2 Actuarial Models, Mixtures and Risk

Definition 2.1. Let X be a random variable which follows some distribution. This distribution is a **scale distribution** if for any constant $c > 0$, cX is also a random variable that has the same distributional form as X (with possibly different parameters). □

The definition of a scale distribution is most easily seen through an example.

Example 2.2. Let $X \sim$ exponential(θ), so $F(x) = 1 - e^{-x/\theta}$. Now let $Y = cX$. Then,

$$F_Y(y) = Pr(Y \leq y) = Pr(cX \leq y) = Pr(X \leq \frac{y}{c}) = 1 - e^{-y/(c\theta)}$$

Thus, $F_Y(y) = 1 - e^{-y/(c\theta)}$, which implies $Y \sim exp(c\theta)$, and that the exponential distribution is a scale distribution. The form of the distribution of Y is still exponential, although the parameter is has been rescaled by a factor of c. □

2.1 Variables Related to Insurance

Now that we have defined some basic statistical terms that will help us "describe" the data mathematically in terms of shape and location, we will now proceed to investigate some variables that have direct insurance applications. Let us start with some definitions which you should recall from Exam 3.

Definition 2.3. The **survival function** for a random variable X is $S(x) = Pr(X > x) = 1 - F(x)$.

The **hazard function** for a random variable X is

$$h(x) = \frac{f(x)}{S(x)}$$

□

Note that taking the derivative of both sides in our definition for $S(x)$ above, we get another useful relationship: $S'(x) = -f(x)$. Since $F(x)$ increases monotonically from 0 to 1, we know that $S(x)$ decreases monotonically from 1 to 0. In fact, $S(\infty) = 0$ (if X is used to model a person's lifetime, then this simply reminds us that no one can live forever!). A useful fact is the following result, which should be memorized.

Theorem 2.4.

$$S(x) = e^{-\int_0^x h(t)dt}$$

Proof. Observe that

$$\begin{aligned} \frac{d}{dx}\ln S(x) &= \frac{1}{S(x)} \cdot \frac{d}{dx}S(x) \\ &= \frac{1}{S(x)} \cdot \frac{d}{dx}(1 - F(x)) \\ &= \frac{1}{S(x)} \cdot -f(x) \\ &= -\frac{f(x)}{S(x)} \end{aligned}$$

Since $\frac{d}{dx}\ln S(x) = -f(x)/S(x) = -h(x)$, we get the desired result. □

Example 2.5. Suppose that we have a constant hazard rate of $h(t) = 1$. Find the survival function and calculate the probability of surviving to time $t = 5$.

Answer We know that $S(t) = exp\{-\int_0^t h(x)dx\}$. Plugging in the appropriate values, we get $S(t) = exp\{-\int_0^t 1dx\} = e^{-t}$. To find the probability of surviving to time $t = 5$, we substitute $t = 5$ to get $S(5) = e^{-5} = 0.00674$.

□

Definition 2.6. For a given value of d such that $Pr(X > d) > 0$ (i.e., X is larger than d with positive probability), the **excess loss variable** or **per-payment random variable** is denoted $Y^P = X - d|X > d$.

The expected value of Y^P is called the **excess loss function** and is denoted $e_X(d)$ or $e(d)$:

$$e_X(d) = E(Y^P) = E(X - d|X > d)$$

□

Remark. Some notes:

1. Other names for the excess loss function include **mean excess loss function**, **mean residual life function**, and the **complete expectation of life**. The per-payment random variable is also known as the **left-truncated-and-shifted variable**.
2. $X - d|X > d$ is read "$X - d$ given $X > d$" and signifies conditioning. In other words, $X - d|X > d$ is the random variable which represents the amount that X exceeds some value d given that X exceeds d.

□

Why is such a variable useful? We give two examples. First, let X be a random variable that represents a person's lifespan. Then $e_X(d)$ is how much longer a person can expect to live given that they've survived to age d (hence the term mean residual life function and complete expectation of life). Now, imagine that X is a variable of loss. In this case, $e_X(d)$ is the additional amount an insurer is expected to pay given that the loss has exceeded a certain deductible d (hence the term mean excess loss function). This variable allows us to quantify what we expect to pay out or how long we expect a person to live given that some threshold has already been achieved.

Since we've already settled the fact that Y^P is a random variable, we might very well want to compute moments of Y^P. Since the mean, or first raw moment, of Y^P is denoted $e_X(d)$, we continue this notation and write $e_X^k(d)$ for the k-th raw moment of Y^P. To calculate $e_X^k(d)$, we compute:

$$e_X^k(d) = \int_d^\infty \frac{(x-d)^k f(x)}{1-F(d)} dx \text{ if } X \text{ is continuous} \tag{2.1}$$

or

$$e_X^k(d) = \sum_{x_j > d} \frac{(x_j - d)^k\, p(x_j)}{1-F(d)} \text{ if } X \text{ is discrete} \tag{2.2}$$

(2.1) or (2.2) can be quite messy to compute. However, there is a unique relationship between the survival function and the mean excess loss function (remember the mean excess loss function only refers to the first moment of the excess loss variable) which makes things easier.

Theorem 2.7.

$$e_X(d) = \int_d^\infty \frac{S(x)}{S(d)} dx \textit{ if } X \textit{ is continuous } \underline{\textit{or}} \textit{ discrete} \tag{2.3}$$

Note: *(2.3)show a very important relationship. Below, we will derive the continuous form in detail. You should read through this derivation and understand it. Of greater importance is knowing how to apply it. It's also key to note that even in the case of discrete variables, the integral (and* <u>*not*</u> *the sum) still applies.*

Proof. Recall the formula for integration by parts:

$$\int u\, dv = uv - \int v\, du.$$

Let $u = x - d$ so that $du = dx$, and let $v = -S(x)$ so that $dv = f(x)dx$. Then,

$$
\begin{aligned}
e_X(d) &= \int_d^\infty \frac{(x-d)f(x)}{1-F(d)}\,dx \\
&= \int_d^\infty \frac{u}{S(d)}\,dv \\
&= \frac{[(x-d)(-S(x)]|_d^\infty - \int_d^\infty -S(x)\,dx}{S(d)} \\
&= \frac{-(x-d)S(x)|_d^\infty + \int_d^\infty S(x)\,dx}{S(d)} \\
&= \int_d^\infty \frac{S(x)}{S(d)}\,dx
\end{aligned}
$$

□

Definition 2.8. The **left-censored-and-shifted variable** , or **per-loss random variable** , denoted Y^L is defined as:

$$Y^L = (X-d)_+ = \begin{cases} 0 & \text{if } X \le d, \\ X-d & \text{if } X > d. \end{cases}$$

□

We can think of Y^L as the compensation (issued by the insurer) per **L**oss. This is different from the definition of Y^P which could be thought of as the compensation per **P**ayment, hence the superscripts. Thus, while Y^L exists for any random variable X, it turns out Y^P exists only if $X > d$ (some money was actually paid by the insurer).

The expectation $E(Y^L)$ can be computed in the usual way:

$$E(Y^L) = \int_d^\infty (x-d)f(x)\,dx \text{ if } X \text{ is continuous} \tag{2.4}$$

or

$$E(Y^L) = \sum_{x_j > d} (x_j - d)\,p(x_j) \text{ if } X \text{ is discrete} \tag{2.5}$$

There is a direct relationship between $E[Y^P]$ and $E[Y^L]$:

$$E[Y^L] = E[Y^P](1-F(d)) \tag{2.6}$$

Now, we will introduce the limited loss variable. This is useful if an insurance policy has a maximum payout, such as a policy that pays up to \$100,000 dollar-for-dollar, but losses above that amount will only be reimbursed up to \$100,000.

Definition 2.9. The **limited loss random variable**, or **right-censored variable**, is

$$X \wedge u = \begin{cases} X & \text{if } X \le u, \\ u & \text{if } X > u. \end{cases}$$

□

To calculate $E(X \wedge u)$ (known as **limited expected value**), we compute the following:

$$E[(X \wedge u)^k] = \int_{-\infty}^u x^k f(x)\,dx + u^k[1-F(u)] \text{ if } X \text{ is continuous} \tag{2.7}$$

or

$$E[(X \wedge u)^k] = \sum_{x_j \le u} x_j^k\, p(x_j) + u^k[1-F(u)] \text{ if } X \text{ is discrete} \tag{2.8}$$

Example 2.10. Suppose losses are uniform from 0 to 20,000, ie. $X \sim \text{Unif}(0,\ 20,000)$.

1. Compute the 4th raw moment.
2. Compute the variance of X.
3. Compute the expected excess of the loss over 2,000 (i.e. compute $e_X(2,000)$).
4. Compute $E(Y^L) = E[(X-2,000)_+]$.
5. Compute $E(X \wedge 15,000)$.

Answer Note that X is defined on the domain 0 to 20,000, so in our solutions where we evaluate an integral from $-\infty$ to ∞, we can simply change the limits of integration to 0 to 20,000. This is because $f(x) = 0$ for $x < 0$ and $x > 2000$. Everywhere else, $f(x) = 1/20,000$.

1. We are given $k = 4$ and need to compute μ_k. Thus, referring back to (1.4),

$$\mu_4 = E[X^4] = \int_0^{20,000} \frac{x^4}{20,000} dx = \left. \frac{x^5}{5(20,000)} \right|_0^{20,000} = 3.2 \times 10^{16}$$

2. We compute the variance in a similar way, but first we need to find the mean μ, as it is part of the formula for the variance.

$$\mu = E(X) = \int_0^{20,000} \frac{x}{20,000} dx = \left. \frac{x^2}{2(20,000)} \right|_0^{20,000} = 10,000$$

Now, we are ready to plug this into our formula for variance.

$$Var(X) = E[(X-10,000)^2] = \int_0^{20,000} \frac{(x-10,000)^2}{20,000} dx = \left. \frac{(x-10,000)^3}{3(20,000)} \right|_0^{20,000} = 33,333,333$$

3. Recall that for a $\text{Unif}(0,\ 20,000)$ distribution, the CDF $F(x) = x/20,000$. Using (2.1)

$$\begin{aligned} e_X(2000) &= E(Y^P) = E(X-2000|X>2000) \\ &= \int_{2000}^{\infty} \frac{(x-2000)f(x)dx}{1-(2000/20,000)} \\ &= \int_{2000}^{20,000} \frac{(x-2,000)f(x)dx}{1-(0.1)} \\ &= \int_{2000}^{20,000} \frac{(x-2000)}{0.9} \cdot \frac{1}{20,000} dx \\ &= 9000 \end{aligned}$$

4.

$$\begin{aligned} E(Y^L) &= \int_{2000}^{20,000} \frac{x-2000}{20,000} dx \\ &= 8100 \end{aligned}$$

5.

$$\begin{aligned} E(X \wedge 15,000) &= \int_0^{15,000} xf(x)dx + u[1-F(u)] \\ &= \int_0^{15,000} \frac{x}{20,000} dx + 15,000[1-15,000/20,000] \\ &= 5625 + 3750 \\ &= 9375 \end{aligned}$$

□

Example 2.11. Suppose losses, X, are constant at 500. For $d = 100$, calculate $e_x(d)$ using:

1. equation (2.2)

2. equation (2.3)

Answer 1. We begin by looking at equation (2.2):

$$e_X(d) = \sum_{x_j > d} \frac{(x_j - d)\,p(x_j)}{1 - F(d)}$$

We know that every loss is above 100, so that means $F(100) = Pr(X \leq 100) = 0$ or $1 - F(100) = 1$. Now for the numerator, we know that there is only one loss possible: namely $X = 500$. This implies $p(500) = 1$. Substituting in, we get:

$$\begin{aligned} e_X(d) &= \sum_{x_j > d} \frac{(x_j - d)\,p(x_j)}{1 - F(d)} \\ &= \frac{(500 - 100) \cdot 1}{1} \\ &= 400 \end{aligned}$$

2. Now we'll use equation (2.3) to arrive at the same result. Recall that (2.3) states:

$$e_X(d) = \int_d^\infty \frac{S(x)}{S(d)} dx$$

$S(100) = 1 - F(100) = 1 - Pr(X \leq 100) = 1 - 0 = 1$. Therefore, we just need to solve $\int_d^\infty S(x)$. Due to the discrete nature of our data, we can write $F(x)$ as follows:

$$F(x) = \begin{cases} 0 & x < 500 \\ 1 & x \geq 500 \end{cases}$$

This is equivalent to:

$$S(x) = \begin{cases} 1 & x < 500 \\ 0 & x \geq 500 \end{cases}$$

Finally, $\int_d^\infty S(x)\ dx = \int_{100}^\infty S(x)\ dx = \int_{100}^{500} 1\ dx + \int_{500}^\infty 0\ dx = 400 + 0 = 400$. The final answer comes to be $\frac{400}{1} = 400$, as before.

□

2.2 Distribution Tail Comparisons

The right tail of a distribution shows what the distribution does at large values of a random variable. This is of great concern to the actuary because insuring losses where large values of the random variable have high probabilities means paying much more in total losses. Thus, we need to be able to talk about whether a distribution is **heavy-tailed** (where high probabilities are concentrated at large values of a random variable) or **light-tailed** (where low probabilities are concentrated at large values of a random variable). We also want to be able to compare distributions of different random variables to determine which one has a heavier tail. Below are four ways in which we evaluate the tails.

2.2.1 Using Raw Moments

As mentioned earlier, the k-th raw moment is $E(X^k) = \int x^k f(x)dx$. For some distributions and values of k, $E(X^k)$ may not exist (be infinite).

It is generally agreed upon that distributions for which $E(X^k) < \infty$ for all $k > 0$ are considered light tailed while distributions for which $E(X^k) < \infty$ only up to some specific k are considered heavy tailed.

Example 2.12. Consider the gamma distribution with

$$\begin{aligned} E(X^k) &= \int_0^\infty x^k \frac{x^{\alpha-1}e^{-x/\theta}}{\Gamma(\alpha)\theta^\alpha}dx \quad (\text{Let } y = \frac{x}{\theta} \Rightarrow \frac{dy}{dx} = \frac{1}{\theta} \Rightarrow dx = \theta dy) \\ &= \frac{1}{\Gamma(\alpha)\theta^\alpha}\int_0^\infty (y\theta)^k (y\theta)^{\alpha-1}e^{-y}\theta dy \quad (\text{Observe as } x \to 0 \text{ to } \infty,\ y \to 0 \text{ to } \infty.) \\ &= \frac{\theta^k}{\Gamma(\alpha)}\int_0^\infty y^{k+\alpha-1}e^{-y}dy \\ &= \frac{\theta^k}{\Gamma(\alpha)}\Gamma(\alpha+k) \quad (\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt \text{ from the Tables}) \\ &< \infty \quad \text{for all } k > 0 \end{aligned}$$

Thus the gamma distribution is light tailed. □

Example 2.13. Consider the Pareto distribution in the same manner.

$$\begin{aligned} E(X^k) &= \int_0^\infty \frac{x^k \alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}dx \quad (\text{Let } y = x+\theta \Rightarrow \frac{dy}{dx} = 1 \Rightarrow dy = dx) \\ &= \int_\theta^\infty (y-\theta)^k \frac{\alpha\theta^\alpha}{y^{\alpha+1}}dy \quad (\text{as } x \to 0 \text{ to } \infty, y \to \theta \text{ to } \infty) \\ &= \alpha\theta^\alpha \int_\theta^\infty \frac{(y-\theta)^k}{y^{\alpha+1}}dy \end{aligned}$$

From here, one can see that not all moments exist. For example, if $k > \alpha + 1$, then the integral will blow up as y approaches infinity. Thus, the Pareto distribution is considered heavy-tailed. □

Note that one could have easily used moments from the Exam Tables for the above examples rather than computing them from scratch, although the calculus presented here is a great workout! The Exam Tables further state that only some moments exist for the Pareto distribution.

2.2.2 Limiting Tail Behavior

If a two random variables X_1 and X_2 have corresponding survival functions $S_1(x)$ and $S_2(x)$ which satisfy

$$\lim_{x\to\infty} \frac{S_1(x)}{S_2(x)} = \infty$$

then the distribution of X_1 has a heavier tail than the distribution of X_2. By an application of L'Hospital's rule, an equivalent criterion is

$$\lim_{x\to\infty} \frac{f_1(x)}{f_2(x)} = \infty$$

Example 2.14. Let us again consider the case of the Pareto and gamma distributions. We will write the parameters of the gamma distribution using τ and λ in order to avoid using the same symbols in both densities. Observe that

$$\begin{aligned} \lim_{x\to\infty} \frac{f_{\text{pareto}}(x)}{f_{\text{gamma}}(x)} &= \lim_{x\to\infty} \frac{\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}}{\frac{x^{\tau-1}e^{-x/\lambda}}{\Gamma(\tau)\lambda^\tau}} \\ &= \lim_{x\to\infty} \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} \frac{\Gamma(\tau)\lambda^\tau}{x^{\tau-1}e^{-x/\lambda}} \\ &= \lim_{x\to\infty} C \cdot \frac{e^{x/\lambda}}{(x+\theta)^{\alpha+1}x^{\tau-1}} \end{aligned}$$

where $C = \alpha\theta^{\alpha}\Gamma(\tau)\lambda^{\tau}$ is simply a constant. Observe that the numerator is exponential while the denominator is a polynomial. Recalling that exponentials tend to ∞ much faster than polynomials do, we see that

$$\lim_{x\to\infty} C \cdot \frac{e^{x/\lambda}}{(x+\theta)^{\alpha+1}x^{\tau-1}} = \infty$$

Thus, the Pareto distribution is heavier-tailed than the gamma distribution. □

2.2.3 Hazard Rate Functions

One can also compare the hazard rate function $h(x) = \frac{f(x)}{S(x)}$. If $h(x)$ decreases as x increases, then X is heavy-tailed. If $h(x)$ increases as x increases, then X is light-tailed. If $h(x)$ is a constant with respect to x, then no conclusions can be made.

Furthermore, a distribution has a lighter tail than another if its hazard rate function is increasing at a faster rate.

Example 2.15. Let us now see if the Pareto distribution is heavy-tailed using the hazard rate.

$$h(x) = \frac{\frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}}{(\frac{\theta}{x+\theta})^{\alpha}} = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}}\frac{(x+\theta)^{\alpha}}{\theta^{\alpha}} = \frac{\alpha}{x+\theta}$$

Thus $h(x)$ decreases as x increases. This shows the distribution to be heavy-tailed.

□

2.2.4 Mean Excess Loss Function

If the mean excess loss function $e(d)$ is increasing in d, then the distribution is heavy-tailed. If the mean excess loss function is decreasing in d, then it is considered light-tailed. Similarly, if $e(d)$ is a constant with respect to d, then no conclusions can be drawn.

Example 2.16. For this, let us consider the exponential distribution.

$$\begin{aligned} e(d) &= \frac{\int_d^\infty S(x)dx}{S(d)} \\ &= \frac{\int_d^\infty e^{-x/\theta}dx}{e^{-d/\theta}} \\ &= \frac{-\theta e^{-x/\theta}\big|_d^\infty}{e^{-d/\theta}} \\ &= \frac{-\theta}{e^{-d/\theta}}\left(-\frac{1}{e^{d/\theta}}\right) \\ &= \theta \end{aligned}$$

This is a constant, independent of d, yielding no information regarding whether the tail is heavy or light, suggesting we seek alternate methods. For the exponential distribution, the hazard rate calculation gives

$$h(x) = \frac{f(x)}{S(x)} = \frac{\frac{e^{-x/\theta}}{\theta}}{e^{-x/\theta}} = \frac{1}{\theta}$$

Note that this method also yields a constant. It is worth trying several different methods if one proves inconclusive or too difficult to apply. In fact, we could easily take a look at the k-th raw moment on the Exam Tables to find that $E(X^k) = \theta^k k! < \infty$ for all k. Thus, we conclude by Method 1 that the distribution is light-tailed. □

2.3 Measures of Risk

Actuaries, along with other financial professionals, often need to know the risk they are exposing their company to. Of course, the actuary can look at the entire book of business (and should periodically!), but that is oftentimes a long process and the results are difficult to communicate with others. To this extent, something called *value-at-risk* or **VaR** has been created. It helps to summarize the risk exposure in one number that can then easily be presented to others.

However, there is of course no such thing as a free lunch, and VaR suffers from several drawbacks. To help negate these drawbacks, something called *tail value-at-risk* or TVaR has been created (other names for TVaR include **conditional value-at-risk** (cVaR), **conditional tail expectation** (CTE) and **expected shortfall** (ES)). We will look at both of these measures and see their strengths and weaknesses.

A word of caution: although not on the syllabus, the real estate bubble around 2006 was in part caused by an over-reliance on the VaR measure. Although VaR is useful and informative, one must be keenly aware of its weaknesses.

2.3.1 Coherence

Basically, a risk measure ρ quantifies the level of risk inherent in a random variable X by mapping X to some number on the real line $\rho(X)$.

Standard deviation or some multiple of it is an example of a risk measure because it intuitively provides a measure of uncertainty.

We will now consider some desired properties that risk measures should possess.

Definition 2.17. Let there be two random variables X and Y. A **coherent** risk measure satisfies the following 4 properties:

1. Subadditivity : $\rho(X+Y) \leq \rho(X)+\rho(Y)$
2. Monotonicity : If $X \leq Y$, then $\rho(X) \leq \rho(Y)$.
3. Positive homogeneity : $\rho(cX) = c\rho(X)$
4. Translation invariance : $\rho(X+c) = \rho(X)+c$, where c is a constant.

□

Let's tackle each part of the definition in turn:

1. **Subadditivity** simply implies that there is usually a diversification benefit. In essence, by combining risks, the overall risk goes down. If this were not the case, large insurers would never exist, because it would be more beneficial to handle losses individually rather than in aggregate.
2. **Monotonicity**: This is mostly self-evident: a random variable associated with greater loss ought to be riskier.
3. **Positive Homogeneity**: This condition deals with currencies and proportional risks. First, the risk measure should be independent of currency. For example, if we insure a house in \$ or €, the underlying uncertainty which is being measured by the risk measure should not change (except for the conversion rate for \$ to €-assuming no currency risk).

   Furthermore, if we double our exposure, the risk measure should double as well. Comparing two policies for insuring a \$500,000 house (one that pays out \$500,000 versus one that pays out \$250,000), the risk measure for the \$500,000 policy should be double that of the \$250,000 policy. This is because the *inherent risk* of total loss is the same. The only thing different is the payout. Thus, $\rho(2X) = 2\rho(X)$.
4. **Translation Invariance**: This simply means if you add a risk for which there is no inherent uncertainty, the risk simply increases linearly. If we pay \$500,000 if the house burns down and \$1000 at the end of the year regardless of what happens, $\rho(X+c) = \rho(X)+c$, where $c = \$1000$ and

$$X = \begin{cases} 500,000 & \text{if total loss} \\ 0 & \text{otherwise} \end{cases}$$

In other words, since we know we have to pay \$1000 at the end of the year, there is no *inherent uncertainty* for the \$1000 payout.

2.3.2 Value-at-Risk (VaR)

VaR was created in order to compute the amount of capital an insurer must hold to guarantee with a high probability that the insurer does not become insolvent. Thus, first you pick your guarantee probability (99%, 95%, etc.), and then you use this probability to compute the VaR.

Definition 2.18. Suppose X is a loss random variable that takes on only positive values. The **value-at-risk (VaR)** of X at the $100p\%$ level (denoted $VaR_p(x)$ or $\pi_p(x)$) is equal to the $100p$-th percentile of the distribution of X. □

Refer back to our discussion on percentiles in Section 1.1 if necessary.

Example 2.19. Suppose $X \sim \text{Unif}(0,\ 100{,}000)$. Compute $VaR_{95\%}(x)$.

Answer Since $F(95{,}000) = 0.95$, we have $VaR_{95\%}(x) = \$95{,}000$. If X represents claim amounts, this implies that with 95% certainty, by holding \$95,000, we will have enough money to pay all claims.

□

While VaR is useful, it fails to satisfy the coherence properties described in Definition 2.17. Specifically, VaR is not subadditive. To show this, let us consider a somewhat contrived example which we construct solely for the purpose of demonstrating this fact.

Example 2.20. Construct a random variable Z such that:

$$F_Z(10) = 0.90$$
$$F_Z(90) = 0.975$$
$$F_Z(100) = 0.995$$

Observe that $VaR_{97.5\%}(Z) = 90$. Suppose Z represents a policy covering total fire losses, which we can equivalently split up into 2 policies: X is a policy which covers fire losses $\leq$ \$100 and Y is a policy which covers fire losses exceeding \$100. We define $Z = X + Y$ via

$$X = \begin{cases} Z & Z \leq 100 \\ 0 & Z > 100 \end{cases}$$

$$Y = \begin{cases} 0 & Z \leq 100 \\ Z & Z > 100 \end{cases}$$

Then, the following statements are consistent with our definition of the CDF of Z:

$$F_X(10) = 0.975$$
$$F_X(90) = 0.99$$
$$F_X(100) = 1$$

These statements are not the only set of possible values for $F_X(10)$, $F_X(90)$, and $F_X(100)$, but they do not contradict our definition of the CDF of Z and serve to prove our point. Note that the only constraints we have for the CDF $F_X(x)$ is that $F_X(x) \geq F_Z(x)$ and $F_X(100) = 1$ since $Pr(X > 100) = 0$.

From the above statements, we conclude $VaR_{97.5\%}(X) = 10$. For Y, we know that $F_Y(0) = 0.995$, since

$$F_Y(0) = Pr(Y \leq 0) = Pr(Z \leq 100) = F_Z(100) = 0.995$$

Thus $VaR_{97.5\%}(Y) = 0$, from which we discover

$$VaR_{97.5\%}(X) + VaR_{97.5\%}(Y) = 10 + 0 = 10 \leq VaR_{97.5\%}(Z)$$

We have reached a violation of the subadditivity property. □

We readily admit that this example is contrived. However, this also shows that an unscrupulous CFO or risk manager can potentially break up a risk into smaller risks (with the same total risk) but come up with a lower required VaR at the same probability level. This leaves the company in a poor situation if large losses occur. Thus, a better metric is needed. TVaR (tail Value-at-Risk) was developed for exactly this reason.

2.3.3 TVaR

Initially, VaR was developed by the financial industry to help estimate exposure on financial trades. Since the risk was estimated over a short and fixed timeframe, the normal distribution was often used to model gains or losses. If one restricts the VaR calculations to the normal distribution, all the coherency requirements hold. Unfortunately, insurance losses are not normally distributed but are actually skewed. As such, TVaR was invented to mitigate the shortcomings of VaR for skewed distributions.

Definition 2.21. Suppose X is a loss random variable that takes on only positive values. The **tail value-at-risk (TVaR)** of X at the $100p$-th level (denoted $TVaR_p(X)$) is the expected loss given that the loss exceeds the $100p$-th percentile of the distribution of X. □

TVaR is a measure of expected loss given that the losses are high enough. In other words, it measures how bad things can get if the situation has already turned sour.

Theorem 2.22. *Below are 3 equivalent formulas for TVaR:*

$$TVaR_p(X) = E(X|X > \pi_p) = \frac{\int_{\pi_p}^{\infty} xf(x)dx}{1-p} \tag{2.9}$$

$$= \frac{\int_p^1 VaR_t(X)dt}{1-p} \tag{2.10}$$

$$= VaR_p(X) + e_X(\pi_p) \tag{2.11}$$

where $e_X(\pi_p)$ is defined as in Definition 2.6.

Proof. By the linearity property of expectations,

$$E(X|X > \pi_p) = E(X - \pi_p|X > \pi_p) + E(\pi_p|X > \pi_p) = e_X(\pi_p) + \pi_p = e_X(\pi_p) + VaR_p(X)$$

This shows (2.11). Now, substituting $d = \pi_p$ in (2.1),

$$e(\pi_p) = \frac{\int_{\pi_p}^{\infty}(x - \pi_p)f(x)dx}{1-p} \tag{2.12}$$

Note that

$$\pi_p = \pi_p \cdot \underbrace{\frac{\int_{\pi_p}^{\infty} f(x)dx}{1 - F(\pi_p)}}_{1} = \frac{\int_{\pi_p}^{\infty} \pi_p f(x)dx}{1-p}$$

Adding our expression for $e_X(\pi_p)$ in (2.12) to π_p gives us (2.9). Now, we show that (2.9) is equivalent to (2.10), by letting $t = F(x)$ which implies $dt = f(x)dx$ and $x = F^{-1}(t) = VaR_t(x)$. Thus, we make a change of variable in the numerator. Note that the limits of integration change as follows: as x goes from π_p to ∞, u goes from $F(\pi_p) = p$ to $F(\infty) = 1$.

$$\frac{\int_{\pi_p}^{\infty} xf(x)dx}{1-p} = \frac{\int_p^1 VaR_t(x)dt}{1-p}$$

□

Example 2.23. Let X be an exponential random variable with mean θ. From the Exam Tables we know that

$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}} \quad \text{and} \quad F(x) = 1 - e^{-\frac{x}{\theta}}$$

Thus $VaR_p(X)$ is the value of π_p such that $Pr(X > \pi_p) = 1 - p$.
Observe that $Pr(X > \pi_p) = 1 - (1 - e^{-\frac{\pi_p}{\theta}}) = e^{-\frac{\pi_p}{\theta}}$. Thus,

$$\begin{aligned} Pr(X > \pi_p) = 1 - p \quad &\Rightarrow \quad e^{-\frac{\pi_p}{\theta}} = 1 - p \\ &\Rightarrow \quad -\frac{\pi_p}{\theta} = \ln(1-p) \\ &\Rightarrow \quad \pi_p = -\theta \ln(1-p) \end{aligned}$$

To compute $TVaR_p(X)$, we now use formula (2.10).

$$\frac{\int_p^1 VaR_t(X)dt}{1-p} = \frac{\int_p^1 -\theta\ln(1-t)dt}{1-p} = \frac{-\theta}{1-p}\int_p^1 \ln(1-t)dt \tag{2.13}$$

Now we need to compute this integral. To do this there are 2 steps, first substitution and then integration by parts (but don't worry, we'll show a trick at the end so that you won't have to do this every time).

Part 1: Substitution

Let $x = 1-t \Rightarrow dx = -dt$. Note as t goes from p to 1, x goes from $1-p$ to 0. Thus,

$$\int_p^1 \ln(1-t)dt = -\int_{1-p}^0 \ln x dx$$

Part 2: Integration by parts

Let u = $\ln x \Rightarrow du = \frac{1}{x}dx$ and $v = x \Rightarrow dv = dx$. Thus

$$\begin{aligned}
-\int_{1-p}^0 \ln x dx &= -\int_{1-p}^0 udv \\
&= -\left[uv|_{1-p}^0 - \int_0^{1-p} vdu\right] \\
&= -\left[x\ln x|_{1-p}^0 - \int_{1-p}^0 x\frac{1}{x}dx\right] \\
&= -\left[x\ln x|_{1-p}^0 - \int_{1-p}^0 dx\right] \\
&= -\left[x\ln x|_{1-p}^0 - x|_{1-p}^0\right] \\
&= -\left[x\ln x|_{1-p}^0 + (1-p)\right]
\end{aligned}$$

Now to evaluate $x\ln x$ as $x \to 0$, we use L'Hospital's rule. Observe that $x\ln x = \frac{\ln x}{1/x}$. Differentiating the numerator and denominator gives

$$\frac{\frac{1}{x}}{-\frac{1}{x^2}} = -x$$

As $x \to 0$, $-x \to 0$. Thus, we see that $-[x\ln x|_{1-p}^0 + (1-p)]$ becomes

$$-[0-(1-p)\ln(1-p)+(1-p)] = (1-p)\ln(1-p)-(1-p)]$$

Lastly, plugging into (2.13), we get

$$\frac{-\theta}{1-p}[(1-p)\ln(1-p)-(1-p) = -\theta\ln(1-p)+\theta = VaR_p(x)+\theta$$

Whew! That was a long and hard computation. Although instructive to see it once (and to practice a few times), it is way, way too tedious to do on the exam. If you find yourself doing something this involved, reconsider your attack plan. The easier way is to make use of (2.11), recalling that the exponential distribution is memoryless, so that $e(\pi_p) = \theta$. Remember that the memoryless property implies $P(X > \pi_p + x | X > \pi_p) = P(X > x)$.

Then using formula (2.11),

$$TVaR_p(X) = VaR_p(X) + e(\pi_p) = VaR_p(X) + \theta$$

See this is much shorter, easier, and has less room for mistakes. We cannot stress enough the importance to know various ways to attack the same problem. Here is a perfect example. In short, know all 3 formulas.

Of course, note that $TVaR_p(X)$ for the exponential distribution is written in your formula packet, so as always, check the formula sheet first. □

We will now present one last formula for $TVaR_p(X)$ which only uses $\pi_p(X), E(X)$, and $E(X\wedge\pi_p)$. Since these parts are shown in the formula sheet for almost all the distributions, this often becomes the simplest way to use when one of the first 3 formulas does not immediately become a trivial calculation.

Theorem 2.24.

$$TVaR_p(X) = \pi_p + \frac{E(X)-E(X\wedge\pi_p)}{1-p} \tag{2.14}$$

Proof.

$$\begin{aligned}
TVaR_p(X) &= E(X|X>\pi_p)\\
&= \pi_p + \frac{\int_{\pi_p}^{\infty}(x-\pi_p)f(x)dx}{1-p}\\
&= \pi_p + \frac{\int_{-\infty}^{\infty}(x-\pi_p)f(x)dx - \int_{-\infty}^{\pi_p}(x-\pi_p)f(x)dx}{1-p}\\
&= \pi_p + \frac{\int_{-\infty}^{\infty}xf(x)dx - \int_{-\infty}^{\pi_p}xf(x)dx - \pi_p(\int_{-\infty}^{\infty}f(x)dx - \int_{-\infty}^{\pi_p}f(x)dx)}{1-p}\\
&= \pi_p + \frac{E(X) - \int_{-\infty}^{\pi_p}xf(x)dx - \pi_p(1-F(\pi_p))}{1-p}\\
&= \pi_p + \frac{E(X)-E(X\wedge\pi_p)}{1-p}
\end{aligned}$$

□

Thus, when in doubt, use this formula and simply plug in. You should absolutely memorize this formula. Let's do a numerical example for the exponential distribution in a few ways to see how this works.

Example 2.25. Let us take an exponential distribution with $\theta = 100$ (thus the expected loss is \$100) and compute $VaR_{99\%}(X)$ and $TVaR_{99\%}(X)$.

We will use (2.11) and (2.14). This way, you see both methods via example.

Using (2.11):

$$VaR_{99\%}(X) = -100\ln(1-0.99) = 460.517$$
$$TVaR_{99\%}(X) = VaR_{99\%}(X) + \theta = 460.517 + 100 = 560.517$$

Using (2.14):
Recalling that $TVaR_p(X) = VaR_p(X) + \frac{E(X)-E(X\wedge\pi_p)}{1-p}$, we need to compute: $\frac{E(X)-E(X\wedge\pi_p)}{1-p}$. Observe from the formula sheet that $E(X\wedge x) = \theta(1-e^{-x/\theta})$. Thus, in our case $E(X\wedge\pi_p) = 100(1-e^{-460.517/100}) = 99$. Hence,

$$\frac{E(X)-E(X\wedge\pi_p)}{1-p} = \frac{100-99}{0.01} = 100$$

Thus, $TVaR_{99\%}(X) = 460.517 + 100 = 560.517$.

Observe that this is the same result as the one obtained using method 1. Although in this case, when one has the formula sheet it is easier to use method 1, in many cases when the formula for $TVaR_p(X)$ is not directly provided, (2.14) provides the easiest computation method. □

2.3.4 Extreme Value Distributions

The exam requires a conceptual understanding of two types of extreme value distributions. It is unlikely that any problems involving these distributions will require computations.

Extreme value distributions are used to model things that are "worst-case", in some sense. For example, *VaR* and *TVaR* are examples of extreme values. There are 2 types of extreme value distributions you should know, and we describe them below.

1. **The maximum observation from a random sample.**
   This is equivalent to $VaR_{100\%}(x)$. The Fréchet distribution (more familiarly known as, and included in your formula sheet as, the inverse Weibull distribution) is most often used for modeling this random variable. One reason is that $VaR_{100\%}(x)$ tends to have a heavy tail, which is a characteristic of the inverse Weibull distribution.

2. $\mathbf{E}[(\mathbf{X}-\mathbf{d})_+]$ **as** $\mathbf{d} \to \infty$.
   In an insurance setting, the limiting distribution of this mean excess loss random variable is modeled using the exponential distribution if X is light-tailed, and is modeled via the Pareto distribution if X is heavy-tailed.

   The reason why insurance companies care about this distribution is that some insurance policies have high deductibles (i.e. d is large), and there is a need to understand their average payouts in this setting.

2.4 Mixtures

Definition 2.26. A random variable Y is a k**-point mixture** of random variables $X_1, \ldots, X_k$ if its cumulative distribution function is given by

$$F_Y(y) = a_1 F_{X_1}(y) + \cdots + a_k F_{X_k}(y) \quad \text{where } \sum_{i=1}^{k} a_i = 1 \tag{2.15}$$

□

Thus, the $\{a_i\}$ can be thought of as weights for a mixture of k distributions.

Example 2.27. Let $X_1 \sim exp(\theta_1)$ and $X_2 \sim exp(\theta_2)$. Define a random variable Y as follows: Y follows the distribution of X_1 with a 30% probability and follows the distribution of X_2 with a 70% probability. The 2-point mixture has a cumulative distribution function of

$$\begin{aligned} F_Y(y) &= 0.3F_{X_1}(y) + 0.7F_{X_2}(y) \\ &= 0.3(1-e^{-y/\theta_1}) + 0.7(1-e^{-y/\theta_2}) \\ &= 0.3 + 0.7 - 0.3e^{-y/\theta_1} - 0.7e^{-y/\theta_2} \\ &= 1 - 0.3e^{-y/\theta_1} - 0.7e^{-y/\theta_2} \end{aligned}$$

□

Definition 2.28. A random variable Y is a **variable-component mixture** if it is a k-point mixture where $k \in \{1,2,3,\ldots\}$ and is not predetermined.

$$F_Y(y) = \sum_{j=1}^{k} a_j F_j(x) \quad \text{where } \sum_{j=1}^{k} a_j = 1 \quad \text{and } k \in \{1,2,3,\ldots\}$$

□

A variable-component mixture is so named because k, the number of components is not fixed. This is something to keep in the back of your mind; we wouldn't worry too much about these as most problems will have k set beforehand.

One fact to note is that a k-point mixture can be equivalently written as a mixture of probability density functions instead of a mixture of cumulative distribution functions as in (2.15). Observe that if

$$F_Y(y) = a_1 F_{X_1}(y) + \cdots + a_k F_{X_k}(y)$$

then taking the derivatives of both sides,

$$\begin{aligned} f_Y(y) = F'_Y(y) &= a_1 F'_{X_1}(y) + \cdots + a_k F'_{X_k}(y) \\ &= a_1 f_{X_1}(y) + \cdots + a_k f_{X_k}(y) \end{aligned}$$

2.5 Problems for Section 2

1. Let $X \sim exp(\theta)$. Compute $e(d)$.
2. Suppose losses are Pareto($\alpha = 3, \theta = 100$). Compute the mean excess loss at 5,000.
3. Compare the tail weight of Pareto and exponential distributions using various methods. Do not assume specific values for the parameters of the distributions.
4. Compute the $VaR_{95\%}(X)$ and $TVaR_{95\%}(X)$ for the single parameter Pareto distribution.
5. 20% of claims follow a $N(\mu = 5,000, \sigma^2 = 500,000)$ distribution and 80% of claims folow a $N(\mu = 7,000, \sigma^2 = 1,000,000)$ distribution. What is the probability that a random claim will exceed \$6,000?
6. Let claims be distributed $exp(\theta)$ this year. Claims in the next two years will experience 5% and 7% inflation, respectively. (Hint: You will need to rescale the random variable for claims in 2 years.) Define α as follows:

$$\alpha = \frac{Pr(\text{claims} > d \text{ in 2 years})}{Pr(\text{claims} > d \text{ this year})}$$

   Compute α for $\theta = 100$ and $d = 150$.

2.6 Solutions for Section 2

1.

$$\begin{aligned} e(d) &= \frac{\int_d^\infty S(x)dx}{S(d)} \\ &= \frac{\int_d^\infty e^{-x/\theta}dx}{e^{-d/\theta}} \\ &= \frac{-\theta}{e^{-d/\theta}}\left(-\frac{1}{e^{d/\theta}}\right) \\ &= \theta \end{aligned}$$

2. First observe that we can rewrite the formula for $e(d)$ as follows:

$$e(d) = \frac{\int_d^\infty (x-d)f(x)dx}{S(d)} = \frac{E(x) - E(x \wedge d)}{S(d)}$$

   You should memorize this as another expression for $e(d)$ as it is extremely useful. Notice that the formula sheet has expressions already for $E(X), E(X \wedge d)$, and $F(x)$ (and thus trivially $S(x)$). Thus, you can use this formula to plug in and directly compute values for $e(d)$.

   Having the above derivation we now proceed as follows:

$$\begin{aligned} e(d) &= \frac{E(X) - E(X \wedge d)}{S(d)} \\ &= \frac{\frac{\theta}{\alpha-1} - \frac{\theta}{\alpha-1}\left[1 - \left(\frac{\theta}{\theta+d}\right)^{\alpha-1}\right]}{\left(\frac{\theta}{\theta+d}\right)^{\alpha}} \\ &= \frac{\theta}{\alpha-1}\frac{\theta+d}{\theta} \\ &= \frac{\theta+d}{\alpha-1} \end{aligned}$$

   Using this formula, we plug in $d = 5000$ and values for α and θ to get

$$e(5000) = \frac{100+5000}{3-1} = 2550$$

3. We can use various methods to explore the tail weight.

Method 1: Moments
We compute $E(X^k)$ for the Pareto distribution and the exponential distribution separately. For the Pareto distribution, we have

$$\begin{aligned} E(X^k) &= \int_0^\infty \frac{x^k \alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} dx \quad (\text{Let } y = x+\theta \Rightarrow \frac{dy}{dx} = 1 \Rightarrow dy = dx) \\ &= \int_\theta^\infty (y-\theta)^k \frac{\alpha\theta^\alpha}{y^{\alpha+1}} dy \quad (\text{as } x \to 0 \text{ to } \infty, y \to \theta \text{ to } \infty) \\ &= \alpha\theta^\alpha \int_\theta^\infty \frac{(y-\theta)^k}{y^{\alpha+1}} dy \end{aligned}$$

As seen earlier, this final integral implies that not all moments exist for the Pareto distribution. If $k > \alpha+1$, the integral blows up as $y \to \infty$. Hence, the Pareto distribution is heavy-tailed.

In contrast, for the exponential distribution, we have $E(X^k) = \theta^k k!$ from the Exam Tables. Hence, the exponential distribution is light-tailed by our criterion in Method 1.

Method 2: Limiting Tail Behavior
We compute the limit of the ratio of the PDFs of the two distributions as $x \to \infty$.

$$\begin{aligned} \lim_{x\to\infty} \frac{f_{\text{pareto}}(x)}{f_{\text{exponential}}(x)} &= \lim \frac{\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}}{\frac{1}{\theta}e^{-x/\theta}} \\ &= \lim \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} \cdot \theta e^{x/\theta} \\ &= C \cdot \lim \frac{e^{x/\theta}}{(x+\theta)^{\alpha+1}} \end{aligned}$$

Note that the numerator contains an exponential function, which goes to infinity faster than the polynomial in the denominator. Hence, the limit is infinite. We conclude that comparatively, the Pareto distribution has a heavier tail.

Method 3: Hazard Rate
Next, we can try the hazard rate method by computing for each distribution

$$h(x) = \frac{f(x)}{S(x)}$$

For the Pareto distribution, this becomes

$$h(x) = \frac{\frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}}{(\frac{\theta}{x+\theta})^\alpha} = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}} \frac{(x+\theta)^\alpha}{\theta^\alpha} = \frac{\alpha}{x+\theta}$$

$h(x)$ gets smaller as x increases. We conclude the distribution to be heavy-tailed.

As for the exponential distribution,

$$h(x) = \frac{1/\theta e^{-x/\theta}}{e^{-x/\theta}} = \frac{1}{\theta}$$

Since this is a constant with respect to x, no conclusion can be drawn from this method regarding the tails of the exponential distribution. We rely on the other methods to determine that the exponential distribution is actually light-tailed.

4. Recall that for the single-parameter Pareto distribution, $x > \theta$. For the $VaR_p(X) = c$, we use the Exam Tables to find

$$P(X \leq c) = F_X(c) = p = 1 - \left(\frac{\theta}{c}\right)^{\alpha}$$

Solving for c in the equation, we get $VaR_p(X) = c = \frac{\theta}{(1-p)^{1/\alpha}}$. Hence, setting $p = 0.95$, we get

$$VaR_{95\%}(X) = \frac{\theta}{0.05^{1/\alpha}}$$

We can use one of several formulas to get $TVaR_p(X)$. Since we already know the $VaR_p(X)$, the easiest to use is

$$\begin{aligned}
TVaR_p(X) &= \frac{\int_p^1 VaR_p(X)}{1-p} \\
&= \frac{\int_p^1 \theta(1-t)^{-1/\alpha} dt}{1-p} \\
&= \frac{\theta}{1-p} \int_p^1 (1-t)^{-1/\alpha} dt \\
&= \frac{\theta}{1-p} \cdot -\frac{(1-t)^{-1/\alpha+1}}{-1/\alpha+1} \Bigg|_p^1 \quad (\text{if } \alpha > 1) \\
&= \frac{\theta}{1-p} \frac{(1-p)^{-1/\alpha+1}}{-1/\alpha+1} \\
&= \frac{\alpha\theta(1-p)^{-1/\alpha}}{\alpha - 1} \quad (\alpha > 1)
\end{aligned}$$

Plugging in our value of 0.95 for p, we get

$$TVaR_{95\%}(X) = \frac{\alpha\theta(0.05)^{-1/\alpha}}{\alpha - 1} \quad (\alpha > 1)$$

Note that oftentimes the commonly used distributions will have the VaR and $TVaR$ formula in the exam formula sheet. Still, it is good practice to try calculating these on your own, in case you get thrown a distribution that is not shown in the formula sheet and need to calculate it from scratch.

5. Observe that since we have a mixture distribution, to compute $F(6000)$ we first compute the probabilities separately in each component of the mixture, and then weight them appropriately. Let $X_1 \sim N(\mu = 5{,}000, \sigma^2 = 500{,}000)$ and let $X_2 \sim N(\mu = 7{,}000, \sigma^2 = 1{,}000{,}000)$. Then we have that:

$$\begin{aligned}
F(6000) &= 0.2P(X_1 < 6000) + 0.8P(X_2 < 6000) \\
&= 0.2\Phi\left(\frac{6000-5000}{\sqrt{500{,}000}}\right) + 0.8\Phi\left(\frac{6000-7000}{\sqrt{1{,}000{,}000}}\right) \\
&= 0.2\Phi(1.414) + 0.8\Phi(-1.00) \\
&= 0.2(.9213) + 0.8(1-.8413) \\
&= 0.3112 \\
P(X > 6000) &= 1 - 0.3112 \\
&= 0.6888
\end{aligned}$$

6. Let X be the random variable denoting claim amounts for this year (Year 0). In Year 0, we have $P(X > d) = e^{-d/\theta}$. In Year 2, our claims can be expressed as $1.05(1.07)X$. We can take advantage of the fact that the exponential distribution is a scale distribution and simply adjust our θ for the calculation. Namely, claims in Year 2 will also be exponential with parameter $\theta' = 1.05(1.07)\theta = 1.1235\theta$. Hence, after 2 years:

$$P(X > d) = e^{-d/(1.1235\theta)}.$$

Hence, dividing the two probabilities gives:

$$\alpha = \frac{e^{-d/(1.1235\theta)}}{e^{-d/\theta}}$$

Setting $\theta = 100$ and $d = 150$ yields

$$\alpha = \frac{e^{-150/(1.1235(100))}}{e^{-150/100}} \approx 1.18$$