Impact of Large Likelihood Distribution Shift on Bayesian Estimation

Michael Hubig[1*], Holger Muggenthaler[2], and Gita Mall[3]

[1]Department of biomechanics, Institute of Legal Medicine, Jena University Hospital – Friedrich Schiller University Jena

[2]Head, Department of biomechanics, Institute of Forensic Medicine at the University Hospital Jena of the Friedrich Schiller University Jena

[3]Director, Institute of Forensic Medicine at the University Hospital Jena of the Friedrich Schiller University Jena

Abstract

A major problem arising frequently in Bayesian estimation (BE) is to cope with a shift (e.g. a bias) of the likelihood probability. This paper investigates the impact of a severely shifted continuous likelihood probability on the output of BE in the real numbers. The result can be interpreted as a classification of asymptotic conditional probability distributions induced on a bounded interval by the far tail of probability distributions. It can be applied to a wide range of probability distributions. A hypothetical example of BE in death time estimation in a homicide trial, which is designed similar to a real case, illustrates the practical relevance of the results.

Keywords: Bayesian estimation; Likelihood shift; Bias as shift; Asymptotic classification; Auto Asymptotic Limit (AAL)

[*]Corresponding author: Michael Hubig, Mathematician, Department of biomechanics, Institute of Legal Medicine, Jena University Hospital – Friedrich Schiller University Jena, Fürstengraben 23 07743 Jena, Germany, Tel: +49-3461-935551; Fax: +49-3461-935552; E-mail: Michael.Hubig@med.uni-jena.de

1.Introduction

Bayesian estimation is a well known technique in applied science and as such was addressed by a large number of theoretic studies. Besides notorious problems in prior determination, the method is subjected to errors in the likelihoods. One class of likelihood error is induced by bias or shift in the argument of the likelihood leading to a shift of the functional graph as well as of the expectation. The present study investigates the effect of likelihood shifts. It focuses on the case of large shifts of an estimators error probability distribution (pd) on the real numbers and deals with the question, which conditional pd is induced on a bounded interval in the far tail of the estimators error pd. For a large pd class, containing many important pds for application problems, the induced conditional pds on a bounded interval in the far tail are computed and classified.

Our study is motivated by a forensic science expertise in a murder trial where a Bayesian technique called Conditional Probability Distribution (CPD), published by Biermann and Potente [1], was used to include witness reports into confidence interval estimation for temperature based death time estimation. We published a study [2] on the errors caused by large input biases in the CPD method, dealing with the special case of the Gaussian distribution. The present article generalizes the asymptotic result of [2] to a wide spectrum of distributions and gives a classification theorem.

2.Terminology

Our considerations use the terminology of general probability theory, which defines the general conditional probability $P(A|B)$ of an event A, given an event B, where A and B are sets or families of sets of a common probability space O which is never mentioned explicitly. The sets A and B are specified by stating arithmetical or analytical formulae for random variables e.g. "A = {w in O | t(w)=a}" means, that A comprises all w in O for which the random variable t takes the value a. We adopt the convention to write e.g. P(t^ | t) with two random variables t and t^ given for a type of probability distribution of the random variable t^ under a fixed but not specified condition given by an arithmetic statement (equivalent to a set A in O) concerning the random variable t as e.g. A = {w in O | t(w)=a}. If the condition is specified, we write e.g. t=a abbreviating the term for the event A above. For basic terminology see e.g. the textbook of Papoulis [3].

2.1 Bayesian Estimation

Let t be a random variable taking values in the real numbers IR and let t^ be an estimator of t. The pd of t^ given the true value of the variable t is denoted in a somewhat sloppy style by P(t^ | t) and is called the likelihood as usual in Bayesian estimation. Generally we will use this sloppy notation and specify values only if the reader would be mislead else. The probability distribution density (pdd) of t^, given the true value of the variable t, is labelled f(t^ | t) and will be called likelihood density.

The (unconditional) pds of the variables t and t^ are named the prior P(t) and the marginal distribution P(t̂) and we use g(t) and h(t^) as symbols for their pdds. We can now introduce the pdd f(t | t^) of the posterior P(t | t^). Since the posterior density f(t | t^) belongs to a conditional pd, it may be written:

$$f\left(t\,|\,t^{\,\hat{}}\right)=f\left(t^{\,\hat{}}\,|\,t\right)\cdot g\left(t\right)\Big/\,h\left(t^{\,\hat{}}\right) \qquad (1)$$

The purpose of Bayesian estimation (BE) is to compute an estimator of the posterior f(t | t^) given an expression f(t^ | t) for the likelihood and another expression for the prior g(t). Expressing h(t^) as the integral of f(t^ | t) g(t) over all t makes it possible to eliminate h(t^) from (1) yielding:

$$f\left(t\,|\,t^{\,\hat{}}\right)=f\left(t^{\,\hat{}}\,|\,t\right)\cdot g\left(t\right)\Big/\left[\int_{IR}f\left(t^{\,\hat{}}\,|\,t\right)\cdot g\left(t\right)dt\right] \qquad (2)$$

Let E(t^ | t = t$_0$) be the conditional expectation of the variable t^ given a value t0 of the variable t in case of existence of the expectation with respect to the likelihood pdd f(t^ | t = t$_0$) given a fixed value t$_0$ of the variable t. We will frequently omit the assignment of the value " = t$_0$", sometimes the condition "t = t$_0$" as a whole and sometimes even the argument brackets and the symbol t^ – writing simply E^ in cases where no confusion can occur. The symbol f(t^ | E^, t) will be used in case E(t^ | t) exists and if we want to emphasize or use this fact. Sometimes we even skip the argument t^ and write f(E^, t) or f(t) for f(t^ | E^, t). The symbols f(t^| t) or f(t^) will be used for short in more general considerations, not depending on the existence of E(t^ | t). Writing g(t) for the prior pdd we are now able to represent (2) in a form, where the conditional expectation E(t^ | t) explicitly appears:

$$f\left(t\,|\,t^{\,\hat{}}\right)=f\left(t^{\,\hat{}}\,|\,E^{\,\hat{}},t\right)\cdot g\left(t\right)\Big/\left[\int_{IR}f\left(t^{\,\hat{}}\,|\,E^{\,\hat{}},t\right)\cdot g\left(t\right)dt\right] \qquad (3)$$

2.2 Shift Families of Probability Densities

Let f: IR→IR be a pdd on the real numbers IR and let Δt be any real number. Let further be f$_{\Delta}$t:(s → f(s + Δt)): IR → IR, the Δt–shift of f and Γ be the shift family of f or the location family of f, which is the set of all possible Δt–shifts of f:

$$\Gamma:=\left\{\,f_{\Delta t}\,|\,\Delta t\in IR\right\} \qquad (4)$$

Let E be the expectation of f, then we yield the expectation E$_{\Delta t}$ of any f$_{\Delta t}$ by straight forward computation and via substitution:

$$E_{\Delta t}=E-\Delta t \qquad (5)$$

So each f$_{\Delta t}$ is uniquely determined by its expectation E$_{\Delta t}$ = E – Δt and therefore the dependency on the expectation E$_{\Delta t}$ can be used to reparameterize the whole family Γ. This gives rise to a change of terminology integrating the expectation E$_{\Delta t}$ as second argument in the symbol for every family member f$_{\Delta t}$ of Γ:

$$\forall s\in IR:f\left(s\,|\,E_{\Delta t}\right):=f_{\Delta t}\left(s\right) \qquad (6)$$

As a direct consequence of (5) and (6) we have the following formula, which allows to push the shift Δt from the second to the first argument of f changing Δt's sign:

$$\forall s\in IR:f\left(s\,|\,E-\Delta t\right)=f\left(s\,|\,E_{\Delta t}\right)=f_{\Delta t}\left(s\right)=f\left(s+\Delta t\,|\,E\right) \qquad (7)$$

2.3 Auto Asymptotic Limits

In the following we present some general assumptions to specify a certain class A of pdds f, defined on open infinite intervals in the real numbers, which is investigated here. We use the symbol "↓↓" to express "strictly monotonic decreasing" and "↑↑" to represent "strictly monotonic increasing" and "↕↕" to denote strict monotonicity.

A pdd f, defined on an open infinite interval J in the real numbers IR is an element of the class A, iff there is a positive real number R with:

(A1) The function f is continuous on the rest J –]–R,R[of its range J with the interval]–R,R[removed.

(A2) (a) If +∞ lies in J, then f is strictly monotonic decreasing to 0 on the interval J ∩]R,+∞[:

$$x\uparrow\uparrow+\infty\Rightarrow f\left(x\right)\downarrow\downarrow 0$$

b) If −∞ lies in J, then f is strictly monotonic decreasing to 0 on the interval J ∩ [−∞, -R[:

$$x\downarrow\downarrow-\infty\Rightarrow f\left(x\right)\downarrow\downarrow 0$$

(A3) f is differentiable and f' / f is continuous on J – [-R, R] and fulfills the condition:

(a) If +∞ lies in J, then f' / f is strictly monotonic on J ∩]R,+∞[:

$$x\uparrow\uparrow+\infty\Rightarrow f'\left(x\right)\Big/f\left(x\right)\updownarrow\updownarrow$$

(b) If −∞ lies in J, then f' / f is strictly monotonic on J ∩]-∞, -R[:

$$x\downarrow\downarrow-\infty\Rightarrow f'\left(x\right)\Big/f\left(x\right)\updownarrow\updownarrow$$

2.3.1 Remark

Condition (A3)(a) is equivalent to the existence of a positive constant e and a function α: J ∩]R,+∞[→ IR which is continuous and strictly monotonic and fulfills:

$$\forall x\in J\cap\,]\,R,+\infty[\,:f\left(x\right)=e\cdot exp\left(\int_{R}^{x}\alpha\left(s\right)ds\right)$$

The fact that f is a pdd implies an upper boundary to the descending rate of the integral over α.

During the investigation of the impact of large likelihood shifts on the result of BE a certain quantity – we call it the auto asymptotic limit – evolves, which measures the asymptotic decline intensity of a pdd on an unbounded interval J of the real numbers IR. We define:

2.3.2 Definition

Let J be an unbounded interval in the real numbers IR. The symbol $x \to \underline{+}\infty$ means $x \to +\infty$ in J if J is unbounded in positive direction and means $x \to -\infty$ in J if J is unbounded in negative direction. In case J = IR the symbol is used for both limit processes alternatively. Let f: J → IR be a pdd on J. The auto-asymptotic a[f] (AA) of the pdd f is defined:

$$\forall x \in J : \forall x + \delta x \in J : a[f](x, \delta x) := f(x + \delta x)/f(x) \quad (8)$$

The auto-asymptotic limit $A\underline{\pm}[f](\delta x)$ (AAL) is defined in case of existence as:

$$\left\{\begin{matrix} If +\infty \in J \\ If -\infty \in J \end{matrix}\right\} : \forall \delta x \in IR : A_{\pm}[f](\delta x) := \left\{\begin{matrix} \lim_{x \to +\infty} \\ \lim_{x \to -\infty} \end{matrix}\right\} a[f](x, \delta x) = \left\{\begin{matrix} \lim_{x \to +\infty} \\ \lim_{x \to -\infty} \end{matrix}\right\} f(x + \delta x)/f(x) \quad (9)$$

It is well known (see Pericchi and Sanso [4]) that the following proposition on AALs can be derived directly from their definition:

2.3.3 Proposition

For the pdd f on J there are λ_+ in $IR_{\geq 0}$ U {∞} and λ_- in $IR_{\leq 0}$ U {-∞} with: In case of existence:

$$\forall \delta x \in IR : A_{\pm}[f](\delta x) = exp(-\lambda_{\pm}\delta x) \quad (10)$$

The following remarks are direct consequences of proposition 2.3.3 and of definition 2.3.2.

2.3.4 Remark

For a pdd f in the class A defined by (A1), (A2), (A3):

(a) $\forall \delta x \in IR \cup \{+\infty, -\infty\} : A_{\pm}[f](\delta x) = 1/A_{\pm}[f](-\delta x) \quad (11)$

(b) $\forall \delta x > 0 : 0 \leq A_+[f](\delta x) < 1 < A_-[f](\delta x)$
$\forall \delta x < 0 : 0 \leq A_-[f](\delta x) < 1 < A_+[f](\delta x) \quad (12)$

(c) $\exists \delta x \neq 0 : A_{\pm}[f](\delta x) = 0 \Rightarrow \forall \delta x' \neq 0 : A_{\pm}[f](\delta x') = 0 \quad (13)$

(d) $\delta x = 0 \Rightarrow A_{\pm}[f](\delta x) = 1 \quad (14)$

2.3.5 Examples

A) The normal distribution $\nu(\mu, \sigma^2)$

$$\forall x \in IR : \nu(\mu, V)(x) := \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\forall x \in IR : \nu(\mu, \sigma^2)'(x) := \nu(\mu, \sigma^2)(x)\left(-\frac{x-\mu}{\sigma^2}\right)$$

$$\forall x \in IR : \alpha\left[\nu(\mu, \sigma^2)\right](x) := -\frac{x-\mu}{\sigma^2}$$

$$\forall d \in IR_{>0} : \forall x \in IR : a\left[\nu(\mu, \sigma^2)\right](x, \pm d) = exp\left(\mp\frac{d}{\sigma^2}x\right)exp\left(\pm\frac{\mu d}{\sigma^2} - \frac{d^2}{2\sigma^2}\right)$$

$$\forall d \in IR_{>0} : A_{\pm}\left[\nu(\mu, \sigma^2)\right](d) = 0$$

B) The log-normal distribution L(ν):

$$\forall x > 0 : L[\nu](x) = \frac{1}{\sqrt{2\pi V}\,x}exp\left(-\frac{1}{2\sigma^2}(ln(x) - M)^2\right)$$

$$\forall x > 0 : L[\nu]'(x) = L[\nu](x)\cdot\frac{1}{x}\left[\frac{M - ln(x)}{V^2} - 1\right]$$

$$\forall x > 0 : \alpha[L](x) = \frac{1}{x}\left[\frac{M - ln(x)}{V^2} - 1\right]$$

$$\forall d \in IR : \forall x > |d| :$$

$$a[L](x, d) = \frac{x}{x+d}\frac{exp\left(-\frac{1}{2\sigma^2}(ln(x+d) - \mu)^2\right)}{exp\left(-\frac{1}{2\sigma^2}(ln(x) - \mu)^2\right)}$$

$$= \frac{x}{x+d}exp\left(-\frac{1}{2\sigma^2}\left(ln(x+d)^2 - ln(x)^2 - 2\mu ln\left(\frac{x+d}{x}\right)\right)\right)$$

$$\forall d \in IR :$$

$$A_+[L](d) = \lim_{x \to \infty}\frac{x}{x+d}exp\left(-\frac{1}{2\sigma}\left(ln(x+d)^2 - ln(x)^2 - 2\mu ln\left(\frac{x+d}{x}\right)\right)\right)$$

$$= 1$$

C) The Gamma distribution $\Gamma_{a,b}$:

$$\forall x > 0 : \Gamma_{a,b}(x) = \frac{1}{a^b \cdot \Gamma(b)}\cdot x^{b-1}\cdot exp\left(-\frac{x}{a}\right)$$

$$\forall x > 0 : \Gamma_{a,b}'(x) = \Gamma_{a,b}(x)\cdot\left[\frac{(b-1)}{x} - \frac{1}{a}\right]$$

$$\forall x > 0 : \alpha\left[\Gamma_{a,b}\right](x) = \frac{(b-1)}{x} - \frac{1}{a}$$

$$\forall d \geq 0 : \forall x > 0 : a\left[\Gamma_{a,b}\right](x, d) = \left(\frac{x+d}{x}\right)^{b-1}exp\left(-\frac{d}{a}\right)$$

$$\forall d \geq 0 : A_+\left[\Gamma_{a,b}\right](d) = exp\left(-\frac{d}{a}\right)$$

D) The t-distribution t_n

$$\forall x \in IR : t_n(x) := \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)}\cdot\frac{1}{\left(1 + \frac{x^2}{n}\right)^{(n+1)/2}}$$

$$\forall x \in IR : t_n'(x) = t_n(x)\cdot\left(-\frac{(n+1)\cdot x}{n+x^2}\right)$$

$$\forall x \in IR : \alpha[t_n](x) = -\frac{(n+1)\cdot x}{n+x^2}$$

$$\forall d \geq 0 : \forall x \in IR : a[t_n](x, d) = \left(\frac{1 + \frac{(x\pm d)^2}{n}}{1 + \frac{x^2}{n}}\right)^{\frac{n+1}{2}}$$

$$\forall d \geq 0 : A_{\pm}[t_n](d) = 1$$

E) The Laplace distribution La:

$$\forall x \in IR : La(x) := \frac{1}{2} \cdot exp(-|x|)$$

$$\forall x \in IR : La'(x) = La(x) \cdot (\mp 1)$$

$$\forall x \in IR : \alpha[La](x) = -1$$

$$\forall x \in IR : \forall d \ge 0 : a[La](x, \pm d) = exp(-|x \pm d| + |x|)$$

$$\forall d \ge 0 : A_\pm[La](d) = exp(\mp d)$$

F) The logistic distribution lc:

$$\forall x \in IR : lc(x) := \frac{exp(-x)}{(1 + exp(-x))^2}$$

$$\forall x \in IR : lc'(x) = lc(x) \cdot \frac{exp(-x) - 1}{1 + exp(-x)}$$

$$\forall x \in IR : \alpha[lc](x) = \frac{exp(-x) - 1}{1 + exp(-x)}$$

$$\forall x \in IR : \forall d \ge 0 : a[lc](x, \pm d) = \left(\frac{1 + exp(-x \mp d)}{1 + exp(-x)}\right)^2 exp(\mp d)$$

$$\forall d \ge 0 : A_\pm[lc](d) = exp(\mp d)$$

We will show now that every function f in the function class A, defined by the conditions (A1), (A2), (A3) above, guarantees the existence of the AALs.

2.3.6 Remark

Let J be an interval in IR containing $+\infty$ and / or $-\infty$. Let f: J $\longrightarrow$ IR be a pdd with (A1), (A2), (A3). This implies for all δx in IR the unambiguous existence of the AAL $A_\pm[f](\delta x)$ in $IR_{\ge 0}$ U $\{+\infty\}$.

Remark 2.3.6 is proven in Appendix A.

3. Large Shifts

The following proposition investigates the case of large likelihood shifts which is caused e.g. by large bias. It provides asymptotic formulae for the probability $P_{\Delta t}(t \in [a,b] | t \in [c,d])$ in case $\Delta t \to \pm\infty$.

3.1 Proposition

Let J be an open interval in the real numbers IR and let be $+\infty$ in J or $-\infty$ in J. Let further be f: J $\longrightarrow$ IR a pdd which fulfills (A1), (A2), (A3) and let $[c,d]$ be a nonempty interval in the real numbers IR and $[a,b]$ a partial interval of $[c,d]$. For all Δt in IR let $f_{\Delta t}(t) := f(t + \Delta t)$ and let $P_{\Delta t}(t \in [a,b] | t \in [c,d])$ be the conditional probability of t in $[a,b]$ under the condition t in $[c,d]$ with respect to the pdd $f_{\Delta t}$. Let $\lambda\pm$ be the factor in the exponent of the AAL $A_\pm[f](\delta t)$ from (10). The following expressions can be derived for the asymptotic probability $\lim_{\Delta t \to \pm\infty} P_{\Delta t}(t \in [a,b] | t \in [c,d])$:

Case (1): One of the following alternative conditions (15) shall be fulfilled:

$$\left.\begin{array}{l} A_+[f](d-c) \ne 1 \\ A_-[f](d-c) \ne 1 \end{array}\right\} \quad (15)$$

Case (1a): If additionally to (15) the matching condition of the alternatives (16) is realized:

$$\left.\begin{array}{l} \forall \delta t > 0 : A_+[f](\delta t) \ne 0 \\ \forall \delta t < 0 : A_-[f](\delta t) \ne 0 \end{array}\right\} \quad (16)$$

we yield:

$$\lim_{\Delta t \to +/-\infty} P_{\Delta t}(t \in [a,b] | t \in [c,d]) = \frac{exp(-\lambda_\pm b) - exp(-\lambda_\pm a)}{exp(-\lambda_\pm d) - exp(-\lambda_\pm c)} \quad (17)$$

Case (1b): If else additionally to (15) a matching alternative condition of (18) is given:

$$\left.\begin{array}{l} \forall \delta t > 0 : A_+[f](\delta t) = 0 \\ \forall \delta t < 0 : A_-[f](\delta t) = 0 \end{array}\right\} \quad (18)$$

The following expression represents our probability of interest:

$$\lim_{\Delta t \to +/-\infty} P_{\Delta t}(t \in [a,b] | t \in [c,d])$$
$$= \left\{\begin{array}{ll} 0 & if \left[(\Delta t \to \pm\infty) \wedge (b \ne d) \wedge (a \ne c)\right] \\ 1 & if \left[(\Delta t \to +\infty) \wedge (a = c)\right] \vee \left[(\Delta t \to -\infty) \wedge (b = d)\right] \end{array}\right\} \quad (19)$$

Case (2): If in contraposition to (15) one of the following conditions (20) is valid:

$$\left.\begin{array}{l} A_+[f](d-c) = 1 \\ A_-[f](d-c) = 1 \end{array}\right\} \quad (20)$$

the conditional probability of the true value of t lying in a partial interval $[a,b]$ of the greater interval $[c,d]$ under the additional condition of the true value of t lying in $[a,b]$ can now be rewritten as:

$$\lim_{\Delta t \to +/-\infty} P_{\Delta t}(t \in [a,b] | t \in [c,d]) = \frac{b-a}{d-c} \quad (21)$$

Proposition 3.1 is proven in appendix B.

The results developed in proposition 3.1 are easily applicable to the case of estimator biases in BE with constant priors:

3.2 Remark

Let the presuppositions of proposition 3.1 be given for the conditional pdd $f(t^\wedge | t)$ of the unbiased estimator $t^\wedge$ of a real value t. Let for every real number Δt the shifted pdd $f_{\Delta t}$ be the pdd of an estimator $t^\wedge_{\Delta t}$, which therefore has a bias of Δt. Let $P_{\Delta t}(t \in [a,b] | t^\wedge, t \in [c,d])$ – which is computed using the shifted pdd $f_{\Delta t}$ – be the conditional probability of t in $[a,b]$ under the conditions of t in $[c,d]$ and the estimation value $t^\wedge$. The values of the variable t are realized by the transformation $t = t_0 - s$, where t_0 is an arbitrarily fixed value of t and s is a matching real random variable. This leads via application of (7) and the definition $v := t^\wedge + s$ to:

$$P_{\Delta t}\left(t \in [a,b] \,|\, t\hat{}\,,\ t \in [c,d]\right) := \frac{\int\limits_a^b f\left(t\hat{}\,|\,t-\Delta t\right)\,dt}{\int\limits_c^d f\left(t\hat{}\,|\,t-\Delta t\right)\,dt}$$

$$= \frac{-\int\limits_{t_0-a}^{t_0-b} f\left(t\hat{}\,|\,t_0-s-\Delta t\right)\,ds}{-\int\limits_{t_0-c}^{t_0-d} f\left(t\hat{}\,|\,t_0-s-\Delta t\right)\,ds} = \frac{-\int\limits_{t_0-a}^{t_0-b} f\left(t\hat{}+s\,|\,t_0-\Delta t\right)\,ds}{-\int\limits_{t_0-c}^{t_0-d} f\left(t\hat{}+s\,|\,t_0-\Delta t\right)\,ds} \qquad (22)$$

$$= \frac{\int\limits_{t_0-t\hat{}-b}^{t_0-t\hat{}-a} f\left(v\,|\,t_0-\Delta t\right)\,dv}{\int\limits_{t_0-t\hat{}-d}^{t_0-t\hat{}-c} f\left(v\,|\,t_0-\Delta t\right)\,dv} = \frac{\int\limits_{t_0-t\hat{}-b}^{t_0-t\hat{}-a} f\left(v+\Delta t\,|\,t_0\right)\,dv}{\int\limits_{t_0-t\hat{}-d}^{t_0-t\hat{}-c} f\left(v+\Delta t\,|\,t_0\right)\,dv}$$

$$=: P_{\Delta t}\left(v \in [t_0-t\hat{}-b, t_0-t\hat{}-a]\,|\,v \in [t_0-t\hat{}-d, t_0-t\hat{}-c]\right)$$

Proposition 3.1 may now be applied to the probability $P_{\Delta t}(t \in [a,b] \,|\, t\hat{}\,, t \in [c,d])$, where v plays the role of t and the interval limits a, b, c, d are realized by: $t_0 - t\hat{} - b$, $t_0 - t\hat{} - a$, $t_0 - t\hat{} - d$, $t_0 - t\hat{} - c$. Note that in all cases the results do not depend on t_0 or on $t\hat{}$.

4 Application Example: Forensic Death Time Estimation

Death time estimation (DE) means reconstruction of the time difference t_D between real death time t and time t_M of measuring a value X_M of a quantity $X = X(s, \theta)$ – e.g. in case of temperature based DE (TDE) the rectal temperature $T_M = X_M$ – which monotonously depends on time s and whose value $X_0 = X(t)$ at time of death t is known. The parameter θ refers to any vector of measurable quantities (e.g. in the TDE approach of Marshall and Hoare [5] with the parameter definition of Henssge [6,7] the rectal temperature T_0 at death time, the ambient temperature T_A, the body mass m, and Henssges corrective factor cf), influencing the time evolution of the quantity X. Reconstruction is performed by solving the following equation system for t_D:

$$X_0 = X(t)$$
$$X_M = X(t_M) = X(t + t_D) \qquad (23)$$

The time of death estimator $t\hat{}$ can now easily be calculated by computing:

$$t\hat{} := t_M - t_D \qquad (24)$$

Since DE uses real world measurements, it is prone to measurement errors of its input variables θ, t_M, X_M and X_0 and to systematic errors of the model X(t) used. To cope with the resulting errors of the death time estimator $t\hat{}$, one usually represents $t\hat{}$ as a random variable associated with a pd $P(t\hat{})$ or with its pdd $f(t\hat{})$ respectively. Since the pd $P(t\hat{})$ is nearly always determined under the assumption of a fixed real value of death time t, one refers to the conditional pd $P(t\hat{} \,|\, t)$ of the estimator $t\hat{}$ under the condition of real death time t assuming a particular, though usually unknown, time value. This is the distribution most frequently stressed in application cases. The pd $P(t\hat{} \,|\, t)$ is also used as likelihood distribution for BE.

4.1 Temperature Based Death Time Determination

Choosing a special body temperature T as the quantity X leads to temperature based death time estimation (TDE). The most frequently used method of temperature based death time determination is the model based approach of Marshall and Hoare [4] with the parameter definition of Henssge [5,6]. It will be referred to here for short as the Henssge method. Relying on the central limit theorem it is implicitly taken for granted that the death time estimator $t\hat{}$ of the Henssge method is associated with a Gaussian distribution as its conditional pd $P(t\hat{} \,|\, t)$:

$$\forall [r,s] \subseteq IR : P\left(t\hat{} \in [r,s]\,|\,t\right) = \int\limits_r^s \frac{1}{\sqrt{2\pi V}} exp\left(-\frac{\left(t\hat{}-t\right)^2}{2V}\right) dt\hat{} = N\left(\frac{s-t}{\sqrt{V}}\right) - N\left(\frac{r-t}{\sqrt{V}}\right) \quad (25)$$

with the associated conditional pdd $f(t\hat{} \,|\, t)$:

$$f\left(t\hat{}\,|\,t\right) = \frac{1}{\sqrt{2\pi V}} exp\left(-\frac{\left(t\hat{}-t\right)^2}{2V}\right) \quad (26)$$

Where t is the true death time value, N is the probability function of the standard normal distribution or Gaussian, V is the variance value and $[r,s]$ is any interval in the real numbers IR. It should be emphasized here, that the assumptions of (25) and (26) imply the estimator $t\hat{}$ to be unbiased, which means, that the expected value of $P(t\hat{} \,|\, t)$ is the true value t:

$$E\left(t\hat{}\,|\,t\right) = t \quad (27)$$

Usually there is only one temperature value T_M measured at a time t_M. Therefore the estimated value $t\hat{} = t_M - t_D(T_M, t_M)$ has to be taken as the only valid estimator of the expectation in formula (25) for the likelihood. In practice the following likelihood formula is used:

$$f\left(t\hat{}\,|\,t\right) \approx \frac{1}{\sqrt{2\pi V}} exp\left(-\frac{\left(t\hat{}-t_M+t_D\left(T_M,t_M\right)\right)^2}{2V}\right) \qquad (28)$$

To compute the expression (3) for the application of the CPD‑method.

4.2 Influence of Biased Death Time Estimators on Bayesian Estimation

We now assume the scenario of the time interval $[c,d]$ with the particular partial interval $[a,b]$ on the time axis and ask for the posterior probability $P(t \in [a,b] \,|\, t\hat{}\,, t \in [c,d])$ of the true death time t lying in the partial interval $[a,b]$ of the time interval $[a,b]$ under the conditions that the estimator $t\hat{}$ takes a fixed value and the true death time lying in $[c,d]$. The limit c of the last condition may be motivated by testimonies of witnesses who had seen the deceased at time c alive whereas the upper limit d is the time where the body was found. The question for $P(t \in [a,b] \,|\, t\hat{}\,, t \in [c,d])$ can be answered by a BE.

In cases where the conditional likelihood distribution $P(t\hat{} \,|\, t)$ used for BE is biased, the results are distorted in the typical way described in paragraph 3. Integration of the pdd $f(t\hat{} \,|\, t - \Delta t)$ over the intervals $[a, b]$ and $[c, d]$ of the true time of death t and taking the quotient and assuming a constant prior on $[c, d]$ yields the conditional probability $P_{\Delta t}(t \in [a,b] \,|\, t\hat{}\,, t \in [c,d])$, which bears the index Δt to remind the reader of the existence of the likelihood probabilities' bias:

$$P_{\Delta t}\left(t \in [a,b] \mid t^{\hat{}}, t \in [c,d]\right) = \frac{\displaystyle\int_a^b exp\left(-\frac{\left((t-\Delta t)-t^{\hat{}}\right)^2}{2V}\right)dt}{\displaystyle\int_c^d exp\left(-\frac{\left((t-\Delta t)-t^{\hat{}}\right)^2}{2V}\right)dt} \qquad (29)$$

Assuming a large bias Δt, we have to take into account the fact that with example (A) we have $A_+[v(E,V)](\delta t)$ = 0 for all $\delta t > 0$ and $A_-[v(E,V)](\delta t)$ = 0 for all $\delta t < 0$. We apply formula (19) and yield:

$$\lim_{\Delta t \to +/-\infty} P_{\Delta t}\left(t \in [a,b] \mid t^{\hat{}}, t \in [c,d]\right)$$
$$= \begin{cases} 0 & if \left[(\Delta t \to \pm\infty) \wedge (b \neq d) \wedge (a \neq c)\right] \\ 1 & if \left[(\Delta t \to +\infty) \wedge (a=c)\right] \vee \left[(\Delta t \to -\infty) \wedge (b=d)\right] \end{cases} \qquad (30)$$

4.3 Bayesian Estimation in Temperature Based Death Time Estimation

The following hypothetical example (E) was constructed to show the power and the risks of the CPD [3] in case of its application to TDE in a court hearing of a homicide charge. Example (E) is similar to a real homicide case where we were took part as additional experts of the defense counsel.

Example (E)

There is non-temperature information from testimonies that the deceased was still alive at a time c = 10:00 a.m. and his body was found at d = 4:00 p.m. the same day, which makes the time of death lying in the interval [c,d] with a 100% probability a priori. The prime suspect has no alibi in a time interval [a,b] lying in the 100%-interval [c,d]. For a TDE approach there is a measurement consisting of the measurement time t_M, the ambient temperature T_A and the rectal temperature T_M at t_M. Backcalculation using Henssges TDE yields a time difference t_D between death and measurement which leads to an estimated time of death $t^{\wedge}$: Now the judge asks for the probability of the real time of death lying in the interval [a,b] with respect to the back calculation value $t^{\wedge}$. In our terminology this is the conditional probability P(t ∈ [a,b] | $t^{\wedge}$, t ∈ [c,d]). As usual the likelihood distribution P($t^{\wedge}$ | t) of the TDE result $t^{\wedge}$, given the fixed time of death t, is assumed to be a Gaussian with expectation E = t and variance V. Let us assume Henssge's TDE yields a 95% confidence interval radius r = 2.8 h, which results in a variance V = $(r / 2)^2$ = 1.96 h² of the likelihood distribution P($t^{\wedge}$ | t).

Let us assume further that there are two alternatives for the alibi time interval [a,b]:

(A)   Between a = 10:00 a.m. and b = 11:00 a.m.
(B)   Between a = 10:30 a.m. and b = 11:30 a.m.

There is only one back calculation value $t^{\hat{}}$, but we present three versions (1), (2), (3) of this value in three cases of an estimator bias e.g. by measurement errors of the ambient temperature T_A, – of the rectal temperature T_0 at time of death or by nonstandard conditions, which were not taken into account by choosing an adequate corrective factor cf. In each one of the three alternatives, TDE backcalculation produces an estimated value $t^{\wedge}$, which lies outside the interval [c,d]:

(1) $t^{\wedge}$ = 9:00 a.m.
(2) $t^{\wedge}$ = 7:00 a.m.
(3) $t^{\wedge}$ = 5:00 a.m.

We compute the conditional probability P = $P_{\Delta t}$(t ∈ [a,b] | $t^{\wedge}$, t ∈ [c,d]) for each of the six possible combinations of [a,b] and $t^{\hat{}}$ without taking into account the estimators bias Δt which is mostly unknown in real casework. This yields the following results:

(A)^(1) $a = 10:00 a.m., b = 11:00 a.m., t^{\hat{}} = 9:00 \Longrightarrow P = 0.678$
(A)^(2) $a = 10:00 a.m., b = 11:00 a.m., t^{\hat{}} = 7:00 \Longrightarrow P = 0.867$
(A)^(3) $a = 10:00 a.m., b = 11:00 a.m., t^{\hat{}} = 5:00 \Longrightarrow P = 0.949$
(B)^(1) $a = 10:30 a.m., b = 11:30 a.m., t^{\hat{}} = 9:00 \Longrightarrow P = 0.442$
(B)^(2) $a = 10:30 a.m., b = 11:30 a.m., t^{\hat{}} = 7:00 \Longrightarrow P = 0.346$
(B)^(3) $a = 10:30 a.m., b = 11:30 a.m., t^{\hat{}} = 5:00 \Longrightarrow P = 0.231$

It is of some interest to have a look at the erroneous posterior probability P = $P_{\Delta t}$(t ∈ [a,b] | $t^{\wedge}$, t ∈ [c,d]) for different values of the bias Δt. Since the bias is unknown in this situation, we present plots of the probability P as functions of Δt for scenario (A) in figure 1 and for scenario (B) in figure 2 where we assume a 'true value' (which means an unbiased value) of $t^{\wedge}$ = 9:00 a.m. to be able to compute P.

In scenario (A) the P value falls from 1 down to 0 in an S-shaped curve crossing the x-axis at 0.678 while the Δt values rise from −10 h up to 10 h (figure 1). The result of scenario (B) over an interval [−8 h, 8 h] of biases Δt is a peak of P rising moderately from left to right from ca. P = 0.1 to a maximum value of P = 0.45 at Δt = 1 h. The curve then takes a steeper course down to a value of ca. P = 0.005 at Δt = 8 h (figure 2).

The figures 1 and 2 illustrate the two types of scenarios possible for the long range biases:

(A) Equal lower limits a = c of the two intervals [a,b] and [c,d]:

$P_{\Delta t}$ *rises for lower negative* values of the bias Δt to 1.

$P_{\Delta t}$ *falls for higher positive* values of the bias Δt to 0.

(B) Different lower limits a > c and b < d of the two intervals [a,b] and [c,d]:

$P_{\Delta t}$ *falls for higher positive and for lower negative* values of the bias Δt to 0.

$P_{\Delta t}$ *reaches a maximum* for a value of the bias Δt near 0 h.

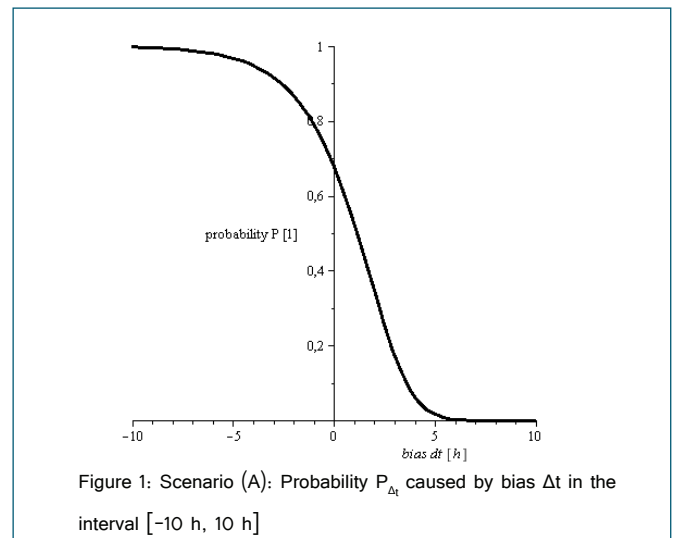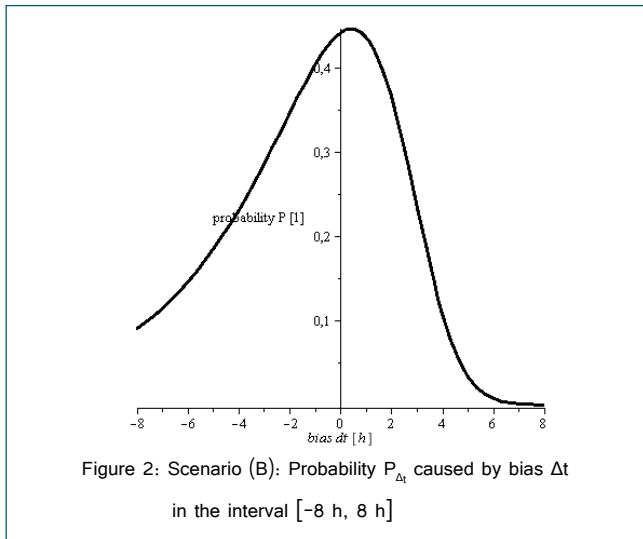Figure 1: Scenario (A): Probability $P_{\Delta t}$ caused by bias Δt in the interval [−10 h, 10 h]

Figure 2: Scenario (B): Probability P_{Δ_t} caused by bias Δt

in the interval $[-8\,h,\,8\,h]$

5 Discussion

This paper presents an investigation of the influence of likelihood shifts on the results of BE. It was inspired by a work [1] of Biermann and Potente from 2011 who intended their Bayesian approach – which they called Conditional Probability Distribution (CPD) method – to calculate probabilities for time intervals in temperature based death time determination.

We analyze the important cases of large systematic errors in the BE input. Systematic errors which result in biases Δt can lead to major deviations of the estimated probabilities. The analysis of the large bias cases reveals that the set A of all pdds investigated (for a definition see: (A1), (A2), (A3)) is divided into three disjoint subsets according to three families of asymptotic limit distributions which are in a one to one correspondence to the three possible range sets for the AAL-values: $A\pm[f](d-c)$ in $\{0\}$, $\{1\}$, $]0,1[$ U $]1,+\infty[$.

In the cases (1b) of proposition 3.1 only the position of the interval $[a,\,b]$ of interest in the large 100% interval matters: For large biases $\Delta t \rightarrow +\infty$ (which is equivalent to the expectation $E(t^\wedge)$ of the estimator $t^\wedge$ being shifted to $-\infty$ if the interval $[c,d]$ is considered fixed) the conditional probability of $[a,\,b]$ under the condition of the backcalculated value t_D tends to the limit 1 if a = c and to the limit 0 if c < a. For a large negative bias $\Delta t \rightarrow -\infty$, the situation is vice versa: The limit is 1 if b = d and 0 if b < d. This establishes the paradox that a huge bias can make the probability of an interval higher, the farer the estimated value $t^\wedge$ lies away from this interval.

Case (2) of proposition 3.1 shows a dependency on the relative length of the interval $[a,b]$ in the interval $[c,d]$ but no dependency on the position of the interval $[c,d]$ in IR or on the position of $[a,b]$ in the larger interval $[c,d]$. In case (1a) of proposition 3.1 shows there is a dependency on the absolute position of c, d as well as on the absolute position of a, b in IR.

The fact that large biases can lead to dramatic overestimation of partial intervals establishes a warning post for practical work using Bayesian estimation approaches as e.g. the CPD method. The probabilities calculated are *useful only in case the likelihood probability P(t^ | t) is unbiased.*

We illustrated the importance of the issue by a forensic science example (E) representing a typical error induced in the results of the CPD method by a TDE bias. Example (E) was designed to demonstrate the typical sensitivness of Bayesian estimation to biases in the likelihood probability $P(t^\wedge \mid t)$ used. TDE, having a Gaussian likelihood, implies case (1b) of proposition 3.1 in CPD.

From an abstract point of view, the result can be interpreted as a classification of asymptotic conditional probability distributions induced on a bounded interval by the far tail of a probability distribution which can be chosen from a wide range containing many probability distributions which are important in science statistical problems.

Up to now we are not aware of any further applications of our results apart from the usage in death time determination. In the latter case the approach resolved merely a misunderstanding than yielding a new statistical method. There is no concrete statistical application to our knowledge yet, but we believe that our mainly theoretical article might stimulate further statistical applications. Particularly the classification result could be used for designing tests which can differentiate between classes of probability distributions by using samples from their far tails only. This might be interesting for experimental researchers if they can access the far tails of a distribution only for experimental or financial reasons.

References

1. Biermann FM, Potente S (2011) The deployment of conditional probability distributions for death time estimation. Forensic Sci Int 210: 82-86.
2. Hubig M, Muggenthaler H, Mall G (2014) Conditional probability distribution (CPD) method in temperature based death time estimation: Error propagation analysis. Forensic Sci Int 238: 53-58.
3. Papoulis A (1990) Probability & Statistics. Prentice Hall, Englewood Cliffs, New Jersey.
4. Pericchi LR, Sanso B (1995) A Note on Bounded Influence in Bayesian-Analysis. Biometrika 82: 223-225.
5. Marshall TK, Hoare FE (1962) Estimating the time of death – The rectal cooling after death and its mathematical expression. J Forensic Sci 7: 56-81.
6. Henssge C (1979) Precision of estimating the time of death by mathematical expression of rectal body cooling (author's transl). Z Rechtsmed 83: 49-67.
7. Henssge C (2000) Experiences with a compound method for estimating the time since death in practical casework at the scene. I Rectal temperature time of death nomogram. Int J Legal Med 113: 320-331.