STOCHASTIC ORDERING OF ORDER STATISTICS

A.D. BARBOUR,* *Universität Zürich,*

T. LINDVALL, ** *University of Göteborg,*

L.C.G. ROGERS, *** *University of Cambridge.*

Abstract

If X_i, $i = 1, \ldots, n$ are independent exponential random variables with parameters $\lambda_1, \ldots, \lambda_n$, and if Y_i, $i = 1, \ldots, n$ are independent exponential random variables with common parameter equal to $(\lambda_1 + \cdots + \lambda_n)/n$, then there is a *monotone coupling* of the order statistics $X_{(1)}, \ldots X_{(n)}$ and $Y_{(1)}, \ldots Y_{(n)}$; that is, it is possible to construct on a common probability space random variables X_i', Y_i', $i = 1, \ldots, n$, such that for each i, $Y_{(i)}' \leq X_{(i)}'$ a.s., where the law of the X_i' (repectively, the Y_i') is the same as the law of the X_i (respectively, the Y_i.) This result is due to Ball. We shall here offer a new proof, extended to a more general class of distributions for which the failure rate function $r(x)$ is decreasing, and $xr(x)$ is increasing. This very strong order relation allows comparison of properties of epidemic processes where rates of infection are not uniform with the corresponding properties for the homogeneous case. We further prove that for a sequence Z_i, $i = 1, \ldots, n$ of independent random variables whose failure rates at any time add to 1, the order statistics are stochastically larger than the order statistics of a sample of n independent exponential random variables of mean n, but that the strong monotone coupling referred to above is impossible in general.

---

* Postal Address: Institut für Angewandte Mathematik, Universität Zürich, Rämistrasse 74, CH-8001, Zürich, Switzerland.

** Postal Address: Department of Mathematics, University of Göteborg, Sven Hultins gata 6, S-412 96 Göteborg, Sweden.

*** Postal Address: Statistical Laboratory, 16 Mill Lane, Cambridge CB2 1SB, Great Britain.

ORDER STATISTICS; HOMOGENEOUS SAMPLE; HETEROGENEOUS SAMPLE; EPIDEMIC; STOCHASTIC ORDERING

Stochastic ordering of order statistics

1. Introduction.

A model describing the spread of an epidemic among a population of size N can be constructed as follows [Sellke (1983), Ball (1985)]. Each individual i independently samples the length of his infectious period T_i from a distribution F and his resistance U_i to the disease from a distribution G. Initially, one or more members of the population are deemed to be infective. Anyone who becomes infected transmits infection at a fixed rate α to every other member of the population for as long as his infectious period lasts, and an individual becomes infected when the total amount of infection that he has received exceeds his resistance. Thus the course of the epidemic is determined by the order statistics of the T's and U's, the permutation of $\{1, 2, ..., N\}$ which associates the correct U-value with each T-value, and the ranks of the T-values belonging to the initial infectives.

Detailed analysis of the model is difficult, though a number of asymptotic results are known. However, if the model is further generalized, allowing the U_i's, for instance, to be drawn from different distributions F_i, corresponding to the realistic assumption of the existence of groups of individuals of differing susceptibility, even the asymptotic analysis becomes intractable. This led Ball (1985), who was working with exponential distributions, to compare the behaviour of the heterogeneous model with a suitably chosen homogeneous model. The remarkable result that he was able to prove is as follows: if $Z_1, \ldots, Z_N$ are independent negative exponential random variables with means $\lambda_1^{-1}, \ldots, \lambda_N^{-1}$, and if $\hat{Z}_1, \ldots, \hat{Z}_N$ are independent and identically distributed negative exponential random variables with mean $\hat{\lambda}^{-1}$, where $\hat{\lambda} = n^{-1} \sum_i \lambda_i$, then the vector of the order statistics of the $\hat{Z}$'s is stochastically smaller than that of the Z's. Thus the epidemic is stochastically more severe than the homogeneous counterpart defined in this way, if the T's are heterogeneous, but less severe if the U's are heterogeneous.

The purpose of this paper is to examine the comparison between the order statistics of a set of independent random variables and those of a homogenized set, in greater generality. Two types of result are proved. In the first, Ball's result for negative exponential random variables is extended to a much wider class of scale families.

Let $V_1, \ldots, V_n$ be independent identically-distributed non-negative random variables with common distribution function F. Define two samples $X_i \equiv V_i/\lambda_i, Y_i \equiv V_i/\hat{\lambda}, i = 1, \ldots, n$, where the λ_i are positive constants with average $\hat{\lambda}$, and let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order-statistics of $X_1, \ldots, X_n$ (with the corresponding notation for the Y-sample.) This we refer to as Model 1. Then, under suitable conditions on F, we can prove the existence of a *monotone coupling* of the X and Y *order statistics* showing that it is possible to construct on some probability space random variables $X'_1, \ldots, X'_n$ and $Y'_1, \ldots Y'_n$ such that

$$
\begin{aligned}
&(X'_1, \ldots, X'_n) \stackrel{\mathcal{D}}{=} (X_1, \ldots, X_n) : \\
(1) \qquad &(Y'_1, \ldots, Y'_n) \stackrel{\mathcal{D}}{=} (Y_1, \ldots, Y_n) : \\
&Y'_{(j)} \leq X'_{(j)} \text{ a.s. for } j = 1, \ldots, n.
\end{aligned}
$$

This is a very strong statement, which trivially implies the stochastic ordering of $(Y_{(j)}, X_{(j)})$ for each j. We prove this as Theorem 1, under the conditions that F has a decreasing failure-rate function r, and that the function $x \mapsto xr(x)$ is increasing. The paradigm example is the exponential distribution, considered by Ball, for which r is constant.

Of course, if random variables $\{X_i\}$ are related by scale change as in Model 1, so are their α-th powers $\{X_i^\alpha\}$, being constructed from the independent identically distributed sequence $\{V_i^\alpha\}$ using the constants $\{\lambda_i^\alpha\}$. Now, for this new set of constants, there is a new parameter $\hat{\lambda}_\alpha = n^{-1} \sum_i \lambda_i^\alpha$, which is in general different from $\hat{\lambda}^\alpha$. Hence, for any $\alpha > 0$ for which Theorem 1 holds for the X_i^α, the stochastic ordering between the order statistics of the X's and Y's holds true, with the Y's defined using $(\hat{\lambda}_\alpha)^{1/\alpha}$ in place of $\hat{\lambda}$. The sharpest comparison is thus obtained by choosing α to make $(\hat{\lambda})^{1/\alpha}$ as small as possible: that is, by taking α as small as possible. Now, if the V_i have failure rate function

r, the V_i^α have failure rate function $t \to r_\alpha(t) \propto t^{1/\alpha-1}r(t^{1/\alpha})$. Hence the conditions to be satisfied if Theorem 1 is to be applied to the V_i^α are that $s^{1-\alpha}r(s)$ should be decreasing, and that $sr(s)$ should be increasing. In particular, if the V's are exponential, r is constant, and thus $\alpha = 1$ is the best choice: Ball's result cannot be sharpened in this way. However, if the V's have decreasing failure rate $r(t) \propto t^{-1/\beta}$ for $0 < \beta < 1, \alpha$ can be taken to be $1-\beta$, giving a sharper comparison than that obtained directly with $\alpha = 1$. Note that the case $\beta = 1$ is not included, because then $r(t)$ is not finitely integrable at 0.

The second type of result which we consider is motivated by the following observation. Let $X_1, \ldots, X_N$ be independent random variables with distributions

$$P(X_j > t) = \exp\{-R_j(t)\},$$

so that R_j denotes the integrated failure rate for the variable X_j. In Model 1 above, $R_j(t) = R(\lambda_j t)$, where $R(\cdot) = \int_0^\cdot r(s)ds$, and, if r is decreasing, the concavity of R implies that

$$(**) \qquad \sum_i R_i(t) \le nR(\hat{\lambda}t)$$

for all t. Now, in view of the formulae

$$P(\min_i X_i > t) = \exp\{-\sum_i R_i(t)\}; \qquad P(\min_i Y_i > t) = \exp\{-nR(\hat{\lambda}t)\},$$

it is clear that (**) is necessary for the minimum $Y_{(1)}$ of the Y's to be stochastically smaller than $X_{(1)}$. However, except in the case of exponential random variables, there are values of t for which

$$\sum_i R_i(t) = \sum_i R(\lambda_i t) < nR(\hat{\lambda}t),$$

and, if r is strictly decreasing, strict inequality holds for all t. Thus, requiring r to be decreasing is rather too strong, if only the stochastic ordering of $Y_{(1)}$ and $X_{(1)}$ is of interest. This suggests the following question. Suppose, in Model 2, that the R_j's are now allowed to be arbitrary, and that the independent and identically distributed random variables Y_i are defined to have distribution given by

$$P(Y_i > t) = \exp\{-n^{-1} \sum_j R_j(t)\} :$$

are all the order statistics of the Y's still stochastically smaller than those of the X's? The answer turns out to be that $Y_{(j)}$ is indeed stochastically smaller than $X_{(j)}$ for each j, but that the strong monotone coupling result (1) is in general impossible. Although this result is not strong enough to be useful in the epidemic context above, it has obvious application in reliability theory, where, for instance, component failure determined by the failure of k out of n elements is a property of the k-th order statistic alone. Note that, once again, Ball's setting is recovered in the case of exponential random variables.

2. Monotone Coupling

LEMMA 1. *Let $\xi_1, \xi_2, \eta_1, \eta_2$ be independent non-negative random variables with distributions*

$$P(\xi_i > t) = \overline{F}(\lambda_i t), P(\eta_i > t) = \overline{F}(\hat{\lambda} t) \quad , i = 1, 2,$$

where λ_1, λ_2 are positive constants and $\hat{\lambda} = \frac{1}{2}(\lambda_1 + \lambda_2)$. Assume that $\overline{F}(t) = \exp\{-\int_0^t r(s)ds\}$, where $x \mapsto r(x)$ is decreasing, and $x \mapsto xr(x)$ is increasing. Then on some probability space there exist random variables $\xi'_1, \xi'_2, \eta'_1, \eta'_2$ such that $(\xi'_1, \xi'_2) \stackrel{\mathcal{D}}{=} (\xi_1, \xi_2), (\eta'_1, \eta'_2) \stackrel{\mathcal{D}}{=} (\eta_1, \eta_2)$ and

$$\eta'_{(1)} \leq \xi'_{(1)} \, , \, \eta'_{(2)} \leq \xi'_{(2)} \quad \text{a.s.}$$

Proof. Firstly, we compute

$$\begin{aligned} P(\xi_{(1)} > t) &= \overline{F}(\lambda_1 t)\overline{F}(\lambda_2 t) \\ &= \exp\{-R(\lambda_1 t) - R(\lambda_2 t)\} \\ &\geq \exp\{-2R(\hat{\lambda} t)\} \\ &= P(\eta_{(1)} > t), \end{aligned} \tag{2}$$

exploiting the concavity of $R(\cdot) \equiv \int_0^{\cdot} r(s)ds$.

Next, we take any $0 \leq \upsilon \leq u$ such that

$$2R(\hat{\lambda}\upsilon) = R(\lambda_1 u) + R(\lambda_2 u)$$

(so that $P(\eta_{(1)} > \upsilon) = P(\xi_{(1)} > u)$), and show that for $t \geq u$,

$$P(\xi_{(2)} \geq t | \xi_{(1)} = u) \geq P(\eta_{(2)} \geq t | \eta_{(1)} = \upsilon). \tag{3}$$

Indeed, writing $\alpha_i = \lambda_i r(\lambda_i u), \theta_i = \alpha_i/(\alpha_1 + \alpha_2)$ for $i = 1, 2$, we have that

$$\begin{aligned} \log P(\xi_{(2)} \geq t | \xi_{(1)} = u) &= \log \left\{ \theta_1 e^{-R(\lambda_2 t) + R(\lambda_2 u)} + \theta_2 e^{-R(\lambda_1 t) + R(\lambda_1 u)} \right\} \\ &\geq -\left\{ \theta_1 (R(\lambda_2 t) - R(\lambda_2 u)) + \theta_2 (R(\lambda_1 t) - R(\lambda_1 u)) \right\}, \end{aligned} \tag{4}$$

and also

$$\begin{aligned} \log P(\eta_{(2)} \geq t | \eta_{(1)} = \upsilon) &= -R(\hat{\lambda} t) + R(\hat{\lambda}\upsilon) \\ &= -R(\hat{\lambda} t) + \tfrac{1}{2}(R(\lambda_1 u) + R(\lambda_2 u)). \end{aligned} \tag{5}$$

Subtracting (4) and (5) yields

$$\log P(\xi_{(2)} \geq t | \xi_{(1)} = u) - \log P(\eta_{(2)} \geq t | \eta_{(1)} = \upsilon)$$

$$\geq R(\hat{\lambda} t) - \tfrac{1}{2} R(\lambda_1 t) - \tfrac{1}{2} R(\lambda_2 t) - \delta\{R(\lambda_1 t) - R(\lambda_2 t)\}$$

$$+ \delta\{R(\lambda_1 u) - R(\lambda_2 u)\},$$

where $\delta \equiv \theta_2 - \frac{1}{2}$

$$\geq \delta\{R(\lambda_2 t) - R(\lambda_1 t)\} - \delta\{R(\lambda_2 u) R(\lambda_1 u)\}$$

$$\geq 0,$$

because, under the hypothesis that $xr(x)$ is increasing, we have $\delta \geq 0$ and $R(\lambda_2 t) - R(\lambda_1 t) \geq R(\lambda_2 u) - R(\lambda_1 u)$.

Inequality (3) follows. □

We now construct $\xi'_{(i)}, \eta'_{(i)}$ from four independent $U[0,1]$ random variables U_1, U_2, U_3, U_4 in the obvious manner: obtain $\xi'_{(1)}, \eta'_{(1)}$ by applying the inverse distribution functions of $\xi_{(1)}, \eta_{(1)}$ to U_1, and then obtain $\xi'_{(2)}, \eta'_{(2)}$ by applying the inverse conditonal distribution functions of $\xi_{(2)}|\xi_{(1)}, \quad \eta_{(2)}|\eta_{(1)}$ to U_2. Inequalities (2) and (3) guarantee $\eta'_{(1)} \leq \xi'_{(1)}, \eta'_{(2)} \leq \xi'_{(2)}$. Now to get ξ_1, ξ_2, we set

$$\xi_1 = \xi'_{(1)} \mathrm{I}_{\{U_3 \leq P(\xi_1 < \xi_2 | \xi_{(1)}, \xi_{(2)})\}} + \xi'_{(2)} \mathrm{I}_{\{U_3 > P(\xi_1 < \xi_2 | \xi_{(1)}, \xi_{(2)})\}},$$

disposing of the η_i in analogous fashion.

THEOREM 1. *Let $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ be independent non-negative random variables with distributions*

$$P(X_j > t) = \overline{F}(\lambda_j t) \equiv \exp(-\int_0^{\lambda_j t} r(s) ds), \tag{6}$$

$$P(Y_j > t) = \overline{F}(\hat{\lambda} t), \tag{7}$$

where $\lambda_1, \ldots, \lambda_n$ are positive constants with average $\hat{\lambda}$ and $\overline{F}$ is the tail of the distribution function $F \equiv 1 - \overline{F}$, with failure-rate function r. Assume

$$x \mapsto r(x) \quad \textit{is decreasing;} \tag{8.i}$$

$$x \mapsto x r(x) \quad \textit{is increasing.} \tag{8.ii}$$

Then it is possible to construct on some probability space random variables $(X'_1, \ldots X'_n) \stackrel{\mathcal{D}}{=} (X_1, \ldots, X_n)$ and $(Y'_1, \ldots, Y'_n) \stackrel{\mathcal{D}}{=} (Y_1, \ldots, Y_n)$ such that

$$Y'_{(j)} \leq X'_{(j)} \quad \textit{for} \quad j = 1, \ldots, n \quad \text{a.s.}$$

Proof. The proof proceeds by repeated application of Lemma 1. Start with an independent sequence $X_1, \ldots, X_n$ with distributions given by (6), and assume that the probability space supports an infinite sequence of independent $U[0,1]$ random variables. We now describe the first step of the procedure, which 'consolidates' X_1 and X_2. Make two independent $U[0,1]$ random variables W_1, W_2 by the recipe

$$W_1 = F_{(1)}(\tilde{X}_1) \quad , W_2 = F_{(2)}(\tilde{X}_2 | \tilde{X}_1)$$

where $F_{(1)}$ is the distribution function of $\tilde{X}_1 \equiv \min(X_1, X_2)$ and $F_{(2)}(\cdot|x)$ is the conditional distribution function of $\tilde{X}_2 \equiv \max(X_1, X_2)$ given $\tilde{X}_1$. Now apply the corresponding inverse distribution functions calculated using $\lambda_1 = \lambda_2 = \frac{1}{2}(\lambda_1+\lambda_2) \equiv c$ to generate two new random variables $\tilde{X}'_1, \tilde{X}'_2$ such that $\tilde{X}'_1 \le \tilde{X}_1, \tilde{X}'_2 \le \tilde{X}_2$, and such that $P(\tilde{X}'_1 > t) = \overline{F}(ct)^2, P(\tilde{X}'_2 \le t) = F(ct)^2$, as in the proof of the Lemma 1. Using one of the available $U[0,1]$ random variables, we can pick one of $\tilde{X}'_1, \tilde{X}'_2$ to be ξ_1 and the other to be ξ_2, each of ξ_1, ξ_2 having distribution function $t \mapsto F(ct)$. Because of Lemma 1, we have $\min\{\xi_1, \xi_2\} \le \tilde{X}_1, \max\{\xi_1, \xi_2\} \le \tilde{X}_2$. *Thus we have changed* $X_1, X_2, X_3, \ldots, X_n$ *into* $\xi_1, \xi_2, X_3, \ldots, X_n$ *in such a way that*

(9.i) $\xi_1, \xi_2, X_3, \ldots, X_n$ *are independent*

(9.ii) ξ_1, ξ_2 *have the common distribution* $P(\xi_i \le t) = F(ct)$;

(9.iii) $\min\{\xi_1, \xi_2\} \le \min\{X_1, X_2\}, \max\{\xi_1, \xi_2\} \le \max\{X_1, X_2\}$ a.s..

So the random variables with distributions parametrised by $\mathbf{\Lambda} \equiv \{\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n\}$ have been replaced by a sequence with distributions parametrised by $\mathbf{\Lambda}' \equiv \{\frac{1}{2}(\lambda_1+\lambda_2), \frac{1}{2}(\lambda_1+\lambda_2), \lambda_3, \ldots, \lambda_n\}$, without increasing the order-statistics. We now just go on doing this, successively 'consolidating' pairs of terms in the sequence with the greatest and smallest λ-value, always reducing the order-statistics, and bringing the λ_i closer to the average $\hat{\lambda}$ in that $\sum_{i \ne j} |\lambda_i - \lambda_j|$ is reduced at least geometrically fast towards zero, with common ratio $c_n \equiv [\binom{n}{2} - 1]/\binom{n}{2}$. Thus the order-statistics decrease to some limits $Z_1 \le Z_2 \le \ldots \le Z_n$, which have the same law as the order-statistics of a sample with distribution function $F(\hat{\lambda}t)$. Re-ordering $Z_1, \ldots, Z_n$ at random gives a sample $Y_1, \ldots, Y_n$ with distribution given by (7), whose order-statistics satisfy $Y_{(j)} \le X_{(j)}$ a.s..

3. Stochastic ordering of order-statistics.

We turn now to the second problem. By changing the time scale, it may be assumed without loss of generality that $\sum_i R_i(t) = t$ for all t, so that the Y's are negative exponentially distributed.

THEOREM 2. *Suppose that* $X_1, \ldots, X_n$ *are independent,* $P(X_i > t) = \exp\{-R_i(t)\}$, *where*

$$\sum_{i=1}^{n} R_i(t) = t \quad \textit{for all} \quad t.$$

If $Y_1, \ldots, Y_n$ *are independent, with common* $\exp(1/n)$, *distribution, then for each* $k = 1, \ldots, n$

$$Y_{(k)} \overset{\mathcal{D}}{\leq} X_{(k)}.$$

($X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are the order statistics).

Proof. Whatever n, $Y_{(1)}$ and $X_{(1)}$ have an $\exp(1)$ distribution. Also,

$$\begin{aligned} P(X_{(2)} \geq t) &= \sum_{i=1}^{n} \big(1 - e^{-R_i(t)}\big) e^{-(t - R_i(t))} + e^{-t} \\ &= \sum_{i=1}^{n} \big(e^{R_i(t) - t} - e^{-t}\big) + e^{-t} \\ &\geq n e^{-t} e^{t/n} - (n-1) e^{-t} \qquad \text{by AM-GM inequality} \\ &= P(Y_{(2)} \geq t). \end{aligned}$$

Now suppose that $Y_{(k)} \overset{\mathcal{D}}{\leq} X_{(k)}$ for all $k \leq n$, when $n = 1, 2, \ldots, N$. We proceed inductively to the case $N + 1$.

Let $C(t) \equiv \sum_{i=1}^{N} R_i(t)$. Then the variables $\tilde{X}_j \equiv C(X_j), j = 1, \ldots, N$, are N independent r.v.s. whose rates sum to 1. Let $\tilde{X}_{(1)} < \ldots < \tilde{X}_{(N)}$ be the order-statistics of this sample, $\tilde{X}'_{(1)} < \ldots < \tilde{X}'_{(N+1)}$ be the order-statistics of the sample enlarged by the inclusion of $\tilde{X}_{N+1} \equiv C(X_{N+1})$. Fixing t and temporarily abbreviating $C(t)$ to C, we have that for any $3 \leq k \leq N + 1$,

$$P(X_{(k)} \leq t) = P(\tilde{X}'_{(k)} \leq C)$$

$$= \begin{cases} P(\tilde{X}_{(k)} \le C)P(\tilde{X}_{N+1} > C) + P(\tilde{X}_{(k-1)} \le C)P(\tilde{X}_{N+1} \le C), & \text{if } k \le N; \\ P(\tilde{X}_{(N)} \le C)P(\tilde{X}_{N+1} \le C) & \text{if } k = N+1, \end{cases}$$

$$= \begin{cases} P(\tilde{X}_{(k)} \le C)e^{-R_{N+1}(t)} + P(\tilde{X}_{(k-1)} \le C)\big(1 - e^{-R_{N+1}(t)}\big) & \text{if } k \le N; \\ P(\tilde{X}_{(N)} \le C)\big(1 - e^{-R_{N+1}(t)}\big) & \text{if } k = N+1. \end{cases}$$

Notice that $t = C(t) + R_{N+1}(t)$ so, taking the easy case $k = N+1$ first, we have from the inductive hypothesis that (with $C(t)$ abbreviated to C)

$$\begin{aligned} P(\tilde{X}_{(N)} \le C)(1 - e^{-t+C}) &\le P(\tilde{Y}_{(N)} \le C)(1 - e^{-t+C}) \\ &= (1 - e^{-C/N})^N (1 - e^{-t+C}), \end{aligned}$$

where $\tilde{Y}_{(1)} < \ldots < \tilde{Y}_{(N)}$ are the order statistics of $(\frac{N}{N+1})Y_j$, $j = 1, 2, ..., N$, a sequence of i.i.d. $\exp(\frac{1}{N})$ random variables. Elementary calculus shows that this last expression, considered as a function of C, attains its unique maximum when $C = Nt/(N+1)$ and therefore $R_{N+1}(t) = t/(N+1)$, showing that

$$P(X_{(N+1)} \le t) \le (1 - e^{-t/(N+1)})^{N+1} = P(Y_{(N+1)} \le t).$$

Now we turn to the case $k \le N$ and once again estimate via the inductive hypothesis:

$$\begin{aligned} P(X_{(k)} \le t) &= P(\tilde{X}_{(k)} \le C)e^{-t+C} + P(\tilde{X}_{(k-1)} \le C)\big(1 - e^{-t+C}\big) \\ &\le P(\tilde{Y}_{(k)} \le C)e^{-t+C} + P(\tilde{Y}_{(k-1)} \le C)\big(1 - e^{-t+C}\big) \\ &= P(\tilde{Y}_{(k-1)} \le C) - e^{-t+C}P\big(\tilde{Y}_{(k-1)} \le C < \tilde{Y}_{(k)}\big) \\ &= P\big[B(N,p) \ge k-1\big] - e^{-t+C}P\big[B(N,p) = k-1\big] \end{aligned} \tag{10}$$

where $p = 1 - e^{-C/N}$ is the probability that $\frac{N}{N+1}Y_1$ is at most C. Now it is useful to note that

$$\frac{\partial}{\partial p}P\big[B(N,p) \ge k-1\big] = NP\big[B(N-1,p) = k-2\big],$$

so if we differentiate (10) with respect to C we get

$$\begin{aligned} &NP\big[B(N-1,p) = k-2\big]\frac{\partial p}{\partial C} - e^{-t+C}P\big[B(N,p) = k-1\big] \\ &\quad - e^{-t+C} \cdot \frac{\partial p}{\partial C} \cdot N\big\{P\big[B(N-1,p) = k-2\big] - P\big[B(N-1,p) = k-1\big]\big\} \end{aligned}$$

$$= NP\big[B(N-1,p)=k-2\big]\frac{1}{N}e^{-C/N}\big(1-e^{-t+C}\big) - e^{-t+C}P\big[B(N,p)=k-1\big]$$
$$+ e^{-t+C}e^{-C/N}P\big[B(N-1,p)=k-1\big]$$

$$= \big(1-e^{-t+C}\big)\binom{N-1}{k-2}p^{k-2}q^{N-k+2} - e^{-t+C}\binom{N}{k-1}p^{k-1}q^{N-k+1}$$
$$+ e^{-t+C}\binom{N-1}{k-1}p^{k-1}q^{N-k+1}$$

$$= \frac{(N-1)!}{(k-1)!(N-k+1)!}p^{k-2}q^{N-k+1}\Big\{(1-e^{-t+C})q(k-1) - e^{-t+C}Np + e^{-t+C}(N-k+1)p\Big\}$$
$$= \binom{N-1}{k-2}p^{k-2}q^{N-k+1}\Big\{q(1-e^{-t+C}) - pe^{-t+C}\Big\}$$
$$= \binom{N-1}{k-2}p^{k-2}q^{N-k+1}\Big\{q - e^{-t+C}\Big\}.$$

Thus the derivative vanishes if and only if $C = Nt/(N+1)$, which is easily seen to yield the unique maximum of the function.

The inductive hypothesis for $N+1$ now follows, since for this choice of C,

$$P(\tilde{Y}_{(k)} \le C)e^{-t+C} + P(\tilde{Y}_{(k-1)} \le C)(1-e^{-t+C}) = P(Y_{(k)} \le t).$$

□

To see that there can in general be no monotone coupling, consider the following example, where $n = 2$, and the failure-rate functions are given by

$$r_1(t) = \begin{cases} 1 & \text{for } t \le 1; \\ \eta & \text{for } t > 1, \end{cases}$$

$r_2(t) \equiv 1 - r_1(t)$, where η should be thought of as small. Suppose that there was a monotone coupling. Since $X_{(1)} \overset{\mathcal{D}}{=} Y_{(1)}$, this would imply that $X_{(1)} = Y_{(1)}$ a.s.. But now for $\tau < 1, \quad t > 1$

$$P(X_{(2)} > t | X_{(1)} = \tau) = \exp\big(-(t-1)(1-\eta)\big),$$
$$P(Y_{(2)} > t | Y_{(1)} = \tau) = \exp\big(-\tfrac{1}{2}(t-\tau)\big)$$
$$\le P\big(X_{(2)} > t | X_{(1)} = \tau\big)$$

for large enough t. Thus the conditional laws of $X_{(2)}, Y_{(2)}$ given $X_{(1)}$ are not stochastically ordered, so no monotone coupling is possible. This makes Theorem 2 seem even more surprising!

Acknowledgement. This work was assisted by conversations held between the first and last authors at the IMS meeting at Sheffield, August 1989.

References

BALL, F. (1985), Deterministic and stochastic epidemics with several kinds of susceptibles, *Adv. Appl. Probability* **17**, 1–22.

SELLKE, T. (1983), On the asymptotic distribution of the size of a stochastic epidemic, *J. Appl. Probability* **20**, 390–394.