Random problems with R

Kellie Ottoboni

University of California, Berkeley
Department of Statistics

BIDS
BERKELEY INSTITUTE
FOR DATA SCIENCE

```
> sample(x = 1:100, size = 1)
[1] 67
>
```

The simplest command in R…

right?

Multiply-and-floor method

to generate a random integer on $\{1, ..., m\}$

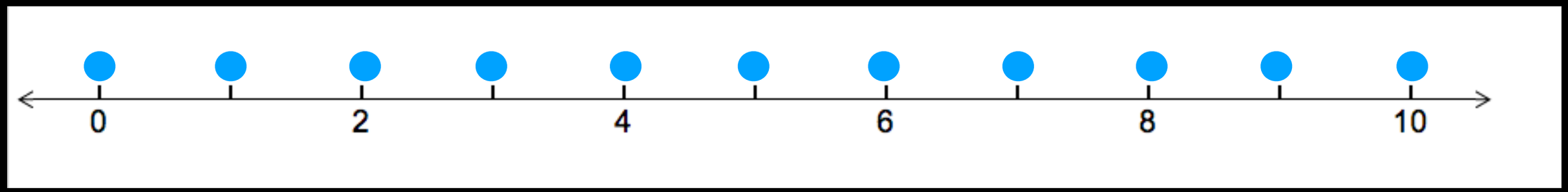
start with $X \sim U[0, 1)$and define $Y \equiv 1 + \lfloor mX \rfloor$

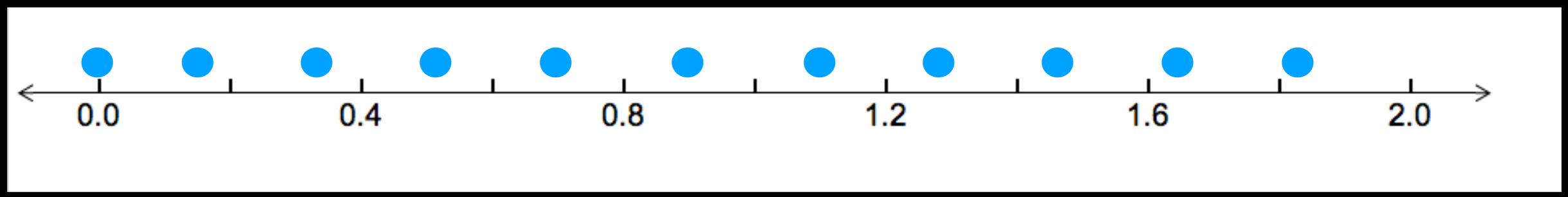
Multiply-and-floor method

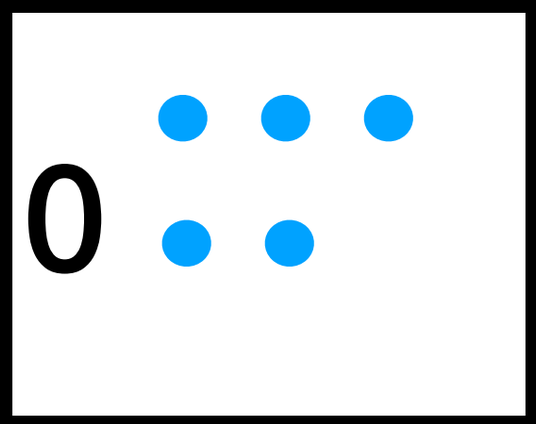
to generate a random integer on $\{1, ..., m\}$

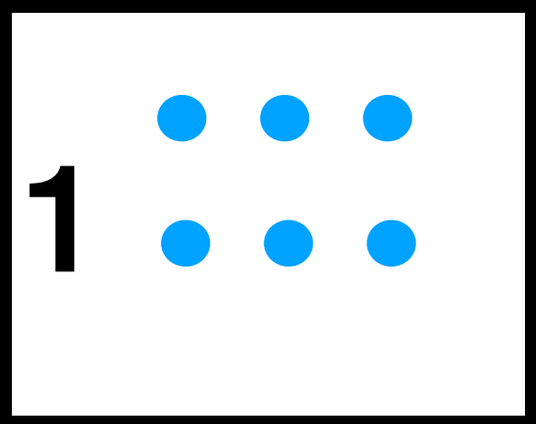
start with $X \sim U[0, 1)$ and define $Y \equiv 1 + \lfloor mX \rfloor$

Multiply-and-floor method

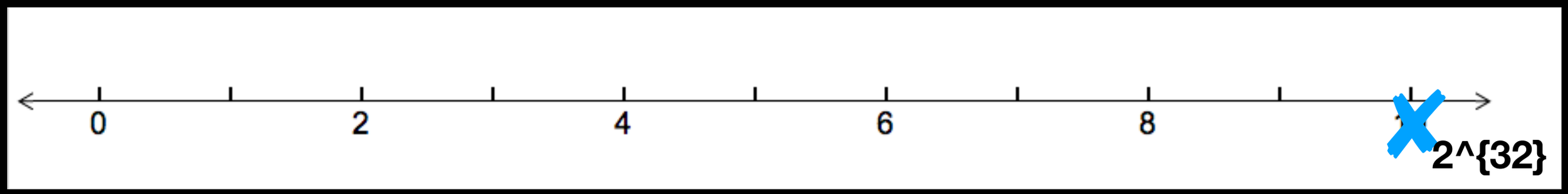








What R does is even more complicated.



Theorem

The sampling probabilities for {1,...,m} are LEAST uniform when m is just below 2^{31}

Some numbers are nearly twice as likely to be selected as others.

Evens and odds

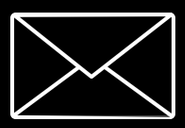
```
> m <- (2/5)*2^32
> m > 2^31
[1] FALSE
> x <- sample(m, 1000000, replace = TRUE)
> table(x %% 2)

     0      1
399827 600173
```

https://stat.ethz.ch/pipermail/r-devel/2018-September/076827.html

Should you worry?

https://arxiv.org/pdf/1809.06520.pdf

Thanks!



kellieotto@berkeley.edu

www.kellieottoboni.com

@kellieotto