4. Characterizations of Probability Distributions

A characterization is a certain distributional or statistical property of a statistic or statistics that uniquely determines the associated stochastic model. This chapter provides a brief survey of the huge literature on this topic. Characterizations based on random (complete or censored) samples from common univariate discrete and continuous distributions, and some multivariate continuous distributions are presented. Characterizations that use the properties of sample moments, order statistics, record statistics, and reliability properties are reviewed. Applications to simulation, stochastic modeling and goodness-of-fit tests are discussed. An introduction to further resources is given.

Suppose X is the population random variable (RV) with cumulative distribution function (CDF) $F(x) = \Pr(X \leq x)$ from which the data are generated according to a specified sampling scheme. Let $\mathcal{F}$ be the family of probability distributions used as an initial model to describe X and $\mathcal{F}_0$ be a subclass of $\mathcal{F}$ that is of interest to the modeler. If $T = T(X)$ is a statistic arising from F, and if a certain distributional property it possesses implies that $F \in \mathcal{F}_0$, then this property of $T(X)$ produces a characterization of $\mathcal{F}_0$. If the concerned property of $T(X)$ holds if and only if $F \in \mathcal{F}_0$, then we have a complete characterization of $\mathcal{F}_0$ and it is particularly helpful when this subfamily has a single member or members from a single parametric family of distributions. This is the essence of the abundant literature on this fascinating area of characterizations, produced by probabilists and mathematical statisticians, mostly over the past half-century.

In Sect. 4.1 we describe characterizing functions and their role in characterizations. Various types of data settings and associated characterizing properties discussed in the literature are reviewed in Sect. 4.2. A general classification of characterization results is given in Sect. 4.3. Treatment of continuous distributions begins with that of the exponential in Sect. 4.4, and continues with the normal in Sect. 4.5, and other distributions in Sect. 4.6. Characterizations for the Poisson distribution and Poisson process are given in Sect. 4.7, and Sect. 4.8 contains characterizations for other common discrete distributions. A brief look at the characterizations of multivariate distributions is provided in Sect. 4.9, where some characterizations based on conditional specification are presented. Special attention is drawn to the Marshall–Olkin bivariate exponential model and the multivariate normal distribution. Stability of characterization results is discussed in Sect. 4.10 and applications of characterizations in Sect. 4.11. The last section (Sect. 4.12) contains a listing of major resources.

4.1 Characterizing Functions

There are several functions associated with a probability distribution that uniquely identify it. We call these characterizing functions and describe a few of them.

4.1.1 Cumulative Distribution Function (CDF)

The CDF $F(x)$ of a RV X, defined for all real x, describes $P(X \in A)$ for any Borel set A on the real line. It is right-continuous, nondecreasing, and $F(-\infty) = 0$ and $F(+\infty) = 1$. If X is a discrete RV, $\Pr(X = x) = F(x) - F(x-) > 0$ for any possible value x. A closely associated function, the *survival function* (SF), is defined by

$$S(x) = \Pr(X \geq x) = 1 - F(x-)\,. \tag{4.1}$$

4.1.2 Probability Density Function (PDF)

The PDF $f(x)$ is a nonnegative function with the property that, for any Borel set A, $\Pr(X \in A)$ can be obtained by either summing or integrating $f(x)$ over A. When X is absolutely continuous, $f(x) = F'(x)$ almost everywhere (a. e.). In many cases, the PDF provides the most convenient way to describe the probability assignment.

4.1.3 Quantile Function

Also known as the inverse CDF, it is often defined as

$$F^{-1}(u) = \inf\{x : F(x) \geq u\},\, 0 < u < 1\,. \tag{4.2}$$

The quantile function is nondecreasing and, in this form, is left-continuous. It describes the probability assignment in terms of the quantiles of the distribution. When X is absolutely continuous, $F^{-1}(u)$ is differentiable with derivative $1/f[F^{-1}(u)]$.

The *probability integral transformation* $U = F(X)$ transforms any continuous RV X with CDF $F(x)$ into a standard uniform RV U that has PDF $f(u) = 1$, $0 < u < 1$. For *any* RV X with CDF $F(x)$,

$$X \stackrel{d}{=} F^{-1}(U)\,, \tag{4.3}$$

where $\stackrel{d}{=}$ stands for equality in distribution. This valuable distributional identity is helpful in simulating arbitrary RVs and also plays an important role in the theoretical developments associated with moments of X and of order statistics generated from random samples.

For a nonnegative RV with finite mean $\mu = \mathrm{E}(X) = \int_0^1 F^{-1}(u)\,\mathrm{d}u$, the quantile function yields the *Lorenz curve* used for describing the discrepancy in income distributions. The Lorenz curve is described by the function

$$L(t) = \frac{1}{\mu}\int_0^t F^{-1}(u)\,\mathrm{d}u\,, \quad 0 \leq t \leq 1\,. \tag{4.4}$$

It identifies F^{-1} and consequently F, up to the scale parameter.

4.1.4 Characteristic Function (CF) and Other Generating Functions

For any RV X, a CF exists and is defined by $\psi(t) = \mathrm{E}[\exp(\mathrm{i}tX)]$, where t is real and $\mathrm{i} = \sqrt{-1}$. This complex-valued function (it is real for F symmetric about 0) uniquely determines the CDF ([4.1], p. 104). It has played a major role in the proof of the central limit theorem and in numerous characterizations, especially of the normal distribution. An associated function, the moment generating function (MGF), given by $\mathrm{E}[\exp(tX)]$, may not exist for $t \neq 0$. If the MGF exists for t in a neighborhood of 0, then moments of X of all orders exist. Furthermore, the jth moment of X, $\mathrm{E}(X^j)$, can be obtained by a Taylor-series expansion of either the CF or the MGF of X. A one-to-one function of the MGF is its logarithm, the cumulant generating function. A discrete distribution with support over nonnegative integers is uniquely determined by its probability generating function, given by $\mathrm{E}\{s^X\}$, defined at least for $|s| \le 1$ ([4.1], p. 101).

4.1.5 Reliability Considerations

Several characterizing functions that are based on reliability properties have appeared in the area of life-testing experiments. There are many classes of distributions in the reliability literature that are based on the properties of such functions [4.2]. We introduce three such functions where we assume X to be nonnegative for practical reasons, and to be absolutely continuous for convenience.

Failure Rate and Hazard Functions

The failure (or hazard) rate function is defined for all x in the support of the RV X and is given by

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1-F(x)} , \qquad (4.5)$$

where $S(x)$ is the SF defined in (4.1), and in this case, equals $1 - F(x)$. Clearly,

$$H(x) \equiv \int_0^x h(w)\,\mathrm{d}w = -\log[1-F(x)] , \qquad (4.6)$$

and thus $h(x)$ uniquely determines $F(x)$. Further, $H(x)$, the integrated failure-rate function, also called the *hazard function*, identifies $F(x)$ as well, through the relationship $F(x) = 1 - \exp[-H(x)]$. The family of distributions for which the failure rate (FR) $h(x)$ is increasing (decreasing) is the IFR (DFR) family. If $H(x)/x$ is increasing (decreasing) one obtains the IFRA (DFRA) family since the average failure rate in that case will be increasing (decreasing).

Mean Residual Life (MRL) Function

The MRL is of practical interest in life-testing experiments and is defined whenever $\mathrm{E}(X)$ is finite. For a nonnegative X, with $t > 0$, it is given by

$$\begin{aligned} m(t) &\equiv \mathrm{E}(X - t | X > t) \\ &= \frac{1}{1-F(t)} \int_t^\infty [1 - F(w)]\,\mathrm{d}w , \end{aligned} \qquad (4.7)$$

if the CDF F is continuous. Equation (4.7) can be used to recover $F(x)$ from $m(t)$. In fact, when F is continuous, and $F^{-1}(0) = 0$,

$$1 - F(x) = \frac{m(0)}{m(x)} \exp\left(-\int_0^x \frac{\mathrm{d}y}{m(y)}\right) , x > 0 , \qquad (4.8)$$

where $m(0) = \mathrm{E}(X)$. Another closely related characterizing function is the truncated mean $\mathrm{E}(X|X > t) \equiv t + m(t)$, sometimes referred to as the conditional tail expectation. Increasing (decreasing) $m(t)$ produces the family of IMRL (DMRL) life distributions. Yet another characterizing function is the total time on test transform ([4.3], p. 91), defined as

$$\tau(t) = \int_0^{F^{-1}(t)} [1 - F(x)]\,\mathrm{d}x, 0 \le t \le 1 . \qquad (4.9)$$

This is concave if and only if (iff) F is IFR.

Characterization of a probability distribution then refers to the identification of any of $F(x)$, $F^{-1}(x)$, $f(x)$, $\psi(t)$, $h(x)$, $H(x)$, $m(t)$, $\tau(t)$, or of families for which these functions possess a certain property.

Example. Consider the exponential distribution with rate parameter λ whose PDF is given by

$$f(x) = \lambda\,\mathrm{e}^{-\lambda x}, x \ge 0 . \qquad (4.10)$$

If (4.10) holds, we say that X is an $\exp(\lambda)$ RV. For such an RV, the CDF is given by $F(x) = 1 - \exp(-\lambda x)$, $x \ge 0$; the quantile function is given by $F^{-1}(u) = -\log(1-u)/\lambda$, $0 \le u \le 1$; the characteristic function by $\psi(t) = 1/(1 - \mathrm{i}t/\lambda)$; the failure-rate function by $h(x) = \lambda, x \ge 0$; the hazard function by $H(x) = \lambda x$, $x \ge 0$; the MRL function by $m(t) = 1/\lambda$, $t \ge 0$; and the total time on test transform by $\tau(t) = t/\lambda$, $t \ge 0$.

4.2 Data Types and Characterizing Conditions

A characterization result is based on the assumed model for the data and the specified property of a particular statistic of interest. We now describe some common scenarios.

4.2.1 Data Models

The most basic data set consists of a single observation X from the CDF $F(x)$. Next, one may encounter either a random sample $X_1, \dots, X_n$ from $F(x)$, i. e., the X_j are independent, identically distributed (IID) RVs with CDF $F(x)$. Or the X_j are just independent and their CDFs belong to a family $\mathcal{F}_0$. Type II right-censored samples from (absolutely) continuous distributions that appear naturally in life-testing experiments are considered as well.

Let $X_{1:n} \le \cdots \le X_{k:n} \le \cdots \le X_{n:n}$ be the order statistics of the random sample [4.3]. Then the CDF of $X_{k:n}$ is given by

$$\sum_{j=k}^{n} \binom{n}{j} F^j(x)[1-F(x)]^{n-j} = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1}(1-t)^{n-k}\,\mathrm{d}t\,, \tag{4.11}$$

and the joint PDF of $X_{1:n}, \cdots, X_{k:n}$, a type II right censored sample, is given by

$$\frac{n!}{(n-k)!} f(x_1)\cdots f(x_k)[1-F(x_k)]^{n-k}, \quad x_1 < \cdots < x_k\,. \tag{4.12}$$

Models that use the properties of upper or lower record values for characterizations do exist. For an infinite sequence $\{X_j, j \ge 1\}$ from a continuous parent, for $j \ge 1$, X_j is called an *upper record value* of this sequence if $X_j = \max(X_1, \dots, X_j)$ [4.4]. By convention, X_1 is the first upper record value (or reference value). The joint PDF of the first m upper record values, $R_1, \dots, R_m$, is given by

$$f(r_1, \dots, r_m) = f(r_m) \prod_{j=1}^{m-1} h(r_j) \quad r_1 < \cdots < r_m\,, \tag{4.13}$$

where $h(x)$ is given by (4.5). From (4.12) and (4.13), we can, respectively, conclude that (see, e.g., [4.4], p. 114), for a continuous parent

$$\Pr(X_{k+1:n} > y | X_{k:n} = x) = [\Pr(X > y | X > x)]^{n-k}, x < y\,, \tag{4.14}$$

and

$$\Pr(R_{m+1} > y | R_m = x) = \Pr(X > y | X > x) = \frac{1-F(y)}{1-F(x)}, x < y\,. \tag{4.15}$$

4.2.2 Characterizing Conditions

There are many types of characterizing properties. For example, the property could be based on identical distribution of two statistics, or independence of them, or on constancy of regression of one on the other. Alternatively, the characterization may be based on properties of admissibility and optimality of certain estimators based on random samples [4.5]. Characterizations based on fixed or random sample sizes and on damaged/missing observations do exist. Basically, the assumed condition is shown to yield a functional equation satisfied by a characterizing function, leading to the identification of the parent distribution. Other conditions involve recurrence relations satisfied by the ordinary moments, or moments of order statistics. There are characterizations based on certain inequalities for the moments, Fisher information or entropy measures, where equality holds for a unique family of distributions.

4.2.3 General Techniques

For the normal distribution, identification via the characteristic function is the common approach. The proofs involve results on complex variables. Numerous results for the exponential or geometric distributions have been based on the SF and rely on the Cauchy functional equation (CFE) ([4.6], Sect. 1.1), namely,

$$g_0(x+y) = g_0(x)g_0(y), \text{ for all } x, y \ge 0\,. \tag{4.16}$$

The only continuous solution is the exponential SF. The well-known lack-of-memory property (LMP) of the exponential distribution, namely

$$\Pr(X > x + y | X > x) = \Pr(X > y), \quad \text{for all } x, y \ge 0\,, \tag{4.17}$$

leads to (4.16). In (4.17), the exponential SF $\Pr(X > x) = \exp(-\lambda x)$ is recovered even if the equation holds for all $x \ge 0$ and for two y values y_1, y_2 such that y_1/y_2 is irrational.

In recent years, the integrated CFE (ICFE) whose solution goes back to Deny's Theorem ([4.6, 7], Chapt. 2) has been used for several characterizations of the exponential and geometric distributions. We say that g_1 satisfies an ICFE if it is a function defined on $[0, \infty)$ and satisfies the condition (4.18) (see below). The following characterization result due to *Lau* and *Rao* [4.8] is taken from [4.7] (p. 29).

Theorem 4.1
Let g be a nonnegative locally integrable function on $[0, \infty)$ that is not a function identically equal to 0 a.e. ν_0, the Lebesgue measure defined on $[0, \infty)$. Suppose g satisfies (4.18) given below where ν is a σ-finite measure

satisfying the condition $\nu(\{0\}) < 1$:

$$g(x) = \int_0^\infty g(x+y)\,\mathrm{d}\nu(y), \qquad \text{a.e. } \nu_0 \text{ for } x \ge 0\,. \tag{4.18}$$

Then either

$$g(x+n\delta) = g(x)b^n,\, n \ge 0, \qquad \text{a.e. } \nu_0 \text{ for } x \ge 0\,, \tag{4.19}$$

where b is such that $\sum_{n=0}^\infty b^n \nu(\{n\delta\}) = 1$ for some $\delta > 0$, or

$$g(x) \propto \mathrm{e}^{\alpha x} \quad \text{a.e. } \nu_0 \text{ for } x \ge 0\,, \tag{4.20}$$

where α satisfies the condition $\int_0^\infty \mathrm{e}^{\alpha y}\,\mathrm{d}\nu(y) = 1$.

If g_0 satisfying (4.16) is integrable with respect to a σ-finite measure ν_1 on $[0,\infty)$, the integral being a positive quantity c, then it also satisfies (4.18) with $\nu = c^{-1}\nu_1$.

The above result has provided elegant unified proofs of several characterizations of exponential and geometric distributions. See the review in [4.9] and the monograph [4.7]. A nice account of the role of functional equations in probability theory and in characterization theorems is given in [4.6].

A handy technique is based on complete families of functions. Suppose $\{f_n(x),\ n \ge 1\}$ is a sequence such that if, for any integrable function $g(x)$, the condition

$$\int_A g(x) f_n(x)\,\mathrm{d}x = 0 \qquad \text{for all } n \ge 1 \tag{4.21}$$

implies that $g(x) = 0$ a.e. on A, where A is an interval. Then we say that the sequence of functions $f_n(x)$ is complete on A. For example, $f_n(x) = x^n$ is complete on [0, 1] and leads to the characterization of F based on the sequence $\mathrm{E}(X_{n:n})$. A classical completeness result due to Müntz and Szász shows that even an appropriately chosen subsequence would do. (See [4.10] for a good summary.) The *method of intensively monotone operators* is another general method of showing uniqueness of solutions to the functional equations generated by the characterizing conditions [4.11].

An approach involving inequalities such as the Cauchy–Schwarz (or Cramér–Rao) has produced some characterizations. The distribution being characterized corresponds to the case where equality holds. For example,

$$\mathrm{Cor}(X_{j:n}, X_{k:n}) \le \sqrt{\frac{j(n-k+1)}{k(n-j+1)}},\ j<k\,, \tag{4.22}$$

where "Cor" represents the correlation coefficient, and equality holds iff F is a uniform CDF.

The *method of limit laws*, elaborated in [4.12], is found to be helpful in establishing some characterizations.

4.3 A Classification of Characterizations

We attempt below a general classification of characterization results.

4.3.1 Uniqueness Conditions

These are properties that provide a one-to-one correspondence with the parent CDF F. For example, from (4.11) it is clear that the CDF of $X_{k:n}$ for any fixed k and n identifies $F(x)$, and thus provides a trivial characterization.

A more interesting question is considered in the *classical moment problem* [4.13]. It is concerned with the determination of the CDF $F(x)$ from the sequence of population moments $\{\mathrm{E}(X^j),\ j \ge 1\}$, which are assumed to exist. Two distinct distributions can have the same moment sequence. Under certain conditions, however, the associated CDF is unique. One such condition is that

$$\sum_{j=1}^\infty \frac{\mathrm{E}(X^j)}{j!} t^j \tag{4.23}$$

is absolutely convergent for some $t > 0$ ([4.1], p. 106).

In the context of order statistics, the moment problem has the goal of identifying F based on the moment sequence $\{\mathrm{E}(X_{k(n):n}),\ 1 \le k(n) \le n,\ n \ge 1\}$. There is no loss of generality in choosing $k(n)$ [or $n-k(n)$] to be a constant. It is known that the subsequence $\{\mathrm{E}(X_{n_j:n_j}),\ n_1 < n_2 < \cdots,\ j \ge 1\}$ characterizes F provided that $\sum_{j=1}^\infty 1/n_j$ diverges [4.14]. Such a subsequence will not suffice when we consider $\mathrm{E}(R_n)$—the moment sequence of record values. One needs the entire sequence and identification can be achieved only within the family of continuous distributions ([4.4], Sect. 4.2.1).

It is clear from (4.15) that $E(R_{n+1} - R_n | R_n = x) = m(x)$ a.s., and hence, in view of (4.8), one can recover $F(x)$ from the regression function $E(R_{n+1}|R_n)$. Similarly $E(X_{k+1:n}|X_{k:n})$ or even $E[h(X_{k+1:n})|X_{k:n}]$ where $h(.)$ is a real, continuous and strictly monotonic function, would identify F in the family of *arbitrary* distributions [4.15].

4.3.2 Characterizations of Families of Distributions

There are characterizations of distributions defined by their reliability properties, of the exponential-family distributions, or of spherically symmetric (multivariate) distributions, and other large families. Such results identify the general properties possessed by the members without actually providing an explicit form for the CDF or other characterizing functions. For example, certain properties of order statistics characterize various classes of life distributions [4.16]. A specific characterization of this sort is the following. Among nonnegative distributions with finite mean, a CDF F is IFR iff $E[(n-k+1)(X_{k:n} - X_{k-1:n})]$ is decreasing in k $(2 \le k \le n)$ for infinitely many n. A one-parameter exponential family is characterized (under some conditions) by the fact that, for a single random sample of size $n \ge 3$, the sample mean $\bar{X}$ is the maximum-likelihood estimator (MLE) of the mean of the distribution (see also [4.17]). Such results are interesting theoretically and provide some reality checks for the model assumptions.

4.3.3 Characterizations of Specific Parametric Families

Of maximal interest in terms of applications and creation of goodness-of-fit tests are the characterizations that identify specific parametric family of distributions. In subsequent sections we will discuss a few such results for some common distributions; in fact, we consider only a tiny subset of the voluminous, ever-growing literature. We only list the characterization results with no details on the other needed conditions, some of which could be very technical and hard to verify. We refrain from proving any of our claims.

4.4 Exponential Distribution

Arnold and *Huang* [4.18], in their survey of characterizations of the exponential distribution, mentioned about 275 citations 10 years ago, and one could safely add 50 more to the list, making it perhaps the most popular distribution on this topic. In recent years results based on the properties of order statistics and record values have been in vogue. In most cases, the condition imposed will result in the CFE or the ICFE discussed in Sect. 4.2. Apart from the exponential characterizations stemming from the general results [e.g., if $E(X_{1:n}) = \lambda/n$, $n \ge 1$, then F must be an $\exp(\lambda)$ CDF], the conditions imposed could be based on: (i) truncated moments or regression functions associated with order statistics or (upper) records, (ii) distributional identities among order statistics, their spacings, or the spacings of records, (iii) independence of certain linear functions of order statistics or records, (iv) reliability properties of order statistics or records, or (v) *geometric compounding*, where the RV of interest is a random sum of IID RVs where the number in the sum is determined by an independent geometric RV.

Consider the order statistics case. For the $\exp(\lambda)$ parent: (a) $X_{j:n}$ and $X_{k:n} - X_{j:n}$, $j < k$, are independent, (b) $nX_{1:n} \stackrel{d}{=} X_1$, $n \ge 2$, and the normalized spacings, $Y_j = (n-j+1)(X_{j:n} - X_{j-1:n})$, $j = 1, \ldots, n$ (with $X_{0:n} = 0$), have the following properties: (c) the Y_j are independent, (d) they are identically distributed as X_1 for all $j = 1, \ldots, n$, and (e) $E(Y_j) = 1/\lambda$, $1 \le j \le n$. Under some mild conditions each of these provides a characterization of the exponential CDF. In (b), it is known that, if the distributional equality holds for two values of n, say n_1 and n_2 such that $\log n_1 / \log n_2$ is irrational, then the exponential RV is identified in the family of nonnegative RVs.

If F is an exponential CDF, from the joint PDF of the record values given in (4.13), it can be shown that, for $j \le m < n$:

1. R_m and $R_n - R_m$ are independent,
2. R_j and $R_n - R_m$ are independent,
3. $E(R_{n+1} - R_n | R_n)$ does not depend on R_n,
4. $\mathrm{Var}(R_{n+1} - R_n | R_n)$ does not depend on R_n,
5. $R_n - R_m$ and R_{n-m} are identically distributed,
6. $E((R_n - R_m)^s | R_j)$ does not depend on R_j,
7. $E(R_n) = n/\lambda$, and
8. $E(R_{n+1}|R_n)$ and $E(R_n|R_{n+1})$ are both linear in conditioning variables.

Each of these provides a characterization of the exponential distribution (possibly with a location shift) in an appropriately chosen $\mathcal{F}$ (such as continuous CDF). There are many more that dwell on the properties of the spacings of upper record values or order statistics.

Using (4.13) and (4.14), characterizations involving record or order statistic spacings can be linked to those based on truncated distributions. For example, assuming that $\mathrm{E}(R_{n+1} - R_n | R_n)$ does not depend on R_n is basically the same as saying $\mathrm{E}(X - x | X > x)$ is free of x. This means that the MRL is a constant—a characterizing property of the exponential distribution. Conditions such as $X_{j+1:n} - X_{j:n} \stackrel{d}{=} X_{1:n-j}$ or $R_n - R_m \stackrel{d}{=} R_{n-m}$ result in the ICFE (see Theorem 4.1), yielding an exponential characterization in the family of continuous CDFs.

In the context of life-testing experiments, the statistic representing the *total time on test* by time $X_{i:n}$, $T_{i,n} = \sum_{j=1}^{i} Y_j = \sum_{j=1}^{i} X_{j:n} + (n-i) X_{i:n}$, is of considerable interest. For an exponential parent, for $2 \leq k \leq n$,

$$\left(\frac{T_{1,n}}{T_{k,n}}, \ldots, \frac{T_{k-1,n}}{T_{k,n}}\right) \tag{4.24}$$

behaves like the vector of order statistics from a random sample of size $k-1$ from the standard uniform distribution. The converse is shown to be true assuming that the order-statistic property holds for some k and n such that $5 \leq k \leq n$ [4.9].

There is a characterization of the exponential distribution based on records that is similar to that of the uniform distribution based on the order statistics [see (4.22)]. If $\mathrm{Var}(R_m)$ and $\mathrm{Var}(R_n)$ are both finite, $\mathrm{Cor}(R_m, R_n)$ does not exceed $\sqrt{m/n}$, $m < n$. Furthermore, the upper bound is attained if F is an exponential CDF, possibly with a location shift.

Numerous characterizations of the exponential distribution exist when $\mathcal{F}$ is restricted to families defined by the reliability properties (such as the *new better/worse than used* families). Characterizations of the exponential distribution arising from queueing models are rare; for a few results, see [4.19].

The survey of characterizations using order statistics by *Gather* et al. [4.20] contains an excellent compendium of results for the exponential distribution based on their properties. Other important references include the monograph by *Azlarov* and *Volodin* specializing on exponential characterizations [4.21], the survey [4.18] mentioned earlier, and [4.22], Chapt. 19.

Remarks. The memoryless property (LMP) of the exponential distribution is also shared by the geometric distribution, which can be extracted as $[X]$ where $[\cdot]$ represents the greatest integer function. This results in many parallel characterizations for the geometric parent among distributions on nonnegative integers (see Sect. 4.8; [4.23]). Homogeneous Poisson processes, characterized by the fact that the inter-arrival times of the events are IID exponential, can be identified by the characterizing properties of the exponential distributions. Also, for any RV X with continuous CDF F, $-\log[1 - F(X)]$ is standard exponential and thus exponential characterizations naturally lead to characterizations of such distributions. For example, results based on exponential spacings will lead to the characterizations of the uniform, power-function or Pareto distributions on the basis of the properties of the ratios of order statistics.

4.5 Normal Distribution

The earliest characterization results were for the normal distribution. In 1923 Pólya showed that, if X_1 and X_2 are IID with finite variance, and $X_1 \stackrel{d}{=} a_1 X_1 + a_2 X_2$, then the X_j must necessarily be normal. *Cramér* showed in 1937 that, if X_1 and X_2 are independent and the sum is assumed to be normal, then each of them must be normal (see [4.24], p. 53). *Skitovich* and *Darmois* established in 1953 the following result:

Theorem 4.2
If $X_1, \ldots, X_n$ are independent and the linear functions

$$L_1 = \sum_{j=1}^{n} a_j X_j, \; L_2 = \sum_{j=1}^{n} b_j X_j \tag{4.25}$$

are independent, then the RVs X_j for which $a_j b_j \neq 0$ must all be normal.

The work on normal characterizations accelerated from the 1950s and it was an extremely active area of research until the late 1970s. Extensive accounts are available, particularly in the influential book on characterizations by *Kagan* et al. [4.5], and also in the monograph on the applications of characteristic functions by *Lukacs* and *Laha* [4.25], all major contributors to the area. A short book devoted to normal characterizations by *Mathai* and *Pederzolli* [4.26] also provides a good account. An excellent brief summary of the normal characterizations and an extensive reference list is provided in [4.22], Chapt. 13. Another recent account of normal characterizations is provided by the monograph [4.27]. We now record a listing of assorted types of characterizations, mostly those that are easy to describe. These hold under usually mild conditions. Many normal characterizations hold under technical conditions that are hard to verify and often it is difficult to assess their practical implications.

Characterizations

We begin with characterizations based on the properties of a single observation from the parent distribution.

Among absolutely continuous distributions with support $(-\infty, \infty)$ and specified mean and variance, the entropy $-\int_{-\infty}^{\infty} f(x) \log f(x)\, dx$ is the largest for the normal distribution ([4.5], p. 410).

Assume $\mathcal{F}$ is the location family of absolutely continuous distributions with support $(-\infty, \infty)$ and differentiable density $f(x-\theta)$, where θ is the location parameter. Then the Fisher information in this location family is given by

$$I_f(\theta) = \int\limits_{f(x)>0} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\, dx\,. \tag{4.26}$$

The distribution in this family with the smallest $I_f(\theta)$ is normal ([4.5], p. 406).

The following properties based on a random sample $X_1, \dots, X_n$ of size n, characterize the normal parent:

1. X_1 has zero mean and unit variance, and for some $a_1, a_2 \neq 0$, $(a_1 X_1 + a_2 X_2)^2/(a_1^2 + a_2^2) \stackrel{d}{=} (a_1 X_1 - a_2 X_2)^2/(a_1^2 + a_2^2) \stackrel{d}{=} \chi^2_{(1)}$ ([4.22], p. 105).
2. X_1 has zero mean and is nondegenerate and with $ab = -1$ and $n = 2$, $\mathrm{E}(X_1 + aX_2 | X_1 + bX_2) = \mathrm{E}(X_1 + bX_2 | X_1 + aX_2) = 0$ ([4.5], p. 158).
3. The sample mean $\bar{X}$ and sample variance S^2 are independent for some $n \geq 2$ ([4.5], p. 103).
4. For some $n \geq 5$, the vector
$$\left(\frac{X_1 - \bar{X}}{(n-1)S^2}, \dots, \frac{X_n - \bar{X}}{(n-1)S^2}\right) \tag{4.27}$$
is uniformly distributed on the $(n-2)$-dimensional sphere $\left\{(w_1, \dots, w_n) : \sum_{i=1}^n w_i = 0, \sum_{i=1}^n w_i^2 = 1\right\}$ ([4.7], p. 142).
5. $\mathcal{F}$ is the location family of distributions $F_0(x-\theta)$, with mean θ, the location parameter, and finite variance, and, for some $n \geq 3$, $\bar{X}$ is admissible under squared error loss among all unbiased estimators of θ ([4.5], p. 228).
6. In the above location family (with finite variance and θ as the mean), take the null hypothesis $H_0 : \theta = 0$ and the alternative $H_1 : \theta > 0$. Suppose that, for some $n \geq 3$, the critical region $\{\bar{X} > c_\alpha\}$ is the uniformly most powerful among all tests with level of significance $\leq \alpha$ for all $\alpha \in (0, 1)$, where $c_\alpha = \max\{c : \Pr(\bar{X} > c | \theta = 0) = \alpha\}$ ([4.5], p. 451).
7. In the above location family with θ as the mean, $\bar{X}_n$ is the MLE of θ for $n = 2, 3$ ([4.5], p. 411).
8. $\mathcal{F}$ is the location-scale family of continuous distributions $F_0[(x-\theta)/\sigma]$, θ real, $\sigma > 0$, and $(\bar{X}, S)$ is a sufficient statistic for (θ, σ) ([4.22], p. 106).
9. In the above location-scale family with mean θ, $\bar{X}_n$ is a best linear unbiased estimation (BLUE) of θ for *all* $n \geq 1$ [4.28].
10. In the linear regression model under the Bayesian framework, the distributions with posterior expectation is linear in data values [4.29].

When the X_j are independent RVs, some of the above results have appropriate generalizations. For example,

1. In (4.25), if the sequences $\{a_j/b_j\}$ and $\{b_j/a_j\}$ for which $a_j b_j \neq 0$ are bounded and as $n \to \infty$ L_1 and L_2 converge with probability 1 to independent RVs, then the X_j for which $a_j b_j \neq 0$ are normal ([4.5], p. 94).
2. If $X_1 \stackrel{d}{=} L_1$ in (4.25), where $a_1 \neq -1, 0, 1$, $n \geq 2$, and the X_j have finite variance, then X_1 is normal ([4.22], p. 104).
3. Suppose X_1, X_2, X_3 are independent symmetric RVs with median 0 and CDF that are continuous at 0. If $W_1 = X_1/X_3$, $W_2 = X_2/X_3$ have joint characteristic function $\mathrm{E}[\exp(\mathrm{i}t_1 W_1 + \mathrm{i}t_2 W_2)] = \exp(t_1^2 + t_2^2)^{1/2}$, then the X_j are IID normal ([4.22], p. 104). (Here W_1, W_2 are identically distributed Cauchy RVs.)

Remarks. Some of the above results have parallel versions that characterize the gamma populations when the support is restricted to positive values or when $\mathcal{F}$ is the scale family distribution. Some provide characterizations of the Poisson distribution when the support is restricted to nonnegative integers.

4.6 Other Continuous Distributions

4.6.1 Uniform

As noted earlier, some characterizations of the exponential lead to similar results for the uniform. In addition, several characterizations based on the properties of order statistics and sums of IID RVs do exist. We have noted earlier [see (4.18)] that the maximum correlation for any two order statistics is attained only for the uniform distribution. Below are some other conditions that characterize the uniform distribution over an interval whose endpoints are chosen appropriately. (See [4.30], p. 282–285 for original references.)

1. $\mathrm{Cov}(X_{1:2}, X_{2:2}) = (1/3)\mathrm{Var}(X)$ or equivalently $\mathrm{E}(X_{2:2} - X) = [(1/3)\mathrm{Var}(X)]^{1/2}$.
2. $[\mathrm{E}(X_{2:2})]^2 = (4/3)\mathrm{E}(X^2)$. Here and above, the right side represents the maximum possible value for the left side expression.
3. $X_{2:2} - X_{1:2} \overset{d}{=} X_{1:2}$.
4. For some $n \geq 3$, $\mathrm{E}(X_1 | X_{1:n}, X_{n:n}) = (1/2)(X_{1:n} + X_{n:n})$ almost surely and F is continuous.

4.6.2 Gamma

The following conditions characterize the gamma distribution, whose PDF is given by

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \exp(-x/\beta) x^{\alpha-1}, x > 0\,, \tag{4.28}$$

where $\alpha > 0$ and $\beta > 0$ are the shape and scale parameters, respectively ([4.22], p. 350–354; [4.5], p. 407–410).

1. X_1 and X_2 are independent nondegenerate RVs, and $X_1 + X_2$ and X_1/X_2 are independent.
2. For independent RVs $X_1, \ldots, X_n$, and $T = \sum_{j=1}^n X_j$, $\left(\frac{X_1}{T}, \ldots, \frac{X_n}{T}\right)$ and T are independent.
3. In a random sample of size n from a positive RV X for which $\mathrm{E}(1/X)$ is finite, the conditional expectation $\mathrm{E}(\sum_{j=1}^n X_j^{-1} | X_1 - \bar{X})$ is a constant.
4. The distribution in the scale family of distributions with support $(0, \infty)$ that has the smallest attainable Fisher information measure.
5. The distribution with the maximum entropy among distributions with support $(0, \infty)$ and have specified $\mathrm{E}(X)$ and $\mathrm{E}(\log(X))$.

4.6.3 Weibull

Some exponential characterizations easily lead to characterizations of the Weibull distribution with the PDF

$$f(x) = \frac{\alpha x^{\alpha-1}}{\beta^\alpha} \mathrm{e}^{-(x/\beta)^\alpha}, x > 0\,, \tag{4.29}$$

where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the scale parameter. This is also a distribution that is *min-stable* and is one of the extreme-value distributions to which the sample minima from random samples may converge. Further, each of the following properties characterize it ([4.22], Sect. 21.9).

1. X_1, X_2 are independent nonnegative RVs and for some $a, b \in (0, 1)$
$$\min(X_1, X_2) \overset{d}{=} aX_1 \overset{d}{=} bX_2\,. \tag{4.30}$$
2. $\mathrm{Var}(X^\alpha_{k+1:n} | X_{k:n} = x)$ is a constant.
3. $X_1, \ldots,$ are IID and N is an independent RV with support $\{2, 3, \ldots, \}$, and $N^{1/\alpha} X_{1:N} \overset{d}{=} X_1$.
4. In the scale family of distributions, the distribution for which the Fisher information in the right-censored sample $X_{1:n}, \ldots, X_{k:n}$ is the same as in a random sample of size k for all n and $k \leq n$ ([4.3], p. 226).

4.6.4 Gumbel and Other Extreme-Value Distributions

For an arbitrary parent distribution, if $X_{n:n}$ has a nondegenerate limiting distribution, possibly after suitable standardization, then it is known that the limiting CDF has one of the following forms, except for a change of location and scale:

$$\text{(Fréchet)} \quad \begin{aligned} G_1(x;\alpha) &= 0 \quad x \leq 0, \alpha > 0\,, \\ &= \exp(-x^{-\alpha}) \quad x > 0\,; \end{aligned} \tag{4.31}$$

$$\text{(Weibull)} \quad \begin{aligned} G_2(x;\alpha) &= \exp\left[-(-x)^\alpha\right] \\ &\qquad x \leq 0, \alpha > 0\,, \\ &= 1 \quad x > 0\,; \end{aligned} \tag{4.32}$$

$$\text{(Gumbel)} \quad \begin{aligned} G_3(x) &= \exp\left(-\mathrm{e}^{-x}\right) \\ &\qquad -\infty < x < \infty\,. \end{aligned} \tag{4.33}$$

The CDF $G_2(x;\alpha)$ above is that of $-X$ where X has the Weibull PDF given in (4.29). These distributions form the class of *max-stable* distributions characterized by the following result ([4.31], p. 38).

Theorem 4.3
Let F be a nondegenerate CDF and a_n and b_n be two sequences such that $(X_{n:n}-a_n)/b_n \overset{d}{=} X_1$ for integers n_1 and n_2 such that $b_{n_2} \neq 1$ and $\log(b_{n_1})/\log(b_{n_2})$ is irrational. Then there are constants $a, b > 0$ and $\alpha > 0$ such that $F(a+bx)$ is either (4.31) or (4.32). If $b_{n_2} = 1$ and a_{n_1}/a_{n_2} is irrational, then $F(a+bx)$ is given by (4.33).

Additional results for the Gumbel distribution are available. Here are two characterizing properties.

1. $\mathrm{E}(R'_n - R'_{n+1}|R'_{n+1})$ is a constant for some n, where the R'_n is the nth lower record value.
2. $R'_n - R'_m$ and R'_m are independent for some $n < m$.

4.6.5 Pareto

The Pareto distribution, used as a model for income distributions and as the limit distribution of residual lifetime, has the CDF

$$F(x) = 1 - \left\{1 + [(x-\mu)/\sigma]^{1/\gamma}\right\}^{-\alpha}, \quad x \geq \mu, \sigma > 0, \gamma > 0, \alpha > 0\,. \tag{4.34}$$

Several characterizations exist for this distribution, especially when $\gamma = 1$, in which case the resulting distribution is called the Pareto distribution of the second kind (II); one obtains the Pareto distribution of the first kind (I) when, in addition, $\mu = \sigma$. (See [4.32], Sect. 3.7. or [4.22], Sect. 21.9.) Since $-\log[1-F(X)] = \alpha \log\{1 + [(X-\mu)/\sigma]^{1/\gamma}\}$ is the standard exponential, its numerous characterizations provide simple counterparts for the Pareto distribution. For example, independence of exponential spacings is equivalent to the independence of the ratios of order statistics from a Pareto I distribution. The Pareto II (with $\alpha > 1$) is the only distribution for which $h(x)m(x)$ is a constant where the failure rate $h(x)$ and the MRL function $m(x)$ are given by (4.5) and (4.7), respectively. Another characterization is that, if W is a continuous RV with support (0, 1) and is independent of X, and $\{WX|WX \geq \mu\} \overset{d}{=} X$, then X must be a Pareto II RV. Here W can be seen as the proportion *underreported* or *undamaged*. For some recent results on generalized Pareto distributions, see [4.33].

4.6.6 Inverse Gaussian (IG)

This distribution has the PDF

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda}{2\nu^2}\frac{(x-\nu)^2}{x}\right), \quad x > 0\,, \tag{4.35}$$

where the parameters are $\nu > 0$ and $\lambda > 0$. It arises as the waiting time to cross a certain threshold in Brownian motion. Its characterizations often mimic those for the normal distribution. For example, the IG distribution has the maximum entropy subject to certain restrictions on $\mathrm{E}(X)$ and $\mathrm{E}(1/X)$. For a random sample $X_1, \ldots, X_n$, let $Y = (1/n)\sum_{i=1}^n X_i^{-1} - \bar{X}^{-1}$. Then the population is IG if either $\bar{X}$ and Y are independent, or the regression $\mathrm{E}(Y|\bar{X})$ is a constant ([4.34], Chapt. 3).

4.7 Poisson Distribution and Process

The Poisson distribution, commonly known through its PDF

$$\Pr(X = j) = \mathrm{e}^{-\lambda}\frac{\lambda^j}{j!}, \quad j = 0, 1, 2, \ldots; \lambda > 0\,, \tag{4.36}$$

appears often in the engineering literature as a model for rare events and in queueing or reliability studies. We write X is Poi(λ) if (4.36) holds. The earliest result seems to be due to *Raikov* (1938) ([4.35], Sect. 4.8) who showed that, if X_1 and X_2 are independent and $X_1 + X_2$ is Poisson, then each of them should be Poisson RVs. It is known that, if X_1 is Poi(λ_1) and X_2 is Poi(λ_2) and they are independent, the conditional distribution of X_1 given $X_1 + X_2 = n$ is Bin$[n, \lambda_1/(\lambda_1+\lambda_2)]$, i.e., binomial with n trials and success probability $p = \lambda_1/(\lambda_1+\lambda_2)$. This property has led to various characterizations of the Poisson distribution. For example, if the conditional distribution is binomial, then X_1 and X_2 are both Poisson RVs.

One particularly interesting set up that identifies the Poisson distribution is the *damage model* due to *Rao*. The associated characterization result due to *Rao* and *Rubin* is the following ([4.7], p. 164):

Theorem 4.4
Let X and Y be nonnegative integer-valued RVs such that $\Pr\{X = 0\} < 1$ and, given $X = n$, Y is $\text{Bin}(n, p)$ for each $n \geq 0$ and a fixed $p \in (0, 1)$. Then the Rao–Rubin condition, given by

$$\Pr(Y = j) = \Pr(Y = j|Y = X), \quad j = 0, 1, \ldots , \tag{4.37}$$

holds iff X is Poisson.

The condition (4.37) is equivalent to the condition

$$\Pr(Y = j|Y = X) = \Pr(Y = j|Y < X) \tag{4.38}$$

which can be interpreted as follows. Suppose X is the number of original counts and Y is the number actually available, the remaining being lost due to damage according to the binomial model. Then if the probability distribution of the actual counts remains the same whether damage has taken place or not, the number of original counts must be Poisson. Incidentally, the number of observations that survived is also Poisson. The above damage model can be seen as binomial *splitting* or *thinning* and a similar notion is that of binomial *expanding*. It also leads to a characterization of the Poisson distribution.

A weaker version of (4.37) can be used to characterize the Poisson distribution by restricting the family $\mathcal{F}$ under consideration. Let X belong to the family of the power-series distribution, i. e., it has the PDF

$$\Pr(X = j) = \frac{a_j \theta^j}{A(\theta)}, \quad j = 0, 1, \ldots \tag{4.39}$$

Suppose that given $X = n$, Y has support $0, \ldots , n$, and has mean np and variance $np(1 - p)$, where p does not depend on θ. Then $E(Y|Y = X) = E(Y)$ and $\text{Var}(Y|Y = X) = \text{Var}(Y)$ iff X is Poisson.

Poisson characterizations based on the properties of the sample mean $\bar{X}$ and variance S^2 from a random sample are known. In the power-series family (4.39), if $E(S^2|\bar{X} > 0) = 1$, then the population is necessarily Poisson. When X is assumed to be nonnegative, if $E(S^2|\bar{X}) = \bar{X}$, the parent is Poisson also. See [4.35], Sect. 4.8, for relevant references. Characterizations based on the discrete analogue of the Skitovich–Darmois theorem (Theorem 4.2) are available [4.36]. It is also known that, in a wide class of distributions on the set of integers, the Poisson distribution is characterized by the equality sign in a discrete version of the Stam inequality for the Fisher information; the continuous version yields a normal characterization [4.37]. Another *normal-like* result is the Poisson characterization by the identity $E(X)E[g(X + 1)] = E[Xg(X)]$ assumed to hold for every bounded function $g(.)$ on the integers [4.38].

Poisson Process

A renewal process is a counting process $\{N(t), t \geq 0\}$ where the inter-arrival times of events are IID with CDF F. The (homogeneous) Poisson process is characterized by the fact that F is exponential. Several characterizations of a Poisson process in the family of renewal processes do exist. For example, if a renewal process is obtained by the superposition of two independent renewal processes, then the processes must be Poissonian. Several are tied to the exponential characterizations from random samples. Other characterizations of interest are based on the properties of the *current age* and *residual lifetime* distributions.

Let X_i represent the IID inter-arrival times and $S_n = X_1 + \cdots + X_n$, so that $N(t) = \sup(m : S_m \leq t)$. Then $A(t) = t - S_{N(t)}$ represents the current age or *backward recurrence time* at t and $W(t) = S_{N(t)+1} - t$ is the residual lifetime or *forward recurrence time* at t, $t \geq 0$. A good summary of the available results is provided in [4.39], p. 674–684. Chapter 4 of [4.31] contains an early account of various characterizations of the Poisson process that include thinned renewal processes and geometric compounding. We state below a few simple characterizing properties of the Poisson process:

1. Either $E[W(t)]$ or $\text{Var}[W(t)]$ is a finite constant for all $t > 0$.
2. F is continuous with $F^{-1}(0) = 0$, and for some fixed t, $A(t)$ and $W(t)$ are independent.
3. F is continuous and $E[A(t)|N(t) = n] = E[X_1|N(t) = n]$ for all $t > 0$ and all $n \geq 1$.
4. F is continuous and $E[A(t)] = E[\min(X_1, t)]$ for all $t > 0$.

4.8 Other Discrete Distributions

4.8.1 Geometric

Numerous versions of the LMP of the geometric distribution have led to several characterizations of the geometric distribution with PDF

$$\Pr(X = j) = (1-p)^j p, \; j = 0, 1, \dots \qquad (4.40)$$

Here the LMP means $\Pr(X > x + j | X \geq x) = \Pr(X > j)$, $j, x = 0, 1, \dots$ When X has the above PDF, the following properties hold:

1. $\mathrm{E}(X - x | X \geq x) = \mathrm{E}(X)$, $x = 0, 1, \dots$.
2. $|X_1 - X_2| \stackrel{d}{=} X$.
3. $\Pr(X_{1:n} \geq 1) = \Pr(X_1 \geq n)$, $n \geq 1$.
4. $X_{j+1:n} - X_{j:n} \stackrel{d}{=} X_{1:n-j}$, $1 \leq j < n$.
5. $(X_{k:n} - X_{j:n} | X_{j+1:n} - X_{j:n} > 0) \stackrel{d}{=} 1 + X_{k-j:n-j}$, $1 \leq j < k \leq n$.
6. $X_{1:n}$ and $X_{j:n} - X_{1:n}$ are independent.

Each of these is shown to be a characteristic property of the geometric or slightly modified versions of that distribution, under mild conditions [4.40].

In terms of the upper record values, the following properties hold for the geometric parent and characterize it ([4.4], Sect. 4.6).

1. $R_1, R_2 - R_1, R_3 - R_2, \dots$ are independent.
2. $E(R_{n+1} - R_n \mid R_n)$, $E(R_{n+2} - R_{n+1} \mid R_n)$, and $E[(R_2 - R_1)^2 \mid R_1]$ are constants.
3. $R_{n+1} - R_n \stackrel{d}{=} R_1$, $n \geq 1$.

4.8.2 Binomial and Negative Binomial

The damage model, discussed in Theorem 4.4, also produces a characterization of the binomial distribution in that, if (4.37) holds and X is Poisson, then the damage process is binomial. Another characterization of the binomial distribution assumes that the RVs X and Y are independent, and that the conditional distribution of X given $X + Y$ is hypergeometric [4.41]. When the conditional distribution is negative hypergeometric, a similar result for the negative binomial distribution is obtained.

Remarks. Characterizations of other discrete distributions are limited. For results on hypergeometric and logarithmic distributions, see [4.35]. Characterizations of discrete distributions based on order statistics are discussed in [4.40]. See [4.42] for characterizations based on weighted distributions when $\mathcal{F}$ is the power-series family in (4.39).

4.9 Multivariate Distributions and Conditional Specification

Characterization results are less common for multivariate distributions. Notable exceptions are the multivariate normal and the Marshall–Olkin multivariate exponential distribution. First we discuss another dimension to multivariate characterizations, namely the specification of the properties of the conditional distribution(s). For example, can one identify the joint PDF $f(x, y)$ using the properties of the conditional PDFs $f(x|y)$ and $f(y|x)$? This has been an active area of research in recent years. See [4.43] for an excellent account of the progress. We present one such result.

Theorem 4.5

Let $f(x, y)$ be a bivariate PDF where conditional PDFs belong to natural parameter exponential families with full rank given by

$$f(x|y) = r_1(x)\beta_1[\boldsymbol{\theta}_1(y)] \exp\left[\boldsymbol{\theta}_1(y)'\boldsymbol{q}_1(x)\right] \qquad (4.41)$$

and

$$f(y|x) = r_1(y)\beta_2[\boldsymbol{\theta}_2(x)] \exp\left[\boldsymbol{\theta}_2(x)'\boldsymbol{q}_2(y)\right] \qquad (4.42)$$

where $\boldsymbol{\theta}_1(y)$ and $\boldsymbol{q}_1(x)$ are $k_1 \times 1$ vectors, and $\boldsymbol{\theta}_2(x)$ and $\boldsymbol{q}_1(y)$ are $k_2 \times 1$ vectors, and the components of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are linearly independent. Then the joint PDF is of the form

$$f(x, y) = r_1(x) r_2(y) \exp[\boldsymbol{A}(x, y)] \qquad (4.43)$$

where $\boldsymbol{A}(x, y) = [1, \boldsymbol{q}_1(x)']\boldsymbol{M}[1, \boldsymbol{q}_2(y)']'$ for a suitable matrix $\boldsymbol{M} = (m_{ij})$, whose elements are chosen so that $f(x, y)$ integrates to 1.

When both the conditional distributions are normal, this result implies that

$$f(x, y) \propto \exp[(1, x, x^2)\boldsymbol{M}(1, y, y^2)'] \qquad (4.44)$$

and the classical bivariate normal corresponds to the condition $m_{23} = m_{32} = m_{33} = 0$ [4.44].

Instead of the conditional PDF, the conditional distribution may be specified using regression functions, say $\mathrm{E}(Y|X=x)$. Then the joint distribution can be determined in some cases. For example, suppose X given $Y=y$ is $\mathrm{N}(\alpha y, 1)$, i. e., normal with mean αy and unit variance, and $\mathrm{E}(Y|X=x)=\beta x$. Then $0<\alpha\beta<1$ and (X, Y) is bivariate normal [4.44].

The conditional specification could be in terms of the conditional SF $\Pr(Y>y|X>x)$, or in the form of the marginal distribution of X and the conditional distribution of X given $Y=y$. Sometimes these together can also identify the joint distribution.

4.9.1 Bivariate and Multivariate Exponential Distributions

Marshall and *Olkin* [4.45] introduced a bivariate exponential (BVE) distribution to model the component lifetimes in the context of a shock model. Its (bivariate) SF, with parameters $\lambda_1>0$, $\lambda_2>0$, and $\lambda_{12}\geq 0$, is given by

$$\Pr(X>x, Y>y)=\mathrm{e}^{-\lambda_1 x-\lambda_2 y-\lambda_{12}\max(x,y)}, \quad x, y>0\,. \tag{4.45}$$

Here X is $\exp(\lambda_1+\lambda_{12})$ and Y is $\exp(\lambda_2+\lambda_{12})$. The joint distribution is characterized by the following conditions: (a) X and Y are marginally exponential. (b) $\min(X, Y)$ is exponential, and (c) $\min(X, Y)$ and $|X-Y|$ are independent.

The LMP in (4.18) that characterized the univariate exponential can be extended as

$$\begin{aligned}&\Pr(X>x+t_1, Y>y+t_2|X>x, Y>y)\\&=\Pr(X>t_1, Y>t_2)\,.\end{aligned} \tag{4.46}$$

If (4.46) is assumed to hold for all $x, y, t_1, t_2\geq 0$, then X and Y are necessarily independent exponential RVs. The SF (4.45) would satisfy (4.46) for all $x, y\geq 0$ and $t_1=t_2=t\geq 0$. This condition, often referred to as *bivariate* LMP, is equivalent to assuming that both (b) and (c) above hold. While the Marshall–Olkin BVE distribution has exponential marginals and bivariate LMP, it is not absolutely continuous. If one imposes the LMP and absolute continuity, the marginal distributions will no longer be exponential [4.46].

There are other multivariate distributions that are characterized by the multivariate versions of the failure-rate function (see [4.47], p. 403–407).

4.9.2 Multivariate Normal

An early characterization of the classical multivariate normal (MVN) random vector, known as Cramér–Wold Theorem, is that every linear combination of its components is univariate normal. Most of the characterizations of the univariate normal distribution discussed in Sect. 4.5 easily generalize to the MVN distribution. For example, the independence of nonsingular transforms of independent random vectors [see (4.25) for the univariate version], independence of the sample mean vector and sample covariance matrix, maximum entropy with a given mean vector and covariance matrix, are all characteristic properties of the MVN distribution. There are, of course, results based on conditional specifications. We mention two.

For an m-dimensional RV $\boldsymbol{X}$, let $\boldsymbol{X}_{(i,j)}$ be the vector $\boldsymbol{X}$ with coordinates i and j deleted. If, for each i, j the conditional distribution of (X_i, X_j) given $\boldsymbol{X}_{(i,j)}=\boldsymbol{x}_{(i,j)}$ is BVN for each $\boldsymbol{x}_{(i,j)}$, then $\boldsymbol{X}$ is MVN ([4.43], p. 188).

If $X_1, \cdots, X_m$ are jointly distributed RVs such that $(X_1, \cdots, X_{m-1})\overset{d}{=}(X_2, \cdots, X_m)$, and X_m given $\{X_1=x_1, X_2=x_2, \cdots, X_{m-1}=x_{m-1}\}$ is N $(\alpha+\sum_{j=1}^{m-1}\beta_j x_j, \sigma^2)$, then $(X_1, \cdots, X_m)$ are jointly m-variate normal ([4.47], p. 157).

Excellent summaries of characterizations of the bivariate and multivariate normal distributions are available, respectively, in Sect. 46.5 and Sect. 45.7 of [4.47] (see also, the review [4.48]).

4.9.3 Other Distributions

Characterization results for other multivariate distributions are not common. A few characterizations of the *multinomial distribution* are available ([4.49], Sect. 35.7), and these are natural extensions of the binomial characterizations. One result is that, if the sum of two independent vectors is multinomial, then each is multinomial. There are also a few characterizations of the Dirichlet distribution, a multivariate extension of the beta distribution over $(0, 1)$ ([4.47], Sect. 49.5). It has the characteristic property of *neutrality*, which can be described for $m=2$ components as follows. For two continuous RVs X and Y such that $X, Y\geq 0$ and $X+Y\leq 1$, neutrality means X and $Y/(1-X)$ are independent, and Y and $X/(1-Y)$ are independent.

The multivariate Pareto distribution due to Mardia, having the multivariate SF

$$\Pr(X_1 > x_1, \ldots, X_m > x_m) = \left[1 + \sum_{i=1}^{m}\left(\frac{x_i}{\sigma_i}\right)\right]^{-\alpha}, \quad x_1, \ldots, x_m > 0\,, \tag{4.47}$$

accepts characterizations that are based on conditional specifications. Marginally, the X_i here are Pareto II RVs.

A few papers that characterize bivariate distributions with geometric marginals do exist. Some are related to the Marshall–Olkin BVE.

4.10 Stability of Characterizations

Consider the LMP in (4.18) that characterizes the exponential distribution. Now suppose the LMP holds approximately in the sense

$$\sup_{x\geq 0,\, y\geq 0} \left|\Pr(X > x+y|X > x) - \Pr(X > y)\right| \leq \epsilon\,. \tag{4.48}$$

The question of interest is how close the parent CDF F is to an exponential CDF. It is known ([4.21], p. 7) that, when X is nondegenerate and $F^{-1}(0) = 0$, if (4.48) holds then $\mathrm{E}(X)$ is finite and, with $\mathrm{E}(X) = 1/\lambda$,

$$\sup_{x\geq 0} \left|\Pr(X > x) - \exp(-\lambda x)\right| \leq 2\epsilon\,. \tag{4.49}$$

This result provides an idea about the stability of the LMP of the exponential distribution. There are many such results—mostly for the exponential and normal distributions. Such results involve appropriate choices of metrics for measuring the distance between: (a) the characterizing condition and the associated perturbation, and (b) the CDF being characterized and the CDF associated with the perturbed condition. We will mention below a few simple stability theorems. It is helpful to introduce one popular metric measuring the distance between two distributions with associated RVs X and Y:

$$\rho(X, Y) \equiv \sup_{-\infty<x<\infty} |\Pr(X \leq x) - \Pr(Y \leq x)|\,. \tag{4.50}$$

Note that (4.49) basically says that $\rho(X, Y) \leq 2\epsilon$, where Y is an $\exp(\lambda)$ RV.

For the exponential parent, the constancy of $\mathrm{E}[(X_{k+1:n} - X_{k:n})|X_{k:n}]$ is a characterizing property. The associated stability result is the following [4.22, p. 545], [4.50]:

If F has support $(0, \infty)$ and is strictly increasing in its domain, and $\gamma > 0$ is such that

$$\left|\mathrm{E}\left[(X_{k+1:n} - X_{k:n})\middle|X_{k:n} = x\right] - \gamma\right| \leq \epsilon\,, \quad \text{for almost all } x \geq 0\,, \tag{4.51}$$

then, there exist positive constants λ_1, λ_2 and c that depend only on γ and $n-k$ such that $\rho(X, Y) \leq c\epsilon^{1/(n-k)} \exp(-\lambda_2 x)$, where Y is $\exp(\lambda_1)$.

For the normal distribution, we state stability results for two classical characterizations.

1. *Pólya's characterization* [4.5, p. 298]. Let X_1 and X_2 be IID with zero mean, unit variance and $\mathrm{E}(|X_1|^3) \leq M < \infty$. Let $Y = (X_1 + X_2)/\sqrt{2}$. If $\rho(X_1, Y) \leq \epsilon$, then $\rho(X_1, Z) \leq c\epsilon^{1/3}$, where c depends only on M and Z is standard normal. (See the recent work [4.51] for another metric of comparison.)
2. *Independence of $\bar{X}$ and S^2* [4.52]. Suppose that for all x, y,
$$\left|\Pr(\bar{X} \leq x, S^2 \leq y) - \Pr(\bar{X} \leq x)\Pr(S^2 \leq y)\right| \leq \epsilon\,. \tag{4.52}$$
Then there exists a (possibly degenerate) $\mathrm{N}(\mu, \sigma^2)$ RV Y such that $\rho(X, Y) \leq c[\log(1/\epsilon)]^{-1/2}$, where c depends only on n, and ρ is given by (4.50).

Many such results are available in the several monographs on stability of characterizations edited by *Kalashnikov* and *Zolotarov* and others (see e.g., [4.52, 53]). These have come from the periodic conferences held in Eastern Europe.

4.11 Applications

A characterization can be of use in the construction of goodness-of-fit tests and in the examination of the consequences of the modeling assumptions. It can be helpful in some simulation studies. A full characterization states that a condition C on the sample data is necessary and sufficient for the condition $F \in \mathcal{F}_0$ to hold. The *neces-*

sity part can be used for the simulation study of the properties of the sample statistics involved in the condition C. For example, if one is interested in the properties of a function of the random sample mean $\bar{X}$ and variance S^2 from a standard normal population, one can start with two *independent* statistics, one normal and the other a χ^2. We can simulate these statistics directly, and study the empirical properties of the function of interest. The characterization of the exponential based on the independence and exponentially distributed nature of spacings of order statistics, can be used to simulate all or a selected set of exponential order statistics without any sorting.

The *sufficiency* part can be used for checking implications of modeling assumptions or of their compatibility. Again consider the situation where the researcher is willing to accept, based on past data, say, that $\bar{X}$ and variance S^2 are independent. This is equivalent to assuming that the sample is from a normal population. As another example, consider the early characterization result, proved by *Cramér* in 1937, which states that, if X and Y are independent and the sum is assumed to be normal, then each of them must be normal (see [4.24], p. 53). Thus, if a researcher is willing to assume the independence of these RVs and that the sum is normal, this is the same as assuming that both X and Y are individually normal and are independent. This implication can be gainfully employed to infer that $aX + bY$ is $N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$.

Characterizing properties, when used as necessary and sufficient conditions, naturally lead to both informal (graphical) and formal goodness-of-fit tests. The various plots such as the quantile–quantile (Q–Q) plots, hazard function plots, or the MRL plots are all based on the characterizing properties of these functions. For example, in Q–Q plots, the sample quantile $X_{j:n}$ is plotted against the corresponding hypothesized population quantile $F_0^{-1}(p_j)$, where F_0 is the standardized form of F, and p_j is sometimes chosen as $(j - (1/2))/n$. (See [4.3], p. 270 for other choices for p_j.) If a linear fit is unreasonable, then one may infer that the assumption that F is the parent CDF is untenable. Probability plots (see [4.54], Chapt. 3), and the hazard-function plots ([4.54], Chapt. 4), popular in reliability studies, also provide similar informal checks of the fit of the assumed distribution. MRL plots, where $\frac{1}{k}\sum_{j=n-k+1}^{n}(x_{j:n} - x_{n-k:n})$ is plotted against the $x_{n-k:n}$ (see, [4.55], p. 296), are used to determine the appropriate domain of attraction for the extreme order statistics.

In theory, a formal goodness-of-fit test can be constructed from any characterization. For example, the independence of spacings of order statistics of a random sample from a continuous distribution implies the CDF is exponential, and thus can be used to construct a goodness-of-fit test even with a (type II) censored sample. Such a test does have good power properties. Another example is the test based on the maximum-entropy property of the normal distribution [4.56]. See [4.57–59] for some goodness-of-fit tests inspired by characterizations. A nice overview is provided in [4.60]. However, not all characterizations lead to powerful tests. Also, creation of goodness-of-fit tests that exploit the concerned characterizing property to the full extent may not be easy. Perhaps this explains the sparsity of such applications of characterizations.

4.12 General Resources

The literature on characterization is extensive. Over 1000 papers have appeared to date; however, several review articles, monographs (e.g., [4.5, 7, 31, 43]) and encyclopedic books on distributions by *Johnson*, *Kotz*, and their coauthors ([4.22, 30, 35, 47, 49]) have served as excellent filters. Other sources include the volume [4.61], which contains papers on characterizations presented at the Calgary conference in 1974, and the monograph [4.62], which treats characterization (identification) in a broader context. For results based on order statistics see [4.3, Sect. 6.7]; see also [4.63].

Here, we have given an informal introduction to characterizations using a small fraction of the available results. For economy, we have cited the above general and other (distribution-specific) secondary resources in this survey. To get a full appreciation of this active area of research, one should consult many of the primary as well as secondary sources.

References

4.1 C. R. Rao: *Linear Statistical Inference and Its Applications*, 2nd edn. (Wiley, New York 1973)
4.2 R. E. Barlow, F. Proschan: *Statistical Theory of Reliability and Life Testing: Probability Models* (To Begin With, Silverspring 1981)
4.3 H. A. David, H. N. Nagaraja: *Order Statistics*, 3rd edn. (Wiley, New York 2003)
4.4 B. C. Arnold, N. Balakrishnan, H. N. Nagaraja: *Records* (Wiley, New York 1998)
4.5 A. M. Kagan, Yu. V. Linnik, C. R. Rao: *Characterization Problems in Mathematical Statistics* (Wiley, New York 1973)
4.6 B. Ramachandran, K.-S. Lau: *Functional Equations in Probability Theory* (Academic, Boston 1991)
4.7 C. R. Rao, D. N. Shanbhag: *Choquet–Deny Type Functional Equations with Applications to Stochastic Models* (Wiley, Chichester 1994)
4.8 K.-S. Lau, C. R. Rao: Integrated Cauchy functional equations, characterizations of the exponential law, Sankhyā Ser. A **44**, 72–90 (1982)
4.9 C. R. Rao, D. N. Shanbhag: Extensions of a characterization of an exponential distribution based on a censored ordered sample. In: *Advances in the Theory and Practice of Statistics*, ed. by N. L. Johnson, N. Balakrishnan (Wiley, New York 1997) pp. 431–440
4.10 U. Kamps: Characterizations of distributions by recurrence relations and identities for moments of order statistics. In: *Order Statistics: Theory and Methods*, Handbook of Statistics, Vol. 16, ed. by N. Balakrishnan, C. R. Rao (North-Holland, Amsterdam 1998) pp. 291–311
4.11 A. V. Kakosyan, L. B. Klebanov, J. A. Melamed: *Characterization of distributions by the method of intensively monotone operators*, Lect. Notes in Math., Vol. 1088 (Springer, Berlin Heidelberg New York 1984)
4.12 J. Galambos: Characterizations of distributions. In: *Encyclopedia of Statistical Sciences*, Vol. 1, ed. by S. Kotz, N. L. Johnson (Wiley, New York 1982) pp. 422–428
4.13 J. A. Shohat, J. D. Tamarkin: *The Problem of Moments*, Math. Surveys, Vol. 1 (American Mathematical Society, New York 1943)
4.14 J. S. Huang: Moment problem of order statistics: a review, Int. Stat. Rev. **57**, 59–66 (1989)
4.15 M. Franco, J. M. Ruiz: Characterization based on conditional expectations of adjacent order statistics: a unified approach, Proc. Am. Math. Soc. **127**, 861–874 (1999)
4.16 N. A. Langberg, R. V. León, F. Proschan: Characterization of nonparametric classes of life distributions, Ann. Probab. **8**, 1163–1170 (1980)
4.17 L. Bondesson: A generalization of Poincaré's characterization of exponential families, J. Stat. Plan. Infer. **63**, 147–155 (1997)
4.18 B. C. Arnold, J. S. Huang: Characterizations. In: *The Exponential Distribution: Theory, Methods and Applications*, Lect. Notes Math., ed. by N. Balakrishnan, A. P. Basu (Gordon Breach, Amsterdam 1995) pp. 185–203
4.19 J.-L. Xu: Characterizations of the exponential distribution via the blocking time in a queueing system with an unreliable server, J. Appl. Probab. **35**, 236–239 (1998)
4.20 U. Gather, U. Kamps, N. Schweitzer: Characterizations of distributions via identically distributed functions of order statistics. In: *Order Statistics: Theory and Methods*, Handbook of Statistics, Vol. 16, ed. by N. Balakrishnan, C. R. Rao (North-Holland, Amsterdam 1998) pp. 257–290
4.21 T. Azlarov, N. A. Volodin: *Characterization Problems Associated With the Exponential Distribution* (Springer, Berlin Heidelberg New York 1986)
4.22 N. L. Johnson, S. Kotz, N. Balakrishnan: *Continuous Univariate Distributions*, Vol. 1 (Wiley, New York 1994)
4.23 T. S. Ferguson: On a Rao–Shanbhag characterization of the exponential/geometric distribution, Sankhyā A **64**, 246–255 (2002)
4.24 H. Cramér: *Random Variables and Probability Distributions*, 3rd edn. (Cambridge Univ. Press, Cambridge 1970)
4.25 E. Lukacs, R. G. Laha: *Applications of Characteristic Functions* (Griffin, London 1964)
4.26 A. M. Mathai, G. Pederzoli: *Characterizations of the Normal Probability Law* (Halsted, New York 1977)
4.27 W. Bryc: *The Normal Distribution. Characterizations with Applications*, Lect. Notes Stat., Vol. 100 (Springer, Berlin Heidelberg New York 1995)
4.28 L. Bondesson: When is the sample mean BLUE?, Scand. J. Stat. **3**, 3116–120 (1976)
4.29 P. K. Goel, M. H. DeGroot: Only normal distributions have linear posterior expectations in linear regression, J. Am. Stat. Assoc. **75**, 895–900 (1980)
4.30 N. L. Johnson, S. Kotz, N. Balakrishnan: *Continuous Univariate Distributions*, Vol. 2, 2nd edn. (Wiley, New York 1995)
4.31 J. Galambos, S. Kotz: *Characterizations of Probability Distributions*, Lect. Notes Math., Vol. 675 (Springer, Berlin Heidelberg New York 1978)
4.32 B. C. Arnold: *Pareto Distributions* (International Cooperative, Burtonsville 1983)
4.33 M. Asadi, C. R. Rao, D. N. Shanbhag: Some unified characterization results on generalized Pareto distributions, J. Stat. Plan. Infer. **93**, 29–50 (2001)

4.34 V. Seshadri: *The Inverse Gaussian Distribution. A Case Study in Exponential Families* (Oxford Univ. Press, Oxford 1993)
4.35 N. L. Johnson, S. Kotz, A. W. Kemp: *Univariate Discrete Distributions*, 2nd edn. (Wiley, New York 1992)
4.36 E. McKenzie: Linear characterizations of the Poisson distribution, Statist. Probab. Lett. **11**, 459–461 (1991)
4.37 A. Kagan: A discrete version of the Stam inequality, a characterization of the Poisson distribution, J. Stat. Plan. Infer. **92**, 7–12 (2001)
4.38 P. Diaconis, S. Zabell: Closed form summation for classical distributions: variations on a theme of de Moivre, Stat. Sci. **6**, 284–302 (1991)
4.39 B. L. S. Prakasa Rao: Characterization, identifiability for stochastic processes. In: *Stochastic Processes: Theory and Methods*, Handbook of Statistics, Vol. 19, ed. by D. N. Shanbhag, C. R. Rao (Elsevier, Amsterdam 2001) pp. 643–691
4.40 H. N. Nagaraja: Order statistics from discrete distributions (with discussion), Statistics **23**, 189–216 (1992)
4.41 P. Patil, V. Seshadri: Characterization theorems for some univariate probability distributions, J. Roy. Statist. Soc. Ser. B **26**, 286–292 (1964)
4.42 A. G. Pakes, J. Navarro, J. M. Ruiz, Y. del Aguila: Characterizations using weighted distributions, J. Stat. Plan. Infer. **116**, 389–420 (2003)
4.43 B. C. Arnold, E. Castillo, J. M. Sarabia: *Conditional Specification of Statistical Models* (Springer, Berlin Heidelberg New York 1999)
4.44 B. C. Arnold: Characterizations involving conditional specification, J. Stat. Plan. Infer. **63**, 117–131 (1997)
4.45 A. W. Marshall, I. Olkin: A multivariate exponential distribution, J. Am. Stat. Assoc. **62**, 30–44 (1967)
4.46 H. W. Block: A characterization of a bivariate exponential distribution, Ann. Stat. **5**, 808–812 (1977)
4.47 S. Kotz, N. Balakrishnan, N. L. Johnson: *Continuous Multivariate Distributions*, Vol. 1, 2nd edn. (Wiley, New York 2000)
4.48 G. G. Hamedani: Bivariate, multivariate normal characterizations: a brief survey, Commun. Stat. Theory Methods **21**, 2665–2688 (1992)
4.49 N. L. Johnson, S. Kotz, N. Balakrishnan: *Discrete Multivariate Distributions* (Wiley, New York 1997)
4.50 R. V. Yanushkevichius: *Stability of Characterizations of Probability Distributions* (Mokslas, Vilnius 1991) In Russian
4.51 R. Yanushkevichius: Stability of the characterization of normal distribution in the Laha–Lukacs theorem, Stat. Probab. Lett. **49**, 225–233 (2000)
4.52 R. Yanushkevichius: On ϵ-independence of sample mean, sample variance. In: *Stability Problems for Stochastic Models*, Lect. Notes Math., ed. by V. V. Kalashnikov, B. Penkov, V. M. Zolotarev (Springer, Berlin Heidelberg New York 1978) pp. 207–223
4.53 V. V. Kalashnikov, V. M. Zolotarev (eds.): *Stability Problems for Stochastic Models*, Lect. Notes in Math., Vol. 982 (Springer, Berlin Heidelberg New York 1983)
4.54 W. Nelson: *Applied Life Data Analysis* (Wiley, New York 1982)
4.55 P. Embrechts, C. Klüppelberg, T. Mikosch: *Modelling Extremal Events for Insurance and Finance* (Springer, Berlin Heidelberg New York 1997)
4.56 O. Vasicek: A test for normality based on sample entropy, J. R. Stat. Soc. Ser. B **38**, 54–59 (1976)
4.57 M. Csörgő, V. Seshadri, M. Yalovsky: Applications of characterizations in the area of goodnees of fit. In: *Statistical Distribution in Scientific Work*, Vol. 2, ed. by G. P. Patil, S. Kotz, J. K. Ord (Reidel, Dordrecht 1975) pp. 79–90
4.58 F. J. O'Reilly, M. A. Stephens: Characterizations, goodness of fit tests, J. R. Stat. Soc. Ser. B **44**, 353–360 (1982)
4.59 K. W. Morris, D. Szynal: Tests for uniformity, exponentiality using a characterization, J. Math. Sci. **111**, 3572–3581 (2002)
4.60 C. E. Marchetti, G. S. Mudholkar: Characterization theorems and goodness-of-fit tests. In: *Goodness-of-Fit Tests and Model Validity*, ed. by C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, M. Mesbah (Birkhäuser, Boston 2002) pp. 125–142
4.61 G. P. Patil, S. Kotz, J. K. Ord (eds.): A Modern Course on Statistical Distributions in Scientific Work. In: *Characterizations and Applications*, Vol. 3 (Reidel, Dordrecht-Boston 1975)
4.62 B. L. S. Prakasa Rao: *Identifiability in Stochastic Models – Characterization of Probability Distributions* (Academic, Boston 1992)
4.63 C. R. Rao, D. N. Shanbhag: Recent approaches to characterizations based on order statistics and record values. In: *Order Statistics: Theory and Methods*, Handbook of Statistics, Vol. 16, ed. by N. Balakrishnan, C. R. Rao (North-Holland, Amsterdam 1998) pp. 231–256