The Probability Index

by David Giovannoni
(7 pages)

Originally published as:

Giovannoni, David. "The Probability Index." Audience Research Report. Washington, DC: National Public Radio, April 6, 1981.

THE PROBABILITY INDEX

by David Giovannoni

Introduction

Doing audience research is a lot like being a bookie. In each profession, the practitioner must possess an intimate understanding of the odds that any event will happen. Yet while professional integrity, the rules of the game, and circumstance preclude the direct control over any single outcome's chance of occurrence, persons unable to apply total understanding of the odds to their own advantage will not long prosper in their chosen profession.

In this Audience Research Report, we will examine how audiences can be studied in "probability" terms. We argue that our ability to predict and understand our stations' appeals and our audiences' behaviors can be improved substantially by knowledge of other attributes, sometimes seemingly unrelated. This report relies heavily on fictitious data for illustrative purposes. However, the subsequent understanding of the probability index will provide a valuable tool, which will be used extensively in following editions to examine actual audience data.

The Probability Index

A census taker knocks on the door of a statistician's house. The statistician welcomes her in and they begin the interview. After the first few questions, the following dialogue ensues:

Census taker:	How many children do you have?
Statistician:	Two — Chris and Pat.
Census taker:	And their sexes?
Statistician:	Chris is a boy.

What is the probability that Pat is a girl? Traditional wisdom tells us that "the odds are even" — Pat has a 50% chance of being a girl and a 50% chance of being a boy. But we have one other piece of information: Chris is a boy. Does this change the odds of Pat being a girl? Yes it does. Here's how.

Given two children, the complete list of sex-combination possibilities is as follows:

BOY — BOY	25%
BOY — GIRL	25%
GIRL — BOY	25%
GIRL — GIRL	25%

Each of these combinations has an equal chance of occurring. If all we know is that the statistician has two children, the odds are 25% that both children are boys, 25% that both children are girls, and 50% that they are of opposite sexes.

So far so good. But in the opening puzzle we are given the knowledge that Chris is a boy. This fact cuts out one of the above combinations — girl — girl — and now our possibility table looks like this:

BOY - BOY	33%
BOY - GIRL	33%
GIRL - BOY	33%

Each combination is equally as likely. But now Pat has two chances of being a girl (boy — girl and girl — boy) and only one chance of being a boy (boy — boy). Therefore, the odds that Pat is a girl, *given the fact that Chris is a boy*, are two to one, or 66%. The odds of Pat being a girl have increased![1]

By knowing one child's gender, we have substantially increased our odds of correctly predicting the other child's gender. We can create an index to show how much our odds have increased. Without knowledge of Chris' gender, the odds of Pat being a girl are 50%. Knowing Chris' gender increases Pat's feminine odds to 66%. The index of improved odds is calculated as follows:

$$\frac{\text{Odds given additional information}}{\text{Odds without this information}} \times 100 = \text{Index.}$$

In our particular case, $\frac{66\%}{50\%} \times 100 = 133.$

In other words, the knowledge that Chris is a boy has increased Pat's chances of being a girl by 33%. Note that 133 can be used as a multiplier — i.e., 50% x 1.33 = 66% — which is the probability of Pat being a girl.

Also note that this index can quantify the amount by which odds of one outcome *decrease* with knowledge of another. The index of the odds of Pat being a boy is 67:

$$\frac{33\%}{50\%} \times 100 = 67.$$

In other words, knowing that Chris is a boy decreases Pat's odds of being a boy by 33% (100 - 67 = 33).

Two things are important to understand about this index of increased odds, which we will call the **Probability Index**. First, it is based on a scale of 100. If a new piece of information did not have any effect on the probability of another, then

$$\frac{\text{Odds given additional information}}{\text{Odds without this information}}$$

is always equal to 1, and the index is equal to 1 x 100 = 100. In other words, when an additional piece of information does not affect the predicted outcome of another event, we have not gained predictive ability and the probability index remains at 100.

Second, it is important to distinguish between increasing predictive ability and changing the odds of a given outcome. For example, in the above puzzle, the odds of Pat being a boy *decreased* once we found out Chris' gender; however, this knowledge *increased* our predictive power. The probability of Pat being a boy is the *outcome* (which has decreased), yet our *power to predict* this outcome was increased with additional information.

Interrelation

The reason why our power to predict Pat's gender increased with the knowledge of Chris' gender is due to the fact that the sexes of the two children are *interrelated*.[2] This should not be interpreted to imply that Chris' gender *caused* Pat's — far from it. Interrelation simply means that knowing one fact increases our chances of guessing, or predicting, another.

The interrelations among sets of variables are widely used in ways that affect our everyday lives. Last year, when I turned 26, my car insurance premiums dropped substantially. The reason: 26-34 year old males wreck fewer

[1] For many people, this formal proof still belies intuition. This example is really asking the probability of a set of occurrences. If we knew the position of birth (i.e., Chris was the first child or the oldest child) then the odds of Pat being a girl are 50%. A quick examination of the list of combinations verifies this.

[2] We refrain from using the word "correlated," as this has a similar, yet different, statistical meaning.

cars than 19-25 year old males. Age and car wrecks are highly interrelated. Another example: as one gets older, premiums for health insurance increase. The reason is the same — increased age and the propensity to have doctor and hospital bills are highly interrelated. Just as insurance companies look for the interrelations between demographic variables and the number of claims to be expected, so do scientists look for interrelations among important variables: television viewing and tendencies toward violence, smoking or the consumption of food additives and cancer, education and success in society, etc.

Again it must be emphasized that interrelations do not necessarily imply cause and effect. Just as Chris' gender did not cause Pat's, age does not *cause* car wrecks or illness; television viewing does not *cause* violent behavior; we can not say with absolute certainty that smoking or food additives *cause* cancer. All we need to really understand is that variables are often associated with the occurrence of outcomes; if we understand the degree to which the variables and outcomes are related, then we can better predict the outcomes when given the associated variables.

This relates to examining radio audience in some very important ways. We can, for example, examine the use of radio formats (the outcomes) in light of other demographic or radio usage variables. For instance, given that a person is a teenager (our additional piece of information, the equivalent of knowing Chris' gender in the opening problem) we would guess that s/he is more likely to listen to a rock and roll station than if s/he were not a teenager. In other words, being a teenager and listening to rock and roll radio are highly related. Similarly, given that a person listens to an NPR member station, we are able to predict that s/he is more likely than the general public (who are not NPR member station users) to be well educated. Listening to NPR and being educated are highly related (as we will see in a subsequent Audience Research Report). But let's assume that a person listens to country music on the radio. Is s/he more or less likely to listen to classical music? To jazz? To an NPR member station?

The answers to this last set of questions are very pertinent to the readers of this Report, for if we can understand the use of radio and its formats in light of interrelated demographic and radio usage variables, then we understand to a much greater degree then, now, how audiences perceive and use these formats (and the stations carrying them). The end product is a very powerful tool that will help us to better program, position, and promote our radio station(s). And this is what using audience research in public radio is all about.

The remainder of this Report will trace the creation and application of the **Format Sharing Index**, which is simply our probability index put to use analyzing how radio formats share listening audiences. In a subsequent edition of the Audience Research Report we will examine format sharing in the Chicago and San Francisco markets, as well as format sharing nationally.

The Format Sharing Index

Let's apply what we now know about interrelation of variables and outcomes to a simplified audience research case study. Our analysis will examine the format sharing in the small market of Notown, USA (Population 40,000). Notown has three radio stations and each station carries a single format. The three formats are Rock, News, and Classical.

The first thing we do is look in Notown's

Spring Arbitron book. We find the following information for the Monday-Sunday, 6am-12m daypart:

Table 1
Station Audience in Notown USA
Spring 1981

Station	Cume Persons	Cume Rating
Rock	25,000	62.5%
News	20,000	50.0%
Classical	5,000	12.5%

The book tells us the relative popularity of each format in Notown. As far as format *sharing* goes, there is little the book can tell us. There is no way to tell from these data which format pairs are most highly interrelated.

Because we have taken the book data as far as it will go, we call Tom Church and ask him to do an AID[3] run ascertaining the number of people listening to each pair of formats. AID tells us the following:

Table 2
Number of Persons Who Listen to Two Formats in Notown, USA
Spring 1981

Format Pair	Cume Persons
Rock and News	2,500
News and Classical	4,000
Rock and Classical	1,000

Rock and News are the dominant formats in town. They also have the greatest number of persons listening to both. But this is to be expected — they have the greatest number of listeners to begin with. How much crossover is this in terms of the percent of each format's audience? Some simple number crunching yields the following table:

[3]Arbitron Information on Demand.

Table 3
Format Crossover in Notown USA
Spring 1981

Percent of A's audience who also listen to B:

	Format B		
	Rock	News	Classical
Format A			
Rock	100	50	4
News	63	100	20
Classical	20	80	100

The greatest percentage of sharing is from Classical to News. Eighty percent of all Classical listeners also listen to News sometime during the week. We might be tempted to say that the News station is the Classical station's main competitor; but in keeping with the intention of understanding audience behavior, we emphasize instead the symbiotic, rather than the competitive, nature of the formats. Listening to these two formats seems to be interrelated, i.e., listening to Classical music is highly associated with listening to News.

Two-thirds of the News audience also listens to the Rock station. Twenty percent of the News audience also listens to the Classical station. Given the information from our first table, we would expect these sorts of numbers. The Rock station has the largest single audience in town, so it is not surprising to see more News listeners using it than the Classical station. Similarly, we might also expect to see more Classical listeners also using the Rock station than using the News station for the same reason. However, this is not the case. More Classical listeners listen to the News station than to the Rock station.

What's happening here? Apparently Classical and News have much more "in common" than

Classical and Rock. There seems to be an underlying similarity in audience attraction, or "appeal," between Classical and News which we would not have guessed by chance. Are News and Classical more compatible than Rock and Classical? If so, to what extent? And may other format combinations be even *more* compatible?

We are now beginning to ask the questions which get at the heart of how people split listening across formats; more basically, how are people using radio? What formats are the most compatible to the audience's ears? Note here that we are allowing the audience to define "compatibility" among formats. This is an important concept. As professional broadcasters, we make our own distinctions of format compatibility based on how formats sound, whether they play music or give information, or what type of music they play or information they give, how often, in what style, with what emphasis, etc. These sorts of distinctions may have some applications for us, but we are professional broadcasters — most members of our audiences are not. Therefore, while our definitions of compatibility among format types may be useful to us, *they are meaningless to our audience*. Much more important to our understanding of radio is the way the *audiences* define the formats *by how they use them*. In order to understand how audience listening patterns are associated with format use, we must quantify the degree to which listening to one format relates with listening to another. Sound familiar? Right — this is a problem for our probability index. It works like this.

Suppose you are walking down the main street of Notown. You meet your old friend Sam, who is now a telephone repair person. Sam is on her way to the Notown Cafe and she invites you along for some of that famous Notown chili. How can you refuse?

As you sit on the stool next to Sam waiting for your chili to cool, you begin to wonder... What are the odds that Sam has listened to the News station this week? Since 20,000 of the town's 40,000 persons listen to the News station each week, you guess that she has a 50% chance of being one of them.

So far so good. But instead of asking if she listens to the News station, you ask instead if she listens to the Classical station. She replies that both she and her husband (the Notown Constable) do. Now the odds that she listens to the News station are **80%**. Why? Because 80% of all Classical listeners are also News listeners. The odds have now increased substantially in your favor. In other words, given the knowledge that she is a Classical listener, she is 60% more likely to listen to the News station.[4]

"Sam," you say, "I bet you've listened to the News station in the last seven days."

"As a matter of fact I have," says Sam, in awe of your predictive powers. "Gosh, you audience research people know everything!"

You smile and begin to chew your chili. Due to her husband's occupation, you think better of pointing out the similarities of being an audience researcher and being a bookie.

Yet because of our research, we really do "know the odds." The increase in the odds of predicting a listener of one format, given listening to another format, is our probability index in use as a **Format Index**. In Notown, the odds of listening to News are 50%. Given knowledge of Classical listening, the odds increase to 80%. Our format index is 80/50 — 160. Notice that this works when approached

[4]Index = 80/50 x 100 = 160.

from the opposite direction. 12.5% of all people in town listen to the Classical station, yet 20% of all News listeners also listen to the Classical station. Again, our index is 160 (20/12.5 x 100 = 160). The index is telling us that, in Notown, News and Classical share many more people than expected. As a format pair, they have a much stronger mutual appeal than Rock and Classical. In fact, given that a person listens to Rock, s/he is *less* likely to listen to Classical than expected. A complete set of format sharing indices is expressed in the following table.

Table 4
Format Sharing Indices in Notown USA
Spring 1981

Format Pair	Sharing Index
News – Classical	160
Rock – News	100
Rock – Classical	64

Notice that Rock and News have an "even" index of 100 — knowledge of a person's listening to one does not increase or decrease the chances of his or her listening to the other.

Of course we have used a simple example in our description of the format sharing index. But what we have accomplished — the understanding of how to better examine format sharing among listeners — is a substantial advance from the raw, almost meaningless, numbers in Table 1. Our next step is to ascertain this knowledge on a national level and to apply it to everyday decision making in our own shops. This will be the subject of a subsequent Audience Research Report.

The Simmons Index

The following six issues of the Audience Research Report will present detailed information about the characteristics of the national audience for NPR member stations. These reports rely heavily on the syndicated data gathered by Simmons Market Research Bureau. Since Simmons makes extensive use of the probability index, it is appropriate to explain the Simmons index in terms of the probability index just presented.

Recall the meaning of the probability index: Given one attribute, we are better able to predict the chances of another. The index is created as follows:

$$\text{Index} = \frac{\text{Odds given additional information}}{\text{Odds without this information}} \times 100$$

With the Simmons index, our additional piece of information is that a person is an NPR listener. For instance, we know that 22% of the women in America are 25-34 years old. But surveys of the NPR audience show that 25% of its female listeners are 25-34 years old. Our index of 25/22 x 100 = 115 is read as follows: Given that a woman is an NPR listener, she is 15% more likely to be between the ages of 25-34 years old. This can also be read another way:[5] Given that a woman is 25-34 years old, she is 15% more likely to listen to an NPR member station.

[5] The mathematics are not presented here, as they require knowledge of other information not readily available. Page 9 in the Report shows why this works.

Appendix:
More on Calculating the Format Sharing Index

Many of you may wish to use AID to analyze format/station sharing in your market. Or you may be sitting on the results of a local survey which can stand analysis by this method. For either case, the mathematics of the Format Sharing Index are presented below for your use.

You need to know the following in order to create the index:

A = # of cume persons using Format A.
B = # of cume persons using Format B.
T = # of persons in the market.
S = # of cume persons listening to *both* Format A *and* Format B.

The Format Sharing Index (I) is equal to the Percent of Format A's audience also listening to Format B (S/A) over the percent of people in the market listening to Format B (B/T).[1]

Mathematically:

$$I = \frac{S/A}{B/T}$$

Some algebraic manipulation provides an easier-to-use calculating formula:

$$I = \frac{TS}{BA}.$$

[1] Note that B/T is the format's cume rating.