Probabilities for a Period and Its Subperiods: Theoretical Relations for Forecasting

ROMAN KRZYSZTOFOWICZ

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

(Manuscript received 6 August 1997, in final form 2 March 1998)

ABSTRACT

Consider an event definable in terms of two subevents as, for example, the occurrence of precipitation within a 24-h period is definable in terms of the occurrence of precipitation within each of the 12-h subperiods. A complete forecast must specify three probabilities; these may be marginal probabilities, one for the period and two for subperiods. Theoretical relations between these probabilities are investigated and solutions are presented to three problems encountered in operational forecasting: (i) guaranteeing that the marginal probabilities jointly obey the laws of probability, (ii) structuring admissible procedures for adjusting the initial (guidance) probabilities by forecasters, and (iii) formulating optimal estimators of the probability for period in terms of the probabilities for subperiods.

1. Introduction

a. Nested forecast paradigm

It is common in operational meteorology to forecast an event within a *period* and, independently, to forecast subevents within two *subperiods*; such subperiods are nonoverlapping and their concatenation equals the period. For example, forecasts of precipitation occurrence (0.01 in. or more of precipitation observed at a gauge) produced via model output statistics (MOS) specify probabilities of precipitation (PoPs) for a 12-h period and its two 6-h subperiods (Carter et al. 1989). Each PoP is estimated from a separate system of equations although predictors may be the same in each system.

In general, let A and $\overline{A}$ denote the occurrence and nonoccurrence of an event within a period. Let A_i and $\overline{A}_i$ denote the occurrence and nonoccurrence of a subevent within subperiod i, $i = 1, 2$. The event and subevents are *nested* as follows: the event does not occur if and only if neither subevent occurs; symbolically, $\overline{A} \Leftrightarrow (\overline{A}_1, \overline{A}_2)$. With P denoting the probability, a forecast specifies $\pi = P(A)$ and $\pi_i = P(A_i)$ for $i = 1, 2$, where $0 \leq \pi \leq 1$ and $0 \leq \pi_i \leq 1$.

b. Operational questions

Three questions arise in the context of the nested forecast paradigm. (i) Suppose that each probability is estimated irrespective of the other two. Does probability theory impose any coherence conditions that (π_1, π_2, π) must satisfy in order to constitute a proper probabilistic forecast? (ii) Suppose that probabilities (π_1, π_2, π) are coherent and a forecaster adjusts one of them. Does probability theory impose any constraint on such an adjustment, and does it imply any rule for adjusting the other two probabilities? (iii) Suppose a forecast system produces only π_1 and π_2. Does probability theory imply a procedure for estimating π?

The answer to the first question flows directly from basic probability relations; however, the importance of coherence conditions needs to be elucidated and the problem of reconciling (π_1, π_2, π) that are incoherent may have to be addressed. For example, MOS-produced PoPs are not guaranteed to be coherent.

The second question has not been studied, to the best of our knowledge. It arose recently in the context of developing software that will support operational forecasting in the modernized National Weather Service (NWS). The software displays a gridded (in space) field of MOS-produced PoPs for a period or a subperiod. A forecaster may interactively adjust this field (Ruth and Peroutka 1993). In effect, he may easily destroy the coherence conditions that must be satisfied by the three fields (π_1, π_2, π) at each grid point. Therefore, a theory is needed to correctly structure the adjustment process.

The third question has been addressed by Hughes and Sangster (1979) and Wilks (1990), who developed heuristic estimators of a 24-h PoP in terms of two 12-h PoPs. Such estimators were needed because MOS does not produce the 24-h PoP. The same need exists currently, as the 24-h PoP becomes part of a probabilistic quantitative precipitation forecast (Krzysztofowicz and Sigrest 1997). However, for a general forecasting pro-

Corresponding author address: Prof. Roman Krzysztofowicz, University of Virginia, Thornton Hall, SE, Charlottesville, VA 22903.

TABLE 1. Subevents, marginal probabilities, and transition probabilities from subperiod i to subperiod j.

	$\overline{A}_j$	A_j	
$\overline{A}_i$	$1 - q_{ij}$	q_{ij}	$1 - \pi_i$
A_i	$1 - p_{ij}$	p_{ij}	π_i
	$1 - \pi_j$	π_j	

cedure, one desires estimators that are derived theoretically.

This article investigates theoretical relations between π_1, π_2, and π. These relations provide a basis for answering all three questions in a comprehensive manner and for designing forecasting procedures. Although the theory is general, in that it holds for any events, interpretations throughout the paper are in terms of rain and no-rain events.

2. Basic probability relations

When the forecast specifies marginal probabilities of event and subevents, the conditional, transition, and joint probabilities can be derived via the laws of probability. Table 1 depicts the subevents, their marginal probabilities, and the transition probabilities. In all expressions, the indices of subperiods are interchangeable; that is, either ($i = 1$, $j = 2$) or ($i = 2$, $j = 1$).

a. Derived probabilities

Let $r_i = P(A_i \mid A)$ denote the probability of rain in subperiod i, conditional on the hypothesis that it rains during the period. It is given by

$$r_i = \frac{\pi_i}{\pi}. \tag{1}$$

Let $p_{ij} = P(A_j \mid A_i)$ denote the first transition probability: the probability of rain in subperiod j, conditional on the hypothesis that it rains in subperiod i. It is given by

$$p_{ij} = \frac{\pi_i + \pi_j - \pi}{\pi_i}. \tag{2}$$

Let $q_{ij} = P(A_j \mid \overline{A}_i)$ denote the second transition probability: the probability of rain in subperiod j, conditional on the hypothesis that no rain occurs in subperiod i. It is given by

$$q_{ij} = \frac{\pi - \pi_i}{1 - \pi_i}. \tag{3}$$

The joint probabilities of subevents are given by

$$P(A_i, A_j) = \pi_i + \pi_j - \pi,$$
$$P(A_i, \overline{A}_j) = \pi - \pi_j,$$
$$P(\overline{A}_i, A_j) = \pi - \pi_i,$$
$$P(\overline{A}_i, \overline{A}_j) = 1 - \pi. \tag{4}$$

In summary, forecast (π_1, π_2, π) completely characterizes stochastic dependence between the subevents.

b. Coherence conditions

Suppose each forecast probability is proper in the sense that $0 \le \pi \le 1$ and $0 \le \pi_i \le 1$ for $i = 1, 2$. The transition probability p_{ij} is proper if and only if $0 \le p_{ij} \le 1$. The bounds on p_{ij}, together with (2), yield two conditions. First, $0 \le p_{ij}$ if and only if $\pi \le \pi_i + \pi_j$. Second, $p_{ij} \le 1$ if and only if $\pi_j \le \pi$ for $j = 1, 2$. All these inequalities can be summarized as follows:

$$0 \le \pi_i \le 1, \qquad i = 1, 2, \tag{5a}$$

$$\max\{\pi_1, \pi_2\} \le \pi \le \min\{\pi_1 + \pi_2, 1\}. \tag{5b}$$

In conclusion, forecast (π_1, π_2, π) of nested events is proper if it satisfies the *coherence conditions* (5). (These conditions also guarantee that probabilities r_i and q_{ij} are proper.) And vice versa, any three probabilities (π_1, π_2, π) that satisfy the coherence conditions (5) form a proper forecast of nested events.

3. Coherent adjustment of probabilities

The theory can be harnessed into an interactive forecasting procedure for the NWS forecasters. The input to such a procedure is a guidance forecast (π_1, π_2, π), as produced by MOS or other technique. The guidance is assumed to be coherent, as in (5). The task of the forecaster is to adjust the guidance probabilities based on local information. This task has a strong empirical justification: the adjusted (subjective) PoPs have been shown to be better calibrated and more informative than the guidance (objective) PoPs (Murphy and Brown 1984; Krzysztofowicz and Long 1991). However, the theory implies that an adjustment to any one of the probabilities (π_1, π_2, π) must not be made irrespective of the other two probabilities. Two *admissible adjustment procedures* are detailed next.

a. Procedure with fixed marginals

The forecaster adjusts one marginal probability, while the other two marginal probabilities are held fixed. The fixed probabilities impose constraints on the adjustment. These constraints, which follow from (5), ensure that the adjusted probabilities remain coherent. When π is being adjusted to new value π', the constraints are

$$\max\{\pi_1, \pi_2\} \le \pi' \le \min\{\pi_1 + \pi_2, 1\}. \tag{6}$$

When π_i is being adjusted to new value π_i', the constraints are

$$\pi - \pi_j \le \pi_i' \le \pi, \tag{7}$$

where ($i = 1$, $j = 2$) or ($i = 2$, $j = 1$). The procedure can be applied to each of the three probabilities, in any order. It can also be iterated as many times as desired.

TABLE 2. Procedure for adjusting marginal probabilities that preserves conditional probabilities and coherence conditions. Subperiod indices take values ($i = 1, j = 2$) or ($i = 2, j = 1$).

Option	Preserved probability	Adjusted probability	Constraint on adjustment	Inferred probabilities	
1a	r_i, r_j	π'	$0 \le \pi' \le 1$	$\pi'_i = r_i \pi'$	$\pi'_j = r_j \pi'$
1b	p_{ij}, q_{ij}	π'	$q_{ij} \le \pi' \le 1$	$\pi'_i = (\pi' - q_{ij})/(1 - q_{ij})$	$\pi'_j = \pi' - (1 - p_{ij})\pi'_i$
2a	p_{ij}, r_i	π'_i	$0 \le \pi'_i \le r_i$	$\pi' = \pi'_i/r_i$	$\pi'_j = \pi' - (1 - p_{ij})\pi'_i$
2b	p_{ij}, q_{ij}	π'_i	$0 \le \pi'_i \le 1$	$\pi' = \pi'_i + q_{ij}(1 - \pi'_i)$	$\pi'_j = \pi' - (1 - p_{ij})\pi'_i$
3a	r_j	π'	$0 \le \pi' \le 1$		$\pi'_j = r_j \pi'$
		π'_i	$\pi' - \pi'_j \le \pi'_i \le \pi'$		
3b	p_{ij}	π'	$0 \le \pi' \le 1$		
		π'_i	$0 \le \pi'_i \le \pi'$		$\pi'_j = \pi' - (1 - p_{ij})\pi'_i$
4a	r_i	π'_i	$0 \le \pi'_i \le r_i$	$\pi' = \pi'_i/r_i$	
		π'_j	$\pi' - \pi'_i \le \pi'_j \le \pi'$		
4b	q_{ij}	π'_i	$0 \le \pi'_i \le 1$	$\pi' = \pi'_i + q_{ij}(1 - \pi'_i)$	
		π'_j	$\pi' - \pi'_i \le \pi'_j \le \pi'$		

The procedure has two properties. (i) All three marginal probabilities may undergo significant adjustments, while retaining the property of coherence; also, the forecaster is in control of the adjustment to each marginal probability π_1, π_2, and π. (ii) The conditional probabilities (r_i, p_{ij}, q_{ij}) may be altered as a result of adjustments to marginal probabilities. In other words, the adjustments may change the degree, and even the sign (positive or negative), of the *temporal stochastic dependence* between the event and subevents, and between the subevents. Thus the procedure is suitable when the forecaster has a reason to effect such a change.

b. Procedure with fixed conditionals

The key properties of this procedure are as follows. (i) The forecaster adjusts one or two of the marginal probabilities (π_1, π_2, π) within constraints that ensure coherence, while theoretical relations are used to infer the remaining marginal probabilities; thus the forecaster controls the adjustment of only one or two marginal probabilities. (ii) One or two of the conditional probabilities (r_i, p_{ij}, q_{ij}) are held fixed as calculated from the initial (or guidance) marginal probabilities. Thereby the procedure preserves the sign and the degree of the temporal stochastic dependence between the event and subevents, or between the subevents, as specified by the initial forecast.

The procedure has been derived from relations (1), (2), (3), (5) and is summarily detailed in Table 2. It consists of five tasks.

1) The forecaster selects the marginal probabilities to be adjusted; there are four options: π, π_i, (π, π_i), and (π_i, π_j), where ($i = 1, j = 2$) or ($i = 2, j = 1$).
2) The forecaster selects the conditional probabilities to be preserved. For example, when π_i is chosen for adjustment, one can preserve either (p_{ij}, r_i) or (p_{ij}, q_{ij}).
3) Every conditional probability to be preserved is calculated from the initial marginal probabilities according to an appropriate equation, (1), (2), or (3).
4) The forecaster adjusts the chosen marginal probability. The adjusted value must satisfy a constraint that derives from the coherence conditions (5).
5) The remaining marginal probability is inferred via an equation that derives from (1), (2), or (3).

Tasks 4 and 5 should be executed in the order of columns (rightward) and lines (downward) in Table 2. The procedure can be repeated as many times as desired, each time with the same or a different option; however, only options 1b and 2b ensure that the conditional probabilities p_{ij} and q_{ij} specified by the initial forecast are preserved regardless of the number and the order of executions. The options listed in Table 2 are not exhaustive, but they are deemed primary for operational forecasting.

In summary, there are two admissible procedures for adjusting the marginal probabilities in a manner that guarantees their coherence, as required by probability theory. Each procedure, the one with fixed marginals or the one with fixed conditionals, offers different properties. Switching between the two procedures in the course of forecast preparation is admissible.

4. Characterization of dependence

The expressions for transition probabilities, given earlier, will now be used to characterize stochastic dependence between the subevents. The purpose of such a characterization is twofold. First, it should be valuable in training forecasters because it elucidates the connection between marginal probabilities (π_1, π_2, π) and the notion of the temporal stochastic dependence. Second, it is prerequisite to constructing an estimator of π in terms of π_1 and π_2, which is one of our objectives.

a. Extreme dependence

All elements involved in the characterization of stochastic dependence between the subevents are shown in Table 1. As before, the transition probabilities apply to either ($i = 1, j = 2$) or ($i = 2, j = 1$).

Definition 1. The subevents are said to be (i) *stochastically independent* if and only if

$$P(A_j|A_i) = P(A_j) \quad \text{and} \quad P(A_j|\overline{A}_i) = P(A_j); \quad (8a)$$

(ii) *extremely positive dependent* if and only if

$$P(A_j\,|\,A_i) = 1 \quad \text{and} \quad P(\overline{A}_j\,|\,\overline{A}_i) = 1; \quad (8b)$$

(iii) *extremely negative dependent* if and only if

$$P(A_j\,|\,\overline{A}_i) = 1 \quad \text{and} \quad P(\overline{A}_j\,|\,A_i) = 1. \quad (8c)$$

When expressions (2) and (3) are substituted for the conditional probabilities, the following fact is readily proven.

Fact 1. The subevents are (i) stochastically independent if and only if

$$\pi = \pi_1 + \pi_2 - \pi_1\pi_2; \quad (9a)$$

(ii) extremely positive dependent if and only if

$$\pi = \pi_1 = \pi_2; \quad (9b)$$

(iii) extremely negative dependent if and only if

$$\pi = \pi_1 + \pi_2 = 1. \quad (9c)$$

The potential utility of this fact in operational forecasting can be illustrated via three scenarios. (i) In the warm season, a forecaster prepares PoPs for the 12-h period, 0600–1800, and its two 6-h subperiods. He judges it probable to observe two thunderstorms, one in early morning, 0600–1200, and another in late afternoon, 1200–1800. He also judges that the second thunderstorm may occur (or may not occur) regardless of whether or not the first one has occurred. In essence, he predicts that the subevents are stochastically independent. Consequently, he needs to assess only two PoPs whereas the third can be calculated via (9a). For example, he may assess π_1 and π_2 such that $0 \leq \pi_1 \leq 1$ and $0 \leq \pi_2 \leq 1$, and then calculate π. Or he may assess π and π_1 such that $0 \leq \pi_1 \leq \pi \leq 1$, and then calculate $\pi_2 = (\pi - \pi_1)/(1 - \pi_1)$. (ii) In the cold season, an approaching front may or may not pass over an area. If it does, then both 6-h subperiods will be wet. If it does not, then both 6-h subperiods will be dry. This is the case of extreme positive dependence between the subevents. Consequently, the forecaster needs to assess only one PoP because, per (9b), PoPs for subperiods are identical with PoP for the period. (iii) In the warm season, a forecaster is almost certain to observe a thunderstorm, which may occur either during subperiod 1200–1800 or during subperiod 1800–0000. In essence, the forecaster predicts that the subevents are extremely negative dependent. His judgment of almost certain occurrence implies $\pi \simeq 1$, whereas his judgment of extreme negative dependence implies (9c); consequently, he needs to assess only one PoP, say, π_1, and then calculate $\pi_2 = 1 - \pi_1$.

In summary, the notion of stochastic dependence between subevents can be incorporated into the meteorologic thought process in a way that helps the forecaster to assess PoPs for the period and subperiods.

b. Quadrant dependence

The weakest, and therefore most general, characterization of stochastic dependence between subevents, in cases which are not extreme, rests on the following definition (Lehmann 1966).

Definition 2. The subevents are said to be (i) *positive quadrant dependent* if and only if

$$P(A_j|A_i) > P(A_j) \quad \text{and} \quad P(\overline{A}_j|\overline{A}_i) > P(\overline{A}_j); \quad (10a)$$

(ii) *negative quadrant dependent* if and only if

$$P(A_j|A_i) < P(A_j) \quad \text{and} \quad P(\overline{A}_j|\overline{A}_i) < P(\overline{A}_j). \quad (10b)$$

When expressions (2) and (3) are substituted for the conditional probabilities, the following fact is readily proven.

Fact 2. The subevents are (i) positive quadrant dependent if and only if

$$\pi < \pi_1 + \pi_2 - \pi_1\pi_2; \quad (11a)$$

(ii) negative quadrant dependent if and only if

$$\pi > \pi_1 + \pi_2 - \pi_1\pi_2. \quad (11b)$$

In essence, these inequalities establish the connection between the marginal probabilities and the sign, positive or negative, of stochastic dependence between the subevents. When instead of an inequality, the equality holds, the subevents are stochastically independent based on fact 1.

c. Bounds on probability for period

It is possible now to establish a relationship between the sign of stochastic dependence and the bounds on π. Toward this end, we introduce abbreviations IN for independence, PD for positive dependence, and ND for negative dependence; and define quantities

$$b = \max\{\pi_1, \pi_2\},$$
$$\beta = \pi_1 + \pi_2 - \pi_1\pi_2,$$
$$B = \min\{\pi_1 + \pi_2, 1\}. \quad (12)$$

From (5b), (9), and (11) it follows that $0 \leq b \leq \beta \leq B \leq 1$ and that

$$\text{PD} \Leftrightarrow b \leq \pi < \beta,$$
$$\text{IN} \Leftrightarrow \pi = \beta,$$
$$\text{ND} \Leftrightarrow \beta < \pi \leq B. \quad (13)$$

To complete this characterization of stochastic dependence between the subevents, one could derive and examine a correlation coefficient. This is done in the appendix.

TABLE 3. Examples of bound-based estimates and heuristic estimates of probability π for period.

Marginal probabilities		Bounds on probability π			Estimate of π from bounds		Heuristic estimates of π		Bounds on correlation ρ	
π_1	π_2	b	β	B	Sign	π^*	HS[a]	HSW[b]	r	R
0.1	0.1	0.10	0.19	0.20	PD	0.145	0.172	0.147	−0.111	1.0
0.5	0.5	0.50	0.75	1.00	IN	0.750	0.659	0.745	−1.0	1.0
0.1	0.9	0.90	0.91	1.00	ND	0.955	0.906	0.905	−1.0	0.111

[a] HS: Hughes–Sangster; coefficient $k = 0.55$.
[b] HSW: Hughes–Sangster–Wilks; coefficients $k = 0.55$ and $l = 7$.

5. Estimators of probability for period

a. Estimator from bounds

Relations (12)–(13) suffice to answer the third question: How to estimate π when only π_1 and π_2 are given? The proposed estimator is justified by a principle of Bayesian inference (e.g., DeGroot 1970) and is composed of two rules.

First, the sign of stochastic dependence between the subevents is inferred according to the rule

$$\beta - b > B - \beta \Rightarrow \text{PD}, \tag{14a}$$

$$\beta - b = B - \beta \Rightarrow \text{IN}, \tag{14b}$$

$$\beta - b < B - \beta \Rightarrow \text{ND}. \tag{14c}$$

Second, an estimate π^* of probability π is obtained according to the rule

$$\text{PD} \Rightarrow \pi^* = (b + \beta)/2, \tag{15a}$$

$$\text{IN} \Rightarrow \pi^* = \beta, \tag{15b}$$

$$\text{ND} \Rightarrow \pi^* = (\beta + B)/2. \tag{15c}$$

The three bounds on π define two subintervals of length $\beta - b$ and $B - \beta$. If the subinterval of π values implying PD is longer than the subinterval of π values implying ND, then the inferred sign is PD. This inference is optimal whenever the meteorologist has no additional knowledge about π and expresses this uncertainty in a Bayesian fashion by assigning a uniform density to all values of π between bounds b and B. Consequently, the likelihood of a particular sign (PD or ND) is proportional to the length of the subinterval. The sign having higher likelihood is chosen. Next, conditional on this choice, the mean value of π under the uniform density is taken as the estimate π^*. When the subintervals are of equal length, no choice between PD and ND can be made. Because the uncertainty about the sign cannot be resolved, the unconditional mean value of π under the uniform density on the interval (b, B) is taken as the estimate π^*. This estimate is β, which is equivalent to inferring IN.

Table 3 shows three numerical examples. In the first example, the marginal probabilities have a common value, $\pi_1 = \pi_2 = 0.1$, which is the necessary condition for extreme positive dependence (fact 1). Of course, such dependence cannot be confirmed without knowing probability π. However, a comparison of subintervals, $\beta - b = 0.09$ and $B - \beta = 0.01$, reveals that PD is nine times as likely as ND. Consequently, π^* is calculated as the mean of b and β. In the second example, $\pi_1 = \pi_2 = 0.5$, which is the necessary condition for extreme positive dependence, and $\pi_1 + \pi_2 = 1$, which is the necessary condition for extreme negative dependence (fact 1). Moreover, the identical length of subintervals leaves the inference of sign inconclusive. Consequently, π^* is set to β. In the third example, $\pi_1 + \pi_2 = 1$, which is the necessary condition for extreme negative dependence (fact 1). Based on a comparison of subintervals, $\beta - b = 0.01$ and $B - \beta = 0.09$, ND is nine times as likely as PD. Consequently, π^* is calculated as the mean of β and B.

To sum up, the proposed estimator has three properties: (i) it allows for both positive and negative dependence between the subevents, (ii) it is parameter-free because only π_1 and π_2 are needed to estimate π, and (iii) it is general in the sense that it applies to any events and to periods of any length. What remains to be examined is its performance. While an empirical evaluation is left for future studies, a comparison of this estimator with known heuristic estimators is undertaken next.

b. Previous heuristic estimators

The problem of estimating π from π_1 and π_2 was studied by Hughes and Sangster (1979), who came up with the following estimator:

$$\hat{\pi} = \pi_1 + \pi_2 - [\min\{\pi_1, \pi_2\}][\max\{\pi_1, \pi_2\}]^k, \tag{16}$$

where k is a parameter such that $0 \le k \le 1$. If $k = 0$, then $\hat{\pi} = \max\{\pi_1, \pi_2\} = b$, which is the lower bound on π. If $k = 1$, then $\hat{\pi} = \pi_1 + \pi_2 - \pi_1\pi_2 = \beta$, which is the case of stochastic independence between the subevents. Thus $b \le \hat{\pi} \le \beta$, which means that the estimator allows for only positive dependence between the subevents, regardless of the values of π_1 and π_2.

The estimator was used to calculate 24-h PoP from two 12-h PoPs, as well as to calculate 36-h PoP from the calculated 24-h PoP and a given 12-h PoP. Parameter k was estimated from data for eight stations, separately for the cool season (October–March) and the warm season (April–September).

Wilks (1990) tested the performance of estimator (16) on a large set of historical forecasts for 100 stations throughout the United States and found that the calculated $\hat{\pi}$ tended to overestimate low values of π. He suggested a modification to (16) whereby parameter k is replaced by a function m of (π_1, π_2) such that

$$m(\pi_1, \pi_2) = k[1 - \exp(-l \min\{\pi_1, \pi_2\})], \quad (17)$$

where k is the original parameter and l is a second parameter to be estimated heuristically. A test confirmed that the modification alleviated the bias of the original estimator.

Table 3 shows three numerical examples of estimates $\hat{\pi}$ according to (16) and (17). In the first and second examples, the Hughes–Sangster (HS) estimates are different from the bound-based estimates, π^*, falling above $\pi^* = 0.145$ and below $\pi^* = 0.750$. On the other hand, the Hughes–Sangster–Wilks (HSW) estimates are close to the bound-based estimates. In appraising this result, one must bear in mind that estimates π^* are obtained sans parameters, whereas the HSW estimates are obtained using two parameters that had to be estimated from historical forecasts.

In the third example, the HS and HSW estimates are close to each other, but each is smaller than the bound-based estimate. The inferred sign of dependence between the subevents is ND. None of the heuristic estimators recognizes ND, but always assumes PD. As a result, each heuristic estimate $\hat{\pi}$ falls within the short subinterval, $0.90 < \hat{\pi} < 0.91$, even though the likelihood of π being inside this subinterval is nine times smaller than the likelihood of π being between 0.91 and 1.0, where the bound-based estimate π^* falls.

The shortcomings of the HSW estimator were keenly recognized by Wilks (1990). First, this estimator allows for only positive dependence between the subevents; in reality, negative dependence is possible, as explained in section 4a. Second, its parameters lack a formal statistical interpretation; in effect their estimation was rather ad hoc. Third, its functional form was chosen heuristically; consequently, no assurance could be given that this is the correct form.

Despite its shortcomings, the HSW estimator performed robustly in tests reported by Wilks (1990). Nevertheless, Wilks cautioned the reader that this estimator ''was developed for use specifically with the 12-h PoP forecasts'' and ''should not be used with other types of probability forecasts without previous demonstration of its validity.'' This raises a methodological question: Are there general estimators, applicable to any events and to periods of any length?

It turns out that general estimators can be deduced from theoretical relations between marginal probabilities π_1, π_2, and π. Two types of estimators have been formulated herein, one parameter-free (already described in section 5a) and one functional (to be presented next).

c. *Estimator from correlation*

A functional estimator of π follows directly from relation (A2) in the appendix. It takes the form

$$\pi^* = \pi_1 + \pi_2 - \pi_1\pi_2 - \alpha(\pi_1, \pi_2) \times [\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}, \quad (18)$$

where α is a *correlation function,* defined for all (π_1, π_2) such that $0 < \pi_1, \pi_2 < 1$, and bounded as

$$r(\pi_1, \pi_2) \le \alpha(\pi_1, \pi_2) \le R(\pi_1, \pi_2), \quad (19)$$

with r and R being specified by (A4) and (A5) in the appendix. Loosely speaking, parameter $\alpha(\pi_1, \pi_2)$ is a conditional correlation coefficient between the subevents. Formally, let V_i be a Bernoulli variate indicating the occurrence and nonoccurrence of a subevent: $V_i = 0$ if and only if $\overline{A}_i$, $V_i = 1$ if and only if A_i, for $i = 1, 2$. Then $\alpha(\pi_1, \pi_2) = \mathrm{cor}(V_1, V_2 \mid \pi_1, \pi_2)$ is the Pearson's correlation coefficient between V_1 and V_2 on those occasions on which the forecast probabilities of subevents are (π_1, π_2).

To explain the estimator, a practical procedure for finding α is outlined for the case wherein each probability π_i is a continuous variable. (A procedure for discrete probabilities is somewhat simpler.) The procedure consists of five steps.

1) Collect joint observations of forecast probabilities (π_1, π_2) and subevent indicators (υ_1, υ_2), where $\upsilon_i = 0$ if $\overline{A}_i$ and $\upsilon_i = 1$ if A_i for $i = 1, 2$.
2) Partition the probability interval (0, 1) into m exclusive and exhaustive subintervals. Apply the subintervals to π_1 and π_2. As a result, m^2 subsets for (π_1, π_2) are created. Index these subsets by k, so that $k = 1, \ldots, m^2$.
3) Stratify the sample of joint observations $(\pi_1, \pi_2, \upsilon_1, \upsilon_2)$ into m^2 subsamples based on the observed (π_1, π_2), which falls into one of the m^2 subsets.
4) For each subset k, estimate correlation coefficient $\alpha_k = \mathrm{cor}(V_1, V_2)$ using observations (υ_1, υ_2) from the corresponding subsample. The estimate must satisfy the constraint $r(\pi_1, \pi_2) \le \alpha_k \le R(\pi_1, \pi_2)$ for every possible (π_1, π_2) within the subset.
5) Assign values to the correlation function as follows: $\alpha(\pi_1, \pi_2) = \alpha_k$ for every possible (π_1, π_2) within subset k. Alternatively, use subsets to form a grid of (π_1, π_2) values, attach correlation values α_k to the grid points, and estimate a continuous function α.

It should be apparent that the estimator will perform best when the correlation between subevents is stable across all forecasting occasions on which the same probabilities (π_1, π_2) are issued. For this reason, function α could have different estimates for different regions, seasons, storm types, and possibly other predictors of the correlation between subevents.

To sum up, the estimator has three main properties. (i) It allows for both positive and negative dependence between the subevents; the degree of dependence can vary with marginal probabilities and within bounds specified by the theory. (ii) It incorporates a parameter that has a familiar statistical interpretation and that can be optimally estimated from observations. (iii) It is general in the sense that it applies to any events and to periods of any length.

d. Connection between estimators

Estimator (16) was developed by postulating a parametric model for the conditional probability $P(A_j \mid A_i)$, as explained by Hughes and Sangster (1979). Because the functional form of estimator (18) is exact, it can reveal still another interpretation of the HS estimator. By equating (16) with (18), one finds that

$$\alpha(\pi_1, \pi_2) = \frac{[\min\{\pi_1, \pi_2\}][\max\{\pi_1, \pi_2\}]^k - \pi_1\pi_2}{[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}}. \tag{20}$$

It follows that the HS estimator, and likewise the HSW estimator, constitutes a special case of general estimator (18), arising when correlation function α is assumed to have parametric form (20).

In conclusion, parametric estimation of π from π_1 and π_2 hinges on modeling the correlation between subevents as a function of the marginal probabilities. This function can be estimated in a discrete form, as outlined in section 5c, or in a functional form. In the latter case, a model must be hypothesized and tested. Expression (20), with k replaced by (17), is one possible model for positively dependent subevents.

Finally, an obvious question arises. Will the effort expended on modeling and estimation of the correlation function lead to estimates of π that are significantly better than the estimates obtained from bounds, which require no parameters? The examples shown in Table 3 suggest a possibility of a negative answer. Only empirical tests can confirm or refute this conjecture.

6. Closure

Forecast probabilities of nested events, such as PoPs, have for years been used intuitively by the general public and decision makers. Increasingly, these probabilities will be used analytically in models forecasting related events and in models supporting decision making. Analytic uses require that each of the three marginal probabilities be estimated optimally and that jointly the three probabilities obey the laws of probability. The theoretical relations investigated in this article provide a normative basis for designing forecasting procedures that meet these requirements.

Acknowledgments. This article is based upon work supported by the National Oceanic and Atmospheric Administration under Award NA67WD0486, ''Probabilistic Hydrometeorological Forecast System.''

APPENDIX

Correlation Coefficient

Stochastic dependence between the subevents can also be characterized in terms of the Pearson's correlation coefficient. Toward this end, we define a Bernoulli variate V_i for each subperiod $i = 1, 2$ as follows:

$$V_i = 0 \Leftrightarrow \overline{A}_i,$$

$$V_i = 1 \Leftrightarrow A_i. \tag{A1}$$

With the joint probability function of (V_1, V_2) specified by (4), it is straightforward to derive the correlation coefficient between V_1 and V_2, denoted $\rho = \text{cor}(V_1, V_2)$. It is defined for $0 < \pi_1, \pi_2 < 1$ and takes the form

$$\rho = \frac{\pi_1 + \pi_2 - \pi_1\pi_2 - \pi}{[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}}. \tag{A2}$$

The coherence condition (5b), which imposes bounds on π in terms of π_1 and π_2, implies that ρ is bounded as well. Specifically, for any fixed π_1 and π_2,

$$r \leq \rho \leq R, \tag{A3}$$

where

$$r = \frac{\pi_1 + \pi_2 - \pi_1\pi_2 - \min\{\pi_1 + \pi_2, 1\}}{[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}}, \tag{A4}$$

$$R = \frac{\pi_1 + \pi_2 - \pi_1\pi_2 - \max\{\pi_1, \pi_2\}}{[\pi_1(1 - \pi_1)\pi_2(1 - \pi_2)]^{1/2}}. \tag{A5}$$

Two observations are noteworthy. First, when probabilities π_1 and π_2 for the subperiods are fixed, the correlation between subevents decreases linearly with probability π for the period. Second, when π_1 and π_2 are fixed, the correlation can vary within the interval $[r, R]$, which may be narrower than the interval $[-1, 1]$ because $-1 \leq r \leq R \leq 1$. Examples are shown in Table 3.

The notion of stochastic dependence is formalized in definitions 1 and 2, section 4. The correlation coefficient ρ turns out to be the sufficient and necessary measure of stochastic dependence. This fact is stated without proof.

Fact A. The subevents are (i) stochastically independent if and only if $\rho = 0$, (ii) extremely positive dependent if and only if $\rho = 1$, (iii) extremely negative dependent if and only if $\rho = -1$, (iv) positive quadrant dependent if and only if $0 < \rho \leq 1$, (v) negative quadrant dependent if and only if $-1 \leq \rho < 0$.

REFERENCES

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

DeGroot, M. H., 1970: *Optimal Statistical Decisions.* McGraw-Hill.

Hughes, L. A., and W. E. Sangster, 1979: Combining precipitation probabilities. *Mon. Wea. Rev.,* **107,** 520–524.

Krzysztofowicz, R., and D. Long, 1991: Forecast sufficiency characteristic: Construction and application. *Int. J. Forecasting,* **7,** 39–45.

——, and A. A. Sigrest, 1997: Local climatic guidance for probabilistic quantitative precipitation forecasting. *Mon. Wea. Rev.,* **125,** 305–316.

Lehmann, E. L., 1966: Some concepts of dependence. *Ann. Math. Stat.,* **37,** 1137–1153.

Murphy, A. H., and B. G. Brown, 1984: A comparative evaluation of objective and subjective weather forecasts in the United States. *J. Forecasting,* **3,** 369–393.

Ruth, D. P., and M. R. Peroutka, 1993: The interactive computer worded forecast. Preprints, *Ninth Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology,* Anaheim, CA, Amer. Meteor. Soc., 321–326.

Wilks, D. S., 1990: On the combination of forecast probabilities for consecutive precipitation periods. *Wea. Forecasting,* **5,** 640–650.