Published by

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

AN ESTIMATION OF THE RISK PREMIUM FOR INSURANCE AGAINST FLOODS

Photo: The World Bank Photo Collection

Tudor Moldovan

February 2021

About the author

Tudor Moldovan is a member of the Supervisory Board of Generali Romania – company of Generali Assicurazioni (Italy). He was formerly CEO of the said company and also served as Executive Director of Generali Bulgaria - company of Generali Assicurazioni (Italy).

Mr. Moldovan received his Actuarial education from the Institute of Actuaries, UK.

Acknowledgement

The author wishes to thank GIZ RFPI Asia for their cooperation in completing this document. At GIZ RFPI Asia office Dr. Antonis Malagardis made a decisive contribution to the completion of this study. Dr. Antonis Malagardis' long and outstanding support, for developing micro-insurance and helping the poor in Asia, and especially in Philippines, has to be acknowledged and mentioned.

Table of contents

Motto: "The future isn't what it used to be" (Yogi Berra)

Summary

This paper analyses different actuarial methods for risk premium calculation, and, consequently, for establishing a tariff premium for a defined risk, focusing, in the second part, to floods insurance. The methods discussed are appropriate for many non-life lines of business, but the document refers mainly to household insurance and/or business interruption of small and medium sized enterprises.

The first part of the document is dedicated to statistical analysis and models. It starts with the definitions of the main terms used in insurance industry and actuarial literature for risk premium calculation: *frequency*, *severity*, *exposure* and finally the almost agreed formula for *risk premium* calculation. In insurance business, statistical data for a specific insured event (household, business interruption, accidents, etc.) consist of loss data $x_1, \dots x_n$ over a specified period of time. Depending on the type of the insured event provoking the losses, a condition of the type "$x_1, \dots x_n$ come from an *independent* and *identically distributed* (iid) sample $X_1, \dots X_n$ with common *distribution function* (df) F" may or may not be justified. Considering this assumption correct, the next chapters introduce the methods for data sample analysis and propose some functions specific for this analysis, like *mean excess function* or *limited expected value function*. Different distribution functions for loss data are presented, especially *extreme value distributions*. Using the loss data sample and the statistics provided, the risk premium can be calculated using the frequency and the expected value of loss distribution function. Alternative sample functions are defined, too. The subject of simulating data is also discussed.

The second part is dedicated to floods risks for household insurance and/or business interruption. In both cases, the indemnity paid is a cash amount as a percentage of the sum insured. Different aspects of floods insurance, its difficulties in defining a precise cover and the links to different other risks are discussed. The risk premium depends on loss distribution and return periods of floods events. A theoretical example is provided, using rainfall height data provided by Catarman[1] Meteorological Station.

This study is useful for underwriters and actuaries of insurance companies. More details about the methods proposed in this study can be found in the Bibliography presented at the end of the document. The interested reader can find more ideas in the first document of the Bibliography about the subject of analyzing frequency and loss data.

The Annex at the end of the document provides details about formulas, theorems (without proof) and concepts used in this document.

[1] Catarman is the capital Municipality of the province of Northern Samar located in the central part of the Philippines

1. Introduction

Losses caused by occurrence of unexpected events are problems both for individuals and for society as a whole. Insurance is a mechanism for spreading these losses. Examples of insured events and their consequences are property damage or business interruption due to floods, typhoons, fire, theft, hails; car accidents (for replacement cost); disability or death (loss of future income and support); illness (cost of medical treatment); and personal injury resulting from accidents. These are just a few of many basic insured events, more complicated ones can be designed also.

The actuary wants to evaluate somehow the probability of an insured event occurring. In particular, it is important to know the expected number of occurrences for a specific measure of exposure to risk. For a simple example, we might observe the number of claims occurring during the next year for a certain group of insured people against car accidents. Upon dividing by the number of cars, we obtain an estimate of the *expected number of claims* for one person in one policy year. This ratio is the **mean frequency** and is defined as:

$$\frac{number\ of\ occurences}{exposure\ (to\ risk)\ basis}$$

From usage in insurance industry, the ratio is usually called simply **frequency**.

Frequency is one of the principal topics of this document, being an extremely important measure in placing a value on an insurance contract. It should be noted that the terms of the insurance contract itself have a determining effect on frequency, as is the case of deductibles, which reduce the number of reported claims.

Together with frequency, another important measure in placing a value on an insurance contract is how much money will cost the insurer if the insured event occurs: “if the insured event occurs, what will be the cost to the insurer?”. The cost (i.e., the loss) is a *random variable*. One characteristic of several such costs, arising from the occurrence of similar insured events, is the *mean value*, sometimes referred to as a **mean severity**, or to conform with insurance usage, simply **severity**, which is the ratio:

$$\frac{total\ losses\ from\ all\ occurences}{number\ of\ occurences}$$

It is important to note that severity estimates the *expected value* of the individual loss random variable, and sometimes severity refers to that expected value when there is no chance of confusion. Since the amount of an insured loss normally is expressed in dollars (or other currency), the severity normally is expressed as an average dollar amount per loss.

Severity, like frequency, is an extremely important measure in placing a value on an insurance contract. Almost without exemption the terms of the insurance contract have a determining effect on severity: introducing a *deductible* or a *limit* on the payment reduces the severity.

The **risk premium** (**pure premium**) is the average amount of losses per exposure to risk. Thus, the risk premium is the ratio:

$$\frac{total\ amount\ from\ all\ occurences}{exposure\ (to\ risk)\ basis}$$

which equals the product:

(frequency) x (severity)

But the client is interested in the amount of money (dollars) he will pay for concluding an insurance contract for a defined event, i.e., the **insurance premium** or **premium** (tariff premium or gross premium). The relationship between the premium and the risk premium may be expressed as follows:

Tariff Premium = (risk premium) x (exposure) + expenses (of doing business) + risk charge

Before discussing modelling techniques, we need to clarify the topic of loss. In insurance context, *the loss* is the value of the actual damage caused by an insured event. It always is represented by the random variable (rv) *X* and its support is always represented by nonnegative real numbers. The amount paid by the insurer is a function of *X* determined by the policy. This amount is called the *payment,* or the *amount paid*. In most cases the insurer had data on the payment but desire to model the loss.

1.1 Modelling Techniques

In the application of mathematics to the real world, the selection of a suitable model is extremely important. For the actuary, models for distributions of *random variables* (rvs) are of great concern, for example the losses due to fire or due to floods. The actuary has, mainly, two big options for models' selection:

a) Standard distribution models
b) Probabilistic models describing *extremal events,* e.g., major insurance claims (property losses or business interruption due to typhoons, floods, earthquakes etc.), flood levels of rivers, rainfall values in a specific area, extreme levels of environmental indicators such as ozone or carbon monoxide etc. The corresponding distributions are called *standard extremal value distributions* and the corresponding rvs *standard extremal* rvs.

In both cases, the real world informs the insurer about losses and/or extremal events through *statistical data*. Statistical mathematics tries to offer to the insurer/actuary the necessary set of tools in order to deduce scientifically sound conclusions from data. It is also about *reporting correctly*: the data have to be presented in a clear and objective way, precise questions have to be formulated, model-based answers given, always stressing the underlying assumptions.

In insurance business, statistical data for a specific insured event (household, fire, business interruption, accidents, etc.) consist of loss data $x_1, \ldots x_n$ over a specified period of time in a well-defined portfolio. Depending on the type of the insured event provoking the losses, a condition of the type "$x_1, \ldots x_n$ come from an *independent* and *identically distributed* (iid) sample $X_1, \ldots X_n$ with common *distribution function* (df) *F*" may or may not be justified.

In both cases a) and b) from above, the statistical techniques used for fitting the unknown distribution *F* are the same, but the parameters of these distributions are different, and, consequently, their estimators.

1.2 Exploratory Data Analysis

We assume that the *n* sample items $X_1, \ldots X_n$ are iid with common df *F*: if $f = F'$ (first derivative) is the *probability distribution function* (pdf) of the unknown distribution *F*, then the joint pdf of the sample observations is:

$$f(x_1)\cdot f(x_2)\cdot \ldots \cdot f(x_n)$$

The main subject of this point is:

Find a distribution function F which is a good model for the iid data $X_1, \ldots X_n$.

Given a set of data to be analyzed, the actuary usually starts with a *histogram*, one or more *box-plots* or a *plot of empirical* df. Or he can start by defining other characteristics of the sample, named *statistics*, for instance the *sample mean*, *sample variance* or other higher moments:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

(S^2 is an unbiased estimator of the population's variance σ^2).

1.3 Empirical Distribution Function

We can define the *empirical df* or *sample df*:

$$F_n(x) = \frac{1}{n} card\{i : 1 \le i \le n, X_i \le x\} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}, \qquad x \in R$$

where I_A stands for the indicator function of the set A. The corresponding pdf, $f_n(x)$, is a discrete function with a weight of $1/n$ on each X_i.

As an application of the *strong law* of *large numbers* (SLLN), Glivenko-Cantelli theorem says that $F_n(x)$ converges almost sure (a.s.) to the true (unknown) df $F(x)$, see details in [3].

F_n, as defined above is a discrete distribution of the sample and gets close to the true df *F*, which usually is of continuous type. That is why it is better to smooth F_n in some manner: drawing connecting line segments through F_n. The first line segment begins at some point, say $x = c_0$, which is less than or equal to the smallest x_i, and ends at $x = c_k$, which is greater or equal to the largest x_i. Suppose the intermediate points are $c_1 < c_2 < \dots < c_{k-1}$, which are usually taken so as not equal to any given x_i. These points $c_0, c_1, c_2 \dots c_{k-1}, c_k$ can be selected as not equal spaced, but they should be selected so that the line segments joining the points:

$$[c_0, F_n(c_0) = 0], [c_1, F_n(c_1)], \dots\dots\dots, [c_{k-1}, F_n(c_{k-1})], [c_k, F_n(c_k) = 1]$$

fit F_n(x) quite well. Obviously, the more points the better the fit, but the principle of parsimony dictates that we should use as few line segments as possible without losing too much fit.

1.4 Ordered Sample, Quantile Function and Quantile Plots

Define the *ordered sample* of the data sample: if *F* is a continuous function, for the sample $X_1, X_2, \dots..X_n$ we may assume the ordered sample $X_{n,n} < \cdots < X_{1,n}$. For the sample $X_1, X_2, \dots..X_n$ we denote *the empirical quantile function* by $F_n^{\leftarrow}$. In this case $F_n^{\leftarrow}$ is a simple function of the order statistics, namely:

$$F_n^{\leftarrow}(t) = X_{k,n}\ for\ 1 - \frac{k}{n} < t \le 1 - \frac{k-1}{n}$$

and for $k = 1, \dots, n$.

For the case $t = 1 - \frac{k-1}{n}$ we obtain $F_n^{\leftarrow}\left(1 - \frac{k-1}{n}\right) = X_{k,n}$.

The *quantile transformation* is the theoretical basis which defines *probability plots* (PP plots):

$$\{\left(F\left(X_{k,n}\right), \frac{n-k+1}{n+1}\right), k = 1 \dots \dots n\}$$

or *quantile plots* (QQ plots):

$$\{(X_{k,n}, F^{\leftarrow}\left(\frac{n-k+1}{n+1}\right)), k = 1 \dots \dots n\}$$

There exist various variants of QQ plot defined above of the type:

$$\{(X_{k,n}F^{\leftarrow}(p_{k,n})), k = 1 \dots \dots n\}$$

where $p_{k,n}$ is a certain *plotting position,* see details in Annex or [3].

In both cases, the approximate linearity of the plot is justified by Glivenko-Cantelli theorem.

1.5 Mean Excess Function and Mean Excess Plot

Another useful graphical tool, in particular for discrimination in the tails, is the *mean excess function e(u)*, which is usually known as *mean excess over the threshold value u*:

$$e(u) = E(X - u / X > u)$$

e(u) is the expected value of excess losses over the threshold *u*, subject to the condition that the loss *X* is higher than the threshold *u*.

In reinsurance context, *e(u)* can be interpreted as the expected loss in the unlimited layer, over priority *u* (for excess of loss reinsurance): here *e(u)* is also called *mean excess loss* function. *e(u)* is calculated for almost all loss distribution functions *F* and a graphical test for tail behaviour can be based on the *empirical mean excess function* $e_n(u)$.

Define $\Delta_n(u) = \{i : i = 1, \ldots, n, X_i > u\}$ then, for $u \geq 0$:

$$e_n(u) = \frac{1}{card\Delta_n(u)} \sum_{i\in\Delta_n(u)} (X_i - u)$$

A continuous df *F* is uniquely determined by its main excess function: if you find a model for the distribution function, then the value of *e(u)*, calculated using this df, should be closer to $e_n(u)$.

A *mean excess plot (ME plot)* consists of the graph:

$$\{X_{k,n}, e_n\left(X_{k,n}\right) : k = 1, \ldots, n\}$$

ME plot is a graphical method used mainly for distinguishing between light and heavy tailed models. Some caution is needed when interpreting such plots. Due to spareness of the data available for calculating $e_n(u)$ for large values of *u*, the resulting plots are extremely sensitive to changes to data towards the end of the range.

1.6 Limited Expected Value Function

Similar with mean excess function *e(u)* is the *limited expected value function E(X;d)* of the rv *X*:

$$E(X; d) = \int_0^d xf(x)dx + d \cdot [1 - F(d)] = \int_0^d xf(x)dx + d \cdot \bar{F}(d)$$

where *f* is the pdf of *F*, $\bar{F}$ is the tail of *F* and *d* is the threshold, as above.

If the mean residual function *e(d)* exists, *E(X;d)* and *e(d)* are related through the equality:

$$E(X) = E(X;d) + e(d)\cdot[1 - F(d)] = E(X ; d) + e(d) \cdot \bar{F}(d)$$

where *E(X)* is the *expected value* of rv *X* and $\bar{F}$ is the *tail* of the df *F*, $\bar{F}(d) = 1 - F(d)$.

It is important to note that *E(X;d)* exists for all distributions and is calculated in textbooks, see [4].

Similar with the empirical mean excess function, we can define an *empirical limited expected value function* for a sample:

$$E_n(d) = \frac{1}{n}\sum_{i=1}^{n} \min(x_i, d)$$

Before accepting any model as providing a reasonable description of the loss data, the actuary should verify that $E(X;d)$ and $E_n(d)$ are essentially in agreement for all values of d.

Since

$$\lim_{d\to\infty} E(X\,;d) = E(X)\ (if\ it\ exists)\quad and\quad \lim_{d\to\infty} E_n(d) = \bar{X}$$

comparing $E(X;d)$ and $E_n(d)$ is similar with method of moments approach, only much more restrictive: not only the means value agree, but also the limited expected value functions essentially agree for all values of d.

$E(X;d)$ has another interesting meaning: as it was mentioned, for the insurer an important quantity is the severity. When a coverage modification is in effect, the insurer is interested in the expected loss that is eliminated (for the insurer) by the modification. If, for example, the modification is a *deductible* of d, the expected loss eliminated is just $E\,(X;d)$, which in this case can be viewed as the expected value of rv $Y = min(X\,;d)$, that is the mean of a rv *censored* (*truncated*) at d.

1.7 Risk Premium

Using the model of losses truncated at d, let us consider the rv X with df F_X and pdf f_x representing the loss variable without truncation. We define the rv Y as the truncated X losses at d, with df F_Y and pdf f_Y:

$$F_Y(x) = \begin{Bmatrix} 0 & if\ x \le d \\ P(X \le x|X > d) & if\ x > d \end{Bmatrix}$$

But

$$P(X \le x|X > d) = \frac{F_X(x) - F_X(d)}{1 - F_X(d)}$$

where P stands for probability.

And

$$f_Y(x) = \begin{Bmatrix} 0 & if\ x \le d \\ \dfrac{f_X(x)}{1 - F_X(d)} & if\ x > d \end{Bmatrix}$$

Let p being the frequency of a loss before imposing the deductible d. With this definition, the risk premium before introducing the deductible d is:

$$p \cdot E[X]$$

where $E[X]$ is the expected value of X (expected loss). With deductible d as defined above, the new frequency becomes $p \cdot P(X > d) = p \cdot [1 - F_X(d)]$ (only losses higher than d are considered).

The new risk premium is:

$$p \cdot [1 - F_X(d)] \cdot E[Y]$$

or

$$p[1 - F_X(d)]\left\{d + \frac{E[X] - E[X;d]}{1 - F_X(d)}\right\} = p\{E[X] - E[X;d] + d[1 - F_X(d)]\}$$

(see details in [4]).

1.8 Estimating Distributions by Simulation

Sometimes it is theoretically impossible to find the df or pdf of a function of random variable in a convenable mathematical way: the mathematics might be too difficult and require some numerical methods or it may be that the distributions of rvs are known only approximatively.

One way of solving this important problem is through *simulation*. We consider rvs of continuous type. If we want *X* to have the distribution with df *F(x)* of the continuous type and *Y* has a *uniform distribution* on $0 \leq y < 1$ then *Y = F(x)* or equivalently, $X = F^{-1}(Y)$ yields a rv *X* with df *F(x)*. The distribution of *X* generated in this way is:

$$P(X \leq x) = P(F(X) \leq F(x))\ since\ F(\cdot)\ is\ nondecreasing\ function$$

By definition *Y = F(x)* and so

$$P(X \leq x) = P(Y \leq F(x)) = \int_0^{F(x)} (1)dy = F(x)$$

since the pdf of the uniform distribution on $0 \leq y < 1$ is *g(y)* = 1. That is, the df of *X* is exactly *F(x)*. Uniform random variables are easy to generate with a computer.

For example, suppose we want to generate a rv with a *Pareto* df *F*:

$$F(x) = 1 - \frac{\lambda^\alpha}{(\lambda + x)^\alpha},\ \ 0 \leq x < \infty$$

Let *Y* be the uniform rv on $0 \leq y < 1$. Writing:

$$Y = 1 - \frac{\lambda^\alpha}{(\lambda + X)^\alpha}$$

and solving for *X* in terms of *Y* we obtain

$$X = \lambda\left[\frac{1}{(1 - Y)^{1/\alpha}} - 1\right] = F^{-1}(Y)$$

X defined by the latter expression has the desired Pareto distributions and Y, as a uniform rv, can be simulated on a computer. Running many simulations on Y the actuary can obtain the data sample of random variables X having the desired Pareto distribution.

For ordered statistics, the concept of *quantile transformation* is extremely useful since it often reduces the problem concerning order statistics to one concerning the corresponding order statistics from a *uniform* sample. The same technique of generating uniform rvs with a computer applies, details in the Annex.

1.9 Fitting a Model to the Data

The distributions used for modelling loss data usually depend on one or more parameters (up to three). The actuary's mission is to find an estimation (estimator) of these parameters using the data sample. The mission is accomplished by using different methods, for instance: *percentile matching, method of moments, minimum distance, minimum chi-square* or *maximum likelihood.* The first two ones are crude methods which have the advantage of being easy to apply, but the penalty is a significant lack of accuracy. The last three methods are formal procedures with well-defined statistical properties. They produce reliable estimators but are computationally complex. For each of the standard distributions used in insurance, iterative procedures are needed to obtain the corresponding estimators. For extreme value distributions, there are specific estimators defined, which can be calculated using ordered statistics, for details see [3].

The standard distributions used for modelling loss data in insurance are Pareto, Weibull, Lognormal, Burr, Gamma, Transformed Gamma, Loggamma.

Regarding the extremal events, the fundamental *Fischer-Tippett theorem*, which establishes limit laws for maxima, has the following content:

Let (X_n) be a sequence of iid random variables and $M_n = \max(X_1, X_2, \ldots, X_n)$. If there exist constants $c_n > 0 \ and \ d_n \in R$ such that

$$c_n^{-1}(M_n - d_n) \ converges \ in \ distribution \ to \ H$$

for some non-degenerate distribution H, then H must be of the type of one of the three so called *standard extreme value distributions* (Fréchet, Weibull, Gumbel).

The standard extreme value distributions are limited to only three types (Fréchet, Weibull, Gumbel), but the *maximum domain* of *attraction* (MDA) of the extreme value distributions provides more options. For instance, the following standard dfs belong to:

- Fréchet's MDA: Cauchy, Pareto, Burr, Loggamma
- Weibull's MDA: uniform, power law, Beta
- Gumbel's MDA: exponential, Weibull, Gamma, Normal, Lognormal, Benktander-type-I, Benktander-type-II.

The three standard types of extreme value distributions can be unified in so called *standard generalized extreme value distribution (GEV)* H_ξ.

Another important distribution in the extremal domain is the *generalized Pareto distribution* (GPD) $G_{\xi,\beta}$.

One of the main results of the extreme value theory is:

The GEV H_ξ, for $\xi \in R$ describes the limit distributions of normalized maxima (of losses).

The GPD $G_{\xi,\beta}$, $\xi \in R$, ß > 0 is the limit distribution of scaled excesses over high thresholds.

1.10 Tail and Quantile Estimation

Now, assume that we consider our sample $X_1, \dots X_n$ of iid rvs with df F belonging to the maximum domain of attraction of H_ξ:

$$F \in MDA(H_\xi)\ for\ some\ \xi \in R$$

Let $0 < p < 1$ and x_p denote the corresponding *p-quantile*, $x_p = H_\xi^{\leftarrow}(p)$.

The whole point behind the domain of attraction condition $F \in MDA(H_\xi)$ is to be able to estimate quantiles *outside* the range of data, i.e., for $p > 1 - 1/n$, which is equivalent to finding estimators for the tail $\bar{F}(x)$ with x large. It is extremely important to note that estimation outside the range of our data can be made only if extra model assumptions are imposed. Estimating quantile outside the range of our data is similar to moving into the territory of "unknown unknowns", as Donald Rumsfeld, the US Secretary of Defense, said in 2002 (within insurance community, these are the unidentified risks which carry the greatest uncertainty and can have the most consequential financial impacts). There is no magical technique which yields reliable results for free.

Let's define:

$$U(t) = F^{\leftarrow}(1 - t^{-1})\ so\ that\ x_p = U\left(\frac{1}{1-p}\right) is\ the\ p - quantile$$

Denoting $U_n(t) = F_n^{\leftarrow}(1 - t^{-1})$, where $F_n^{\leftarrow}$ is the empirical quantile function, we get:

$$U_n\left(\frac{n}{k-1}\right) = F_n^{\leftarrow}\left(1 - \frac{k-1}{n}\right) = X_{k,n},\ \ k = 1, \dots, n$$

Hence $X_{k,n}$ appears as a natural estimator of the $(1 - (k - 1)/n)$ – quantile. The range $[X_{n,n}, X_{1,n}]$ of the data allows one to make a within-sample estimation up to the $(1 - 1/n)$ – quantile.

The *extreme value theory* (EVT) offers techniques allowing for extrapolation outside the range of available data. But such estimates should be treated with extreme care. Using these models, estimates for p-quantiles $x_p\ for\ every\ p \in (0,1)$ can be given, but the reliability of these estimates becomes very difficult to judge in general. The EVT model-based estimates could form the basis for a detailed discussion with the actuary/underwriter responsible for these data. One can use them to calculate the so called *actuarial fair premiums* or *technical premiums*, which can be interpreted as those premiums which the actuary believes to reflect the information available most honestly from the data. But many other factors enter into discussion: economic conditions, management strategy, market forces etc. have to be considered, and by using all these inputs the actuary will be able to come up with a premium acceptable both for the insurer and the insured – more details about this point in [3].

2. Flood Insurance – Methodology for Premium Calculation

2.1 Background

The provision of floods cover by insurers is becoming more common, although is still far from the norm. Recent advances in flood risk data have made risk rating of floods cover more achievable for insurers. The availability of flood risk data is expected to increase further, following public-private partnerships initiatives to provide to the industry flood risk data on individual properties within the country (Philippines).

In general, the main types of flood risk to properties include (but are not limited to):

- Flash flooding – caused by high intensity (but short) duration storms that produce localized flooding conditions, sometimes as a result of inappropriate drains.
- Riverine flooding – typically occurs as a result of overflow of rivers and creeks following long duration rainfall over large areas.
- Storm surge flooding – caused by rising coastal waters associated with a storm event.
- Tsunami.

There is no standard definition of floods used by insurers. Traditionally (and at risk of oversimplifying the position), householders' policies have included coverage for flash flooding, but have excluded coverage for riverine flooding. This distinction, together with the variety of policy wordings, can provide complication for policyholders and insurers alike. In practice, it is sometimes impossible to distinguish between riverine and flash flooding. In other cases, insurers have made ex-gratia payments to policyholders, rather than risk damage to their reputations. Some flood coverage has been provided for larger commercial risks.

In this paper, we address the issue of pricing ***flash flooding*** and ***riverine flooding*** covers.

2.2 Flash Flooding

The provision of flash flooding coverage is similar with other natural perils insured (such as typhoons, for example). Reasons for this include:

- flash flooding, as typhoons and other natural perils impact, to some extent, over a large area.
- if risk rated, the cost to individual insureds can become affordable. This reflects that the exposures are significant and shared across a large proportion of risks.

For example, for Philippines, the data provided by Catarman Meteorological Station, containing monthly rainfall data for the period 1972 – 2018, can be used, with additional assumptions, to define and calculate the basic risk premium for flash flooding of buildings and/or contents or business interruption. This approach has to be used with caution, only the method has significance, the data are not appropriate and can lead to big errors and mispricing.

The information needed to estimate the cost of flash flooding includes:

- Flood risk data – especially maxima of rainfall heights which is susceptible to produce flash flooding in specific location(s), due to inefficient drainage systems (typically in terms of the depth of water likely from floods of certain frequencies).
- Loss curves – providing the relationship between flood depth and the extent of damage to property. In theory, the loss curve will differ according to the location of the property exposed to flash flooding and the characteristics of the dwelling.

Risk data for flash flooding, for each location or area (properly defined):

- depth above ground level of flood at various return periods (available for at least 20 years, 100 years and, if possible, probably maximum flood).

2.3 Riverine Flooding

The provision of riverine flooding coverage is more problematic than other natural perils insured (such as typhoon, for example). Reasons for this include:

- the difficulties in identifying properties at risk, whereas typhoons and other natural perils impact, to some extent, over a large area, the riverine flooding risk can vary materially from house to house.
- if risk rated, the cost to individual insureds can become exceptionally large and unaffordable. This reflects that the exposures are significant and shared across a small proportion of risks, with a large part of properties having no riverine flooding risk at all.

The information needed to estimate the cost of riverine flooding includes:

- Flood risk data – providing for each location its susceptibility to riverine flooding (typically in terms of the depth of water likely from floods of certain frequencies).
- Loss curves – providing the relationship between flood depth and the extent of damage to property. In theory, the loss curve will differ according to the location of the property exposed to riverine flooding and the characteristics of the dwelling.

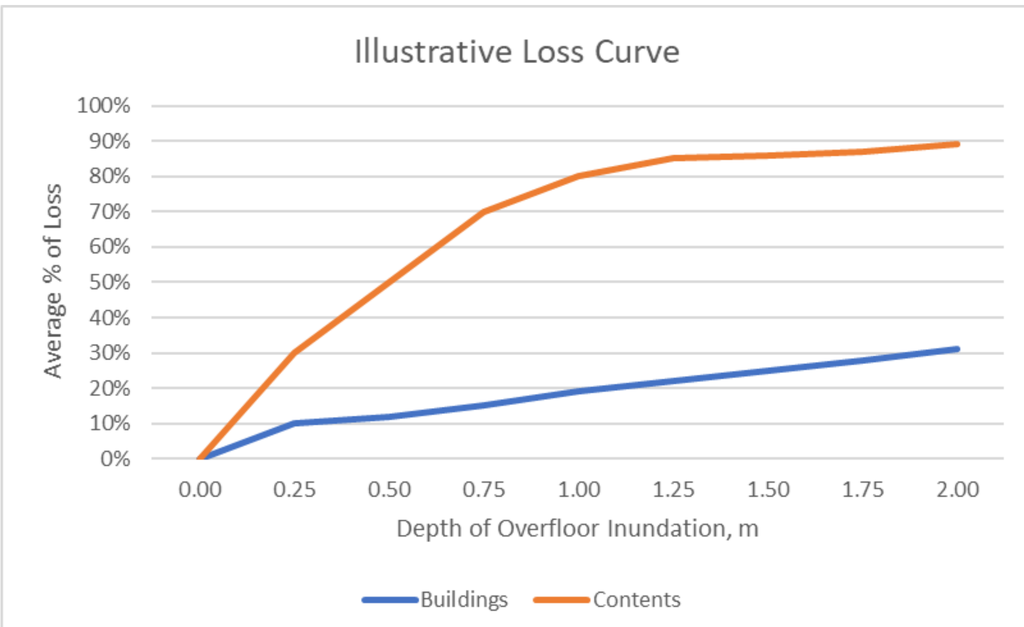
Risk data for riverine flooding, for each location or area (properly defined):
- depth above ground level of flood at various return periods (available for at least 20 years, 100 years, and probably maximum flood).

Unfortunately, for Catarman, the data about riverine flooding does not exist, but the method developed for flash flooding applies also to riverine one.

2.4 Loss Curves

Loss curves describe the relationship between the level of above the ground inundation and the damage to an individual property. The depth of flooding is sometimes expressed in terms of the extent of above floor flooding (as distinct from above ground). The water level used to assess the damage is the highest water level recorded during the flood event.

The average loss may be expressed as either a monetary amount (dollar or PHP), or a % of sum insured (which is preferable).



Typically, the increase in cost is less than the proportionate increase in flood depth (e.g., the cost of a 2m flood is less than double that of a 1m flood).

In theory, different loss curves should be used by:

- *region and the type of flood likely in that area*. This reflects that the severity of a flood is defined by more than just its depth, with factors such as the velocity of the floodwaters and the duration of flooding also relevant to the cost likely to emerge. In this regard it is necessary to consider the topographical characteristics of each catchment area.
- *type of buildings* (e.g., construction type, number of floors, floor height).
- *product* (e.g., commercial versus home).
- *coverage* (e.g., buildings versus contents, indemnity versus replacement value).

- *rating factors* (e.g., for commercial, the nature of business would be an important criteria).
- *socio-economic profile of region* (which is linked to the nature of the properties insured).

Flood loss curves may be developed based on empirical analysis, as described in the first part of this document. However, this is not always possible, and some curves may be developed synthetically by examining properties and estimating the types of damage that will be sustained at various water depths.

2.5 Other Data Sources

Besides the information available on flood risk, there are a number of other data sources that may be required by an insurer to comprehensively rate floods. These include specific policyholder information, for instance:

- building construction type
- whether it is a Buildings or Contents policy
- whether coverage is Residential or Commercial
- number of floors
- whether there are any rooms below ground level
- height of building from ground level, including if the building is built on stilts
- what external buildings are covered, e.g., sheds, swimming pools, fences.

2.6 Methodology

"The future isn't what it used to be" is the motto of this document: the gains in our ability to model (and predict) the world may be dwarfed by the increase in its complexity − implying a greater and greater role for the unpredicted.

The ***law of iterated expectations*** can help:

> *Strong form of the law of iterated expectations*: if I anticipate expecting *something* at some date in the future, then I already expect that *something* at present.
>
> *Weaker form of the law of iterated expectations*: to understand the future to the point of being able to predict it, you need to incorporate elements from this future itself.

This weaker form is useful in insurance business.

This section discusses the methodology for basic risk premium calculation for households or businesses exposed to flooding, flash, or riverine ones: the method is the same, only the source of data is different, and of course, the definition of covers. The method applies to buildings, but also to contents. The indemnity paid is a cash amount.

As with most pricing tasks described in the first part of this document, the risk premium (or technical price) for flooding risk involves understanding the (*expected*) ***frequency*** of events and (*expected*) ***loss*** or ***severity***. As with many natural hazards, events of different magnitude have different expected frequencies – or return periods.

The scope of the calculation is to combine the information available at the risk level about the depth of floods at various return intervals with the loss curve, to derive the costs that will emerge from those floods.

The steps to follow for risk premium calculation are outlined below:

1. Start with the analyses of the data's sample for rainfall/floods provided by different sources, like local government units, meteorological stations, web-based data, reinsurers, own portfolio, for the specific region/province/location. If the data sample is large, it is useful to analyze the maxima of data, because maximum rainfall heights can produce flash flooding or even riverine flooding – see the first part of this document.
2. Use these data to calculate frequency distributions or fitting distribution functions like Pareto, Log-normal or Gumbel (for maxima). If it is not possible to fit distribution functions, the empirical frequency distribution can be used, but it is likely that this approach will underestimate the likelihood of very rare events, with high impact on losses.
3. Using these distributions, the probability of exceedances of different floods level can be established.
4. It is useful to work with *return period*, defined as the inverse of the exceedance probability. Why it is useful to work with return periods? Because very rare events, with an extremely low frequency like 0.001, are not visible in the plotted tail of the distributions; 0.001 is plotted as almost zero on a chart describing the distribution function. But a return period of 1/0.001 = 1,000 years can be very easily plotted on a chart (using logarithmic scale, for instance). A return period of 1,000 years does not mean that you have to wait 1,000 years for such an event to happen, it can happen next year or in any year in the next period of 1,000 years, with an annual probability of 0.001.
5. Following the steps 1 - 4, the main result is a table with 2 columns, return period (years) on x-axis and flood height (m) on y-axis.
6. Using this table, the chart of flood occurrence risk can be plotted: on x-axis is the frequency of annual exceedance and on y-axis the flood height.
7. The most difficult step is to design loss curves for properties proposed for insurance. It means to find a correspondence between the flood height (which was analyzed at points 1-4), the depth of the overflow inundation at buildings' location and the damage likely to be produced (as a monetary amount or percentage of sum insured, mostly preferred).
8. Now, using the flood occurrence risk curve with the loss curve we get a loss curve depending on the return periods (due to the reason explained at point 4): on x-axis is the return period and on y-axis the loss produced, as a percentage of the sum insured.

9. *If it were possible to estimate, **the risk premium** would be calculated as the sum of various event sizes with their expected exceedance probabilities (the inverse of return periods). Using the loss charts designed at point 8, **the expected cost of flood (risk premium) can then be estimated by determining the area under that curve*** – see the above comments about the law of iterated expectations, weaker form.
10. The calculation of the area is a matter of basic mathematics or can be done using advanced mathematical tools on a computer: using basic mathematics, split the chart area in common and known geometric figures, for which the area formulas are well known (the method proposed is using only facilities provided by Excel: the tables are drafted in Excel, the charts are designed using these tables and Excel chart tab). Of course, using more advanced software, the calculation can be carried straightforward. One important remark: the length of one side of the geometric figure represents the return period (x-axis), but we are interested in the exceedance probabilities, which are the inverse of the return period. Great care should be taken on this point! Also, decomposing the area in common geometric figures can introduce some errors. But these errors are mitigated by a loading factor for errors when the tariff premium is established.
11. Adding up the areas of these simple geometric figures, one gets the risk premium (as a percentage of the sum insured).

We will demonstrate how to evaluate the cost of flooding, following the above steps.
We will also calculate a risk premium for (flash) flooding of Buildings. A similar exercise can be carried out for Contents or business interruption.

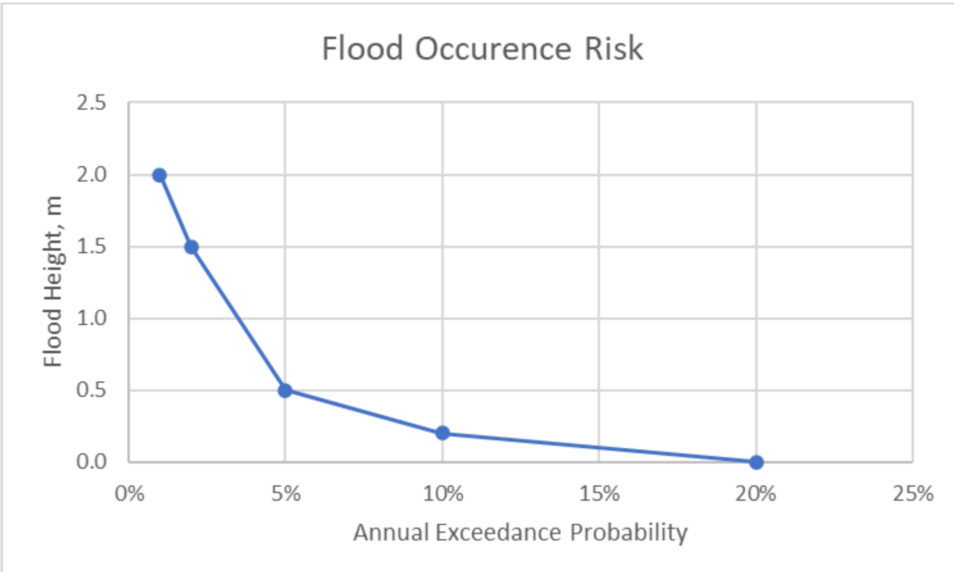
Note that all the examples given in this section are purely illustrative and are not necessarily indicative of the true cost of flooding risks.

For the property exposed to flooding risk, let us suppose that we have information on the expected flood height at various return intervals. This table, as an example, is obtained after the steps 1-5 above:

Return Period, years	Flood Height, m
5	0.00
10	0.20
20	0.50
50	1.50
100	2.00

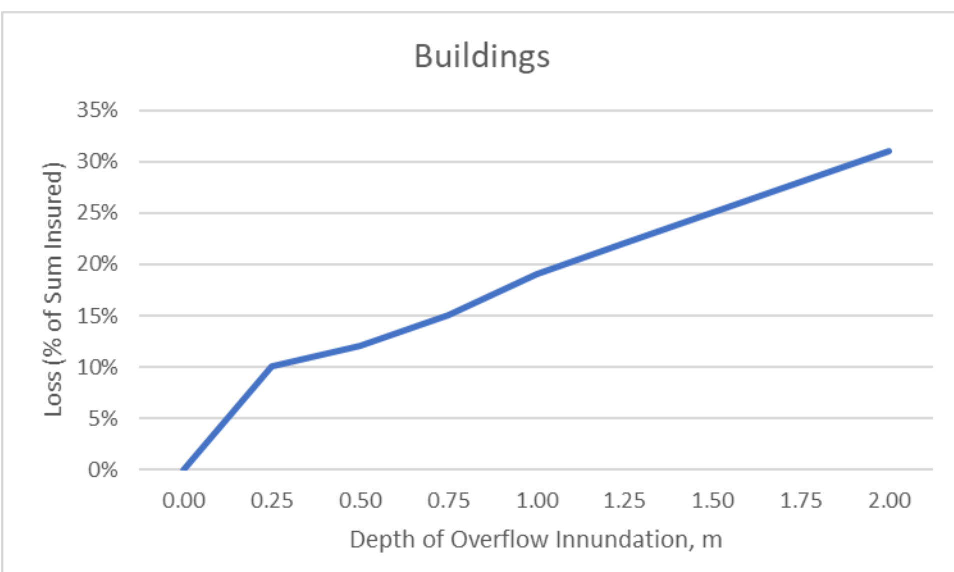
The inverse of the return period gives the *annual exceedance probability* – that is, the probability that flood heights will exceed a particular height. A curve is fitted to the data points above to determine the flood height at probabilities between the given data points, with a minimum height of zero. In this example, the following figure demonstrates this fit.

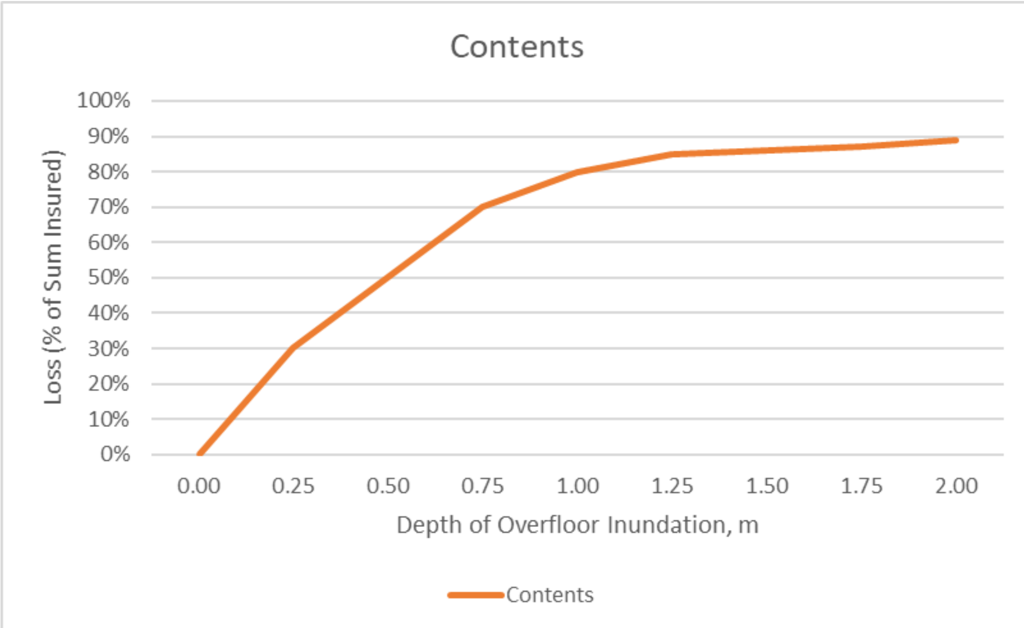
Step 6: define flood occurrence risk.



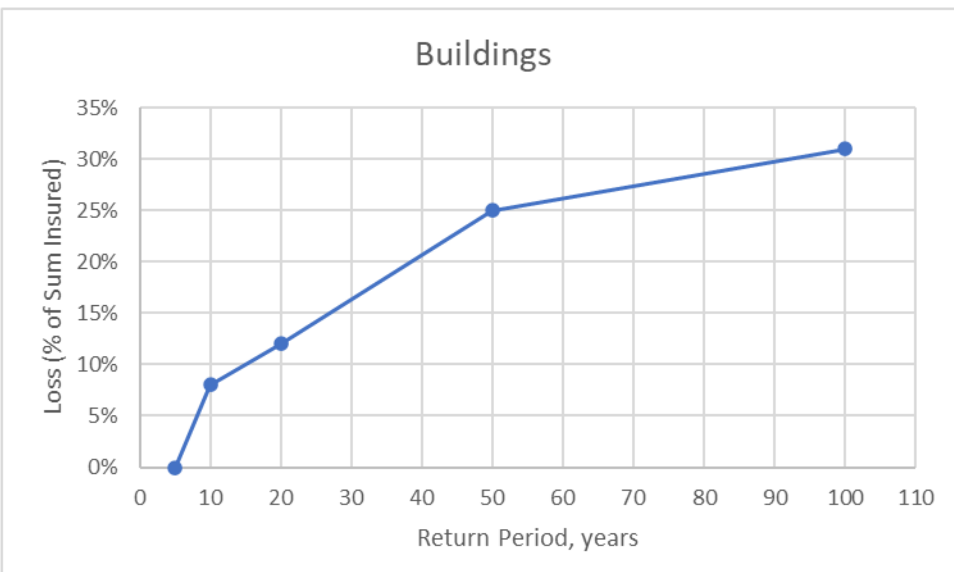
Step 7: loss curves, plotting the depth of overflow inundation and the damage produced, as a percentage of sum insured.

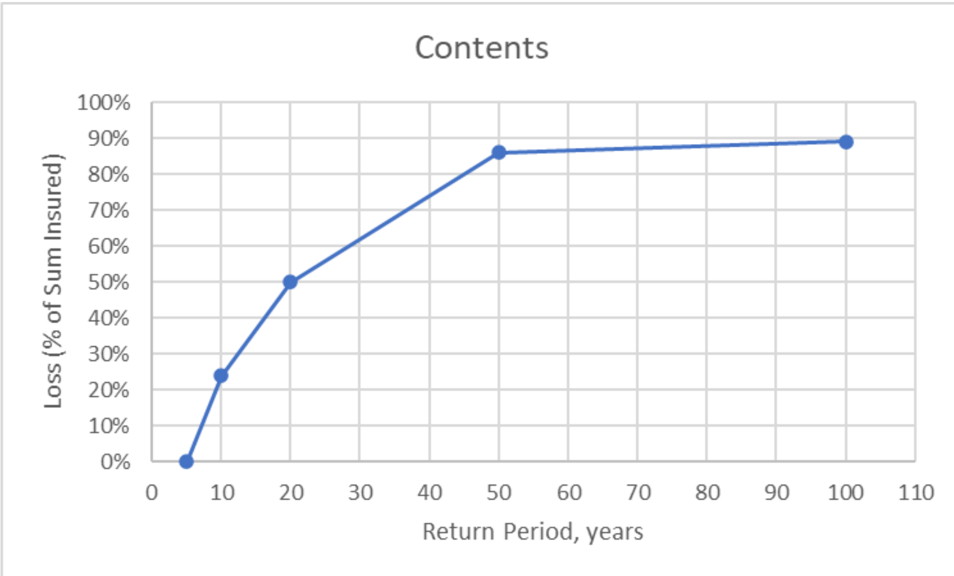
The Buildings and Contents loss curves which shows the damage incurred for differing flood heights. Illustrative curves for Buildings and Contents insurance are shown in the two charts below.





Step 8: it is straightforward to combine the fitted flood occurrence risk curve with the loss curve to obtain a ***loss curve*** by return period.





Step 9: if it were possible to estimate, the risk premium would be calculated as the sum of various event sizes with their expected exceedance probabilities (the inverse of return periods).

Using the above chart for Buildings, the expected cost of flooding (risk premium) can then be estimated by determining the area under that curve (similar for Contents). The chart area is quite simple, and the result is straightforward.

Step 10: for Buildings, the area under the corresponding curve includes one triangle and two trapezes:

- right-angle triangle area: $1/2 * (10 - 5)^{-1} * 8\% = 1/2 * 1/5 * 8\% = 0.5 * 0.2 * 8\% = 0.008 = 0.8000\%$
- 1st trapeze: $(8\% + 25\%)/2 * (50 - 10)^{-1} = 16.5\% * 1/40 = 16.5\% * 0.025 = 0.004125 = 0.4125\%$
- 2nd trapeze: $(25\% + 31\%)/2 * (100 - 50)^{-1} = 28\% * 1/50 = 28\% * 0.02 = 0.0056 = 0.5600\%$

Step 11: Total = 1.7725% or approx. 1.8% of the sum insured.

For a building insured for $1,000, the annual risk premium should be $18.

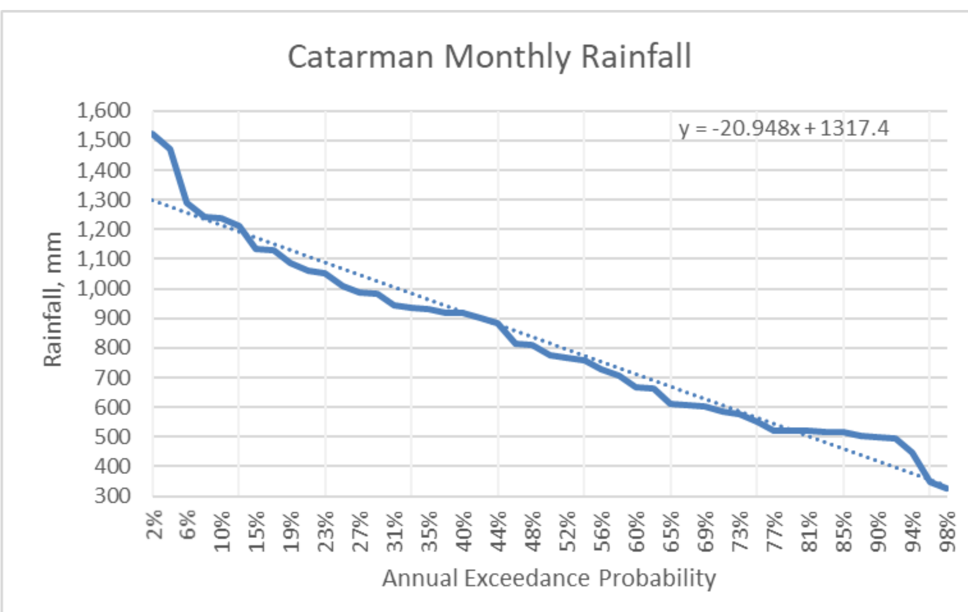
2.7 Flash Flooding in Catarman

Catarman Meteorological Station provides monthly rainfall data for the complete period 1972 - 2018, i.e., 564 monthly rainfall heights. These observations were registered at the meteorological station and contain information about the monthly rainfall heights at that location. Using these data, you can find statistics about the rainfall heights, including frequency and probability distributions, exceedance probabilities and return periods, but *only at the* *location of meteorological station*. Loss curves, for buildings and/or contents in different locations, cannot be designed, due to lack of data. More details about Catarman data can be found in [1].

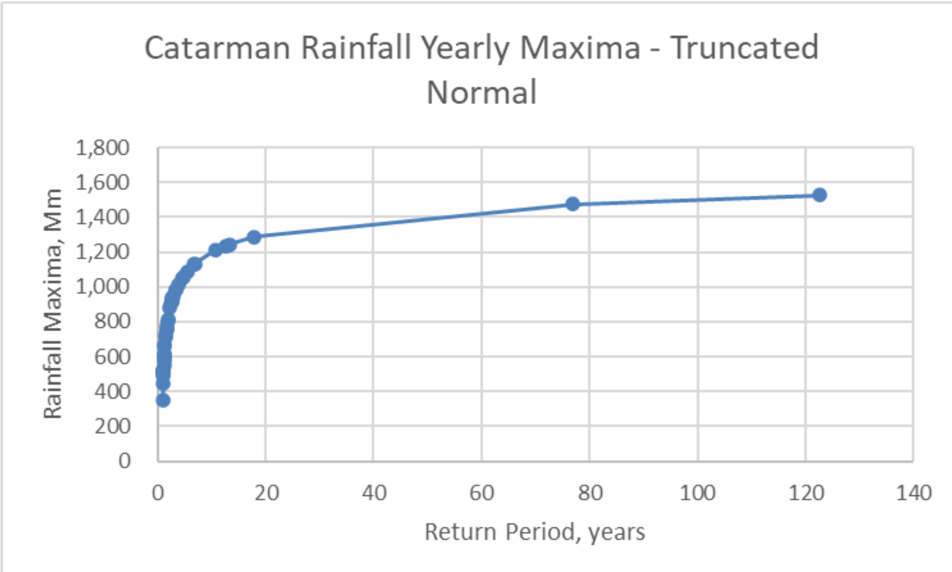
The largest frequency of monthly rainfall data, 84.4%, representing 476 monthly observations, are in the interval 0 - 500mm. It means that in these 476 months of observations (representing approx. 40 years out of 47 years of data) the monthly rainfall level was in the interval 0 – 500mm.
It is unlikely that a monthly rainfall height of 500mm will produce damages to households or businesses (500mm of monthly rainfall height means an average of 17mm rainfall height per day). But one-off event like one day of flash flooding of 500mm can have an impact on households and current activities in a specified location, depending on the quality of the drainage system in that location(s). Greater values, which are more relevant for the design of flash flooding insurance, were registered for only a few months: 5 values over 1,200mm represent just 0.88% of the whole sample.

The data provided allow to establish probability distributions of monthly rainfall and of annual maxima (47 years of annual maxima). Using these distributions, it is possible to determine exceedance probabilities and the corresponding return periods.

But unfortunately, there is no information which would allow designing loss curves, neither for buildings nor for contents. Designing these curves requires additional assumptions and underwriter's judgement.



Using the fit of *Normal distribution* to yearly maxima sample, the return period for different annual maxima can be obtained:



The difficulty is to establish loss curves for buildings and/or contents in different locations. As long as there are no such data, assumptions and models have to be used, because all information we have is about rainfall data (monthly rainfall height or yearly maxima) at the meteorological station.

We do not know:

- What is the depth of overflow inundation produced at a defined location in Catarman like downtown or suburbs, for instance, by a yearly maximum registered at meteorological station, assuming that this yearly maximum produced a flash flooding?
- What is the damage produced at that location to buildings/contents?

For instance, in the table below what we know is the information in first column and second column, return period and yearly maximum, respectively, registered at the meteorological station. The columns (3), (4) and (5) have to be filled in, using appropriate data, assumptions, and/or models.

Assuming that a yearly maximum registered at meteorological station was a flash flooding, estimation is necessary for the overflow inundation produced at the defined location and for the damages produced to buildings and/or contents.

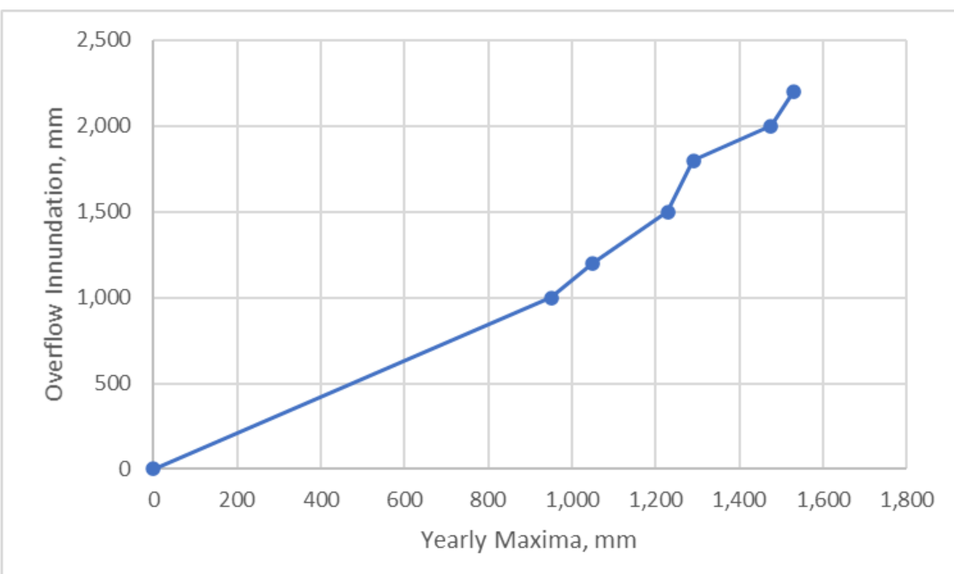
Return Period, years	Yearly Maxima, mm	Overflow Innundation, mm	Damage to Buildings, %SI	Damage to Contents, %SI
(1)	(2)	(3)	(4)	(5)
1	0			
3	950			
5	1,050			
13	1,230			
18	1,290			
77	1,475			
120	1,530			

The case can be more complicated: it is likely that not all the yearly maxima have produced flash floodings, and more details could reduce the occurrence of such floodings.
It is also true that the past data do not always represent a prediction for the future.

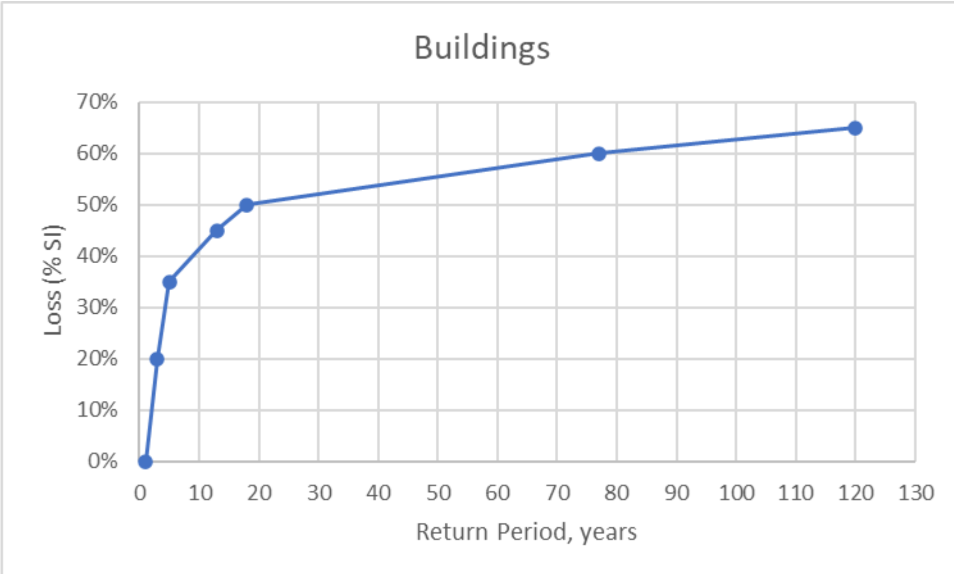
Taking into consideration these assumptions and trying to fill in the above table with reasonable values, based on pure intuition and for example purposes, we can have the following case:

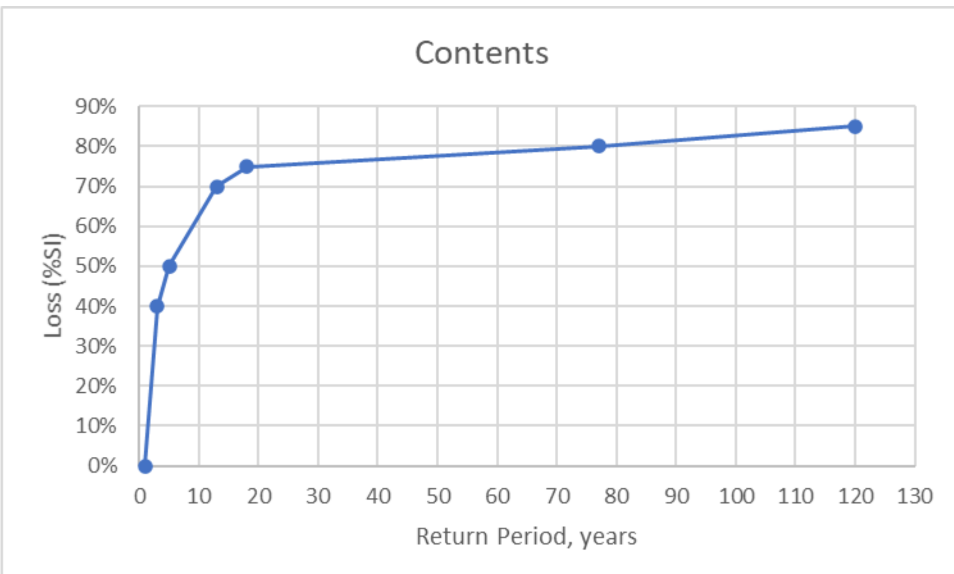
Return Period, years	Yearly Maxima, mm	Overflow Innundation, mm	Damage to Buildings, %SI	Damage to Contents, %SI
(1)	(2)	(3)	(4)	(5)
1	0	0	0%	0%
3	950	1,000	20%	40%
5	1,050	1,200	35%	50%
13	1,230	1,500	45%	70%
18	1,290	1,800	50%	75%
77	1,475	2,000	60%	80%
120	1,530	2,200	65%	85%

The influence of yearly maxima to overflow inundation is shown in the following chart:



The corresponding loss curve looks like:

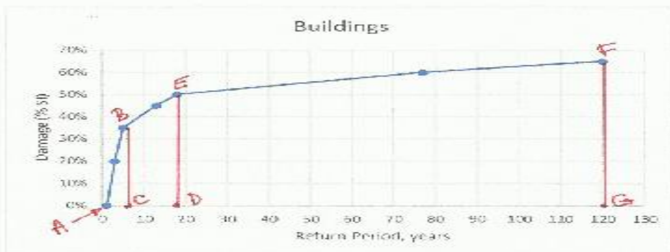


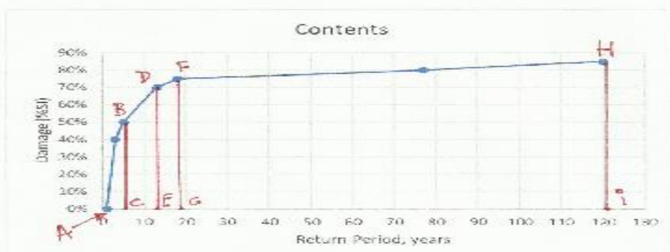


The cover provided for flash flooding is on indemnity basis, which means that if the flash flooding is produced, the insured person receives, usually, a percentage of the sum insured stated in the policy. No other costs are included in the premium calculation, like replacement costs, repairs, or accommodation costs during the repairs, evaluated and calculated by a loss adjuster, based on the actual loss.

The expected cost of flooding can then be estimated by determining the area under each of these curves. The areas under each of these curves are more complicated than in the previous theoretical example, more details are needed.

Detailed calculation is provided for Buildings, similar result can be obtained for Contents.





Calculation for Buildings

The area has been split into 3 classic geometric figures:

- Right-angle triangle ABC.
- Trapeze BCDE.
- Trapeze DEFG.

We need the coordinates [return period, %SI] for each of these points:

A(1,0), B(5, 35%), C(5, 0%), D(18, 0%), E(18, 50%), F(120, 65%), G(120, 0%)

Area of right-angle triangle ABC = 1/2 * AC * BC = 1/2 * $(5 - 1)^{-1}$ * (35% − 0%) = 1/2 * 1/4 * 35% = 0.043750

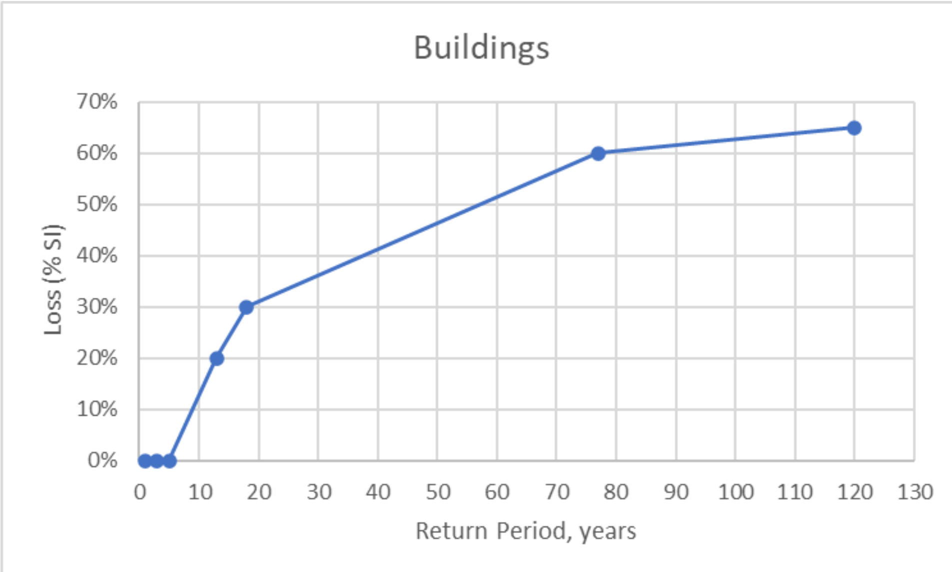
Area of trapeze BCDE = 1/2 *(BC + DE) * DC = 1/2 * (35% + 50%) * $(18 - 5)^{-1}$ = 1/2 * 85% * 1/13 = 0.032692
Area of trapeze DEFG = 1/2 * (DE + FG) * GD = 1/2 * (50% + 65%) * $(120 - 18)^{-1}$ = 1/2 * 115% * 1/102 = 0.005637
Total area = 0.082079 or 8.2%.

This (high) value is mainly due to low return periods of the right-angle triangle ABC and trapeze BCDE (low return period means high exceedance probabilities, likely to happen in Catarman, which is a rainfall zone). The trapeze DEFG has an insignificant contribution to the total area, only 0.56%, due to the large return period of 102 years.

The risk premium shows the high risk of flash flooding in Catarman zone, which is exposed to frequent rainfalls of different magnitudes. An insurer willing to underwrite the risk of flash flooding in Catarman has to be aware of the risk magnitude and has to find appropriate loss curves for the corresponding risk and locations (don't forget, this was just a theoretical example, the real world will be different).

Another case is considering flood levels with return periods higher than 13 years, with nil damages produced to buildings for short return periods:

Return Period, years	Yearly Maxima, mm	Overflow Innundation, mm	Damage to Buildings, %SI
(1)	(2)	(3)	(4)
1	0	0	0%
3	950	1,000	0%
5	1,050	1,200	0%
13	1,230	1,500	20%
18	1,290	1,800	30%
77	1,475	2,000	60%
120	1,530	2,200	65%



Applying the same method as above, the area under the curve is 1.46%, which means that for a sum insured of 1,000 (USD or PHP), the annual risk premium is 14.6 (USD or PHP).

2.8 Tariff or Commercial Premium

Regardless of how it is determined, once a flood level is retrieved, the impact on the premiums is determined by the underlying rating formula in the insurer's pricing methodology, which should include the rating factors defined by the underwriter.

Normally this formula is a function of a number of rating and loading factors. Some of these factors apply to *basic rate* of risk premium (L1, L2, L3), the others to *tariff premium* (L4 – L9). The rating and loading factors can be positive or negative (discounts).
Ignoring other risk perils, we can present this form as follows, for flood risk in the context of home insurance for Buildings cover or for business interruption:

Tariff Premium = **Basic Rate of Risk Premium**
x rating factor for location adjustment, L1
x rating factor for construction type, L2
x rating factor for risk reduction, L3 (see Community Rating System)
+ Tariff Premium
x loading for errors, L4
x loading for profit, L5
x loading for cost of reinsurance, L6
x loading for cost of capital, L7
x loading for commission, L8
x loading for administrative costs, L9

The formula for tariff premium will be:

$$Tariff\ Premium = Basic\ Rate\ of\ Risk\ Premium * \frac{(1 + L1)(1 + L2)(1 + L3)}{1 - L4 - L5 - L6 - L7 - L8 - L9}$$

There would be a different version of this formula for Contents insurance.

Of course, further interactions of the above variables may be appropriate and other minor factors may also be introduced. Typically, post-event analyses can be used to determine the impact of the other rating factors for the floods that have been observed.

Whilst the specific focus of the function above was on just the flood premium, often it is not possible to separate this calculation from the broader rating calculation. Indeed, it is often the case that flood can simply be included as a 'loading' on the rest of the premium (in case of bundling products, for instance), where the loading is dependent on the flood level retrieved in a manual or automatic process.

It will become important for an insurer to have the ability to efficiently *geo-code* risks whenever a quote is required in order for flood risk rating to occur. Without this ability, an insurer would have to aggregate risks up to a higher level, which would reduce the degree of differentiation for flood risk between individual properties. Acquiring the ability to geo-code risks at point of sale could be a significant hurdle for some insurers who may not be able to afford the infrastructure required. Also, the insurer has to consider the difficulties of implementation such changes over a number of IT systems.

2.9 Definition of Cover

In theory, insurers' decision as to when they would market coverage for a specific risk is based on the possibility of maximizing expected profits subject to satisfying constraints related to the survival of the company and solvency requirements. It is also important to consider the stability of the insurer's operations. However, insurers have traditionally not focused on these constraints in dealing with catastrophic risks.

Following the disasters in latest years and regulatory requirements, insurers focused on the survival constraint in determining the amount of catastrophe coverage they wanted to provide. Moreover, insurers were caught off guard with respect to the magnitude of the losses from typhoon Haiyan (Yolanda) and other (similar) extremal events. In conjunction with the losses that resulted from these disasters, the demand for coverage increased.

Whilst the positions taken by competitors will influence the approach, each individual company will weigh the considerations differently. The actual cover being offered, impact of pricing on existing and potential new customers (and overall growth), methods of distribution, portfolio goals (reduce exposure to catastrophes vs price for risk vs some other goals), and the nature and sophistication of pricing and administration systems all contribute to the appropriate outcome for the class of business in question for a company.

On top of these considerations, the approach to pricing depends on the insurer's view of the data available and its fitness for the purpose. The fact that a 'perfect' pricing model is difficult, if not impossible to achieve, means that actuaries approaching this problem need to be prepared to apply a degree of pragmatism. Designing an approach that allows a limited amount of manual intervention will also improve the chances of a successful implementation. However, a preparedness to adjust the pricing soon after it first goes live, in response to any unforeseen issues, is still required.

As with any pricing exercise, once the objective is established it is critical to determine exactly what is being priced. In the case of flood cover, this is particularly tricky due to interplay with other perils such as storms or typhoons.

Where 'full' flood cover is optional, it is still common to have 'flash flooding' (perhaps with a % sum insured limit) included in the standard cover. In these situations, separating the pricing into the two components can be challenging, as most modelling does not distinguish these types of events.

Whilst modelling enhancements could be envisaged to tackle the two problems separately, there is a remarkably high correlation between them. Indeed, even if flash flood is defined broadly as 'connected' flood if the rain falls within some time period of the inundation (commonly 24 hours), there are situations where the distinction is meaningless.

Determining the exact cost of riverine flood is complicated by the interplay with other perils such as storms and typhoons. For example, assume that a large upstream flood takes some time to travel down to a town and that around the time the flood peak hits, it also rains on the town. Separating the loss into that pertaining to the top 'millimeters' of water, and the rest below will create claims problems and legal issues. A solution for insurers would be to pay these claims ex-gratia.

Where the cover is optional, it may be necessary to place an additional loading on the cover due to the information asymmetry that exists – people who know they are exposed to flood risks will buy the cover. This is particularly relevant in locations where it is believed by the insurer that little or no risk of flood inundation exists.

A distinction has to be made between flash flood associated with storms or typhoons and other severe rain activities (Type I). The other possibility refers to floods associated with a regular water course (Type II).

As many policy wordings consider Type I flooding to be part of the standard 'storm' cover (water off the ground), it is important to understand what would be embedded in observed claims data, to ensure that there is not a double count with the pricing of the main storm costs.

Some flooding cost will also be buried in normal storm claim data due to the difficulty distinguishing 'water coming down' from 'water coming up'. For example, when storms cause roofing damage, thereby leading to rain penetration from above, it is common for inundation to also occur. It is seldom the concern of the claim officers to distinguish this whilst they are processing large volumes of claims. So again, care must be taken in determining exactly what the scope of the pricing exercise is, and to ensure that this does not lead to an inadvertent double count.

Key challenges that exist in determining and maintaining the customer prices include:

- Ensuring that, beyond the initial pricing work, efforts are made to encourage mitigation or avoidance of flood risk in the first place.
- Educating the market about the extent of cover and what they are paying for.
- Establishing the appropriate monitoring processes to assess the impacts of the pricing approach on the mix of business.

2.10 Reinsurance Costs

Usually, the reinsurers indicated that they would support riverine flooding coverage or flash flooding as long as the direct insurer can demonstrate a sound pricing structure and as a result, are charging sufficient premium for the risk.

Much of the cost of flood lies in the extremal events, and as such a large proportion of the cost borne by insurers would ultimately be paid for by reinsurers and reflected in the premiums they charge.

One challenge that flows from the significant costs being borne by reinsurers is the allocation of reinsurance premiums for pricing. To put this challenge in context, it is interesting to find out how successful insurers have been in allocating the costs of reinsurance to properties more exposed to typhoon risk (for example, those closest to the coast).

2.11 Mitigation Activities

It is also important for the insurance industry to participate in the discussion about flood mitigation. This includes actively monitoring new housing developments and discouraging developments in flood prone land. Much of this is at local government level, so it should be important for flood insurance provider(s) in a specific area to work actively with the local government for proper management of flood risk (local Disaster Risk Management – DRM activities).

Encouraging mitigation works by quickly responding and reducing premiums once works are completed will also provide an incentive for the local residents to lobby government (both local and state) to provide funding for mitigation works in areas where a problem already exists.

2.12 Community Rating System

The Community Rating System is a voluntary incentive program that recognizes and encourages community floodplain management activities. Usually, this incentive program is part of a national flood insurance program (like in the US, where this model is applied).

As a result, flood insurance premium rates are discounted to reflect the reduced flood risk resulting from the community actions meeting the three goals of the Community Rating System:

- Reduce flood damage to insurable property.
- Strengthen and support the insurance aspects of the national flood insurance program, and
- Encourage a comprehensive approach to floodplain management.

As an example of such program (US program), for Community Rating System participating communities, flood insurance premium rates are discounted in increments of 5 percent. Class 10 is not participating in the Community Rating System and receives no discount. Class 9 community would receive a 5 percent discount to a Class 1 community which would receive a 45 percent premium discount.

The Community Rating System classes for local communities are based, in this example, on 18 creditable activities, organized under four categories:

1. Public Information
2. Mapping and Regulations
3. Flood Damage Reduction; and
4. Flood Preparedness.

As an example, the table below shows the credit points earned, classification awarded, and premium reductions given for communities in a Community Rating System:

Credit Points	Class	SFHA* Premium reduction, %	Non SFHA* Premium Reduction, %
4,500+	1	45	10
4,000 – 4,499	2	40	10
3,500 – 3,999	3	35	10
3,000 – 3,499	4	30	10
2,500 – 2,999	5	25	10
2,000 – 2,499	6	20	10
1,500 – 1,999	7	15	5
1,000 – 1,499	8	10	5
500 – 999	9	5	5
0 – 499	10	0	0

SFHA = Special Flood Hazard Area

For instance, with a risk premium of 8.2% for a special flood hazard area and Class 1 of community rating, 45% discount means that the risk premium will be reduced to 4.5%.
The premium reduction for non-SFHA is lower because the premiums for this area are already lower than the premiums for SFHA.

More details about Community Rating System can be found on

https://www.fema.gov/national-flood-insurance-program-community-rating-system

The company's pricing philosophy may be more one of community rating or cross-subsidization. A company that chooses to rate this way, however, needs to monitor the mix of business to ensure that the subsidy is funded. As discussed above, a more community-rated approach may be desirable for existing customers in order to limit the loss of businesses. Where a company previously had an element of flash flood cover, there might already be an implicit level of community rating in the existing rates.

The market share objectives of the company are therefore important to understand in this process, as the decision on pricing approach can materially impact market share, especially if the approach deviates from market treatment. Further related to market share is the *market image* of the company.

A company that deviates from market practice or is seen to 'turn its back' on some areas or customers, may see brand deterioration which ultimately manifests itself in market share reductions.

2.13 Actuarial Control Cycle

Beyond the task of pricing for flooding, it is vital that appropriate monitoring systems be set up to understand the mix of business, and the degree of increase or decrease in flood exposure.

This monitoring would not stop a few months after initial implementation. Indeed, over the longer term it is even more important as competitors adjust their strategy, or weaknesses in the control framework manifest themselves.

The portfolio outcomes from the pricing implemented should approximately match the original objectives that were set. Deviation from this would of course result in re-work, in the traditional control cycle framework.

2.14 Parametric Insurance

The basic concept of parametric solutions is quite simple: traditionally, instead of indemnifying for the actual loss incurred, parametric insurance covers the probability of a predefined event happening (e.g., a major typhoon, earthquake, or flash flooding, in our case), and pays out according to a predefined scheme. Events may refer to an index-based trigger (e.g., crop shortfall or flash flooding level) or an event within a defined area. Exemplifying this concept, a policy might be structured to pay out 50%, 75%, or 100% of a predefined limit for a typhoon category 3, 4, or 5, respectively, happening within a defined radius around the client's address. Or, for riverine flooding, a policy might be structured to pay predefined percentages of sum insured for riverine flooding happening within a predefined area. Similar with flash flooding, the policy may be structured to pay 30%, 75% or 100% of the sum insured, depending on the overflow inundation.

Bibliography

[1] GIZ and MEFIN NETWORK – *Statistical Analysis on Rainfall Data in Catarman Province, Philippines*

[2] Spiegelhalter, D. − *The Art of Statistics (Learning from Data)*, Penguin Random House UK, 2020

[3] Embrechts, P. and Kluppelberg, C, Mikosch,T. − *Modelling Extremal Events*, Springer, New York, 2012

[4] Hogg, R. and Klugman, S. − *Loss Distributions*, John Wiley & Sons, 1984

[5] Hossack, I.B. and Pollard J.H., Zehnwirth B. − *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, 1983

Annex

Definition (generalized inverse of a monotone function)

Suppose *h* is a non-decreasing function on R. The generalized inverse of *h* is defined as:

$$h^{\leftarrow}(t) = \inf\{x \in R : h(t) \geq t\}$$

Definition (tail of the distribution function (df))

$$P(X > x) = \bar{F}(x) = 1 - F(x)$$

where *P* stands for probability.

Definition (quantile function)

The generalized inverse of the df F

$$F^{\leftarrow}(t) = \inf\{x \in R : F(t) \geq t\}$$

is called the quantile function *of df F. The quantity* $x_t = F^{\leftarrow}(t)$ *defines the* t-quantile *of F.*

Order Statistics

Let $X_1, X_2, \dots .. X_n$ denote a sequence of independent and identically distributed (iid) random variables (rvs) with common df *F*.

We define the *order sample:*

$$X_{n,n} \leq \cdots \leq X_{k,n} \leq \cdots X_{1,n}$$

Here

$$X_{n,n} = \min(X_1, X_2 \dots, X_n), \quad X_{1,n} = M_n = \max(X_1, X_2, \dots, X_n)$$

The rv $X_{k,n}$ is called the *k-th upper order statistic* - details in [3].

The exact df of the maximum M_n is:

$$P(M_n \leq x) = P(X_1 \leq x, \dots X_n \leq x) = F^n(x)$$

Extreme values occur "near" the upper end of the support of the distribution, hence intuitively the asymptotic behaviour of M_n must be related to the df *F* at its right tail near the right end point. We denote by:

$$x_F = \sup\{x \in R: \ F(x) < 1\}$$

the *right end point* of *F*. We immediately obtain, for all $x < x_F$,

$$P(M_n \leq x) = F^n(x) \to 0, \qquad n \to \infty$$

And, in the case $x_F < \infty$, we have for $x \geq x_F$ that:

$$P(M_n \leq x) = F^n(x) = 1$$

Thus $M_n \overset{P}{\rightarrow} x_F \; as \; n \rightarrow \infty, where \; x_F \leq \infty$ (convergence in probability). Since the sequence (M_n) is non-decreasing in *n*, it converges *almost sure* (a.s.) and we conclude that:

$$M_n \overset{a.s.}{\rightarrow} x_F$$

This fact does not provide a lot of information. More insight into the order of magnitude of maxima is given by weak convergence results for centred and normalized maxima. This is one of the main topics of classical extreme value theory (see the fundamental Fischer-Tippett theorem).

The relationship between the order statistics and the *empirical df* or *sample df*:

$$F_n(x) = \frac{1}{n} card\{i : 1 \leq i \leq n, X_i \leq x\} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}}, \qquad x \in R$$

where I_A stands for the *indicator function* of the set *A*.

Now

$$X_{k,n} \leq x \; if \; and \; only \; if \; \sum_{i=1}^{n} I_{\{X_i \geq x\}} < k$$

Which implies that:

$$P(X_{k,n} \leq x) = P(F_n(x) > 1 - \frac{k}{n})$$

where *P* is the probability, as before.

Upper order statistics estimate tails and quantiles, and also excess probabilities over certain thresholds. For a sample $X_1, X_2, \ldots \ldots X_n$ we denote *the empirical quantile function* by $F_n^{\leftarrow}$. If *F* is a continuous function, we may assume $X_{n,n} < \cdots < X_{1,n}$. In this case $F_n^{\leftarrow}$ is a simple function of the order statistics, namely:

$$F_n^{\leftarrow}(t) = X_{k,n} \; for \; 1 - \frac{k}{n} < t \leq 1 - \frac{k-1}{n}$$

and for *k = 1, …, n*.

For the case $t = 1 - \frac{k-1}{n}$ we obtain $F_n^{\leftarrow}\left(1 - \frac{k-1}{n}\right) = X_{k,n}$.

The concept of *quantile transformation* is extremely useful since it often reduces the problem concerning order statistics to one concerning the corresponding order statistics from a *uniform sample*.

Lemma (Quantile transformation)

Let $X_1, X_2, \dots\dots\dots X_n$ be iid random variables with distribution function *F*. Furthermore, let $U_1, U_2, \dots\dots\dots U_n$ be iid random variables uniformly distributed on (0,1) and denote by $U_{n,n} < \cdots < \cdots U_{1,n}$ the corresponding order statistics. Then the following results hold:

(a) $F^{\leftarrow}(U_1) \overset{d}{=} X_1$ (have the same distribution)

(b) For every $n \in N$

$$(X_{1.n,} \dots \dots \dots \dots X_{n,n}) \overset{d}{=} (F^{\leftarrow}(U_{1,n}), \dots \dots \dots \dots F^{\leftarrow}(U_{n,n}))$$

(c) The random variable $F(X_1)$ has a uniform distribution on (0,1) if and only if *F* is a continuous function.

The lemma on quantile transformation implies that for *F* continuous, the random variables $U_i = F(X_i)$, for $i = 1, 2\dots, n$ are iid uniform on (0,1). Moreover,

$$(F(X_{k,n}))_{k=1\dots\dots n} \overset{d}{=} (U_{k,n})_{k=1\dots\dots n}$$

From this it follows that the expected value of $F(X_{k,n})$ is:

$$EF(X_{k,n}) = \frac{n-k+1}{n+1}, k = 1, \dots., n$$

It is important to note that:

$$F_n(X_{k,n}) = \frac{n-k+1}{n}$$

where F_n stands for the empirical df of *F*.

The graph

$$\{\left(F(X_{k,n}), \frac{n-k+1}{n+1}\right), k = 1 \dots \dots n\}$$

is called a *probability plot (PP plot).*

More common is to plot the graph:

$$\{(X_{k,n}, F^{\leftarrow}\left(\frac{n-k+1}{n+1}\right)), k = 1 \dots \dots n\}$$

referred to as *the quantile plot (QQ plot).*

In both cases, the approximate linearity of the plot is justified by Glivenko-Cantelli theorem.

There exist various variants of QQ plot defined above of the type:

$$\{(X_{k,n} F^{\leftarrow}(p_{k,n})), k = 1 \dots \dots n\}$$

where $p_{k,n}$ is a certain *plotting position.* Typical choices are:

$$p_{k,n} = \frac{n - k + \delta_k}{n + \gamma_k}$$

with (δ_k, γ_k) appropriately chosen allowing for some *continuity correction* - details in [3].

We shall mostly consider:

$$p_{k,n} = \frac{n - k + 0.5}{n}$$

QQ Plot for Gumbel Distribution

For Gumbel extreme value distribution

$$\Lambda(x) = \exp\{-e^{-x}\}, x \in R$$

assume that we want to test whether the sample $X_1, X_2, \ldots\ldots\ldots\ldots X_n$ comes from Λ. For this purpose, we take the ordered sample and plot $X_{k,n}$ (more precisely, the *k*th largest observation $x_{k,n}$) against

$$\Lambda^{\leftarrow}(p_{k,n}) = -\ln\,(-ln(p_{k,n}))$$

where $p_{k,n}$ is a plotting position defined above.

The QQ plot for Gumbel extreme value distribution is the graph:

$$\{(X_{k,n};\ -\ln\,(-\ln(p_{k,n})), k = 1 \ldots\ldots n\}$$

If the Gumbel extreme value distribution provides a good fit to our data, then the QQ plot should look *roughly linear* - details in [3].

Fischer-Tippett theorem

The fundamental *Fischer-Tippett theorem* (limit laws for maxima) has the following content:

Let (X_n) be a sequence of iid random variables. If there exist constants $c_n > 0\ and\ d_n \in R$ such that

$$c_n^{-1}(M_n - d_n)\ converges\ in\ distribution\ to\ H$$

for some non-degenerate distribution *H*, then *H* must be of the type of one of the three so called *standard extreme value distributions* (Fréchet, Weibull, Gumbel):

Fréchet: $$\Phi_{\propto}(x) = \begin{Bmatrix} 0, & x \leq 0 \\ \exp\{-x^{-\propto}\} & x > 0,\ \propto > 0 \end{Bmatrix}$$

Weibull: $$\Psi_{\alpha}(x) = \begin{Bmatrix} \exp\{-(-x)^{-\alpha}\} & x \leq 0, \propto > 0 \\ 1 & x > 0 \end{Bmatrix}$$

Gumbel: $$\Lambda(x) = \exp(-e^{-x}), x \in R$$

Definition (extreme value distribution and extremal rv)

The dfs Φ_α, Ψ_α and Λ as presented in the Fischer-Tippett theorem are called *standard extreme value distributions,* the corresponding rvs are *standard extremal rvs*. Dfs of the types Φ_α, Ψ_α and Λ are *extreme value distributions*; the corresponding rvs *extremal rvs*.

Consequently, we should consider probabilities of the form:

$$P(c_n^{-1}(M_n - d_n) \leq x), which\ can\ be\ written\ as\ P(M_n \leq u_n), u_n(x) = c_n x + d_n$$

u_n can be considered a threshold and $P(M_n > u_n) = 1 - P(M_n \leq u_n)$ the probability of exceeding this threshold - details in [3].

Definition (maximum domain of attraction)

The rv *X* (or the df *F* of *X*) belongs to the *maximum domain* of *attraction* of the extreme value distribution *H* if there exist constants $c_n > 0\ and\ d_n \in R$ such that:

$$c_n^{-1}(M_n - d_n) \overset{d}{\to} H$$

We write $F \in MDA(H)$.

The standard extreme value distributions, together with their types, provide the only non-degenerate limit laws for affinely transformed maxima of iid rvs. The three standard cases can be represented in one family of dfs. They can be represented by introducing a parameter ξ so that:

$$\xi = \alpha^{-1} > 0\ corresponds\ to\ Fréchet\ distribution\ \Phi_\alpha$$

$$\xi = 0\ corresponds\ to\ Gumbel\ distribution\ \Lambda$$

$$\xi = -\alpha^{-1} < 0\ corresponds\ to\ Weibull\ distribution\ \Psi_\alpha$$

Definition (representation of the extreme value distributions: the generalized extreme value distribution (GEV))

Define the df H_ξ by:

$$H_\xi(x) = \begin{Bmatrix} \exp\left\{-(1+\xi x)^{-1/\xi}\right\}\ if\ \xi \neq 0 \\ \exp\{-\ exp\{-x\}\}\ \ if\ \xi = 0 \end{Bmatrix}$$

where $1 + \xi x > 0$.

H_ξ is called a *standard generalized extreme value distribution (GEV).*

We consider the df H_0 as the limit of H_ξ for $\xi \rightarrow 0$. With this interpretation:

$$H_\xi(x) = \exp\left\{-(1+\xi x)^{-1/\xi}\right\}, 1 + \xi x > 0$$

serves as a representation for all $\xi \in R$.

The support of H_ξ corresponds to:

$$x > -\xi^{-1} \; for \; \xi > 0$$

$$x < -\xi^{-1} for \; \xi < 0$$

$$x \in R \; for \; \xi = 0$$

The GEV provides a convenient and unifying representation of the three value types Fréchet, Weibull, Gumbel; its introduction is motivated by statistical applications.

Theorem (characterization of MDA(H_ξ)

For $\xi \in R$ the following assertion are equivalent:

(a) $F \in MDA(H_\xi)$
(b) There exists a positive, measurable function a(·) such that for 1 + ξ x > 0

$$\lim_{u \uparrow x_F} \frac{\bar{F}\left(u + xa(u)\right)}{\bar{F}(u)} = \begin{Bmatrix} (1+\xi x)^{-1/\xi} & if \;\; \xi \neq 0 \\ e^{-x} & if \;\; \xi = 0 \end{Bmatrix}$$

(c) For x, y > 0, $y \neq 1$

$$\lim_{s \to \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{Bmatrix} \dfrac{x^\xi - 1}{y^\xi - 1} & if \quad \xi \neq 0 \\ \dfrac{\ln x}{\ln y} & if \quad \xi = 0 \end{Bmatrix}$$

where $U(t) = F^{\leftarrow}(1 - t^{-1})$.

Remark 1

For x > 0, condition (b) of the above theorem has a probabilistic interpretation: let *X* be a rv with df $F \in MDA(H_\xi)$, then we can write:

$$\lim_{u \uparrow x_F} P\left(\frac{X-u}{a(u)} > x \middle| X > u\right) = \begin{Bmatrix} \left((1+\xi x)^{-1/\xi}\right) & if \quad \xi \neq 0 \\ e^{-x} & if \quad \xi = 0 \end{Bmatrix}$$

which gives a distributional approximation for the scaled *excesses over* the (*high*) *threshold u*. The appropriate scaling factor is *a(u)*.

Definition (Excess distribution function, mean excess function)

Let *X* be a rv with df *F* and right end point x_F. For a fixed $u < x_F$:

$$F_u(x) = P(X - u \leq x | X > u), \qquad x \geq 0$$

is the *excess df* of the rv *X* (of the df *F*) over the threshold *u*. The function

$$e(u) = E(X - u | X > u)$$

is called the *mean excess function* of X (E stands for expected value). Using the definition of $e(u)$ and partial integration, it is easy to calculate the mean excess function for many dfs.

Remark 2

Excesses over thresholds play a fundamental role in many fields. Different names arise from specific applications. For instance, F_u is known as *excess-life* or *residual lifetime* df in reliability theory and medical statistics. In an insurance context, F_u is usually referred to as the *excess-of-loss* df.

The (characterization of MDA(H_ξ) provides the following definition:

Definition (The generalized Pareto distribution (GPD))

Define the df G_ξ by:

$$G_\xi(x) = \begin{Bmatrix} 1-(1+\xi x)^{-1/\xi} & if\ \xi \neq 0 \\ 1-e^{-x} & if\ \ \xi = 0 \end{Bmatrix}$$

where:

$$x \geq 0\ \ if\ \xi\ \geq 0$$

$$0 \leq x \leq {}^{-1}/_{\xi}\ \ if\ \xi < 0$$

G_ξ is called a standard *generalized Pareto distribution* (GPD).

As in the case of H_0 we consider the df G_0 as the limit of G_ξ for $\xi \rightarrow 0$. We shall denote:

$$G_{\xi,\beta}(x) = 1-(1+\xi\frac{x}{\beta})^{-1/\xi},\ \ x \in D(\xi,\beta)$$

where $x \in D(\xi,\beta) = \begin{Bmatrix} [0,\infty) & if\ \xi \geq 0 \\ \left[0, {}^{-\beta}/_{\xi}\right] & if\ \xi < 0 \end{Bmatrix}$

Summary

The GEV H_ξ, for $\xi \in R$ describes the limit distributions of normalized maxima.
The GPD $G_{\xi,\beta}$, $\xi \in R$, $\beta > 0$ is the limit distribution of scaled excesses over high thresholds.

Glivenko- Cantelli Theorem (without proof)

If $F(x)$ is the theoretical distribution function and $F_n(x)$ is the empirical distribution function of an independent and identically distributed (iid) sample X_1, X_2, ... X_n, then:

$$P\left[\underbrace{sup}_{x}|F_n(x)-F(x)| \rightarrow 0\right] = 1 \text{ (}P\text{ stands for probability)}$$

It is one of the fundamental results in non-parametric statistics (see details in [3]).

Definition (limited expected value function)

If X has a df F and a pdf function f whose support is $0 < x < \infty$, the *limited expected value function* of F is:

$$E(X;d) = \int_0^d xf(x)dx + d \cdot [1 - F(d)] = \int_0^d xf(x)dx + d \cdot \bar{F}\,(d)$$

where $\bar{F}$ is the tail of df F.

Published by
Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn, Germany

Regulatory Framework Promotion of Pro-poor Insurance Markets in Asia (RFPI Asia)
10th Floor Bank of Makati Building
Ayala Avenue Extension near corner Metropolitan Avenue
1029 Makati City
Philippines
T+63 2 8651 5164
https://www.giz.de/en/html/index.html
https://www.inclusiveinsuranceasia.com/index.html

As of
February 2021

Design and Layout
Alexander Tabbada
Sunshine Gervacio

Photo credit
The World Bank Photo Collection

Author
Tudor Moldovan

Editors
Dr. Antonis Malagardis
Dante Portula

URL links
https://mefin.org/index.html
https://www.inclusiveinsuranceasia.com/index.html

GIZ is responsible for the content of this publication.

On behalf of the
German Federal Ministry for Economic Cooperation and Development (BMZ)

Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH
Regulatory Framework Promotion of Pro-poor
Insurance Markets in Asia (RFPI Asia III)
10th Floor, Bank of Makati Building,
Ayala Avenue Extension near corner Metropolitan Avenue
1209 Makati City, Philippines

T: +632 2 8651 51 64
E: antonis.malagardis@giz.de
I: https://mefin.org/index.html
https://www.inclusiveinsuranceasia.com/index.html