4.1.2.2.2

Random Numbers and Distributions
Session 2

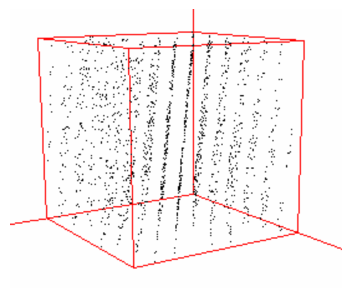
- probability distributions

last time

- computers cannot generate real random numbers, only pseudo-random numbers
  - PRNs are drawn from a deterministic sequence, with possibly a very large period
    - Marsenne Twister has a period of ~220,000
  - linear congruential generators (LCG) are a very common method of generating PRNs
    - modified LCGs are at the heart of many PRNGs like ANSI C rand(), drand48() and others

$$r_{i+1} = (ar_i + c) \bmod m$$

  - LCG values do not fill space evenly
    - choose a,c,m carefully
    - do not choose a,c,m yourself
    - random.mat.sbg.ac.at/~charly/server/node3.html



  - LCG sequences can fail randomness tests well before the end of their period
    - Park-Miller minimal standard fails chi-squared after 10^7 numbers (<1% of its period)

probability distributions

- the statistical outcome of random processes can frequently be described by using a probability distribution
- if the observed outcomes are $x_1, x_2, x_3, \ldots$, then the probability distribution gives the chance of observing any one of the outcomes.

$$P(x_1), P(x_2), P(x_3), \ldots$$

$$P(x_i) \geq 0 \quad \sum P(x_i) = 1$$

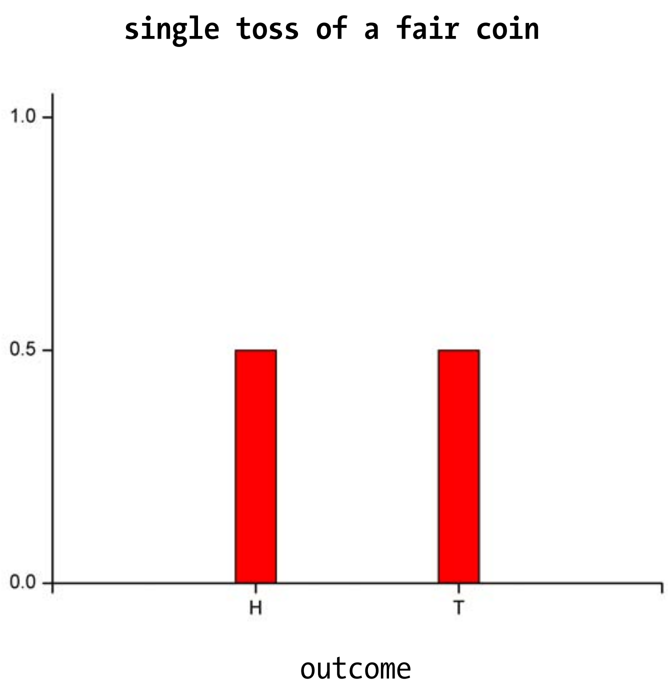
- for a fair coin, there are two outcomes, heads (H) or tails (T)

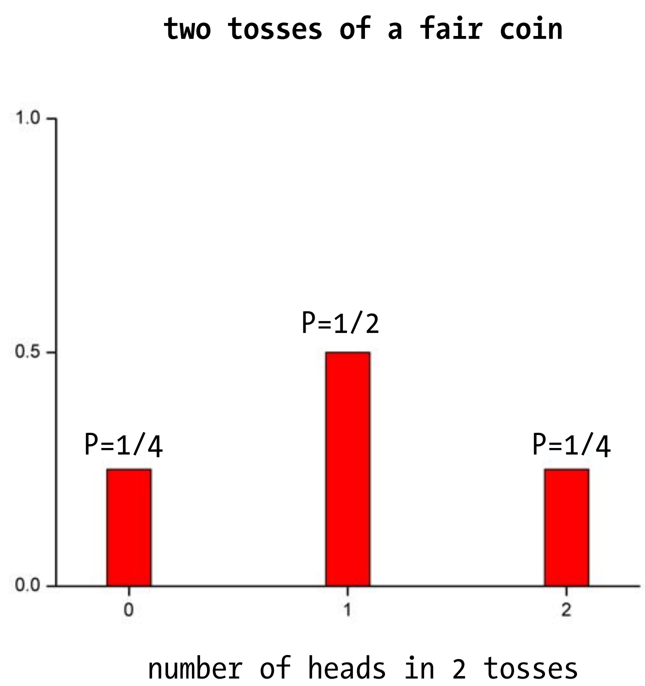
$$P(H) = P(T) = \frac{1}{2}$$

- these are examples of a finite random variable – number of outcomes are finite

graphing probability distributions

- probability distributions are drawn as column graphs when the number of outcomes is small





expectation value

- the average or expectation value of a random variable is given in terms of the probability distribution

$$E(X) = \sum x_i P(x_i)$$

- E(x) is a weighted average, with the weights of each outcome being the probability of that outcome
  - for two tosses of a coin, the average number of heads in two tosses is

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

number of heads, nh

probability of observing nh heads

expectation value

- if the random variable is uniformly distributed, then P(x) is constant
- if there are n outcomes, the requirement $\sum P(x_i) = 1$ implies $P(x_i) = \frac{1}{n}$
- the expectation value reduces to the familiar form of the average

$$E(X) = \sum x_i P(x_i) = \frac{\sum x_i}{n}$$

- commonly μ is used to indicate the expectation value

variance

- the variance gives a sense of the dispersion of the random values away from the expectation value
- the variance is given by

$$\mathrm{var}(X) = \sum (x_i - E(X))^2 P(x_i)$$

- this looks just like an expectation value – and it is, but not of the variable but of a transform of the variable – the square of the distance from the average

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

- the standard deviation, σ, is related to the variance

$$\sigma^2 = \mathrm{var}(X)$$

standardized random variable

- the overall shape of the probability distribution function is of most importance
  - for different values of the mean, the distribution will be "centered" at a different value
  - for different values of the variance, the distribution will be "stretched" differently
  - mean and variance are parameters, not fundamental descriptors of the distribution

- standardized random variable is a transformation

$$Z = \frac{X - E(X)}{\sqrt{\mathrm{var}(X)}} = \frac{X - \mu}{\sigma}$$

- Z is used to describe the underlying nature of a process (e.g. cars arriving at a traffic light), whereas X describes a particular instance (e.g. cars arriving at a traffic light in rush hour)

$$E(Z) = 0 \quad \mathrm{var}(Z) = 1$$

chebyshev's inequality

- the variance measures the spread of values about the mean
  - the smaller the variance, the more tightly values are grouped around the mean

- Chebyshev's inequality puts a lower bound on the probability of finding a random variable within a multiple of the standard deviation
  - recall that for a normal distribution, a value will land within σ 68% of the time (2σ 95%)

$$P(\mu - k\sigma \leq x \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

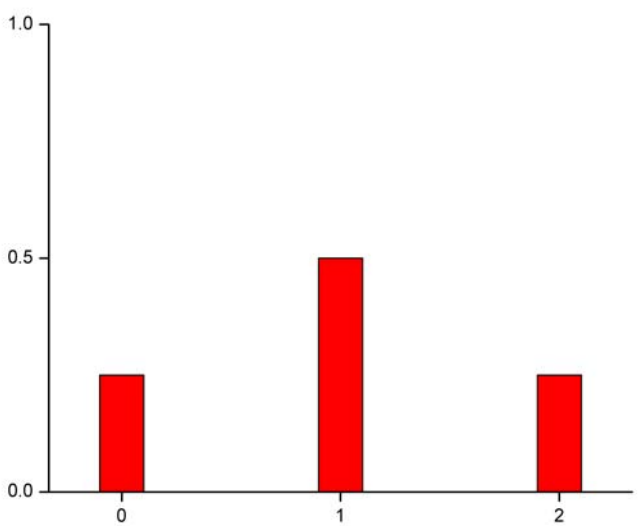
- for any random variable,
  - probability of falling within 2σ is at least 75%
  - ... within 3σ is at least 89%
  - ... within 4σ is at least 94%

cumulative probability distribution

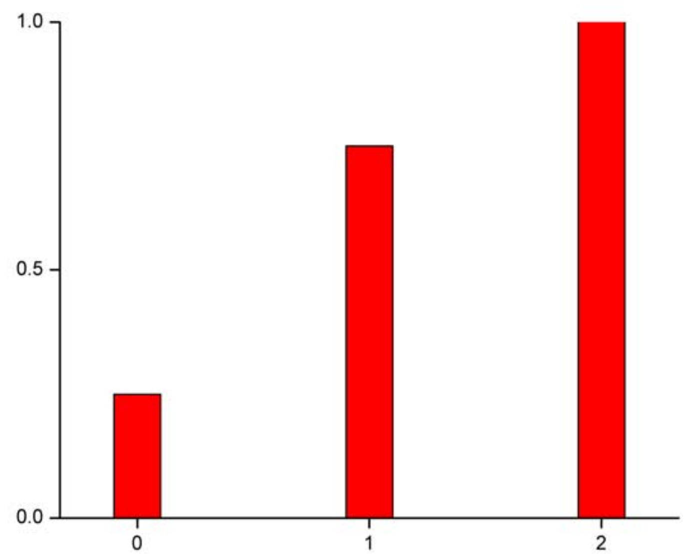
- the probability distribution P(x) gives you the chance of observing an event, x
- the cumulative distribution F(x) gives you the chance of observing any event ≤x

$$F(x_i) = \sum_{x_j \le x_i} P(x_j)$$

probability of observing exactly nh heads



probability of observing nh or fewer heads

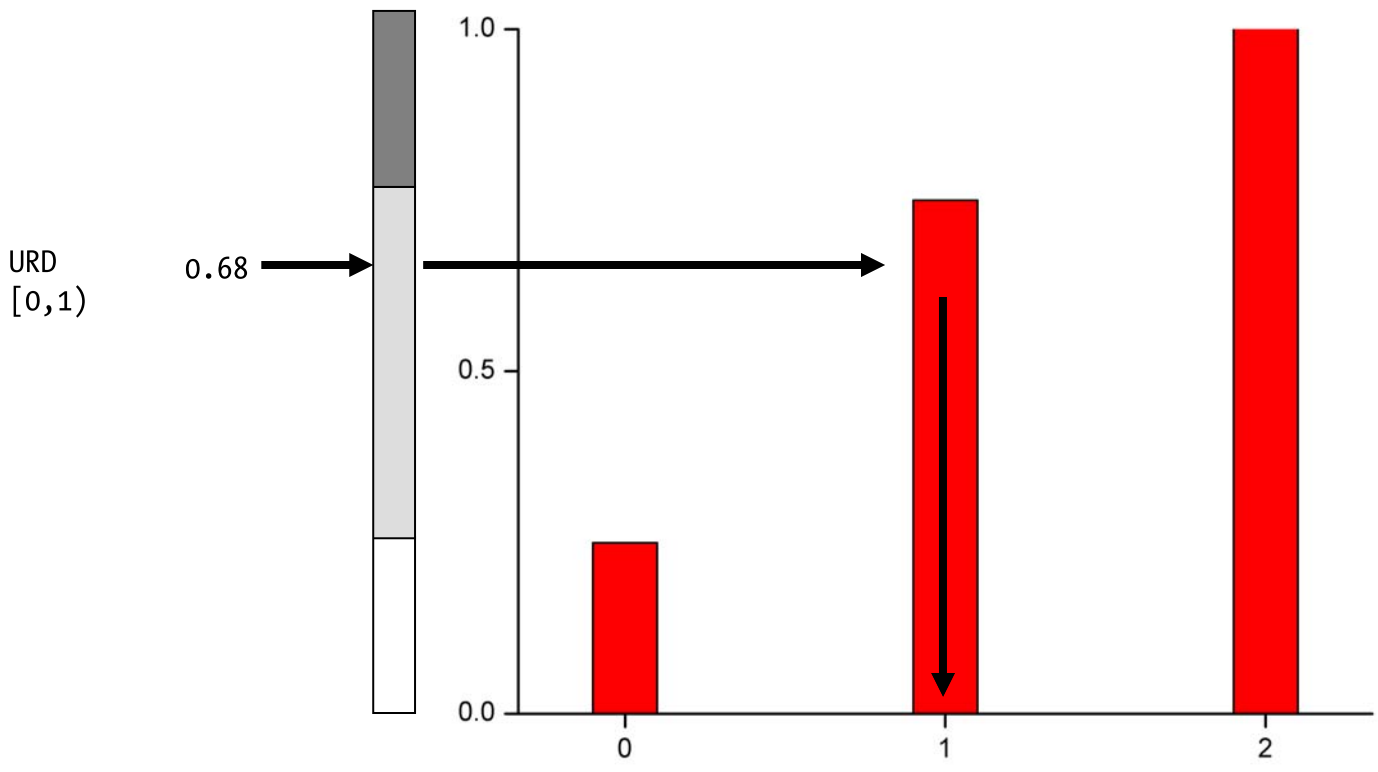


number of heads, nh, in 2 tosses

generating numbers from arbitrary distributions

- let's use the coin example to generate random numbers from a non-uniform distribution
  - the distribution will correspond to the number of heads in two tosses of a coin
    - generate 0   25% of the time
    - generate 1   50% of the time
    - generate 2   25% of the time
- we can use the cumulative distribution and a uniform random number generator
  - each URD will be mapped onto an outcome that is distributed according to our probability distribution

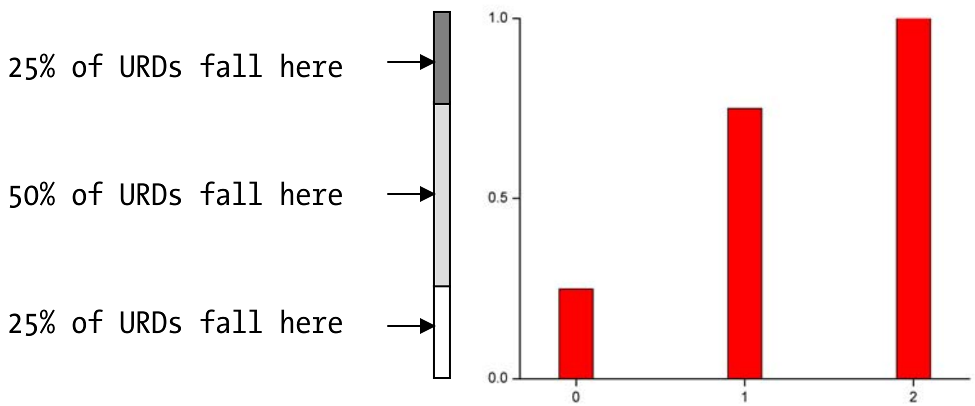
using cumulative distribution to generate random values



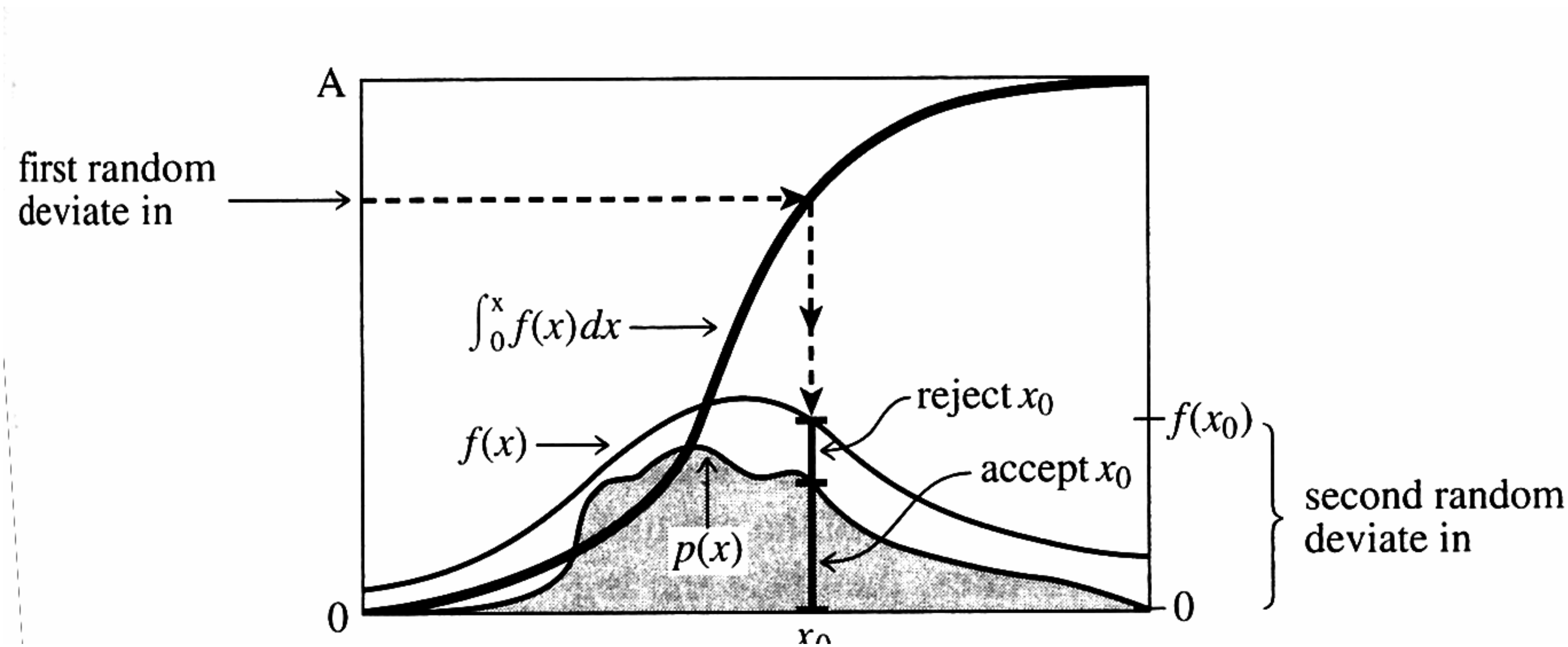
0.68 maps onto 1

using cumulative distribution to generate random values

```
# generate a uniform random deviate
my $urd = rand();
# define the cumulative distribution function
# c(0) = 0.25, c(1) = 0.75, c(2) = 1
my @c = (0.25,0,75,1);
# find the smallest i for which urd <= c(i)
for my $i (0..@c-1) {
  return $i if $urd <= $c[$i];
}
```



rejection method



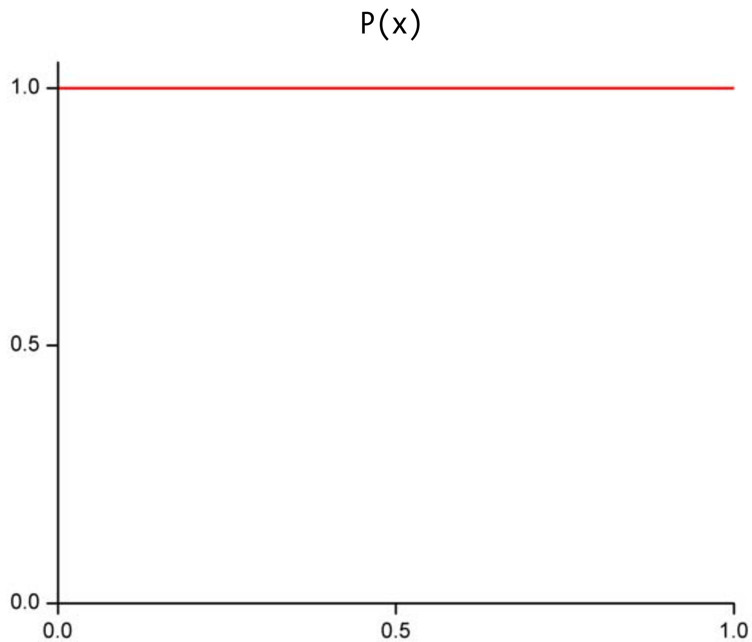
continuous random variables

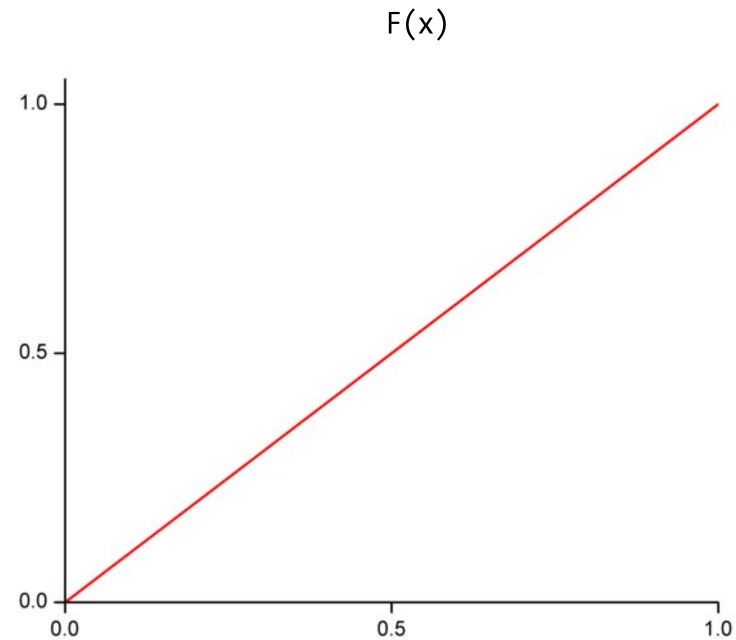
- not all random variables are finite
- toss of a coin or die is finite
- heights of individuals is not finite
  - height can be any real number in a practical range, e.g. 0 – 3 meters
  - number of different heights in this range is infinite
- the probability and cumulative distribution functions are replaced by continuous equivalents
  - sums are now integrals

$$P(a \leq x \leq b) = \int_a^b p(x)dx \qquad F(x) = P(x \leq b) = \int_{-\infty}^{b} p(x)dx$$

uniform distribution

- the uniform distribution is the distribution from which PRNGs sample their values





$$P(a \le x \le b) = \int_a^b dx$$

$$F(x) = P(x \le b) = \int_0^b dx$$

distributions

- bernoulli
- geometric
- binomial
- normal
- poisson
- exponential

bernoulli distribution

- an experiment in which there can be only two outcomes is a Bernoulli trial
  - typically labeled as success (value 1) or failure (value 0)
  - probability of success is p
  - probability of failure is 1-p=q

- E(X)=p   var(X)=p(1-p)

- to generate a Bernoulli variable, compare an URD to the success probability
  - return 1 if URD is smaller than success
  - return 0 otherwise

```
my $brd1 = rand() < $p;
# or equivalently
my $brd2 = rand() > $q;
```

geometric distribution

- given a Bernoulli trial with probability of success p, the geometric distribution describes the probability of obtaining a success (S) after exactly n failures (F)
  - n=0 : S
  - n=1 : FS
  - n=2 : FFS
  - n=3 : FFFS, etc

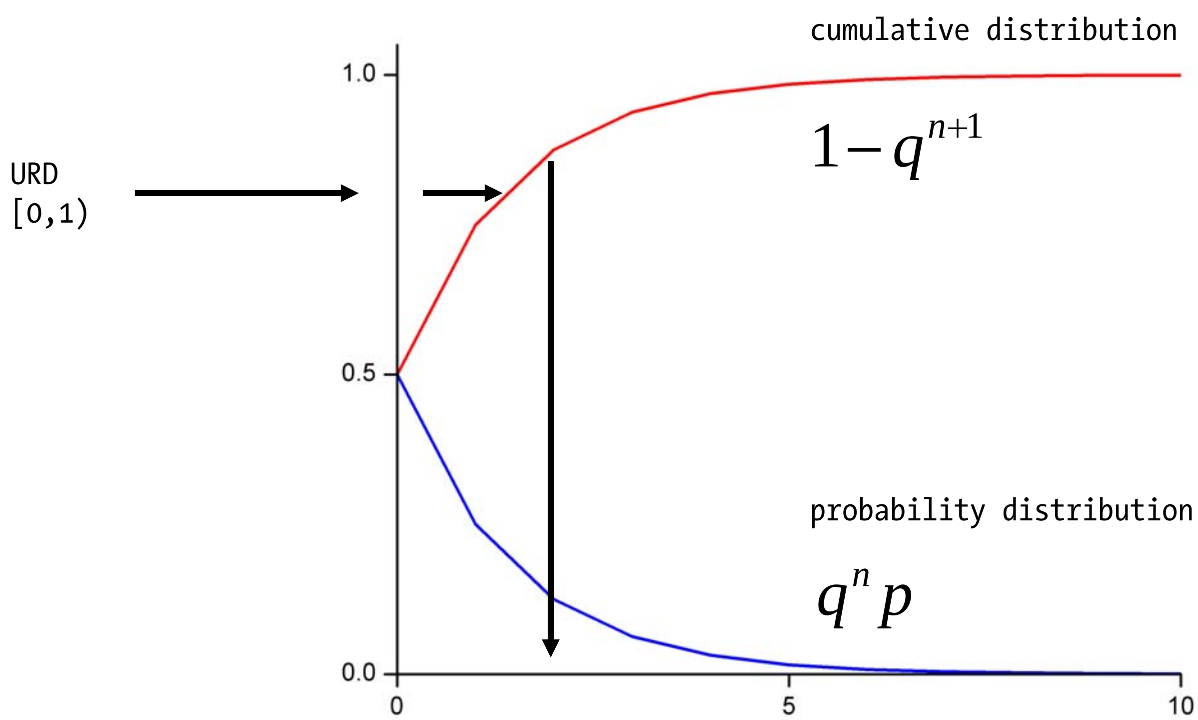
- $P(X=n)=(1-p)^n p$ $E(X) = 1/p$ $P(X \le n)=1-q^{n+1}$

- given a die, the probability or tossing a "1" is 1/6
  - the probability of having to toss the die 9 times before seeing a 1 (on the 10th toss) is

$$\left(1-\frac{1}{6}\right)^9 \frac{1}{6} = 0.032$$

generating geometric distribution

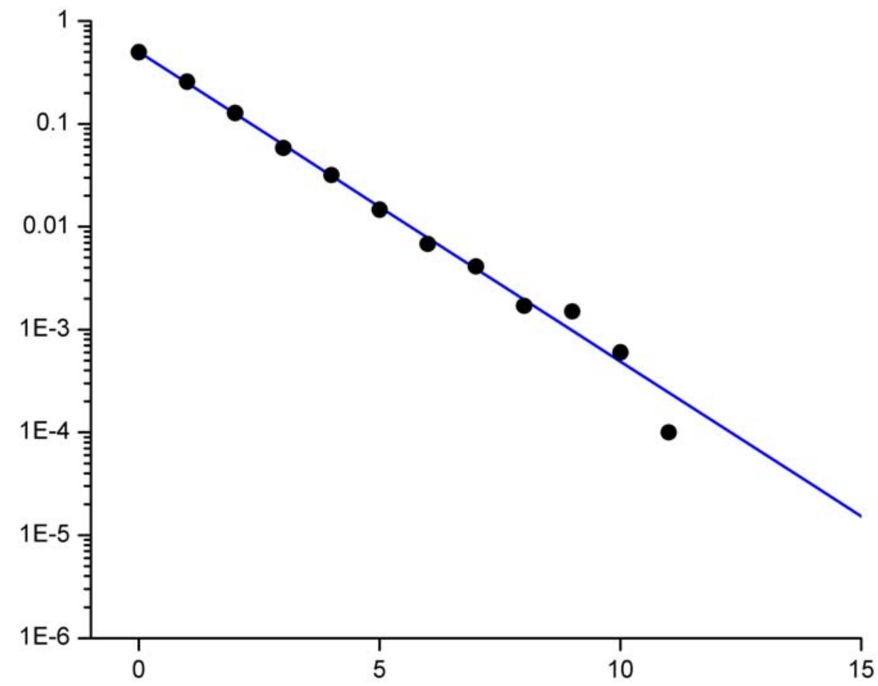
- transforming a uniform distribution to geometric distribution can be done via the cumulative form of the geometric distribution



generating geometric distribution

- I generated 10,000 values from the geometric distribution with p=q=0.5

```
# generate a uniform random deviate
my $urd = rand();
# walk along cumulative distribution until
# the URD is smaller
my $i = 0;
while( $urd > 1-$q**($i+1)) {
  $i++;
}
print $urd,$i;
```



binomial distribution

- the geometric distribution gives the probability of success after n failures, but...
- the binomial distribution gives the probability of k success after n trials in a Bernoulli process with success probability p

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

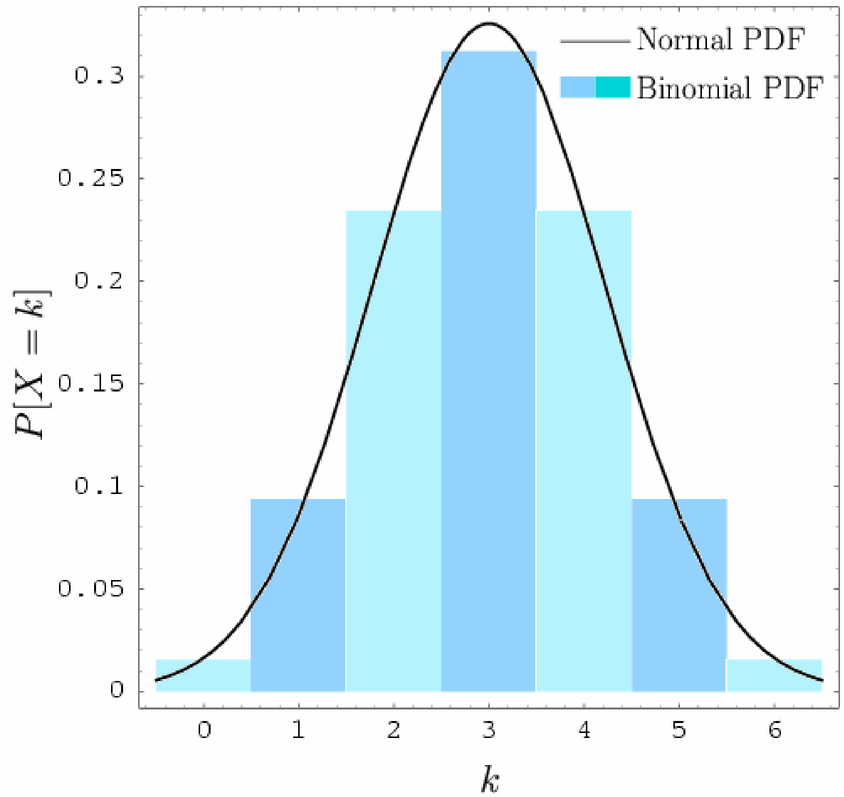
- $\binom{n}{k}$: number of ways k successes can appear in n trials
- p^k: probability of obtaining k successes
- $(1-p)^{n-k}$: probability of obtaining n-k failures

binomial distribution

- consider k=2 and n=3 and p=0.5
  - seek the probability of 2 successes out of 3 trials
  - there are three ways in which this can happen
    - SSF
    - SFS
    - FSS
    - the binomial coefficient for C(3,2)=3 multiples the probability $p^k q^{n-k}$ to correct for the fact that the outcome may manifest itself in more than one way
- E(X) = np   var(X) = npq
- consider a box of 12 lighbulbs - if the chance that any one bulb is broken is 0.01
  - 89% of the time there will be no broken bulbs, P(n=12,k=0,p=0.01)
  - 99.4% of the time there will be no more than one broken bulb, P(12,0,0.01)+P(12,1,0.01)
  - 99.98% of the time there will be no more than two broken bulbs

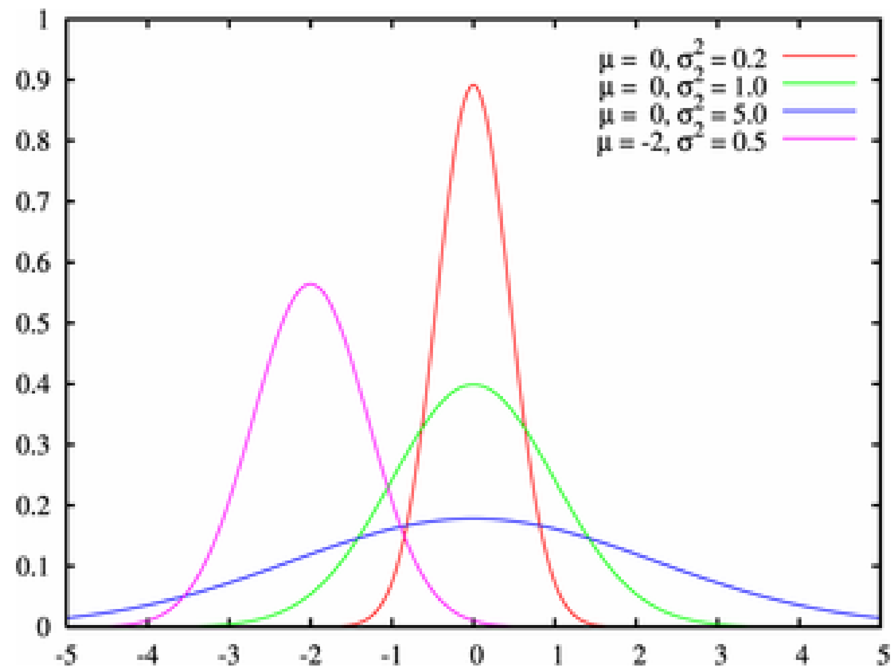
normal distribution

- the binomial distribution approaches the normal distribution when
  - n is very large
  - p is fixed
  - regime for np,nq>5 and
- plot at right shows normal and binomial distributions for n=6 and p=0.5



normal distribution

$$\frac{1}{\sigma\sqrt{2\pi}}\,\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



normal distribution

- the normal distribution is extremely common in physical and psychological sciences
  - underlying causes of phenomena are unknown, but small effects are added into an observable score
- central limit theorem popularizes the normal distribution
  - take a collection of random values from the same distribution which has a given mean and standard deviation
  - compute the average of these values
  - if you repeat this experiment, the average will be normally distributed

poisson distribution

- the binomial distribution is approximated by the poisson distribution when
  - n is very large
  - p is very small
  - $\lambda = np$

- Poisson distribution describes the number of events in unit time, if the events occur at a fixed rate

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

- consider cars arriving at a traffic light at the rate of 1 per minute. In a 10 minute period, you expect 10 cars (this is the average and the value of λ above)
  - what is the probability that you'll see only 5 cars in this time period (10 minutes)?

$$P(X = 5) = \frac{e^{-10}10^5}{5!} = 0.038$$

poisson distribution

- if λ is taken to be a rate, per unit time, then Poisson gives the probability of a given number of occurrences before time t

$$P(N_t = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

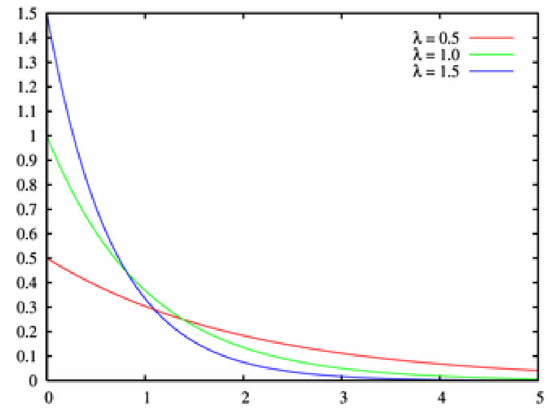
- in the example before, the rate was λ=1 car per minute and the probability to calculate was waiting t=10 minutes and seeing only 5 cars

- many other occurrences of Poisson exist
  - number of dead squirrels per unit distance of highway
  - number of spelling mistakes on a page
  - number of hits to a web server per minute
  - number of randomly selected points in a volume of space

exponential distribution

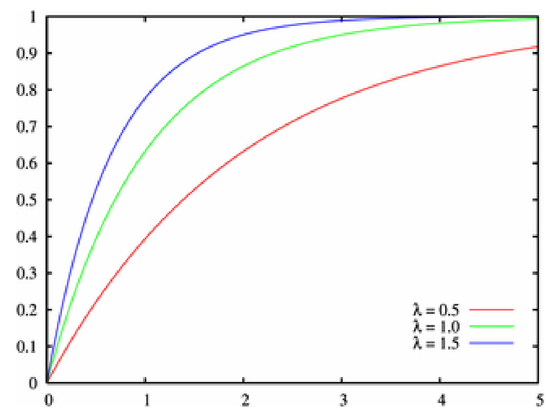
- this is a continuous version of the geometric distribution we've already seen
  - geometric distribution gave the probability of seeing a success after n failures of a Bernoulli trial

- exponential distribution gives the probability of having to wait a given amount of time before an event happens
  - before your next phone call
  - before your next email arrives
  - before your next car accident



$$P(X = x) = \lambda e^{-\lambda x}$$

$$P(X \le x) = 1 - e^{-\lambda x}$$

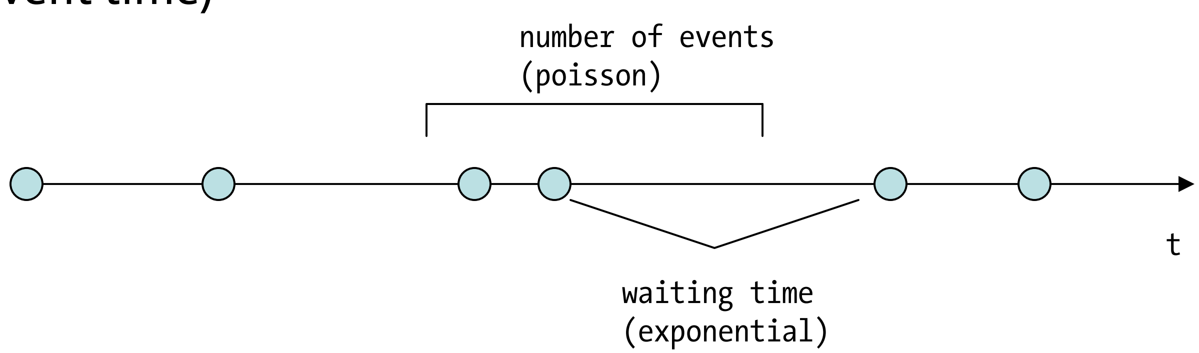


waiting for him/her to call

- suppose your boy/girl-friend calls you at a rate of once per 12 hour period (λ=1/12), what is the probability that you'll have to wait more than 24 hours before their call?

- the event (phone call) happens at a rate of λ=1/12
  - Poisson would tell us how many calls we can expect in a given time
  - e.g. probability of receiving 2 calls in 1 hour, 2 calls in 2 hours, 3 calls in 10 hours etc

- exponential distribution tells us how long we need to wait before the next event (inter-event time)

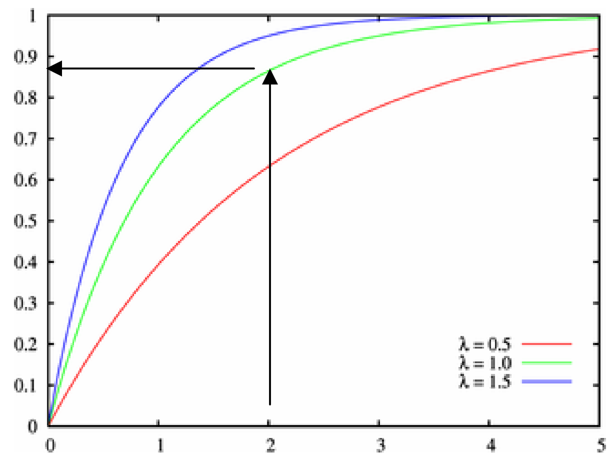


waiting for him/her to call

- the cumulative form of the exponential distribution gives us the probability that the waiting time is less than a certain value
  - p = probability of waiting more than 24 hours
  - 1 – p = probability of waiting less than 24 hours

$$P(X \leq x) = 1 - e^{-\lambda x} = 1 - e^{-\left(\frac{1}{12}\right)(24)} = 1 - e^{-2} = 0.86$$

- thus the probability of waiting more than 24 hours without a call is 0.14.



Math::CDF

- this module gives both probability and cumulative distributions
  - cumulative probability PXXXX
  - quantile probability QXXXX

```
pbeta(), qbeta() [Beta Distribution]
pchisq(), qchisq() [Chi-square Distribution]
pf(), qf() [F Distribution]
pgamma(), qgamma() [Gamma Distribution]
pnorm(), qnorm() [Standard Normal Dist]
ppois(), qpois() [Poisson Distribution]
pt(), qt() [T-distribution]
pbinom() [Binomial Distribution]
pnbinom() [Negative Binomial Distribution]
```

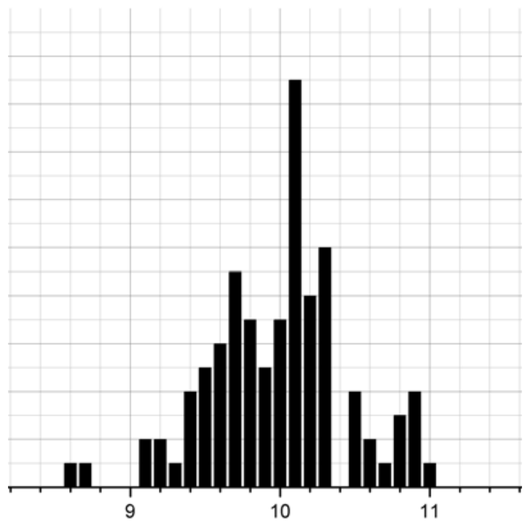
```
# -1.96 – value at which probability is 0.025 that (X-μ)/σ (X normally distributed) is smaller
qnorm(0.025)
```

Math::Random

- provides random values sampled from variety of distributions

```
random_beta
random_chi_square
random_exponential
random_f
random_gamma
random_multivariate_normal
random_multinomial
random_noncentral_chi_square
random_noncentral_f
random_normal
random_permutation
random_permuted_index
random_uniform
random_poisson
random_uniform_integer
random_negative_binomial
random_binomial
random_seed_from_phrase
random_get_seed
random_set_seed_from_phrase
random_set_seed
```

```
# generate 100 normally distributed random numbers
# with average 10 and stdev 0.5
random_normal(100, 10, 0.5)
```



4.1.2.2.2

Random Numbers and Distributions
Session 2

- lots of distributions exist
- search for “random” on CPAN