Kvantifisering av operasjonell risiko basert på kombinering av hendelsesdata og subjektive risikovurderinger

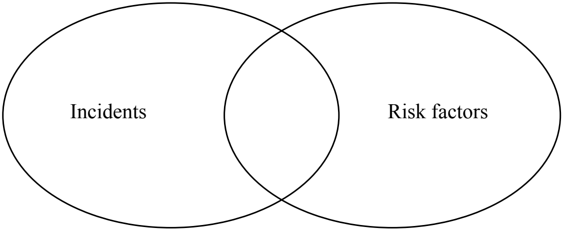
Arne Bang Huseby [1] and Jan Thomsen [2]

[1]University of Oslo, Norway

[2]Norges Bank Investment Management, Norway

ESRA-seminar, Universitet i Oslo, 3. februar 2016

Kombinering av hendelsesdata and subjektive risikovurderinger



- HENDELSER: Observasjoner av hendelser som har inntruffet med tilhørende konsekvenser (f.eks. økonomisk tap)
- RISIKO FAKTORER: Et ekspertpanel identifiserer potensielle risiko faktorer. For hver risikofaktor anslås *hendelsesfrekvens* og *konsekvensfordeling*.

Hovedutfordringer

- Bare nyere hendelser kan betraktes som relevante når man skal vurdere mulige fremtidige hendelser. Dette betyr at vi har relativt begrenset med data.
- Hovedpoenget med å inkludere de subjektivt identifiserte risikofaktorene er å ta hensyn til potensielle risikohendelser som ennå ikke har inntruffet.
- Det vil typisk være overlapp mellom type hendelser som har inntruffet, og de identifiserte risikofaktorene.
- Hendelsesdataene er i en form som gjør det umulig å sammenholde disse direkte med de subjektivt identifiserte risikofaktorene. Dette betyr at graden av overlapp er usikker.
- Det er erfaringsvis store forskjeller mellom de observerte hendelsene og de identifiserte risikofaktorene.

Sammensatte Poisson-prosesser

$$N(t) = \text{Antall hendelser i intervallet } [0, t)$$

$$X_i = \text{Konsekvensen av } i\text{te hendelse}$$

$$Z(t) = \sum_{i=1}^{N(t)} X_i$$

Hovedantagelser

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \ldots$$

$X_1, X_2, \ldots$ er uavhengige og identisk fordelte stokastiske variable med felles kumulativ sannsynlighetsfordeling F.

Summer av uavhengige sammensatte Poisson-prosesser

$$\text{Antall hendelser: } N_s(t) \sim Po(\lambda_s t), \qquad s = 1, \ldots, r$$

$$\text{Konsekvenser: } X_{s1}, X_{s2}, \ldots \sim F_s, \qquad s = 1, \ldots, r$$

$$Z_s(t) = X_{s1} + \cdots + X_{s,N_s(t)}, \qquad s = 1, \ldots, r$$

$$Z(t) = Z_1(t) + \cdots + Z_r(t)$$

Da er $Z(t)$ en sammensatt Poisson-prosess. Antall hendelser er $Po(\lambda t)$-fordelt, der $\lambda = \lambda_1 + \cdots + \lambda_r$, og konsekvensfordeling:

$$F(x) = \sum_{s=1}^{r} \frac{\lambda_s}{\lambda} F_s(x),$$

dvs. en blanding av konsekvensfordelingene for $Z_1(t), \ldots, Z_r(t)$.

MC simulering av uavhengige sammensatte Poisson-processer

METODE 1

STEP 1. Trekk $N(t) \sim \mathrm{Po}(\lambda t)$

STEP 2. Trekk $X_1, \ldots, X_{N(t)} \sim F(x) = \sum_{s=1}^{r} \frac{\lambda_s}{\lambda} F_s(x)$

$$Z(t) = \sum_{i=1}^{N(t)} X_i$$

METODE 2

STEP 1. Trekk $N_s(t) \sim \mathrm{Po}(\lambda_s t)$, $s = 1, \ldots, r$.

STEP 2. Trekk $X_{s1}, \ldots, X_{s,N_s(t)} \sim F_s(x)$, $s = 1, \ldots, r$

$$Z(t) = \sum_{s=1}^{r} Z_s(t) = \sum_{s=1}^{r} \sum_{i=1}^{N_s(t)} X_{si}$$

Usikkerhet knyttet til hendelsesintensiteter

Apriorifordeling for λ

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0.$$

Aposteriorifordeling for λ

$$\pi(\lambda|\tau,\nu) = \frac{(\beta+\tau)^{\alpha+\nu}}{\Gamma(\alpha+\nu)} \lambda^{\alpha+\nu-1} e^{-(\beta+\tau)\lambda}, \quad \lambda > 0.$$

der τ er lengden på intervallet prosessen har blitt observert i og ν er antall observerte hendelser i dette intervallet.

BEMERK: Parameteren β kan fortolkes som et mål for *styrken* i den apriori innsikten. Mer presist: den apriori innsikten er sammenlignbar med å ha observert prosessen over et intervall av lengde β.

Usikkerhet knyttet til konsekvensfordelingen

La F betegne den *sanne* konsekvensfordelingen for en gitt hendelse, la $Y_1, \ldots, Y_\nu$ betegne de tilgjengelige data, og innfør:

$$S_\nu(x) = \sum_{j=1}^{\nu} \mathrm{I}(Y_j \leq x) = \nu \hat{F}_\nu(x).$$

Da følger det at:

$$S_\nu(x) | F(x) \sim \mathrm{Bin}(\nu, F(x)).$$

For hver $x > 0$ har vi også et apriori estimat for $F(x)$, betegnet $\phi(x)$. For å kombinere dette med data, må vi utvide dette til en full apriorifordeling:

$$F(x) \sim \mathrm{Beta}(c\phi(x), c(1 - \phi(x))), \qquad x > 0,$$

der $c > 0$ en passende valgt konstant.

Usikkerhet knyttet til konsekvensfordelingen (forts.)

Vi får da at:

$$\mathrm{E}[F(x)] = \phi(x),$$

$$\mathrm{Var}[F(x)] = \frac{\phi(x)(1 - \phi(x))}{c + 1}.$$

Bemerk at høye verdier for c medfører at $\mathrm{Var}[F(x)]$ er liten. Hvis dette er tilfelle, indikerer dette at den apriori innsikten er sterk.

Usikkerhet knyttet til konsekvensfordelingen (forts.)

Ved å benytte Bayes' teorem får vi dermed at:

$$F(x)|S_\nu(x) \sim \mathrm{Beta}(c\phi(x)+S_\nu(x), c(1-\phi(x))+\nu-S_\nu(x)), \qquad x > 0.$$

Forventningen i aposteriorifordelingen er da:

$$\mathrm{E}[F(x)|S_\nu(x)] = \frac{c\phi(x) + S_\nu(x)}{c+\nu} = \frac{c}{c+\nu}\phi(x) + \frac{\nu}{c+\nu}\hat{F}_\nu(x),$$

For enkelhets skyld, betegner vi $\mathrm{E}[F(x)|S_\nu(x)]$ med $\phi'(x)$, og bruker dette som vå oppdaterte konsekvensfordeling.

Usikkerhet knyttet til konsekvensfordelingen (forts.)

Apriori-estimatet for konsekvensfordelingen kan selv være en blandingsfordeling:

$$\phi(x) = \sum_{s=1}^{r} \frac{\lambda_s}{\lambda} \phi_s(x),$$

der $\lambda = \lambda_1 + \cdots + \lambda_r$.

Vi får dermed følgende aposteriori-estimat for konsekvensfordelingen:

$$\phi'(x) = \frac{c}{c+\nu} \sum_{s=1}^{r} \frac{\lambda_s}{\lambda} \phi_s(x) + \frac{\nu}{c+\nu} \hat{F}_\nu(x).$$

Simulering av s uavhengige sammensatte Poisson-prosesser med usikkerhet - APRIORI

I hver iterasjon i Monte Carlo simuleringen gjøres følgende:

STEP 1. Trekk λ_s fra $\text{Gamma}(\alpha_s, \beta_s)$, $s = 1, \ldots, r$.

STEP 2. Trekk $N_s(t)$ fra Po$(\lambda_s t)$, $s = 1, \ldots, r$.

STEP 3. Trekk $X_{s1}, \ldots, X_{s,N_s(t)}$ fra $\phi_s(x)$, $s = 1, \ldots, r$.

STEP 4. Beregn:

$$Z(t) = \sum_{s=1}^{r} \sum_{i=1}^{N_s(t)} X_{si}.$$

Simulering av s uavhengige sammensatte Poisson-prosesser med usikkerhet - APOSTERIORI

I hver iterasjon i Monte Carlo simuleringen gjøres følgende:

STEP 1. Trekk λ_s from $\mathrm{Gamma}(\alpha_s + \nu_s, \beta_s + \tau)$, $s = 1, \ldots, r$.

STEP 2. Beregn $\tilde{\lambda}_s = \frac{\lambda_s c}{(c+\nu_s)}$ og $\tilde{\mu}_s = \frac{\lambda_s \nu_s}{(c+\nu_s)}$, $s = 1, \ldots, r$.

STEP 2. Trekk $N_s(t)$ fra $\mathrm{Po}(\tilde{\lambda}_s t)$, $s = 1, \ldots, r$.

STEP 3. Trekk $M_s(t)$ fra $\mathrm{Po}(\tilde{\mu}_s t)$, $s = 1, \ldots, r$.

STEP 4. Trekk $X_{s1}, \ldots, X_{s,N_s(t)}$ fra $\phi_s(x)$, $s = 1, \ldots, r$.

STEP 5. Trekk $Y_{s1}, \ldots, Y_{s,M_s(t)}$ fra $\hat{F}_{\nu_s}(x)$, $s = 1, \ldots, r$.

STEP 6. Beregn:

$$Z(t) = \sum_{s=1}^{r} \Big(\sum_{i=1}^{N_s(t)} X_{si} + \sum_{i=1}^{M_s(t)} Y_{si} \Big).$$

Simulering av s uavhengige sammensatte Poisson-prosesser med usikkerhet (forts.)

Bemerk at metoden vi nettopp har beskrevet forutsetter at vi er i stand til å fordele de observerte hendelsene på de ulike risikofaktorene, slik at ν_s hendelser tilskrives $s$te risikofaktor.

I praksis vil det ikke være mulig å fordele de observerte hendelsene på denne måten. I stedet må man ta hensyn til at det er usikkerhet knyttet til en slik klassifisering. Løsningen på dette problemet blir for komplisert å beskrive her, så vi henviser til forskningsrapporten vår for en detaljert gjennomgang av denne situasjonen.

To separate delmodeller

Den totale simuleringsmodellen for kombinering av hendelsesdata og risikofaktorer består av to separate delmodeller:

DELMODELL 1: Den rene hendelsesdatamodellen
Denne modellen er ment å inkludere data for alle hendelser som ikke relateres til de subjektivt identifiserte risikofaktorene.

DELMODELL 2: Risikofaktormodellen
Denne modellen inkluderer de subjektivt identifiserte risikofaktorene, og oppdateres i lys av data som relateres til disse (dvs. de data som ikke er inkludert i den rene hendelsesdatamodellen).

Fordeling av hendelsesdata mellom den rene hendelsesmodellen

Vi antar at vi har observert ν hendelser med tilhørende konsekvenser: $W_1, \ldots, W_\nu$.

La så $R_i = 1$ dersom $i$te hendelse skal inkluderes i risikofaktormodellen, og 0 ellers, $i = 1, \ldots, \nu$.

$R_1, \ldots, R_\nu$ er stokastiske variable med fordeling:

$$P(R_i = 1) = 1 - e^{-\rho W_i}, \quad i = 1, \ldots, \nu$$

Vi innfører også følgende indeksmengder:

$$\Omega_R = \{i : R_i = 1\}, \qquad \Omega_I = \{i : R_i = 0\}$$

Ω_R er indeksmengden for de data som skal inkluderes i risikofaktormodellen, og Ω_I er indeksmengden for de data som skal inkluderes i den rene hendelsesdatamodellen.

Simulering av totalmodellen

I hver iterasjon i Monte Carlo simuleringen gjøres følgende:

STEP 1. Trekk R_i slik at $P(R_i = 1) = 1 - e^{-\rho W_i}$, $i = 1, \ldots, \nu$, and bestem indeksmengdene Ω_R and Ω_I.

STEP 2. Simuler risikofaktormodellen basert på observasjoner med indekser i Ω_R.

STEP 3. Simuler den rene hendelsesdatamodellen basert på observasjoner med indekser i Ω_I.

STEP 4. Beregn den totale konsekvensen ved å legge sammen resultatene fra begge modeller.

Numeriske eksempler

- Hendelsesdatabasen inneholder data fra en periode på $\tau = 5$ år med totalt $\nu = 460$ observerte hendelser.
- Risikofaktormodellen inneholder $r = 30$ risikofaktorer.
- Vi antar at $\lambda_s \sim \text{Gamma}(\alpha_s, \beta_R)$, $s = 1, \ldots, 30$.
- To tilfeller betraktes:
  $\beta_R = 0.2$ (svak apriori innsikt)
  $\beta_R = 1.0$ (moderat apriori innsikt).
- I sannsynlighetsfordelingen $P(R_i = 1) = 1 - e^{-\rho W_i}$ ser vi på tre tilfeller:
  $\rho = 0.005$ (nesten alle dataene benyttes i den rene hendelsesdatamodellen)
  $\rho = 0.5$ (dataene fordeles omtrent likt mellom de to modellene)
  $\rho = 50$ (nesten alle dataene benyttes i risikofaktormodellen).

Numeriske eksempler

Table: Mean and standard deviations for the case when $\beta_R = 0.2$.

		$\rho = 0.005$	$\rho = 0.5$	$\rho = 50$
Incidence model	Mean	172.07	32.72	0.05
	St.dev.	67.14	7.81	0.05
Risk factor model	Mean	40.47	174.43	207.1
	St.dev.	68.71	91.45	91.48
Combined model	Mean	212.54	207.15	207.17
	St.dev.	93.23	91.76	91.48

Simulerte resultater når $\beta_R = 0.2$

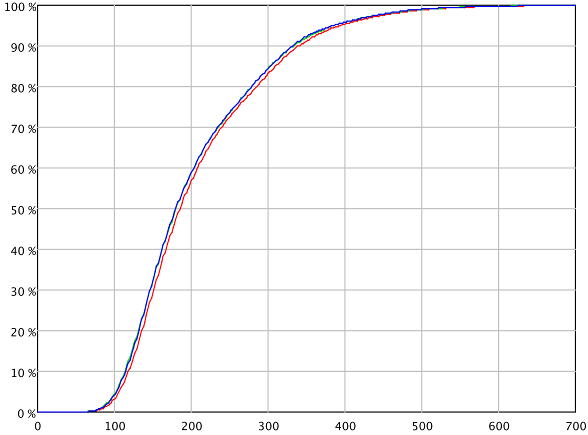


Figure: Estimated cumulative distributions for the accumulated consequences using $\rho = 0.005$ (red curve), $\rho = 0.5$ (green curve) and $\rho = 50$ (blue curve), $\beta_R = 0.2$.

Numeriske eksempler

Table: Mean and standard deviations for the three models for the case when $\beta_R = 1.0$

		$\rho = 0.005$	$\rho = 0.5$	$\rho = 50$
Incidence model	Mean	172.07	32.72	0.05
	St.dev.	67.14	7.81	0.05
Risk factor model	Mean	84.06	199.40	227.87
	St.dev.	92.67	105.86	105.86
Combined model	Mean	256.13	232.11	227.93
	St.dev.	112.23	106.10	105.86

Simulerte resultater når $\beta_R = 1.0$

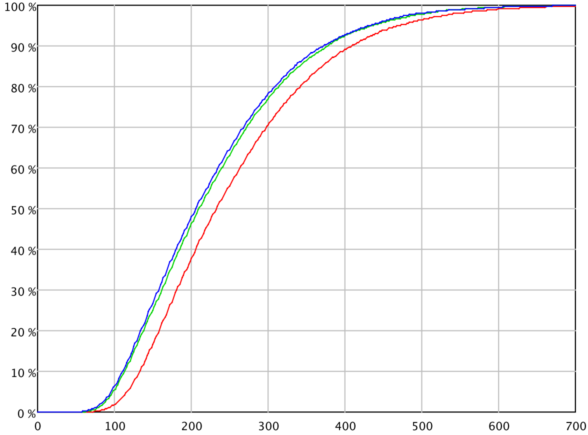


Figure: Estimated cumulative distributions for the accumulated consequences using $\rho = 0.005$ (red curve), $\rho = 0.5$ (green curve) and $\rho = 50$ (blue curve), $\beta_R = 1.0$.

Konklusjoner

- Vi har utviklet en modell for hvordan man kan kombinere hendelsesdata med subjektivt identifiserte risikofaktorer. Modellen består av to delmodeller: *den rene hendelsesdatamodellen* og *riskofaktormodellen*.
- Modellen kan håndtere usikkerhet knyttet til graden av overlapp mellom de to delmodellene.
- Modellen kan håndtere usikkerhet knyttet til hendelsesrater for de ulike risikofaktorene, og lar oss oppdatere denne usikkerheten basert på tilgjengelige observerte hendelser.
- Modellen kan håndtere usikkerhet knyttet til konsekvenser for de ulike risikofaktorene, og lar oss oppdatere denne usikkerheten basert på tilgjengelige observerte konsekvenser.