

A PROBABILITY DISTRIBUTION FOR CROP YIELDS IN POLAND

TADEUSZ GÓRSKI

Department of Agrometeorology and Applied Informatics,
Institute of Soil Science and Plant Cultivation – State Research Institute
Czartoryskich 8, 24-100 Puławy, Poland
E-mail: tgorski@iung.pulawy.pl

Abstract: An important source of risk in agricultural production is the variability to crop yields reflecting irregularly changing weather. This variability may be described as a stochastic process that has a function of density. Analyses of historical data on crop yields reveal that the function of density changed from right-skewed to left-skewed, along with increasing mean yields. All examined yields of crops cultivated recently in Poland demonstrate the left skew, which does not diminish with the aggregation of acreage. A fairly good approximation of the probability distribution for actual yields may be obtained using the log-normal distribution with an inverted abscissa.

Keywords: crop yields, density function, log-normal distribution, area aggregation, scaling of variance

INTRODUCTION

Irregular inter-annual changes in yield are usually ascribed to weather conditions. This variability may be formalized and calculated on the basis of residuals in detrended multi-year series of yields. Where a trend cannot be approximated by a low-order function (as during the economic transition in Poland), use may be made of the residuals of production function (based mainly on agrochemicals applied) (Fig. 1). Another method usable in determining residuals is the weather index (dimensionless WI, Fig. 2), which is a function of meteorological variables (Górski *et al.* 1994; Górski and Górska 1999), and may be treated as a percentage deviation from the expected value.

The climatic variability of yields may be described as a stochastic process having a function of density, with the mean as the expected value. A knowledge of this function

(i.e. the type of distribution and its parameters) is necessary for the quantitative assessment of the probability of extreme situations (yield risk), because all the risk calculations are sensitive to distributional assumptions (Nelson 1990).

TYPES OF STATISTICAL DISTRIBUTION

As revealed by extensive studies of historical and more recent data on yields (Day 1965; Gallagher 1986; Ramirez *et al.* 2003; Sherrick *et al.* 2004), the form of the distribution depends on the crop, and may depend on the average yield level and many local or regional conditions (mainly climatic). Biological and technical progress has allowed mean values and standard deviations for yields to increase considerably over the last two centuries. It may be stated that the coefficients of variability remain fairly stable (Austin and Arnold 1989; Górski and Górska 1999).

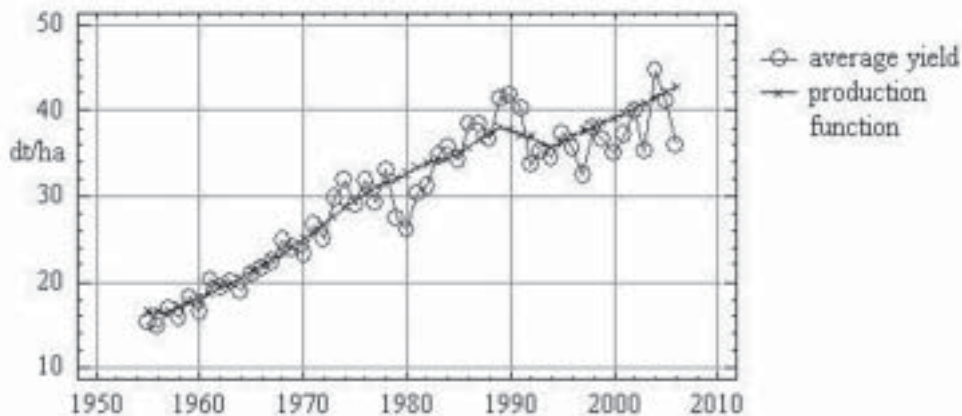


Figure 1. Winter wheat yields in Poland.

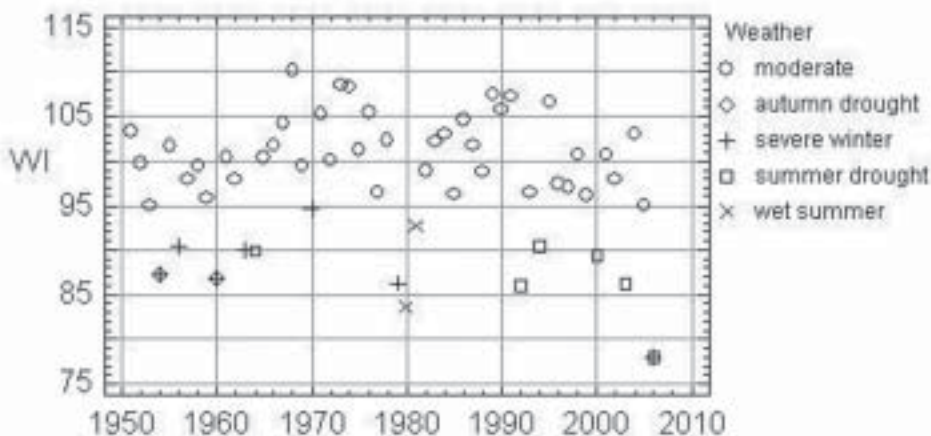


Figure 2. Weather index (WI) for winter wheat in Poland. Main causes of low yields are indicated.

The determination of yield probability is also complicated by the type of statistical distribution, which is often far from normal (Gaussian). This problem has been discussed widely (Gallagher 1986; Just and Weninger 1999; Atwood *et al.* 2003; Ramirez *et al.* 2003), with many cases of skewed distributions (both negative and positive) being described.

In the case of Poland, the density functions for low cereal yields in the 18th century were often skewed positively (Fig. 3), but later – at higher mean yields – the function tended towards normality (Fig. 4). The same trend might be observed elsewhere, for example in the Kansas series of wheat yields over the years 1866-2004 (Kansas Department of Agriculture 2004). By the end of the

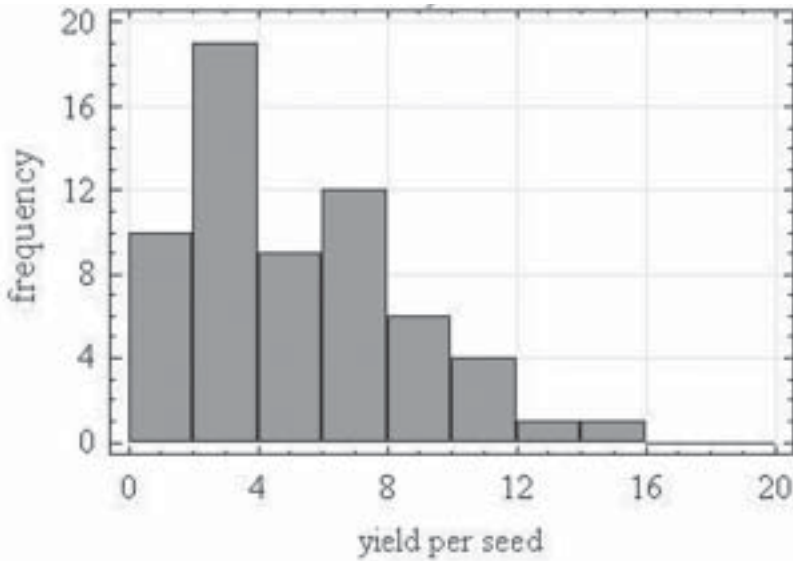


Figure 3. Wheat yields on three Podolian estates (1770-1790).
Source: Data after Rychlikowa (1968)

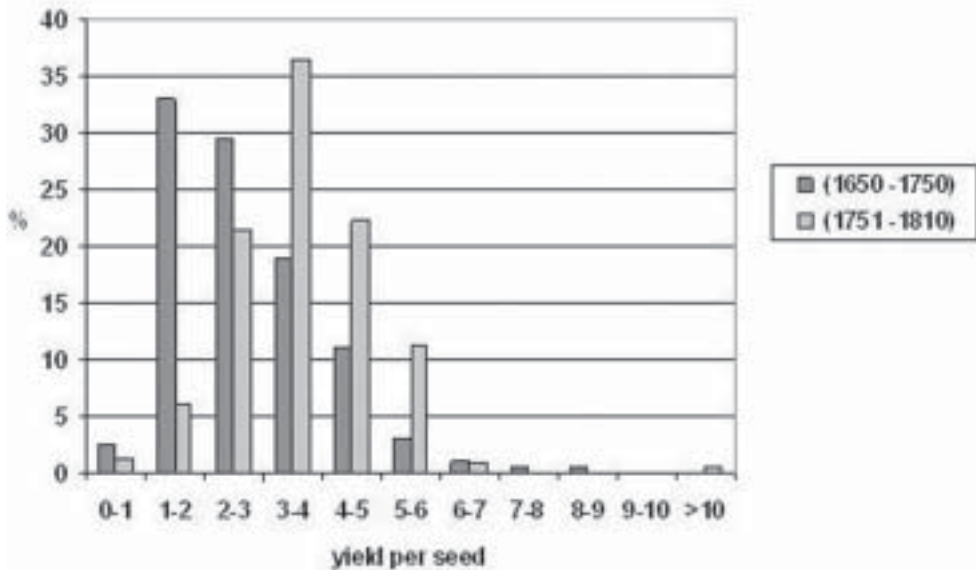


Figure 4. Relative frequency (%) of the Silesian yields of oats.
Source: Data after Żytkowicz (1970)

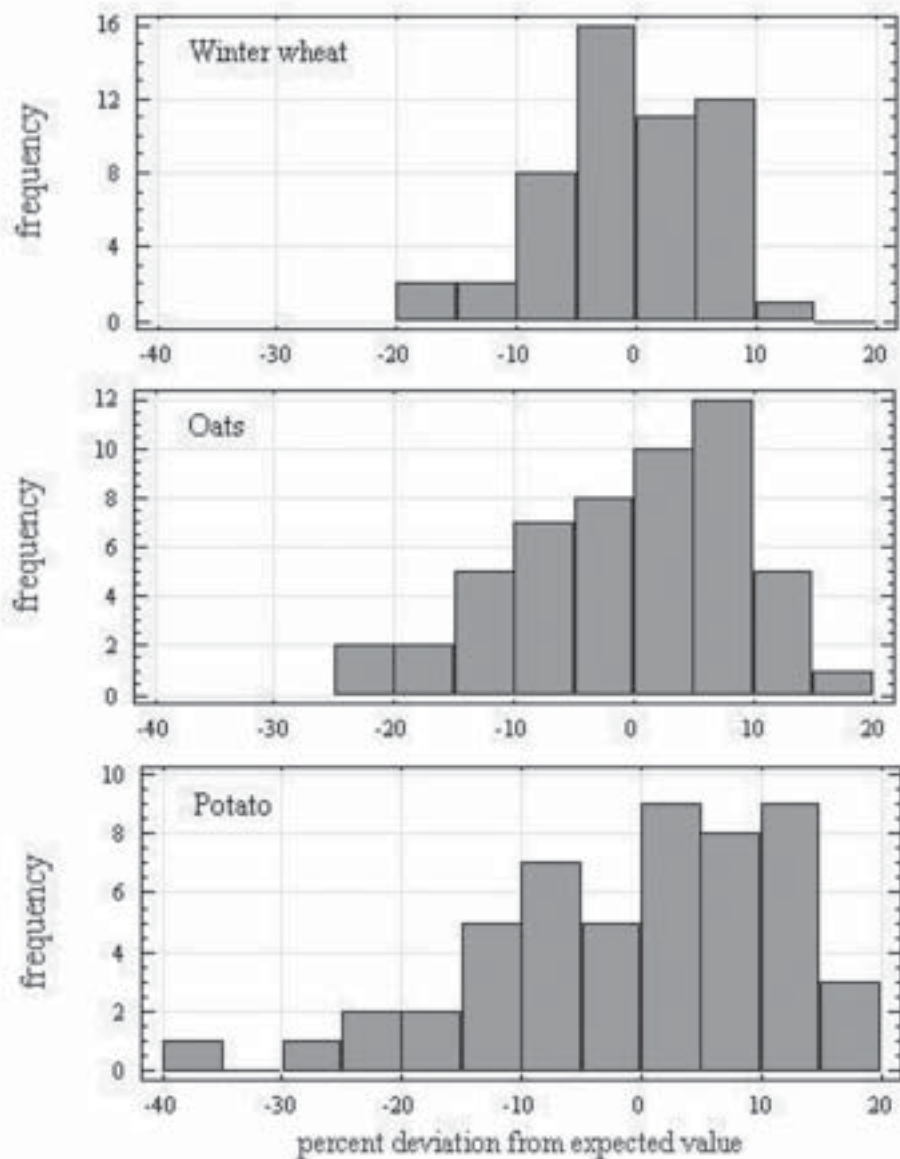


Figure 5. Climatic variability of mean yields in Poland (1955-2006).

20th century, crop yields (now several times higher) usually had a negatively skewed distribution (Skees *et al.* 1997; Górski and Górski 1999). It may be argued that the shape of the distribution is an effect of limiting conditions (Park and Sinclair 1993). In the first case, the natural limit was zero yield, now-

days – the upper limit may be determined by the biological potential of the crop. In Poland, all crop species examined demonstrate greater or lesser negative skew for their distributions (Fig. 5). All the recent data on crop yields came from the Central Statistical Office (GUS), with the exception of the results

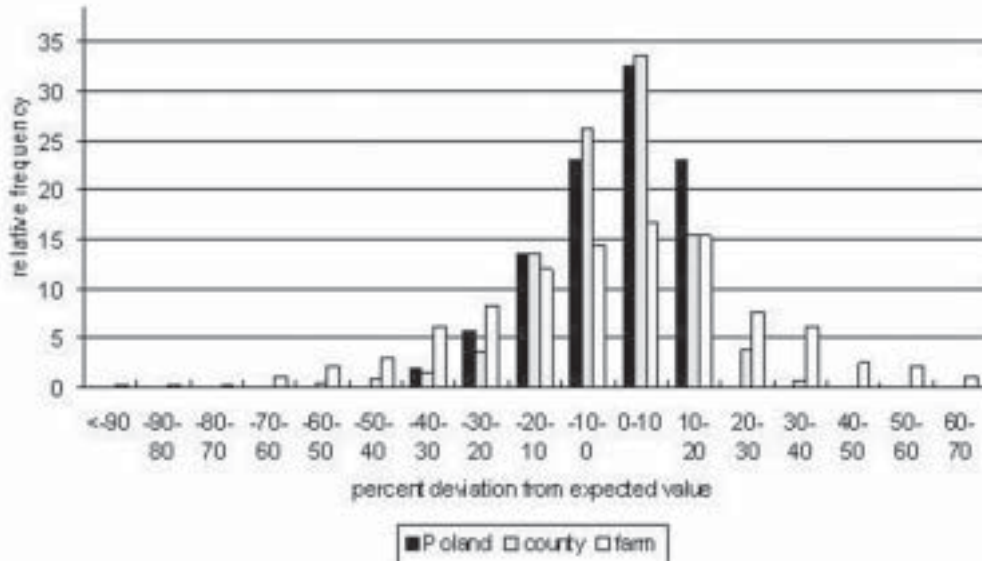


Figure 6. Climatic variability of potato yields in Poland.

for farms which derived from the Institute of Soil Science and Plant Cultivation (IUNG).

IMPACT OF AGGREGATION SCALE

The yield variance is shown to depend markedly on yield aggregation in space: from farm to national level. This regularity may be presented (Górski and Górka 2003) as:

$$\ln V = \ln v - b(n-1)^{0.2}$$

where V is the coefficient of yield variability in an aggregated area; v per unit area; n – the number of unit areas; b – the empirical parameter

Thus, for example, the coefficient of variability for winter wheat yield in Poland in relation to acreage may be calculated after the equation:

$$\ln V = 3,0 - 0,034 (n-1)^{0.2}$$

where n is area (ha)

However, the calculations of probability are complicated by the lasting asymmetry of the density function. Because the weather is strongly correlated throughout the country,

the central limit theorem of statistics does not apply in the case of aggregation of farm yields, and the distribution does not tend towards normality. The negative skew is apparent also at the region (county, voivodship) and country level. Moreover, the negative skew in Poland is commonly greater at the country than at the farm level (Fig. 6).

Table 1 shows the main quantiles of distribution function of winter wheat and potato yields in three scales of area aggregation. It may be seen that the crops differ in their spatial characteristics. The interpretation of such differences is possible on the basis of ecological features of the species.

CONCLUDING REMARKS

The three parameter log-normal distribution (Aitchison and Brown 1957) with an inverted abscissa gives a fairly good assessment of the density function of actual yields.

The random variable:

$$X = \ln (A - Y)$$

where A is the empirical parameter and Y the yield, has a normal distribution and

Table1. Extreme quantiles of crop yield distribution (percent probability of non - exceedance). Farm - about 300 ha; voivodship - about 600 kha. Table values are percent deviations from expected (mean) yield

Percentile	Winter wheat			Potato		
	Farm	Voiv.	Poland	Farm	Voiv.	Poland
1	-59	-34	-20	-64	-44	-37
5	-36	-19	-12	-44	-28	-25
95	32	17	9	41	19	17
99	54	21	14	58	23	20

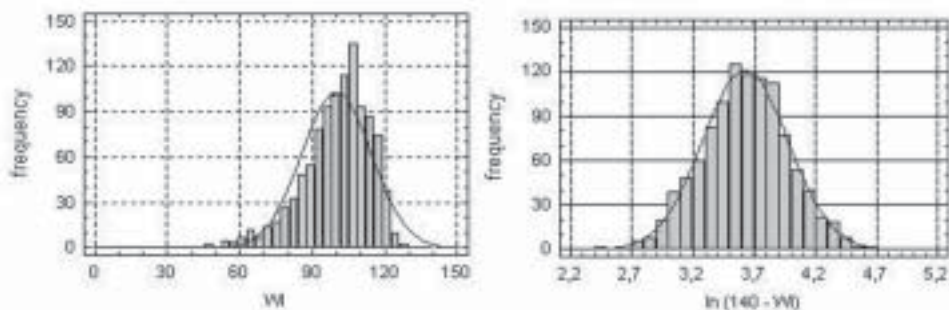


Figure 7. Histograms of potato weather indices (WI) in Poland (20 meteorological stations in the years 1951-2003). The lines show the density function of the normal distribution.

therefore a cumulative distribution function can readily be obtained from statistical tables. The parameter A (the starting point) may be determined using the method of quantiles (Q).

The following equation assumes symmetry of distribution:

$$\ln(A - Q_{50}) - \ln(A - Q_5) = \ln(A - Q_{95}) - \ln(A - Q_{50})$$

where Q_5, Q_{50}, Q_{95} are empirical quantiles

Hence, after simple transformations:

$$A = (Q_5 - Q_{95} - Q_{50}^2) / (Q_5 + Q_{95} - 2 Q_{50})$$

Quantiles Q_5 and Q_{95} can be replaced by another pair, in line with the quantity of empirical data and planned applications. The parameter A changes with the scale of aggregation and should be calculated separately for each scale. The log-normal function of density approximates well the empirical histograms of yields or WI values, particularly at the extremes (Fig. 7).

Knowing the parameters of the log-normal distribution, calculations of yield risk due to extreme weather conditions become relatively precise, as long as the statistical

distribution of climate elements influencing yields may be treated as stationary. Climate change will complicate the issue. It seems that in such a situation special models of weather impact on crop yield (among them weather indices) may improve risk calculations.

ACKNOWLEDGEMENTS

The work was supported by the Polish Ministry of Education and Science under Grant PBZ-KBN-086/PO4/2003.

REFERENCES

- Aitchison, J., Brown, J.A.C. (1957), *The log-normal distribution*, Cambridge University Press, Cambridge.
- Atwood, J., Shaik, S., Watts, M. (2003), Are crop yields normally distributed? A re-examination, *American Journal of Agricul-*

- tural Economics*, 85: 888-901.
- Austin, R.B., Arnold, M.H. (1989), *Variability of wheat yields in England: analysis and future prospects*, in Anderson, J.R., Hazel, P.B.R. (eds.), *Variability in grain yields*, Baltimore, London.
- Day, R.H. (1965), Probability distribution of field crop yields, *Journal of Farm Economics*, 47: 713-741.
- Gallagher, P. (1986), U.S. corn yield capacity and probability: estimation and forecasting with nonsymmetric disturbances, *North Central Journal of Agricultural Economics*, 8: 109-122.
- Górski, T., Demidowicz, G., Deputat T., Górka, K., Marcinkowska, I., Spoz-Pač, W., Krakowiak, A. (1994), Agrometeorological quantification of agricultural year in Poland, *Zeszyty Problemowe Postępów Nauk Rolniczych*, 405: 81-87.
- Górski, T., Górka, K. (1999), Zmienność klimatyczna plonów pszenicy ozimej w Polsce i wpływ agregacji powierzchni na ocenę ryzyka [Climatic Variability of winter wheat crop yields in Poland, and impact of surface aggregation on risk calculation], *Folia Universitatis Agriculturae Stetinensis, Agricultura*, 202, (79): 73-80.
- Górski, T., Górka, K. (2003), The effects of scale on crop yield variability, *Agricultural Systems*, 78: 425-434.
- Just, R.E., Weninger, Q. (1999), Are crop yields normally distributed? *American Journal of Agricultural Economics*, 81: 287-304.
- Kansas Department of Agriculture (2004), *Kansas wheat history*, Statistical Division, November 2004.
- Nelson, C.H. (1990), The influence of distributional assumptions on the calculation of crop insurance premia, *North Central Journal of Agricultural Economics*, 12: 71-78.
- Park, W., I., Sinclair, T.R. (1993), Consequences of climate and crop yield limits on the distribution of corn yields, *Review of Agricultural Economics*, 15: 483-493.
- Ramirez, O.A., Misra, S., Field J. (2003), Crop-yield distributions revisited, *American Journal of Agricultural Economics*, 85: 108-120.
- Rychlikowa, I. (1968), Metody statystyki matematycznej w praktyce badawczej historyka [Methods of mathematical statistics in the research practice of historian], *Kwartalnik Historii Kultury Materialnej*, 16, Instytut Historii Kultury Materialnej, PAN: 317-335.
- Sherrick, B.J., Zanini, F.C., Schnitkey, G.D., Irwin, S.H. (2004), Crop insurance valuation under alternative yield distributions, *American Journal of Agricultural Economics*, 86: 406-419.
- Skees, J.R., Black, J.R., Barnett, B.J. (1997), Designing and rating an area field crop insurance contract, *American Journal of Agricultural Economics*, 79: 430-438.
- Żytkowicz, L. (1970), Plony zbóż w Polsce, Czechach, na Węgrzech i Słowacji w XVI-XVIII w [Cereal crops in Poland, Czech, Hungary and Slovakia throughout the 16th and 18th centuries], *Kwartalnik Historii Kultury Materialnej*, 18, Instytut Historii Kultury Materialnej, PAN: 227-253.
- Paper first received: March 2007
In final form: May 2009

