

<PartB-1: Covid-19 사망자 예측>

1) 2020.3.7 ~ 2020.8.31 영국의 일별 Covid-19 사망자 수(rolling 7-day average)

	date	St
1	2020-03-07	0.286
2	2020-03-08	0.286
3	2020-03-09	0.429
4	2020-03-10	1.000
5	2020-03-11	1.000
6	2020-03-12	1.286
7	2020-03-13	1.286

(a) 일별 신규 Covid-19 사망자 수(St), 누적 사망자 수(Yt) 시계열도표

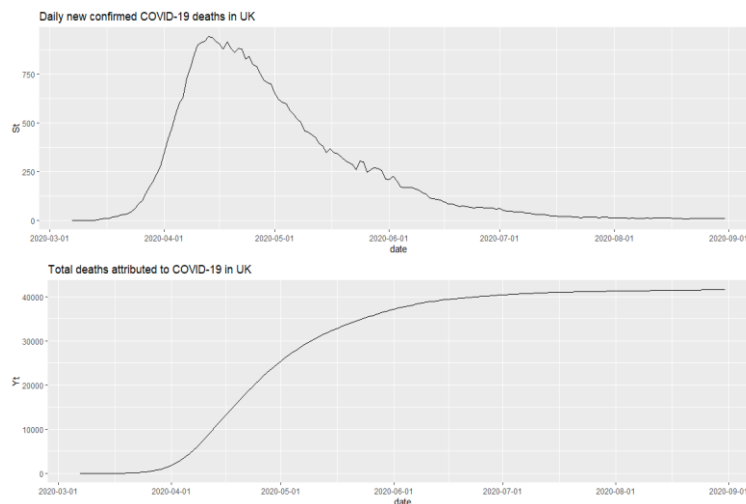


그림 1 영국 일일 신규 Covid-19 사망자 수 시계열 도표 (St , Yt)

2020 년 3 월 7 일 영국에서 코로나 19 로 인한 사망자가 발생한 후 2020 년 4 월 13 일까지 일일 신규 사망자가 급격하게 증가하다가 이후 천천히 감소하며 오른쪽 꼬리가 긴 분포를 보이고 있다. 또한 누적 사망자 수도 시간이 지남에 따라 수렴하는 양상을 보인다.

(b) 영국 내 총 사망자 수 추정

i. OLS 추정법

각각 $n=20, 30, 50$ 일 때 Bass, Logistic, Gumbel model 을 가정하여 OLS 추정법을 통해 모수(m, p, q)를 추정하였다. 아래 표는 모수 추정 결과와 영국의 2020 년 8 월 31 일의 누적 사망자 수($m=41569$)와 비교한 상대 오차를 구한 결과이다. Covid-19 사망자 수는 혁신 계수(p)보다 모방 계수(q)가 더 크다. Covid-19 감염은 접촉에 의해 발생하기 때문이다. 영국의 신규 코로나 사망자 수가 2020 년 4 월 중순에 정점을 찍는다. n 이 작을 경우 학습자료에 정점을 포함하지 못하므로 m 이 작게 추정된다. 또한 신규 코로나 사망자 수는 정점 전후의 증가세(감소세)가 다르다. 이러한 분포를 normal 분포의 모양을 가지는 Bass 와 Logistic model 은 고려하지 못하므로 해당 데이터에는 적절하지 않다. 이로 인해 상대 오차가 Gumbel 에 비해 크다. 따라서 꼬리가 긴 분포를

가지며 정점을 학습자료에 포함하는 $n=50$ 이고 Gumbel model 을 이용한 모형을 최적으로 선택한다.

표 1. OLS 추정 결과

n	model	p	q	m	상대 오차
20	bass	0	0.284	-47098.6	-213.302
	logistic	0	0.312	5014.32	-87.937
	gumbel	·	0.024	49433347	118818.8
30	bass	0	0.276	9694.136	-76.679
	logistic	0	0.281	9354.834	-77.496
	gumbel	·	0.052	122390.9	194.428
50	bass	0.002	0.157	25462.15	-38.747
	logistic	0	0.166	25013.73	-39.826
	gumbel	·	0.078	32854.69	-20.963

ii. Q-Q plot 추정법

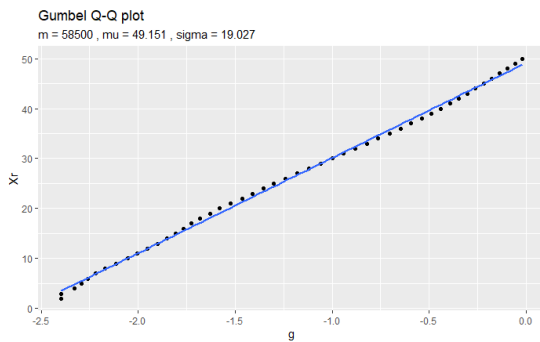


그림 3 Q-Q plot

58500 이다. 그 때의 $\mu = 49.151$, $\sigma = 19.027$ 이다.

OLS 추정법에서 선택한 모형인 $n=50$, Gumbel 모형을 이용하여 Q-Q plot 추정법으로 모수를 추정하였다.

OLS 추정법에서 $m=32854.69$ 으로 추정된 것을 고려하여 $m=30000, 30500, 31000, 31500, \dots, 70000$ 일 때 r 번째 사망자의 사망시간을 $X(r)$ 을 종속변수로

$$\frac{G^{-1}(U_r)}{G^{-1}(r/(m+1))}$$

를 독립변수로 하여 선형회귀모형을

적합시켰다. 그 결과 R^2 를 극대화시키는 m 값은

iii. NLSE

$n=50$, Gumbel 모형을 기반으로 OLS 추정치와 Q-Q plot 추정치를 초기값으로 하여 NLSE 를 구하였다. 모형 1 로 추정한 결과는 $m = 32854.69, q = 0.078$ 이고 모형 2 의 경우 $m = 58486.77, \mu = 49.147, \sigma = 19.025$ 이다.

모형 1: $S_t = aY_{t-1} + bY_{t-1} \cdot \ln Y_{t-1} + e_t, a = q \cdot \ln(m), b = -q$

모형 2: $X_{(r)} = \mu + \sigma(-\ln(-\ln(U_r))) + e_r, U_r = \frac{r}{m+1}$

iv. MLE

$n=50$, Gumbel 모형의 log-likelihood 를 구하여 optim 함수를 통해 이를 maximize 하는 모수 MLE 를 추정하였다. 그 결과 $m = 58459.011, \mu = 49.945, \sigma = 20.007$ 이다.

v. 비교

모형 1 을 기반으로 추정한 OLS 와 NLSE 는 상대 오차는 약 -21%이고 모형 2 를 기반으로 추정한 Q-Q plot, NLSE, MLE 상대 오차가 약 41% 이다. 전자의 경우 전체 사망자 수를 적게 추정하였고 후자의 경우 더 많게 추정하였다. 절대값은 전자가 더 작으므로 정확도는 더 높았다. 또한 같은 모형을 이용할 때 추정 방법 간의 차이는 크게 없다.

2) 2020.02.21 ~ 2020.08.31 이탈리아의 일별 Covid-19 사망자 수

(a) 이탈리아 내 총 사망자 수 추정

이탈리아의 Covid-19 사망자 수 분포도 영국과 동일하다. Covid-19 일일 사망자 수는 급격하게 증가하다가 2020 년 4 월 중순에 정점을 찍고 비교적 완만하게 감소한다. 따라서 오른쪽 꼬리가 긴 분포로 gumbel 모형이 적절할 것이다. $n=20, 30, 50$ 일 때 Bass, Logistic, Gumbel 모형을 이용하여 OLS 추정법으로 모수를 추정하였다. 2020 년 8 월 31 일의 이탈리아의 누적 사망자 수($m=35461.29$)과 비교하였을 때 상대 오차는 n 이 커질수록 절대값이 작아졌고 gumbel 모형이 가장 작았다. 영국과 이탈리아의 코로나 사망자 수의 분포는 거의 동일한 양상을 보이기에 이러한 결과가 나왔을 것이다.

$N=50$ 까지를 학습데이터로, gumbel 모형을 이용하여 Q-Q plot, NLSE, MLE 를 구하였다. 그 결과 m 은 각각 48000 (Q-Q plot), 26636.6 (NLSE-모형 1), 48037.7 (NLSE-모형 2), 47752.096 (MLE)이다. 상대 오차는 OLS, NLSE-모형 1 의 경우 약 -25%, Q-Q plot, NLSE-모형 2, MLE 의 경우 약 35%이다. 영국과 동일하게 전자의 경우 총 사망자 수를 적게 추정하나 상대 오차는 적고, 후자의 경우 상대 오차는 크나 총 사망자 수를 많이 추정한다.