NWS-CR-TA-90-20 CRH SSD
JUNE 1990

# PROBABILITY THEORY AND ITS IMPLICATION ON PRECIPITATION FORECASTING

CENTRAL REGION TECHNICAL ATTACHMENT 90-20

Jeffrey K. Last
Meteorological Services Division
National Weather Service Central Region Headquarters
Kansas City, Missouri

James Skowronski
Weather Training Division
3350th Technical Training Group
Chanute Air Force Base, Illinois

## 1. Introduction

Operational meteorologists often express the likelihood of a meteorological event occurring during a forecast period by assigning the event a probability. This is done for two main reasons. First, the use of probabilities is a convenient way to deal with the inherent uncertainty involved in weather prediction. A forecast expressed in probabilistic terms can represent the forecaster's confidence in the forecast. Second, users require estimates of the likelihood of occurrence of weather events in order to make important weather-related decisions.

The most common probability in weather forecasts is the Probability of Precipitation (PoP). At first glance, the PoP seems like a straight forward concept. A 60% PoP means that there is a 60% chance measurable precipitation will fall and a 40% chance it will not. The situation is not quite so clear if PoP forecasts are carried in two consecutive forecast periods.

Consider a forecast with a 50% PoP in both the first and second time periods. Suppose a farmer must apply a fertilizer to his corn field soon, but it must rain within 24 hours of application for the fertilizer to work properly. What is the chance the needed rain will come within the 24-hour period? You might guess 50% since that is the probability of rain for each 12-hour period. You might guess 100% since that is 50% plus 50%. You might guess 25% since that is 50% times 50%. Unfortunately, it is not possible to determine the 24-hour PoP accurately without an understanding of the meteorological situation leading to the two 50% PoP forecasts and some elementary concepts of probability theory.

## 2. Probability Theory and Its Application

Mendenhall et al. (1981) define an experiment as the process by which an observation is made. In weather forecasting, the experiment is tomorrow's weather. Just as the outcome of a laboratory experiment is not known for certain until after the experiment has been completed, tomorrow's weather will not be known for certain until tomorrow. All laboratory experiments have a set of possible outcomes. When forecasting precipitation, the set of all possible outcomes are grouped into two categories. Either measurable precipitation does occur or it does not occur. Probability is defined as the proportion of the time that a specific outcome to an experiment will occur over the long run (Miller and Freund, 1985). The probability of any specific outcome (or a group of outcomes) to an experiment must be between 0 and 1, inclusive. Also, the probability of all possible outcomes to an experiment must be exactly 1.

The PoP is typically expressed as a percentage rather than a decimal. The PoP for a 12-hour period can be interpreted as the percentage of time that measurable precipitation would occur if many such meteorological situations could be sampled. For example, a 50% "chance" (the same as "probability") means that if the same weather situation were to occur ten times, it would produce measurable precipitation five times. In the National Weather Service (NWS), the "average point" probability over the forecast area is used. Many users need the chance of precipitation in a very small area (a farm, construction site, or campground, for example) and point probabilities suit these needs. As shown by Schaefer and Livingston (1990), however, the average point probability is equal to the "expected areal coverage" of the precipitation. (The reader is encouraged to read Schaefer and Livingston to examine the mathematics of this equality, which are beyond the scope of the paper presented here.)

The unconditional probability of an outcome from an experiment is its probability without any knowledge about what happened during the experiment. Each 12-hour PoP is an unconditional probability. While a forecaster uses every available piece of information and skill to develop the forecast, future events remain uncertain. The PoP for each 12-hour period is still based only on what is known before the experiment (tomorrow's weather) occurs. The unconditional probability of an event, A, is denoted $P\{A\}$.

The conditional probability of an outcome from an experiment is its probability given some information about what happened during the experiment. If the additional information is denoted by B, then the conditional probability of A given B is expressed as $P\{A \mid B\}$ and read "the probability of A given B." In terms of a PoP, it is reasonable to assume that the probability in the second 12-hour period could depend on whether or not precipitation occurs in the first period. In other words, given information about whether precipitation will occur in the first 12-hour period, the PoP for the second 12-hour period can be updated. Below are three scenarios where observing measurable precipitation in the first period would affect the likelihood of precipitation in the second period. In all three cases, the PoP in both periods is 50%. Outcome A is the occurrence of measurable precipitation in the first period and outcome B is the occurrence of measurable precipitation in the second period. In these scenarios, we will assume a period is 12 hours long. The same concept, however, applies to 6-hour periods, which are commonly used in enhanced local forecasts.

SCENARIO 1. A slow moving storm system is expected to track just south of the forecast area. The northward extent of the rain is uncertain. If rain occurs in the first period (A occurs) it is virtually guaranteed in the second period. In other words, P{B | A} is nearly 100%.

SCENARIO 2. A fast moving cold front is advancing toward the forecast area. It is scheduled to arrive at about the end of the first period or the beginning of the second period. A narrow band of showers will almost certainly develop along the front. If the showers occur during the first period, they will be over before the second period. In other words P{B | A} is nearly 0%.

SCENARIO 3. Scattered instability showers are developing under cold cyclonic flow aloft. A shower occurring in the first period has nothing to do with whether one occurs in the second period. Here P{B | A} is the same as P{B}. This is a special case where events A and B are INDEPENDENT. Two events are independent if the occurrence of one event in no way alters the likelihood of the other.

All that is needed to combine the 12-hour probabilities of precipitation into a 24-hour PoP are two laws of probability. Mendenhall <u>et al</u>. express the Additive Law of Probability as

$$P\{A \text{ or } B\} = P\{A\} + P\{B\} - P\{\text{both } A \text{ and } B\} \quad (1)$$

and the Multiplicative Law of Probability as

$$P\{\text{both } A \text{ and } B\} = P\{A\} \times P\{B \mid A\}. \quad (2)$$

Substituting (2) into (1) yields

$$P\{A \text{ or } B\} = P\{A\} + P\{B\} - (P\{A\} \times P\{B \mid A\}). \quad (3)$$

Equation 3 allows estimation of 24-hour probabilities of precipitation for the three scenarios described above since all terms on the right hand side of (3) are known. A 24-hour PoP is the probability that measurable precipitation occurs within the 24-hour period. This can happen if precipitation occurs in the first 12 hours OR the second 12 hours, which is what P{A or B} denotes. In scenario 1, P{A or B} is 50%; in scenario 2, P{A or B} is 100%; and in scenario 3, P{A or B} is 75%.

Situations 1 and 2 are the extreme cases. In both scenarios, simplifying assumptions led to conditional probabilities of 0% and 100%. While these estimates are not totally realistic, they do serve to define the limits of the 24-hour PoP and illustrate that two sets of identical 12-hour probabilities, taken in different meteorological contexts, can yield widely different 24-hour PoP estimates. Additional insight can be gained from examining each situation in detail.

In scenario 1, the main uncertainty in the forecast is WHETHER PRECIPITATION WILL OCCUR AT ALL. In this situation, the PoP reflects the uncertainty in the evolution of the synoptic scale flow pattern. In other words, the situation is too close to call and the only recourse is to wait and see if the system affects the area. In situations like these, the 12- and 24-hour PoP are identical.

In scenario 2, the main uncertainty in the forecast is WHEN PRECIPITATION WILL OCCUR. Here the overall weather situation is understood, and precipitation is likely, but the timing is in question. In this type of situation, the 24-hour PoP is much larger than each individual 12-hour PoP since confidence in precipitation is high, but confidence in time of occurrence is low.

In situation 3, the main uncertainty in the forecast is WHERE PRECIPITATION WILL OCCUR (more specifically, how widespread the precipitation will be). It is a safe bet (high probability) that scattered showers will occur in the area, but the PoP reflects the likelihood that a shower will occur at a given location. This is a function of the areal coverage and temporal distribution of the precipitation. The larger the areal coverage and/or the forecast period, the larger the probability that rain will affect any given location. Since scattered showers are reasonably random events, their occurrence at one time does not affect their likelihood at another. The 24-hour PoP in this case can be expressed as

$$P\{A \text{ or } B\} = P\{A\} + P\{B\} - P\{A\} \times P\{B\}. \qquad (4)$$

In general, the 24-hour PoP must be greater than the larger of the two 12-hour probabilities and smaller than the sum of the two 12-hour probabilities. If the sum of the two 12-hour PoP's exceeds 100%, then the most the 24-hour PoP can be is unity. For example, if two successive 12-hour POP forecasts were 40% and 70%, then the 24-hour PoP must be between 70% and 100%. Note, however, that it would not make meteorological sense to insert these probabilities into scenarios 1 and 2. Each case must be evaluated based on the meteorological situation at hand.

The 12- and 24-hour probabilities of precipitation can differ widely. Often forecast wording can help clarify this somewhat complicated situation for the general public.

## 3. Operational Significance

It is important that forecasters understand the implications of their PoP forecasts and know how to correctly use them. According to the Weather Service Operations Manual, Chapter C-11, paragraph 8.3.1.b., the PoP is:

> "The likelihood of occurrence of a precipitation event at any given point in the forecast area.... The time period to which the PoP applies must be clearly stated."

An important and often misunderstood element of this concept is the relationship between the PoP and the forecast time period.

A common misuse of the PoP and time periods is to combine them by saying, for example, "Probability of rain...40 percent tonight and tomorrow." Does the forecaster mean a 40% chance of rain tonight AND a 40% chance tomorrow? If this is the case, then the 24-hour PoP is somewhere between 40% and 80%. If both events are independent, the PoP for the 24-hour period can be computed explicitly and is 64% (see the Appendix for the calculation). Maybe the forecaster meant that the 24-hour PoP was 40%. In this case, the 12-hour PoP's can be anywhere between 20% and 40%. In either case, considerable ambiguity exists over exactly what message the forecaster intended to convey. The PoP must pertain to a clearly defined time period that is understood by the user. In this example, the forecast should have read, "Probability of rain...40 percent BOTH tonight and tomorrow." What appears to be an insignificant point actually makes a significant difference in the forecast's meaning.

During the summer, convective precipitation is common. It is reasonable to treat convective probabilities of precipitation in successive time periods as independent events (as in Scenario 3). Consider the following forecast: "A chance of thunderstorms through tomorrow. Rain chance...20 percent today...20 percent tonight and 20 percent tomorrow." What the forecaster is saying is that there is nearly a 50 (48.8) percent chance of measurable rain falling sometime during the 36-hour period.

Although the next example may not be realistic, it dramatically demonstrates probability theory in action. A forecaster places a 10% chance of rain over 14 consecutive 12-hour periods (seven days). The probability of measurable rain falling anytime during the week turns out to be 77% (the Appendix contains the calculation). This is probably not what the forecaster intended at the beginning of the week.

## 4. Summary

A survey done by Murphy and Winkler (1974) has shown that the public wants probabilities in forecasts. Murphy (1977) has shown that the value of probabilistic forecasts generally is equal to or greater than the value of categorical forecasts (i.e., descriptive words) to users. Decisions by farmers, constructions workers, golfers, etc., are often based on the probabilities given in weather forecasts.

Elementary probability theory demonstrates various relationships among probabilities, and forecasters should be aware of these relationships. It is equally important that the user understand what a forecaster means when the forecast includes a probability. The correct use of the PoP ensures that the public has the most useful product possible.

## 5. Acknowledgement

The authors would like to thank Dr. Joseph Schaefer, Warren Sunkel and Kenneth Rizzo for their very helpful comments.

6. References

Court, A., 1952: Some new statistical techniques in geophysics. Advances in Geophysics, Vol. 1. Academic Press, Inc., 45-85.

Mendenhall, W., R. L. Scheaffer, and D. D. Wackerly, 1981: Mathematical Statistics with Applications. Duxbury Press, 686 pp.

Miller, I., and J. E. Freund, 1985: Probability and Statistics for Engineers. Prentice-Hall, Inc., 530 pp.

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. Mon. Wea. Rev., 105, 803-816.

___________, and R. L. Winkler, 1974: Probability forecasts: a survey of National Weather Service forecasters. Bull. Amer. Meteor. Soc., 55, 1449-1453.

Schaefer, J. T., and R. L. Livingston, 1990: Operational implications of the "probability of precipitation." Wea. and Forecasting, in press.

## APPENDIX

Probabilities for the occurrence of at least one in a series of successive independent events can be computed easily. This is useful for determining the chance of any measurable precipitation during a series of forecast periods where the individual probabilities of precipitation are independent.

A series of 12-hour precipitation forecasts can be grouped so that, overall, two possible outcomes exist. Either rain does occur sometime during at least one of the forecast periods, or it does not occur during any of the forecast periods. The total probability of all outcomes to an experiment must be exactly 1. Since we have categorized all possible outcomes into two groups, the

$$P\{\text{Rain}\} = 1 - P\{\text{No Rain}\}. \qquad (1)$$

Mathematically, P{No Rain} is much easier to calculate than the various scenarios leading to some rain. There is only one way to get no rain at all--each period must be dry. The probability that a period is dry is 1 - PoP.

If only two periods are considered, the probability of no rain in the first period (A) AND no rain in the second period (B) is

$$P\{\text{both A and B}\} = P\{A\} \times P\{B \mid A\}. \qquad (2)$$

But, since the events are independent, $P\{B \mid A\} = P\{B\}$, thus (2) reduces to

$$P\{\text{both A and B}\} = P\{A\} \times P\{B\}. \qquad (3)$$

Similarly, the probability of a series of n independent events occurring can be found by multiplying all the individual unconditional probabilities together

$$P\{\text{A and B and C}\ldots\} = P\{A\} \times P\{B\} \times P\{C\}\ldots\ . \qquad (4)$$

Applying (4) to precipitation forecasts as described above,

$$P\{\text{No Rain}\} = (1-PoP_1) \times (1-PoP_2) \times (1-PoP_3)\ldots\ . \qquad (5)$$

By combining (1) and (5), we have a generic equation for a series of n successive forecast periods

$$P\{\text{Rain}\} = 1 - (\ (1-PoP_1) \times (1-PoP_2) \times (1-PoP_3)\ldots). \qquad (6)$$

Applying (6) to two successive 12-hour probabilities of 40%, the results are

$$\begin{aligned} P\{\text{Rain}\} &= 1 - (\ (1-PoP_1) \times (1-PoP_2)\ ) \\ &= 1 - (\ (1-.40) \times (1-.40)\ ) \\ &= 1 - (\ (.60) \times (.60)\ ) \\ &= 1 - .36 \\ &= .64 \text{ or } 64\%. \end{aligned}$$

As noted by Court (1952), if all of the PoP's are equal (i.e., $PoP_1 = PoP_2 = \ldots = PoP_n$), then

$$P\{Rain\} = 1 - (1-PoP)^n. \qquad (7)$$

Thus, for 14 consecutive 12-hour periods of 10% PoP,

$$P\{Rain\} = 1 - (1-.10)^{14} = 1 - (.90)^{14} = .77 \text{ or } 77\%.$$

Caution must be exercised as (6) and (7) are valid only for INDEPENDENT events.