# ML Engineer’s look at insurance


## Peter Zvirinský

peter.zvirinsky@ceai.io

@zvirisk

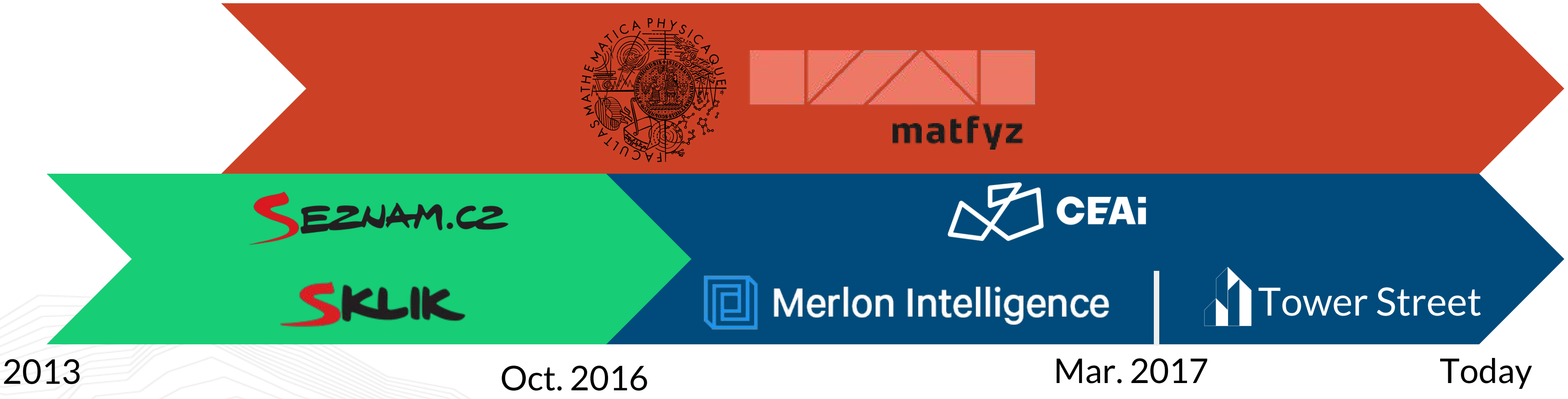

CEAi

# Disclaimer

- This presentation is more about problem setting than individual models.
- No fancy models will be presented, mostly old school statistics.
- Expect to see some formulas.
- I will not use the word **deep** during the entire presentation.

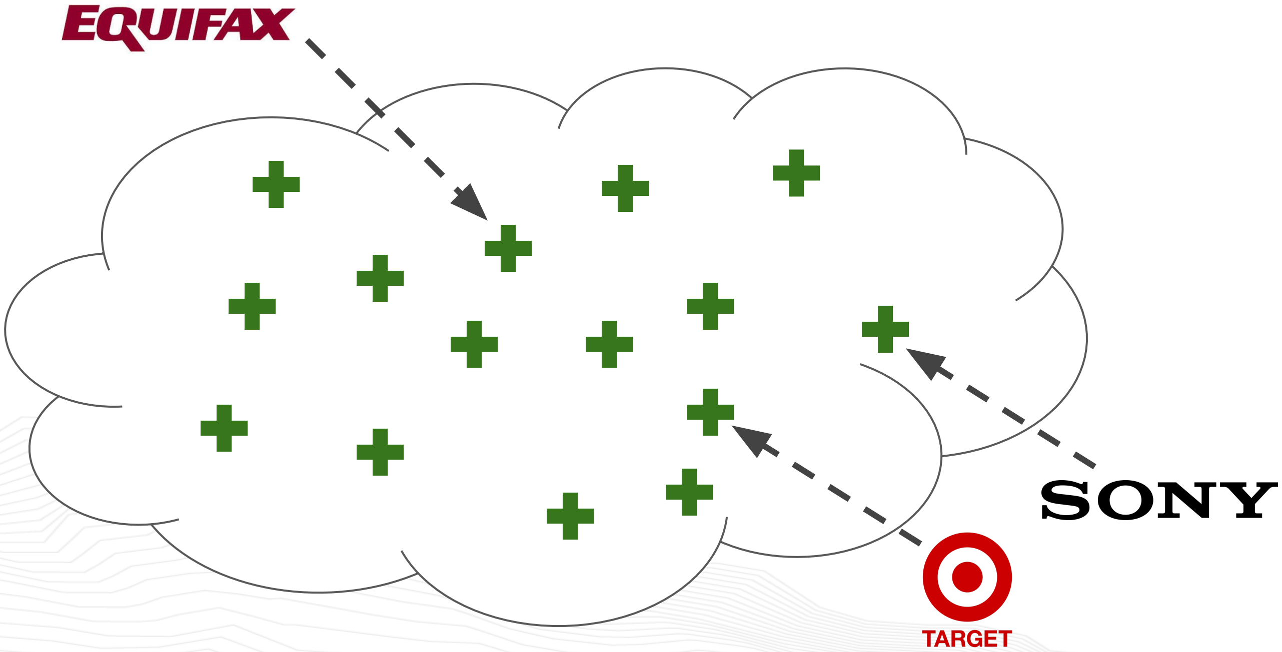
# Who am I?

- **ML Engineer** at CEAi
- **PhD student** at MFF CUNI
- Previously **ML Researcher** at Seznam.cz



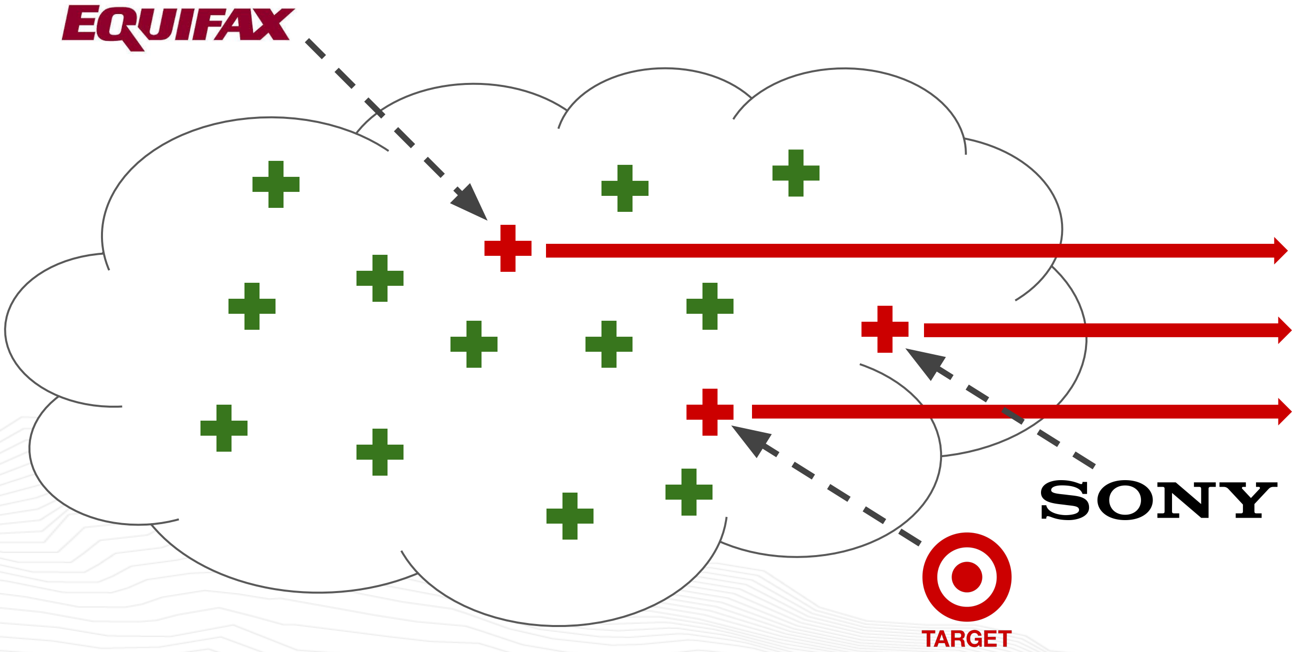
# The Problem

- Imagine you have a set of companies.



# The Problem

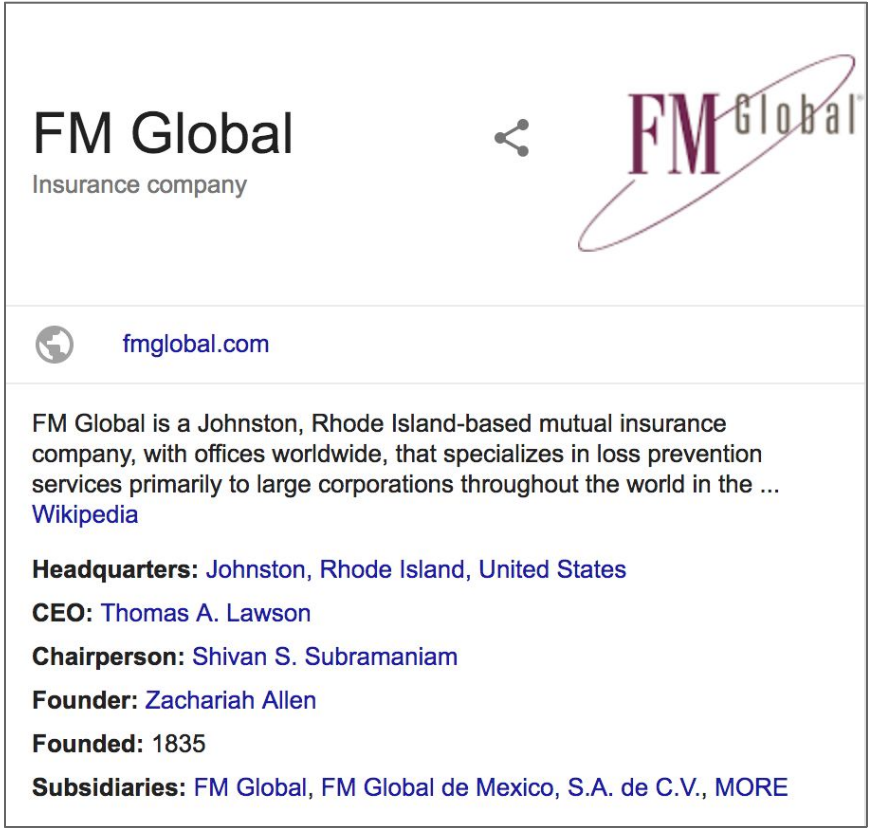
- Imagine you have a set of companies.
- Imagine some of them get breached and generate a **claim**.




| **Claims** | |
|---|---|
| Claim 1: | $1M |
| Claim 2: | $12M |
| Claim 3: | $8M |
| **Sum:** | $21M |

# Engineering solution

1. All companies agree to evenly contribute to a **$21M** (+ some buffer) “risk” fund.
2. If a **claim occurs**, it is payed out from this fund.
3. On the end of each coverage period, the remaining amount in the fund is refunded to all companies.

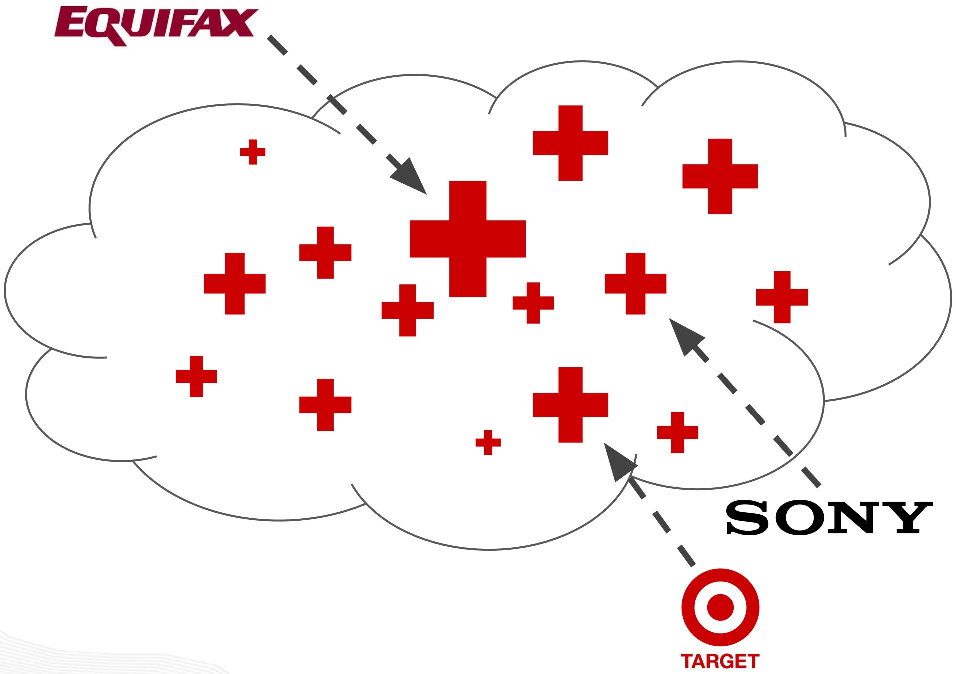


# Engineering solution - issues

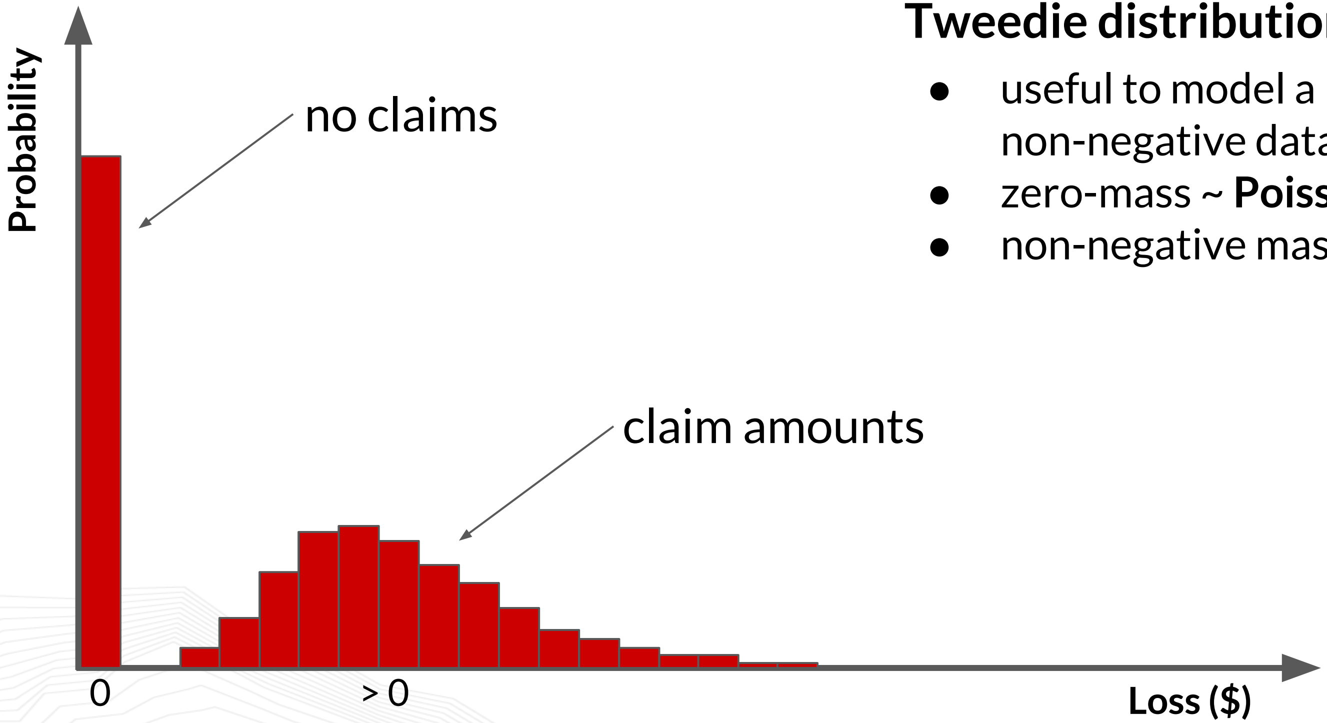
**Not all companies are born equal, i.e. different companies pose different risks.**

**Main issues:**

- unfair pricing:
  - low risk companies overpay
  - high risk companies underpay
- incremental portfolio rollout
- adverse selection



# Distribution of losses




**Tweedie distribution**

- useful to model a mixture of zeros and non-negative data point
- zero-mass ~ **Poisson(λ)**
- non-negative mass ~ **Gamma(α, θ)**

# Frequency-severity models

$$Pr(loss \mid company) = Pr(claim \mid company) \times Pr(claim\ size \mid claim, company)$$

**Frequency model**

Modeling as probability
or count response.

**Severity model**

Continuous response.

# Frequency model (1)
## Logistic regression

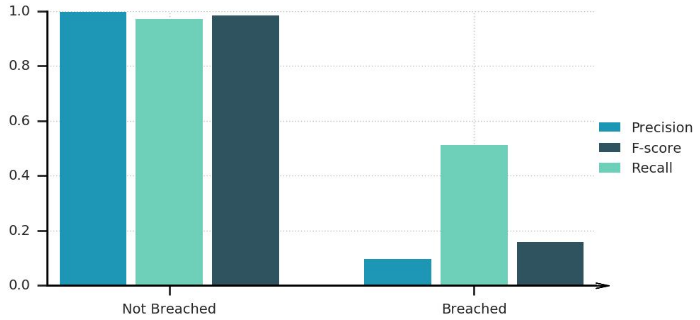
Formulating as claim probability prediction for a certain time period:

- two classes: claim (~**breach)** / no claim (**~ no breach)**

Training Logistic Regression for TowerStreet:

- using 200+ Financial features from:
  - Bureau van Dijk
  - KLD Stats
- regularizations:
  - L1 (~ Lasso) induces sparsity
    - implicit feature selection
  - L2 (~ Ridge) induces stronger shrinkage
    - prevents overfitting
  - L1 + L2 (~ Elastic Net)
    - linear combination of both to control trade-off

**Evaluation**




Most predictive features:

- Market Cap
- Total Assets
- Number of Branches
- Solvency Ratio

# Frequency model (2)

## Poisson regression

**Poisson regression** assumes the response variable has a Poisson distribution.

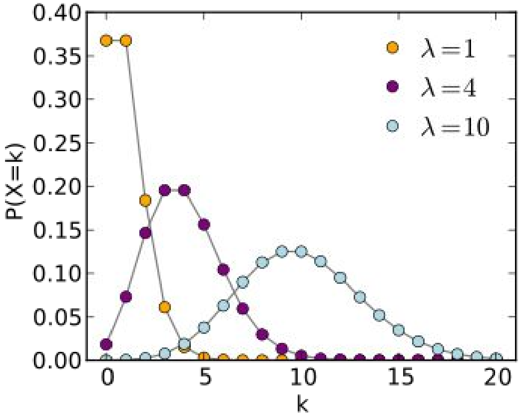
**Regression model:**

$$\lambda = exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k)$$

where:

- $x_1, x_2, \ldots, x_k$ are regressor variables,
- $\beta_0, \beta_1, \ldots, \beta_k$ are regression coefficients.

**Poisson distribution**




**PMF:** $\frac{\lambda^k e^{-\lambda}}{k!}$

**λ** ~ expected number of occurrences within a time period

# Frequency model (3)

## Poisson regression - how do you fit it?

**FREQUENTISTS**

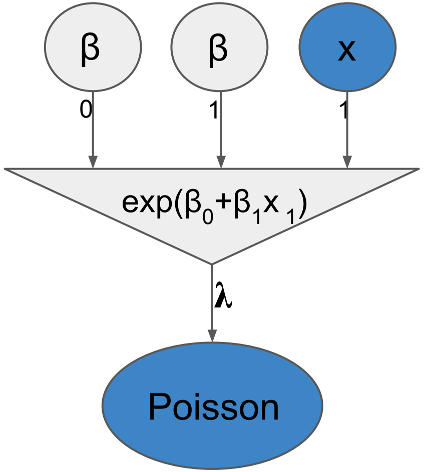
1. derive **maximum-likelihood estimates (MLE)** for the regression parameters
2. construct a system of nonlinear equations
   a. no closed-form solution
3. solve equations using some numerical method:
   a. Newton-Raphson method

**This is included in most statistical tools:**

- **R, SAS, SPSS, NCSS...**

**BAYESIAN**

**Bayesian hierarchical model:**




**Fit parameters using Markov chain Monte Carlo:** PyMC3

# Frequency model (4)

## Poisson regression - what about overfitting?

### FREQUENTISTS

**Comparing models using a statistical test:**

- most common is F-test

**sklearn.feature_selection.f_regression**

`sklearn.feature_selection.f_regression(X, y, center=True)` [source]

Univariate linear regression tests.

Linear model for testing the individual effect of each of many regressors. This is a scoring function to be used in a feature selection procedure, not a free standing feature selection procedure.

### BAYESIAN

**Comparing models using "quality" metrics:**

- Akaike information criterion (AIC)
  - considers number of parameters and data fit
- Bayesian information criterion (BIC)
  - adds sample size
- Deviance information criterion (DIC)
- Bayesian predictive information criterion (BPIC)

# Frequency model (5)

## Overdispersion

**Overdispersion** is the presence of greater variability in a data than what can be explained by the given statistical model.

**Poisson example**:

- having only one free parameter **λ** does not allow to adjust mean and variance independently
- basic assumption: **mean == variance**

**Poisson usually does not fit!**

# Frequency model (6)

## Zero-inflated Poisson regression

### Zero-inflated Poisson distribution

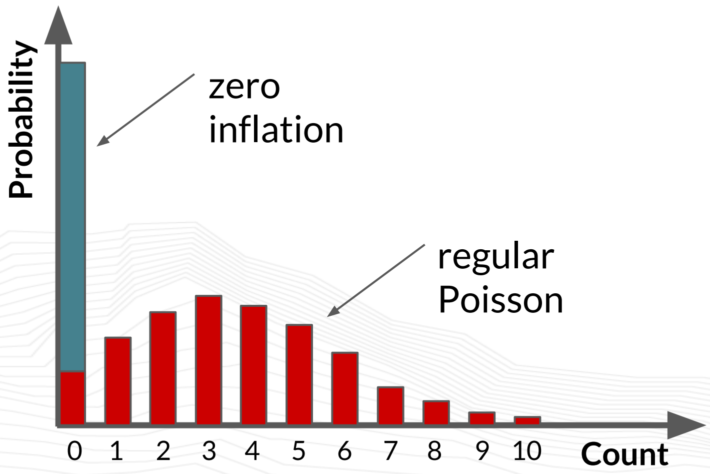
Probability of exactly ***k*** occurrences:

$$Pr(k \mid \lambda, \pi) = \pi + (1-\pi)\, exp(-\lambda), \quad k = 0$$

$$Pr(k \mid \lambda, \pi) = (1-\pi)\, \frac{\lambda^{k}\, exp(-\lambda)}{k!}, \quad k \in \{1, \ldots, \infty\}$$

where

- **$\lambda$** > 0 is the Poisson rate parameter,
- **$\pi$**>0 is the zero mass




### Regression model:

$$\pi = \frac{\alpha}{1+\alpha}$$

$$\lambda = exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$$

$$\alpha = exp(\gamma_0 + \gamma_1 z_1 + \ldots + \gamma_m z_m)$$

where:

- $x_1, x_2, \ldots, x_k$ and $z_1, z_2, \ldots, z_m$ are respective regressor variables,
- $\beta_0, \beta_1, \ldots, \beta_k$ and $\gamma_0, \gamma_1, \ldots, \gamma_m$ are respective regression coefficients,
- $\alpha$ is an additional parameter.

# Frequency model (7)

## Negative-binomial regression

**Negative-binomial distribution** arises as a continuous mixture of **Poisson distributions** where the mixing distribution is a **Gamma distribution.**

**Probability mass function:**

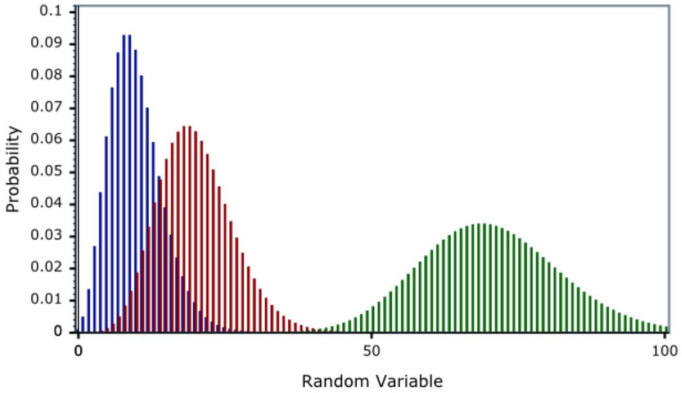
$$Pr(k \mid \mu, \alpha) = \frac{\Gamma(k+\alpha^{-1})}{\Gamma(k+1)\,\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1}+\mu}\right)^{k}$$

where

- $\mu = E(k)$ is rate parameter,
- $\alpha = \frac{1}{v}$ is a parameter controlling overdispersion.

The parameter $\boldsymbol{\mu}$ is again parametrized by a simple linear model, as in previous cases.

**Negative-binomial distribution**



# Severity model
## Gamma regression

**Probability density function:**

$$Pr(\theta \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \ \theta > 0$$

where:

- $\theta$ is a random variable that follows gamma distribution,
- $\alpha > 0$ is a shape parameter,
- $\beta > 0$ is a scale parameter.

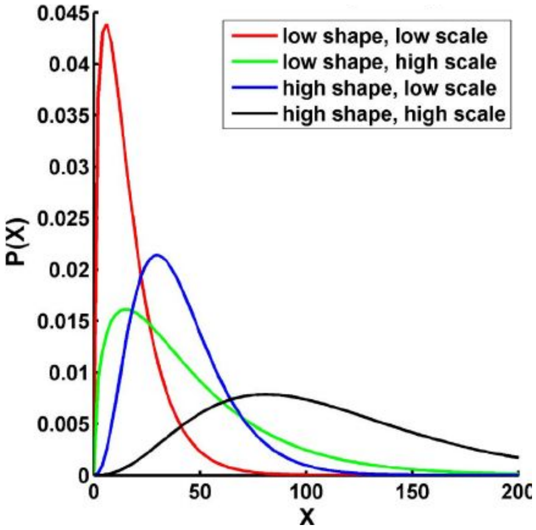
Gamma distribution has the following mean and variance: $E(\theta) = \frac{\alpha}{\beta}$ and $Var(\theta) = \frac{\alpha}{\beta^2}$.

**Regression model:**

$Pr(loss \mid claim) \sim gamma(\mu_t, \sigma_t^{2})$,

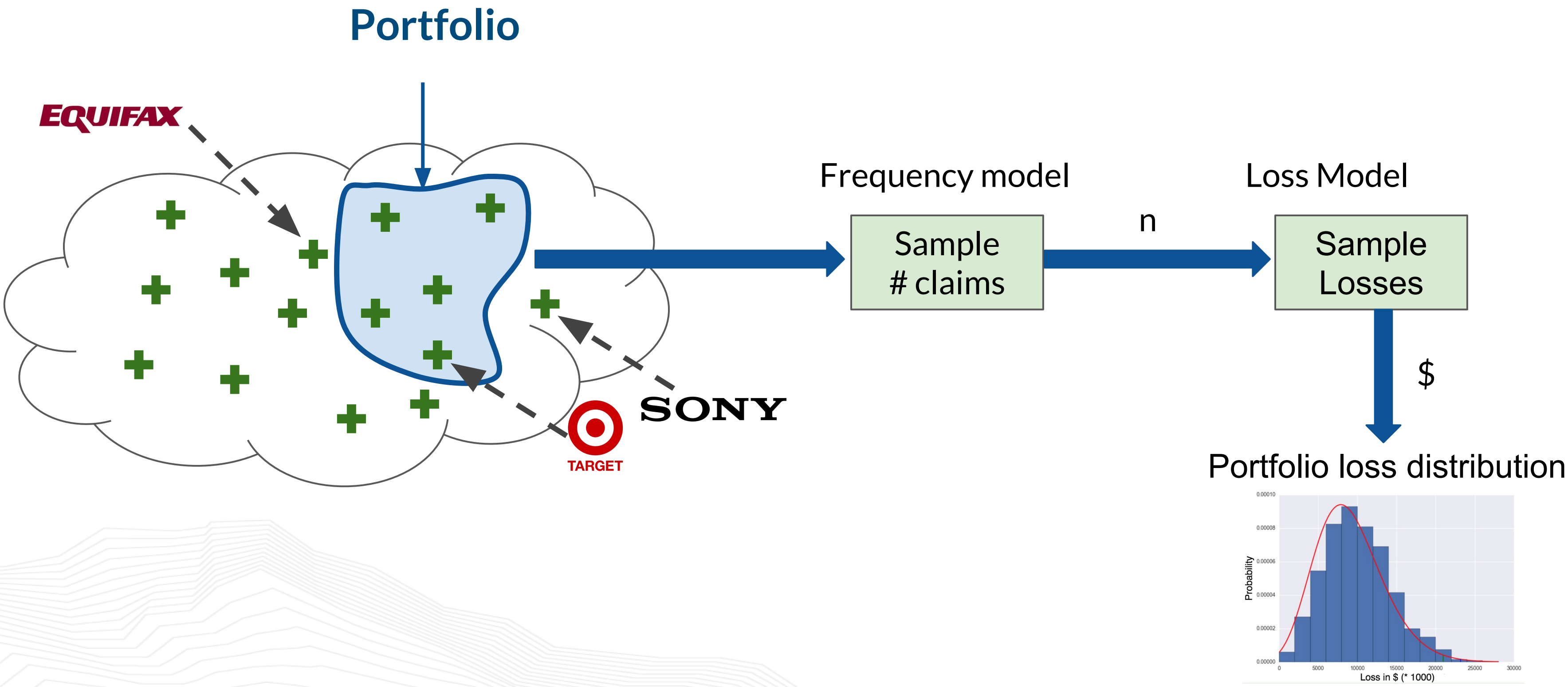
where the gamma distribution is parametrized by its mean and variance, and

- $\mu_t = exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$,
- $\sigma_t^{2} = \frac{\mu_t^{2}}{\alpha}$, where $\alpha$ is an unknown parameter.



# Pricing (1)

## Monte Carlo simulation

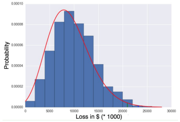


# Pricing (2)

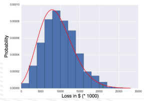
## Monte Carlo simulation - output

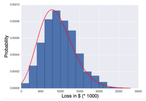
| | Company 1 | Company 2 | Company3 | Company4 | ... | Sum |
|---|---|---|---|---|---|---|
| **Iteration 1** | $0 | $30M | $0 | $15M | ... | $45M |
| **Iteration 2** | $0 | $25M | $25M | $25M | ... | $75M |
| **Iteration 3** | $0 | $0 | $15M | $0 | ... | $15M |
| **Iteration 4** | $0 | $30M | $0 | $0 | ... | $30M |
| **Iteration 5** | $10M | $15M | $0 | $0 | ... | $25M |
| ... | ... | ... | ... | ... | ... | ... |
| **Sum** | $10M | $100M | $40M | $40M | ... | ... |

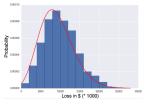
Portfolio loss distribution

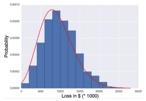



Loss distribution of companies:





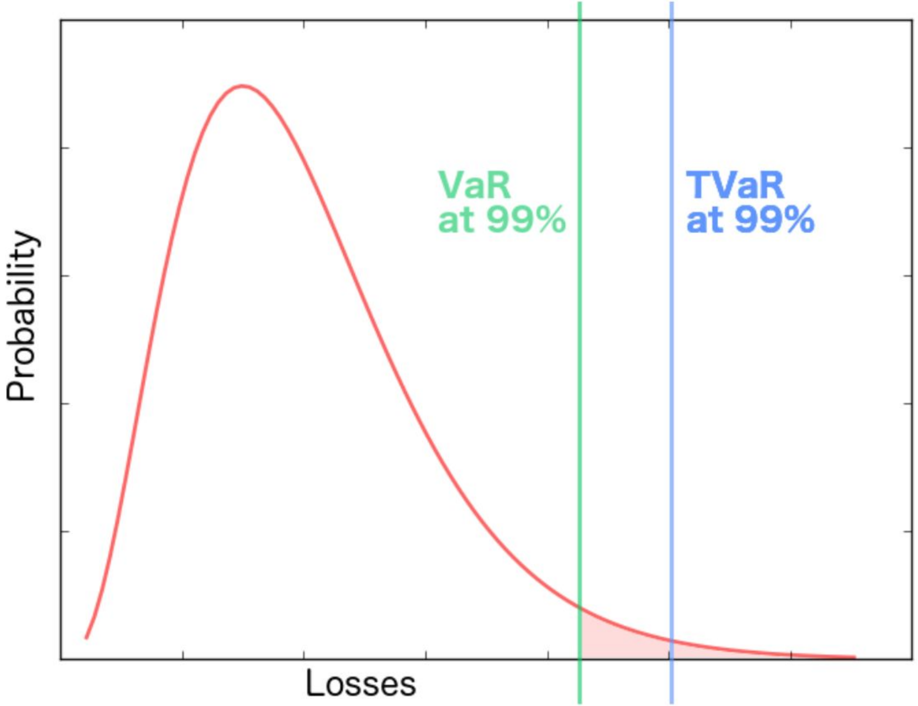




# Pricing (3)
Var & TVaR

**VaR**
- Value at Risk
- worst **x%** of losses

**TVaR**
- Tail Value at Risk
- average worst case in the tail



# Pricing (4)

## Excel sheet calculation

**Input:**

- portfolio TVaR and expected loss
- company TVaR
- bunch of bulgarian constants

**Calculation:**

- spread the desired profit across companies proportionally to their risk (TVaR)

**Output:**

- premium in $$$, finally!

### Full formula

| Profit and Contingency Load | Variable | Value |
|---|---|---|
| **Portfolio Level** | | |
| TVaR benchmark | tvar | 1.00% |
| TVar at Benchmark* | tvar_val | $1,000,000 |
| Cost of capital reserved | capital | $50,000 |
| | | |
| Expected (mean loss) for the portfolio | portfolio_loss | $2,500,000 |
| Profit load as % of losses | profit | 2.00% |
| **Policy Level** | | |
| Expected (mean) loss in $ for given policy (company) | L | $1,000 |
| | | |
| Loss adjustment expense as % of losses | a1 | 12.00% |
| | | |
| Agent or broker commision as % of premium | v1 | 15.00% |
| Premium tax as % of premium | v2 | 3.00% |
| | | |
| Fixed expenses ($ per policy) | F | $33.00 |
| | | |
| Profit load as % of losses | a2 | 2.00% |
| | | |
| Expenses that vary with losses, as a % of losses | a | 14.00% |
| Variable expenses, as a % of premium | v | 18.00% |
| | | |
| **Policy premium** | P | **$1,430** |

Is this it?

# When shit hits the fan (1)

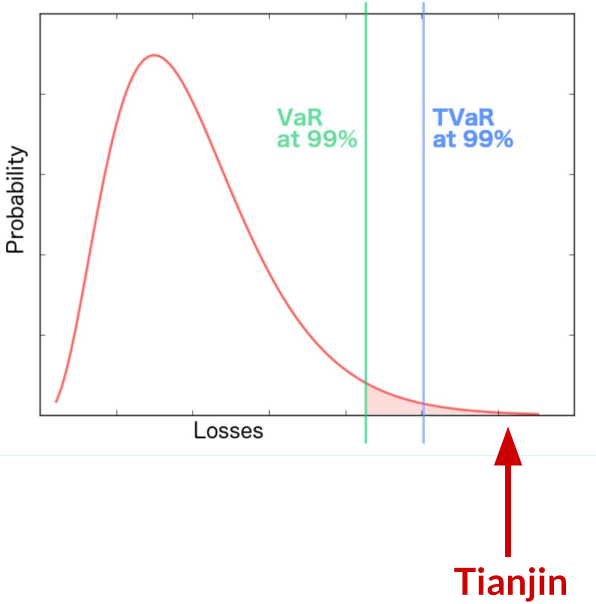
## Tianjin Blast Could Be Largest Marine Insurance Loss Ever

Image courtesy U.S. EPA

By **MarEx** 2016-02-04 18:12:51

Claims related to the massive explosion at the port of Tianjin, China may grow to as much as $6 billion, says the International Union of Marine Insurance (IUMI). More than half of the claims reportedly fall within marine insurance or reinsurance lines – potentially making it the largest single marine disaster (by claim value) in history, surpassing Hurricane Sandy.

Tail event within the portfolio



# When shit hits the fan (2)

## Ex-CEO Of Largest Swiss Insurer Commits Suicide, Three Years After CFO Hanged Himself

by Tyler Durden
May 30, 2016 3:46 PM



In the latest tragic news from the world of finance, earlier today Zurich Insurance, the largest Swiss insurer which employs 55,000 people and provides general insurance and life insurance products in more than 170 countries, reported that Martin Senn, the company's former chief executive officer who stepped down in a December reshuffle, has committed suicide. He was 59.

Senn had been a long-time employee of the insurer, serving as its chief executive for six years before stepping down in December.

The family informed Zurich Insurance that Senn had taken his own life on Friday, according to the statement. "We are profoundly shocked by the news of the sudden death," the company said. According to Bloomberg, Senn was found in his holiday house in Klosters, a Swiss ski resort, Blick newspaper reported. The cantonal police of Grisons wouldn’t confirm the death but said officers had been deployed on Friday in connection with Senn.

# Accumulation of risk (1)

## Solution 1: Hawkes process

Extending our existing approach by a ***correlation factor***.

**Hawkes process**

Self-exciting point process - generalization of a Poisson process.

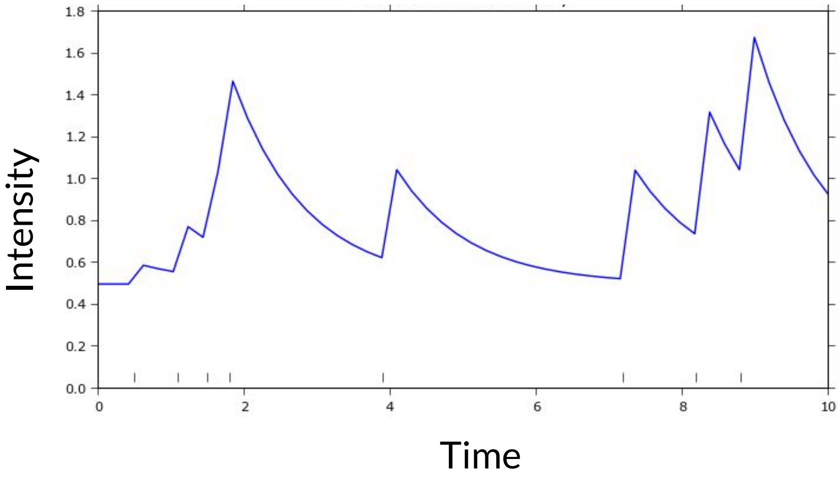
**Parameters:**

- $\boldsymbol{\mu}$ - base rate the process reverts to
- $\boldsymbol{\alpha}$ - intensity increase after an event occurrence
- $\boldsymbol{\beta}$ **-** exponential intensity decay

The **conditional intensity** at time ***t***:

$$\lambda(t) = \mu + \sum_{t_i < t} \alpha e^{-\beta(t - t_i)}$$

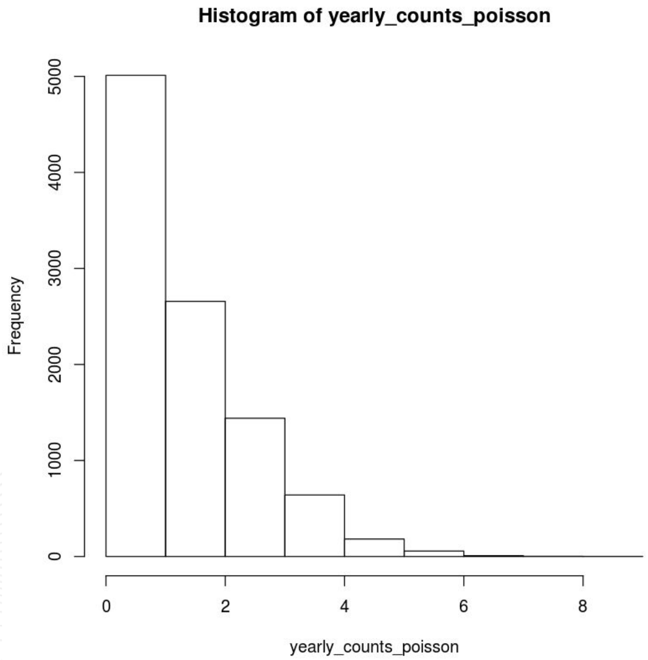
**Simulating Hawkes**

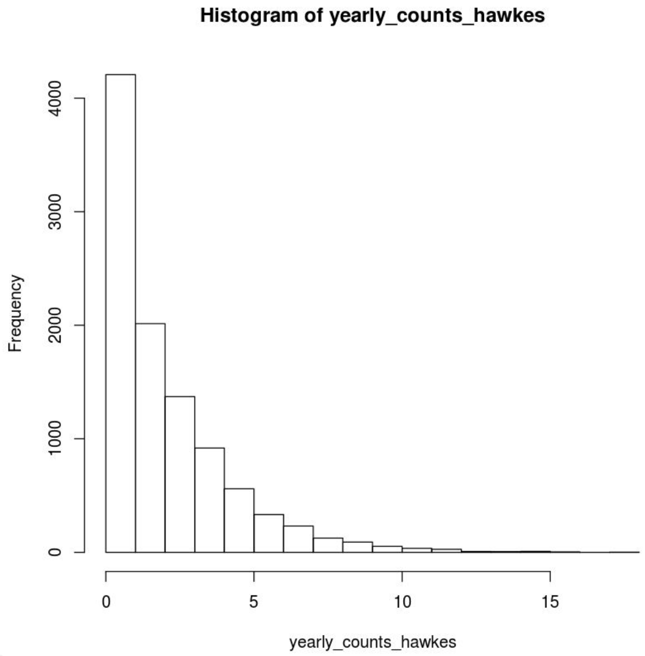



**Branching ratio:** $n = \int_0^\infty \alpha e^{-\beta t} dt = \frac{\alpha}{\beta}$

# Accumulation of risk (2)
## Hawkes vs Poisson

Simulation of a Poisson and Hawkes process with the same base rate.



# Accumulation of risk (3)

## Scenario approach for Cyber Insurance

**Main idea:** Simulate specific accumulation of risk scenarios that might occur within our portfolio.

| LEAKOMANIA | CLOUD SERVICE COMPROMISE |
|---|---|
| Systemic release of confidential customer records from many corporate enterprises.<br><br>***Example****: Three rare 'zero-day' vulnerabilities provide a criminal gang with the capability to scale data exfiltration attacks across thousands of companies.* | Mass release of confidential customer records from a specific cloud storage/database.<br><br>***Example:*** *A vulnerability in the Amazon Metadata Service allows attackers to create temporary credentials than can be used to access all your data stored on S3. Thousands of Amazon's customers are affected.* |

**Billions of confidential data records are leaked in a few months, more than the total number of confidential data records leaked in the past ten years.**

# Accumulation of risk (3)

## Scenario approach for Cyber Insurance

### Approach:

**Phase 1:**

- build a suite of scenarios that should cover the **worst vulnerabilities** that were disclosed historically and affected the largest amount of companies.

**Phase 2:**

- add scenarios of unprecedented scale that have not been witnessed yet
- need to extrapolate from historical events (phase 1) and other technological trends, e.g. increased dependence of companies on cloud provider

**Phase 3:**

- perform a stochastic simulation on top of factors shared by multiple companies

**Worst known vulnerabilities**

| Name | Year | Scale | Latest state |
|---|---|---|---|
| **Heartbleed** | 2014 | Over 600 000 websites. | 200 000 devices still affected in Jan. 2017 (source: Shodan) |
| **ShellShock** | 2014 | Estimated impact anywhere from 20% - 50% of all global servers supporting web pages. | roughly 10% of all servers still remain unpatched in 2017 (source: IBM) |
| **Stagefright** | 2015 | Nearly a billion of android devices. | N/A |
| **Poodle** | 2014 | Any web client on a public network. | N/A |
| **MS Server Service Vulnerability** | 2008 | Any instance running Microsoft Windows 2000 SP4, XP SP2 and SP3, and a few more... | N/A |

# Accumulation of risk (3)

## Scenario approach for Cyber Insurance

### Approach:

**Phase 1:**

- build a suite of scenarios that should cover the **worst vulnerabilities** that were disclosed historically and affected the largest amount of companies.

**Phase 2:**

- **add scenarios of unprecedented scale that have not been witnessed yet**
- need to extrapolate from historical events (phase 1) and other technological trends, e.g. increased dependence of companies on cloud provider

**Phase 3:**

- perform a stochastic simulation on top of factors shared by multiple companies

**Worst known vulnerabilities**

| Name | Year | Scale | Latest state |
|---|---|---|---|
| **Heartbleed** | 2014 | Over 600 000 websites. | 200 000 devices still affected in Jan. 2017 (source: Shodan) |
| **ShellShock** | 2014 | Estimated impact anywhere from 20% - 50% of all global servers supporting web pages. | roughly 10% of all servers still remain unpatched in 2017 (source: IBM) |
| **Stagefright** | 2015 | Nearly a billion of android devices. | N/A |
| **Poodle** | 2014 | Any web client on a public network. | N/A |
| **MS Server Service Vulnerability** | 2008 | Any instance running Microsoft Windows 2000 SP4, XP SP2 and SP3, and a few more... | N/A |

# Trend (1)

## Hype and heavy tails

**Trends reported in cyber:**

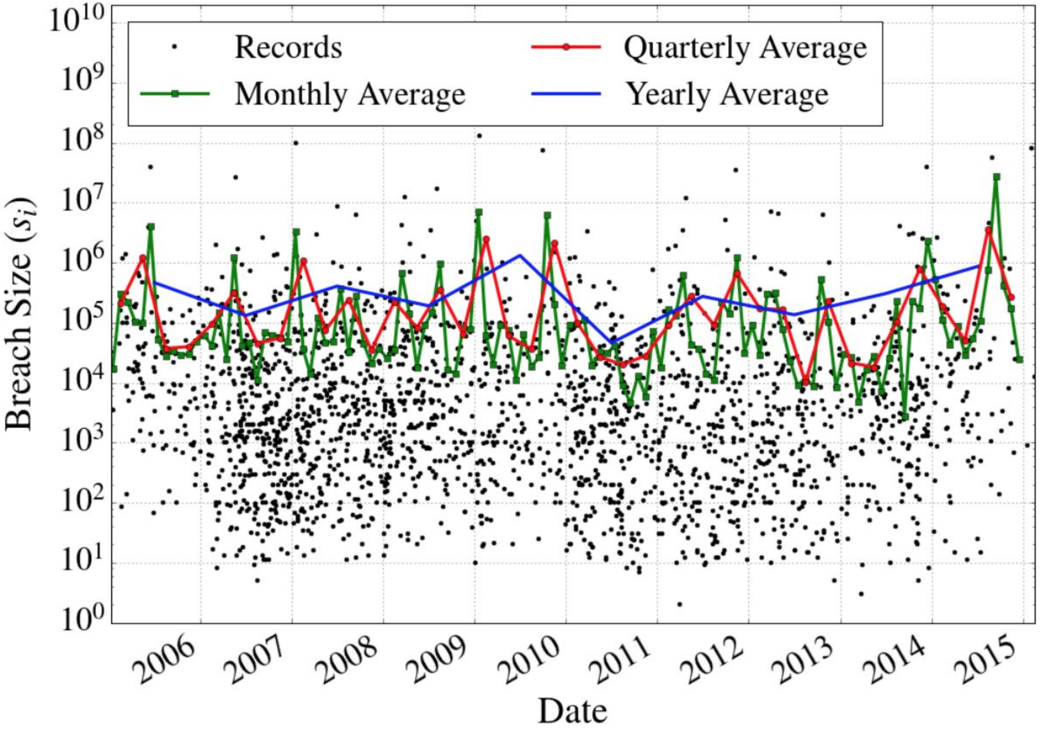
- in 2014,  Symantec reported a five-fold increase in the number of exposed records
- in 2013, Redspin reported 29% increase in the number of breaches and 148% increase in the number of exposed records.

**Are we all going to hell?**

**Issue:** The data used to produce these kinds of reports have very high variance, so simply reporting average values, can be misleading.

# Trend (2)

## Hype and heavy tails




*Source: Benjamin Edwards, Steven Hofmey, Stephanie Forrest (2015): [link]*

# Trend (3)

## Hype and heavy tails

**Approach:**

1. Figure out which distribution fits your data using e.g. **Kolmogorov–Smirnov test**.
2. Model the dependence of your distribution's mean on time:

$$
\begin{aligned}
S_n &\sim Lognormal(\mu, \tau) \\
\mu &= \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_d t^d \\
\beta_0 &\sim \mathcal{N}(\overline{\log(S_n)}, 1) \\
\beta_i &\sim \mathcal{N}(0, \frac{1}{Var[t^i]}) \\
\tau &\sim Gamma(1, 1)
\end{aligned}
$$

3. Try polynomials of different degrees and select the simplest model by BIC/DIC/BPIC.
4. If there is a significant trend in your data, a model with the time parameter(s) should be selected.

**Surprise, surprise: the constant model fits the data best for both breach sizes and breach frequencies!**

# Conclusion

Going old school still makes sense in some areas!

# Thank you