Characteristics of Generalized Extreme Value Distributions

Michel Bierlaire, Institute of Mathematics, Swiss Institute of Technology
Denis Bolduc, Department of Economics, Laval University
Daniel McFadden, Econometrics Laboratory, University of California, Berkeley

March 2003

This note is concerned with joint probability distributions whose one-dimensional marginal distributions are Extreme Value Type 1 (EV1); i.e., $\text{Prob}(U_j \le u_j) = \exp(-\exp(\mu(u_j - v_j))$, where the v_j are location parameters and μ is a common scale factor. Call these *Generalized Extreme Value* (GEV) *distributions*. GEV distributions have application in the study of discrete choice behavior, and were initially studied by McFadden (1978,1981,1984,2001).

Let $\mathbf{C} = \{1,...,J\}$. Let $\mathbf{1}_j$ denote the unit vectors for $j \in \mathbf{C}$, and for $\mathbf{A} \subseteq \mathbf{C}$, let $\mathbf{1}_\mathbf{A}$ denote a vector with components that are one for the elements of $\mathbf{A}$, zero otherwise. Define a *GEV generating function* $H(w_1,...,w_J)$ on $w = (w_1,...,w_J) \ge 0$ to have the properties that it is non-negative, homogeneous of degree $\mu > 0$, and differentiable, with its mixed partial derivatives for $j = 1,...,J$ satisfying $(-1)^j \partial^j H/\partial w_1 ... \partial w_j \le 0$. A GEV generating function H is *proper* with respect to a subset $\mathbf{A}$ of $\mathbf{C}$ if $H(\mathbf{1}_j) > 0$ for $j \in \mathbf{A}$ and $H(\mathbf{1}_{\mathbf{C}\backslash\mathbf{A}}) = 0$. Let $\mathcal{H}_\mu$ denote the family of GEV generating functions that are homogenous of degree μ, and let $\mathcal{H}_\mu(\mathbf{A})$ denote the subfamily that is proper with respect to $\mathbf{A}$. Let $\gamma = 0.5772156649$ denote Euler's constant.

Theorem 1. If a random vector $U = (U_1,...,U_J)$ has a GEV distribution $F(u) = \text{Prob}(U \le u)$, then this distribution has the form

[1] $$F(u) = \exp(-H(\exp(-u_1 + v_1),...,\exp(-u_J + v_J))),$$

where $(v_1,...,v_J)$ are location parameters and $H(w_1,...,w_J)$ is a non-negative function of $w \ge 0$ which is homogeneous of degree $\mu > 0$ and satisfies $H(\mathbf{1}_j) > 0$ for $j \in \mathbf{C}$. Conversely, a sufficient condition for the function [1] to be a GEV distribution is that $H \in \mathcal{H}_\mu(\mathbf{C})$. GEV distributions have the properties:

A. $f(u) = \partial^J F(u)/\partial w_1 ... \partial w_J \ge 0$, $F(u) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_J} f(u)du$, and $0 \le F(u) \le 1$.

B. The U_j for $j = 1,...,J$ are EV1 with common variance $\pi^2/6\mu^2$, means $v_j + \mu^{-1} \log H(\mathbf{1}_j) + \gamma/\mu$, and moment generating functions $\exp(tv_j) H(\mathbf{1}_j)^{t/\mu} \Gamma(1-t/\mu)$.

C. $U_0 = \max_{i=1,...,J} U_i$ is EV1 with variance $\pi^2/6\mu^2$, mean $(\log H(\exp(v_1),...,\exp(v_J))) + \gamma)/\mu$, and moment generating function $H(\exp(v_1),...,\exp(v_J))^{t/\mu} \Gamma(1-t/\mu)$.

D. Letting $H_j(w) = \partial H(w)/\partial w_j$, the probability P_j that alternative j maximizes U_i for $i \in \mathbf{C}$ satisfies

[2] $$P_j = \exp(v_j) \cdot H_j(\exp(v_1),...,\exp(v_J))/\mu H(\exp(v_1),...,\exp(v_J)).$$

Proof: Recursive mixed differentiation of F and the derivative property of H, plus its limits, establish result A, so that F is a CDF. Results B to D follow by direct computation, with the moment generating function of the EV1 distribution in result C given by Johnson and Kotz (1970, chap. 21).

Not every function $H(w_1,...,w_J)$ associated with a GEV distribution function via [1] and the inverse mapping $H(w_1,...,w_J) = -\log F(v_1 - \log w_1,...,v_J - \log w_J)$ is a GEV generating function, as the signed derivative property $(-1)^j\partial^j H/\partial w_1...\partial w_j \leq 0$ for $j = 1,...,J$ may fail. For example, probability mixtures of GEV distributions obtained from GEV generating functions are again GEV distributions, but these mixtures have inverse mappings that can fail to satisfy the signed derivative property.

In application, interpret U as a vector of random payoffs or utilities for alternatives in C. Then the P_j given by [2] are choice probabilities from random utility maximization. The closed form for these probabilities facilitates computation. The linear function $H(w) = w_1 + ... + w_J$ is a GEV generating function; the vector U with the distribution function [1] for this H has independent extreme value distributed components. The choice probabilities [2] for this case have a *multinomial logit* (MNL) form,

[3] $$P_j = \exp(v_j)/\Sigma_{i\in\mathbf{C}}\exp(v_i).$$

The next result gives operations on GEV generating functions that can be applied recursively to generate additional GEV generating functions.

Lemma 2. The family $\mathcal{H}_\mu$ of GEV generating functions is closed under the following operations:

A. If $H(w_1,...,w_J) \in \mathcal{H}_\mu(\mathbf{A})$, then $H(\alpha_1 w_1,...,\alpha_J w_J) \in \mathcal{H}_\mu(\mathbf{B})$ for $\alpha_1,...,\alpha_J \geq 0$ and $\mathbf{B} = \{j\in\mathbf{A}|\alpha_j>0\}$.
B. If $H(w_1,...,w_J) \in \mathcal{H}_\mu(\mathbf{A})$ and $s \geq 1$, then $H(w_1^s,...,w_J^s)^{1/s} \in \mathcal{H}_\mu(\mathbf{A})$.
C. If $H^{\mathbf{A}}(w_1,...,w_J) \in \mathcal{H}_\mu(\mathbf{A})$ and $H^{\mathbf{B}}(w_1,...,w_J) \in \mathcal{H}_\mu(\mathbf{B})$, where **A** and **B** are subsets of **C**, not necessarily disjoint, then $H^{\mathbf{A}}(w_1,...,w_J) + H^{\mathbf{B}}(w_1,...,w_J) \in \mathcal{H}_\mu(\mathbf{A}\cup\mathbf{B})$.

Proof: Direct computation.

Note that the operations in Lemma 2 all apply to and preserve a common degree of homogeneity μ; the result in Theorem 1 giving a closed form for the choice probabilities requires this common scale factor for the extreme value components U_j. It is possible through the normalization $H(w_1,...,w_J) = H^*(w_1^{1/\mu},...,w_J^{1/\mu})$ to convert a family $H^* \in \mathcal{H}_\mu$ into a linear homogeneous family $H \in \mathcal{H}_1$; however, the literature on GEV models has adopted the case of general μ. An example of a GEV generating function built up using operations B and C in Lemma 2 is a three-level nested MNL model generated by a function H of the form

[4] $$H(w_1,...,w_J) = \sum_{m=1}^{M}\left[\sum_{k=1}^{K}\left[\sum_{i\in A_{mk}} w_i^{s_m' s_k}\right]^{\frac{1}{s_k}}\right]^{\frac{1}{s_m'}},$$

where the $\mathbf{A}_{mk}$ partition $\mathbf{C}$ and $s_k, s_m' \geq 1$. This form corresponds to a decision tree: m indexes major branches, k indexes limbs from each branch, and i indexes the final twigs. The larger s_k or s_m', the more substitutable the alternatives in $\mathbf{A}_{mk}$, and the more rapidly the choice probabilities will respond to differences in the location parameters v_i within partition sets. If $s_k = s_m' = 1$, this model reduces to a MNL model. The function [4] remains a GEV generating function when the sets $\mathbf{A}_{mk}$ are not necessarily a partition, and overlap. In this case, the choice process can be described by a directed graph, with terminal nodes j reached by one or more possible paths.

GEV families are closed under location shifts; this is the property established by operation A in Lemma 2. This property has particular application to a property of choice models estimated from stratified samples and/or with analysis restricted to alternatives sampled from the full set faced by a subject..

Suppose choices $j \in \mathbf{C}$ and covariates z are distributed in a population with probability $P_j(z,\theta_0)p(z)$, where θ is a parameter vector with true value θ_0, and that a sample is drawn in which a subject with configuration (j,z) in the population has a probability R(j,z) of qualification. For example, stratification on j or z or a combination of the two, or attenuation that depends on (j,z), will produce a qualification probability that varies over the configurations. The qualification probability may reflect the composition of subsamples drawn from various strata. Important cases are exogenous stratification, with R(j,z) independent of j, and choice-based sampling, with R(j,z) independent of z. The subset of alternatives that is sampled is denoted $\mathbf{A}(z) = \{i \in \mathbf{C} | R(i,z)>0\}$. The joint probability of (j,z) in the sample, also termed the sample data generation process, is $P_j(z,\theta_0)p(z)R(j,z)/r(\theta_0)$, where $r(\theta_0) = \sum_z \sum_j P_j(z,\theta_0)p(z)R(j,z)$. Next suppose that for an observation with configuration (j,z), the analyst draws a set $\mathbf{B} \subseteq \mathbf{A}(z)$ with probability $S(\mathbf{B}|j,z)$, and analyzes choice *as if* it were limited to **B.** The reason for doing this is to limit data collection and computation. We will assume that any set **B** that is drawn with positive probability contains the chosen alternative j, and at least one non-chosen alternative. We will also assume a *positive conditioning property* that $S(\mathbf{B}|j,z) > 0$ for the observed choice j implies $S(\mathbf{B}|i,z) > 0$ for each $i \in \mathbf{B}$; i.e., **B** could have been drawn conditioned on any of its elements as the observed choice. In the presence of a sampling protocol described by R(j,z), and analysis conducted as if choice were restricted to a set **B** drawn with probability $S(\mathbf{B}|j,z)$, the joint probability of $(j,z,\mathbf{B})$ is $S(\mathbf{B}|j,z)P_j(z,\theta_0)p(z)R(j,z)/r(\theta_0)$. From this, the conditional probability in the sample of j given z and **B** is

$$[5] \qquad \Pr(j|z,\mathbf{B},\theta) = P_j(z,\theta)R(j,z)S(\mathbf{B}|j,z)/\textstyle\sum_{i\in\mathbf{B}} P_i(z,\theta)R(i,z)S(\mathbf{B}|i,z).$$

This probability describes the data generation process for the sample as observed and collected (including the restriction of alternatives considered to the set **B**), and can be used as a basis for maximum likelihood of the parameter vector θ_0. Note that if R(j,z) or $S(\mathbf{B}|j,z)$ have multiplicative factors that depend only on z, then these factors cancel out of [5]. Hence, it is sufficient for analysis to consider only the kernels of R(j,z) and $S(\mathbf{B}|j,z)$ in which j and z interact.

Now suppose the choice probabilities $P_j(z,\theta_0)$ are obtained from a GEV generating function $H(w_1,...,w_J)$ of degree μ and location parameters $v_j(z,\theta_0)$, so that [2] gives

$$[6] \quad P_j(z,\theta_0) = \exp(v_j(z,\theta_0))H_j(\exp(v_1(z,\theta_0)),...,\exp(v_J(z,\theta_0)))/\mu H(\exp(v_1(z,\theta_0)),...,\exp(v_J(z,\theta_0))).$$

Substituted into [5], this probability yields

[7] $$\Pr(j|z,\theta_0) = \frac{\exp(v_j(z,\theta_0))H_j(\exp(v_1(z,\theta_0)),...,\exp(v_J(z,\theta_0)))R(j,z)S(\boldsymbol{B}|j,z)}{\sum_{i\in\boldsymbol{B}}\exp(v_i(z,\theta_0))H_i(\exp(v_1(z,\theta_0)),...,\exp(v_J(z,\theta_0)))R(i,z)S(\boldsymbol{B}|i,z)}\ .$$

Define $\rho(i,z) = 1/R(i,z)$ if $R(i,z) > 0$, and $\rho(i,z) = 0$ otherwise. Similarly, define $\sigma(\mathbf{B}|j,z) = 1/S(\mathbf{B}|j,z)$ for $j \in B$, and $\sigma(\mathbf{B}|j,z) = 0$ otherwise. From Lemma 2, since $H(w_1,...,w_J)$ is a GEV generating function of degree μ that is proper with respect to **C**, the function

[8] $$H^*(w_1,...,w_J) = H(w_1\rho(1,z)\sigma(\mathbf{B}|1,z),...,w_J\,\rho(J,z)\sigma(\mathbf{B}|J,z))$$

is a GEV generating function that is proper with respect to $\mathbf{B} \subseteq \mathbf{A}(z)$. Consider the choice probabilities generated by H^* with location parameters $v_j(z,\theta) + \log R(j,z) + \log S(\mathbf{B}|j,z)$ for $j \in \mathbf{B}$. These are

[9] $P^*_j(z,\theta_0)$

$$= \frac{\exp(v_j(z,\theta))H^*_j(\exp(v_1(z,\theta))R(1,z)S(\boldsymbol{B}|1,z),...,\exp(v_J(z,\theta))R(J,z)S(\boldsymbol{B}|J,z))R(j,z)S(\boldsymbol{B}|j,z)}{\mu H^*(\exp(v_1(z,\theta))R(1,z)S(\boldsymbol{B}|1,z),...,\exp(v_J(z,\theta))R(J,z)S(\boldsymbol{B}|J,z))}$$

$$= \frac{\exp(v_j(z,\theta))H^*_j(\exp(v_1(z,\theta))R(1,z)S(\boldsymbol{B}|1,z),...,\exp(v_J(z,\theta))R(J,z)S(\boldsymbol{B}|J,z))R(j,z)S(\boldsymbol{B}|j,z)}{\sum_{i\in\boldsymbol{B}}\exp(v_i(z,\theta))H^*_i(\exp(v_1(z,\theta))R(1,z)S(\boldsymbol{B}|1,z),...,\exp(v_J(z,\theta))R(J,z)S(\boldsymbol{B}|J,z))}$$

$$= \frac{\exp(v_j(z,\theta))H_j(\exp(v_1(z,\theta)),...,\exp(v_J(z,\theta)))R(j,z)S(\boldsymbol{B}|j,z)}{\sum_{i\in\boldsymbol{B}}\exp(v_i(z,\theta))H_i(\exp(v_1(z,\theta)),...,\exp(v_J(z,\theta)))R(i,z)S(\boldsymbol{B}|i,z)}\ ,$$

where the second equality follows from homogeneity of degree μ and the last equality from the construction of H^*. Then, $P^*_j(z,\theta) = \Pr(j|z,\theta)$, and the conditional probability of j given z in the sample is of GEV form on the apparent choice set **B**, which may be a proper subset of **C**, with the shifted locations

[10] $$v^*_j(z,\theta) = v_j(z,\theta) + \log R(j,z) + \log S(\mathbf{B}|j,z), \quad j \in \mathbf{B}.$$

Then, maximum likelihood methods for GEV choice probabilities in simple random samples can also be applied to non-random samples characterized by qualification probabilities $R(j,z)$ and alternative selection probabilities $S(\mathbf{B}|j,z)$ simply by incorporating location shifts $\log R(j,z) + \log S(\mathbf{B}|j,z)$ into the model for choice from the sampled choice set **B**. If the location functions $v_j(z,\theta)$ that determine the GEV population choice probabilities $P_j(z,\theta)$ are linear and additive in functions of j and z, with coefficients that are components of θ, and the location shift $\log R(j,z) + \log S(\mathbf{B}|j,z)$ is contained in the subspace spanned by these functions, then the effect of the sampling and selection is absorbed by the coefficients on these functions. In this simple case, the population GEV model [2] and the sampling-adjusted GEV model [9] are identical except for shifts in coefficients of functions that enter as explanatory variables, and all corrections for sampling can be made by first estimating the GEV model as if the sample were random (with choice set **B**), and then correcting the resulting estimates for the effects of the location shifts in [10].

The property that the family of choice probabilities obtained from GEV generating functions is closed under the effects of stratification and sampling of alternatives with positive conditioning allows analysis of stratified samples in which the qualification probabilities vary with interactions of j and z as well as with j alone. It also allows R(j,z) to contain unknown parameters, provided restrictions on $v_j(z,\theta)$ and R(j,z) are sufficient to identify the parameter vector θ and parameters embedded in R(j,z). In the case of no selection of alternatives, or *uniform conditioning* where S(**B**|j,z) is the same for every j $\in$ **B**, the correction term log S(**B**|j,z) drops out of [9], and in the case of exogenous or pseudo-exogenous sampling where R(j,z) is independent of j, the correction term log R(j,z) drops out of [9].

References

Johnson, N.; Kotz, S. (1970) CONTINUOUS UNIVARIATE DISTRIBUTIONS - 1, Holden-Day: San Francisco.

McFadden, D. (1978) "Modeling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), SPATIAL INTERACTION THEORY AND PLANNING MODELS, 75-96, North Holland: Amsterdam, Reprinted in J. Quigley (ed.), THE ECONOMICS OF HOUSING, Vol. I, 531-552, Edward Elgar: London, 1997.

McFadden, D. (1981) "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden (eds), STRUCTURAL ANALYSIS OF DISCRETE DATA, MIT Press: Cambridge, 198-272.

McFadden, D. (1984) "Econometric Analysis of Qualitative Response Models," in Z. Griliches and M. Intriligator (eds) HANDBOOK OF ECONOMETRICS, 2, Elsevier Science: Amsterdam, 1395-1457.

McFadden, D. (2001) "Observational Studies: Choice-based Sampling," INTERNATIONAL ENCYCLOPEDIA OF SOCIAL AND BEHAVIORAL SCIENCES, Vol. 2.1, Article 92, Elsevier Science: Amsterdam.