

일반화된 로렌츠 곡선을 기반으로 한 Gumbel 분포의 적합도 검정

이경준¹

¹대구대학교 전산통계학과

접수 2017년 6월 29일, 수정 2017년 7월 11일, 게재확정 2017년 7월 13일

요약

통계학에서 사용되어지고 있는 Gumbel 분포는 환경과학, 시스템 신뢰성, 수문학과 같은 분야에서 많이 응용되고 있다. 따라서 환경과학, 시스템 신뢰성, 수문학과 관련된 자료를 분석함에 있어서 분석에 사용되어지는 자료가 Gumbel 분포를 따르는지 확인하는 것은 매우 중요하다. 이를 확인하기 위해 본 논문에서는 새로운 두 가지의 Gumbel 분포의 적합도 검정통계량을 일반화된 로렌츠 곡선을 기반으로 하여 제안하였고, Anderson - Darling 검정, Cramer - vonMises 검정, 수정된 Anderson - Darling 검정과 비교하였다. 그 결과 새롭게 제안한 검정통계량은 기존의 검정방법에 비하여 우수한 것을 확인할 수 있었다. 또한 새롭게 제안한 변형된 표본 일반화된 로렌츠 곡선을 이용하여 두 가지의 새로운 적합도 검정 그래프 방법을 제안하였고, 새롭게 제안된 그래프를 통하여 손쉽게 데이터가 Gumbel 분포를 따르는지를 파악할 수 있었다. 또한 호주 시드니의 연간 일 최대 강수량 자료를 사용하여 새롭게 제안한 검정통계량과 그래프 방법을 이용하여 적용해 보았다.

주요용어: 검정력, 로렌츠 곡선, 일반화된 로렌츠 곡선, 적합도 검정, Gumbel 분포.

1. 서론

통계학에서 사용되어지고 있는 Gumbel 분포는 환경과학 (environmental sciences), 시스템 신뢰성 (system reliability), 수문학 (hydrology)과 같은 분야에서 많이 응용되고 있다. Gumbel 분포의 확률밀도함수 (probability density function, pdf)와 누적분포함수 (cumulative distribution function, cdf)는 다음과 같다.

$$f(x, \alpha, \beta) = \frac{1}{\beta} \exp \left[-\frac{x - \alpha}{\beta} - \exp \left(-\frac{x - \alpha}{\beta} \right) \right],$$
$$F(x, \alpha, \beta) = \exp \left[-\exp \left(-\frac{x - \alpha}{\beta} \right) \right], \quad -\infty < x < \infty, \quad -\infty < \alpha < \infty, \quad \beta > 0.$$

이러한 Gumbel 분포는 수문학에서 지난 몇 년 동안의 강 (river)의 최소 수위를 나타내는 분포로서 사용되어진다. 또한 지진, 홍수, 기타 자연 재해의 발생을 예측하는 데에 Gumbel 분포가 사용된다.

따라서 환경과학, 시스템 신뢰성, 수문학과 관련된 자료를 분석함에 있어서 분석에 사용되어지는 자료가 Gumbel 분포를 따르는지 확인하는 것은 매우 중요하다. 이를 확인하기 위해 우리는 적합도 검정을 실시하는데 이는 히스토그램 (histogram), Q-Q 플롯 (Q-Q plot), P-P 플롯 (P-P plot)과 같은 그래프

¹ (38453) 경상북도 경산시 진량읍 대구대로 201, 대구대학교 전산통계학과, 조교수.
E-mail: leekj@daegu.ac.kr

를 이용하는 방법과 Anderson - Darling 검정 (AD)과 Cramer - vonMises 검정 (CVM)과 같은 통계량을 이용하는 방법이 있다. 이 외에도 Kinnison (1989)는 상관계수 형태의 새로운 적합도 검정통계량을 제안하였고, Arizono와 Ohata (1989)는 Kullback-Leibler 정보 (Kullback-Leibler information)의 추정량을 기반으로 한 적합도 검정통계량을 새롭게 제안하였다. 그리고 Song (2002)은 Kullback-Leibler 정보를 기반으로 하여 근사적으로 분포 무관한 (asymptotically distribution-free) 적합도 검정 방법을 제안하였다. Zempleni (2004)는 Anderson - Darling test을 개량한 수정된 Anderson - Darling 검정 (ZAD)을 제안하였다. Cho와 Lee (2014)는 일반화된 로렌츠 곡선을 이용하여 정규분포의 적합도 검정통계량을 새롭게 제안하였고, Kang 등 (2014a)은 로렌츠 곡선을 이용하여 제 2종 다중 중도절단 하에서 로지스틱분포의 적합도 검정통계량을 새롭게 제안하였다. 또한 Kang 등 (2014b)은 로렌츠 곡선을 이용하여 제 2종 다중 중도절단 하에서 역 와이블분포의 적합도 검정통계량을 새롭게 제안하였다.

본 논문은 Shorrocks (1983)이 제안한 일반화된 로렌츠 곡선을 기반으로 한 Gumbel 분포의 적합도 검정통계량을 새롭게 제안하고자 한다. 본 논문의 구성은 2절에서 일반화된 로렌츠 곡선에 대하여 설명한 후 이를 기반으로 한 변형된 일반화된 로렌츠 곡선을 새롭게 제안하고자 한다. 그리고 변형된 일반화된 로렌츠 곡선을 활용하여 두 가지 형태의 새로운 검정통계량을 제안하고, 두 가지 형태의 새로운 검정통계량을 활용하여 Gumbel 분포의 적합도를 검정하는 새로운 그래프 방법을 제안하고자 한다. 3절에서 다양한 분포에 대하여 몬테카를로 시뮬레이션 (Monte Carlo simulation)을 통하여 새로운 검정통계량의 검정력을 확인하고 실제 사례 데이터에 적용한 후 4절에서 결론을 내리고자 한다.

2. 새로운 적합도 검정

2.1. 일반화된 로렌츠 곡선

로렌츠 곡선 (Lorenz curve)은 Lorenz (1905)에 의해 처음 제안된 것으로 이는 소득의 불균형 정도를 비교하기 위한 방법으로 제안되어졌다. 이러한 로렌츠 곡선은 증가하는 오목한 형태 (strictly concave)의 효용함수 (utility function)이다. Atkinson (1970)은 로렌츠 곡선을 이용하여 더 위쪽에 있는 로렌츠 곡선이 아래쪽에 있는 로렌츠 곡선에 비하여 더 소득 또는 부의 분배가 잘 되어져 있다고 판단하였다. 하지만 이러한 부나 소득의 분배를 비교하기 위해서는 비교대상이 되는 두 집단의 평균이 서로 같다고 할 때 가능하다. 따라서 이러한 문제점을 해결하기 위해 Shorrocks (1983)이 평균이 같지 않은 두 집단의 경우에도 부나 소득의 분배를 비교하기 위해 새롭게 일반화된 로렌츠 곡선 (generalized Lorenz curve)를 제안하였다. 일반화된 로렌츠 곡선은 단순히 로렌츠 곡선의 각 지점을 평균만큼 곱하여 그 크기를 증가시킨 형태이다. 따라서 일반화된 로렌츠 곡선의 최종 높이는 그 집단의 평균 부의 크기를 나타내고, 그 곡률은 부의 불균형 정도를 나타낸다.

Gastwirth (1971)는 로렌츠 곡선을 연속형 뿐만 아니라 이산형 분포의 누적분포함수의 역의 형태를 이용하여 다음과 같이 새롭게 정의하였다.

$$L(p) = \frac{1}{\mu} \int_0^{F^{-1}(p)} x dF(x), \quad 0 \leq p \leq 1,$$

여기서 $F^{-1}(p) = \inf\{y | F(y) \geq p\}$, μ 는 누적분포함수 F 의 평균을 나타낸다. 이러한 로렌츠 곡선을 그림으로 나타내면 Figure 2.1의 (a)와 같다. 일반화된 로렌츠 곡선은 로렌츠 곡선을 평균만큼 그 크기를 증가시킨 형태이므로 일반화된 로렌츠 곡선은 다음과 같이 정의할 수 있다.

$$GL(p) = \mu L(p) = \int_0^{F^{-1}(p)} x dF(x), \quad 0 \leq p \leq 1.$$

이러한 일반화된 로렌츠 곡선을 그림으로 나타내면 Figure 2.1의 (b)와 같다.

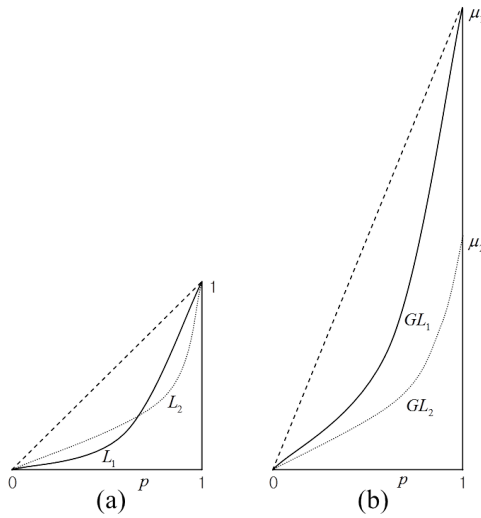


Figure 2.1 (a) Lorenz curve and (b) generalized Lorenz curve

2.2. 새로운 적합도 검정통계량과 그래프

새로운 적합도 검정통계량을 제안하기에 앞서 먼저 변형된 일반화된 로렌츠 곡선 (modified generalized Lorenz curve)을 제안하고자 한다. X_1, X_2, \dots, X_n 을 순서 통계량 $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ 을 가지고 있는 확률변수라고 하자. 그런데 일반화된 로렌츠 곡선의 식은 소득의 자료에 대하여 정의하고 있기 때문에 확률변수 X 는 모두 비음 (non-negative)의 값을 가지고 있다고 가정하고 있다. 하지만 Gumbel 분포의 경우 비음의 값을 가질 수 있음으로 이를 해결하기 위해 모든 순서 통계량 값에서 첫 번째 순서 통계량 값을 빼는 것으로 해결하였다. 그리고 일반화된 로렌츠 곡선은 증가하는 오목한 형태의 곡선이므로 분포의 형태의 특성을 제대로 나타낼 수 없다. 따라서 이를 해결하기 위해 모든 순서 통계량에서 첫 번째 순서 통계량 값을 빼 결과를 모두 더하여 $(1 - p_j)$ 를 곱함으로써 해결하였다. 이러한 점을 반영한 변형된 일반화된 로렌츠 곡선은 다음과 같이 정의할 수 있다.

$$mGLC(p_j) = \frac{1}{X_{n:n} - X_{1:n}} \left[\sum_{i=1}^j (X_{i:n} - X_{1:n}) + (1 - p_j) \sum_{i=1}^n (X_{i:n} - X_{1:n}) \right],$$

여기서 $p_j = j/n, j = 1, 2, \dots, n$ 이다. 이렇게 정의된 $mGLC$ 는 위치모수 (location parameter)와 척도모수 (scale parameter)에 대하여 무관함을 손쉽게 알 수 있다. $mGLC(p_j)$ 를 다양한 분포 (Gumbel 분포 (Gumbel distribution), 표준정규분포 (standard normal distribution), 로그-감마분포 (log-gamma distribution), t -분포 (t -distribution), 이중지수분포 (double exponential distribution), 균일 분포 (uniform distribution), 로지스틱분포 (logistic distribution)의 경우 그 형태를 알아보기 위하여 각 분포를 따르는 50개의 데이터를 누적분포함수의 역함수 ($X_{i:n} = F^{-1}(i/(n + 1))$)를 이용하여 추출하였다. 그 결과 Figure 2.2와 같다.

이러한 변형된 일반화된 로렌츠 곡선을 이용하여 새로운 적합도 검정통계량을 제안하고자 한다. 앞서 정의한 변형된 일반화된 로렌츠 곡선은 각 분포마다 고유한 형태를 가지고 있다. 따라서 만약 샘플 데이터가 정확하게 Gumbel 분포를 따르고 있다면, 정확히 Gumbel 분포를 따르는 값에 의해 계산되

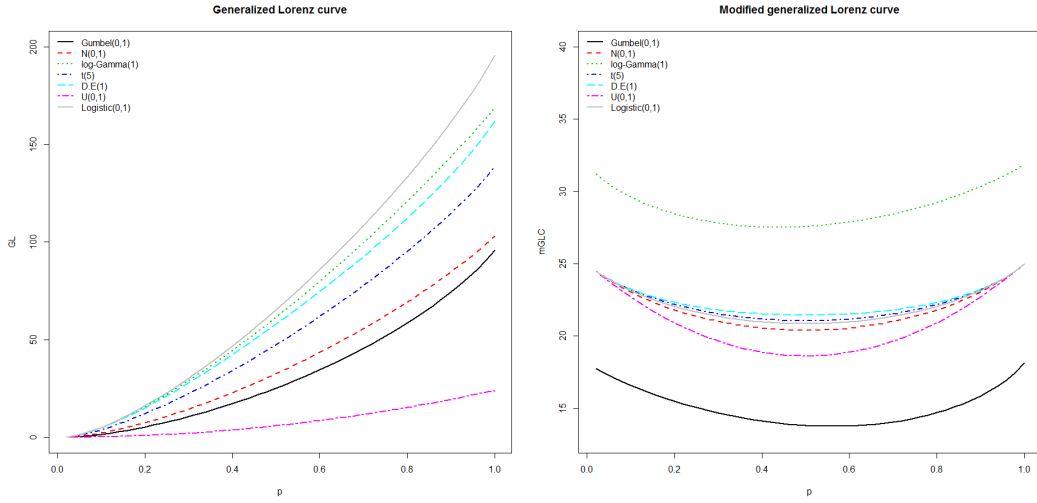


Figure 2.2 Generalized Lorenz curve and modified generalized Lorenz curve of various distributions

어떤 $mGLC$ 와 샘플 데이터를 대입하여 계산한 $mGLC$ 는 서로 같은 값을 가지게 되어 모두 1의 값을 가지게 될 것이다. 따라서 이러한 점을 고려하여 변형된 표본 일반화된 로렌츠 곡선 (modified sample generalized Lorenz curve)을 다음과 같이 정의할 수 있다.

$$msGLC_1(p_j) = \frac{mGLC(p_j)}{mGLC_F(p_j)}, \tag{2.1}$$

여기서

$$mGLC_F(p_j) = \left[F^{-1} \left(\frac{n}{n+1} \right) - F^{-1} \left(\frac{1}{n+1} \right) \right]^{-1} \left[\sum_{i=1}^j \left\{ F^{-1} \left(\frac{i}{n+1} \right) - F^{-1} \left(\frac{1}{n+1} \right) \right\} + (1-p_j) \sum_{i=1}^n \left\{ F^{-1} \left(\frac{i}{n+1} \right) - F^{-1} \left(\frac{1}{n+1} \right) \right\} \right],$$

F^{-1} 는 Gumbel 분포의 누적분포의 역함수이다.

또한 샘플 데이터가 정확하게 Gumbel 분포를 따른다면 $msGLC_1(p_j)$ 의 값을 제공할 하여도 모두 1의 값을 가지게 될 것이므로, 이러한 점을 고려하여 변형된 표본 일반화된 로렌츠 곡선을 다음과 같이도 정의할 수 있다.

$$msGLC_2(p_j) = \left[\frac{mGLC(p_j)}{mGLC_F(p_j)} \right]^2. \tag{2.2}$$

따라서 $msGLC_1(p_j)$ 과 $msGLC_2(p_j)$ 모두 1에서 값이 멀어질수록 Gumbel 분포를 따르지 않는다는 것을 알 수 있으므로 이러한 점을 고려하여 다음과 같은 두 가지 새로운 적합도 검정통계량을 제안할

수 있다.

$$GLC_1 = \max \left[\left| 1 - \frac{mGLC(p_j)}{mGLC_F(p_j)} \right| \right], \tag{2.3}$$

$$GLC_2 = \max \left[\left| 1 - \left[\frac{mGLC(p_j)}{mGLC_F(p_j)} \right]^2 \right| \right]. \tag{2.4}$$

즉, 만약 데이터가 Gumbel 분포를 따른다면 GLC_1 과 GLC_2 는 매우 작은값을 가지게 될 것이라고 기대할 수 있다. 따라서 만약 GLC_1 과 GLC_2 의 값이 기각값 (critical value)보다 크게 된다면 귀무가설을 기각하게 되어 Gumbel 분포를 따르지 않는다고 판단할 수 있다. 하지만 GLC_1 과 GLC_2 은 정확하게 어떠한 분포를 따르는지 이론적으로 밝히기는 매우 어려우므로 몬테카를로 시뮬레이션을 통하여 표본의 크기가 10에서 50까지일 때 5% 유의수준에서의 기각값은 Table 2.1과 같이 나타낼 수 있다.

Table 2.1 The critical values for goodness-of-fit test for Gumbel distribution

n	10	20	30	40	50
GLC_1	.5957059	.5090583	.4645597	.4389910	.4207487
GLC_2	1.3599460	1.0829440	.9379359	.8741304	.8329360

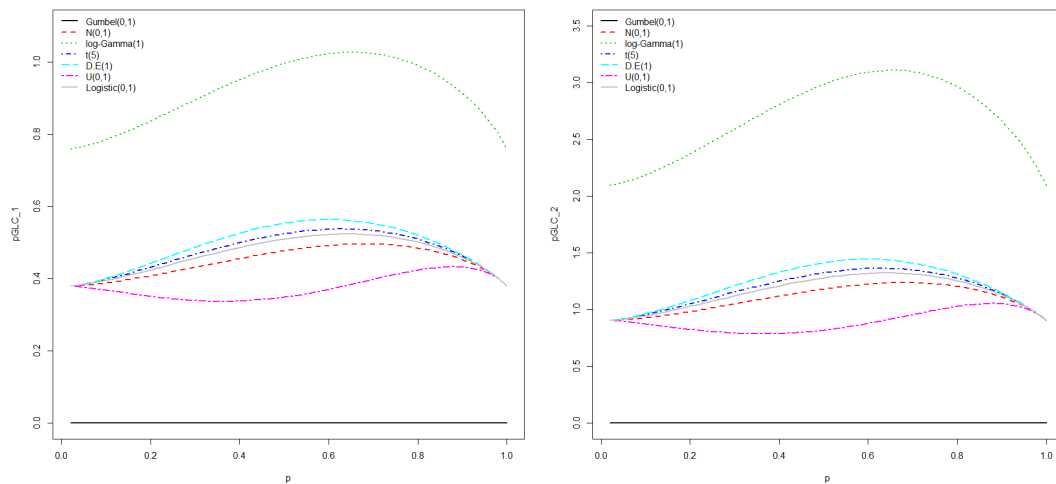


Figure 2.3 $pGLC_1$ and $pGLC_2$ of various distributions

또한 변형된 표본 일반화된 로렌츠 곡선을 이용하면 새로운 적합도 검정 그래프를 생각할 수 있다. 만약 데이터가 Gumbel 분포를 따른다면 $msGLC_1(p_j)$ 과 $msGLC_2(p_j)$ 는 모두 1의 값을 가지게 될 것이

다. 따라서 다음과 같은 새로운 적합도 검정 그래프를 제안할 수 있다.

$$pGLC_1(p_j) = \left| 1 - \frac{mGLC(p_j)}{mGLC_F(p_j)} \right|, \quad (2.5)$$

$$pGLC_2(p_j) = \left| 1 - \left[\frac{mGLC(p_j)}{mGLC_F(p_j)} \right]^2 \right|. \quad (2.6)$$

위의 $pGLC_1(p_j)$ 과 $pGLC_2(p_j)$ 값들은 데이터가 Gumbel 분포를 따른다면 플롯되는 점들은 모두 x 축과 일치하게 될 것이다. 따라서 플롯되는 점들이 x 축과 떨어져 있는 정도를 통해서 Gumbel 분포를 따르는지 따르지 않는지 그 여부를 판단할 수 있다.

$pGLC_1(p_j)$ 과 $pGLC_2(p_j)$ 를 다양한 분포 (Gumbel 분포, 표준정규분포, 로그-감마분포, t 분포, 이중지수분포, 균일분포, 로지스틱분포)의 경우 그 형태를 알아보기 위하여 각 분포를 따르는 50개의 데이터를 누적분포함수의 역함수 ($X_{i:n} = F^{-1}(i/(n+1))$)를 이용하여 추출하였다. 그 결과 Figure 2.3과 같다.

3. 모의 실험과 실제 사례

3.1. 모의 실험

새롭게 제안한 GLC_1 과 GLC_2 검정통계량의 검정력을 평가하기 위해 몬테카를로 시뮬레이션을 수행하였다. 비교를 위한 분포로 표준정규분포, 로그감마분포, t 분포, 이중지수분포, 균일분포, 로지스틱분포를 사용하였다. GLC_1 과 GLC_2 검정통계량의 검정력 비교를 위해 Anderson - Darling 검정 (AD), Cramer - vonMises 검정 (CVM), 수정된 Anderson - Darling 검정 (ZAD)을 사용하였다. 또한 검정력 비교를 위해 표본의 크기가 10, 20, 30, 40, 50인 경우를 각각의 분포에서 10,000번 샘플링하였고, Table 2.1의 기각역 안에 들어가는 샘플의 수를 계산하여 검정력을 측정하였다. 그 결과 유의수준 5%에서 측정된 각 검정통계량의 검정력은 Table 3.1과 같다. Table 3.1을 살펴보면, 일반적으로 모든 검정통계량에서 표본의 크기가 커질수록 검정력이 커지는 것을 확인할 수 있고, 로그감마분포의 경우 다른 분포에 비해 검정력의 크기가 큰 것을 확인할 수 있다. 그리고 검정통계량 간의 검정력 비교에서는 GLC_1 의 경우 Anderson - Darling 검정, Cramer - vonMises 검정, 수정된 Anderson - Darling 검정에 비하여 총 30가지 경우 중 17가지가 검정력이 우수하다고 결과가 나타났고, GLC_2 의 경우 Anderson - Darling 검정, Cramer - vonMises 검정, 수정된 Anderson - Darling 검정에 비하여 총 30가지 경우 중 23가지가 검정력이 우수하다고 결과가 나타났다. 즉, GLC_1 과 GLC_2 검정통계량이 기존의 Anderson - Darling 검정, Cramer - vonMises 검정, 수정된 Anderson - Darling 검정에 비하여 Gumbel 분포의 적합도 검정에는 더 우수한 결과를 나타낸다는 것을 알 수 있다. 또한 GLC_1 과 GLC_2 검정통계량 간의 비교에서는 GLC_2 검정통계량이 GLC_1 검정통계량에 비해서 모든 경우에서 검정력이 우수하다는 결과가 나타나 두 검정통계량 중 GLC_2 검정통계량을 사용하는 것이 Gumbel 분포를 잘 검출할 수 있다고 할 수 있다. 하지만 균일분포의 경우 GLC_1 과 GLC_2 검정통계량 보다는 수정된 Anderson - Darling 검정이 더 우수한 것으로 나타났다.

3.2. 실제 사례

앞서 제안한 GLC_1 과 GLC_2 검정통계량과 $pGLC_1$ 과 $pGLC_2$ 그래프를 실제 사례 데이터에 적용하기 위해 1859년부터 1945년까지 호주 시드니의 연간 일 최대 강수량 자료를 사용하였다 (Hall Jr. 등,

Table 3.1 Estimated power of the GLC_1 and GLC_2 test the alternative distributions at the significance level 5%

Distribution	n	AD	CVM	ZAD	GLC_1	GLC_2
$N(0,1)$	10	.1634	.1134	.1581	.1659	.2132
	20	.2919	.1805	.2519	.3138	.3989
	30	.4046	.2597	.3501	.4492	.5620
	40	.5101	.3458	.4372	.5522	.6644
	50	.6131	.4333	.5204	.6432	.7458
log-gamma(0,1)	10	.4519	.3478	.4008	.4576	.5257
	20	.7948	.6370	.6972	.8058	.8562
	30	.9261	.8201	.8630	.9376	.9627
	40	.9791	.9246	.9232	.9794	.9902
	50	.9941	.9727	.9513	.9938	.9975
$t(5)$	10	.2216	.1645	.1525	.2227	.2641
	20	.3758	.2539	.1956	.3921	.4546
	30	.4981	.3507	.2311	.5092	.5831
	40	.5882	.4435	.2496	.5872	.6546
	50	.6675	.5196	.2665	.6384	.7018
D.E(0,1)	10	.2688	.2126	.1490	.2529	.2929
	20	.4492	.3441	.1549	.4061	.4624
	30	.5882	.4690	.1700	.5295	.5982
	40	.6800	.5735	.1689	.5930	.6669
	50	.7763	.6836	.1861	.6689	.7377
$U(0,1)$	10	.1648	.1464	.2654	.1168	.1568
	20	.3808	.3172	.6190	.1959	.2862
	30	.6067	.5130	.8695	.3070	.4570
	40	.7747	.6800	.9681	.4202	.5924
	50	.8879	.8092	.9939	.5379	.7228
logistic(0,1)	10	.2021	.1393	.1460	.2039	.2431
	20	.3420	.2216	.2049	.3606	.4318
	30	.4585	.3020	.2551	.4846	.5699
	40	.5457	.3817	.2936	.5720	.6548
	50	.6282	.4653	.3368	.6366	.7153

1989). 1859년 부터 1945년까지 호주 시드니의 연간 일 최대 강수량 자료의 분포형태는 Figure 3.1과 같다. 분포의 형태를 살펴보면 오른쪽으로 긴 꼬리를 가진 형태를 나타내고 있다.

이를 $pGLC_1$ 과 $pGLC_2$ 그래프를 활용하여 나타내본 결과 $pGLC_1$ 과 $pGLC_2$ 그래프 모두 x 축과 가까운 지점에서 그래프가 플롯되는 것을 확인할 수 있다. 또한 데이터의 분포가 균일분포를 가정하여 플롯했을 때보다 Gumbel 분포를 가정하여 플롯했을 때 x 축과 가까운 지점에서 그래프가 플롯되는 것을 확인할 수 있다. 따라서 위의 호주 시드니의 연간 일 최대 강수량 자료는 Gumbel 분포를 따른다는 것을 알 수 있다.

다음으로 GLC_1 과 GLC_2 검정통계량을 사용하여 Gumbel 분포의 적합도 검정을 실시하였다. 그 결과 $n = 87$ 인 경우 GLC_1 과 GLC_2 검정통계량의 기각값은 각각 0.3825402와 0.7327600으로 나타났고, 그 때의 GLC_1 과 GLC_2 검정통계량 값은 각각 0.2210454와 0.3292876으로 기각값보다 작게 나타나 위의 호주 시드니의 연간 일 최대 강수량 자료는 Gumbel 분포를 따르는 것을 알 수 있다. 이는 Hall Jr. 등 (1989)의 결과와 같음을 알 수 있다.

4. 결론

통계학에서 사용되어지고 있는 Gumbel 분포는 환경과학, 시스템 신뢰성, 수문학과 같은 분야에서 많이 응용되고 있다. 따라서 환경과학, 시스템 신뢰성, 수문학과 관련된 자료를 분석함에 있어서 분석에

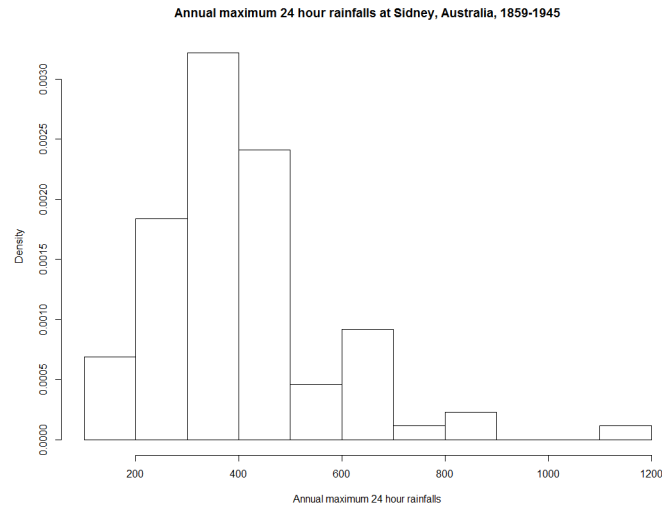


Figure 3.1 Histogram of annual maximum 24 hour rainfalls at Sidney, Australia, 1859-1945

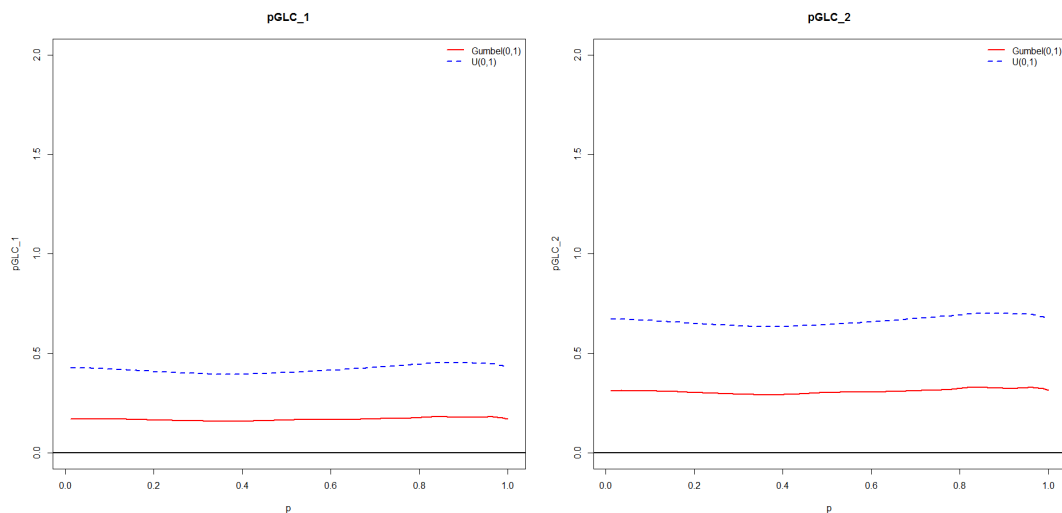


Figure 3.2 $pGLC_1$ and $pGLC_2$ of annual maximum 24 hour rainfalls at Sidney, Australia, 1859-1945

사용되어지는 자료가 Gumbel 분포를 따르는지 확인하는 것은 매우 중요하다. 이를 확인하기 위해 본 논문에서는 새로운 두 가지의 Gumbel 분포의 적합도 검정통계량을 일반화된 로렌츠 곡선을 기반으로 하여 제안하였고, Anderson - Darling 검정, Cramer - vonMises 검정, 수정된 Anderson - Darling 검정과 비교하였다. 그 결과 두 적합도 검정통계량 모두 기존의 검정방법보다 우수하다고 할 수 있었고, 특히 GLC_1 검정통계량은 훨씬 우수함을 알 수 있었다.

또한 새롭게 제안한 변형된 표본 일반화된 로렌츠 곡선 (*msGLC*)을 이용하여 두 가지의 새로운 적합도 검정 그래프 방법을 제안하였다. 새롭게 제안된 그래프를 통하여 손쉽게 데이터가 Gumbel 분포를 따르는지를 파악할 수 있었다.

하지만 새롭게 제안한 두 검정통계량 모두 균일분포의 경우 기존의 검정통계량에 비하여 검정력이 좋지 못하였고, 로그 감마분포를 제외한 대부분의 분포에서 검정력이 상대적으로 우수한 것이지 절대적으로 우수하다고 할 수 없으므로 이를 더욱더 향상시킬 연구가 필요할 것으로 생각한다.

References

- Arizono, I. and Ohata, H. (1989). A test for normality based on Kullback-Leibler information. *American Statistician*, **43**, 20-22.
- Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, **2**, 244-263.
- Cho, Y. and Lee, K. (2014). Goodness-of-fit test for the normality based on the generalized Lorenz curve. *Communications for Statistical Applications and Methods*, **21**, 309-316.
- Gastwirth, J. L. (1971). A general definition of the Lorenz curve. *Econometrica*, **39**, 1037-1039.
- Hall Jr, C. E., Lehnigk, S. H., and Viswanath, G. R. (1989). *Maximum-likelihood parameter estimation of a generalized gumbel distribution* (No. AMSMI/TR-RD-RE-89-5), Army Missile Command Redstone Arsenal AI Research Directorate.
- Kang, S. B., Han, J. T. and Cho, Y. S. (2014a). Goodness-of-fit test for the logistic distribution based on multiply type-II censored samples. *Journal of the Korean Data & Information Science Society*, **25**, 195-209.
- Kang, S. B., Han, J. T., Seo, Y. J. and Jeong, J. (2014b). Goodness-of-fit tests for the inverse Weibull or extreme value distribution based on multiply type-II censored samples. *Journal of the Korean Data & Information Science Society*, **25**, 903-914.
- Kinnison, R. (1989). Correlation coefficient goodness of fit test for the extreme value distribution. *American Statistician*, **43**, 98-100.
- Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**, 209-219.
- Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, **50**, 3-17.
- Song, K. S. (2002). Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory*, **48**, 1103-1117.
- Zempleni, A. (2004). *Goodness-of-fit test in extreme value applications*, Technical Report, University Munchen, Munchen.

Goodness-of-fit test for the gumbel distribution based on the generalized Lorenz curve

Kyeongjun Lee¹

¹Department of Computer Science and Statistics, Daegu University

Received 29 June 2017, revised 11 July 2017, accepted 13 July 2017

Abstract

There are many areas of applications where Gumbel distribution are employed such as environmental sciences, system reliability and hydrology. The goodness-of-fit test for Gumbel distribution is very important in environmental sciences, system reliability and hydrology data analysis. Therefore, we propose the two test statistics to test goodness-of-fit for the Gumbel distribution based on the generalized Lorenz curve. We compare the new test statistic with the Anderson - Darling test, Cramer - vonMises test, and modified Anderson - Darling test in terms of the power of the test through by Monte Carlo method. As a result, the new test statistics are more powerful than the other test statistics. Also, we propose new graphic method to goodness-of-fit test for the Gumbel distribution based on the generalized Lorenz curve.

Keywords: Generalized Lorenz curve, goodness-of-fit test, Gumbel distribution, Lorenz curve, power.

¹ Assistant professor, Department of Computer Science and Statistics, Daegu University, Gyeongsan 38453, Korea. E-mail: leekj@daegu.ac.kr