TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

A tale of two centres

Leigh Roberts
School of Economics and Finance,
Victoria University of Wellington, New Zealand
`leigh.roberts@vuw.ac.nz`

WWPMS, 19 November 2015

Thanks to Estate Khmaladze for helpful comments. Any errors remain the responsibility of the author.

Overview

A tale of two centres OR A tale of two Brownian Bridges

In fact, lots of twos

Two centrings of basic gof statistics: duality between

- conventional centring
- less conventional, but more intuitive, centring
  - BB1
  - BB2

Two applications: Pareto and Weibull

Two main mathematical tools: projections, rotations

Ancillary tools: two Taylor series

- gof statistic
- score function

$$v_n(x,\theta) = \sqrt{n}[F_n(x) - F_\theta(x)] \tag{1}$$

First Taylor series, $\dot{\ell} = \psi$ the score function, ℓ = log likelihood

$$0 = \dot{\ell}(\widehat{\theta}) = \dot{\ell}(\theta) + \frac{1}{2}\Gamma(\widehat{\theta} - \theta) + \dots$$

$$h_\theta(x) = \frac{\psi_\theta(x)}{\sqrt{\Gamma(\theta)}}$$

$$\sqrt{n}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{\Gamma(\theta)}} \int_{-\infty}^{\infty} h_\theta(x) v_n(dx,\theta) + o_P(1)$$

Second Taylor series expands (1) to give

$$v(x,\widehat{\theta}) = v(x,\theta) - \int_{-\infty}^{x} h_\theta(y)\, dF_\theta(y) \int_{-\infty}^{\infty} h_\theta(x)\, v(dx,\theta)$$

The two centrings

$$v_n(x,\theta) = \sqrt{n}[F_n(x) - F_\theta(x)]$$
$$w_n(x,\theta) = \sqrt{n}\left[F_n(x) - \int_{-\infty}^{x} \frac{1-F_n(y)}{1-F_\theta(y)}\, dF_\theta(y)\right]$$

second centring is unconventional, but well motivated.

Intuition:
applying theoretical hazard rate to actual exposed to risk
matching expected and actual deaths over short time intervals.

For simple hypotheses, i.e. given θ,
$v_n(x,\theta)$ is a BB,
$w_n(x,\theta)$ is a BM.

Extra notation

$$h_\theta(x) = \frac{\psi_\theta(x)}{\sqrt{\Gamma(\theta)}} \qquad H_\theta(x) = \int_{-\infty}^{x} h_\theta(y) dF_\theta(y)$$

$$g_\theta(x) = h_\theta(x) + \frac{H_\theta(x)}{1 - F_\theta(x)}$$

Duality

$$v(x, \widehat{\theta}) = v(x, \theta) - \int_{-\infty}^{x} h_\theta(y)\, dF_\theta(y) \int_{-\infty}^{\infty} h_\theta(x)\, v(dx, \theta)$$

$$w_n(x, \widehat{\theta}) = w_n(x, \theta) - \int_{-\infty}^{x} g_\theta(y)\, dF_\theta(y) \int_{-\infty}^{\infty} g_\theta(x)\, w_n(dx, \theta) + o_P(1)$$

The limiting process is the g projected F_θ BM (Khmaladze, 2016):

$$w(x, \widehat{\theta}) = w(x, \theta) - \int_{-\infty}^{x} g_\theta(y)\, dF_\theta(y) \int_{-\infty}^{\infty} g_\theta(x)\, w(dx, \theta) \qquad (2)$$

The first BB is derived from (2).

$$\int_{-\infty}^{x} g(y,\widehat{\theta})w(dy,\widehat{\theta}) = \int_{-\infty}^{x} g(y,\theta)w(dy,\theta)$$

$$-\int_{-\infty}^{x} g_{\theta}^{2}(y)\,dF_{\theta}(y)\int_{-\infty}^{\infty} g_{\theta}(x)\,w(dx,\theta)$$

which is a BB in time $\int_{-\infty}^{x} g_{\theta}^{2}(y)\,dF_{\theta}(y)$, with $\int_{-\infty}^{\infty} g_{\theta}^{2}(y)\,dF_{\theta}(y) = 1$.

The second BB is a rotation of the q projected F-BM in (2), where $q = g$, and we are mapping onto a q_0 projected F-BM, with $F = F_\theta$ and q_0 the uniform distribution. (Khmaladze, 2016).

Tangent: rotation to distribution free statistics of the form

$$U_{a,b} = I - (a-b)(a-b)^T$$

but with extra terms for normalising a and b to be unit length. $U_{a,b}$ swaps a and b around, leaves unchanged vectors orthogonal to them. Khmaladze (2013) more easily understood than Khmaladze (2016).

Same pattern as $U_{a,b}$ for rotations of projection (2), to obtain BB2.

The second BB is a rotation of the q projected F-BM in (2), where $q = g$, and we are mapping onto a q_0 projected F-BM, with $F = F_\theta$ and q_0 the uniform distribution (Khmaladze, 2016).

$$w_n(x, \widehat{\theta}) - \frac{\int_{-\infty}^{x} g_{\widehat{\theta}}(y) dF_{\widehat{\theta}}(y) - F_{\widehat{\theta}}(x)}{1 - \int_{-\infty}^{\infty} g_{\widehat{\theta}}(y) dF_{\widehat{\theta}}(y)} w_n(\infty, \widehat{\theta})$$

has as its limiting process a Brownian bridge in the simpler time $F_\theta(x)$.

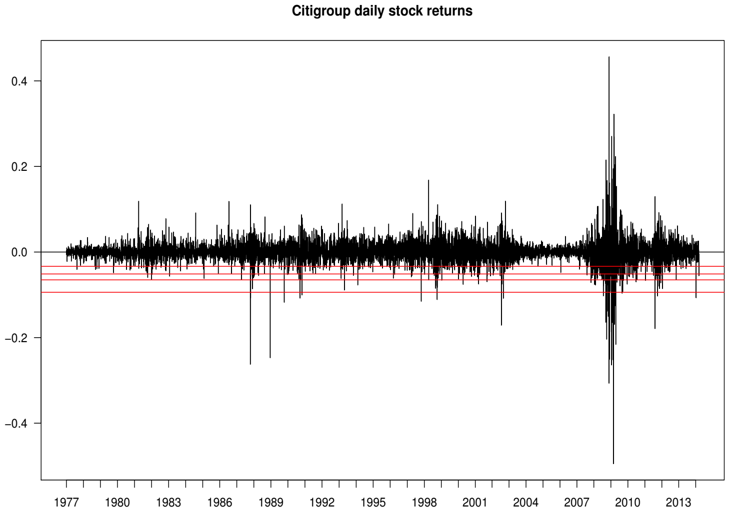
Applications

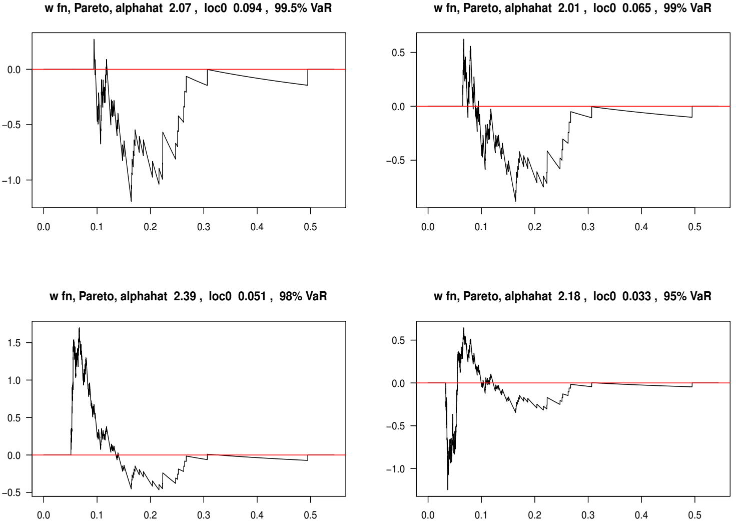
Weibull
Data from Smith and Naylor (1987)
Two BBs are in Figure 3

Pareto
data from yahoo: returns on Citigroup over the last 35 years,
(business) daily data concatenated over weekends and holidays
We model exceedances of greater losses than Value at Risk

It turns out that $g = 1$, so BB1 is a BB in time $F_\theta(x)$, and BB2 is not well defined.

data and results in Figures 1 and 2


Citigroup daily stock returns
0.4
0.2
0.0
−0.2
−0.4
1977
1980
1983
1986
1989
1992
1995
1998
2001
2004
2007
2010
2013


w fn, Pareto, alphahat 2.07 , loc0 0.094 , 99.5% VaR
w fn, Pareto, alphahat 2.01 , loc0 0.065 , 99% VaR
w fn, Pareto, alphahat 2.39 , loc0 0.051 , 98% VaR
w fn, Pareto, alphahat 2.18 , loc0 0.033 , 95% VaR
0.0
0.1
0.2
0.3
0.4
0.5
−0.5
−1.0
1.0
1.5

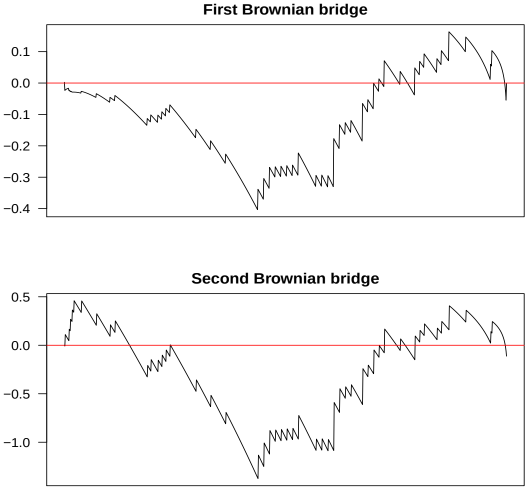


Figure : The two Brownian bridges for the Weibull distribution

Conclusion

- approach a useful alternative to Khm transform of $v(x, \widehat{\theta})$.
- Pareto and Weibull both commonly used
  - Weibull is an extreme value distribution
- BB2 convenient because could transform to other distributions.
- 2 parameters?: messy; rotations more awkward to effect
- drawback: g not necessarily easy to calculate
- not easily adapted to mixtures.

Thank You.

Khmaladze, E. V. (2013). Note on distribution free testing for discrete distributions. *Annals of Statistics*, 41(6):2979–2993.

Khmaladze, E. V. (2016). Unitary transformations, empirical processes and distribution free testing. *Bernoulli*, 22(1).

Roberts, L. A. (2015). Distribution free testing of goodness of fit in a one dimensional parameter space. *Statistics & Probability Letters*, 99:215–222.

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36(3):358–369.