TTIC 31150/CMSC 31150 Mathematical Toolkit (Fall 2024)

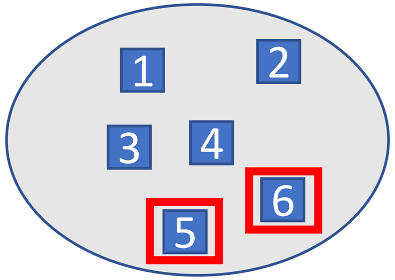
Avrim Blum

Lecture 9: Probability basics

Note: Homework 3 is on webpage. Due Nov 6.

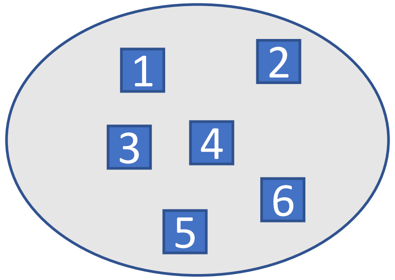
Probability basics

E.g., roll a die, $\Omega = \{1,2,3,4,5,6\}$



- A probabilistic setting (finite case) is defined by a sample space Ω and a probability distribution $\nu: \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} \nu(\omega) = 1$.
- An event A is a subset of the sample space, and we define $\mathbb{P}[A] = \sum_{\omega \in A} \nu(\omega)$. The elements $\omega \in \Omega$ are called elementary events and associated with their singleton sets.
- A real-valued random variable (R.V.) X is a function $X: \Omega \to \mathbb{R}$ (can also talk about vector-valued R.V.s, etc).
  - ➢E.g., roll two dice. Let X_1 = value of die #1, X_2 = value of die #2, $X = X_1 + X_2$.
- Often convenient to go back and forth between events and R.V.s.
  - ➢Given R.V. X and a value b, define event "$X = b$" as $\{\omega: X(\omega) = b\}$.
  - ➢Given event A, define indicator R.V. $X(\omega) = \begin{cases} 1\ if\ \omega \in A \\ 0\ if\ \omega \notin A \end{cases}$

Probability basics

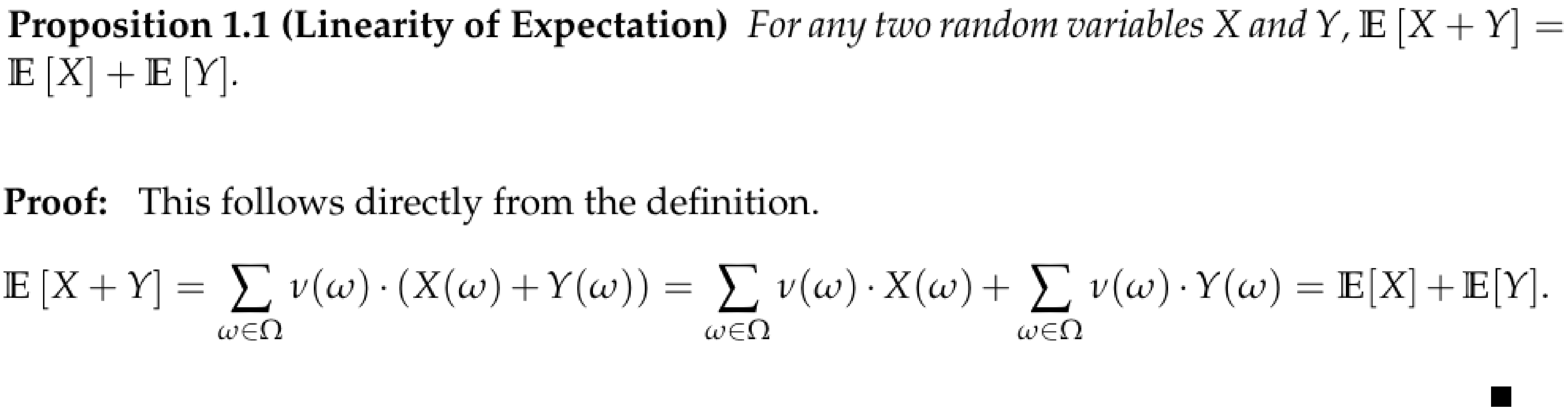


- The expectation $\mathbb{E}[X]$ of a random variable X is $\sum_{\omega \in \Omega} \nu(\omega) X(\omega)$.
  - In other words, it is the probability-weighted average value. E.g., if X_1 is the R.V. for the roll of a die, then $\mathbb{E}[X_1] = 3.5$.

- Often it is convenient to group elementary events by the value of X giving us $\mathbb{E}[X] = \sum_a \mathbb{P}(X = a) \cdot a$.

- An extremely useful property of expectation is that it is a linear transformation. In particular, $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Linearity of Expectation



Proposition 1.1 (Linearity of Expectation) *For any two random variables X and Y,* $\mathbb{E}\left[X+Y\right] = \mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right]$.

Proof: This follows directly from the definition.

$$\mathbb{E}\left[X+Y\right] = \sum_{\omega \in \Omega} \nu(\omega) \cdot (X(\omega) + Y(\omega)) = \sum_{\omega \in \Omega} \nu(\omega) \cdot X(\omega) + \sum_{\omega \in \Omega} \nu(\omega) \cdot Y(\omega) = \mathbb{E}[X] + \mathbb{E}[Y].$$

■

Linearity of Expectation: Example

Card Shuffling:

- Unwrap a deck of n cards and shuffle until ordering is completely random.
- What is expected number of cards that end up in the same position as they started?
- What do we get for $n = 1$? $n = 2$? General n?

Answer: 1. Proof:

- Define X_i to be the indicator R.V. for the event that card i ends up in position i.
- Let $X = X_1 + \cdots + X_n$ be the number of cards that end in their starting position.
- We know $\mathbb{E}[X_i] = 1/n$, so by linearity of expectation, $\mathbb{E}[X] = 1$.

Conditioning

Conditioning on an event A is equivalent to restricting the probability space to A.

- Define $\nu_A(\omega) = \begin{cases} \nu(\omega)/\mathbb{P}[A] & if\ w \in A \\ 0 & if\ w \notin A \end{cases}$

"probability of ω conditioned on A"

- $\mathbb{P}[B|A] = \frac{\mathbb{P}[A \wedge B]}{\mathbb{P}[A]} = \sum_{\omega \in A \cap B} \frac{\nu(\omega)}{\mathbb{P}[A]} = \sum_{\omega \in A \cap B} \nu_A(\omega)$.

- For an R.V. X and event A, define $\mathbb{E}[X|A] = \sum_{\omega \in A} \nu_A(\omega) \cdot X(\omega)$.

- For any partition of Ω into $A_1, A_2, \ldots$, we can rewrite $\mathbb{E}[X]$ as:

$$\mathbb{E}[X] = \sum_i \sum_{\omega \in A_i} \nu(\omega) X(\omega) = \sum_i \mathbb{P}[A_i] \mathbb{E}[X\ |A_i]$$

Example: Random walk stock market

Imagine there is a stock that each day goes up or down by $1 with equal probability (unless it hits $0, in which case it stays there forever).

You begin with $m. At the start of each day you can buy or sell as much as you like. At the end of the year, you must cash out. What strategy maximizes your expected gain?

Answer: it doesn't matter. For any strategy, your expected gain is $0.

- Define X_t to be the gain of your algorithm on day t. Let $X = \mathrm{X}_1 + \cdots + \mathrm{X}_{365}$ be your gain at the end of the year.
- By Linearity of expectation, $\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_{365}]$. So, what is $\mathbb{E}[X_t]$?
- Let A_{ti} be the event that you own i shares on day t. For all i, we have $\mathbb{E}[X_t | A_{ti}] = 0$. So, $\mathbb{E}[X_t] = 0$.

Independence

Two events A and B are independent if $\mathbb{P}(A \wedge B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

- If they have nonzero probability, can write as $\mathbb{P}(A|B) = \mathbb{P}(A)$, or $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Two random variables X, Y are independent if the events “$X = x$” and “$Y = y$” are independent for all x, y.

k events $A_1, \dots, A_k$ are independent if for all $S \subseteq \{1, \dots, k\}$, $\mathbb{P}(\bigwedge_{i \in S} A_i) = \prod_{i \in S} \mathbb{P}(A_i)$.

k R.V.s $X_1, \dots, X_k$ are independent if for all $x_1, \dots, x_k$ the events $X_i = x_i$ are independent

- Also called *mutual* independence.
- Weaker condition: pairwise independence (above holds for all $|S| \leq 2$).

Can you think of 3 events or 3 RVs that are pairwise independent but not mutually indep?

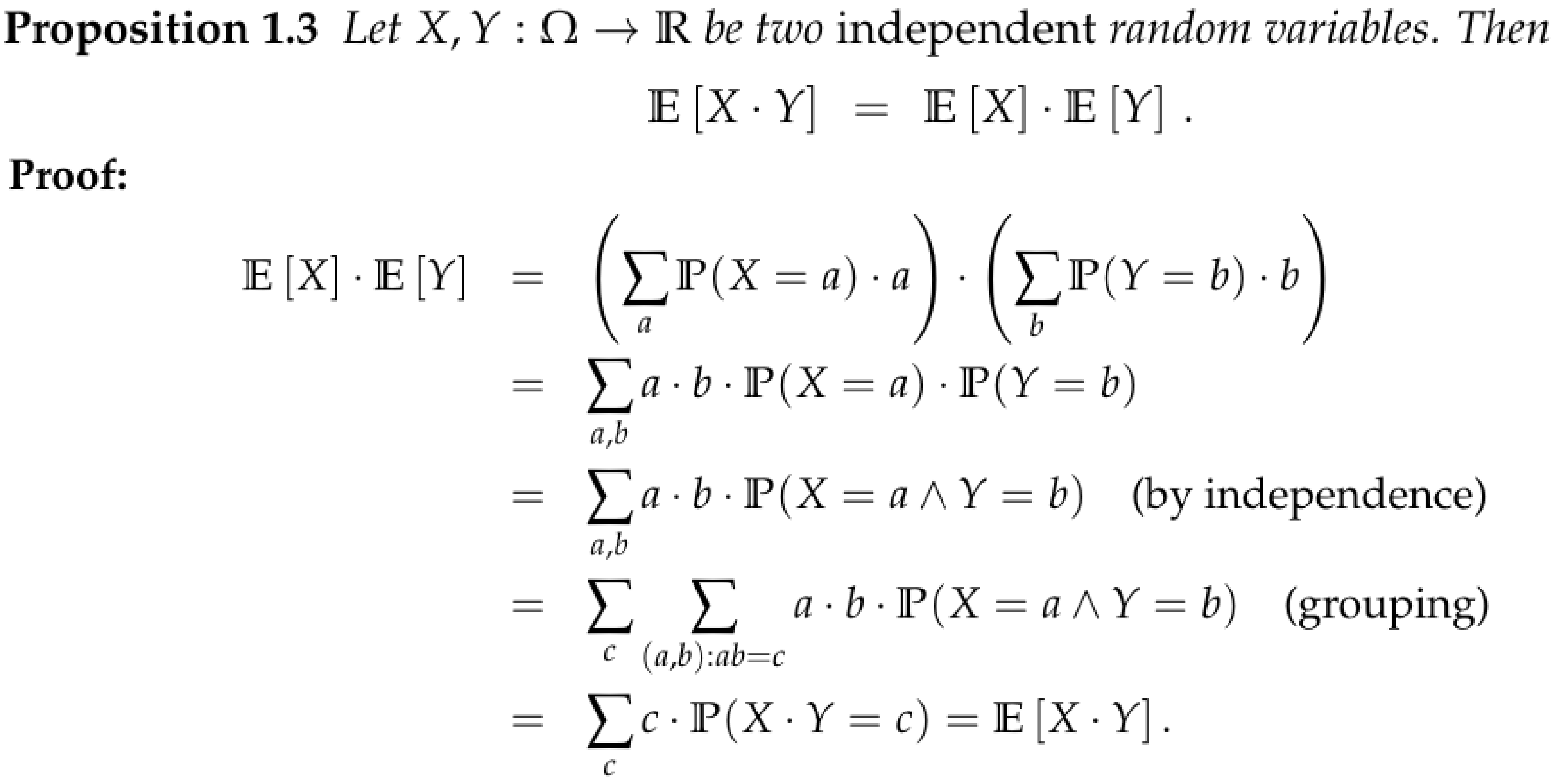
Independence

We saw that for any two R.V.'s X and Y, we have $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, regardless of whether X and Y are independent.

Can you think of two R.V.'s X and Y where $\mathbb{E}[X \cdot Y] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y]$?

But we do get $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$ when X and Y are independent.

Independence



Proposition 1.3 *Let* $X, Y : \Omega \to \mathbb{R}$ *be two* independent *random variables. Then*

$$\mathbb{E}\left[X \cdot Y\right] \;=\; \mathbb{E}\left[X\right] \cdot \mathbb{E}\left[Y\right] .$$

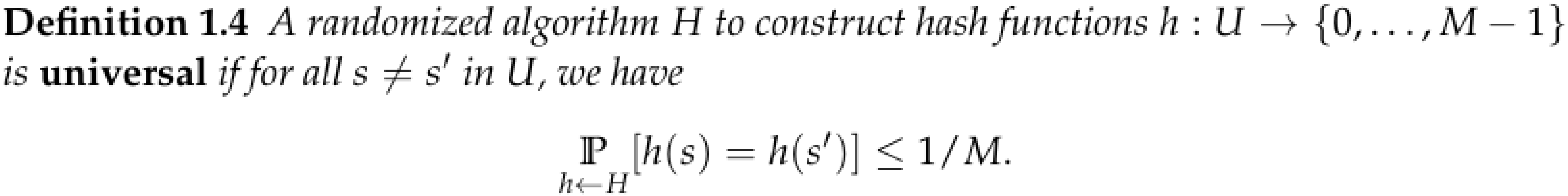
Proof:

$$\begin{aligned}
\mathbb{E}\left[X\right] \cdot \mathbb{E}\left[Y\right] &= \left(\sum_a \mathbb{P}(X = a) \cdot a\right) \cdot \left(\sum_b \mathbb{P}(Y = b) \cdot b\right) \\
&= \sum_{a,b} a \cdot b \cdot \mathbb{P}(X = a) \cdot \mathbb{P}(Y = b) \\
&= \sum_{a,b} a \cdot b \cdot \mathbb{P}(X = a \wedge Y = b) \quad \text{(by independence)} \\
&= \sum_c \sum_{(a,b):ab=c} a \cdot b \cdot \mathbb{P}(X = a \wedge Y = b) \quad \text{(grouping)} \\
&= \sum_c c \cdot \mathbb{P}(X \cdot Y = c) = \mathbb{E}\left[X \cdot Y\right].
\end{aligned}$$

Universal hashing

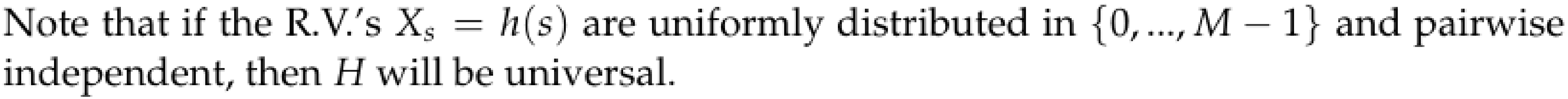
- A hash function is a function $h: U \to \{0, \ldots, M-1\}$ where U is an input space of size typically much larger than M. E.g., hashing strings to the range 1,…,10000.
- One property you want is that for the subset $S \subseteq U$ of inputs you actually care about (e.g., English words) you don't get too many collisions, especially when $|S| \approx M$.
- A convenient way to construct such a function is using randomization. (Randomization in the choice of h. The function h itself is deterministic.)
  - For all $s \in U, X_s = h(s)$ is a random variable
  - Asking for these to be mutually independent would be too much (h would essentially have to be a huge lookup table).
  - But *pairwise independence* will be sufficient, and implementable with simple h's.

Universal hashing

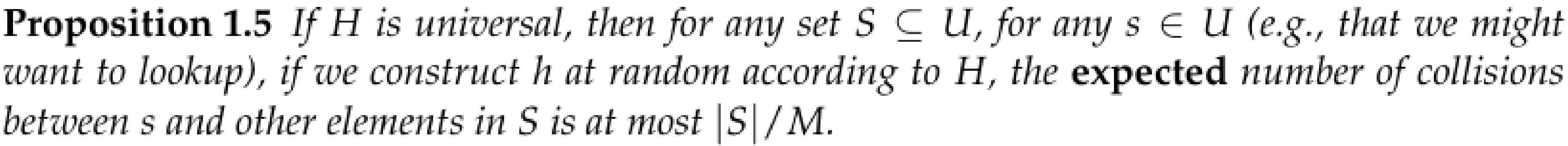


Definition 1.4 *A randomized algorithm H to construct hash functions $h : U \to \{0, \ldots, M-1\}$ is* **universal** *if for all $s \neq s'$ in U, we have*

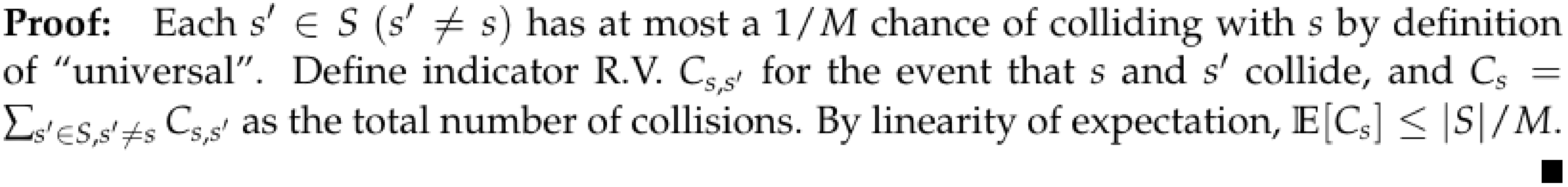
$$\mathop{\mathbb{P}}_{h \leftarrow H}[h(s) = h(s')] \leq 1/M.$$



Note that if the R.V.'s $X_s = h(s)$ are uniformly distributed in $\{0, ..., M-1\}$ and pairwise independent, then H will be universal.



Proposition 1.5 *If H is universal, then for any set $S \subseteq U$, for any $s \in U$ (e.g., that we might want to lookup), if we construct h at random according to H, the* **expected** *number of collisions between s and other elements in S is at most $|S|/M$.*

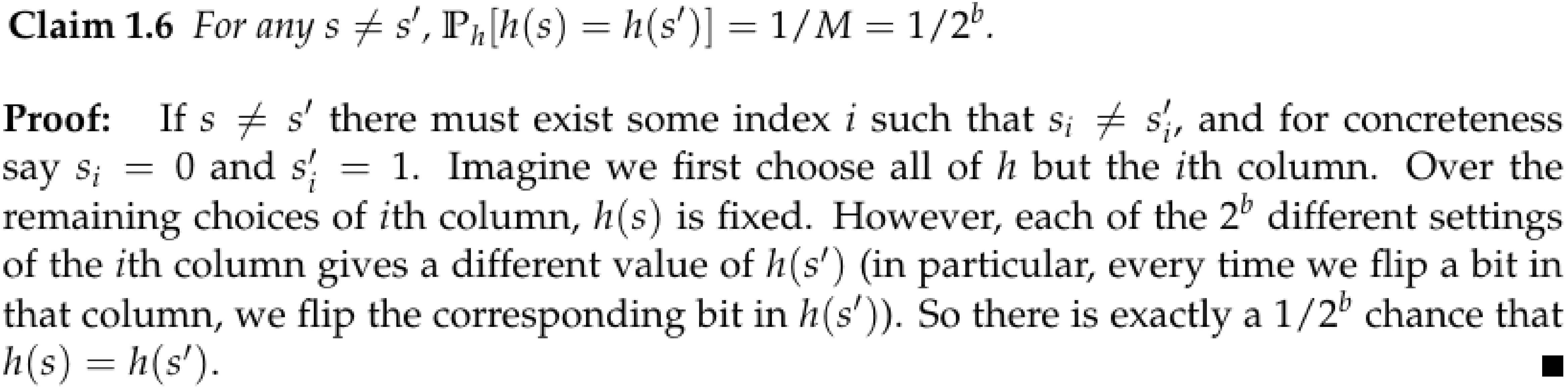


Proof: Each $s' \in S$ $(s' \neq s)$ has at most a $1/M$ chance of colliding with s by definition of "universal". Define indicator R.V. $C_{s,s'}$ for the event that s and s' collide, and $C_s = \sum_{s' \in S, s' \neq s} C_{s,s'}$ as the total number of collisions. By linearity of expectation, $\mathbb{E}[C_s] \leq |S|/M$. ■

Constructing a universal hash function

One approach:

- Say inputs are u bits long ($|U| = 2^u$), table size $M = 2^b$.
- Choose h to be a random linear transformation from $\mathbb{F}_2^u$ to $\mathbb{F}_2^b$ (i.e., a random $b \times u$ matrix over $\mathbb{F}_2$).



Claim 1.6 *For any* $s \neq s'$, $\mathbb{P}_h[h(s) = h(s')] = 1/M = 1/2^b$.

Proof: If $s \neq s'$ there must exist some index i such that $s_i \neq s'_i$, and for concreteness say $s_i = 0$ and $s'_i = 1$. Imagine we first choose all of h but the ith column. Over the remaining choices of ith column, $h(s)$ is fixed. However, each of the 2^b different settings of the ith column gives a different value of $h(s')$ (in particular, every time we flip a bit in that column, we flip the corresponding bit in $h(s')$). So there is exactly a $1/2^b$ chance that $h(s) = h(s')$. ■

Bernoulli and Binomial Random Variables

A Bernoulli(p) R.V. takes value 1 with probability p, and 0 with probability 1-p.

(Will also use the common terminology of “tossing a coin of bias p”.)

Let $X_1, \ldots, X_n$ be n independent (iid) Bernoulli(p) random variables and let $Z_n = X_1 + \cdots + X_n$. Z_n is called a Binomial(n,p) random variable.

- ➢ $\mathbb{E}[Z_n] = pn.$
- ➢ $\mathbb{P}[Z_n = k] = \binom{n}{k} p^k (1-p)^{n-k}.$

Infinite Bernoulli sequence and Geometric R.V.s

Consider an infinite sequence of iid Bernoulli(p) random variables $X_1, X_2, X_3, \ldots.$

Let Y be the index of the first $X_i = 1$ (i.e., the number of coin tosses until the first heads).

Then Y is a Geometric(p) R.V., with $\mathbb{P}(Y = i) = p \cdot (1-p)^{i-1}$.

To calculate $\mathbb{E}[Y]$, can use: $\mathbb{E}[Y] = \mathbb{E}[Y \,| X_1 = 1] \cdot \mathbb{P}[X_1 = 1] + \mathbb{E}[Y|X_1 = 0] \cdot \mathbb{P}[X_1 = 0]$.

- First term is $1 \cdot p = p$.
- Second term is $(1 + \mathbb{E}[Y]) \cdot (1-p)$.
- Overall, get $p \cdot \mathbb{E}[Y] = p + (1-p)$, so $\mathbb{E}[Y] = 1/p$.

Homework 3 is now available

Due Nov 6 (in 1 week)