# On Probabilistic Inference about the Parameters of Sampling Distributions

Tomaž Podobnik

*Faculty of Mathematics and Physics, University of Ljubljana, and ”Jožef Stefan” Institute, Ljubljana, Slovenia.*

and Tomi Živko

*”Jožef Stefan” Institute, Ljubljana, Slovenia.*

**Summary**. The problem of the so-called non-informative prior probability distributions has been solved. The solution of the problem allows for construction of a consistent and calibrated theory of probabilistic inference about the parameters of sampling distributions. The constructed theory speaks in favour of a complete reconciliation between the Bayesian and the frequentist schools of inference.

*Keywords*: Bayesian parametric inference; non–informative priors; frequentist interpretation; predictive distributions

## 1. Introduction

We make a probabilistic inference about a parameter of a sampling distribution by specifying a probability distribution that corresponds to the distribution of our belief in different values of the parameter. Such a distribution is always conditional upon information at hand that is relevant to the inferred parameter. When new pieces of information arrive, the distribution can sequentially be updated by using the well-known Bayes’ Theorem.

The problems arise, however, when the Theorem is used not only for updating the assigned probability distribution, but also at the starting point of the inference, i.e., in attempts to assign the probability distribution that is later to be updated. For then, we must specify the so-called *non-informative prior probability distribution*, reflecting our belief in different values of the inferred parameter when we are in a state of complete ignorance about these values. But since all probability distributions are, by definition, conditional upon the relevant information, the non-informative prior distribution, conditional upon the complete lack of such information, is simply *a contradiction in terms*. Apart from the problems on the conceptual level, the non-informative prior distributions have also been representing an insurmountable practical problem, since the question about the explicit form of these distributions remains unanswered (for a survey of the topic see, for example, Kass and Wasserman, 1996, and the references quoted therein).

The difficulties with prior distributions have been serving as the main argument against subjecting the possible systems for inference about the parameters to the axioms of probability. As realized long ago by Jeffreys (1961, § 3.1, p. 120), ”a succession of authors have said that the prior probability is nonsense and therefore that the principle of inverse probability, which cannot work without it, is nonsense too.” Fisher (1922), for example, wrote: ”During the rapid development of practical statistics in the past few decades, the theoretical foundations of the subject have been involved in great obscurity. This 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 committed itself)”. Here, the inverse probability stands for the probability, assigned to the certain value of an inferred parameter.

In this way two, at first glance fundamentally distinct, schools of inductive reasoning emerged. The first one, usually referred to as the *Bayesian school* due to the central role of the Bayes’ Theorem in the process of inference, recognizes probability as a degree of reasonable belief and applies probability theory in the course of inductive reasoning. The second one, usually referred to as the *frequentist school* due to its strict frequency interpretation of probability, advocates the usage of the calculus of probability only for treatment of so-called *random phenomena*. The aim of the frequentist school is to avoid the supposed mistakes and inconsistencies of the probabilistic inductive inference, so they relegate the problems of inductive inference, e.g., the problem of inference about parameters, to a new field, *statistical inference*.

The purpose of the present article is twofold. Our first and main goal is to demonstrate that the problems, both the conceptual and the practical ones, of the non-informative prior distributions *can* be solved and that a consistent theory of probabilistic inference about the parameters of sampling distributions can be deduced from very general Principles of scientific reasoning. Second, we are aiming at a complete reconciliation between the frequentist and the Bayesian approaches to the parametric inference.

## 2. Direct probabilities

### 2.1. Sampling variates and their distributions

Probability $p(x_i|I) \equiv P(x = x_i|I)$ for *a sampling* (or *random*) *variate* $x$ from a discrete *sample space* $X$ to take a value $x_i$ is defined as a long run relative frequency,

$$p(x_i|I) = \lim_{N_0 \to \infty} \frac{N_i}{N_0} \ ,$$

where $N_0$ is the total number of recorded values of $x$, while $N_i$ is the number of the outcomes $x = x_i$. On the other hand, for (*absolutely*) *continuous sampling distributions*, with sample spaces coinciding with subintervals of real numbers, $X \subseteq \mathbb{R}$, $p(x_i|I)$ denotes the probability $P\big(x \in (x_i, x_i + dx)|I\big)$ for $x$ to take a value in an interval $(x_i, x_i + dx)$, and can be expressed by a non-negative function $f(x_i|I)$, called *the probability density function* (*pdf*),

$$p(x_i|I) \equiv f(x_i|I)\, dx \ .$$

Every probability distribution $p(x|I)$ is *conditional* upon *the* (*state of*) *information* or *knowledge* $I$ about the variate $x$. Let $I_\circ$ denote a family of sampling distributions and let $\theta$ be *a parameter* whose value specifies a unique distribution within the family. Then, for parameters from a discrete *parameter space* $\Theta$, our knowledge $I = \theta_j I_\circ$ about the distribution $p(x|I)$ consists of the known family $I_\circ$ and the value $\theta = \theta_j$ of the parameter. Similarly, for parameters from the parameter spaces $\Theta = (\theta_a, \theta_b) \subseteq \mathbb{R}$, the sampling distribution from a family $I_\circ$ is uniquely determined by knowing that an (infinitesimal) interval $(\theta_j, \theta_j + d\theta)$ contains the so-called *true value* of the parameter. For example, a sampling distribution of a variate $t$ from the exponential family $I_\circ$ is uniquely determined by the value of the parameter in the interval $(\tau, \tau + d\tau)$.

ASSUMPTION 1. *In the present paper, we restrict our considerations to sampling distributions whose parameter spaces are (possibly infinite) subintervals of positive length of real numbers,* $\Theta \subseteq \mathbb{R}$*. For* $m$*-dimensional parameters,* $\boldsymbol{\Theta} \subseteq \mathbb{R}^m$ *are Cartesian products of* $m$ *such intervals.*

*Second, we require the sampling distributions be continuous functions of their parameters. Third, the sampling distributions are to be strictly positive for all $x \in X$ and for all $\theta \in \Theta$, i.e.,*

$$p(x|\theta I_\circ) > 0 \quad ; \quad \forall\, x \in X \quad \textit{and} \quad \forall\, \theta \in \Theta$$

*for discrete sampling distributions, and*

$$f(x|\theta I_\circ) > 0 \quad ; \quad \forall\, x \in X \quad \textit{and} \quad \forall\, \theta \in \Theta$$

*for continuous sampling distributions.*

For example, Gaussian (normal) distribution with the pdf $f(x|\mu\sigma I_\circ)$ is strictly positive for all real $x$ and $\mu$ and for all positive $\sigma$, while binomial distribution $p(n|\theta n_0 I_\circ)$ with positive integer $n_0$ and with integer $n$, $0 \le n \le n_0$, is strictly positive only if the parameter space $\Theta = (0,1)$ is open on both sides.

Sampling probability distributions are subject to the *axioms of conditional probability*. First, every joint probability $p(x_0 y_0|\theta I_\circ) \equiv P\big(x \in (x_0, x_0+dx) \wedge y \in (y_0, y_0+dy)|\theta I_\circ\big)$ can be decomposed according to the so-called *product rule*:

$$p(x_0 y_0|\theta I_\circ) = p(x_0|\theta I_\circ)\, p(y_0|x_0\theta I_\circ) = p(y_0|\theta I_\circ)\, p(x_0|y_0\theta I_\circ)\ ,$$

where $p(y_0|x_0\theta I_\circ)$ is the probability for the sampling variate $y$ to take a value in the interval $(y_0, y_0+dy)$, given the family $I_\circ$ of the sampling distribution of $x$ and $y$, the value of the parameter of the family in an interval $(\theta, \theta+d\theta)$, and the sampling variate $x$ being observed in an interval $(x_0, x_0+dx)$. The product rule can also be expressed in terms of the appropriate pdf's

$$f(xy|\theta I_\circ) = f(x|\theta I_\circ)\, f(y|x\theta I_\circ) = f(y|\theta I_\circ)\, f(x|y\theta I_\circ)\ , \tag{1}$$

which for independent $x$ and $y$ reduces to

$$f(xy|\theta I_\circ) = f(x|\theta I_\circ)\, f(y|\theta I_\circ)\ . \tag{2}$$

Second, every sampling probability distribution is subject to the normalization

$$\int\limits_X f(x'|\theta I_\circ)\, dx' = 1\ , \tag{3}$$

where the integration is performed over the entire sample space $X$.

Third, suppose there exists a one-to-one transformation $s$ of a scalar sampling variate $x$, $y \equiv s(x)$. The corresponding pdf's of $x$ and $y$ are then related as

$$f(y|\theta I_\circ') = f(x|\theta I_\circ)\, \big|s'(x)\big|^{-1}\ , \tag{4}$$

where $s'(x) \equiv dy/dx$, while for multidimensional variates $\mathbf{x}$ and $\mathbf{y}$ the derivative must be replaced by the corresponding Jacobian $J \equiv \partial\mathbf{y}/\partial\mathbf{x}$. By using the symbol $I_\circ'$ instead of $I_\circ$ on the left-hand side of (4) we stress that the form (the family) of the sampling distribution may in general be altered by the transformation of a sampling variate.

Besides of the pdf's, distributions of sampling variates can equivalently be presented by the (*cumulative*) *distribution functions* (*cdf*'s). The cdf of a sampling variate $x$ from a continuous sample space $X = (x_a, x_b)$ is defined as

$$F(x, \theta, I_\circ) \equiv \int_{x_a}^{x} f(x'|\theta I_\circ)\, dx'\ , \tag{5}$$

or, inversely,

$$f(x|\theta I_\circ) \equiv F_1(x,\theta,I_\circ)\ , \tag{6}$$

where $F_i$ denotes differentiation with respect to the $i$-th argument of $F$ (we adhere to this notation throughout the present paper, whatever the function and the arguments may be). A cdf coincides with the probability of the sampling variate to take the value less or equal to $x$. Consequently, the cdf is a nondecreasing function of $x$ whose values are limited within

$$F(x_a,\theta,I_\circ) = 0 \quad \text{and} \quad F(x_b,\theta,I_\circ) = 1\ .$$

### 2.2. Location, scale and dispersion parameters

A parameter $\mu$ of a sampling distribution is a location parameter, and a parameter $\sigma$ is a dispersion parameter, if the pdf of $x$ takes the form

$$f(x|\mu\sigma I_\circ) = \frac{1}{\sigma}\,\phi\Big(\frac{x-\mu}{\sigma}\Big)\ , \tag{7}$$

with the ranges of $x$ and $\mu$, $(x_a,x_b)$ and $(\mu_a,\mu_b)$, stretching over the entire real axis, and with the range of $\sigma$, $(\sigma_a,\sigma_b)$, coinciding with $(0,\infty)$.

According to the product rule (2), a bivariate pdf of two independent random variates, $x^{(1)}$ and $x^{(2)}$, both being subject to the same pdf of the form (7), is equal to the product of univariate pdf's, $f(x^{(1)}|\mu\sigma I_\circ)$ and $f(x^{(2)}|\mu\sigma I_\circ)$:

$$f(x^{(1)}x^{(2)}|\mu\sigma I_\circ) = f(x^{(1)}|\mu\sigma I_\circ)\,f(x^{(2)}|\mu\sigma I_\circ) = \frac{1}{\sigma^2}\,\phi\Big(\frac{x^{(1)}-\mu}{\sigma}\Big)\,\phi\Big(\frac{x^{(2)}-\mu}{\sigma}\Big)\ .$$

The pdf of transformed variates $x^{(1)}$ and $x^{(2)}$, $\bar{x}$ and $s$,

$$\bar{x} \equiv \frac{x^{(1)}+x^{(2)}}{2} \quad \text{and} \quad s \equiv \frac{x^{(1)}-x^{(2)}}{2}\ \ ;\ \ \bar{x},s \in (-\infty,\infty)\ ,$$

can be calculated according to (4):

$$\begin{aligned} f(\bar{x}s|\mu\sigma I_\circ') &= f\big(x^{(1)}(\bar{x},s)\,x^{(2)}(\bar{x},s)|\mu\sigma I_\circ\big)\,\big|J\big|^{-1} \\ &= \frac{2}{\sigma^2}\,\phi\Big(\frac{\bar{x}-\mu}{\sigma}+\frac{s}{\sigma}\Big)\,\phi\Big(\frac{\bar{x}-\mu}{\sigma}-\frac{s}{\sigma}\Big) \\ &\equiv \frac{1}{\sigma^2}\,\tilde{\phi}\Big(\frac{\bar{x}-\mu}{\sigma},\frac{s}{\sigma}\Big)\ . \end{aligned} \tag{8}$$

When inferring the parameters of a sampling distribution of the form (7), it may happen that the value of one of the two parameters is known to a high precision. Then, the parameter with the precisely determined value is *fixed* and we only make an inference about the remaining one. Let first the dispersion parameter $\sigma$ be fixed, say to 1, so that the pdf of $x$, given the possible value of $\mu$ and the fixed $\sigma$,

$$f(x|\mu\sigma I_\circ) = \frac{1}{\sigma}\,\phi\Big(\frac{x-\mu}{\sigma}\Big) = \phi(x-\mu)\ , \tag{9}$$

is a function of $x$ and $\mu$ only. The fixed parameter ($\sigma$ in the present case) is usually (though not always) omitted from explicit expressions.

According to (5) and (9), the cdf of $x$ reads:

$$F(x,\mu,\sigma,I_\circ) = \int_{-\infty}^{x} f(x'|\mu\sigma I_\circ)\,dx' = \int_{-\infty}^{x}\phi(x'-\mu)\,dx' = \int_{-\infty}^{x-\mu}\phi(u)\,du = \Phi(x-\mu)\ . \tag{10}$$

Note that the form $\Phi(x-\mu)$ of the above cdf implies the corresponding pdf to be of the form (9), i.e., implies $\mu$ to be a location parameter of a sampling probability distribution of $x$. Indeed:

$$f(x|\mu\sigma I_\circ) = \frac{\partial}{\partial x} F(x,\mu,\sigma,I_\circ) = \frac{\partial}{\partial x}\Phi(x-\mu) = \frac{d}{du}\Phi(u)\,\frac{\partial u}{\partial x} = \frac{d}{du}\Phi(u) = \phi(u)\ ,$$

where $u \equiv x-\mu$. Since the integrand in (10), i.e., the pdf of $x$, is a positive function, and the upper bounds of the integral are strictly decreasing with the increase in the parameter, the cdf is obviously strictly decreasing in $\mu$.

If, on the other hand, $\mu$ is fixed, while $\sigma$ is not, the sample space $X = (-\infty,\infty)$ can be split into

$$X_1 \equiv (\mu,\infty)\ ,\quad X_2 \equiv (-\infty,\mu)\quad\text{and}\quad X_3 \equiv \{\mu\}\ ,$$

and the pdf of $x$, given $\sigma$ and a fixed $\mu$,

$$f(x|\sigma\mu I_\circ) = \frac{1}{\sigma}\,\phi\Big(\frac{x-\mu}{\sigma}\Big)\ , \tag{11}$$

can be split into two pdf's,

$$f^{(1,2)}(x|\sigma\mu I_\circ) \equiv \frac{1}{C^{(1,2)}}\, f(x|\sigma\mu I_\circ)\ ;\quad x\in X_{1,2}\ , \tag{12}$$

where $C^{(1,2)}$ are the appropriate normalization constants

$$C^{(1,2)} = \int\limits_{\mu(-\infty)}^{\infty(\mu)} f(x'|\sigma\mu I_\circ)\,dx' = \int\limits_{0(-\infty)}^{\infty(0)} \phi(u)\,du\ .$$

In this way, the problem of inferring a dispersion parameter can be split into two separate problems. Dispersion parameters of sampling distributions with the sample space being bound either to the positive or to the negative half of the real axis are referred to as the *scale* parameters.

The reasons for discarding $X_3$ in the above definitions are given below in this section, as well as in § 4.4 and § 4.7. Here we would just like to point out that since the probability measure of $X_3$, i.e., of a single point within a continuous interval, is zero (i.e., the probability for observing exactly $x=\mu$ is zero), this can always be done with no possible loss of generality.

Any sampling probability distribution determined by a scale parameter $\sigma$ and a fixed location parameter $\mu$, can be further transformed into a probability distribution determined by a location parameter $\nu$ (see, for example, Ferguson, 1967, § 4.4, p. 144):

$$\nu = \ln\sigma\ ,$$

and a fixed dispersion parameter $\lambda$, say $\lambda = 1$. Namely, substitutions

$$y_{1,2} = \ln\{\pm(x-\mu)\}$$

yield:

$$f^{(1,2)}(y_{1,2}|\nu\lambda_0 I_\circ) = f^{(1,2)}(x|\sigma\mu I_\circ)\left|\frac{dy_{1,2}}{dx}\right|^{-1} \propto e^{y_{1,2}-\nu}\,\phi\big(e^{y_{1,2}-\nu}\big) \equiv \widetilde{\phi}(y_{1,2}-\nu)\ . \tag{13}$$

Note, however, that the corresponding variate $y_3$ for $x=\mu$ cannot be defined.

The cdf's $F^{(1,2)}(x,\mu,\sigma,I_\circ)$, corresponding to $f^{(1,2)}(x|\sigma\mu I_\circ)$, read:

$$F^{(1,2)}(x,\mu,\sigma,I_\circ) = \int\limits_{\mu(-\infty)}^{x} f^{(1,2)}(x'|\sigma\mu I_\circ)\,dx' \propto \int\limits_{\mu(-\infty)}^{x} \frac{1}{\sigma}\,\phi\Big(\frac{x'-\mu}{\sigma}\Big)\,dx' = \int\limits_{0(-\infty)}^{(x-\mu)/\sigma} \phi(u)\,du\ .$$

While $F^{(1)}(x,\mu,\sigma)$ is monotonically decreasing with increasing value of $\sigma$, $F^{(2)}(x,\mu,\sigma)$ is monotonically increasing.

Some of the most important continuous sampling distributions are determined by one or more parameters of the above mentioned types (see Eadie et al., 1971, § 4.2, pp. 58-83). In addition, all distributions determined by either location, dispersion or scale parameters share a very important property: they all belong to the *invariant families of distributions*.

### 2.3. Invariant distributions

Let

$$f(x|\theta I_\circ) = \phi(x,\theta) \tag{14}$$

be a pdf of a random variable $x$ from a continuous sample space $X$ that is determined by the value of parameter $\theta$ from the parameter space $\Theta$. Let there exist *a group* $\mathcal{G}$ of transformations $g_a$ of the sample space into itself:

$$\begin{aligned} g_a:\ X &\longrightarrow X\ , \\ g_a:\ x &\longrightarrow g_a(x) \equiv y\ , \end{aligned} \tag{15}$$

where index $a$ denotes a particular element of the group. Since $\mathcal{G}$ is a group, it is closed under composition of transformations, i.e., a composition $g_c$ of every pair of transformations $g_a,\ g_b \in \mathcal{G}$, $g_c = g_b g_a$, such that

$$g_c(x) = g_b g_a(x) = g_b\big[g_a(x)\big]\ ,$$

is also contained in $\mathcal{G}$. In addition, the group also contains an identity $g_e$ such that

$$g_e(x) = x\ ,\ \forall\, x \in X\ ,$$

and the inverse transformation $g_a^{-1}$ for any $g_a$ such that

$$g_a^{-1} g_a = g_a g_a^{-1} = g_e\ .$$

As a consequence, the transformations $g_a$ are one-to-one, i.e., $g_a(x_1) = g_a(x_2)$ implies $x_1 = x_2$, and *onto* $X$, i.e., *on* (entire) $X$ and *to* (entire) $X$: for every $x_1 \in X$ and $g_a \in \mathcal{G}$ there exists an $x_2 \in X$ such that $g_a(x_2) = x_1$ (see, for example, Ferguson, 1967, § 4.1, p. 143). The sample space $X$ is said to be *invariant under the group* $\mathcal{G}$

Since the transformation $y = g_a(x)$ is one-to-one, the pdf of the transformed variate according to (4) reads:

$$f(y|\theta I_\circ') = f(x|\theta I_\circ)\,\big|g_a'(x)\big|^{-1} = \phi(x,\theta)\big|g_a'(x)\big|^{-1}\ .$$

In addition to $\mathcal{G}$, let there exist also a set $\bar{\mathcal{G}}$ of transformations $\bar{g}_a$ of the parameter space $\Theta$ into itself,

$$\begin{aligned} \bar{g}_a:\ \Theta &\longrightarrow \Theta\ , \\ \bar{g}_a:\ \theta &\longrightarrow \bar{g}_a(\theta) \equiv \nu\ . \end{aligned}$$

Then, the pdf of $y$ can be re-expressed in terms of the parameter $\nu$, instead of $\theta$,

$$f(y|\nu I_\circ'') = \phi(x,\theta)\big|g_a'(x)\big|^{-1} = \phi\big(g_a^{-1}(y),\bar{g}_a^{-1}(\nu)\big)\big|g_a'(x)\big|^{-1} \equiv \widetilde{\phi}\big(g_a(x),\bar{g}_a(\theta)\big) \ .$$

If for all $x \in X,$ and $\theta \in \Theta$, and for every $g_a \in \mathcal{G}$ there exists $\bar{g}_a \in \bar{\mathcal{G}}$ such that

$$\widetilde{\phi}\big(g_a(x),\bar{g}_a(\theta)\big) = \phi\big(g_a(x),\bar{g}_a(\theta)\big) \ , \tag{16}$$

the family of distributions $f(x|\theta I_\circ)$ is said to be *invariant under the group* $\mathcal{G}$ (Ferguson, 1967, § 4.1, p. 144; Stuart et al., 1999, § 23.10, pp. 300-301).

If a family of distributions is invariant under $\mathcal{G}$, then the set $\bar{\mathcal{G}}$ of transformations $\bar{g}_a$ is also a group, usually referred to as *the induced group* (Stuart et al., 1999, § 23.10, p. 300). Namely, according to the definition of invariance, if the pdf of $x$ is given by $\phi(x,\theta)$, the pdf for $g_a(x)$ is given by $\phi\big(g_a(x),\bar{g}_a(\theta)\big)$. Hence, the pdf of $g_b\big(g_a(x)\big) = g_bg_a(x)$ is given by both $\phi\big(g_b(g_a(x)),\bar{g}_b(\bar{g}_a(\theta))\big)$ and $\phi\big(g_bg_a(x),\overline{g_bg_a}(\theta)\big)$. From the equality of the two it follows that

$$\overline{g_bg_a} = \bar{g}_b\bar{g}_a \ .$$

This shows that $\bar{\mathcal{G}}$ is closed under composition. It also shows that $\bar{\mathcal{G}}$ is closed under inverses if we let $g_b = g_a^{-1}$ and note that $\bar{g}_e$ is the identity in $\bar{\mathcal{G}}$.

As for $\mathcal{G}$ and $X$, the parameter space $\Theta$ is *invariant under the induced group* $\bar{\mathcal{G}}$. *The invariance of the sample and the parameter spaces under groups* $\mathcal{G}$ *and* $\bar{\mathcal{G}}$*, respectively, is therefore inherent to every invariance of a sampling probability distribution under a group* $\mathcal{G}$.

For example, a sampling distribution of $\bar{x}$ and $s$ (8), determined by the values of a location parameter $\mu$ and a dispersion parameter $\sigma$, is invariant under the group of simultaneous location and scale transformations:

$$\begin{aligned} g_{a,b}: &\left\{\begin{array}{lll} \bar{x} & \longrightarrow & g_{a,b}(\bar{x}) = a\bar{x}+b \\ s & \longrightarrow & g_{a,b}(s) = as \end{array}\right. , \\ \bar{g}_{a,b}: &\left\{\begin{array}{lll} \mu & \longrightarrow & \bar{g}_{a,b}(\mu) = a\mu+b \\ \sigma & \longrightarrow & \bar{g}_{a,b}(\sigma) = a\sigma \end{array}\right. , \end{aligned} \tag{17}$$

where

$$a \in (0,\infty) \ \text{ and } \ b \in (-\infty,\infty) \ .$$

By fixing the dispersion parameter, we are left with the remaining symmetry of the sampling distribution under simultaneous translations of $x$ and $\mu$ by an arbitrary real number $b$:

$$\begin{aligned} g_b:\ x \quad &\longrightarrow \quad g_b(x) = x+b \ , \\ \bar{g}_b:\ \mu \quad &\longrightarrow \quad \bar{g}_b(\mu) = \mu+b \ . \end{aligned} \tag{18}$$

When, on the other hand, the location parameter is fixed while the dispersion of the distribution is unknown, the appropriate pdf's $f(x|\sigma\mu I_\circ)$ (11) and $f^{(1,2)}(x|\sigma\mu I_\circ)$ (12) are still invariant under the scale transformation:

$$\begin{aligned} g_a:\ x \quad &\longrightarrow \quad g_a(x) = ax+(1-a)\mu \\ \bar{g}_a:\ \sigma \quad &\longrightarrow \quad \bar{g}_a(\sigma) = a\sigma \ . \end{aligned}$$

The above transformation of the sampling variate $x$ is identical to the transformation

$$x-\mu \quad \longrightarrow \quad a(x-\mu)$$

of the difference between the sampling variate $x$ and the fixed location parameter $\mu$.

LEMMA 1. *Let a distribution of a sampling variate $x$, parameterized by $\theta$, be invariant under $\mathcal{G}$, with $\bar{\mathcal{G}}$ being the corresponding induced group. Then, the distribution of $y \equiv s(x)$, parameterized by $\nu \equiv \bar{s}(\theta)$, is invariant under $\mathcal{H} = s\mathcal{G}s^{-1}$, while $\bar{\mathcal{H}} = \bar{s}\bar{\mathcal{G}}\bar{s}^{-1}$ is the appropriate induced group. Here, $s$ and $\bar{s}$ are arbitrary continuous and differentiable one-to-one transformations of $x$ and $\theta$, respectively.*

### 2.4. Invariance under Lie groups

Let $F(x, \theta, I_\circ)$ be the cdf of a sampling variate $x$, subject to the distribution (14). The cdf of $y = s(x)$, given $\nu = \bar{s}(\theta)$, reads:

$$F(y, \nu, I'_\circ) = F\big(s(x), \bar{s}(\theta), I'_\circ\big) = \int_{y_a}^{y} f(y'|\nu I'_\circ)\, dy' = \int_{y_a}^{s(x)} \widetilde{\phi}\big(s(x'), \bar{s}(\theta)\big)\, d[s(x')] \;,$$

where $y_a$ is the lower bound of the range of $y$, while $s$ and $\bar{s}$ are arbitrary continuous and differentiable one-to-one functions. Then,

LEMMA 2. *The lower and the upper bound of the range of $x$, $x_a$ and $x_b$, become transformed into the bounds of $y$, $y_a$ and $y_b$:*

$$y_a = \begin{cases} s(x_a) & ; \quad s'(x) > 0 \\ s(x_b) & ; \quad s'(x) < 0 \end{cases} \qquad \textit{and} \quad y_b = \begin{cases} s(x_b) & ; \quad s'(x) > 0 \\ s(x_a) & ; \quad s'(x) < 0 \end{cases} \qquad ,$$

*and the cdf of $y$, given $\nu$, is related to the cdf of $x$, given $\theta$, as:*

$$F(y, \nu, I'_\circ) = F\big(s(x), \bar{s}(\theta), I'_\circ\big) = \begin{cases} \quad F(x, \theta, I_\circ) & ; \; s'(x) > 0 \\ 1 - F(x, \theta, I_\circ) \, ; & s'(x) < 0 \end{cases} \quad . \tag{19}$$

COROLLARY. *Let a sampling distribution with the cdf $F(x, \theta, I_\circ)$ be invariant under $\mathcal{G}$. Then, the cdf of $g_a(x)$ can be expressed as*

$$F\big(g_a(x), \bar{g}_a(\theta), I_\circ\big) = \begin{cases} \quad F(x, \theta, I_\circ) & ; \; g'_a(x) > 0 \\ 1 - F(x, \theta, I_\circ) \, ; & g'_a(x) < 0 \end{cases} \qquad ; \quad \forall\, g_a \in \mathcal{G} \; . \tag{20}$$

Note that for invariant distributions the information $I'_\circ$ that $F\big(g_a(x), \bar{g}_a(\theta), I'_\circ\big)$ is based upon (the family that the sampling distribution belongs to), is identical to the information $I_\circ$ of $F(x, \theta, I_\circ)$,

$$F\big(g_a(x), \bar{g}_a(\theta), I'_\circ\big) = F\big(g_a(x), \bar{g}_a(\theta), I_\circ\big) \;,$$

so that the parameter $a$ enters $F\big(g_a(x), \bar{g}_a(\theta), I_\circ\big)$ only through $g_a(x)$ and $\bar{g}_a(\theta)$, but *not* through $I_\circ$, i.e., $I_\circ \neq I_\circ(a)$.

Let now $\mathcal{G}$ be a *Lie group* of transformations (15), so that the partial derivative

$$\frac{\partial}{\partial a} g_a(x)$$

exists for every $g_a \in \mathcal{G}$ and every $x \in X$ (see, for example, Elliot and Dawber, 1986, § 7.1-7.2, pp. 126-130). Here, $a$ is a scalar parameter of the group while $X$ is a set of real scalar variates (objects of group transformations). Under these circumstances, we state the following

LEMMA 3. *For all $x \in X$ for which*

$$\frac{\partial}{\partial a} g_a(x)\Big|_{a=e} \tag{21}$$

*vanishes, all group transformations are trivial, i.e.,*

$$g_a(x) = x \quad ; \ \forall\, g_a \in \mathcal{G} \ .$$

Clearly, if (21) vanishes for all $x \in X$, then the action of the group $G$ on the entire $X$ is trivial:

$$g_a(x) = x \quad ; \ \forall\, g_a \in \mathcal{G} \quad \text{and} \quad \forall\, x \in X \ .$$

LEMMA 4. *Let $F(x, \theta, I_\circ)$ be a cdf of a strictly positive continuous sampling distribution that is invariant under a Lie group $\mathcal{G}$ whose action is not identically trivial on the entire sample space $X$. In addition, let $F(x, \theta, I_\circ)$ be differentiable in the second argument (differentiability in the first argument is guaranteed by definition* (6)*). Then, the partial derivative*

$$\frac{\partial}{\partial a} \bar{g}_a(\theta)\Big|_{a=e} \tag{22}$$

*does not vanish for any $\theta \in \Theta$.*

An important statement - an Existence Theorem - can be deduced from the above Lemmata. Let therefore a probability distribution of a scalar sampling variate $x$ be invariant under a Lie group $\mathcal{G}$ (consequently, $\bar{\mathcal{G}}$ is also a Lie group). Then, differentiation of (20) with respect to $a$ yields

$$F_1\big(g_a(x), \bar{g}_a(\theta), I_\circ\big)\, \frac{\partial}{\partial a} g_a(x) + F_2\big(g_a(x), \bar{g}_a(\theta), I_\circ\big)\, \frac{\partial}{\partial a} \bar{g}_a(\theta) = 0 \ .$$

On the subspace $\widetilde{X} \subseteq X$ with non-vanishing derivative (21) (derivative (22) is strictly different from zero by Lemma 4), the above differential equation reduces to

$$F_1(x, \theta, I_\circ)\, \bar{s}'(\theta) + F_2(x, \theta, I_\circ)\, s'(x) = 0 \ , \tag{23}$$

where the derivatives $s'(x)$ and $\bar{s}'(\theta)$ of functions $s(x)$ and $\bar{s}(\theta)$ are defined as:

$$s'(x) \equiv \frac{ds(x)}{dx} \equiv \left[\frac{\partial}{\partial a} g_a(x)\Big|_{a=e}\right]^{-1} \quad \text{and} \quad \bar{s}'(\theta) \equiv \frac{d\bar{s}(\theta)}{d\theta} \equiv \left[\frac{\partial}{\partial a} \bar{g}_a(\theta)\Big|_{a=e}\right]^{-1} \ . \tag{24}$$

By defining a function $G(x, \theta)$,

$$G(x, \theta) \equiv s(x) - \bar{s}(\theta) \ , \tag{25}$$

(23) can be further rewritten as

$$F_1(x, \theta, I_\circ)\, G_2(x, \theta) - F_2(x, \theta, I_\circ)\, G_1(x, \theta) = 0 \ ,$$

or as a functional determinant (see Aczél, 1966, § 7.2.1, p. 325),

$$\begin{vmatrix} F_1(x, \theta, I_\circ) & F_2(x, \theta, I_\circ) \\ G_1(x, \theta) & G_2(x, \theta) \end{vmatrix} = 0 \ .$$

The Jacobian vanishes for all $x \in X$ and $\theta \in \Theta$ if and only if the cdf $F(x, \theta, I_\circ)$ is a function of a single variable $G(x, \theta)$ (25) (see, for example, Courant, 1962, § 1, p. 5),

$$F(x, \theta, I_\circ) = \Phi[G(x, \theta)] = \Phi[s(x) - \bar{s}(\theta)] = \Phi(y - \mu) \ , \tag{26}$$

where we introduced

$$y \equiv s(x) \text{ and } \mu \equiv \bar{s}(\theta) .$$

Then, by equation (19), the cdf $F(y, \mu, I'_\circ)$ is of the form

$$F(y, \mu, I'_\circ) = \begin{cases} \Phi(y-\mu) \,; s'(x) > 0 \\ \widetilde{\Phi}(y-\mu) \,; s'(x) < 0 \end{cases} ,$$

where

$$\widetilde{\Phi}(y-\mu) \equiv 1 - \Phi(y-\mu) .$$

That is, $\mu$ is a location parameter of the sampling distribution of $y$ (see eq. (10)), and the above reasoning can be summarized as

THEOREM 1. *Let $f(x|\theta I_\circ)$ be a pdf of a continuous scalar sampling variate $x \in X$ and $\theta \in \Theta$ a continuous scalar parameter of the distribution that is invariant under a one-parameter Lie group $\mathcal{G}$, and let the cdf $F(x, \theta, I_\circ)$ be differentiable in $\theta$. Then, on the subspace $\widetilde{X} \subseteq X$ where the derivative* (21) *does not vanish, the distribution is necessarily reducible (by separate one-to-one transformations $x \to z$ and $\theta \to \mu$) to a sampling distribution of $z$ with the parameter $\mu$ being a location parameter.*

In the sequel (Theorem 4) we shall further demonstrate that the subspace $X - \widetilde{X} \subseteq X$ of the sample space with vanishing derivative (21), is irrelevant to probabilistic parametric inference, since for the observed $x$ from $X - \widetilde{X}$ a pdf cannot be assigned to the inferred parameter of the sampling distribution.

The following Theorem also proves to be relevant to our derivations:

THEOREM 2. *If a sampling distribution* (9) *of $x$, determined by a location parameter $\mu$ and a fixed dispersion parameter $\sigma$, is invariant under a Lie group $\mathcal{G}$, then $\mathcal{G}$ is the group of translations.*

## 3. Inverse probabilities

### 3.1. Plausibilities and inverse probabilities

Let now information about a random variate $x$ consist only of a family $I_\circ$ of possible distributions of $x$, while the true value of the parameter $\theta$ that uniquely determines the sampling distribution, is unknown. Then, an inference about the true distribution of $x$ is equivalent to an inference about the parameter $\theta$ of the family $I_\circ$.

An inference about the parameter is made by specifying a real number, called (*degree of*) *plausibility*, $(\theta|x_1 x_2 \ldots I_\circ)$, representing our degree of belief that, given a set of observations $x \in (x_1, x_1 + dx)$, $x \in (x_2, x_2 + dx)$, $\ldots$ of the sampling variate, an interval $(\theta, \theta + d\theta)$ covers the true value of the parameter. Cox (1946) showed that a system for manipulating plausibilities is either isomorphic to the probability system or inconsistent with very general requirements, referred to as the Cox-Pólya-Jaynes Desiderata (see, for example, Jaynes, 2003, § 1.7, pp. 17-19, or Van Horn, 2003). Motivated by the Cox's Theorem, we therefore once and for all choose probabilities $p(\theta|x_1 x_2 \ldots I_\circ)$ among all possible plausibility functions $(\theta|x_1 x_2 \ldots I_\circ)$ to represent our degree of belief in particular values of inferred parameters:

ASSUMPTION 2. *The so-called* inverse probabilities, *$p(\theta|x_1 x_2 \ldots I_\circ)$, and the so-called* direct *(or sampling)* probabilities *$p(x|\theta I_\circ)$, are subjected to identical rules, i.e., to the aforementioned axioms of conditional probability.*

These, expressed in terms of non-negative functions $f(\theta|x_1x_2\ldots I_\circ)$, called *the probability density functions* (*pdf's*) *for the inferred parameters*, include the product rule,

$$\begin{aligned} f(\theta\lambda|x_1x_2\ldots I_\circ) &= f(\theta|x_1x_2\ldots I_\circ)\, f(\lambda|\theta x_1x_2\ldots I_\circ) \\ &= f(\lambda|x_1x_2\ldots I_\circ)\, f(\theta|\lambda x_1x_2\ldots I_\circ)\ , \end{aligned} \tag{27}$$

the requirement of normalization,

$$\int\limits_{\Theta} f(\theta'|x_1x_2\ldots I_\circ)\, d\theta' = 1\ , \tag{28}$$

and the rule for transformations of the pdf's, induced by one-to-one transformations of their arguments,

$$f(\nu|xI'_\circ) = f(\theta|xI_\circ)\, \left|\bar{s}'(\theta)\right|^{-1}\ , \tag{29}$$

where $\nu \equiv \bar{s}(\theta)$ and $\bar{s}'(\theta) \equiv d\nu/d\theta$, while $f(\theta\lambda|x_1x_2\ldots I_\circ)$ is a a joint pdf for the parameters $\theta$ and $\lambda$ of a family $I_\circ$ of two-parametric sampling distributions.

Let a family $I_\circ$ contain the (unknown) true distribution of a sampling variate $x$ whose first value was observed in an interval $(x_1, x_1 + dx)$, and let $p(\theta_1x_2|x_1I_\circ)$ be the joint probability that an interval $(\theta_1, \theta_1 + d\theta)$ covers the true value of the parameter $\theta$ of the family $I_\circ$ and that the second observation of the random variate $x$, independent of the first one, will be recorded in an interval $(x_2, x_2 + dx)$. Then, due to the imposed equivalence of rules for manipulating probabilities of sampling variates and those for manipulating probabilities for inferred parameters, $p(\theta_1x_2|x_1I_\circ)$ can be decomposed as

$$p(\theta_1x_2|x_1I_\circ) = p(\theta_1|x_1I_\circ)\, p(x_2|\theta_1I_\circ) = p(x_2|x_1I_\circ)\, p(\theta_1|x_1x_2I_\circ)\ ,$$

so that the product rule for $f(\theta x_2|x_1I_\circ)$ reads

$$f(\theta x_2|x_1I_\circ) = f(\theta|x_1I_\circ)\, f(x_2|\theta I_\circ) = f(x_2|x_1I_\circ)\, f(\theta|x_1x_2I_\circ)\ . \tag{30}$$

We add this product rule to the axioms, listed within the Assumption 2.

ASSUMPTION 3. *In the same way as we did for the sampling distributions (see Assumption 1), we impose continuity in $\theta$ also to the pdf's $f(\theta|x_1x_2\ldots I_\circ)$ for the inferred parameters.*

As for the cdf's $F(x,\theta,I_\circ)$ of sampling variates, we can also define the cdf's $H(x,\theta,I_\circ)$ for inferred parameters,

$$H(x,\theta,I_\circ) \equiv \int_{\theta_a}^{\theta} f(\theta'|xI_\circ)\, d\theta'\ ,$$

or, equivalently,

$$f(\theta|xI_\circ) \equiv H_2(x,\theta,I_\circ)\ . \tag{31}$$

By subjecting degrees of plausibility to the axioms of probability, the domain of the probability calculus, originally being consisted only of sampling variates, has been extended. In other words, the sampling variates represent only a subset of all possible arguments of probability functions $p$ and pdf's $f$. It is important, however, to distinguish between the concept of probability distribution of a sampling variate, and the concept of probability distribution for an inferred parameter. By definition, the probability for a sampling variate $x$ to take a value in an interval $(x_1, x_1 + dx)$ coincides with the long run relative frequency of occurrence of $x$ in that interval. The probability distribution for a parameter, on the other hand, only represents a distribution of our belief in different values of the inferred parameter within the parameter space. In a vast majority of situations, the inferred parameter is assumed to be fixed (though unknown), so that in general $p(\theta|x_1x_2\ldots I_\circ)$ does *not* coincide with the frequency distribution of the true values of $\theta$. The conceptual difference becomes of high practical importance when the interpretation of verifiable predictions, based on the probability distributions for $\theta$, is concerned. We will come back to this important point in Section 5.

### *3.2. The procedure of marginalization and the Bayes' Theorem*

We can make use of the product rules (27) and (30) for deducing the procedure of marginalization and the Bayes' Theorem. Let us therefore integrate the pdf $f(\theta\lambda|x_1x_2\ldots I_\circ)$ (27) over either of the spaces $\Theta$ and $\Lambda$ of the two inferred parameters $\theta$ and $\lambda$, respectively. With the pdf's $f(\lambda|\theta x_1x_2\ldots I_\circ)$ and $f(\theta|\lambda x_1x_2\ldots I_\circ)$ being properly normalized, we obtain

$$\boxed{\begin{aligned}\int\limits_{\Theta} f(\theta'\lambda|x_1x_2\ldots I_\circ)\,d\theta' &= f(\lambda|x_1x_2\ldots I_\circ)\;,\\ \int\limits_{\Lambda} f(\theta\lambda'|x_1x_2\ldots I_\circ)\,d\lambda' &= f(\theta|x_1x_2\ldots I_\circ)\;,\end{aligned}} \tag{32}$$

where $f(\lambda|x_1x_2\ldots I_\circ)$ and $f(\theta|x_1x_2\ldots I_\circ)$ are the so-called *marginal pdf's*.

The Bayes' Theorem (Bayes, 1763; Laplace, 1774), on the other hand, is obtained simply by rearranging the product rule (30):

$$\boxed{f(\theta|x_1x_2I_\circ) = \frac{f(\theta|x_1I_\circ)\,f(x_2|\theta I_\circ)}{f(x_2|x_1I_\circ)}}\;. \tag{33}$$

The Theorem is also referred to as *the principle of inverse probability* (see Jeffreys, 1961, § 1.22, p. 28) and is interpreted in the following way (O'Hagan, 1994, § 1.3, p. 2). We are interested in the probability distribution for $\theta$ and begin with the initial or *prior* pdf $f(\theta|x_1I_\circ)$, representing the distribution of our belief in different values of $\theta$ prior to taking evidence $x_2$ into account, while the *posterior* pdf $f(\theta|x_1x_2I_\circ)$ represents the distribution of our belief posterior to adding evidence $x_2$ to our previous information about $\theta$. According to Bayes' Theorem, the only consistent way for updating the probability distribution, assigned to the inferred parameter, is by multiplying the prior pdf by the so-called *likelihood density* $f(x_2|\theta I_\circ)$, corresponding to the probability for observing $x \in (x_2, x_2 + dx)$, given the true value of the parameter in the interval $(\theta, \theta + d\theta)$.

The denominator $f(x_2|x_2I_\circ)$ is obtained *via* the normalization requirement (28),

$$f(x_2|x_1I_\circ) = \int\limits_{\Theta} f(\theta'|x_1I_\circ)\,f(x_2|\theta'I_\circ)\,d\theta' \equiv \zeta_\theta(x_1, x_2)\;,$$

and is independent of the value of the inferred parameter.

### *3.3. The difference between an assignment and an update of a probability distribution*

While being a unique tool for sequential *updating* of a pdf $f(\theta|x_1I_\circ)$, assigned to an inferred parameter $\theta$ prior to its updating, Bayes' Theorem says nothing about *assigning* the pdf $f(\theta|x_1I_\circ)$ that is later to be updated. Consequently, the existing system of the adopted rules (27-30) for manipulating the pdf's for inferred parameters, together with the applications (32) and (33) of these rules, need be amended in order to allow for assignments of probability distributions to the parameters, with such assignments representing natural and indispensable starting points in every sequential updating of probability distributions.

The set of rules for assigning probability distributions to inferred parameters is based on the two *fundamental Principles of scientific reasoning* (Popper, 1959, § 24, pp. 91-92):

***I (Principle of Consistency)*** *The theory of inference about the parameters of sampling distributions must be internally consistent. In particular, if within the rules of the theory, a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result. Similarly, identical states of knowledge about a problem that is solvable within the theory, must always lead to identical solutions of the problem.*

***II (Operational Principle)*** *The theory must specify operations that ensure falsifiability of its predictions.*

In what follows, the entire system for assigning inverse probabilities is deduced exclusively by observing these two rules.

### 3.4. Consistency Theorem

Suppose that before we made the first observation of a sampling variate $x$, we had been completely ignorant about the value of the parameter $\theta$ that determines the distribution of $x$: we had only known the family $I_\circ$ of sampling distributions that the distribution of $x$ belongs to. In this context we can prove the following proposition, henceforth referred to as the Consistency Theorem:

THEOREM 3. *Suppose that a sampling variate $x$ from a strictly positive continuous distribution has been observed in an (infinitesimal) interval $(x, x + dx)$, and that, based on the observation and on information $I_\circ$ about the form of the sampling distribution, a pdf $f(\theta|xI_\circ)$ can be assigned to the parameter $\theta$ of the distribution. When positive the pdf must be, in order to meet the Consistency Principle I, directly proportional to the likelihood density $f(x|\theta I_\circ)$,*

$$\boxed{f(\theta|xI_\circ) = \frac{\pi(\theta)}{\eta_\theta(x)}\, f(x|\theta I_\circ)}\,, \tag{34}$$

*where $\pi(\theta)$ is the so-called* consistency factor, *while $\eta_\theta(x)$ is* the normalization factor *that is determined by invoking normalization* (28) *of the pdf, assigned to $\theta$:*

$$\eta_\theta(x) = \int\limits_{\Theta} \pi(\theta')\, f(x|\theta' I_\circ)\, d\theta' \ . \tag{35}$$

*For those $\theta$ for which $f(\theta|xI_\circ)$ vanishes, however, the pdf must be zero regardless the recorded value $x$ of the sampling variate.*

Since both $f(x|\theta I_\circ)$ and $f(\theta|xI_\circ)$ are assumed to be continuous in $\theta$ (recall Assumptions 1 and 3), *$\pi(\theta)$ is also continuous.* In addition we note that *the consistency factor can only be determined up to an arbitrary constant factor, say $k$.* That is, multiplying $\pi(\theta)$ by $k$ clearly implies multiplication of $\eta_\theta(x)$ by the same factor, which then cancels out in the ratio on the right-hand side of (34). The factors $\pi(\theta)$ may therefore be either strictly positive or strictly negative, but must *not* switch sign within $\Theta$, since the latter would imply negative values for pdf's $f(\theta|xI_\circ)$ whose values are non-negative by definition.

For discrete sampling variates, the Consistency Theorem is obtained simply by replacing the likelihood density $f(x|\theta I_\circ)$, both in (34) and in (35), with the corresponding likelihood $p(x|\theta I_\circ)$. Note, however, that the consistency factors cannot be uniquely determined in such problems (see § 4.3 and § 5.1 below), so it is impossible to make consistent and calibrated probabilistic inferences about the parameters of discrete sampling distributions.

The form of the Consistency Theorem (34) is remarkably similar to that of Bayes' Theorem (33): in both Theorems, within a specified model $I_\circ$, the complete information about the inferred parameter $\theta$ of the model that can be extracted from a measurement $x$, is contained in the value of the appropriate likelihood density, $f(x|\theta I_\circ)$. But there is also a fundamental and *very important difference* between the two Theorems: while $f(\theta|x_1 I_\circ)$ in Bayes' Theorem represents the pdf for $\theta$ prior to including observation $x \in (x_2, x_2 + dx)$ in our inference about $\theta$, the consistency factor $\pi(\theta)$ in the Consistency Theorem is just a proportionality coefficient between the pdf for $\theta$ and the appropriate likelihood density.

In the sequel we show how and under what conditions the basic Principles of scientific reasoning uniquely determine a consistency factor $\pi(\theta)$. *The form of the latter depends on the only relevant information that we possess before the first datum $x \in (x_1, x_1 + dx)$ is collected: it depends only on the specified family $I_\circ$ of possible sampling distributions of $x$.* Therefore, in this particular concept, the Principle of Consistency reads:

> *Inferences about the parameters of sampling distributions whose forms, sample spaces and parameter spaces are identical, must be made by using the consistency factors of the forms that are identical up to multiplication constants.*

Note that the above formulation of the Principle of Consistency coincides with the *Principle of Relative Invariance*, stated by Hartigan (1964).

## 4. Determination of the consistency factors

### 4.1. Consistency factors under transformations of the inferred parameters

Let $f(x|\theta I_\circ)$ of the form (14) be a sampling pdf of $x$ whose parameter $\theta$ we would like to infer by specifying the pdf $f(\theta|xI_\circ)$. We saw in the foregoing section that when this can be done in a consistent way, the pdf for $\theta$ must take the form (34). Let $s$ be a one-to-one transformation of the sampling variate $x$, $y = s(x)$, so that the pdf (34) for $\theta$ can be expressed as

$$f\big(\theta|s(x)I'_\circ\big) = \frac{|s'(x)|}{\eta_\theta(x)}\,\pi(\theta)\,\phi(x,\theta)\,\big|s'(x)\big|^{-1} = \frac{|s'(x)|}{\eta_\theta(x)}\,\pi(\theta)\,f(y|\theta I'_\circ)\;,$$

where

$$f\big(y|\theta I'_\circ\big) \equiv \phi(x,\theta)\,\big|s'(x)\big|^{-1}\;.$$

Let there also exist a one-to-one transformation $\bar{s}$ of the parameter $\theta$, $\nu = \bar{s}(\theta)$. According to (29), the pdf's for $\nu$ and $\theta$ are related as

$$f(\nu|xI''_\circ) = f(\theta|xI_\circ)\,\big|\bar{s}'(\theta)\big|^{-1}\;,$$

so that

$$
\begin{aligned}
f\big(\bar{s}(\theta)|x(y)I_\circ'''\big) &= \frac{\big|s'(x)\big|}{\eta_\theta(x)}\,\frac{\pi(\theta)}{\big|\bar{s}'(\theta)\big|}\,\phi(x,\theta)\,\big|s'(x)\big|^{-1} \\
&= \frac{\big|s'(x)\big|}{\eta_\theta(x)}\,\frac{\pi(\theta)}{\big|\bar{s}'(\theta)\big|}\,\widetilde{\phi}\big(s(x),\bar{s}(\theta)\big) \\
&= \frac{\widetilde{\pi}[\bar{s}(\theta)]}{\widetilde{\eta}_\nu[s(x)]}\,\widetilde{\phi}\big(s(x),\bar{s}(\theta)\big)\ ,
\end{aligned}
\tag{36}
$$

where:

$$
\begin{gathered}
f(y|\nu I_\circ''') = \widetilde{\phi}\big(s(x),\bar{s}(\theta)\big) \equiv \phi(x,\theta)\,\big|s'(x)\big|^{-1}\ , \\
\boxed{\widetilde{\pi}[\bar{s}(\theta)] \equiv \tilde{k}\,\pi(\theta)\,\big|\bar{s}'(\theta)\big|^{-1}}
\end{gathered}
\tag{37}
$$

and, for $\bar{s}'(\theta) > 0$,

$$
\widetilde{\eta}_\nu[s(x)] \equiv \tilde{k}\,\frac{\eta_\theta(x)}{\big|s'(x)\big|} = \int_{\bar{s}(\theta_a)}^{\bar{s}(\theta_b)} \widetilde{\pi}[\bar{s}(\theta')]\,f\big(s(x)|\bar{s}(\theta')I_\circ'''\big)\,d[\bar{s}(\theta')]\ , \tag{38}
$$

while for $\bar{s}'(\theta) < 0$ the limits of the above integral are to be interchanged. When dealing with multidimensional parameters, the derivative $\big|\bar{s}'(\theta)\big|$ in (37) must be substituted by the appropriate Jacobian. Thus, the transformations of consistency factors, induced by transformations of parameters, are very similar to those of pdf's (4) and (29).

By using two different symbols, $\pi$ and $\widetilde{\pi}$, it is stressed that the consistency factors for $\theta$ and for the transformed parameter $\bar{s}(\theta)$ may, in general, be different functions. However, for $s(x) = g_a(x)$ with $g_a$ being an element of a group $\mathcal{G}$ of transformations, for the sampling distribution being invariant under $\mathcal{G}$, and for $\bar{s}(\theta) = \bar{g}_a(\theta)$ with $\bar{g}_a$ being an element of the corresponding induced group $\bar{\mathcal{G}}$, the form of the consistency factor must also be invariant under $\bar{\mathcal{G}}$: according to the Consistency Principle, $\pi$ and $\widetilde{\pi}$ must be the same functions up to an arbitrary multiplication factor, say $k(a)$:

$$
\widetilde{\pi}[\bar{g}_a(\theta)] = \frac{\pi[\bar{g}_a(\theta)]}{k(a)}\ . \tag{39}
$$

Note that in general the value of the multiplication constant $k$, up to which the above invariant consistency factor is uniquely determined, may depend on the value of the transformation parameter $a$. When combined with (37), (39) implies:

$$
\boxed{\pi[\bar{g}_a(\theta)] = k(a)\,\pi(\theta)\,\big|\bar{g}_a'(\theta)\big|^{-1}}\ , \tag{40}
$$

with $\tilde{k}$ being contained in $k(a)$. The above functional equation for $\pi(\theta)$ is the cornerstone of the entire theory of consistent *assignment* of probabilities to parameters of sampling distributions.

When a sampling distribution, determined by a two-dimensional parameter $\boldsymbol{\theta} = \big(\theta^{(1)},\theta^{(2)}\big)$, is invariant under a two-parametric group $\mathcal{G}$ of transformations $g_{a,b}$, the corresponding functional equation for the consistency factor $\pi(\theta^{(1)},\theta^{(2)})$ reads:

$$
\pi[\bar{g}_{a,b}(\theta^{(1)}),\bar{g}_{a,b}(\theta^{(2)})] = k(a,b)\,\pi(\theta^{(1)},\theta^{(2)})\,|J|^{-1}\ , \tag{41}
$$

where

$$
J \equiv \frac{\partial\big(\bar{g}_{a,b}(\theta^{(1)}),\bar{g}_{a,b}(\theta^{(2)})\big)}{\partial\big(\theta^{(1)},\theta^{(2)}\big)}\ .
$$

We will come across the above functional equation in § 4.6, during a simultaneous inference about a location and a dispersion parameter.

When equation (39) holds, the normalization factor $\widetilde{\eta}_\nu[g_a(x)]$ is equal (up to the usual factor $k(a)$) to $\eta_\theta[g_a(x)]$ (index $a$ in $g_a$ and $\bar{g}_a$ denotes particular elements of transformation groups, while in $x_a$ and $\theta_a$ it indicates the lower bounds of the sample and the parameter space, respectively),

$$\widetilde{\eta}_\nu[g_a(x)] = \int_{\bar{g}_a(\theta_a)}^{\bar{g}_a(\theta_b)} \frac{1}{k(a)}\, \pi[\bar{g}_a(\theta')]\, \phi\big(g_a(x), \bar{g}_a(\theta')\big)\, d[\bar{g}_a(\theta')] = \frac{\eta_\theta[g_a(x)]}{k(a)}\ . \tag{42}$$

Consequently, the pdf for the parameter $\theta$ of a sampling distribution that is invariant under $\mathcal{G}$, is invariant under $\bar{\mathcal{G}}$. For if $\psi(x,\theta)$ denotes the pdf (68) for $\theta$, given $x$, and if $\widetilde{\psi}\big(g_a(x), \bar{g}_a(\theta)\big)$ denotes the pdf (36) for $\bar{g}_a(\theta)$, given $g_a(x)$, then the invariance of the sampling distribution, combined with the equations (39) and (42), implies

$$\widetilde{\psi}\big(g_a(x), \bar{g}_a(\theta)\big) = \psi\big(g_a(x), \bar{g}_a(\theta)\big)\ ,$$

the latter coinciding with the definition (16) of invariant distributions. Then, according to Lemma 2 and equation (20), the cdf for $\theta$, $H(x,\theta,I_\circ)$, is also invariant under $\bar{\mathcal{G}}$, so that

$$H\big(g_a(x), \bar{g}_a(\theta), I_\circ\big) = \begin{cases} \quad H(x,\theta,I_\circ) & ;\ \bar{g}'_a(\theta) > 0 \\ 1 - H(x,\theta,I_\circ) & ;\ \bar{g}'_a(\theta) < 0 \end{cases}\ . \tag{43}$$

### 4.2. On consistency of the adopted rules

We obtained equation (37) as a direct consequence of the rule (29) for transformations of pdf's $f(\theta|xI_\circ)$, induced by one-to-one transformations of inferred parameters, while (40) was deduced by applying the Consistency Principle. As a test of consistency of the two rules, we shall verify the compatibility of the two equations.

In the same way as we obtained the functional equation (40) for $\pi(\theta)$, we arrive also at the corresponding equation for the consistency factor $\widetilde{\pi}(\nu)$ for $\nu \equiv \bar{s}(\theta)$,

$$\widetilde{\pi}[\bar{h}_a(\nu)] = l(a)\, \widetilde{\pi}(\nu)\, \left|\bar{h}'_a(\nu)\right|^{-1}\ , \tag{44}$$

where, due to Lemma 1, $\bar{h}_a = \bar{s}\bar{g}_a\bar{s}^{-1}$, so that $\bar{h}_a(\nu) = \bar{s}[\bar{g}_a(\theta)]$, $\bar{h}'_a(\nu) = \bar{s}'[\bar{g}_a(\theta)]\, \bar{g}'_a(\theta)\, [\bar{s}'(\nu)]^{-1}$ and $\widetilde{\pi}[\bar{h}_a(\nu)] = \widetilde{\pi}\{\bar{s}[g_a(\theta)]\}$. The latter can then be rewritten by invoking (37) and (40),

$$\widetilde{\pi}[\bar{h}_a(\nu)] = \tilde{k}\, k(a)\, \pi(\theta)\, \left|g'_a(\theta)\right|^{-1} \left|s'[g_a(\theta)]\right|^{-1}\ ,$$

which, when inserted to (44), implies $k(a) = l(a)$. That is, equations (37) and (40) are perfectly compatible if in equation (40) the same proportionality factor $k(a)$ is used for all parameters that are related *via* one-to-one transformations.

### 4.3. Invariance under a discrete group of transformations

Under what circumstances does a unique solution of the functional equation (40) exist? Let us consider a problem with a sampling distribution being invariant under a discrete group of transformations

$$g_a :\ x \quad \longrightarrow \quad g_a(x) = ax\ ,$$

$$\bar{g}_a :\ \theta \quad \longrightarrow \quad \bar{g}_a(\theta) = a\theta\ ,$$

where $a$ can only take two values,

$$a = \{1, -1\}$$

for both groups, $\mathcal{G}$ and $\bar{\mathcal{G}}$. That is, the considered distribution possesses parity under simultaneous inversion of the sample space and the parameter space coordinates. By combining functional equations

$$\pi[\bar{g}_a(\theta)] = k(a)\,\pi(\theta) \quad \text{and} \quad \pi[\bar{g}_a^2(\theta)] = k(a)\,\pi[\bar{g}_a(\theta)]\ ,$$

we obtain for $a = -1$

$$\pi(-\theta) = k(a = -1)\,\pi(\theta) \quad \text{and} \quad \pi(\theta) = k^2(a = -1)\,\pi(\theta)\ ,$$

so that

$$k^2(a = -1) = 1\ .$$

This, when inability of $\pi$ to switch sign is invoked (see § 3.4), further implies

$$\pi(-\theta) = \pi(\theta)\ . \tag{45}$$

That is, the consistency factor that corresponds to a sampling distribution being invariant under simultaneous inversions of sampling and parameter space coordinates, must itself have *positive parity* under the inversion of the parameter space coordinates. But apart from this, it can take any form and so in this case the solution of (40) is clearly *not* unique.

It is not difficult to understand that this is a common feature of all solutions based on invariance of the sampling distributions under *discrete* groups. If the symmetry group is discrete, the sample and the parameter spaces break up in intervals, the so-called *fundamental regions* or *domains* of the group (Wigner, 1959, § 19.1, p. 210; Jaynes, 2003, § 10.9, p. 332), with no connections in terms of group transformations within the points of the same interval. We are then free to choose the form of $\pi(\theta)$ in one of these intervals (e.g., we can choose $\pi(\theta)$ for the positive values of $\theta$ in the above example), so it is evident that *it is impossible to determine uniquely the form of consistency factors for problems that are invariant only under discrete groups of transformations*. This is also why it is impossible to make consistent probabilistic inferences about the parameters of discrete sampling distributions.

### *4.4. Consistency factors and homogenous parameter spaces*

If, on the other hand, for every $\theta_1$ and $\theta_2$ from a parameter space $\Theta$ there exists an element $\bar{g}_a$ from a group $\bar{\mathcal{G}}$ of transformations such that $\theta_2 = \bar{g}_a(\theta_1)$ (i.e., if all points of $\Theta$ are connected *via* transformations $\bar{g}_a$), then the fundamental domain of the parameter space reduces to a single point and we say that $\Theta$ is *a homogenous space for the group* $\bar{\mathcal{G}}$, or equivalently, that the entire $\Theta$ is *a single* $\bar{\mathcal{G}}$*-orbit*. In what follows we show that homogeneity of parameter spaces for Lie groups plays a decisive role in determination of consistency factors by using symmetry arguments.

According to the Existence Theorem 1 of § 2.4, on the subspace $\widetilde{X} \subseteq X$ with non-vanishing derivative (21), every sampling distribution of a continuous sampling variate $x$ that is invariant under a single-parametric Lie group $\mathcal{G}$, is necessarily reducible (by separate transformations $x \to y$ and $\theta \to \mu$) to a sampling distribution of $z$ with the parameter $\mu$ being a location parameter. For the subspace $\widetilde{X} \subseteq X$ it is therefore sufficient to determine the consistency factor $\widetilde{\pi}(\mu)$ for $\mu$, which can subsequently be transformed (by means of (37)) to the corresponding consistency factor $\pi(\theta)$ for the original parameter $\theta$. Note, however, that implications of Theorem 1 may be extended to the subspaces $X - \widetilde{X} \subseteq X$ with vanishing derivative (21):

THEOREM 4. *Consider a continuous scalar sampling variate $x$ with the cdf $F(x,\theta,I_\circ)$ that is differentiable in the second argument, and with the corresponding pdf $f(x|\theta I_\circ)$ that is strictly positive and invariant under a Lie group $\mathcal{G}$ whose action is not identically trivial on entire $X$. Then, for the observed $x \in X - \widetilde{X} \subseteq X$ with vanishing derivative* (21)*, the probability distribution whose cdf $H(x,\theta,I_\circ)$ is differentiable in the first argument, cannot be assigned to $\theta$. (Existence of derivatives $F_1(x,\theta,I_\circ)$ and $H_2(x,\theta,I_\circ)$ is assured by definitions* (6) *and* (31)*, respectively.)*

Reducibility of the sampling distribution of $x$ to a distribution that is determined by a location parameter $\mu$, is therefore a necessary condition that is to be met in order to solve (40) exclusively by using symmetry arguments. The only Lie group of invariant transformations of sampling distributions, determined by the location parameters, is the one of translations (Theorem 2) and the space of the location parameters, consisting of the entire real axis, is homogenous for that group. Below we shall demonstrate that the axioms of probability, imposed to inverse probabilities (Assumption 2), together with the Principle of Consistency, uniquely determine the consistency factors for location parameters. In this way, the reducibility of a problem of parametric inference to the inference about a location parameter will be proved also a sufficient condition for a consistent probabilistic parametric inference.

### 4.5. Inference about location parameters

We saw in § 2.3 that a sampling distribution, parameterized by a location parameter $\mu$ and by a fixed dispersion parameter $\sigma$, is invariant under the group of translations (18). Then, the functional equation (40) for the appropriate consistency factor $\pi(\mu|\sigma)$ for $\mu$ reads:

$$\pi(\mu+b|\sigma) = k(b)\,\pi(\mu|\sigma) \quad ; \quad \forall\, \mu, b \in \mathbb{R}\,, \tag{46}$$

with the notation $\pi(\mu|\sigma)$ stressing that the dispersion parameter is being fixed.

LEMMA 5. *By setting $\pi(0) = 1$,*

$$\pi(\mu|\sigma) = \exp\{-q(\sigma)\,\mu\} \tag{47}$$

*becomes the most general solution of* (46)*.*

Note that at this point the value of the constant $q$ in (47) may, at least in principle, depend on the value of the fixed parameter $\sigma$.

For sampling distributions, symmetric under simultaneous inversions of the sampling and the parameter space, equation (45) implies $q = 0$, i.e., implies uniform consistency factors for location parameters. This argument was used by Hartigan (1964) to determine a unique form of the so-called non-informative prior distributions in problems of inference about the location parameter of a Gaussian distribution with $\sigma$ being fixed. However, in § 4.6 we demonstrate that $q$ must vanish also in problems without the space-inversion symmetry.

Based on a measured value $x_1$, the pdf for a location parameter $\mu$ therefore reads:

$$f(\mu|x_1\sigma I_\circ) = \frac{\pi(\mu|\sigma)}{\eta_\mu(x_1,\sigma)}\, f(x_1|\mu\sigma I_\circ) = \frac{e^{-q(\sigma)\,\mu}}{\eta_\mu(x_1,\sigma)}\, \frac{1}{\sigma}\phi\Big(\frac{x_1-\mu}{\sigma}\Big)\,.$$

Now, as an example, we want to update our inference about the parameter $\mu$ by including additional information $x_2$ in our inference, where $x_2$ is a result of a measurement of $x$ that is also subject to the same sampling distribution and independent of $x_1$. We can write the likelihood density of $x_2$,

$$f(x_2|\mu\sigma x_1 I_\circ) = f(x_2|\mu\sigma I_\circ) = \frac{1}{\sigma}\phi\Big(\frac{x_2-\mu}{\sigma}\Big)\,,$$

and the updated pdf for $\mu$,

$$\begin{aligned} f(\mu|x_1x_2\sigma I_\circ) &\propto f(\mu|x_1\sigma I_\circ)\, f(x_2|\mu\sigma I_\circ) \\ &\propto \pi(\mu|\sigma)\, f(x_1|\mu\sigma I_\circ)\, f(x_2|\mu\sigma I_\circ) \\ &= \pi(\mu|\sigma)\, f(x_1x_2|\mu\sigma I_\circ)\ , \end{aligned} \tag{48}$$

where the update is made in accordance with Bayes' Theorem (33). According to (8), the pdf for $\mu$ (48) can equivalently be expressed in terms of the likelihood density $f(\bar{x}s|\mu\sigma I_\circ)$ of the variates $\bar{x} \equiv (x_1 + x_2)/2$ and $s \equiv (x_1 - x_2)/2$:

$$f(\mu|\sigma\bar{x}sI_\circ) = \frac{\pi(\mu|\sigma)}{\zeta_\mu(\bar{x},s,\sigma)}\, f(\bar{x}s|\mu\sigma I_\circ) = \frac{e^{-q(\sigma)\,\mu}}{\zeta_\mu(\bar{x},s,\sigma)}\,\frac{1}{\sigma^2}\,\widetilde{\phi}\Big(\frac{\bar{x}-\mu}{\sigma},\frac{s}{\sigma}\Big)\ . \tag{49}$$

The findings of the present example will become of particular importance in the following two subsections, where we determine the form of the consistency factors $\pi(\mu,\sigma)$ for simultaneous estimation of a location and a dispersion parameter, and $\pi(\sigma|\mu)$ for estimation of a dispersion parameter with $\mu$ being fixed.

### *4.6. Simultaneous inference about a location and a dispersion parameter*

By fixing neither the location nor the dispersion parameter, an inference about the two parameters is invariant under a simultaneous location and scale transformation (17). The symmetry of the problem implies the following form of the functional equation (41) for the appropriate consistency factors $\pi(\mu,\sigma)$:

$$\pi(a\mu+b,a\sigma) = h(b,a)\,\pi(\mu,\sigma) \quad ; \quad \forall\,\mu,b\in\mathbb{R} \ \text{ and } \ \forall\,\sigma,a\in\mathbb{R}^+\ , \tag{50}$$

where

$$h(b,a) \equiv k(a,b)\left|\frac{\partial(a\mu+b,a\sigma)}{\partial(\mu,\sigma)}\right|^{-1} = \frac{k(a,b)}{a^2}\ .$$

LEMMA 6.

$$\pi(\mu,\sigma) = \sigma^{-r} \tag{51}$$

*is the most general solution of* (50), *compatible with a condition* $\pi(0,1) = 1$.

The pdf for $\mu$ and $\sigma$, given $\bar{x}$ and $s$, therefore reads:

$$f(\mu\sigma|\bar{x}sI_\circ) = \frac{\pi(\mu,\sigma)}{\eta_{\mu,\sigma}(\bar{x},s)}\, f(\bar{x}s|\mu\sigma I_\circ) = \frac{\sigma^{-(r+2)}}{\eta_{\mu,\sigma}(\bar{x},s)}\,\widetilde{\phi}\Big(\frac{\bar{x}-\mu}{\sigma},\frac{s}{\sigma}\Big)\ . \tag{52}$$

Then, according to the product rule (27), the pdf (52) can be written as

$$f(\mu\sigma|\bar{x}sI_\circ) = f(\mu|\sigma\bar{x}sI_\circ)\, f(\sigma|\bar{x}sI_\circ)\ , \tag{53}$$

where $f(\sigma|\bar{x}sI_\circ)$ is a marginal pdf (see equation (32)),

$$f(\sigma|\bar{x}sI_\circ) = \int_{-\infty}^{\infty} f(\mu'\sigma|\bar{x}sI_\circ)\, d\mu' = \frac{\sigma^{-r}}{\eta_{\mu,\sigma}(\bar{x},s)}\int_{-\infty}^{\infty} f(\bar{x}s|\mu'\sigma I_\circ)\, d\mu'\ ,$$

while $f(\mu|\sigma\bar{x}sI_\circ)$ denotes the pdf (49) for $\mu$ with the value of the dispersion parameter assumed to be fixed at $\sigma$. Expressing $f(\mu\sigma|\bar{x}sI_\circ)$ and $f(\mu|\sigma\bar{x}sI_\circ)$ in equation (53) in terms of (52) and (49) yields

$$\frac{\sigma^{-r}}{\eta_{\mu,\sigma}(\bar{x},s)} = \frac{\pi(\mu|\sigma)}{\zeta_\mu(\bar{x},s,\sigma)}\, f(\sigma|\bar{x}sI_\circ)\ ,$$

implying $\pi(\mu|\sigma)$ (47) to be independent of $\mu$. Consequently, the value of $q(\sigma)$ in (47) must identically be zero, so that the consistency factor for $\mu$, given fixed dispersion parameter $\sigma$, must be a constant, e.g.,

$$\pi(\mu|\sigma) = 1 \ , \tag{54}$$

regardless the explicit form of the sampling distribution, as well as the value of the fixed dispersion parameter $\sigma$.

### 4.7. Inference about dispersion parameters

As for $\pi(\mu|\sigma)$ in the previous subsection, also the consistency factors $\pi(\sigma|\mu)$ and $\pi(\mu,\sigma)$ need be uniquely determined. In § 2.2 and § 2.3 we stressed that an assignment of a pdf to a dispersion parameter $\sigma$, given datum $x$ and a fixed location parameter $\mu$, can be split into two separate assignments of (the same) scale parameter, each of the latter two being further reducible to an assignment of a pdf to a location parameter $\bar{s}(\sigma) = \ln\sigma$, given a fixed dispersion parameter (see equation (13)). Then, according to the findings of the previous section (see eq. (54)), we can immediately write the appropriate consistency factor for $\bar{s}(\sigma)$:

$$\widetilde{\pi}[\bar{s}(\sigma)|\mu] = 1 \ ,$$

so that (37) implies the factor for the original parameter $\sigma$ to be of the form

$$\pi(\sigma|\mu) = \widetilde{\pi}[\bar{s}(\sigma)|\mu] \left|\bar{s}'(\sigma)\right| = \sigma^{-1} \ , \tag{55}$$

again regardless the explicit form of the particular sampling distribution, as well as the value of the fixed location parameter $\mu$.

The general form of $\pi(\sigma|\mu) \propto \sigma^{-r(\mu)}$ could have been obtained by solving functional equation (40) for scale transformations $\bar{g}_a(\sigma) = a\sigma$,

$$\pi(a\sigma|\mu) = h(a)\,\pi(\sigma|\mu) \ ,$$

where

$$h(a) \equiv \frac{k(a)}{a} \ .$$

When the observed $x$ is equal to $\mu$, the pdf, assigned to $\sigma$,

$$f(\sigma|\mu x I_\circ) = \frac{\pi(\sigma|\mu)}{\eta_\sigma(x,\mu)}\, f(x=\mu|\mu\sigma I_\circ) = \frac{1}{\sigma^{r+1}}\,\frac{\phi(0)}{\eta_\sigma(x,\mu)} \ ,$$

cannot be normalized since the integral

$$\eta_\sigma(x,\mu) = \int_0^\infty \pi(\sigma|\mu)\, f(x=\mu|\mu\sigma' I_\circ)\, d\sigma' = \phi(0)\int_0^\infty \frac{d\sigma'}{\sigma'^{r+1}}$$

clearly does *not* exist for any real $r$. This, for if $x=\mu$ the dispersion parameter $\sigma$ cannot be reduced to a location parameter (recall § 2.3), is in perfect agreement with Theorem 4.

In the limit of complete prior ignorance about its value, the pdf for the *scale* parameter $\sigma$, given fixed $\mu$ and observed $x_1 \neq \mu$, therefore reads:

$$f(\sigma|\mu x_1 I'_\circ) \propto \pi(\sigma|\mu) \times \begin{cases} f^{(1)}(x_1|\mu\sigma I'_\circ)\ ; x_1 > \mu \\ f^{(2)}(x_1|\mu\sigma I'_\circ)\ ; x_1 < \mu \end{cases} \ ,$$

which is equivalent to the pdf for the *dispersion* parameter $\sigma$,

$$f(\sigma|\mu x_1 I_\circ) \propto \pi(\sigma|\mu)\, f(x_1|\mu\sigma I_\circ) = \frac{1}{\sigma^2}\,\phi\Big(\frac{x_1-\mu}{\sigma}\Big) \qquad \forall x_1 \neq \mu\ .$$

Following the steps of the example at the end of § 4.5, we update the pdf for the dispersion parameter $\sigma$ by including a result $x_2$ of an additional measurement in our inference. The updated pdf, expressed in terms of $\bar{x}$ and $s$, reads:

$$f(\sigma|\mu\bar{x}sI_\circ) = \frac{\pi(\sigma|\mu)}{\zeta_\sigma(\bar{x},s,\mu)}\, f(\bar{x}s|\mu\sigma I_\circ) = \frac{\sigma^{-3}}{\zeta_\sigma(\bar{x},s,\mu)}\,\widetilde{\phi}\Big(\frac{\bar{x}-\mu}{\sigma},\frac{s}{\sigma}\Big)\ . \tag{56}$$

The value of $r$ in the consistency factor $\pi(\mu,\sigma)$ (52) is then uniquely determined by invoking the product rule (27) that relates the pdf's $f(\mu\sigma|\bar{x}sI_\circ)$ and $f(\sigma|\mu\bar{x}sI_\circ)$ as:

$$f(\mu\sigma|\bar{x}sI_\circ) = f(\sigma|\mu\bar{x}sI_\circ)\, f(\mu|\bar{x}sI_\circ)\ , \tag{57}$$

with the marginal distribution $f(\mu|\bar{x}sI_\circ)$ standing for

$$f(\mu|\bar{x}sI_\circ) = \int_0^\infty f(\mu\sigma'|\bar{x}sI_\circ)\, d\sigma' = \frac{1}{\eta_{\mu,\sigma}(\bar{x},s)} \int_0^\infty (\sigma')^{-r}\, f(\bar{x}s|\mu\sigma' I_\circ)\, d\sigma'\ .$$

Expressing $f(\mu\sigma|\bar{x}sI_\circ)$ and $f(\sigma|\mu\bar{x}sI_\circ)$ in (57) according to (52) and (56) yields

$$\frac{\sigma^{-r}}{\eta_{\mu,\sigma}(\bar{x},s)} = \frac{\sigma^{-1}}{\zeta_\sigma(\bar{x},s,\mu)}\, f(\mu|\bar{x}sI_\circ)\ ,$$

which, since it is to be true for all $\sigma \in (0,\infty)$, implies $r = 1$, i.e., implies the consistency factor $\pi(\mu,\sigma)$ to be

$$\pi(\mu,\sigma) = \sigma^{-1}\ . \tag{58}$$

Throughout Subsections 4.5–4.7 we thus proved

THEOREM 5. *Axioms of probability (Assumption 2) and the Principle of Consistency combined determine the form of consistency factors $\pi(\mu|\sigma)$, $\pi(\sigma|\mu)$ and $\pi(\mu,\sigma)$ uniquely up to an arbitrary multiplication constant:*

$$\boxed{\pi(\mu|\sigma) = 1 \quad \textit{and} \quad \pi(\sigma|\mu) = \pi(\mu,\sigma) = \sigma^{-1}}\ .$$

### *4.8. On uniqueness and integrability of consistency factors*

Provided the sampling distribution (7) is normalizable, it is straightforward to verify that all the pdf's, involved in the derivations of the consistency factors, are normalizable and thus satisfy the basic requirements (3) and (28), imposed to probability distributions. No requirement of normalizability, however, has ever been imposed to consistency factors, since the factors conceptually differ from pdf's. Moreover,

THEOREM 6. *None of the consistency factors $\pi(\theta)$ for scalar parameters $\theta$ that can be deduced exclusively on the grounds of invariance of the sampling pdf's under Lie groups of transformations, is normalizable, demonstrating in this way unambiguously that the factors do* not *represent any kind of probability distribution, neither the sampling one nor that of our belief.*

The second important property of consistency factors that we want to address, is uniqueness:

THEOREM 7. *The consistency factors for inference about a parameter $\theta$ are unique in that if a family $I_\circ$ of sampling distributions* (14) *is invariant under two Lie groups, say $\mathcal{G}$ and $\mathcal{H}$, then the two groups lead to the consistency factors $\pi_g(\theta)$ and $\pi_h(\theta)$ that are identical up to a multiplication constant,*

$$\pi_h(\theta) = k\,\pi_g(\theta)\ . \tag{59}$$

## 5. Calibration

Thus far, the theory of plausible inference about parameters has been developed by following only the Principle of Consistency, while the implications of the Operational Principle *II* have not yet been considered. According to the latter, in order to exceed the level of a mere speculation, our theory of inference about parameters must be exposed, i.e., must be able to make predictions that can be verified (or falsified) by experiments.

Let therefore several values $x_i$ of a scalar random variate $x$ be sampled from a family $I_\circ$ (14) of sampling distributions. The value $\theta_i$ of the scalar parameter $\theta$ of the family may arbitrarily vary from one sampling to another. The predictions of the theory are then made in terms of probabilities

$$P\big(\theta \in (\theta_{i,1},\theta_{i,2})|x_i I_\circ\big) = \int_{\theta_{i,1}}^{\theta_{i,2}} f(\theta'|x_i I_\circ)\,d\theta' = \delta \tag{60}$$

that given measured value $x_i$ of the sampling variate, an interval $(\theta_{i,1},\theta_{i,2})$ contains the actual value $\theta_i$ of the parameter.

The interval for the inference of a particular value $\theta_i$ is not unique: it can be the shortest of all possible intervals, the central interval with $P(\theta_i \le \theta_{i,1}|x_i I_\circ) = P(\theta_i > \theta_{i,2}|x_i I_\circ) = (1-\delta)/2$, the lower-most interval with $\theta_{i,1} = \theta_a$, the upper-most interval with $\theta_{i,2} = \theta_b$, or any other interval as long as the probability (60) is equal to $\delta$.

Our probability judgments are said to be *calibrated* if the fraction of inferences with the specified interval $(\theta_{i,1},\theta_{i,2})$ covering the true value $\theta_i$ of the parameter in the particular sampling $x_i$, coincides with $\delta$.

For sampling distributions whose cdf $F(x,\theta,I_\circ)$ is either strictly increasing or strictly decreasing in $\theta$, a necessary and sufficient condition for calibrated inferences reads (Fisher, 1956, § 3.6, p. 70):

$$f(\theta|xI_\circ) = \mp F_2(x,\theta,I_\circ)\ , \tag{61}$$

where the upper (lower) sign is for cdf's that are strictly decreasing (increasing) in $\theta$. It is easy to verify that for the pdf's, assigned to location and scale parameters by using the consistency factors (54) and (55), the condition (61) is satisfied.

### 5.1. Lindley's Theorem

The probability distributions for location, scale or dispersion parameters that were assigned by following the Consistency Principle, passed an important test: they are all calibrated. The question can be raised whether there are any other types of parameters that are also in accordance with the calibration requirement (61)? We restrict the answer only to parameters whose pdf can be assigned according to the Consistency Theorem (34). By combining the two equations we obtain:

$$\pi(\theta)\,F_1(x,\theta,I_\circ) \pm \eta_\theta(x)\,F_2(x,\theta,I_\circ) = 0\ , \tag{62}$$

where the upper (lower) sign stands for cdf's which are strictly decreasing (increasing) in $\theta$. By defining function $G(x,\theta)$ as a difference (sum),

$$G(x,\theta) \equiv s(x) \mp \bar{s}(\theta) \ ,$$

with $s(x)$ and $\bar{s}(\theta)$ being related to $\pi(\theta)$ and $\eta_\theta(x)$ as

$$s'(x) = \eta_\theta(x) \ \text{ and } \ \bar{s}'(\theta) = \pi(\theta) \ ,$$

equation (62) can be rewritten as

$$F_1(x,\theta,I_\circ)\, G_2(x,\theta) - F_2(x,\theta,I_\circ)\, G_1(x,\theta) = 0 \ ,$$

with $G_1(x,\theta) = \eta_\theta(x)$ and $G_2(x,\theta) = \pi(\theta)$ being strictly positive functions (see § 3.4). But as we saw in § 2.3, the general solution of such a differential functional equation implies a cdf $F(y,\mu,I'_\circ)$ of the form

$$F(y,\mu,I'_\circ) = \Phi(y-\mu) \ ,$$

corresponding to a cdf of $y \equiv s(x)$ with $\mu \equiv \pm\bar{s}(\theta)$ being a location parameter (10). Therefore, in the limit of complete prior ignorance, *an inference about a parameter $\theta$ that is subject to the calibration condition* (61)*, is necessarily reducible to an inference about a location parameter*. Note that this result was first obtained by Lindley (1958) by combining the calibration condition (61) and the Bayes' Theorem with a prior pdf $f(\theta|I_\circ)$ which is independent of data $x$.

Imagine that for a particular parameterization of a sampling variate $x$ and the corresponding parameter $\theta$, say $y \equiv s(x)$ and $\mu \equiv \bar{s}(\theta)$, a calibrated inference about $\bar{s}(\theta)$ exists, i.e., the cdf of $s(x)$, $F\big(s(x), s(\theta), I'_\circ\big)$, solves equation (62),

$$\widetilde{\pi}[\bar{s}(\theta)]\, F_1\big(s(x),\bar{s}(\theta),I'_\circ\big) \pm \widetilde{\eta}_\mu[s(x)]\, F_2\big(s(x),\bar{s}(\theta),I'_\circ\big) = 0 \ . \tag{63}$$

Here, $\widetilde{\pi}[\bar{s}(\theta)]$ is the consistency factor for $\bar{s}(\theta)$, while $\widetilde{\eta}_\mu[s(x)]$ is the appropriate normalization factor for $f\big(\bar{s}(\theta)|s(x)I'_\circ\big)$. By differentiating equation (19) separately with respect to $x$ and $\theta$, (63) can be expressed in terms of $F_{1,2}(x,\theta,I_\circ)$, instead of $F_{1,2}\big(s(x),\bar{s}(\theta),I'_\circ\big)$:

$$\widetilde{\pi}[\bar{s}(\theta)]\, \big|\bar{s}(\theta)\big|\, F_1(x,\theta,I_\circ) \pm \widetilde{\eta}_\mu[s(x)]\, \big|s(x)\big|\, F_2(x,\theta,I_\circ) = 0 \ . \tag{64}$$

Inference about $\theta$ will thus be calibrated (cf. (64) and (62)) if and only if

$$\widetilde{\pi}[\bar{s}(\theta)] = \pi(\theta)\, \big|\bar{s}(\theta)\big|^{-1} \quad \text{and} \quad \widetilde{\eta}_\mu[s(x)] = \eta_\theta(x)\, \big|s(x)\big|^{-1} \ ,$$

which coincides with the rules (37) and (38) for transformation of the consistency and normalization factors, with the arbitrary constant $\tilde{k}$ in (37) and (38) being set to unity. Calibration of inference about a parameter of a sampling distribution is therefore invariant under arbitrary one-to-one transformations of the sampling variate and the inferred parameter.

Then, since every problem of inference about a parameter $\theta$ that is uniquely solvable within the Principle of Consistency, is reducible to an inference about a location parameter $\mu = \bar{s}(\theta)$, and since the uniform consistency and normalization factors $\widetilde{\pi}[\bar{s}(\theta)]$ and $\widetilde{\eta}_\mu[s(x)]$ provide a calibrated inference about $\mu$, the inference about $\theta$ will also be automatically calibrated.

The Principle of Consistency and the Operational Principle are thus equivalent concepts for determination of the consistency factors. First, they are applicable under identical circumstances, i.e., when the problem of inference is reducible to a problem of inference about a location parameter. Second, the consistency factor that solves the functional equation (40), based on the requirement of consistency, is identical to the solution of the calibration requirement (62). The equivalence of the two Principles speaks in favour of complete reconciliation between the (objective) Bayesian school and the frequentist school of inference, the former paying attention primarily to logical consistency and the latter stressing the importance of verifiable predictions.

In order to avoid a frequent misunderstanding we should stress that a pdf, assigned to an inferred parameter $\theta_i$, does *not* necessarily imply that the parameter is distributed according to $f(\theta_i|x_i I_\circ)$: what *is* distributed is our belief in different values of $\theta$ within the parameter space $\Theta$. In practice, the inferred parameters are usually fixed while unknown, but can also, at least in principle, arbitrarily vary from one inference to another. In general, the assigned $f(\theta_i|x_i I_\circ)$ will therefore differ from the true distribution of the parameters $\theta_i$. Still, the calibrated pdf's $f(\theta_i|x_i I_\circ)$ will correctly predict the fraction $\delta$ (60) of the confidence intervals $(\theta_{i,1}, \theta_{i,2})$, covering the true values $\theta_i$.

## 5.2. Calibration and symmetry preserved under updating

In previous subsections we saw that a calibrated assignment of probability distribution to an inferred scalar parameter $\theta$ can be assured if $\theta$ is reducible to a location parameter $\mu$. The calibration of the probability distributions, assigned to the inferred parameters, is preserved under updating:

THEOREM 8. *Every updating of a calibrated probability distribution for an inferred parameter that is performed in accordance with Bayes' Theorem, preserves the calibration of the distribution. The preservation of calibration can be connected to the preserved translation invariance under the updating.*

## 5.3. Predictive distributions

Imagine now a slightly different problem. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a sequence of recorded values of a continuous sampling variate $x$ with the pdf $f(x|\theta I_\circ)$. In the present subsection we are interested in predicting values of the sampling variate that are yet to be observed, rather than in inferring the (unknown) value of the parameter $\theta$ of its distribution. That is, given the collected $\mathbf{x}$, we are aiming at assigning a pdf $f(x_{n+1}|\mathbf{x} I_\circ)$ to the possible values of $x_{n+1}$.

Suppose that the family $I_\circ$ is reducible to a family $I'_\circ$ that is parameterized by a location parameter, so that the consistency factor $\pi(\theta)$ can uniquely be determined, and that a pdf, based on $\mathbf{x}$, can be assigned to $\theta$:

$$f(\theta|\mathbf{x} I_\circ) = \frac{\pi(\theta)}{\zeta_\theta(\mathbf{x})}\, f(\mathbf{x}|\theta I_\circ) = \frac{\pi(\theta)}{\zeta_\theta(\mathbf{x})} \prod_{i=1}^{n} f(x_i|\theta I_\circ) \; .$$

Then, the joint pdf $f(\theta x_{n+1}|\mathbf{x} I_\circ)$ can be factorized (cf. equation (30)),

$$f(\theta x_{n+1}|\mathbf{x} I_\circ) = f(\theta|\mathbf{x} I_\circ)\, f(x_{n+1}|\theta \mathbf{x} I_\circ) = f(\theta|\mathbf{x} I_\circ)\, f(x_{n+1}|\theta I_\circ) \; ,$$

and the pdf $f(x_{n+1}|\mathbf{x} I_\circ)$ can be obtained simply by a convolution, i.e., by the marginalization of $f(\theta x_{n+1}|\mathbf{x} I_\circ)$,

$$f(x_{n+1}|\mathbf{x} I_\circ) = \int_\Theta f(\theta' x_{n+1}|\mathbf{x} I_\circ)\, d\theta' \; . \tag{65}$$

It is important to note that $f(x_{n+1}|\mathbf{x} I_\circ)$ stands for the distribution of our degree of belief in different values of $x_{n+1}$: in the context of the prediction of the future value of $x$, $x_{n+1}$ is *not* a sampling variate, so that $f(x_{n+1}|\mathbf{x} I_\circ)$ will in general differ from the observed distribution of the future values of the sampling variate $x$. However, in the same way as for Theorem 8, we can prove that the future values of $x$, predicted according to (65), will still be calibrated: the probabilities

$$P(x_{n+1} \in (x_c, x_d)|\mathbf{x}I_\circ) = \int_{x_c}^{x_d} f(x'_{n+1}|\mathbf{x}I_\circ)\, dx'_{n+1}$$

will always coincide with the relative frequency of intervals $(x_c, x_d)$, covering the observed future values of $x$ in the long run.

## 6. Conclusions

This article presents a theory of probabilistic inference about the parameters of sampling distributions. A special attention has been payed to assignment of probability distributions to the inferred parameters, with such an assignment representing natural and indispensable starting point in every inference about the parameters. In order to be internally consistent, the assignments must be made in accordance with the Consistency Theorem (34). The form of the Theorem is very similar to the form of Bayes' Theorem (33) that is used for *updating* the assigned probability distributions, but we stressed an important difference between the two. While in Bayes' Theorem the prior probability $f(\theta|x_1 I_\circ)$ represents a distribution of credibility among different values of the inferred parameter $\theta$, $\pi(\theta)$ in the Consistency Theorem is just a proportionality factor that *by no means* represents any kind of probability distribution.

The requirement of consistency uniquely determines the form of the consistency factors only in those inferences that are reducible to inferences about location parameters of sampling distributions. Since, according to Lindley's Theorem, correct verifiable predictions can only be assured under the very same condition, the requirement of reducibility does *not* restrict the class of sampling distributions with possible consistent and calibrated inference about their parameters.

The theory is operational in the sense that it is verifiable from long range consequences. Within the theory, all inferences are calibrated: in the long run, the fraction of the confidence intervals, constructed on the basis of (posterior) probabilities, that cover the true values of the inferred parameters, always coincides with the probability contents, assigned to these intervals. This is a very important feature that permits for a reconciliation between the frequentist and the Bayesian approaches to inference, probably the same kind of reconciliation that Kendall (1949) had in mind: "Neither party can avoid ideas of the other in order to set up and justify a comprehensive theory." In this way, the distinction between the *theory of probability* and that of *statistical inference* may be removed, leaving a logical unity and simplicity.

## Acknowledgment

Prof. Gabrijel Kernel, Prof. Alojz Kodre, Prof. Louis Lyons, Dr. Igor Mandić, Prof. Aleš Stanovnik and Prof. Peter Šemrl read a preliminary draft of this paper. We wish to thank them for their kindness and for their help in correcting several inaccuracies. They are not responsible, of course, for any errors that remain as to the opinion expressed concerning the nature of consistent plausible reasoning.

## Appendix: Proofs of Theorems and Lemmata

### *A.1. Proof of* Lemma 1

If a pdf of a sampling variate $x$ is of the form $\phi(x,\theta)$, the assumed invariance of the distribution under $\mathcal{G}$ implies the same type of distribution, $\phi\big(g_a(x),\bar{g}_a(\theta)\big)$, of $g_a(x)$, for all $g_a\in\mathcal{G}$ and $\bar{g}_a\in\bar{\mathcal{G}}$. The distribution of the transformed variate $y$, on the other hand, reads

$$f(y|\nu I'_\circ)=\phi(x,\theta)\,\big|s'(x)\big|^{-1}\equiv\widetilde{\phi}(y,\nu)\;. \tag{66}$$

Then, since

$$h_a(y)=s\{g_a[s^{-1}(y)]\}=s[g_a(x)]\quad\text{and}\quad\bar{h}_a(\nu)=\bar{s}\{\bar{g}_a[\bar{s}^{-1}(\nu)]\}=\bar{s}[\bar{g}_a(\theta)] \tag{67}$$

for all $h_a\in\mathcal{H}$ and $\bar{h}_a\in\bar{\mathcal{H}}$, the pdf of $h_a(y)$, given $\bar{h}_a(\nu)$, $f\big(h_a(y)|\bar{h}_a(\nu)I''_\circ\big)$, is equal to

$$f\big(s[g_a(x)]|\bar{s}[\bar{g}_a(\theta)]I''_\circ\big)=\phi\big(g_a(x),\bar{g}_a(\theta)\big)\,\big|s'[g_a(x)]\big|^{-1}\;,$$

which, by (66) and (67), is further equal to $\widetilde{\phi}\big(s[g_a(x)],\bar{s}[\bar{g}_a(\theta)]\big)=\widetilde{\phi}\big(h_a(y),\bar{h}_a(\nu)\big)$. □

### A.2. *Proof of* Lemma 2

Indeed:

$$\begin{aligned}F\big(s(x),\bar{s}(\theta),I'_\circ\big)-F\big(s(x_a),\bar{s}(\theta),I'_\circ\big)&=\int_{s(x_a)}^{s(x)}f\big(y'|\bar{s}(\theta)I'_\circ\big)\,dy'\\&=\int_{s(x_a)}^{s(x)}\phi(x',\theta)\,\big|s'(x')\big|^{-1}\,d[s(x')]\\&=\pm\int_{x_a}^{x}\phi(x',\theta)\,dx'\\&=\pm F(x,\theta,I_\circ)\;,\end{aligned}$$

where the positive and the negative sign correspond to $s'(x)>0$ and to $s'(x)<0$, respectively. Setting $x$ to the upper bound $x_b$ of its range, the above equation reads:

$$F\big(s(x_b),\bar{s}(\theta),I'_\circ\big)-F\big(s(x_a),\bar{s}(\theta),I'_\circ\big)=\pm F(x_b,\theta,I'_\circ)=\pm1\;.$$

Since the cdf's are limited within $[0,1]$, this completes the proof of the Lemma by implying

$$F\big(s(x_a),\bar{s}(\theta),I'_\circ\big)=\begin{cases}0\;;\;g'(x)>0\\1\;;\;g'(x)<0\end{cases}\quad\text{and}\quad F\big(s(x_b),\bar{s}(\theta),I'_\circ\big)=\begin{cases}1\;;\;g'(x)>0\\0\;;\;g'(x)<0\end{cases}\;.$$ □

### A.3. *Proof of* Lemma 3

Let $g(\,\cdot\,,b\circ a)\equiv g_{b\circ a}\in\mathcal{G}$ denote a composition of group elements $g(\,\cdot\,,a)\equiv g_a$ and $g(\,\cdot\,,b)\equiv g_b$, such that

$$g(x,b\circ a)=g[g(x,a),b]\;.$$

When differentiated with respect to $a$, the above equation reads:

$$g_2(x,b\circ a)\,\frac{d}{da}(b\circ a)=g_1[g(x,a),b]\,g_2(x,a)\;,$$

with existence of the derivative of $(b\circ a)$ being guaranteed by the requirement on $\mathcal{G}$ to be a Lie group (see, for example, Elliot and Dawber, 1986, § 7.1, p. 126). Since the above equation is valid for arbitrary $a$ and $b$, it is also to be valid for $b = a^{-1}$ (here, $a^{-1}$ is the index of the inverse of $g_a$), so that

$$g_2(x,c)\Big|_{c=e} \frac{\partial}{\partial a}(b \circ a)\Big|_{b=a^{-1}} = g_1[g(x,a),b]\Big|_{b=a^{-1}} g_2(x,a) \ ,$$

where $c \equiv b \circ a$. The left-hand side of the above equation is identically zero due to the premise of the Lemma,

$$g_2(x,c)\Big|_{c=e} \equiv \frac{\partial}{\partial c} g_c(x)\Big|_{c=e} = 0 \ .$$

On the right-hand side, however, the first term,

$$g_1[g(x,a),b]\Big|_{b=a^{-1}} = g_1(y,a^{-1}) = \frac{\partial}{\partial y} g_{a^{-1}}(y) \neq 0 \ ,$$

is non-vanishing for all admissible values of the index $a$ and of the variate $y \equiv g_a(x) \in X$ (all group transformations (15) are necessarily one-to-one – recall § 2.3), so that

$$g_2(x,a) = \frac{\partial}{\partial a} g_a(x) = 0$$

is implied for all permissible $a$, i.e., $g_a(x)$ is permitted to be a function of $x$ only, say $h(x)$. When $g_e(x) = x$ is invoked, this further means $h(x) = x$ and the Lemma is proved. □

### *A.4. Proof of* LEMMA 4

The proof of the Lemma is accomplished by *reductio ad absurdum* so let suppose that there exists a value of $\theta_0 \in \Theta$ for which the partial derivative (22) vanishes. Then, since the sampling distribution is invariant under $\mathcal{G}$, equation (20) applies which, when differentiated with respect to $a$ and set afterwards $a = e$, yields

$$F_1(x,\theta,I_\circ)\, \frac{\partial}{\partial a} g_a(x)\Big|_{a=e} = -F_2(x,\theta,I_\circ)\, \frac{\partial}{\partial a} \bar{g}_a(\theta)\Big|_{a=e} \ .$$

The second term on right-hand side of the above equation vanishes for $\theta = \theta_0$ which, when strict positivity of $F_1(x,\theta,I_\circ) = f(x|\theta I_\circ)$ is invoked, implies

$$\frac{\partial}{\partial a} g_a(x)\Big|_{a=e} = 0 \quad ; \quad \forall\, x \in X \ .$$

This means, according to Lemma 3, that all transformations $g_a \in \mathcal{G}$ of $X$ are trivial, which is in direct contradiction with the premises of the Lemma, so that the proof is completed. □

### *A.5. Proof of* THEOREM 2

The assumed invariance of distribution (9) implies the existence of $y \equiv g_a(x)$ and $\nu \equiv \bar{g}_a(\theta)$, such that

$$f(y|\nu I_\circ) = \phi(y - \nu) \ ,$$

with $g_a$ and $\bar{g}_a$ being elements of the group $\mathcal{G}$ and the corresponding induced group $\bar{\mathcal{G}}$, respectively. Then, due to equation (10), the cdf of $y$ reads

$$F(y,\nu,\sigma,I_\circ) = \Phi(y - \nu) \ .$$

In addition, according to Lemma 2, the cdf (10) of $x$ and that of $y$ are related as

$$\Phi(y-\nu) = \begin{cases} \Phi(x-\mu) & ;\ g'_a(x) > 0 \\ 1-\Phi(x-\mu) & ;\ g'_a(x) < 0 \end{cases} \quad .$$

If differentiated with respect to $a$, the relation between the cdf's yields

$$\Phi'(y-\nu)\ \left[\frac{\partial}{\partial a} g_a(x) - \frac{\partial}{\partial a}\bar{g}_a(\mu)\right] = 0 \ ,$$

implying

$$\frac{\partial}{\partial a} g_a(x) = \frac{\partial}{\partial a}\bar{g}_a(\mu)$$

for all $y$ and $\nu$ with non-vanishing $\Phi'(y-\nu) = f(y|\nu\sigma I_\circ)$. Then, the two derivatives can only depend on $a$, but not on $x$ or $\mu$. Parameterizing this dependence by $h'(a) \equiv dh(a)/da$ yields

$$g_a(x) = h(a) + k(x) \quad \text{and} \quad \bar{g}_a(\mu) = h(a) + l(\mu) \ ,$$

which, when $a$ is set to index $e$ of the unity element, further gives

$$k(x) = x - h(e) \quad \text{and} \quad l(\mu) = \mu - h(e) \ .$$

Finally, since the elements of the groups $\mathcal{G}$ and $\bar{\mathcal{G}}$ can be re-enumerated according to $b \equiv h(a) - h(e)$, we obtain

$$g_b(x) = x + b \quad \text{and} \quad \bar{g}_b(\mu) = \mu + b \ ,$$

while the invariance of the sample and the parameter spaces implies $X = (-\infty, \infty)$ and $\Theta = (-\infty, \infty)$. □

### *A.6. Proof of* THEOREM 3

Let the pdf of $x$ given $\theta = \theta_1$, $f(x|\theta_1 I_\circ)$, be denoted by $\phi(x, \theta_1)$, and let the pdf for $\theta$ at $\theta = \theta_1$, given the observation $x \in (x_1, x_1 + dx)$, whose general form we would like to determine, be denoted by $\psi(x_1, \theta_1)$:

$$\psi(x_1, \theta_1) \equiv f(\theta_1 | x_1 I_\circ) \ . \qquad (68)$$

In addition, let another value of $x$, independent of the first one, be recorded in an interval $(x_2, x_2 + dx)$, with the appropriate likelihood density being

$$f(x_2|\theta_1 x_1 I_\circ) = f(x_2|\theta_1 I_\circ) = \phi(x_2, \theta_1) \ .$$

In § 3.2 we saw that the only way of updating pdf for $\theta$ that is consistent with the adopted rules, in particular with the product rule (30), is the one in accordance with Bayes' Theorem (33). With $f(\theta|x_1 I_\circ)$ taking the role of the prior pdf for $\theta$, the pdf posterior to including $x_2$ into our reasoning about $\theta$ is thus written as:

$$f(\theta_1|x_1 x_2 I_\circ) = \frac{\psi(x_1, \theta_1)\,\phi(x_2, \theta_1)}{\zeta_\theta(x_1, x_2)} \ , \qquad (69)$$

with the normalization constant $\zeta_\theta(x_1, x_2)$ being

$$\zeta_\theta(x_1, x_2) = \int\limits_{\Theta} \psi(x_1, \theta')\,\phi(x_2, \theta')\,d\theta' \ .$$

Nothing prevents us from reversing the order of taking the two pieces of information, $x_1$ and $x_2$, into account, which results in the following pdf for $\theta$:

$$f(\theta_1|x_2 x_1 I_\circ) = \frac{\psi(x_2,\theta_1)\,\phi(x_1,\theta_1)}{\zeta_\theta(x_2,x_1)}\ . \tag{70}$$

Moreover, the Consistency Principle *I* requires equality of the two results, (69) and (70):

$$f(\theta_1|x_1 x_2 I_\circ) = f(\theta_1|x_2 x_1 I_\circ)\ ,$$

i.e., it requires

$$\frac{\psi(x_1,\theta_1)\,\phi(x_2,\theta_1)}{\zeta_\theta(x_1,x_2)} = \frac{\psi(x_2,\theta_1)\,\phi(x_1,\theta_1)}{\zeta_\theta(x_2,x_1)}\ . \tag{71}$$

We distinguish two cases:

CASE 1. None of the pdf's, assigned to the inferred parameter, vanishes in (71). Then, due to the imposed continuity of $\psi(x,\theta)$ (Assumptions 3), and due to the normalization condition (28), there exists $\theta_2 \neq \theta_1$ for which none of the terms in

$$\frac{\psi(x_1,\theta_2)\,\phi(x_2,\theta_2)}{\zeta_\theta(x_1,x_2)} = \frac{\psi(x_2,\theta_2)\,\phi(x_1,\theta_2)}{\zeta_\theta(x_2,x_1)} \tag{72}$$

vanishes, either. Dividing equations (71) and (72) results in

$$\frac{\kappa(x_1,\theta_1)}{\kappa(x_1,\theta_2)} = \frac{\kappa(x_2,\theta_1)}{\kappa(x_2,\theta_2)}\ , \tag{73}$$

where

$$\kappa(x,\theta) \equiv \frac{\psi(x,\theta)}{\phi(x,\theta)}\ .$$

Clearly, in order to ensure equality in (73) for all possible values of $x_1$ and $x_2$, the left-hand and the right-hand side of the equation must be independent of $x_1$ and $x_2$, respectively, but may depend on the values $\theta_1$ and $\theta_2$ of the parameter $\theta$. Taking this dependence into account by introducing a function $h(\theta_1,\theta_2)$, we obtain

$$\frac{\kappa(x,\theta_1)}{\kappa(x,\theta_2)} = h(\theta_1,\theta_2)\ ,$$

further implying factorizability of $h(\theta_1,\theta_2)$,

$$h(\theta_1,\theta_2) \equiv \frac{\pi(\theta_1)}{\pi(\theta_2)}\ ,$$

so that

$$\frac{\kappa(x,\theta_1)}{\pi(\theta_1)} = \frac{\kappa(x,\theta_2)}{\pi(\theta_2)} \equiv \frac{1}{\eta(x)}\ ,$$

and finally

$$\psi(x,\theta) = \frac{\pi(\theta)}{\eta_\theta(x)}\,\phi(x,\theta)\ ,$$

which, when written in terms of generic pdf's, reduces to (34).

CASE 2. One of the two pdf's assigned to the inferred parameter, e.g., $\psi(x_2,\theta_1)$, is zero. Then, according to (71), $\psi(x_1,\theta_1)$ must vanish for all $x_1 \in X$.

Note that within the theorem and its proof, both $x$ and $\theta$ may be multi-dimensional variates. □

### A.7. *Proof of* THEOREM 4

Suppose for a moment that a pdf for $\theta$, $f(\theta|x_0 I_\circ)$, *can* be assigned to $\theta \in \Theta$ based on $x_0 \in X - \widetilde{X} \subseteq X$ for which the partial derivative (21) vanishes. Then, since the sampling distribution is invariant under a Lie group $\mathcal{G}$ of transformations, the distribution assigned to $\theta$ is invariant under the induced Lie group $\bar{\mathcal{G}}$ so that the equation (43), concerning the cdf $H(x_0, \theta, I_\circ)$ for $\theta$, is valid. When differentiated with respect to $a$ and set afterwards $a = e$, (43) further implies

$$H_1(x_0, \theta, I_\circ)\, \frac{\partial}{\partial a} g_a(x_0)\Big|_{a=e} = -H_2(x_0, \theta, I_\circ)\, \frac{\partial}{\partial a} \bar{g}_a(\theta)\Big|_{a=e} \quad ; \quad \forall\, \theta \in \Theta \ ,$$

whose left-hand side vanishes due to the premise, adopted at the beginning of the proof. Since, by Lemma 4, the second term on the right-hand side does not vanish anywhere on $\Theta$, $H_2(x_0, \theta, I_\circ) = f(\theta|x_0 I_\circ)$ must vanish for all $\theta \in \Theta$, which is incompatible with the normalization requirement (28). Therefore, the assumed existence of $f(\theta|x_0 I_\circ)$, based on $x_0$ with vanishing derivative (21), inevitably leads to inconsistencies and is thus ruled out. □

### A.8. *Proof of* LEMMA 5

The consistency factors can only be determined up to an arbitrary multiplication constant, so that no generality is lost by choosing the factor such that $\pi(0|\sigma) = 1$. In addition, if (46) is to be true for all $\mu, b \in (-\infty, \infty)$, it must also be true for $b = -\mu$ when it reads

$$\pi(0) = k(-\mu)\, \pi(\mu|\sigma) = 1 \ .$$

By construction, consistency factors do *not* vanish anywhere where defined (recall Theorem 3 and its proof). Non-vanishing $\pi(\mu|\sigma)$ thus implies

$$k(b) = \frac{1}{\pi(-b|\sigma)} \ ,$$

which, when inserted to (46), further yields

$$\pi(u + v|\sigma) = \pi(u|\sigma)\, \pi(v|\sigma) \ ,$$

where $u \equiv x + b$ and $v \equiv -b$. The latter functional equation is of Cauchy's type (Cauchy, 1897, Part 1, Chapter V, § I, pp. 98-105; see also Aczél, 1966, § 2.1.1-2.1.2, pp. 31-42) and its most general continuous solutions are

$$\pi(\mu|\sigma) = \exp\{-q\, \mu\} \quad \text{and} \quad \pi(\mu|\sigma) = 0 \ ,$$

where $q$ is an arbitrary constant. Finally, the latter of the two solutions is ruled out by requirement (28). □

### A.9. *Proof of* LEMMA 6

For $b = -a\mu$ equation (50) reads

$$\pi(0, a\sigma) = h(-a\mu, a)\, \pi(\mu, \sigma) \tag{74}$$

which, for $a = \sigma^{-1}$, further reduces to

$$\pi(0,1) = h\left(-\mu\sigma^{-1}, \sigma^{-1}\right)\,\pi(\mu,\sigma)\ .$$

Choosing $\pi(0,1) = 1$ leads to

$$h(u,v) = \frac{1}{\pi\left(-uv^{-1}, v^{-1}\right)} \quad ; \quad \forall\, u \in \mathbb{R} \ \text{ and } \ \forall\, v \in \mathbb{R}^+\ ,$$

which, when inserted to (74), yields

$$\pi(\mu,\sigma) = \pi\left(\mu, a^{-1}\right)\,\pi(0, a\sigma)\ .$$

By setting $a = 1$ we obtain

$$\pi(\mu,\sigma) = \Xi(\mu)\,\Omega(\sigma)\ ,$$

where

$$\Xi(\mu) \equiv \pi(\mu,1) \ \text{ and } \ \Omega(\sigma) \equiv \pi(0,\sigma)\ .$$

That is, $\pi(\mu,\sigma)$ is factorizable and, consequently, $h(b,a)$ is factorizable, too:

$$h(b,a) = \frac{1}{\Xi\left(-ba^{-1}\right)\,\Omega\left(a^{-1}\right)}\ .$$

Applying the factorizability to (50) and setting $a = \sigma^{-1}$ and $b = 0$ then yields

$$\Xi\left(\mu\sigma^{-1}\right) = \Xi(\mu)\ ,$$

which implies $\Xi(\mu)$ be a constant, e.g., $\Xi(\mu) = 1$, and so

$$\pi(\mu,\sigma) = \Omega(\sigma) \ \text{ and } \ h(b,a) = \frac{1}{\Omega\left(a^{-1}\right)}\ .$$

As a result, (50) reduces to

$$\Omega(uv) = \Omega(u)\,\Omega(v)$$

($u \equiv a\sigma$ and $v \equiv a^{-1}$), which is again a Cauchy's functional equation (Cauchy, 1897, Part 1, Chapter V, § I, pp. 104-105) whose most general continuous solutions are

$$\Omega(\sigma) = \sigma^{-r} \ \text{ and } \ \Omega(\sigma) = 0$$

($r$ is an arbitrary real constant). As before, the trivial solution $\Omega(\sigma) = 0$ is ruled out by invoking normalization requirement (28). □

### *A.10. Proof of* THEOREM 6

To verify the Theorem, suppose for a moment that the opposite is true, i.e., that

$$\int\limits_{\Theta} \pi(\theta')\,d\theta' < \infty\ . \tag{75}$$

Then, according to Theorems 1 and 4, every scalar parameter $\theta$, with the corresponding consistency factor determined uniquely by the underlying symmetry under a Lie group of transformations, is reducible to a location parameter $\mu = \bar{s}(\theta)$, with the two consistency factors, $\pi(\theta)$ and $\widetilde{\pi}(\mu|\sigma)$, being related according to (37),

$$\widetilde{\pi}(\mu|\sigma)\, d\mu = k\, \pi(\theta)\, d\theta\ ,$$

where $k$ is an arbitrary (finite) constant. The supposed existence of the integral (75) thus implies integrability of $\widetilde{\pi}(\mu)$ over the corresponding parameter space $M$:

$$\int\limits_M \widetilde{\pi}(\mu'|\sigma)\, d\mu' = k \int\limits_\Theta \pi(\theta')\, d\theta' < \infty\ . \tag{76}$$

If, on the other hand, $\widetilde{\pi}(\mu|\sigma)$ can be determined exclusively by invoking the translation symmetry, the domain $M$ of $\widetilde{\pi}(\mu|\sigma)$ must be invariant under the group of translations, i.e., it must range over the entire real axis. But then, since $\widetilde{\pi}(\mu|\sigma) \propto 1$, the integral

$$\int\limits_M \widetilde{\pi}(\mu'|\sigma)\, d\mu' = \int_{-\infty}^{\infty} d\mu' \tag{77}$$

clearly does *not* exist. By realizing an evident contradiction between (76) and (77), we can conclude that the supposed normalizability of consistency factors (75) inevitably leads to inconsistencies and is thus ruled out. □

### *A.11. Proof of* THEOREM 7

Recall Theorems 1 and 4, stating that if a consistency factor is to be deduced solely by invoking invariance of a sampling pdf $f(\theta|xI_\circ)$, the problem of inference must be deducible by one-to-one transformations to an inference about a location parameter (recall also (24) and (26)). Let $y = y(x)$, $\mu = \mu(\theta)$, $z = z(x)$ and $\nu = \nu(\theta)$ be such transformations, the former two corresponding to the invariance of the pdf under $\mathcal{G}$, and the latter two to the invariance under $\mathcal{H}$, so that, according to (26),

$$F(x, \theta, I_\circ) = \Phi(y - \mu) = \widetilde{\Phi}(z - \nu)\ . \tag{78}$$

By transitivity of one-to-one relations, $z = z(y)$ and $\nu = \nu(\mu)$ are also one-to-one. Consequently, by differentiating (78) separately with respect to $y$ and $\mu$, and by dividing the two resulting equations, we obtain

$$\frac{\partial z}{\partial y} = \frac{\partial \nu}{\partial \mu}\ .$$

If this is to be true for all $y$ and $\mu$, it must further be equal to a constant, say $c$.

Above we learned that sampling distributions (9), determined by location parameters, can only be invariant under one Lie group, that of translations (Theorem 2), and that such an invariance uniquely determines the form of the consistency factors (54). Consequently, $\widetilde{\pi}_g(\mu|\sigma_\mu)$ exists and is unique, and so is $\widetilde{\pi}_h(\nu|\sigma_\nu)$. The two consistency factors are related *via* (37),

$$\widetilde{\pi}_h(\nu|\sigma_\nu) = \tilde{k}\, \widetilde{\pi}_g(\mu|\sigma_\mu) \left|\frac{\partial \nu}{\partial \mu}\right|^{-1} = \frac{\tilde{k}}{|c|} \widetilde{\pi}_g(\mu|\sigma_\mu)\ . \tag{79}$$

The factors $\widetilde{\pi}_g(\mu|\sigma_\mu)$ and $\pi_g(\theta)$, as well as the factors $\widetilde{\pi}_h(\nu|\sigma_\nu)$ and $\pi_h(\theta)$, are also related by the same equation (37):

$$\widetilde{\pi}_g(\mu|\sigma_\mu) = k_1\, \pi_g(\theta) \left|\frac{\partial \mu}{\partial \theta}\right|^{-1}\ , \tag{80}$$

$$\widetilde{\pi}_h(\nu|\sigma_\nu) = k_2\, \pi_h(\theta) \left|\frac{\partial \nu}{\partial \theta}\right|^{-1}\ . \tag{81}$$

By invoking (79) and

$$\frac{\partial \nu}{\partial \theta} = \frac{\partial \nu}{\partial \mu}\,\frac{\partial \mu}{\partial \theta} = c\,\frac{\partial \mu}{\partial \theta}\ ,$$

(81) can be rewritten as

$$\tilde{k}\,\widetilde{\pi}_g(\mu|\sigma_\mu) = k_2\,\pi_h(\theta)\,\left|\frac{\partial \mu}{\partial \theta}\right|^{-1}\ ,$$

which, when divided by (80), finally yields (59). □

### *A.12. Proof of* THEOREM 8

It is enough to prove the Theorem for the location parameter $\mu$ of the sampling distribution of the form (9), where the dispersion parameter $\sigma$ is fixed, say $\sigma = 1$. Suppose, therefore, that we have collected a set $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ $(n \geq 2)$ of independent measurements from (9). By using the Consistency Theorem (34) for the assignment of a pdf to the inferred parameter $\mu$, and the Bayes' Theorem (33) for its sequential updating, we obtain

$$f(\mu|\mathbf{x}\sigma I_\circ) \propto \pi(\mu|\sigma)\,f(\mathbf{x}|\mu\sigma I_\circ)\ , \tag{82}$$

where

$$f(\mathbf{x}|\mu\sigma I_\circ) = \prod_{i=1}^{n} f(x_i|\mu\sigma I_\circ) = \prod_{i=1}^{n} \phi(x_i - \mu)\ ,$$

while the appropriate consistency factor $\pi(\mu|\sigma)$ is independent of either $\mu$ or $\sigma$ (54).

We introduce $n$ linear combinations, $\bar{x}$ and $\mathbf{s} = (s_1, \ldots, s_{n-1})$, of the measured set $\mathbf{x}$,

$$\bar{x} \equiv \frac{1}{n}\sum_{j=1}^{n} x_j \quad \text{and} \quad s_i \equiv x_i - \bar{x}\ ;\ i = 1, \ldots, n-1\ ,$$

so that

$$\begin{aligned} x_i - \mu &= \bar{x} - \mu + s_i\ ;\ \ i = 1, \ldots, n-1 \\ x_n - \mu &= \bar{x} - \mu - \sum_{j=1}^{n-1} s_j\ . \end{aligned}$$

Since

$$f(\mathbf{x}|\mu\sigma I_\circ) \propto f(\bar{x}\mathbf{s}|\mu\sigma I_\circ) \propto \phi\Big(\bar{x} - \mu - \sum_{j=1}^{n-1} s_j\Big) \cdot \prod_{i=1}^{n-1} \phi(\bar{x} - \mu + s_i) \tag{83}$$

the pdf (82) can be rewritten in terms of $\bar{x}$ and $\mathbf{s}$:

$$f(\mu|\bar{x}\mathbf{s}\sigma I_\circ) = \frac{\pi(\mu|\sigma)}{\zeta_\mu(\bar{x}, \mathbf{s}, \sigma)}\,f(\bar{x}\mathbf{s}|\mu\sigma I_\circ)\ .$$

According to the product rule (1), the likelihood density $f(\bar{x}\mathbf{s}|\mu\sigma I_\circ)$ can be decomposed as

$$f(\bar{x}\mathbf{s}|\mu\sigma I_\circ) = f(\mathbf{s}|\mu\sigma I_\circ)\,f(\bar{x}|\mathbf{s}\mu\sigma I_\circ)\ .$$

By introducing $u \equiv \bar{x}' - \mu$, it becomes obvious that $f(\mathbf{s}|\mu\sigma I_\circ)$ is independent of $\mu$,

$$f(\mathbf{s}|\mu\sigma I_\circ) = \int_{-\infty}^{\infty} f(\bar{x}'\mathbf{s}|\mu\sigma I_\circ)\, d\bar{x}' = \int_{-\infty}^{\infty} \phi\Big(u - \sum_{j=1}^{n-1} s_j\Big) \cdot \prod_{i=1}^{n-1} \phi(u + s_i)\, du = f(\mathbf{s}|\sigma I_\circ) \; ,$$

and can be included in the normalization constant $\zeta_\mu(\bar{x}, \mathbf{s}, \sigma)$:

$$f(\mu|\bar{x}\mathbf{s}\sigma I_\circ) = \frac{\pi(\mu|\sigma)}{\zeta_\mu(\bar{x}, \mathbf{s}, \sigma)}\, f(\mathbf{s}|\sigma I_\circ)\, f(\bar{x}|\mathbf{s}\mu\sigma I_\circ) \equiv \frac{\pi(\mu|\sigma)}{\widetilde{\zeta}_\mu(\bar{x}, \mathbf{s}, \sigma)}\, f(\bar{x}|\mathbf{s}\mu\sigma I_\circ) \; .$$

The remaining likelihood density, $f(\bar{x}|\mathbf{s}\mu\sigma I_\circ)$, is of the form

$$f(\bar{x}|\mathbf{s}\mu\sigma I_\circ) = \frac{f(\bar{x}\mathbf{s}|\mu\sigma I_\circ)}{f(s|\mu\sigma I_\circ)} \propto \phi\Big(\bar{x} - \mu - \sum_{j=1}^{n-1} s_j\Big) \cdot \prod_{i=1}^{n-1} \phi(\bar{x} - \mu + s_i) \equiv \widetilde{\phi}(\bar{x} - \mu, \mathbf{s}) \; ,$$

so that the updated pdf for $\mu$ reads

$$f(\mu|\bar{x}\mathbf{s}\sigma I_\circ) \propto \pi(\mu|\sigma)\, f(\bar{x}|\mathbf{s}\mu\sigma I_\circ) = \pi(\mu|\sigma)\, \widetilde{\phi}(\bar{x} - \mu, \mathbf{s}) \; . \tag{84}$$

The cdf of $\bar{x}$, given $\mu$, s and $\sigma$,

$$F(\bar{x}, \mu, \mathbf{s}, \sigma, I_\circ) = \int_{-\infty}^{\bar{x}} f(\bar{x}'|\mathbf{s}\mu\sigma I_\circ)\, d\bar{x}' = \int_{-\infty}^{\bar{x}} \widetilde{\phi}(\bar{x}' - \mu, \mathbf{s})\, d\bar{x}' \; ,$$

can be explicitly written as

$$F(\bar{x}, \mu, \mathbf{s}, \sigma, I_\circ) = \int_{-\infty}^{\bar{x}-\mu} \widetilde{\phi}(u, \mathbf{s})\, du \equiv \Phi(\bar{x} - \mu, \mathbf{s}) \; .$$

Then, equation (84) can be rewritten as

$$F_2(\bar{x}, \mu, \mathbf{s}, \sigma, I_\circ) = -\frac{\pi(\mu|\sigma)}{\widetilde{\zeta}_\mu(\bar{x}, \mathbf{s}, \sigma)}\, F_1(\bar{x}, \mu, \mathbf{s}, \sigma, I_\circ) \; ,$$

which, with the appropriate $\pi(\mu|\sigma) = \widetilde{\zeta}_\mu(\bar{x}, \mathbf{s}, \sigma) \propto 1$, satisfies the calibration condition (62).

Note that the updated pdf for $\mu$ is in general also a function of s, i.e., except for some very special sampling distributions like, for example, the Gaussian, $\bar{x}$ is *not a sufficient statistic* for $\mu$. However, the pdf is calibrated for *every* possible s, so that the proof of invariance of the calibration under updating is completed.

The preservation of calibration can also be connected to the preserved translation invariance of inference under updating. Namely, simultaneous location transformations

$$x_i \longrightarrow x_i + b \; ; \quad i = 1, \ldots, n \; ; \quad b \in (-\infty, \infty) \; , \tag{85}$$

imply

$$\bar{x} \longrightarrow \bar{x} + b \; ,$$

$$\mathbf{s} \longrightarrow \mathbf{s} \; ,$$

so that (83) is clearly invariant under simultaneous translations (85) of the measured values of the sampling variate, and of the inferred parameter,

$$\mu \longrightarrow \mu + b \; .$$

□

## References

Aczél, J. (1966). *Lectures on Functional Equations and Their Applications*. Academic Press.

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

Cauchy, A. (1897). *Œuvres Complètes – II$^{e}$ Série, Tome III*. Paris: Gauthier-Villars.

Courant, R. (1962). *Methods of Mathematical Physics, Vol. II – Partial Differential Equations*. Interscience Publishers.

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

Eadie, W. T., Drijard, D., James, F. E., Roos, M., and Sadoulet, B. (1971). *Statistical Methods in Experimental Physics*. North–Holland.

Elliot, J. P. and Dawber, P. G. (1986). *Symmetry in Physics, Vol. 1 – Principles and Simple Applications*. London: Macmillan.

Ferguson, T. S. (1967). *Mathematical Statistics – A Decision Theoretical Approach*. Academic Press.

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

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

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

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

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

Kass, R. E. and Wasserman, L. (1996). Formal Rules for Selecting Prior Distributions: A Review and Annotated Bibliography. *J. Amer. Statist. Assoc.*, **91**:1343–1370.

Kendall, M. G. (1949). On the reconciliation of theories of probability. *Biometrika*, **36**:101–116.

Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les évènemens. *Mem. Acad. Roy. Sci. Paris*, **6**:621–656.

Lindley, D. V. (1958). Fiducial Distributions and Bayes' Theorem. *J. R. Stat. Soc.*, **B 20**:102–107.

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

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.

Van Horn, K. S. (2003). Constructing a Logic of Plausible Inference: A Guide to Cox's Theorem. *Int. J. Approx. Reas.*, **34**:3–24.

Wigner, E. P. (1959). *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra*. Academic Press.