Confidence Intervals and Probability Testing

PO7001: Quantitative Methods I
Kenneth Benoit

3 November 2010
Recall that using the $\mu$ and $\sigma$ from a normal curve, we can then assess the probability of finding specific scores along this distribution (as we did last week).

The question: *If we assume $\mu$ equals some specific value, then how likely was it to have drawn a given sample mean $\bar{X}$?*

A variation would be: If we draw two sample means, what is the probability that they have identical population means?
Confidence intervals

- Recall that a confidence interval is the range of values within which our population mean is most likely to fall.
- Confidence intervals are expressed as a percentage of confidence: If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter.
- Confidence intervals can be:
  - parametric: based on either assumed probability distributions
  - nonparametric: obtained through simulation (e.g. bootstrapping)
Confidence intervals for proportions

- Often we may seek to estimate a population proportion on the basis of a random sample: e.g. proportion to vote Yes on the Lisbon Treaty referendum
- We use a version of the standard error known as the standard error of the proportion:

$$s_p = \sqrt{\frac{p(1 - p)}{N}}$$

where
- $s_p$ is the standard error of the proportion
- $p$ is the sample proportion
- $N$ is the total number of cases in the sample

- Because proportions have only a single parameter $\pi$ of which $p$ is an estimate, we can use the normal distribution and not $t$
- So: 95% CI for proportions $= p \pm 1.96s_p$
Say, we want to look at the voter support for Vladimir Putin. 66% of 1600 respondents in a survey say they will vote for Putin. 

\[
SE(\hat{p}) = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{(0.66 \times 0.34)/1600} = 0.012
\]

\[
CI_{95\%} = [\hat{p} - 2 \times SE(\hat{p}); \hat{p} + 2 \times SE(\hat{p})] = [0.637; 0.683]
\]
Confidence intervals for proportions: example

**Figure:** Confidence interval of support for Putin
Confidence intervals for proportions: example

\texttt{qnorm(c(.025, .975), .66, sqrt(.66 \times .34 / 1600))}

\texttt{prop.test(.66 \times 1600, 1600)}
Hypothesis testing

Hypothesis testing refers to a method of making statistical decisions using experimental data: a result is called statistically significant if it is unlikely to have occurred by chance.

The null hypothesis ($H_0$) is the proposition that we will try to reject. It usually states the assumption that there is no effect or difference (depending on the research question).

The null hypothesis is stated as the opposite of what we would like to prove (more shortly).

The alternative hypothesis ($H_1$ or $H_a$) states the research hypothesis which we will test assuming the null hypothesis.
Steps in hypothesis testing

1. Establish a hypothesis about population(s)

2. Collect sample data

3. See how likely the sample data is, given hypothesis about the population

4. If sample results are reasonably plausible under the hypothesis about the population, then retain hypothesis

5. If sample results are very unlikely given hypothesis, we reject our (null) hypothesis (more on this shortly)
Comparisons of groups

- An extremely common goal of political/social: to assess differences between groups
- Examples:
  - Do women vote more often than men?
  - Do Fianna Fail members support the Lisbon Treaty more than Fine Gael members?
  - Do proportional representation electoral systems tend to result in more parties in parliament than single-member district electoral systems?
- Basic method: to use the $t$-distribution theory of sampling means to compare how likely it is that two sample means differ only because of sampling error
Sampling error revisited

- Recall: We may *expect* a sample statistic to equal a population parameter, but in each individual sample, we are likely to observe differences.

- Recall: These differences will vary each time a sample is drawn, even if in the long run, the average difference will be zero.

- Recall: These differences are referred to as **sampling error**: the difference caused by chance differences between the sample’s characteristics and those in the population.
### Example

<table>
<thead>
<tr>
<th>Method A</th>
<th>Method B</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>78</td>
</tr>
<tr>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>80</td>
<td>78</td>
</tr>
<tr>
<td>83</td>
<td>76</td>
</tr>
</tbody>
</table>

Mean = 82 Mean = 77  
SD = 1.22 SD = 1.00

- **Question:** are the population means for method A and B different, or in fact the same?

- To conclude they are in fact the same is to conclude that observed differences arise *only from sampling error*.

- Because both sets of scores are fairly tightly clustered around their means, we can reject the (null) hypothesis that they are in fact from the same populations.
Example cont.

First set

N = 5   Bandwidth = 0.4868
Example (version 2)

<table>
<thead>
<tr>
<th>Method A</th>
<th>Method B</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>98</td>
<td>90</td>
</tr>
<tr>
<td>63</td>
<td>91</td>
</tr>
<tr>
<td>74</td>
<td>56</td>
</tr>
<tr>
<td>85</td>
<td>78</td>
</tr>
</tbody>
</table>

Mean = 82  Mean = 77  
SD = 13.73  SD = 14.63

▶ Same question: are the population means for method A and B different, or in fact the same?

▶ Here the two groups have the same means, but much larger standard deviations – and hence larger sampling variances

▶ Because of the overlap of the confidence intervals, we cannot reject the (null) hypothesis that the samples are in fact from the same populations
Example (version 2) cont.

Second set

N = 5   Bandwidth = 7.789
Comparing the two sample distributions

Note: This type of plot is known as a strip plot

The means of A1, A2 are same, as are B1, B2
Null hypothesis: no difference between means

- Remember: the null hypothesis asserts something you would like to disprove
- In this context:
  \[ H_0 : \mu_1 = \mu_2 \]
- Note: This does not mean we expect to find no differences in sample means – quite the contrary, \( H_0 \) asserts that any difference will be due only to sampling error
- We will see if we can amass enough evidence to reject the null hypothesis; otherwise, we will retain it
- If we reject the null hypothesis, then we (automatically) accept the research or alternative hypothesis that a true difference exists:
  \[ H_A : \mu_1 \neq \mu_2 \]
Sampling distribution of differences between means

- Just as the $t$-distribution is the sampling distribution of sample means, we can also derive a sampling distribution of the difference in sample means.

- (Naturally) we would expect the probability to become smaller and smaller as we move away from observed differences of zero, given $H_0 : \mu_1 = \mu_2$.

- We can express the quantities as:
  - $\bar{X}_1$ = sample mean of group 1
  - $\bar{X}_2$ = sample mean of group 2
  - $\sigma_{\bar{X}_1 - \bar{X}_2}$ as the standard deviation of differences between sample means of group 1 and group 2.
Probability distribution of differences between means
Differences between means: Formulas

- Observe a set of sample mean differences \((\bar{X}_1 - \bar{X}_2)\)
- Convert these into units of standard deviations using:

\[
z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}
\]

where:

- \(\bar{X}_1\) is the mean of sample 1
- \(\bar{X}_2\) is the mean of sample 2
- \((\mu_1 - \mu_2) = 0\) is the value of the mean of the sampling distribution of differences between means, assuming that \((\mu_1 - \mu_2) = 0\)
- \(\sigma_{\bar{X}_1 - \bar{X}_2}\) is the standard deviation of the sampling distribution of differences between means

- Simplified:

\[
z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}
\]
Statistically significant differences

- To establish whether observed differences represent real differences (in underlying populations) we will need some standard.
- The standard tells us how many times we are willing to reject null hypotheses that are actually true.
- Put another way, we need a standard for how strict we are going to be about the possibility of observing sample differences due to chance, if we decide to reject the null hypothesis of no difference.
- This concept is known as statistical significance.
- The level of significance is denoted by $\alpha$, which is the level of probability at which $H_0$ can be rejected with confidence and $H_A$ accepted with confidence.
- By convention, this is usually set at $\alpha = .05$.
- The probability of rejecting $H_0$ with $\alpha = .05$ is $p \leq .05$. 
Type I and Type II errors

- Whenever we decide to reject \( H_0 \) at a given \( \alpha \), we risk wrongly rejecting null hypothesis \( H_0 \) that is actually \textit{true}.
- **Type I error**: rejecting \( H_0 \) when we should have retained it.
- The counterpart is the risk of retaining \( H_0 \) when in fact it is \textit{false} – this is known as **Type II error** and is denoted by \( \beta \).
- Type I and Type II errors are inversely related.
## Type I and Type II errors

<table>
<thead>
<tr>
<th>Reality</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null hypothesis is true</td>
<td>Retain null hypothesis</td>
</tr>
<tr>
<td>Null hypothesis is false</td>
<td>Reject null hypothesis</td>
</tr>
<tr>
<td>Null hypothesis is false</td>
<td>Type II error $P$(Type II error) = $\beta$</td>
</tr>
</tbody>
</table>
Errors and statistical decision-making

- Solution to Type II error is to increase the sample size so that we are more able to detect true population differences – otherwise this is not really within the researcher’s control.
- Type I error is within researcher’s control – it is our decision rule for rejecting $H_0$.
- We *always* run the risk of Type I or Type II errors, depending on our decision about unknown parameters.
- We balance between the two types of errors depending on which is more costly or damaging and therefore riskier.
  - For example: differences in abilities (on standardized tests) between men and women? (Type I error worse)
  - Another example: Alcohol’s negative effects on foetuses in pregnant women (Type II error worse)
- Convention in social science is $\alpha = 0.05$. 
Hypothesis testing: typical critical values

The alpha-level refers to the probability that we make a Type-I error. Typical alpha values:

- $\alpha = .05$: most common in the social sciences
- $\alpha = .01$: more common in pharmaceutical research
- $\alpha = .10$: less picky social scientists
Difference between $p$ and $\alpha$

- $p$ is the exact probability that we would have obtained our sample data given $H_0$
- $\alpha$ is the threshold for $p$ below which is considered so small that we ought to reject $H_0$
- $\alpha$ is decided ahead of time by the researcher
- $p$ comes from the sample data and refers to the size of the tail regions (e.g. under a $t$ or $z$) distribution that determine whether we reject $H_0$
- We do not need the exact $p$-value to decide whether to accept or reject $H_0$, but exact $p$-values are usually reported anyway
Standard error of the difference between means

- The formula is:

\[ \sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma^2_{X_1}}{n_{X_1}} + \frac{\sigma^2_{X_2}}{n_{X_2}}} \]

- Just as with single samples, we do not actually know any of the population variances \( \sigma_{X_1}, \sigma_{\bar{X}_1}, \) etc.

- So we estimate this from the sample, just as with single-sample tests. This quantity is known as the standard error of the difference between means denoted \( s_{\bar{X}_1-\bar{X}_2} \)

- The precise formula depends on whether we can assume equal population variances

- Varies according to how much each sample varies, as well as the size of the difference between the two means

- In R: \( \text{t.test}(x, y, \text{var.equal}=\text{TRUE}) \) or \( \text{t.test}(x, y, \text{var.equal}=\text{FALSE}) \)
Standard error of the difference between means: Equal population variances

If the samples are assumed to be drawn from populations with equal variances (where we are simply testing for a difference in means) then, the formula pools the data and estimates $\hat{\sigma} = \hat{\sigma}_{x_1} = \hat{\sigma}_{x_2} = s_{\bar{x}}$:

$$s_{\bar{x}} = \sqrt{\frac{(n_{x_1} - 1)s_{x_1}^2 + (n_{x_2} - 1)s_{x_2}^2}{n_{x_1} + n_{x_2} - 2}}$$

and then

$$s_{\bar{x}_1 - \bar{x}_2} = s_{\bar{x}} \sqrt{\frac{1}{n_{x_1}} + \frac{1}{n_{x_2}}}$$

to get a sample statistic from the t-test that has $n_{x_1} + n_{x_2} - 2$ degrees of freedom.
Standard error of the difference between means: Unequal population variances

If the samples are assumed to be drawn from populations with different variances, then we estimate $\sigma_{x_1}$ and $\sigma_{x_2}$ separately using $s_{\bar{x}_1}$ and $s_{\bar{x}_1}$, using the sample standard deviations, e.g.

$$s_{\bar{x}_1} = \sqrt{\frac{s_{x_1}}{n_1}}$$

where $s_{x_1}$ is the sample standard deviation of the sample data $x_1$ and then

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_{\bar{x}_1}^2}{n_{\bar{x}_1}} + \frac{s_{\bar{x}_1}^2}{n_{x_2}}}$$

to get a sample statistic from the $t$-test whose degrees of freedom are computed using the Welch approximation (see Verzani p202)
Testing the difference between means (same sample twice)

- Frequently occurs when we have a before-after design, for instance
- The computation of standard deviation of the difference is:
  \[ s_D = \sqrt{\frac{\sum_i D_i^2}{N} - (\bar{X}_1 - \bar{X}_2)^2} \]

  where \( D_i \) is the difference between post and pre score values
- From this we compute the standard error of the difference between means:
  \[ s_{\bar{D}} = \frac{s_D}{\sqrt{N - 1}} \]
Two-sample test of proportions

Rather than the \( z \) formula we used for testing the difference between means (with known \( \sigma_1 \) and \( \sigma_2 \)), we use

\[
z = \frac{P_1 - P_2}{s_{P_1-P_2}}
\]

where \( P_1 \) and \( P_2 \) are the respective sample proportions. The standard error of the difference in proportions is given by

\[
s_{P_1-P_2} = \sqrt{P^*(1 - P^*)\left(\frac{N_1 + N_2}{N_1N_2}\right)}
\]

where \( P^* \) is the combined sample proportion,

\[
P^* = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}
\]
Assumptions for tests of two means

Question: When is the $t$-test appropriate? The $t$-test is only appropriate when:

1. Performing a comparison between two means
2. Interval data
3. Random sampling
4. Normal distribution of population data, at least for small samples
5. Equal variances (although an alternative formula does exist for unequal variances)
Additional tests on two samples

- Test of two variances.
  - Fisher’s test of equal variances: divide larger by smaller variance. If the ratio $> 1$ then they must be different, where “different” is determined by critical values of the $F$ distribution.
  - In R, this is `var.test(a, b)`.

- If two samples are not normally distributed, instead of the $t$-test, we use non-parametric equivalent called the Wilcoxon rank sum test (or signed-rank test) (see `wilcox.test(a, b)`).

- There is also a single-sample median signed-rank test,

  $$T = \sum_{i: X_i > m} \text{rank}(|X_i - m|)$$

  where $m$ is a specific median such that $H_0 : \text{median} = m$. 

Two-sample tests in R

- *t*-test is `t.test(a,b,paired = FALSE, var.equal = FALSE)
  - `paired` is for independent versus before-after designs
  - `var.equal` is for assumption of equal variances (false by default) or not

- Test for equality of proportions is `prop.test(x, n)`

- Comparing two samples without normality assumption using Wilcoxon rank sum test is `wilcox.test(a,b)`

- Comparing the variances of two samples using Fisher’s test is `var.test(a,b)`
Confidence intervals and sample sizes

- By the Central Limit Theorem: $SE(\bar{x}) = \frac{s}{\sqrt{n}}$
- The parametric confidence interval is then

$$CI_{95\%} = [\bar{x} - 1.96 \times SE(\bar{x}); \bar{x} + 1.96 \times SE(\bar{x})]$$

- So the larger the sample, the smaller the confidence interval:

$$CI_{95\%} = \bar{x} \pm 1.96 \times \frac{s}{\sqrt{n}}$$
Calculating sample size

- Using the mathematical computation for confidence intervals, we can determine how many observations we will need to estimate a quantity for a given confidence level.
- Very common in experimental or clinical studies, although much less common (virtually non-existent) in social science statistics.
Calculating sample size: Example

- Suppose that want to estimate the number of seats the Dutch Labour Party (PvdA) will get in the next elections.
- Solution: find an appropriate sample size such that with probability .95, the error will not exceed more than 3 seats. (There are 150 seats in total.)

The margin of error is $2 \times SE(p)$.

Thus, solve: $\frac{3}{150} = 2 \sqrt{p(1 - p)}/n$

$$n = \frac{(2)^2 p(1 - p)}{\frac{3}{150}^2}$$
Sample sizes

\[ n = \frac{(2)^2 p(1 - p)}{\frac{3}{150}} \]

- \( n \) is a function of \( p \), the true proportion
- \( n \) is largest for \( p = .5 \)
- i.e. when a proportion is close to .5, it is estimated less precisely (it needs a larger sample size) than if it is closer to 0 or 1.
Sample sizes

\[ n = \frac{(2)^2 \cdot 5(1 - .5)}{3^2} \frac{2}{150} = 2500 \]

\[ n = \frac{(2)^2 \cdot .25(1 - .25)}{3^2} \frac{2}{150} = 1875 \]
Difference in proportions

For example, the electoral ratings for political parties in Russia has 54% for Unified Russia; 6% for the LDPR; 5% the KPRF; and 3% for Rodina. What is the confidence interval for the difference between LDPR and KPRF, given that the sample size is 1600?

\[
SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \sqrt{\frac{.06 \times .94}{1600} + \frac{.05 \times .95}{1600}}
\]

prop.test(c(.05*1600,.06*1600), c(1600,1600))
Null hypothesis as basis

Why do we need to assume one of the two hypotheses to test the other? ...

... Because we need to have some estimate of the sampling distribution to determine how likely our outcome is. For this estimate, we use the null hypothesis.
Hypothesis testing

Given the sampling distribution, depending on the distribution and test, we can calculate a test statistics.

We can then calculate the probability of observing the observed value or more extreme, in the direction of the alternative hypothesis, given the assumptions of the null hypothesis.

This is the p-value.
Hypothesis testing: old-fashioned

We used to calculate the critical value given the test statistic and the sampling distribution where we would reject the null hypothesis.

This critical value you can find in tables for the specific probability distribution.

It’s easier to just calculate p-values.

Note that you can reject a null hypothesis, but you can never “accept” a null hypothesis. You either reject or you do not.
Hypothesis testing: p-values

Figure: Calculating p-values
Conditional probabilities

$P(A)$ refers to the probability that $A$ happens or is true.

$P(A|B)$ refers to the probability that $A$ happens, given that $B$ happens.

Conditional probabilities can most easily be understood using Venn diagrams.
Using Venn diagrams, it is straightforward to see that:

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \]

I.e. the probability of \( A \) given \( B \) is the probability of both \( A \) and \( B \) happening, as a proportion of the probability that \( B \) happens.
Bayes’ theorem

We can rewrite that equation as:

\[ P(A \cap B) = P(A|B)P(B) \]

\( P(A \cap B) \) and \( P(B \cap A) \) are the same thing, so we can combine the two equations:

\[ P(A|B)P(B) = P(B|A)P(A) \]

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]
Note that this also implies that $P(A|B) \neq P(B|A)$, unless $P(A) = P(B)$.

In other words, be careful not to inverse conditional probabilities - which is common (!).
Conditional probabilities

For example, imagine you get a positive result on a medical test (95% accuracy) for a rare disease (only 1 in a thousand have the disease) - what is the probability you have the disease?

\[
P(disease|positive) = \frac{P(positive|disease)P(disease)}{P(positive)},
\]

so the base rate of the disease, \(P(disease)\), matters - if it is small (a rare disease), \(P(disease|positive)\) is also small.
Conditional probabilities

From the description we know:

\[ P(\text{positive} | \text{disease}) = 0.95 \]

\[ P(\text{disease}) = \frac{1}{1000} = 0.001 \]

\[ P(\text{positive}) = P(\text{positive} | \text{disease}) P(\text{disease}) \]
\[ + P(\text{positive} | \neg \text{disease}) P(\neg \text{disease}) \]
\[ = 0.95 \times 0.001 + 0.05 \times 0.999 \]
\[ = 0.0509 \]

\[ P(\text{disease} | \text{positive}) = \frac{P(\text{positive} | \text{disease}) P(\text{disease})}{P(\text{positive})} = \frac{0.095}{0.0509} = 0.19 \]
Frequentist reasoning

The CLT tells us what the probability distribution is of an imaginary set of sample means, given the true mean of a random variable.

By knowing the probability distribution given the true mean, we can deduce how likely it is to get the result we get in our sample, given a particular assumption about the true mean.

For example, we can ask: “Given that I assume the mean is zero, how likely is it that I got the sample mean I got from my sample?”

From this the frequentist derives the uncertainty of the estimate of the mean.
Frequentist reasoning

So the probability distribution the frequentist is related to how likely the observed data would be given a particular assumed situation.

In other words, given a particular estimate of the parameter, how likely is it to observe this data?

\[ P(data|estimate) \]
Frequentist reasoning

Hence, we try to find the estimate that makes the data we observe most likely.
Bayesian reasoning

But ... we had just observed that we should be careful not to invert conditional probabilities.

If we are interested in the most likely correct estimate, should we not worry about

\[ P(\text{estimate}|\text{data}) = \frac{P(\text{data}|\text{estimate})P(\text{estimate})}{P(\text{data})} \]
Bayesian reasoning

So a Bayesian wants to maximise $P(\text{estimate}|\text{data})$, rather than $P(\text{data}|\text{estimate})$.

This requires knowledge of $P(\text{estimate})$ and $P(\text{data})$, which we of course do not have.

The frequentist solution is the twist of reasoning that leads to the CLT; the Bayesian solution is a different conception of what $P(A)$ means.
A Bayesian sees $P(A)$ as the belief that $A$ is true (or will happen). Or, in a slightly different conceptualisation, $P(A)$ represent the willingness to bet on $A$. 
Bayesian inference

Given that we are interested in the estimate that maximises $P(data|estimate)$, and given that $P(data)$ is not affected by the value of the estimate, we can conclude that

$$\maxarg_{estimate} \frac{P(data|estimate)P(estimate)}{P(data)} = \maxarg_{estimate} P(data|estimate)P(estimate)$$

So the only issue is $P(estimate)$, which, for the Bayesian, is the prior belief in a particular estimate you have, before seeing the data.
Bayesian inference

We can rephrase the medical test example in these terms - we try to estimate whether the person is ill:

- \( P(\text{disease}) \) is the prior belief, before we see the test result, that the person has the disease - without test, we would say that the person has a 1/1000 chance of being ill;
- \( P(\text{positive} | \text{disease}) \) is the likelihood of observing the positive test result, given that the patient has the disease;
- \( P(\text{disease} | \text{positive}) \) is the posterior belief, after seeing the test result, and given our prior belief, that the person is ill.

... We updated our belief that \( P(\text{disease}) = 0.001 \) to believing that \( P(\text{disease}) = 0.019 \).
In the previous example, the variable of interest was a categorical variable - you either have the disease or you do not.

Bayesian inference can similarly be applied with continuous variables.

The \( P(X) \) then refer to probability distributions rather than probabilities \textit{per se}. 

Bayesian inference
Priors

The prior is more influential on the posterior when:

▶ The data set is small - we have few observations to update our beliefs;
▶ The prior has very high precision.

Priors with very low precision are called *uninformative priors*, or at its extreme, *flat priors*.

There is also substantial research in *empirical priors*, whereby the priors themselves are based on the data rather than subjective beliefs.
Whereas frequentists generally talk about statistical significance and confidence intervals, Bayesians prefer to talk about high probability density regions and generally present the actual posterior distribution, rather than a few summary statistics of this posterior (as in frequentist work).
Bayesians like to talk about congruence. This refers to models whereby the distribution of $P(\text{estimate})$ is of a similar form to the distribution of $P(\text{estimate} | \text{data})$, in which case you can use an updating algorithm whereby beliefs are continuously updated as more information flows in.

This is particularly useful in machine learning applications.
Computation

Only in a few very specific cases can Bayesian estimates be analytically derived. Generally, we need computational algorithms to numerically approach the posterior distribution.

Bayesian inference is older than frequentist inference, but lack of computing power prohibited extensive use.

After the invention of the Gibbs sampler and the Metropolis algorithm, and the rise of desktop computing power, Bayesian estimation has become quite common.