Stat310

Discrete random variables

Hadley Wickham

1. Homework
2. Random variables
   1. Bernoulli
   2. Binomial
3. Mean and variance

Homework

- Model answers up on website.
- Homeworks back on Thursday.
- Homework help session: Tuesday DH3092, Wednesday MechLab 251.
- **New**: practice problems with worked solutions.

Random variables

Definitions

A **random variable** is a random experiment with a numeric sample space. Usually given a capital letter like X, Y or Z.

(More formally a random variable is a function that converts outcomes from a random experiment into numbers)

The **space** (or **support**) of a random variable is the range of the function (cf. sample space)

Definitions

If the size of the support is **finite** or **countably infinite**, then the random variable is **discrete**.

If the size of the support is **uncountably infinite**, then the random variable is **continuous**.

pmf/pdf

Every discrete random variable as **probability mass function** (**pmf**).

Every continuous random variable has **probability density function** (**pdf**).

Different ways of defining the function that says how **likely** each outcome is.

Now: **discrete**
Later: **continuous**

This diverges from the book, but I think it's easier to work with one set of mathematical tools at a time

Notation

Normally call pmf f

If we have multiple rv's and want to make clear which pmf belongs to which rv, we write:

$f_X(x)$ $f_Y(y)$ $f_Z(z)$ for X, Y, Z

$f_1(x)$ $f_2(x)$ $f_3(x)$ for X_1, X_2, X_3

But sometimes we are sloppy...

$$P(a < X < b) = \sum_{x_i \in (a,b)} f(x_i)$$

$$P(X = x) = f(x)$$

probability pmf

To be a pmf, f must satisfy two conditions:

$$\sum_{x_i \in S} f(x_i) = 1$$

$$f(x_i) \geq 0, \forall\ x_i \in S$$

x	f(x)
-1	0.3
0	0.3
2	0.3

x	f(x)
10	-0.1
20	0.9
30	0.2

x	f(x)
1	0.35
2	0.25
3	0.2
4	0.1
5	0.1

x	f(x)
5	1

x	f(x)
10	0.1
20	0.9
30	0.2

Which of these is a pmf?

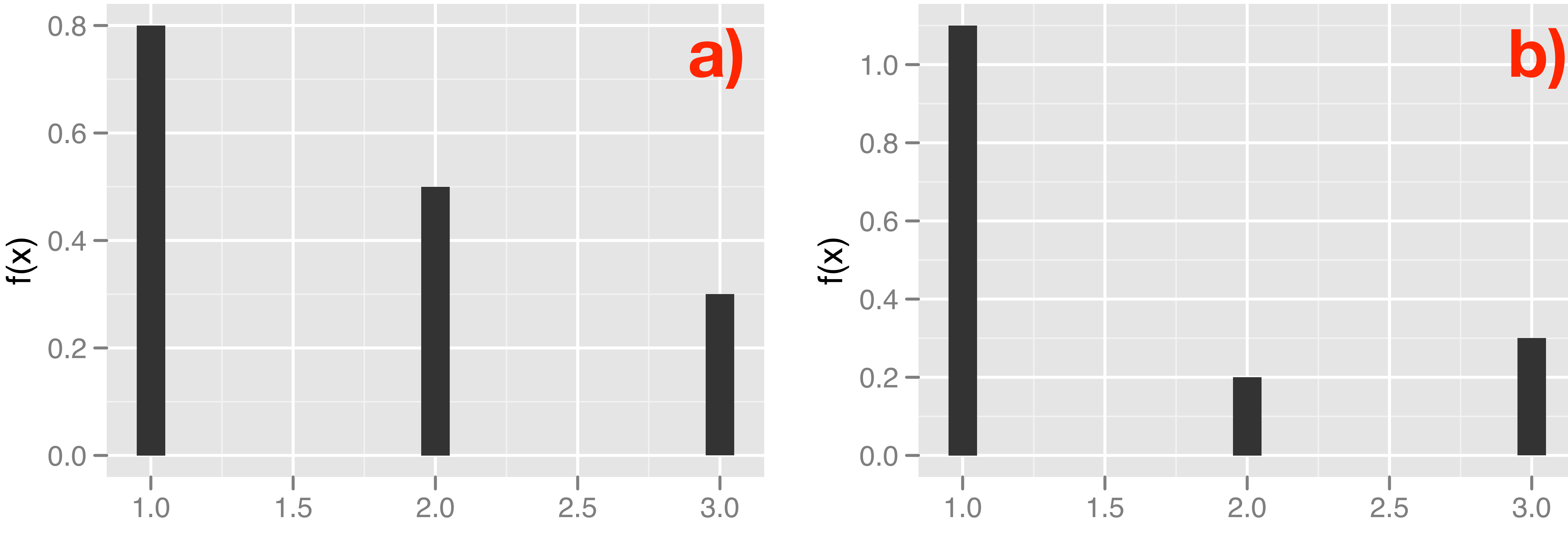
Notation

Can give pmf in two ways:

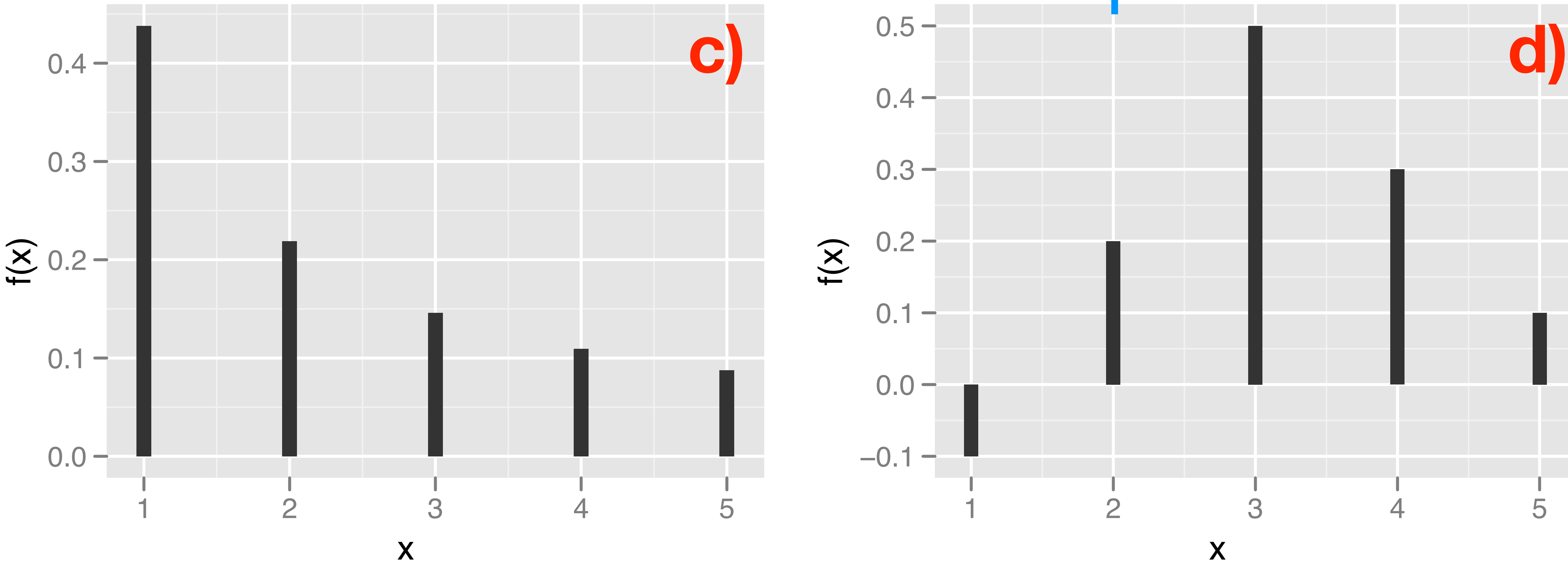
- List of numbers (for small n)
- Function (for large n)

These are equivalent!

Also useful to display visually.



Which of these is a pmf?



Distributions

In practice, many real problems can be solved with just a few different families of pmf/pdfs. These are called **distributions**.

A distribution has parameters which control how it acts. If a random variable has a named distribution, then we write it as:

X ~ DistributionName(parameters)

Bernoulli

Bernoulli distribution

Single binary event: either happens (with probability p) or doesn't happen.

Let X be a random variable that takes the value 1 if the event happens, 0 otherwise. Then **X ~ Bernoulli(p)**

$f(1) = P(X = 1) = p$

$f(0) = ?$

Bernoulli distribution

Single binary event: either happens (with probability p) or doesn't happen.

Let X be a random variable that takes the value 1 if the event happens, 0 otherwise. Then **X ~ Bernoulli(p)**

$f(1) = P(X = 1) = p$

$f(0) = ?$

Is that a pmf?

Binomial distribution

n **independent** Bernoulli trials with the same probability of success. Let X be the number of successes.

Then we say X ~ Binomial(n, p)

$P(X = x) = f(x) = ??$

Wait, is that a pmf?

Random mathematical fact.

Need to check the two conditions.

First easy, second a bit harder.

(If I ever give you a random mathematical fact you can expect to use it. Main challenge is recognising where it is needed)

Your turn

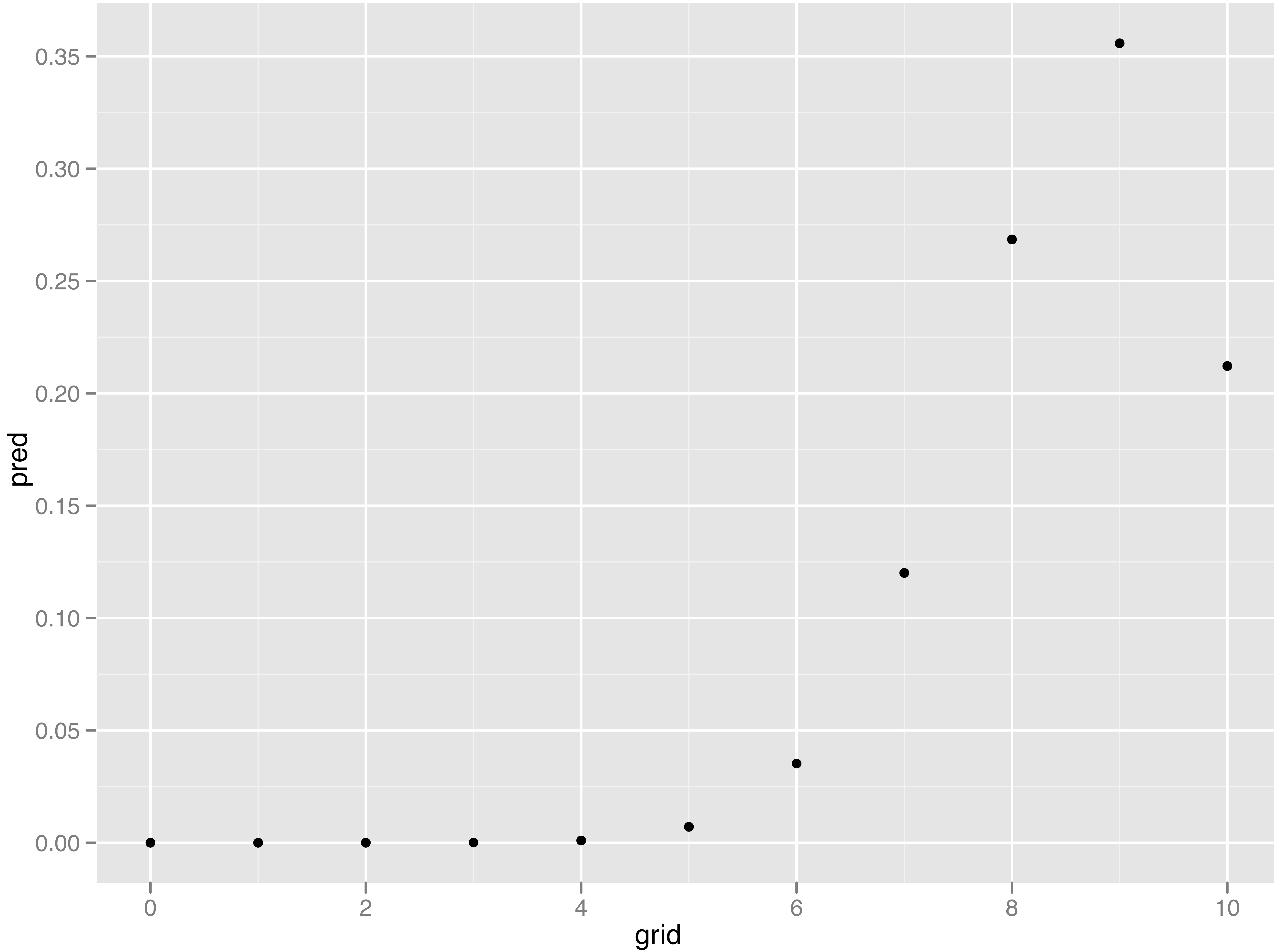
Complete the gaps on the first page of the proof sheet.

Your turn

In the 2008–09 season, Kobe Bryant attempted 564 field goals and made 483 of them.

In the first game of the next season he makes 10 attempts. What's the probability he doesn't make any? What's the probability he makes them all?

What assumptions did you make?



0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
pred
0
2
4
6
8
10
grid

Mean & Variance

Mean & variance

The **mean** summarises the "middle" of the distribution. The **variance** summarises the "spread" of the distribution.

Mean = E(X) = "Sum" of all outcomes, weighted by their probability.

Variance = Var(X) = $E[\ (X - E[X])^2)\ ]$ = expected squared distance from mean

Intuition for mean

Imagine the number line as a beam with weights of f(x) at position x. The balance point is the mean.

Example

Assume 95% of you have 0 stds. 4% of you have 1 std. 1% have 2 stds. What is the average number of stds per person?

http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5806a1.htm

Expectation

Expectation is a **linear operator**:

Expectation of a sum =
sum of expectations (additive)

Expectation of a constant * a function =
constant * expectation of function (homogenous)

Expectation of a constant is a constant.

T 2.6.2 p. 95

Your turn

Write (or recall) the mathematical description of these properties.

Work in pairs for two minutes.

Mean of a binomial random variable

For named distributions we can usually work out the mean (and variance) as functions of the parameters.

This is typically a little tricky, but once we've done it, we can use a simple formula every time we see that distribution.

Your turn

Complete the gaps on the second page of the proof sheet.

Another way

http://www.wolframalpha.com/input/?i=sum_(x%3D0)^(n)+(x+n!+/+(x!+(n-x)!)+p^x+(1-p)^(n-x))