PROBABILISTIC ANALYSIS OF THE EFFICACY OF PERIODIC TESTING OF EMPLOYEES

SIMEON M. BERMAN,* New York University

Abstract

Some major companies have the policy of annually giving numerical scores to their employees according to their performance, firing those whose performance scores are below a given percentile of the scores of all employees, and then recruiting new employees to replace those who were fired. We introduce a probabilistic model to describe how this practice affects the quality of employee performance as measured over time by the annual scores. Let *n* be the number of years that the policy has been in effect, and let $F_n(x)$ be the distribution function of the evaluation scores in year *n*. We show, under certain technical assumptions, that the sequence $(F_n(x))$ satisfies a particular nonlinear difference equation, and furnish estimates of the solution of the equation and expressions for the quantiles of F_n . The mathematical tools that are used include convex functions, difference equations, and extreme value theory for independent and identically distributed random variables.

Keywords: Convex function; difference equation; extreme value; performance test; quantile

2010 Mathematics Subject Classification: Primary 60K99; 62G32 Secondary 90B70

1. Introduction and summary

In recent years some major companies have introduced annual performance testing of employees for the purpose of evaluating and improving performance. When the first annual test is given, all employees are evaluated in some uniform manner and are given a numerical test score. On the basis of the scores, a predetermined proportion of employees are fired and are replaced by new employees over the course of the year. The employees who are fired are those with scores that fall below a predetermined quantile of the distribution of the scores. During the year following the annual test, some employees who passed the test leave voluntarily, and are also replaced by new employees. Thus, at the end of the second year, the employees who take the second annual test fall into one of two categories:

- (A) those taking the annual test for the first time;
- (B) those who passed the previous test and are now taking the second annual test.

For a specified α , $0 < \alpha < 1$, let α and $1 - \alpha$ represent the proportions of employees in categories (A) and (B), respectively.

This process is repeated every year: the employees who fail an annual test are fired and replaced by new employees, and those who pass an annual test and voluntarily leave during the

Received 3 June 2010; revision received 7 September 2010.

^{*} Postal address: Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. Email address: sberman@cims.nyu.edu

year are also replaced by new employees. The proportions α and $1 - \alpha$ in categories (A) and (B), respectively, are assumed to remain the same for each year.

Let F(x) be the distribution function of the numerical scores of those taking the annual test for the first time. We call this the 'basic score distribution':

$$F(x)$$
 = proportion of employees with scores less than or equal to x.

The population of employees is assumed to be very large so that, as is the convention in the mathematical analysis, we assume, for convenience, that F(x) is strictly increasing and continuous on $\{x : 0 < F(x) < 1\}$.

We recursively define a sequence of distribution functions $(F_n(x))$, where $F_n(x)$ represents the score distribution for the *n*th annual test, for $n \ge 1$. By definition we have $F_1(x) = F(x)$. We propose a mathematical model that defines $F_n(x)$ in terms of F(x). Consider the case in which n = 2. The distribution function F_2 of the scores on the second test is a weighted sum of the distribution functions of scores in categories (A) and (B), with weights α and $1 - \alpha$, respectively.

The score distribution for employees in category (A), i.e. those taking the test for the first time, is assumed to be the basic score distribution F. Now we define the score distribution for employees in category (B), i.e. those who passed the previous test. It is the composite function G(F(x)), where G is a distribution function satisfying

$$G(y), 0 \le y \le 1$$
, is continuous and convex, $G(0+) = 0$, and $G(1-) = 1$. (1.1)

It now follows that the weighted sum representing the score distribution for the second test is

$$F_2(x) = \alpha F(x) + (1 - \alpha)G(F(x)).$$
(1.2)

Now we explain the relevance of the assumption that G is convex. The convexity of G implies that $[G(x_2) - G(x_1)]/(x_2 - x_1)$, $x_1 < x_2$, is nondecreasing in x_1 for fixed x_2 and nondecreasing in x_2 for fixed x_1 . (See [4, p. 3].) Therefore, for any distribution function F(x),

$$\frac{G(F(x_2)) - G(F(x_1))}{F(x_2) - F(x_1)}$$
(1.3)

is similarly nondecreasing in x_1 and x_2 for fixed x_2 and x_1 , respectively, for $x_1 < x_2$. If X and Y are random variables with distribution functions F and G(F), respectively, then (1.3) represents the ratio

$$\frac{\mathsf{P}(x_1 < Y \le x_2)}{\mathsf{P}(x_1 < X \le x_2)}$$

The fact that this ratio is nondecreasing in x_1 and x_2 for fixed x_2 and x_1 , respectively, represents a strong form of a stochastic order relation between the random variables X and Y. These two random variables represent randomly drawn individuals from categories (A) and (B), respectively, and the bimonotonicity property of $P(x_1 < Y \le x_2)/P(x_1 < X \le x_2)$ signifies that Y is stochastically larger than X in the sense that we defined the order relation.

Remark. The conventional definition of the relationship 'Y is stochastically larger than X' is defined by the inequality $G(F(x)) \leq F(x)$, a weaker assumption than the one introduced through (1.3).

Set

1176

 $F_n(x) =$ score distribution on the *n*th annual test.

We will extend (1.2) to

$$F_n(x) = \alpha F(x) + (1 - \alpha)G(F_{n-1}(x)) \text{ for } n \ge 2,$$
(1.4)

with $F_1(x) = F(x)$. This relation describes the score distribution as a weighted sum of the score distribution of employees taking the test for the first time, which we assume to be the same as the basic score distribution F, and the score distribution of employees who passed the previous annual test (the (n-1)th test). The latter distribution is defined as $G(F_{n-1}(x))$, where $F_{n-1}(x)$ is the score distribution on the (n-1)th test. Here we assume that the weights $(\alpha, 1 - \alpha)$ remain the same for all tests and that the score distribution on the *n*th test for employees who passed the (n-1)th test is the composition $G(F_{n-1})$, with the same G. The latter assumption is meant to express the fact that the scores on an annual test for employees who *passed* the previous test are stochastically larger than the scores of all employees who *took* the previous test. (The latter group includes employees who failed the previous test.)

Since, by assumption, F(x) is continuous and increasing on the closure of $\{x: 0 < F(x) < 1\}$, and G(y) is continuous and convex on the closure of $\{y: 0 < G(y) < 1\}$, it follows from (1.2) that $F_2(x)$ is continuous and increasing on the same set. By (1.4) and induction on *n*, it follows that the same holds for all F_n , $n \ge 1$; hence, for every *y*, 0 < y < 1, the *y*-quantile $F_n^{-1}(y)$ is well defined.

Now we define the 'critical quantile', denoted by q: it is the smallest solution y of the equation

$$G(y) = \frac{y - \alpha}{1 - \alpha}, \qquad \alpha \le y \le 1.$$
(1.5)

Note that such a solution always exists because y = 1 is a solution and the right-hand member of (1.5) is negative for $y < \alpha$.

Set

$$z = \sup(x : F(x) < 1),$$
 (1.6)

where $z \le \infty$. It will be shown that $z = \sup(x : F_n(x) < 1)$ for all $n \ge 2$. Thus, we can think of z as the common perfect score on all the tests. It will be shown that $(F_n(x))$ is a nonincreasing sequence for each x, and so has a limit Q(x). The critical quantile q plays a crucial role in the form of Q. Indeed, Q(x) is continuous and increasing for all x such that Q(x) < q, and it is equal to 1 for all x such that $Q(x) \ge q$; hence, it has a jump of magnitude 1 - q at the point $x = Q^{-1}(q)$. The latter result can be formulated in an equivalent way in terms of the quantile functions: $F_n^{-1}(y)$, 0 < y < 1, is a nondecreasing sequence and has the limit $Q^{-1}(y)$, which is an increasing function of y for 0 < y < q, and $Q^{-1}(y) = z$ for $q \le y \le 1$. The statistical interpretation of this result is that the quantile function $F_n^{-1}(y)$ converges to a continuous limit for y-values below the critical quantile q and converges to the 'perfect score' z for y-values at or above the critical quantile. This implies that the long-run proportion of perfect (or nearly perfect) scores is 1 - q, and the complementary set of scores is distributed according to the distribution function Q for Q(x) < q. The relation between the quantiles of Q and the basic score distribution is given by (2.4) below.

While these limit results are of mathematical interest, the rate of convergence to the limits is of more practical interest because employers are generally seeking improvement over relatively short time periods, e.g. five or ten years. Thus, we investigate the solutions of the system of equations (1.4) for fixed n. Since these equations are generally nonlinear (because G is not

necessarily linear), the system does not have an explicit solution. However, upper and lower bounds for the solution can be obtained by replacing G by linear upper and lower bounds, and explicitly solving the corresponding linear equations. Formulae (3.15) and (3.16) below express these bounds in terms of α , F, and the Radon–Nikodym derivative of G. In the special case where G(y) is linear for $w \le y \le 1$ for some w > 0, we obtain an explicit solution for F_n and provide numerical illustrations in the extreme cases $q = \alpha$ and q = 1.

Next we turn to the determination of the asymptotic form of the quantile $F_n^{-1}(y)$ for $n \to \infty$ and for y at least equal to the critical quantile q. (As indicated above, $F_n^{-1}(y) \to z \le \infty$.) The analysis is based on the classical extreme value theory concerning the limiting distributions of the partial maxima of independent and identically distributed (i.i.d.) random variables. For this purpose, we assume that the basic score distribution F belongs to the domain of attraction of one of the three types of limiting extreme value distributions. Let H(x) represent an extreme value distribution of a given type; the well-known forms are given in Section 5. In Section 4 we construct sequences (A_n) , $A_n > 0$, and (B_n) , based on α , G, and F, such that

$$\limsup_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} \le H^{-1}(\mathrm{e}^{y-1})$$

for $q \le y < 1$. Similarly, we construct sequences (A_n) , $A_n > 0$, and (B_n) such that, under an additional condition on *G*, the inequality above, in reversed form, holds with lim sup replaced by lim inf. It follows as a corollary that if there exists w, $q \le w < 1$, such that if *G* is linear on [w, 1] then, for all $y, w \le y < 1$,

$$\lim_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} = H^{-1}(e^{y-1}).$$

The actual sequences are constructed in Section 5.

We note that the difference equations (1.4) represent an extended form of the classical difference equations for the extinction probabilities x_n in the basic branching process, $x_n = G(x_{n-1})$, where G is a generating function of the distribution of the number of offspring of a single individual. (See [3, p. 296].) Set $y_n = F_n(x)$ in (1.4), so that the equation is of the form $y_n = \alpha y_1 + (1 - \alpha)G(y_{n-1})$. This represents a generalization of the branching process equation. Indeed, the latter is a restriction of (1.3) to the particular case where $\alpha = 0$ and G(x) is a probability generating function (which is necessarily continuous, convex, and nondecreasing), while (1.3) holds for any G satisfying (1.1) and for α not necessarily equal to 0.

The problem that is analyzed in this work was motivated by an article that the author read in the *New York Times* [1] about a court case that arose as a result of the use of annual testing by large companies. The author first considered the simple mathematical model where *G* is the uniform distribution on $[\alpha, 1]$, and proposed it as a problem to Hartono Tjoe, a master's degree candidate at the Courant Institute, New York University, whose thesis [5] contained the solution of the difference equations in that particular case. His result is included in the special case considered in Example 3.1.

Explicit calculations in Section 6 illustrate the progression of selected quantiles of the score distributions F_n , n = 1, ..., 10, for two particular convex functions G. These examples indicate that (i) the improvement in scores is strongly influenced by the choice of the convex function G, and (ii) the rate of improvement seems to be relatively low, so that there are limits to the cost effectiveness of the practice in actual applications.

2. The limit of F_n

In this section we show that the sequence $F_n(x)$ defined by (1.4) has a limit Q(x), and we describe its essential features.

Theorem 2.1. Let (F_n) be defined recursively by (1.4) with $F_1 = F$. Then $F_n(x)$, n = 1, 2, ..., is nonincreasing for each x, so that $Q(x) = \lim_{n \to \infty} F_n(x)$ exists, and it satisfies the equation

$$Q(x) = \alpha F(x) + (1 - \alpha)G(Q(x)).$$
(2.1)

Furthermore, 0 < Q(x) < 1 if and only if 0 < F(x) < 1 for each x.

Proof. A convex function G satisfying G(0) = 0 and G(1) = 1 also satisfies $G(x) \le x$ for 0 < x < 1. It then follows from (1.3) that

$$F_2(x) \le \alpha F(x) + (1 - \alpha)F(x) = F_1(x).$$

We extend this by induction to $F_{n+1}(x) \leq F_n(x)$. Assume that $F_n(x) \leq F_{n-1}(x)$. Then, by (1.4),

$$F_n(x) - F_{n+1}(x) = (1 - \alpha)[G(F_{n-1}(x)) - G(F_n(x))] \ge 0.$$

It follows that the limit Q(x) exists, and, by the continuity of G, (2.1) follows from (1.4).

Since (F_n) is nonincreasing, it follows that $Q(x) \le F(x)$; hence, F(x) < 1 implies that Q(x) < 1. Conversely, if F(x) = 1 then (1.2) implies that $F_2(x) = 1$, and, by induction, $F_n(x) = 1$ for all $n \ge 1$; hence, Q(x) = 1, and, consequently, Q(x) < 1 if and only if F(x) < 1. If F(x) = 0 then Q(x) = 0 because $Q(x) \le F(x)$ for all x. Conversely, if Q(x) = 0 then, by (2.1), F(x) is necessarily equal to 0. Therefore, Q(x) > 0 if and only if F(x) > 0. This completes the proof.

Theorem 2.2. Let z be defined by (1.6). Then Q(x) is strictly increasing and continuous on $\{x : 0 < Q(x) < 1\}$, q (the smallest solution of (1.5)) satisfies

$$Q(z-) = \lim_{x \uparrow z} Q(x) = q,$$
 (2.2)

and Q(x) has a jump of magnitude 1 - q at x = z.

Proof. By (2.1), Q(x) increases at x if F(x) does; hence, Q(x) increases on the set $\{x : 0 < F(x) < 1\}$ because F is assumed to increase on that set. By Theorem 2.1, that set is identical to $\{x : 0 < Q(x) < 1\}$ and so Q(x) increases on the latter set.

Let *y* be any number satisfying 0 < Q(y) < q. Then, by the definition of *q*,

$$G(q) = \frac{q-\alpha}{1-\alpha}, \qquad G(Q(y)) > \frac{Q(y)-\alpha}{1-\alpha},$$

so that γ , defined as

$$\gamma = (1 - \alpha) \frac{G(q) - G(Q(y))}{q - Q(y)},$$

satisfies $0 < \gamma < 1$. By the convexity of *G* we have

$$(1-\alpha)\frac{G(s)-G(r)}{s-r} \le \gamma \quad \text{for all } r < s < Q(y).$$
(2.3)

To prove that Q(x) is continuous at every x < y, assume the contrary, i.e. that x is a point of discontinuity of Q, so that Q(x-) < Q(x+). Since, by assumption, F is continuous, (2.1) implies that

$$Q(x+) - Q(x-) = (1-\alpha)[G(Q(x+)) - G(Q(x-))]$$

Since Q is strictly increasing, it follows that G(Q(x-)) < G(Q(x+)) < G(Q(y)); hence, applying (2.3) with r = Q(x-) and s = Q(x+), we find that

$$(1-\alpha)[G(Q(x+)) - G(Q(x-))] \le \gamma [Q(x+) - Q(x-)],$$

which contradicts the previously displayed equation because $\gamma < 1$. Since y is an arbitrary number with Q(y) < q and x is an arbitrary number less than y, it follows that x is a continuity point for any x such that Q(x) < q.

Now we prove (2.2). If x < z then F(x) < 1, and so, by (2.1),

$$G(Q(x)) = \frac{Q(x) - \alpha F(x)}{1 - \alpha} > \frac{Q(x) - \alpha}{1 - \alpha}$$

and so Q(x) is not a solution of (1.5), so Q(x) < q. Since x < z is arbitrary, it follows that $Q(z-) \le q$. On the other hand, Q(z-) is, in fact, a solution of (1.5), because by letting $x \uparrow z$ in (2.1) we obtain $Q(z-) = \alpha + (1-\alpha)G(Q(z-))$. Since q is, by definition, the smallest solution, we have $q \le Q(z-)$. From this and the previous (reverse) inequality, we obtain (2.2).

Since, by Theorem 2.1, the sets $\{x : 0 < F(x) < 1\}$ and $\{x : 0 < Q(x) < 1\}$ are identical, Q(z) = F(z) = 1, and so Q has a jump of magnitude 1 - q at z. This completes the proof.

Theorem 2.3. *For every* y, 0 < y < q,

$$Q^{-1}(y) = F^{-1}\left(\frac{1}{\alpha}[y - (1 - \alpha)G(y)]\right).$$
(2.4)

Proof. By Theorem 2.2, $Q^{-1}(y)$ is well defined for 0 < y < q. For each such y, set $x = Q^{-1}(y)$ in (2.1):

$$y = \alpha F(Q^{-1}(y)) + (1 - \alpha)G(y),$$

which is equivalent to

$$F(Q^{-1}(y)) = \frac{1}{\alpha} [y - (1 - \alpha)G(y)],$$

and (2.4) follows upon applying F^{-1} to both sides of the latter equation.

Equation (2.4) relates the quantiles of *F* and *Q*: the *y*-quantile of *Q* is the quantile of order $(1/\alpha)[y - (1 - \alpha)G(y)]$ for *F*.

Theorem 2.4. If $G(\alpha) = 0$ then

$$Q(x) = \begin{cases} \alpha F(x), & x < z, \\ 1, & x \ge z. \end{cases}$$

Proof. If $G(\alpha) = 0$ then $y = \alpha$ is a solution of (1.5) and is necessarily the smallest solution, so $q = \alpha$. If x < z then, by (2.2), Q(x) < Q(z-) = q; hence, $G(Q(x)) \le G(q) = G(\alpha) = 0$, and, by (2.1), $Q(x) = \alpha F(x)$. If $x \ge z$ then $Q(x) \ge Q(z) = 1$, which implies that Q(x) = 1. This completes the proof.

Theorem 2.5. The sequence $F_n^{-1}(y)$ satisfies

$$\lim_{n \to \infty} F_n^{-1}(y) = \begin{cases} Q^{-1}(y), & 0 < y < q, \\ z, & q \le y \le 1. \end{cases}$$
(2.5)

Proof. By Theorem 2.1, $F_n(x)$ is nonincreasing for each x, so $F_n^{-1}(y)$ is nondecreasing for each $0 < y \le 1$; thus, $\lim_{n\to\infty} F_n^{-1}(y)$ exists. Since Q(x) and $F_n(x)$ are continuous and strictly monotonic on $\{x : 0 < Q(x) < q\}$, the convergence of $F_n(x)$ to Q(x) on this set implies the first equation in (2.5).

Since $Q^{-1}(y)$ is monotonic for 0 < y < 1, it suffices, for the proof of the second equation in (2.5), to show that $Q^{-1}(q) = z$, which is equivalent to $F(Q^{-1}(q)) = 1$. The latter follows by substituting $x = Q^{-1}(q)$ into (2.1), i.e. we obtain

$$q = \alpha F(Q^{-1}(q)) + (1 - \alpha)G(q),$$

which, by (1.5), is equivalent to

$$0 = \alpha F(Q^{-1}(q)) - \alpha,$$

which implies that $F(Q^{-1}(q)) = 1$. This completes the proof.

Corollary 2.1. For $q \le y < 1$,

$$\lim_{n \to \infty} F_n(F_{n-1}^{-1}(y)) = \alpha + (1 - \alpha)G(y).$$
(2.6)

Proof. Set $x = F_{n-1}^{-1}(y)$ in (1.4). Then

$$F_n(F_{n-1}^{-1}(y)) = \alpha F(F_{n-1}^{-1}(y)) + (1-\alpha)G(y).$$

Take the limit for $n \to \infty$, and apply Theorem 2.5, i.e. $\lim_{n\to\infty} F(F_{n-1}^{-1}(y)) = F(z) = 1$. Then (2.6) follows.

Example 2.1. Suppose that $G(y) = y^2$, $0 \le y \le 1$. Then (2.1) takes the form $Q = \alpha F + (1 - \alpha)Q^2$. Solving for Q we obtain

$$Q(x) = \frac{1 - [1 - 4\alpha(1 - \alpha)F(x)]^{1/2}}{2(1 - \alpha)}$$
(2.7)

as the only solution that is nondecreasing in x. By (1.5), q satisfies the equation

$$q = \alpha + (1 - \alpha)q^2$$

whose solutions are

$$q = \frac{1 \pm [1 - 4\alpha(1 - \alpha)]^{1/2}}{2(1 - \alpha)} = \frac{1 \pm |2\alpha - 1|}{2(1 - \alpha)}.$$
(2.8)

If $\alpha > \frac{1}{2}$ then the smaller solution is equal to 1 and follows by choosing the minus sign. If $\alpha = \frac{1}{2}$ then q = 1 is the only solution, and (2.7) takes the form $Q(x) = 1 - [1 - F(x)]^{1/2}$. If $\alpha < \frac{1}{2}$ then the two solutions of q in (2.8) are $\alpha/(1 - \alpha)$ and 1, and so the smaller one is $q = \alpha/(1 - \alpha)$, which is less than 1. It follows from this analysis that the smallest solution of (2.8) is

$$q = \min\left(\frac{\alpha}{1-\alpha}, 1\right).$$

3. Estimates of the solution of (1.4)

For each x, the system of equations (1.4) recursively express $F_n(x)$ in terms of $F_j(x)$, j = 1, ..., n - 1, so that $F_n(x)$ is expressible in terms of F(x). In the case where G is linear on some interval [r, 1], 0 < r < 1, the system (1.4) is first-order linear and can be explicitly solved for $F_n(x)$. In this section we consider the case where G is majorized or minorized by linear functions, and we obtain upper and lower bounds, respectively, for $F_n(x)$ in terms of the bounding linear functions and F(x).

Lemma 3.1. Let $y_1, y_2, ..., y_n$ satisfy

$$y_j \le a + by_{j-1}, \qquad j = 2, \dots, n,$$
 (3.1)

for fixed a > 0, b > 0, and $b \neq 1$. Then

$$y_n \le a \frac{b^{n-1} - 1}{b - 1} + b^{n-1} y_1.$$
(3.2)

The relation between (3.1) and (3.2) also holds with the reversed inequalities: if

$$y_j \ge a + by_{j-1}, \qquad j = 2, \dots, n_j$$

then

$$y_n \ge a \frac{b^{n-1} - 1}{b - 1} + b^{n-1} y_1.$$
(3.3)

If b = 1 then the right-hand members of (3.2) and (3.3) are replaced by

$$a(n-1) + y_1.$$
 (3.4)

Proof. Starting with j = 2, we apply (3.1) to obtain

$$y_2 \le a + by_1,$$

 $y_3 \le a + b(a + by_1) = a(1 + b) + b^2y_1, \dots, y_n \le a(1 + b + \dots + b^{n-2}) + b^{n-1}y_1.$

Noting that

$$a\sum_{j=0}^{n-2} b^j + b^{n-1}y_1 = \begin{cases} a\frac{b^{n-1}-1}{b-1} + b^{n-1}y_1 & \text{for } b \neq 1, \\ a(n-1) + y_1 & \text{for } b = 1, \end{cases}$$

we obtain (3.2) and (3.4). The reverse inequality (3.3) is obtained by reversing the inequalities in (3.1). This completes the proof.

For the purpose of the next several proofs, we note that (1.4) is equivalent to

$$1 - F_n(x) = \alpha [1 - F(x)] + (1 - \alpha) [1 - G(F_{n-1}(x))].$$
(3.5)

For arbitrary w, $0 < w \le 1$, define g(w) as

$$g(w) = \begin{cases} \lim_{y \downarrow w} \frac{G(y) - G(w)}{y - w}, & 0 < w < 1, \\ \lim_{y \uparrow 1} \frac{1 - G(y)}{1 - y}, & w = 1. \end{cases}$$

These limits exist because G is convex. The convexity of G also implies that

$$1 - G(y) \le (1 - y)g(1), \qquad 0 \le y \le 1,$$

$$1 - G(y) \ge (1 - y)g(w), \qquad w \le y \le 1.$$
(3.6)

For $0 < \alpha < 1$, $n \ge 2$, and $0 \le y \le 1$, define

$$R_n(y) = \begin{cases} \alpha \frac{[(1-\alpha)g(y)]^{n-1}-1}{(1-\alpha)g(y)-1} + [(1-\alpha)g(y)]^{n-1} & \text{for } (1-\alpha)g(y) \neq 1, \\ 1+(n-1)\alpha & \text{for } (1-\alpha)g(y) = 1. \end{cases}$$
(3.7)

Theorem 3.1. Let F_n be defined by (1.4). Then,

$$1 - F_n(x) \le [1 - F(x)]R_n(1) \quad for \ all \ x, \tag{3.8}$$

and, for every w, $0 \le w < 1$,

$$1 - F_n(x) \ge [1 - F(x)]R_n(w) \tag{3.9}$$

for all x satisfying

$$F_{n-1}(x) \ge w. \tag{3.10}$$

Proof. Setting $y = F_{n-1}(x)$, and applying the first inequality in (3.6) to the right-hand member of (3.5), we obtain

$$1 - F_n(x) \le \alpha [1 - F(x)] + (1 - \alpha) [1 - F_{n-1}(x)]g(1)$$

for all $n \ge 2$. Now apply Lemma 3.1 (first part) with $a = \alpha [1 - F(x)]$ and $b = (1 - \alpha)g(1)$. This proves (3.8).

Again setting $y = F_{n-1}(x)$, and noting assumption (3.10), we apply the second inequality in (3.6) to the right-hand member of (3.5) to obtain

$$1 - F_n(x) \ge \alpha [1 - F(x)] + (1 - \alpha) [1 - F_{n-1}(x)] g(w)$$
(3.11)

for all x satisfying (3.10). Since $F_1(x) \ge F_2(x) \ge \cdots \ge F_{n-1}(x)$ for each x (Theorem 2.1), the reasoning used to prove (3.11) implies that the latter inequality can be extended to

$$1 - F_j(x) \ge \alpha [1 - F(x)] + (1 - \alpha) [1 - F_{j-1}(x)] g(w), \qquad j = 2, \dots, n.$$

Now apply the second part of Lemma 3.1 with $a = \alpha [1 - F(x)]$ and $b = (1 - \alpha)g(w)$ to obtain (3.9). This completes the proof.

Corollary 3.1. *For* 0 < y < 1 *and* $n \ge 2$ *,*

$$1 - F(F_n^{-1}(y)) \ge \frac{1 - y}{R_n(1)},$$
(3.12)

and, for all 0 < w < 1 and $n \ge 2$,

$$1 - F(F_n^{-1}(y)) \le \frac{1 - y}{R_n(w)} \quad \text{for } w \le y \le 1.$$
(3.13)

Proof. For 0 < y < 1, set $x = F_n^{-1}(y)$ in (3.8) to obtain $1 - y = 1 - F_n(x) = 1 - F_n(x)$ $F_n(F_n^{-1}(y)) \le [1 - F(F_n^{-1}(y))]R_n(1)$; this implies (3.12). For any 0 < w < 1, and with $w \le y < 1$, set $x = F_{n-1}^{-1}(y)$. Then $F_{n-1}(x) = y \ge w$, and

so (3.9) implies that

$$1 - F_n(x) \ge [1 - F(x)]R_n(w). \tag{3.14}$$

Now, for any $y, w \le y \le 1$, set $x = F_n^{-1}(y)$. Then $F_{n-1}(x) = F_{n-1}(F_n^{-1}(y)) \ge F_n(F_n^{-1}(y)) = y \ge w$. Therefore, (3.14) holds for such x, and we obtain (3.13) by this substitution. This completes the proof.

Note that (3.12) is clearly equivalent to

$$F_n^{-1}(y) \le F^{-1}\left(1 - \frac{1 - y}{R_n(1)}\right), \qquad 0 < y \le 1,$$
(3.15)

and (3.13) is equivalent to

$$F_n^{-1}(y) \ge F^{-1}\left(1 - \frac{1 - y}{R_n(w)}\right), \qquad w \le y \le 1.$$
 (3.16)

If G(y) is linear for $w \le y \le 1$ then (3.15) and (3.16) imply that

$$F_n^{-1}(y) = F^{-1}\left(1 - \frac{1 - y}{R_n(1)}\right), \qquad w \le y \le 1,$$
(3.17)

because g(w) = g(1), and so $R_n(w) = R_n(1)$.

Lemma 3.2. If q < 1 then either $(1 - \alpha)g(1) > 1$ or $(1 - \alpha)g(y) = 1$ for $q \le y \le 1$.

Proof. If q < 1 then, by (1.5), the definition of q, and the convexity of G,

$$\frac{1}{1-\alpha} = \frac{1-G(q)}{1-q} \le \frac{1-G(y)}{1-y} \le g(1), \qquad q \le y < 1.$$

If either of the two inequalities above is strict then $(1 - \alpha)g(1) > 1$. If both inequalities are equalities then

$$\frac{1-y}{1-\alpha} = 1 - G(y), \qquad q \le y \le 1,$$

and, by differentiation, $(1 - \alpha)g(y) = 1$. This completes the proof.

Example 3.1. Let G(x) be the uniform distribution on $[\alpha, 1]$. It is clear that $q = \alpha$ and that $g(y) = G'(y) = (1 - \alpha)^{-1}$ for $\alpha \le y \le 1$. It follows from (3.7) that $R_n(y) = 1 + (n - 1)\alpha$.

If $F_{n-1}(x) \leq \alpha$ then $G(F_{n-1}(x)) \leq G(\alpha) = 0$, and so, by (1.4), $F_n(x) = \alpha F(x)$. If $F_n(x) \ge \alpha$, it follows from Theorem 3.1 and the fact that $R_n(y)$ is constant on $\alpha \le y \le 1$ that $F_n(x) = 1 - [1 - F(x)][1 + (n - 1)\alpha]$. Hence, F_n is given by

$$F_n(x) = \begin{cases} \alpha F(x) & \text{for } F_{n-1}(x) \le \alpha, \\ 1 - [1 - F(x)][1 + (n-1)\alpha] & \text{for } F_{n-1}(x) \ge \alpha. \end{cases}$$
(3.18)

By (3.17),

$$F_n^{-1}(y) = F^{-1}\left(1 - \frac{1 - y}{1 + (n - 1)\alpha}\right), \qquad \alpha \le y \le 1;$$
(3.19)

hence,

$$F_{n-1}^{-1}(\alpha) = F^{-1}\left(\frac{(n-1)\alpha}{1+(n-2)\alpha}\right).$$

It follows that (3.18) has the equivalent form,

$$F_n(x) = \begin{cases} \alpha F(x) & \text{for } F(x) \le (n-1)\alpha/(1+(n-2)\alpha), \\ 1 - [1 - F(x)][1 + (n-1)\alpha] & \text{for } F(x) \ge (n-1)\alpha/(1+(n-2)\alpha). \end{cases}$$

Example 3.2. Let G(x) be the uniform distribution on $[\beta, 1]$, where $0 < \beta < \alpha$; *G* is linear on $[\beta, 1]$ and $g(y) = (1 - \beta)^{-1}$ for $\beta \le y \le 1$. Here q = 1 because the smallest solution of (1.5) for $G(y) = (y - \beta)/(1 - \beta)$ is y = 1. The function $R_n(y)$ in (3.7) takes the form

$$R_n(y) = R_n = \alpha \frac{1 - \delta^{n-1}}{1 - \delta} + \delta^{n-1} \text{ for } \beta \le y < 1,$$
 (3.20)

where $\delta = (1 - \alpha)/(1 - \beta) < 1$. It follows from (1.4), Theorem 3.1, and the form of G that

$$F_n(x) = \begin{cases} \alpha F(x) & \text{for } F_{n-1}(x) \le \beta, \\ 1 - R_n(1 - F(x)) & \text{for } F_{n-1}(x) \ge \beta. \end{cases}$$
(3.21)

Since G(x) is linear on $[\beta, 1]$, (3.17) holds for $w = \beta$, and, in particular, we obtain

$$F_{n-1}^{-1}(\beta) = F^{-1}\left(1 - \frac{1-\beta}{R_{n-1}}\right), \qquad n \ge 2.$$
(3.22)

The inequality $F_{n-1}(x) \ge \beta$, which is equivalent to $x \ge F_{n-1}^{-1}(\beta)$, implies, by (3.22), that $F(x) \ge 1 - (1 - \beta)/R_{n-1}$, so (3.21) is equivalent to

$$F_n(x) = \begin{cases} \alpha F(x) & \text{for } F(x) \le 1 - (1 - \beta)/R_{n-1}, \\ 1 - R_n(1 - F(x)) & \text{for } F(x) \ge 1 - (1 - \beta)/R_{n-1}. \end{cases}$$
(3.23)

It follows from (3.20) that $R_n \to \alpha/(1-\delta)$ for $n \to \infty$, and from the definition of δ that $\alpha/(1-\delta) = \alpha(1-\beta)/(\alpha-\beta)$. By taking the limits in (3.23), we obtain

$$\lim_{n \to \infty} F_n(x) = \begin{cases} \alpha F(x) & \text{for } F(x) \le \beta/\alpha, \\ 1 - \frac{\alpha(1-\beta)}{\alpha-\beta} [1-F(x)] & \text{for } F(x) \ge \beta/\alpha. \end{cases}$$

4. Asymptotic bounds for $F_n^{-1}(y)$ when F belongs to the domain of attraction of an extreme value distribution

For a given distribution function F, the limiting behavior of $F^{-1}(1-h)$, $h \downarrow 0$, has a central role in the analysis of the limiting distribution of the partial maximum in a sample of i.i.d. random variables with distribution function F. A brief summary of facts from extreme value theory that are relevant to this work are included in Section 5. Under the condition that $R_n(w) \rightarrow \infty$, $w \leq 1$, the right-hand members of (3.15), (3.16), and (3.17) are of the same asymptotic form, $F^{-1}(1-h)$ for $h \downarrow 0$. Hence, when F belongs to the domain of attraction of an extreme value distribution, the corresponding theory can be applied to give explicit asymptotic forms for $F^{-1}(1-h)$.

1184

Suppose that *F* is in the domain of attraction of the extreme value distribution H(x). Then there are sequences of numbers (a_n) and (b_n) , with $a_n > 0$, such that, for every *x*,

$$\lim_{n \to \infty} F^n(a_n x + b_n) = H(x),$$

or, equivalently,

$$\lim_{n \to \infty} n[1 - F(a_n x + b_n)] = -\log H(x) \quad \text{for all } x, \tag{4.1}$$

or, equivalently, for $-\log H(x) = t$,

$$\lim_{n \to \infty} n[1 - F(a_n H^{-1}(e^{-t}) + b_n)] = t \quad \text{for all } t > 0.$$
(4.2)

If (r_n) is a positive integer-valued sequence such that $r_n \to \infty$, then (4.1) and (4.2) continue to hold if the index *n* in these relations is replaced by r_n . Hence, if we define

$$A_n = a_{r_n} \quad \text{and} \quad B_n = b_{r_n},\tag{4.3}$$

then

$$\lim_{n \to \infty} r_n [1 - F(A_n x + B_n)] = -\log H(x)$$

and

$$\lim_{t \to \infty} r_n [1 - F(A_n H^{-1}(e^{-t}) + B_n)] = t.$$
(4.4)

Theorem 4.1. Assume that q < 1 and that F is in the domain of H. Set

$$r_n = integer \ part \ of \ R_n(1). \tag{4.5}$$

Then, for $q \leq y < 1$, and with A_n and B_n in (4.3),

$$\limsup_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} \le H^{-1}(e^{y-1}).$$
(4.6)

Proof. By Lemma 3.2, and the definition of $R_n(y)$ in (3.7), it follows that $R_n(1) \to \infty$ under the condition q < 1; the same holds for r_n in (4.5).

For fixed y, $q \le y < 1$, let y' be an arbitrary number such that y < y' < 1. Then, by (4.4), with t = 1 - y',

$$\lim_{n \to \infty} r_n [1 - F(A_n H^{-1}(e^{y'-1}) + B_n)] = 1 - y'.$$

By (3.12),

$$r_n[1 - F(F_n^{-1}(y))] \ge 1 - y \text{ for } n \ge 2$$

From the last two relations, it follows that

$$\limsup_{n \to \infty} \frac{1 - F(A_n H^{-1}(e^{y'-1}) + B_n)}{1 - F(F_n^{-1}(y))} \le \frac{1 - y'}{1 - y} < 1,$$

so that, for all sufficiently large *n*,

$$F_n^{-1}(y) < A_n H^{-1}(e^{y'-1}) + B_n;$$

hence,

$$\limsup_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} \le H^{-1}(e^{y'-1}).$$

Conclusion (4.6) follows from this because $H^{-1}(e^{y-1})$ is monotonic and continuous in y, and y' > y is arbitrary.

Our next step is to derive a reverse inequality corresponding to (4.6), with lim inf in the place of lim sup, and where the index sequence (r_n) is replaced by one that increases more slowly and the domain of the quantile function $F_n^{-1}(y)$ is restricted to a specified subinterval $w \le y < 1$, where w is a number in [q, 1] determined by α and G. This is shown in the following lemma.

Lemma 4.1. Let g(y) be the right-hand derivative of G(y), defined in Section 3. If q < 1 then there exists $w, q \le w < 1$, such that

$$\lim_{n \to \infty} R_n(w) = \infty. \tag{4.7}$$

Proof. By the definition of $R_n(y)$ in (3.7), it suffices to show that there exists $w, q \le w < 1$, such that $(1-\alpha)g(w) \ge 1$. Assume the contrary, namely, that $(1-\alpha)g(y) < 1$ for all $q \le y < 1$, or, equivalently, $1 - (1 - \alpha)g(y) > 0$ for $q \le y < 1$. By integration we obtain

$$0 < \int_{q}^{1} [1 - (1 - \alpha)g(y)] \, \mathrm{d}y = 1 - q - (1 - \alpha)[1 - G(q)].$$

It follows from (1.5) that the right-hand member of the last relation is equal to 0, which obviously contradicts the stated inequality.

Let w be a number satisfying the conditions in the conclusion of Lemma 4.1, and define, by analogy to (4.5),

$$r_n(w) =$$
integer part of $R_n(w)$.

Then define (A_n) and (B_n) as in (4.3), but with $r_n(w)$ in place of r_n .

Theorem 4.2. Assume, as in Theorem 4.1, that q < 1 and F is in the domain of H. Then, for any $w, q \le w < 1$, satisfying (4.7), and for (A_n) and (B_n) as in (4.3) with $r_n(w)$ in place of r_n ,

$$\liminf_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} \ge H^{-1}(e^{y-1})$$
(4.8)

for $w \leq y < 1$.

Proof. The proof is analogous to that of Theorem 4.1, with the modification that inequality (3.13) is used in the place of (3.12). Given y, we choose an arbitrary y' < y, and find that

$$\liminf_{n \to \infty} \frac{1 - F(A_n H^{-1}(\mathbf{e}^{y'-1}) + B_n)}{1 - F(F_n^{-1}(y))} \ge \frac{1 - y'}{1 - y} > 1,$$

so that, for all sufficiently large *n*,

$$F_n^{-1}(y) > A_n H^{-1}(e^{y'-1}) + B_n,$$

which implies that

$$\liminf_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} \ge H^{-1}(e^{y'-1}),$$

and conclusion (4.8) follows from the continuity and monotonicity of H^{-1} .

Corollary 4.1. Assume that the conditions of Theorem 4.2 hold. If, in addition, G(y) is linear on [w, 1], then, for $w \le y < 1$,

$$\lim_{n \to \infty} \frac{F_n^{-1}(y) - B_n}{A_n} = H^{-1}(e^{y-1}).$$

Proof. Since G(y) is linear on [w, 1], g(y) is constant on [w, 1], and so $r_n = r_n(w)$.

5. Applications to the three domains of attraction

The following results on extreme values appear in many papers and books, but here we follow one of the classical sources, De Haan [2]. Our notation differs from his in that we replace his parameter α with γ because we have used α to represent the population proportion described in Section 1, and we replace his upper bound x_0 of the support of F with z, defined in Section 1.

There are three possible types of distribution H(x). The first type is

$$\Phi_{\gamma}(x) = \begin{cases} 0, & x \le 0, \\ \exp[-x^{-\gamma}], & x > 0. \end{cases}$$
(5.1)

The sequences (a_n) and (b_n) are

$$a_n = F^{-1} \left(1 - \frac{1}{n} \right), \qquad b_n = 0.$$
 (5.2)

An easy calculation shows that

$$H^{-1}(e^{y-1}) = (1-y)^{-1/\gamma}, \qquad 0 < y < 1.$$

The second type is

$$\Psi_{\gamma}(x) = \begin{cases} \exp\left[-(-x)^{r}\right], & x \le 0, \\ 1, & x \ge 0. \end{cases}$$
(5.3)

The sequences (a_n) and (b_n) are

$$a_n = z - F^{-1} \left(1 - \frac{1}{n} \right), \qquad b_n = z.$$
 (5.4)

For H as in (5.3), we clearly have

$$H^{-1}(e^{y-1}) = -(1-y)^{1/\gamma}, \qquad 0 < y < 1.$$

The third type is

$$\Lambda(x) = \exp[-e^{-x}]. \tag{5.5}$$

The sequences (a_n) and (b_n) are

$$a_n = n \int_{F^{-1}(1-1/n)}^{z} [1 - F(u)] \,\mathrm{d}u, \qquad b_n = F^{-1} \bigg(1 - \frac{1}{n} \bigg), \tag{5.6}$$

and $H^{-1}(e^{y-1})$ is given by

$$H^{-1}(e^{y-1}) = -\log(1-y), \qquad 0 < y < 1$$

In order to illustrate the applications of these results to Theorems 4.1 and 4.2, and to Corollary 4.1, we derive the forms of the ratio

$$\frac{F_n^{-1}(y) - B_n}{A_n}$$
(5.7)

for the three domains of attraction, respectively.

For the domain of Φ_{γ} in (5.1), relations (5.2) imply that $B_n = 0$ and

$$A_n = \begin{cases} F^{-1} \left(1 - \frac{1}{r_n} \right) & \text{in Theorem 4.1,} \\ F^{-1} \left(1 - \frac{1}{r_n(w)} \right) & \text{in Theorem 4.2.} \end{cases}$$

For the domain of Ψ_{γ} in (5.3), relations (5.4) imply that $B_n = z$ and

$$A_n = \begin{cases} z - F^{-1} \left(1 - \frac{1}{r_n} \right) & \text{in Theorem 4.1,} \\ z - F^{-1} \left(1 - \frac{1}{r_n(w)} \right) & \text{in Theorem 4.2.} \end{cases}$$

For the domain of Λ in (5.5), relations (5.6) imply that

$$A_n = \begin{cases} r_n \int_{F^{-1}(1-1/r_n)}^{z} [1-F(u)] \, du & \text{in Theorem 4.1,} \\ r_n(w) \int_{F^{-1}(1-1/r_n(w))}^{z} [1-F(u)] \, du & \text{in Theorem 4.2.} \end{cases}$$

Example 5.1. Suppose that $1 - F(x) \sim x^{-\gamma}$ for $x \to \infty$ and fixed $\gamma > 0$. Then *F* is in the domain of attraction of Φ_{γ} . Then, by (5.2), $b_n = 0$ and $a_n \sim n^{1/\gamma}$ for $n \to \infty$. Thus, the ratio in (5.7) is asymptotically equal to $F_n^{-1}(y)/r_n^{1/\gamma}$ or $F_n^{-1}(y)/(r_n(w))^{1/\gamma}$ under Theorem 4.1 or 4.2, respectively.

Example 5.2. Suppose that $z < \infty$ and that *F* satisfies $1 - F(z - x) \sim x^{\gamma}$ for $x \to 0+$. Then *F* is in the domain of attraction of Ψ_{γ} , and, by (5.4), $b_n = z$, and it can be shown that

$$a_n = z - F^{-1}\left(1 - \frac{1}{n}\right) \sim n^{-1/\gamma} \quad \text{for } n \to \infty.$$

Thus, the ratio in (5.7) is asymptotically equal to $[r_n \text{ or } r_n(w)]^{1/\gamma} [F_n^{-1}(y) - z]$.

Example 5.3. In contrast to the domains of attraction of (5.1) and (5.3), the asymptotic form of 1 - F(x) for $x \to z$ is not conveniently related to the conditions for membership in the domain of attraction of Λ in (5.5). Hence, we present two examples where F is known to belong to the domain of Λ and where $F^{-1}(1 - h)$ has a sufficiently precise asymptotic estimate for $h \to 0$. First, let F(x) be the standard exponential distribution. It is known to belong to the domain of Λ , and, by (5.6), $b_n = F^{-1}(1 - 1/n) = \log n$ and $a_n \equiv 1$. It follows that the ratio in (5.7) is equal to

$$F_n^{-1}(y) = \log(r_n \text{ or } r_n(w)).$$

Next, let F(x) be the standard normal distribution, which is also known to belong to the domain of Λ . A well-known asymptotic formula for $F^{-1}(1-1/n)$ is

$$(2\log n)^{1/2} - \frac{1}{2} \frac{\log\log n + \log 4\pi}{(2\log n)^{1/2}} + o((2\log n)^{-1/2}) \quad \text{for } n \to \infty;$$

furthermore, it is known that a_n in (5.6) has the asymptotic form $(2 \log n)^{-1/2}$. Hence, the ratio in (5.7) takes the form

$$(2\log r_n)^{1/2} \bigg[F_n^{-1}(y) - F^{-1} \bigg(1 - \frac{1}{r_n} \bigg) \bigg],$$

or with $r_n(w)$ in place of r_n .

6. Numerical illustrations of Examples 3.1 and 3.2

In Example 3.1 take $\alpha = 0.1$ and let *F* be the standard normal distribution function. (This accords with the practice of treating test scores as normally distributed.) For the purpose of illustrating the effect of annual testing on the corresponding test scores, Table 1 provides the three quartiles of the score distribution F_n for $1 \le n \le 10$, based on (3.19) with y = 0.25.

Table 2 provides the three quartiles for the score distribution F_n for $1 \le n \le 10$ in Example 3.2 for $\alpha = 0.1$, $\beta = 0.05$, and F the standard normal distribution.

TABLE 1: The first three quartiles for F_n , n = 1, ..., 10.

п	Quartile		
	First	Second	Third
1	-0.67	0.00	0.67
2	-0.47	0.11	0.74
3	-0.32	0.21	0.81
4	-0.19	0.29	0.88
5	-0.10	0.37	0.92
6	0.00	0.43	0.95
7	0.08	0.49	0.99
8	0.15	0.55	1.05
9	0.20	0.58	1.08
10	0.28	0.64	1.13

TABLE 2: The first three quartiles for F_n , $n = 1, ..., 10, \infty$.

n	Quartile		
	First	Second	Third
1	-0.67	0.00	0.67
2	-0.58	0.05	0.71
3	-0.58	0.05	0.71
4	-0.49	0.10	0.74
5	-0.42	0.15	0.77
6	-0.36	0.19	0.80
7	-0.29	0.22	0.82
8	-0.26	0.26	0.84
9	-0.21	0.28	0.86
10	-0.18	0.38	0.88
∞	0.27	0.63	1.12

References

- [1] ABELSON, R. (2001). Companies turn to grades, and employees go to court. New York Times (March 19), p. A1.
- [2] DE HAAN, L. (1970). On Regular Variation and Its Application to the Weak Convergence of Sample Extremes (Math. Centre Tracts 32). Mathematisch Centrum, Amsterdam.
- [3] FELLER, W. (1968). An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd edn. John Wiley, New York.
- [4] KRASNOSEL'SKIĬ, M. A. AND RUTICKIĬ, YA. B. (1961). Convex Functions and Orlicz Spaces. P. Noordhoff, Groningen.
- [5] TJOE, H. (2005). The distribution of performance scores of employees subject to annual pass-or-be-fired tests. Master's Thesis, Courant Institute, New York University.