Тема доклада

Почему распределения параметров сложных систем:

1. часто имеют высокую неоднородность (Принцип 80/20)
2. часто имеют степенные хвосты с низкими показателями степени (обычно от 1 до 3)

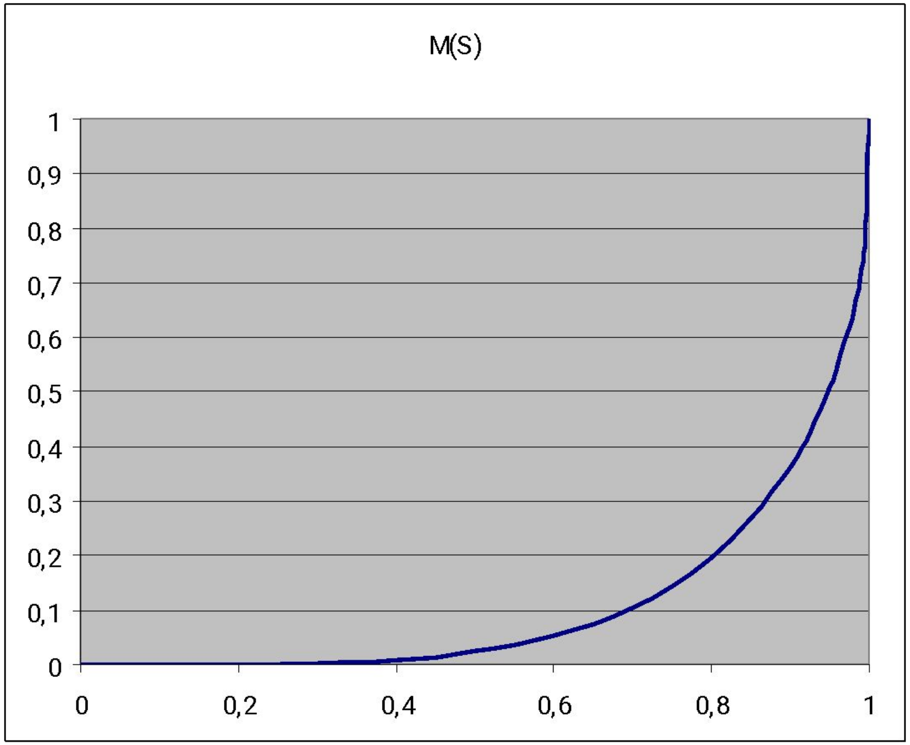
Причины:

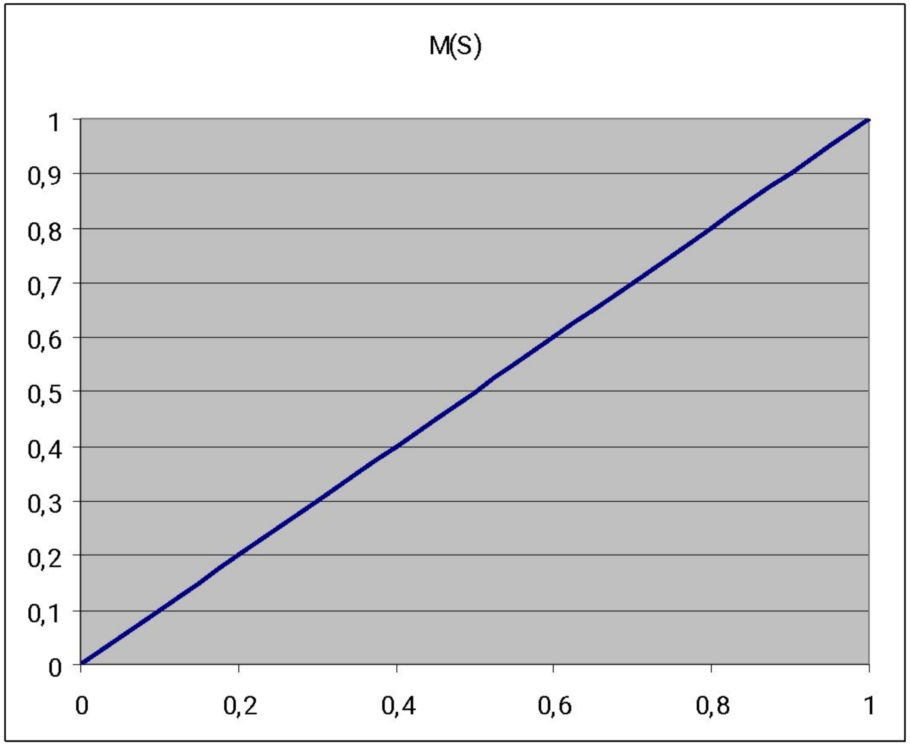
- эффект случайного смещения результата (1)
- эффект естественного отбора (1+2)
- "чистые распределения" (1+2)

Принцип 80/20

- 20% ассортимента продукции - 80% от общего объема продаж
- 20% покупателей и клиентов - 80% от общего объема продаж
- 20% ассортимента продукции или 20% покупателей - 80% прибыли
- 20% преступников - 80% преступлений
- 20% водителей - 80% дорожно-транспортных происшествий
- 20% вступивших в брак - 80% разводов
- 20% детей - 80% возможностей, предоставляемых системой образования в данной стране
- 20% площади ковров - 80% воздействий, ведущих к их износу
- 80% всего времени - 20% имеющейся у вас одежды
- 80% всех ложных тревог при срабатывании противоугонной сигнализации - 20% возможных причин

Кривая Лоренца





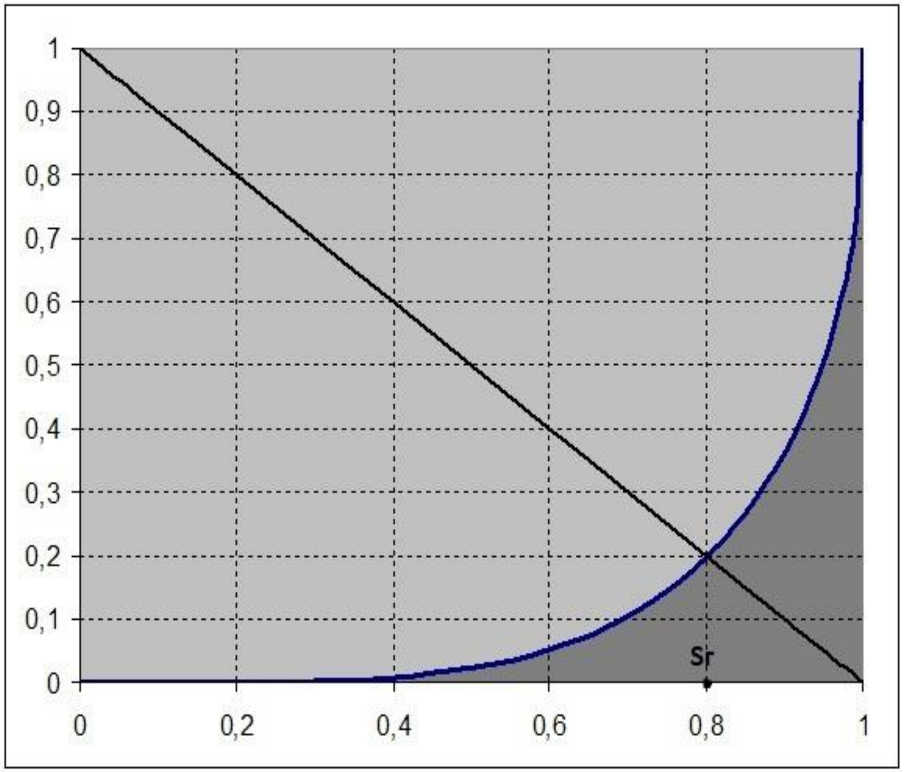
$$S(r) = \int_0^r p(t) \cdot dt$$

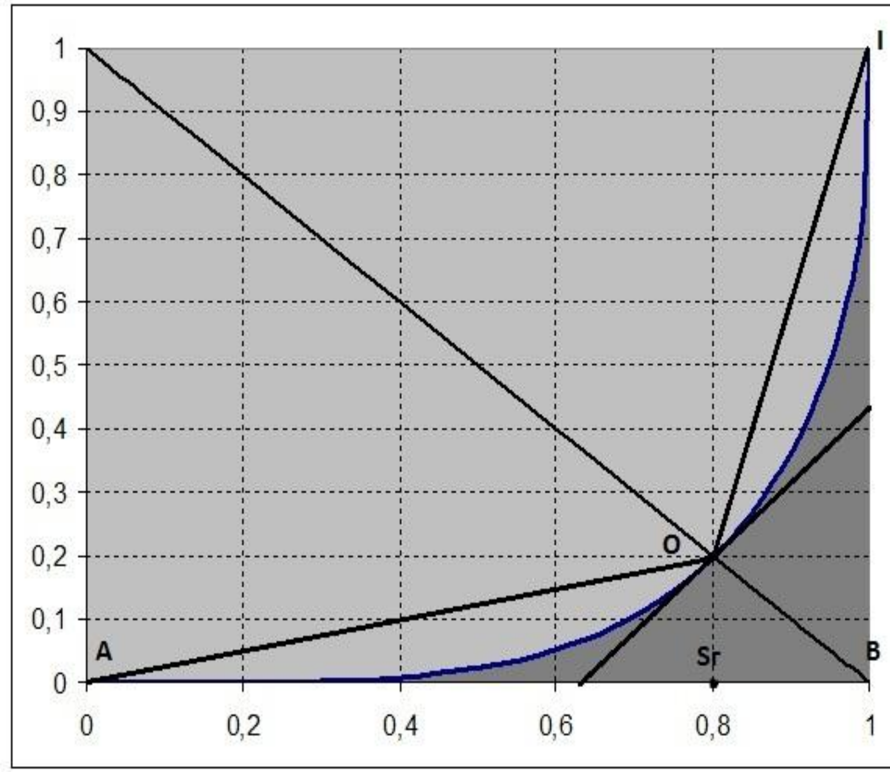
$$M(r) = \frac{\int_0^r t \cdot p(t) \cdot dt}{\int_0^\infty t \cdot p(t) \cdot dt} = \frac{\int_0^r t \cdot p(t) \cdot dt}{< r >}$$

Меры однородности G и Г

$$\Gamma = 2 \cdot (1 - S_\Gamma)$$

$$G = 2 \cdot \Sigma$$





$0 \le \Gamma \le 1$ $\quad 0 \le G \le 1$

$$\Gamma^2 \le G \le \Gamma$$

Три меры неоднородности

$$\overline{\Gamma} = 1 - \Gamma$$

$$\overline{G} = 1 - G$$ - коэффициент Джини

$$\Phi = D\left[\frac{r}{< r >}\right] = \frac{D[r]}{< r >^2} = \frac{< r^2 >}{< r >^2} - 1$$

- Меры Г и G "работают" только для неотрицательных "r" - мера Ф работает всегда
- Свойства меры Ф проще анализировать

Свойства меры однородности G

$$G = \frac{\int\limits_0^\infty \overline{S}^2(r)\cdot dr}{\int\limits_0^\infty \overline{S}(r)\cdot dr} = \frac{\int\limits_0^\infty \overline{S}^2(r)\cdot dr}{<r>} = 2\cdot\frac{<r\cdot\overline{S}>}{<r>}$$

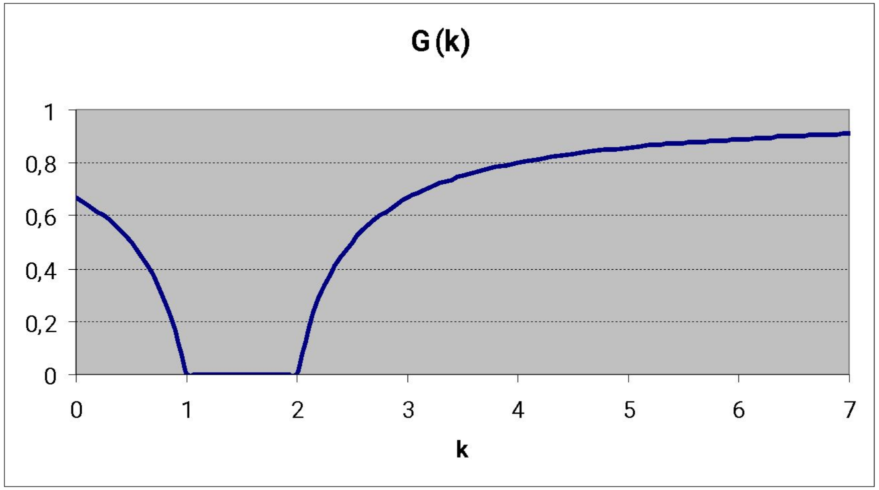
$$\overline{S}(r) = \int\limits_r^\infty p(t)\cdot dt = 1 - S(r) = 1 - \int\limits_0^r p(t)\cdot dt$$

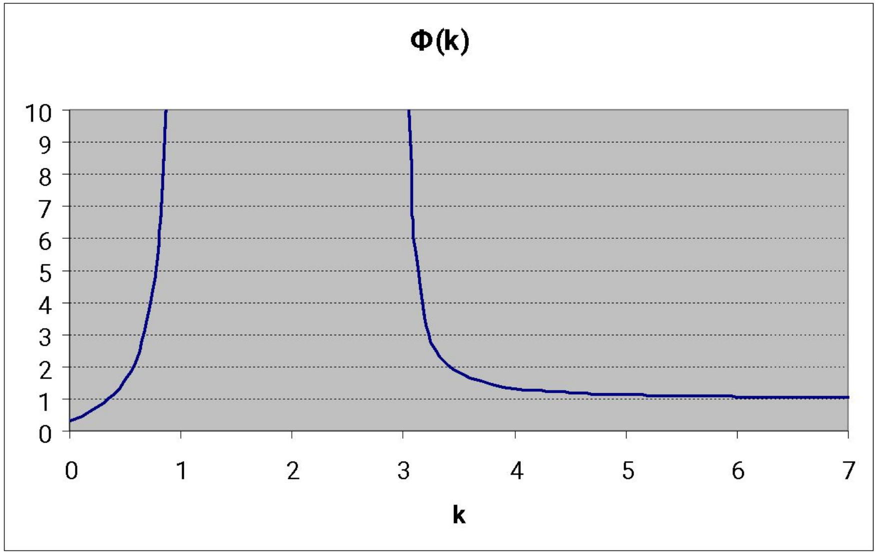
$$<r> = \int\limits_0^\infty p(r)\cdot r\cdot dr = \int\limits_0^\infty (1 - S(r))\cdot dr = \int\limits_0^\infty \overline{S}(r)\cdot dr$$

Однородность G разных распределений вероятности

Распределение ($r \geq 0$)	**G**	**Г**	**Ф**
дельта-функция	1	1	0
экспоненциальное ("водораздел")	1/2	0,64	1
равномерное	$\geq 2/3$	$\geq 0{,}62$	$\leq 1/3$
"две дельта-функции", случай "$p(r) \to \delta(r)$" с сохранением двух дельта-функций	0	0	$+\infty$
логнормальное	1 при $\sigma=0$ 0,03 при $\sigma=3$	1 при $\sigma=0$ 0,13 при $\sigma=3$	0 при $\sigma=0$ 8102 при $\sigma=3$
чисто степенное: $p(r)=C/r^k$, $0<r<\infty$	0 при $1 \leq k \leq 2$ 1 при $k \to +\infty$	0 при $1 \leq k \leq 2$ 1 при $k \to +\infty$	0 при $1 \leq k \leq 3$ 1 при $k \to +\infty$

Неоднородность чисто степенного распределения





Модель эффекта случайных смещений результата

Можно условно считать, что:

1. добавление нового параметра "y" увеличивает дисперсию и
2. среднее значение результата не меняется

1-й случай: r=x+y, где r≥0

1) если у "y" корреляция с "x" не отрицательная, то рост всегда будет:

$$G_{x+y} = G_x - \frac{\sigma^2}{12 \cdot < r >_x} \cdot \int_0^\infty p_x{}^2(r) \cdot dr$$

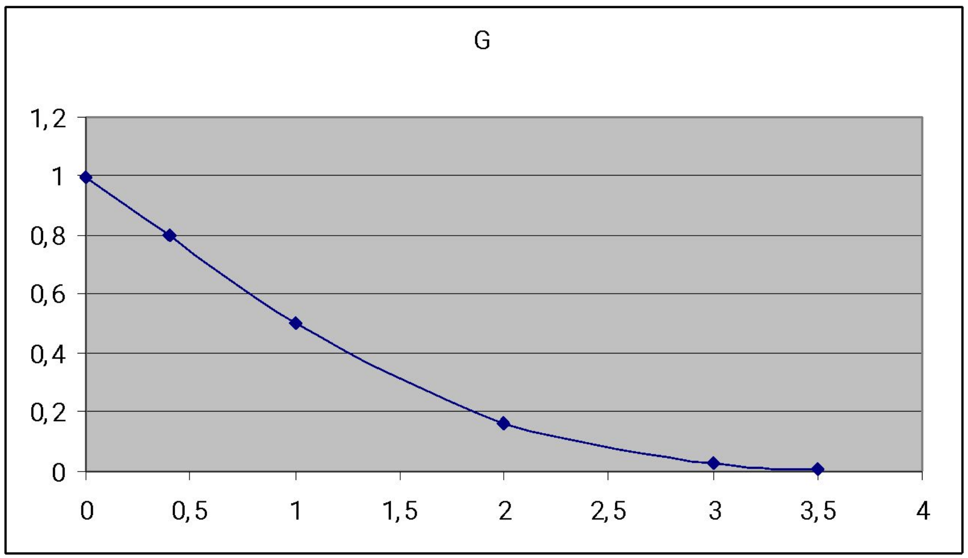
$$\Phi_{x+y} = \frac{D_y}{< r_x >^2} + \Phi_x + 2 \cdot \Omega_{x,y} \cdot \sqrt{\frac{D_y}{< r_x >^2} \cdot \Phi_x}$$

2) возможен случай p(r) → δ(r), G →1,

иначе G≥2/3

2-й случай: r=x+y, где -∞≤r≤+∞ – неоднородность Ф растет,
но
пропорционально числу параметров: σ~N

3-й случай: r=x*y –
σ~N, но дисперсия растет гораздо быстрее.
Для логнормального распределения:



$$\Phi = \frac{D}{<r>^2} = \exp[\sigma^2] - 1$$

Эффект естественного отбора

"Побеждают сильнейшие": $w(r) = \int u(r, \vec{y}) \cdot d^k y$

1. G = результат усреднения однородности G_i "гармоник" с весовым коэффициентом $k_i \sim (\langle r_i \rangle)^{1/2}$
2. Если среди суммируемых "гармоник", т.е. распределений с фиксированным значением параметров , существуют распределения со степенными "хвостами", то итоговое распределение тоже будет иметь степенной хвост, потому что он убывает медленнее.

ЭЕО: Большая неоднородность

Скалярное произведение порождает сепарабельное гильбертово пространство:

$$u_1 \otimes u_2 = \int_0^{\infty} \overline{S}_1 \cdot \overline{S}_2 \cdot dr$$

$$\overline{S}(r) = \int_r^{\infty} u(t) \cdot dt$$

$$L = \| u \| = \sqrt{u \otimes u} = | S(\infty) | \cdot \sqrt{G \cdot < r >} \geq 0$$

Поведение меры однородности G при суммировании гармоник

Сумма гармоник: $w(r) = \int u(r, \overrightarrow{y}) \cdot d^k y, \;\; w(r) = \sum_i u(r, \overrightarrow{y_i})$

$L = \| u \| = \sqrt{u \otimes u} = \sqrt{G \cdot < r >} \geq 0$ - "длина вектора"

$$L_\Sigma = \sqrt{G_\Sigma} = \frac{| \sum_i \vec{L}_i |}{W_\Sigma \cdot \sqrt{< r >_\Sigma}} = \frac{\left| \sum_i \vec{e}_i \cdot \sqrt{W_i} \cdot \sqrt{W_i \cdot < r >_i} \cdot \sqrt{G_i} \right|}{\sqrt{\sum_i W_i} \cdot \sqrt{\sum_i W_i \cdot < r >_i}}$$

То есть побеждают наибольшие $\sqrt{< r >_i}$

ЭЕО: Появление "степенных хвостов" распределений

Эффективный показатель степени:

$$m(r) = m[f(r)] = \frac{-d(\ln f(r))}{d(\ln r)} = \frac{-r \cdot \frac{d}{dr} f(r)}{f(r)}$$

$$m_r[u(r, \vec{y})] = \frac{-r \cdot \frac{\partial}{\partial r} u(r, \vec{y})}{u(r, \vec{y})}$$

"Степенное усреднение":

$$\overline{m} = m[w(r)] = \frac{-r \cdot \frac{\partial}{\partial r} \int u(r, \vec{y}) \cdot d^k y}{\int u(r, \vec{y}) \cdot d^k y} = \frac{\int m_r[u(r, \vec{y})] \cdot u(r, \vec{y}) \cdot d^k y}{\int u(r, \vec{y}) \cdot d^k y}$$

В дискретном случае:

$$\overline{m} = m[w(r)] = \frac{\sum_i m[u_i(r)] \cdot u_i(r)}{\sum_i u_i(r)} = \frac{\sum_i m_i \cdot u_i}{\sum_i u_i}$$

Дрейф показателя суммы гармоник вниз вплоть до наименьшего из показателей

$$\frac{d}{dr}\overline{m} = \overline{\frac{d}{dr}m} - \frac{1}{r}\cdot\overline{\left(m-\overline{m}\right)^2} = \overline{\frac{d}{dr}m} - \frac{1}{r}\cdot D$$

$$\rho = \ln\frac{r}{r_0}: \qquad \frac{d}{d\rho}\overline{m} = \overline{\frac{d}{d\rho}m} - \overline{\left(m-\overline{m}\right)^2} = \overline{\frac{d}{d\rho}m} - D$$

"Степенная дисперсия":

$$D = \overline{\left(m-\overline{m}\right)^2} = \overline{m^2} - \left(\overline{m}\right)^2 \geq 0$$

Отклонение эфф-го показателя степени от минимального

$$m_{\min}(r) \to_{r\to\infty} \min_{\vec{y}} m(r,\vec{y}) \qquad \min_{\vec{y}} m(r,\vec{y}) \le m_{\min}(r) = ?$$

$$u(r_2,\vec{y}) = u(r_1,\vec{y}) \cdot \exp\left(-\int_{r_1}^{r_2} m_r[u(r,\vec{y})] \cdot d(\ln r) \right) = u(r_1,\vec{y}) \cdot e^{-S_{12}(\vec{y})}$$

$$\ln u(r_2,\vec{y}) = \ln u(r_1,\vec{y}) - S_{12}(\vec{y})$$

$$S_{12}(\vec{y}) = \int_{r_1}^{r_2} m_r[u(r,\vec{y})] \cdot d(\ln r) = \int_{r_1}^{r_2} m_r[u(r,\vec{y})] \cdot d\rho = \int_{\rho_1}^{\rho_2} m(\rho,\vec{y}) \cdot d\rho$$

Эффективный показатель степени для конечных результатов

$$m_{\min}(r_2) + \ln(10 \cdot A) \cdot \left(\frac{\Delta S_{12}}{\Delta m_2} \right)^{-1}$$

"Чистые" законы распределения

- не объясняются эффектом естественного отбора
- объясняются эффектом случайных смещений параметра только для некоторых распределений (логнормальное, логнормальное)

Условие сокрытия параметров для марковских процессов

$$\hat{A} = \hat{N} \cdot \hat{G}$$

$$\hat{N} 0(r, y) = 0(r, y)$$

$$G(u_s, r', \vec{y}', r, \vec{y}) = (u_s(r, \vec{y}) - w_s(r) \cdot \delta(\vec{y})) \cdot g_1(r', \vec{y}', r) + G_2(w_s(r), r', \vec{y}', r, \vec{y})$$

$$\int G_2(w_s(r), r', \vec{y}', r, \vec{y}) \cdot d^n y = \overline{G}(w_s, r', \vec{y}', r) = G_0(w_s, r', r)$$

$g_1(r', \vec{y}', r)$ - произвольная функция

Чисто степенные распределения для марковских процессов

Разумно предположить, что если в результате эволюции получается чисто степенное распределение, то оператор эволюции **почти всегда** можно представить в виде "BA", где "A" – масштабно-инвариантный оператор (глобально или только локально), а оператор "B" нулевую функцию переводит в нулевую. Исключения представляют собой вырожденные случаи и, видимо, не должны часто встречаться.

Оператор "A" является масштабно-инвариантным:

- глобальная масштабная инвариантность может существовать только если объекты не взаимодействуют или эволюционирует один объект
- локальная масштабная инвариантность может возникнуть, только если на объекты влияют параметры всей совокупности других объектов как целое: "температура", число, "давление"

Оператор "B" в общем случае не является масштабно-инвариантным и, соответственно, не будет масштабно-инвариантным и итоговый оператор "BA".

Цикл статей "Доказательный менеджмент"

1. Общий подход к расчету проектов
2. Критика книги "От хорошего к великому" Филом Розенцвейгом – насколько она обоснована?
3. Расчет эффективностей и их использование
4. Тестирование в бизнесе
5. Инновация как враг прибыльного бизнеса
6. Как объяснить Принцип 80/20 с помощью эффектов "естественного отбора" и случайных смещений результата
7. Оценка вероятностей человеком – дважды неожиданные эффекты
8. Система Тойоты и реинжиниринг – чем могут помочь численные модели?