What is probability?

- Let's look at the throw of a dice:
- Six *discrete* outcomes: x_i =1,2,3,4,5,6
- Here, x_i is a discrete random variable ($i = 1..6$)

If we roll the dice N times, what fraction of the trials will be 6 ?

Is it just 1/6 ?

Use a random number generator to create a sequence of N trials.
e.g. N=12: **6, 4, 5, 1, 2, 4, 6, 1, 1, 2, 5, 6**

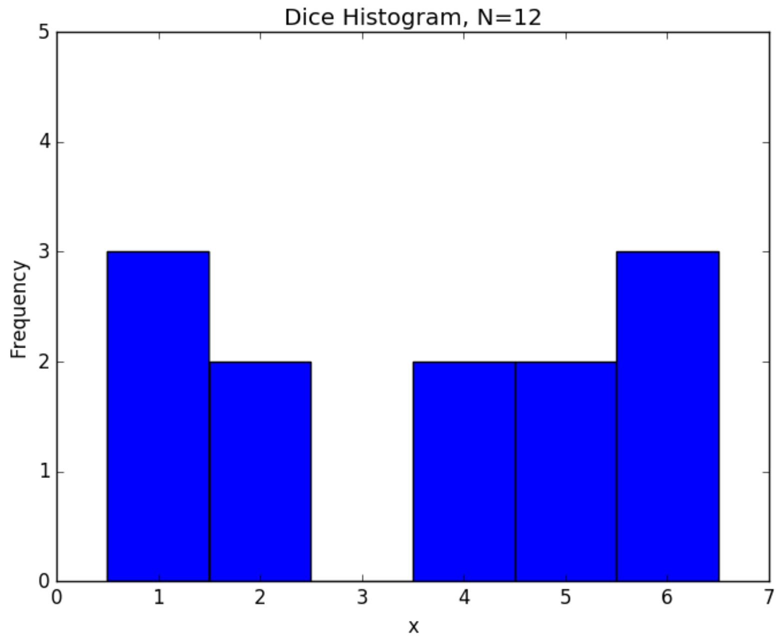
Fraction of '6's = 3/12 = 1/4

Why? – Sample size too small !

Visualizing random trials: Histograms

Can visualize the outcome of our trials in a histogram:

x-axis denotes *all* the possible outcomes x_i, y-axis denotes frequency f_i - the number of times x_i has occurred in N trials.



N=12: **6, 4, 5, 1, 2, 4, 6, 1, 1, 2, 5, 6**

Histogram normalization

We can convert the frequencies into fractions:

$$\sum_{i=1}^{n} f_i = N \rightarrow \sum_{i=1}^{n} \frac{f_i}{N} = 1$$

Notation:

N = number of trials,

n = number of outcomes (for dice $n = 6$).

Confusion alert: Subscript of x may refer to outcome or trial number !

e.g. $\sum_{i=1}^{n} x_i = 1 + 2 + 3 + 4 + 5 + 6$ for dice

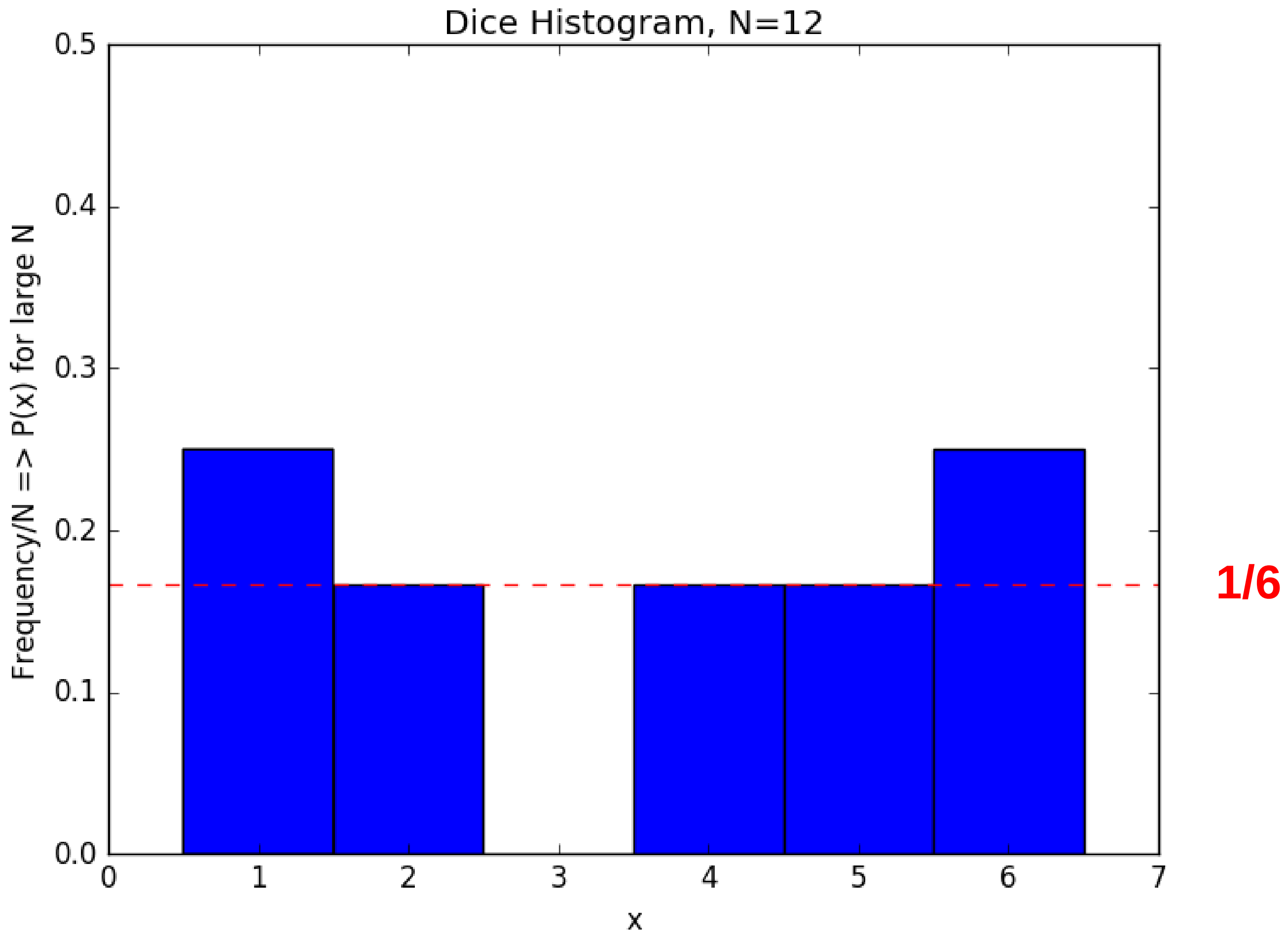
but $\sum_{j=1}^{N} x_j = \underbrace{2 + 5 + 1 + 3 + 4 + ..}_{\text{N trials}}$ (summing over N random trials of a dice)

- Normalization condition: $\sum_{i=1}^{n} \frac{f_i}{N} = 1$. All the fractions have to sum up to 1 (100%)
- A normalized histogram shows the probability distribution (in the limit of large N).

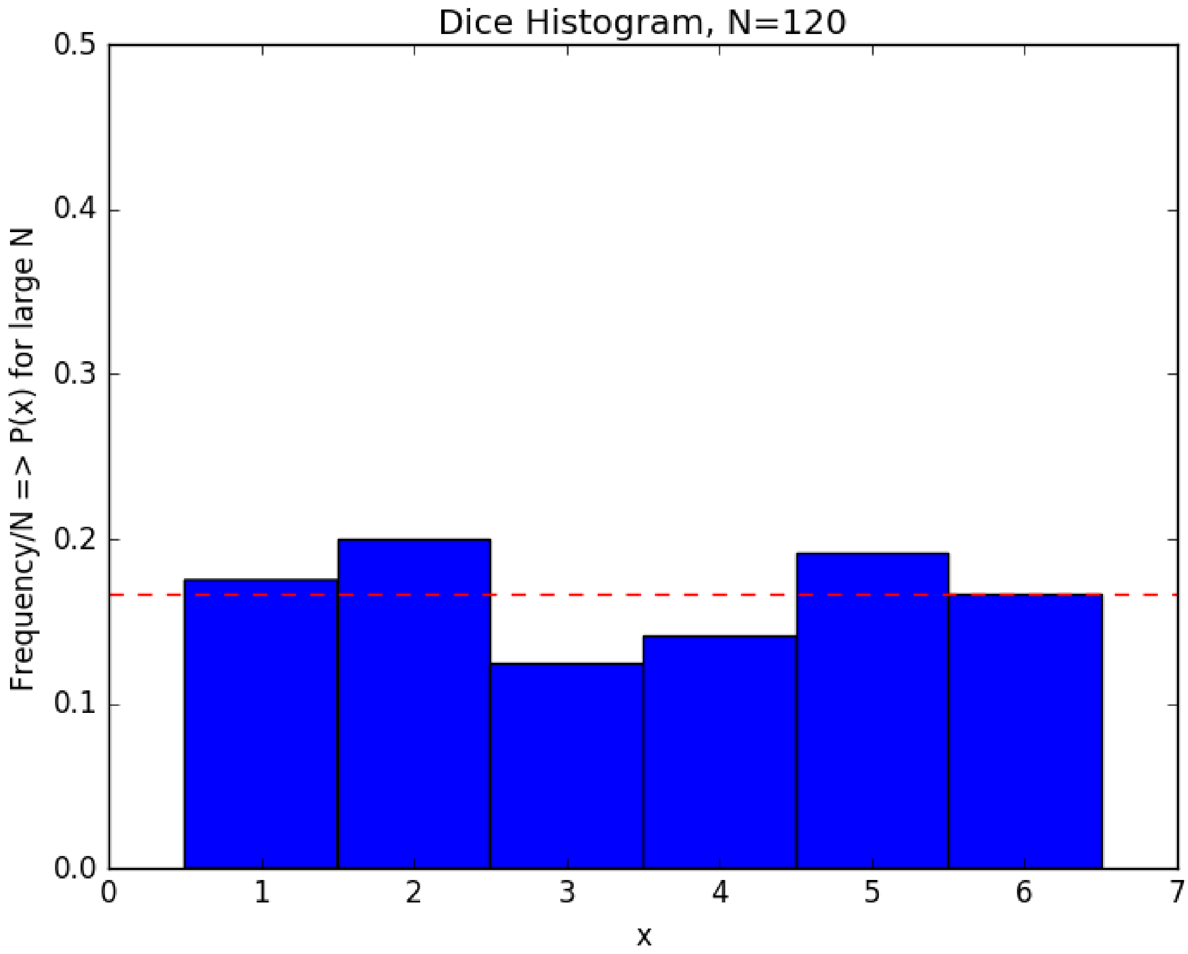
Limiting (or parent) distribution

- The true shape of the distribution will only emerge as number of trials $N \to \infty$.
- The **true probability** $\boldsymbol{P(x_i) = \lim_{N\to\infty} \frac{f_i}{N}}$, is the fraction of trials that have an outcome x_i in the **limit of infinite number of trials ($\boldsymbol{N \to \infty}$).**
- In this limit, the histogram will converge to the **limiting/parent distribution P(x).**
- E.g. true distribution of dice roll is $P(x_i) = \frac{1}{6}$.
- Need large N to see parent distribution emerge

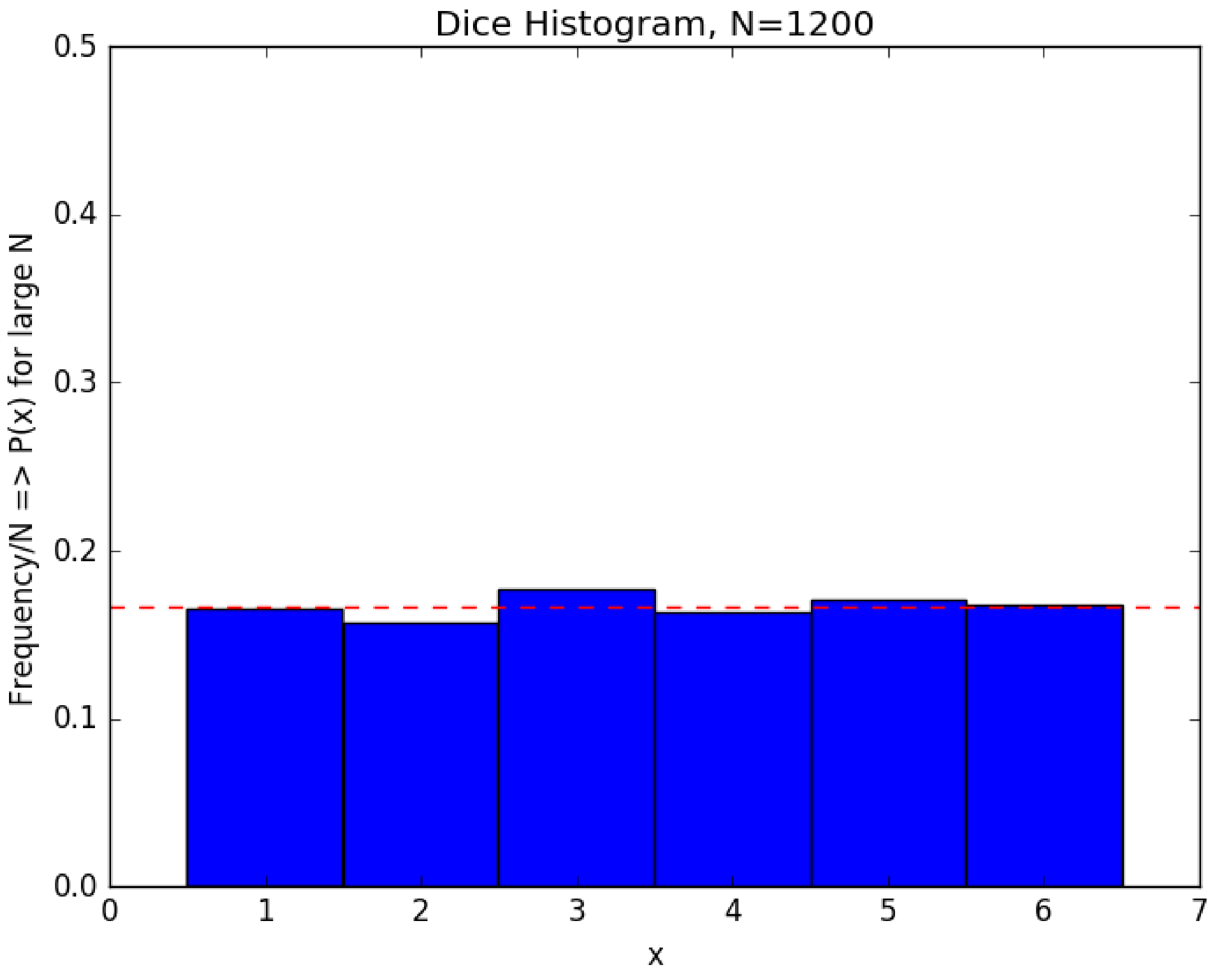
Evolution towards parent distribution



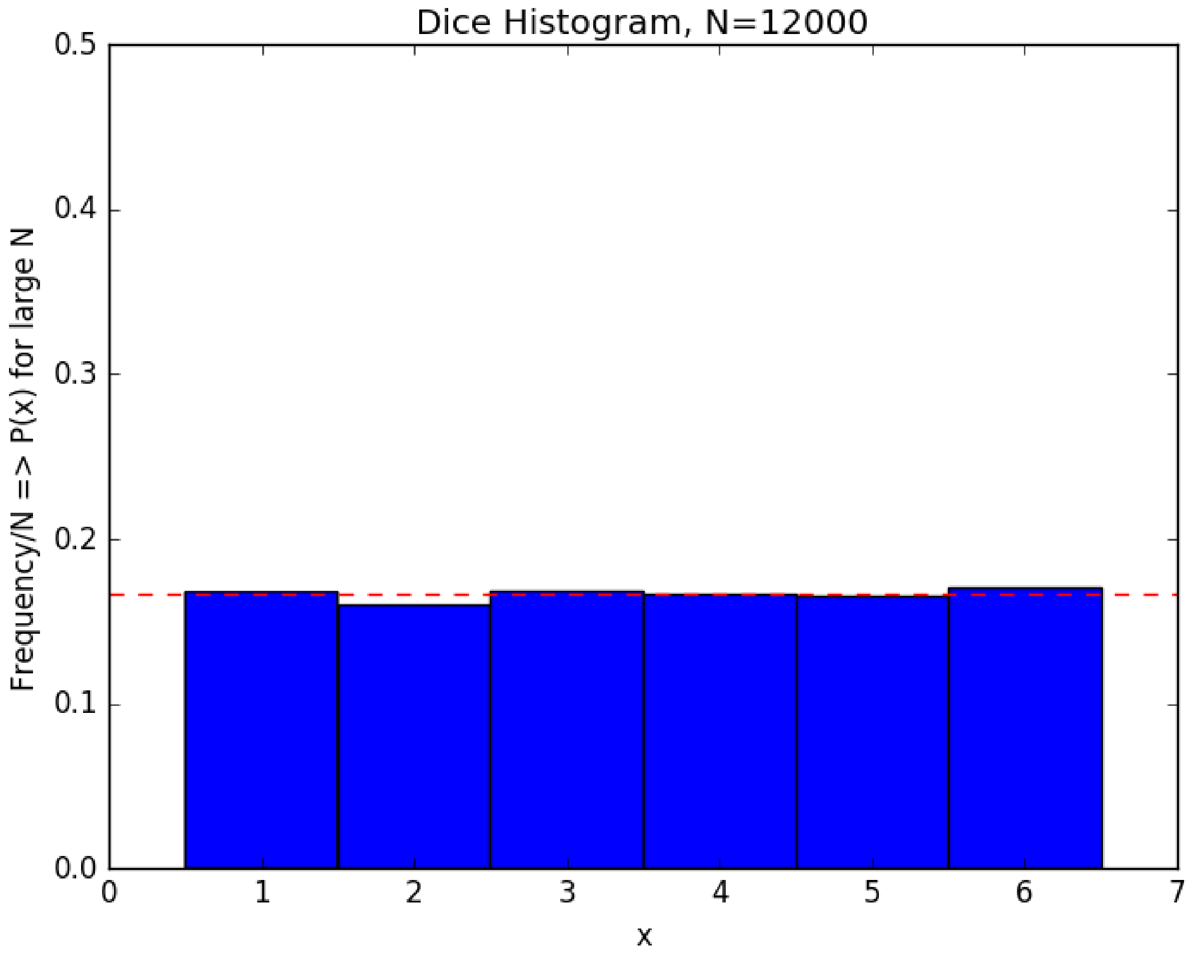
Evolution towards parent distribution



Evolution towards parent distribution

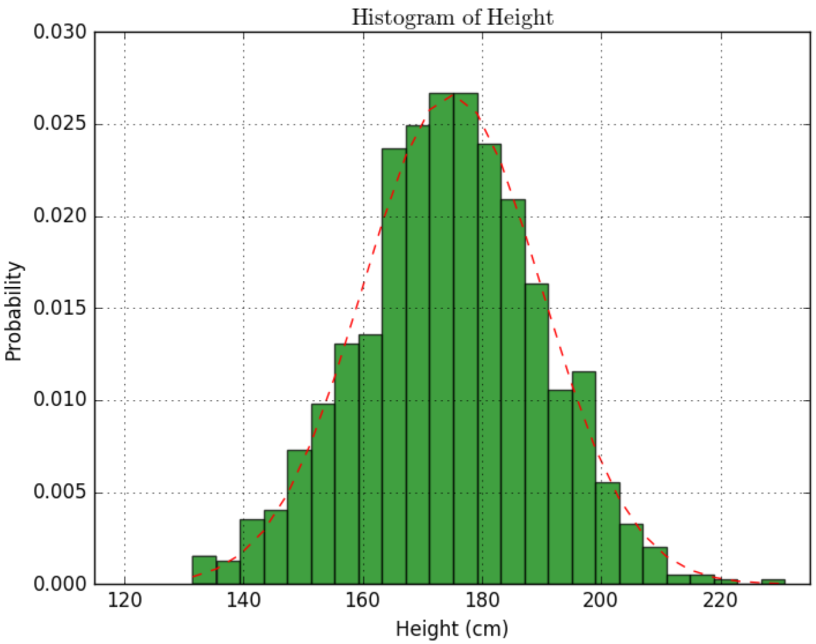


Evolution towards parent distribution

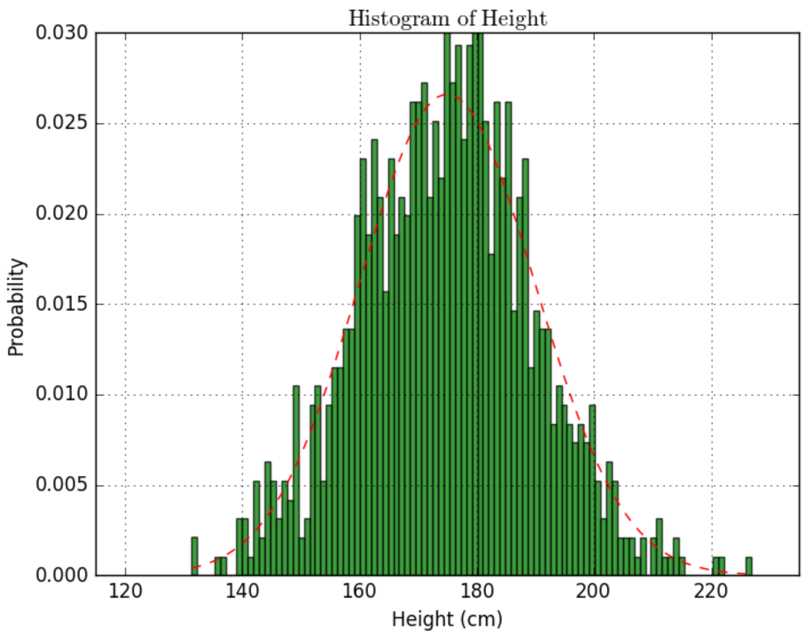


Histogram of a continuous random variable

- Histograms can also be used for *continuous* random variables. E.g. height of people.
- In this case, data is *'binned'* – each bin counts number of occurrence within its *'bin size'.*



Bin size = 4cm, N=1000



Bin size = 1cm, N=1000

Bin size matters. If too small, histogram noisy. Larger N allows for smaller bin size

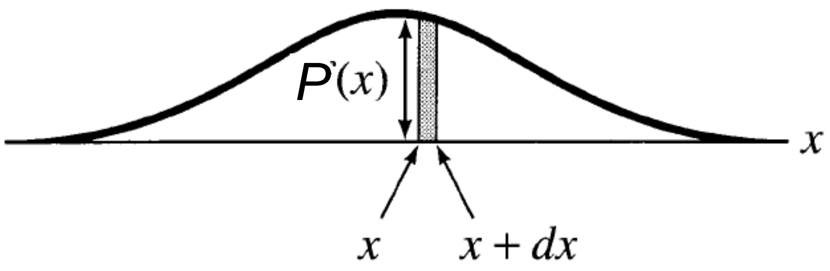
Histogram vs Probability density

- As bin size Δx becomes infinitesimally small, we obtain a probability density. In both cases the area under the curve is 1 due to the normalization condition:

$$\sum_{i=1}^{n} P(x_i)\Delta x = 1 \rightarrow \int_{-\infty}^{+\infty} P(x)dx = 1$$

- Note: Unlike discrete probability distribution $P(x_i)$, the probability density can have units [1/x].

- Probability at x is $P(x)dx$:



The mean and variance of a distribution

The most common way to characterize a distribution P(x) is through its **Mean** and **Variance.**

Mean

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i$$

The mean corresponds to the average value of x_i.

Can also be denoted by $\langle x \rangle$ or $\bar{x}$.

Variance

$$Var(x) = \sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

The variance characterizes the deviations from the mean.

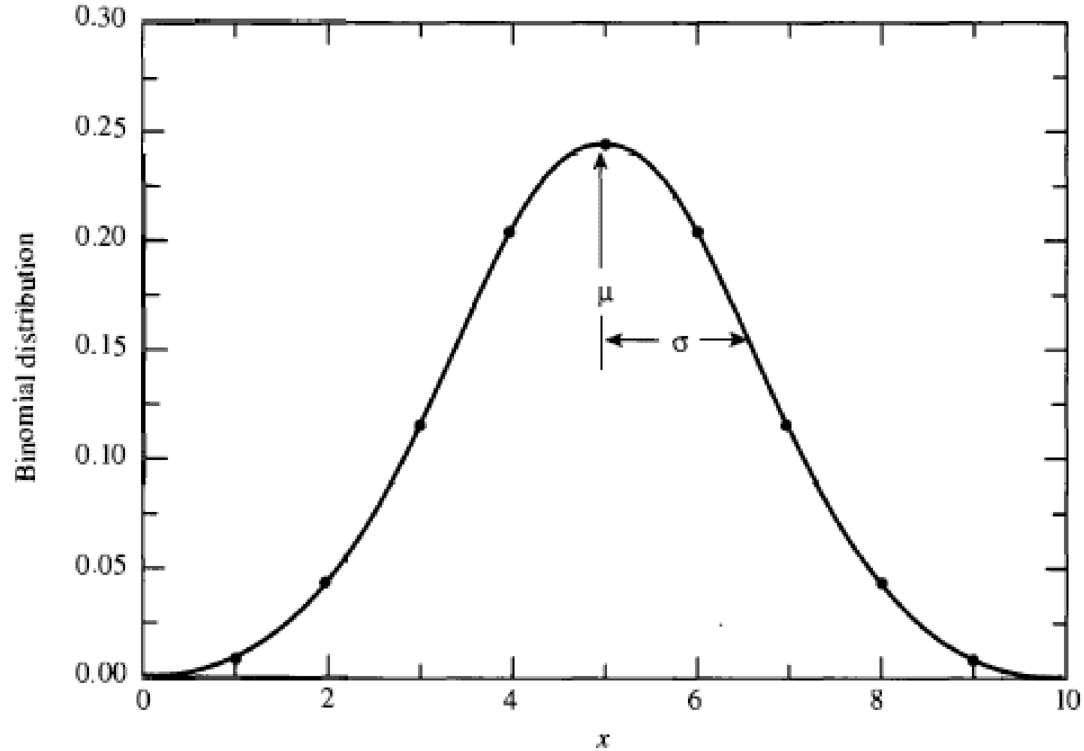
Standard deviation

The standard deviation of P(x) is simply

$$\sigma = \sqrt{Var(x)}$$

It is related to the width of the distribution P(x).

If P(x) is bell shaped, then full-width-half-maximum (FWHM) is $\approx 2\sigma$



Sample mean and standard deviation

$$\mu = \cancel{\lim_{N\to\infty}} \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\cancel{\lim_{N\to\infty}} \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}$$

Note that in practice N is finite when calculating the mean and standard deviation. Therefore, **the measured $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ will differ from the true value.**

Later in the course will discuss the uncertainty associated with it.

Computing the mean from the parent distribution

For a given P(x), mean μ (or $\langle x \rangle$) is given by :

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{n} x_j f_j = \lim_{N\to\infty} \sum_{j=1}^{n} x_j \left(\frac{f_j}{N}\right)$$

Sum over N trials

Sum over n outcomes, weighted by frequency f_j

Fraction converges to $P(x_j)$ as $N \to \infty$

$$\boxed{\mu = \sum_{j=1}^{n} x_j P(x_j)}$$

for continuous random variable:

$$\boxed{\mu = \int_{-\infty}^{+\infty} x \cdot P(x)\, dx}$$

Computing the Variance from the parent distribution

For a given P(x) with mean $\langle x \rangle$, variance is given by :

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - \langle x \rangle)^2$$

Using the same arguments as on previous slide this can be written as

$$\sigma^2 = \sum_{i=1}^{n} \left(x_j - \langle x \rangle\right)^2 P(x_j)$$

or for continuous x

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 P(x) dx$$

Moments of a distribution and Variance

Can express variance in terms of moments of $P(x)$:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 P(x) dx$$

$$= \int_{-\infty}^{+\infty} (x^2 - 2x\langle x \rangle + \langle x \rangle^2) P(x) dx$$

$$= \int_{-\infty}^{+\infty} x^2 P(x) dx - 2\langle x \rangle \int_{-\infty}^{+\infty} x P(x) dx + \langle x \rangle^2 \int_{-\infty}^{+\infty} P(x) dx$$

$$= \langle x^2 \rangle - 2\langle x \rangle^2 + \langle x \rangle^2$$

$$\boxed{\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2}$$ (valid for discrete and continuous x)

n[th] moment of $P(x)$: $\langle x^n \rangle = \int_{-\infty}^{+\infty} x^n P(x) dx$

Binomial distribution

The binomial distribution governs the likelihood of having a certain number of successes in a fixed number of *independent* binary (yes/no) experiments.

Examples:

- Coin flip: what is probability of getting 2 heads in 5 coin flips?
- Throwing a dice: Probability of rolling a '6' in three trials.
- Drugs: 80% of patients respond to a certain medication (success or failure). It is given to 10 new patients. What is the probability that it will work for less than 5 patients?

Gambler’s fallacy

We assume that the trials are independent from each other.

Example: You watch your friend flip a (fair) coin. He gets 3 Heads in a row. He gives you the coin. What is the probability of you flipping a Head?

50%

Past trials do not affect the current one !

Example: Throwing a dice

Let's throw a dice three times.

What is the probability of getting three 6's ?

$$P(3\ \times\ 6) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \approx 0.5\%$$

(use multiplication to combine probabilities)

What is the probability of getting two 6's ?

Example: Throwing a dice

First need to enumerate the number of different permutations.

6,6, not 6

6, not 6, 6

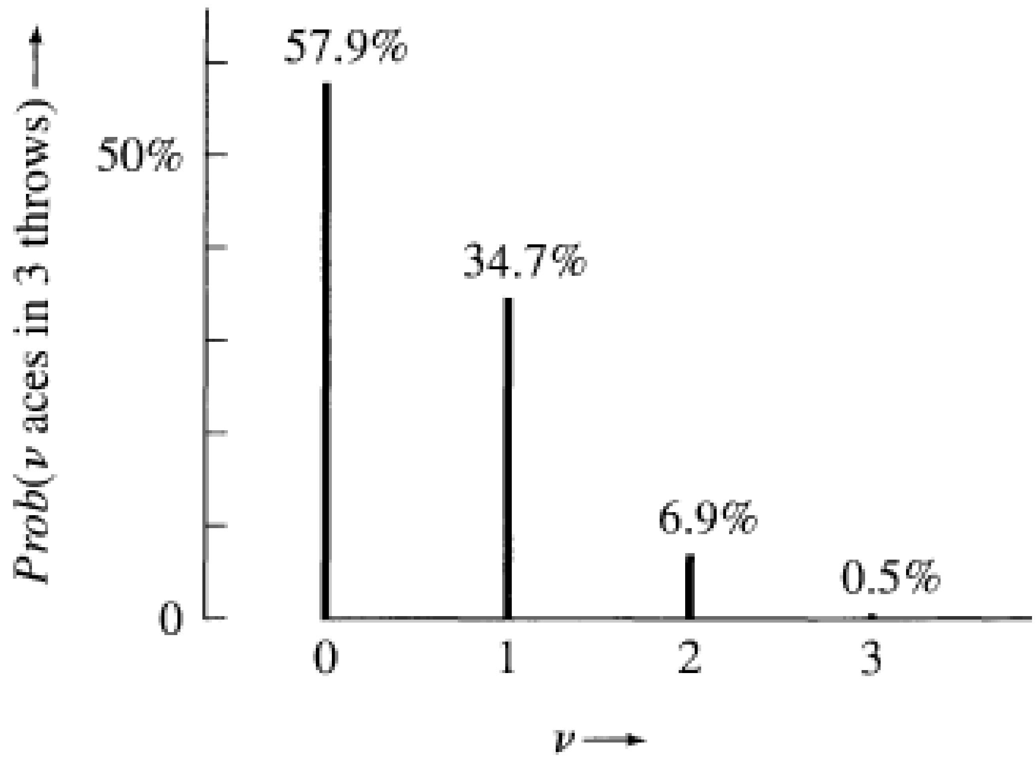
not 6, 6, 6

Three permutations. The probability of not throwing a '6' is just $1 - \frac{1}{6} = \frac{5}{6}$. Therefore,

$$P(2 \times 6s) = 3 \times \left(\frac{1}{6}\right)^2 \times \frac{5}{6} \approx 6.9\%$$

Example: Throwing a dice

Similarly, can show that $P(1\ \times 6) = 34.7\%$ $\quad P(0\ \times 6) = 57.9\%$



The general case

Let's toss a coin N times. What is the probability of getting x heads?

Let the success probability (Heads in this case) be p and failure $q = 1 - p$.

In order to solve this problem, enumerate all possible combinations of having x successes in N trials

Let's imagine N place holders that can be populated with Heads or Tails. In how many ways can we arrange the Heads?

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ … _ (N place holders)

The general case

For the first Head we have N possibilities.

For the second Head, we have $(N-1)$ placeholders to choose from.
The third has $(N-2)$ etc.

Can express this in terms of factorials:

$$N! = N \times (N-1) \times (N-2) \times (N-3) \times \cdots \times 1$$

So the number of permutations of x successes in N trials is

$$\frac{N!}{(N-x)!} = \underbrace{N \times (N-1) \times \cdots (N-x+1)}_{x \text{ terms}}$$

The general case

For $N = 3$ and $x = 2$ (e.g. 2 x 6s in three throws), we get $\frac{3 \times 2 \times 1}{1} = 6$ permutations, but there are only three ! (see previous example).

When calculating the permutations, have implicitly assumed that the successes are ordered:

e.g. for throwing 2 x 6's:

6a,6b, not 6

6b,6a, not 6

Instead of just 6,6, not 6

In this example, overcounted by a factor of 2

The general case

- To fix this, need to consider the number of permutations of sequence of x distinct objects.
- E.g. How many ways can A,B,C be reordered ?
- # of Permutations $= 3 \times 2 \times 1 = 3! = 6$
- In general the number of permutations is $x!$

The binomial coefficient

The number of combinations of having x successes in N trials is therefore

$$\frac{N!}{(N-x)!\,x!} = \binom{N}{x}$$

This is called the binomial coefficient.
Why?

It appears in the binomial expansion:

$$(p+q)^N = p^N + np^{N-1}q + \cdots + q^N = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x}$$

e.g. $N = 2$: $(p+q)^2 = p^2 + 2pq + q^2$

The binomial distribution

We can now write down the Binomial distribution

$$B_{N,p}(x) = \binom{N}{x} p^x q^{N-x}$$

$B_{N,p}(x)$ is the probability of having x successes in N trials for a given success probability p and q is the probability of failure $q = 1 - p$

Using the binomial expansion can show that this satisfies the normalization condition. Since $p + q = 1$

$$(p+q)^N = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x} = 1$$

Mean of binomial distribution

The mean of binomial distribution is given by

$$\langle x \rangle = \sum_{x=0}^{N} x \cdot B_{N,p}\,(x) = \sum_{x=0}^{N} x \binom{N}{x} p^x q^{N-x}$$

Solve this using a trick. Let's write down the binomial expansion:

$$(p+q)^N = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x}$$

Mean of binomial distribution

$$(p+q)^N = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x}$$

Differentiate w.r.t. p, i.e. $\frac{\partial}{\partial p}$ on both sides

$$N(p+q)^{N-1} = \sum_{x=0}^{N} \binom{N}{x} x p^{x-1} q^{N-x}$$

Set $(p+q) = 1$ and multiply both sides with p

$$Np = \sum_{x=0}^{N} x \binom{N}{x} p^x q^{N-x} = \sum_{x=0}^{N} x \cdot B_{N,p}\,(x)$$

This is just the definition of the mean: $\boxed{\langle x \rangle = Np}$

Standard deviation

Find variance by calculating the moments:

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$$

Already know: $\langle x \rangle = Np$, what is $\langle x \rangle^2$?

Binomial expansion:

$$(p+q)^N = \sum_{x=0}^{N} \binom{N}{x} p^x q^{N-x}$$

Differentiate w.r.t. p twice, i.e. $\frac{\partial}{\partial p}$ on both sides,

$$N(N-1)(p+q)^{N-2} = \sum_{x=0}^{N} \binom{N}{x} x(x-1) p^{x-2} q^{N-x}$$

Set $(p+q) = 1$ and multiply both sides with p^2

Standard deviation

$$N(N-1)p^2 \;= \sum_{x=0}^{N} \binom{N}{x} x(x-1)p^x q^{N-x}$$

$$N(N-1)p^2 = \sum_{x=0}^{N} x^2 \binom{N}{x} p^x q^{N-x} - \sum_{x=0}^{N} x \binom{N}{x} p^x q^{N-x})$$

$$N(N-1)p^2 = \langle x^2 \rangle - \langle x \rangle = \langle x^2 \rangle - \mathrm{Np}$$

$$\langle x^2 \rangle = Np + N(N-1)p^2$$

Therefore, $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = Np + N^2p^2 - Np^2 - N^2p^2$

$$= Np(1-p)$$

$$\boxed{\sigma = \sqrt{Np(1-p)} = \sqrt{Npq}}$$

Binomial distribution - summary

Binomial distribution: $B_{N,p}(x) = \binom{N}{x} p^x q^{N-x}$

N = number of trials, p = success probability, $\mathrm{q} = 1 - p$ failure probability

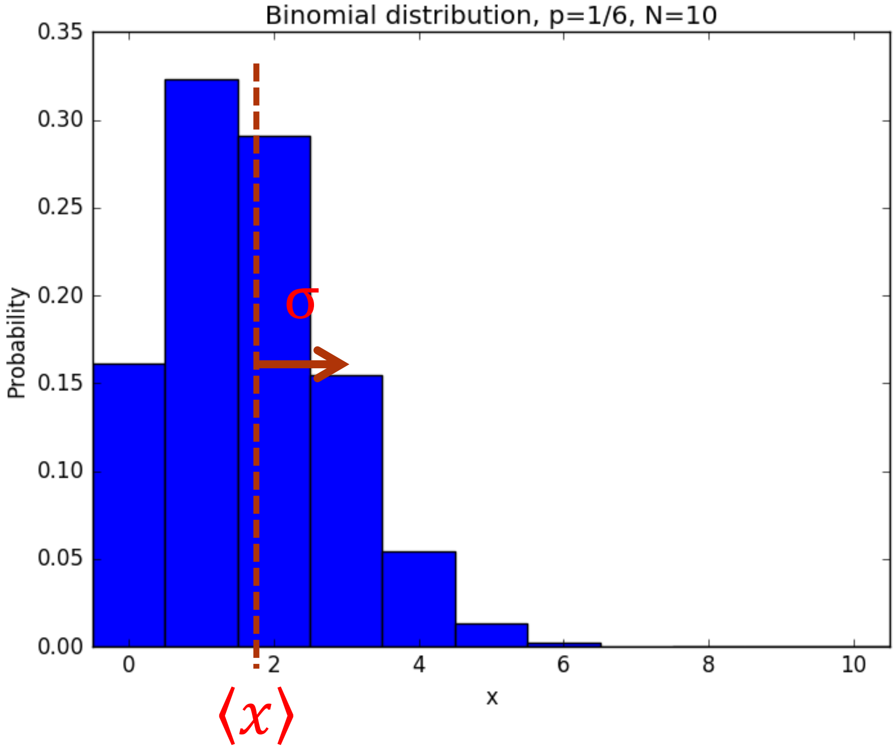
Binomial coefficient: $\binom{N}{x} = \frac{N!}{(N-x)!x!}$

Mean: $\langle x \rangle = Np$

Standard deviation: $\sigma = \sqrt{Np(1-p)} = \sqrt{Npq}$

Mean versus most probable value

In general binomial distribution is asymmetric. The peak value is not equal to the mean! Only the same for symmetric distributions.



$N = 10, p = \frac{1}{6},\qquad \langle x \rangle = \frac{10}{6} = 1.66$ but most probable outcome is 1