1 Tails of distributions

In 1998 freezing rains fell on eastern Ontario, south-western Quebec and lasted for six days. The event doubled the amount of precipitation in the area experienced in any prior ice storm, and resulted in a catastrophe that produced excess of 840,000 cases of insurance claims. This number is 20% more than that of the claims caused by the Hurricane Andrew - one of the largest natural disasters in the history of North America. After all, the catastrophe caused approximately 1.44 billion Canadian dollars insurance settlements which is the highest loss burden in the history of Canada. More examples of similar catastrophic events that caused extremal insurance losses are Hurricanes Harvey and Sandy, the 2011 Japanese earthquake and tsunami, and so forth.

In the context of insurance, a few large losses hitting a portfolio and then converting into claims usually represent the greatest part of the indemnities paid by insurance companies. The aforementioned losses, also called 'extremes', are quantitatively modelled by the tails of the associated probability distributions. From the quantitative modelling standpoint, relying on probabilistic models with improper tails is rather daunting. For instance, periods of financial stress may appear with higher frequency than expected, and insurance losses may occur with worse severity. Therefore, the study of probabilistic behavior in the tail portion of actuarial models is of utmost importance in the modern framework of quantitative risk management. For this reason, this section is devoted to the introduction of a few mathematical notions that characterize the tail weight of random variables (r.v.'s). The applications of these notions will benefit us in the construction and selection of appropriate models with desired mathematical properties in the tail portion, that are suitable for a given task.

Formally, define X to be the (random) obligations that arise from a collection (portfolio) of insurance contracts. We are particularly interested in studying the right tail of the distribution of X, which represents the occurrence of large losses. Speaking plainly, a r.v. is said to be heavier-tailed if higher probabilities are assigned to larger values. Unwelcome outcomes are more likely to occur for an insurance portfolio that is described by a loss r.v. possessing heavier (right) tail. Tail weight can be an absolute or a relative concept. Specifically, for the former, we may consider a r.v. to be heavy-tailed if certain mathematical properties of the probability distribution are met. For the latter, we can say the tail of one distribution is heavier than the other if some tail measures are larger/smaller.

In the statistics and probability literature, there are several quantitative approaches have been proposed to classify and compare tail weight. Among most of these approaches, the

survival functions serve as the building block. In what follows, we are going to introduce two simple yet useful tail classification methods, in which the basic idea is to study the quantities that are closely related to behavior of the survival function of X.

1.1 Classification Based on Moments

One possible way of classifying the tail weight of distribution is by assessing the existence of raw moments. Since our major interest lies in the right tails of distributions, we henceforth assume the obligation/loss r.v. X to be positive. At the outset, let us recall that the $k-$th raw moment of a continuous r.v. X, for $k \geq 0$, can be computed via

$$\mu'_k = k \int_0^\infty x^{k-1} S(x) dx,$$

where $S(\cdot)$ denotes the survival function of X. It is a simple matter to see that the existence of the raw moments depends on the asymptotic behavior of the survival function at infinity. Namely, the faster the survival function decays to zero, the higher the order of finite moment the associated r.v. possesses. Hence the maximal order of finite moment, denoted by $k^* := \sup\{k \in \mathbf{R}_+ : \mu'_k < \infty\}$, can be considered as an indicator of tail weight. This observation leads us to the moment-based tail weight classification method, which is defined formally next.

Definition 1. *For a positive loss random variable X, if all the positive raw moments exist, namely the maximal order of finite moment $k^* = \infty$, then X is said to be light-tailed based on the moment method. If $k^* = a \in (0, \infty)$, then X is said to be heavy-tailed based on the moment method. Moreover, for two positive loss random variables X_1 and X_2 with maximal orders of moment k_1^* and k_2^* respectively, we say X_1 has a heavier (right) tail than X_2 if $k_1^* \leq k_2^*$.*

It is noteworthy that the first part of Definition 1 is an absolute concept of tail weight, while the second part is a relative concept of tail weight which compares the (right) tails between two distributions. Next, we are going to present a few examples that illustrate the applications of the moment-based method for comparing tail weight. Some of these examples are borrowed from Klugman et al., 2012.

Example 1. *Let $X \sim Gamma(\alpha, \theta)$, with $\alpha > 0$ and $\theta > 0$, then for all $k > 0$,*

$$
\begin{aligned}
\mu'_k &= \int_0^\infty x^k \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha} dx \\
&= \int_0^\infty (y\theta)^k \frac{(y\theta)^{\alpha-1} e^{-y}}{\Gamma(\alpha)\theta^\alpha} \theta dy \\
&= \frac{\theta^k}{\Gamma(\alpha)} \Gamma(\alpha + k) < \infty.
\end{aligned}
$$

Since all the positive moments exist, i.e., $k^* = \infty$, in accordance with the moment-based classification method in Definition 1, the gamma distribution is light-tailed.

Example 2. *Let $X \sim Weibull(\theta, \tau)$, with $\theta > 0$ and $\tau > 0$, then for all $k > 0$,*

$$
\begin{aligned}
\mu'_k &= \int_0^\infty x^k \frac{\tau x^{\tau-1}}{\theta^\tau} e^{-(x/\theta)^\tau} dx \\
&= \int_0^\infty \frac{y^{k/\tau}}{\theta^\tau} e^{-y/\theta^\tau} dy \\
&= \theta^k \Gamma(1 + k/\tau) < \infty.
\end{aligned}
$$

Again, due to the existence of all the positive moments, the Weibull distribution is light-tailed.

We notice in passing that the gamma and Weibull distributions have been used quite intensively in the actuarial practice nowadays. Applications of these two distributions are vast which include, but are not limited to, insurance claim severity modelling, solvency assessment, loss reserving, aggregate risk approximation, reliability engineering and failure analysis. We have thus far seen two examples of using the moment-based method to analyze light-tailed distributions. We document a heavy-tailed example in what follows.

Example 3. *Let $X \sim Pareto(\alpha, \theta)$, with $\alpha > 0$ and $\theta > 0$, then for $k > 0$*

$$
\begin{aligned}
\mu'_k &= \int_0^\infty x^k \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} dx \\
&= \alpha\theta^\alpha \int_\theta^\infty (y-\theta)^k y^{-(\alpha+1)} dy.
\end{aligned}
$$

Consider a similar integration:

$$
g_k := \int_\theta^\infty y^{k-\alpha-1} dy = \begin{cases} < \infty, & \text{for } k < \alpha; \\ = \infty, & \text{for } k \geq \alpha. \end{cases}
$$

Meanwhile,

$$
\lim_{y \to \infty} \frac{(y-\theta)^k y^{-(\alpha+1)}}{y^{k-\alpha-1}} = \lim_{y \to \infty} (1 - \theta/y)^k = 1.
$$

3

Application of the limit comparison theorem for improper integrals yields μ'_k is finite if and only if g_k is finite. Hence we can conclude that the raw moments of Pareto r.v.'s exist only up to $k < \alpha$, i.e., $k^* = \alpha$, and thus the distribution is heavy-tailed. What is more, the maximal order of finite moment depends only on the shape parameter α and it is an increasing function of α. In other words, based on the moment method, the tail weight of Pareto r.v.'s is solely manipulated by α – the smaller the value of α, the heavier the tail weight becomes. Since $k^* < \infty$, the tail of Pareto distribution is heavier than those of the gamma and Weibull distributions.

We are going to conclude this current section by an open discussion on the limitations of the moment-based method. Despite its simple implementation and intuitive interpretation, there are certain circumstances in which the application of the moment-based method is not suitable. First, for more complicated probabilistic models, the k-th raw moment may not be simple to derive, and thus the identification of the maximal order of finite moment can be challenging. Second, the moment-based method does not well comply with main body of the well established heavy tail theory in literature. Specifically, the existence of moment generating functions (MGF's) is arguably the most popular method for classifying heavy tail versus light tail within the community of academic actuaries. However, for some r.v's such as the log normal r.v.'s, their MGF's do not exist even that all the positive moments are finite. In these cases, applications of the moment-based and the MFG-based methods can lead to different tail weight assessment. Third, when we need to compare the tail weight between two light-tailed distributions both having all positive moments exist, the moment-based method is no longer informative (see, e.g., Examples 1 and 2).

1.2   Comparison Based on Limiting Tail Behavior

In order to resolve the aforementioned issues of the moment-based classification method, an alternative approach for comparing tail weight is to directly study the limiting behavior of the survival functions.

Definition 2. *For two r.v.'s X and Y, and let*

$$\gamma := \lim_{t \to \infty} \frac{S_X(t)}{S_Y(t)}.$$

We say that

- *X has a heavier right tail than Y if $\gamma = \infty$;*

- X and Y are proportionally equivalent in the right tail, if $\gamma = c \in \mathbf{R}_+$;

- X has a lighter right tail than Y if $\gamma = 0$.

Example 4. *Let $X \sim Pareto(\alpha, \theta)$ and $Y \sim Weibull(\tau, \theta)$, for $\alpha > 0$, $\tau > 0$, and $\theta > 0$, we have*

$$
\begin{aligned}
\lim_{t \to \infty} \frac{S_X(t)}{S_Y(t)} &= \lim_{t \to \infty} \frac{(1 + t/\theta)^{-\alpha}}{\exp\{-(t/\theta)^\tau\}} \\
&= \lim_{t \to \infty} \frac{\exp\{t/\theta^\tau\}}{(1 + t^{1/\tau}/\theta)^\alpha} \\
&= \lim_{t \to \infty} \frac{\sum_{i=0}^{\infty} \left(\frac{t}{\theta^\tau}\right)^i / i!}{(1 + t^{1/\tau}/\theta)^\alpha} \\
&= \lim_{t \to \infty} \sum_{i=0}^{\infty} \left(t^{-i/\alpha} + \frac{t^{(1/\tau - i/\alpha)}}{\theta}\right)^{-\alpha} / \theta^{\tau i} i! \\
&= \infty.
\end{aligned}
$$

Therefore, the Pareto distribution has a heavier tail than the Weibull distribution. One may also realize that exponentials go to infinity faster than polynomials, thus the aforementioned limit must be infinite.

For some distributions of which the survival functions do not admit explicit expressions, we may find the following alternative formula useful:

$$
\begin{aligned}
\lim_{t \to \infty} \frac{S_X(t)}{S_Y(t)} &= \lim_{t \to \infty} \frac{S'_X(t)}{S'_Y(t)} \\
&= \lim_{t \to \infty} \frac{-f_X(t)}{-f_Y(t)} \\
&= \lim_{t \to \infty} \frac{f_X(t)}{f_Y(t)}.
\end{aligned}
$$

given that the density functions exist.

Example 5. *Let $X \sim Pareto(\alpha, \theta)$ and $Y \sim Gamma(\alpha, \theta)$, for $\alpha > 0$ and $\theta > 0$, we have*

$$
\begin{aligned}
\lim_{t \to \infty} \frac{f_X(t)}{f_Y(t)} &= \lim_{t \to \infty} \frac{\alpha \theta^\alpha (t + \theta)^{-\alpha - 1}}{t^{\tau - 1} e^{-t/\lambda} \lambda^{-\tau} \Gamma(\tau)^{-1}} \\
&\propto \lim_{t \to \infty} \frac{e^{t/\lambda}}{(t + \theta)^{\alpha + 1} t^{\tau - 1}} \\
&= \infty,
\end{aligned}
$$

as exponentials go to infinity faster than polynomials.

2 Risk measures

In this previous section, we studied two methods for classifying the weight of distribution tails. We may claim that the risk associated with one distribution is more dangerous (asymptotically) than the others if the tail is heavier. However, knowing one risk is more dangerous (asymptotically) than the others may not provide sufficient information for a sophisticated risk management purpose, and in addition, one is also interested in quantifying how much more. In fact, the magnitude of risk associated with a given loss distribution is an essential input for many insurance applications, such as actuarial pricing, reserving, hedging, insurance regulatory oversight, and so forth.

To compare the magnitude of risk in a practically convenient manner, we aim to seek a function that maps the loss r.v. of interest to a numerical value indicating the level of riskiness, which is termed the risk measure. Putting mathematically, denoted by $\mathcal{X}$ a set of insurance loss r.v.'s, a risk measure is a functional map $H : \mathcal{X} \to \mathbf{R}_+$. In principle, risk measures can admit an unlimited number of functional formats. Classical examples of risk measures include the mean $\mathbf{E}[X]$, the standard deviation $\mathbf{SD}(X) := \sqrt{\mathbf{Var}(X)}$, the standard deviation principle

$$H_{\mathbf{SD}}(X) := \mathbf{E}[X] + \alpha \mathbf{SD}(X), \text{ for } \alpha \geq 0, \tag{1}$$

and the variance principle

$$H_{\mathbf{Var}}(X) := \mathbf{E}[X] + \alpha \mathbf{Var}(X), \text{ for } \alpha \geq 0.$$

It is a simple matter to check that all the aforementioned functions are risk measures in which we input the loss r.v. and the functions output a numerical value. On a different note, the function $H^*(X) := \alpha X^\beta$ for any real-valued $\alpha, \beta \neq 0$, is not a risk measure since H^* produces another r.v. rather than a single numerical value.

Since risk measures are scalar measures which aim to use a single numerical value to describe the stochastic nature of loss r.v.'s, it should not be surprising to us that there is no risk measures can capture all the risk information of the associated r.v.'s. Therefore, when seeking useful risk measures, it is important for us to keep in mind that the measures should be at least

- interpretable practically;

- computable conveniently; and

- being able to reflect the most critical information of risk underpinning the loss distribution.

A vast number of risk measures have been developed in the literature of actuarial mathematics. Unfortunately, there is no best risk measure that can outperform the others, and the selection of appropriate risk measure depends mainly on the application questions at hand. In this respect, it is imperative to emphasize that 'risk' is a subjective concept, and thus even given the same problem, there are multifarious approaches to assess risk. However, for many risk management applications, there is a wide agreement that economically sounded risk measures should satisfy four major axioms which we are going to describe them in detail next. Risk measures that satisfy these axioms are termed *coherent* risk measures.

Consider in what follows a risk measure $H(\cdot)$, then H is a coherent risk measure if the following axioms are satisfied.

- **Axiom 1.** *Subadditivity:* $H(X + Y) \leq H(X) + H(Y)$. The economic implication of this axiom is that diversification benefits exist if different risks are combined.

- **Axiom 2.** *Monotonicity:* if $\mathbf{P}[X \leq Y] = 1$, then $H(X) \leq H(Y)$. Recall that X and Y are r.v.'s representing losses, the underlying economic implication is that higher losses essentially leads to a higher level of risk.

- **Axiom 3.** *Positive homogeneity:* $H(cX) = cH(X)$ for any positive constant c. A potential economic implication about this axiom is that risk measure should be independent of the monetary units in which the risk is measured. For example, let c be the currency exchange rate between the US and Canadian dollars, then the risk of random losses measured in terms of US dollars (i.e., X) and Canadian dollars (i.e., cX) should be different only up to the exchange rate c (i.e., $cH(x) = H(cX)$).

- **Axiom 4.** *Translation invariance:* $H(X + c) = H(X) + c$ for any positive constant c. If the constant c is interpreted as risk-free cash, this axiom tells that no additional risk is created for adding cash to an insurance portfolio, and injecting risk-free capital of c can only reduce the risk by the same amount.

Verifying the coherent properties for some risk measures can be quite straightforward, but it can be very challenging sometimes. For example, it is a simple matter to check that the mean is a coherent risk measure since for any pair of r.v.'s X and Y having finite means and constant $c > 0$,

- validation of *subadditivity*: $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$;

- validation of *monotonicity*: if $\mathbf{P}[X \leq Y] = 1$, then $\mathbf{E}[X] \leq \mathbf{E}[Y]$;

- validation of *positive homogeneity*: $\mathbf{E}[cX] = c\mathbf{E}[X]$;

- validation of *translation invariance*: $\mathbf{E}[X + c] = \mathbf{E}[X] + c$.

On a different note, the standard deviation is not a coherent risk measure. Specifically, one can check that the standard deviation satisfies

- validation of *subadditivity*:

$$
\begin{aligned}
\mathbf{SD}[X + Y] &= \sqrt{\mathbf{Var}(X) + \mathbf{Var}(Y) + 2\mathbf{Cov}(X, Y)} \\
&\leq \sqrt{\mathbf{SD}(X)^2 + \mathbf{SD}(Y)^2 + 2\mathbf{SD}(X)\mathbf{SD}(Y)} \\
&= \mathbf{SD}(X) + \mathbf{SD}(Y);
\end{aligned}
$$

- validation of *positive homogeneity*: $\mathbf{SD}[cX] = c\,\mathbf{SD}[X]$.

However, the standard deviation does not comply with translation invariance property as for any positive constant c,

$$
\mathbf{SD}(X + c) = \mathbf{SD}(X) < \mathbf{SD}(X) + c.
$$

Moreover, the standard deviation also does not satisfy the monotonicity property. To see this, consider the following two r.v.'s

$$
X = \begin{cases} 0, & \text{with probability } 0.25; \\ 4, & \text{with probability } 0.75, \end{cases} \tag{2}
$$

and Y is a degenerated r.v. such that

$$
\mathbf{P}[Y = 4] = 1. \tag{3}
$$

It is easy to check that $\mathbf{P}[X \leq Y] = 1$, but $\mathbf{SD}(X) = \sqrt{4^2 \cdot 0.25 \cdot 0.75} = \sqrt{3} > \mathbf{SD}(Y) = 0$.

We have so far checked that $\mathbf{E}[\cdot]$ is a coherent risk measure, but not $\mathbf{SD}(\cdot)$. Let us now proceed to study the coherent property for the standard deviation principle (1) which is a linear combination of two coherent and incoherent risk measures. To this end, for a given $\alpha > 0$, we check the four axioms for $H_{\mathbf{SD}}(X + Y)$ one by one:

- validation of *subadditivity:*

$$
\begin{aligned}
H_{\mathbf{SD}}(X + Y) &= \mathbf{E}[X + Y] + \alpha\mathbf{SD}(X + Y) \\
&\leq \mathbf{E}[X] + \mathbf{E}[Y] + \alpha[\mathbf{SD}(X) + \mathbf{SD}(Y)] \\
&= H_{\mathbf{SD}}(X) + H_{\mathbf{SD}}(Y);
\end{aligned}
$$

- validation of *positive homogeneity:*

$$H_{\mathbf{SD}}(cX) = c\mathbf{E}[X] + c\alpha\mathbf{SD}(X) = cH_{\mathbf{SD}}(X);$$

- validation of *translation invariance:*

$$H_{\mathbf{SD}}(X + c) = \mathbf{E}[X] + c + \alpha\mathbf{SD}(X) = H_{\mathbf{SD}}(X) + c.$$

It only remains to verify the monotonicity property, which may or may not be satisfied depending on the value of α. To see this, consider again the setup of (2) and (3) in which $\mathbf{P}[X \leq Y] = 1$. Let $\alpha = 0.1 \cdot \sqrt{3}$, then $H_{\mathbf{SD}}(X) = 3 + 0.3 = 3.3 < H_{\mathbf{SD}}(Y) = 4$ and the monotonicity condition is met. On the other hand, let $\alpha = \sqrt{3}$, then $H_{\mathbf{SD}}(X) = 3 + 3 = 6 > H_{\mathbf{SD}}(Y) = 4$ and the monotonicity condition is not satisfied. More precisely, by setting

$$H_{\mathbf{SD}}(X) = 3 + \alpha\sqrt{3} \leq 4 = H_{\mathbf{SD}}(Y),$$

we find that the monotonicity condition is only satisfied for $0 \leq \alpha \leq 1/\sqrt{3}$, and thus the standard deviation principle $H_{\mathbf{SD}}$ is coherent. This result appears to be very intuitive to us since the standard deviation principle $H_{\mathbf{SD}}$ is a linear combination two risk measures of which one is coherent and the other is incoherent. If $\alpha \leq 1/\sqrt{3}$, then the coherent measure dominates the incoherent one, thus the resulting measure $H_{\mathbf{SD}}$ is coherent and vice versa.

The literature on risk measures has been growing rapidly in popularity and importance. In the succeeding subsections, we are going to introduce two indices which have recently earned unprecedented amount of interest among theoreticians, practitioners, and regulators. They are namely the *Value-at-Risk* (VaR) and the *Tail Value-at-Risk* (TVaR) measures. The economic rationale behind these two popular risk measures is similar to that for the tail classification methods introduced in the previous section, with which we hope to capture the risk of extremal losses represented by the distribution tails.

2.1 Value-at-Risk

At the outset, we offer the formal definition of VaR.

Definition 3. *Consider an insurance loss random variable X. The Value-at-Risk measure of X with confidence level $q \in [0,1]$ is formulated as*

$$VaR_q[X] := \inf\{x \in \mathbf{R} : F_X(x) \geq q\}. \tag{4}$$

Speaking bluntly, the VaR measure outputs the smallest value of X such that the associated c.d.f. first excesses or equates to q. In the fields of probability and statistics, the VaR is also known as the percentiles.

Here is how we should interpret VaR in the lingo of actuarial mathematics. VaR is a forecast of the 'maximal' probable loss for a insurance product/portfolio or a risky investment occurring $q \times 100\%$ of times, over a specific time horizon (typically, one year). For instance, let X be the annual loss r.v. of an insurance product, $VaR_{0.95}[X] = 100$ million means that there is a 5% chance that the loss will exceed 100 million over a given year. Owing to the meaningful interpretation, VaR has become the industrial standard to measuring financial and insurance risks since 1990's. Financial conglomerates, regulators, and academics often utilize VaR to price insurance products, measure risk capital, ensure the compliance with regulatory rules, and disclose the financial positions.

Next, we are going to present a few examples about the computation of VaR.

Example 6. *Consider an insurance loss r.v. $X \sim Exp(\theta)$ for $\theta > 0$, then the c.d.f. of X is given by*

$$F_X(x) = 1 - e^{-x/\theta}, \ \text{for } x > 0.$$

Since exponential distribution is a continuous distribution, the smallest value such that the c.d.f. first exceeds or equates to $q \in [0,1]$ must be at the point x_q satisfying

$$q = F_X(x_q) = 1 - \exp\{-x_q/\theta\}.$$

Thus

$$VaR_q[X] = F_X^{-1}(q) = -\theta[\log(1-q)].$$

The result reported in Example 6 can be generalized to any continuous r.v.'s having strictly increasing c.d.f. Specifically, the VaR of any continuous r.v.'s is simply the inverse of the corresponding c.d.f. Let us consider another example of continuous r.v. which has the support from negative infinity to positive infinity.

Example 7. *Consider an insurance loss r.v. $X \sim Normal(\mu, \sigma^2)$ with $\mu \in \mathbf{R}$ and $\sigma > 0$. In this case, one may interpret the negative values of X as profit or revenue. Because normal distribution is a continuous distribution, the VaR of X must satisfy*

$$\begin{aligned} q &= F_X(VaR_q[X]) \\ &= \mathbf{P}\left[(X - \mu)/\sigma \le (VaR_q[X] - \mu)/\sigma\right] \\ &= \Phi((VaR_q[X] - \mu)/\sigma). \end{aligned}$$

Therefore, we have

$$VaR_q[X] = \Phi^{-1}(q)\ \sigma + \mu.$$

In many insurance applications, we have to deal with transformations of r.v.'s. For instance, in Example 7, the loss r.v. $X \sim Normal(\mu, \sigma^2)$ can be viewed as a linear transformation of a standard normal r.v. $Z \sim Normal(0,1)$, namely $X = Z\sigma + \mu$. By setting $\mu = 0$ and $\sigma = 1$, it is straightforward for us to check $VaR_q[Z] = \Phi^{-1}[q]$. A useful finding revealed from Example 7 is that the VaR of a linear transformation of the normal r.v.'s is equivalent to the linear transformation of the VaR of the original r.v.'s. This finding can be further generalized to any r.v.'s as long as the transformations are strictly increasing. The next example highlights the usefulness of the abovementioned finding.

Example 8. *Consider an insurance loss r.v. $Y \sim Log - Normal(\mu, \sigma^2)$, for $\mu \in \mathbf{R}$ and $\sigma > 0$. That is $\log Y \sim Normal(\mu, \sigma^2)$, or equivalently let $X \sim Normal(\mu, \sigma^2)$, then $Y \overset{d}{=} e^X$ which is strictly increasing transformation. Here, the notation '$\overset{d}{=}$' means equality in distribution. The VaR of Y is thus given by the exponential transformation of the VaR of X. Precisely, for $q \in [0,1]$,*

$$VaR_q[Y] = e^{VaR_q[X]} = \exp\{\Phi^{-1}(q)\ \sigma + \mu\}.$$

We have thus far seen a number of examples about the VaR for continuous r.v.'s, let us consider an example concerning the VaR for a discrete r.v.

Example 9. *Consider an insurance loss r.v. with the following probability distribution:*

$$\mathbf{P}[X = x] = \begin{cases} 1, & with\ probability\ 0.75; \\ 3, & with\ probability\ 0.20; \\ 4, & with\ probability\ 0.05. \end{cases}$$

The corresponding c.d.f. of X is

$$F_X(x) = \begin{cases} 0, & x < 1; \\ 0.75, & 1 \le x < 3; \\ 0.95, & 3 \le x < 4; \\ 1, & 4 \le x. \end{cases}$$

By the definition of VaR, we thus have, for example,

- $VaR_{0.6}[X] = 1;$

- $VaR_{0.9}[X] = 3$;

- $VaR_{0.95}[X] = 3$;

- $VaR_{0.950001}[X] = 4$.

Let us now conclude this current subsection by an open discussion of the VaR measure. Some advantages of utilizing VaR include

- possessing a practically meaningful interpretation;

- relatively simple to compute for many distributions with closed-form distribution functions;

- no additional assumption is required for the computation of VaR.

On the other hand, the limitations of VaR can be particularly pronounced for some risk management practices. We report some of them herein:

- the selection of the confidence level $q \in [0, 1]$ is highly subjective, while the VaR can be very sensitive to the choice of q (e.g., in Example 9, $VaR_{0.95}[X] = 3$ and $VaR_{0.950001}[X] = 4$);

- the scenarios/loss information that are above the $(1 - p) \times 100\%$ worst event, are completely neglected;

- VaR is not a coherent risk measure (specifically, the VaR measure does not satisfy the subadditivity axiom, meaning that diversification benefits may not be fully reflected).

2.2   Tail Value-at-Risk

Recall that the VaR represents the $(1 - p) \times 100\%$ chance maximal loss. As we mentioned in the previous section, one major drawback of the VaR measure is that it does not reflect the extremal losses occurring beyond the $(1 - p) \times 100\%$ chance worst scenario. For an illustration purpose, let us consider the following slightly unrealistic yet inspiring example.

Example 10. *Consider two loss r.v.'s $X \sim Uniform[0, 100]$, and $Y \sim Exp(31.71)$. We use VaR at 95% confidence level to measure the riskiness of X and Y. Simple calculation yields (see, also, Example 6),*

$$VaR_{0.95}[X] = VaR_{0.95}[Y] = 95,$$

*and thus these two loss distributions have the same level of risk according to $VaR_{0.95}$. How-
ever, it is clear that Y is more risky than X if extremal losses are of major concern since X
is bounded above while Y is unbounded. Simply quantifying risk by using VaR at a specific
confidence level could be misleading and may not reflect the true nature of risk.*

As a remedy, the *Tail Value-at-Risk* (TVaR) was proposed to measure the extremal
losses that are above a given level of VaR as an average. We document the definition of
TVaR in what follows. For the sake of simplicity, we are going to confine ourselves to
continuous positive r.v's only, which are more frequently used in the context of insurance
risk management. We refer the interested reader to Hardy (2006) for a more comprehensive
discussion of TVaR for both discrete and continuous r.v.'s.

Definition 4. *Fix $q \in [0, 1]$, the Tail Value-at-Risk of a (continuous) r.v. X is formulated
as*

$$TVaR_q[X] := \mathbf{E}[X|X > VaR_q[X]],$$

given that the expectation exists.

In light of Definition 4, the computation of TVaR typically consists of two major com-
ponents - the VaR and the average of losses that are above the VaR. The TVaR can be
computed via a number of formulas. Consider a continuous positive r.v. X, for notional con-
venience, henceforth let us write $\pi_q := VaR_q[X]$. By definition, the TVaR can be computed
via

$$TVaR_q[X] = \frac{1}{(1-q)} \int_{\pi_q}^{\infty} x f_X(x) dx. \tag{5}$$

Example 11. *Consider an insurance loss r.v. $X \sim Normal(\mu, \sigma^2)$ with $\mu \in \mathbf{R}$ and $\sigma > 0$.
Denoted by Z the standard normal r.v., for $q \in [0, 1]$, the TVaR of X can be computed via*

$$\begin{aligned}
TVaR_q[X] &= \mathbf{E}[X|X > VaR_q[X]] \\
&= \mathbf{E}[\sigma Z + \mu | \sigma Z + \mu > VaR_q[X]] \\
&= \sigma \mathbf{E}[Z|Z > (VaR_q[X] - \mu)/\sigma] + \mu \\
&\overset{(1)}{=} \sigma \mathbf{E}[Z|Z > VaR_q[Z]] + \mu,
\end{aligned}$$

where '$\overset{(1)}{=}$' holds because of the results reported in Example 7. Next, we turn to study

$TVaR_q[Z] = \mathbf{E}[Z|Z > VaR_q[Z]]$. Let $\omega(q) = (\Phi^{-1}(q))^2/2$, we have

$$
\begin{aligned}
(1-q)\ TVaR_q[Z] &= \int_{\Phi^{-1}(q)}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= \int_{\omega(q)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\omega(q)} \\
&= \phi(\Phi^{-1}(q)).
\end{aligned}
$$

Thus,

$$
TVaR_q[X] = \sigma \frac{\phi(\Phi^{-1}(q))}{1-q} + \mu.
$$

We mentioned earlier in the previous subsection that the VaR of a strictly increasing function of r.v. is equal to the function of VaR of the original r.v. Motivated by the results in Example 11, one can show that the TVaR of a strictly increasing linear transformation of r.v. is equal to the function of VaR of the original r.v. This is due to the linearity property of expectation. However, the aforementioned finding can not be extended to non-linear functions. The following example of log normal r.v. serves as a counter example.

Example 12. *Consider an insurance loss r.v. $X \sim Log - Normal(\mu, \sigma^2)$, with $\mu \in \mathbf{R}$ and $\sigma > 0$. Recall that the p.d.f. of log normal distribution is formulated as*

$$
f_X(x) = \frac{1}{\sigma\sqrt{2\pi}x} \exp\{-(\ln x - \mu)^2/2\sigma^2\}, \ for\ x > 0.
$$

Fix $q \in [0,1]$, then the TVaR of X can be computed via

$$
\begin{aligned}
TVaR_q[X] &= \frac{1}{(1-q)} \int_{\pi_q}^{\infty} x f_X(x) dx \\
&= \frac{1}{(1-q)} \int_{\pi_q}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\} dx \\
&\overset{(1)}{=} \frac{1}{(1-q)} \int_{\omega(q)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2 + \sigma w + \mu} dw \\
&= \frac{e^{\mu+\sigma^2/2}}{(1-q)} \int_{\omega(q)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(w-\sigma)^2} dw \\
&= \frac{e^{\mu+\sigma^2/2}}{(1-q)} \Phi(\omega(q) - \sigma), \quad\quad\quad (6)
\end{aligned}
$$

where '$\overset{(1)}{=}$' holds by applying change of variable $w = (\log x - \mu)/\sigma$, and $\omega(q) = (\log \pi_q - \mu)/\sigma$. Evoking the formula of VaR for log normal r.v. reported in Example 7, we can simplify the

expression (6) into

$$TVaR_q[X] = \frac{e^{\mu+\sigma^2/2}}{(1-q)}\Phi(\Phi^{-1}(q)-\sigma).$$

Clearly, the TVaR of log normal r.v. is not the exponential of the TVaR of normal r.v.

For distributions of which the distribution functions are more tractable to work with, we may apply integration by parts technique to rewrite Equation (5) as

$$TVaR_q[X] = \left[-xS_X(x)\Big|_{\pi_q}^{\infty} + \int_{\pi_q}^{\infty} S_X(x)dx\right]\frac{1}{(1-q)}$$

$$= \pi_q + \frac{1}{(1-q)}\int_{\pi_q}^{\infty} S_X(x)dx.$$

Example 13. *Consider an insurance loss r.v. $X \sim Exp(\theta)$ for $\theta > 0$, we have seen from the previous subsection that*

$$\pi_q = -\theta[\log(1-q)].$$

Let us now consider the TVaR:

$$TVaR_q[X] = \pi_q + \int_{\pi_q}^{\infty} e^{-x/\theta}dx/(1-q)$$

$$= \pi_q + \theta e^{-\pi_q/\theta}dx/(1-q)$$

$$= \pi_q + \theta.$$

In the SOA exam, we may use the expectation formulas provided in the distribution table to compute TVaR. Specifically, we have

$$TVaR_q[X] = \int_{\pi_q}^{\infty} (x - \pi_q + \pi_q)f_X(x)dx/(1-q)$$

$$= \pi_q + \frac{1}{(1-q)}\int_{\pi_q}^{\infty}(x - \pi_q)f_X(x)dx$$

$$= \pi_q + e_X(\pi_q)$$

$$= \pi_q + \frac{(\mathbf{E}[X] - \mathbf{E}[X \wedge \pi_q])}{(1-q)}, \tag{7}$$

where $e_X(d) := \mathbf{E}[X - d|X > d]$ for $d > 0$ denotes the mean excess loss function, and for many commonly used parametric distributions, the formulas for calculating $\mathbf{E}[X]$ and $\mathbf{E}[X \wedge \pi_q]$ can be found in the distribution table.

Example 14. *Consider a loss r.v. $X \sim Pareto(\theta, \alpha)$ with $\theta > 0$ and $\alpha > 0$. The c.d.f. of X is given by*

$$F_X(x) = 1 - \left(\frac{\theta}{\theta + x}\right)^{\alpha}, \text{ for } x > 0.$$

Fix $q \in [0,1]$ and set $F_X(\pi_q) = q$, we readily obtain

$$\pi_q = \theta \left[(1-q)^{-1/\alpha} - 1 \right]. \tag{8}$$

According to the distribution table provided in the SOA exam C, we know

$$\mathbf{E}[X] = \frac{\theta}{\alpha - 1},$$

and

$$\mathbf{E}[X \wedge \pi_q] = \frac{\theta}{\alpha - 1} \left[1 - \left(\frac{\theta}{\theta + \pi_q} \right)^{\alpha - 1} \right].$$

Evoking Equation (7) yields

$$
\begin{aligned}
TVaR_q[X] &= \pi_q + \frac{\theta}{\alpha - 1} \frac{(\theta/(\theta + \pi_q))^{\alpha - 1}}{(\theta/(\theta + \pi_q))^{\alpha}} \\
&= \pi_q + \frac{\theta}{\alpha - 1} \left(\frac{\pi_q + \theta}{\theta} \right) \\
&= \pi_q + \frac{\pi_q + \theta}{\alpha - 1},
\end{aligned}
$$

where π_q is given by (8).

Via change of variable, we can also rewrite Equation (5) as

$$TVaR_q[X] = \frac{1}{(1-q)} \int_q^1 VaR_\alpha[X] \, d\alpha. \tag{9}$$

What this alternative formula (9) tells is that TVaR in fact is the average of $VaR_\alpha[X]$ with varying degree of confidence level over $\alpha \in [q, 1]$. Therefore, the TVaR effectively resolves most of the limitations of VaR outlined in the previous subsection. First, due to the averaging effect, the TVaR may be less sensitive to the change of confidence level compared with VaR. Second, all the extremal losses that are above the $(1-q) \times 100\%$ worst probable event are taken in account. In this respect, it is a simple matter for us to see that for any given $q \in [0,1]$

$$TVaR_q[X] \geq VaR_q[X].$$

Third and perhaps foremost, TVaR is a coherent risk measure and thus is able to more accurately capture the diversification effects of insurance portfolio. Herein, we do not intend to provide the proof of the coherent feature for TVaR, which is considered to be challenging technically.

References

Hardy M.R., 2006. An introduction to risk measures for actuarial applications. CAS and SOA.

Klugman, S.A., Panjer, H.H., Willmot, G.E., 2012. Loss Models: From Data to Decisions, 4th ed. Wiley, New Jersey.