Stochastic orders of variability

Hans Colonius

Revision August 2015

Introduction: Some Variability Orders

The fact that the response to one and the same stimulus, under seemingly identical conditions, varies from one instance to the next is a fundamental observation of reaction time (RT) research and, arguably, of behavioral science in general. Moreover, recent neurophysiological modeling approaches seek to identify the sources of variability at the neuronal level (e.g., Lin et al., 2015; Goris et al., 2014; van den Berg et al., 2012) and often try relate them to trial-to-trial variability in behavior. In psychology, the recognition that the complete distribution function of RTs may contain critical information about the underlying stochastic mechanism not accessible by only considering their mean, has led to the development of many RT models predicting specific distribution functions, e.g. race, counter, and diffusion models (Luce, 1986). However, these models are often testable only if one takes specific distributional assumptions for granted, even if these assumption are not part of the models' core assumptions (for a recent discussion, see Jones and Dzhafarov, 2014).

Much of Townsend's work is in a different tradition, however (e.g., Townsend and Eidels, 2011; Townsend and Wenger, 2004; Townsend and Nozawa, 1995). Order relations between response time (RT) distributions that compare the "location" or the "magnitude" of random variables –like stochastic order, hazard rate order, or likelihood order– have played an important role in his work (e.g., Townsend, 1990). More recently, we addressed the issue of comparing RT distributions with respect to measures of variability (Townsend and Colonius, 2005). In this chapter, I recollect some results of the latter paper and relate them to some recent developments in statistics.

While specific distributions are typically described by certain moments (mean and variance) and transforms thereof, in particular skewness and kurtosis, this approach has certain shortcomings. For example, the variance may not be finite. Several notions of variability, weaker and stronger than the variance, have been investigated in the statistical literature (cf., Shaked and Shantikumar, 2007)[1]. The first variability order considered here requires the definition of a *convex function* $\phi : \Re \longrightarrow \Re$, i.e.,

$$\phi(\alpha x + (1-\alpha)y) \leq \alpha\phi(x) + (1-\alpha)\phi(y),$$

for all $x, y \in \Re$ and $\alpha \in [0, 1]$.

Definition 1 (Convex order). [2] Let X and Y be two random variables such that

$$\mathrm{E}\phi(X) \leq \mathrm{E}\phi(Y) \ \text{ for all convex functions } \phi : \Re \longrightarrow \Re,$$

provided the expectations exist. Then X is said to be *smaller than Y in the convex order*, denoted as

$$X \leq_{cx} Y.$$

---

[1]If not indicated otherwise, proofs of all statement made here can be found in this monograph.
[2]Definitions, examples, and propositions are numbered jointly and consecutively.

Roughly speaking, convex functions take on their (relatively) larger values over regions of the form $(-\infty, a) \cup (b, +\infty)$ for $a < b$. Therefore, when the condition in the above definition holds, Y is more likely to take on "extreme" values than X, that is, Y is "more variable" than X.[3]

The convex order has some strong implications. Using the convexity of functions $\phi(x) = x, \phi(x) = -x$, and $\phi(x) = x^2$, it is can be shown that

$$X \leq_{cx} Y \quad \text{implies} \quad \mathrm{E}X = \mathrm{E}Y \quad \text{and} \quad \mathrm{Var}X \leq \mathrm{Var}Y.$$

Moreover, when $\mathrm{E}X = \mathrm{E}Y$, then a condition equivalent to $X \leq_{cx} Y$ is

$$\mathrm{E}[\max(X, a)] \leq \mathrm{E}[\max(Y, a)], \quad \text{for all } a \in \Re.$$

The convex order only compares random variables that have the same means. One way to drop this requirement is to introduce a so-called *dilation order* by

Definition 2 (Dilation order).

$$X \leq_{dil} Y, \quad \text{if} \quad [X - \mathrm{E}X] \leq_{cx} [Y - \mathrm{E}Y].$$

For nonnegative random variables X and Y with finite means, one can alternatively define the *Lorenz order* by

$$X \leq_{Lorenz} Y, \quad \text{if} \quad \frac{X}{\mathrm{E}X} \leq_{cx} \frac{Y}{\mathrm{E}Y},$$

which can be used to order random variables with respect to the *Lorenz curve* used, e.g., in economics to measure the inequality of incomes.

For the next order type, we need the concept of a *quantile function*, closely related to that of a distribution function.

Definition 3 (Quantile function). Let X be a real-valued random variable with distribution function $F(x)$. Then the *quantile function* of X is defined as

$$Q(u) = F^{-1}(u) = \inf\{x \mid F(x) \geq u\}, \quad 0 \leq u \leq 1. \tag{1}$$

For every $-\infty < x < +\infty$ and $0 < u < 1$, we have

$$F(x) \geq u \quad \text{if, and only if,} \quad Q(u) \leq x.$$

Thus, if there exists x with $F(x) = u$, then $F(Q(u)) = u$ and $Q(u)$ is the smallest value of x satisfying $F(x) = u$. If $F(x)$ is continuous and strictly increasing, $Q(u)$ is the unique value x such that $F(x) = u$. The specification of a distribution through its quantile

---

[3]Clearly, it is sufficient to consider only functions ϕ that are convex on the union of the supports of X and Y rather than over the whole real line.

function takes away the need to describe a distribution through its moments. Alternative measures in terms of quantiles are available, e.g. the median as a measure of location, defined as $md(X) = Q(0.5)$, or the interquartile range, as a measure of dispersion, defined as $IQR = Q(.75) - Q(.25)$.

The following order is based on comparing difference between quantiles of the distribution functions.

Definition 4 (Dispersive order)**.** Let X and Y be random variables with quantile functions F^{-1} and G^{-1}, respectively. If

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha), \quad \text{whenever} \quad 0 < \alpha \leq \beta < 1,$$

then X is said to be *smaller than Y in the dispersive order*, denoted by $X \leq_{disp} Y$.

In contrast to the convex order, the dispersive order is clearly *location free*:

$$X \leq_{disp} Y \Leftrightarrow X + c \leq_{disp} Y, \quad \text{for any real } c.$$

The dispersive order is also *dilative*, i.e., $X \leq_{disp} aX$ whenever $a \geq 1$ and, moreover,

$$X \leq_{disp} Y \Leftrightarrow -X \leq_{disp} -Y.$$

However, it is not closed under convolutions. Its characterization requires two more definitions.

First, a random variable Z is called *dispersive* if $X + Z \leq_{disp} Y + Z$ whenever $X \leq_{disp} Y$ and Z is independent of X and Y. Second, a density (more general, any nonnegative function) g is called *logconcave* if $\log g$ is concave, or, equivalently,

$$g[\alpha x + (1 - \alpha)y] \geq [g(x)]^{\alpha}[g(y)]^{1-\alpha}.$$

Proposition 5. *The random variable X is dispersive if, and only if, X has a logconcave density.*[4]

It can be shown that the dispersion order implies the dilation order, that is, for random variables with finite means,

$$X \leq_{disp} Y \Longrightarrow X \leq_{dil} Y,$$

from which it immediately follows that also $\text{Var}X \leq \text{Var}Y$.

---

[4]Note that a logconcave density follows from its distribution function G being *strongly unimodal*, i.e., if the convolution $G * F$ is unimodal for every unimodal F.

The Quantile Spread

All variability orders considered so far are strong enough to imply a corresponding ordering of the variances. Moreover, if X and Y have the same finite support and satisfy $X \leq_{disp} Y$, then they must have the same distribution. Hence a weaker concept is desirable. In Townsend and Colonius (2005), we developed such a weaker concept of variability order in an attempt to describe the effect of sample size on the shape of the distribution of the extreme order statistics $X_{1:n}$ (minimum) and $X_{n:n}$ (maximum). It is based on the notion of *quantile spread*:

Definition 6. The *quantile spread* of random variable X with distribution F is

$$QS_X(p) = F^{-1}(p) - F^{-1}(1 - p),$$

for $0.5 < p < 1$.

With $S = 1 - F$ (the survival function), $F^{-1}(p) = S^{-1}(1 - p)$ implies

$$QS_X(p) = S^{-1}(1 - p) - S^{-1}(p).$$

It turns out that the concept of a *spread function*, as already defined by Balanda and MacGillivray (1990), is equivalent to the quantile spread. The quantile spread of a distribution describes how the probability mass is placed symmetrically about its median and hence can be used to formalize concepts such as peakedness and tailweight traditionally associated with kurtosis. This way, it allows us to separate concepts of kurtosis and peakedness for asymmetric distributions.

Example 7 (Weibull)**.** For the Weibull distribution (see Figure 1) defined by

$$F(x) = 1 - \exp(-\beta x^\alpha) \text{ with } \alpha, \beta > 0,$$

the quantile spread is

$$QS_X(p) = \beta^{-1/\alpha} \left[\left(\ln \frac{1}{1-p} \right)^{1/\alpha} - \ln \left(\frac{1}{p} \right)^{1/\alpha} \right]$$

for $0.5 < p < 1$.

For $\alpha = 1$ (exponential distribution with parameter β) one gets

$$QS_X(p) = (1/\beta) \ln \frac{p}{1-p}.$$

For $\beta = 1/2$, this equals the quantile spread function for the (standard) *logistic distribution*,

$$F(x) = \frac{1}{1 + \exp(-x)}, \quad -\infty < x < +\infty.$$

NOTE: This show that the quantile spread does not uniquely characterize the distribution function from which it has been generated.
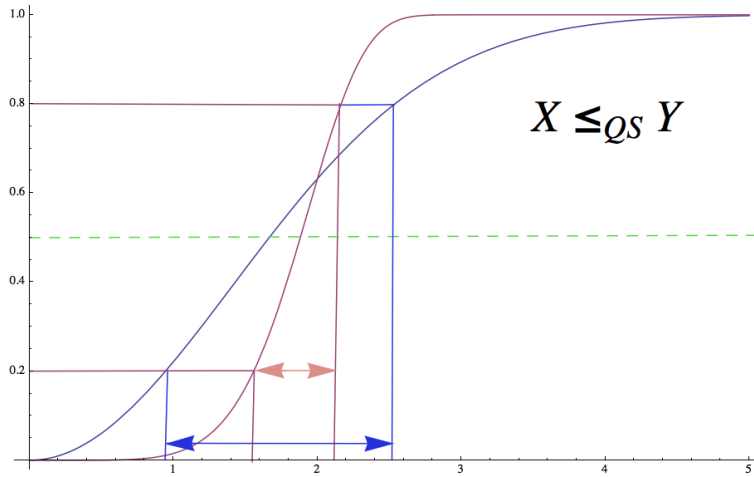
Figure 1: Two Weibull distributions with $\alpha = 2$ and $\alpha = 6$, and $\beta = 2$. The horizontal arrows indicate their quantile spread at $p = 0.8$. The order $\leq_{QS}$ is defined in the next section.

Example 8 (Cauchy distribution). The Cauchy distribution with density

$$f(x) = \frac{\gamma}{\pi(x^2 + \gamma^2)}, \quad \gamma > 0,$$

has no finite variance; the quantile function is

$$F^{-1}(p) = \gamma \tan[\pi(p - 0.5)],$$

yielding the quantile spread

$$F^{-1}(p) - F^{-1}(1 - p) = \gamma\{\tan[\pi(p - 0.5)] - \tan[\pi(0.5 - p)]\},$$

for $0.5 < p < 1$.

Example 9 (Extreme order statistics). For the extreme order statistics, one gets simple expressions for the quantile spread, making it easy to describe their behavior as a function of sample size n:

$$QS_{min}(p) = S^{-1}[(1 - p)^{1/n}] - S^{-1}(p^{1/n})$$

and

$$QS_{max}(p) = F^{-1}(p^{1/n}) - F^{-1}[(1 - p)^{1/n}].$$

Not surprisingly, the case of a symmetric parent distribution leads to a simplification with respect to the order statistics:

Proposition 10. *For a symmetric parent distribution, the quantile spread for the maximum is identical to the quantile spread for the minimum.*

Proof. (For a proof assuming symmetry around zero, see Townsend and Colonius, 2005, p. 765. The general case follows similarly.). $\square$

The Quantile Spread Order

Definition 11. Let X and Y be random variables with quantile spreads QS_X and QS_Y, respectively. Then X is called *smaller than Y in quantile spread order*, denoted as $X \leq_{QS} Y$, if

$$QS_X(p) \leq QS_Y(p), \quad \text{for all } p \in (0.5, 1).$$

The following properties of the quantile spread order can be determined:

1. The order $\leq_{QS}$ is *location-free*, i.e.,

$$X \leq_{QS} Y \;\; \text{iff} \;\; X + c \leq_{QS} Y, \text{ for any real } c$$

2. The order $\leq_{QS}$ is *dilative*, i.e, $X \leq_{QS} aX$, whenever $a \geq 1$.

3.
$$X \leq_{QS} Y \;\; \text{if, and only if} \;\; -X \leq_{QS} -Y.$$

4. Assume F_X and F_Y are symmetric, then

$$X \leq_{QS} Y \;\; \text{if, and only if,} \;\; F_X^{-1}(p) \leq F_Y^{-1}(p) \text{ for } p \in (0.5, 1).$$

5. $\leq_{QS}$ implies ordering of the *mean absolute deviation around the median*, MAD,

$$\text{MAD}(X) = \text{E}[|X - \text{md}(X)|]$$

(md the median), i.e.,

$$X \leq_{QS} Y \;\; \text{implies} \;\; \text{MAD}(X) \leq \text{MAD}(Y).$$

The last point follows from writing

$$\text{MAD}(X) = \int_{0.5}^{1} [F^{-1}(p) - F^{-1}(1-p)] \, dp.$$

For the Cauchy distribution with quantile spread

$$QS(p) = \gamma\{\tan[\pi(p - 0.5)] - \tan[\pi(0.5 - p)]\},$$

it is obvious that parameter γ clearly defines the QS order for this distribution.

The next example demonstrates how the QS order can be used to describe the effect of sample size on the minimum order statistic.

Example 12. The quantile spread for the Weibull minimum (cf. Example 7) is

$$QS_{min}(p;n) = (n\lambda)^{-1/\alpha} \left[\left(\ln \frac{1}{1-p} \right)^{1/\alpha} - \left(\ln \frac{1}{p} \right)^{1/\alpha} \right].$$

For $p \in (0.5, 1)$, this is decreasing in n. Thus,

$$X_{1:n} \leq_{QS} X_{1:n-1},$$

i.e., the quantile spread for the Weibull minimum decreases as a function of sample size n.

The next section illustrates a recent application of the quantile spread order to define the dispersive properties of a not yet well known distribution.

The Quantile Spread Order for the Kumaraswamy Distribution

The density for the Beta distribution, frequently used in modeling, is:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1,$$

where Γ is the gamma function, $\alpha > 0$, and $\beta > 0$. It is flexible and has simple closed-form expressions for the moments. However, it has no closed-form solution for its quantile function $F^{-1}(p)$, which is often a drawback in simulation and estimation studies.

An alternative distribution is the *Kumaraswamy distribution* which is, in many respects, very similar to the Beta distribution. Its dispersion properties have recently been studied by Mitnik and Baek (2013). The distribution function is:

$$F(x) = 1 - \left[1 - \left(\frac{x-c}{b-c} \right)^p \right]^q$$

where $c < x < b$ and p, q are shape parameters with $p > 0, q > 0$ (Kumaraswamy, 1980). Let us denote this general form of the distribution by $K(p, q, c, b)$. In contrast to the Beta distribution, it has no closed-form expressions for the mean and variance, but a simple quantile function:

$$x = Q(u) = c + (b-c) \left[1 - (1-u)^{\frac{1}{q}} \right]^{1/p},$$

for $0 < u < 1$. One question addressed in Mitnik and Baek (2013) was how to reparametrize the distribution so that members of its distribution family can be compared with respect to their variability. From the quantile function, one obtains for the median

$$\text{md}(x) = c + (b-c)(1 - 0.5^{\frac{1}{q}})^{1/p}.$$

From this, the authors derive two re-parametrizations of the distribution, $K_p(\omega, d_p, c, b)$ and $K_q(\omega, d_q, c, b)$, where $d_p = p^{-1}$, $d_q = q^{-1}$, and $\omega = \text{md}(x)$ (for details, see Mitnik and Baek,

2013, p. 181). They go on to show that d_p and d_q are in fact dispersion parameters with reference to a well-defined dispersion order. Note that this order cannot be the dispersive order $\leq_{disp}$ (cf. Definition 4): the Kumaraswamy distribution has a finite support, so any two distributions with the same b and c parameters ordered by $\leq_{disp}$ would be identical (see Shaked and Shantikumar, 2007, Theorem 3.B.14). This argument, however, does not rule out the dilation order which is derived from the convex order (see Definition 2). In an additional study (see supplement to Mitnik and Baek, 2013), they show, via numerical computations, that the dilation order is not an order consistent with values of parameters d_p and d_q but that the quantile spread order is. They prove the following

Proposition 13. *Let X be distributed as $K_r(\omega, d_{rx}, c, b)$ and Y as $K_r(\omega, d_{ry}, c, b)$; then $X \leq_{QS} Y$ if, and only if, $d_{rx} \leq d_{ry}$, with $r = p, q$.*

Trivially, as well the interquartile range, defined as $QS_X(.75)$ is consistent with values of the parameters d_p and d_q. Finally, it is also follows from the above that

$$\text{if} \quad d_{rx} \leq d_{ry}, \quad \text{then} \quad \text{MAD}(X) \leq \text{MAD}(Y),$$

for $r = p, q$.

Thus, Mitnik and Baek (2013) conclude that the only known variability order that is appropriate for the Kumaraswamy distribution is the quantile spread order defined in Townsend and Colonius (2005) (see also p. 151, Shaked and Shantikumar, 2007).

Acknowledgment

Bibliography

Balanda, K. and MacGillivray, H. (1990). Kurtosis and spread. *The Canadian Journal of Statistics*, 18(1):17–30.

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–865.

Jones, M. and Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121:1–32.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46:79–88.

Lin, I.-C., Okun, M., Carandini, M., and Harris, K. D. (2015). The nature of shared cortical variability. *Neuron*, 87:644–656.

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press, New York, NY.

Mitnik, P. and Baek, S. (2013). The kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54:177–192.

Shaked, M. and Shantikumar, J. (2007). *Stochastic orders.* Springer Science+Business Media, New York, NY.

Townsend, J. (1990). Truth and consequences of ordinal differences in statistical distributions: toward a theory of hierarchical inference. *Journal of Mathematical Psychology*, 108(551–567).

Townsend, J. and Colonius, H. (2005). Variability of the max and min statistic: a theory of the quantile spread as a function of sample size. *Psychometrika*, 70(4):759–772.

Townsend, J. and Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin and Review*, 18:659–681.

Townsend, J. and Nozawa, G. (1995). Spatio-temporal properties of elementary perception: an investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39(4):321–359.

Townsend, J. and Wenger, M. (2004). A theory of interactive parallel processing: new capacity measures and predictions for a response time inequality series. *Psychological Review*, 111(4):1003–1035.

van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22):8780–8785.