Random phenomena and relation to cog. sc. experiments I

- A random phenomenon is one where the outcome of an experiment is uncertain - that is cannot be predicted with certainty.
- Examples: a) outcome of tossing a coin, b) outcome of rolling a die, c) predicting the next day's stock market index value, d) predicting a finite time interval when a radio-active nucleus will decay, etc.
- While an individual outcome is uncertain the distribution of outcomes over the long term is stable. Example: if the coin and 6-sided die are fair then over a large number n of tosses and throws (actually as $n \to \infty$) the fraction of heads (and also tails) will be very close to 0.5 and the fraction for any die face will be very close to $\frac{1}{6}$.

Random phenomena and relation to cog. sc. experiments II

- Over the short term (that is when n is small) the variability in outcomes is larger. But in all cases we can calculate the probability of different types of outcomes after making some assumptions (like coin is fair).
- In cog. sc. experiments randomness enters in multiple ways: a) choosing a sample from a population b) individual variability even while controlling for some parameters c) errors in measurements.
- Ultimately, we want to predict for the population. This can only be done with a probability measure and we need to be able to calculate this rigourously based on the assumptions we make about the nature of randomness in various aspects of the experiment.

Sample space, Event

- A **sample space** is the set of all possible outcomes of an experiment that involves some random phenomenon. Examples: coin toss: $\{H, T\}$; die roll:$\{1, 2, 3, 4, 5, 6\}$, drawing a card from a shuffled pack: 52 values, 13 of each suit.
- An **event** is a subset of the sample space. Examples: die roll with an even number, picture cards in a pack of shuffled cards, ten coin tosses that have at least 2 heads. We can associate a probabilty with an event.

Random variable (RV or rv)

Definition 1 (Random variable)

A random variable (RV) is a variable whose values depend on the outcome of a random phenomenon.
Formally it is defined as a function from the sample space to a measureable space E. If X is a random variable then $X : \mathcal{S}(X) \to E$, where $\mathcal{S}(X)$ is the sample space of X. E is either a) some countable set (discrete) or b) an interval of the real line (continous.)
Random variables are normally symbolized by capital letters.
$\triangleleft$

Discrete and continuous random variables I

Definition 1 (Discrete, Continuous RV)

An RV is discrete if the range of values it can take is countable and is continuous if the values it can take fall within an interval (a subset of $\mathbb{R}$).

Examples:

- X - Birth year of a randomly chosen student in the class.
- B - The exact time of birth of a randomly chosen student in the class.
- H - The exact height of a randomly chosen student in the class.
- Y - The number of hair follicles (or hair) on a randomly chosen students head.

Discrete and continuous random variables II

- Z - The winnings of a player (say P1) in the following game after 100 rounds. In one round two players P1, P2 randomly draw a card from separate packs. The player with the lower card value pays the one with the higher card value Re 1/-. If values are same there is no payment.

Distributions I

- Given a population (or sample) and measurements of some attribute or variable (say X) a distribution describes how frequently particular values occur.
- Discrete case: If the variable takes discrete values then it is the relative frequency with which each discrete value occurs. This is called the **probability mass function** or pmf. For any particular value the pmf gives the probability that X has that value. The probabilities for all possible values add up to 1.
- Continuous case: Two possibilities; the **cumulative distribution function** (cdf) or the **probability density function** (pdf). For any value v the cdf gives the probability that $X \leq v$. The pdf is interpreted in terms of intervals. For any two values a, b, $a < b$ the probability that $a \leq X \leq b$ is the area under the pdf curve between a and b. The total probability, that is area under the pdf curve for $-\infty \leq x \leq \infty$ is 1. Also, $cdf(X = a) = \int_{-\infty}^{a} p(x)dx$, where $p(x)$ is the pdf.

Distributions II

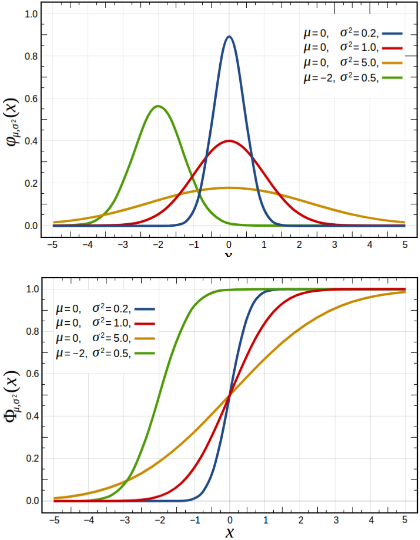
- Since distributions are closely tied up with probability they are called probability distributions.
- An RV is always associated with a probability distribution.

Distribution example - discrete

Two fair dice are rolled and X is sum of the values. Then the sample space is: $(i, j), 1 \leq i, j \leq 6$ and the range space of X is E. $E = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and the pmf is:

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Distribution example - continuous pdf, cdf[3]



[3]Source: Wikipedia

Parameters of a distribution

- Distributions are defined by parameters. For example, the Gaussian or normal distribution is defined by the mean μ and the standard deviation σ or variance σ^2.
- A discrete distribution like the Binomial distribution is defined by parameters p and n. Example, the number of times Head occurs when a coin is tossed n times. p is the probability of a Head in a single toss.
- Populations can consist of mixtures. For example, a mixture of Gaussians. For example, if we look at weight distribution in the population we expect a mixture of two Gaussians - one for males and another for females.
- One major question in statistical inference is how to infer parameter values of the population given the values for a sample.