THE PROBABILITY OF RAIN

Robert G. Miller *Life Insurance Agency Management Association*

THE NOTED mathematician Norbert Wiener has referred to meteorology as one of the semi-exact sciences. In his book *Cybernetics* he states:

> . . . in meteorology, the number of particles concerned is so enormous that . . . if all the readings of all the meteorological stations on earth were simultaneously taken, they would not give a billionth part of the data necessary to characterize the actual state. . . . [A]ll that we can predict at any future time is a probability distribution . . . and even this predictability fades out with the increase of time.

An example of one of Professor Wiener's probability distributions might be the 70% issued by a weather forecaster on radio or television as his estimate of the probability of rain.

Some people feel lost in the face of a probability forecast. "Just tell me the answer!" they say. But, like many other forecasts of events in life, the future state of the weather is uncertain; no one knows for sure what

will happen. Consequently, a forecast that admits to uncertainty seems appropriate. Such a forecast gives extra information to people who may want to take action in the face of weather threatening their pocketbooks. For example, suppose that it costs $100 to take preventive measures against a threatening storm and that the loss if the storm occurs is likely to be about $300. If the chance of the storm is only 0.1, then the expected loss is $30, which is less than the cost of the preventive measure; it would be uneconomical to take steps. But if the chance of the storm were 0.9, the expected loss is $270, and preventive measures look worthwhile.

The ordinary citizen, deciding whether to go out equipped for rain, will consider inconvenience versus risk of drenching and find a probability to use as a cutoff point. For example, he may decide to equip for rain when the probability of rain is 0.5 or higher, otherwise not.

Probability distributions may be arrived at in various ways. The most common method is to use the human judgment of an experienced weather expert. He considers all the evidence and on the basis of his experience chooses a number that he thinks expresses the chance of rain. Another way of generating a probability distribution is to apply statistical methods to weather data stored in government archives. This essay describes a method for arriving at such a probability distribution based on the statistical evidence of past years.

POSSIBLE FORECASTS

As an example, we may want to estimate at 7:00 AM (0700 EST) each day the probabilities for each of five possible precipitation conditions at Hartford, Connecticut, during the next six hours. The five conditions listed in Table 1 are: dry, a little rain, a little snow, rain, and snow. The prediction consists of five numbers, adding to 1.0 and representing the probabilities of each of the five possible outcomes. For example, the numbers 0.4, 0.2, 0.2, 0.1, 0.1 mean a 40% chance of dry weather, a 20% chance each of a little rain or snow, and a 10% chance each of substantial rain or snow. The prediction might group the last four numbers together and report a 60% precipitation probability. Of course, one or the other of the five weather possibilities *must* occur, and we do not report more than one for the period, even though we sometimes have "snow changing to rain."

LEANING ON DATA FROM OTHER PLACES

Meteorologists have found from experience that observing present weather conditions over a fairly large region enables them to predict future weather conditions at points within the region. They have not ordinarily made forecasts with quantitative probabilities, though "a slight chance of rain" was a common way of describing a small probability without quantifying it. The

TABLE 1. Detailed Definition of Five Precipitation Categories for Hartford, Connecticut

CATEGORY	CONDITIONS
(1) Dry	No precipitation of any kind over the period 0701 EST–1300 EST.
(2) Little rain	Rain or freezing rain reported at some time over the period 0701 EST–1300 EST in the amount of at least a trace but not more than 0.05 inch. No snow or sleet reported at any time over this six-hour interval of time.
(3) Little snow	Snow or sleet reported at some time over the period 0701 EST–1300 EST in the amount of at least a trace but not more than 0.05 inch of melted water equivalent.
(4) Rain	Rain or freezing rain reported at some time over the period 0701 EST–1300 EST in the amount of greater than 0.05 inch. No snow or sleet reported at any time over this six-hour interval of time.
(5) Snow	Snow or sleet reported at some time over the period 0701 EST–1300 EST in the amount of greater than 0.05 inch of melted water equivalent.

statistical approach, in attempting to refine the meteorologist's forecast, will also use data from a fairly large region around the point of interest; in the case of our example, from weather stations with Hartford, Connecticut, considerably east of center because of the weather's general movement from West to East in the Northern Hemisphere. The dots on the map in Figure 1 indicate the locations in the network; Table 2 lists the stations in the network.

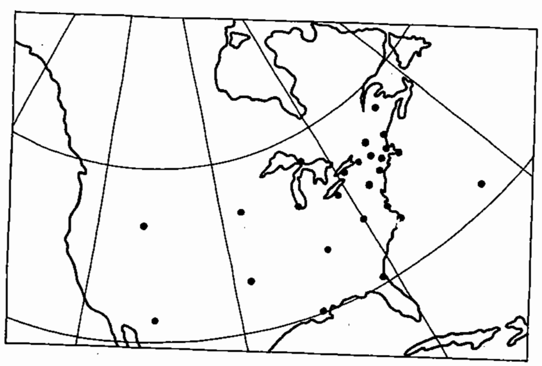

FIGURE 1
25-station network

TABLE 2. 25 Network Locations

Hartford, Conn.	Hatteras, N. C.
Caribou, Me.	Jacksonville, Fla.
Portland, Me.	Sault Ste. Marie, Mich.
Boston, Mass.	Chicago, Ill.
Nantucket, Mass.	Cleveland, O.
Burlington, Vt.	Knoxville, Tenn.
Albany, N. Y.	New Orleans, La.
Buffalo, N. Y.	Sioux Falls, S. D.
New York, N. Y.	Oklahoma City, Okla.
Syracuse, N. Y.	Pocatello, Ida.
Harrisburg, Pa.	Tucson, Ariz.
Norfolk, Va.	Saint George, Bermuda
Roanoke, Va.	

VARIABLES USED IN THE FORECAST

Out of the many weather elements observed hourly at each of these locations, seven were used to characterize the state of the weather at forecast time, 0700 EST. These seven, listed in Table 3, are: barometer reading and its three-hour change, temperature, moisture, wind direction and velocity, and cloud cover. They were chosen because as a group they seemed to predict the weather at least as well as others that were available. The inclusion of 0700 EST rain or snow conditions would have enhanced the predictions but they were not among the elements available in this sample study.

As data for developing the procedure we used the 1096 daily 0700 EST observations of each weather element for the years 1951–53 for all locations. (These are stored in the U.S. Weather Bureau files in Asheville, North Caro-

TABLE 3. Meteorological Elements

Sea level pressure (millibars)
Past three-hour change in sea level pressure (millibars)
Dry bulb temperature (degrees Fahrenheit)
Temperature-dew point depression (degrees Fahrenheit)
East-West wind component (knots)
North-South wind component (knots)
Total cloud cover (tenths of the sky covered)

lina.) An additional sample of 221 observations from 1954 are used in an independent test of the chosen forecasting procedure, described later.

A typical 0700 EST observation at one of the 25 locations, say, Hatteras, North Carolina, for one of the 1317 days in the two samples might have been:

Element	*Measurement*
Sea level pressure	1000.2 millibars
Three-hour pressure change	−1.0 millibars
Dry-bulb temperature	55° Fahrenheit
Dew-point depression	0° Fahrenheit
Wind	Northeast at 25 knots
Cloud cover	10/10

The pressure measurement of 1000.2 millibars (1000.2 $\times$ 1000 dynes per square centimeter) indicates low pressure conditions because the normal value is 1017 millibars.

A drop in pressure of one millibar was observed over the last three hours, suggesting the approach of even lower pressure and a worsening of weather conditions.

The air temperature was observed to be 55° Fahrenheit. The dew-point temperature (the temperature at which the air becomes saturated) and the air temperature were precisely the same (a zero depression), so that it was a humid morning.

The wind blew from the Northeast (45°, at 25 knots. We can, with a little trigonometry, express the wind in terms of its East-West and North-South components:

u (East-West component) = 17.68 knots
v (North-South component) = 17.68 knots

On this morning, moist air from over the ocean was carried over Hatteras generally making for precipitation. A sky cover of ten-tenths of clouds was also observed, further enhancing the chance of precipitation.

THE STATISTICAL TECHNIQUE

The statistical method used to predict the probability distribution of precipitation at Hartford, Connecticut, has the technical name of *multiple discriminant analysis*. Elsewhere in this book (see the essay by Howells) the general idea of discriminant analysis is described. Ordinarily we find several variables, or measurements, each of which is related to the presence or absence of the categories we are predicting—here dry, a little rain, a little snow, rain, and snow. Then it may be possible to make, out of the several measurements, a single measure that is a better predictor than any individual one. Often the better predictor is a weighted sum of the individual variables. Two important steps are required: first, to select a few from among all the variables and, second, to determine the weights to get a good predictor.

Howells wanted to decide or "predict" whether a bone was human or not and he used a single weighted sum of properties of the bone. We have more categories and our forecasts will be improved by using more than one weighted sum of the same variables to summarize the predictive data.

Let us describe our steps in more detail:

Step 1: Selecting the Variables. Using statistical procedures, we selected out of the original 175 weather variables (7 elements at each of 25 locations) those 16 that contained most of the predictive information about the subsequent six-hour precipitation conditions at Hartford, Connecticut (see Table 4). One way to do this is to provide a high-speed computer with the data and give it a rule for deciding which among several collections of variables makes the best predictions. In one approach, the computer tries all 175 variables one at a time and chooses the best of these. Next it tries combining each of the remaining 174 variables with the first one selected, and chooses the second variable that goes best with the first one chosen, and so on until adding further variables doesn't help much. If we had used data from years different than the ones we chose, no doubt a somewhat different set of variables would have arisen, and the order would probably have been different.

TABLE 4. Selected Predictors in Order of Selection

STATION	ELEMENT
Boston, Mass.	Total cloud cover
Portland, Me.	Past three-hour pressure change
Sault Ste. Marie, Mich.	Dry-bulb temperature
Hartford, Conn.	Temperature–dew-point depression
Buffalo, N. Y.	Dry-bulb temperature
Boston, Mass.	East-West wind component
Hatteras, N. C.	North-South wind component
Norfolk, Va.	Past three-hour pressure change
New York, N. Y.	Dry-bulb temperature
Portland, Me.	North-South wind component
Nantucket, Mass.	North-South wind component
Norfolk, Va.	Dry-bulb temperature
Oklahoma City, Okla.	North-South wind component
Caribou, Me.	Dry-bulb temperature
Boston, Mass.	Dry-bulb temperature
Albany, N. Y.	North-South wind component

Because Hartford is only 100 miles from Boston, cloud cover at the latter is a reasonable variable. That temperatures appear as six variables is not remarkable, though most of us would not have expected a Michigan temperature to appear so early on the list, even though weather usually moves from West to East. Having temperatures at a variety of places gives the discriminant a way of measuring the changing temperatures from one station to another. That changes in the barometer matter is to be expected, but that the computer would choose stations to the East and to the South is not at all obvious. Lacking a direct measure of 0700 EST rain or snow, the selection of a moisture variable (dew-point depression) at Hartford seems a very reasonable choice. The reader will not be surprised to see that among these 16 variables, six measure wind direction and velocity, four at stations near Hartford. The selection of these wind measurements indicates an attempt to include circulation characteristics as well as the location of weather fronts.

While selecting the 16 variables to be used, the computer also computed two sets of 16 weights to be attached to them. They are used to give two weighted sums, or *discriminant functions*, to be used as predictors. In principle, more than two sets of weights could be used, but we stopped with two because they seemed to wrap up nearly all the information.

Step 2: Finding the Weights. The discriminants are scores that predict the weather. The first discriminant forecasts *dryness*—a high score forecasts very dry, a low score considerable precipitation—rain or snow. The second discriminant helps forecast the kind of precipitation—high scores implying snow, low scores rain, middle scores dry; let us call this the *snowiness* score.

For each historical observation, the numerical score for dryness and the one for snowiness were plotted on a graph (see Figure 2) and each point was labeled to show the actual weather that occurred. Once this was done, we found that the (dryness, snowiness) scores for a given weather, say snow, clustered around a point and that larger and larger ellipses drawn around such a central point included more and more of the observations of that weather. Figure 2 shows all the points corresponding to historical instances where a little rain fell. It also shows the ellipse that includes 50% of these points. Each category has such a scatter of points and has a corresponding ellipse. The choice of this particular ellipse and its orientation are guided by some calculational considerations we need not go into here.

Figure 3 shows the placement and shape of five ellipses, each of which contains 50% of the points corresponding to its kind of weather. These 50% ellipses overlap considerably. The overlap illustrates why one may not be able to give a simple "it will rain" as a forecast. Only the ellipse for snow stands a bit off from the others. Remember, these are only 50% ellipses, so the snow points fall among the others more than Figure 3 may suggest; nevertheless, "heavy snow" observations are found mostly in the upper left-hand region of the graph (in and around the ellipse labeled 5). Most of

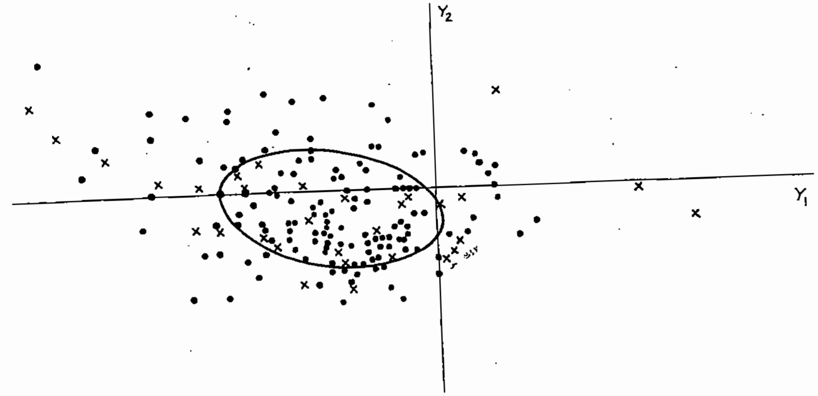

FIGURE 2

Empirical distribution of sample points (dots) from the original 1096 observations and sample points (crosses) from the later 221 observations for those cases of a little rain, group 2. About half the points fall inside the ellipse

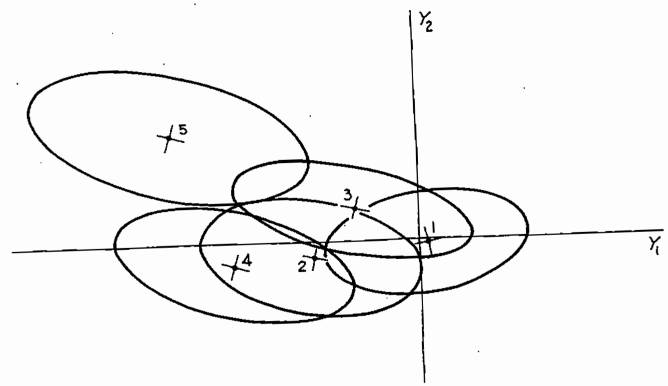

FIGURE 3

The ellipses shown encompass about 50% of the observations of the corresponding precipitation categories: (1) dry, (2) a little rain, (3) a little snow, (4) rain, (5) snow the observations located near the intersection of the axes (inside ellipse 1) are for "no precipitation."

Step 3: Calculating the Probabilities. The next step is to calculate the probabilities of the precipitation categories for each of the 1096 observations and for the 221 independent sample observations to be used to test the method. This is accomplished as follows.

As we have seen above, the set of observed conditions of the 16 selected weather elements at 0700 EST on any particular day corresponds to a point in Figure 3. The probability of each kind of weather is determined by where the point lies. If it lies inside the ellipse labeled 1, then the probability that there will be no precipitation within the next six hours will be high, say, 0.8 or 0.9; whereas if the point lies inside the ellipse labeled 5, the chances for snow will be high. Naturally, computing the probabilities requires a complicated procedure, but we can give the idea if not the exact method.

Suppose, for a moment, our 16 variables produced a point somewhere in Figure 3. You could swing a small circle around that point just large enough to include, say, 100 of the 1096 historical points. Suppose 60 were in category 1, 20 in category 2, 10 in category 3, 9 in category 4, and 1 in category 5. Then we could forecast 0.6 chance of no precipitation, 0.2 chance of a little rain, and so on, and practically 0.0 chance of heavy snow. The actual method uses more mathematics, but this idea is adequate for understanding.

In place of providing all of these predicted probabilities, we give two tables that summarize the predictions of the 1096-case sample (see Table 5) and the 221-case sample (see Table 6). These tables, for particular ranges of predicted probabilities, include: the number of forecasts F made within each group of weather conditions and within each range of probability (notice that across each row the Fs sum to 1096, because in each case *some* probability was assigned to each group), the number of actual occurrences U of the designated group when the predicted probability was in that range, and the sum of the probability over all F forecasts, ΣP, gives the expected or average number of occurrences.

Let us give a few examples of reading Table 5. In row F of the "dry" group, the first entry tells us that 50 of the 1096 historical points would have forecast "dry" with probabilities at least 0 but less than 0.1; the next entry tells us that 80 points would have forecast "dry" with probability at least 0.1 but less than 0.2, and so on. In row U of the dry group, the first entry tells us that in 2 of the 50 cases in which we predicted "dry" with probability between 0 and 0.1, "dry" actually occurred. If we look at the total U for the "dry" group, we see that, in all, "dry" occurred in 817 of the 1096 observations. Comparing this line with the third gives us an idea of the comparison between occurrences and probabilities as we pointed

TABLE 5. Summary of Probability Predictions (1096-Case Sample)

	Group		$0.0 \le P < 0.1$	$0.1 \le P < 0.2$	$0.2 \le P < 0.3$	$0.3 \le P < 0.4$	$0.4 \le P < 0.5$	$0.5 \le P < 0.6$	$0.6 \le P < 0.7$	$0.7 \le P < 0.8$	$0.8 \le P < 0.9$	$0.9 \le P \le 1.0$	Total
Dry	1	F	50	80	37	58	37	42	22	82	145	543	1096
		U	2	10	8	25	16	23	14	68	122	529	817
		ΣP	2.36	12.88	9.60	21.36	16.84	23.12	14.60	63.48	125.76	528.28	818.28
Little rain	2	F	616	212	89	130	30	18	1	0	0	0	1096
		U	11	30	22	45	17	10	0	—	—	—	135
		ΣP	12.88	33.24	23.00	46.84	13.64	9.76	0.64	—	—	—	140.00
Little snow	3	F	1034	46	15	1	0	0	0	0	0	0	1096
		U	19	6	4	0	—	—	—	—	—	—	29
		ΣP	17.56	7.04	3.92	0.32	—	—	—	—	—	—	28.84
Rain	4	F	821	100	54	63	20	20	12	6	0	0	1096
		U	8	11	16	20	11	13	10	3	—	—	92
		ΣP	6.84	15.52	13.88	22.16	9.20	11.32	8.00	4.32	—	—	91.24
Snow	5	F	1031	38	12	12	3	0	0	0	0	0	1096
		U	4	5	2	9	3	—	—	—	—	—	23
		ΣP	3.28	5.56	3.08	4.32	1.40	—	—	—	—	—	17.64

TABLE 6. Summary of Probability Predictions (221-Case Sample)

	Group		$0.0 \le P < 0.1$	$0.1 \le P < 0.2$	$0.2 \le P < 0.3$	$0.3 \le P < 0.4$	$0.4 \le P < 0.5$	$0.5 \le P < 0.6$	$0.6 \le P < 0.7$	$0.7 \le P < 0.8$	$0.8 \le P < 0.9$	$0.9 \le P \le 1.0$	Total
Dry	1	F	12	17	7	11	10	10	4	22	27	101	221
		U	—	1	5	4	1	7	3	17	23	94	155
		ΣP	0.56	2.52	1.76	3.92	4.60	5.60	2.60	17.00	23.28	98.20	160.04
Little rain	2	F	114	48	22	21	11	5	—	—	—	—	221
		U	6	8	4	8	4	3	—	—	—	—	33
		ΣP	2.52	7.24	5.68	7.28	5.00	2.72	—	—	—	—	30.44
Little snow	3	F	211	6	4	—	—	—	—	—	—	—	221
		U	4	2	2	—	—	—	—	—	—	—	8
		ΣP	3.68	0.96	1.04	—	—	—	—	—	—	—	5.68
Rain	4	F	159	24	12	11	3	9	2	1	—	—	221
		U	3	5	1	5	1	3	1	1	—	—	20
		ΣP	1.48	3.64	3.08	3.76	1.36	4.92	1.28	0.72	—	—	20.24
Snow	5	F	203	10	3	4	1	—	—	—	—	—	221
		U	1	1	—	2	1	—	—	—	—	—	5
		ΣP	0.52	1.40	0.76	1.44	0.48	—	—	—	—	—	4.60

out in Table 5. On 2 occasions it was actually "dry" when the probability of "dry" was only between 0 and 0.1. The sum of the computed probabilities for the 50 cases falling into this category was 2.36. That means the method forecasts 2.36 instances of dry weather even though it forecasts none of the 50 as likely. The number observed, 2, is very close to the computed count 2.36. The largest discrepancy is in category 5, snow, for probability between 0.3 and 0.4, where there were 9 occurrences versus 4.3 predicted. By and large the agreement is good. This shows that we can find a way to forecast the past based on that same past. But will the same method work for another period of time?

Step 4: Validating the Approach. In Table 6, we show the result of applying the same technique to the new 221 observations which were not used to construct the discriminant functions. Again the agreement between the second and third lines is usually close, and this validates the method. Thus, Tables 5 and 6 show us a lot about the weather in Hartford and the forecasts. Seventy-five percent (817/1096) of the time it is dry, 12% (135/1096) of the time there is some rain, 2% of the time it snows a little, 8% it rains heavily, and 2% it snows heavily. As for the forecasts, often "dry" can be predicted with high probability; indeed about half the forecasts are "dry" with probability greater than 0.9. Because the sum of the probabilities must be 1.0 for each forecast, it follows that forecasts of low probability can be given frequently for the other four categories. But it is a general feature of both tables that a specific category of precipitation is never forecast with probability even as great as 0.8. One requirement for obtaining higher probabilities of these rare events is a much larger data sample. A smaller, more concentrated ellipse would then be capable of encompassing the 100 or so sample points needed to estimate the probabilities.

CONCLUDING REMARKS

From the results obtained the following observations can be made:

(1) From Figure 3, it can be seen that the horizontal axis orients the precipitation categories such that larger amounts of precipitation, irrespective of type, lie to the left, and lower amounts to the right. This is contrasted with the vertical axis, which appears to "discriminate" the snow groups (categories 3 and 5) from the rain groups (categories 2 and 4).

(2) It may be of interest to know how many correct predictions were made of the conditions—precipitation or no precipitation—in the test sample. Taking those situations in which the probability of no precipitation was 0.5 or greater as an arbitrary criterion for categorically predicting no precipitation, the number of forecasts of "no precipitation" was 164

out of 221, with 144 correct. The number of forecasts of "precipitation" was 57 out of 221 with 46 correct. Altogether there were 190 out of 221 correct, or 86% accurate. This compares favorably with the 957 out of 1096, or 87% for the sample used to develop the relationships.

(3) There appears to be good agreement between U, the number of observed precipitation events, and ΣP, the sum of the probabilities in each of the five categories in each of the ten columns of Tables 5 and 6. For example, Table 6 shows that for the 21 times that category 2 (light rain) was predicted with a probability from 0.3 to 0.4 the light rain was observed to happen eight times while its expected number of occurrences was 7.28 times. This good agreement can be seen in all of the independent data results of Table 6. The discrepancies that do exist can be largely attributed to sampling fluctuations.[1]

The method of discriminant analysis has been instrumental in enabling the prediction of weather probabilities objectively. However, operational weather prediction utilizing the method requires a computer to do the calculations. Alternative discriminant methods have been developed which do not have such requirements. These alternative methods possess the following features:

(1) Qualitative as well as quantitative predictors can be used.

(2) Probabilities can be predicted for categories which are not necessarily mutually exclusive.

(3) More varied shapes of clustered points than those characterized by simple ellipses can be dealt with.

(4) Operational probabilities can be obtained directly by merely adding a small set of numbers together.

(5) Large numbers of variables with many sample cases may be processed with ease.

(6) Results are more easily interpreted.

(7) Missing and erroneous or incomplete data are handled systematically.

Prospects for the future are that discriminant type methods will make it possible for users to request weather probabilities by telephone and to receive a voice response directly from a computer.

[1] For a detailed exposition of this and one other weather forecasting example using statistical methods see R. G. Miller. 1962. *Statistical Prediction by Discriminant Analysis*. Meteorological Monographs, No. 25. Boston: American Meteorological Society.