Probability and Bayes Theorem

Biology 683

Lecture 3

Heath Blackmon

Last week

1. What is the difference in a population and a sample?

2. What is the difference in a parameter and a statistic?

3. How does the central limit theorem help us?

Today

1. Probability
2. Frequentist vs Bayesian

Covariance and Correlation

The covariance shows the extent to which the two variables are not statistically independent

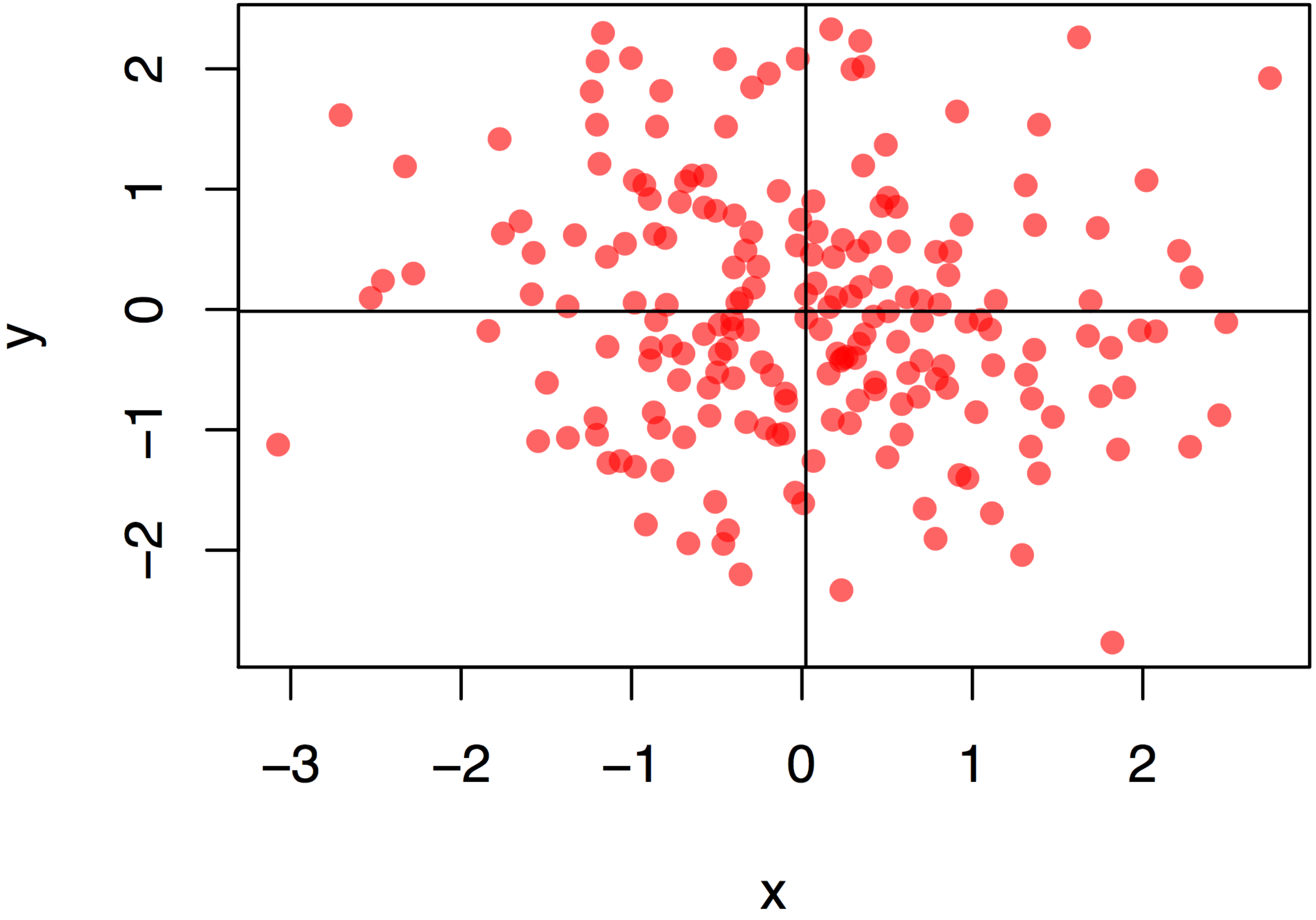
$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

The correlation is the covariance, standardized to fall between -1 and 1.

$$r(X,Y) = \frac{cov(X,Y)}{s_x s_y}$$

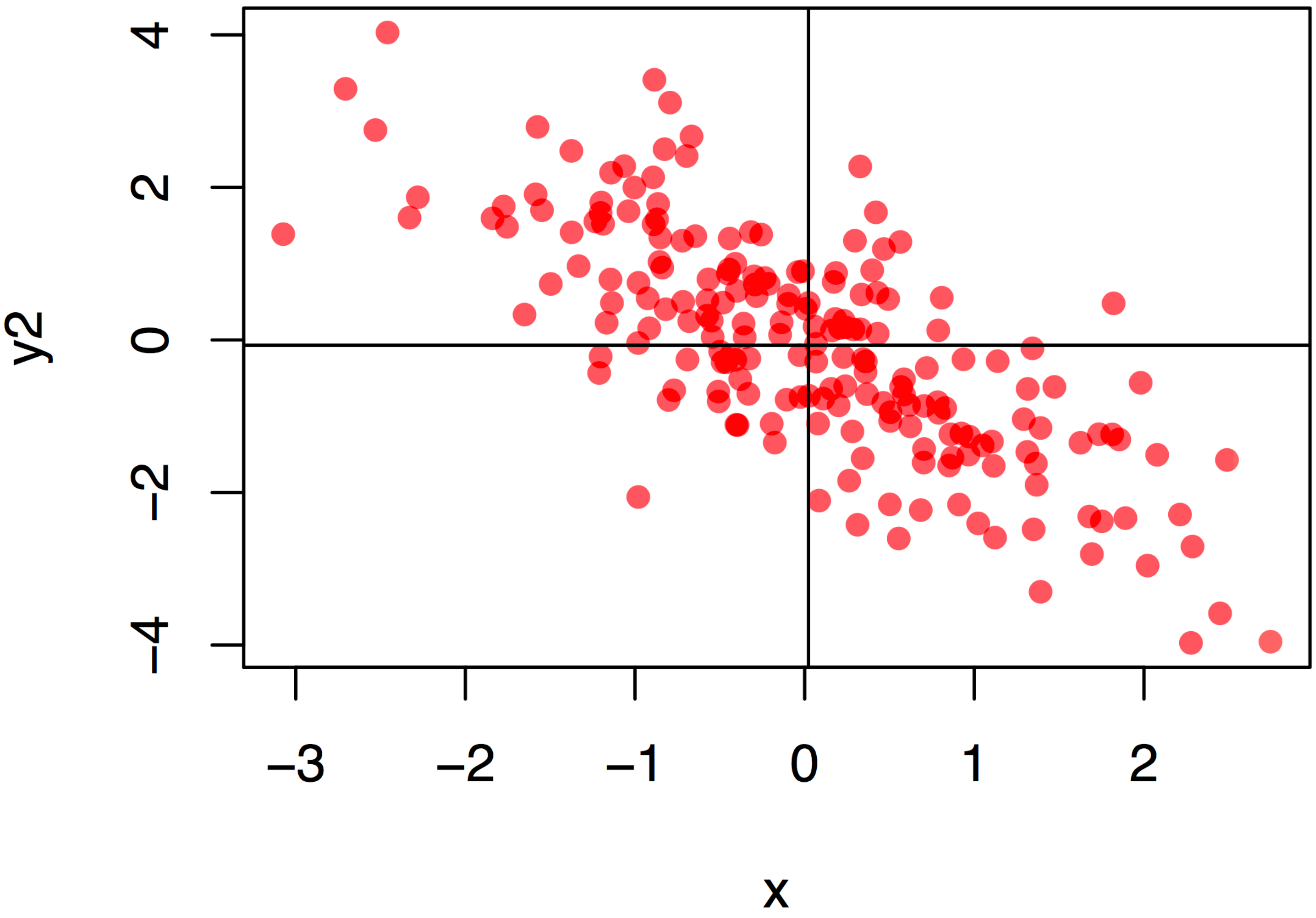
Covariance and Correlation

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

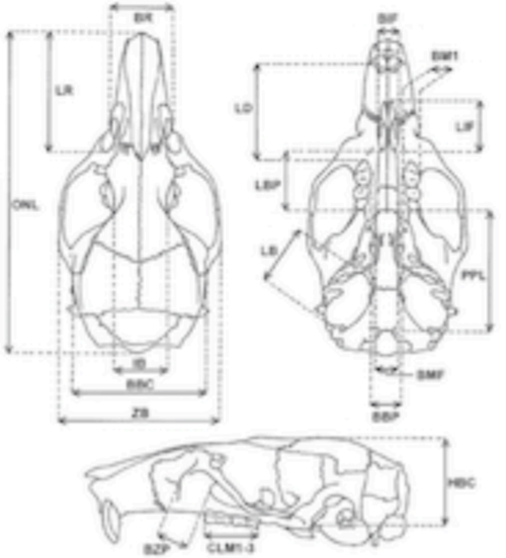


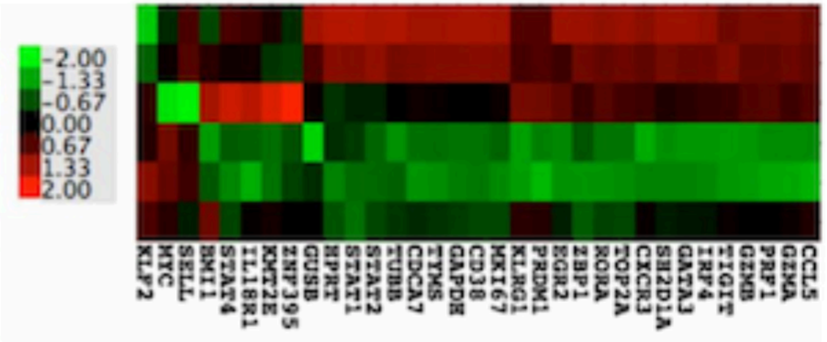
Covariance and Correlation

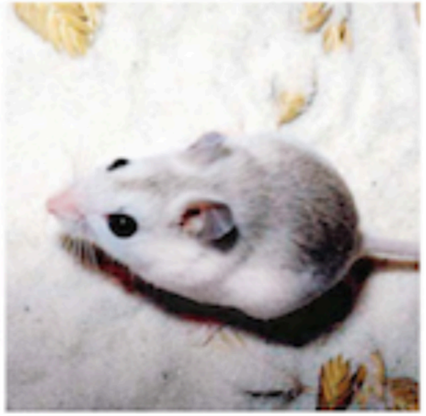
$$cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$



Biological examples of correlations

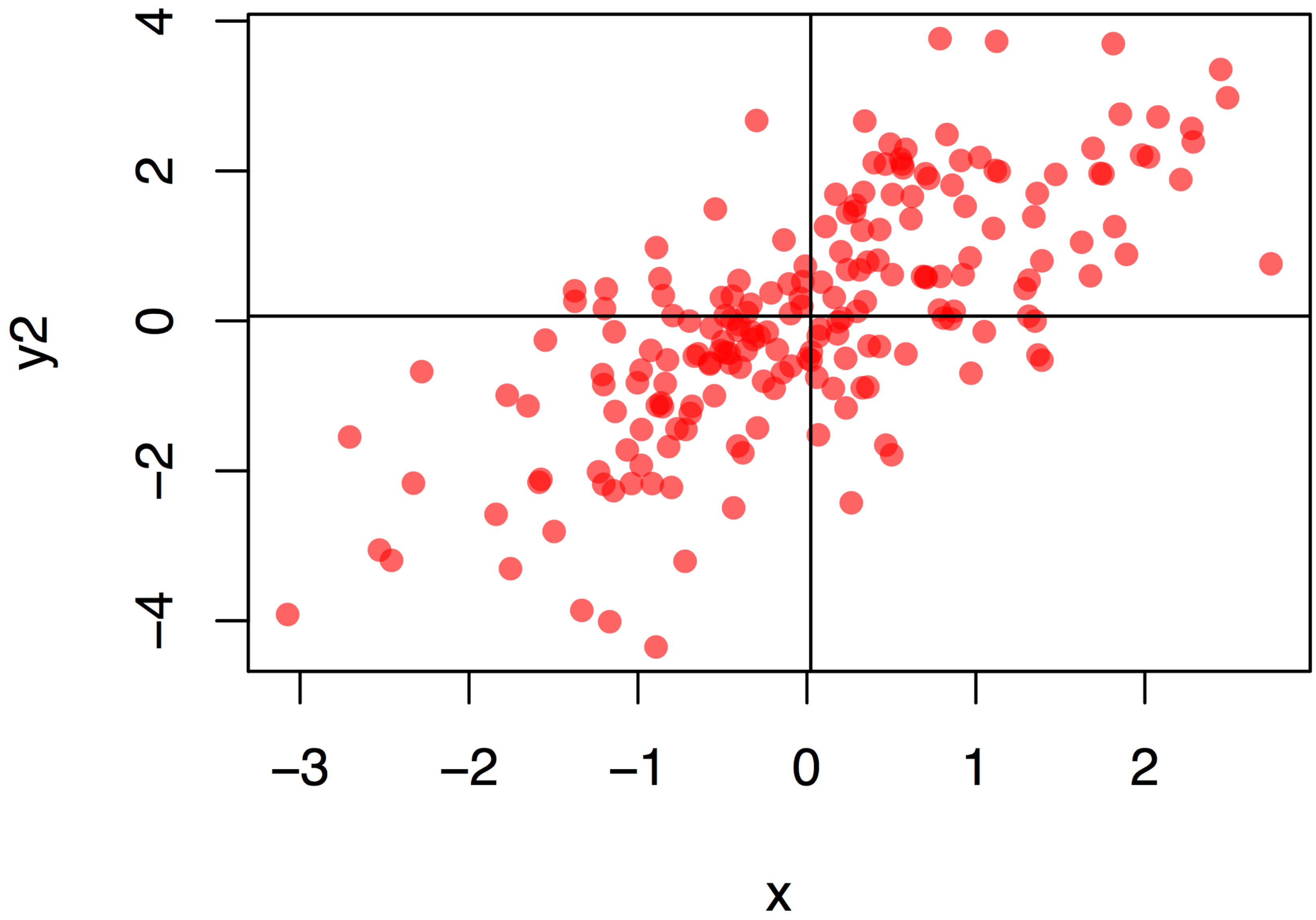






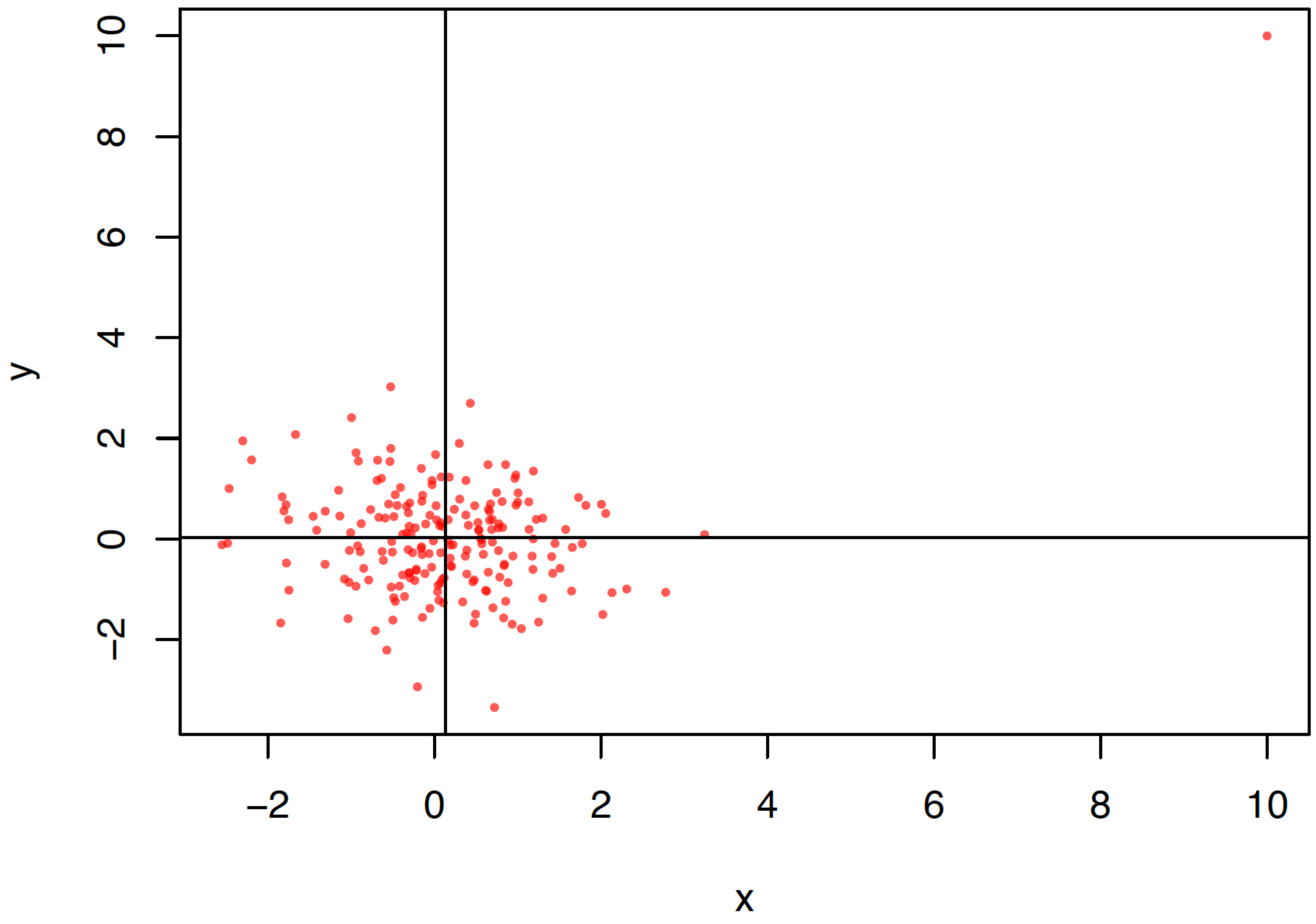
Covariance and Correlation

$$cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$



Covariance and Correlation

$$cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$



Mutually exclusive events

What is the probability of drawing either an ace or a king from a deck cards?

addition rule: if two events are mutually exclusive then the probability that either will occur is just the sum of the individual probabilities

$$\Pr(A \ or \ B) = \Pr(A) + \Pr(B)$$

$$\Pr(ace \ or \ king) = 15\% = \frac{4}{52} + \frac{4}{52}$$

assumes standard deck with no jokers

Independent events

What is the probability of rolling a 1 and then a 2 on a pair of dice?

multiplication rule: if two events are independent of one another then the probability of both occurring is just the product of the individual probabilities.

$$\Pr(A\ then\ B) = \Pr(A) \times \Pr(B)$$

$$\Pr(1\ then\ 2) = 3\% = \frac{1}{6} \times \frac{1}{6}$$

Non-independence

non-independent events- conditional probability: is the probability of an event given that another event has already occurred.

For instance, the probability of surviving the sinking of the Titanic was very different than the probability of surviving if you were a female

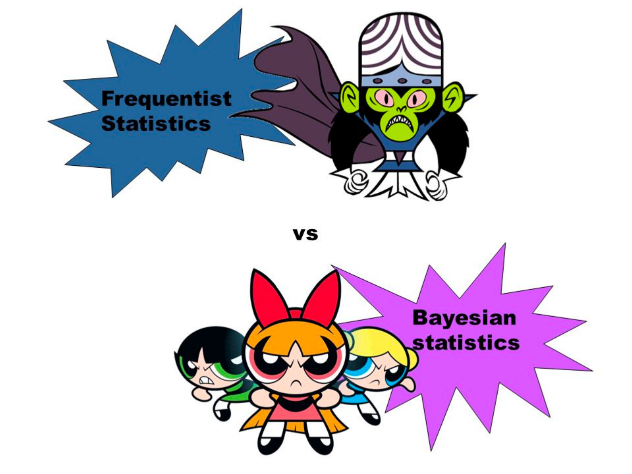
	died	lived
male	1364	367
female	126	344

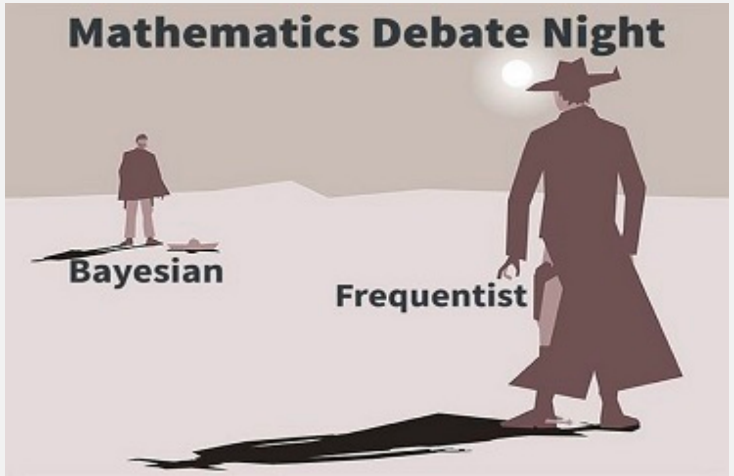
Conditional probability

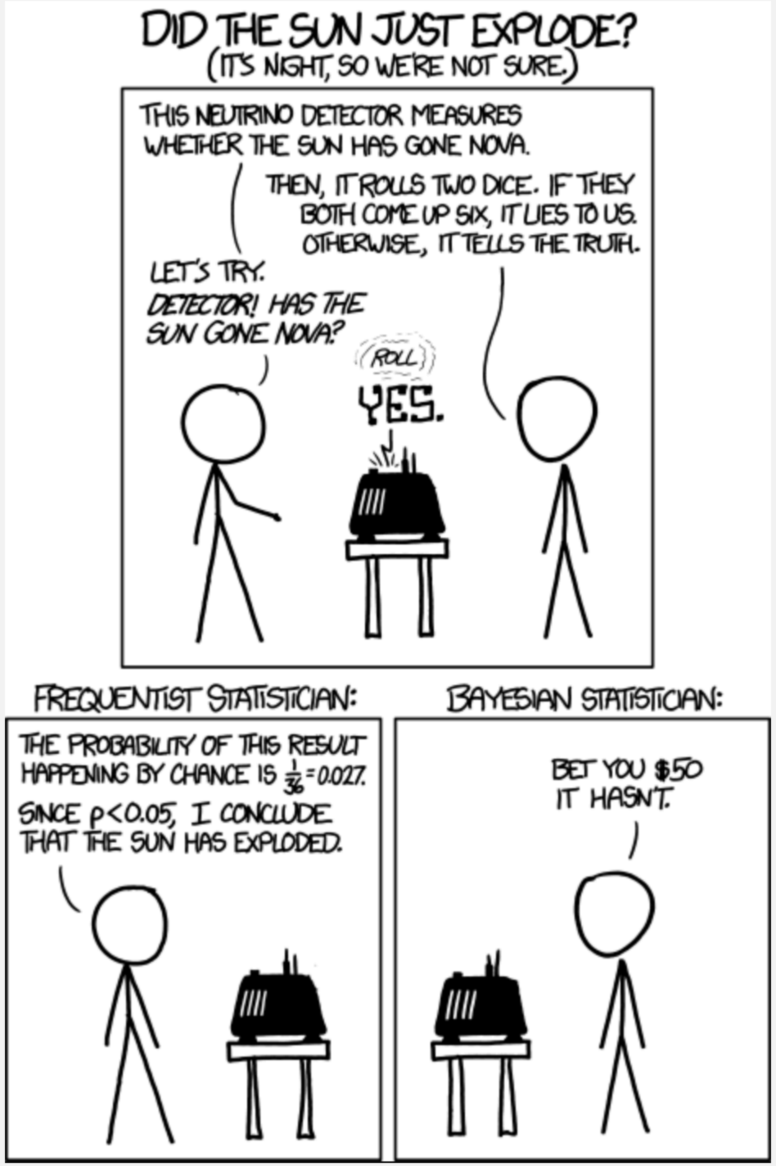
$$\Pr(survivial) = 32\% = \frac{711}{2201}$$

$$\Pr(survivial \mid female) = 73\% = \frac{344}{470}$$

Frequentist and Bayesian







Frequentist and Bayesian

	Frequentist	**Bayesian**
Hypothesis test	P-value	Bayes factors
Parameter estimation	Maximum likelihood with confidence interval	Posterior distribution

Frequentist and Bayesian

	Frequentist	**Bayesian**
Hypothesis test	P-value	Bayes factors
Parameter estimation	Maximum likelihood with confidence interval	Posterior distribution

What is a p-value?

Is the probability of finding the observed, or more extreme, results when the null hypothesis is true.

````
```{r}
cor.test(dat.f$offspring, dat.f$body)
```
````

```
	Pearson's product-moment correlation

data:  dat.f$offspring and dat.f$body
t = 3.3693, df = 105, p-value = 0.001055
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1302191 0.4740979
sample estimates:
      cor
0.3123551
```

$$t = \frac{r\sqrt{df}}{\sqrt{1 - r^2}}$$

Frequentist and Bayesian

	Frequentist	**Bayesian**
Hypothesis test	P-value	Bayes factors
Parameter estimation	Maximum likelihood with confidence interval	Posterior distribution

Frequentist and Bayesian

$$Bayes\ Factor = \frac{Pr(D|M_1)}{Pr(D|M_2)}$$

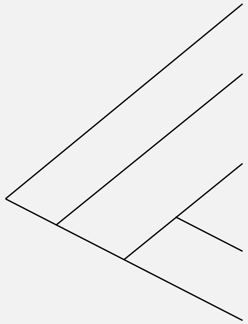
Bayes factor	Interpretation
>100	Decisive evidence for H_A
30–100	Very strong evidence for H_A
10–30	Strong evidence for H_A
3–10	Substantial evidence for H_A
1–3	Anecdotal evidence for H_A
1	No evidence
1/3–1	Anecdotal evidence for H_0
1/10–1/3	Substantial evidence for H_0
1/30–1/10	Strong evidence for H_0
1/100–1/30	Very strong evidence for H_0
<1/100	Decisive evidence for H_0

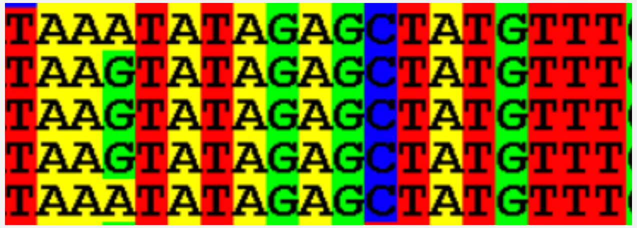
Frequentist and Bayesian

	Frequentist	**Bayesian**
Hypothesis test	P-value	Bayes factors
Parameter estimation	Maximum likelihood with confidence interval	Posterior distribution

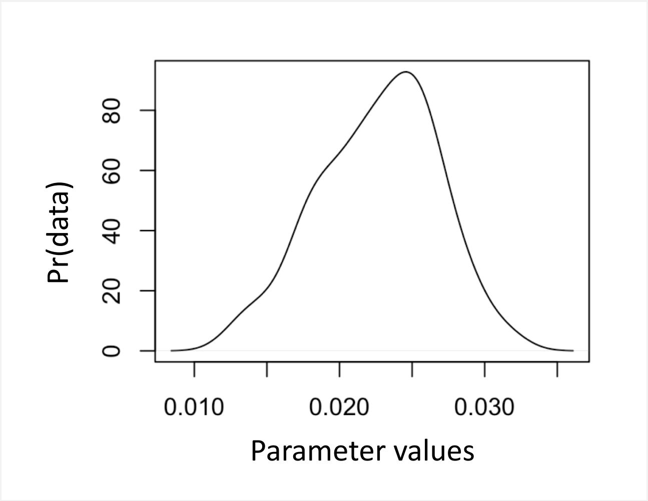
Frequentist parameter estimate

Everything is fixed except parameter of interest. Shape of likelihood determines estimate of the statistic





	A	G	C	T
A		β	α	α
G	β		α	α
C	α	α		β
T	α	α	β	

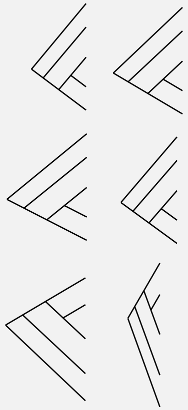


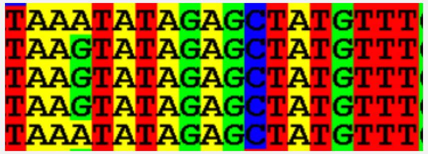
Frequentist and Bayesian

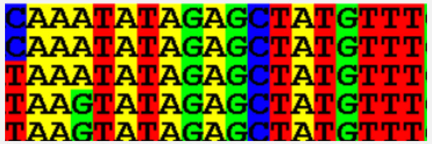
	Frequentist	**Bayesian**
Hypothesis test	P-value	Bayes factors
Parameter estimation	Maximum likelihood with confidence interval	Posterior distribution

Bayesian parameter estimate

We create a marginal estimate of the statistic of interest



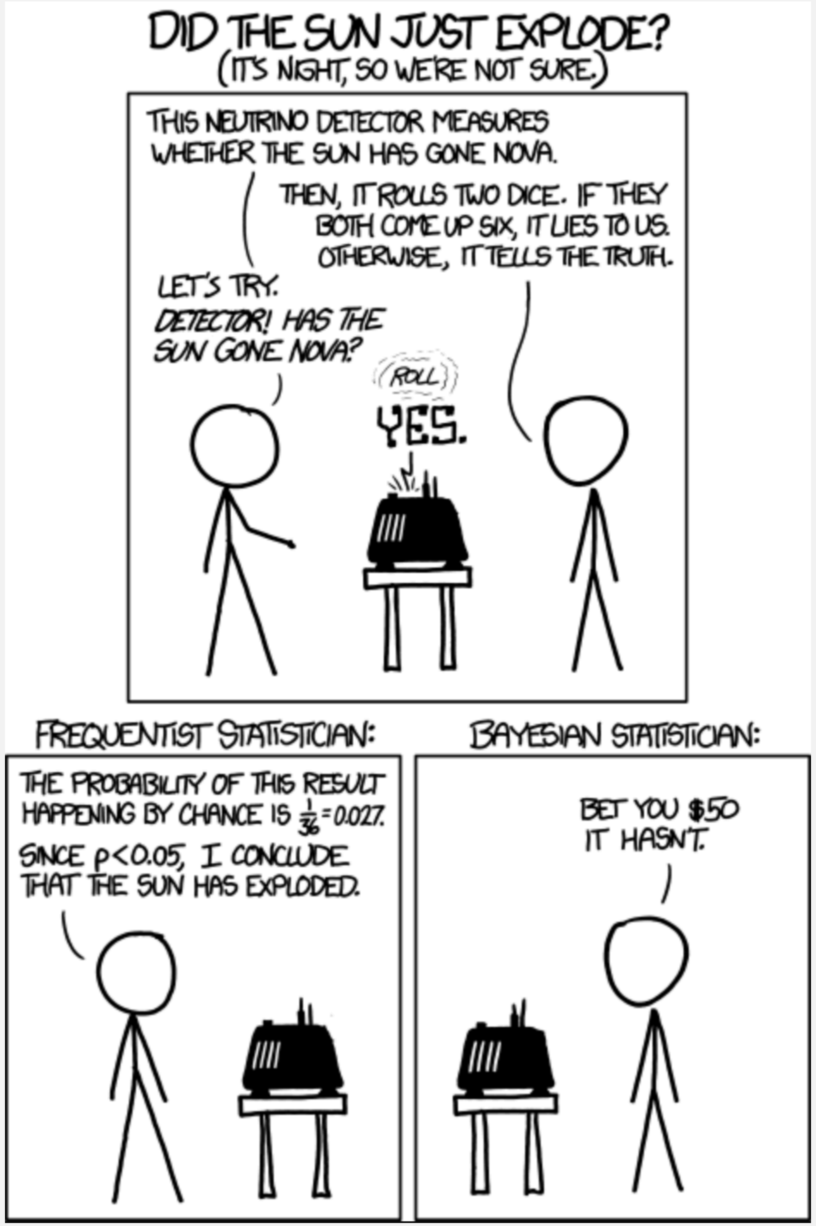




	A	G	C	T
A		β	α	α
G	β		α	α
C	α	α		β
T	α	α	β	

Tree4	alignment8	model6	1.0391601	2.25225446
Tree8	alignment4	model6	1.04903411	2.11859404
Tree10	alignment7	model5	1.03117276	2.88413248
Tree10	alignment3	model4	1.02935202	2.26183042
Tree9	alignment8	model3	1.04700046	2.00114012
Tree2	alignment10	model2	1.01892741	2.60891431
Tree2	alignment6	model3	1.01524482	2.11437827
Tree8	alignment2	model1	1.02230859	2.30319668
Tree5	alignment7	model1	1.02593538	2.13290411
Tree8	alignment10	model5	1.00762542	2.13242719
Tree0	alignment4	model5	1.01884236	2.42583388
Tree9	alignment8	model4	1.00252288	2.71467137
Tree4	alignment0	model8	1.02480351	2.5452598
Tree8	alignment8	model10	1.03066934	2.38768005
Tree1	alignment3	model5	1.046732	2.8957133
Tree9	alignment5	model6	1.0226392	2.79771149
Tree10	alignment0	model8	1.03210411	2.08787975
Tree2	alignment6	model8	1.00233853	2.15930312
Tree4	alignment6	model3	1.0266862	2.21811627
Tree2	alignment0	model6	1.02813474	2.01792657
Tree1	alignment1	model6	1.00101131	2.0549765
Tree7	alignment7	model2	1.00596282	2.20547933
Tree5	alignment3	model9	1.00241339	2.62082855
Tree5	alignment9	model9	1.03602053	2.25983463
Tree0	alignment5	model4	1.00629136	2.86463569
Tree8	alignment5	model7	1.02207157	2.79890432
Tree8	alignment3	model5	1.01522322	2.66694394
Tree9	alignment8	model9	1.02316269	2.75648806
Tree4	alignment5	model10	1.0229115	2.28025282
			1.02547847	

Frequentist and Bayesian



Frequentist approach

You have a null hypothesis and calculate the probability of observing your data under that hypothesis.

Null: The sun has not exploded

$$\Pr(det\ yes \mid sun\ is\ ok) = 0.027 = \frac{1}{6} \times \frac{1}{6}$$

this is less than the typical $\propto$ level of 0.05 so we reject the null that the sun is ok.

Bayesian approach

$$\Pr(A|B) = \frac{\Pr(B|A)\times\Pr(A)}{\Pr(B)}$$

$$\Pr(nova|det\ yes) = \frac{\Pr(det\ yes|nova)\times\Pr(nova)}{\Pr(\det yes)}$$

$$\Pr(nova|det\ yes) << 0.00000000037\ \frac{1\times0.0000000001}{0.027}$$

Bayesian approach

Imagine a genetic mutation present in 1 person per 100,000. If you take a test that correctly returns a positive result 99.9% of the time when someone is infected but has a false positive rate of 0.5%.

How likely are you to have the disease?

Should you be concerned?

Bayes' theorem provides a natural way to think about this.

Lets do this!

Bayesian approach

$$\Pr(inf|pos.test) = \frac{\Pr(pos.test|inf) \times \Pr(inf)}{\Pr(pos.test)}$$

$$\Pr(inf|pos.test) = 0.001 = \frac{0.999 \times 0.00001}{0.00501}$$

Which means that you have only a 0.1% chance of having the disease even if you test positive.

What if your doctor noticed a symptom that made them give you this test what changes?

Priors

$$\Pr(A|B) = \frac{\Pr(B|A) \times \boldsymbol{\Pr(A)}}{\Pr(B)}$$

Pr(A) this is the prior. It is powerful because it allows you to incorporate previous knowledge into your analysis, but it can lead to very bad inference if you are careless.

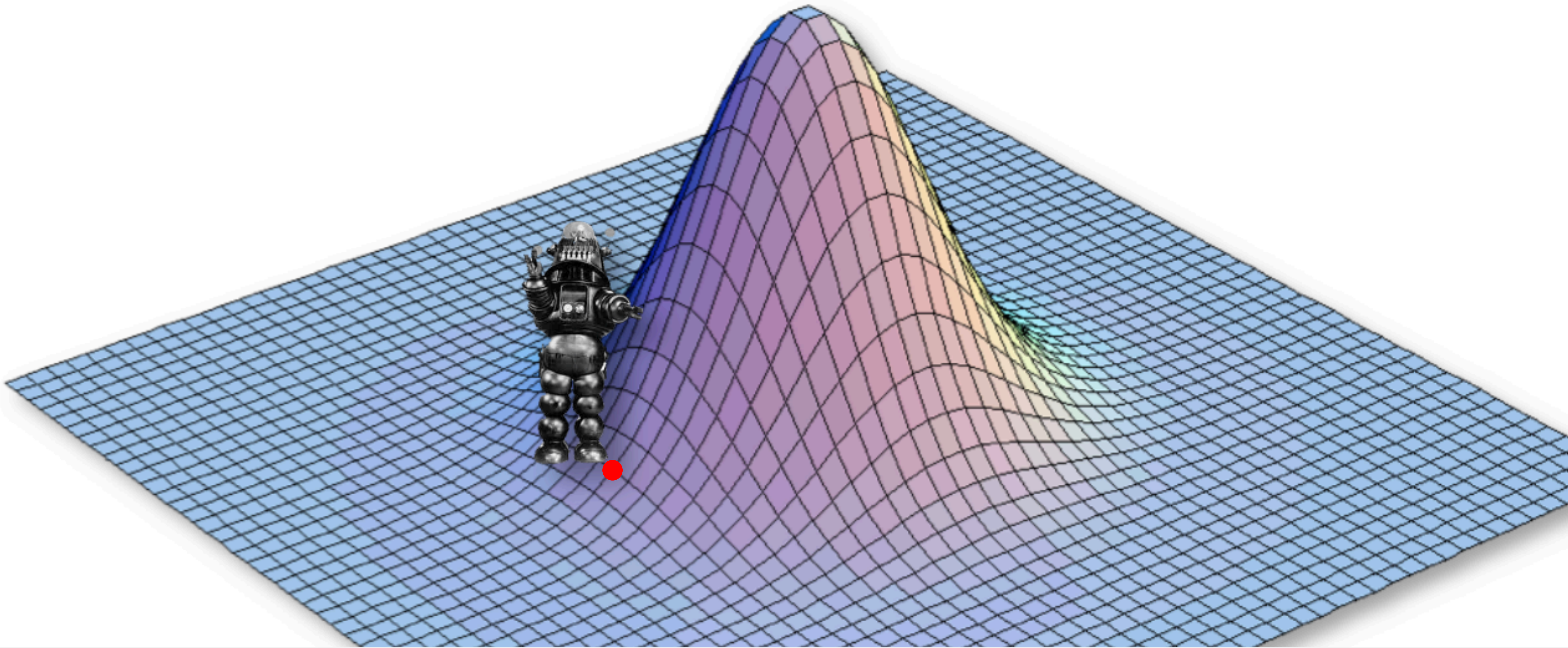
Bayesian methods in practice (MCMC)

In practice our models are quite complex with many parameters we would like to estimate. Each of these has its own prior distribution. The way that we explore the space of solutions is using an MCMC.

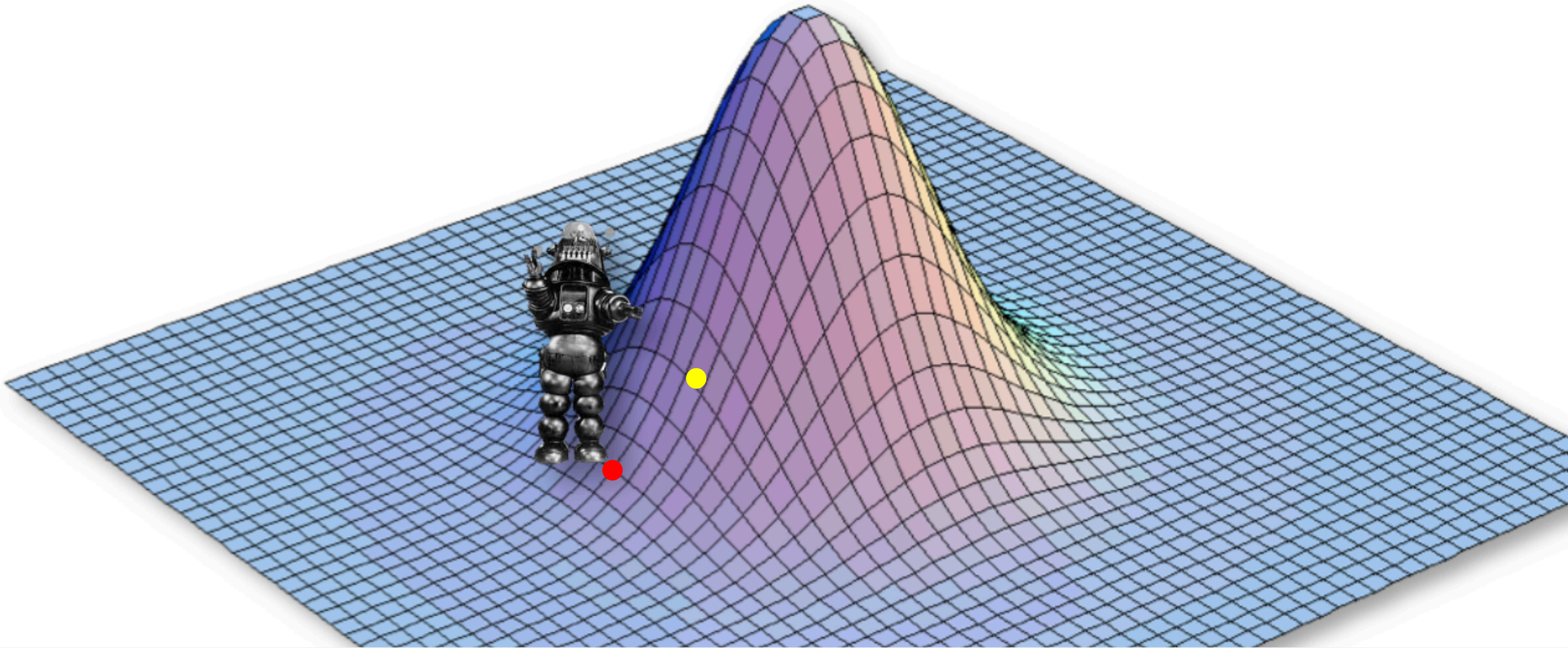
$$\Pr(A|B) = \frac{Pr(B|A)\times\Pr(A)}{Pr(B)}$$

1) Pick an arbitrary starting point
2) Calculate the current probability p_t
3) Generate a change to one of our parameters
4) Calculate the new probability p_{t+1}
5) Accept new parameter value with probability $\frac{p_{t+1}}{p_t}$
6) Return to step 3 (repeat for a really long time)

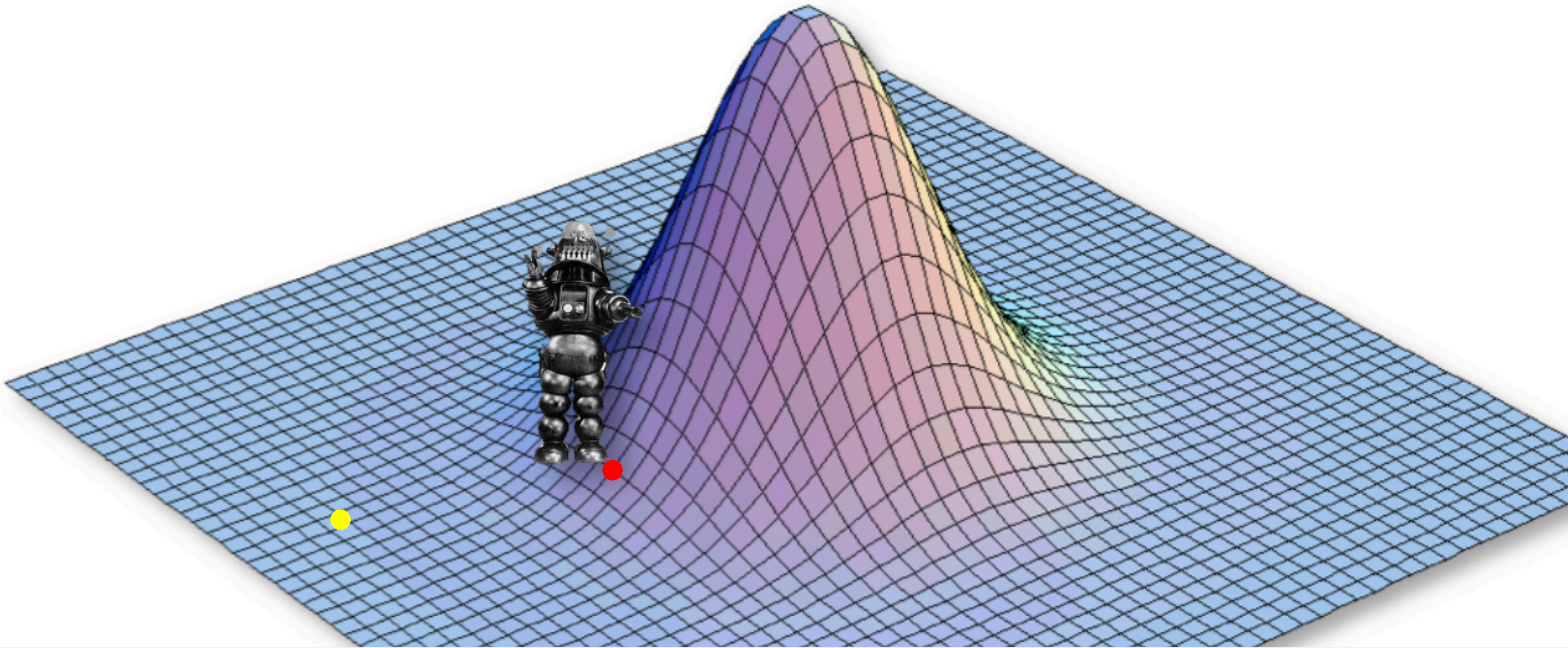
Bayesian methods in practice (MCMC)



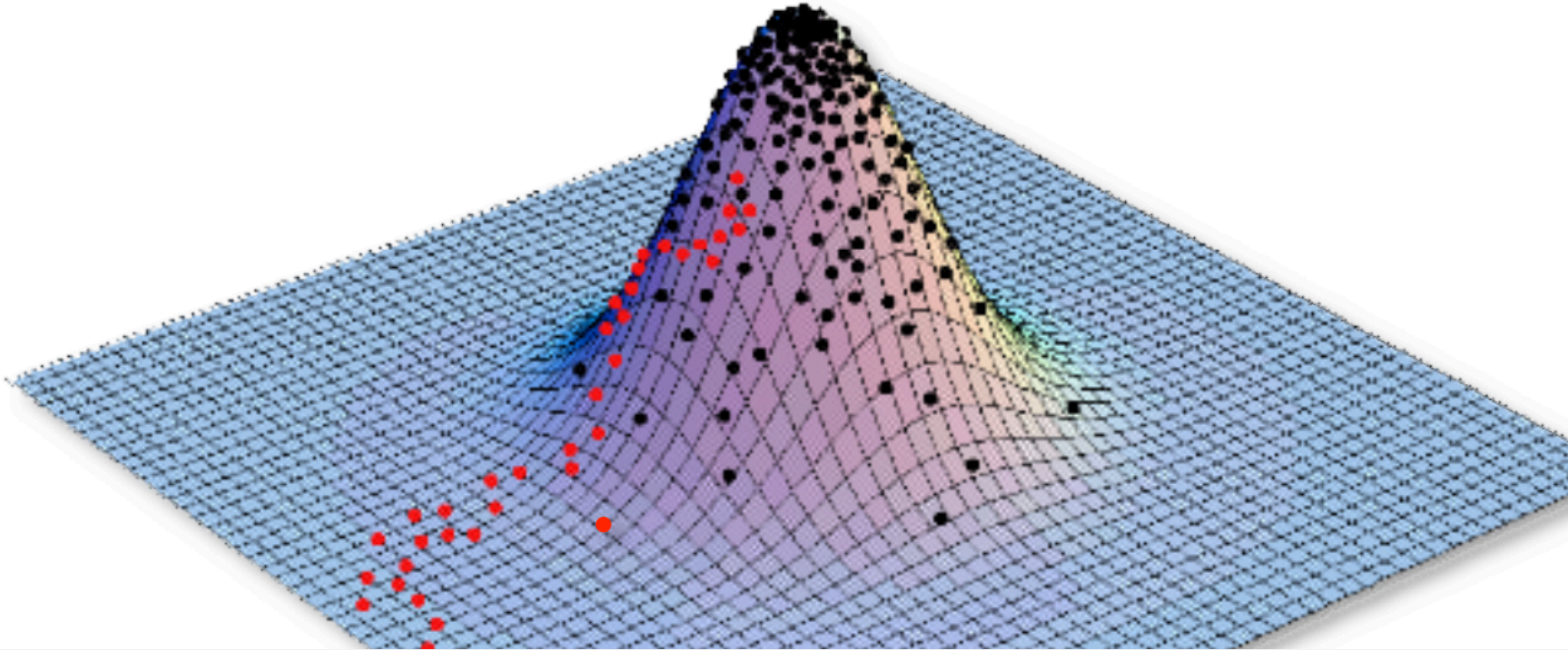
Bayesian methods in practice (MCMC)



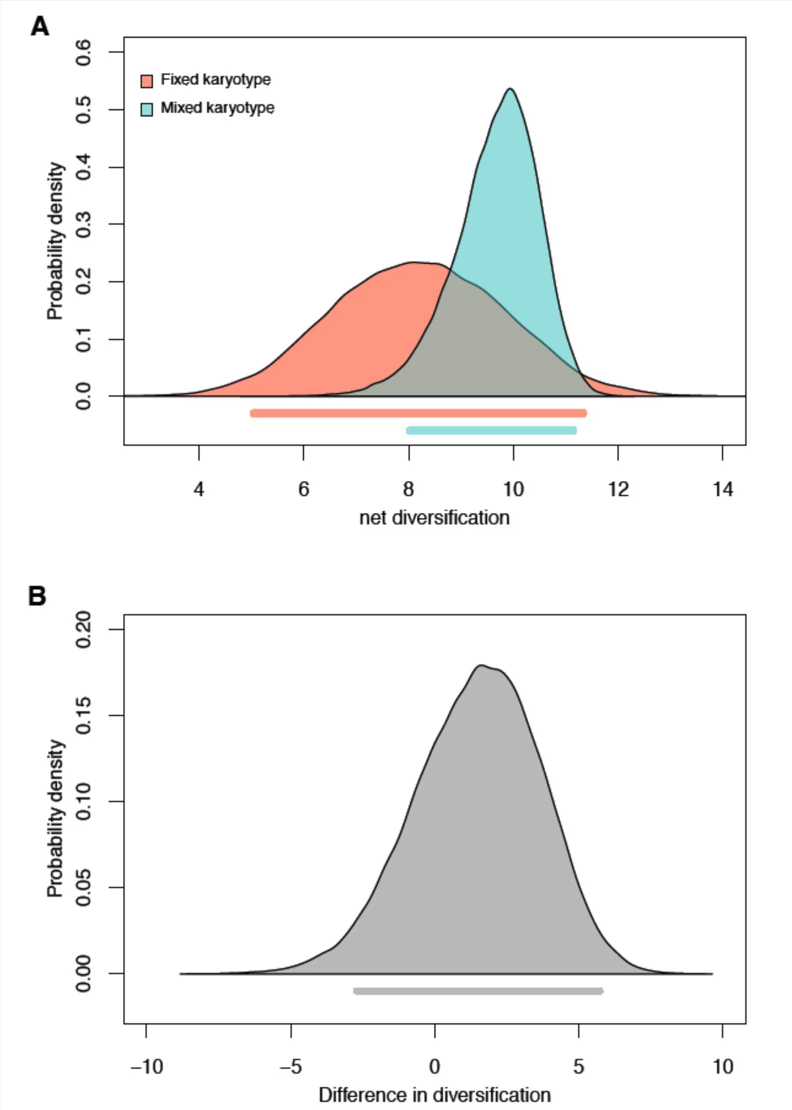
Bayesian methods in practice (MCMC)



Bayesian methods in practice (MCMC)



The power of Bayesian approaches



With a Bayesian approach we can take into account all types of uncertainty and be more conservative.

- Phylogenetic
- Model selection uncertainty
- Parameter value uncertainty
- Uncertainty in measurements

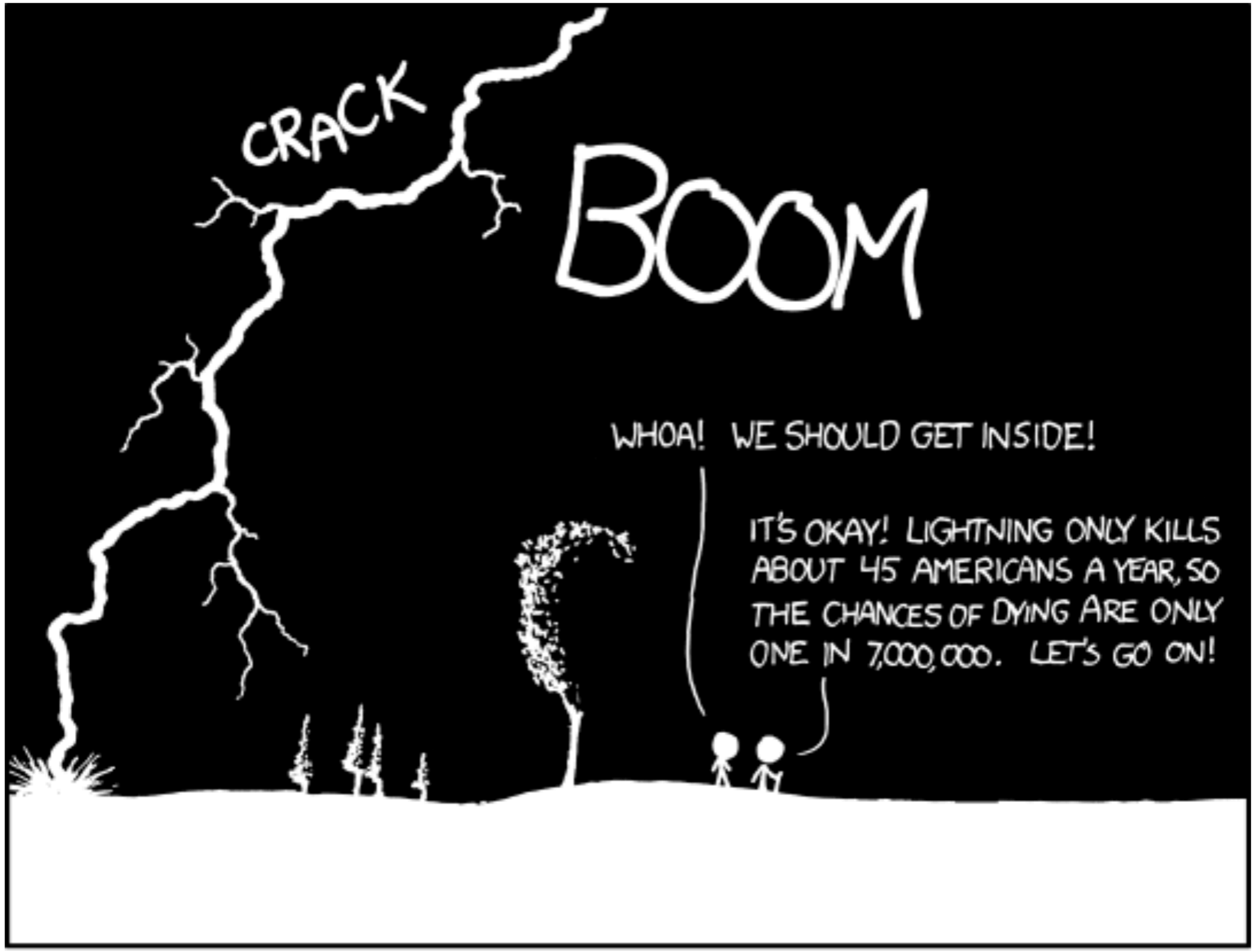
For Thursday

Read chapter WS 5 and chapter 2 of McElreath

Bring laptop to class!

Heath Blackmon
BSBW 309
coleoguy@gmail.com
@coleoguy

Setting matters



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.