

THE NUMBERS SPEAK

16 3 6 21 4 21 11 27 13 10 9 9 15 10 21 7 5 7 13 6 8 9 5 7 6 5 3 5 2 9 3 9
9 8 3 4 2 3 3 3 2 2 2 4 3 4 6 4 4 1 2 1 2 3 2 1 2 4 2 6 2 10 4 6 3 7
11 26 73 5 2 3 2 6 8 11 7 16 2 3 1 14 8 6 6 8 4 2 12 19 3 7 6 2 3 7 3
4 2 9 3 10 4 5 2 3 14 15 3 7 9 2 10 4 12 116 3 16 9 4 2 3 3 6 21 7 14 5 10

THIS MAY LOOK LIKE A MEANINGLESS JUMBLE OF NUMBERS AT FIRST SIGHT, BUT LOOK AT IT AGAIN. THE NUMBERS IN THE SECOND ROW ARE NEARLY ALL APPRECIABLY SMALLER THAN THOSE IN THE FIRST ROW. THERE ARE ALSO A FEW BRIEF SEQUENCES OF THE SAME NUMBERS, SUCH AS THREE THREES AND THREE TWOS IN THE SECOND ROW. THESE LATTER ARE LESS SIGNIFICANT THAN YOU MIGHT THINK. THE NUMBERS ACTUALLY REPRESENT A SERIES OF SHOT LENGTHS IN THE MIDDLE OF *The Adventures of Robin Hood* (1938), AND THEIR LENGTHS HAVE BEEN RECORDED TO THE NEAREST FOOT OF 35 MM. FILM. IN FACT, THE REAL LENGTHS OF THE SHOTS MAKING UP THE SEQUENCE OF THREE SHOTS IN THE SECOND ROW WITH NOMINAL TWO FOOT LENGTH ARE 39 FRAMES, 34 FRAMES, AND 33 FRAMES. THAT IS, TWO FEET PLUS SEVEN FRAMES, TWO FEET PLUS TWO FRAMES, AND TWO FEET PLUS ONE FRAME, SINCE THERE ARE 16 FRAMES FOR EACH FOOT OF 35 MM FILM. IN FILMS IN GENERAL IT IS EXTREMELY RARE TO FIND TWO SHOTS NEXT TO EACH OTHER WITH EXACTLY THE SAME LENGTH, TO THE VERY FRAME.

ANOTHER POINT ABOUT THE SEQUENCE OF NUMBERS AT THE BEGINNING OF THIS PIECE, WHICH IS LESS OBVIOUS, IS THAT THERE ARE MORE SMALL NUMBERS THAN BIG NUMBERS. THERE ARE ONLY 28 SHOT LENGTHS IN DOUBLE FIGURES OUT OF THE TOTAL OF 128. MORE IMPORTANTLY, THERE ARE ONLY 52 SHOTS GREATER IN LENGTH THAN THE AVERAGE SHOT LENGTH (ASL) FOR THE SEQUENCE, WHICH IS 6.4 FEET, WHEREAS THERE ARE 70 SHOTS SHORTER THAN THIS AVERAGE.

There is greater significance in this series of shot lengths if we look at the scenes in which they occur. The sequence

16 3 6 21 4 21 11 27 13 10 9 9

corresponds to the tavern scene in which Lady Marion goes to Robin Hood's men to organize an attempt to rescue him. The ASL for the scene is 13.3 feet, which is much longer than the overall ASL for the whole film of 7.15 feet.

The next section is:

15 10 21 7 5 7 13 6 8 9

which makes up the beginning of the execution scene. Here we see the VIP stand in the town square with the bad guys gloating over Robin's impending death. The ASL for this scene speeds up from the previous scene to 10.1 feet, though still longer than the average.

Then follows an increase in the excitement of the crowd and the dramatic tension accompanied by faster cutting as Robin is brought into the square and led onto the gallows:

5 7 6 5 3 5 2 9 3 9 9 8 3 4 2 3 3 3 2 2 2 4 3
4 6 4 4 1 2 1

The ASL here is 4.13 feet.



The next section of the scene starts with the first of the arrows fired by Robin's men to kill his executioners, and continues with his escape and the resulting chaos in the square. The shot lengths are

2 3 2 1 2 4 2 6 2 10 4 6

and the high speed of the action is accompanied by even faster cutting, with a sectional ASL of 3.66 feet.



The next section is, strictly speaking, a new scene, as it takes place in a different part of the town to the three previous sections, which are all set in the town square. So as Robin and his band flee through the town, and out the town gate, rather surprisingly the cutting rate slows down a bit, with these shot lengths:

6 2 3 7 3 4 2 9 3 10 4 5 2 3 14 15 3 7 9 3

which give a sectional ASL of 5.7 feet. This is still appreciably faster than the overall average for the film, but the longer shots are presumably left entire to show the real natural agility, strength, and grace of the movement of Robin Hood (or rather his double) in riding the end of the rope up the gatehouse, and over the top.



After a fade out and fade in, we come to the scene outside, and then inside, Lady Marion's chamber, as Robin Hood scales the castle wall, and Marion confides her love for Robin to her maid:

7 11 26 73 5 2 3 2 6 8 11 7 16 2 3 1 14 8 6 6 8 4
2 12 19 3 7 2 10 4 12

The ASL for this section of 10 feet, which is of course slower than the average.



Then the real love scene between Robin and Marion follows:

116 3 16 9 4 2 3 3 6 21 7 14 5 10

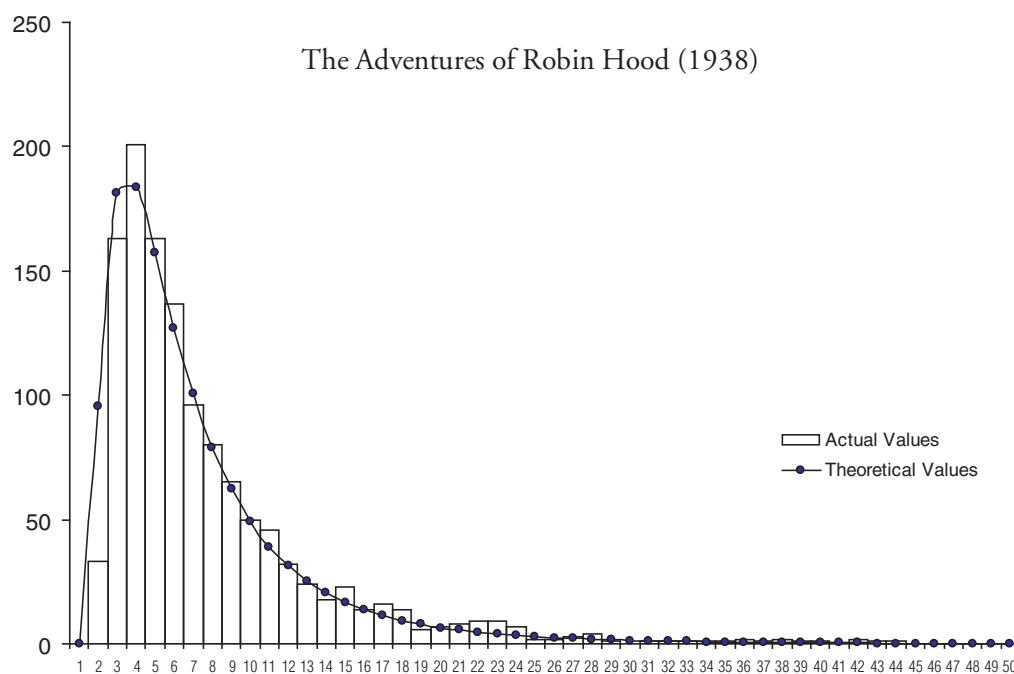
in which the cutting slows down even further, to an average length of 15 feet, so following the normal expressive pattern of film editing, in which cuts are not allowed to disturb the union of the lovers.



The same sort of expressive variation in cutting rate, depending on the nature of the scene, can also be studied in this book on pages 266-267 for the first episode of the Star Trek television series, and on pages 177, 303 and 308 of *Film Style and Technology* for *The Iron Horse*, *Liebelei*, and *Letter from an Unknown Woman*.

Turning to the general use of different lengths of shot for different films, this is best investigated by taking the complete sequence of shots in the film and counting the number of shots with lengths between zero and one foot, between one

foot and two feet, two feet and three feet, and so on. These divisions of the quantity in question are properly called "class intervals", but the popular name for them is "bins". The handy way to display the results of this enumeration is as a graph of the histogram kind. Histograms are a particular type of bar chart in which the bars are in contact with each other, which signifies that the quantity being described covers a continuous range, and is not just measured in simple integers. For *The Adventures of Robin Hood* we get a graph like that following:



The asymmetrical shape of the distribution of shot lengths for *The Adventures of Robin Hood* is described in a broad way by the difference between the mean or average value of length of the 1,254 shots making up the film, which is 7.15 feet, and the median value for the shot lengths; that is, the value that separates the 50% of the shots with larger values from the 50% with smaller values. The median shot length is 5 feet. In other words, most of the shots in the film are shorter than the average shot. (The mean or average only coincides with the median for symmetrical distributions, such as the well-known Normal or Gaussian distribution.) The actual or observed distribution of shot lengths for *the Adventures of Robin Hood* is well fitted by a standard theoretical statistical distribution called the Lognormal distribution. The probability density function $f(x)$ for a shot of any particular length x is given by:

$$f(x) = \frac{1}{x\sigma(2\pi)^{1/2}} \exp\left\{-\frac{[\log(x/m)]^2}{2\sigma^2}\right\}$$

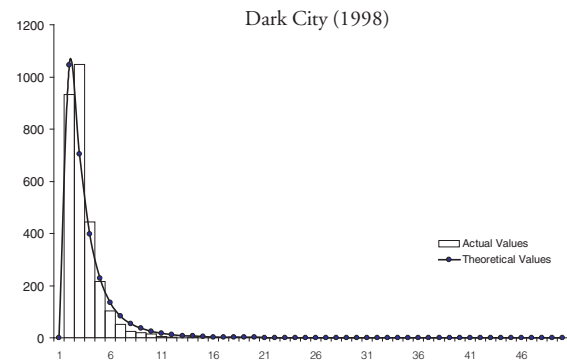
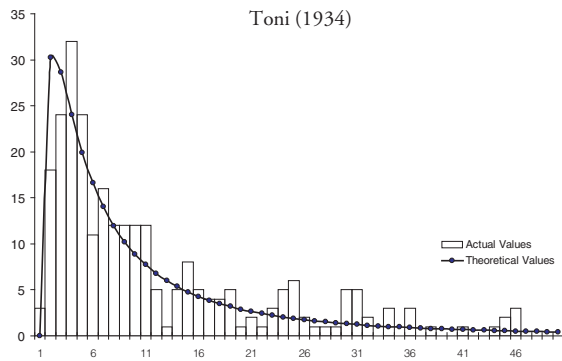
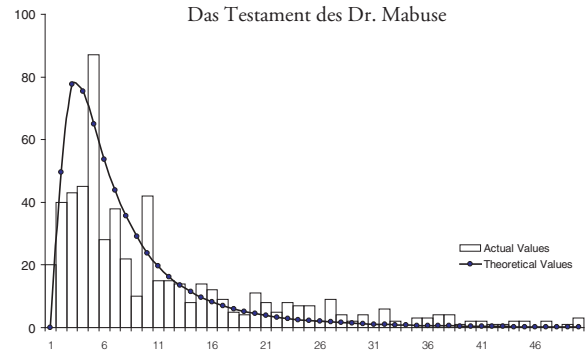
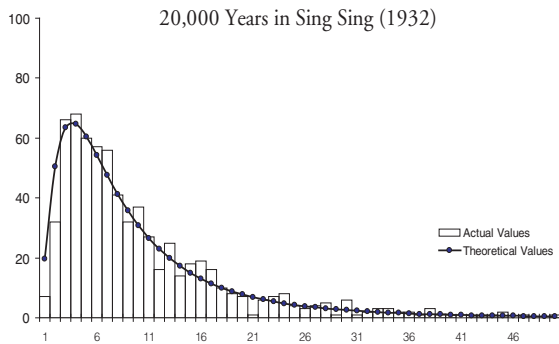
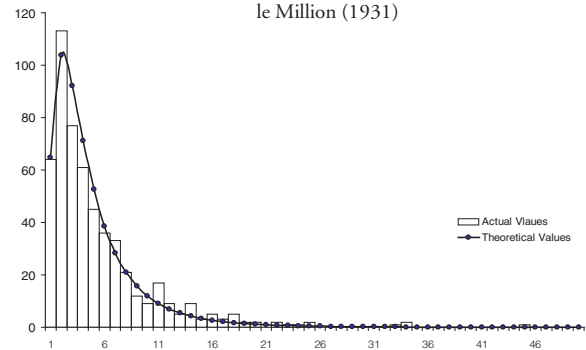
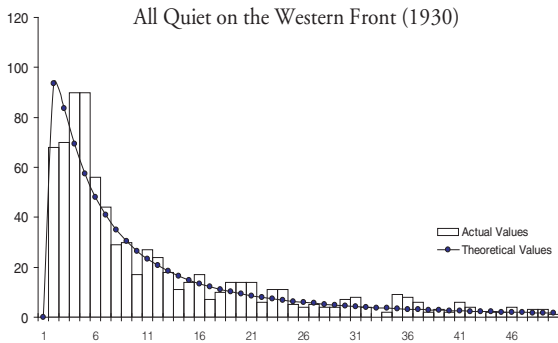
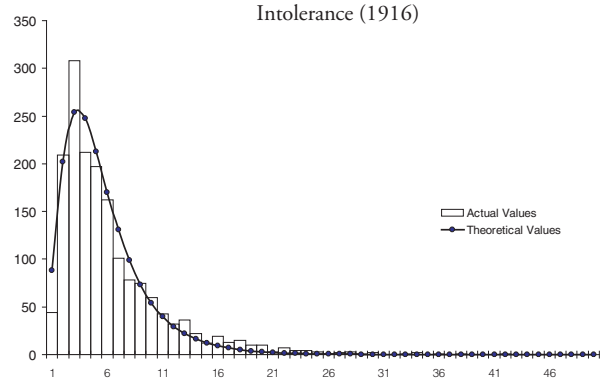
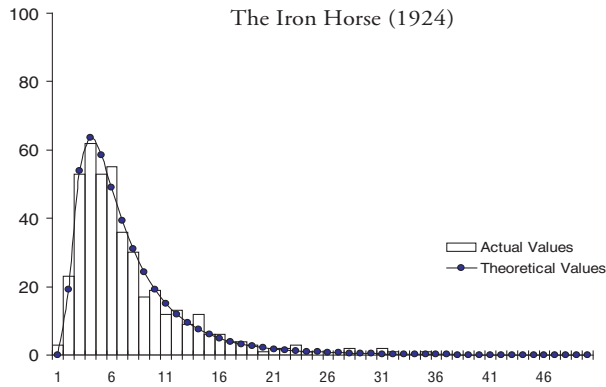
where σ is the shape factor, and m is the median value of the distribution.

$f(x)$ multiplied by the number of shots in the film is plotted on the graph above as a continuous line using the values of the median (the scale parameter) and the shape factor, which is the standard deviation of the logarithm of the shot lengths, both obtained from the actual values of the shot lengths for this film. You can see that it corresponds very closely to the actual values. There is a standard method of giving a value to the accuracy of fit between two sets of quantities; in this case the

observed and theoretical values for the shot lengths, the latter assuming that the distribution is lognormal. It is called the correlation coefficient, and is best used with its value squared, written R^2 , for comparison purposes. If the correspondence was perfect, its value would be 1, but in this case it is 0.956, which is still very good as these things go.

The Lognormal distribution is found to apply to many varied phenomena, particularly in economics. One example of this is the numbers of insurance claims for different amounts of money as a result of damage to motor cars in accidents. But more interestingly to us, it has also been found to apply in literary statistics, where the numbers of sentences of various given lengths in a stretch of prose conforms to the Lognormal distribution. Theoretically, the Lognormal distribution results when the quantity under consideration, in our case shot length, is determined as a result of the probabilities associated with a large number of independent causative factors being multiplied together. In films what is presumably concerned in determining the length of a shot is the simultaneous interaction of such factors in the scene being filmed as how the actor moves in the shot with respect to the closeness of the camera, the length of the lines he speaks, and how the other actors react, and so on. The fact that different individuals are usually responsible for these various components of a film, from the scriptwriter to the director to the editor, assists the independence of these causes.

The generality of the Lognormal distribution for shot lengths in movies is illustrated by some examples from films made between 1916 and 1998.



Number of shots with lengths within the given length intervals

These are all taken from the complete lengths of the films, with the exception of *The Iron Horse*, which covers the first 40 minutes of the film. You will notice that nearly all the lengths

of shots are recorded either in 35 mm. feet or 16 mm. feet. This is because the prints that came my way in the nineteen-seventies were in both gauges.

More than that, some of the flat-bed editing machines I was using in those days had footage counters that did not indicate lengths to the exact frame, so in many cases I only took the shot lengths to the nearest half foot of film, either in 35 mm. or 16 mm. However, some of the 16 mm. films were measured on an Acmade picture synchronizer, which did have an indicator for frames as well as feet. It might occur to you that mathematical conversion is always possible between the lengths for the different film gauges, but if this is done when the results are only taken to the nearest foot, the consequent rounding introduces an extra unevenness when the resulting distribution is plotted. So I have left the footages in their original form. Finally, *Dark City* was analysed much more recently, using a DVD copy fed into a non-linear editing program on a computer, so the shot lengths are in seconds in that case.

As you can see, these observed results fit well with the Lognormal distribution, and the values for R^2 quoted in the

table below confirm this. It is noticeable that the values for the shape factor for most of the films is in the region of 0.7 to 0.9. This is the reason for my assertion long ago in *Film Style and Technology* (p. 225) that given the ASL for a film, one can be fairly certain of the proportion of the shots that will be longer than a certain value, and so on. There are exceptions to this, as here in the case of *Toni* and *Good News*, with shape factors of 1.25 and 1.30, and *Intolerance* and *All Quiet on the Western Front*, where it is down near 0.5. This means that in the former case the distribution is not quite so skewed towards the left, and in the two latter case it is more skewed towards the left: that is, towards the y-axis. The first of these deviations seems to be connected with the long ASLs, around 20 seconds, of these two films. Though the reason for the deviation towards a low value of the shape factor in the case of *Intolerance* and *All Quiet on the Western Front* is obscure to me. So the shape factor seems to be fairly independent of the ASL up to around 20 seconds. The Godard films are in a world of their own.

Title	ASL	Shape Factor (σ)	R^2
Intolerance (1916)	6.00	0.529	0.963
The Iron Horse (1924)	6.00	0.708	0.984
All Quiet on the Western Front (1930)	9.30	0.501	0.884
le Million (1931)	13.00	0.754	0.885
20,000 Years in Sing Sing (1933)	6.40	0.829	0.955
Das Testament des Dr. Mabuse (1933)	12.00	0.876	0.948
Toni (1934)	19.00	1.253	0.832
The Adventures of Robin Hood (1938)	5.00	0.783	0.956
Dark City (1998)	1.87	0.811	0.938
Good News (1947)	20.00	1.299	0.760
Carmen Jones (1954)	46.00	0.932	0.206
Black God, White Devil (1964)	21.08	1.670	0.955
Une femme mariée (1964)	21.00	1.309	0.945
Catherine the Great (1934)	8.50	0.792	0.950
The Scarlet Empress (1934)	10.00	0.916	0.728
Vivre sa vie (1962)	20.95	1.339	0.848
Vivre sa vie (reduced)	27.16	1.067	0.839
Week-end (1967)	26.50	2.037	0.898
Week-end (reduced)	38.12	1.847	0.573
Sauve qui peut (la vie) (1980)	6.30		
Sauve qui peut (la vie) (reduced)	19.40	1.642	0.82

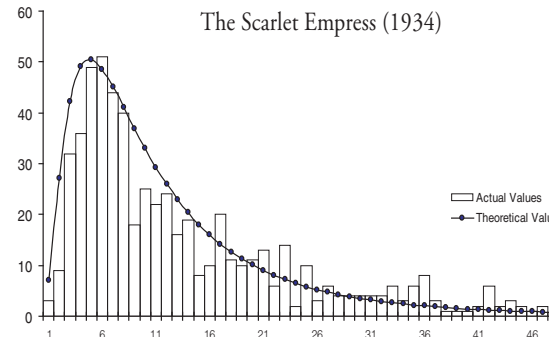
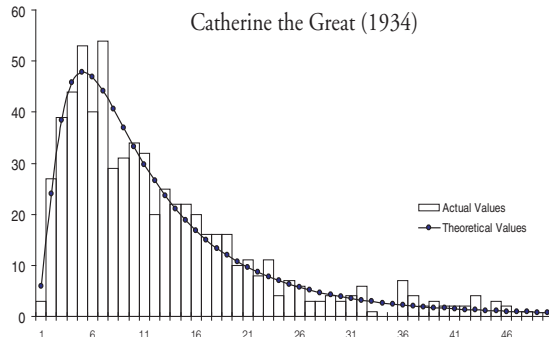
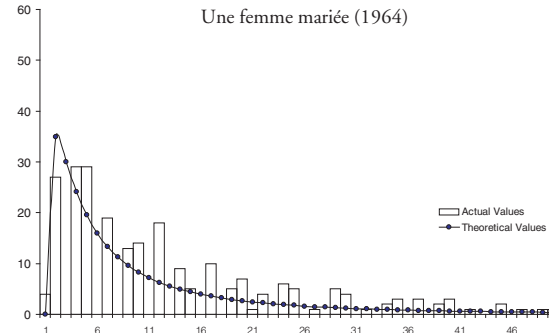
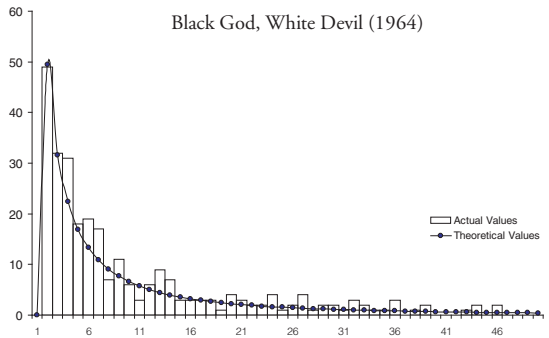
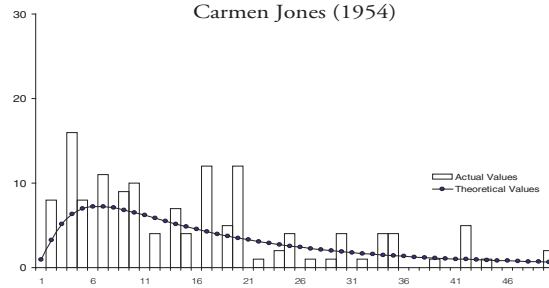
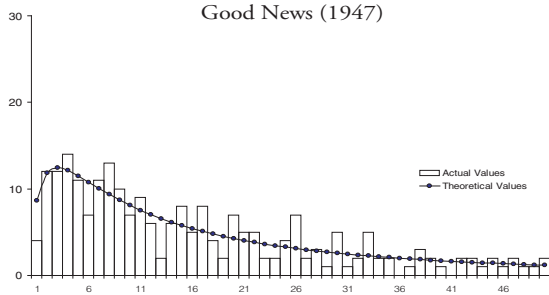
For ASLs above 20 seconds, the fitting of the actual shot length distributions to the lognormal distribution is markedly less good, and an example of this is *Carmen Jones* (1954),

which has an ASL of 46 seconds, as can be seen from the value of R^2 for goodness of fit in the table above, and the graph with comparative lognormal distribution on the next page.

For *Carmen Jones*, the Gamma distribution proves a slightly better fit (not illustrated), but is still not really good. In any case, films with very long ASLs are only a small minority of all films: in my collection of just over 12,000 Average Shot Lengths, only about 120 are greater than 25 seconds, which is around one film in one hundred.

It is possible to get quite good correlations of observed

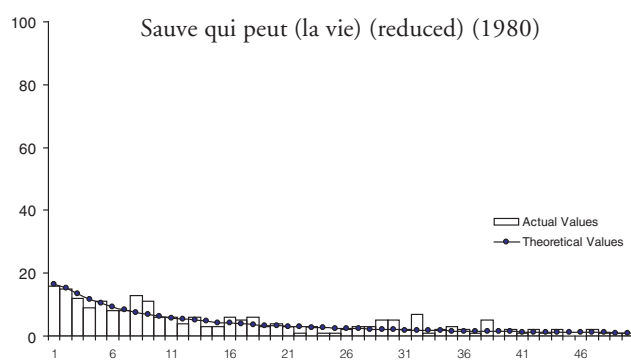
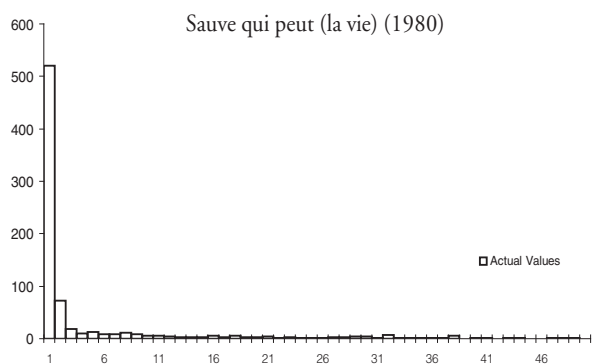
distributions with the theoretical distribution for films with ASLs of around 20 seconds, as the cases of *Black God, White Devil* (1964) and *Une femme mariée* (1964) demonstrate. The values of R^2 are 0.955 and 0.945 for these two films. Glauber Rocha's *Antonio das mortes* which was made immediately after his *Black God, White Devil*, is very similar indeed in style, and hence statistics, though I am not illustrating it here.



Number of shots with lengths within the given length intervals

The comparison between *The Scarlet Empress* (1934) and *Catherine the Great* (1934) is particularly interesting. These two films, made on the same subject almost simultaneously in the United States and England, have already been compared in other respects in the article *Sternberg's Heart Beats in Black and White* earlier in this book. As far as their shot length distributions are concerned, *Catherine the Great* is an appreciably better fit to the lognormal distribution, as is obvious from a glance at the two distributions, not to mention the values quoted in the table for R^2 on the previous page. Very noticeable in *The Scarlet Empress* graph is the way the numbers of shots of certain lengths are far below the expected value.

This is particularly true of shots of 9 feet, 11 feet, 14 feet, 16 feet, 24 feet, and 26 feet, though there are some other values that are rather low as well. I think this is very probably due to the lengths of the shots in the film being cut to accommodate the regular pulsation of light and dark at the centre of the frame, as is described in *Sternberg's Heart Beats in Black and White*; remember that the main frequency detected there was 160 frames (10 feet), with an occasional double frequency of 80 frames (5 feet), and also a section at 127 frames (7 feet 13 frames). The way this effect is observed in the shot length distribution is a confirmation of the remarkable phenomenon described in that previous article.



Number of shots with lengths within the given length intervals

Jean-Luc Godard's films mostly conform to the lognormal distribution, despite their peculiarities and large ASLs. The one exception to this is *Sauve qui peut (la vie)*, where the hiccuping freeze frame effect applied to Nathalie Baye's periodic bicycle excursions are counted as shots. Each series of these has a different equal length for its constituent shots; 8 frames, 10 frames, 11 frames, and so on. The resulting distribution has an extremely large peak for shots of 0 to 1 foot in length, and certainly cannot be fitted by a Lognormal distribution, or indeed any other standard statistical distribution. However, if one does not count these freeze frames as separate shots, the resulting shot length distribution, which I call a "reduced" distribution, is then a pretty satisfactory Lognormal distribution.

One might expect that the apparently random title cards announcing things like "Un film égaré dans le monde" and "Un film trouvé sur le ferroviaire" or more pointedly "???" etc. cut into *Week-end* might disturb its Lognormal purity, but this is not so. If one removes them from the listing of shot lengths, one still gets another Lognormal distribution. Strangely enough, the reverse happens if one removes the many inserts of images from women's magazines cut into *Une femme mariée*. This reduces the film's fit with the Lognormal distribution. For *Vivre sa vie*, which does not have lots of intertitles or insert shots extraneous to the story, but does have two sequences from Dreyer's *Passion of Joan of Arc* cut in when the heroine of the film watches it in the cinema, the fit with the Lognormal distribution is about as good without as with. That is Godard for you. Always perverse.

Inside the Atom

The patterning in shot lengths described in *The Adventures of Robin Hood* is undoubtedly due to a degree of standardization in the way that scenes in a film script are broken down into separate shots when the scene is filmed, and then standardization in the way it is later edited. The standard approach to film script-writing developed from the example of the "well-made play" of the late nineteenth century stage, and here the practice was to alternate sections involving much dramatic tension or action with sections of exposition or light

relief or romance. This happens both within scenes, and also from one scene to the next. Because film scenes are much shorter than play scenes, it often happens that the action in a whole film scene completely falls into one emotional category, but in such cases it is likely to be followed and preceded by a scene of the other kind. Starting from around the First World War, the standard practice in making films from film scripts was to break down scenes (or parts of scenes), involving dramatic tension, and also action scenes, into a large number of shots of short length, while the alternating romantic, or comic, or expository scenes (or parts of scenes), were shot with a smaller number of shots of longer duration. The intention here, which is quite consciously formulated by film-makers, is for an "expressive" intensification of the dramatic material in the film script being filmed.

The way in which there are long strings of shots of fairly similar length in *The Adventures of Robin Hood*, and indeed in other films is best measured by the autocorrelation coefficient, which computes a kind of average over the length of the series for the difference between the size of one value and that of the next in the series, or between one value and the value after the next, or between one value and that two values after the next, and so on. These different degrees of separation give autocorrelation coefficients of the 1st. order, 2nd. order, 3rd. order, and so on. (The terms lag 1, lag 2, lag 3, etc., are also used as alternative names to describe these orders). For sequences in which the successive values are very close indeed to each other, and hence to their average value, the autocorrelation coefficient approaches 1, while for sequences in which each successive value is as different as possible from the one before (i.e., a very large value is always followed by a very small value), the autocorrelation coefficient approaches -1, and for sequences in which the values are distributed completely at random, the autocorrelation coefficient is zero.

When the autocorrelation coefficient is computed it confirms what is usually visible to the experienced eye when it runs down the sequence of actual values of shot lengths in a film, which is that there are a few films which do not conform to the general pattern I have described and illustrated above for *The Adventures of Robin Hood*.

The usual value of the autocorrelation coefficient for films constructed on the standard pattern of shot breakdown (“classical cinema”) is around 0.2, or as a statistician would put it, the shot lengths show weak autocorrelation, but there are a few groups of films in my sample for which there is very nearly no autocorrelation at all in shot lengths. Of these, the two Frederick Wiseman films, *Law and Order* and *Hospital* are not fiction features, but *cinéma vérité* documentaries, for which it is no surprise that the standard pattern for fictional films does not hold, as different considerations apply in the shooting and editing of such films. In particular, the notion of varying the cutting rate according to the nature of the scene cannot apply when a single camera is used running continuously all the way through the scene. If multiple cameras were used shooting simultaneously all the way through all the real events recorded in the documentary, it would be possible to impose different cutting rates on different scenes.

For *The Scarlet Empress*, the very peculiar structural principles being used, which I have previously described, have an effect on autocorrelation of shot lengths, but I can think of no reason for the anomalous value for *The Front Page* (1931).

Amongst Jean-Luc Godard’s films in the sample, *le Petit soldat* and *Vivre sa vie* have about the usual degree of autocorrelation in their shots lengths, while *Pierrot le fou* is a bit on the low side at 0.110, but *Une femme mariée* and *Week-end* have close to zero autocorrelation. The latter instance is not surprising, given the way that fairly short title cards are stuck into the middle of the long takes that the action of *Week-end* is carried through. In *Une femme mariée*, there are not only the magazine page inserts, but also brief inserts of signs and details of posters making plays on words such as the famous “danger” -- “ange”, etc. The result of removing these inserts from the statistics of these two films gives the autocorrelations for these “reduced” versions tabulated at the right. In *Sauve qui peut (la vie)*, the very high autocorrelation of 0.396 results from the chains of equal length freeze frames, but if we remove these, it goes back to a normal 0.146.

Glauber Rocha’s *Black God, White Devil* and *Antonio das mortas* were consciously made in a variation of the latest advanced or “art film” style of the period, and to some extent influenced by Godard’s work. However, their formal style is very different in the aspect I am considering, for they alternate shots of very long duration with long sequences of very short shots with very similar lengths indeed, and these latter are cut

Title	Auto-correlation
The Iron Horse (1924)	0.290
All Quiet on the Western Front (1930)	0.208
le Million (1931)	0.199
The Front Page (1931)	0.047
La chienne (1931)	0.251
Boudu sauvé des eaux (1932)	0.186
20,000 Years in Sing Sing (1932)	0.133
Das Testament des Dr. Mabuse (1933)	0.163
Toni (1934)	0.228
The Adventures of Robin Hood (1938)	0.174
Dark City (1998)	0.287
Good News (1947)	0.297
Carmen Jones (1954)	0.205
Black God, White Devil (1964)	0.347
Une femme mariée (1964)	0.013
Catherine the Great (1934)	0.116
The Scarlet Empress (1934)	0.052
Law and Order (1969)	0.022
Hospital (1970)	0.062
le Petit soldat (1960)	0.185
Vivre sa vie (1962)	0.169
Vivre sa vie (reduced)	0.164
Pierrot le fou (1965)	0.110
Week-end (1967)	-0.047
Week-end (reduced)	0.050
Sauve qui peut (la vie) (1980)	0.396
Sauve qui peut (la vie) (reduced)	0.146

to length without much regard for the action contained in each shot. It is these sequences with strings of equal length short shots which push the autocorrelation coefficient up to about 0.34 for the whole film, just as happens in Godard’s *Sauve qui peut (la vie)*.

Autocorrelation statistics represent the lowest order of Markov chain analysis, which I mentioned as a means of analysis in *Let a Hundred Flowers Bloom*, long ago. The full development of this method requires a complete tabulation of the significant variables for each shot in a film, taken in order.