STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

SP 86
(S 346)

A priori distributions in industry, or conditional and unconditional moments.

J. Hemelrijk

1965

Proceedings of the IXth EOQC Conference, Rotterdam, 6-9 September 1965

A PRIORI DISTRIBUTIONS IN INDUSTRY, OR CONDITIONAL AND UNCONDITIONAL MOMENTS

by

PROF. DR. J. HEMELRIJK

(Mathematical Centre, Amsterdam)

Prof. Hemelrijk studied mathematics at the Amsterdam University, graduated in 1948 and obtained his degree in 1950. After having lectured at the Technological University Delft he was appointed Professor Mathematical Statistics at his Alma Mater.
Professor Hemelrijk is Director of the Institute for the Application of Mathematics of the University, Head of the Department for Mathematical Statistics of the Mathematical Centre in Amsterdam and Editor in Chief of the periodical "Statistica Neerlandica" of the Netherlands Statistical Society.

1. INTRODUCTION

There are many situations in industrial production where it seems reasonable to use a mathematical model containing a priori distributions. A well known example is the production of a sequence of batches, consisting of items which are either "good" or "bad". In setting up a sampling inspection scheme for accepting or rejecting individual batches, one may try to incorporate an a priori distribution of the probability of producing a bad item, into the model. Most sampling inspection schemes are not based on this idea; they are derived from the theory of testing statistical hypotheses and are valid for any a priori distribution and even without assuming the existence of such a distribution at all.

But if one wants to select an optimum scheme, e.g. with regard to costs, one has to introduce an a priori distribution and if and when it is reasonable to do so it is also possible to use a priori information in the sampling inspection scheme itself. Using reliable a priori information can, of course, only increase the efficiency of the inspection.

This means, however, that the a priori distribution under consideration has to be estimated from observations which will usually be inspection-data from the past. It is this problem of estimation which led to the writing of this paper. Its aim is to draw the attention to a set of useful formulae, useful with respect to the problem indicated above and similar problems, but also with respect to a much wider field. The formulae are not new, but they seem to have escaped the attention of many statisticians who could use them to their advantage. They are expounded and explained in the following section and illustrated by means of examples in the third one.

2. CONDITIONAL AND UNCONDITIONAL MOMENTS

Notational convention: random variables are indicated by underlined symbols: $\underline{x}$, $\underline{y}$, $\underline{v}$, $\underline{w}$ etc., and the same symbols, without bar, are used for values assumed: $P(\underline{x} = x)$ will denote the probability that the random variable $\underline{x}$ assumes the value x. The expected value (or mean) of a random variable $\underline{x}$ is denoted by $E\underline{x}$ (or occasionally by $\mu_{\underline{x}}$), its variance by $\sigma^2(\underline{x})$.

Throughout this section the ideas and formulae will, for the sake of convenience, be explained for discrete probability distributions, but it all holds for the continuous case too. The moments and conditional distributions mentioned will all

* Report nr. SP 86 of the Dept. of Mathematical Statistics of the Mathematical Centre, Amsterdam.

be supposed to exist and be finite. Measure-theoretical details are omitted.

Now let $\underline{v}$ and $\underline{w}$ have a joint probability distribution, meaning that for every v and w the joint probability

(1) $$P(\underline{v} = v \text{ and } \underline{w} = w)$$

exists. Then the (unconditional) mean of $\underline{v}$ is given by

(2) $$E\underline{v} = \sum_v v\, P(\underline{v} = v)$$

where the summation is over all values v which $\underline{v}$ can assume. This convention about summation holds throughout this paper. The conditional mean of $\underline{v}$, given $\underline{w}$ = w, is

(3) $$E(\underline{v} \mid \underline{w} = w) = \sum_v v\, P(\underline{v} = v \mid \underline{w} = w),$$

where

(4) $$P(\underline{v} = v \mid \underline{w} = w) = \frac{P(\underline{v} = v \text{ and } \underline{w} = w)}{P(\underline{w} = w)}$$

represents the conditional distribution of $\underline{v}$ under the condition $\underline{w}$ = w. Of course this only has sense for values w with $P(\underline{w} = w) \neq 0$. It is clear from (3) that $E(\underline{v} \mid \underline{w} = w)$ is a function of w and we therefore write

(5) $$\varphi(w) = E(\underline{v} \mid \underline{w} = w).$$

Often this function $\varphi(w)$ can easily be obtained and if we then substitute the random variable $\underline{w}$ into it we get a new random variable $\varphi(\underline{w})$ which now leads easily to the unconditional mean $E\underline{v}$ by means of the relation

(6) $$E\underline{v} = E\ \varphi(\underline{w})$$

This can be seen as follows. The relation between the (marginal) distribution of $\underline{v}$ and the joint distribution (1) of $\underline{v}$ and $\underline{w}$ is given by

(7) $$P(\underline{v} = v) = \sum_w P(\underline{v} = v \text{ and } \underline{w} = w)$$

and from (2) and (7) we have

(8) $$E\underline{v} = \sum_v \sum_w v\, P(\underline{v} = v \text{ and } \underline{w} = w).$$

Further (5), (3) and (4) respectively give

$$\begin{aligned} E\ \varphi(\underline{w}) &= \sum_w \varphi(w)P(\underline{w} = w) = \\ &= \sum_w E(\underline{v} \mid \underline{w} = w)P(\underline{w} = w) = \\ &= \sum_w \sum_v vP(\underline{v} = v \mid \underline{w} = w)P(\underline{w} = w) = \\ &= \sum_v \sum_w vP(\underline{v} = v \text{ and } \underline{w} = w) = E\underline{v} \end{aligned}$$

according to (8). Thia proves (6). Writing

(9) $$E_v(\underline{v} \mid \underline{w})$$

for $\varphi(\underline{w})$, where the subscript v indicates that the expected value is taken over $\underline{v}$ but not over $\underline{w}$, we obtain the first of our formulae:

(10) $$\boxed{E\underline{v} = E_w E_v(\underline{v} \mid \underline{w})}$$

In words: <u>in order to obtain the unconditional expectation of $\underline{v}$ one may first take the conditional expectation of $\underline{v}$, given $\underline{w}$, and then the expectation of the result with respect to the variable $\underline{w}$.</u> The subscripts w and v can be omitted without causing ambiguity, but for clarity's sake we will maintain them, where this seems advisable.

Using this formula others may be obtained for higher moments of $\underline{v}$. For the variance we prove

(11) $$\boxed{\sigma^2(\underline{v}) = E_w \sigma^2_v(\underline{v} \mid \underline{w}) + \sigma^2_w\left\{E_v(\underline{v} \mid \underline{w})\right\}}$$

<u>Proof</u>. Starting from the well known formula

(12) $$\sigma^2(\underline{v}) = E\underline{v}^2 - (E\underline{v})^2$$

we apply (10) to $\underline{v}^2$ instead of $\underline{v}$. This gives

(13) $$E\underline{v}^2 = E_w E_v(\underline{v}^2 \mid \underline{w}).$$

Since (12) holds for every probability distribution (with finite second moment) we can also apply it to the conditional distribution of $\underline{v}$, given w. This gives

(14) $$E_v(\underline{v}^2 \mid w) = \sigma_v^2(\underline{v} \mid w) + \left\{E_v(\underline{v} \mid w)\right\}^2$$

and substituting this into (13) we get

(15) $$E\underline{v}^2 = E_w \sigma_v^2(\underline{v} \mid \underline{w}) + E_w \left\{E_v(\underline{v} \mid \underline{w})\right\}^2.$$

In (12) the last term may, according to (10), also be written as $\left\{E_w E_v(\underline{v} \mid \underline{w})\right\}^2$ and then we obtain, by also substituting (15) into (12):

(16) $$\sigma^2(\underline{v}) = E_w \sigma_v^2(\underline{v} \mid \underline{w}) + \left[E_w \left\{E_v(\underline{v} \mid \underline{w})\right\}^2 - \left\{E_w E_v(\underline{v} \mid \underline{w})\right\}^2\right].$$

Recalling (5) we see that the term between square brackets is nothing but the variance of $\varphi(\underline{w})$ (cf. (12) with $\varphi(\underline{w})$ for $\underline{v}$) and thus this term equals the last term of (11), which completes the proof.
Formula (11) may be expressed in words as follows: in order to find the variance of $\underline{v}$ one may first compute the conditional expectation and variance of $\underline{v}$, given $\underline{w}$, and then add the variance of the former to the expectation of the latter, both with respect to $\underline{w}$.

For the third reduced moment

(17) $$\tilde{\mu}_3(\underline{v}) = E(\underline{v} - E\underline{v})^3$$

the following formula may be obtained in a similar way:

(18) $$\tilde{\mu}_3(\underline{v}) = E_w \tilde{\mu}_{3,v}(\underline{v}\ \ \underline{w}) + \tilde{\mu}_{3,w}\left\{E_v(\underline{v} \mid \underline{w})\right\} + 3\mathrm{cov}_w \left\{E_v(\underline{v} \mid \underline{w}), \sigma_v^2(\underline{v} \mid \underline{w})\right\}$$

The last term in this equation is the covariance of the random variables $E_v(\underline{v} \mid \underline{w})$ and $\sigma_v^2(\underline{v} \mid \underline{w})$, which are both functions of the random variable $\underline{w}$ and thus have a joint distribution.

If we consider three random variables $\underline{u}$, $\underline{v}$ and $\underline{w}$ the covariance of the first pair may be expressed in conditional moments with respect to $\underline{w}$ as follows:

(19) $$\mathrm{cov}\,(\underline{u}, \underline{v}) = E_w \mathrm{cov}_{u,v}(\underline{u}, \underline{v} \mid \underline{w}) + {} + \mathrm{cov}_w \left\{E_u(\underline{u} \mid \underline{w}), E_v(\underline{v} \mid \underline{w})\right\}$$

The proof of (18) and (19) is left to the reader. This proof, though a little more laborious than the above proof of (11), is of exactly the same character. It is easiest to prove (19) first and to use it in the proof of (18). Formula (11) is a special case of (19), i.e. the case where $\underline{u}$ is identically equal to $\underline{v}$.
The examples in the next section are restricted to the formulae (10) and (11).

3. EXAMPLES

Now for the application of these formulae to some examples. As a first example we consider the situation indicated in section 1. The model is as follows. For any batch of items the probability of an item being bad is equal to p, but p is different from batch to batch; as a matter of fact $\underline{p}$ is a random variable with an unknown distribution, the a priori distribution in question. Taking a random sample of size n from a batch the number of bad items, $\underline{x}$, in the sample has a binomial distribution with parameters n and p, where p is the value of $\underline{p}$ realised for this batch. Now applying (10) and (11) with

$$\underline{v} = \underline{x} \quad \text{and} \quad \underline{w} = \underline{p};$$

we find from (10)

$$E\underline{x} = E_p E_x(\underline{x} \mid \underline{p});$$

but, given p, the mean of $\underline{x}$ is np, thus

$$E\underline{x} = E_p(n\underline{p}) = n \,.\, E\underline{p}.$$

In (11) we also meet the conditional variance $\sigma_x^2(\underline{x} \mid p) = np(1 - p)$, thus

$$\sigma^2(\underline{x}) = E_p \left\{n\underline{p}(1-\underline{p})\right\} + \sigma_p^2(n\underline{p}) =$$
$$= n.E\,\underline{p} - n.E\,\underline{p}^2 + n^2.\,\sigma^2(\underline{p}) =$$
$$= n.E\,\underline{p} - n\left\{\sigma^2(\underline{p}) + (E\underline{p})^2\right\} + n^2.\sigma^2(\underline{p}) =$$
$$= n.E\,\underline{p}(1-E\underline{p}) + n(n-1).\sigma^2(\underline{p}).$$

Or, writing μ instead of E:

(20) $\mu_{\underline{x}} = n\,\mu_{\underline{p}}\ ;\quad \sigma^2(\underline{x}) = n\,\mu_{\underline{p}}(1-\mu_{\underline{p}}) + n(n-1)\,\sigma^2(\underline{p}).$

Now $\underline{x}$ can be observed and thus $\mu_{\underline{x}}$ and $\sigma^2(\underline{x})$ can be estimated from a sample of observations of $\underline{x}$ in the usual way by means of the sample mean $\underline{\bar{x}}$ and the sample variance $\underline{s}^2$. But (20) can also be written as

(21) $\mu_{\underline{p}} = \mu_{\underline{x}}/\,n;\ \sigma^2(\underline{p}) = \dfrac{\sigma^2(\underline{x}) - \mu_{\underline{x}}\,(1-\mu_{\underline{x}}/n)}{n(n-1)},$

and, substituting $\underline{\bar{x}}$ for $\mu_{\underline{x}}$ and $\underline{s}^2$ for $\sigma^2(\underline{x})$ we obtain estimators of $\mu_{\underline{p}}$ and $\sigma^2(\underline{p})$. Similarly (18) can be used to estimate the third reduced moment of $\underline{p}$. Note that the unconditional distribution of $\underline{x}$ is not a binomial one, it has an extra term in its variance; note further that all observations of $\underline{x}$ have been assumed to be based on inspection samples of the same size n.

In the case of unequal sample sizes the data are of the following form: k samples of sizes $n_1, \ldots, n_k$ yield observations $\underline{x}_1, \ldots, \underline{x}_k$. With $\underline{X} = \Sigma\underline{x}_i$ and $N = \Sigma n_i$ we find by means of (10)

(22) $E\,\underline{X} = N.E\,\underline{p},$

which means that $\underline{X}/N$ can be used as estimator for $\mu_{\underline{p}}$. To estimate the variance of $\underline{p}$ we now apply (10) to $\underline{x}_i^2$, which gives

(23) $E\,\underline{x}_i^2 = n_i.E\,\underline{p} + n_i(n_i-1)E\,\underline{p}^2.$

Summation over i gives, after some rearrangement

(24) $E\,\underline{p}^2 = E\,(\Sigma\underline{x}_i^2 - \underline{X})\,/\Sigma n_i(n_i-1),$

which means that $(\Sigma\underline{x}_i^2 - \underline{X})\,/\Sigma n_i(n_i-1)$ can be used as estimator of $E\,\underline{p}^2$.

According to (12) this leads to

(25) $\dfrac{\Sigma\underline{x}_i^2 - \underline{X}}{\Sigma n_i(n_i-1)} - \dfrac{\underline{X}^2}{N^2}$

as estimator for $\sigma^2(\underline{p})$. For equal samples ($n_i = n$, $N = kn$) this reduces to the former case if $\underline{s}^2$ is defined as $k^{-1}\,\Sigma(\underline{x}_i - \underline{\bar{x}})^2$.

A modification of this example is obtained if we do not consider the probability of a defective item as the parameter with an a priori distribution, but the number of defective items in a batch. If the size of a batch is denoted by N, the number of bad items it contains by $\underline{A}$ and the number of bad items in a sample without replacement by $\underline{a}$, then the conditional distribution of $\underline{a}$, given A, is a hypergeometric distribution and this leads to the following result, corresponding to (20):

(20[a]) $\mu_{\underline{a}} = \dfrac{n}{N}\mu_{\underline{A}}$

$$\sigma^2(\underline{a}) = \frac{n(N-n)}{N^2(N-1)}\mu_{\underline{A}}\,(N-\mu_{\underline{A}}) + \frac{n\,(n-1)}{N\,(n-1)}\,\sigma^2(\underline{A}).$$

The proof of this is completely analogous to the proof of (20). The elaboration of the case when the samples are of unequal size is left to the reader. It may be pointed out, however, that in this model the batches must be of equal size N, for otherwise there is no sense to one a priori distribution of $\underline{A}$ for all batches considered; in the first model there is no need for this, the batch-size does not matter at all, provided p may be assumed to be constant during the production of every batch separately.

As a second example, which has nothing to do with a priori distributions but which is meant to show the wider scope of the method, we consider the sum of a random number of random variables:

$$\underline{X} = \underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_{\underline{n}}, \tag{26}$$

where, given n, the terms $\underline{x}_1, \ldots, \underline{x}_n$ form a random sample of size $\underline{n}$. Let the mean and variance of the $\underline{x}_i$ and $\underline{n}$ respectively be denoted by $\mu_{\underline{x}}$ and $\sigma^2(\underline{x})$ and by $\mu_{\underline{n}}$ and $\sigma^2(\underline{n})$ respectively, then we want to find the mean and the variance of $\underline{X}$. This situation occurs in decision problems: if $\underline{n}$ is, e.g. a random number of customers buying quantities $\underline{x}_1, \underline{x}_2, \ldots$ of a certain commodity, then $\underline{X}$ is the total quantity sold (if the stock holds out); $\underline{n}$ may also be the number of claims for damage paid during a year by an insurance company and $\underline{x}_1, \underline{x}_2, \ldots$ the amounts of the claims, thus $\underline{X}$ being the total amount paid.

We now take $\underline{w} = \underline{n}$ and $\underline{v} = \underline{X}$ in (10) and (11). This gives

$$\mu_{\underline{X}} = E\underline{X} = E_n\, E_X(\underline{X} \mid \underline{n}) = E_n\, (\underline{n}\mu_{\underline{x}}) = \mu_{\underline{n}} \cdot \mu_{\underline{x}}$$

and

$$\sigma^2_X(\underline{X} \mid \underline{n}) = \underline{n} \cdot \sigma^2(\underline{x}),$$

thus

$$\sigma^2(\underline{X}) = E\, \underline{n} \cdot \sigma^2(\underline{x}) + \sigma^2_n\, (\underline{n}\mu_{\underline{x}}) =$$
$$= \mu_{\underline{n}} \cdot \sigma^2(\underline{x}) + \mu^2_{\underline{x}} \cdot \sigma^2(\underline{n}).$$

The result is

$$\mu_{\underline{X}} = \mu_{\underline{n}} \cdot \mu_{\underline{x}} \quad \text{and} \quad \sigma^2(\underline{X}) = \mu_{\underline{n}} \cdot \sigma^2(\underline{x}) + \mu^2_{\underline{x}} \cdot \sigma^2(\underline{n}). \tag{27}$$

This means that the mean of $\underline{X}$ is just the product of the means of $\underline{n}$ and $\underline{x}$, but that the variance of $\underline{X}$ consists of two terms: the first is of the same structure as when n is fixed but the second does not occur in that case and this term may be very large. To see this better we express (27) in coefficients of variation V. This gives

$$V^2_{\underline{X}} = V^2_{\underline{x}} / \mu_{\underline{n}} - V^2_{\underline{n}}, \tag{28}$$

where the second, additional, term in the right hand member has denominator 1 and may, therefore, by far be the most important of the two.

As our last example we consider the effect of observational errors which are proportional to the value to be observed. To be more precise let $\underline{x}$ be a positive random variable of which we want to estimate the moments. This variable cannot be observed without error; the error $\underline{w}$ has zero expectation but its standard deviation is proportional to the value x to be observed, i.e. for every $x > 0$

$$E\,(\underline{w} \mid x) = 0\ ; \quad \sigma\,(\underline{w} \mid x) = cx. \tag{28}$$

An observation $\underline{y}$ has the form

$$\underline{y} = \underline{x} + \underline{w} \tag{29}$$

and we can now apply (10) and (11). This gives

$$E_y\,(\underline{y} \mid x) = E_w\,(x + \underline{w} \mid x) = x,$$

thus

$$E\underline{y} = E_x\, E_y\, (\underline{y} \mid \underline{x}) = E\underline{x};$$

and

$$\sigma^2_y\, (\underline{y} \mid x) = \sigma^2(\underline{w} \mid x) = c^2x^2,$$

thus

$$\sigma^2(\underline{y}) = E_x(c^2\underline{x}^2) + \sigma^2_x(\underline{x}) =$$
$$= c^2 \left\{\sigma^2(\underline{x}) + (E\underline{x})^2\right\} + \sigma^2(\underline{x}) =$$
$$= (1 + c^2)\,\sigma^2\,(\underline{x}) + c^2(E\underline{x})^2.$$

Together, with μ instead of E:

$$\mu_{\underline{y}} = \mu_{\underline{x}} \; ; \; \sigma^2(\underline{y}) = (1 + c^2)\,\sigma^2(\underline{x}) + c^2 \mu_{\underline{x}}^2 \; . \tag{30}$$

If a sample of observations of $\underline{y}$ is given, the parameters $\mu_{\underline{y}}$ and $\sigma^2(\underline{y})$ can be estimated, but this does not lead to an estimator of $\mu_{\underline{x}}$ and $\sigma^2(\underline{x})$ unless c is known. If this is not the case one would e.g. need duplicate observations, with respect to the same value of $\underline{x}$, in order to estimate c too.

4. REMARKS

Many more examples of the method in different fields of application can be given. One field which seems especially appropriate is the theory of sampling; (10), (11) and (19) can e.g. be found in the book of M.H. HANSEN, W.H. HURWITZ and W.G.MADOW (1953), "Sampling Survey Methods and Theory", part II, with applications to stratified sampling etc. The author has not tried to find out when and where the formulae were published first; (10) is fundamental in probability theory and is therefore found in many textbooks, but (11), (18) and (19) are usually not mentioned. The author met (11) for the first time in the university lectures of D. VAN DANTZIG, given in 1947.