

Michael Scott Cuthbert

**Counting our Losses:
The Missing Polyphonic Works of the Trecento¹**

Fully understanding a repertory of music involves, above all, having a grasp of its extent. We need to view the repertory as a whole in our minds in order to distill its salient features, its internal subdivisions, and, perhaps above all, the distinctive and wonderful exceptions which give life and development to music. Getting a handle on a repertory is especially difficult when what survives for us to study is distant, or worse, incomplete. We know that our perspective is obscured, our understanding partial. Our conclusions are subject to revision; they are in short, inconclusive.

We would be more assured about our work if we were convinced that we lacked only a little from the repertory, and that what we lacked was similar to what we already had. But understanding the extent of our losses has been considered difficult or impossible by musicologists.

In this section, I consider the size, measured in number of pieces, of various sub-genres of the trecento. I suspected that the information we already had for certain repertories could substantially lessen our uncertainty about the extent of our losses. This section discusses some ways we conceive of missing pieces in a repertory, and ways we might develop methods for estimating the number of missing pieces. It then applies these methodologies to the subject at hand: the various polyphonic genres of the fourteenth century. The section concludes by remarking on other uses of these methods and their applicability to other branches of music scholarship and humanistic studies.

There are several reasons why we should consider the total size of an incomplete repertory. The number of missing pieces gives us an estimate of how fruitful we expect searches for new manuscripts to be. As is noted elsewhere in this thesis, the rate of discovery of fragments has increased rather than declined over the last forty years, and we have no reason to expect that the rate will drop off in the near future. As important as the discovery of new manuscripts is for the study of scribal concordances and notational features, given that these discoveries are time-consuming and often require expensive excursions to study distant “leads,” it is fair for scholars, and those who fund scholars, to ask if we expect new manuscript finds to result in new pieces of music. More importantly (and less materialistically), if we suspected a single source or small group of sources to be representa-

¹ This article originally appeared as a section in “Trecento Fragments and Polyphony Beyond the Codex” (Ph.D. Dissertation, Harvard University, 2006), pp. 44–76. I wish to acknowledge Lisa Friedland (Department of Computer Science, University of Massachusetts, Amherst) for conversations and advice which resulted in many of the mathematical models used in this section, and David Tabak (National Economic Research Associates) for first noting the similarities to animal capture/recapture sampling methods. I owe a special thanks Prof. William Bossert of the Department of Biophysics, Harvard University for spending time in discussion with me about this project.

tive of a much larger collection of music we would be inclined to grant that source or group more weight in our analyses. A source that represented many missing sources would carry more force in preparing descriptions of typical music of a time, than sources that represented in themselves the full extent of the genre. The monophonic instrumental compositions in the **London** codex (29987) are examples of pieces to which we have given further weight and study because they are presumed to stand in for a much larger repertory.²

We should also consider the missing repertory because its size and composition affect how we view sources that do exist. As has already been mentioned, the majority of fragmentary manuscripts seem to have originally been similar in size to those few sources which do survive in complete or mostly complete state. Our losses are represented by the disembodied folio numbers which stand in for so many lost pages:

TABLE 1.11: HIGHEST EXTANT FOLIO NUMBER FOR SOME TRECENTO FRAGMENTARY MANUSCRIPTS³

Parma 75	243 (233?)
Perugia 15755	171 ⁴
Stresa 14	141
Florence 5	138 ? (see Chapter 3)
Frosinone 266/267	133 ⁵
Ciliberti	97
Todi Carità	93 ?
Brescia 5	71
Siena Ravi 3	70
Vatican 1969	60
Padua 1475	50
Munich 3223	22
Florence Conservatorio	19

² I should add that serious questions can be raised at least in this case about whether these pieces are similar to the unwritten instrumental pieces; this is taken up in more detail within my discussion of keyboard music in the fragments in the following chapters and also the discussion of the possible instrumental work “Sones ces Nachares” from **Pad A** in Chapter 2.

³ These sorts of loss are not confined to the main period of this study: among slightly later manuscripts, one should recall the **Boorman** fragment’s preserved foliation of 125, or the earlier **Venice Giorgio**’s folio 86.

⁴ Oliver Huck, review of *Frammenti Musicali Del Trecento nell’incunabolo Inv. 15755 N. F.*, edited by Biancamaria Brumana and Galliano Ciliberti (Florence: Olschki, 2004), forthcoming in *Plainsong and Medieval Music*. Brumana and Ciliberti did not notice this folio number on binding strip VIa, thus their highest identified folio number is 36.

⁵ A second, arabic foliation of 217 appears on the bifolio with signature 267, but it is unclear whether this foliation is original.

We should not forget that these numbers do not represent the original length of these manuscripts, but merely the highest numbered folio which currently survives. For instance, the gathering structure of **Pad A**, discussed in Chapter 2, shows that although our last folio number is 50 (on **Padua** 1475), we can be fairly certain that the original manuscript contained at least 70 folios. The order of works in **Florence 5** gives another hint at the original length of a manuscript. Its seemingly-alphabetical presentation of Francesco's ballate ends with ballate beginning with the letter "C" (*Che pen'è quest'al cor*, *Chogli ochi assai ne miro*, and *Cosa nulla*). Even supposing that Francesco were the last composer in the manuscript (unlikely) and that it preserved only half of his 113 known ballate which begin with the letters *D* through *V*,⁶ we would still need forty folios to complete the manuscript.⁷

As tempting as it might be to suppose that manuscripts were often dismembered from their extremes, we have little evidence for this mode of destruction. It would therefore be more prudent to suppose that these folios represent random samples of the original manuscripts. The expected length of the manuscripts, as an average, would then be twice the highest surviving folio number.⁸

But what was on these lost pages? We return to the problem of the missing pieces within these missing sources. There are several other lost pieces (or at least, lost concordances) which are

⁶ The transmission rate of fifty percent seems appropriate since, of Francesco's thirteen known ballate between *Benché ora* and *Cosa nulla*, **Florence 5** provides readings for seven.

⁷ In the case of **Florence 5**, however, we would have less reason than for other manuscripts to suppose that the lost pages represent otherwise unknown works, because of its high concordance rate.

⁸ For a manuscript with j folios, the expected folio value, that is, the likely average folio over repeated random discoveries, given by:

$$EV = \sum_{i=1}^j p_i i$$

where p_i is the probability of drawing folio i . If each page is equally likely to be preserved then the expected

value reduces to:

$$EV = \frac{1}{j} \sum_{i=1}^j i = \frac{1}{j} \cdot \frac{j(1+j)}{2} = \frac{1+j}{2} \approx \frac{j}{2}$$

(It is not always the case that one can reverse a formula like this one to get the estimated book length. In fact, the field of parameter estimation is controversial enough that it accounts for perhaps half of all theoretical statistical research. However as a general rule for the average length of a manuscript, the inversion of this formula would raise few eyebrows. It should not be considered an accurate way of estimating the length of any one particular manuscript given a surviving folio number).

The average of the entries on Table 1.11 is 100, so we might predict an average book length of 200. For another way of considering the expected length of a manuscript, we can compare with the lengths of the surviving Florentine codices, **Panciaticchi** 115, **London 29987** 185 (palimpsest numbering), **Pit.** 150, **Squarcialupi** 216, and **San Lorenzo 2211** 188 (highest surviving folio), which average 171 folios. These two estimates accord well, and strongly suggest that the fragments were originally similar in length to the larger, surviving Florentine codices.

tantalizingly close to being available to us. Four trecento flyleaves are still attached to their host manuscripts, leaving a face undiscovered, or visible only as show through. Librarians have good reason to be cautious about lifting flyleaves: in several cases, much of the ink is lifted from the page, and the cover (with a mirror) becomes the more important source for that face of the manuscript.

TABLE 1.12 : POLYPHONIC SOURCES STILL PASTED DOWN WITH AT LEAST ONE FACE HIDDEN.

Houghton 122	1v, Marian motet. 2r, Credo
Oxford 56	Back pastedown: unknown work, probably in <i>tempus imperfectum cum prolatione maiori</i> . ⁹
Padua 1027	Half of the front and back folios are attached to the cover. As the rest of the fragment is blank, and there is no show-through, the hidden sections are probably blank also.
Ivrea 105	No description

Works which are unidentifiable despite being revealed are another glimpse into the problems of lost sources. The following table, Table 1.13 lists only those works not included in the previous and does not begin to consider the problem of identification of certain works from **San Lorenzo 2211**:

⁹ For the identification of the front pastedown of **Oxford 56** as Ciconia's *Gloria: Suscipe, Trinitas*, see Chapter 2, below.

TABLE 1.13 : POLYPHONIC SOURCES WITH ILLEGIBLE FACES OR FRAGMENTS TOO SMALL TO IDENTIFY

Cividale 98	Ballade tenor (?) f. 1r bottom.
Cortona 2	Gloria, f. 1r., Sanctus, f. 1v B.
Grottaferrata/Dartmouth	Two offsets from missing folios.
Krakow 40582	One side of each of the two folios is an illegible Gloria.
London 29987	Erased Credo, f. 1r. ¹⁰
Oxford 16	Erased work.
Oxford 56	Several unidentified and mostly illegible works.
Perugia 15755	Several motets and music with no surviving texts.
Rome 1067	Small work on f. 42v.
Seville 25	Unidentified compositions, ff. 23v and 39r.
Vatican 171	Four unidentified Glorias.
Vatican 1790	Mensural voice at the bottom of f. 1r.
Vatican 1969	Three voice virelai, f. 49r.

This table should not be read as implying that all other sources have satisfactory readings.

Some hints as to the extent of our musical losses can be found in references to musical compositions in other works, such as poems in text sources where composers' names have been added, or texts which make obvious that they are discussing specific musical compositions. These pieces are in a sense then only semi-lost. Their music and their poetry are not available to us, but their one-time existence is documented. An example of a poem documenting lost musical works is Simone de' Prodenzani's thirty-fifth sonnet of *Il Saporetto*:¹¹

¹⁰ Michael Long, "Musical Tastes in Fourteenth-Century Italy: Notational Styles, Scholarly Traditions, and Historical Circumstances," (Ph.D. dissertation, Princeton University, 1981), pp. 172–73. The visible parts of the Credo, transcribed by Long on p. 176, are compatible with Zachara's Credo in **Cividale 98**. Further investigation is warranted. Another unidentified, erased early-fifteenth century Credo can be found on f. 1v of the probably Viennese manuscript **Nuremberg 9a**, f. 1v. The voice has been erased in favor of Zachara's Credo, "Cursor." (Mentioned in Fischer and Gallo, *PMFC 13*, p. 264.)

¹¹ Edition from Fabio Carboni, *Simone De' Prodenzani: Rime* (Manziana: Vecchiarelli, 2003), computer file 3, p. 15. In Carboni's new numbering of the sonnets, this sonnet is no. 24. I have added italics to the full title of Rosetta in line two and inverted the order of "partir da te mi" from "da te partir me" in line four. This reading accords with the versions of Boccaccio's text found in **Bologna Archivio Covers**, year 1337 and 1338. Although not present in the Bologna versions of this text, in *Filostrato*, the text continues asking, "Perché mi toglì il sollazzo e la pace?" Perhaps Prodenzani selected this poem because of the potential for a pun on the name of the central character of *Il Saporetto* or the title of his other major work. The version of the poem given above can be compared with Santorre Debenedetti, editor, *Il "Sollazzo" e il "Saporetto," con altre rime di Simone Prudenzenani d'Orvieto*, supplement to *Giornale Storico della Letteratura Italiana* 15 (Torino: Loescher, 1913), which includes as songs, "Cançon di maggio" (1), "dolçe sapore" (3), and considers as a title, "El dolce raggio" rather than the shorter "Raio."

Titles definitely to be associated with works which survive today are shown in bold type.

Colla vivola fe' cançon di maio,
Rosetta che non cambi mai colore,
Le sui nafres tam fort, dolce sapore,
Comme partir da te mi degio oma'io?
D'amor languire e puoi el dolce *Raio,*
O rosa bella, che m'alegrie 'l core,
Legiadra donna e poi *Donna d'amore,*
Un fior gientile del qual mi 'namoraio,
Questa mirabil donna, Margarita,
Con lagrime bagnando el suo bel viso,
Ditutto se' e fè Sella mia vita,
Costei sarebbe bella in Paradiso,
Non credo, donna, O giemme incolorita
 del Cicognia una parte fu l'aviso.

Of the works or possible works cited, we have copies of the nine in bold in Example 1.14. All of these works are by Antonio Zachara da Teramo except *O rosa bella*, and the three works with “donna” in their incipits. John Nádas has equated “El dolce Raio” with Ciconia’s *Le Ray au Soleyl* and has tentatively connected *Questa mirabil donna, Margarita* with the refrain of the ballade *N’a pas longtemps* which discusses the pleasing and beautiful Margarite.¹² We are still left with at least two lost works (*Come partir da te me debbo mai* and *Se la mia vita*) and possibly five if we consider “Costei sarebbe bella in Paradiso,” “O gemma incolorata,” and “Cançon di maggio” the titles of lost works. Depending on what mix of these interpretations we use, we have between 56 percent (9 of 16) and 85 percent (11 of 13) of the works mentioned in this poem. Are these typical numbers? Can we generalize from this evidence?

We have other evidence of lost sources which we can use. The poet Franco Sacchetti provided several editions of his works. In later editions, he was careful to note which of his poems had been set to music and by whom. Figure 1.15 lists the works which Sacchetti reports were set by the composer Nicolò:¹³

¹² John Nádas, “A cautious reading,” p. 35. The quotation in *N’a pas longtemps* is “La très plaisant et belle Margarite.” See David Fallows, “Ciconia’s last songs and their milieu,” in *Johannes Ciconia: musicien de la transition*, edited by Philippe Vendrix (Turnhout: Brepols, 2003), p. 114, for a summary of the arguments which allow *Le Ray au Soleyl* to shed the designation “opus dubium.”

¹³ Adapted from F. Alberto Gallo, *Music of the Middle Ages II*, (Cambridge: Cambridge University Press, 1985), pp. 65-66.

FIGURE 1.15: NICOLÒ'S WORKS MENTIONED IN THE CATALOG OF SACCHETTI

M = madrigal, B = Ballata, C = Caccia; works which survive today are shown in bold type

- Come selvaggia fera fra le fronde*** (M)
Come la gru quando per l'aere vola (M)
Correndo giù del monte a le chiar'onde (M)
Di diavol vecchia femmina ha natura (B)
Nel mezzo già del mar la navicella (M)
Passando con pensiero per un boschetto (C)
Una augelletta, Amor, di penna nera (M)
Chi 'l ben sofrir non pò (B)
Povero pelegrin salito al monte (M)
Lasso, s'io fu' già preso (B)
State su, donne!—Che debian noi fare (C)
Chi vide più bel nero (B)

Seven of Nicolò's twelve works on texts by Sacchetti currently survive (58%); of the thirty-four of Sacchetti's texts that were set to music by any composer, only twelve remain (35%).¹⁴ Do these percentages apply to Italian music as a whole, or are the pieces set to Sacchetti's texts, mostly by the oldest generation of composers of the Italian *Ars Nova*, different and unrepresentative?¹⁵

For the remainder of this section, I wish to introduce another possible method for examining repertoires which do not survive. This method uses probabilistic models and simulations in part borrowed from animal biology. These models are most commonly employed to count animal populations whose members are difficult to capture *in toto*. Although there is a fair amount of probability and other mathematics used to get the final numbers presented in this project, the fundamental points can be followed with little background in probability and statistics.

¹⁴ One lost Sacchetti ballata, Francesco's *Né te né altra voglio amore*, possessed at least four different lauda contrafacts, though all are transmitted in the same source, **Chigi 266**. The ascription to "Franciscus de Organis" is from Sacchetti's autograph, **Florence 574**.

¹⁵ There are further documents which might allow us to estimate our losses in a similar fashion for nearby repertoires. Perhaps the most famous to scholars working on French music of the period is the index page formerly in the possession of the Duchess of Trémoille of a lost manuscript of motets. Work on this source was carried out by Martin Staehelin in a short but important article on lost manuscripts of the fourteenth and fifteenth centuries, "Mehrstimmige Repertoires im 14. und 15. Jahrhundert: Das Problem der verlorenen Quellen," in *Atti del XIV congresso della società internazionale di musicologia, Bologna, 27 agosto—1 settembre 1987*, Vol. 1 (Round Tables) (Turin: E.D.T., 1990), pp. 153–59. Through concordances with other French and Italian manuscripts, Staehelin ascertained that 63% of the 114 pieces in the index survive. (Staehelin did not seem to include the concordance in the recent manuscript **Cortona 1**, though this changes the percentage only slightly). His work was concentrated on source losses rather than work losses and, as such, focused on library catalogs, payment records, and assumed omissions in stemmata as his most important evidence.

The first principle to borrow is that the number of unique works in each manuscript source gives us some indication of the size of a repertory. If with every new fragment or book we discover, the majority of works are unknown from other sources, then, all else being equal, we would expect that a large part of the repertory remains undiscovered. Conversely, if new manuscript discoveries were, in general, not bringing with them new works, then we would suspect we have most of the original repertory (if not most of the copies of the original repertory, which is an important distinction). While it might be obvious that more unique works would hint at a larger repertory, this presumption does not tell us *how much* larger one repertory might have originally been than another.

It is even more intuitive, but extremely important to keep at the forefront of our minds, that this principle tells us nothing about whether we have most or all of the contents of *other* repertories. If we have few new pieces of fourteenth-century Italian music accompanying new manuscript discoveries, it does not tell us anything about how much French music there is left to discover. This obvious statement makes the decision of what constitutes a repertory and what does not an important decision. Slicing repertories too thinly can create a problem of overfitting—seeing correlations where there is not enough data to support them, a problem I will return to later.¹⁶

We may begin with a simple example including some assumptions that might sound incorrect. We can modify these assumptions later and see how altering them affects the result. Consider how scientists might count the number of fish in a lake—they could catch 100 fish, tag them with some sort of marker, release them; then they could catch another 100 fish. If 20% of those fish were previously tagged, then we could guess that we had originally tagged 20% of the total fish in the lake. We would then estimate that there were 500 fish in the lake. This is known as a “capture/recapture” method of counting.¹⁷

¹⁶ It follows that even a small amount of data collected on a certain repertory is more important for estimating the size of that repertory than an abundance of data gathered about a different repertory. The weight of this axiom to my work cannot be overstated: there are many large Florentine codices of mostly-Florentine works, nearly exclusively secular, which have a great many pieces in common. As I will show later, the fragments on which I work preserve parts of that repertory but primarily comprise different repertories, mainly sacred and ceremonial, with a much lower rate of retransmission. The fragments therefore preserve the types of music which we should expect future manuscript discoveries to have a higher chance of containing.

¹⁷ The generalized formula for a capture-recapture model with two captures is:

$$\text{Size of population} = \frac{\text{Total number of items tagged in first capture} * 100}{\text{percentage of tagged items in the second capture}}$$

One might note that the size of the second capture does not come into the equation. However, larger captures will usually result in more accurate estimates.

We can use the same method of counting with musical works in manuscripts—we take a certain number of manuscripts as the first catch and we mentally “tag” the pieces in that batch by taking note of which pieces appear. We might then consult other manuscripts and see the amount of overlap among manuscripts. What might seem like a flaw in this method is that we assume each song was equally likely to be transmitted—as if each fish were equally easily caught. Surprisingly, there are several cases where this assumption does not strongly conflict with our data, as will be presented. More importantly, when we adjust for different pieces having different popularities, we find that our unadjusted prediction *underestimates* the number of pieces. So a model assuming equal probability gives us a minimum estimate of the number of missing pieces, which is still extremely useful. It happens that most other refinements to the model—non- or only partially intersecting repertoires is one—affect the model in the same way, increasing the range of possible values but leaving the estimated minimum number alone.¹⁸ It bears repeating, that though the estimates given in this paper can be refined, and will be later, the refined estimates will not be lower than what I present here.

In order to make accurate estimates we must first have a good grasp of the number of pieces which survive in each genre. This study will limit itself to the period in which nearly all our manu-

¹⁸ A list of potential refinements to a capture-recapture model and their effects on the estimated size of the population can be found on p. 57 of Michael Begon’s short introduction, *Investigating Animal Abundance: Capture-Recapture for Biologists* (Baltimore: University Park Press, 1979). Begon gives three situations where this number might be overestimated, none of which is likely to occur in this study. First that the mark on the animal might not be permanent; for our purposes this impermanence means that we might not recognize a piece when it appears in a second manuscript. Second, that marking decreases survival rates, or here that the presence of a piece in one manuscript lessens its likelihood of appearing in a second manuscript. The first case, lack of recognition of a piece, is only possible in the case of poorly researched concordances and tiny fragments which may be different parts of the same piece—these form a near negligible percentage of the total corpus. The second case, that a source would avoid containing the same piece as another source, may be true for fragments which were originally part of the same manuscript but are not today identified as such. In this case, the two fragments would be less likely to have works in common. However, some of the most similar manuscript fragments, for example **Pad A** and **Pad D**, *do* have repertory in common and this sharing has been an important reason for not uniting the fragments. In other cases where scholars might disagree about whether two or more fragments are from one source, in this study I have considered them the same source in order to avoid the possibility of overestimating. **Trent 1563** and **Krakow 40582** are exceptions to this rule, since the different numbers of lines per staff makes it unlikely that they stem from the same source (see Chapter 2). The final possible source of overestimating comes from open populations, where individuals can enter and leave the sample space. One might suppose that the changing repertory over time would be equivalent to this situation, but it is instead equivalent to death and birth within a population which is already accounted for. Since our sample space, that is, our repertory, is the whole of Italian mensural polyphony from the late fourteenth century to the early fifteenth century, it is impossible for such a piece to enter or leave this realm from some other.

scripts stem, 1380–1415, (the only major exclusion is the **Rossi** codex), and will thus consider only those earlier pieces which are retransmitted in a retrospective manuscript.

Table 1.16 gives the number of works in each of five different genres contained in different Italian and foreign manuscripts of the trecento and early quattrocento. The number of pieces in the genre contained in each manuscript is given, as are the number of pieces appearing in one, two, three, etc., manuscripts, and the percentage of *unica*. About half of the madrigals and cacce exist in only a single source. This number increases to about two-thirds for the ballate and three-quarters for the Latin-texted works. As I have mentioned above, this alone hints at a relatively larger lost repertory of sacred music than secular.¹⁹

TABLE 1.16: SURVIVING NUMBERS OF TRECENTO PIECES LISTED BY MANUSCRIPT AND ORGANIZED BY GENRE

[*in the original version, a five-page table appears here. The most important results of which are summarized below*]

Though tangential to this part of the study, a surprising revelation of Table 1.16 is that there are nearly as many sacred and ceremonial works of the trecento as there are madrigals.²⁰ I will return to this observation later when we consider the number of missing madrigals and missing Latin works.

Taking the pieces that exist today as our given, I began with an equal popularity model. I looked at the amount of overlap between manuscripts to estimate the number of works which do not appear in any manuscript. In order to give the details of my method while avoiding obscuring the results for readers uninterested in the more technical aspects, the probability basics necessary to obtain these estimates are given as an appendix to this chapter. Interested readers are invited to follow that argument before continuing.

¹⁹ Instrumental diminutions have been omitted from the present study.

²⁰ This discovery is a side result of the revision I am preparing of Kurt von Fischer's landmark 1956 catalog but will force a major revision of our view of the century as a whole, of which this dissertation is a start. In Table 1.16, pieces which appear twice within the same source are counted once.

TABLE 1.17: ESTIMATES FOR THE NUMBER OF ITALIAN WORKS IN MSS CA.1380–1415

	(a)	(b)	(c)
Cacce	25	28	
Madrigals	167	177	165
Ballate	409	507	384
Liturgical Pieces	116	196	115
(only in trecento MSS	98	168)	
Non-Liturgical Latin Works	47	105	
(only in trecento MSS	38	93)	

(a) total surviving today

(b) estimated lower bound for the number of pieces given a random distribution model

(c) estimated total for today from cross validating the model by removing the fragments and San Lorenzo (for madrigals and ballate) or the five MS with the most liturgical works (Boverio, Mod A, Pad A, Pad D, Grottaferata/Dartmouth). See below on cross validation. No holdout cross validations were performed for cacce or motets since there are fewer of them.²¹

I want to point out some results which can be seen simply from Table 1.16 and column (b) in Table 1.17. Comparing the estimates for madrigals to that for sacred and ceremonial works, the much lower concordance rate for the Latin works gives us reason to believe that more Latin pieces were composed in the trecento than madrigals, that most quintessentially Italian of all genres. (And this estimate still excludes the international repertory which, for the most part, mixed freely with the native Italian sacred music).

An important quality in a model is its ability to be tested and stand up to such testing. One way to test the model is called cross validation. This means running the model with incomplete information and then using the model to predict our current situation, to which we can compare. For instance, I removed the fragmentary sources and San Lorenzo entirely from the data used to make the model and then used the model to predict how many additional pieces would be found if added the number of folios in those fragmentary sources we now have. Without the fragments or San Lorenzo, for instance, we would have 159 madrigals in 314 copies. On the basis of this information, the model then predicted that there were originally 175 madrigals, and further that if we had 65 more copies of madrigals, six of them would be new. So the cross validated model predicted that with the sources we have today we should have 165 madrigals given our source situation. As you can see, we have 167—a close estimate. Running the same model for ballate, we have an estimate of 385 ballate instead of the 409 we do have—not as close but still a good estimate, while the model for li-

²¹ It is important that the works chosen to be removed for holdout cross validation are chosen arbitrarily and that if repeated cross validations are performed with different works the researcher does not choose the one which gives the desired result. Here, I chose to remove the manuscripts which were easiest to delete and recalculate from my spreadsheet version of the Kurt von Fischer catalog—the small manuscripts for the secular tables and the large manuscripts for the liturgical works.

turgical music is off only by one from our observed number, 115 instead of 116, which is amazingly close. Taken as a whole, these tests suggest that the role of popularity in the transmission of music to us today is a supporting one to that played by random chance. (More information about cross validation appears as an appendix to this chapter).

The other standard way to test a model is, unfortunately, more difficult for us to perform: find new sources and see how they accord with the model. We can not just find new trecento sources whenever we want.²² However, since I began this project, four new sources have been discovered. One fragment, **Brescia 5**, I was able to incorporate into this study; two others, **Siena Ravi 3**, **Bologna Archivio Covers**, and **Perugia 15755** came to my attention too late. However, we can see how they conform to the model's predictions. The fragment in Brescia contains two ballate; as was to be expected, both of them were already known. The **Bologna Archivio Covers** source contains a single ballata—already known. The fragment in Siena contains five Latin-texted works; two are known and three unknown. The Perugia fragments contain Mass movements, motets, and madrigals. All of the madrigals are previously known while all the Mass movements and most of the motets are new discoveries. Of course, not every newly discovered source will conform so well to this model, but we should not be surprised when they do: new sources bringing old secular and new sacred music.

Popularity and Transmission

The study and analysis of medieval music has always been, and will always be, a selective art. Some works and some composers are more studied than others, and this selection informs (at best) or skews (at worst) our view of the period being studied. Given the limited time and resources with which we work, we may wish to focus our efforts on those pieces which were most well-known or most popular in the period in which we study. Unfortunately, as we know too well, determining which pieces were popular at the time they were written is a difficult task, sometimes seen as impossible.

We often think that a work in many sources must by definition have been popular. (Or at least, when we take into account the vast unwritten tradition, we can at least say it was popular among those who copied and read music). We use similar metrics to determine the popularity of pieces today, such as number of performances or record sales. But we should become concerned about the usefulness of such measures when there are extremely few sources. For instance, David Fallows reminds us in a recent paper that although 10 songs by Du Fay are preserved in the 11 sources copied after his death (that is, one song in each manuscript with one piece duplicated), we

²² The inability to create more data samples as needed has been explored in the works of the statistician John Tukey who coined the term “uncomfortable science” for such situations.

are most likely seeing random survival of sources and it would be “dangerous for statistical purposes” to consider these pieces popular.²³

In the previous section, the models were used with the important supposition that each piece was equally likely to be selected (random). We saw before that if pieces are not equally likely to be chosen then our models provide a minimum estimate for the number of lost pieces. We also saw via the cross-validation method that the random model only differs slightly (less than 7%) from the non-random, actual world. But the deviation from the random model, however small, should be investigated.

We can create further models which allow us for the first time to pin down a few pieces of trecento music as being definitely popular for scribes to copy. We often think that a work in many sources must by definition have been popular, or when we take into account the vast unwritten tradition at least popular to copy. But just as a random series of coin flips will occasionally have a long string of heads without having any meaning behind it, so too can a piece of music appear in many different manuscripts purely by the vagaries of preservation.

What we might like to know is how likely it is that a piece which is copied in, say six sources, appears so often out of chance rather than because it was specifically popular. For example, *Tosto che l'alba* and *Usellet(t)o selvag(g)io* are cacce found in five and six sources respectively. No other cacce are found in more than four sources. We might therefore conclude that these were popular cacce. Yet if all cacce were once copied equally, given the surviving manuscript situation, we might still expect to see a cacce appearing in five sources. Thus we cannot say without other testimonies that *Tosto che l'alba* was a popular song for its time, only that it is *fairly likely* that it was popular. However, there is only a 2% chance that any caccia at all would appear randomly in six sources, so it is more likely that *Usellet(t)o selvag(g)io* was popular. Further, we cannot say anything definitive about the popularity of the two cacce which appear in four sources, *Cosi pensoso* and *Nell'acqua chiara*, since a random distribution of surviving sources would predict a couple of pieces appearing in four manuscripts. We simply have too few caccia sources. To put it another way, the number of sources in which a work appears is significant only in relation to the total number of sources available in which it could have appeared.²⁴

²³ Fallows, “Ciconia’s Influence,” paper presented at the Jena Conference, *Kontinuität und Transformation der italienischen Vokalmusik zwischen Due- und Quattrocento*, July 1–3, 2005.

²⁴ That it is difficult to say for sure which pieces were definitely popular does not excuse the injustices done by the lack of performances of many works which survive in four, five, or more sources. David Fallows in 1975 drew attention to a neglect of Bartolino da Padova on disc (since somewhat ameliorated). He admonished that if we use the number of surviving sources as “any yardstick of respect in the 14th century, Bartolino is especially important, for three of the ten most widely distributed trecento pieces are by him.” (“Performing Early Music on Record—1: A Retrospective and Prospective Survey of the Music of the Italian Trecento,” *Early Music* 3.3 (July 1975), pp. 252–53 with evidence in a note on p. 260.) One may

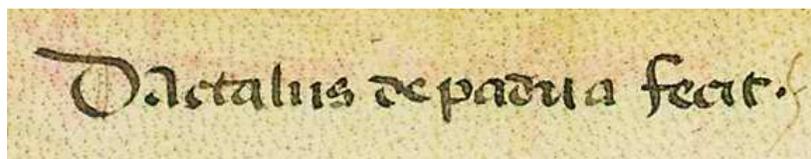
Let us take the liturgical Latin works as a second example. Table 1.18 gives for the sacred Latin works the actual number of pieces copied in six, five four, etc. manuscripts, and gives a comparison to the number predicted if all pieces were equally popular.

TABLE 1.18: COUNT OF LITURGICAL PIECES COMPARED TO THE PREDICTED NUMBER

# of MSS in	Actual # of pieces	Predicted # of pieces	(Titles of actual pieces; Z = Zachara)
Seven	2	.00	(Z. Credos <i>PMFC 13: 21 & 23</i>)
Six	1	.03	(Z. <i>Gloria: Laus, Honor;</i>)
Five	3	.24	(Z. <i>Gloria "Micinella"; Ciconia, Gloria: Suscipe Trinitas;</i> Egardus, <i>Gloria PMFC 12: 7</i>)
Four	1	1.6	
Three	5	8	
Two	15	30	
One	85	71	

The predicted number of pieces differs from the number of pieces we actually possess in two significant respects. First, there are slightly more *unica* relative to the number of pieces with concordances than we would suppose if all pieces were equally popular. This higher percentage is to be expected in cases where some pieces are more popular than others, since (if we hold the total number of copies of pieces constant) each concordance of a popular piece is one fewer concordance of a less popular work. Reducing the number of concordances of less popular works also pushes more works into the “zero-copy” range, that is, the lost works. Thus we can see that our estimate of the total number of lost works should be slightly higher than the model worked out on a supposition of equal popularity.

have to amend Fallows’s statement based on an argument he reports twenty-eight years later that one of these three works, “Imperial sedendo” is not by Bartolino. The argument by his student, Leah Stuttard, is that there is a conflicting attribution between **Squarcialupi** and **Mod A**—where it is attributed to the otherwise unknown Dactalus de Padua—and its style does not accord with Bartolino’s (Fallows, “Ciconia’s last songs,” p. 120). As Fallows points out, it is nearly impossible that Dactalus is a miscopying of Bartolinus. Indeed, the added suffix, “fecit” (to my knowledge never again used in this manuscript), could be read as a reaffirmation of authorship, “Yes, Dactalus, and not someone else, composed this,” (**Mod A**, f. 30r):



It also seems more likely that a work by an unknown composer would be misattributed to a well-known, than vice-versa.

The second significant difference is that we have more pieces with many copies (five or six for the liturgical works) than would be predicted. Only two out of every hundred simulations predicted that there should be even a single piece with six sources, instead we have three such pieces. These pieces that greatly exceed an equal probability model can be identified as the most likely popular pieces (at least for scribes to copy) among works of the trecento and early quattrocento.

We can run the same analysis for the other genres of trecento music. Table 1.19 lists the five works which we can say were possibly or probably popular at their time and the ten pieces which were popular almost without doubt.

TABLE 1.19: [POSSIBLY-]POPULAR WORKS

Liturgical: Undeniably Popular, Seven sources: 1 in 400 probability (0.28%)

(i.e., that the number of copies of any of these is due to chance)

Credo, *PMFC 13.21* (Zachara) Bologna Q 15, Boverio, Grottaferrata/Dartmouth, Pad D, Mod A,
Valladolid 7, Warsaw 378,

Credo, *PMFC 13.23* (Zachara) Boverio, Cividale 98, Grottaferrata/Dartmouth, Kras., Siena 207,
Trent 1563, Warsaw 378

Liturgical: Popular, Six sources: 3% probability that the perceived popularity is only due to chance

Gloria: Laus, Honor (Zachara) Bologna Q 15, Munich Emmeram, Old Hall, Pad D, Siena 207, Warsaw 378

Liturgical: Possibly popular, Five sources: 22% probability

Gloria "Micinella" (Zachara) Atri 17, Bologna Q 1, Bologna Q 15, Bologna 2216,
Grottaferrata/Dartmouth

Gloria: Suscipe, Trinitas (Ciconia) Grottaferrata s.s., Grottaferrata/Dartmouth, Oxford 56, Pad D, Warsaw 378

Gloria, PMFC 12.7 (Egardus) Grottaferrata/Dartmouth, Mod A, Pad D, Udine 22,²⁵ Kras.

Caccia: Popular, Six sources: 2% probability

Usellet(t)o selvaggio (Jacopo da Bologna)

Caccia: Possibly popular, Five sources: 23% probability

Tosto che alba (Gherardello)

Madrigal: Popular, Eight sources: 0.4% probability

La douce çere (Bartolino da Padova)

Madrigals: Probably popular, Seven sources: 6% probability

La bella stella (Giovanni da Cascia)

O dolce appres'un bel pelaro (Jacopo)

O cieco mondo (Jacopo)

Ballata: Undeniably popular, Eight sources: 1 in 500 probability (0.2%)

Donna s'ì'ò fallito (Francesco da Firenze)

²⁵ See Chapter 2 for more information on the Udine 22 version of this Gloria.

Ballate: Popular, Seven sources: 3% probability*Con langreme bagnandome* (Johannes Ciconia)*Gentil aspetto* (Francesco)*Non avrà mai pietà* (Francesco)*S'i'ti so(n) stato* (Francesco)

Francesco's *Donna s'i't'ò fallito* stands out on Table 1.19 for appearing in so many sources (eight, not counting a lauda contrafact and a citation by Prodenzani) that it is nearly impossible to believe that it was not a popular work for scribes to copy *ca.* 1400.²⁶ That all three of the popular ballate are by Francesco should not be too surprising—his popularity has never been seriously challenged in the literature.²⁷ Between the madrigals and the cacce a wider variety of composers are represented, but still without any surprises. Intriguingly, the only sacred works which we can definitely say were popular are all compositions by Antonio Zachara da Teramo, a composer whose popularity in our own time continues to increase as we become more and more fascinated with his bizarrely compelling output.²⁸

It is natural to want to ask why these works, particularly Zachara's, were popular; in doing so we move into a more speculative realm. It may be Zachara had a connection, particularly at the beginning of his works, with more simple polyphony, which had a wide distribution throughout Italy. Evidence of the influence of homophonic mensural polyphony is found in Zachara's *Gloria*, "*Micinella*" (mentioned as possibly popular, above), as well as in an unattributed *Gloria* found in **Warsaw 378** in a similar style to Zachara's. The openings are similar to the mostly-homophonic mensural Mass movements (or rhythmicized *cantus planus binatim*) which flourished during the late

²⁶ There are two released recordings of the work, Thomas Binkley and the Studio der frühen Musik's 1972 edition and Mary Springfels's Newberry Consort recording of 1990. Tiziana Morsanuto, "Discografia di Francesco Landini," in Defino and Rosa-Barezzani, *Col dolce suon* (q.v.), pp. 564 and 581.

²⁷ However, the reasons which have justified Francesco's popularity can be scrutinized. Leo Schrade begins his edition of Francesco's works by describing him as "long recognized as Italy's greatest composer of the fourteenth century." Schrade continues by saying, "Perhaps as a result of such a recognition, the music of Landini has been more comprehensively preserved than the music of any other Italian musician." (Schrade, *PMFC 4*, p. i). It is dangerous to suppose that those music manuscripts which were preserved survive because of the greatness of music collected and not because of the vagaries of time. Our notions of presumed quality and importance in the fourteenth century are already shaped so strongly by what happens to survive. To crown the surviving manuscripts by hinting that they are the products of quasi-Darwinian natural selection over the centuries elevates this bias even further.

²⁸ Unfortunately, only one of these popular Zachara works (*Credo 23*) has been recorded, and that on a nigh impossible-to-find mono-CD released by Quadrivium in 1992 (SCA 027). Fortunately, the Ensemble Micrologus has made Zachara's sacred works part of their repertory, so one might hope for more recordings in the future.

trecento and early quattrocento. An example of which is seen in a Credo (“Regis”) setting from Vatican 657:

FIGURE 1.22: DETAIL OF CREDO FROM VATICAN 657, FF. 419V–20R.²⁹



This movement is perfectly homophonic for the first two lines of music and nearly perfect following. The phrases have a tendency to use longer note values at the beginning and ends, and semibreves and minims in the middle and before cadences.³⁰ Some pieces of homophonic polyphony, such as the first Credo of *Parma 9* (ff. A–D. Cardinalis) even accelerate from their opening longs, through breves, to semibreves, and finally minims before allowing the notes to occur in other orders.

The *Gloria “Micinella”* of Zachara also begins homophonically and may recall the same tradition. The opening is in two voices, almost a trademark of Zachara’s *Glorias*. Two places which are not homophonic set $\blacklozenge\blacklozenge$ ($\bullet\bullet$) in the top voice against $\blacklozenge\blacklozenge$ ($\bullet\bullet$) in the lower voice. This substitution is common in homophonic mensural polyphony (see the Nachtrag to *Wolkenstein A*, on f. 18r for one example):

²⁹ I have touched-up part of this facsimile to remove some show through. (Throughout this dissertation, all altered images are noted.)

³⁰ This connects slightly to the trecento style of having long melismas on the penultimate syllable of a phrase, but unlike the secular styles, such as ballate or especially madrigals, the shorter note values in the sacred works begin several syllables before the cadence.

EXAMPLE 1.23: ZACHARA, GLORIA “MICINELLA” FROM PMFC 13. OPENING

Antonius 'Zacharias' de Teramo
1-Bc 15 (1-Bu 2216)

Glo - ri - a in ex - cel - sis De - o.

Et in Et in ter - ra
(Not in 1-Bu)
Et in Et in ter - ra

5 10 #

pax ho - mi - ni - bus bo - nae vo - lun - ta - tis.

pax ho - mi - ni - bus bo - nae vo - lun - ta - tis.

15

C1 Lau - da - mus te.

C2 Lau - da - mus te.

Contratenor
Laudamus te. ①

T
Tenor
Laudamus te. ①

It might be noted that the only pieces where rhythmicized binatim is recalled before moving to more complex polyphony are Glorias. No known Credo begins like this. This might be an indication that the two repertoires existed alongside each other and (moving further into speculative territory) that Zachara, known for his musical trickery, might have wished to deceive his listeners as long as possible about what type of piece they are about to hear. Since the Gloria was the first Mass movement which seems to have been set polyphonically with frequency in fourteenth and early fifteenth-century Italy—polyphonic Kyries were still rare at this time—it would be the most likely candidate for such deceptively-homophonic treatment. In one of the definitely popular Credos (*PMFC 13.21*), Zachara does prolong the plainchant beyond its traditional ending at “Credo in unum deum” by setting also “Patrem omnipotentem” to the monophonic (but rhythmic) formula of Credo I. This formula was gaining in popularity at the end of the century—it is also used for Zachara’s Credo “du village,” the first of many settings by later composers—and Zachara could have been counting on the listeners to recognize this (and perhaps recognize an old warhorse) before jolting them with something original and at a much increased rhythmic pace. Significantly, the otherwise rhythmically active version in **Mod A** includes no decorations in the opening, as if they are being held in reserve for after the suspense has been lifted:

EXAMPLE 1.24: ZACHARA Credo (PMFC 13.21), INCIPIT

Cre - do in u - num De - um,

1-Bc 15, 1-MOe 5.24 *Antonius 'Zacharias' de Teramo*

C MOe Patrem Pa - trem o - mni - po - ten - tem,

C Bc Patrem Pa - trem o - mni - po - ten - tem,

C MOe fa - cto - rem cae - li et ter - rae, vi - si -

C Bc fa - cto - rem cae - li et ter - rae, vi - si -

Cc Bc (MOe) Factorem fa - cto - rem cae - li et ter - rae, vi - si -

T Bc (MOe) Factorem fa - cto - rem cae - li et ter - rae, vi - si -

FIGURE 1.25: ZACHARA Credo FROM MOD A, FF. 23v–24r (DETAILS)

Zacharias.

Actores ce li et terre visi bili uisibilium omnium et invisibilium. Et in unum dominum iesum christum filium dei unigenitum et ex patre natum ante omnia secula deo deo lumine de lumine deo uero de deo uero

Actores ce li et terre visi bili uisibilium omnium et invisibilium. Et in unum dominum iesum christum filium dei unigenitum et ex patre natum ante omnia secula deo deo lumine de lumine deo uero de deo uero

Actores ce li et terre visi bili uisibilium omnium et invisibilium. Et in unum dominum iesum christum filium dei unigenitum et ex patre natum ante omnia secula deo deo lumine de lumine deo uero de deo uero

I want to end by stressing both the need for and the promise of greater refinements to this system. Our models currently do not exploit the many different sizes of manuscripts at our disposal,

nor do they take a particularly nuanced approach to deviations from the predicted random distributions. Such refinements afford us an opportunity to give back to the world of statistical analysis since there are few statistical models dealing with multiple captures where the captures happen with no particular order. The models presented here could also be employed in any number of other areas of research in the humanities. In musicology, estimates of the total number of chants sung in a region or the number of folk songs recalled by a group of people could be useful in many studies. We could figure out the probability that a Renaissance motet which is unattributed in many sources was unattributed as a result of chance in order to investigate theories of scribal confusion about the composer. Beyond musicology, the methods could be refined and reapplied to answer questions about the number of sonnets of a given poet or the total vocabulary of a particular author.³¹ Numismatists might be interested in how many different types of coin were in circulation in a given region at a certain time, or what certainties we might have about estimating the proportions of coins minted; codicologists could have a better understanding of the economics of papermaking if they possessed more accurate estimates of the total number of watermark types originally produced in a particular region and time. The use of population estimates in musicology and in the humanities is in its earliest infancy; as such, the number of uses for these models can only be guessed.

Appendix 1: Statistical Applications:

a. Probability review

Probability review

A probability is defined as a number between 0 and 1 (inclusive), and represents the likelihood of an event happening. For example, if we roll a fair six-sided die, the chance that we get a five is $1/6$. That is, there are six possible outcomes, of which one gives us the desired outcome. We can write that a is the event “roll 5” and $\Pr(a) = 1/6$.

The probability of something *not* happening is defined as one minus the probability of the event happening. So $\Pr(\text{roll something other than } 5) = \Pr(a \text{ does not occur}) = 1 - 1/6 = 5/6$.

If x and y are independent events, like dice rolls or people working on unrelated manuscripts, then the probability of x and y happening is $\Pr(x \text{ and } y) = \Pr(x) * \Pr(y)$.

In addition to knowing how likely it is that something will occur (probability) we also often want to know how many times an event will occur if we keep performing or observing a certain ac-

³¹ This last problem was approached by two scholars of statistics, Bradley Efron and Ronald Thisted, in their article, “Estimating the number of unseen species: How many words did Shakespeare know?” *Biometrika* 63 (1976), pp. 435–447. Some of the more difficult math in their article which might hamper their methodology’s usefulness for humanists can now be alleviated by using the speed of personal computers to solve exactly equations which previously needed to be estimated.

tion. For instance, if we go back to the example of dice, you may want to know how many times you would *expect* to roll a five if you rolled a die ten times. We call this rational expectation the *expected value* (EV).

Fortunately, for independent events, such as dice rolls, where what you rolled previously does not affect what you are likely to roll next, all you need to do to calculate expected value is multiply the probability of your outcome by how many times you do it.

So on average the expected number of fives if you roll a die ten times is:

$$\text{EV} = \text{number of rolls} * \text{Pr}(\text{roll a five}) = 10 * 1/6 = 10/6 \text{ or } 1.67$$

Of course, it is impossible to roll 1.67 fives. What it means is that, on average, one or two of the ten rolls would be a five.

b. Estimating the number of lost pieces in trecento manuscripts

Let $X = \{x_1, x_2, \dots, x_n\}$ be the set of pieces which we assume may have once existed. We want to estimate n , the original number of pieces. Any given piece x in X might be a work which exists today or one which is no longer extant. All pieces of both types are included in the set X .

Let $M = \{m_1, m_2, \dots, m_y\}$ be the set of manuscripts now available, where we define k_i to be the number of pieces in manuscript m_i . Unlike the set of pieces, this set M only comprises manuscripts or fragments we have now. The total number of surviving manuscripts is y .

The compiler of manuscript m_1 chooses k_1 different pieces to place in it. There are any number of reasons why the person writing the manuscript might choose a given piece to be in the manuscript—the audience of the manuscript, the pieces known to the scribe, forms to be represented, etc.—but among the pieces in a single sub-genre, it can be difficult for us to tell why certain pieces are chosen or not.

We will begin with a model that supposes that within each sub-genre the pieces chosen are as good as random to us; certainly we will check to see how good an assumption this is later. Given this supposition, the probability than any piece (call it x) appears in this manuscript m_1 depends just on the number of pieces in the manuscript and the total number of pieces in the sub-genre. In fact, it is equal to the proportion of all the pieces available which are in the manuscript. Thus, if we use the designation k_1 to represent the number of pieces in manuscript m_1 , then the proportion of all pieces in manuscript m_1 is the ratio of the number of pieces in m_1 , that is, k_1 , to n , written mathematically as:

$$\text{Pr}(\text{item } x, \text{ appears in } m_1) = \frac{k_1}{n}.$$

The probability that x does not appear in m_1 is:

$$\text{Pr}(x_1 \text{ does not appear in } m_1) = 1 - \text{Pr}(\text{item } x, \text{ appears in } m_1) = 1 - \frac{k_1}{n}$$

Or for a different manuscript, m_2 : $\Pr(x \text{ does not appear in } m_2) = 1 - \frac{k_2}{n}$. And so on for any manuscript.

For two manuscripts which are compiled independently of each other (excluding for example the Machaut manuscripts, but not the principal trecento manuscripts), we can multiply probabilities to get the probability that a piece does not appear in either manuscript. For instance the probability that x_1 does not appear in m_1 and also x_1 does not appear in m_2 is the product of the two terms:

$$\Pr(x_1 \text{ does not appear in } m_1) * \Pr(x_1 \text{ does not appear in } m_2) = \left(1 - \frac{k_1}{n}\right) \left(1 - \frac{k_2}{n}\right)$$

which elementary algebra reduces to $\left(\frac{n-k_1}{n}\right) \left(\frac{n-k_2}{n}\right)$ or more simply $\frac{(n-k_1)(n-k_2)}{n^2}$. We can then generalize this statement to find the probability of x not appearing in any extant manuscript:

$$\begin{aligned} \Pr(x_1 \text{ does not appear in any MS}) &= \left(1 - \frac{k_1}{n}\right) \left(1 - \frac{k_2}{n}\right) \cdots \left(1 - \frac{k_y}{n}\right) = \left(\frac{n-k_1}{n}\right) \left(\frac{n-k_2}{n}\right) \cdots \left(\frac{n-k_y}{n}\right) \\ &= \frac{(n-k_1)(n-k_2) \cdots (n-k_y)}{n^y}. \end{aligned}$$

If we have a formula for the probability that any given (original) piece is not known to us, then we can use the principle of expected value (discussed above) to estimate how many pieces we would expect to be missing today, given the manuscripts we have and the number of pieces there once were in the trecento. (Note though that the probability of x not appearing in any MS, and the expected number of such pieces, each depends on n , the original number of pieces in the trecento—which is exactly what we are trying to find in the first place! This obstacle will be worked out soon).

The expected value of the number of pieces not appearing in any MS that survives today is simply the probability that any given piece does not appear in any manuscript multiplied by the total number of pieces, our unknown n :

$$\begin{aligned} \text{EV(missing pieces)} &= n * \Pr(x \text{ does not appear in any ms}) = n * \frac{(n-k_1)(n-k_2) \cdots (n-k_y)}{n^y} \text{ or} \\ &\frac{(n-k_1)(n-k_2) \cdots (n-k_y)}{n^{y-1}} \end{aligned}$$

It looks like we have two unknowns here: the expected number of missing pieces (EV) and the total number of pieces (missing or known), n . But what is the expected number of missing pieces? It is simply the number of pieces that were written originally (n) minus the number we currently have (let us call that number r).

$$\text{EV(missing pieces)} = n - r$$

So we can substitute back into the previous equation:

$$n - r = \frac{(n - k_1)(n - k_2) \cdots (n - k_y)}{n^{y-1}}$$

In this equation, r and k_1, k_2, \dots, k_y are all numbers we know, so n is our only variable. However, solving for n in this equation is still not easy when y is a number above three or four—since y is the number of manuscripts containing pieces in a particular genre, y will be on the order of ten to thirty.

Since the last equation is too complicated to solve directly, reducing it would have required tricky math decades ago. However, the solution can be closely estimated in seconds through computer-assisted “trial and error”. We rewrite the previous equation as:

$$n - r - \frac{(n - k_1)(n - k_2) \cdots (n - k_y)}{n^{y-1}} = 0$$

and then write a program to try various numbers of n (theoretically, from $r + 1$ to infinity, but from $r + 1$ to 2,000 is good enough) until it finds the n which comes closest to solving this equation. By closest, one means which comes closest to making the left side of the equation zero. (We are unlikely to find the exact solution since n can be a fraction rather than a whole number).

In the dissertation, a perl program for solving this equation is attached, the results of which appear in the body of the dissertation. The remainder of the appendix is dedicated to ways of checking whether or not our initial suppositions about the seeming randomness of scribal behavior and manuscript loss are valid (in fact, they are, to an extent that even surprised me).

MIT OpenCourseWare
<http://ocw.mit.edu>

21M.220 Early Music
Fall 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.