Why Every Physicist Should Be a Bayesian

(Towards a Complete Reconciliation between the Bayesian and the Frequentist Schools of Parametric Inference)

Tomaž Podobnik

Physics Department, University of Ljubljana

Jožef Stefan Institute, Ljubljana, Slovenia

Recommended reading:

R. D. Cousins, *"Why Isn't Every Physicist a Bayesian?"*, Amer. J. Phys. **63** (1995) 398.

"Physicists embarking on seemingly routine error analyses are finding themselves grappling with major conceptual issues which have divided the statistical community for years. …The lurking controversy can come as a shock to a graduate student who encounters a statistical problem at some late stage in writing up the Ph.D. dissertation."

Basic Principles of scientific reasoning (Popper, 1959, pp. 91-92):

1. Principle of Consistency: Every theory must be internally consistent: if a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result. Also, identical states of knowledge in a problem must always lead to identical solutions of the problem.

2. Operational Principle: Every theory must specify operations that ensure falsifiability of its predictions.

Direct probabilities (=long term relative frequencies):

$p(x_1 \mid I)$: probability for observing $x = x_1$ (for observing $x \in (x_1, x_1 + dx)$), given information I

$f(x \mid I)$: probability density function (pdf);
$p(x \mid I) = f(x \mid I)\, dx$

$I = \theta I_0$: I_0 = family of sampling distributions
θ = parameter

$F(x, \theta, I_0) \equiv \int_{x_a}^{x} f(x' | \theta I_0)\, dx'$: (cumulative) distribution function (cdf)

Location and scale parameters:

$$f(x \mid \mu\, I_0) = \phi(x-\mu); \qquad x \in (-\infty, \infty)$$

$\mu \in (-\infty, \infty) \equiv$ location parameter

$$f(x \mid \sigma I_0) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right); \qquad x \in (0, \infty)$$

$\sigma \in (0, \infty) \quad \equiv$ scale parameter

$$f(x \mid \mu \sigma I_0) = \frac{1}{\sigma}\phi\left(\frac{x\text{-}\mu}{\sigma}\right); \qquad x \in (-\infty, \infty)$$

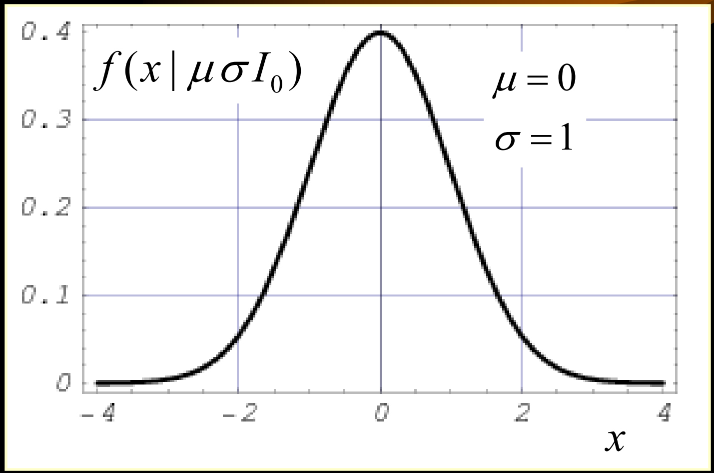
$\mu \in (-\infty, \infty) \equiv$ location parameter

$\sigma \in (0, \infty) \quad \equiv$ scale (dispersion) parameter

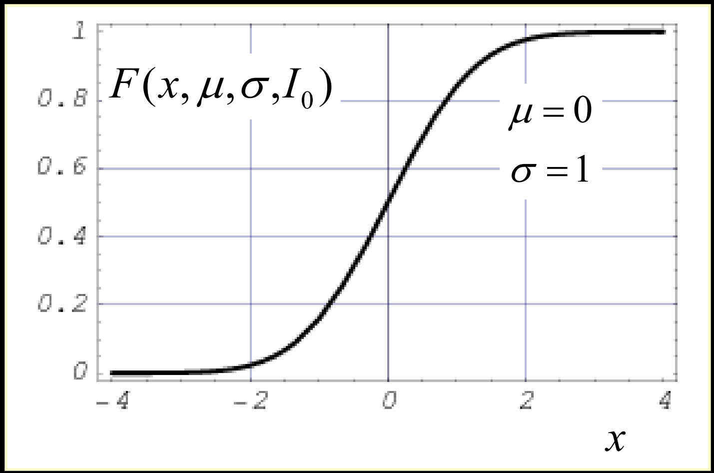
Examples:

$$f(x\,|\,\mu\sigma I_0) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} = \frac{1}{\sigma}\phi\left(\frac{x\text{-}\mu}{\sigma}\right)$$

$\mu \equiv$ location parameter
$\sigma \equiv$ scale (dispersion) parameter
$I_0 \equiv$ Gaussian distribution



$$F(x,\mu,\sigma,I_0) = \int_{-\infty}^{x} f(x'|\mu\sigma I_0)\,dx'$$



Axioms of conditional probability:

- every probability distribution is conditional upon the available (relevant) information.

1. $f(x\,|\,\theta I_0) \geq 0$

2. $$f(xy\,|\,\theta I_0) = f(x\,|\,\theta I_0)\, f(y\,|\,x\theta I_0) = f(y\,|\,\theta I_0)\, f(x\,|\,y\theta I_0)$$

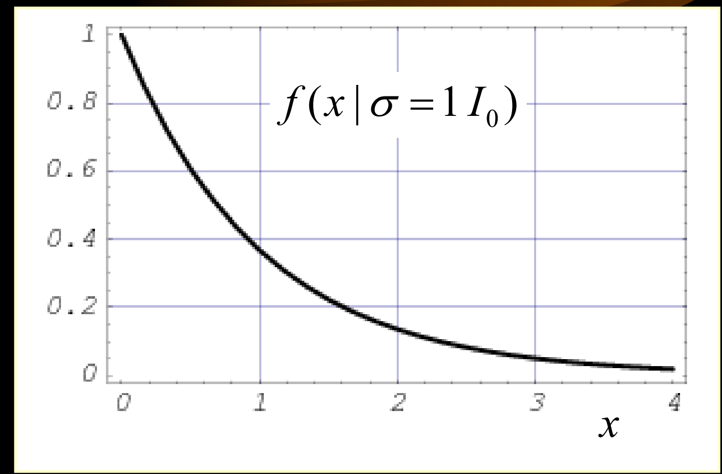
3. $$\int_X f(x\,|\,\theta I_0)\, dx = 1$$

4. $$f(y\,|\,\theta \tilde{I}_0) = f(x\,|\,\theta I_0) \left|\frac{\partial y}{\partial x}\right|^{-1}; \quad y = y(x) \text{ one-to-one}$$

Example:

$$f(x\,|\,\sigma I_0) = \frac{1}{\sigma}\exp\left\{-\frac{x}{\sigma}\right\} = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)\;;$$

$\Rightarrow \sigma =$ scale parameter



$$\left.\begin{array}{l} y \equiv \ln x \\ \mu \equiv \ln \sigma \end{array}\right\} \Rightarrow f(y\,|\,\mu \tilde{I}_0) = f(x\,|\,\sigma I_0)\left|\frac{\partial y}{\partial x}\right|^{-1}$$

$$= e^{(y-\mu)}\exp\left\{-e^{(y-\mu)}\right\}$$

$\equiv \tilde{\phi}(y-\mu) \Rightarrow \mu =$ location parameter

$I_0 \equiv$ exponential distribution $\rightarrow \tilde{I}_0 :$ distribution not exponential

Scale parameter reducible to location parameter!

Parametric inference:

Given measured $x \in (x_1, x_1 + dx)$, specify degree of belief

$$(\theta_1 \mid x_1 I_0): \quad \theta \in (\theta_1, \theta_1 + d\theta)$$

Probabilistic approach (Bayesian school):

$$(\theta \mid x I_0) \to p(\theta \mid x I_0) = f(\theta \mid x I_0)\, d\theta$$

N. b.: $f(\theta \mid x\, I_0)$ distribution of our belief in different values of θ, not (!) distribution of θ.

Axioms of inverse probability:

1. $f(\theta \mid x I_0) \geq 0$

2. $f(\theta \nu \mid x I_0) = f(\theta \mid x I_0)\, f(\nu \mid \theta x I_0)$
$\qquad = f(\nu \mid x I_0)\, f(\theta \mid \nu x I_0)$

3. $\int_{\Theta} f(\theta \mid x I_0)\, d\theta = 1$

4. $f(\nu \mid x \tilde{I}_0) = f(\theta \mid x I_0) \left| \frac{\partial \nu}{\partial \theta} \right|^{-1}; \quad \nu = \nu(\theta)$ one-to-one

5. $f(\theta x_2 \mid x_1 I_0) = f(\theta \mid x_1 I_0)\, f(x_2 \mid \theta x_1 I_0)$
$\qquad = f(x_2 \mid x_1 I_0)\, f(\theta \mid x_2 x_1 I_0)$

Pro's for subjecting degrees of belief to the Axioms of probability:

1. "It is not excluded *a priori* that the same mathematical theory may serve two purposes." (Pólya, 1954, Chapter XV, p. 116)

2. Cox's Theorem: Every theory of plausible inference is either isomorfic to probability theory or inconsistent with very general qualitative requirements (e.g., $(\theta \in (\theta_1, \theta_1 + \theta) | x_1, I_0) \rightarrow (\theta \notin (\theta_1, \theta_1 + \theta) | x_1, I_0)$). (Cox, 1946)

3. Dutch Book Theorem (de Finetti): A "Dutch Book" can be organized against anyone whose betting coefficients violate axioms of probability. (Howson and Urbach, 1991)

Pro's (cont'd):

4. Avoiding adhockeries. (O'Hagan, 2000, p. 20)

5. Powerful tools: marginalization and Bayes' Theorem (Bayes, 1763)

$$f(\theta\nu\,|\,xI_0) = f(\theta\,|\,xI_0)\,f(\nu\,|\,\theta xI_0) \Rightarrow \boxed{f(\theta\,|\,xI_0) = \int_{\mathrm{N}} f(\theta\nu'\,|\,xI_0)\,d\nu'}$$

$$f(\theta\nu\,|\,xI_0) = f(\nu\,|\,xI_0)\,f(\theta\,|\,\nu xI_0) \Rightarrow \boxed{f(\nu\,|\,xI_0) = \int_{\Theta} f(\theta'\nu\,|\,xI_0)\,d\theta'}$$

$$f(\theta\,|\,x_1I_0)\,f(x_2\,|\,\theta x_1I_0) = f(x_2\,|\,x_1I_0)\,f(\theta\,|\,x_2x_1I_0); \quad f(x_2\,|\,\theta x_1I_0) = f(x_2\,|\,\theta I_0)$$

$$\Rightarrow \boxed{f(\theta\,|\,x_2x_1I_0) = \frac{f(\theta\,|\,x_1I_0)\,f(x_2\,|\,\theta I_0)}{f(x_2\,|\,x_1I_0)};} \quad f(x_2\,|\,x_1I_0) = \int_{\Theta} f(\theta'\,|\,x_1I_0)\,f(x_2\,|\,\theta' I_0)\,d\theta'$$

But…(con's): a) how to assign $f(\theta \mid x_1 I_0)$???
b) what are verifiable predictions???

If making use of Bayes' Theorem:

$$f(\theta \mid x_1 I_0) = \frac{f(\theta \mid I_0)\, f(x_1 \mid \theta I_0)}{\int_{\Theta} f(\theta' \mid I_0)\, f(x_1 \mid \theta' I_0)\, d\theta'}$$

$f(\theta \mid I_0)$: non-informative prior (distribution)

"According to Bayesian philosophy it is also possible to make statements concerning the unknown θ in the absence of data, and these statements can be summarized in a prior distribution."
(Villegas, 1980)

Example: The Principle of Insufficient Reason (Bayes, 1763; Laplace, 1886, p. XVII)

$$f(\theta\,|I_0) = C; \quad C^{-1} = \int_{\theta_a}^{\theta_b} f(\theta'|I_0)\,d\theta' = \theta_b - \theta_a$$

Twofold problem:

a) (θ_a,θ_b) infinite (e.g., $\theta_b = \infty$) $\Rightarrow\ \nexists \int_{\theta_a}^{\theta_b} f(\theta'|I_0)\,d\theta'$

b) $f(\theta\,|I_0)$ not invariant under non-linear transformations

$$\nu = \theta^2 \Rightarrow f(\nu\,|\,\tilde{I}_0) = f(\theta\,|\,I_0)\left|\frac{\partial\nu}{\partial\theta}\right|^{-1} \propto \frac{1}{\sqrt{\nu}} \neq \text{const}\ (\rightarrow\!\!\ast\!\!\leftarrow)$$

"A succession of authors have said that the prior probability is nonsense and that the principle of inverse probability, which cannot work without it, is nonsense too." (Jeffreys, 1961, p. 120)

"During the rapid development of practical statistics in the past few decades, the theoretical foundations of the subject have been involved in great obscurity. The obscurity is centred in the so-called 'inverse' methods. … The inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply commited itself)." (Fisher, 1922)

Long-lasting and fierce controversy:

"The essence of the present theory is that no probability, direct, prior, or posterior, is simply a frequency." (Jeffreys, 1961, p. 401)

"Probability is a ratio of frequencies." (Fisher, 1922, p.326)

Twofold aim of the lecture:

1. Overcome conceptual and practical problems concerning assignment of probability distributions to inferred parameters;

2. Reconcile the Bayesian and the frequentist schools of parametric inference.

Consistency Theorem: How to assign $f(\theta|x_1 I_0)$?

Assumptions:

a) x_1 and x_2 two independent measurements from $f(x|\theta I_0)$:

$f(x_2|x_1\theta I_0) = f(x_2|\theta I_0)$ and $f(x_1|x_2\theta I_0) = f(x_1|\theta I_0)$;

b) $f(\theta|x_1 I_0)$ and $f(\theta|x_2 I_0)$ can be assigned.

Then (Bayes' Theorem):

$$f(\theta|x_2 x_1 I_0) = \frac{f(\theta|x_1 I_0)\, f(x_2|\theta I_0)}{f(x_2|x_1 I_0)} \quad ; \quad f(\theta|x_1 x_2 I_0) = \frac{f(\theta|x_2 I_0)\, f(x_1|\theta I_0)}{f(x_1|x_2 I_0)}$$

Consistency:

$$f(\theta \,|\, x_2 x_1 I_0) = f(\theta \,|\, x_1 x_2 I_0)$$

$$\Rightarrow \boxed{f(\theta \,|\, x I_0) = \frac{\pi(\theta)\, f(x \,|\, \theta I_0)}{\eta(x)}}$$

Strikingly similar to Bayes' Theorem, but…

$\pi(\theta)$: consistency factor; not(!!) probability distribution (e.g., need not be normalizable);

$\eta(x)$: normalization factor $\quad \eta(x) \equiv \int_{\Theta} \pi(\theta')\, f(x \,|\, \theta' I_0)\, d\theta'$

Properties of $\pi(\theta)$:

1. Determined only up to a multiplication constant (say k);

2. Transformation $\pi(\theta) \to \tilde{\pi}(\nu)$ under $\theta \to \nu$ (one-to-one):

$$\left.\begin{aligned} f(\nu \mid x\tilde{I}_0) &= f(\theta \mid x I_0) \left|\frac{\partial \nu}{\partial \theta}\right|^{-1} \\ f(\nu \mid x\tilde{I}_0) &= \frac{\tilde{\pi}(\nu)\, f(x \mid \nu \tilde{I}_0)}{\tilde{\eta}(x)} \end{aligned}\right\} \Rightarrow \boxed{\tilde{\pi}(\nu) = k\, \pi(\theta) \left|\frac{\partial \nu}{\partial \theta}\right|^{-1}}$$

3. Depends on I_0 (=the only available information before data $(x_1, x_2, \ldots)$ are collected).

Consistency: $\tilde{I}_0 = I_0 \Rightarrow \tilde{\pi}(\nu) \propto \pi(\nu)$ (a.k.a. The Principle of Relative Invariance; Hartigan, 1964)

Example:

$$f(t \mid \tau I_0) = \frac{1}{\tau}\exp\left\{-\frac{t}{\tau}\right\} = \frac{1}{\tau}\phi\left(\frac{t}{\tau}\right) \quad (\tau = \text{scale parameter});$$

$$g_a \in \mathcal{G} : t \to g_a(t) = at \equiv y; \qquad \text{group } \mathcal{G} : X \to X$$

$$\bar{g}_a \in \bar{\mathcal{G}} : \tau \to \bar{g}_a(\tau) = a\tau \equiv \nu; \ \text{(induced) group } \bar{\mathcal{G}} : \Theta \to \Theta$$

$$\Rightarrow f(y \mid \nu \tilde{I}_0) = f(t \mid \tau I_0)\left|\frac{\partial \nu}{\partial \tau}\right|^{-1} = \frac{1}{\nu}\exp\left\{-\frac{y}{\nu}\right\} = \frac{1}{\nu}\phi\left(\frac{y}{\nu}\right) = f(y \mid \nu I_0)$$

$\Rightarrow f(t \mid \tau\, I_0)$ invariant under $\mathcal{G}$ $\Rightarrow$ $\pi(a\tau) = k(a)\pi(\tau)\frac{1}{a} \equiv h(a)\pi(\tau)$

Similarly:

Distribution	Inv. transformation	Functional equation	Solution
$f(x\mid\mu\sigma I_0)=\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$	$x\to ax+b$ $\mu\to a\mu+b$ $\sigma\to a\sigma$	$\pi(a\mu+b,a\sigma)=h(a,b)\pi(\mu,\sigma)$	$\pi_{LS}(\mu,\sigma)\propto\sigma^{-r}$
$f(x\mid\mu\sigma I_0)=\phi(x-\mu)$	$x\to x+b$ $\mu\to\mu+b$	$\pi(\mu+b)=h(b)\pi(\mu)$	$\pi_L(\mu)\propto e^{-q\mu}$
$f(x\mid\mu\sigma I_0)=\frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)$	$x\to ax$ $\sigma\to a\sigma$	$\pi(a\sigma)=h(a)\pi(\sigma)$	$\pi_S(\sigma)\propto\sigma^{-(q+1)}$

r,q : constants

Product rule and Marginalization:

$$\pi_{LS}(\mu,\sigma) \propto \pi_L(\mu)\,\pi_S(\sigma)$$

$$\Rightarrow \quad \pi_L(\mu) \propto 1 \text{ and } \pi_S(\sigma) \propto \pi_{LS}(\mu,\sigma) \propto \sigma^{-1}$$

Consistency factors determined uniquely (up to an arbitrary multiplication constant) exclusively by observing the Axioms of Probability and the Principle of Consistency.

Examples: Inferring parameters of Gaussian distribution.

$\mathbf{x} \equiv (x_1, x_2, \ldots, x_n)$ independent measurements, sampled from

$$f(x \mid \mu \sigma I_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$\bar{x}_n \equiv \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad s_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

a) Both μ and σ unknown:

$$f(\mu\sigma \mid \mathbf{x} I_0) \propto \pi_{LS}(\mu,\sigma)\, f(\mathbf{x} \mid \mu\sigma I_0) = \pi_{LS}(\mu,\sigma) \prod_{i=1}^{n} f(x_i \mid \mu\sigma I_0) \propto \sigma^{-1} \prod_{i=1}^{n} f(x_i \mid \mu\sigma I_0)$$

Marginalization:

$$f(\mu|\mathbf{x}I_0)=\int_0^\infty f(\mu\sigma'|\mathbf{x}I_0)\,d\sigma'=\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\frac{\left(ns_n^2\right)^{n/2}}{\sqrt{\pi s_n^2}}\frac{1}{\left[ns_n^2+n(\bar{x}_n-\mu)^2\right]^{n/2}}$$

$$f(\sigma|\mathbf{x}I_0)=\int_{-\infty}^\infty f(\mu'\sigma|\mathbf{x}I_0)\,d\mu'=\frac{\left(ns_n^2\right)^{(n-1)/2}}{2^{(n-3)/2}\Gamma((n-1)/2)}\frac{1}{\sigma^n}\exp\left\{-\frac{ns_n^2}{2\sigma^2}\right\}$$

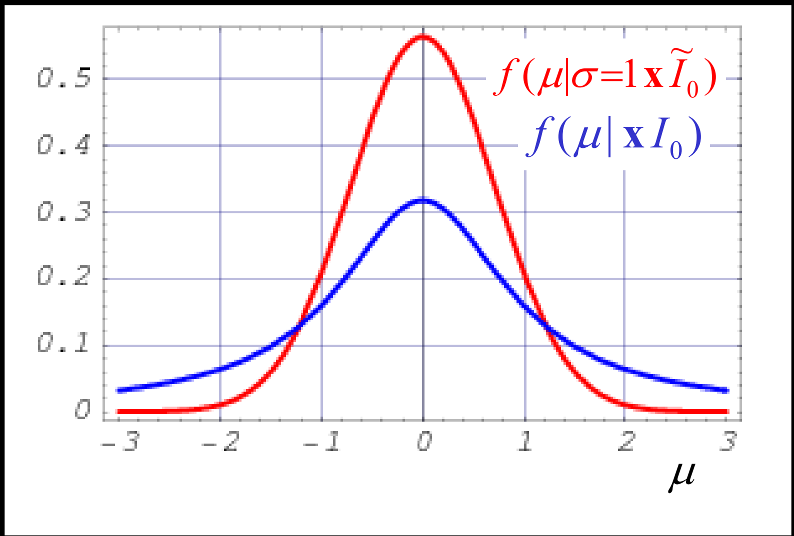
b) Only μ unknown:

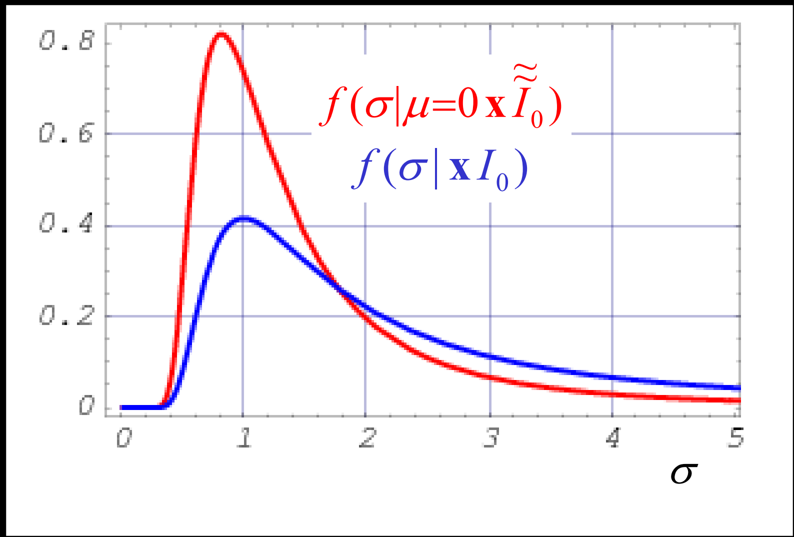
$$f(\mu|\sigma\mathbf{x}\tilde{I}_0)\propto\pi_L(\mu)\,f(\mathbf{x}|\mu\sigma I_0)=\pi_L(\mu)\prod_{i=1}^n f(x_i|\mu\sigma I_0)\propto\sqrt{\frac{n}{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{n(\bar{x}_n-\mu)^2}{2\sigma^2}\right\}$$

c) Only σ unknown:

$$f(\sigma|\mu\mathbf{x}\tilde{\tilde{I}}_0)\propto\pi_S(\sigma)\,f(\mathbf{x}|\mu\sigma I_0)\propto\frac{\left[n(\bar{x}-\mu)^2+ns_n^2\right]^{n/2}}{\Gamma(n/2)2^{n/2-1}}\frac{1}{\sigma^{n+1}}\exp\left\{-\frac{n(\bar{x}-\mu)^2+ns_n^2}{2\sigma^2}\right\}$$

For $n=2,\ \bar{x}_2=0$ and $s_2^2=1$:





Comments:

a) Consistency factors not normalizable, e.g., $\nexists \int_{-\infty}^{\infty} \pi(\mu')\, d\mu'$,

$\Rightarrow \pi(\theta)$ not a probability distribution!!!

b) Consistency factors for the parameters of distributions that are invariant under Lie groups of transformations.

Necessary condition: reducibility of θ to location parameter (not a disaster; see below).

$\Rightarrow$ enough to determine $\pi(\mu)$.

"The most striking achievement of physical sciences is prediction."
(Pólya, 1954, p. 64)

Calibration (coverage):

- $f(\theta \mid x\, I_0)$ calibrated if coverage of confidence intervals (θ_1, θ_2) coincides with probability

$$P\big(\theta \in (\theta_1, \theta_2) \mid x I_0\big) = \int_{\theta_1}^{\theta_2} f(\theta' \mid x I_0)\, d\theta'$$

- Fiducial theory: $\boxed{f(\theta \mid x I_0) = \left|\frac{\partial}{\partial \theta} F(x, \theta, I_0)\right|}$ ($F(x,\theta,I_0)$ monotone in θ; Fisher, 1956, p. 70)

Important:

1. $\pi_L(\mu)=1$ and $\pi_S(\sigma)=\pi_{LS}(\mu,\sigma)=\sigma^{-1}$ ensure calibrated inferences;

2. Exact calibration $\Rightarrow$ "Dutch Book" impossible;

3. Consistency theorem and Fiducial argument combined $\Rightarrow$ θ necessarily reducible to a location parameter (Lindley, 1958).

Therefore:

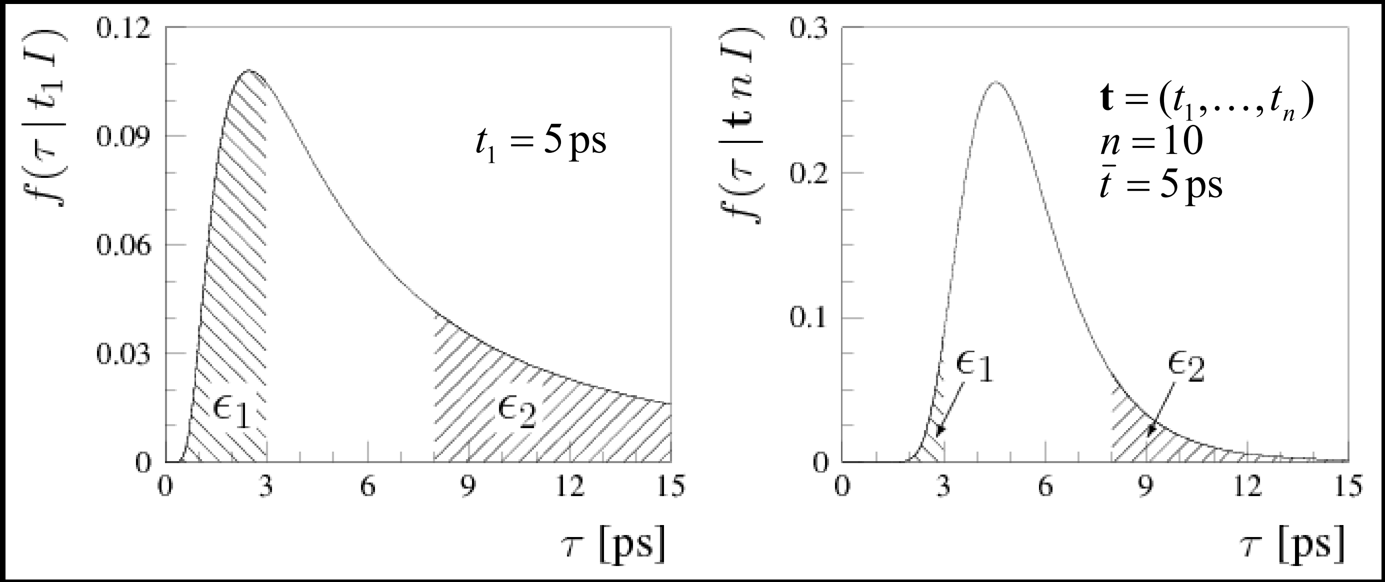
The Principle of Consistency and The Operational Principle are equivalent (identical consistency factors & applicable under identical circumstances).

⇒ complete reconciliation between the Bayesian and the Frequentist schools of parametric inference!!!

Probabilistic parametric inference not universal (e.g., pre-constrained parameters, counting experiments).

Remedy (under fairly general conditions): “Repetitio est mater studiorum.” (Latin proverb)

Example: inferring pre-constrained τ of an exponential distribution.

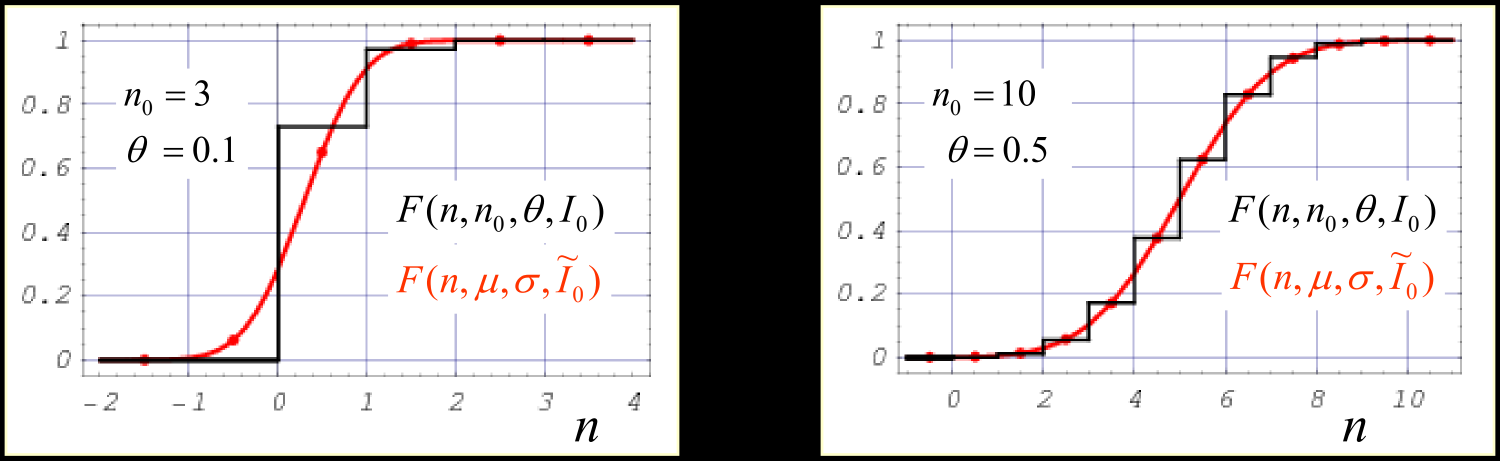


Example: inferring parameter θ of a binomial distribution

$$p(n|\theta n_0 I_0) = \binom{n_0}{n} \theta^n (1-\theta)^{n_0-n} \;; \quad n_0 \in \mathbb{N}, \quad n \in \mathbb{N}_0 ; n \le n_0$$

$$n_0\theta, n_0(1-\theta) \gg 1: \quad F(n, n_0, \theta, I_0) = \sum_{i=0}^{n} p(i|\theta n_0 I_0) \simeq \int_{-\infty}^{n+0.5} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x'-\mu)^2}{2\sigma^2}\right\} dx'$$

$$\mu = n_0\theta \,, \quad \sigma = \sqrt{n_0\theta(1-\theta)}$$



Conclusions:

1. Consistency Theorem (instead of Bayes' Theorem) for assigning $f(\theta|x_1 I_0)$;

2. Equivalence of the Consistency Principle and the Operational Principle for determination of $\pi(\theta)$;

3. Equivalence of the Bayesian and the frequentist schools of parametric inference.

Applications:

1. Simple parametric inference;
2. Inference about the parameters of linear models (e.g., histogram fitting and partial wave analyses) (Stuart, Ord and Arnold, 1999);
3. Inference about the parameters of dynamical models: $\theta=\theta(t)$ (e.g., Kalman filter (Brown and Hwang, 1983));
4. Predictive distributions ($\mathbf{x}=(x_1,x_2,\ldots,x_n)$ from $f(x\,|\,\theta\,I_0) \rightarrow f(x_{n+1}\,|\,\mathbf{x}\,I_0)$).

Warning:

Several "Principles" for determination of $f(\theta | I_0)$: the Laplace Principle of Insufficient Reason (Bayes, 1763; Laplace, 1886, p. XVII), the Principle of Maximum Entropy (Jaynes, 2003, pp. 343-377), Reference Priors (Bernardo, 1979), the Principle of Group (Form) Invariance (Harney, 2003), the Principle of Reduction (Dawid, 1977):

a) resulting $f(\theta | I_0)$ not unique;

b) "Principles" inconsistent with Axioms of inverse probability;

c) Non-calibrated inferences.

Which kind of approach has been being advocated, frequentist or Bayesian?

Depends....

If:

1. Frequentist ≡ axioms of conditional probability only applicable to sampling distributions.

2. Bayesian ≡ (non-informative) prior probability distributions indispensable in the process of inference.

…then none of the two.

If:

1. Frequentist ≡ observing the Operational Principle.
2. (Objective) Bayesian ≡ observing the Principle of Consistency.

…then both.

Bibliography:

T.P. and Živko, T. (2006). Towards Reconciliation between Bayesian and Frequentist Reasoning. In Lyons, L. and Ünel, M. K. (eds.). *Statistical Problems in Particle Physics, Astrophysics and Cosmology (Proceedings of PHYSTAT05).* London: Imperial College Press.

Erratum: Inference about the parameters of Weibull distribution can be reduced to a location-scale problem.

T.P. and Živko, T. On Probabilistic Inference about the Parameters of Sampling Distributions.

References:

Bayes, Rev. T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philos. Trans. R. Soc. Lond.*, **53**: 370-418.

Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *J. R. Statist. Soc.*, **B 41**: 113-147.

Brown, R. G. and Hwang, P. Y. C. (1983). *Introduction to Random Signals and Applied Kalman Filtering.* John Wiley & Sons, Inc.

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation. *Amer. J. Phys.*, **14**: 1-13.

Dawid, A. P. (1977). Conformity of inference patterns. In Barra, J. R., van Cutsen, B., Brodeau, F. and Romier, G. (eds.). *Developments in Statistics.* Amsterdam: North-Holland.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. Lond.*, **A 222**: 309-368.

References (cont'd):

Fisher, R. A. (1956). *Satistical Methods and Scientific Inference.* Edinbourgh: Oliver & Boyd.

Harney, H. L. (2003). *Bayesian Inference.* Springer.

Hartigan, J. A. (1964). Invariant Prior Distributions. *Ann. Math. Statist.,* **35**: 836-845.

Howson, C. and Urbach, P. (1991). Bayesian Reasoning in Science. *Nature,* **350**: 371-374

Jaynes, E. T. (2003). *Probability Theory – The Logic of Science.* Cambridge University Press.

Jeffreys, H. (1961). *Theory of Probability.* Oxford: Clarendon Press.

Laplace, P. S. (1886) *Œvres Complètes – Tome Septième: Théorie Analitique des probabilités.* Paris: Gauthier-Villars.

Lindley, D. V. (1958). Fiducial Distribution and Bayes' Theorem. *J. R. Statist. Soc.,* **B 20**: 102-107.

References (cont'd):

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics, Vol. 2B – Bayesian Inference.* London: Arnold.

Pólya, G. (1954). *Mathematics and Plausible Reasoning, Vol.2 – Patterns of Plausible Inference.* Princeton: Princeton University Press.

Popper, K. R. (1959). *The Logic of Scientific Discovery.* London: Hutchinson & Co. Publishers.

Stuart, A., Ord., K. and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics, Vol. 2A – Classical Inference and the Linear Model.* London: Arnold.

Villegas, C. (1980). Inner Statistical Inference II. *Ann. Statist.*, **9**: 768-776.