Kompozit Tweedie-Pareto dağılımı: Aktüeryal Hasar Modellemesi örneği

İsmail GÜR[1,2] Kasırga YILDIRAK[1]

[1]Hacettepe Üniversitesi,Aktüerya Bilimleri Bölümü

[2]Fırat Üniversitesi, İstatistik Bölümü

İçerik

Aktüeryal Anlamda Hasar Bileşenleri

Hasar Dağılımları 2 bileşenden oluşmaktadır.

- Hasar sayısı(frekans)=>N
- Hasar tutarı(şiddeti)=>X [1]

Bireysel Risk Modeli çerçevesinde Toplam Hasar

$$S_N = X_1 + X_2 + \dots + X_N$$

$$E(S) = \sum_{i=1}^{N} E(X_i)$$

$$Var(S) = \sum_{i=1}^{N} Var(X_i)$$

$$X_i = I_i \times B_i = \begin{cases} 0, I_i = 0 \\ B_i, I_i = 1 \end{cases}$$

Kolektif Risk Modeli

Kolektif Risk Modeli çerçevesinde Toplam Hasar

$$\mu = E(X_i)$$
$$\sigma^2 = Var(X_i)$$
$$E(S|N) = E(X_1 + X_2 + ... + X_N|N) = \mu N$$
$$Var(S|N) = Var(X_1 + X_2 + ... + X_N|N) = \sigma^2 N$$
$$E(S_N) = E_N(E(S|N)) = E_N(\mu N) = \mu E(N)$$
$$\begin{aligned} Var(S_N) &= E_N(Var(S|N)) + Var_N(E(S|N)) \\ &= E_N(\sigma^2 N) + Var_N(\mu N) \\ &= \sigma^2 E_N(N) + \mu^2 Var(N) \end{aligned}$$

Tweedie Dağılımı

- Üstel Dağılım ailesinin bir üyesidir.
- Tweedie dağılımı,sıfır değerinde yığılmalı veriye sahip,diğer değerleri ise sürekli olan, Bileşik Poisson- Gamma dağılımının alternatif parametreleri elde edilmiş özel bir yapısıdır [2].

Üstel dağılım ailesi, olasılık dağılımlarının kümelerinden sadece bir tanesidir ve Eşitlik 1 şeklinde özel bir yapıya sahiptir.

$$f(x;\theta,\lambda) = c(x,\lambda)\exp(\lambda(\theta x - \kappa(\theta)), \theta \in \Theta, \lambda \in \Lambda \qquad (1)$$

Denklemde θ kanonik parametre
λ kesinlik parametresi
κ kümülant yaratan fonksiyon
$c(x,\lambda)$ ise negatif olmayan normalleştirme parametresi olarak geçmektedir.

Üstel Dağılım Ailesi

- Bazı ÜDA üyeleri Binom, Poisson, Gamma and Normal dağılımlarıdır.
- Eğer herhangi X rd, ÜDA yapısına sahipse, moment çıkaran fonksiyonu Eşitlik 2 gibi elde edilir.

$$m(t) = \exp(\lambda(\kappa(\theta + t/\lambda) - \kappa(\theta)) \quad (2)$$

- m fonksiyonunun t'ye göre türevlenmesi ile beklenen değer ve varyans elde edilmektedir.

$$E(X) = \kappa'(\theta) \quad V(X) = \frac{\kappa''(\theta)}{\lambda}$$

- Varyansın, ortalamanın bir fonksiyonu olduğu görülmektedir.
- Eğer ortalama $\mu = E(X)$ and $\tau = \kappa'$ olarak ifade edilirse,varyans fonksiyonu $V(\mu) = (\kappa''^{\circ}\tau^{-1})(\mu)$ olacaktır.

Tweedie Ailesi

- Tweedie ailesi ise, varyans fonksiyonun değişimi ile farklı yapılara evrilebilen, ÜDA ailesi üyelerinden biridir.
- Eğer varyans fonksiyonun tanım aralığı $(0,\infty)$ ve

$$V(\mu) = \mu^p, p \in R$$

tanımlanıyorsa, ÜDA ailesi Tweedie ailesi olarak adlandırılabilmektedir.

Tablo 1: Farklı p değerleri ile Tweedie Dağılımları

P değerleri	Dağılımlar
p=0	Normal
p=1	Poisson
$p \in (1,2)$	Compound Poisson- Gamma
p=2	Gamma
p=3	Inverse Gaussian

Tweedie ve Bileşik Poisson-Gamma İlişkisi

Tweedie dağılımı, Tw(p,μ,λ) olarak gösterilsin. Bu dağılım, ortalama μ ve kesinlik parametresi λ olan varyans fonksiyonu ise p üsteline sahip olacaktır.
Bileşik Poisson-Gamma dağılımı ise Poisson hızına ve Gamma sıçrama büyüklüğü dağılımına sahip olan,CPG(μ_N,α,τ) olarak gösterilsin. S rd'nin dağılımı CPG dağılıma sahip olacaktır.

$$S = \sum_{i=0}^{N} Y_i$$

Bu ifadede, $Y_0 = 0$ olan ve N, Poisson dağılıma sahip, $Y_1, Y_2 \ldots$ ise bağımsız ve aynı dağılımlı Gamma dağılıma sahip olsun.
Eğer $p \in (1,2)$,$\mu > 0$ ve $\lambda > 0$ olması durumunda,

$$Tw(p,\mu,\lambda) = CPG(\frac{\lambda\mu^{2-p}}{2-p}, -\frac{p-2}{p-1}, \frac{\lambda\mu^{1-p}}{p-1})$$

[2]

Pareto Dağılımı

- Temel uçdeğer ölçüm dağılımlarından biridir.
- Hasar tutarı modellemesinde geniş yer kaplamaktadır.

Pareto Dağılımı olasılık yoğunluk fonksiyonu

$$f(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}$$

Birikimli Dağılım fonksiyonu

$$F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha}$$

for $x > 0, \alpha > 0, \theta > 0$ [3].

Kompozit Dağılımlar

Kompozit dağılımlar aşağıdaki şekilde türetilebilmektedir.

$$f(x) = \begin{cases} w \times f_1(x) & , x \leq \theta \\ (1-w) \times f_2(x), & x > \theta \end{cases}$$

Süreklilik ve türevlenebilirlik koşulunun sağlanması gerekmektedir.

$$\int f(x)dx = 1$$

$$f_1(\theta) = f_2(\theta)$$

$$f_1{}'(\theta) = f_2{}'(\theta)$$

Kompozit Dağılımlar

Bu kapsamda dağılımlar, kesiklendirilerek işlem yapılmaktadır.

$$f_1 * (x) = \frac{f_1(x)}{F_1(x)}$$

$$f_2 * (x) = \frac{f_2(x)}{1 - F_2(x)}$$

Ağırlık değeri

$$w = \frac{1}{1 + \varphi}$$

olarak ifade edildiğinde, birikimli dağılım fonksiyonu,

$$F(x) = \begin{cases} \frac{1}{1+\varphi} \frac{F_1(x)}{F_1(\theta)}, x \leq \theta \\ \frac{1}{1+\varphi} \left(1 + \varphi \frac{F_2(x) - F_2(\theta)}{1 - F_2(\theta)}\right), x > \theta \end{cases}$$

olacaktır.

R'da Kullanılan Paketler

"tweedie" paketi

tweedie::dtweedie(y = y, power = power, mu = mu, phi = phi) [4]

"actuar" paketi

actuar::dpareto(x, shape, scale, log = FALSE) [5]

"gendist" paketi

gendist::dcomposite(x, spec1, arg1, spec2, arg2, initial = 1, log = FALSE) [6]

Kompozit Tweedie-Pareto Dağılımı

- f_1 dağılımının Tweedie yapısında
- f_2 dağılımının Pareto yapısında

olması durumunda kullanılabilir.

Dağılımın Kısıtı

p değerinin 1 ve 2 arasında olması gereklidir=> Süreklilik varsayımı

Bu nedenle nlminb optimizasyon yöntemi kullanılmıştır.

Workers Compensation veriseti

Bu veriseti,

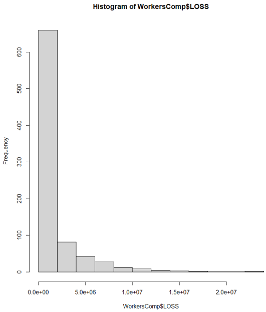
- Stuart Klugman'ın "Bayesian Statistics in Actuarial Science" kitabında [7].
- Edward Frees'in "Regression Modeling with Actuarial and Financial Applications" kitabında [8].
- *insuranceData* paketi içerisinde "*WorkersComp*" ismiyle yer almaktadır [9].

Farklı çalışma grupları için, kısmi yada kalıcı engellilik ödemelerini içermektedir. Toplam 7 yıllık ve 121 meslek grubu,risk bileşeni için analiz edilmiştir.

Workers Compensation veriseti

Tablo 2: Özet İstatistikleri

Min	1st Q	Median	Mean	3rd Q	Max
0	120864	428505	1564540	1566068	22540025

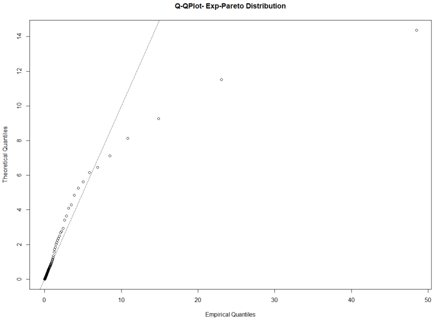


Şekil 1: Workers Comp-Histogram

Kompozit Üstel-Pareto Sonuçları

Karşılaştırma amaçlı Üstel-Pareto dağılımı incelenmiştir.

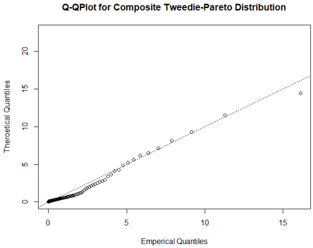
Parameters for Composite Exponential-Pareto Distribution			NLL
rate	shape	scale	
5	0.9427823	0.3695873	946.5582



Şekil 2: Exp-Pareto-QQPlot

Kompozit Tweedie-Pareto Sonuçları

Parameters for Composite Tweedie-Pareto Distribution					NLL
p	mu	phi	scale	shape	
1.729033	1	1.539689	2.521914	2.202256	1290.574



Şekil 3: Tweedie-Pareto-QQPlot

Sonuç ve Değerlendirme

- Kompozit dağılımın toplam hasar modellemesinde kullanılışı gösterilmiştir.
- Tweedie-Pareto yapısının uygulanabilirliği, bazı dağılımlardan daha iyi çalıştığı gösterilmiştir.
- R vasıtasıyla "tweedie" ve "actuar" paketi uygulanmıştır.
- "gendist" paketi yardımıyla kompozit dağılım uygulaması yapılmıştır.

Gelecek çalışmalar

- Bu yapının tek bir değişken için değil, çok değişkenli yapıda incelenmesi
- Farklı yapıdaki kompozit dağılımlar ile karşılaştırma yapılması
- Farklı uçdeğer dağılımlarının etkisinin ölçülebilmesi
- Bu eşik değerin aktüeryal problemlerde bir kısıt yada girdi değişkeni olarak kullanılabilmesi

Kaynaklar I

- S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss models: from data to decisions*, vol. 715.
John Wiley & Sons, 2012.

- O. A. Q. Xacur and J. Garrido, "Generalised linear models for aggregate claims: to tweedie or not?," *European Actuarial Journal*, vol. 5, no. 1, pp. 181–202, 2015.

- M. Fackler, "Reinventing pareto: Fits for all losses, small and large," *Small and Large (January 28, 2021)*, 2021.

- P. K. Dunn and M. P. K. Dunn, "Package 'tweedie'," *R Packag. version. doi*, vol. 10, 2017.

- C. Dutang, V. Goulet, and M. Pigeon, "actuar: An r package for actuarial science," *Journal of Statistical software*, vol. 25, pp. 1–37, 2008.

Kaynaklar II

- S. A. Abu Bakar, S. Nadarajah, Z. A. ABSL Kamarul Adzhar, and I. Mohamed, "Gendist: An r package for generated probability distribution models," *PloS one*, vol. 11, no. 6, p. e0156537, 2016.
- S. A. Klugman, *Bayesian statistics in actuarial science: with emphasis on credibility*, vol. 15. Springer Science & Business Media, 1991.
- E. W. Frees, *Regression modeling with actuarial and financial applications*. Cambridge University Press, 2009.
- A. Wolny and M. A. Wolny, "Package 'insurancedata'," 2014.

Katılımınız için teşekkür ederim.
Herhangi soru veya görüşünüz için ismail.gur@hacettepe.edu.tr adresine mail atabilirsiniz.
Github hesabım: `https://github.com/ismail-gur/WhyR_Turkey`