

Reply

ROMAN KRZYSZTOFOWICZ

*Department of Systems Engineering and Department of Statistics, University of Virginia,
Charlottesville, Virginia*

25 October 1999 and 18 January 2000

1. Overview

Kivman (2000, hereafter GAK) comments on the estimators of a probability developed by Krzysztofowicz (1999) and presents a new estimator. This reply points out erroneous comments, shortcomings of the new estimator, and methodological incoherences.

2. Erroneous comments

a. *Is the inference procedure Bayesian?*

In reference to my *estimator from bounds*, GAK alleges that the inference procedure “has nothing in common with the Bayesian analysis.” Let us note, therefore, that an enlightened taxonomy recognizes 46 656 varieties of Bayesian analyses (Good 1983, chapter 3). In the viewpoint adopted herein, Bayesian analysis is tantamount with decision making (DeGroot 1970). The estimator from bounds is obtained via a two-stage decision procedure. First, one decides the sign of stochastic dependence between the subevents. Second, conditional on the first decision, one decides the value of probability π . Sequential decision procedures are as old as Bayesian analysis itself. Classic paradigms include stopping-control problems and stochastic dynamic programming. In such procedures, the probability distribution of uncertain quantity at a stage is always conditional on decisions from the preceding stages. The claim—that only observations (“what actually did happen”) can condition distributions in Bayesian analyses—is naive vis-à-vis the vast literature on sequential decision procedures.

GAK ponders “using another partitioning” of the admissible domain of probability π . It is needless pondering because there is no other sensible partitioning for this decision problem. The sign of stochastic dependence can be either positive or negative, and these hypotheses uniquely prescribe the partitioning.

b. *Is information ignored?*

In reference to my *estimator from correlation*, GAK states that this estimator “does not extract all information contained in the observations but only the correlation coefficient.” This statement is misleading. Given the marginal probabilities π_1 and π_2 of subevents, the probability π of the event is given by (18) in Krzysztofowicz (1999):

$$\pi = \pi_1 + \pi_2 - \pi_1\pi_2 - \alpha(\pi_1, \pi_2)[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}, \quad (1)$$

where $\alpha(\pi_1, \pi_2)$ is the Pearson’s correlation coefficient between the subevents, conditional on the values of the marginal probabilities π_1 and π_2 . Because (1) is the exact equation (it is derived from probability theory), it implies unequivocally that, given π_1 and π_2 , the conditional correlation coefficient $\alpha(\pi_1, \pi_2)$ is *sufficient* to calculate probability π . More generally, the bivariate correlation function α is sufficient to encode all information that is *relevant* to estimating π for all values of (π_1, π_2) .

c. *Which information is used?*

GAK incorrectly states that “a priori information on correlations” is used in (1). The correlation function α is to be estimated from joint observations of forecast probabilities and subevent indicators. This is the *likelihood record* in Bayesian terminology (Krzysztofowicz 1983). The *prior record* (or prior information) contains only observations of subevent indicators. Typically, the prior record is longer than the likelihood record. A full-fledged Bayesian estimator of π would use both records. The estimation procedure outlined in Krzysztofowicz (1999) uses the likelihood record only. GAK’s estimator uses the prior record only.

It is untrue that “the only possible way of reconstructing the whole statistics is to involve prior knowledge.” The theory of limiting posterior distributions teaches that the prior record (or prior knowledge) becomes irrelevant once the likelihood record is large enough (DeGroot 1970, chapter 10).

Corresponding author address: Prof. Roman Krzysztofowicz, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747.

TABLE 1. Examples of estimates of probability π obtained from PME and from bounds.

Forecast marginal probabilities		Bounds† on probability π			Estimate†† of π from PME		Estimate of π from bounds	
π_1	π_2	b	β	B	sign	$\hat{\pi}$	sign	π^*
0.1	0.9	0.90	0.91	1.00	ND	0.918	ND	0.955
0.9	0.9	0.90	0.99	1.00	ND	0.995	PD	0.945
0.1	0.1	0.10	0.19	0.20	ND	0.195	PD	0.145

† Bounds: $b = \max\{\pi_1, \pi_2\}$, $\beta = \pi_1 + \pi_2 - \pi_1\pi_2$, $B = \min\{\pi_1 + \pi_2, 1\}$.
 †† Climatic joint probabilities: $\mu_{11} = \mu_{22} = 0.2$, $\mu_{12} = \mu_{21} = 0.3$.

3. Shortcomings of the PME estimator

A new estimator of π is derived by GAK from the principle of maximum entropy (PME). Its advertised virtues are exaggerated while shortcomings limit its usefulness.

a. Fixed sign of dependence

Let V_i be a Bernoulli variate indicating the occurrence and nonoccurrence of a subevent: $V_i = 0 \Leftrightarrow \bar{A}_i$ and $V_i = 1 \Leftrightarrow A_i$, for $i = 1, 2$. Let μ_{ij} , for $i = 1, 2$ and $j = 1, 2$, denote the climatic joint probabilities of subevents, as introduced by GAK:

$$\begin{aligned} \mu_{11} &= P(\bar{A}_1, \bar{A}_2), & \mu_{22} &= P(A_1, A_2), \\ \mu_{12} &= P(A_1, \bar{A}_2), & \mu_{21} &= P(\bar{A}_1, A_2). \end{aligned} \quad (2)$$

The climatic covariance of V_1 and V_2 is $\text{cov}(V_1, V_2) = d$, where

$$d = \mu_{11}\mu_{22} - \mu_{12}\mu_{21}. \quad (3)$$

It may now be shown that any feasible solution π to (11) in GAK has this property:

$$\begin{aligned} d > 0 &\Rightarrow \pi < \pi_1 + \pi_2 - \pi_1\pi_2, \\ d = 0 &\Rightarrow \pi = \pi_1 + \pi_2 - \pi_1\pi_2, \\ d < 0 &\Rightarrow \pi > \pi_1 + \pi_2 - \pi_1\pi_2. \end{aligned} \quad (4)$$

By referring to facts 1 and 2 in Krzysztofowicz (1999, section 4), we conclude the following: the PME estimator prescribes the sign of stochastic dependence between the subevents that is always identical to the sign of the climatic covariance, regardless of the forecast probabilities π_1 and π_2 . As explained in Krzysztofowicz (1999, sections 4a and 5b) this is a shortcoming.

The purpose of a probabilistic forecast is to quantify

TABLE 2. Forecasts issued by a clairvoyant.

Event	Forecast probabilities			Dependence sign	Frequency of issuing
	π_1	π_2	π		
\bar{A}_1, \bar{A}_2	0	0	0	PD	μ_{11}
A_1, \bar{A}_2	1	1	1	PD	μ_{22}
\bar{A}_1, A_2	1	0	1	ND	μ_{12}
A_1, A_2	0	1	1	ND	μ_{21}

the degree of uncertainty that exists on a particular occasion. Because the degree of uncertainty varies from one occasion to the next, so may vary the sign of stochastic dependence between the subevents. Table 1 shows three numerical examples. In each example, the PME estimate of π assumes negative dependence (ND) because the climatic covariance is negative, $d = -0.05$. In the first example, ND may be reasonable because $\pi_1 + \pi_2 = 1$, which is a necessary condition for extreme ND. In each of the other examples, ND may be unreasonable because $\pi_1 = \pi_2$, which is a necessary condition for extreme positive dependence (PD). In contrast to the PME estimates, the estimates from bounds specify either ND or PD, depending on the values of forecast marginal probabilities π_1 and π_2 .

GAK postulates that “no property of μ_{ij} can generally survive” after updating with a forecast. The PME estimator violates his postulate: it forever retains the sign of the covariance calculated from μ_{ij} .

b. Limited usefulness

The fixed sign of stochastic dependence between the subevents limits the usefulness of the PME estimator. This can be demonstrated as follows. A naive forecaster issues the same climatic probabilities on each occasion: $\pi_1 = \mu_{12} + \mu_{22}$, $\pi_2 = \mu_{21} + \mu_{22}$, and $\pi = \mu_{12} + \mu_{21} + \mu_{22}$. The sign of stochastic dependence is fixed and identical to the sign of the climatic covariance d , given by (3). A clairvoyant issues one of the four possible forecasts, depending on the occasion (Table 2). The sign of stochastic dependence is either PD or ND, based on fact 1 in Krzysztofowicz (1999, section 4). In the long run, the two forecasts implying PD are issued with frequency $\mu_{11} + \mu_{22}$, and the two forecasts implying ND are issued with frequency $\mu_{12} + \mu_{21}$.

The characteristics of any real forecast system fall somewhere between the naive forecaster and the clairvoyant. A good forecaster will predict sometimes PD and sometimes ND, irrespective of the sign of the climatic covariance. The PME estimator fails to mimic this characteristic. Hence it is useless for estimating π based on π_1 and π_2 output from a good forecast system.

c. Single parameter

To avoid trivial cases, suppose all climatic joint probabilities are nondegenerate, that is, $\mu_{ij} > 0$ for $i = 1, 2$ and $j = 1, 2$. Then define parameter

$$c = \frac{\mu_{12}\mu_{21}}{\mu_{11}\mu_{22} - \mu_{12}\mu_{21}}. \tag{5}$$

The PME estimator may now be reparametrized as follows. The condition $d \neq 0$ is equivalent to $-\infty < c < \infty$. The condition $D \geq 0$ is equivalent to $D/d^2 > 0$, which holds if and only if $(\pi_1 + \pi_2 + c)^2 - 4\pi_1\pi_2(1 + c) \geq 0$. Finally, (16) in GAK is identical to

$$\pi = \frac{1}{2}\{\pi_1 + \pi_2 - c + \text{sign}(c)[(\pi_1 + \pi_2 + c)^2 - 4\pi_1\pi_2(1 + c)]^{1/2}\}. \tag{6}$$

This reveals that the PME estimator has a single parameter c . In other words, the PME estimator does not require the whole climatic joint probability function (which is specified by three parameters, say μ_{11} , μ_{22} , and μ_{12}), but only the single statistic c . Thus, it is an exaggeration to claim that the PME estimator “enables one to take into account the total information contained in the meteorological data.”

d. Just another correlation function

Equation (6) may be transformed further so that probability π is given by (1) with the correlation function being specified by

$$\alpha(\pi_1, \pi_2) = \frac{A(\pi_1, \pi_2) - \pi_1\pi_2}{[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}}, \tag{7}$$

where

$$A(\pi_1, \pi_2) = \frac{1}{2}\{\pi_1 + \pi_2 + c - \text{sign}(c)[(\pi_1 + \pi_2 + c)^2 - 4\pi_1\pi_2(1 + c)]^{1/2}\}. \tag{8}$$

Hence, the PME estimator is nothing else but a particular case of the general estimator from correlation (1). Like earlier estimators of Hughes and Sangster (1979) and Wilks (1990), the PME estimator prescribes a parametric model for the correlation function α .

GAK critiques estimator (1), which supposedly “does not extract all information contained in the observations but only the correlation” function. Ironically, it turns out that the PME estimator extracts even less: a single parameter of the correlation function.

4. Methodological incoherences

a. Misapplying PME

The PME estimator boils down to a special case of the correlation function. Thus whatever limitations of the correlation function (as a means of calculating probability π), they apply to the PME estimator as well. But the PME estimator has additional shortcomings: (i) it fixes the sign of the correlation function regardless of the values of the forecast marginal probabilities, and (ii) it prescribes the functional form of the correlation function regardless of any data.

The assertions that the PME is inherently superior lack coherence: they are contradicted by the shortcomings of the derived estimator. These shortcomings need not be surprising. The primary intent of the PME in Bayesian analyses has been to determine an (essentially) noninformative prior distribution that reflects some initial “objective” information. Some consider this usage of the PME convincing (Bernardo and Smith 1994, section 5.6). However, the task of determining π is unlike the task of determining the noninformative prior probability. Rather, it is the task of determining the posterior probability π , given forecast (π_1, π_2) and an informative prior probability function $\{\mu_{ij} : i = 1, 2; j = 1, 2\}$. GAK’s approach to this task misapplies the PME.

b. Blaming data

A pragmatic view might be that the PME is merely a tool for obtaining analytic expressions, especially when knowledge and data are insufficient to pursue physically based and/or empirically based modeling. Then one should remember that the PME is not deduced from the axioms of probability theory (as, for instance, Bayes theorem is), but is appended to the theory. It brings its own axioms (Shore and Johnson 1980), which bear on the derived expressions. Meaningless, therefore, is the claim that probabilities “calculated by means of the PME are the most ‘objective’ among others.”

The calculated probabilities must be verified against data. GAK concedes the inevitability of experimentation. But should the PME estimator perform poorly, GAK has already prepared the answer: an incorrect prediction “will give evidence that the inputs (π_i, μ_{ik}) are wrong.” Now, let us recall that π_1 and π_2 are probabilities from any well-calibrated forecast system, and μ_{ik} are climatic probabilities of elementary rain events. When GAK will declare these probabilities to be wrong, should we outlaw the forecast system, distort the climatic record, or change the climate?

Hopefully modern meteorology will continue to evolve its predictive capabilities despite an occasional folly.

REFERENCES

Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory*. Wiley, 586 pp.

- DeGroot, M. H., 1970: *Optimal Statistical Decisions*. McGraw-Hill, 490 pp.
- Good, I. G., 1983: *Good Thinking: The Foundations of Probability and Its Applications*. University of Minnesota Press, 332 pp.
- Hughes, L. A., and W. E. Sangster, 1979: Combining precipitation probabilities. *Mon. Wea. Rev.*, **107**, 520–524.
- Kivman, G. A., 2000: Comments on “Probability for a period and its subperiods: Theoretical relations for forecasting.” *Mon. Wea. Rev.*, **128**, 3011–3013.
- Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes theorem. *Water Resour. Res.*, **19** (2), 327–336.
- , 1999: Probabilities for a period and its subperiods: Theoretical relations for forecasting. *Mon. Wea. Rev.*, **127**, 228–235.
- Shore, J. E., and R. W. Johnson, 1980: Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inform. Theory*, **IT-26** (1), 26–37.
- Wilks, D. S., 1990: On the combination of forecast probabilities for consecutive precipitation periods. *Wea. Forecasting*, **5**, 640–650.