A General Model of Random Variation

Haim Shore

Department of Industrial Engineering and Management,
Ben-Gurion University, Israel

ICOSDA 2013. Mt. Pleasant. Central Michigan University. October 11-12

Why

is the Box-Cox transformation

so effective?

The Inverse Box-Cox Transformation (IBCT)

$$y = [1 + \lambda(\eta + \sigma z)]^{(1/\lambda)}$$

- η is a linear predictor (linear combination of co-variates);
- $\{\lambda,\sigma\}$ are parameters

- This model represents three separate models:
  - Linear (λ=1)
  - Power ($\lambda \neq 1$, $\lambda \neq 0$)
  - Exponential ($\lambda \rightarrow 0$)

- Since λ is a parameter, the IBCT spans an infinite number of monotone convex models, of which the linear and exponential are just two points on this **continuous** spectrum.

The Ladder of Monotone Convex Functions

EXPONENTIAL-EXP.-EXP.-LINEAR

EXPONENTIAL-EXP.-POWER

EXPONENTIAL-EXP.-LINEAR

EXPONENTIAL-POWER

EXPONETNAIL-LINEAR

POWER

LINEAR

- First three steps of the Ladder are given by IBCT model.
- First four steps of the Ladder are given by the RMM model.
- An additional "cycle" of linear-power-exponential may be added for a "cost" of two parameters.
- Ladder is unbounded from above.

Some Examples for Ladder's Models

- **Kinetic Energy:** $E_k(V) = M (V^2/2)$ (power fun.)
- **Einstein's:** $E = MC^2 / [1-(v/C)^2]^{1/2}$ (power)
- **Radioactive Decay:** $R(t) = R_0 \exp(-kt)$ (exponential)
- **Antoine Eq.:** $\log(P) = A + B / (T+C)$ (exp.- power)
- **Arrhenius:** $R_e(T) = A \exp[-E_a/(k_B T)]$ (exp.- power)
- **Gompertz Grth-Model:** $Y= \beta_1 \exp[-\beta_2 \exp(-\beta_3 x)]$(exp.- exp.)

The RMM Model

$$Y = g(\eta, \varepsilon_1, \varepsilon_2; \theta) = \exp\{(\alpha/\lambda)[(\eta + \varepsilon_1)^{\lambda} - 1] + \varepsilon_2\}$$

Special Cases (from the "ladder"):

- Linear: $\lambda=0$, $\alpha=1$
- Power: $\lambda=0$, $\alpha\neq1$
- Exponential-linear: $\lambda=1$
- Exponential-power: $\lambda\neq0$, $\lambda\neq1$
- Exponential-exponential: ??

θ – Vector of parameters

"Exponential-exponential" (??) and further "climbing up" the "ladder"??

$$g(\eta,\varepsilon_1,\varepsilon_2;\theta) = \exp\{(\alpha / \lambda)[(\eta + \varepsilon_1)^{\lambda} - 1] + \varepsilon_2\}$$

Insert: $\exp\{(\beta/\kappa)[(\eta+\varepsilon_1)^{\kappa} - 1]\}$

For $\kappa=0$, $\beta=1$:

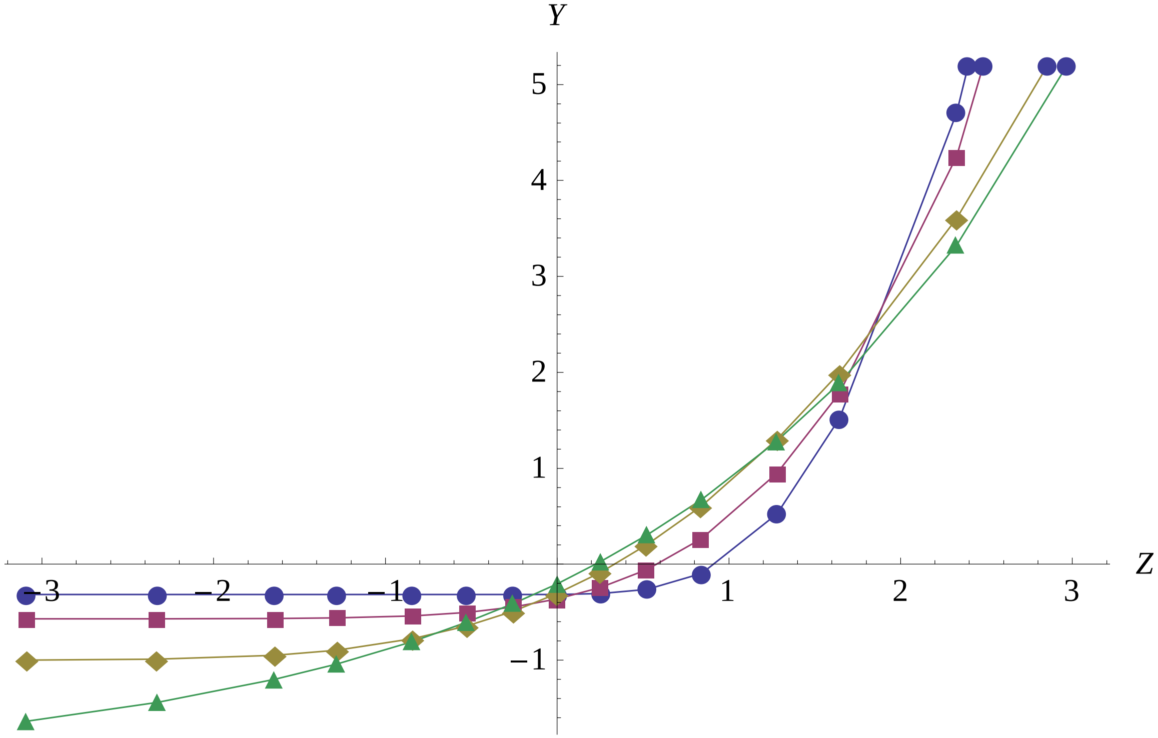
$$\exp\{(\beta/\kappa)[(\eta+\varepsilon_1)^{\kappa} - 1]\}= \eta+\varepsilon_1$$

Important result: Adding two parameters allows a repeat of the cycle "Linear-power-exponential", while "climbing" the ladder!!

Conclusion:

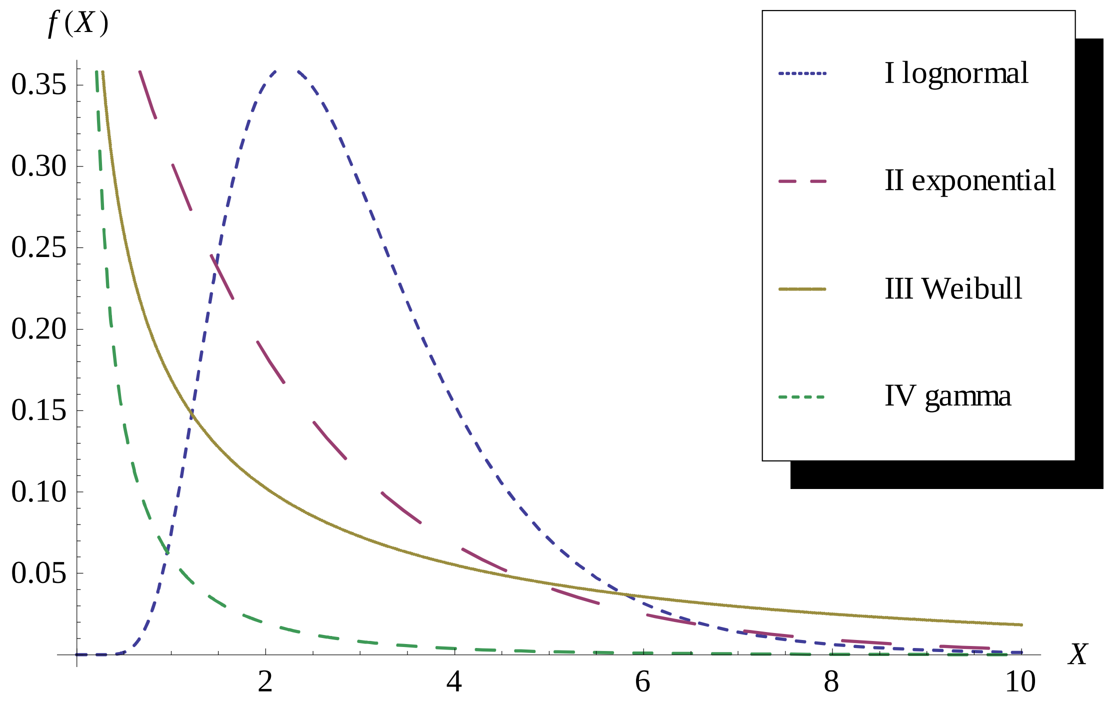
All monotone-convex relationships, traditionally represented by **“single-point” models**, belonging to the **“Ladder”,** may now be represented by a single RMM model, which delivers **“monotone convexity” as a continuous spectrum**!!

Four Monotone Convex Funtions

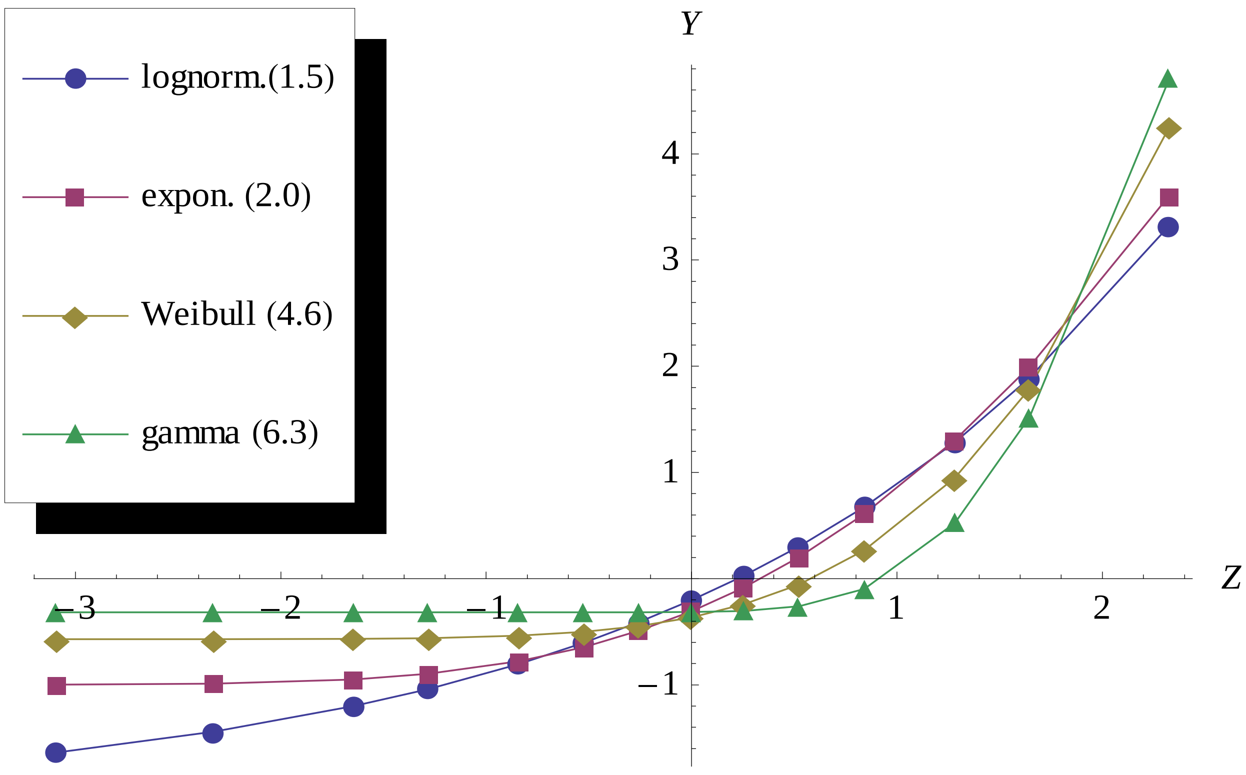


What models may these curves represent?

Four asymmetrical distributions



Respective Quantile Functions in terms of Standard Normal Quantiles (with Sk values)



Basic idea of a “General model of random variation”:

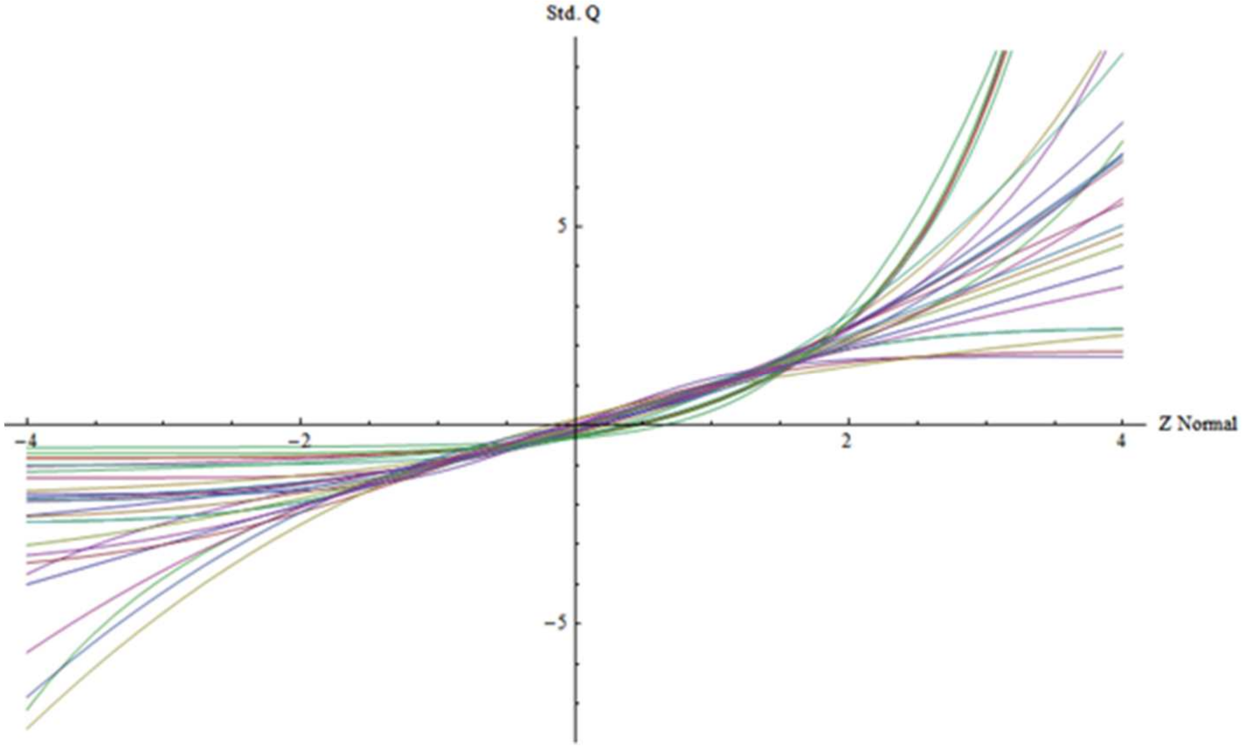
Any Z-based quantile function (of a unimodal distribution) may be modeled well via RMM due to its unique property of

Continuous Monotone Convexity (CMC)

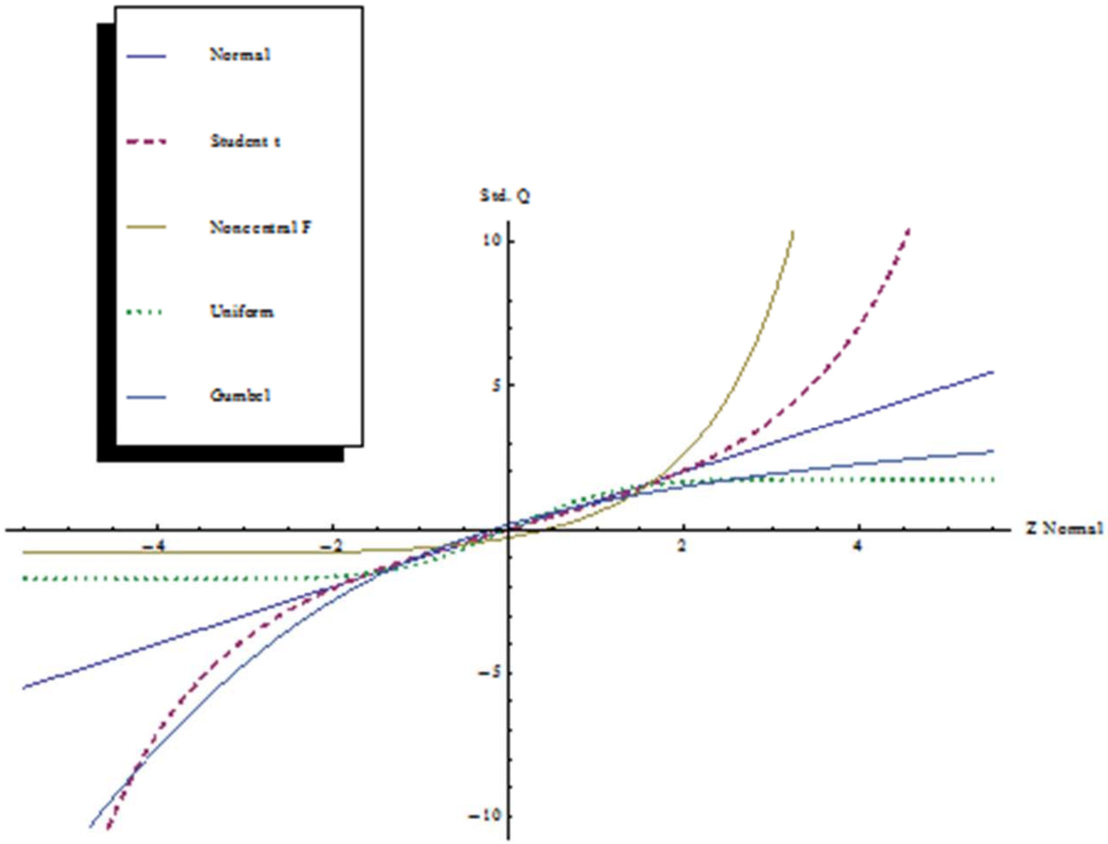
Some references (where this idea was examined or implemented):

- Shore, H. (2005). *Response Modeling Methodology- Empirical Modeling for Engineering and Science*. World Scientific Publishing Co. Ltd., Singapore.
- Shore, H. (2005). Accurate RMM-based Approximations for the CDF of the Normal Distribution. *Communications in Statistics (Theory&Methods)*, 34(2). 507-513.
- Shore, H. (2007). Comparison of Generalized Lambda Distribution (GLD) and Response Modeling Methodology (RMM) as general platforms for distribution fitting. *Communications in Statistics (Theory&Methods)*. 36(15). 2805-2819.
- Shore, H. (2008). Distribution fitting with response modeling methodology (RMM) - some recent results, *American Journal of Mathematical and Management Sciences, 28, 1&2, 3-18.*
- Shore, H., A'wad, F. (2010). Statistical comparison of the goodness-of-fit delivered by five families of distributions used in distribution fitting. *Communications in Statistics (Theory&Methods)*. 39(10), 1707-1728.

Z-based QF of 27 Distributions



Z-based Quantile Functions of 4 distributions



Conclusions

Skewness and kurtosis jointly determine the *general* shape of $Q_Y(z;\theta)$ as follows:

- All symmetric r.v.s have S-shaped $Q_Y(z;\theta)$, for negative kurtosis, inverted S-shaped, for positive kurtosis, and linear, for Y normal;
- All non-symmetric r.v.s have $Q_Y(z;\theta)$ that, as skewness departs from zero, gradually transforms from S-shaped to monotone convex, for positively skewed distributions, or to monotone concave, for negatively skewed distributions;

Final General Model

A weighted average of the Z-based standardized QF of an inherently-symmetric distribution (ISD) and of a non-ISD distribution:

$$\tilde{Q}_x(z;\tilde{\theta}) = (w)\frac{g(z;\beta) - g(-z;\beta)}{(2\sigma_Y)} + (1-w)\frac{g(z;\beta) - \mu_Y}{\sigma_Y}\ .$$

g(z;β) – An RMM model for Z-based quantile function (next slide)

w – weight, determined to preserve the median of the distribution

For ISD (w=1) – Basic structure of Taylor series is preserved (odd-order powers)

RMM Z-based Quantile Function (QF)

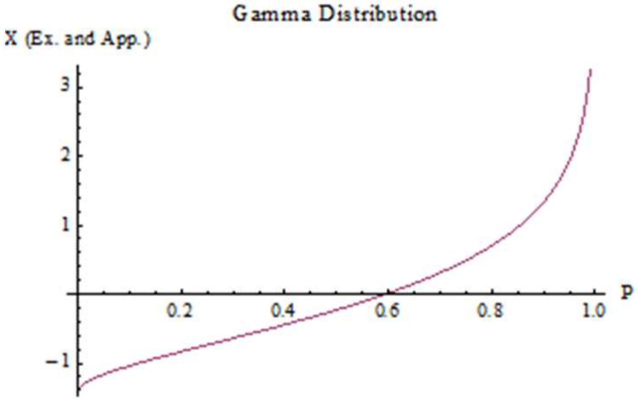
- Original RMM QF (M_Y is median):

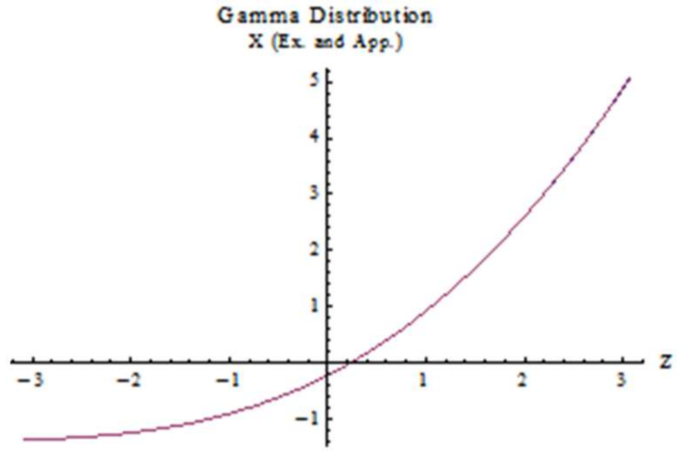
$$\log(y-L)=g(z;\beta)=\log(M_Y-L)+\left(\frac{a}{b}\right)[(1+cz)^b-1]+dz+\varepsilon,$$

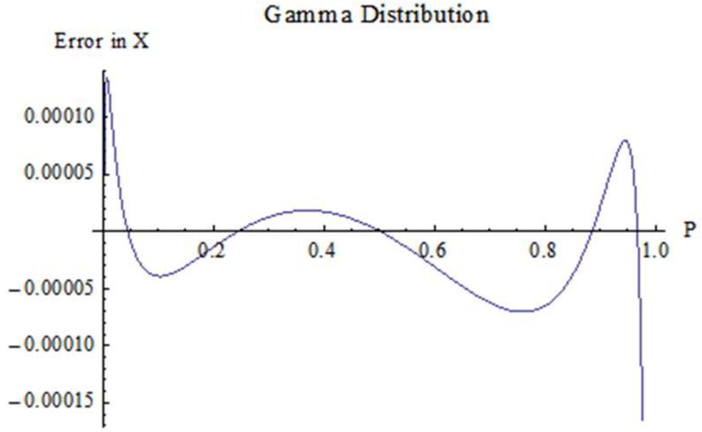
- An approximate expression (cz << 1):

$$\log(y-L)=\log(M_Y-L)+\left(\frac{a}{b}\right)(e^{bz}-1)+cz+\varepsilon\ .$$

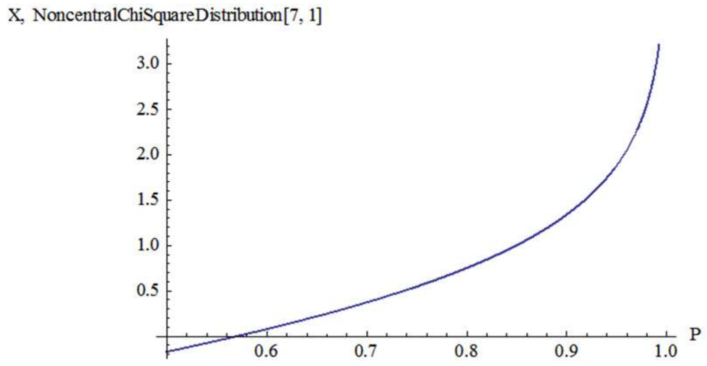
Example: Gamma

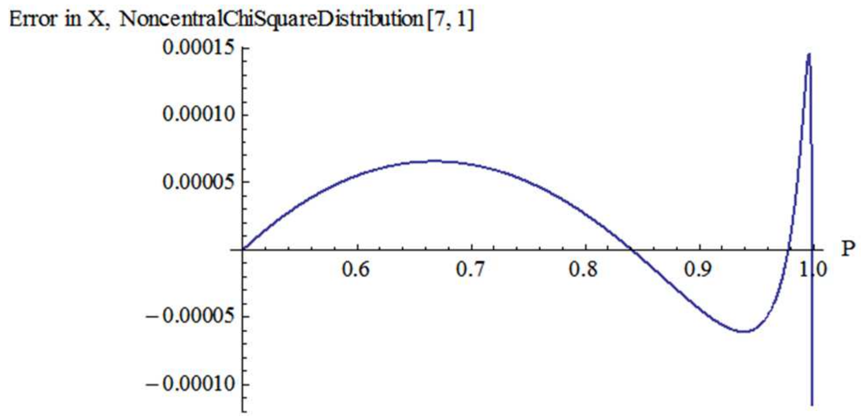




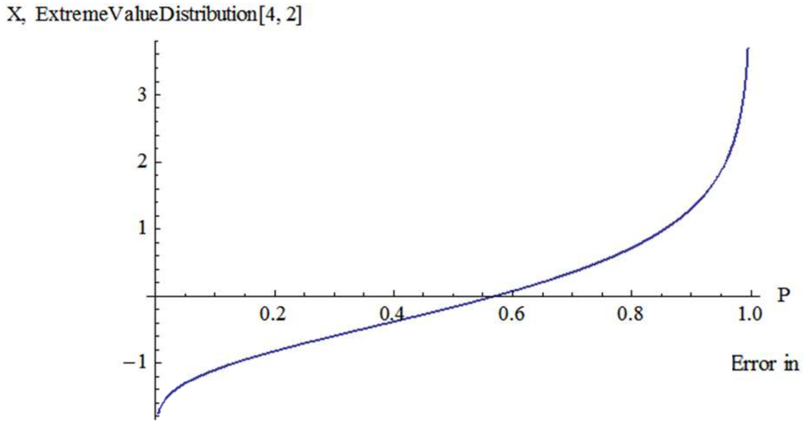


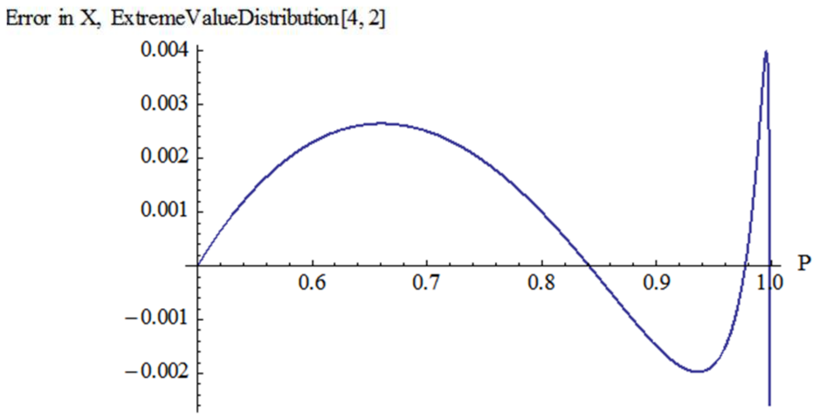
Example: Non-central Chi-square (7,1)





Example: Extreme Value (4,2)





Further Details:

Shore, H. (2013). A General Model of Random Variation. *Comm. in Statistics (Theory&Methods)*. Forthcoming.

Thank you

(shor@bgu.ac.il)