

# Case Study: Heavy-Tailed Distribution and Reinsurance Rate-making

October 28, 2016

The purpose of this case study is to give a brief introduction to a heavy-tailed distribution and its distinct behaviors in contrast with familiar light-tailed distributions in standard texts. You will learn about QQ-plot, which is a popular tool for checking goodness-of-fit for a particular statistical model. You will also work on a real-life application of heavy-tailed distributions in reinsurance rate-making.

Reinsurance is a very important component of the global financial market. It allows insurers to take on risks that they would otherwise not be able to. Did you know that NASA buys insurance contracts for every rocket it launches and every satellite and probe it sends to the outer space? These equipments are so expensive that typical insurers would not be able to cover on their own. Therefore, they can go to the reinsurance market, slide up the coverage and transfer partial coverage to reinsurers that exceed their financial capabilities. By the end of this case study, you will be able to learn basic principles of pricing an reinsurance contract.

## **Learning Objectives:**

- Visualize the concepts in the Central Limit Theorem;
- Identify cases where the Central Limit Theorem does not apply;
- Reinforce the concept of cumulative distribution function;
- Understand why and how QQ-plot works for the assessment of goodness-of-fit;
- Reinforce the concepts of conditional probability and conditional expectation;
- Apply basic integration technique to compute mean excess function;
- Learn about behaviors of a heavy-tailed distribution;
- Learn how to use order statistics to estimate quantiles and mean excess function;
- Develop intuition behind point estimators.

# 1 Background

## 1.1 Central limit theorem

This section is to provide visualization of central limit theorem which you should already be familiar with. We provide examples on both discrete random variable and continuous random variable.

**Example 1.1.** (Bernoulli random variables) Suppose we intend to test the fairness of a coin, i.e. whether the coin has equal chance of landing on a head or a tail. We can do so by counting the number of heads in a sequence of coin tosses. The number of heads in each toss is a Bernoulli random variable, denoted by  $X_1$ . Let  $p$  be the probability of a head and  $q = 1 - p$  be the probability of a tail. Then, its probability mass function is given by

$$\mathbb{P}(X_1 = x) = \begin{cases} p, & x = 1, \\ q, & x = 0. \end{cases}$$

We let  $X_k$  be the number of heads in the  $k$ -th coin toss,  $k = 1, 2, \dots, n$ . Then we count the total number of heads after  $n$  coin tosses.

$$S_n := \sum_{k=1}^n X_k.$$

Then it is easy to show that  $S_n$  is a binomial random variable with parameters  $n$  and  $p$  and its probability mass function is given by

$$\mathbb{P}(S_n = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \dots, n.$$

For example, suppose that we have an unfair coin with  $p = 0.3$ . Figure 1 shows the probability mass functions of the number of heads,  $S_n$ , where the number of coin tosses  $n = 1, 2, 3, 10, 20, 50$ . Since  $p < 0.5$ , we are more likely to see a smaller number of heads than that of tails in any given  $n$  tosses. In general, the probability mass function of  $S_n$  tends to skew towards to the right. However, as one can see in the later graphs in Figure 1, the probability mass function becomes more and more symmetric<sup>1</sup> as  $n$  gets bigger and bigger. This phenomenon is present for any  $p \in (0, 1)$ , no matter how extreme is  $p$ . Why is this happening? The answer is the **Central Limit Theorem**, which we have already learned in class.

Let us consider the sample average

$$\bar{X}_n := \frac{1}{n} S_n$$

---

<sup>1</sup>Note, however, this is not to suggest that the coin becomes fair.

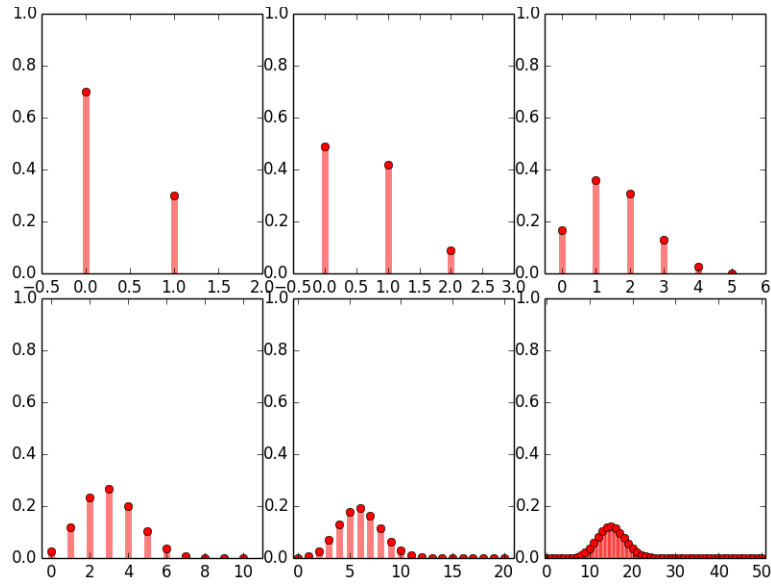


Figure 1: Probability mass function of  $S_n$

Since the expectation of the sample average

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = p,$$

the sample mean provides an unbiased estimator of the unknown parameter of fairness  $p$ .

**Exercise 1.1.** The Central Limit Theorem tells us that the estimator  $\bar{X}_n$  is asymptotically normal. In particular, we can construct the following random variable,  $Y_n$ , such that

$$Y_n := \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{pq}} \rightarrow \mathcal{N}(0, 1),$$

where  $\mathcal{N}(\mu, \sigma^2)$  is a normal random variable with mean  $\mu$  and variance  $\sigma^2$ . Explain why the estimator  $\bar{X}_n$  behaves roughly like  $\mathcal{N}(p, \frac{pq}{n})$ .

[**HINT:** First, note that  $Y_n$  is approximately  $\mathcal{N}(0, 1)$ . Now, reformulate the given equation to express  $\bar{X}_n$  as a function of  $Y_n$ ; what kind of function do you get? Combine these two facts to determine the distribution of  $\bar{X}_n$  by computing  $\mathbb{E}(\bar{X}_n)$  and  $\text{Var}(\bar{X}_n)$ .]

As the sample size  $n$  gets big, the variance is so small that the sample average gives very good estimate of the actual parameter  $p$ . That is why in practice we use the value of  $\bar{X}_n$  as an estimate, despite the fact that it is in fact a random variable.

**Exercise 1.2.** What is the exact distribution of  $\bar{X}_n$ ?

[**HINT:** First, determine the set of all values of  $X_n$  that can occur. Then, note that  $\bar{X}_n$  as a function of  $S_n$ , and apply the probability mass function of  $S_n$  to derive the probability mass function of  $\bar{X}_n$ . As a reminder, your probability mass function should give  $\mathbb{P}(\bar{X}_n = k)$  for all possible outcomes of  $\bar{X}_n$ . ]

Let us verify numerically the conclusion of Central Limit Theorem. Similar to what you showed in Exercise 1.2, one can show that the exact probability mass function of  $Y_n$  is given by

$$\mathbb{P}(Y_n = y) = \binom{n}{h} p^h q^{n-h}, \quad h := \sqrt{npq}y + np,$$

where  $y = (k - np)/\sqrt{npq}$  for  $k = 0, 1, \dots, n$ .

We can draw graphs of the probability mass functions and see how they converge to a normal distribution as  $n$  increases. Figure 2 below is an illustration of the central limiting theorem. The blue bars visually depict how a point mass function of a binomial random variable behave over an interval. The red dashed lines indicate the normal density function. From left to right, top to bottom we have the densities for binomial random variables with sample size  $n=1, 2, 5, 20, 100, 1000$  respectively, with probability of success being once again 30%.

**Example 1.2.** (Exponential random variables) Recall that the probability density function of an exponential random variable is given by

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0$$

where  $\mathbb{E}(X_i) = 1/\lambda$ . If we redefine  $\bar{X}_n$ ,  $Y_n$ , and  $S_n$  according to this new random variable, then the Central Limit Theorem tells us that

$$Y_n := \sqrt{n}(\lambda \bar{X}_n - 1) \longrightarrow \mathcal{N}(0, 1).$$

Let us compute this analytically.

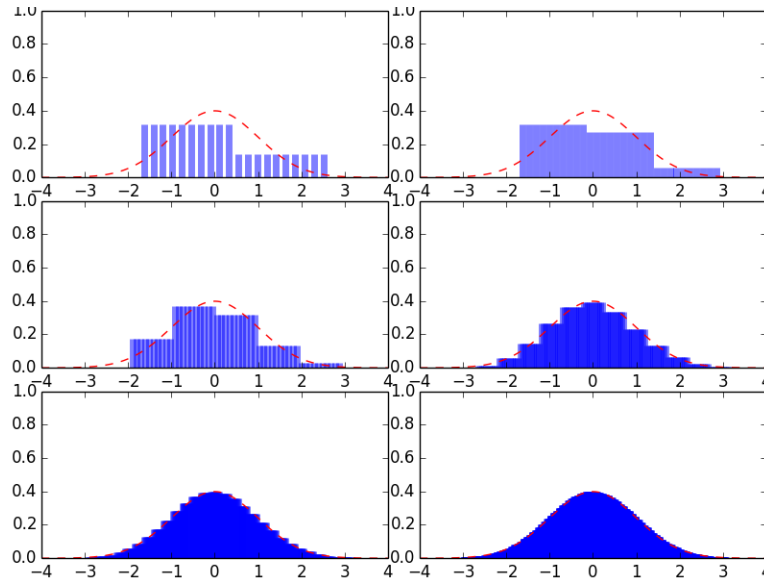


Figure 2: Binomial probability mass converging to normal density

**Exercise 1.3.** It can be shown that the probability density function of  $S_n$  is given by

$$f_{S_n}(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad x \geq 0 \quad (1)$$

Using this density function, show that the probability density function of  $Y_n$  is

$$f_{Y_n}(y) = \frac{n^{n-1/2}}{(n-1)!} \left(1 + \frac{y}{\sqrt{n}}\right)^{n-1} \exp\left\{-n\left(1 + \frac{y}{\sqrt{n}}\right)\right\}, \quad y > -\sqrt{n} \quad (2)$$

[**HINT:** First, write  $Y_n$  as a function of  $S_n$ . Then, use this function to determine both (a) the set of possible outcomes for  $Y_n$  based on the set of potential outcomes for  $S_n$ , and (b) the probability density function of  $Y_n$  over these outcomes, based on the probability density function of  $S_n$ .]

Figure 3 is a visual illustration of the probability density function of  $Y_n$  for various choices of  $n$ . We can see how the probability density function for  $Y_n$  converges to the standard normal density function. Again the red dashed lines is the standard normal density function while the blue lines are the densities for  $Y_n$ , given above, for  $n = 1, 2, 3, 5, 10, 100$ .

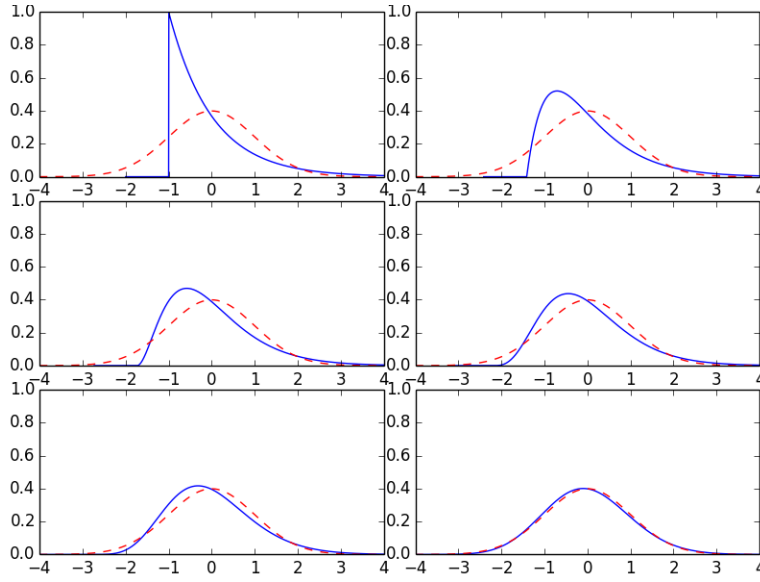


Figure 3: Exponential densities converging to normal density

**Exercise 1.4.** Use *matplotlib* in Python to create plots showing how  $f_{Y_n}(y)$  converges to the normal distribution as  $n$  increases. Your plots should contain an outline of the normal distribution like Figure 3. Use  $n = 1, 2, 10$ .

In general, we can conclude that *Central Limit Theorem* tells us that the distribution of an average tends to be Normal, even when the distribution from which the average is computed is non-Normal. However, there are certain cases in which the average deviate from Normal behavior. When does the Central Limit Theorem not hold?

## 1.2 Heavy-tailed Distributions

Let us introduce heavy-tailed distributions which are probability distributions with a heavier tail than the exponential. We will see how the extremes produced by heavy-tailed distributions will corrupt the average so that an asymptotic behavior different from the Normal behavior is obtained. Formally, a random variable  $X$  is said to have a **heavy-tailed** if

$$\lim_{x \rightarrow \infty} \frac{e^{-\lambda x}}{\bar{F}(x)} = 0, \quad \text{for all } \lambda > 0 \quad (3)$$

where  $\bar{F}(x) := \mathbb{P}(X > x)$  (here,  $\bar{F}(x)$  is often referred to as a **survival function**).

An example of a heavy-tailed distribution is the Pareto distribution. Consider the *strict Pareto* random variable whose density is given by

$$f(x) = \alpha x^{-\alpha-1}, \quad x > 1$$

where  $\alpha$  is a positive number, called the *Pareto index*. The Pareto distribution is very important in reinsurance so we will study it closely.

**Exercise 1.5.** Show that the Pareto may not have finite mean or variance by calculating the mean and variance of a strict Pareto random variable. Are there any values of  $\alpha$  for which either does not take a finite value?

[**HINT:** Examine the case when  $\alpha \in (0, 1]$  and  $\alpha > 1$ ]

**Exercise 1.6.** The Pareto distribution is closely related to the exponential. Given  $X$  is a strict Pareto random variable, show that  $Y = \ln(X)$  is exponentially distributed with mean  $\frac{1}{\alpha}$ .

**Exercise 1.7.** Using the definition of heavy-tailed from Equation 3, show the following:

- (a) A strict Pareto random variable is heavy-tailed.
- (b) An exponential random variable with rate  $\alpha$  is not heavy-tailed.

[**HINT:** You may need to apply L'Hospitôl's rule when taking the limit for (a). For part (b), consider selecting  $\lambda < \alpha$ .]

## 2 QQ-plot

Quantile-quantile plot, also known as QQ-plot for short, is a visual tool to check if a proposed model provides a plausible fit to the distribution of the random variable at hand. It is also a good visualization tool to see if a distribution is heavy-tailed. Let us first define what is a **quantile function** with an example.

**Example 2.1.** (Exponential distribution) Recall the cumulative density function for exponential with rate  $\lambda$  is given by

$$F_\lambda(x) := 1 - \exp(-\lambda x), \quad x > 0$$

and the survival function of the exponential is

$$\bar{F}_\lambda(x) := 1 - F_\lambda(x) = \exp(-\lambda x), \quad x > 0$$

Suppose we have real data  $x_1, x_2, \dots, x_n$  which we suspect might be exponentially distributed with some  $\lambda > 0$ . We can order these observations from the smallest to the largest. Denote the  $i$ -th smallest observation by  $x_{(i)}$ .

The quantile function for the exponential function has the form

$$Q_\lambda(p) = -\frac{1}{\lambda} \ln(1 - p), \quad p \in (0, 1).$$

Note that, of  $\lambda = 1$ , we have  $Q_1(p) = -\ln(1 - p)$ . Hence, there exists a simple linear relation between the quantiles of any exponential distribution and the corresponding standard exponential quantiles

$$Q_\lambda(p) = \frac{1}{\lambda} Q_1(p), \quad p \in (0, 1).$$

Note that we do not know the exact distribution of the unknown random variable, let alone the quantile. Nevertheless, we can replace the unknown quantile  $Q_\lambda$  by the empirical distribution  $\hat{Q}_n$  where

$$\hat{Q}_n(p) = x_{(i)}, \quad \text{for } \frac{i-1}{n} < p \leq \frac{i}{n}$$

In two-dimension, we essentially plotting the points with values

$$(-\ln(1 - p), \hat{Q}_n(p))$$

for several values of  $p \in (0, 1)$ . A typical choice of values of  $p$  is given by

$$p \in \left\{ \frac{1-1/2}{n}, \frac{2-1/2}{n}, \dots, \frac{n-1-1/2}{n}, 1 - \frac{1/2}{n} \right\}$$

In other words,  $p = (j - 1/2)/n$  for  $j = 1, 2, \dots, n$ . We then expect that a straight line pattern will appear in the scatter plot if  $\hat{Q}_n(p)$  indeed resembles  $Q_\lambda(p)$ , in other words, if the exponential model provides a plausible statistical fit for the given data set. When a straight line pattern is obtained, the slope of a fitted line can be used as an estimate of the parameter  $1/\lambda$ .

## 2.1 Interpreting QQ-plots

The discussion in the previous section compares our observed data with the exponential distribution using QQ-plots. If our QQ-plot looks like a straight line, then our data fit the exponential distribution, since our data grow proportionally with the quantiles of the exponential distribution; in this case the specific slope of our line will depend on the rate



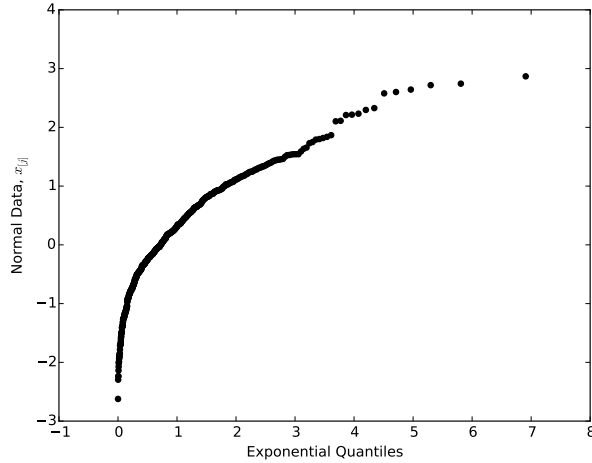


Figure 4: QQ-plot comparing normal data to an exponential distribution

λ. But what if our QQ-plot does not look like a straight line? In this case, we are not sure what distribution best fits our data, but may be able to gain some insights on how its shape differs from the exponential. Since we are interested in heavy-tailed distributions, we will focus on behavior in the tails of our distribution.

Consider the QQ-plot in Figure 4. Observe that this plot does not fit well to a straight line; as we move from left to right, the slope of this plot decreases. First, let's focus on the right side of the plot, which corresponds to the right tail of the distribution. If our observed data in the right tail were increasing proportionally with the quantiles of the exponential distribution, then the right portion of our QQ-plot would look like a straight line. However, we see that our plot is curving downward; therefore, the observations in our right tail are growing more slowly than the quantiles of the exponential distribution. This behavior suggests that our data have a lighter right tail than the exponential distribution. Now, let's focus on the left side of our plot, which corresponds to the left tail of our distribution. As we move to the left, we see that our plot is curving downward, which suggests that our data are decreasing more quickly than the exponential distribution. This behavior suggests that our data have a **heavier left tail** than the exponential distribution. Together, these two facts imply that our data show a **left skew** when compared to the exponential distribution.

To give insight into the reason for our observations, note that the observed data depicted in Figure 4 are taken from a standard normal population. The probability density functions of the normal and exponential distributions are depicted in Figure 5. Comparing the two, we can clearly see why the normal distribution has a heavier left tail than the exponential; the exponential distribution has a minimum value of zero, while the normal distribution can take any negative value. Comparisons in the right tail are not as clear cut, because

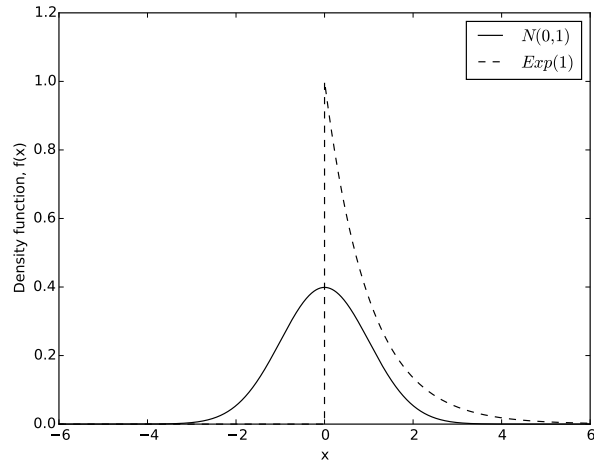


Figure 5: Probability density functions of the  $N(0,1)$  and  $Exp(1)$  distributions

both distributions can take arbitrarily large values. However, as we look out further to the right, you can see that the exponential is larger than the normal pdf when  $x$  is non-negative. Hence, **the exponential distribution has as heavier right tail** than the normal distribution, **since it is more likely to produce values farther to the right**. Again, these two facts show that the normal distribution has a left skew when compared to the exponential.

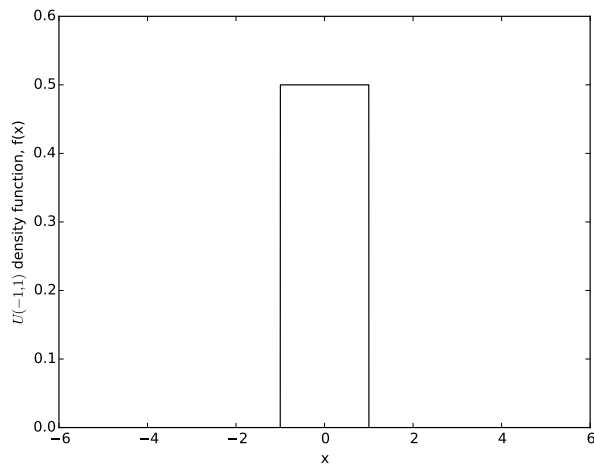


Figure 6: Probability density function of the  $U(-1,1)$  distribution

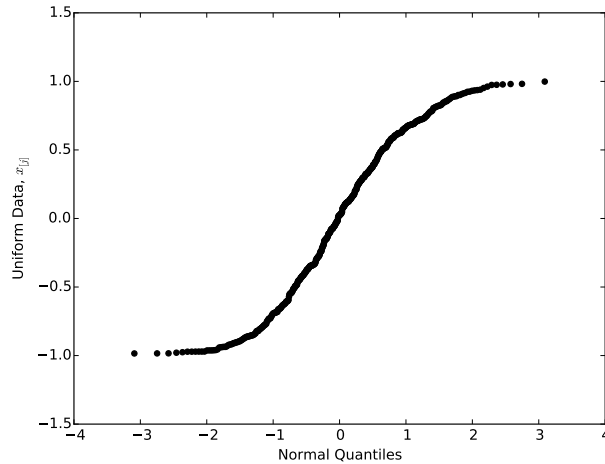


Figure 7: QQ-plot comparing  $U(-1, 1)$  data to a normal distribution

Suppose that our data were instead taken from a  $U(-1, 1)$  population, whose pdf is depicted in Figure 6. Comparing this distribution to the  $N(0, 1)$  pdf in Figure 5, we see that the  $U(0, 1)$  distribution can only generate observations between  $-1$  and  $1$ , while the  $N(0, 1)$  distribution can generate arbitrarily large positive and negative observations. Hence, the  $N(0, 1)$  distribution has heavier left and right tails than the  $U(0, 1)$  distribution. To observe this fact from data, consider Figure 7, which shows a QQ-plot created by generating 500 observations from the  $U(-1, 1)$  distribution and comparing them to the normal distribution. As we look to the left, we see that the data decrease more slowly than the normal quantiles, suggesting that the normal distribution has a heavier left tail than our (uniform) data. Similarly, as we look to the right, the data increase more slowly than the normal quantiles, suggesting that the normal distribution also has a heavier right tail than our data. Hence, our conclusions from the data agree with our earlier conclusions from observing the uniform and normal pdfs.

**Exercise 2.1.** Generate 500 observations from a  $U(0, 1)$  distribution, and create a QQ-plot comparing these observed data to the exponential distribution. Does your data show a heavier or lighter right tail than the exponential distribution? Create a plot of both the  $U(0, 1)$  and  $Exp(1)$  probability density functions distributions and explain why these plots support your conclusions.

Note that, in the above discussion, we did not say that a particular distribution has a

heavy tail, we only said that one distribution has a heavier tail *than another distribution*. To think of this another way, you would say that two sheets of paper are heavier than one sheet of paper, but it is unlikely that you would say that two sheets of paper are heavy. In our case, we say that a distribution has a heavy right tail if its tail is heavier than the exponential distribution, and has a heavy left tail if its tail is heavier than  $-X$ , where  $X$  is an exponential random variable.

**Example 2.2.** Figure 8 depicts a normal QQ-plot of 100 observations drawn from a Pareto distribution with  $\alpha = \frac{3}{2}$ . It shows that the sample average is far from a normal distribution. This illustrates numerically that the classical Central Limit Theorem does not apply for the Pareto distribution.

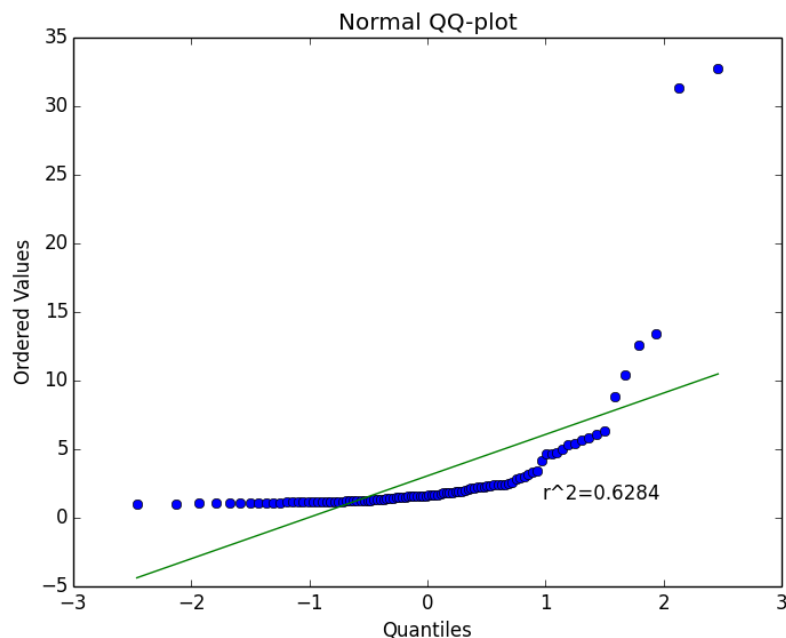


Figure 8: Normal QQ-plot for Example 2.2

**Exercise 2.2.** Give a justification why the QQ-plot in Example 2.2 fails to exhibit a pattern of normality even with a relatively large sample size of 100. [HINT: Consider the requirements for applying the Central Limit Theorem.]

## 2.2 Using real insurance claims data

Let us create a QQ-plot using real data. we investigate the insurance claim data from a reinsurance company. The data set *insurance.txt* contains automobile claims from 1988 until 2001 which are greater than 1,200,000 euro. This data set contains  $n = 371$  observations.

**Exercise 2.3.** Using Python, generate the QQ-plot by doing the following:

1. Read the data and select only the largest 270 observations
2. Take a logarithmic transform of the selected observations
3. Using *matplotlib* and *scipy.stats.probplot*, create exponential and normal QQ-plots of the data.

Does the data fit an exponential or normal distribution? Justify your answer. Note that, if the exponential provides a good fit to your data after a logarithmic transformation, then your result from Exercise 1.6 suggests that the data fit a Pareto distribution.

## 3 Mean excess function and reinsurance

Reinsurance is an insurance policy purchased by an insurance company from one or more other insurance companies, known as the “reinsurer”, as a means of risk management. It is a very common market practice when insurance companies undertake high risk profiles with potential catastrophic losses. A reinsurance agreement details the conditions upon which the reinsurer would pay a share of the claims incurred by the insurer and the reinsurer is paid a “reinsurance premium” by the insurer, which issues insurance policies to its own policyholders. A diagram of cash flows among the participants in an insurance market can be found in Figure 9.

A common form of reinsurance is the excess of loss (XL) reinsurance, where the insurer covers insurance claims from policyholders up to the maximum of its retention level and any amount beyond the retention will be reimbursed by the reinsurer. For example, an insurance company might insure commercial property risks with policy limits up to \$10 million, and then buy reinsurance of \$5 million in excess of \$5 million. In this case a loss of \$6 million on that policy will result in the recovery of \$1 million from the reinsurer.

The modeling of the XL reinsurance relies on an important mathematical concept, called the *mean excess function*. Suppose a ceding insurer enters into an XL treaty with a retention level  $t$ . Let  $X$  be a random variable governing the size of a particular policy-

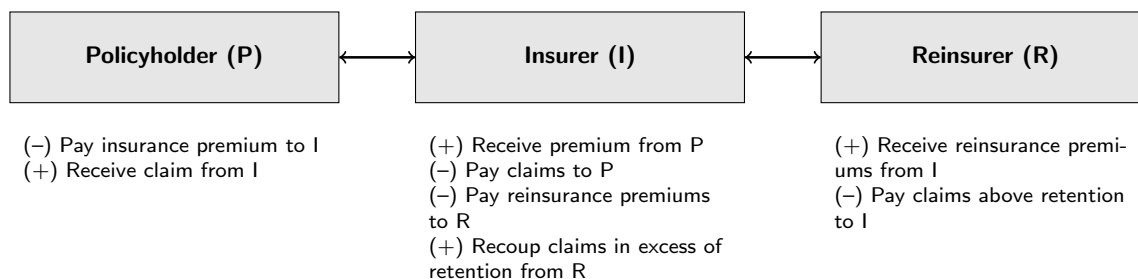


Figure 9: Participants in an insurance market

holder's claim. After claim investigation, the ceding insurer will make the payment and the reinsurer has to pay  $X - t$  if  $X > t$ . Pricing actuaries from reinsurance companies would want to know the average cost of such claims, which is theoretically determined by the mean excess function

$$e(t) = \mathbb{E}(X - t | X > t) = \mathbb{E}(X | X > t) - t. \quad (4)$$

**Exercise 3.1.** Show that the mean excess function of an exponential random variable with mean  $1/\lambda$  is given by

$$e(t) = \frac{1}{\lambda}, \quad t > 0.$$

[**HINT:** Note that  $f_{X|X>t}(x)$ , the probability density function of  $X$  given that  $X > t$ , can be expressed as  $f_{X|X>t}(x) = f_X(x)/P(X > t)$  if  $X > t$  (and 0 otherwise), where  $f_X(x)$  is the PDF of random variable  $X$ . Apply the definition of expectation, using this distribution, to compute  $\mathbb{E}(X - t | X > t)$ ].

**Exercise 3.2.** Show that the mean excess function of a strict Pareto random variable with the Pareto index  $\alpha$  is given by

$$e(t) = \frac{t}{\alpha - 1}, \quad t > 1.$$

[**HINT:** See the hint for the previous exercise].

Observe that the mean excess function can also be written as<sup>2</sup>

$$e(t) = \frac{\mathbb{E}[XI(X > t)]}{\mathbb{E}[I(X > t)]} - t,$$

where  $I(A) = 1$  if the event  $A$  is true and 0 otherwise. In practice, we replace the theoretical mean by its empirical counterpart. Given the sample data  $x_1, x_2, \dots, x_n$ , the mean excess function is estimated by

$$\hat{e}_n(t) = \frac{\sum_{i=1}^n x_i I(x_i > t)}{\sum_{i=1}^n I(x_i > t)} - t.$$

Often the empirical function  $\hat{e}_n$  is evaluated at  $t = x_{(n-k)}$ , the  $(k+1)$ -th largest observation. Then the numerator equals  $\sum_{i=1}^n x_i I(x_i > t) = \sum_{j=n-k+1}^n x_j$ , while the denominator equals  $k$ . The estimates of the mean excesses are then given by

$$e_k := \hat{e}_n(x_{(n-k)}) = \frac{1}{k} \sum_{j=n-k+1}^n x_j - x_{(n-k)}. \quad (5)$$

Consider an XL reinsurance contract with a retention level  $R$ . The reinsurer is obligated to pay for the claim amount in excess over the limit  $R$ . The fair net premium<sup>3</sup> is given by

$$\Pi(R) = \mathbb{E}[(X - R)_+], \quad (6)$$

where  $(x)_+ = \max\{x, 0\}$ . Equivalently, we obtain

$$\Pi(R) = e(R)\bar{F}(R). \quad (7)$$

Since reinsurance contracts are meant to transfer risks of catastrophic losses, claims of small and medium sizes provide no useful information for the valuation of reinsurance contracts. Let us consider a reinsurance contract with various retention levels, for example  $R = 5,000,000$  euro, which is typically used in practice. Observe that only 12 observations are larger than that level in the given data set. We use two methods in this case to determine the net premium.

---

<sup>2</sup>To see this result, note that  $XI(X > t)$  is a function of random variable  $X$ , such that  $XI(X > t) = X$  if  $X > t$  (and 0 otherwise). Hence, we can compute  $\mathbb{E}[XI(X > t)]$  as we do any other function of  $X$ .

<sup>3</sup>Net premium refers to the pure cost of insurance coverage. It is used in contrast with gross premium, which includes commission, policy expenses and other administrative costs.

**Exercise 3.3.** (Estimator #1: non-parametric) The simplest way to estimate the net premium  $\Pi(R)$  is to use an empirical estimator of (6) given by

$$\hat{\Pi}_1(R) := \frac{1}{n} \sum_{i=1}^n (x_i - R)_+.$$

Develop a computer algorithm in Python to estimate the net premiums for various retention levels in Table 1.

**Exercise 3.4.** (Estimator #2: non-parametric) Another way to determine the net premium is to make use of the identity (7)

$$\hat{\Pi}_2(R) := \hat{e}_n(R) \hat{\bar{F}}_n(R),$$

where  $\hat{\bar{F}}_n(R)$  is some estimator of the tail probability  $F(R) = \mathbb{P}(X > R)$ . If  $R$  is fixed at one of the sample points, that is,  $R = x_{(n-k)}$ , the non-parametric estimator is given by

$$\hat{\Pi}_2(X_{(n-k)}) = \frac{k}{n} e_k.$$

If  $R$  is not fixed at one of the sample points, then we have to introduce an estimator for  $\bar{F}(R)$ . There are many different ways of defining an estimator. For example, one can estimate  $\bar{F}(R)$  as

$$\hat{\bar{F}}_n(R) := \frac{1}{n} \sum_{i=1}^n I(x_i > R).$$

Substituting this back into our estimator of  $\hat{\Pi}_2(R)$ , we find

$$\hat{\Pi}_2(R) := \frac{\hat{e}_n(R)}{n} \sum_{i=1}^n I(x_i > R).$$

Show that, if we define  $\hat{\Pi}_2(R)$  in this way, then this estimator for the net premium is the same as  $\hat{\Pi}_1(R)$ . [**HINT:** You should be able to demonstrate that these are equal by algebraic manipulation.]

Note that in the previous two pricing models we did not use any assumption of a parametric model. As much as we prefer simplicity, we should also keep in mind a general rule from statistical theory that statistical estimators do not provide accurate results when the size of sample data for estimation is very small. When  $R = 5,000,000$  euros, only 12 observations enter into the calculations of  $\hat{\Pi}_1$  and  $\hat{\Pi}_2$ . It is essentially a waste of information



from other data points below 5 million euros.

We learned from Exercise 2.3 that Pareto distribution provides a good fit for the large observations in the data set. Hence, we should take advantage of this extra information extracted from the data set. One should keep in mind, however, that the strictly Pareto fits the large observations well but not necessarily for small data points. It indicates that the data set can be modeled by a *Pareto-type* distribution, whose tail behaves like a Pareto distribution. Without getting into too much technical details regarding Pareto-type distributions, we consider a scaled Pareto distribution for our analysis. Suppose the large observations all came from a random variable  $X$ , whose survival probability function is given by

$$\bar{F}(x) = \left(\frac{x}{C}\right)^{-\alpha}, \quad x > C,$$

for some large number  $C > 0$ . It means that  $X$  is  $C$  times a strictly Pareto random variable. Hence it is easy to show that the mean excess function of the scaled Pareto random variable remains the same as in Exercise 3.2. This indicates that we can get around estimating the constant  $C$  when using estimates of the mean excess function.

Hence it remains to estimate the unknown parameter  $\alpha$ , whose reciprocal,  $\gamma := 1/\alpha$ , is known as the extreme value index, in the extreme value theory. A well-known estimator of the extreme value index, called the Hill estimator, was proposed in Hill [1]. The intuition behind this estimator can be easily explained. It follows from Exercise 1.6 and Exercise 3.2 that the mean excess function of the logarithmic of the strict Pareto random variable (and that of the scaled Pareto random variable) is  $1/\alpha$ . According to (5), it would be natural to consider

$$\hat{\gamma}_{k,n} := \frac{1}{k} \sum_{j=n-k+1}^n \ln x_j - \ln x_{(n-k)}.$$

This is the Hill estimator. It enjoys a high degree of popularity thanks to some nice theoretical properties, which we do not intend to discuss in this class. For example, it has been proven that the Hill estimator is a consistent estimator for  $\gamma$ . ([https://en.wikipedia.org/wiki/Consistent\\_estimator](https://en.wikipedia.org/wiki/Consistent_estimator)) One should note that the Hill estimator is based on  $k$ -largest observations and for each choice of  $k$  there would be an estimate of  $\gamma$ . Many researchers have developed strategies to determine the optimal  $k$  under various statistical criteria. The discussion of such procedures can be quite evolved and hence should be omitted here.

**Exercise 3.5.** (Estimator #3: parametric) By combining our result from Exercise 3.2 with equation 7, we can also write

$$\Pi(R) = \frac{R}{\alpha - 1} \bar{F}(R).$$

If the retention level  $R$  is situated within the sample, say  $R = x_{(n-k)}$ , then we can use the estimator

$$\hat{\Pi}_3(R) := \left( \frac{x_{(n-k)}}{1/\hat{\gamma}_{k,n} - 1} \right) \frac{k}{n}.$$

If  $R$  is not fixed at one of the sample points, then we have manipulate the net premium formula to adopt a different form. By applying the definition of conditional probability with  $c > 1$ , we can show that

$$\mathbb{P} \left( \frac{X}{t} > c \mid X > t \right) = c^{-\alpha}.$$

Therefore, by applying the definition of conditional probability, we must have  $\mathbb{P}(X > R) = \mathbb{P}(\frac{X}{t} > \frac{R}{t}) = \mathbb{P}(\frac{X}{t} > \frac{R}{t} \mid X > t) \mathbb{P}(X > t)$ , which we can write as

$$\bar{F}(R) = \bar{F}(t) \left( \frac{R}{t} \right)^{-\alpha}, \quad R > t,$$

which implies

$$\Pi(R) = \frac{R}{\alpha - 1} \bar{F}(t) \left( \frac{R}{t} \right)^{-\alpha}, \quad R > t.$$

Let  $k$  be an appropriate choice for the number of largest observations and set  $t = x_{(n-k)}$ . If we estimate  $\alpha$  as the inverse of the Hill estimator and estimate  $\bar{F}(x_{(n-k)})$  as  $k/n$ , another estimator of net premium is given by

$$\hat{\Pi}_3(R) = \left( \frac{R}{1/\hat{\gamma}_{k,n} - 1} \right) \left( \frac{k}{n} \right) \left( \frac{R}{x_{(n-k)}} \right)^{-1/\hat{\gamma}_{k,n}}.$$

Suppose the optimal choice of  $k$  is 95. Compute the estimators in Table 1 based on the parametric model.

Now you have completed a pricing exercise that an actuary would typically do for an reinsurance contract on luxury commodities. You can see that there is no “one-size-fits-all” magic formula for pricing a product. Keep in mind that each of the estimators we developed has its own merits and limitations. Observe from your solutions to Table 1 that the estimates of net premiums are close for small retention levels, as more data are used in all

R	$\hat{\Pi}_1$	$\hat{\Pi}_3$
3,000,000		
3,500,000		
4,000,000		
5,000,000		
7,500,000		

Table 1: Estimates for  $\Pi(R)$

estimators. But they are far apart for high retention levels, due to the “low quality” results from the non-parametric statistics with limited sample data. An advanced knowledge of heavy-tailed distribution allows us to have better use of data information and provide more trust-worthy solutions in this example, which would be very important for billion-dollar businesses like reinsurance.

## References

- [1] Hill, B.M. (1975). A simple approach to inference about the tail of a distribution. *Annals of Statistics*. **3**: 1163–1174.