Laplace: direct and inverse probabilities

Hans Fischer

Mathematisch-Geographische Fakultät
Katholische Universität Eichstätt

Workshop München

Pierre Simon de Laplace 1749–1827

1774–1786: first probabilistic period
1796: Exposition du système du monde
1799–1805: 4 vols of mécanique céleste
1810–: second probabilistic period
1812: Théorie analytique des probabilités
1825: 5th vol. of mécanique céleste

Direct Probability

Based on a hypothetic assumption ("cause") H the conditional probability

$$P_H(E) \text{ or } P(E|H)$$

of a certain event E is determined.

Depending on the value of this probability we are able, if E is observed, to make a judgement about the "plausibility" of H. If $P(E|H)$ is very small, preference is given to another hypothesis H_1.

$P(E|H)$ with an observable event E is called "direct" probability. The mode of inference from E to H is rather in-direct.

Inverse Probability

Bayes 1763, Laplace 1774: (Direct) inference from (observable) "phenomena" E to (un-observable) hypothetic causes:
Given s hypothetic causes $H_1, \ldots, H_s$, we have

$$P(H_j|E) = \frac{P(H_j \cap E)}{P(E)} = \frac{P(E|H_j)P(H_j)}{P(E)}$$

If $P(H_j)$ is the same for all possible hypotheses and E is fixed:

$$P(H_j|E) \sim P(E|H_j)$$

The left side probability is the "inverse probability" with respect to the right side probability.

Binomial models

Given a Bernoulli process of n trials with success probability p as the hypothetic "cause," and the event $T = t$ (t successes). Under the assumption of equiprobable success probabilities in the interval $(0, 1)$ we have (analogously to the discrete case above):

$$dP(p|T = t) \sim dP(T = t|p),$$

$$dP(p|T = t) \sim p^t(1-p)^{n-t}dp,$$

and therefore

$$P(a \leq p \leq b|T = t) = \frac{\int_a^b x^t(1-x)^{n-t}dx}{\int_0^1 x^t(1-x)^{n-t}dx}$$

The estimation of the success probability

In particular: $a = h_n - \omega$, $b = h_n + \omega$ with $h_n = t/n$:

$$P(h_n - \omega \leq p \leq h_n + \omega | T = t) = \frac{\int_{h_n-\omega}^{h_n+\omega} x^t(1-x)^{n-t}dx}{\int_0^1 x^t(1-x)^{n-t}dx}$$

$$= \mathrm{Beta}(h_n + \omega, t+1, n-t+1) - \mathrm{Beta}(h_n - \omega, t+1, n-t+1).$$

Numerical example

Within 50 trials are 28 successes. Then:

$$P(h_n - 0.135 \leq p \leq h_n + 0.135 | T = 28) = 0.953$$

The unknown "parameter" p lies in [0.425, 0.695] with a probability of more than 95%.

Laplace's innovations 1774

Laplace's innovations 1774

- Approximation for large n and large $t = nh_n$:

$$P(p \leq r | T = nh_n) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(r-h_n)a(h_n,n)} e^{-x^2/2} dx \quad \text{or}$$

$$P(h_n - \omega \leq p \leq h_n + \omega | T = nh_n) \approx \sqrt{\frac{2}{\pi}} \int_0^{\omega a(h_n,n)} e^{-x^2/2} dx,$$

$$a(h_n, n) = \sqrt{\frac{n}{h_n(1-h_n)}}$$

Laplace's innovations 1774

- Approximation for large n and large $t = nh_n$:

$$P(p \leq r | T = nh_n) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(r-h_n)a(h_n,n)} e^{-x^2/2} dx \quad \text{or}$$

$$P(h_n - \omega \leq p \leq h_n + \omega | T = nh_n) \approx \sqrt{\frac{2}{\pi}} \int_{0}^{\omega a(h_n,n)} e^{-x^2/2} dx,$$

$$a(h_n, n) = \sqrt{\frac{n}{h_n(1-h_n)}}$$

- $n \to \infty : \; P(h_n - \omega \leq p \leq h_n + \omega | T = nh_n) \to 1 \quad \forall \omega > 0$
  "Second" law of large numbers

$P(h_n - \omega \leq p \leq h_n + \omega | T = nh_n) \to 1$

Interpretation:

- In the beginning there is a total – subjective – ignorance of the experimental framework, which is expressed by an "a priori" uniform distribution of the success probabilities; with a growing number of experiments, absolute certainty about "objective" success probabilities („possibilités") can be reached "in the limit".
- h_n approaches more and more the unknown success probability (frequentist interpretation of probabilities)
- Laplace is convinced of the compatibility of subjective and objective probabilities.

Birth rates of boys and girls, 1786

There is a slight excess of male births within all large samples from different European countries.
Ex.1: There are 251527 new born boys and new born 241945 girls in Paris from 1745 to 1770. Based on his data, what is the probability that the "possibilité" of a boy's birth is greater than 1/2?

$$h_n = 0.5097, a(h_n, n) = 1405, P(p > 0.5 | T = 251527) \approx \Phi(0.097 * 1405).$$

Ex. 2: In Viteaux (Bourgogne) 203 boys versus 212 girls have been born within 5 years. What is the probability that the "possibilité" of a boy's birth is less than 1/2?

$$P(p < 0.5 | T = 203) = 0.67.$$

Extension to two samples: Laplace calculates the probability that the "possibilité" of a boy's birth is greater in London than in Paris.

Uniform apriori distributions

Uniform apriori distributions

- Laplace's pragmatism: If nothing is "given" a priori with respect to the "possibility" of a certain event, then all possible values between 0 and 1 have to be considered equally probable.

Uniform apriori distributions

- Laplace's pragmatism: If nothing is "given" a priori with respect to the "possibility" of a certain event, then all possible values between 0 and 1 have to be considered equally probable.
- Uniform priors correspond to an epistemic model of inference which starts from total ignorance, and arrives, by means of repeated experiments corresponding to association of ideas, at a (partial) knowledge.

Uniform apriori distributions

- Laplace's pragmatism: If nothing is "given" a priori with respect to the "possibility" of a certain event, then all possible values between 0 and 1 have to be considered equally probable.
- Uniform priors correspond to an epistemic model of inference which starts from total ignorance, and arrives, by means of repeated experiments corresponding to association of ideas, at a (partial) knowledge.
- Cournot (1848): Assumptions on priors are always "subjective". Plausibility consideration: In the long run the influence of particular priors is vanishing.

Uniform apriori distributions

- Laplace's pragmatism: If nothing is "given" a priori with respect to the "possibility" of a certain event, then all possible values between 0 and 1 have to be considered equally probable.
- Uniform priors correspond to an epistemic model of inference which starts from total ignorance, and arrives, by means of repeated experiments corresponding to association of ideas, at a (partial) knowledge.
- Cournot (1848): Assumptions on priors are always "subjective". Plausibility consideration: In the long run the influence of particular priors is vanishing.
- Von Mises 1919:
  $P(h_n-\omega \leq p \leq h_n+\omega | T = nh_n) - \sqrt{\frac{2}{\pi}} \int_0^{\omega a(h_n,n)} e^{-x^2/2} dx \to 0$
  for (almost) any a-priori-distributions of p.

Uniform apriori distributions

- Laplace's pragmatism: If nothing is "given" a priori with respect to the "possibility" of a certain event, then all possible values between 0 and 1 have to be considered equally probable.
- Uniform priors correspond to an epistemic model of inference which starts from total ignorance, and arrives, by means of repeated experiments corresponding to association of ideas, at a (partial) knowledge.
- Cournot (1848): Assumptions on priors are always "subjective". Plausibility consideration: In the long run the influence of particular priors is vanishing.
- Von Mises 1919:
  $P(h_n-\omega \leq p \leq h_n+\omega | T = nh_n) - \sqrt{\frac{2}{\pi}} \int_0^{\omega a(h_n,n)} e^{-x^2/2} dx \to 0$
  for (almost) any a-priori-distributions of p.
- Was Laplace already aware of the full generality of this theorem?

Asymptotic normality of inverse probabilities

Asymptotic normality of inverse probabilities

- Binomial model: $p(x|s_1, s_2) \sim x^{s_1}(1-x)^{s_2}$, numerous applications

Asymptotic normality of inverse probabilities

- Binomial model: $p(x|s_1, s_2) \sim x^{s_1}(1-x)^{s_2}$, numerous applications
- Multinomial model: $p(x, y|s_1, s_2, s_3) \sim x^{s_1}y^{s_2}(1-x-y)^{s_3}$, only abstract theorem

Asymptotic normality of inverse probabilities

- Binomial model: $p(x|s_1, s_2) \sim x^{s_1}(1-x)^{s_2}$, numerous applications
- Multinomial model: $p(x, y|s_1, s_2, s_3) \sim x^{s_1} y^{s_2}(1-x-y)^{s_3}$, only abstract theorem
- Direct observations $x_i = a + \epsilon_i$:
  $p(a|x_1, x_2, \ldots, x_n) \sim f(a-x_1)f(a-x_2)\ldots f(a-x_n)$ if f is the symmetric error law and $x_1, \ldots, x_n$ are observed values.

Asymptotic normality of inverse probabilities

- Binomial model: $p(x|s_1, s_2) \sim x^{s_1}(1-x)^{s_2}$, numerous applications
- Multinomial model: $p(x, y|s_1, s_2, s_3) \sim x^{s_1}y^{s_2}(1-x-y)^{s_3}$, only abstract theorem
- Direct observations $x_i = a + \epsilon_i$: $p(a|x_1, x_2, \ldots, x_n) \sim f(a-x_1)f(a-x_2)\ldots f(a-x_n)$ if f is the symmetric error law and $x_1, \ldots, x_n$ are observed values.
- In all those cases the mean of the approximating normal distribution is the mode of the likelihood function on the right side.

Asymptotic aposteriori estimation of the location parameter *a*, 1812

Basic idea: $\int_{-\infty}^{+\infty} p(a|x_1, x_2, \ldots, x_n)\ell(\hat{a} - a)da = \min$

Asymptotic aposteriori estimation of the location parameter a, 1812

Basic idea: $\int_{-\infty}^{+\infty} p(a|x_1, x_2, \ldots, x_n)\ell(\hat{a} - a)da = \min$

Even approximately this problem is only solvable if the parameters of the asymptotic normal distribution, which depend on the particular error density f, are known.

Sums of independent r.v.'s, 1776, 1781

Ab initio, Laplace deals with direct probabilities.

Ingenious but cumbersome method for calculating densities of sums of independent identically distributed random variables whose density has a compact support; main device: convolutions.
Examples: Sums of observational errors whose density corresponds to a parabola; sums of rectangularly distributed angles of inclination of celestial bodies.

The comet problem I

Under consideration of the two directions of revolution around the sun there are angles of inclination between -90° and 90°, if the revolution is according to the earth rotation, and from -180° to -90° and 90° to 180°, corresponding to a „retrograde“ revolution.

ca. 1770: 6 planets and 6 months known, 63 comets.

Simple example: All planets have the same direction of revolution. Can this be by "hazard"? Probability for this configuration 2^{-12} if both directions equiprobable. $H_0 : p = 1/2$, $P_{H_0}(T = 12) = 2^{-12}$.

The comet problem I

Under consideration of the two directions of revolution around the sun there are angles of inclination between -90° and 90°, if the revolution is according to the earth rotation, and from -180° to -90° and 90° to 180°, corresponding to a „retrograde“ revolution.

ca. 1770: 6 planets and 6 months known, 63 comets.

Simple example: All planets have the same direction of revolution. Can this be by "hazard"? Probability for this configuration 2^{-12} if both directions equiprobable. $H_0 : p = 1/2$, $P_{H_0}(T = 12) = 2^{-12}$.

Why does Laplace not apply inverse probabilities?
$P(p > 0.8 | T = 12) = 0.95$

The comet problem II

Are the orbital planes of particular celestial bodies distributed "at random"?

"Null hypothesis": Equiprobability of all possible absolute values of inclinations

Test statistics: arithmetic mean m of the absolute values of inclinations

Decision rule: Depending on the probability that under H_0 the test statistics m surpasses the observed arithmetic mean m':
$P(m > m')$ "large" $\rightarrow$ rejection of H_0.

Rescaling of angles and rectangular distribution on $[0, 1]$:

$$P(s_1 \leq m \leq s_2) = \frac{1}{n!} \left(\sum_{i=0}^{[ns_2]} \binom{n}{i} (-1)^i (s_2 - ih)^n - \right.$$
$$\left. - \sum_{i=0}^{[ns_1]} \binom{n}{i} (-1)^i (s_1 - ih)^n \right)$$

The comet problem III

In 1776 the "exact" formula can, due to its numerical complexity, only be applied to the 12 most recently observed comets. L. determines the probability that $m \leq m'$. Since this probability is rather large, he infers that there exists no hidden "cause" entailing the coincidence of comet's orbits with that of the earth.

1810 by means of the central limit theorem:
For 97 comets the arithmetic mean of the absolute values of angles of inclination is ca. 46.69°. Laplace infers from $P(m \leq 46.69^\circ) \approx 3/4$ that the hypothesis of equiprobability can not be rejected. Therefore it is not possible to assume a "cause primitive" exerting influence on the angles.

Critical question

What if all comets lay within a joint plane with an angle of inclination of ca. 47°?

Laplace $\sim$ 1810, CLT

Laplace $\sim$ 1810, CLT

Not a general theorem but only stated for each specific case of application

Laplace $\sim$ 1810, CLT

Not a general theorem but only stated for each specific case of application
"translation":
$X_1, X_2, \ldots, X_n$ independent identically distributed r.v.'s (density or lattice distr.)
$\mu := \mathsf{E}X_1$, $\sigma^2 := \mathsf{E}(X_1 - \mu)^2$, n "very large"

$$P\left(n\mu + r_1\sqrt{n} \leq \sum X_k \leq n\mu + r_2\sqrt{n}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{r_1}^{r_2} \frac{1}{\sigma} exp\left(-\frac{t^2}{2\sigma^2}\right) dt.$$

Simple linear model: $Y_i = ax_i + \epsilon_i$ $(i = 1, \ldots, n)$, x_i known, ϵ_i unknown errors, a to be estimated.

Ansatz: $\sum \lambda_i Y_i = \hat{a} \sum \lambda_i x_i$, $\hat{a} - a = (\sum \lambda_i \epsilon_i)/(\sum \lambda_i x_i)$

CLT: $P(|(\sum \lambda_i \epsilon_i)/(\sum \lambda_i x_i)| > r) \stackrel{!}{=} 1 - \Phi_{0,s^2}(r)$,
$s^2 = (\sum \lambda_i^2 \sigma^2)/(\sum \lambda_i x_i)^2$

For all $r > 0$ this prob. gets minimum if s^2 gets minimum. This is the case for $\lambda_i = kx_i$, or

$$\hat{a} = \left(\sum x_i Y_i\right)/\left(\sum x_i^2\right)$$

Least squares estimator.

Hypothesis: If there were no stochastic oscillations, then the air pressure at 4 p.m. would always be above that at 9 a.m.

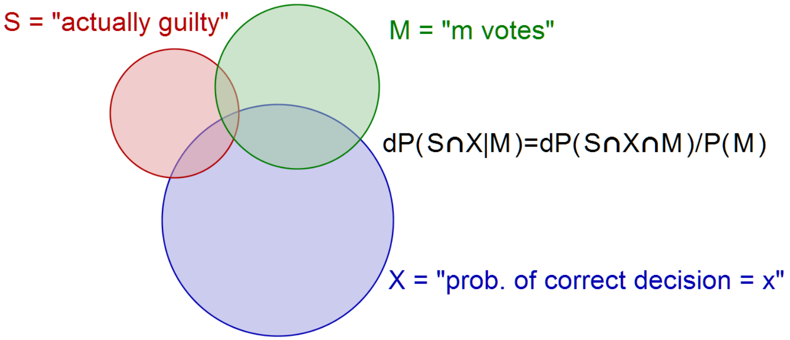
L. refers to 400 observations of differences d_i ("afternoon - morning", $-4\mathrm{mm} \leq d_i \leq 4\mathrm{mm}$) with the result $\sum d_i = 400$mm. He assumes that d_i have a unimodal and symmetric density with expectation 0 and variance σ^2. In this case: $\sigma^2 \leq 16/3$.

Then:

$$P(\sum d_i < 400) \approx \Phi_{0,\sigma^2}(400) \geq \Phi_{0,16/3}(400) = 1 - 10^{-18}.$$

With moral certainty the sum of differences would be less than the observed. Therefore ...

Probabilities of court decisions, 1820



$$dP(S \cap X|M) = \frac{x^m(1-x)^{n-m}f(x)dx}{\int_{0.5}^{1}(x^m(1-x)^{n-m} + x^{n-m}(1-x)^m)f(x)dx}$$

$$P(S|M) = \frac{\int_{0.5}^{1} x^m(1-x)^{n-m}dx}{\int_{0}^{1} x^m(1-x)^{n-m}dx}.$$

Résumé

Résumé

- Ab initio Laplace investigates both direct and inverse probabilities

Résumé

- Ab initio Laplace investigates both direct and inverse probabilities
- He chooses the methods in a pragmatic way

Résumé

- Ab initio Laplace investigates both direct and inverse probabilities
- He chooses the methods in a pragmatic way
- A canonized theory of estimating and testing is not visible yet

Résumé

- Ab initio Laplace investigates both direct and inverse probabilities
- He chooses the methods in a pragmatic way
- A canonized theory of estimating and testing is not visible yet
- L's pragmatism corresponds to his consideration of subjective as well as objective probabilities

Résumé

- Ab initio Laplace investigates both direct and inverse probabilities
- He chooses the methods in a pragmatic way
- A canonized theory of estimating and testing is not visible yet
- L's pragmatism corresponds to his consideration of subjective as well as objective probabilities
- L's application of inverse probabilities is essentially restricted to binomial models.

Résumé

Résumé

- "Typical" applications of inverse probabilities, as to testimonies or court decisions, are more and more opposed during the 19th century as an "abuse" of probability theory. This process significantly contributes to the decline of inverse probabilities after Laplace

Résumé

- "Typical" applications of inverse probabilities, as to testimonies or court decisions, are more and more opposed during the 19th century as an "abuse" of probability theory. This process significantly contributes to the decline of inverse probabilities after Laplace
- The CLT instantly amplifies the range of application of direct probabilities; in this way a development begins that, eventually, in the 20th century leads to the predominance of so called "frequentist" statistics.