A GENERALIZED ERROR FUNCTION*

By

ALBERT WERTHEIMER

I. INTRODUCTION

Given a set of observed values l_i ($i = 1, 2, 3, \ldots, n$.) obtained from n observations assumed to be made on the same quantity, l, under the same conditions. We seek to determine two functions $f(P, l_i)$ and $\phi(P, l_i)$ such that

$$f(P, l_i) = 0, \qquad (i = 1, 2, 3, \ldots, n)$$

defines p as a unique value assigned to the observed quantity; and $\phi(P, l_i)\,dl_i$ gives to within infinitesimals of higher order the probability that if another observation is made, the observed value will lie in the interval

$$l_i \leq l \leq l_i + dl_i.$$

Gauss determined the ϕ function to be the so-called normal error law namely,

$$\phi(P, l_i) = c\,e^{-h^2(P - l_i)^2}$$

on the basis of the following assumptions.

(a) The product $\prod_i \phi(P, l_i)$ is to be a maximum with respect to p. Thus

$$\sum_i \frac{\partial}{\partial p} \log \phi(P, l_i) = 0,$$

$$\sum_i \frac{\partial^2}{\partial p^2} \log \phi(P, l_i) < 0.$$

---

*Presented to the American Mathematical Society, December 28, 1931.

(b) The unique value p is the arithmetic mean of the observations. Thus

$$f(P, \ell_i) = \sum_i (P - \ell_i).$$

(c) The probability function is a function of $(P - \ell_i)$. Thus

$$\phi(P, \ell_i) \equiv \phi(P - \ell_i).$$

Poincaré[1] on the basis of the first two assumptions only obtained the error function

$$\phi(P, \ell_i) = \theta(\ell_i) e^{W(P) + \ell_i V(P)},$$

where

$$\frac{dW}{dP} + p\frac{dV}{dP} = 0.$$

In this paper we assume the unique value p to be defined by a function satisfying certain conditions, and obtain on the basis of assumption (a) a more general error function from which the so-called normal error law, the Poincare function, and other forms of the error function as well as the Pearson curves are obtained as special cases.

2. *The unique value p.*

We now make the following assumptions:

I: The unique value p is defined explicitly as a function of the observed values in the region $a \le \ell_i \le b$. Thus

$$P - F(\ell_1, \ell_2, \ell_3, \cdots, \ell_n) = 0,$$

where F is single valued, continuous with continuous derivatives up to the second order.

II: The value of p is independent of the order in which the observations are obtained. Thus F is a symmetric function.

III: The change in p due to a change in one of the observed values, say ℓ_i, is a function of p and ℓ_i only. Thus

$$F_{\ell_i} = F_{\ell_i}(F, \ell_i).$$

[1] H. Poincare, *Calcul des Probabilites* (1912), p. 171.

IV: If p is regarded as a function of a single variable, say t_i, while all the others are regarded as constants, then with respect to this variable p is a monotonic function and is not constant in any portion of the interval in which it is defined. Thus

$$F_{t_i} \neq 0$$

for all i's.

We have then for the determination of the ϕ function the two equations

$$\sum_i \frac{\partial}{\partial P} \log \phi(P, t_i) = 0, \tag{1}$$

$$P - F(t_1, t_2, t_3, \cdots, t_n) = 0, \tag{2}$$

which must be simultaneously satisfied for any set of values in the region defined.

3. *The g function.*

We will now show by means of the following theorems that if F satisfies the given conditions, then there exists a unique function $g(P, t_i)$ such that the equation

$$\sum_i g(P, t_i) = 0,$$

is identical with equation (2).

THEOREM I.

Given a function of n variables,

$$F(x^1, x^2, x^3, \cdots, x^n),$$

continuous with continuous non-vanishing first derivatives in the region defined, such that

$$F_{x^i} = \psi^i(F, x^i);$$

then $\psi^i(F, x^i)$ must be in the form of a product of a function of F and a function of t_i. Thus

$$\psi^i(F, x^i) = \omega(F)\beta^i(x^i).$$

Proof:
We have

$$F_{x^i x^j} = \psi^i_F(F, x^i)\psi^j(F, x^j),$$

and

$$F_{x^j x^i} = \psi^j_F(F, x^i)\psi^i(F, x^i).$$

Hence:

$$\frac{\psi^i_F(F, x^i)}{\psi^i(F, x^i)} = \frac{\psi^j_F(F, x^j)}{\psi^j(F, x^j)} = \ldots = \frac{\psi^n_F(F, x^n)}{\psi^n(F, x^n)} = \eta(F).$$

Integrating, we get

$$\log \psi^i(F, x^i) = \int \eta(F)\,dF + \xi^i(x^i).$$

from which it follows that

$$\psi^i(F, x^i) = \omega(F)\beta^i(x^i).$$

THEOREM II.

Given a function of n variables,

$$F(x^1, x^2, x^3, \ldots, x^n),$$

continuous with continuous non-vanishing first derivatives in the region defined, then in order that there shall exist a unique function $\mathcal{E}(F)$ such that

$$\xi(F) \equiv \sum_i u^i(x^i),$$

it is necessary and sufficient that

$$\frac{F_{x^i}}{F_{x^j}} = \alpha^i(x^i)\,\alpha^j(x^j).$$

Proof:

Necessary conditions:—

If the ξ function exists then the functional matrix

$$\left\| \begin{matrix} u^1_{x^1} & u^2_{x^2} & u^3_{x^3} & \cdots & u^n_{x^n} \\ F_{x^1} & F_{x^2} & F_{x^3} & \cdots & F_{x^n} \end{matrix} \right\|$$

must be of rank one. Hence

$$\frac{F_{x^i}}{F_{x^j}} = \frac{u^i_{x^i}}{u^j_{x^j}} = \alpha^i(x^i)\,\alpha^j(x^j).$$

Sufficient conditions:—

We assume that

$$\frac{F_{x^i}}{F_{x^j}} = \alpha^i(x^i)\,\alpha^j(x^j).$$

Then we have the following identities:

a) $$\frac{\partial^2}{\partial x^i\,\partial x^j} \log \frac{F_i}{F_j} = 0,$$

b) $$\frac{F_i}{F_j} = \frac{F_{ik}}{F_{jk}} = \frac{F_{ik\ell}}{F_{jk\ell}},$$

for $k, \neq i$ or j and

c) $$F_{ij}\left\{F_{ki}\,F_i - F_{ii}\,F_k\right\} = F_i\left\{F_{ijk}\,F_i - F_{iij}\,F_k\right\},$$

where for convenience of notation,

$$F_i \equiv F_{x^i}, \quad F_{ik} = F_{x^i x^k}, \qquad \text{etc.}$$

Making use of a), b), and c), it is easily shown that the functional matrix

$$\left\| \begin{matrix} \frac{\partial}{\partial x^1}\left(\frac{F_{ij}}{F_i F_j}\right) & \frac{\partial}{\partial x^2}\left(\frac{F_{ij}}{F_i F_j}\right) & \cdots & \frac{\partial}{\partial x^n}\left(\frac{F_{ij}}{F_i F_j}\right) \\ F_1 & F_2 & \cdots & F_n \end{matrix} \right\|$$

is of rank one. It follows that

d) $$\frac{F_{ij}}{F_i F_j} = \lambda(F).$$

Now the differential equation that defines the ξ function is

$$\xi_{ij} \equiv \xi_{FF} F_i F_j + \xi_F F_{ij} = 0,$$

or

$$\frac{\xi_{FF}}{\xi_F} = -\frac{F_{ij}}{F_i F_j}, = -\lambda(F) \text{ from d).}$$

Hence $\xi(F)$ is uniquely determined, namely,

$$\xi(F) = K \int e^{-\int \lambda(F) dF} dF + H.$$

where K and H are constants of integration.

Now, for our problem, if F satisfies the given conditions, we can apply the two theorems in succession and we have that there exists a unique function

$$\xi(F) \equiv u^i(\ell_i).$$

But due to the symmetry of F all the u^i functions will be the same and we have

$$\xi(F) \equiv u(\ell_i).$$

If we now define

$$\varsigma(p, \ell_i) \equiv \frac{1}{n}\xi(p) - u(\ell_i),$$

we have

$$\sum_i g(p, t_i) = \xi(p) - \xi(F) = 0.$$

4. *General Error Function*

We may now write equations (1) and (2) in the form respectively

$$\sum_i \frac{\partial}{\partial p} \log \Phi(p, t_i) = 0,$$

$$\sum_i g(p, t_i) = 0.$$

These equations must be simultaneously satisfied for an arbitrary set of values t_i in the region defined. It follows that they are identical. Thus

$$\frac{\partial}{\partial p} \log \Phi(p, t_i) = \psi(p) g(p, t_i),$$

where $\psi(p)$ is an arbitrary function.
Integrating, we get

$$(3) \qquad \Phi(p, t_i) = \theta(t_i) e^{\int \psi(p) g(p, t_i) dp}$$

where $\theta(t_i)$ is an arbitrary function. This is our general error function. In order to insure a maximum we must have

$$(4) \qquad \psi(p) g_p \neq 0.$$

5. *A Generalized Normal Function*

If we now make the additional assumption that

$$\Phi(p, t_i) \equiv \Phi\{g(p, t_i)\},$$

we have

$$\begin{vmatrix} \Phi_p & \Phi_{t_i} \\ g_p & g_{t_i} \end{vmatrix} = 0$$

Expanding and simplifying, we get

$$\frac{\theta_{\ell_i}}{\theta g_{\ell_i}} = \frac{\psi(p)g(p,\ell_i)}{g_p} - \int \psi(p)dp.$$

Differentiating with respect to ℓ_i, we get

$$\frac{1}{g_{\ell_i}} \frac{\partial}{\partial \ell_i} \left(\frac{\theta_{\ell_i}}{\theta g_{\ell_i}}\right) = \frac{\psi(p)}{g_p} = k.$$

Integrating and substituting in (3), we get

$$\varphi(g) = c e^{kg^2}$$

From (4) we have

$$Kg_p^2 \ngtr 0,$$

Hence

$$\varphi(g) \doteq c e^{-h^2 g^2}. \tag{5}$$

We shall refer to this function as the "Generalized Normal Error Function".

5. *Application to Special Cases*

If p is defined as the arithmetic mean, then the region considered is $-\infty < \ell_i < +\infty$, and

$$g(p, \ell_i) = p - \ell_i.$$

The normal law is obtained directly from (5), and from (4) we have

$$\varphi(p, \ell_i) = \theta(\ell_i) e^{\int \psi(p)(p - \ell_i) dp}$$
$$= \beta(\ell_i) e^{w(p) + \ell_i v(p)},$$

where

$$\frac{dW}{dp} + p\frac{dV}{dp} = 0,$$

which is the same as the Poincare′ function.

For the geometric mean, the region considered is

$$0 < l_i < \infty$$

and

$$g(p, l_i) = \log p - \log l_i.$$

Hence, from (3)

$$\Phi(p, l_i) = \theta(l_i) e^{\int \psi(p)\{\log p - \log l_i\} dp},$$

and from (4)

$$\Phi\{\log p - \log l_i\} = ce^{-h^2\{\log p - \log l_i\}^2}$$

The Geometric mean as the most probable value, as well as its generalized normal curve are used for certain astronomical photometric measurements.[1]

For the harmonic mean, the region considered is

$$(0 < l_i < \infty)$$

and

$$g(p, l_i) = \left(\frac{1}{p} - \frac{1}{l_i}\right).$$

Then from (3), we have

$$\Phi(p, l_i) = \theta(l_i) e^{\int \psi(p)\{\frac{1}{p} - \frac{1}{l_i}\} dp}$$

and from (4), we have

$$\Phi\left\{\frac{1}{p} - \frac{1}{l_i}\right\} = ce^{-h^2\{\frac{1}{p} - \frac{1}{l_i}\}^2}.$$

7. *Remarks About the Generalized Normal Curves*

Let us consider briefly some characteristics of the generalized normal curves corresponding to the following three special cases.

[1] Whittaker & Robinson, Calculus of observations (1924), p. 218.

(a) *Arithmetic mean:* Here

$$\Phi(p,\ell) = ce^{-h^2(p-\ell)^2}$$

From this equation we see that

$$\Phi(p, p+\epsilon^2) = \Phi(p, p-\epsilon^2),$$
$$\Phi(p, \ell) = \Phi(p+\epsilon^2, \ell+\epsilon^2),$$
$$\Phi(p, 0) = ce^{-h^2p^2},$$
$$\Phi(p, \infty) = 0.$$

(b) *Harmonic mean:* In this case

$$\Phi(p,\ell) = ce^{-h^2\{\frac{1}{p}-\frac{1}{\ell_i}\}^2},$$

from which we see that

$$\Phi(p, p-\epsilon^2) < \Phi(p, p+\epsilon^2)$$
$$\Phi(p, \ell) < \Phi(p+\epsilon^2, \ell+\epsilon^2),$$
$$\Phi(p, 0) = 0,$$
$$\Phi(p, \infty) = ce^{-\frac{h^2}{p^2}}$$

(c) *Geometric Mean:* Here

$$\Phi(p,\ell) = ce^{-h^2\{\log p - \log \ell_i\}^2}$$

and

$$\Phi(p, p-\epsilon^2) < \Phi(p, p+\epsilon^2),$$
$$\Phi(p, \ell) < \Phi(p+\epsilon^2, \ell+\epsilon^2),$$
$$\Phi(p, 0) = 0,$$
$$\Phi(p, \infty) = 0,$$

Instead of treating these normal functions as three distinct error laws referred to the same measuring scale, we can regard them as a single error law with reference to three different measuring scales (see sketch). This viewpoint helps to explain the above mentioned characteristics of these laws.

The law for the arithmetic mean applies when an object is measured with a uniformly graduated scale in p. The characteristics for this law follow directly from the consideration that the scale is everywhere the same.

The law for the harmonic mean holds when an object is measured with a reciprocally graduated scale, as for instance measuring the volume of a gas with a pressure gauge graduated for volume. In this case the scale becomes crowded as p increases, and hence

$$\phi(p+\epsilon^2, t+\epsilon^2) > \phi(p,t),$$

and also

$$\phi(p, p+\epsilon^2) > \phi(p, p-\epsilon^2).$$

For large values of p it would take only a small error in the reading of the scale to make an infinitely large error in the value of p and hence $\phi(p,\infty)$ does not necessarily vanish. On the other hand the zero point is at an infinite distance and hence $\phi(p,0)=0$.

The law for the geometric mean holds for measuring objects with a logarithmically graduated scale. The same remarks as for the harmonic mean apply here, except that in this case it would take an infinitely large error in the reading of the scale to make an infinitely large error in the value of p. Hence $\phi(p,\infty)=0$.

8. *The Pearson Curves*

Leaving out the subscripts in (3), we have for the general error function

$$\phi(p,t) = \theta(t)e^{\int \psi(p) g(p,t)\, dp}$$

Remembering that

$$g(p,\ell) \equiv \frac{1}{n}\xi(p) - u(\ell),$$

we have

$$\frac{\partial\varphi}{\partial\ell} = \varphi\left\{\frac{\theta_\ell}{\theta} + u(\ell)\int\psi(p)\,dp\right\}.$$

Thus for a given p the curve approaches the ℓ axis asymptotically.

Let us now impose the condition that

$$\left.\frac{\partial\varphi}{\partial\ell}\right|_{\ell=p} = 0,$$

then

$$\frac{\theta_\ell}{\theta} = u(\ell)\int\psi(\ell)\,d\ell.$$

Integrating, we get

$$\theta(\ell) = ce^{-\int u(\ell)\{\int\psi(\ell)d\ell\}d\ell}$$

so that

(6) $$\varphi(p,\ell) = ce^{-\int u(\ell)\{\int\psi(\ell)d\ell\}d\ell + \int\psi(p)g(p,\ell)dp}$$

and

$$\frac{\partial\varphi}{\partial\ell} = \varphi u_\ell \int_\ell^p \psi(t)\,dt,$$

where t is a variable of integration. If we now take as a special case

$$\psi(t) = 1,$$
$$u_\ell = \{b_0 + b_1\ell + b_2\ell^2 + \dots\}^{-1},$$

we have

$$\frac{\partial \varphi}{\partial t} = \varphi \frac{p - t}{b_0 + b_1 t + b_2 t^2},$$

which is the differential equation defining the Pearson system of frequency curves. For this case, (6) reduces to

$$\varphi(p, t) = c e^{\int \frac{p - t}{b_0 + b_1 t + b_2 t^2} dt + \frac{1}{n} \int \xi(p) dp}$$

from which we see that by a proper choice of ξ (p) we can choose the value of p for which the product $\prod_i \varphi(p, t_i)$ shall be a maximum.

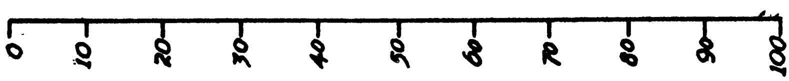
It may be noted that the differential equation defining the Pearson curves is often derived on the basis of the assumptions that the curve shall approach the t axis asymptotically, and have only one maximum point.

In conclusion, it appears that if we restrict the function that defines to satisfy the assumptions given in this paper, and also impose the condition that p shall be the most probable value in the sense, that the product

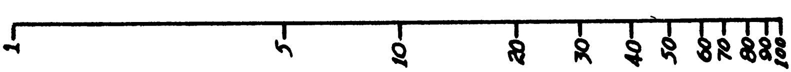
$$\prod_i \varphi(p, t_i)$$

shall be a maximum with respect to p, then (3) is the most general form of the error function.

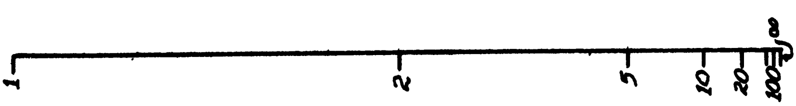
Scales, corresponding to the arithmetic, geometric, and harmonic means.



Arithmetic Mean — Uniform Scale



Geometric Mean — Logarithmic Scale



Harmonic Mean — Reciprocal Scale

Albert Wertheimer