PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 7 (1207) 2008

TAKSONOMIA 15
Klasyfikacja i analiza danych – teoria i zastosowania

Katarzyna Kuziak

Uniwersytet Ekonomiczny we Wrocławiu

MODELOWANIE ROZKŁADU STRAT W POMIARZE RYZYKA OPERACYJNEGO

1. Ryzyko operacyjne

Nie jest łatwo podać jedną definicję ryzyka operacyjnego. W tej pracy[1] przyjmiemy, że ryzyko operacyjne to możliwość poniesienia strat wynikających z błędów ludzkich (złe zarządzanie, brak nadzoru, brak kontroli) i nieprawidłowo działających systemów wewnętrznych podmiotu (technologii). Ryzyko operacyjne nie jest zjawiskiem nowym, nowe jest jedynie rozpatrywanie tego ryzyka w wymiarze ilościowym, jakim jest prezentowana metoda rozkładu strat (*loss distribution approach* – LDA).

Komitet Bazylejski ds. Nadzoru Bankowego zaleca w zakresie pomiaru ryzyka operacyjnego stosowanie trzech podejść (por. [Basle Committee... 2003]): wskaźnika podstawowego (*basic indicator approach* – BIA), metodą standardową (*standardized approach* – SA) i metodą zaawansowaną (*advanced approach: internal measurement approach, loss distribution approach*). Jednym z podejść zaawansowanych w pomiarze ryzyka operacyjnego jest modelowanie stochastyczne, w którym estymacja parametrów rozkładów częstości i rozkładów wielkości strat może być dokonana za pomocą estymacji parametrycznej (LDA).

Celem niniejszego artykułu jest przedstawienie podejścia LDA do modelowania strat z tytułu ryzyka operacyjnego, a następnie kalkulacja miary *capital-at-risk* (obciążenie kapitału ryzykiem operacyjnym). Wskazane zostaną problemy pojawiające się w analizie LDA. Podejście to zostanie zilustrowane hipotetycznym przykładem numerycznym (z uwagi na problem z dostępem do danych dotyczących strat z tytułu zdarzeń operacyjnych).

Ryzyko operacyjne ze względu na źródła powstawania nie jest jednolite – jest zbiorem różnego rodzaju czynników. Ta różnorodność utrudnia stosowanie jedno-

---

[1] Por. definicje w: [Basle Committee... 2003; Jajuga, Jakóbczak 2004; Jakóbczak 2003].

litej metodologii w procesie zarządzania ryzykiem, szczególnie na etapie pomiaru i sterowania. W analizie ryzyka operacyjnego można wyróżnić dwa rodzaje zdarzeń:

– rzadkie zdarzenia, ale o dużej skali (*low frequency, high severity*);
– zdarzenia pojawiające się często, ale na niewielką skalę (*high frequency, low severity*) – łatwiej je modelować, zazwyczaj nie prowadzą od razu do upadłości przedsiębiorstwa.

Model pomiaru ryzyka operacyjnego musi uwzględniać oba rodzaje zdarzeń. Zdarzenia te powinny być modelowane oddzielnie. Modelowanie tego rodzaju ryzyka, w porównaniu z innymi rodzajami, napotyka w praktyce następujące problemy:

– ograniczone informacje nt. strat z tytułu ryzyka operacyjnego (problem dostępności danych). Zwykle dysponujemy niewystarczającą liczbą dużych strat do estymacji parametrów ogona rozkładu;
– inne są modele, niż te wykorzystywane w analizie ryzyka rynkowego,
– rozkład danych empirycznych nt. strat z tytułu ryzyka operacyjnego charakteryzuje się grubymi ogonami (w większości przypadków niewielkie straty oraz kilka dużych strat).

W związku z tym w modelowaniu stosowane są techniki wykorzystywane w ubezpieczeniach (metody aktuarialne) i modelowaniu ryzyka kredytowego, a w przypadku braku odpowiednich danych wykorzystywana jest analiza scenariuszy. Analiza scenariuszy jest przeprowadzana np. w sytuacji, kiedy pomiot nie dysponuje wewnętrznymi (tj. pochodzącymi z wnętrza firmy) informacjami dotyczącymi strat z tytułu ryzyka operacyjnego lub nie ma wystarczającej ich liczby. Wówczas generowane są scenariusze na podstawie danych pochodzących z zewnętrznych baz informacji[2].

W przypadku ryzyka operacyjnego można wyróżnić trzy rodzaje strat:

1) straty oczekiwane – mierzone jako iloczyn prawdopodobieństwa i wielkości straty,

2) straty nieoczekiwane – mierzone np. jako różnica między 99-procentowym kwantylem rozkładu strat (koncepcja *value at risk*) a wartością oczekiwaną,

3) straty spowodowane katastrofami – modelowanie zdarzeń rzadkich.

W przypadku LDA interesować nas będą straty nieoczekiwane.

2. *Loss distribution approach*

W podejściu modelowania stochastycznego estymacja parametrów rozkładów częstości i rozkładów dotkliwości (skali zdarzenia, wielkości) następuje osobno. Podejście LDA prowadzi do otrzymania zagregowanego rozkładu strat. W LDA zakładamy typ rozkładu dla pojedynczego zdarzenia. Dla rozkładu częstości w przypadku ryzyka operacyjnego może to być rozkład: Poissona, dwumianowy,

[2] Na przykład w przypadku banków mogą to być dane z the Operational Riskdata eXchange Association.

ujemny dwumianowy, geometryczny (por. [Johnson i in. 1993; Klugman i in. 2004]), natomiast w przypadku dotkliwości rozkład Weibulla, Pareto, wartości ekstremalnych, lognormalny (ewentualnie normalny) lub mieszanka rozkładów. Kolejnym krokiem po estymacji parametrów rozkładów częstości i wielkości jest testowanie parametrów funkcji gęstości (test Kołmogorowa-Smirnova, test Andersona-Darlinga). Do estymacji parametrów rozkładu wykorzystywane są głównie metoda momentów i funkcja największej wiarygodności. Warto podkreślić, że w podejściu LDA w odniesieniu do ryzyka operacyjnego, jak wskazują analizy teoretyczne [Boecker, Klueppelberg 2005], wybór typu rozkładu dotkliwości ma mniejszy wpływ na kapitał ekonomiczny niż wybór rozkładu częstości. Kapitał ekonomiczny obliczany jest jako odpowiedni kwantyl zagregowanego rozkładu strat minus straty oczekiwane. Następnie następuje agregacja strat wynikających z pojedynczych zdarzeń rozpatrywanych dotąd osobno.

Tak więc w metodzie LDA można wyróżnić następujące etapy:

1. Dopasowanie częstości zdarzeń odpowiednim rozkładem.
2. Dopasowanie dotkliwości (wielkości) strat odpowiednim rozkładem.
3. Wyznaczenie rozkładu zaagregowanego strat za pomocą symulacji Monte Carlo.
4. Kalkulacja wymogu kapitałowego jako sumy wymogów wyznaczonych dla każdej kombinacji jednostki biznesowej i typu ryzyka.

Metoda LDA zilustrowana zostanie hipotetycznym przykładem numerycznym (por. [Frachot i in. 2001]). Załóżmy, że w przedsiębiorstwie identyfikujemy i-tą linię biznesową i j-ty rodzaj zdarzenia ryzyka operacyjnego (indeks i oznacza linie, indeks j rodzaj zdarzenia). $\zeta(i, j)$ oznacza zmienną losową, która reprezentuje wielkość jednego zdarzenia straty dla linii biznesowej i oraz rodzaju zdarzenia j. Przyjmijmy w naszym przykładzie, że mamy $i = 2$ – dwie linie produkcyjne oraz $j = 3$ – takie zdarzenia, jak np. kradzież, oszustwo, awaria systemu informatycznego.

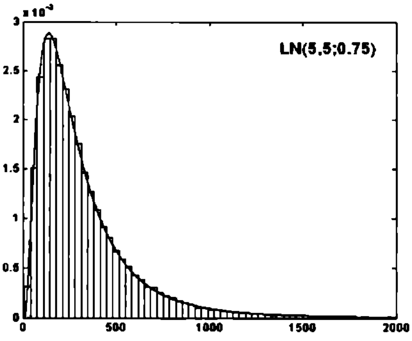
Zakładamy, że liczba zdarzeń między t a $t + \tau$ jest losowa (gdzie τ oznacza przedział czasu, w praktyce często przyjmuje się rok). Odpowiednia zmienna $N(i, j)$ ma funkcję prawdopodobieństwa $p_{i,j}$. Wówczas rozkład częstości strat $P_{i,j}$ dany jest jako:

$$P_{i,j}(n) = \sum_{k=0}^{n} p_{i,j}(k).$$

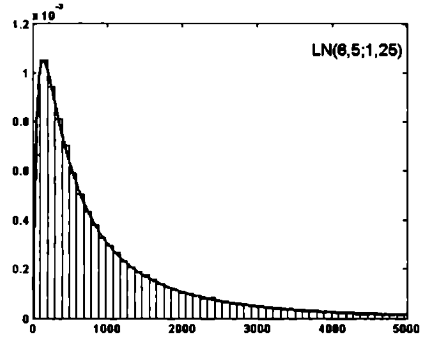
Rozkład częstości strat jest często przybliżany rozkładem Poissona. W przykładzie przyjmiemy rozkład dwumianowy (mamy dwa stany: zdarzenie generujące straty albo zachodzi, albo nie). Z kolei rozkład dotkliwości strat $\zeta(i, j)$ jest oznaczony przez $F_{i;j}$. Przyjęte typy rozkładów dotkliwości strat są następujące:

- rozkład lognormalny: rys. 1 (z parametrami 5, 5 i 0,75) oraz rys. 2 (z parametrami 6, 5 i 1,25),
- rozkład wykładniczy: rys. 3 (z parametrem 55) i rys. 4 (z parametrem 90),

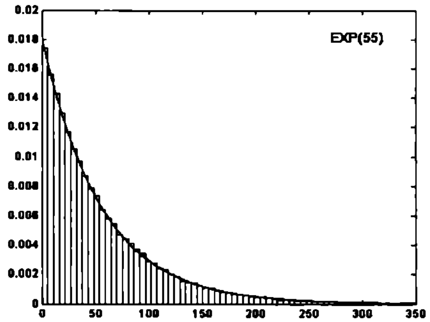
- rozkład normalny: rys. 5 (z parametrami 750 i 150) oraz rys. 6 (z parametrami 400 i 100).



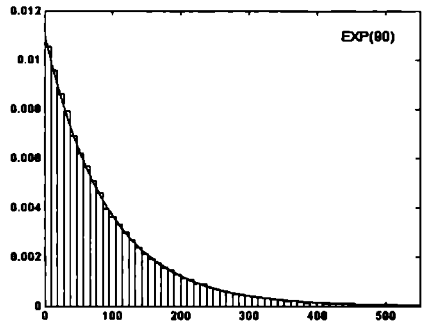
Rys. 1. Rozkład LN(5,5;0,75)
Źródło: obliczenia własne.



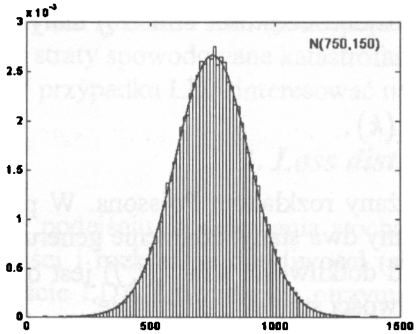
Rys. 2. Rozkład LN(6,5;1,25)
Źródło: obliczenia własne.



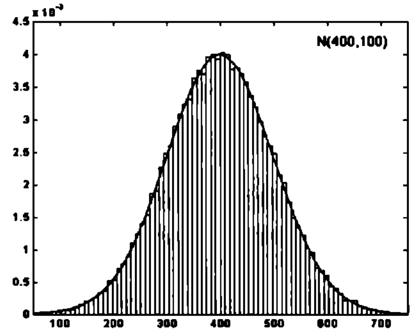
Rys. 3. Rozkład Exp(55)
Źródło: obliczenia własne.



Rys. 4. Rozkład Exp(90)
Źródło: obliczenia własne.



Rys. 5. Rozkład N(750,150)
Źródło: obliczenia własne.



Rys. 6. Rozkład N(400,100)
Źródło: obliczenia własne.

Symulacyjnie otrzymane rozkłady dla zadanych parametrów zapewniają warunki niezależności, w rzeczywistości należałoby tę niezależność testować.

W podejściu LDA strata dla linii biznesowej i oraz zdarzenia j między t a $t+\tau$ jest dana jako:

$$\vartheta(i,j)=\sum_{n=0}^{N(i,j)}\zeta_n(i,j).$$

Niech $G_{i,j}$ będzie rozkładem strat $\vartheta(i,j)$. $G_{i,j}$ jest złożonym rozkładem następującej postaci:

$$G_{i,j}(x)=\begin{cases}\sum_{n=1}^{\infty}p_{ij}(n)F_{i,j}^{n*}(x) & x>0\\ p_{i,j}(0) & x=0\end{cases},$$

gdzie * jest operatorem splotu rozkładów i F^{n*} jest n-krotnym splotem[3] F, który można określić następująco:

$$F^{1*}=F,$$

$$F^{n*}=F^{(n-1)*}*F.$$

W naszym przykładzie będziemy agregować rozkład częstości $N(i, j)$ (przyjęliśmy dwumianowy) oraz rozkład dotkliwości strat $\zeta(i, j)$ (przyjęliśmy lognormalny, wykładniczy i normalny). Przyjęliśmy również założenie, że zmienne $\zeta(i, j)$ mają niezależne rozkłady i są niezależne od liczby zdarzeń. W większości przypadków nie ma analitycznej postaci rozkładu zagregowanego. Konieczne jest często stosowanie algorytmów numerycznych. W przykładzie wykorzystamy symulację Monte Carlo (10 000 powtórzeń), w której rozkład $G_{i,j}$ jest przybliżany przez zbiór:

$$S\langle\vartheta(i,j)\rangle=\{\vartheta_S(i,j),s=1,\ \ldots,\ S\}$$

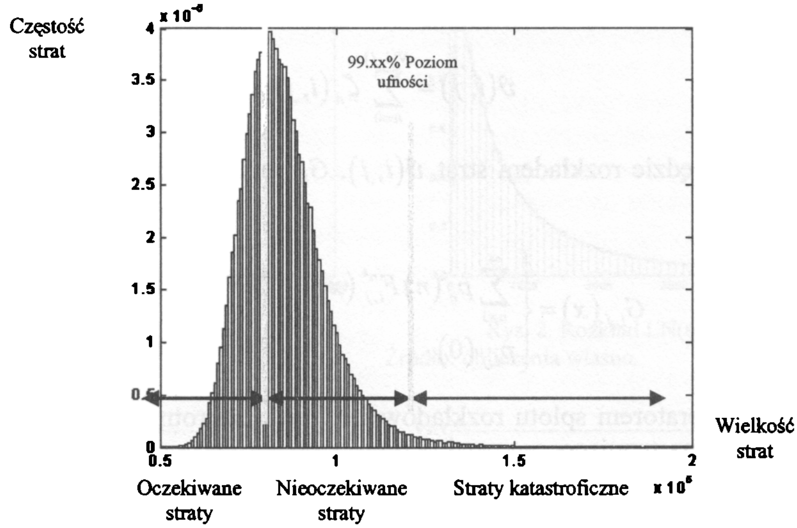
otrzymanych wielkości $\vartheta(i,j)$. Oszacowanie $G_{i,j}$ jest wówczas otrzymywane przez empiryczny rozkład $S\langle\vartheta(i,j)\rangle$. Uzyskany w wyniku symulacji zagregowa-

[3] Splot funkcji φ z funkcją prawdopodobieństwa F jest funkcją zdefiniowaną przez:

$$u(x)=\int_{-\infty}^{+\infty}\varphi(x-y)Fd(y).$$

Można zaznaczyć, że $u=F*\varphi$, jeśli F ma funkcję gęstości f, co można zapisać odpowiednio jako $u=f*\varphi$, oraz $u(x)=\int_{-\infty}^{+\infty}\varphi(x-y)f(y)d(y)$.

ny rozkład strat dla zadanych rozkładów częstości i dotkliwości przedstawiony jest na rys. 7.



Rys. 7. Zagregowany rozkład strat

Źródło: obliczenia własne.

Jeśli straty $\vartheta(i,j)$ są zależne, analizowanie rozkładu zagregowanego wykorzystuje funkcje połączeń[4] (w przypadku ryzyka operacyjnego może być: t-*student copula*, Gumbel *copula*). Wykorzystanie funkcji połączeń umożliwia rozwiązanie problemu dotyczącego rozkładów wielowymiarowych, jakim jest nieznajomość postaci analitycznej empirycznego rozkładu łącznego. Parametryzacja funkcji połączeń jest przeprowadzana na podstawie danych historycznych również za pomocą symulacji [Embrechts i in. 2001]). Na początku szacujemy parametry rozkładów brzegowych dystrybuanty empirycznej (bez założenia o parametryzacji formy każdej z nich) następnie szacujemy za pomocą np. MLE (*maximum likehood estimation*) parametry funkcji połączeń.

Ponieważ nie zawsze zagregowany rozkład strat ϑ ma rozwiązanie analityczne, stosowane są procedury numeryczne, takie jak symulacja Monte Carlo, metoda rekurencyjna Panjera [1981], odwrócenie funkcji charakterystycznej. Postać analityczna zależy od własności rozkładu $G_{i,j}$, np. w przypadku rozkładu ujemnego dwumianowego rozkład zagregowany może być zagregowanym rozkładem Poissona lub logarytmicznym [Feller 1968].

[4] Nieznany n-wymiarowy rozkład łączny może zostać przybliżony funkcją połączeń oraz odpowiednimi rozkładami brzegowymi (twierdzenie Sklara). Problemem jest znalezienie odpowiednio dobrze dopasowanej funkcji połączeń. W praktyce problem ten rozwiązuje się, dopasowując do danych empirycznych kilka wybranych funkcji i następnie, stosując np. kryterium informacyjne Akaike, wykorzystuje się do dalszej analizy najlepiej dopasowaną.

Jeśli uzyskamy zagregowany rozkład strat z tytułu ryzyka operacyjnego (jak na rys. 7), wystarczy jako wymóg kapitałowy – obciążenie kapitału ryzykiem operacyjnym – tzw. CaR (*Capital-at-Risk*) wskazać odpowiedni kwantyl rozkładu. Całkowita strata jest równa sumie:

$$\vartheta = \sum_{i=1}^{I} \sum_{j=1}^{J} \vartheta(i,j).$$

Jeśli przyjmiemy, że straty $\vartheta(i,j)$ są niezależne, rozkład G będzie następującym splotem rozkładów $G_{i,j}$:

$$G(x) = \mathop{*}_{i=1}^{I} \mathop{*}_{j=1}^{J} G_{i,j}(x).$$

Wymóg kapitałowy stanowi wartość kapitału niezbędnego do pokrycia nieoczekiwanych strat (obliczony na ustalonym poziomie ufności) – CaR [Marshall 2001], która dla danej kombinacji jednostki biznesowej i typu ryzyka jest odpowiednim kwantylem tego rozkładu:

$$CaR(\alpha) = G^{-1}(\alpha).$$

Natomiast całkowity wymóg kapitałowy (na poziomie podmiotu) ustalony jest jako suma wymogów wyznaczonych dla każdej kombinacji jednostki biznesowej i typu ryzyka.

Następnie przeprowadza się testowanie wsteczne przyjętego modelu. Testowanie wsteczne w przypadku ryzyka operacyjnego jest skomplikowane z powodu ograniczonej dostępności danych, ale trzeba weryfikować precyzję przyjętego modelu. Na koniec wyznacza się odpowiedni kwantyl rozkładu strat (dla zadanego poziomu). W tym przykładzie wielkości wybranych kwantyli dla zagregowanego rozkładu strat zestawione są w tab. 1. Banki wyznaczają kwantyle na poziomie 99,75, 99,9, 99,95 i 99,97% (w zaleceniach Komitetu Bazylejskiego jest 99%).

Tabela 1. Wielkości kwantyli

Percentyl	Wielkość straty CaR
50	83 493,0081
75	91 359,8207
80	93 598,6527
90	100 321,3938
95	107 105,6813
97,5	114 253,6749
99	125 386,0648
Max	457 631,9083

Źródło: obliczenia własne.

W przykładzie kwantyl na poziomie 0,99 oznacza, że z prawdopodobieństwem 0,99 straty z tytułu ryzyka operacyjnego nie powinny przekroczyć 125 386 j.p. w ustalonym horyzoncie.

3. Podsumowanie

Zagadnienia związane z modelowaniem strat z tytułu ryzyka operacyjnego, przedstawione na hipotetycznym przykładzie, można ująć w następujących punktach:

1. Jakość danych dotyczących strat z tytułu ryzyka operacyjnego jest niska (często występuje sytuacja braku danych wewnętrzych i konieczność korzystania z zewnętrznych baz danych).
2. Ryzyko operacyjne jest trudniej zmierzyć niż ryzyko rynkowe czy kredytowe, z powodu braku lub ograniczonej dostępności danych.
3. Straty z tytułu ryzyka operacyjnego jest bardzo trudno modelować (np. występują obserwacje nietypowe, zdarzenia rzadkie).
4. Problemem w LDA jest wybór prawidłowego rozkładu; może wystąpić sytuacja, w której żaden teoretyczny rozkład nie będzie dobrze dopasowany do danych.
5. Zaletą tego podejścia jest możliwość przeprowadzenia pomiaru ryzyka na poziomie 99,9%.
6. Podejście LDA pozwala podmiotowi, przy kalkulacji kapitału na pokrycie strat z tytułu ryzyka operacyjnego, na korzystanie z własnych baz danych i na ich podstawie – na szacowanie parametrów rozkładu. Jednocześnie pozwala na dużą swobodę wyboru w zakresie zastosowanych narzędzi – modeli szacujących rozkład dotkliwości i częstości strat.
7. LDA uwzględnia specyfikę danego podmiotu (umożliwia podmiotowi samodzielne wyspecyfikowanie jednostek biznesowych oraz typów ryzyka) i tym samym pozwala na dokładniejsze, w porównaniu z metodami wskaźnika podstawowego czy standardową, szacowanie wymogów kapitałowych w zakresie ekspozycji na ryzyko operacyjne.

Literatura

Basle Committee on Bank Supervision (2003), *Supervisory Guidance on Operational Risk Advanced Measurement Approaches for Regulatory Capital.*

Boecker K., Klueppelberg C. (2005), *Operational VAR: a Closed-form Approximation*, RISK, s. 90-93.

Ebnoeter S., Vanini P., McNeil A., Antolinez P. (2002), *Operational Risk: a Practitioner's View*, www.gloriamundi.org.

Embrechts P., Kaufmann R., Samorodnitsky G. (2002), *Ruin Theory Revisited: Stochastic Models for Operational Risk*, www.gloriamundi.org.

Embrechts P., Lindskog F., McNeil A. (2001), *Modelling Dependence with Copulas and Applications to Risk Management*, raport, ETHZ Zurich.
Feller W. (1968), *An Introduction to Probability Theory and Its Applications*, vol. I, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York.
Frachot A., Georges P., Roncalli T. (2001), *Loss Distribution Approach for Operational Risk*, gro.cre ditlyonnais.fr.
Jajuga K., Jakóbczak J. (2004), *Ryzyko operacyjne – nowe tendencje w zakresie pomiaru*, Materiały z konferencji „Innowacje na światowych rynkach finansowych", Warszawa, s. 47-56.
Jakóbczak J. (2003), *Ryzyko operacyjne czy ryzyko korporacyjne*, Materiały z konferencji „Innowacje na światowych rynkach finansowych", Warszawa.
Johnson N.L., Kotz S., Kemp A.W. (1993), *Univariate Discrete Distributions*, John Wiley & Sons, New York.
Klugman S., Panjer H.H., Willmot G. (2004), *Loss Models – from Data to Decisions*, John Wiley & Sons, New Jersey.
Leippold M., Vanini P. (2005), *The Quantification of Operational Risk*, „Journal of Risk", 8(1), s. 59-85.
Marshall C. (2001), *Measuring and Managing Operational Risk in Financial Institutions, Tools, Techniques and other Resources*, John Wiley & Sons, Singapore.
Panjer H.H. (1981), *Recursive Evaluation of Compound Distributions*, „Astin Bulletin" 12, s. 22-26.

LOSS DISTRIBUTION APPROACH FOR OPERATIONAL RISK MEASUREMENT

Summary

In the operational risk measurement we use stochastic modelling approach, where the parameter estimation of probability distribution for the frequency of events and for the severity is made by empirical estimation or parametric estimation (e.g. loss distribution approach, LDA). In this paper the loss distribution approach is introduced. In LDA one assumes the type of distribution for a single event. For frequency distribution Poisson distribution and for severity distribution normal distribution could be used as well. For the estimation methods moments and maximum likelihood methods are generally chosen. Then one obtains the aggregate loss distribution by compounding loss severity and loss frequency distribution. This approach is illustrated by an example.