RANGES AND MIDRANGES[1]

By E. J. Gumbel

New School for Social Research, New York City

1. Introduction. In the following the generating functions of the extremes are studied in order to determine the nature of the distributions of the ranges and the midranges.

A large sample of size n is considered to be drawn from an unlimited symmetrical continuous distribution with zero mean. The difference and the sum of the largest and of the smallest observation, the extremes, are called *range* and *midrange*. R. A. Fisher and L. H. C. Tippett [2] have established the limiting distributions of the largest and of the smallest member of a sample. The exact conditions under which these distributions hold have been given by R. von Mises [4]. For a normal distribution, L. H. C. Tippett [7] has calculated the numerical values of the mean range and the first four moments of the range for sample sizes varying from 2 to 1000. He has shown that, for sample sizes exceeding 200, the correlation between the largest and the smallest observation may be neglected. Later, E. S. Pearson [5] has calculated the probability function of the range for small samples ($n = 2$ to 20) taken from a normal population. These calculations are very laborious. Recently, W. E. Deming [1] has applied the range to quality control.

The concepts "extremes", "range" and "midrange" allow a simple generalization. Let ${}_mx$ and x_m be the mth observation in increasing and in decreasing magnitude, henceforth called mth value "from below" and "from above". As long as the index m is small compared to the sample size n, the mth values under consideration are extremes. The difference and the sum of the mth extreme observations are called the mth *range* and the mth *midrange*. We will investigate the asymptotic distributions of the mth extremes, of the mth range, and of the mth midrange. Assuming that the number of observations is very large, the correlation between the largest and the smallest observation may be neglected. Then the mth range and the mth midrange are the difference and the sum of two independent variates, the mth extremes.

It was found that the distribution of the mth range is skew and the distribution of the mth midrange is of the generalized logistic type, which is symmetrical. For m increasing the distributions of the mth extremes, the mth ranges, and the mth midranges converge toward normality.

2. Generating functions of the mth extremes. Let $\varphi(x)$ be an initial continuous symmetrical distribution with mean zero; let u_m be the most probable mth value from above; let α_m be defined by

$$\alpha_m = \frac{n}{m}\,\varphi(u_m). \tag{1}$$

[1] Research done with the support of a grant from the American Philosophical Society.

Under conditions given in an earlier paper [3] the distributions $f_m(x_m)$ and ${}_mf({}_mx)$ of the mth extremes are, for large n and small m

$$(2)\qquad f_m(x_m) = \alpha_m \frac{m^m}{(m-1)!} e^{-my_m - me^{-y_m}}; \qquad {}_mf({}_mx) = f_m(-x_m)$$

where

$$(3)\qquad y_m = \alpha_m(x_m - u_m)$$

is a reduced extreme mth value.

The moment generating function $G_m(t)$ of the mth extreme value from above obtained from the preceding equations is

$$G_m(t) = \frac{m^m}{(m-1)!} e^{u_m t} \int_{-\infty}^{+\infty} e^{-my_m + y_m(t/\alpha_m) - me^{-y_m}} dy_m .$$

Introducing

$$e^{-y_m} = z$$

the integral becomes

$$\int_0^{\infty} z^{m-(t/\alpha_m)-1} e^{-mz} dz = \Gamma\left(m - \frac{t}{\alpha_m}\right) m^{-m+(t/\alpha_m)}$$

whence

$$(4)\qquad G_m(t) = e^{u_m t} m^{t/\alpha_m} \Gamma\left(m - \frac{t}{\alpha_m}\right) \Big/ \Gamma(m).$$

To obtain the moments of the mth extreme value from above, the fundamental property of the Gamma function is used; whence

$$G_m(t) = e^{(t/\alpha_m)(u_m\alpha_m + lgm)} \prod_{\nu=1}^{m-1} \frac{m - \nu - \dfrac{t}{\alpha_m}}{m - \nu} \Gamma\left(1 - \frac{t}{\alpha_m}\right).$$

Finally, by reversing the order of multiplication the moment generating function of the mth extreme value from above becomes

$$(5)\qquad G_m(t) = e^{(t/\alpha_m)(u_m\alpha_m + lgm)} \prod_{\nu=1}^{m-1} \left(1 - \frac{t}{\nu\alpha_m}\right) \Gamma\left(1 - \frac{t}{\alpha_m}\right).$$

The mean $\bar{x}_m$ of the mth value from above obtained from (5) is

$$(6)\qquad \bar{x}_m = u_m + \frac{1}{\alpha_m}\left(lgm - \sum_{\nu=1}^{m-1} \frac{1}{\nu} + c\right)$$

where $c = .57722$ is Euler's constant.

The seminvariant generating function $L_m(t)$ of the mth extreme value from above becomes from (5) and (6)

$$L_m(t) = \frac{t}{\alpha_m}\left(\sum_{\nu=1}^{m-1} \frac{1}{\nu} - c\right) + \sum_{\nu=1}^{m-1} lg\left(1 - \frac{t}{\nu\alpha_m}\right) + lg\Gamma\left(1 - \frac{t}{\alpha_m}\right)$$

and after expansion

$$L_m(t) = \sum_2^\infty \frac{t^\nu}{\nu\alpha_m^\nu} S_{\nu,m} \tag{7}$$

where the sums

$$S_{\nu,m} = \sum_{k=m}^\infty \frac{1}{k^\nu}; \qquad \nu \geqq 2 \tag{8}$$

are obtained from the sums

$$S_\nu = \sum_{k=1}^\infty \frac{1}{k^\nu} \tag{8'}$$

which are known from the theory of the Gamma function. The numerical values of the seminvariants of the mth extreme value from above $\lambda_{\nu,m}$, being the coefficients of $t^\nu/\nu!$ in the expansion (7) may be calculated from a table of the sums $S_{\nu,m}$ given in an earlier paper [3].

From the generating functions (4) and (7) of the mth extreme value from above the moment generating function ${}_mG(t)$ and the seminvariant generating function ${}_mL(t)$ of the mth extreme value from below are obtained by the symmetric relation (2) as

$${}_mG(t) = G_m(-t); \quad {}_mL(t) = L_m(-t) \tag{9}$$

and the mean ${}_m\bar{x}$ of the mth extreme value from below is

$${}_m\bar{x} = -\bar{x}_m . \tag{6'}$$

The seminvariants $\lambda_{m,\nu}$ and ${}_m\lambda_\nu$ of the mth extreme value from above and from below are linked by

$$\alpha_m^\nu \lambda_{m,\nu} = (\nu - 1)! S_{\nu,m} = (-1)^\nu \alpha_m^\nu {}_m\lambda_\nu . \tag{10}$$

The standard errors σ_m and ${}_m\sigma$ of the mth extreme values are

$$\alpha_m \sigma_m = \sqrt{S_{2,m}} = \alpha_m {}_m\sigma. \tag{11}$$

This procedure for obtaining the moments of the mth extremes from their distribution (2) parallels closely that used by R. A. Fisher and L. C. H. Tippett [2] for obtaining the moments of the largest and smallest value. The special case $m = 1$ of the formulae (4), (9), (6), (6′), (11), and (10) leads to the moment generating functions, the means, the standard errors, and the higher seminvariants of the largest and of the smallest value given by these authors.

The two parameters u_m and α_m which exist in the distribution (2) of the mth extremes may be calculated from the observations by virtue of equations (6) and (11). Thus, the theoretical distributions (2) may be compared to observations even if the initial distribution $\varphi(x)$ is unknown. For increasing m, the distributions of the mth extremes were shown [3] to converge toward normal distributions, their means and standard deviations being given by (6), (6′) and 11.

3. Generating functions of the *m*th range and the *m*th midrange. To obtain the characteristics for the mth range and the mth mid-range we state first some general properties of the sum and of the difference of two independent variates x and y, with means $\bar{x}$ and $\bar{y}$ and standard deviations σ_x and σ_y. Let the distributions be $\varphi(x)$ and $\psi(y)$, and let the generating functions of the seminvariants $\lambda_{x,\nu}$ and $\lambda_{y,\nu}$ be $L_x(t)$ and $L_y(t)$. We write

$$v = x + y; \qquad w = x - y \tag{12}$$

from the sum and the difference of the two variates. Then the means $\bar{v}$ and $\bar{w}$ and the variances σ_v^2 and σ_w^2 are

$$\bar{v} = \bar{x} + \bar{y}; \quad \bar{w} = \bar{x} - \bar{y}; \quad \sigma_v^2 = \sigma_x^2 + \sigma_y^2 = \sigma_w^2 \tag{13}$$

and the seminvariant generating functions $L_v(t)$ and $L_w(t)$ are

$$L_v(t) = L_x(t) + L_y(t); \quad L_w(t) = L_x(t) + L_y(-t). \tag{14}$$

The negative sign in the second equation (14) is obtained through the same well-known derivation as used for the first equation (14). Therefore, the even seminvariants of the sum are equal to the even seminvariants of the difference, whereas the odd seminvariants of the sum and of the difference are

$$\lambda_{v,2\nu+1} = \lambda_{x,2\nu+1} + \lambda_{y,2\nu+1}\,; \qquad \lambda_{w,2\nu+1} = \lambda_{x,2\nu+1} + (-1)^{2\nu+1}\lambda_{y,2\nu+1}\,.$$

If the distributions $\varphi(x)$ and $\psi(y)$ are symmetrical one to another in the sense

$$\psi(y) = \varphi(-x) \tag{15}$$

then the even seminvariants of the two variates x and y coincide in size and sign and the odd seminvariants coincide in size and differ in sign. Under the condition (15), the even seminvariants of the sum v and the even seminvariants of the difference w are twice the even seminvariants of the variates x or y. The odd seminvariants of the sum v are twice the odd seminvariants of the variate x, and the odd seminvariants of the difference w vanish.

We apply these properties to the mth extreme values and write x_m for x and ${}_mx$ for y. According to (2) the distribution of the mth extreme from above is symmetrical to the distribution of the mth extreme from below in the sense (15).

The mth range w_m and the mth mid-range v_m are defined by

$$w_m = x_m - {}_mx; \qquad v_m = x_m + {}_mx. \tag{12'}$$

The mean $\bar{w}_m$ of the mth range, the mean $\bar{v}_m$ of the mth mid-range and the respective variances $\sigma_{v_m}^2$ and $\sigma_{w_m}^2$ are, from (6), (6′) and (11)

$$\bar{w}_m = 2\bar{x}_m\,; \qquad \bar{v}_m = 0; \qquad \sigma_{v_m}^2 = 2\sigma_m^2 = 2{}_m\sigma^2 = \frac{2S_{2,m}}{\alpha_m^2} = \sigma_{w_m}^2\,. \tag{13'}$$

The seminvariant generating functions $L_w(t, m)$ of the mth range and $L_v(t, m)$ of the mth mid-range obtained from (7) and (9) are

$$L_w(t, m) = 2\sum_{\nu=2}^{\infty} \frac{t^\nu}{\nu\alpha_m^\nu}\, S_{\nu,m} = 2L_m(t); \quad L_v(t, m) = 2\sum_{\nu=1}^{\infty} \frac{t^{2\nu}S_{2\nu,m}}{2\nu\alpha_m^{2\nu}}\,. \tag{14'}$$

The seminvariants of the mth range are twice the seminvariants of the mth extreme values from above. The even seminvariants of the mth mid-range and the even seminvariants of the mth range are twice the even seminvariants of the mth extremes. The odd seminvariants of the mth mid-range vanish. Therefore, the distribution of the mth range is skew, and the distribution of the mth midrange is symmetrical. From the convergence of the mth extremes towards normality it follows that the mth range and the mth mid-range tend also, for increasing indices m, toward normality.

The seminvariants of the range and the mid-range are obtained from (13′) and (14′) by putting $m = 1$ and omitting the index 1. Therefore, the standard errors σ_w and σ_v of the range and of the mid-range are

$$(13'') \qquad \alpha\sigma_w = \frac{\pi}{\sqrt{3}} = \sqrt{2}\,\sigma_n = \alpha\sigma_v$$

where α is given by (1) and σ_n stands for the standard error of the first or the last value. The skewness $\beta_{1,w}$ of the range and the excess $\beta_2 - 3$ for the range and the mid-range

$$(16) \qquad \beta_{1w} = .64928; \qquad \beta_{2,w} - 3 = 1.2 = \beta_{2v} - 3$$

are only one half of the corresponding characteristics of the largest value. The distribution of the range is less skew and less concentrated toward the middle than the distribution of the largest value. The moments (13) and (16) of the range and of the midrange may also be obtained directly from Fisher's and Tippett's results [2] when independence of the two extremes is assumed. The numerical values (16) for the limiting distribution of the range differ considerably from the values

$$(16') \qquad \beta_{1w} = .309; \qquad \beta_{2w} - 3 = .54$$

given by Tippett [7] for $n = 1000$ observations. For increasing n the approach of the distribution of the range toward its limiting distribution is very slow. This is reasonable as the approach of the distribution of the largest value toward the limiting distribution is also very slow.

Until now we considered symmetrical initial distributions. The case of an asymmetrical initial distribution may be dealt with briefly. The most probable mth extreme value from below ${}_mu$ differs then from $-u_m$ and

$$(1') \qquad {}_m\alpha = \frac{n}{m}\,\varphi({}_mu)$$

differs from α_m. The distribution of the mth extreme value from below is for large n and small m [3]

$$(2') \qquad {}_mf({}_mx) = {}_m\alpha\,\frac{m^m}{(m-1)!}\,e^{m\,{}_my - me^{\,{}_my}}$$

where

$$(3') \qquad {}_my = {}_m\alpha({}_mx - {}_mu)$$

is a reduced mth extreme value from below. The moment generating function ${}_mG(t)$ of the mth extreme value from below becomes

$$(4') \qquad {}_mG(t) = e^{{}_mut}m^{-(t/{}_m\alpha)}\Gamma\left(m + \frac{t}{{}_m\alpha}\right) \Big/ \Gamma(m)$$

and the moment generating functions $G_w(t, m)$ and $G_v(t, m)$ of the mth range and of the mth mid-range of an asymmetrical distribution obtained from (4), (4′) and in analogy to (14) are

$$(17) \qquad \begin{aligned} G_w(t, m) &= e^{t(u_m - {}_mu)}m^{t((1/\alpha_m)+(1/{}_m\alpha))}\Gamma\left(m - \frac{t}{\alpha_m}\right)\Gamma\left(m - \frac{t}{{}_m\alpha}\right) \Big/ \Gamma^2(m) \\ G_v(t, m) &= e^{t(u_m + {}_mu)}m^{t((1/\alpha_m)-(1/{}_m\alpha))}\Gamma\left(m - \frac{t}{\alpha_m}\right)\Gamma\left(m + \frac{t}{{}_m\alpha}\right) \Big/ \Gamma^2(m). \end{aligned}$$

Thus, the moments of the mth range and the mth mid-range may easily be computed even for an asymmetrical distribution.

4. The distribution of the mth midrange. In the following we return to a symmetrical distribution and establish directly the distribution $f(v_m)$ of the mth midrange. Then, the generating function (14′) and the convergence toward normality will be verified.

From (12′) the distribution of the mth midrange is

$$f(v_m) = \int_{-\infty}^{+\infty} f_m(x_m)\,{}_mf(v_m - x_m)\,dx_m\,.$$

Introducing (2), the equation is written

$$f(v_m) = \alpha_m \frac{m^{2m}}{(m-1)!^2}\int_{-\infty}^{+\infty} e^{-my_m + m\alpha_m(v_m - x_m + u_m) - me^{-y_m} - me^{\alpha_m(v_m - x_m + u_m)}}\,dy_m\,.$$

Using as before

$$e^{-y_m} = z$$

the integral becomes

$$e^{m\alpha_m v_m}\int_0^{\infty} z^{2m-1}e^{-mz(1+e^{\alpha_m v_m})}\,dz = \frac{e^{m\alpha_m v_m}(2m-1)!}{m^{2m}(1 + e^{\alpha_m v_m})^{2m}}\,.$$

The distribution of the mth midrange is therefore

$$(18) \qquad f(v_m) = \alpha_m \frac{(2m-1)!}{(m-1)!^2}\frac{e^{m\alpha_m v_m}}{(1 + e^{\alpha_m v_m})^{2m}}\,.$$

The distribution (18) will be shown to lead to the seminvariant generating function (14′). The generating function of the mth midrange obtained from (18) is

$$\frac{(2m-1)!}{(m-1)!^2}\int_{-\infty}^{\infty} e^{\alpha_m v_m (m+(t/\alpha_m))}(1+e^{\alpha_m v_m})^{-2m}\, d\alpha_m v_m .$$

Introducing

$$1+e^{\alpha_m v_m} = u^{-1}; \qquad du = -u^2 e^{\alpha_m v_m}\, d\alpha_m v_m$$

the integral is rewritten

$$\frac{(2m-1)!}{(m-1)!^2}\int_0^1 (1-u)^{m+(t/\alpha_m)-1} u^{-m-(t/\alpha_m)-1+2m-2}\, du$$

$$= \frac{\Gamma(2m)}{\Gamma^2(m)} \frac{\Gamma\left(m+\frac{t}{\alpha_m}\right)\Gamma\left(m-\frac{t}{\alpha_m}\right)}{\Gamma(2m)} .$$

Consequently, the moment generating function of the mth midrange is

$$G_v(t, m) = \frac{\Gamma\left(m+\frac{t}{\alpha_m}\right)\Gamma\left(m-\frac{t}{\alpha_m}\right)}{\Gamma^2(m)} . \tag{19}$$

This expression is identical with the product of the moment generating functions (4) and (9) of the extreme mth values. Therefore, equation (18) is the distribution, and equation (14′) is the seminvariant generating function of the mth midrange.

For $m = 1$ the distribution (18) of the mid-range becomes the so-called logistic distribution which is commonly written

$$f(v) = \frac{\alpha e^{-\alpha v}}{(1+e^{-\alpha v})^2} . \tag{18′}$$

Accordingly, (18) may be called the *generalized logistic distribution*. The probability $F(v)$ of a value equal to, or less than, v obtained from (18′) is

$$F(v) = 1/(1+e^{-\alpha v}). \tag{20}$$

Therefore the distribution $f(v)$ may be expressed by the probability $F(v)$ through

$$f(v) = \alpha F(v)(1 - F(v)). \tag{18″}$$

Formula (20) considered as a growth function plays a rôle in population statistics where it was introduced about 100 years ago by Verhulst [8]. Recently, it has been widely used by R. Pearl [6]. In his treatment, the value v stands for the time, and the function $F(v)$ stands for the relative size of the population at time v compared to its alleged asymptotic size.

In the following, the influence of m on the distribution (18) is studied. The distribution $\phi_m(z)$ of the reduced mth mid-range

$$z = \alpha_m \sqrt{m}\, v \tag{21}$$

is, from (18)

$$\phi_m(z) = \frac{1}{\sqrt{m}} \frac{(2m-1)!}{(m-1)!^2} \frac{e^{z\sqrt{m}}}{(1 + e^{z/\sqrt{m}})^{2m}}. \tag{22}$$

The probability density of the mean mth reduced mid-range increases with m. Indeed,

$$\frac{\phi_{m+1}(0)}{\phi_m(0)} = \frac{2m+1}{2\sqrt{m(m+1)}} > 1.$$

Therefore, the standard error of the reduced mth mid-range decreases with increasing index m. This is reasonable as the mth mid-range is an estimate of the median for the initial distribution. The larger m, the nearer are the mth extreme values to the median, and the better is the estimate.

To verify that the distribution (18) of the mth mid-range tends toward normality for large indices m equation (22) is rewritten

$$\lg \phi_m(z) = \lg \phi_m(0) + z\sqrt{m} - 2m \lg\left(\frac{1 + e^{z/\sqrt{m}}}{2}\right).$$

Expansion of the exponential and the logarithm leads, if we neglect the third and higher powers of the deviation z, to

$$\lg \phi_m(z) = \lg \phi_m(0) + z\sqrt{m} - 2m\left(\frac{z}{2\sqrt{m}} + \frac{z^2}{4m} - \frac{z^2}{8m} + \cdots\right)$$

whence by virtue of (11)

$$f(v) = \text{Const}\; e^{-\alpha_m^2 m v_m^2/4}.$$

The distribution of the mth mid-range becomes normal for large indices m. The mean is zero, and the standard deviation is

$$\sigma_{v_m} = \frac{\sqrt{2}}{\alpha_m \sqrt{m}}. \tag{23}$$

This is in accordance with the statement (13′), as $S_{2,m}$ tends with increasing m toward $1/m$.

5. Summary. For initial symmetrical distributions the seminvariant generating functions (14′) of the mth range and the mth mid-range are obtained from the seminvariant generating functions (7) and (9) of the mth extreme values from above and from below. As the two mth extreme values are supposed to be independent our results hold only for very large, sample sizes. The

even seminvariants of the mth mid-range and of the mth range coincide. The distribution of the mth range is skew and the distribution of the mth mid-range is the generalized symmetrical logistic distribution. For increasing indices m the distributions of the mth extremes, the mth ranges, and the mth mid-ranges converge toward normality.

REFERENCES

[1] W. E. Deming, "Some principles of the Shewhart methods of quality control," *Mech. Eng.*, Vol. 66 (1944), pp. 173–177.

[2] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. of the Camb. Phil. Soc.*, Vol. 24 (1928), Part 2, pp. 180–190.

[3] E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Ann. de l'Institut Henri Poincaré*, Vol. 4 (1935), pp. 115–158.

[4] R. von Mises, "La distribution de la plus grande de n valeurs," *Revue Math. de l'Union Interbalkanique*, Vol. 1 (1936), pp. 1–20.

[5] E. S. Pearson, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), Parts 3 and 4, pp. 301–308.

[6] R. Pearl, *Medical Biometry and Statistics*. 3rd Edition, W. M. Sanders Co., Philadelphia and London, 1940.

[7] L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925) Parts 3 and 4, pp. 264–387.

[8] P. F. Verhulst, "Recherches mathématiques sur la loi d'accroissement de la population," *Académie de Bruxelles*, Vol. 18 (1845), pp. 1–38.