b) = a^2 \cdot \mathbf{Var}\, X$$