UNIVERSITY OF LINCOLN

SCHOOL OF MATHEMATICS AND PHYSICS

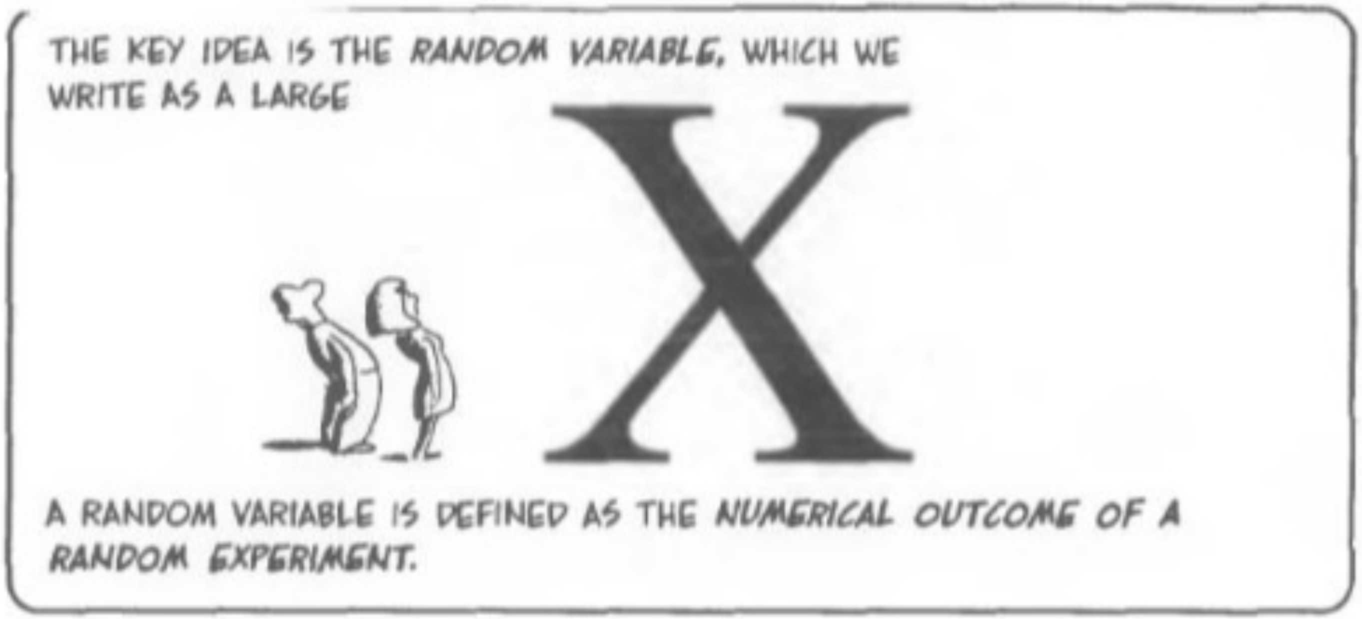
MTH1005
PROBABILITY AND STATISTICS

Semester B
Lecture 5
(27/2/2024)

Danilo Roccatano

Office: INB 3323
Email: droccatano@lincoln.ac.uk

SUMMARY OF THE LAST LECTURE



- Discrete random variables.
- Range and domain of the variables
- Probability (Mass/Density) Function
- Cumulative Distribution Function

CONTINUOUS RANDOM VARIABLES AND CONDITIONAL PROBABILITY

CONTINUOUS CONDITIONAL PROBABILITY

Last week, we introduced discrete probability distributions.

However similar considerations also apply to the case of continuous probabilities.

INFINITE, AND CONTINOUS PROBABILITY SPACES

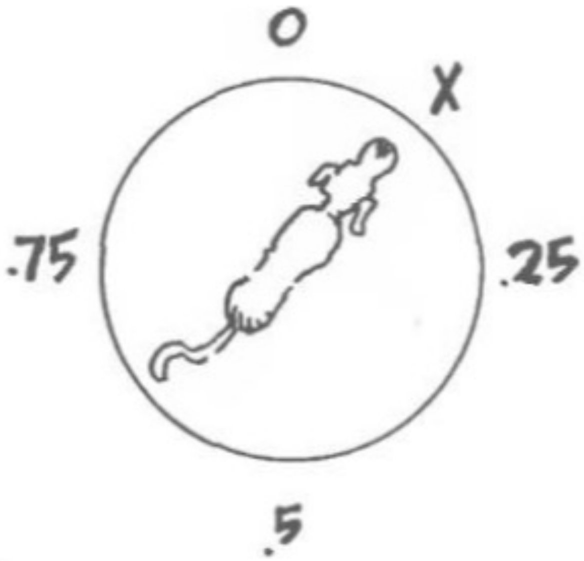
Everything up to this point deals with probability spaces with a finite sample space.

Here we mention two other cases

- when the sample space has discrete points, but an infinite number of them (all integers, for instance),
- or a prevailing situation when the sample space can take on a continuous range of values.

CONTINOUS PROBABILITY SPACES

If we give the spinner a whirl, the pointer will be pointing somewhere a distance x along the circumference.



It seems reasonable that every value *0≤x<1* of the distance between the pointer and the mark on the spinner is equally likely to occur. That means that the sample space is the interval S = [0, 1).

We would like to have a probability model for which every value of the sample space is equally likely. We'll call the result of a spin X for now, and later we'll see that this is a *continuous random variable*.

CONTINOUS PROBABILITY SPACES

Now, we need to characterize the probability of event E = *'the dog's nose pointing between points a and b.'* associated with the random variable X.

Since all points are equally likely, the probability is equivalent to the interval between these points along the circumference, expressed by the formula

$$P(a \leq X < b) = b - a$$

where this relationship holds true for any values of a and b within the event interval E=[a, b].

CONTINOUS PROBABILITY SPACES

The probability can alternatively be conceptualized by integrating a function, denoted as f(x), over the interval of interest E, represented by the integral

$$P(E) = \int_E f(x)dx$$

where *f(x) is* assigned constant value of 1.

This function f(x) is termed the density function of X.

This is the generalisation of the discrete case we saw earlier:

$$P(E) = \sum_{i \epsilon E} P(i)$$

CONDITIONAL CONTINUOUS PROBABILITIES

If we examine a stochastic process governed by a probability density function f(x), and denote E as an event within this process, we proceed to define a conditional density function by

$$f(x \mid E) = \begin{cases} f(x)/P(E) & \text{if } x \in E, \\ 0 & \text{if } x \notin E \end{cases}$$

Then for any event F , we have

$$P(F|E) = \int_F f(x|E)dx$$

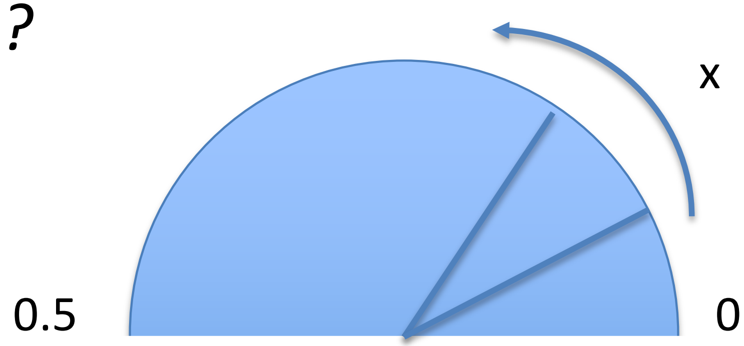
CONDITIONAL CONTINUOUS PROBABILITIES

We call this the conditional probability of F given E . A little manipulation makes the connection to the discrete case:

$$P(F \mid E) = \int_F f(x \mid E)\,\mathrm{d}x = \int_{E\cap F} \frac{f(x)}{P(E)}\mathrm{d}x$$

EXAMPLE OF CONDITIONAL CONTINUOUS PROBABILITY DISTRIBUTION

In the spinner experiment, suppose we know that the spinner has stopped with the head in the upper half of the circle, $0 \le x \le \frac{1}{2}$.
What is the probability that $\frac{1}{6} \le x \le \frac{1}{3}$*?*



Here
E= {[0, 1/2]},
F = {[1/6, 1/3]}

Also, we note that F∩ E = F

Hence

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

which is reasonable, since F is 1/3 the size of E .

EXAMPLE OF CONDITIONAL CONTINUOUS PROBABILITY DISTRIBUTION

The conditional density function here is given by

$$f(x \mid E) = \begin{cases} 2, & \text{if } 0 \leq x < 1/2, \\ 0, & \text{if } 1/2 \leq x < 1. \end{cases}$$

Thus, the conditional density function is non-zero only on [0, 1/2] , and is uniform there.

SUMMARY: PROBABILITY FOR CONTINOUS PROBABILITY SPACES

We introduced the definition of <u>continuous probability</u> as a generalisation of the discrete case :

$$P(E) = \sum_{i \epsilon E} P(i)$$

Using the integral instead of the summation

$$P(E) = \int_E f(x) dx$$

And the function *f(x)* called the *probability density function* of the random variable X.

SUMMARY: CUMULATIVE DISTRIBUTION FUNCTION FOR CONTINUOUS RANDOM VARIABLES

The definition of CDF for a continuous random variable is very similar to the discrete case, but the sum is replaced by the integral, and the integrand is called the (probability) density function.

Cumulative distribution function for a continuous random variable, X

$$F_X(a) = P\{X \leq a\} = \int_{-\infty}^{a} f(x)\mathrm{d}x,$$

note that conversely <u>the fundamental theorem of calculus</u>

$$\frac{d}{da}F(a) = f(a)$$

$F_X(a)$ must be 0 as $a \to -\infty$ and 1 as $a \to \infty$

$F_X(a)$ must be monotonically increasing

$F_X(a)$ $\quad$ *must be right continuous*

when these conditions hold, F_X defines a random variable X .

EXAMPLE

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1/4(y^2 + 3y) & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y > 1 \end{cases}$$

this function fulfils all our requirements:

- as $y \to -\infty$ $we\ see\ that\ F_Y(y) \to 0$
- as $y \to \infty$ $we\ see\ that\ F_Y(y) \to 1$
- our function is right continuous - the central section is continuous (polynomial), and $F_y(0)$ and $F_y(1)$ are the same when approached from either side.

PROBABILITY DENSITY FUNCTION

We can now show that the function

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y}$$

is consistent with our previous short discussions of continuous density functions.

Being y a continuous random variable, instead of the probability mass function we have a probability density function *f(x)* which when integrated over a valid range of *X* values gives the probability that the random variable X lies in that range:

$$P\{a \leq X \leq b\} = \int_a^b f_X(x)\mathrm{d}x = F_X(b) - F_X(a)$$

the last follows from **the fundamental theorem of calculus**.

PROBABILITY DENSITY FUNCTION

By definition

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \lim_{\Delta y \to 0} \frac{F_Y(y+\Delta y) - F_Y(y)}{\Delta y}$$

being a little loose with what should be done with such limits we have

$$\begin{aligned} F_Y(y+\Delta y) - F_Y(y) &= f_Y(y)\Delta y \\ &= P\{y \leq Y \leq y+\Delta y\} \end{aligned}$$

so $f_Y(y)\Delta y$ is the probability that *y* is in the small range [y ,$y + \Delta y$].

Because we multiply $f_Y(y)\Delta$y by a value Δy to get something sensible, it gives a naive explanation of why it is referred to as a density function.

EXAMPLE

Let the probability density function of a random variable X be

$$f(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{2}{5}(4-x) & \text{if } 1 \leq x \leq 2 \\ 0 & \text{if } x > 2 \end{cases}$$

Suppose we want the probability that X is in the interval. From our definition we have

$$P\{\frac{5}{4} \leq X \leq \frac{7}{4}\} = \int_{\frac{5}{4}}^{\frac{7}{4}} \frac{2}{5}(4-x)\mathrm{d}x = \frac{1}{2}.$$

We also note that the total probability, integrating over the non-zero part of *f(x)*, is unity

$$P\{-\infty \leq X \leq \infty\} = \int_{1}^{2} \frac{2}{5}(4-x)\mathrm{d}x = 1.$$

This means that our probability density function make sense and obey the axioms of probability.

TECHNICAL ASIDE: IMPROPER INTEGRALS

In the last example we had a integral

$$P\{-\infty \leq X \leq \infty\} = \int_{-\infty}^{\infty} f(x)\mathrm{d}x$$

This type of integral are so-called **improper integrals** that we will encounter typically have infinite integration limits.
Similarly, to in calculus we should interpret these as a limiting process.
Typically, we will

- split up the integrals into a proper integral, and an improper integral.
- Then show that the improper integral has a well-defined limit - which we take as its value.

TECHNICAL ASIDE: IMPROPER INTEGRALS

EXAMPLE

$$f(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{2}{5}(4-x) & \text{if } 1 \leq x \leq 2 \\ 0 & \text{if } x > 2 \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-\infty}^{1} 0\mathrm{d}x + \int_{1}^{2} \frac{2}{5}(4-x)\mathrm{d}x + \int_{2}^{\infty} 0\mathrm{d}x$$

and we should take $\int_{-\infty}^{1} 0\mathrm{d}x$ to mean $\lim_{a\to-\infty} \int_{a}^{1} 0\mathrm{d}x$. As we take a to be a larger and larger negative number, the value of the integral remains 0, so this is its value.

For most/all functions we will encounter, e.g. $\int_{-\infty}^{\infty} e^{-x^2}$ a similar procedure yields sensible results.

Example - Uniform random variables

Let X be a random variable whose (cumulative) distribution function is

$$F_X(t) = \begin{cases} 0 & t < 0 \\ t & 0 \le t \le 1 \\ 1 & t > 1 \end{cases}$$

our (probability) density function is the derivative with respect to t of $F_X(t)$

$$f_X(t) = \frac{\mathrm{d}F_X(t)}{\mathrm{d}t} = \begin{cases} 0 & t < 0 \\ 1 & 0 \le t \le 1 \\ 0 & t > 1 \end{cases}$$

Since the density function of is equal to 1, the area under the density over any interval between and is equal to the length of the interval.

Example - Uniform random variables

Because we have that

$$P\{a \leq X \leq b\} = \int_a^b f_X(x)\mathrm{d}x$$

this means that the probability that the observed value for X lies in an interval contained in (0, 1) is proportional to the length of the interval.

Example - Uniform random variables

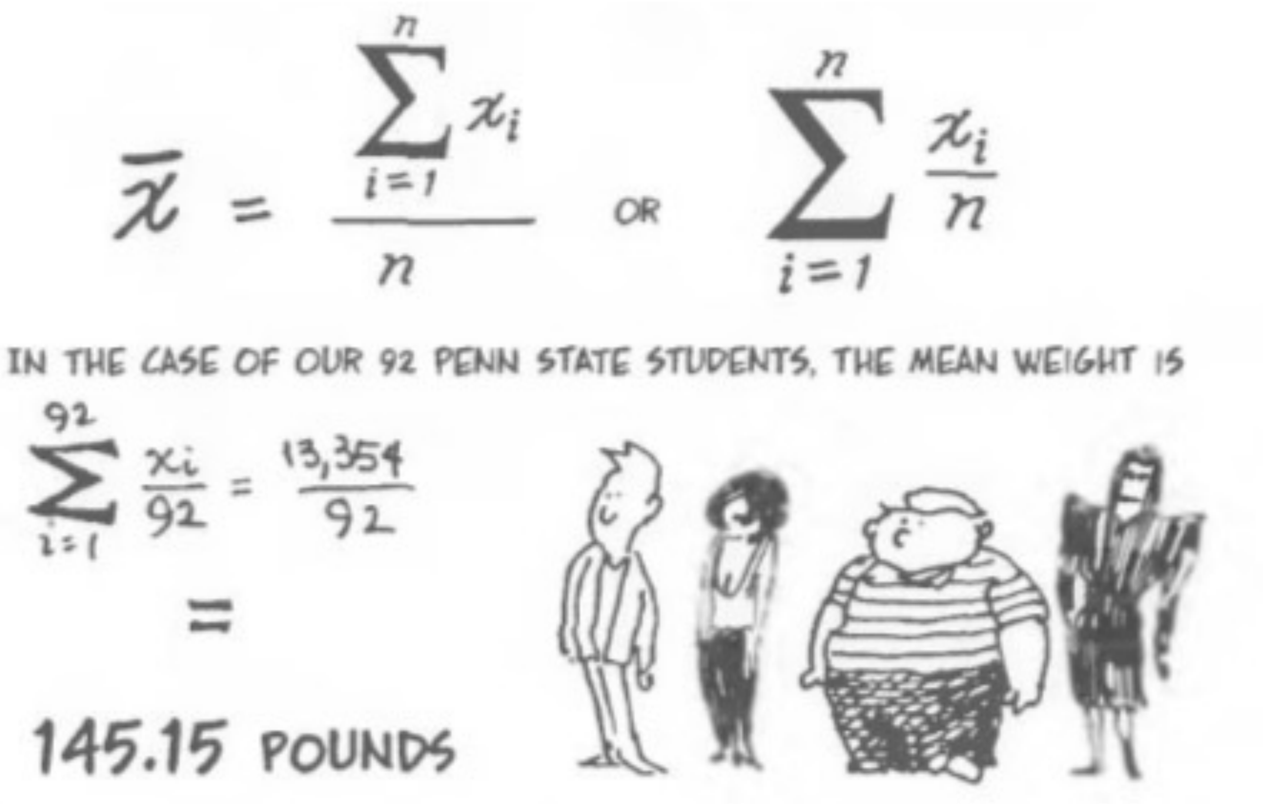
To be a valid probability density function of a random variable, we must have

$$f(t) \geq 0, \text{ for all } t$$
$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1$$

Properties of the random variables

Properties of a discrete random variable

Mean or Expectation value



We will see that the means is also called **expectation value of the random variable** and it is indicated ad E[X}

From: Gonick & Smith The cartoon guide to Statistics

Properties of a discrete random variable

Median

FOR THE $n = 92$ STUDENT WEIGHTS, WE CAN FIND THE MEDIAN FROM THE ORDERED STEM-AND-LEAF DIAGRAM: JUST COUNT TO THE 46TH OBSERVATION. THE MEDIAN IS

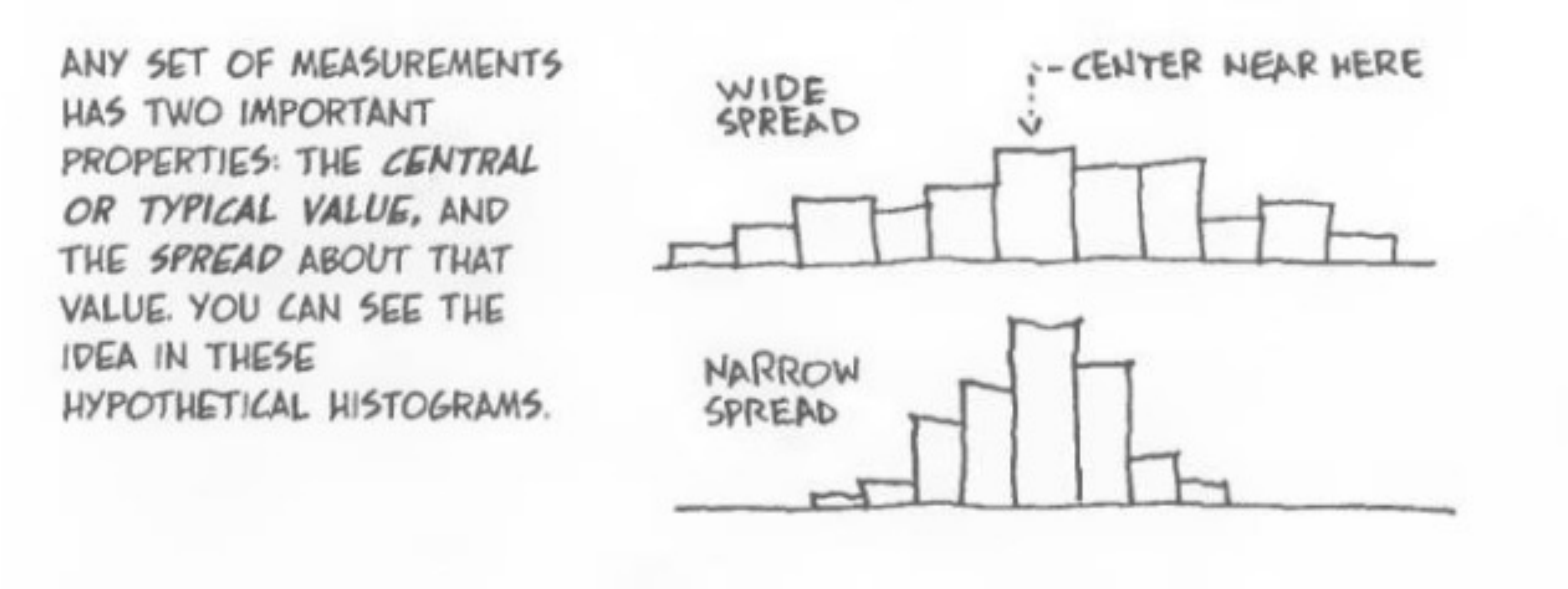
$$\frac{x_{46} + x_{47}}{2} = \frac{145 + 145}{2}$$

$$= 145 \text{ POUNDS}$$

9	5
10	288
11	002556688
12	00012355555
13	0000013555688
14	0000255555**55**8
15	0000000000035555555557
16	000045
17	000055
18	0005
19	00005
20	
21	5

Properties of a discrete random variable

Spread



We are going to learn how to measure these properties and how to predict that the data follow a certain trend in the distribution.

From: Gonick & Smith The cartoon guide to Statistics

EXPECTATION OF A RANDOM VARIABLE

Now we will define the expectation and variance of a random variable. This is a further step in connecting abstract probability distributions to statistical measures and the real world, as needed.

By the end of this part, you should understand:

- How to calculate the expectation value of a random variable (this measures the central tendency of a distribution).
- How to calculate the variance of a random variable (this measures the spread or dispersion of a distribution).
- How to calculate expectation values of a function of a random variable.

Later, we'll study specific probability distributions, and understanding their expectation values will enable us to model and predict the outcomes of measurements taken on objects that can be represented as random variables.

EXPECTATION OF A RANDOM VARIABLE

Definition

Given a random variable X with a probability mass function p(x) we define the expectation of *X* , written as *E[X]* as

$$\mathrm{E}[X] = \sum_{x \in X} x \cdot p(x)$$

We also call the mean of and write it as μ_X

EXAMPLE - DISCRETE RANDOM VARIABLE

Let’s come back to the simulation experiment that we have done in the lab

In a game, 3 dice are rolled. The players bet £1. They gets back £1 if they rolls a single 5, £2 if 2 fives come up, and £3 if 3 fives come up (and his stake is returned).

If no 5s come up he lose their £1 stake.

EXAMPLE - DISCRETE RANDOM VARIABLE

Let us set up a random variable V for the winnings of the player.

The sample space is the cartesian product of rolling a die:

$$S = \{(x_1, x_2, x_3) : x_i = 1, 2, \ldots, 6; i = 1, 2, 3\}$$

Now, we'll define V as the amount the player wins - this can be $-1, 1, 2, 3$ corresponding to 0,1,2, or 3 fives showing:

$$V = \begin{cases} -1 & \text{if '0 dice lands with 5 spots on the top face'} \\ 1 & \text{if '1 dice lands with 5 spots on the top face'} \\ 2 & \text{if '2 dice land with 5 spots on the top face'} \\ 3 & \text{if '3 dice land with 5 spots on the top face'} \end{cases}$$

and our probability mass function will be

$$p_V(v) = \begin{cases} \frac{125}{216} & \text{for } v = -1 \\ \frac{75}{216} & \text{for } v = 1 \\ \frac{15}{216} & \text{for } v = 2 \\ \frac{1}{216} & \text{for } v = 3 \end{cases}$$

note that $p_V(0) > 0.5$.

EXAMPLE - DISCRETE RANDOM VARIABLE

Now, suppose players play the game *n>>1* times.

They win v_1 pounds the first time, v_2 the second time, $\ldots, v_n$ pounds the n^{th} time. The average amount one would then be a standard average

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

Now each of the v_i can only be $-1, 1, 2, 3.$

Lets reorganise our results and say that k_{-1} times they got $v_i = -1$, k_1 times $v_i = 1$, k_2 times $v_i = 2$ and k_3 times $v_i = 3$. Where $k_{-1} + k_1 + k_2 + k_3$ will be equal to n because we are just placing our n values into these boxes, then

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = (-1)\frac{k_{-1}}{n} + (1)\frac{k_1}{n} + (2)\frac{k_2}{n} + (3)\frac{k_3}{n}$$

EXAMPLE - DISCRETE RANDOM VARIABLE

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = (-1)\frac{k_{-1}}{n} + (1)\frac{k_1}{n} + (2)\frac{k_2}{n} + (3)\frac{k_3}{n}$$

is our average value - but as $n \to \infty$ the ratios $\frac{k_i}{n}$ tend to our original frequentist definition of probabilities - the number of times something occurs out of the total number of attempts.

$$\frac{k_{-1}}{n} = p_V(-1), \frac{k_1}{n} = p_V(1), \frac{k_2}{n} = p_V(2), \frac{k_3}{n} = p_V(3)$$

and finally we get

$$\bar{v} = \mu_V = \sum_{i=1}^{n} v_i p_V(i) = \sum_{v \in R_V} v p_V(v)$$

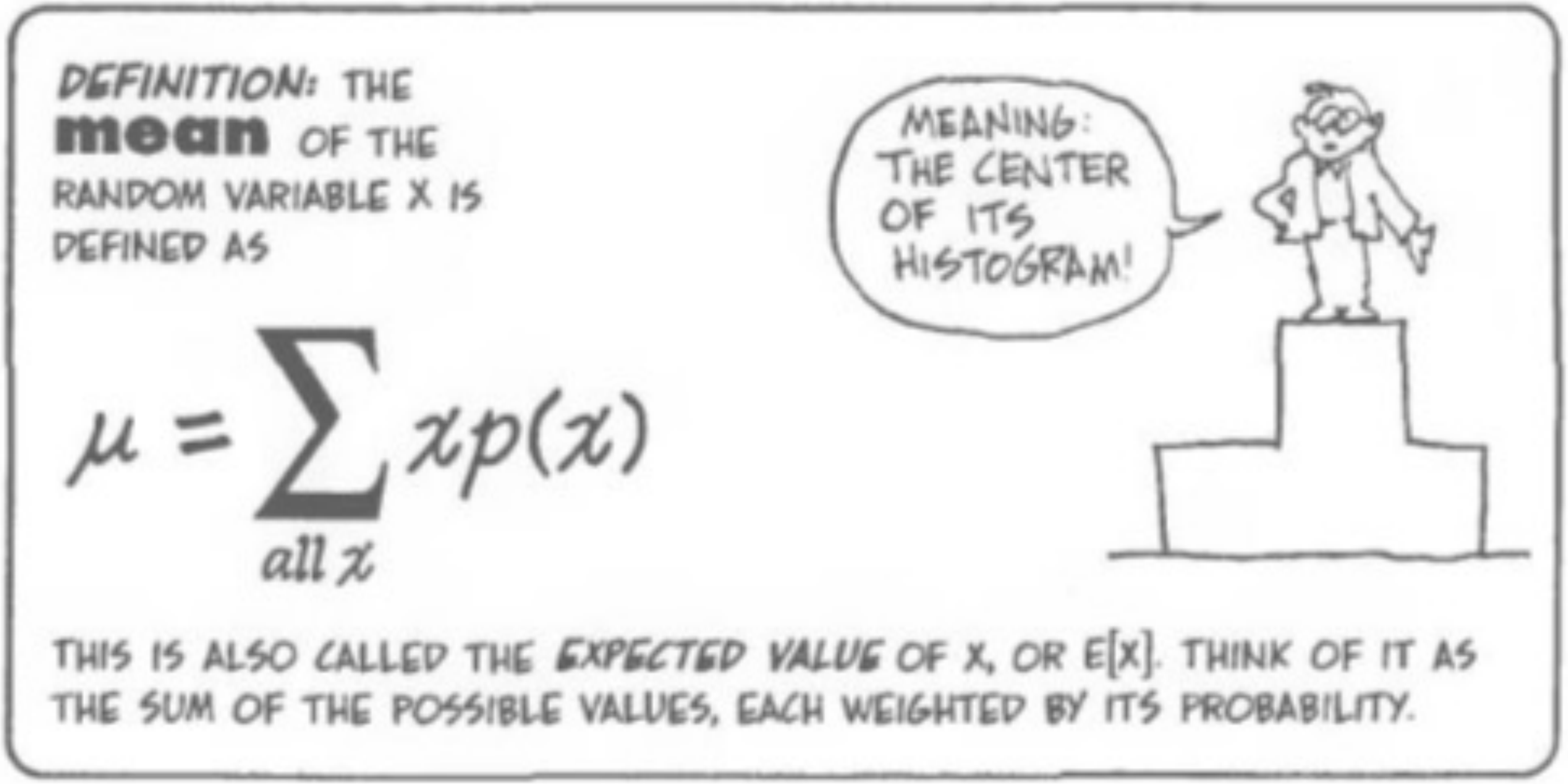
EXAMPLE - DISCRETE RANDOM VARIABLE

in this case

$$\mathrm{E} = \mu_V = \sum_{v \in R_V} v p_V(v) = (-1)\frac{125}{216} + (1)\frac{75}{216} + (2)\frac{15}{216} + (3)\frac{1}{216} = \frac{-17}{216} = -0.08$$

On average the player loses 8p every time they play.

SUMMARY



From: Gonick & Smith The cartoon guide to Statistics

EXPECTATION OF A FUNCTION OF A DISCRETE RANDOM VARIABLE

Definitions

Let $g(X)$ be any function of a discrete random variable X. Then

$$\mathrm{E}[g(X)] = \sum_{x \in X} g(x) \cdot p(x)$$

these can be understood in the same way as *E[X]* , it is the value of *g(X)* times the probability of getting that value **X =** x, and hence *g(x)*.

The derivation using repetitions of experiments can be repeated, but replacing vi with $g(v_i)$.

EXPECTATION OF A FUNCTION OF A DISCRETE RANDOM VARIABLE

In a game the player wins back a sum that is the 'square of the number of heads' on two coins.

We have a new random variable Y with probability mass function $p(0) = \frac{1}{4}, p(1) = \frac{1}{2}, p(4) = \frac{1}{4}$ then the expectation of Y is given by

$$\mathrm{E}[Y] = \sum_{\text{all } y} y \cdot p(y) =$$

$$0 \cdot p(0) + 1 \cdot p(1) + 4 \cdot p(4) =$$

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{3}{2} = \mu_Y$$

note that $Y = X^2$

$$E[X^2] = \sum_{all\ y} x^2 p(x^2)$$

$$0 \cdot p(0) + 1 \cdot p(1) + 2^2 \cdot p(2) =$$

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{3}{2} = \mu_{X^2}$$

EXPECTATION OF A FUNCTION OF A CONTINUOUS RANDOM VARIABLE

EXPECTATION OF A RANDOM VARIABLE

Definition

Given a random variable X with a probability mass function p(x) we define the expectation of *X* , written as *E[X]* as

$$\mathrm{E}[X] = \sum_{x \in X} x \cdot p(x)$$

The equivalent for a continuous random variable is

$$\mathrm{E}[Z] = \int_{-\infty}^{\infty} z \cdot f(z) \mathrm{d}z$$

We see that the relationship is very close between discrete and continuous cases - we replace a sum with an integral.

We also call the mean of and write it as μ_X

EXAMPLE - CONTINUOUS RANDOM VARIABLE

Suppose you meet that friend of yours that is always late. You, of course, arrive on time (T=0) and friend shows up at a random time T > 0 with probability density function

$$f_T(t) = \frac{2}{15} - \frac{2t}{225}, 0 < t < 15$$

We'll ignore other times, where $f_T(t) = 0$

We could break this interval up into a large number, n , of small pieces with length

$$\Delta t = 15/n$$

t_1 , t_2 , t_3 , ... , t_n be the midpoints of these small intervals.

EXAMPLE - CONTINUOUS RANDOM VARIABLE

The probability of your friend arriving between at time t_i is approximately

$$P\,(t_i - \Delta t/2 < T < t_i + \Delta t/2) \approx f_T(t_i)\Delta t\ .$$

We'd expect our average waiting time to be about $\sum_{i=1}^{n} t_i f_T(t_i)\Delta t$ - the sum of the product of length of wait, t_i, times the probability of waiting that long, $f_T(t_i)\Delta t$. The same as for the discrete random variable case.

If we now take *n* larger and large towards infinity, we get

$$\int_0^{15} t f_T(t)\mathrm{d}t = \int_0^{15} t\left(\frac{2}{15} - \frac{2t}{225}\right)\mathrm{d}t = 5\ (\text{minutes})$$

This is the length of time you expect to wait for your friend.

SOME PROPERTIES OF E[X}

E[a] = a, where **a** is a constant for any random variable.

This is a special case of

$$\mathrm{E}[g(X)] = \sum_{x \in X} g(x) \cdot p(x)$$

with *g(X)= a*

$$\begin{aligned} \mathrm{E}[a] &= \sum_{x \in R_X} g(x) \cdot p_X(x) \\ &= \sum_{x \in R_X} a \cdot p_X(x) \\ &= a \sum_{x \in R_X} p_X(x) \\ &= a \end{aligned}$$

because the sum of all the $p_X(x)$ must be 1 for a valid probability mass function.

EXPECTATION OF A FUNCTION OF A RANDOM VARIABLE

Let g(X) be any function of a random variable X . Then

$$\mathrm{E}[g(X)] = \sum_{\text{all } x} g(x) \cdot p(x)$$

or for the continuous random variable Z

$$\mathrm{E}[g(Z)] = \int_{\text{all } z} g(z) \cdot f(z) \mathrm{d}z$$

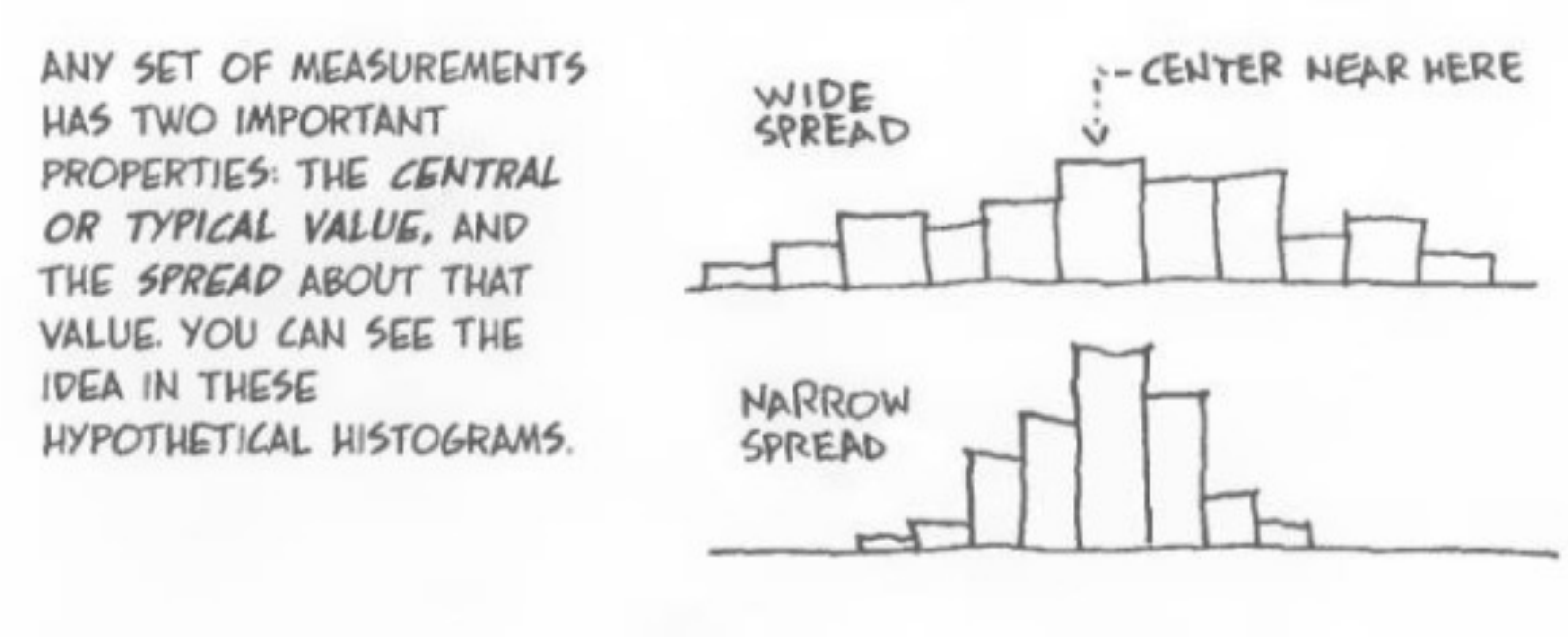
if X is a random variable, then

- $\mathrm{E}[a] = a$
- $\mathrm{E}[aX] = a\mathrm{E}[X]$
- $\mathrm{E}[g_1(X) + g_2(X)] = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)]$, where $g_1(X)$ and $g_2(X)$ are any functions of X.

these define the properties of a linear operator.

Properties of a discrete random variable

The next important property of random variables is the spread around the expectation value



From: Gonick & Smith The cartoon guide to Statistics

VARIANCE OF A RANDOM VARIABLE

The mean describes where a distribution is centred - expected value.

The **variance** describes how widely the values of the distribution are spread around the mean.

Definition

If X is a random variable with mean μ, then the variance of X, denoted by $\mathrm{Var}(X)$ or σ_X^2, is defined by

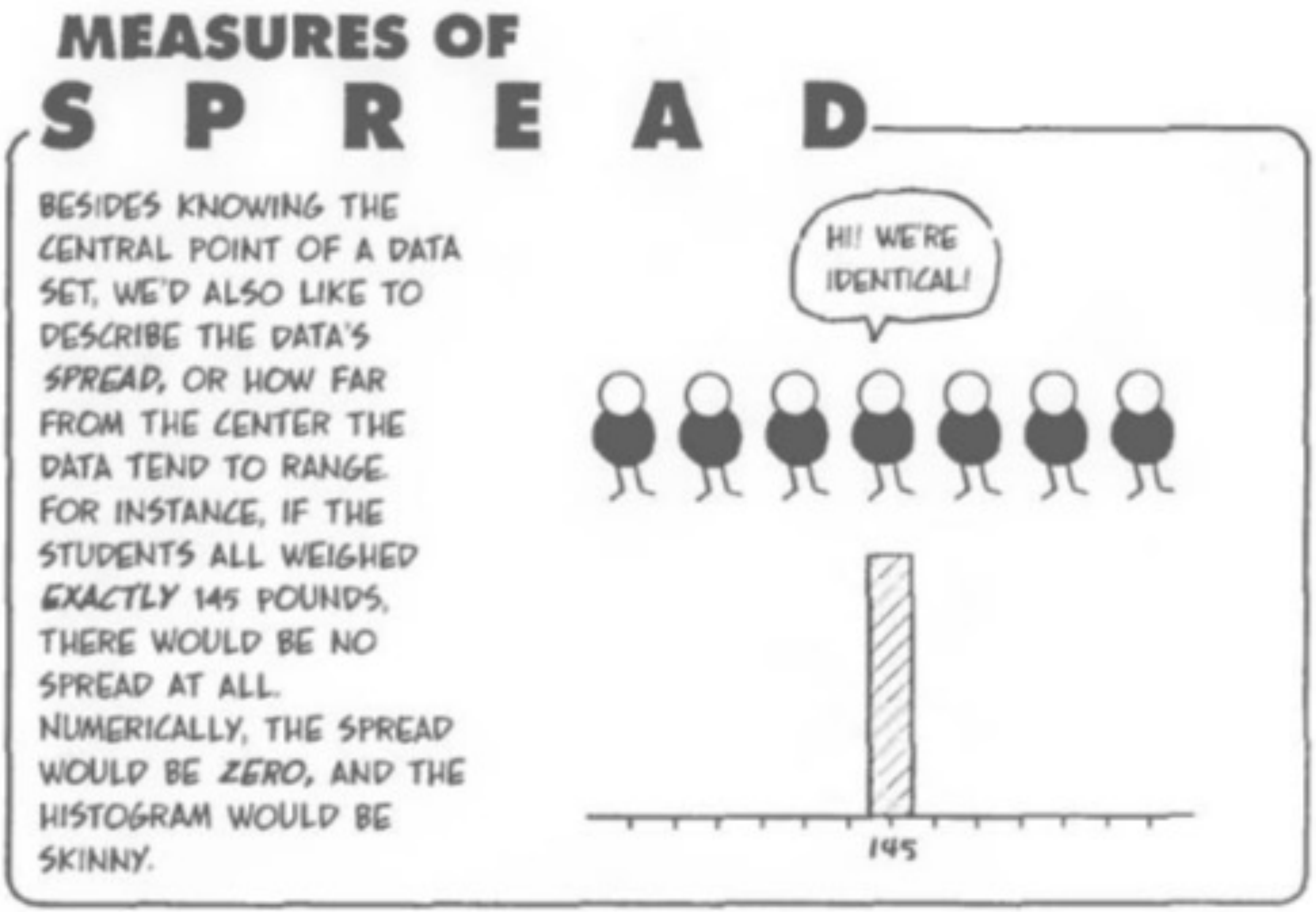
$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2]$$

this can also be written

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

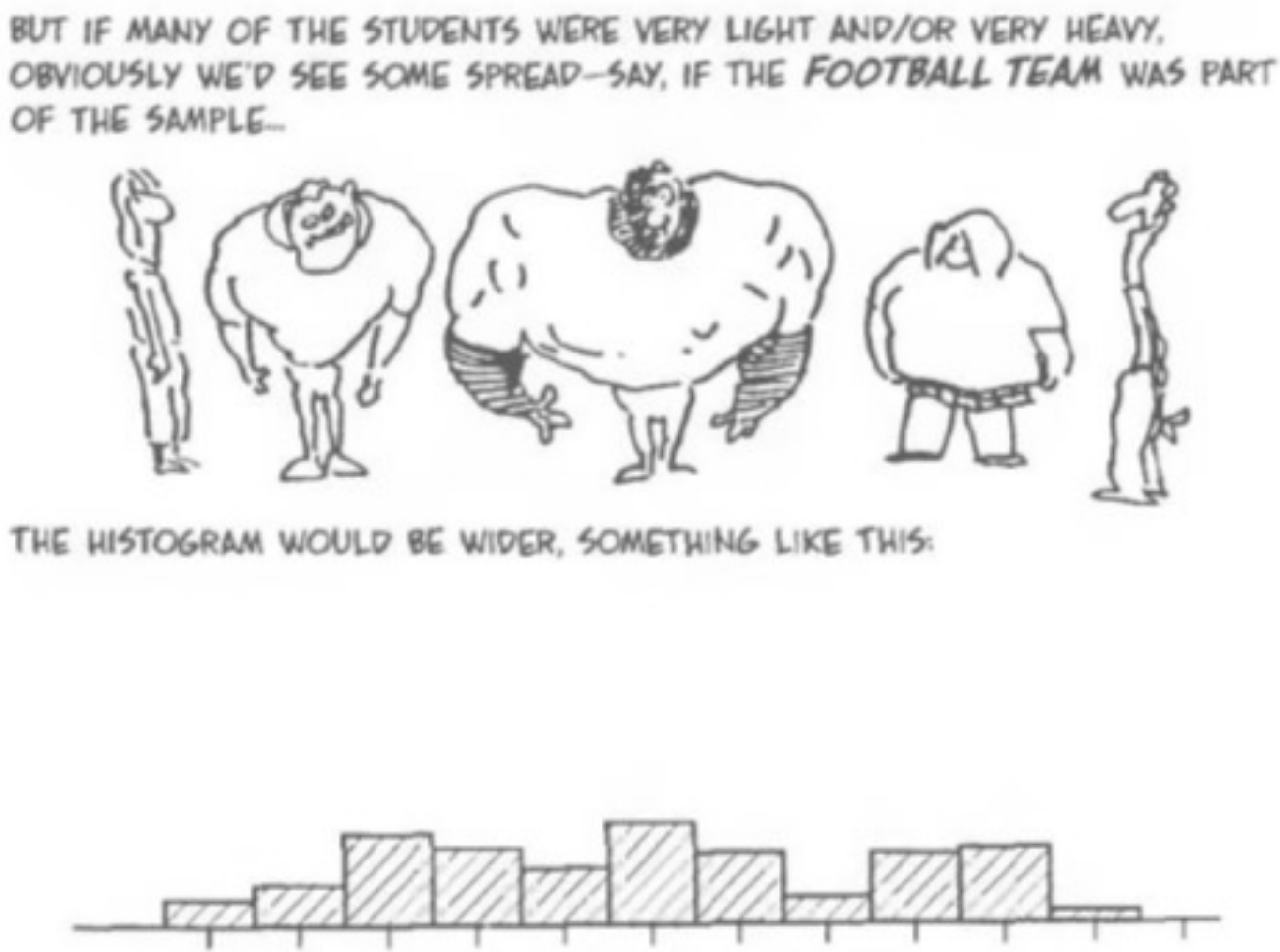
It is the expected squared distance of a value from the centre of the distribution.

RANDOM VARIABLES: Measurement of the spread



From: Gonick & Smith The cartoon guide to Statistics

RANDOM VARIABLES: Measurement of the spread



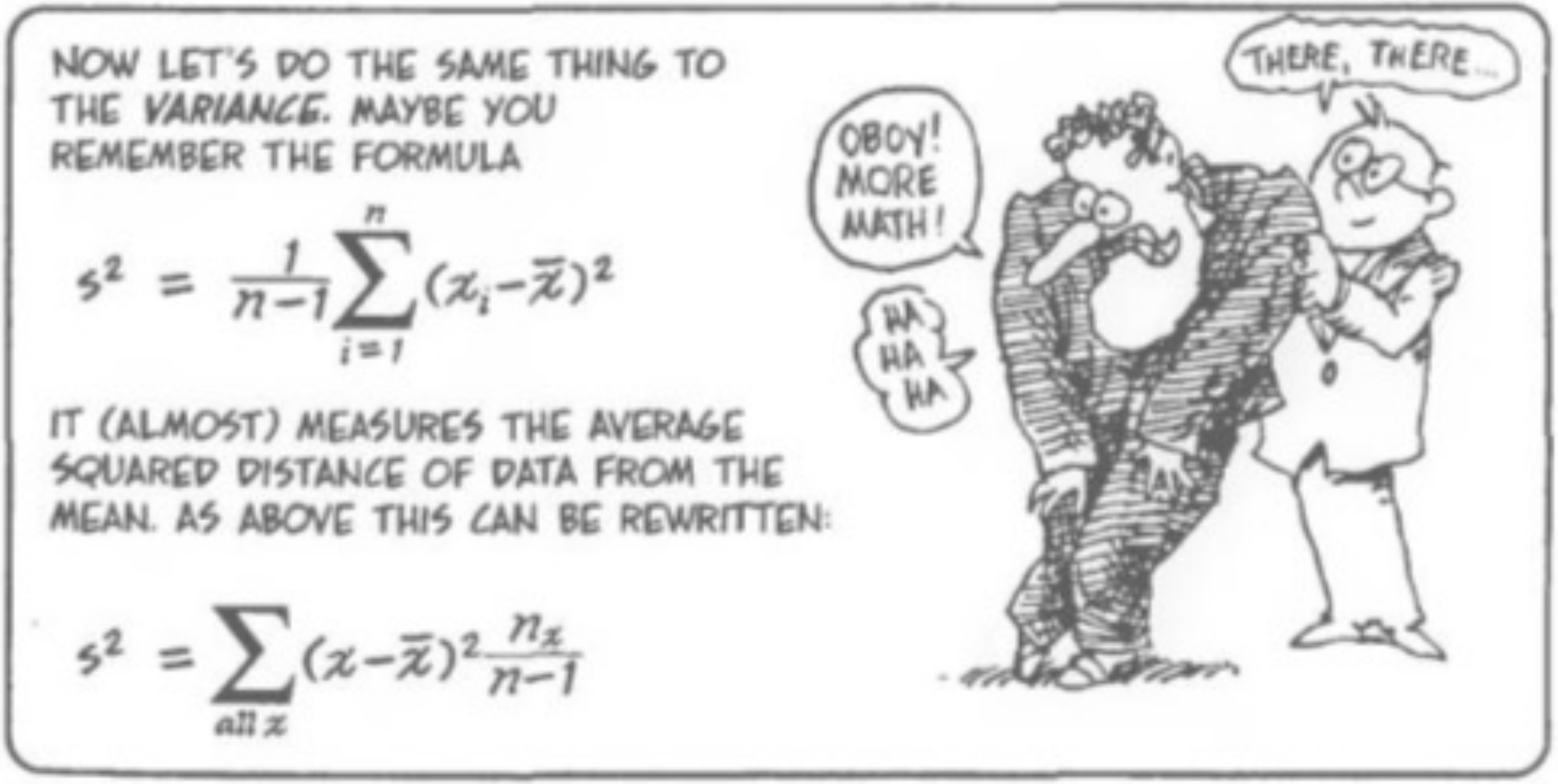
From: Gonick & Smith The cartoon guide to Statistics

VARIANCE OF A RANDOM VARIABLE

It is the expected value of the square of the distance of X to μX .

- If X only took on its average value, the variance would be 0.
- The closer the values of x are to μ the smaller the value of var(X).
- The less likely values of x far from μ are, the smaller the variance.

CALCULATION OF THE VARIANCE OF A DISCRETE RANDOM VARIABLE



From: Gonick & Smith The cartoon guide to Statistics

STANDARD DEVIATION

the standard deviation of a random variable X is the square root of its variance .

$$\sigma_X = \sqrt{\operatorname{Var}(X)} = \sqrt{\operatorname{E}[X^2] - (\operatorname{E}[X])^2} = \sqrt{\operatorname{E}[(X-\mu)^2]}$$

a major advantage of the standard deviation is that it has the same units, if any, as the variable itself.

It gives the expected root mean squared distance of a data point from the centre of the distribution.

The smaller the variance or standard deviation, the more well de ned a distribution is.

VARIANCE OF A CONTINUOUS RANDOM VARIABLE

Find $\mathrm{E}[X^2]$ for the continuous random variable X with probability density function

$$f_X(x) = \frac{3}{4}x(2-x) : 0 \leq x \leq 2$$
$$= 0 \text{ otherwise}$$

taking our general formula

$$\mathrm{E}[h(Z)] = \int_{-\infty}^{\infty} h(z) \cdot f(z)\mathrm{d}z$$

we fill in for our particular $f_X(x)$

$$\mathrm{E}[X^2] = \int_0^2 x^2 \cdot \frac{3}{4}x(2-x)\mathrm{d}x = \frac{6}{5}$$

where we ignored the parts of the improper integral $\int_{-\infty}^{\infty} x^2 \cdot f(x)\mathrm{d}x$ where $f(x)$ is zero.

$\mathrm{E}[g(X)]$ can be interpreted as the 'average' (mean) value of the $g(X)$. **It is a number not a function.**

STANDARD DEVIATION

show how the two formulae for the variance are equivalent

$$\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

we can expand the bracket in E$[(X - \mu)^2]$ to get

$$\mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2 - 2\mu X + \mu^2]$$

and we use the fact that it is a linear operator to write this as

$$\mathrm{E}[X^2 - 2\mu X + \mu^2] = \mathrm{E}[X^2] - \mathrm{E}[2\mu X] + \mathrm{E}[\mu^2]$$

and then that $E[a]= a$ and $E[aX]= aE[X]$

STANDARD DEVIATION

$$\begin{aligned}\mathrm{E}[X^2 - 2\mu X + \mu^2] &= \mathrm{E}[X^2] - \mathrm{E}[2\mu X] + \mathrm{E}[\mu^2] \\ &= \mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2 \\ &= \mathrm{E}[X^2] - 2\mu\mu + \mu^2 \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2\end{aligned}$$

because by definition $\mu = E[X]$.

USEFUL IDENTITY OF THE VARIANCE

$$\boldsymbol{Var(aX+b) = a^2 Var(X)}$$

a shift of the distribution doesn't change its spread.