

Cantor order statistics: without applications.

Barry C. Arnold

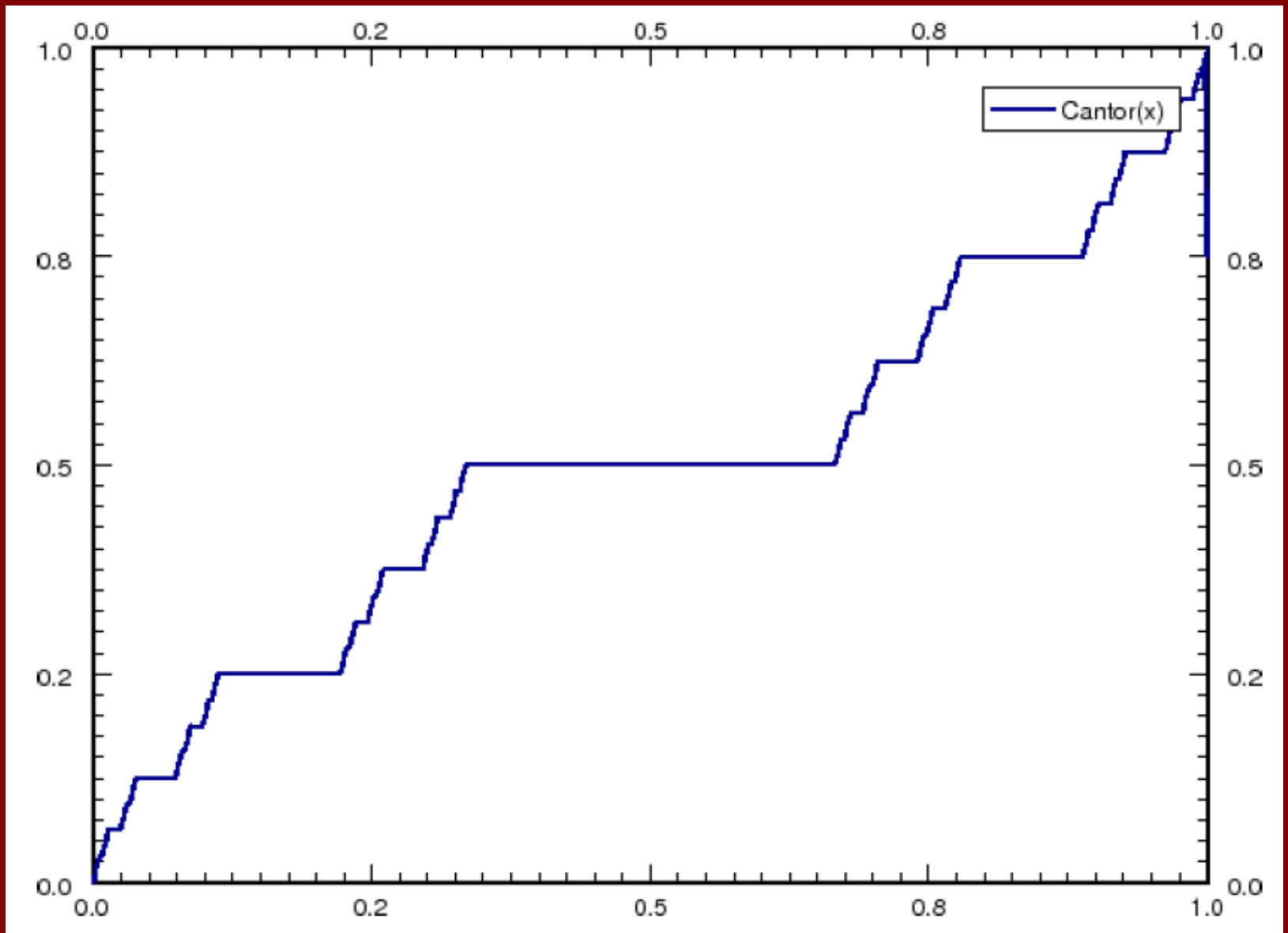
University of California

Riverside, CA, USA



December 16-19, 2011





A classical Cantor r.v. admits the representation:

$$X = \sum_{j=1}^{\infty} \frac{2B_j}{3^j}$$

in terms of i.i.d. Bernoulli(1/2) r.v.'s.

We consider the following generalized Cantor r.v.

$$X = \sum_{j=1}^{\infty} (1 - \phi) \phi^{j-1} B_j$$

In terms of i.i.d. Bernoulli (p) r.v.'s

Notation:

$$X \sim GC(\phi, p)$$

The case in which $p=1/2$ was investigated by Lad and Taylor(1992) and subsequently by Hosking (1994).

The classical Cantor case corresponds to $p=1/2$ and $\phi = 1/3$.

The general model

$$X = \sum_{j=1}^{\infty} (1 - \phi) \phi^{j-1} B_j$$

$$0 < p < 1$$

$$0 < \phi < 1/2$$

The constraint $0 < \phi < 1/2$ is needed to guarantee that we have a singular continuous r.v. , and thus a true analog of the classical Cantor r.v.

Since the generalized Cantor distribution has support in the interval $[0, 1]$, its right continuous inverse function (or quantile function) is also a valid distribution function.

A random variable Y with this quantile function as its distribution function will be said to have an inverse generalized Cantor distribution and

we write: $Y \sim IGC(\phi, p)$

Note this usage is not customary, but it seems reasonable

Compare:

Inverse gamma distribution

Inverse Gaussian distribution

- Properties of the generalized Cantor distr.
- A skewed version (idea:Hosking 1994).
- The corresponding inverse distribution.
- Bivariate and multivariate extensions.

Properties of the generalized Cantor distribution.

$$X = \sum_{j=1}^{\infty} (1 - \phi) \phi^{j-1} B_j$$

$$= (1 - \phi) B_1 + \phi \tilde{X}$$

where B_1 and \tilde{X} are independent

and $X \stackrel{d}{=} \tilde{X}$.

Properties of the generalized Cantor distribution.

The corresponding moment generating function can be easily identified from either of these two representations of X . Thus:

$$M_X(t) = [1 - p + pe^{(1-\phi)t}]M_X(\phi t).$$

$$M_X(t) = \prod_{j=1}^{\infty} [1 - p + pe^{(1-\phi)\phi^{j-1}t}].$$

Properties of the generalized Cantor distribution.

Using the recursive representation we can compute moments (as done by Lad and Taylor when $p=1/2$):

$$E(X^k) = E[\{(1 - \phi)B_1 + \phi\tilde{X}\}^k]$$

$$= \sum_{j=0}^{k-1} \binom{k}{j} p(1 - \phi)^{k-j} \phi^j E(X^j) + \phi^k E(X^k)$$

since $E(B_1^j) = p$ for $j > 0$.

Properties of the generalized Cantor distribution.

Thus:

$$E(X^k) = \frac{p \sum_{j=0}^{k-1} \binom{k}{j} (1 - \phi)^{k-j} \phi^j E(X^j)}{1 - \phi^k}, \quad k = 1, 2, \dots$$

In particular:

$$E(X) = \frac{p(1 - \phi)}{1 - \phi} = p,$$

$$\text{var}(X) = \frac{p(1 - p)(1 - \phi)}{1 + \phi}.$$

Properties of the generalized Cantor distribution.

Though both the mean and variance can be more easily obtained using the series representation

$$X = \sum_{j=1}^{\infty} (1 - \phi) \phi^{j-1} B_j$$

Properties of the generalized Cantor distribution.

Note: $E(X) > \text{var}(X)$

Properties of the generalized Cantor distribution.

Note: $E(X) > \text{var}(X)$

In fact X is more undersdispersed than a Bernoulli (p) r.v.

Properties of the generalized Cantor distribution.

Note: $E(X) > \text{var}(X)$

In fact X is more undersdispersed than a Bernoulli (p) r.v.

Note that if $X \sim GC(\phi, p)$

then $1 - X \sim GC(\phi, 1 - p)$.

Properties of the generalized Cantor distribution.

Consistent asymptotically normal estimates of the parameters are available via the method of moments:

$$\hat{p} = M_1,$$

$$\hat{\phi} = \frac{M_1 - M_2}{M - 1 + M_2 - 2M_1^2},$$

in which $M_1 = \frac{1}{n} \sum_{i=1}^n X_i$ and $M_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$.

Expected values of Order Statistics of the generalized Cantor distribution.

Here we follow Hosking (1994) who dealt
with the case $p=1/2$.

We wish to evaluate

$$\mu_{i:n} = E(X_{i:n})$$

But it will suffice to evaluate

$$\mu_{1:n} = E(X_{1:n})$$

Expected values of Order Statistics of the generalized Cantor distribution.

$$X = \begin{cases} \phi\tilde{X}, & \text{with probability } (1 - p) \\ (1 - \phi) + \phi\tilde{X}, & \text{with probability } p. \end{cases}$$

So

$$X < \phi \text{ with probability } (1 - p)$$

$$X > 1 - \phi \text{ with probability } p.$$

Expected values of Order Statistics of the generalized Cantor distribution.

We then can condition on the number of X 's
for which the corresponding B_1 's are
equal to 0, to obtain:

$$E(X_{1:n}) = \sum_{j=1}^n \binom{n}{j} (1-p)^j p^{n-j} \phi E(X_{1:j}) +$$

$$+ p^n [1 - \phi + \phi E(X_{1:n})].$$

Expected values of Order Statistics of the generalized Cantor distribution.

$$E(X_{1:1}) = E(X) = p,$$

$$E(X_{1:2}) = \frac{p^2(1 + \phi - 2p\phi)}{1 - \phi + 2p(1 - p)\phi}.$$

$$E(X_{2:2}) = 2E(X_{1:1}) - E(X_{1:2}) =$$

$$= p \frac{2 - 2\phi + 3p\phi - 2p^2\phi - p}{1 - \phi + 2p(1 - p)\phi},$$

Gini index of the generalized Cantor distribution

$$G(X) = \frac{E(X_{2:2}) - E(X_{1:2})}{E(X_{2:2}) + E(X_{1:2})}$$

$$= \frac{(1-p)(1-\phi)}{1-\phi + 2p(1-p)\phi}.$$

Gini index of the generalized Cantor distribution

If $p = 1/2$,

$$G(X) = \frac{1 - \phi}{2 - \phi}$$

Gini index of the generalized Cantor distribution

If $p = 1/2$,

$$G(X) = \frac{1 - \phi}{2 - \phi}$$

If $p = 1/2$, and $\phi = 1/3$,

$$G(X) = 2/5.$$

The skew generalized Cantor distribution.

Hosking introduced an asymmetric version of the generalized Cantor distr. In the case $p=1/2$. The extension to $0 < p < 1$ is as follows:

$$X = \begin{cases} \alpha \tilde{X} & \text{with probability } (1 - p), \\ (1 - \beta) + \beta \tilde{X} & \text{with probability } p. \end{cases}$$

$$\alpha > 0, \beta > 0, \alpha + \beta < 1 \text{ and } 0 < p < 1$$

The skew generalized Cantor distribution.

Recursion for moments:

$$E(X^k) = \frac{p \sum_{j=0}^{k-1} \binom{k}{j} (1-\beta)^{k-j} \beta^j E(X^j)}{1 - (1-p)\alpha^k - p\beta^k}$$

$$E(X) = \frac{p(1-\beta)}{1 - (1-p)\alpha - p\beta},$$

$$\begin{aligned} \text{var}(X) = & p \frac{(\beta-1)^2 (\alpha(p-1) + p\beta + 1)}{((p-1)\alpha^2 - p\beta^2 + 1) (\alpha(p-1) - p\beta + 1)} \\ & - \left(\frac{p(1-\beta)}{1 - (1-p)\alpha - p\beta} \right)^2. \end{aligned}$$

The skew generalized Cantor distribution.

Recursion for expected sample minima:

$$E(X_{1:n}) = \sum_{j=1}^n \binom{n}{j} (1-p)^j p^{n-j} \alpha E(X_{1:j}) +$$

$$+ p^n [1 - \beta + \beta E(X_{1:n})]$$

The skew generalized Cantor distribution.

Thus, for example

$$E(X_{1:2}) = \frac{2(1-p)p\alpha \frac{p(1-\beta)}{1-(1-p)\alpha-p\beta} + p^2(1-\beta)}{1 - (1-p)^2\alpha - p^2\beta}$$

Using which, we obtain the Gini index:

$$G(X) = \frac{(1-p)(1-\alpha)}{1 - (1-p)^2\alpha - p^2\beta}$$

Recall

For generalized Cantor distr.

$$G(X) = \frac{(1-p)(1-\phi)}{1-\phi+2p(1-p)\phi}.$$

While for skew generalized Cantor distr.

$$G(X) = \frac{(1-p)(1-\alpha)}{1-(1-p)^2\alpha-p^2\beta}$$

The inverse distribution.

If $X \sim SGC(\alpha, \beta, p)$ then we denote its distribution function by

$$F_{SGC}(x; \alpha, \beta, p)$$

Since $0 \leq X \leq 1$, the corresponding right continuous quantile function

$F_{SGC}^{-1}(x; \alpha, \beta, p)$ is also a valid distr. fn.

The inverse distribution.

If Y has $F_{SGC}^{-1}(x; \alpha, \beta, p)$ as its distr.fn.

then we write $Y \sim SGC^{-1}(\alpha, \beta, p)$ and

say that Y has an inverse-skew-gene

ralized-Cantor distribution.

The inverse distribution.

If $Y \sim SGC^{-1}(\alpha, \beta, p)$ then Y is discrete with a countable number of possible values. Like the corresponding r.v. X , it has a fractal structure.

The inverse distribution.

Thus Y can be described as follows:

$$Y = \begin{cases} (1-p)\tilde{Y} & \text{with probability } \alpha \\ (1-p) & \text{with probability } (1-\alpha-\beta) \\ (1-p) + p\tilde{Y} & \text{with probability } \beta \end{cases}$$

$$\alpha > 0, \beta > 0, \alpha + \beta < 1 \text{ and } 0 < p < 1$$

in which $\tilde{Y} \stackrel{d}{=} Y$.

The inverse distribution.

Using this representation we find:

$$E(Y) = \frac{(1 - \alpha)(1 - p)}{1 - \alpha + \alpha p - \beta p},$$

and eventually

$$\begin{aligned} \text{var}(Y) = & \frac{(1 - p)^2(1 - \alpha)}{1 - \alpha + \alpha p - \beta p} \frac{1 - \alpha + \alpha p + \beta p}{1 - \alpha(1 - p)^2 - \beta p^2} \\ & - \left(\frac{(1 - \alpha)(1 - p)}{1 - \alpha + \alpha p - \beta p} \right)^2. \end{aligned}$$

The inverse distribution.

For the inverse classical Cantor distribution
Corresponding to the choice

$$\alpha = \beta = 1/3 \text{ and } p = 1/2,$$

we have:

$$E(Y) = 1/2$$

$$\text{var}(Y) = 1/20.$$

On Moments of Inverse Distributions

Suppose that X satisfies $0 \leq X \leq 1$ and has d.f. F . Let Y have the corresponding inverse distribution, i.e., its d.f. is F^{-1} . How are the moments of X related to those of Y ?

On Moments of Inverse Distributions

Integrating by parts, one finds

$$\begin{aligned} E(Y^k) &= \int_0^1 y^k dF^{-1}(y) = \int_0^1 [F(x)]^k dx \\ &= x[F(x)]^k \Big|_0^1 - \int_0^1 x d[F(x)]^k \\ &= 1 - E(X_{k:k}) = 1 - E[(1 - X)_{1:k}] \end{aligned}$$

On Moments of Inverse Distributions

With $k=1$, we get

$$E(X) + E(Y) = 1$$

On Moments of Inverse Distributions

In the case of SGC variables, then

$$\begin{aligned} E(Y^k) &= \int_0^1 y^k dF^{-1}(y; \alpha, \beta, p) = \int_0^1 [F(x; \alpha, \beta, p)]^k dx \\ &= x[F(x; \alpha, \beta, p)]^k \Big|_0^1 - \int_0^1 x d[F(x; \alpha, \beta, p)]^k \\ &= 1 - E(X_{k:k}). \end{aligned}$$

$$= E((1 - X)_{1:k}) = E(Z_{1:k})$$

where the Z_i 's are i.i.d. $SGC(\beta, \alpha, 1 - p)$ r.v.'s

A bivariate SGC random variable

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{cases} \begin{pmatrix} \alpha_1 \tilde{X} \\ \alpha_2 \tilde{Y} \end{pmatrix} & \text{with probability } p_{00} \\ \begin{pmatrix} (1 - \beta_1) + \beta_1 \tilde{X} \\ \alpha_2 \tilde{Y} \end{pmatrix} & \text{with probability } p_{10} \\ \begin{pmatrix} \alpha_1 \tilde{X} \\ (1 - \beta_2) + \beta_2 \tilde{Y} \end{pmatrix} & \text{with probability } p_{01} \\ \begin{pmatrix} (1 - \beta_1) + \beta_1 \tilde{X} \\ (1 - \beta_2) + \beta_2 \tilde{Y} \end{pmatrix} & \text{with probability } p_{11} \end{cases} :$$

Multivariate SGC random variables

Multivariate SGC random variables

We'd need a bigger screen !!

Thank you for your attention.

