

# 근사 꼬리분포의 유형별 적용 모형 고찰

## Review of Application Models According to the Classification of Asymptotic Tail Distribution

최 성 운\*

Sung-woon Choi\*

### Abstract

The research classifies three types of asymptotic tail distributions such as long(heavy, thick) tailed distribution, medium tailed distribution and short(light, thin) tailed distribution. The extreme value distributions(EVD) classified in this paper can be used in SPC(Statistical Process Control) control chart and reliability engineering.

**Keywords:** Long-Tailed, Medium-Tailed, Short-Tailed, Distribution EVD

### 1. 서 론

품질향상을 위한 통계적 품질 및 공정관리(Statistical Quality and Process Control), 재무투자 의사결정을 위한 금융공학(Financial Engineering), 댐의 수위조절을 위한 댐 이론(Dam Theory), IFR(Increasing Failure Rate) 유형의 신뢰성 공학(Reliability Engineering)[2,5] 등에서는 평균, 중앙값, 최빈값이 일치하는 좌우대칭인 종 모양(Bell Shaped)의 정규분포를 사용한다.

중심극한정리(Central Limit Theorem)에 의하여 중심인 평균에 관심을 두는 기존의 정규분포 적용과 다르게 최근 꼬리부분에 대한 연구가 활발하게 이루어지고 있다.

---

\* 경원대학교 산업공학과

최근 금융공학 분야에서는 정규분포의 꼬리 위험(Tail Risk)은 발생가능성이 낮거나 예측하기 어려우며 발생하면 충격이 크나 이를 미리 알 수가 없기 때문에 우리의 무지에 대한 인지를 통해 이를 대처해야 한다는 Black Swan [1] 이론이 있다. 이 이론에서는 좌우대칭인 정상적인 꼬리를 갖는 정규분포를 적용한 리스크 금융공학은 충격적인 금융쇼크를 알아 낼 수 없으며 전문가는 발생된 후 이러한 극단적인 꼬리모양을 알았던 것처럼 설명하는 귀납적 오류를 저지르고 있다고 비판하고 있다. 정치경제학 분야에서의 꼬리 위험은 Black Swan에서의 Long Tail과 다르게 사건이 일어날 확률이 반복가능하고 예측이 가능한 Fat Tail이라고 주장하고 있다.[4] 이는 FMEA의 RPN (Risk Priority Number=Occurrence\*Severity\*Detectability)에서 Occurrence는 낮고 Severity는 높다는 점에서는 Long Tail과 Fat Tail이 공통적이나 Detectability에서 차별성을 갖는다고 할 수 있다. 두 꼬리관계는 고객도착시간 간격이 지수분포 즉 도착수가 포아송분포인 Poisson Process와 도착고객이 구매하는 금액이 포아송분포인 Compound Poisson Process와의 관계와 같다. 또한 웹 마케팅 분야에서는 Pareto 분포의 Long Tail에 해당하는 다수의 경미한 구매자가 전체 매출액의 상당부분을 차지한다는 이론이 있다.[6] 이는 매출액의 80%를 소수의 중요 고객 20%가 이룬다는 80:20의 전통적인 Pareto 분포의 마케팅 이론과 대치된다.

정규분포의 꼬리유형은 좌우대칭인 종모양의 pdf(Probability Density Function)를 직선의 정규확률지(Normal Probability Paper)로 나타낼 경우 왜도(Skewness)가 커지다가 작아지는 Right Skew(정규확률지 모양은 Convex), 작아지다가 커지는 Left Skew(정규확률지는 Concave), 첨도가 뾰족하며 꼬리가 얇거나(정규확률지는 S자형), 납작하며 꼬리가 두꺼운(정규확률지 모양은 역S자형) 경우가 있다.

신뢰성 공학에서 사용되는 꼬리분포는 Medium Tail, Long(Heavy, Thick) Tail, Short(Light, Thin) Tail인 경우 3가지 경우가 있으며 금융공학에서는 Heavy Tail의 분포를 주로 다룬다.[3,8] 꼬리에 존재하는 이상값(Outlier)에 강건한 중앙값, 범위, Midrange, Median Range, Trimmed Mean, IQR(Inter Quantile Range) 등을 적용한 SPC 관리도 연구[7]와 Heavy Tail 분포를 가정한 관리도의 연구가 있다.[9]

따라서 본 연구에서는 신뢰성공학과 통계적 공정관리(SPC : Statistical Process Control)에서 적용가능한 근사적 꼬리분포를 유형화하고 고찰한다.

## 2. Heavy Tail 분포의 종류 및 적용

### 2.1 SPC 관리도 적용

최대 강도를 갖는 기계역학 또는 토목구조의 시스템, 고도의 사용 신뢰성이 요구되는 첨단 통신 제품 등은 SPC 관리도에서 가정되는 정규분포의 이론이 적용되지 않는다. 또한 정규분포의 이상값(Outlier)과 가피원인(Assignable Cause)인 꼬리에 Robust한 SPC 관리도 역시 적용될 수 없다.

이러한 경우 극단적인 최대 충격량과 수명을 고려한 오른쪽 꼬리가 긴 최대 분포를 고려하여야 한다. 기존의 SPC 관리도는 정규분포 가정 하에서 OLS(Ordinary Least Squares)의 Shewhart  $3\sigma$  관리용 관리한계를  $\mu \pm 3\sigma/\sqrt{n}$ 으로 설정한다. 그러나 만약 Right Skew된 Heavy(Long) Tail 분포인 경우 이 관리도를 그대로 적용할 경우 과대평가(Overestimate)된다. 이러한 단점을 극복하기 위해 자유도  $\nu$ 인 무한대의 분산을 갖는  $t$ 분포를 고려한다. 이 분포에서는 긴 꼬리를 자유도에 의한 함수

$HT(\nu) = 2(\Gamma((\nu+1)/2))/(\nu\pi)^{1/2}(\Gamma(\nu/2))$ 로 표현한다. 이는 LAD(Least Absolute Deviations)에 의해 유도되었으며[9],  $HT(1) = 2/\pi$ ,  $HT(2) = (0.54)^{1/2}$ 이고 무한대의 분산을 가지므로,  $HT(3) = (0.54)^{1/2}$ 이나 유한대의 분산을 가지므로 긴 꼬리인 경우  $HT(1)$ ,  $HT(2)$ 를 적용한다.[9]  $HT(\nu)$ 는 1보다 작으므로 관리한계는  $\mu \pm (HT(\nu))3\sigma/\sqrt{n}$ 을 이용할 경우 정규분포를 적용한 경우보다 관리한계폭이 좁아져 더욱 정밀한 공정관리를 수행하게 된다.[9]

또한  $y_i = a + \phi y_{i-1} + e_i$ 의  $AR(1)$  프로세스에서 Long Tail  $t$ 분포를 적용하면 관리한계는  $a/(1-\phi) \pm (HT(\nu))3\sigma_e/\sqrt{n}$ 이 된다.[9]

### 2.3 신뢰성 적용

Densities가  $EXP(-x^{-\alpha})$ 의 Frechet Type을 갖는 GEV(Greatest or Maximum Extreme Value)로 신뢰성 공학에서는 Long Tail 분포로 분류한다. Long Tail 분포는 Right Skew된 분포로 Pareto 분포, Log Gamma 분포, Cauchy 분포가 있으며 Scale 척도  $\sigma$ , Shape(Power) 척도  $\alpha$ , Location(Threshold, Guarantee) 척도  $\mu$ 가 고려될 경우 Densities는  $EXP(-((x-\mu)/\sigma)^{-\alpha})$ 이다.

Long Tail 분포는 최대 극단분포로 범람 시 댐의 최대 수위량 조절 또는 금융의 최대 쇼크(Shock, Impulsion) 등에 응용되며 오른쪽으로 꼬리가 긴 수식의 표현은 떨어지지 않는 실수  $e$ 와 오일러 상수 또는  $\pi$ 를 이용한다. 이 분포는 신뢰성 공학에서 인장강도, 파괴강도, 재료의 피로현상, 마모현상과 같은 최대 파괴 역학모형과 고장이 좀처럼 일어나지 않는 장기간 사용 검증된 고 신뢰성의 전자통신 첨단제품과 시스템에 적용된다. 이 분포는 정규분포의 대표적인 속성인 중심극한정리(Central Limit Theorem)를 적용할 수 없을 정도로 분산이 유한하지 않는 긴 꼬리모양을 갖고 있으며 암의 전이처럼 단조성에 충격적인 변동을 갖는다.

### 3. Medium Tail 분포의 종류 및 적용

Densities가  $EXP(-EXP(-x))$ 의 Gumbel Type을 갖는 EVD(Extreme Value Distribution)로 신뢰성 공학에서는 Asymptotic Tail 분포로 분류한다. 이에 대한 분포로는 지수분포, Gamma 분포, Log 정규분포, Log Weibull 분포가 있으며 Scale 척도

$\sigma$ 와 Location(Threshold, Guarantee) 척도  $\mu$ 를 고려할 경우 Densities는  $\text{EXP}(-\text{EXP}((x-\mu)/\sigma))$ 이다.

기존의 좌우대칭인 정규분포의  $\mu \pm 3\sigma$ 의 안쪽 모양에서  $\mu \pm 4\sigma$  이상의 바깥쪽 중간 꼬리모양을 다루는 Medium Tail 분포는 Short Tail, Long Tail의 길고 짧은 꼬리유형에 따라  $\mu \pm 6\sigma$  밖의 꼬리도 고려해야 한다.

#### 4. Short Tail 분포의 종류 및 적용

Densities가  $\text{EXP}(-(-x)^\alpha)$ 의 Weibull Type을 갖는 LEV(Least or Minimum Extreme Value)로 신뢰성 공학에서는 Light Tail 분포로 분류한다. Short Tail 분포는 Left Skew된 분포로 일양분포, Beta 분포, Weibull 분포가 있으며 Medium Tail과 다르게 Shape 척도  $\alpha$ 가 존재한다. 3가지 척도가 고려된 Densities는  $\text{EXP}(-((\mu-x)/\sigma)^\alpha)$ 이다.

Short Tail 분포는 체인같은 신뢰성 직렬구조에서 가장 적은 신뢰도를 갖는 부품이 제품시스템에 영향을 준다는 Weakest Link 이론에 적용되며 가뭄 시 댐의 최저 수위량 조절 또는 미사일 부식성 화학물질에 의한 부품수명 등에 응용된다. Long Tail 분포와 같이 EVD는 떨어지지 않는 실수  $e$ 와 Euler 상수,  $\pi$ 로 구성된 수식으로 표현된다. RBD(Reliability Block Diagram)에서 직렬(Series)구조의 Frailty 모수와 병렬(Parallel)구조의 Resilience와의 Duality 관계성도 꼬리분포의 적용이 가능하다.

#### 5. 결 론

본 연구에서는 중심극한 정리의 정규분포를 적용할 수 없는 근사적 꼬리분포의 모형을 Long Tail, Medium Tail, Short Tail의 3가지 유형으로 분류하여 적용 모형을 고찰하였다. 특히 Heavy형의 Long Tail  $t$ 분포를 이용한 SPC 관리도는 극한 역학모형과 극한수명을 갖는 제품과 시스템에 적용할 수 있고 자동화 생산의 시계열 자기상관 공정에 응용 가능하다. 향후 연구로는 관리도의 용도에 따라 해석용, 관리용 관리도로 구분하여 데이터 가중체계 또는 독립성에 따라 실무에서 적용이 가능한 Long Tail 관리도의 개발에 있다.

#### 6. 참 고 문 헌

- [1] 니심 니콜라스 탈레브, 차익종 번역, 블랙스완, 동녘사이언스, 2008.
- [2] 박동호 외, 공학도를 위한 수명분포 개념과 응용, 영지문화사, 2006.
- [3] 윤석훈, “두터운 꼬리를 갖는 확률분포란?”, 한국통계학회 소식지, 22(5)(2009):16-19.
- [4] 이안 브래머, 프레스트 키트, 한상석 번역, 팻 테일, 현대경제 연구원, 2010.
- [5] 최성운, “신뢰성 척도 및 분포의 적용”, 대한안전경영과학회지, 7(5)(2005):175-184.

- [6] 크리스 앤더슨, 이노무브 그룹 번역, 롱테일 경제학, 랜덤하우스 코리아, 2006.
- [7] Bakir S.T., "A Distribution-Free Shewhart Quality Control Chart Based On Signed-Ranks", *Quality Engineering*, 16(4)(2004):613-623.
- [8] Teugels J.L., "The Class of Subexponential Distributions", *The Annals of Probability*, 3(6)(1975):1000-1011.
- [9] Thaga K., "Control Chart for Autocorrelated Processes with Heavy Tailed Distributions", *Economic Quality Control*, 23(2)(2008):197-206.