Unit III: Statistical distribution and Evaluation

Questions to ask…

- What is a random variable?
- What is a distribution?
- How do u classify distribution
- Where do ‘commonly-used’ distributions come from?
- What distribution does my data come from?
- Do I have to specify a distribution to analyse my data?

2.Ref:ctoolkit.com

Probability Distribution

- Discrete
  - Binomial distribution
  - Poisson Distribution
- Continuous

Binary Distribution

2 outcomes either Pass or Fail

Solve: Probability of 3 flips of a fair coins

P(x=0) =1/8

P(X=1)=3/8

P(X=2)=3/8

P(X=3)=1/8

Plot a graph of outcome and probability

Probability distribution for random variable X which is discreate

A fair coin is tossed twice. Let X be the number of heads that are observed.

Construct the probability distribution of X.

Find the probability that at least one head is observed.

HH, HT,TH,TT

Calculate P(X=0, X=1 and X=2)

"At least one head" is the event $X \geq 1$, which is the union of the mutually exclusive events $X = 1$ and $X = 2$.

- Probability Distributions

Binomial Distribution – Example:

- When sampling, we commonly want to accept a batch if there are (say) **1** or less defective in the sample, and reject it if there are **2** or more.
- In order to determine the probability of acceptance, the individual probabilities for **0** and **1** defectives are summed:
  - P (1 or less) = P(0) + P(1)
    = 0.358 + 0.377 = 0.735

- So the probability of rejection is:
  - 1 – 0.735 = 0.265

- Probability Distributions

Binomial Distribution:

- The probability of 'r' successes P(r) is given by the **binomial formula**:

$$P(r) = n!/(r!(n-r))! * p^r(1-p)^{n-r}$$

p: probability of success
n: number of independent trials
r: number of successes in the n trials

The binomial distribution is fully defined if we know both 'n' & 'p'

Find the probabil

If 2 dice are thrown , 36 outcomes are there

Doublets: same number is obtained on both dice

Possible doublets: 1,1 2,2   3,3   4,4   5,5,   6,6

Probability of getting a doublet

$$\frac{6}{36} = \frac{1}{6}$$

Probability of not getting a doublet

$$1 - \frac{1}{6} = \frac{5}{6}$$

2 dice problem

Dice 2 \ Dice 1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Find the probability of getting at least 9

X>=9

Find the probability that X is even

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

This table has 36 sums. mathlibra

$$P(2) = \frac{1}{36}; P(3) = \frac{2}{36} = \frac{1}{18}; P(4) = \frac{3}{36} = \frac{1}{12};$$

$$P(5) = \frac{4}{36} = \frac{1}{9}; P(6) = \frac{5}{36}; P(7) = \frac{6}{36} = 1/6;$$

$$P(8) = \frac{5}{36}; P(9) = \frac{4}{36} = \frac{1}{9}; P(10) = \frac{3}{36} = \frac{1}{12};$$

$$P(11) = \frac{2}{36} = \frac{1}{18}; P(12) = \frac{1}{36}$$

A service organization in a large town organizes a raffle each month.
One thousand raffle tickets are sold for $1 each. Each has an equal chance of winning.
First prize is $300, second prize is $200, and third prize is $100.
Let *X* denote the net gain from the purchase of one ticket.

1. Construct the probability distribution of *X*.
2. Find the probability of winning any money in the purchase of one ticket.
3. Find the expected value of *X*, and interpret its meaning.

Solution:

1. If a ticket is selected as the first prize winner, the net gain to the purchaser is the $300 prize
2. less the $1 that was paid for the ticket, hence $X = 300 - 1 = 299$.
3. There is one such ticket, so $P(299) = 0.001$.

4. Applying the same "income minus outgo" principle to the second and third prize winners and
5. to the 997 losing tickets yields the probability distribution:

$xP(x)2990.0011990.001990.001−10.997$

1.
2.
3.

1.

Calculating μ and σ.

You are playing a game of chance in which four cards are drawn from a standard deck of 52 cards. You guess the suit of each card before it is drawn. The cards are replaced in the deck on each draw. You pay \$1 to play. If you guess the right suit every time, you get your money back and \$256. What is your expected profit of playing the game over the long term?

Let X = the amount of money you profit. The x-values are –\$1 and \$256.

The probability of guessing the right suit each time is

$(1/4)(1/4)(1/4)(1/4)=1/256=0.0039$

The probability of losing is

$1-1/256=255256=0.9961$

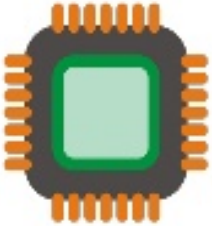
$(0.0039)256 + (0.9961)(-1) = 0.9984 + (-0.9961)$

$= 0.0023$ or 0.23 cents.

- Probability Distributions

Binomial Distribution – Other Examples:

- What is the probability of obtaining exactly **2** heads in the **5** tosses?
- What is the probability that in a random sample of **10** cans there are exactly **3** defective units, knowing that on average there is a **5%** defective product?
- What is the probability that a random sample of **4** units will have exactly **1** unit is defective, knowing that that process will produce **2%** defective units on average?



- Probability Distributions

Binomial Distribution – Example:

- In a sample of **20** drawn from a batch which is **5%** defective, what is the probability of getting exactly **3** defective items?

$$P(r) = n!/(r!(n-r))! * p^r(1-p)^{n-r}$$

$$P(r) = 20!/(3!(20 - 3)!) * 0.05^3(1 - 0.05)^{20-3}$$

$$P(r) = 1{,}140 * 0.05^3 * (0.95)^{17} = \mathbf{0.060}$$

- Probability Distributions

Poisson Distribution:

- It is not always appropriate to classify the outcome of a test simply as pass or fail.
- Sometimes, we have to count the number of defects where there may be several defects in a single item.
- The **Poisson Distribution** is a discrete probability distribution that specifies the probability of a certain number of occurrences over a specified interval.
  - Such as time or any other type of measurements.

Defects
14
8
11
2
13
17
11
9
12

- Probability Distributions

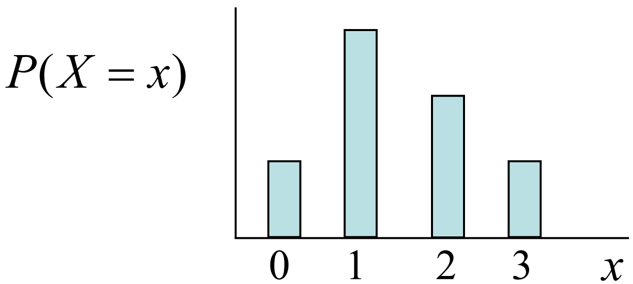
Poisson Distribution:

- This test has a wide range of applications, such as:
  - Counting the number of defects found in a single product.
  - Counting the number of accidents per year in a factory.
  - Counting the number of failures per month for a specific equipment.
  - Counting the number of incoming calls per day to an emergency call center.
  - Counting the number of customers who will walk into a store during the holidays.

What is a distribution?

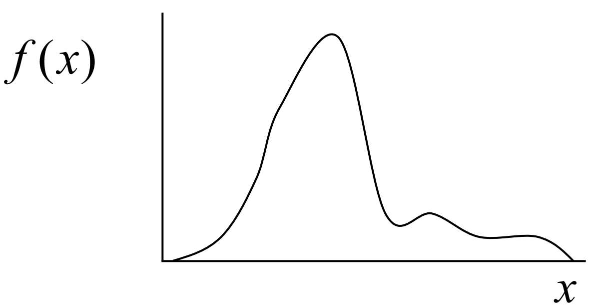
- A distribution characterises the probability (mass) associated with each possible outcome of a stochastic process

- Distributions of discrete data characterised by **probability mass functions**



$$\sum_{x} P(X=x) = 1$$

- Distributions of continuous data are characterised by **probability density functions** (pdf)



$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- For RVs that map to the integers or the real numbers, the **cumulative density function** (cdf) is a useful alternative representation

Some notation conventions

- Instances of random variables (RVs) are usually written in uppercase
- Values associated with RVs are usually written in lowercase
- pdfs are often written as $f(x)$
- cdfs are often written as $F(x)$
- Parameters are often defined as θ
- Hence

$P(X_i = x \mid n, \theta)$	Probability that the ith random variable takes value x given sample size n and parameter(s) θ
$f(x \mid \theta)$	The probability density associated with outcome x given some parameter(s) θ

Expectations and variances

- Suppose we took a large sample from a particular distribution, we might want to summarise something about what observations look like 'on average' and how much variability there is

- The **expectation** of a distribution is the average value of a random variable over a large number of samples

$$E(X) = \sum_x xP(X = x) \quad \text{or} \quad \int xf(x)dx$$

- The **variance** of a distribution is the average squared difference between randomly sampled observations and the expected value

$$Var(X) = \sum_x (x - E(x))^2 P(X = x) \quad \text{or} \quad \int (x - E(x))^2 f(x)dx$$

A men’s soccer team plays soccer zero, one, or two days

A men’s soccer team plays soccer zero, one, or two days a week.

- The probability that they play zero days is 0.2,
- the probability that they play one day is 0.5,
- the probability that they play two days is 0.3.

Find the long-term average or expected value, μ, of the number of days per week the men’s soccer team plays soccer.

Solution

let the random variable X = the number of days the men's soccer team plays soccer per week.

X takes on the values 0, 1, 2.

Construct a PDF table adding a column $x \cdot P(x)$.

In this column, you will multiply each x value by its probability.

Expected Value Table. This table is called an expected value table. The table helps you calculate the expected value or long-term average.

x	P(x)	x.P(x)
0	0.2	0
1	0.5	0.5
2	0.3	0.6

02/Add the last column $x \cdot P(x)$ to find the long term average or expected value:
$(0)(0.2) + (1)(0.5) + (2)(0.3) = 0 + 0.5 + 0.6 = 1.1$

Inference

The expected value is 1.1.

The men's soccer team would, on the average, expect to play soccer 1.1 days per week.

The number 1.1 is the long-term average or expected value if the men's soccer team plays soccer week after week after week.

Hence $\mu = 1.1$.

Exercise

Find the expected value of the number of times a newborn baby's crying wakes its mother after midnight.

The expected value is the expected number of times per week a newborn baby's crying wakes its mother after midnight.

Calculate the standard deviation of the variable as well.

You expect a newborn to wake its mother after midnight 2.1 times per week, on the average.

x	P(X)	X.P(X)	(X-U)^2.P(X)
0	2/50		
1	11/50		
2	23/50		
3	09/50		
4	04/50		
5	01/50		

Add the values in the third column of the table to find the expected value of X:

μ = Expected Value = 105/50= 2.1

Use μ to complete the table.

The fourth column of this table will provide the values you need to calculate the standard deviation. For each value x, multiply the square of its deviation by its probability. (Each deviation has the format $x - \mu$).

Add the values in the fourth column of the table:

0.1764 + 0.2662 + 0.0046 + 0.1458 + 0.2888 + 0.1682 = 1.05

The standard deviation of X is the square root of this sum: $\sigma = \sqrt{1.05} \simeq 1.0247$

iid

- In most cases, we assume that the random variables we observe are **independent and identically distributed**
- The **iid** assumption allows us to make all sorts of statements both about what we expect to see and how much variation to expect
- Suppose *X*, *Y* and *Z* are iid random variables and *a* and *b* are constants

$$E(X+Y+Z) = E(X) + E(Y) + E(Z) = 3E(X)$$

$$\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) = 3\mathrm{Var}(X)$$

$$E(aX+b) = aE(X) + b$$

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

$$\mathrm{Var}\left(\tfrac{1}{n}\sum_i X_i\right) = \tfrac{1}{n}\mathrm{Var}(X)$$

Where do 'commonly-used' distributions come from?

- At the core of much statistical theory and methodology lie a series of key distributions (e.g. Normal, Poisson, Exponential, etc.)
- These distributions are closely related to each other and can be 'derived' as the limit of simple stochastic processes when the random variable can be counted or measured
- In many settings, more complex distributions are constructed from these 'simple' distributions
  - Ratios: E.g. Beta, Cauchy
  - Compound: E.g. Geometric, Beta
  - Mixture models

An aside on Chebyshev's inequality

- Let X be a random variable with mean μ and variance σ^2

- Chebyshev's inequality states that for any $t > 0$

$$P(|X-\mu| > t) \leq \frac{\sigma^2}{t^2}$$

- This allows us to make statements about **any** distribution with finite variance
  - The probability that a value lies more than 2 standard deviations from the mean is less than or equal to 0.25

- Note that this is an upper bound. In reality, the distribution might be considerably tighter
  - E.g. for the normal distribution the probability is 0.046, for the exponential distribution the probability is 0.05

The simplest model

- Bernoulli trials
  - Outcomes that can take only two values: (0 and 1) with probabilities θ and 1 - θ respectively. E.g. coin flipping, indicator functions

- The **likelihood** function calculates the probability of the data

$$P(\mathbf{X} \mid \theta) = \prod_i P(X = x_i \mid \theta) = \theta^k (1-\theta)^{n-k}$$

- What is the probability of observing the sequence (if θ = 0.5)
  - 01001101100111101001?
  - 11111111111000000000?

- Are they both equally probable?

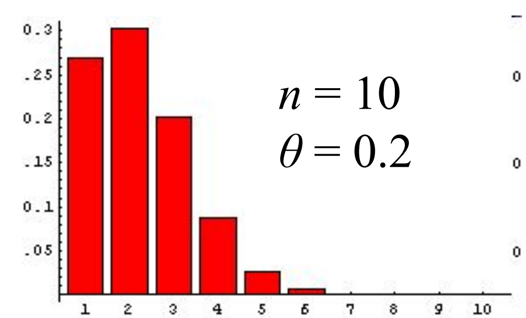
The binomial distribution

- Often, we don't care about the exact order in which successes occurred. We might therefore want to ask about the probability of *k* successes in *n* trials. This is given by the **binomial** distribution

- For example, the probability of exactly 3 heads in 4 coins tosses =
  - P(HHHT)+P(HHTH)+P(HTHH)+P(THHH)
  - Each order has the same Bernoulli probability = $(1/2)^4$
  - There are 4 choose 3 = 4 orders

- Generally, if the probability of success is θ, the probability of *k* successes in *n* trials

$$P(k \mid n,\theta) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$



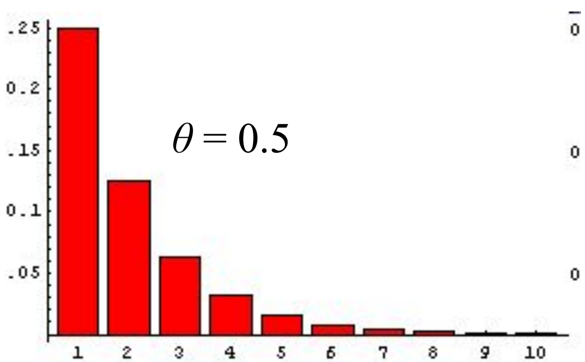
- The expected number of successes is $n\theta$ and the variance is $n\theta(1-\theta)$

The geometric distribution

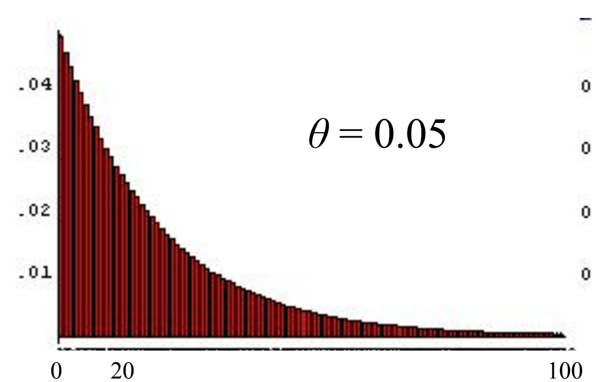
- Bernoulli trials have a **memory-less** property
  - The probability of success ($X = 1$) next time is independent of the number of successes in the preceding trials
- The number of trials between subsequent successes takes a **geometric** distribution
  - The probability that the first success occurs at the k^{th} trial

$$P(k \mid \theta) = \theta(1-\theta)^{k-1}$$

- You can expect to wait an average of $1/\theta$ trials for a success, but the variance is



$$\mathrm{Var}(k) = \frac{1-\theta}{\theta^2}$$



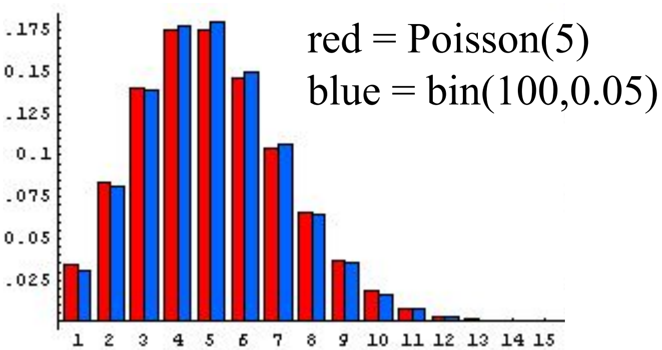
The Poisson distribution

- The Poisson distribution is often used to model 'rare events'

- It can be derived in two ways
  - The limit of the Binomial distribution as $\theta \to 0$ and $n \to \infty$ $(n\theta = \mu)$
  - The number of events observed in a given time for a Poisson process (more later)

- It is parameterised by the expected number of events = μ
  - The probability of k events is

$$P(k \mid \mu) = \frac{e^{-\mu} \mu^k}{k!}$$



- The expected number of events is μ, and the variance is also μ

- For large μ, the Poisson is well approximated by the normal distribution

Other distributions for discrete data

- Negative binomial distribution
  - The distribution of the number of Bernoulli trials until the kth success
  - If the probability of success is θ, the probability of taking m trials until the kth success is

$$P(m \mid k, \theta) = \binom{m-1}{k-1} \theta^k (1-\theta)^{m-k}$$

  - (like a binomial, but conditioning on the last event being a success)

- Hypergeometric distribution
  - Arises when sampling *without* replacement
  - Also arises from Hoppe Urn-model situations (population genetics)

Going continuous

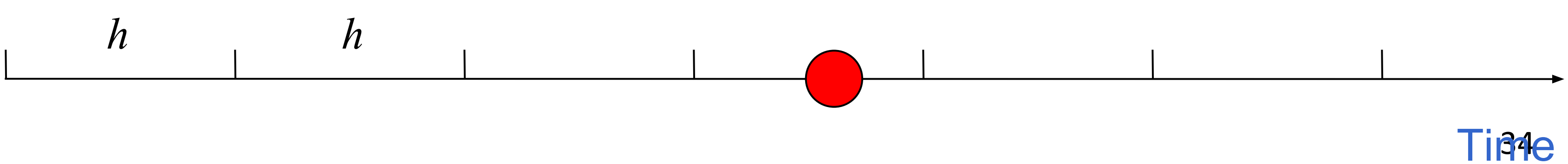
- In many situations while the outcome space of random variables may really be discrete (or at least measurably discrete), it is convenient to allow the random variables to be continuously distributed

- For example, the distribution of height in mm is actually discrete, but is well approximated by a continuous distribution (e.g. normal)

- Commonly-used continuous distributions arise as the limit of discrete processes

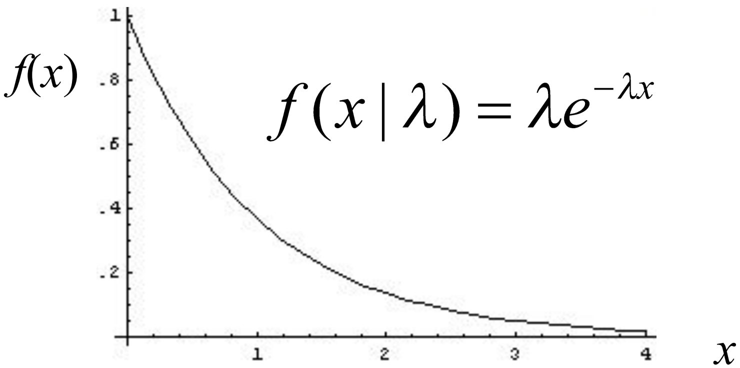
The Poisson process

- Consider a process when in every unit of time some event might occur
- E.g. every generation there is some chance of a gene mutating (with probability of approx 1 in 100,000)
- The probability of exactly one change in a sufficiently small interval $h \equiv 1/n$ is $P = vh \equiv v/n$, where P is the probability of one change and n is the number of trials.
- The probability of two or more changes in a sufficiently small interval h is essentially 0
- In the limit of the number of trials becoming large the total number of events (e.g. mutations) follows the Poisson distribution



The exponential distribution

- In the Poisson process, the time between successive events follows an **exponential** distribution
  – This is the continuous analogue of the geometric distribution
  – It is memory-less. i.e. $f(x + t \mid X > t) = f(x)$



$$f(x \mid \lambda) = \lambda e^{-\lambda x}$$

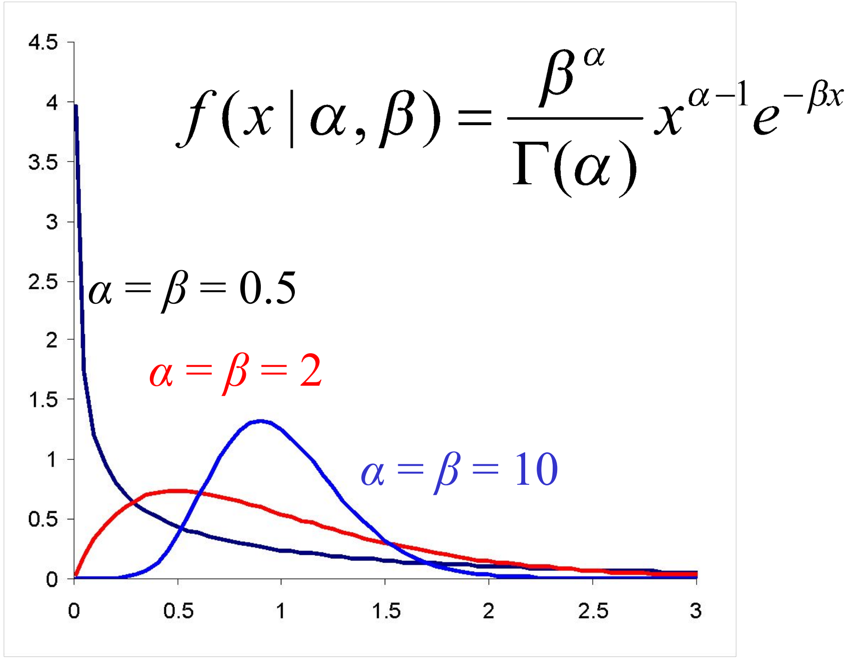
$$E(x) = 1/\lambda$$

$$\mathrm{Var}(x) = 1/\lambda^2$$

The gamma distribution

- The **gamma** distribution arises naturally as the distribution of a series of iid random exponential variables

$$X \sim Exp(\lambda)$$
$$S = X_1 + X_2 + \cdots + X_n$$
$$S \sim \mathrm{Gamma}(n, \lambda)$$



- The gamma distribution has expectation α/β and variance α/β^2

- More generally, α need not be an integer (for example, the Chi-square distribution with one degree of freedom is a $\mathrm{Gamma}(½, ½)$ distribution)

The beta distribution

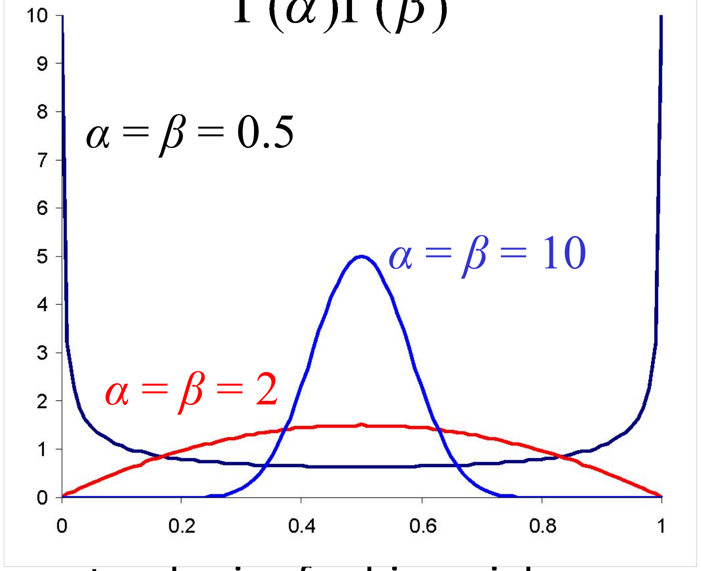
- The **beta** distribution models random variables that take the value [0,1]
- It arises naturally as the proportional ratio of two gamma distributed random variables

$$X \sim \mathrm{Gamma}(\alpha_1, \theta)$$

$$Y \sim \mathrm{Gamma}(\alpha_2, \theta)$$

$$\frac{X}{X+Y} \sim \mathrm{Beta}(\alpha_1, \alpha_2)$$

$$f(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

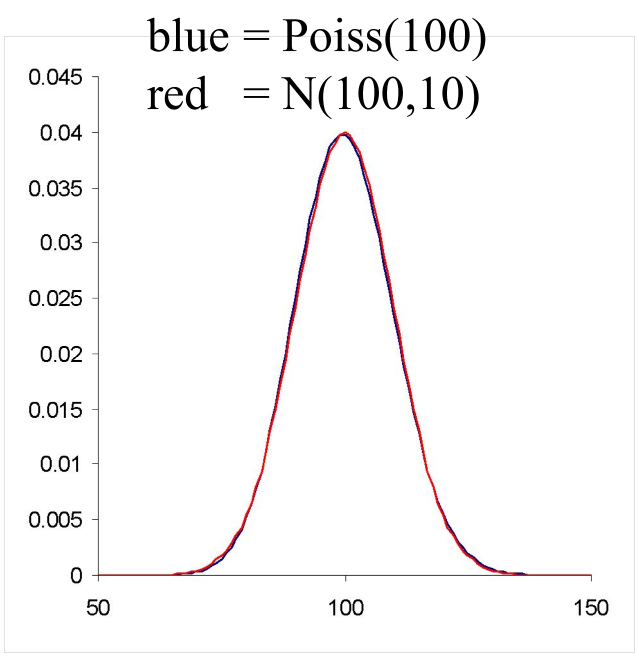


- The expectation is $\alpha/(\alpha + \beta)$
- In Bayesian statistics, the beta distribution is the natural prior for binomial proportions (beta-binomial)
  - The Dirichlet distribution generalises the beta to more than 2 proportions

The normal distribution

- As you will see in the next lecture, the **normal** distribution is related to most distributions through the **central limit theorem**

- The normal distribution naturally describes variation of characters influenced by a large number of processes (height, weight) or the distribution of large numbers of events (e.g. limit of binomial with large *np* or Poisson with large μ)



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

The exponential family of distributions

- Many of the distributions covered (e.g. normal, binomial, Poisson, gamma) belong to the **exponential family** of probability distributions

- a *k*-parameter member of the family has a density or frequency function of the form

$$f(x;\theta) = \exp\left[\sum_{i=1}^{k} c_i(\theta)T_i(x) + d(\theta) + S(x)\right]$$

- E.g. the Bernoulli distribution (*x* = 0 or 1) is

$$P(X = x) = \theta^x (1-\theta)^{1-x} = \exp\left[x \ln\left(\frac{\theta}{1-\theta}\right) + \ln(1-\theta)\right]$$

- Such distributions have the useful property that simple functions of the data, *T*(*x*), contain all the information about model parameter
  - E.g. in Bernoulli case $T(x) = x$

What distribution does my data come from?

- When faced with a series of measurements the first step in statistical analysis is to gain an understanding of the distribution of the data

- We would like to
  - Assess what distribution might be appropriate to model to data
  - Estimate parameters of the distribution
  - Check to see whether the distribution really does fit

- We might refer to the distribution + parameters as being a **model** for the data

Which model?

- Step 1: Plot the distribution of the random variables (e.g. a histogram)
- Step 2: Choose a candidate distribution
- Step 3: Estimate the parameters of the candidate distribution (e.g. by method of moments)
- Step 4: Compare the empirical distribution to that observed (e.g. using a QQplot)
- Step 5: Test model fit
- Step 6: Refine, transform, repeat

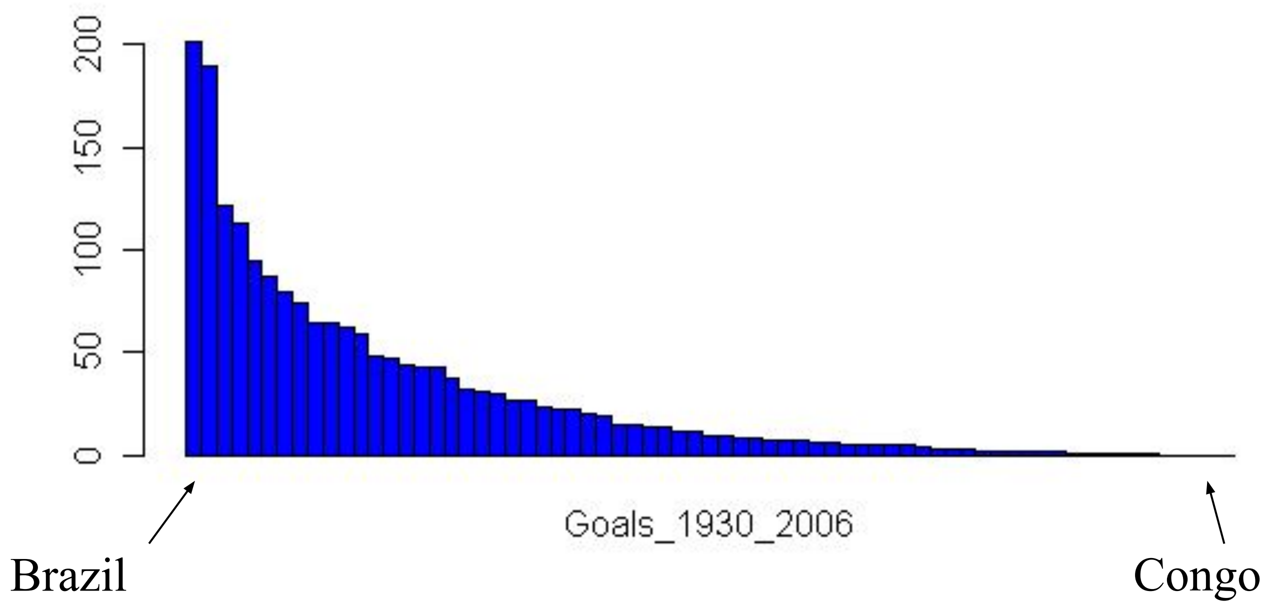
Method of moments

- We wish to compare observed data to a possible model
- We should choose the model parameters such that they match the data
- A simple approach is to match the sample moments to those of the model
  - Start with the lowest moments

Model	**Parameters**	**Matching**
Poisson	μ	sample mean = μ
Binomial	p	sample successes = np
Exponential	λ	waiting time = λ
Gamma	α, β	sample mean = α/β, sample variance = α/β^2

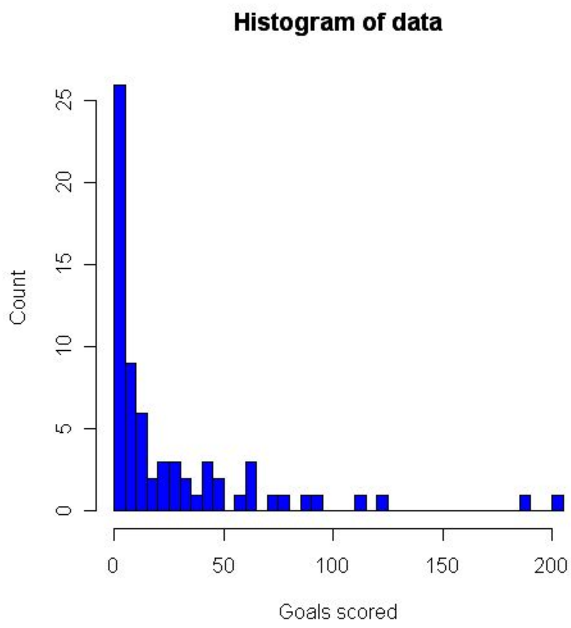
Example: world cup goals 1930-2006

- Total number of goals scored by country over period

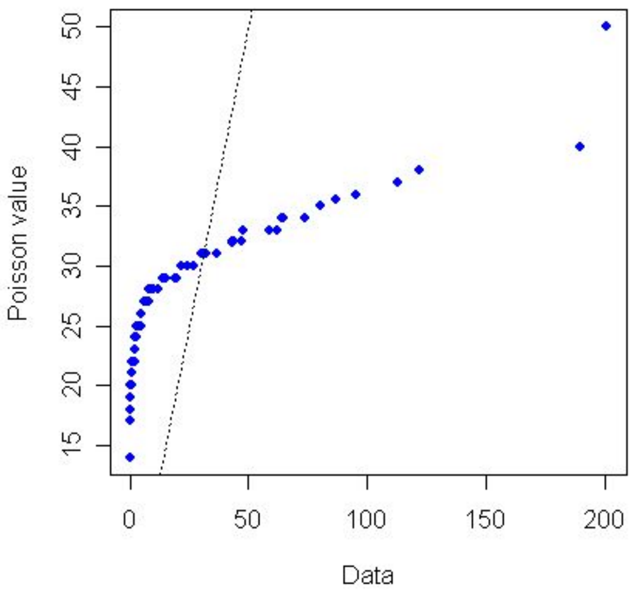


Fitting a model

- The data are discrete – perhaps a Poisson distribution is appropriate
- To fit a Poisson, we just estimate the parameter from the mean (28.0)
- Compare the distributions with histograms and QQplots

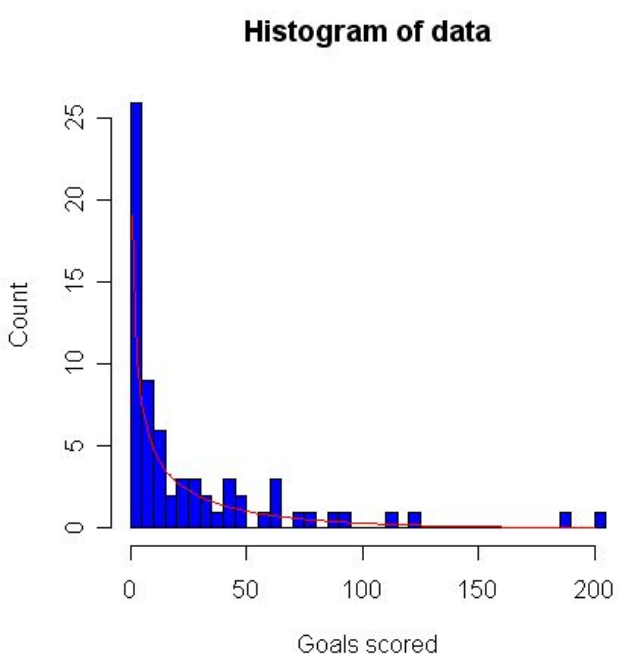


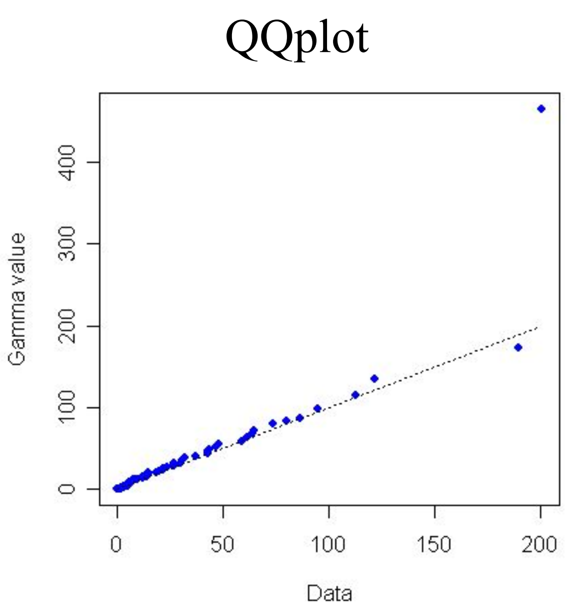
QQplot



A better model

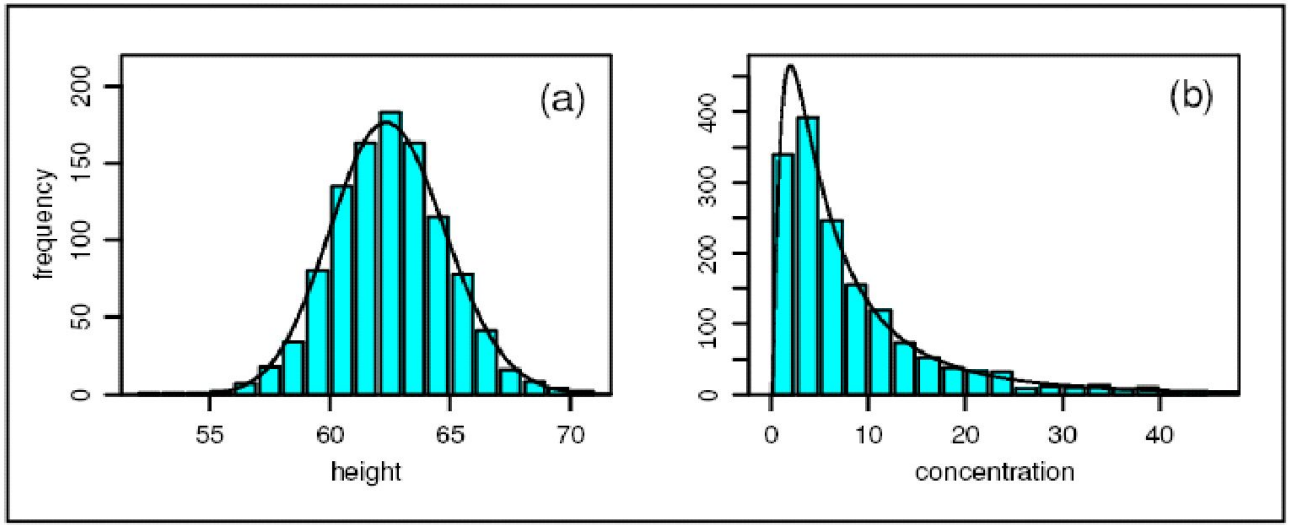
- The number of goals scored is over-dispersed relative to the Poisson
- We could try an exponential? This too is under-dispersed.
- We can generalise the exponential to the gamma distribution. We estimate (by moments) the shape parameter to be 0.47 (approximately the Chi-squared distribution!)





What do I do if I can't find a model that fits?

- Sometimes data needs to be **transformed** before it fits an appropriate distribution
  - E.g. log transformations, power transformations



Female height in inches — Concentration of HMF in honey

Limpert et al (2001). BioScience 51: 341

- Also the removal of (a few!) outliers is a common (and justifiable) approach

Testing model fit

- A QQplot provides a visual inspection of model fit. However, we might also wish to ask whether we can reject the hypothesis that the model is an accurate description of the data

- Testing model fit is a special case of **hypothesis testing**

- Briefly, specify some statistic of the data that is sensitive to model fit and hasn't been used directly to estimate parameters (e.g. location of quantiles) and compare observed data to repeated simulations from distribution

- It is worth noting that a model may be wrong (all models are wrong) but still useful.

Do I have to specify a distribution to analyse my data?

- For some situations in statistical inference it is possible to make inferences without specifying the distribution that data has been drawn from
- Such approaches are called **nonparametric**
- Some examples of nonparametric approaches include
  - Sign tests
  - Rank-based tests
  - Bootstrap techniques
  - Bayesian nonparametrics
- They are typically more robust than parametric approaches, but have lower power
- It is important to stress that these methods are not ‘parameter-free’ – rather they are not tied to specific distributions

Limit theorems and their applications

Questions

- What happens to our inferences as we collect more and more data?
- How can we make statements about our certainty (or uncertainty) in parameter estimates?
- What do the extreme values look like?

Things can only get better - the law of large numbers

- Suppose we have a series of iid samples from a distribution that has a mean μ

$$S_n = X_1 + X_2 + X_3 + \ldots + X_n$$

- The weak **law of large numbers** states that as $n \to \infty$ and for any ε

$$\Pr\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0$$

- The result follows from application of Chebyshev's inequality to the variance of the sample mean

$$\mathrm{Var}\left(\frac{S_n}{n} - \mu \right) = \frac{\sigma^2}{n}$$

Using the law of large numbers

- Monte Carlo integration is widely used in modern statistics where analytical expressions for quantities of interest cannot be obtained

- Suppose we wish to evaluate

$$I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-x^2/2} dx$$

- We can estimate the integral by drawing N pseudorandom $U[0,1]$ numbers

$$I(f) \approx \frac{1}{\sqrt{2\pi}} \tfrac{1}{N} \sum_{i=1}^{N} e^{-X_i^2/2}$$

- More generally, the law of large numbers tells us that any distribution moment (or function of the distribution) can be estimated from the sample

Convergence in distribution

- Suppose that $F_1, F_2, \ldots$ is a sequence of cumulative distribution functions corresponding to random variables $X_1, X_2, \ldots$, and that F is a distribution function corresponding to a random variable X

- X_n converges in distribution to X if (for every point at which F is continuous)

$$\lim_{n\to\infty} F_n(x) = F(x)$$

- A simple example is that the empirical CDF obtained from the sample converges in distribution to the distribution CDF
  - This provides the justification for the nonparametric bootstrap (Efron)

The Bootstrap method of resampling

- Suppose we have *n* observations from a distribution we do not wish to attempt to parameterise. We wish to know the mean of the distribution

- We would like to know something about how good our estimate of some function, e.g. the mean, is from this sample

- We can estimate the sampling distribution of the function simply by repeatedly re-sampling *n* observations from our data set with replacement

- (This will tend to have slow convergence for heavy-tailed distributions)

The central limit theorem

- Suppose we have a series of iid samples from a distribution that has a mean μ and standard deviation σ

$$S_n = X_1 + X_2 + X_3 + \ldots + X_n$$

- The **central limit theorem** states that as $n \to \infty$, the scaled sample mean converges in distribution to the standard normal distribution

Distribution mean

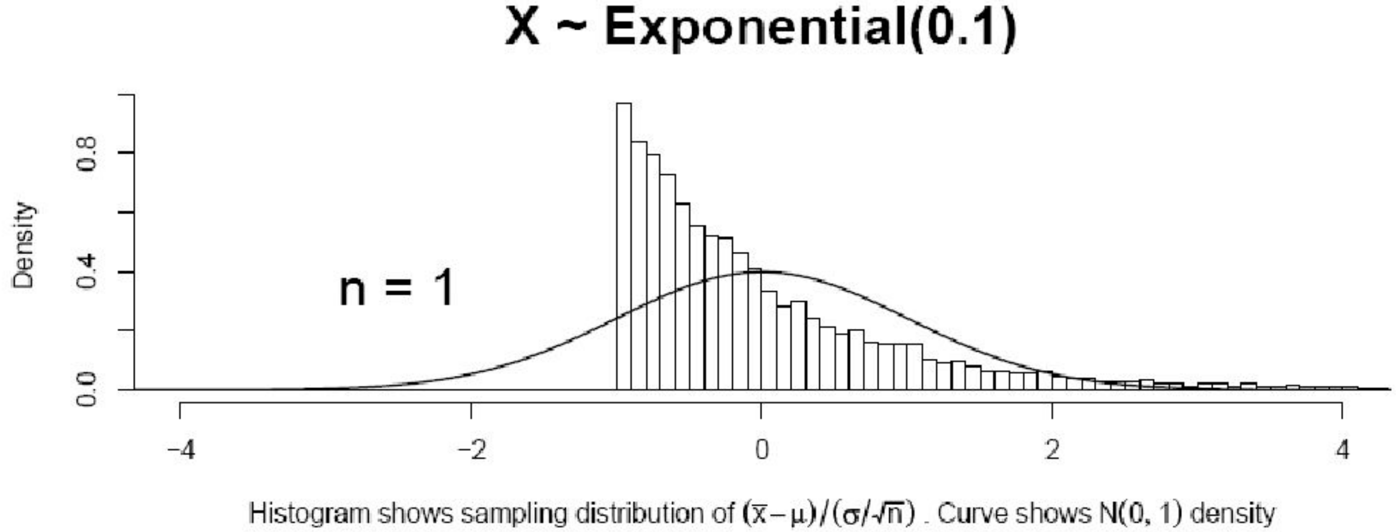
Sample mean

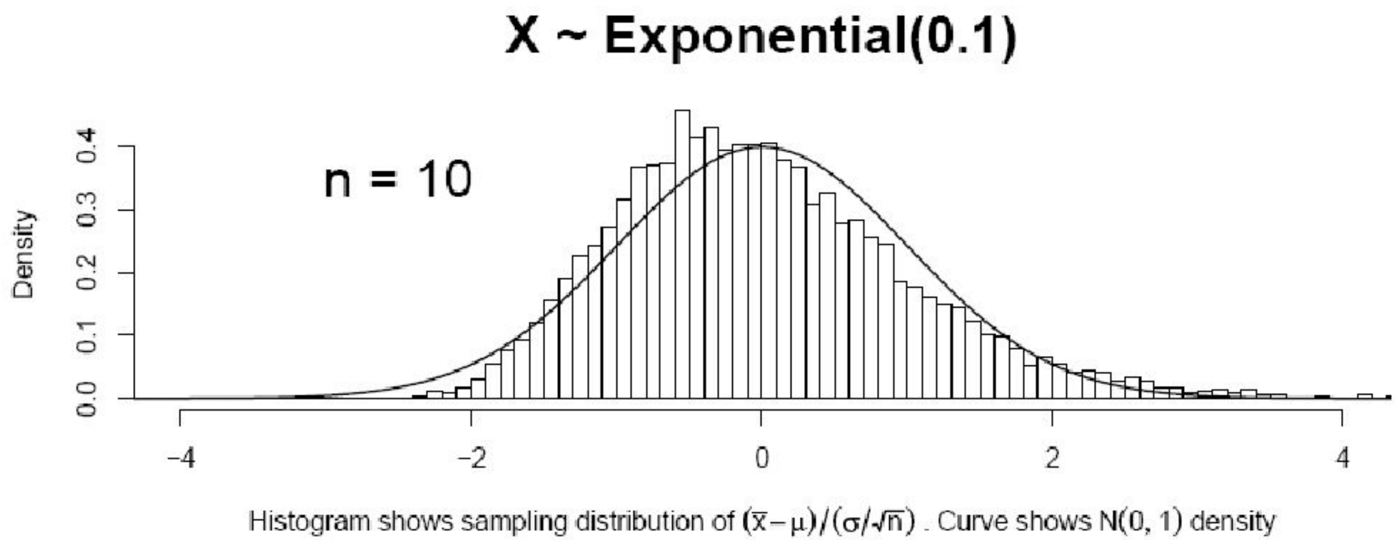
$$\frac{S_n / n - \mu}{\sqrt{\sigma^2 / n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

Variance of the mean

Standard normal distribution

- This result holds for *any* distribution (with finite mean and variance)


X ~ Exponential(0.1)
n = 1
Density
0.0
0.4
0.8
-4
-2
0
2
4
Histogram shows sampling distribution of $(\bar{x}-\mu)/(\sigma/\sqrt{n})$. Curve shows N(0, 1) density

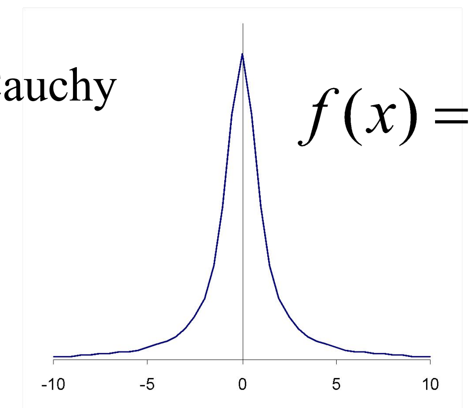

X ~ Exponential(0.1)
n = 10
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
Histogram shows sampling distribution of $(\bar{x}-\mu)/(\sigma/\sqrt{n})$. Curve shows N(0, 1) density

A warning!

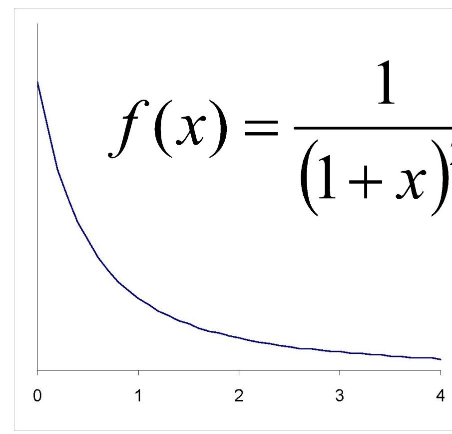
- Not all distributions have finite mean and variance

- For example, neither the Cauchy distribution (the ratio of two standard normal random variables) nor the distribution of the ratio of two iid exponentially distributed random variables have any moments!

Cauchy



$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$



$$f(x) = \frac{1}{(1+x)^2}$$

- For such distributions, the CLT does not hold

Consequences of the CLT

- When asking questions about the mean(s) of distributions from which we havea sample, we can use theory based on the normal distribution
  - Is the mean different from zero?
  - Are the means different from each other?

- Traits that are made up of the sum of many parts are likely to follow a normal distribution
  - True even for mixture distributions

- Distributions related to the normal distribution are widely relevant to statistical analyses
  - χ^2 distribution [Distribution of the sum of squared normal RVs]
  - t-distribution [Sampling distribution of mean with unknown variance]
  - F-distribution [Ratio of two chi-squared RVs]

Properties of the normal distribution

- The sum of two normal random variables also follows a normal distribution

$$X \sim N(\mu, \sigma^2)$$
$$Y \sim N(\lambda, \theta^2)$$
$$X + Y \sim N(\mu + \lambda, \sigma^2 + \theta^2)$$

- Linear transformations of normal random variables also result in normal random variables

$$X \sim N(\mu, \sigma^2)$$
$$Y = aX + b$$
$$Y \sim N(a\mu + b, a^2\sigma^2)$$

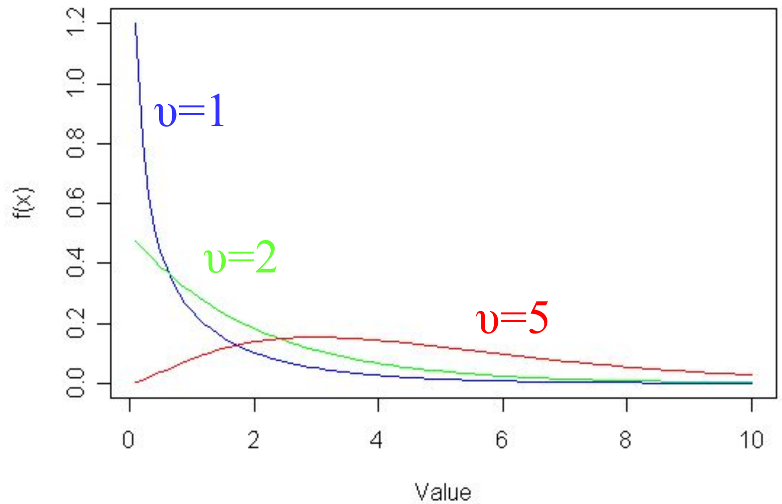
Other functions of normal random variables

- The distribution of the square of a standard normal random variable is the **chi-squared** distribution

$$Z \sim N(0, \sigma^2)$$

$$X = Z^2$$

$$X \sim \chi^2_{v=1}$$



- The chi-squared distribution (χ^2_1) with 1 df is a gamma distribution with α = ½ and β = ½

- The sum of *n* independent chi-squared (1 df) random variables is the chi-squared distribution with *n* degrees of freedom
  - A gamma distribution with $\alpha = n/2$ and $\beta = 1/2$

Uses of the chi-squared distribution

- Under the assumption that a model is a correct description of the data, the difference between observed and expected means is asymptotically normally distributed

- The square of the difference between model expectation and observed value should take a chi-squared distribution

- Pearson's chi-squared statistic is a widely used measure of goodness-of-fit

$$X^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- For example, in a n x m contingency table analysis, the distribution of the test statistic under the null is asymptotically (as the sample size gets large) chi-squared distributed with $(n\text{-}1)(m\text{-}1)$ degrees of freedom

Extreme value theory

- In many situations you may be particularly interested in the tails of a distribution
  - P-values for rare events

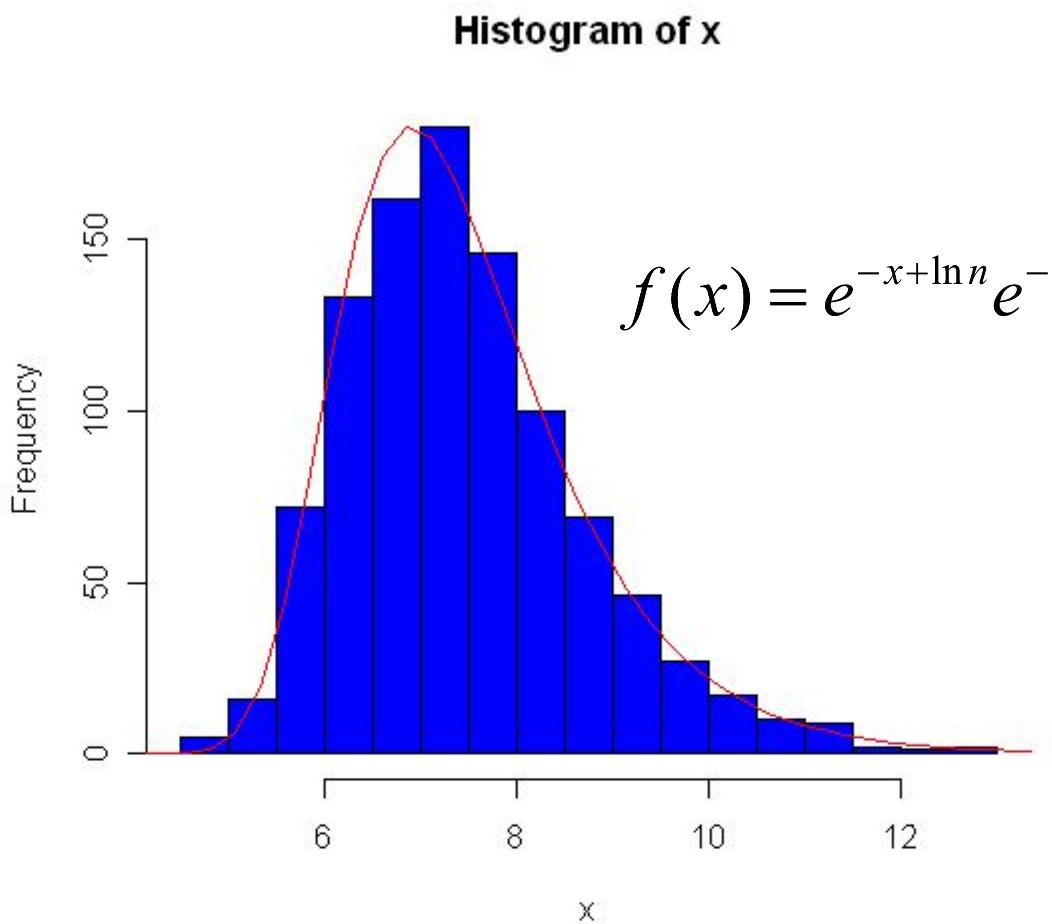
- Remarkably, the distribution of certain rare events is largely independent of the distribution from which the data are drawn

- Specifically, the maximum of a series of iid observations takes one of three limiting forms
  - Gumbel distribution (Type I): e.g. Exponential, Normal
  - Frechet distribution (Type II): Heavy-tailed, e.g. Pareto $X = e^Y, \quad Y \sim Exp(\lambda)$
  - Weibull distribution (Type III): Bounded distributions, e.g. Beta

- These limiting forms can be expressed as special cases of a generalised extreme value distribution

Example: Gumbel distribution

- Distribution of max of 1000 samples from Exp(1)



$$f(x) = e^{-x+\ln n} e^{-e^{-x+\ln n}}$$

More generally..

$$U = \frac{X_{\max} - b_n}{a_n}$$

Re-centered by expected maximum

$$F^{-1}\left(1 - \tfrac{1}{n}\right)$$

Re-scaled by...

$$F^{-1}\left(1 - \tfrac{1}{ne}\right) - b_n$$

$$f(U) = e^{-U} e^{-e^{-U}}$$

e.g. 1000 samples from Normal(0,1)