MS-A0504 First course in probability and statistics

Week 4

Parameter estimation, confidence intervals

Joni Virta

Department of Mathematics and Systems Analysis
School of Science
Aalto University

Spring 2018, IV-period

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Statistical inference

The objective is to make inference based on the observed data.

1. Choose a stochastic model appropriate to the situation.
2. Fit the model to the data
   (estimate the model parameters).
3. Compute the required statistics from the fitted model.
4. Make the conclusions.

Conclusions are usually (educated) guesses:

- What is the true weight of a giraffe if three weightings gave the results 1250 kg, 1300 kg, 1360 kg?
- Does the majority of people living in Espoo think Espoo should stay independent if 509 out of a thousand respondents were in favor of the independence in a poll.
- Does the oil price stay on its current level until the end of the year.

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Unknown parameters

Consider an unknown source of data, the distribution $f(x)$ of which is known apart from a few unknown parameters.

Example (one unknown parameter):

- Bernoulli distribution: $f_p(0) = 1 - p$, $f_p(1) = p$
- Exponential distribution: $f_\lambda(x) = \lambda e^{-\lambda x}, x > 0$

Example (two unknown paameters):

- Uniform distribution on interval $[a, b]$: $f_{(a,b)}(x) = \frac{1}{b-a}$
- Normal distribution: $f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Based on observed data $(x_1, \ldots, x_n)$ what is the best guess for the value of the unknown parameter?

Parameter estimation

Consider an unknown source of data with the distribution $f(x)$ where the value of the parameter θ is unknown.

Observations $x_1, \ldots, x_n$ have been obtained from the data source.

- An estimate $\hat{\theta} = g(x)$ of the parameter θ is a guess for its value calculated based on the data $x = (x_1, \ldots, x_n)$.
- An estimator of θ is the function $(x_1, \ldots, x_n) \mapsto g(x_1, \ldots, x_n)$, which maps the data into the estimate.

There does not usually exists a single "best" choice for the estimator of a certain parameter.

Example: The proportion of faulty products

A production line produces components, of which the proportion p is faulty, independent of each other. When 200 components were inspected, a total of 22 faulty ones were found. Find an estimate for the value of the unknown parameter p.

Intuituvely, a natural estimate is

$$\hat{p} \ = \ \frac{22}{200} \ = \ 11\%$$

Is this the best estimate? Are there other natural alternatives?

Example: Discrete uniform distribution

Military planes of the opposing side are numbered $1, 2, \ldots, N$ where N is an unknown parameter. Scouts have observed the numbers of three planes to be $x_1 = 63$, $x_2 = 17$, $x_3 = 203$. Using the observations determine an estimate for the total number of military planes N.

The data source corresponding to the plane observations has the uniform distribution

$$f_{1,N}(k) \ = \ \begin{cases} \frac{1}{N}, & k = 1, \ldots, N, \\ 0, & \text{else.} \end{cases}$$

What could be a natural estimator $\hat{N}(x)$ for the parameter N?

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Likelihood function

Stochastic model for the data source: $(X_1, \ldots, X_n)$, the components of which are f_θ-distributed and independent of each other.

The probability of observing the values $(x_1, \ldots, x_n)$ under the model is for a discrete distribution:

$$\mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) \ = \ f_\theta(x_1) \cdots f_\theta(x_n)$$

and for a continuous distribution (approximately for a small ϵ)

$$\mathrm{P}(X_1 = x_1 \pm \epsilon/2, \ldots, X_n = x_n \pm \epsilon/2) \ \approx \ \epsilon^n f_\theta(x_1) \cdots f_\theta(x_n).$$

The likelihood function $L(\theta) = f_\theta(x_1) \cdots f_\theta(x_n)$ tells the probability to observe the (approximately) same data under the model f_θ that was actually observed.

Maximum likelihood estimate

The likelihood function $L(\theta) = f_\theta(x_1) \cdots f_\theta(x_n)$ tells the probability to observe the (approximately) same data under the model f_θ that was actually observed.

The larger the value of the likelihood function at θ is, the more likely can we hold the assumption that the observed data comes from a f_θ-distributed data source.

The maximum likelihood estimate (MLE) $\hat{\theta} = \hat{\theta}(x)$ for the parameter θ is the value of the parameter which maximizes the likelihood function.

Example: Estimating the proportion of faulty products

A production line produces components, of which the proportion p is faulty, independent of each other. When 200 components were inspected, a total of 22 faulty ones were found. Find an estimate for the value of the unknown parameter p.

When we inspect a batch of $n = 200$ components, the number N of faulty components has the distribution

$$f(x \mid p) \;=\; \mathrm{P}(N = x \mid p) \;=\; \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, 200$$

For which value of the parameter p does the likelihood function

$$L(p) \;=\; \binom{200}{22} p^{22} (1-p)^{200-22}$$

obtain its maximum value?

Example: Estimating the proportion of faulty products

$$L(p) \;=\; \binom{200}{22} p^{22}(1-p)^{178}$$

is maximized when $\ell(p) = \log L(p)$ is maximized:

$$\ell(p) \;=\; \log f(22 \,|\, p) \;=\; \log \binom{200}{22} + 22 \log p + 178 \log(1-p)$$

$$\ell'(p) \;=\; 22 \frac{1}{p} - 178 \frac{1}{1-p}$$

$$\ell''(p) \;=\; -22 \frac{1}{p^2} - 178 \frac{1}{(1-p)^2} < 0, \quad \forall p \in (0,1)$$

The maximum likelihood estimate for the parameter p is thus found as the value putting the derivative to zero:

$$\ell'(p) = 0 \quad \Longleftrightarrow \quad \frac{22}{p} \;=\; \frac{178}{1-p} \quad \Longleftrightarrow \quad p = \frac{22}{200}$$

ML-estimation for the probability parameter of a binomial distribution

Fact

The maximum likelihood estimate for the unknown parameter p of a $\mathrm{Bin}(n, p)$*-distribution based on an observed point of data x is*

$$\hat{p} \;=\; \frac{x}{n}.$$

Proof.

Repeat the previous computation with $200 \mapsto n$ and $22 \mapsto x$. □

MLE for the continuous uniform distribution

A data source generates independent random numbers from the uniform distribution on the interval $[0, \theta]$ and the values $(1.2, 4.5, 8.0)$ have been observed. Determine the maximum likelihood estimate of θ.

The individual observations have the density function

$$f(x \mid \theta) \;=\; \begin{cases} \frac{1}{\theta}, & x \in [0, \theta], \\ 0, & \text{else}, \end{cases}$$

and the likelihood function of the data sample is

$$L(\theta) = \prod_{i=1}^{3} f(x_i \mid \theta) = \begin{cases} \theta^{-3}, & \theta \geq \max(1.2, 4.5, 8.0), \\ 0, & \text{else}. \end{cases}$$

$L(\theta)$ can be seen to be maximized at $\hat{\theta} = \max(1.2, 4.5, 8.0) = 8.0$ which is then the MLE of θ.

Maximum likelihood estimates for the parameters of normal distribution

The normal distribution density function

$$f_{(\mu,\sigma^2)}(t) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

is known apart from the parameters μ and σ^2.

Fact

The maximum likelihood estimates of the normal distribution parameters (μ, σ^2) *for an observed data* $x = (x_1, \ldots, x_n)$ *are*

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \textit{and} \qquad \hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - m(x))^2,$$

i.e., the mean and the (scaled) sample variance of the data set x*.*

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Unbiased estimators

An estimator $\hat{\theta}(x)$ of the parameter θ of the distribution f_θ is unbiased, if for a stochastic model $X = (X_1, \ldots, X_n)$ coming from the distribution f_θ it holds that

$$\mathsf{E}\,\hat{\theta}(X) \ = \ \theta.$$

Interpretation: assume that the unknown data source really obeys f_θ-distribution and n independent observations are observed from the data source. If we then compute an estimate using an unbiased estimator, repeating this whole experiment multiple times, then the mean of the estimates would be close to the true value of the parameter.

Example: The proportion of faulty products

For a known n the maximum likelihood estimator of the unknown parameter p of a $\mathrm{Bin}(n, p)$-distribution is

$$\hat{p}(x) \;=\; \frac{x}{n}.$$

If the random variable N obeys the $\mathrm{Bin}(n, p)$-distribution then

$$\mathrm{E}\,(\hat{p}(N)) \;=\; \mathrm{E}\left(\frac{N}{n}\right) \;=\; \frac{1}{n}\,\mathrm{E}(N) \;=\; \frac{1}{n} \times np \;=\; p.$$

Thus the function

$$x \mapsto \hat{p}(x)$$

is an unbiased estimator of the parameter p.

Example: ML-estimate of the normal distribution

The maximum likelihood estimate of the expectation parameter μ of the normal distribution is

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

We have for a stochastic model $X = (X_1, \ldots, X_n)$ that

$$\mathsf{E}[\hat{\mu}(X)] \;=\; \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \;=\; \mu,$$

so the function $x \mapsto \hat{\mu}(x)$ is an unbiased estimator of the parameter μ.

Example: ML-estimate of the normal distribution

The maximum likelihood estimate of the variance parameter σ^2 of the normal distribution is

$$\hat{\sigma}^2(x) \;=\; \frac{1}{n}\sum_{i=1}^{n}(x_i - m(x))^2.$$

We have for a stochastic model $X = (X_1, \ldots, X_n)$ that

$$\mathrm{E}[\hat{\sigma}^2(X)] \;=\; \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - m(X)^2\right) \;=\; \cdots \;=\; \frac{n-1}{n}\sigma^2,$$

so $\hat{\sigma}^2(x)$ is biased. An unbiased estimator for the variance parameter is given by the sample variance

$$s^2(x) \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m(x))^2.$$

For large values of n the two are close to each other.

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Linear regression

In linear regression the aim is to predict the behavior of a single variable (response) using a group of other variables (covariates).

In the simplest case a straight line $y = \alpha + \beta x$ is fitted to a set of data $(x_1, y_1), \ldots, (x_n, y_n)$. This can be done by minimizing the mean squared error:

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

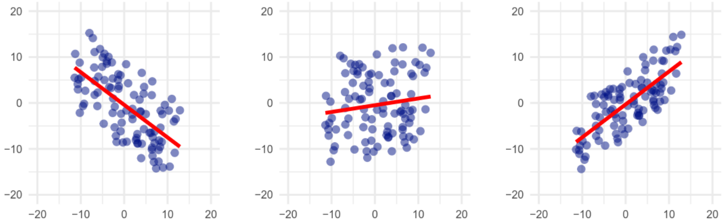
Fact

The line with the smallest MSE is given by the choices

$$\hat{\beta} = \frac{s(y)}{s(x)} cor(x, y), \qquad \hat{\alpha} = m(y) - \hat{\beta} m(x).$$

Linear regression

The smallest MSE lines for three data sets.



Caution: the method will fit a line to data even if there is no linear trend!

Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Example. Coffee machine

In order to find out how much coffee the coffee machine dispenses on average, the researchers dispensed 30 cups of coffee from the machine and measured the amounts.

The following values were observed (cl):
$x = (10.24, 9.94, 10.55, 9.28, 10.57, 9.77, 9.50, 10.03, 10.51, 10.29, 10.92, 10.06, 9.63, 10.83,$
$10.36, 9.17, 10.29, 10.24, 9.73, 10.56, 9.18, 9.84, 9.91, 10.74, 9.57, 10.76, 10.53, 10.52, 10.42, 10.11)$

The mean of the data is $m(x) = 10.14$. Is the expected value μ of the dispensed amount close to 10.14?

Interval estimation

Thus far we have been interested only in point estimation, that is, finding a single, in some sense plausible, value for the unknown parameter.

However, for continuous data a point estimate is almost always different from the true value of the parameter.

A complementary approach is to search for a data-based *interval* $[a(x), b(x)]$ that is expected to contain the true value of the parameter.

How should we measure the goodness of an interval? The event $\theta \in [a(x), b(x)]$ is deterministic.

Confidence interval

Assuming a stochastic model for the data source, $(X_1, \ldots, X_n)$, the probability

$$P(\theta \in [a(X), b(X)])$$

tells us what proportion of intervals computed from samples coming from the data source on average contains the true value.

Any single realized $[a(x), b(x)]$ still either contains or does not contain the true parameter value, with no randomness involved.

A confidence interval (CI) with the confidence level α is an interval $[a(x), b(x)]$ for which

$$P(\theta \in [a(X), b(X)]) \geq \alpha$$

for all values of the parameter θ.

CI for the continuous uniform distribution

A data source generates independent random numbers from the uniform distribution on the interval $[0, \theta]$ and the values $x_1, \dots, x_n$ have been observed. Determine a CI with the confidence level 95 % for θ.

We will first obtain the distribution of the MLE, $\hat{\theta}(X) = \max(X_1, \dots, X_n)$. For a point $t \in [0, \theta]$,

$$P(X_i \leq t) = \int_0^t \theta^{-1} dt = \frac{t}{\theta}.$$

Thus

$$\begin{aligned} P(\hat{\theta}(X) \leq t) &= P(X_1 \leq t, \dots, X_n \leq t) = P(X_1 \leq t) \cdots P(X_n \leq t) \\ &= \left(\frac{t}{\theta}\right)^n \end{aligned}$$

CI for the continuous uniform distribution

For any $0 \leq s \leq 1$ we have

$$P(s\theta \leq \hat{\theta}(X) \leq \theta) = P(\hat{\theta}(X) \leq \theta) - P(\hat{\theta}(X) \leq s\theta) = 1 - s^n,$$

which can be written as

$$P(\hat{\theta}(X) \leq \theta \leq \frac{\hat{\theta}(X)}{s}) = 1 - s^n.$$

If we now choose $s = (1-\alpha)^{1/n}$, a level α confidence interval can be found as

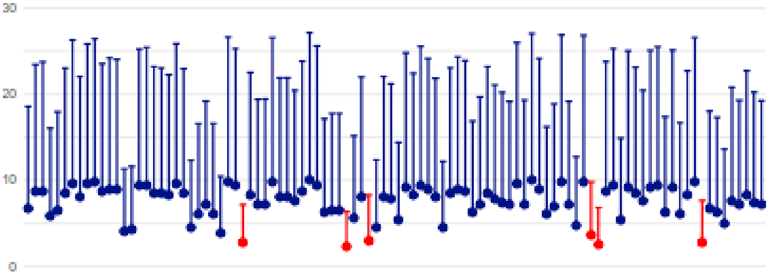
$$\left[\hat{\theta}(x), \frac{\hat{\theta}(x)}{(1-\alpha)^{1/n}}\right].$$

If e.g. the values 8.1, 2.6 and 8.8 have been observed, the confidence interval is

$$\left[(8.8), \frac{(8.8)}{(1-0.95)^{1/3}}\right] = [8.8, 23.9].$$

Simulated CIs

Confidence intervals computed from 100 samples of size $n = 3$ from the uniform distribution on $[0, 10]$. 94 of the intervals contain the true value of the parameter.



Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Confidence interval for μ

In the previous example we needed two things to obtain a confidence interval for the parameter θ.

1. An estimator of θ.
2. The distribution of the estimator under the stochastic model.

In a general situation of a sample $x_1, \ldots, x_n$ from an arbitrary distribution these can be very hard to come by. However, if the parameter of interest in the expected value $\mu = \mathrm{E}(X)$ of the data distribution something universal can be said.

1. $m(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$ is an estimate of μ.
2. The distribution of $m(x)$ is approximately normal.

Approximate confidence interval for μ

A general stochastic model:
$X_1, X_2, \ldots$ independent and identically distributed with the unknown expected value μ

Computing an approximate level α confidence interval:

1. Compute the mean $m(x)$ and standard deviation $s(x)$ of the observed sample
2. Determine using the normal table the number $z > 0$, for which $\mathrm{P}(|Z| \leq z) = 1 - 2\Phi(-z) = \alpha$
   $\implies$ For example, for $\alpha = 0.99$ we get $z \approx 2.58$
3. An approximate level α confidence interval is $m(x) \pm z\,\frac{s(x)}{\sqrt{n}}$.

Proof: By the law of large numbers and the central limit theorem, for large n we have $s(X) \approx \sigma$ and

$$\mathrm{P}\left(|m(X) - \mu| \leq z\frac{s(X)}{\sqrt{n}}\right) \;\approx\; \mathrm{P}\left(\left|\frac{m(X)-\mu}{\sigma/\sqrt{n}}\right| \leq z\right) \;\approx\; \mathrm{P}\left(|Z| \leq z\right) \;=\; \alpha.$$

Approximate confidence interval for μ

Some notes:

- If the standard deviation σ is known we can use $s(x) = \sigma$, making the first approximation in the proof exact.
- If the data is normally distributed then the second approximation in the proof is exact.

The half-width $z \, \frac{s(x)}{\sqrt{n}}$ of the confidence interval is also known as the error margin.

Example. Coffee machine

In order to find out how much coffee the coffee machine dispenses on average, the researchers dispensed 30 cups of coffee from the machine and measured the amounts.

The following values were observed (cl):

$x = (10.24, 9.94, 10.55, 9.28, 10.57, 9.77, 9.50, 10.03, 10.51, 10.29, 10.92, 10.06, 9.63, 10.83, 10.36, 9.17, 10.29, 10.24, 9.73, 10.56, 9.18, 9.84, 9.91, 10.74, 9.57, 10.76, 10.53, 10.52, 10.42, 10.11)$

The mean of the data is $m(x) = 10.14$.

Find an approximate level 95 % confidence interval for the unknown expected value μ of coffee.

Example. Coffee machine

$x = (10.24, 9.94, 10.55, 9.28, 10.57, 9.77, 9.50, 10.03, 10.51, 10.29, 10.92, 10.06, 9.63, 10.83,$
$10.36, 9.17, 10.29, 10.24, 9.73, 10.56, 9.18, 9.84, 9.91, 10.74, 9.57, 10.76, 10.53, 10.52, 10.42, 10.11)$

Stochastic model for the data source: $X_1, \ldots, X_{30}$ are independent and identically distributed with the unknown expected value μ.

Now, $\mathrm{P}(|Z| \leq 1.96) \approx 0.95$ so $z \approx 1.96$.
An approximate 95 % confidence interval for μ is

$$m(x) \pm z\,\frac{s(x)}{\sqrt{n}} = 10.14 \pm 1.96\frac{0.49}{\sqrt{30}} = 10.14 \pm 0.09.$$

- A point estimate for μ is $m(x) = 10.14$
- An interval estimate for μ is $[10.05, 10.23]$

Can we deduce that the interval $[10.05, 10.23]$ contains μ with 95% probability?
We can not.

Interpretation of the estimate

The approximate CI $m(x) \pm 1.96\, s(x)/\sqrt{n}$ is such that:

$$\mathrm{P}(\mu \in [m(X) - 1.96\,\frac{s(X)}{\sqrt{n}}, m(X) + 1.96\,\frac{s(X)}{\sqrt{n}}]) \;\approx\; 95\%.$$

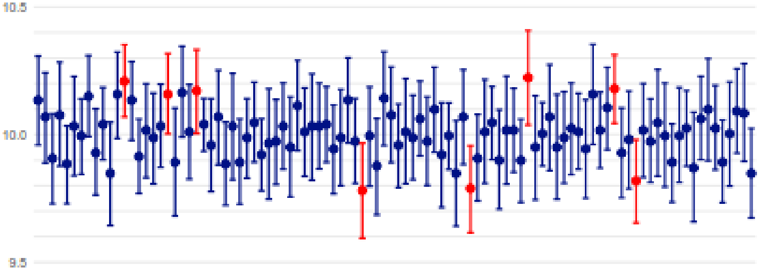
The interval estimate $[10.05, 10.23]$ computed from observed data does not have anything random associated with it.

Someone, who observes multiple samples from the same data source and computes multiple interval estimates using the above formula:

- Knows that 95% of the computed intervals contain the unknown true value of μ (but does not know which)
- Knows that 5% of the computed intervals do not contain the unknown true value of μ (but does not know which)

Simulated CIs

Confidence intervals computed from 100 samples of size $n = 30$ from the normal distribution with $\mu = 10$ and $\sigma = 0.5$. 92 of the intervals contain the true value of the parameter.



Contents

Statistical inference

Parametric distributions

Maximum likelihood estimation

Some properties of estimators

Linear regression

Confidence intervals

Confidence interval for expected value

Confidence interval for the binomial model

Binary model

Data source:
$X_1, X_2, \ldots$ are independent $\{0, 1\}$-valued random numbers with the probability p (unknown) of obtaining the value 1.

The distribution of X_i is determined by p:

$$f(k \mid p) \;=\; \begin{cases} 1 - p, & k = 0, \\ p, & k = 1, \\ 0, & \text{else.} \end{cases}$$

$$\mathrm{E}(X_i) \;=\; 0 \times \mathrm{P}(X_i = 0) + 1 \times \mathrm{P}(X_i = 1) \;=\; \mathrm{P}(X_i = 1) \;=\; p,$$

Recall that this is the Bernoulli-distribution with the parameter p.

Example: Presidential poll

A random sample of size $n = 2000$ from the eligible voters of a certain country was chosen and were asked whether they will vote for the current president in the next election (0=no, 1=yes). 774 of them answered yes.

The results of the measurement $X = (X_1, \ldots, X_{2000})$ obey approximately the binary model with the expected value p, where

$$p \ = \ \mathrm{E}(X_i) \ = \ \mathrm{P}(X_i = 1)$$

is the (unknown) support for the current president in the whole population.

Find a point estimate and a 95% confidence interval for the true support p.

Binary model CI

The MLE of p is

$$m(x) \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i \;=\; \frac{\#\{i : x_i = 1\}}{n},$$

i.e., the proportion of ones in the sample x.

A confidence interval for p could be found using the formula for the approximate level α CI given earlier.

However, under the binary model one of the approximations in the formula can actually be replaced with something more accurate.

Binary model CI

A Bernoulli-distributed X_i with the expected value $\mathrm{E}(X_i) = p$ has the standard deviation

$$\sigma \;=\; \mathrm{SD}(X_i) \;=\; \sqrt{\mathrm{E}(X_i^2) - \mathrm{E}(X_i)^2} \;=\; \sqrt{p - p^2} \;=\; \sqrt{p(1-p)}$$

Derivatives can be used to show that the standard deviation is maximal when $p = 0.5$. Then $\sigma = 0.5$

This time we can write

$$\mathrm{P}\left(|m(X) - p| \leq z\frac{0.5}{\sqrt{n}}\right) \geq \mathrm{P}\left(|m(X) - p| \leq z\frac{\sigma}{\sqrt{n}}\right) \;\approx\; \mathrm{P}\left(|Z| \leq z\right) \;=\; \alpha,$$

where z is such that $\mathrm{P}(|Z| \leq z) = \alpha$

Binary model CI

Data source:
$X_1, X_2, \ldots$ are independent $\{0,1\}$-valued random numbers with the probability p (unknown) of obtaining the value 1.

A conservative approximate (n large) level α confidence interval for the proportion p

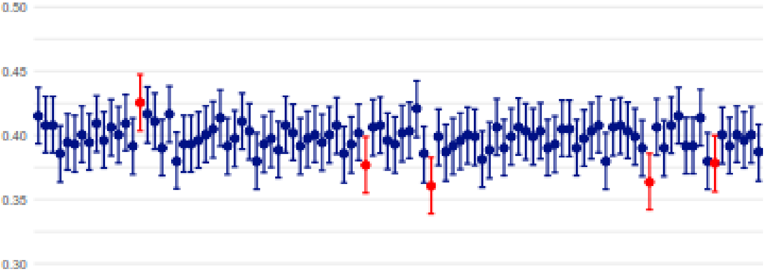
1. Compute the mean $m(x)$ of the observed data
2. Determine using the normal table the number $z > 0$, for which $\mathrm{P}(|Z| \leq z) = 1 - 2\Phi(-z) = \alpha$
3. An approximate level α confidence interval is $m(x) \pm z \, \frac{0.5}{\sqrt{n}}$.

For the poll data:

$$m(x) \pm 1.96 \, \frac{0.5}{\sqrt{n}} = 0.387 \pm 1.96 \frac{0.5}{\sqrt{2000}} = 0.387 \pm 0.022.$$

Simulated CIs

Confidence intervals computed from 100 samples of size $n = 2000$ from the Bernoulli distribution with $p = 0.4$. 95 of the intervals contain the true value of the parameter.



Error margin

Error margin in a poll refers to the half-width of a binary model confidence interval for p computed using a selected confidence level.

- Often the confidence level is not mentioned (usually 95%).
- The used formulas are never told .
- The error margin is almost always defined wrongly in the media.

The width of the conservative CI for binary model is completely determined by the sample size n and the confidence level α and for example for 95 % conservative confidence interval we have:

- $n \approx 1100 \implies m(x) \pm 3\%$
- $n \approx 2400 \implies m(x) \pm 2\%$
- $n \approx 9600 \implies m(x) \pm 1\%$