Fundamental concepts in statistics

Statistical model and the likelihood function

Subhash R. Lele

Department of Mathematical Sciences
University of Alberta
Canada
Email: slele@ualberta.ca

July 13, 2016

Occupancy surveys

1. Divide the study area in equal area cells. Let there be N cells.

Occupancy surveys

1. Divide the study area in equal area cells. Let there be N cells.
2. Take a simple random sample of size n from these cells.

Occupancy surveys

1. Divide the study area in equal area cells. Let there be N cells.
2. Take a simple random sample of size n from these cells.
3. Visit these selected cells and find out if it is occupied by the species or not.

Assumptions

1. Cells are identical to each other.

Assumptions

1. Cells are identical to each other.
2. Occupancy status of one cell does not depend on the occupancy status of the other cells.

Assumptions

1. Cells are identical to each other.
2. Occupancy status of one cell does not depend on the occupancy status of the other cells.
3. Because we cannot visit all selected cells simultaneously, we assume the occupancy status of the cells does not change during the survey period.

Statistical model and notation

We can express the above process in statistical terms as:

$$Y_i \sim Bernoulli(p), i = 1, 2, ..., n. \quad (1)$$

are independent, identically distributed random variables.

$$\text{Sampled cells: } Y_1, Y_2, ..., Y_n$$
$$\text{Unsampled cells: } Y_{(n+1)}, Y_{(n+2)}, ..., Y_N$$

We will not differentiate between data and random variables for the sake of simplicity. The difference should be obvious from the context.

Data: $y_1, y_2, ..., y_n$.

These are the realized values $(0, 0, 1, 1, 0, ..)$

Statistical model: This quantifies the probability of different outcomes. In our case, the possible outcomes are $\{0,1\}$. Hence we use the Bernoulli distribution to model these outcomes. The probability mass function of the Bernoulli random variable is given by $P(Y = y) = p^y(1-p)^{(1-y)}$ where p.

The probability of success, p, is called the parameter of the model. This is generally unknown. However, if we know the value of this parameter, we know the behaviour of the statistical model completely.

One of the goals of the statistical inference is to infer the value of the parameter using the observed data. The **method of maximum likelihood** tells us how this can be done in a very general fashion.

Likelihood function

This is proportional to the probability of observing the data at hand:

$$L(p; y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} \tag{2}$$

We take the product because observations are assumed to be independent of each other. If the observations are dependent, such as in time series, we need to account for dependence appropriately. This will be covered later when we deal with population time series data.

Important properties of the likelihood

1. Likelihood is a function of the parameter.
2. Data are fixed.
3. Likelihood is *not* a probability of the parameter taking a specific value. It represents the probability of observing the data at hand for that particular value of the parameter. Thus, if the parameter is fixed at $\tilde{p}$, then the probability of observing the data at hand is:

$$L(\tilde{p}; y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} \tilde{p}^{y_i}(1-\tilde{p})^{1-y_i}. \tag{3}$$

Important properties of the likelihood (continued)

1. Different data sets lead to different likelihood functions.

Important properties of the likelihood (continued)

1. Different data sets lead to different likelihood functions.
2. As the sample size increases, the likelihood function becomes concentrated around the true value of the parameter. This is an essential property of any estimation procedure. As we get more data, we should have stronger and stronger evidence for the true value.

Important properties of the likelihood (continued)

1. Different data sets lead to different likelihood functions.
2. As the sample size increases, the likelihood function becomes concentrated around the true value of the parameter. This is an essential property of any estimation procedure. As we get more data, we should have stronger and stronger evidence for the true value.
3. Likelihood is an intrinsically relative concept. It answers the question what is the strength of evidence for one parameter values *as compared to* an alternative parameter value. Thus, likelihood *ratio* is a more fundamental concept than the likelihood itself. (Hacking, 1965; Royall 1997).

Goals of statistical inference

1. Given these data, what is the strength of evidence for one hypothesis vis a vis an alternative hypothesis? **(Evidential paradigm)**

Goals of statistical inference

1. Given these data, what is the strength of evidence for one hypothesis vis a vis an alternative hypothesis? **(Evidential paradigm)**
2. Given these data, what decision do I make? **(Neyman-Pearson-Wald paradigm)**

Goals of statistical inference

1. Given these data, what is the strength of evidence for one hypothesis vis a vis an alternative hypothesis? **(Evidential paradigm)**
2. Given these data, what decision do I make? **(Neyman-Pearson-Wald paradigm)**
3. Given these data, how do I change my beliefs? **(Bayesian paradigm)**

Fisherian *p-value* approach falls somewhere between the evidential paradigm and the Neyman-Pearson-Wald paradigm. It lacks the consideration of the alternative hypothesis and hence is inadequate for quantifying evidence (Royall, 1997).

The Maximum Likelihood Estimator (MLE)

1. This is that value of the parameter that is supported the best by the data at hand.

The Maximum Likelihood Estimator (MLE)

1. This is that value of the parameter that is supported the best by the data at hand.
2. For a specific data set, this is simply a number; a point **estimate**.

The Maximum Likelihood Estimator (MLE)

1. This is that value of the parameter that is supported the best by the data at hand.
2. For a specific data set, this is simply a number; a point **estimate**.
3. For different data sets, we get different point estimates. A function that allows us to compute these different values is called an **estimator**.
4. Thus, p is a parameter, $\hat{p} = \bar{Y}$ is an estimator and $\hat{p} = 0.34$ is an estimate.

What is *statistical* about statistical inference?

Cox (1958, Annals of Mathematical Statistics)

- From particular to general: Inductive inference means making statements about population quantities based on the observed sample

What is *statistical* about statistical inference?

Cox (1958, Annals of Mathematical Statistics)

- From particular to general: Inductive inference means making statements about population quantities based on the observed sample
- Quantifying *uncertainty* about such statements is what makes such statements *statistical* inferential statements.

What is *statistical* about statistical inference?

Cox (1958, Annals of Mathematical Statistics)

- From particular to general: Inductive inference means making statements about population quantities based on the observed sample
- Quantifying *uncertainty* about such statements is what makes such statements *statistical* inferential statements.
- How do we quantify uncertainty? Traditionally we use probability as a measure of uncertainty.

Different definitions of probability

1. **Frequency definition** Long range proportion in a repeatable experiment

Different definitions of probability

1. **Frequency definition** Long range proportion in a repeatable experiment
2. **Betting definition** How much would one bet on one event as compared to other event?

Different definitions of probability

1. **Frequency definition** Long range proportion in a repeatable experiment
2. **Betting definition** How much would one bet on one event as compared to other event?

There are other systems of probability *(Stanford encyclopedia of philosophy of science)*. We will use only these two for our discussion.

Frequentist quantification of uncertainty

If we repeat the experiment (whatever that might be), how often would my inferential statement be contradicted? Is my statement replicable?

1. **Confidence intervals** What is the range of values would the statement take?

Frequentist quantification of uncertainty

If we repeat the experiment (whatever that might be), how often would my inferential statement be contradicted? Is my statement replicable?

1. **Confidence intervals** What is the range of values would the statement take?
2. **Hypothesis testing and errors** How often would I make a mistake (Type I) or how often would I be correct (Power)?

Frequentist quantification of uncertainty

If we repeat the experiment (whatever that might be), how often would my inferential statement be contradicted? Is my statement replicable?

1. **Confidence intervals** What is the range of values would the statement take?
2. **Hypothesis testing and errors** How often would I make a mistake (Type I) or how often would I be correct (Power)?
3. **Misleading, Weak and Strong evidence** How often would I be misled? How often would I say 'I do not know'?

1. **Sampling distribution of an estimator**: Suppose we repeat the experiment, we will get different point estimate for each repetition. The histogram of these point estimates is called the *sampling distribution* of the estimator. We may report the mean and quantiles of such a distribution. It characterizes the frequentist uncertainty of the estimator.
2. **Estimated sampling distribution of the estimator**: In practice, of course, we do not have replicated experiments. We can, instead, use the point estimate from the current data as if it is the true parameter value and replicate the experiment under that assumption. This leads to Monte Carlo estimate of the sampling distribution. This is also called the 'parametric bootstrap'.
3. **Non-parametric bootstrap approach**: Instead of assuming the parametric model, if we simply resample with replacement from the data to generate new samples, we obtain a model-robust estimate of the sampling distribution.

Bias, Consistency and validity of the estimated confidence intervals

1. If the estimator on an average equals the true parameter value, it is called an unbiased estimator.

Bias, Consistency and validity of the estimated confidence intervals

1. If the estimator on an average equals the true parameter value, it is called an unbiased estimator.
2. If the estimator converges to the true value as the sample size increases, it is called a 'consistent' estimator. This is an essential property of ANY statistical procedure.

Bias, Consistency and validity of the estimated confidence intervals

1. If the estimator on an average equals the true parameter value, it is called an unbiased estimator.
2. If the estimator converges to the true value as the sample size increases, it is called a 'consistent' estimator. This is an essential property of ANY statistical procedure.
3. An estimator that converges to the true value fastest, is called an 'efficient' estimator. This can be, most of the times, characterized by the variance of the sampling distribution. Smaller the variance, better is the estimator.

Bias, Consistency and validity of the estimated confidence intervals

1. If the estimator on an average equals the true parameter value, it is called an unbiased estimator.
2. If the estimator converges to the true value as the sample size increases, it is called a 'consistent' estimator. This is an essential property of ANY statistical procedure.
3. An estimator that converges to the true value fastest, is called an 'efficient' estimator. This can be, most of the times, characterized by the variance of the sampling distribution. Smaller the variance, better is the estimator.
4. How often the true parameter value lies inside the estimated confidence interval determines the validity of the confidence interval. It should cover the true value close to the stated coverage proportion.

Bayesian paradigm

- Inference should be based on the data at hand and should not depend on what other data I might have observed (roughly, *the likelihood principle*)

Bayesian paradigm

- Inference should be based on the data at hand and should not depend on what other data I might have observed (roughly, *the likelihood principle*)
- Where does the uncertainty come from?

Bayesian paradigm

- Inference should be based on the data at hand and should not depend on what other data I might have observed (roughly, *the likelihood principle*)
- Where does the uncertainty come from?
- Uncertainty is in our beliefs about the values of the parameters.

Bayesian paradigm

- Inference should be based on the data at hand and should not depend on what other data I might have observed (roughly, *the likelihood principle*)
- Where does the uncertainty come from?
- Uncertainty is in our beliefs about the values of the parameters.
- When we have no data, this uncertainty is quantified by the prior distribution $\pi(\theta)$.

Bayesian paradigm

- Inference should be based on the data at hand and should not depend on what other data I might have observed (roughly, *the likelihood principle*)
- Where does the uncertainty come from?
- Uncertainty is in our beliefs about the values of the parameters.
- When we have no data, this uncertainty is quantified by the prior distribution $\pi(\theta)$.
- When we observe some data, this uncertainty is quantified by the posterior distribution

$$\pi(\theta|y_n) = \frac{f(\mathbf{y_n}|\theta)\pi(\theta)}{\int \mathbf{f}(\mathbf{y_n}|\theta)\pi(\theta)d\theta} \tag{4}$$

1. We need to quantify our beliefs in terms of a probability distribution.

1. We need to quantify our beliefs in terms of a probability distribution.
2. We need to be able to compute the posterior distribution.

1. We need to quantify our beliefs in terms of a probability distribution.
2. We need to be able to compute the posterior distribution.
3. This is the uncertainty of our beliefs about different values of the parameters after we observe the data.

1. We need to quantify our beliefs in terms of a probability distribution.
2. We need to be able to compute the posterior distribution.
3. This is the uncertainty of our beliefs about different values of the parameters after we observe the data.
4. The posterior distribution, obviously, depends not just on the data at hand (unlike the likelihood function) but also depends on the prior distribution. Different prior distributions lead to different posterior distributions.

1. We need to quantify our beliefs in terms of a probability distribution.
2. We need to be able to compute the posterior distribution.
3. This is the uncertainty of our beliefs about different values of the parameters after we observe the data.
4. The posterior distribution, obviously, depends not just on the data at hand (unlike the likelihood function) but also depends on the prior distribution. Different prior distributions lead to different posterior distributions.
5. All we need to know is the posterior distribution. Any of the inferential statements can be obtained from the posterior distribution. (Examples later).

Properties of the posterior distribution

1. Same data, different priors lead to different posteriors.

Properties of the posterior distribution

1. Same data, different priors lead to different posteriors.
2. Same prior, different data lead to different posteriors.

Properties of the posterior distribution

1. Same data, different priors lead to different posteriors.
2. Same prior, different data lead to different posteriors.
3. As the sample size increases, the posterior distribution is invariant to the prior and it eventually degenerates at the true value (phew!).

Bayesian analysis of occupancy problem

- Prior distribution: The parameter p takes values in the range (0,1) and hence the prior distribution should also have the same range. Hence we MAY consider

$$\pi(p) \sim Beta(a, b) \tag{5}$$

  We can choose different values for a and b. They will reflect different beliefs about the occupancy probability.

- The model for the observed data is as before:

$$Y_i|p \sim Bernoulli(p) fori = 1, 2, ..., n \tag{6}$$

- The posterior distribution can be analytically computed as:

$$\pi(p|y_{(n)}) \sim Beta(a + \sum y_i, b + (n - \sum y_i)) \tag{7}$$

- Shapes of different priors

- Shapes of different priors
- Shapes of different posteriors for the same data (prior affects the posterior)

- Shapes of different priors
- Shapes of different posteriors for the same data (prior affects the posterior)
- Shapes of the posterior as we increase the sample size (information in the data)

- Nuisance parameters: If we are interested in only one of the parameters, we simply find the marginal posterior distribution for it by integrating over the rest of the parameters.

$$\pi(\theta_2|\mathbf{y_n}) = \int \pi(\theta_2, \theta_1|\mathbf{y_n}) d\theta_1 \tag{8}$$

- Nuisance parameters: If we are interested in only one of the parameters, we simply find the marginal posterior distribution for it by integrating over the rest of the parameters.

$$\pi(\theta_2|\mathbf{y_n}) = \int \pi(\theta_2, \theta_1|\mathbf{y_n}) d\theta_1 \tag{8}$$

- Prediction of a new observation: We can simply write

$$\pi(y_{n+1}|\mathbf{y_n}) = \int f(y_{n+1}|\theta)\pi(\theta|\mathbf{y_n}) d\theta \tag{9}$$

All these operations follow from the standard probability calculus.

Non-informative priors

1. Dependence of the scientific inference on the beliefs of the experimenter is bothersome to many scientists. It can bias the conclusions. Whose belief should we accept: An environmental activist or an oil industry expert or a 'scientist'? How do we justify these beliefs in the court of law where many ecological studies end up.

Non-informative priors

1. Dependence of the scientific inference on the beliefs of the experimenter is bothersome to many scientists. It can bias the conclusions. Whose belief should we accept: An environmental activist or an oil industry expert or a 'scientist'? How do we justify these beliefs in the court of law where many ecological studies end up.
2. Can we construct priors so that the posterior is mostly affected by the data (the likelihood)?

Non-informative priors

1. Dependence of the scientific inference on the beliefs of the experimenter is bothersome to many scientists. It can bias the conclusions. Whose belief should we accept: An environmental activist or an oil industry expert or a 'scientist'? How do we justify these beliefs in the court of law where many ecological studies end up.
2. Can we construct priors so that the posterior is mostly affected by the data (the likelihood)?
3. Such priors are called non-informative priors and the analysis is euphemistically called an 'objective Bayesian' analysis.

Non-informative priors

1. Dependence of the scientific inference on the beliefs of the experimenter is bothersome to many scientists. It can bias the conclusions. Whose belief should we accept: An environmental activist or an oil industry expert or a 'scientist'? How do we justify these beliefs in the court of law where many ecological studies end up.
2. Can we construct priors so that the posterior is mostly affected by the data (the likelihood)?
3. Such priors are called non-informative priors and the analysis is euphemistically called an 'objective Bayesian' analysis.
4. Unfortunately, it is impossible to construct such non-informative priors (even according to the staunchest Bayesians).

Non-informative priors

1. Dependence of the scientific inference on the beliefs of the experimenter is bothersome to many scientists. It can bias the conclusions. Whose belief should we accept: An environmental activist or an oil industry expert or a 'scientist'? How do we justify these beliefs in the court of law where many ecological studies end up.
2. Can we construct priors so that the posterior is mostly affected by the data (the likelihood)?
3. Such priors are called non-informative priors and the analysis is euphemistically called an 'objective Bayesian' analysis.
4. Unfortunately, it is impossible to construct such non-informative priors (even according to the staunchest Bayesians).

Let us look at shapes of some of the 'non-informative' priors commonly used in occupancy models.

- Uniform prior or $Beta(1, 1)$ prior.

Let us look at shapes of some of the 'non-informative' priors commonly used in occupancy models.

- Uniform prior or $Beta(1, 1)$ prior.
- Jeffrey's prior or $Beta(0.5, 0.5)$ prior

Let us look at shapes of some of the 'non-informative' priors commonly used in occupancy models.

- Uniform prior or $Beta(1, 1)$ prior.
- Jeffrey's prior or $Beta(0.5, 0.5)$ prior
- Normal prior with large variance on the logit scale: In most applications, we use logistic regression model to model the probability of occupancy and such priors are specified on the regression coefficients.

Let us look at shapes of some of the 'non-informative' priors commonly used in occupancy models.

- Uniform prior or $Beta(1, 1)$ prior.
- Jeffrey's prior or $Beta(0.5, 0.5)$ prior
- Normal prior with large variance on the logit scale: In most applications, we use logistic regression model to model the probability of occupancy and such priors are specified on the regression coefficients.
- What are the induced priors on the probability scale?

Let us look at shapes of some of the 'non-informative' priors commonly used in occupancy models.

- Uniform prior or $Beta(1, 1)$ prior.
- Jeffrey's prior or $Beta(0.5, 0.5)$ prior
- Normal prior with large variance on the logit scale: In most applications, we use logistic regression model to model the probability of occupancy and such priors are specified on the regression coefficients.
- What are the induced priors on the probability scale?
- What is 'non-informative' on the logit scale is highly informative on the probability scale. This non-invariance to reparameterization is a major problem with the concept of 'non-informative' or 'objective' priors. It has strong implications in scientific inference and decision making.

Hierarchical models

For hierarchical models, we have observables $\mathbf{y_n}$, latent variables$\mathbf{X_n}$ and parameters θ.

- Observation model: $\mathbf{y_n} \sim f(\mathbf{y_n}|\mathbf{X_n}, \theta)$
- Latent variable model: $\mathbf{X_n} \sim g(\mathbf{X_n}|\theta)$
- Prior distribution: $\pi(\theta)$

- We need to define a prior distribution is on ALL unknowns. Bayesian inference does not differentiate between parameters and random variables.
- Information about parameters converges to infinity as the sample size increases.
- Information about random variables does not converge to infinity. There is always uncertainty about the next observation.
- Random effects are NOT parameters. They are variables.

Bayesian inference for hierarchical models is quite easy. All we need is the posterior distribution of the unknowns given the knowns. That is:

$$\pi(\theta, \mathbf{X_n}|\mathbf{y_n}) = \frac{f(\mathbf{y_n}|\theta, \mathbf{X_n})g(\mathbf{X_n}|\theta)\pi(\theta)}{m(\mathbf{y_n})} \tag{10}$$

The MCMC algorithms allow us to generate random numbers from the posterior distribution. These can be used to obtain credible intervals and other inferential quantities.

Complete class theorem says that the best decision is necessarily a Bayesian decision.

Bayesian inference for hierarchical models is quite easy. All we need is the posterior distribution of the unknowns given the knowns. That is:

$$\pi(\theta, \mathbf{X_n}|\mathbf{y_n}) = \frac{f(\mathbf{y_n}|\theta, \mathbf{X_n})g(\mathbf{X_n}|\theta)\pi(\theta)}{m(\mathbf{y_n})} \tag{10}$$

The MCMC algorithms allow us to generate random numbers from the posterior distribution. These can be used to obtain credible intervals and other inferential quantities.

Complete class theorem says that the best decision is necessarily a Bayesian decision.

PROBLEM SOLVED?

- Let us see how to use computer software JAGS and 'dclone' to conduct Bayesian inference using MCMC for the occupancy model we have discussed.

- Let us see how to use computer software JAGS and 'dclone' to conduct Bayesian inference using MCMC for the occupancy model we have discussed.
- Please make sure that the algorithm has converged. You can use the time series plots and R-hat diagnostics; however these are NOT fool-proof tests. There are no such test available.

- The likelihood inference for hierarchical models is computationally very challenging because the evaluating the likelihood function involves evaluation of the high-dimensional integral:

- The likelihood inference for hierarchical models is computationally very challenging because the evaluating the likelihood function involves evaluation of the high-dimensional integral:
- On the other hand, Bayesian inference is relatively computationally easy. Hence the popularity of the Bayesian inference in ecology.

- The likelihood inference for hierarchical models is computationally very challenging because the evaluating the likelihood function involves evaluation of the high-dimensional integral:
- On the other hand, Bayesian inference is relatively computationally easy. Hence the popularity of the Bayesian inference in ecology.
- However, the thorny problem of specifying the prior (informative or non-informative) distribution and its effect on the final scientific inference remains.

- The likelihood inference for hierarchical models is computationally very challenging because the evaluating the likelihood function involves evaluation of the high-dimensional integral:
- On the other hand, Bayesian inference is relatively computationally easy. Hence the popularity of the Bayesian inference in ecology.
- However, the thorny problem of specifying the prior (informative or non-informative) distribution and its effect on the final scientific inference remains.
- Can we trick Bayesian approach into giving frequentist answers? Thus, exploiting the computational simplicity of the Bayesian approach and still retaining the objectivity of the frequentist approach?

- The likelihood inference for hierarchical models is computationally very challenging because the evaluating the likelihood function involves evaluation of the high-dimensional integral:
- On the other hand, Bayesian inference is relatively computationally easy. Hence the popularity of the Bayesian inference in ecology.
- However, the thorny problem of specifying the prior (informative or non-informative) distribution and its effect on the final scientific inference remains.
- Can we trick Bayesian approach into giving frequentist answers? Thus, exploiting the computational simplicity of the Bayesian approach and still retaining the objectivity of the frequentist approach?
- DATA CLONING!

A brief introduction to data cloning

The basic idea behind data cloning is really very simple. Recall that as we increase the sample size, the posterior distribution becomes degenerate at the MLE. Furthermore, it is also known that the posterior distribution converges to $N(\hat{\theta}, I^{-1}(\hat{\theta}))$.

1. Pretend as if we have conducted K independent experiments and each experiment ended up with exactly the same data.

A brief introduction to data cloning

The basic idea behind data cloning is really very simple. Recall that as we increase the sample size, the posterior distribution becomes degenerate at the MLE. Furthermore, it is also known that the posterior distribution converges to $N(\hat{\theta}, I^{-1}(\hat{\theta}))$.

1. Pretend as if we have conducted K independent experiments and each experiment ended up with exactly the same data.
2. Thus, we have K copies of the same data (hence the name 'data cloning'). Now apply the MCMC algorithm on this cloned data.

A brief introduction to data cloning

The basic idea behind data cloning is really very simple. Recall that as we increase the sample size, the posterior distribution becomes degenerate at the MLE. Furthermore, it is also known that the posterior distribution converges to $N(\hat{\theta}, I^{-1}(\hat{\theta}))$.

1. Pretend as if we have conducted K independent experiments and each experiment ended up with exactly the same data.
2. Thus, we have K copies of the same data (hence the name 'data cloning'). Now apply the MCMC algorithm on this cloned data.
3. If the number of clones is large, mean of the posterior distribution is the MLE and the variance multiplied by K is the asymptotic variance of the MLE.

A brief introduction to data cloning

The basic idea behind data cloning is really very simple. Recall that as we increase the sample size, the posterior distribution becomes degenerate at the MLE. Furthermore, it is also known that the posterior distribution converges to $N(\hat{\theta}, I^{-1}(\hat{\theta}))$.

1. Pretend as if we have conducted K independent experiments and each experiment ended up with exactly the same data.
2. Thus, we have K copies of the same data (hence the name 'data cloning'). Now apply the MCMC algorithm on this cloned data.
3. If the number of clones is large, mean of the posterior distribution is the MLE and the variance multiplied by K is the asymptotic variance of the MLE.
4. This is invariant to the choice of the prior distribution. Prior distribution is simply a collection of guesses at the MLE.

Estimability of the parameters

A very important outcome of the data cloning algorithm is a test for estimability of the parameters in the model. Just because you can write the likelihood function and maximize it; it does not imply that the parameters are estimable given the data. We may end up on a ridge in the likelihood. This is difficult to prove mathematically, especially for hierarchical models.

Estimability of the parameters

A very important outcome of the data cloning algorithm is a test for estimability of the parameters in the model. Just because you can write the likelihood function and maximize it; it does not imply that the parameters are estimable given the data. We may end up on a ridge in the likelihood. This is difficult to prove mathematically, especially for hierarchical models.

If the variance of the posterior distribution, instead of converging to 0 converges to a positive number, it implies that some of the parameters in the model are non-estimable. This should raise a very big red flag when interpreting the results of the analysis.

Conducting data cloning based likelihood analysis is quite simple. A very minor modification of the JAGS program and the data file allows us to obtain the MLE and its variance. Let us see how we can do it for the occupancy model.

SUMMARY

1. Statistical inference is an inductive process. Given the data, we want to infer about the mechanism that could have generated the data.

SUMMARY

1. Statistical inference is an inductive process. Given the data, we want to infer about the mechanism that could have generated the data.
2. Estimation corresponds to finding the best supported mechanism out of the proposed alternative mechanisms.

SUMMARY

1. Statistical inference is an inductive process. Given the data, we want to infer about the mechanism that could have generated the data.
2. Estimation corresponds to finding the best supported mechanism out of the proposed alternative mechanisms.
3. An inferential statement is statistical only when we attach a measure of uncertainty to it.

SUMMARY

1. Statistical inference is an inductive process. Given the data, we want to infer about the mechanism that could have generated the data.
2. Estimation corresponds to finding the best supported mechanism out of the proposed alternative mechanisms.
3. An inferential statement is statistical only when we attach a measure of uncertainty to it.
4. There are two ways to attach such uncertainty measures: Frequentist and Bayesian.

1. Frequentist approach requires the researcher to specify how the experiment might be replicated (at least potentially). It tells you how reliable are our statements.

1. Frequentist approach requires the researcher to specify how the experiment might be replicated (at least potentially). It tells you how reliable are our statements.
2. Bayesian approach requires the researcher to specify prior beliefs and then it tells you what your modified beliefs should be.

1. Frequentist approach requires the researcher to specify how the experiment might be replicated (at least potentially). It tells you how reliable are our statements.
2. Bayesian approach requires the researcher to specify prior beliefs and then it tells you what your modified beliefs should be.
3. Both inferences can be applied for hierarchical models. Models are neither Bayesian nor Frequentist; only Bayesian or Frequentist inferences.