Random phenomena

- A random phenomenon is one where the outcome is uncertain. Examples: tossing a coin, rolling a die, predicting the stock market index, predicting 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 $\frac{1}{6}$.
- Over the short term (that is when n is small) things are unpredictable and nothing much can be said.

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, two dice roll where the sum is divisible by 3.

Random variable

Definition 2 (Random variable)

A random variable is a variable whose values depend on the outcome of a random phenomenon.
More formally it is defined as a function from the sample space to the set of real numbers. If X is a random variable then $X : \mathcal{S}(X) \to \mathbb{R}$, where $\mathcal{S}(X)$ is the sample space of X. Random variables are normally symbolized by capital letters.
$\triangleleft$

Distributions I

- Given a population (or sample) and measurements of some attribute or variable (say X) a distribution describes in a broad sense how frequently particular values occur.
- Discrete case: If the measured variable has discrete values then it is simply the relative frequency with which each discrete value occurs. This is often called the probability mass function or pmf. For any particular value the pmf gives the probability that X has that value. The sum for all possible values should add up to 1.
- Continuous case: In this case we can define either 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 probability curve between a and b. The total probability, that is area under the pdf curve

Distributions II

for $-\infty \leq x \leq \infty$ is 1. Also, $cdf(X = a) = \int_{-\infty}^{a} p(x)dx$, $p(x)$ is the pdf.

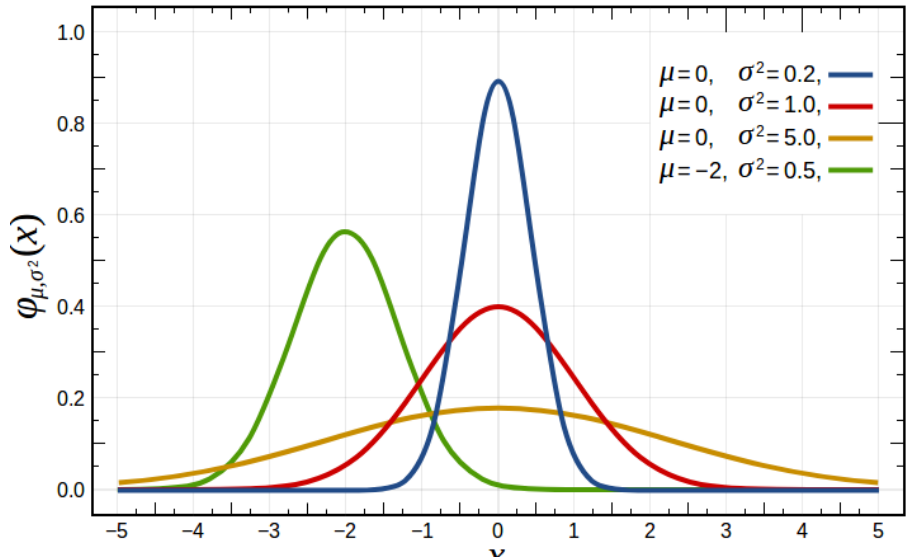
- Since distributions are so closely tied up with probability they are often called probability distributions.

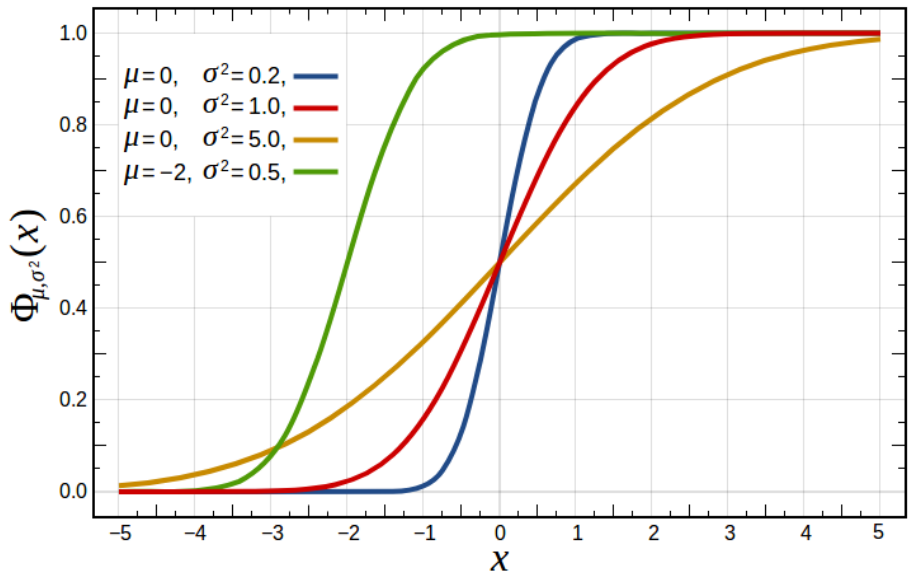
Distribution example - discrete

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

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{16}$	$\frac{1}{36}$

Distribution example - continuous pdf, cdf[3]





[3]Source: Wikipedia

Measures of a distribution I

- Mean. For a discrete distribution $\mu = \sum_{i=1}^{n} x_i p_i$. For a continuous distribution: $\mu = \int_{-\infty}^{\infty} xp(x)dx$, where $p(x)$ is the pdf of X. μ is also called the expected value and often written as $E(X)$.
- Median. The middle value of a distribution. For a discrete distribution assuming the data values are ordered and there are n values then:

$$Median = \begin{cases} x_{\lceil \frac{n}{2} \rceil} & \text{if } n \text{ is odd} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{if } n \text{ is even} \end{cases}$$

For continuous X the median is the value of X for which cdf is 0.5.

Measures of a distribution II

- Mode. Most frequent value for discrete distribution. For a continuous distribution it is the value for which the pdf has the maximum value. More often when there are multiple local maxima it is referred to as a multi-modal distribution.
- Variance. $var(X) = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$. For a continuous distribution $var(X) = \int_{-\infty}^{\infty}(x-\mu)^2 p(x)dx$.
- Standard deviation. $\sigma(X) = \sqrt{var(X)}$

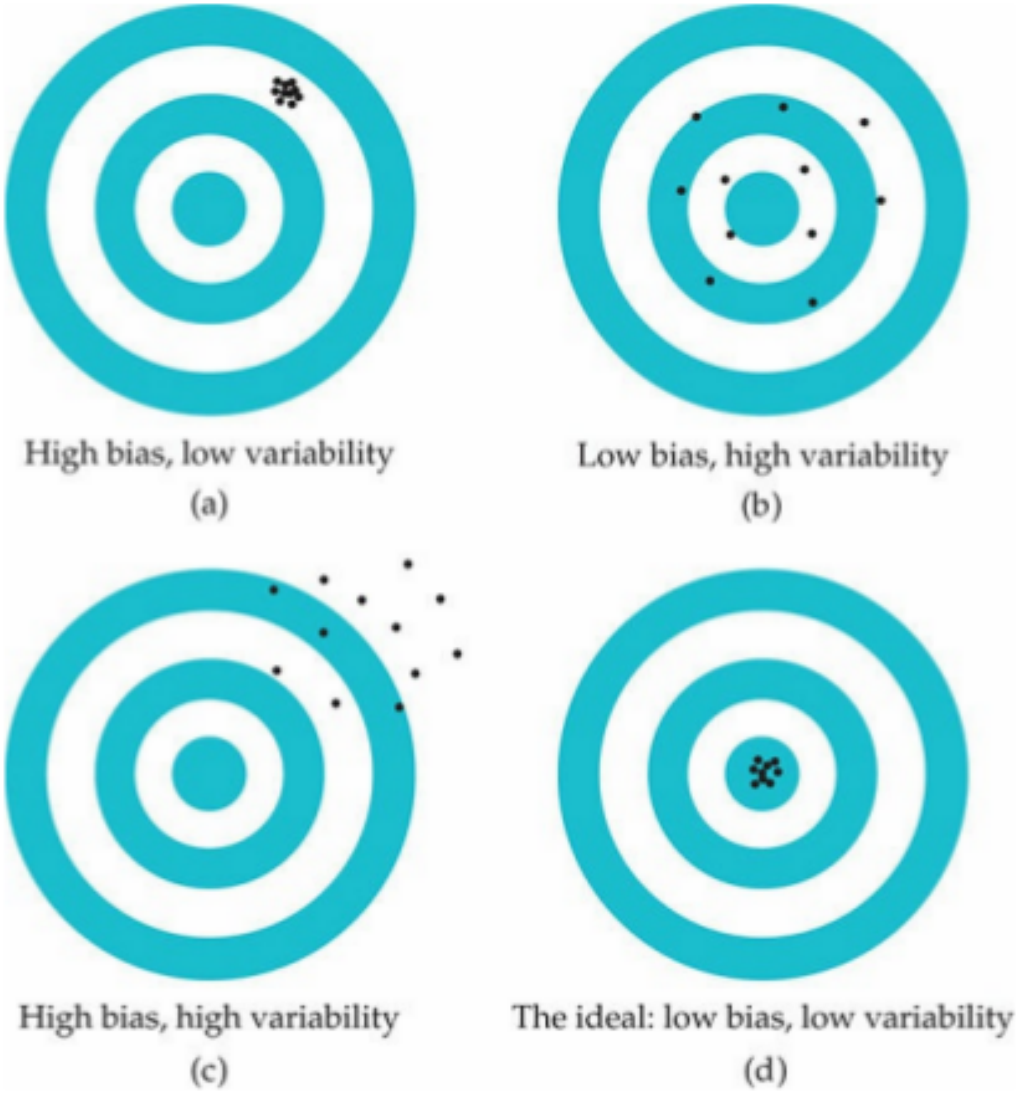
Populations and samples I

- Parameters describe populations.
- A statistic describes the sample.
- Consider picking many random samples the mean (actually any statistic) of these samples will have a distribution called the sampling distribution. Statistics infers unknown population parameters using sample statistics.
- A statistic is unbiased if the mean of the sampling distribution for the statistic is the true value of the parameter for the poplulation. The variability (variance is one measure of variability) or dispersion of a statistic is the spread of the sampling distribution. There are several measures of variability - range, std deviation, inter-quartile range, etc.
- Ideally we want low bias and low variability. (see fig. on next slide).

Populations and samples II

- To reduce bias in a sample use random sampling. To reduce variance of a statistic of the sample distribution increase the size of the sample.
- Variance of a statistic does not depend on the size of the population provided it is large enough (typically greater than 100 times sample size).
- Under-coverage and non-response can bias a sample even when it has been obtained by randomization.

Bias, variability schematic



Basic probability laws or axioms

Let $\mathcal{S}$ be some sample space and $\mathcal{E}$ the event space (that is the set of all subsets of $\mathcal{S}$. Let $P : \mathcal{E} \to \mathbb{R}$. P gives the probability of an event in $\mathcal{E}$ if the following hold:

- $P(E) \geq 0$ for all $E \in \mathcal{E}$.
- $P(\mathcal{S}) = 1$.
- Let $E_i \in \mathcal{E}$ be a countable set of mutually exclusive events (or disjoint sets) then $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$. We will mostly use the finite version of this: $P(\cup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$ where $E_i, i = 1..n$ are n mutually exclusive events.

Consequences of probability laws

- $P(\emptyset) = 0$
- If $A \subseteq B$ then $P(A) \leq P(B)$.
- $0 \leq P(E) \leq 1$ for all $E \in \mathcal{E}$.
- $P(E^c) = 1 - P(E)$. E^c is the complement of E, $E = \mathcal{S} - E$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Independent and conditioned events, Bayes rule I

- Events A and B are **independent** if $P(A)$ is not affected by whether or not B happens and vice, versa ($P(B)$ is not affected by whether or not A happens).
- Product rule: If A and B are independent the $P(A \text{ and } B) = P(A)P(B)$. This is also written as: $P(A, B) = P(A)P(B)$.
- The event that B occurs given that A has occurred is called a conditioned event (B is conditioned on A) and the corresponding probability is written $P(B|A)$. Another way to look at independence is $P(B|A) = P(B)$ since probability of B is unaffected whether or not A occurs.

Independent and conditioned events, Bayes rule II

- The joint probability $P(A, B)$ can be computed the following reasoning: Suppose A has occurred. The probability of this is $P(A)$. For the joint event A and B to occur B must occur given that A has occurred - that is $P(B|A)$ so more generally $P(A, B) = P(A)P(B|A)$. If A, B are independent this gives $P(A, B) = P(A)P(B)$.
- The argument above can be repeated with A, B interchanged so we have: $P(A, B) = P(B)P(A|B)$. This implies: $P(A)P(B|A) = P(B)P(A|B)$ from where we can write: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ - this formula is called **Bayes rule** and is the basis for Bayesian analysis.

Example 1

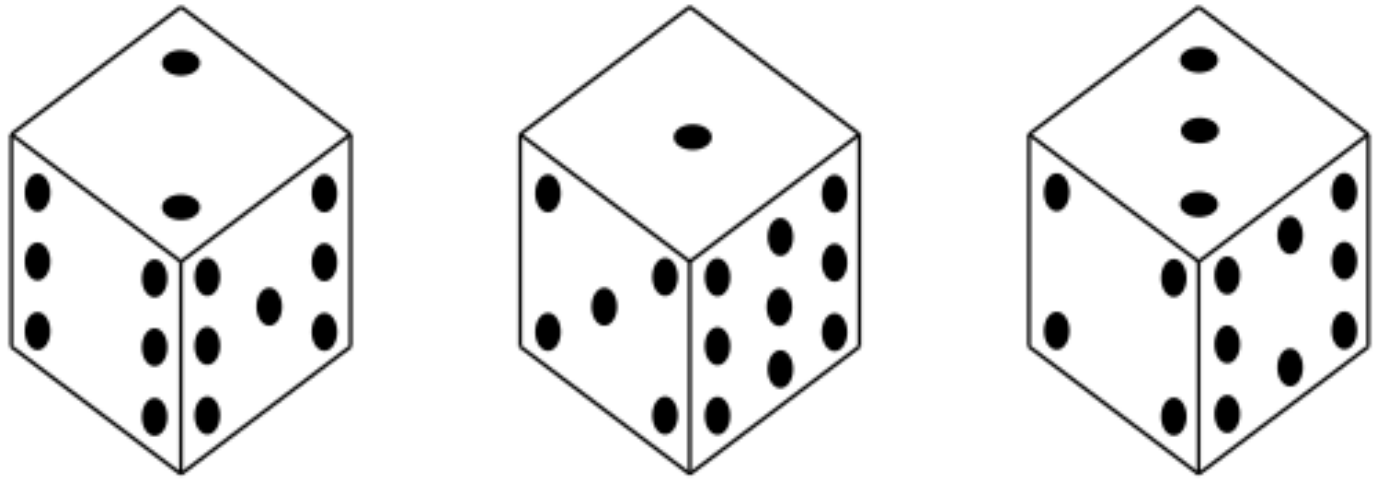


Figure: The number of • is same on opposite sides

The game has two players You and I:
The players select one of the dice and roll the winner is the one who rolls the higher number. Assume You select the die first.