Amadeusz Filipek

Dodawanie liczb losowych

Zagadnienie dodawania liczb losowych odgrywa istotną rolę w opisie statystycznym rozwoju wartości aktywów finansowych w czasie przyszłym. Opis ten realizowany jest przez funkcję gęstości prawdopodobieństwa dla wszystkich możliwych przedziałów czasowych odpowiadających okresom inwestycji. Dla różnych przedziałów czasowych występują odmienne fluktuacje cen danego aktywa. Jeśli założymy, że fluktuacje którym podlega wartość danego aktywa są niezależne i pochodzą z tego samego rozkładu prawdopodobieństwa to możemy zrekonstruować rozkład zmian ceny aktywa w dowolnej skali czasowej analizując rozkład w któtkim okresie czasu.

Rozważmy sumę dwóch niezależnych zmiennych losowych ∂X_1 oraz ∂X_2 o rozkładach prawdopodobieństwa $P_1(\partial x_1)$ oraz $P_2(\partial x_2)$:

$$X = \partial X_1 + \partial X_2$$

Suma ta może przedstawiać zmienność ceny danego aktywa, gdzie ∂X_1 jest różnicą w cenie pomiędzy dniem dzisiejszym a jutrzejszym oraz ∂X_2 jest różnicą w cenie pomiędzy dniem jutrzejszym a pojutrze. Jaki będzie rozkład prawdopodobieństwa sumy tych zmiennych ? Prawdopodobieństwo, że suma znajdzie się w obszarze dx wokół wartości $X = x$, jest równe sumie po wszystkich możliwościach uzykania wartości $X = x$ czyli po wszystkich kombinacjach $\partial X_1 = \partial x_1$ i $\partial X_2 = \partial x_1$ spełniających $\partial x_1 + \partial x_2 = x$ przemnożonych przez wagi prawdopodobieństwa. Ponieważ zmienne są niezależne, łączne prawdopodobieństwo, że $\partial X_1 = \partial x_1$ oraz $\partial X_2 = x - \partial x_1$ wynosi $P_1(\partial x_1)P_2(x - \partial x_1)$, zatem szukany rozkład dla sumy $N = 2$ zmiennych losowych przyjmuje postać :

$$P(x, N = 2) = \int P_1(x')P_2(x - x')dx'$$

Otrzymane równanie jest konwolucją (splotem) funkcji P_1 i P_2 co zapisać można jako $P = P_1 \otimes P_2$. Sumę tą możemy łatwo uogólnić na przypadek N niezależnych zmiennych losowych ∂X_i o rozkładach $P_i(\partial x_i)$:

$$X = \partial X_1 + \partial X_2 + \partial X_3 + \cdots + \partial X_N$$

Natomiast funkcja gęstości prawdopodobieństwa sumy X jest postaci :

$$P(x, N) = \int P_1(x'_1) \dots P_{N-1}(x'_{N-1})P_N(x - x'_1 - \cdots - x'_{N-1}) \prod_{i=1}^{N-1} dx'_i$$

Założenie, że fluktuacje są niezależne i mają ten sam rozkład powoduje, że $P_1 = P_2 = \cdots = P_N$. Zgodnie z tym założeniem, wystarczy znać rozład elementarnych zmian dla danego okresu czasowego aby zrekonstruować rozkład po pewnym czasie N zmian poprzez N krotny splot rozkładu zmiany elementarnej.

W celu znacznych uproszczeń obliczeń można skorzystać z twierdzenia o konwolucji, które mówi, że transformata Fouriera konwolucji dwóch funkcji jest iloczynem ich transformat Fouriera :

$$\mathcal{F}[f(x) \otimes g(x)] = \mathcal{F}[f(x)]\mathcal{F}[g(x)] = F(z)G(z)$$

zatem :

$$\hat{P}(z,N) = \widehat{P_1}(z)\widehat{P_2}(z) \dots \widehat{P_N}(z) = \hat{P}_1^N(z)$$

Aby uzyskać rozkład końcowy wystarczy teraz zastosować transformatę odwrotną Fouriera.

W wyniku procesu dodawania niezależnych liczb losowych z rozkładu $P_1(\partial x_1)$ otrzymujemy rozkład $P(x,N)$, który w ogólności ma inny kształt. Istnieją jednak takie funkcje rozkładu dla których rozkład sumy zachowuje kształt rozkładu elementarnego. Takie rozkłady nazywamy rozkładami stabilnymi. Zachowanie kształtu oznacza, że można wyznaczyć liniową transformację zależną od N zmiennej x taką aby oba rozkłady się pokrywały :

$$x = a_N x_1 + b_N \;\longrightarrow P(x,N)dx = P_1(x_1)dx_1$$

Jeśli zmienna x jest poprawnie przeskalowana to rozkład sumy dla wybranej skali czasowej jest niezależny od skali czasowej. W tym przypadku ewolucja danego aktywa w funkcji czasu ma taką samą strukturę niezależnie od wybranej skali czasowej, różni się jedynie średnie nachylenie krzywej oraz amplituda fluktuacji. Rozkłady te są zatem samopodobne. Rodzina rozkładów stabilnych składa się z rozkładów Lévy'ego, które zawierają rozkład Gaussa oraz Lorentza. Rozkłady Gaussa oraz Lévy'ego są zatem punktami stałymi operacji konwolucji. Okazuje się, że są także atraktorami w przestrzeni rozkładów, oznacza to, że jakikolwiek rozkład spleciony ze sobą wiele razy dąży do rozkładu stabilnego. Innymi słowy rozkład graniczny sumy wielu zmiennych losowych jest rozkładem stabilnym. Powyższe stwierdzenie stanowi treść centralnego twierdzenia granicznego (CTG).

Klasyczne sformułowanie centralnego twierdzenia granicznego dotyczy sumy niezależnych zmiennych losowych o tym samym rozkładzie, który posiada skończoną wariancję σ^2 oraz wartość oczekiwaną m. Wtedy rozkład sumy zmieża do rozkładu Gaussa :

$$\lim_{N\to\infty} P\left(u_1 \le \frac{x - mN}{\sigma\sqrt{N}} \le u_2\right) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Dla wszystkich skończonych u_1 oraz u_2. Warto zwrócić uwagę, że dla skończnej wartości N rozkład sumy na ogonach może znacznie odbiegać od przewidywanego rozkładu Gaussa, jednakże waga tych regionów ekstremalnych zmierza do zera gdy N zmierza do nieskończoności.

Przedyskutujmy podstawowe założenia zapewniające słuszność Gaussowskiego CTG. Zmienne ∂X_i muszą być niezależne, lub przynajmniej dostatecznie słabo skorelowane. Funkcja korelacji $\langle \partial x_i \partial x_j \rangle - m^2$ musi maleć dostatecznie szybko dla rosnących $|i-j|$. W przypadku gdy wszystkie zmienne losowe są idealnie skorelowane (wszystkie są równe), rozkład sumy pokrywa się z rozkładem pojedyńczego ∂X_i po uwzględnieniu czynnika N. Zmienne losowe nie muszą również pochodzić z identycznego rozkładu, jednak wymagane jest aby wariancje zmiennych nie były zbyt odmienne tak aby żadna wariancja nie zdominowała reszty. W takim wypadku wariancja Gaussowskiego rozkładu granicznego jest średnią z wariancji indywidualnych rozkładów. Formalnie twierdzenie jest słuszne dla nieskończonego N. W praktyce N musi być wystarczająco duże aby Gauss mógł dobrze aproksymować rozklad sumy zmiennych. Minimalna wartość N jest zależna od rozkładu elementarnego $P_1(\partial x_1)$ i jego odmienności od rozkładu Gaussa. Minimalna wartość N zależy również od pytania jak daleko na ogonach chcemy aby rozkład sumy był dobrze aproksymowany przez rozkład Gaussa. Zwróćmy uwagę na fakt, że CTG nie mówi nic o ogonach wynikowego rozkładu, tylko centralna część rozkładu jest dobrze opisana rozkładem Gaussa. Szerokość regionu centralnego dobrze aproksymowanego rozkładem Gaussa dla dużych, skończonych N zależy istotnie od rozkładu elementarnego $P_1(\partial x_1)$. Region ten dla „wąskich" rozkładów symetrycznych o wszystkich momentach skończonych jest rzędu

$\sim N^{3/4}\sigma$. Jednak może on być znacznie węższy np. gdy rozkład elementarny ma ogony potęgowe (takie aby σ było skończone), region Gaussowski ma wtedy szerokość rzędu $\sim\sqrt{Nln(N)}$.

Rozpatrzmy teraz przypadek sumy dużej liczby N niezależnych zmiennych losowych o takim samym rozkładzie z asymptotycznymi ogonami potęgowymi :

$$P(x) \sim \frac{\mu A_{\mp}^{\mu}}{|x|^{1+\mu}}, \qquad x \to \mp\infty$$

o parametrze $\mu < 2$ oraz amplitudach ogonowych A^{μ} =$A_{+}^{\mu} = A_{-}^{\mu}$. Wariancja takiego rozkładu jest nieskończona. Dla takiego przypadku rozkładem granicznym przy dużym N jest stabilny rozkład Lévy'ego o eksponencie μ oraz amplitudzie NA^{μ} na ogonach. Jeśli amplitudy ogonów dodatniego i ujemnego są różnej wartości to otrzymujemy asymetryczny rozkład Lévy'ego. Jeśli natomiast ogony potęgowe z lewej i prawej strony rozkładu mają inny parametr $\mu_{-} \neq \mu_{+}$, wtedy wygrywa mniejszy z nich i w efekcie końcowym otrzymujemy całkowicie asymetryczny rozkład Lévy'ego o eksponencie $\mu = \min(\mu_{-}, \mu_{+})$. Uwagi dotyczące działania aproksymacji Gaussowskiego CTG także dotyczą zgeneralizowanego CTG Lévy'ego. Istotną różnicą pomiędzy Gaussowskim CTG a Lévy'ego jest rząd wielkości poszczególnych przyczynków do całej sumy N zmiennych losowych. W przypadku Gaussowskim rząd wielkości sumy (dla granicy dużych N) jest $\sqrt{N}$ razy większy od największego przyczynku z N liczb losowych, zatem żaden ze składników sumy nie przyczynia się do skończonej części sumy. Przeciwnie jest w przypadku zbieżności Lévy'ego przy parametrze $\mu < 2$, cała suma oraz największy przycznek są tego samego rzędu wielkości $N^{1/\mu}$. Największa wartość przyczynia się do skończonej części całej sumy nawet przy $N \to \infty$. Uwidocznia się tutaj istotna cecha rozkładów o ogonach potęgowych dla których waga zdarzeń ekstremalnych przy dużych N nie maleje do zera.

Suma zmiennych losowych nie zawsze dąży do rozkładu Gaussa. Wtedy centralne twierdzenie graniczne uogulnia się na rodzinę rozkładów Lévy'ego. Rozkład sumy zachowuje część informacji o rozkładzie elementarnym na ogonach. Suma ta fizycznie reprezentować może rozwój fluktującej ceny danego aktywa lub portfelu. Istotna jest zatem informacja o Gaussowskiej lub nie-Gaussowskiej naturze tej sumy z punktu widzenia kontrolowania ryzyka, ponieważ ogony potęgowe odpowiadają za najniebezpieczniejsze fluktuacje.

Literatura :

R. Mantegna, H. E. Stanley, *Ekonofizyka Wprowadzenie*, PWN, 2001.

J.-P. Bouchard, M. Potters, *Theory of Financial Risk and Derivative Pricing – From Statistical Physics to Risk Management*, Second Edition, Cambridge University Press.