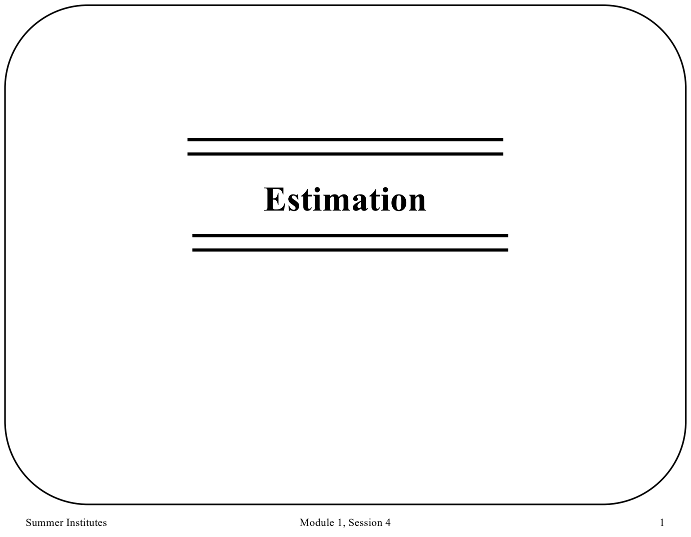
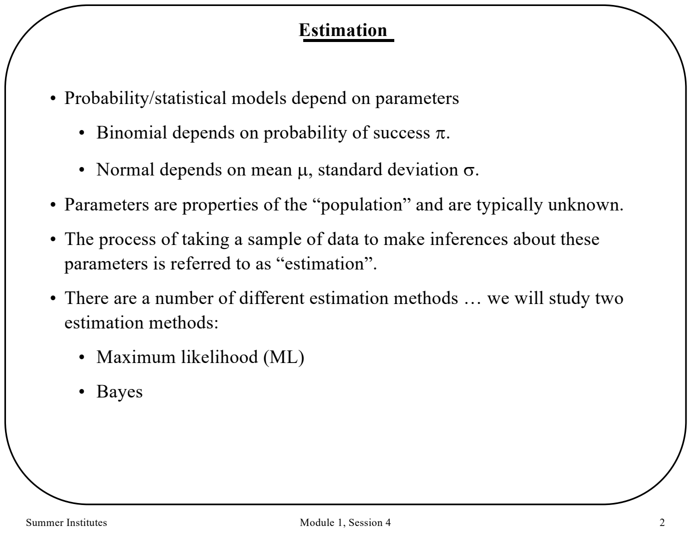


Estimation

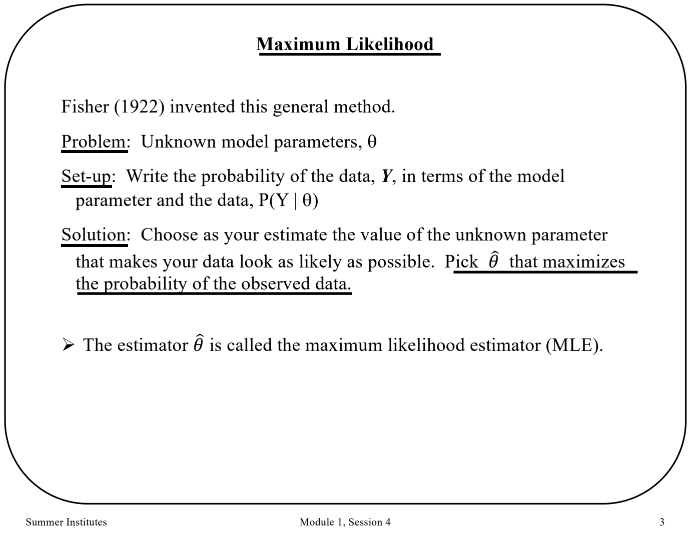
1



Estimation

- Probability/statistical models depend on parameters
  - Binomial depends on probability of success π.
  - Normal depends on mean μ, standard deviation σ.
- Parameters are properties of the "population" and are typically unknown.
- The process of taking a sample of data to make inferences about these parameters is referred to as "estimation".
- There are a number of different estimation methods … we will study two estimation methods:
  - Maximum likelihood (ML)
  - Bayes

2



Maximum Likelihood

Fisher (1922) invented this general method.

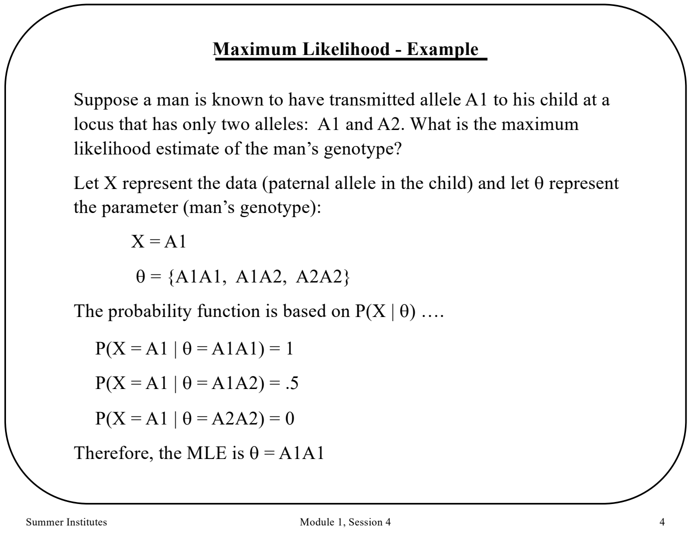
Problem: Unknown model parameters, θ

Set-up: Write the probability of the data, ***Y***, in terms of the model parameter and the data, P(Y | θ)

Solution: Choose as your estimate the value of the unknown parameter that makes your data look as likely as possible. Pick $\hat{\theta}$ that maximizes the probability of the observed data.

➢ The estimator $\hat{\theta}$ is called the maximum likelihood estimator (MLE).

3



Maximum Likelihood - Example

Suppose a man is known to have transmitted allele A1 to his child at a locus that has only two alleles: A1 and A2. What is the maximum likelihood estimate of the man's genotype?

Let X represent the data (paternal allele in the child) and let θ represent the parameter (man's genotype):

X = A1

θ = {A1A1, A1A2, A2A2}

The probability function is based on P(X | θ) ….

P(X = A1 | θ = A1A1) = 1

P(X = A1 | θ = A1A2) = .5

P(X = A1 | θ = A2A2) = 0

Therefore, the MLE is θ = A1A1

4

Maximum Likelihood - Example

Suppose we have a sample of 20 gametes in which the number of recombinants (Z) and nonrecombinants (N-Z) for two loci can be counted. Use these data to estimate the recombination fraction (π) between the two loci.

The probability of the data can be modeled using a binomial distribution. The ***probability distribution function*** is:

$$P(Z;\ \pi) = \binom{20}{Z} \pi^{Z} (1-\pi)^{(20-Z)}$$

where *Z* is the variable and **π is fixed**.

The ***likelihood function*** is the same function:

$$L(\pi; Z) = \binom{20}{Z} \pi^{Z} (1-\pi)^{(20-Z)}$$

except now π is the variable and **Z is fixed**.

Maximum Likelihood - Example

Two ways to look at this:

1) **Probability**: fix π (e.g. π = 0.1) and look at the probability of different values of *Z*:

$\pi = 0.1$

Z	$P(Z, \pi)$
0	0.122
1	0.270
2	0.285
3	0.190
4	0.090
5	0.032

2) Likelihood: fix *Z* (e.g. Z = 3) and look at the "likelihood" under different values of π (this is called the likelihood function):

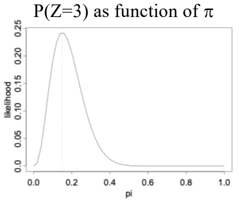
$Z = 3$

π	$P(Z, \pi)$
0.01	0.001
0.05	0.060
0.10	0.190
0.20	0.205
0.30	0.072
0.40	0.012

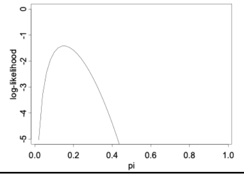
Maximum Likelihood - Example

For the data *Z* = 3 then the likelihood function is shown in the plots below:

P(Z=3) as function of π



log P(Z=3) as function of π



Maximum Likelihood

- We can use calculus to find the maximum of the (log) likelihood function:

$$\begin{aligned} \frac{d\log L}{d\pi} &= 0 \\ \frac{d}{d\pi} Z\log\pi + (20-Z)\log(1-\pi) &= 0 \\ \frac{Z}{\pi} - \frac{(20-Z)}{1-\pi} &= 0 \\ \hat{\pi} &= \frac{Z}{20} \end{aligned}$$

- Not surprisingly, the likelihood in this example is maximized at the observed proportion, 3/20.
- Sometimes (e.g. this example) the MLE has a simple closed form. In more complex problems, numerical optimization is used.
- Computers can find these maximum values!

Maximum Likelihood - Notation

L(θ) = Likelihood as a function of the parameter, θ.

l(θ) = log(L(θ)), the log-likelihood.

- Usually more convenient to work with analytically and numerically.

S(θ) = d*l*(θ)/dθ = the “score”.

- Set d*l*(θ)/dθ = 0 and solve for θ to find the MLE.

I(θ) = -d^2*l*(θ)/dθ^2 = the “information”.

- The inverse of the expected information gives the variance of $\hat{\theta}$

Var(θ) = E(I(θ))$^{-1}$ (in most cases)

Maximum Likelihood - Example

$$L(\pi) = \binom{20}{Z}\pi^Z(1-\pi)^{(20-Z)}$$

$$\ell(\pi) = Z\log(\pi) + (20-Z)\log(1-\pi)$$

$$S(\pi) = \frac{Z}{\pi} - \frac{(20-Z)}{1-\pi} \quad \Rightarrow \quad \pi = \frac{Z}{20}$$

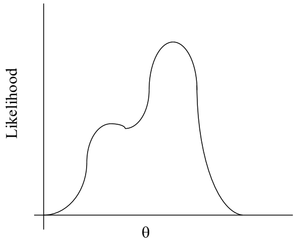
$$I(\pi) = \frac{Z}{\pi^2} + \frac{(20-Z)}{(1-\pi)^2}$$

$$E(I(\pi)) = \frac{20\pi}{\pi^2} + \frac{(20-20\pi)}{(1-\pi)^2} = \frac{20}{\pi(1-\pi)}$$

(note: constant dropped from $\ell(\pi)$)

Numerical Optimization

- In complex problems it may not be possible to find the MLE analytically; in that case we use numerical optimization to search for the value of θ that maximizes the likelihood
- A common problem with maximum likelihood estimation is accidentally finding a local maximum instead of a global one; solution is to try multiple starting values



Maximum Likelihood – Example (numerical)

If you have access to R, here is code to numerically find the mle for the binomial problem that we solved earlier. Try running it.

```
# Numerical mle example
loglike = function(theta,z,n){
# maximize loglike = minimize negative loglike
  -(z*log(theta) + (n-z)*log(1-theta))
}

#initialize theta
init = .5
# numerical optimization with boundaries
# function fails if theta = 0 or 1 so keep away from boundaries
eps = .Machine$double.eps
# optim minimizes function loglike
optim(init,loglike,method="L-BFGS-B",lower=eps,upper=1-eps,z=3,n=20)
```

Maximum Likelihood - Comments

- Maximum likelihood estimates (MLEs) are always based on a probability model for the data.
- Maximum likelihood is the “best” method of estimation for any situation that you are willing to write down a probability model (so generally does not apply to nonparametric problems).
- Maximum likelihood can be used even when there are multiple unknown parameters, in which case θ has several components

  (ie. $\theta_0, \theta_1, \ldots, \theta_p$).
- The MLE is a “point estimate” (i.e. gives the single most likely value of θ). In lecture 5 we will learn about interval estimates, which describe a range of values which are likely to include the true value of θ. We combine the MLE and Var(θ) to generate these intervals.
- The likelihood function lets us compare different models (next).

Model Comparisons

Q: Suppose we have two alternative models for the data; in each case we use maximum likelihood to estimate the parameters. How do we decide which model fits the data “better”?

A: First thought - compare the likelihoods.

- Larger likelihood is better, but …
- the tradeoff is larger likelihood ⇔ more complex model.
- How to choose?

A common approach is to “penalize” the likelihood for more complex models (i.e. more parameters).

The AIC and BIC are two examples of penalized likelihood measures.

Model Comparisons – AIC, BIC

AIC – Akaike’s Information Criterion = $2\ell(\theta) - 2k$

BIC – Bayes Information Criterion = $2\ell(\theta) - k\log(n)$

$\ell(\theta)$ = log-likelihood
k = number of parameters

- Use AIC, BIC to compare a series of models. Pick the model with the largest AIC or BIC
- Larger model ⇒ larger likelihood (typically)
- Therefore, “penalize” the likelihood for each added parameter
- AIC tries to find the model that would have the minimum prediction error on a new set of data.
- BIC tries to find the model with the highest “posterior probability” given the data
- Typically, BIC is more conservative (picks smaller models)

Example – AIC, BIC

Continue with the recombinant example. We have N = 20 gametes and Z = 3 recombinants. Let θ be the recombination fraction between the two loci. Recall that the data can be modeled using the binomial distribution:

$$P(Z; ⍰) = \binom{N}{Z}\theta^{Z}(1-\theta)^{N-Z}$$

The situation of no linkage corresponds to θ = 0.5, so we can express the models as

Model 1: θ = 0.5

Model 2: θ anywhere between 0 and 0.5

Example – AIC, BIC

Model 1: The situation of no linkage corresponds to θ = 0.5. If we substitute this into the likelihood equation, we get

$$\begin{aligned} ln\,L_1 &= Z\,ln\,0.5 + (N-Z)\,ln\,0.5 \\ &= N\,ln\,0.5 \end{aligned}$$

This model has 0 (free) parameters.

Model 2: The log-likelihood when θ is unrestricted is

$$ln\,L_2 = Z\,ln\,\theta + (N-Z)\,ln(1-\theta)$$

This model has 1 parameter. Recall, the mle of θ is

$$\hat{\theta} = \frac{Z}{N}$$

If we substitute this back into the log-likelihood, we get …

$$ln\,L_2 = Z\,ln\frac{Z}{N} + (N-Z)ln\,(1-\frac{Z}{N})$$

Example – AIC, BIC

$$\boxed{\begin{aligned} &\text{AIC} = 2\ell(\theta) - 2k \\ &\text{BIC} = 2\ell(\theta) - k\log(n) \end{aligned}}$$

Here are the AIC and BIC calculations for N = 20, Z = 3

$$ln\,L_1 = N\,ln\,0.5 = -13.86$$

$$ln\,L_2 = Z\,ln\frac{Z}{N} + (N-Z)ln\left(1-\frac{Z}{N}\right) = -8.45$$

	L1 (θ = .5)	L2 (θ arb)
θ	.5	.15
Log likelihood	-13.86	-8.45
k	0	1
AIC	**-27.72**	**-18.91**
BIC	**-27.72**	**-19.90**

Bayes Estimation

Recall Bayes theorem (written in terms of data X and parameter θ):

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_\theta P(X|\theta)P(\theta)}$$

Notice the change in perspective - θ is now treated as a random variable instead of a fixed number.

P(X|θ) is the likelihood function, as before.

P(θ) is called the *prior distribution* of θ.

P(θ | X) is called the *posterior distribution* of θ and is used for estimation

Based on P(θ | X) we can define a number of possible estimators of θ. A commonly used estimate is the maximum a posteriori (MAP) estimate:

$$\hat{\theta}_{\text{MAP}} = \max_\theta P(\theta|X)$$

We can also use P(θ | X) to define "credible" intervals for θ.

Bayes Estimation

- The MAP estimator is a very simple Bayes estimator. More generally, Bayes estimators minimize a "loss function" – a penalty based on how far $\hat{\theta}$ is from θ (e.g. Loss =$(\hat{\theta}\ -\theta)^2$).
- The Bayesian procedure provides a convenient way of combining external information or previous data (through the prior distribution) with the current data (through the likelihood) to create a new estimate.
- As N increases, the data (through the likelihood) overwhelms the prior and the Bayes estimator typically converges to the MLE
- Controversy arises when P(θ) is used to incorporate subjective beliefs or opinions.
- If the prior distribution P(θ) is simply that θ is uniformly distributed over all possible values, this is called an "uninformative" prior, and the MAP is the same as the MLE.

Bayes Estimation

Example

Suppose a man is known to have transmitted allele A1 to his child at a locus that has only two alleles: A1 and A2. What is his most likely genotype?

Soln. Let X represent the paternal allele in the child and let θ represent the man's genotype:

$X = A1$

$\theta = \{A1A1,\ A1A2,\ A2A2\}$

We can write the likelihood function as:

$P(X \mid \theta = A1A1) = 1$

$P(X \mid \theta = A1A2) = .5$

$P(X \mid \theta = A2A2) = 0$

Therefore, the MLE is $\theta = A1A1$.

Bayes Estimation

Suppose, however, that we know that the frequency of the A1 allele in the general population is only 1%. Assuming HW equilibrium we have

$P(\theta = A1A1) = .0001$

$P(\theta = A1A2) = .0198$

$P(\theta = A2A2) = .9801$

Also,

$P(X) = \sum_\theta P(X \mid \theta)\, P(\theta) = .01$

This leads to the posterior distribution

$$P(\theta = A1A1 \mid X) = P(X \mid \theta = A1A1)\, P(\theta = A1A1) / P(X) = 1 * .0001 / .01 = \mathbf{.01}$$

$$P(\theta = A1A2 \mid X) = P(X \mid \theta = A1A2)\, P(\theta = A1A2) / P(X) = .5 * .0198 / .01 = \mathbf{.99}$$

$$P(\theta = A2A2 \mid X) = \mathbf{0}$$

So the Bayesian MAP estimator is $\theta = A1A2$.

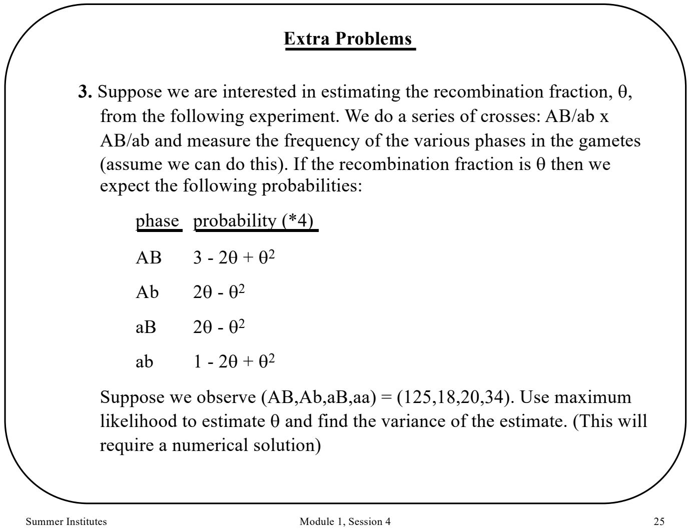
Summary

- Maximum likelihood is a method of estimating parameters from data
- ML requires you to write a probability model for the data
- MLE's may be found analytically or numerically
- (Inverse of the negative of the) second derivative of the log-likelihood gives variance of estimates
- Comparison of log-likelihoods allows us to choose between alternative models
- Bayesian procedures allow us to incorporate additional information about the parameters in the form of prior data, external information or personal beliefs.

Extra Problems

1. Redo the previous problem assuming the man has 2 children who both have the A1 paternal allele.

2. Suppose 197 animals are distributed into five categories with frequencies (95,30,18,20,34). A genetic model for the population predicts the following frequencies for the categories: (.5, .25*p, .25*(1-p), .25*(1-p), .25*p). Use maximum likelihood to estimate p (Hint: use the multinomial distribution).

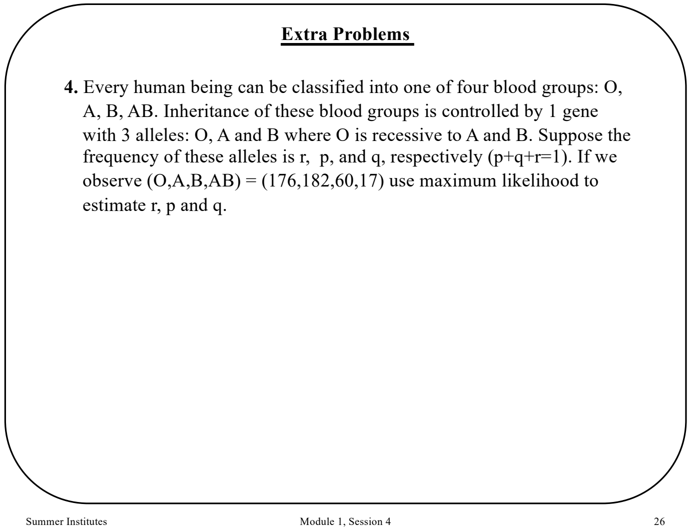


Extra Problems

3. Suppose we are interested in estimating the recombination fraction, θ, from the following experiment. We do a series of crosses: AB/ab x AB/ab and measure the frequency of the various phases in the gametes (assume we can do this). If the recombination fraction is θ then we expect the following probabilities:

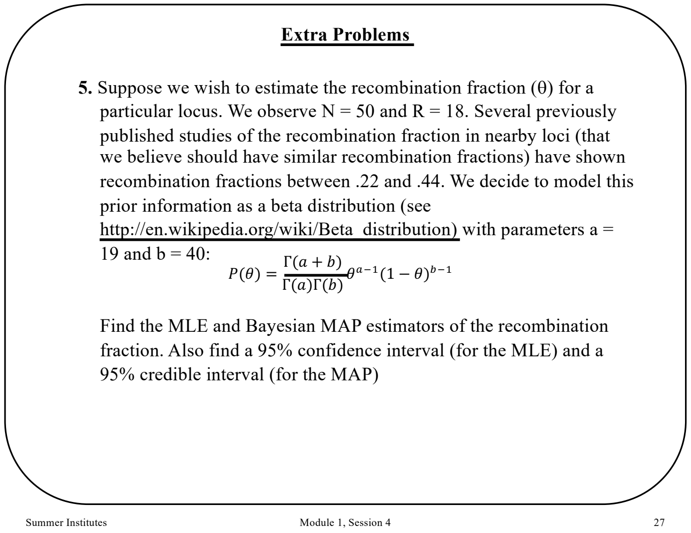
phase	probability (*4)
AB	$3 - 2\theta + \theta^2$
Ab	$2\theta - \theta^2$
aB	$2\theta - \theta^2$
ab	$1 - 2\theta + \theta^2$

Suppose we observe (AB,Ab,aB,aa) = (125,18,20,34). Use maximum likelihood to estimate θ and find the variance of the estimate. (This will require a numerical solution)



Extra Problems

4. Every human being can be classified into one of four blood groups: O, A, B, AB. Inheritance of these blood groups is controlled by 1 gene with 3 alleles: O, A and B where O is recessive to A and B. Suppose the frequency of these alleles is r, p, and q, respectively (p+q+r=1). If we observe (O,A,B,AB) = (176,182,60,17) use maximum likelihood to estimate r, p and q.



Extra Problems

5. Suppose we wish to estimate the recombination fraction (θ) for a particular locus. We observe N = 50 and R = 18. Several previously published studies of the recombination fraction in nearby loci (that we believe should have similar recombination fractions) have shown recombination fractions between .22 and .44. We decide to model this prior information as a beta distribution (see http://en.wikipedia.org/wiki/Beta_distribution) with parameters a = 19 and b = 40:

$$P(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

Find the MLE and Bayesian MAP estimators of the recombination fraction. Also find a 95% confidence interval (for the MLE) and a 95% credible interval (for the MAP)