BSc verslag TECHNISCHE WISKUNDE

"Asymptotisch gedrag van een stochast gegeven een zeldzame gebeurtenis"

(Engelse titel: " Limiting behavior of a random variable conditional on a rare event")

Koen Zomerdijk

Technische Universiteit Delft

Begeleider

Dr. R.C. Kraaij

Overig commissielid

Dr. Y. van Gennip

Juli, 2020 Delft

Abstract

As an insurer you want identify the risks you take to prevent bankruptcy. The theory of large deviations formalizes the study of such rare events. We will use the theorem of Cramér, which is a main theorem in large deviation theory, to investigate the rate at which the probability of large deviations of the sums of random variables decay. Using Sanov's theorem we will derive an expression for large deviations of the empirical measure. Furthermore, we will use Gibbs's principle to derive the distribution of random variables conditional on a large deviation.

Contents

1 Introduction

When rolling a dice, your outcome can be 1, 2, 3, 4, 5, or 6. The mean is then $\frac{1+2+3+4+5+6}{6} = 3.5$. Intuition tells us that the average of multiple throws will be approximately equal to the mean 3.5 and that more throws will probably result in an average value closer to 3.5.
In this thesis we will study the probability of a large deviation of the average, for instance throwing an average of 5 or above. The probability of throwing a 5 or above when rolling one dice is $\frac{2}{6}$. However, when rolling 10 dices the probability of throwing an average amount of 5 or above will be very low. Subsequently, when rolling 100 dices this probability will be even lower. This intuitively explains the fact that the probability of rolling an average amount of 5 or above will decay as you roll more dices. In this thesis, we are interested in what rate the probability of these large deviations decay. The occurrence of such deviations is of importance in applications as, for instance, modelling insurances.

Suppose you are the owner of a non-life insurance company. Then you earn money because your customers pay you a premium periodically. In exchange, you have to pay your customers when they make a claim for damage. There is a possibility that there is a period of extremely many claims or extremely high claims. In that case, the total amount of claims you have to pay out is much higher than you expected and it could be higher than your reserve fund. In other words, the payoff has a large deviation from the expected value of the payoff. This could lead to bankruptcy.
Naturally, an insurer does not want to go bankrupt. By analysing the probability of such a large deviation of the payoff, an insurer can choose a suitable value of the premium such that the probability of bankruptcy will be sufficiently low.

Let us put this in a mathematical framework. Suppose we are modelling a die throw or the payoff of an insurance claim. Let $X_1, X_2, ...$ be i.i.d random variables on a probability space which contain the outcome of the die throws or the values of the payoffs corresponding to the claims. Let

$$\begin{aligned}
&\mathbb{E}[X] = \mu \in \mathbb{R}, \\
&Var[X] = \sigma^2 \in (0, \infty), \\
&S_n := X_1 + ... + X_n.
\end{aligned}$$

The intuition that averages converge is reflected by the Law of large numbers, which states that

$$\frac{1}{n} S_n \xrightarrow{a.s.} \mu, \text{ as } n \to \infty.$$

This implies that

$$\mathbb{P}\left[\frac{1}{n} S_n \geq \mu + c\right] \longrightarrow 0, \quad \text{for } c > 0. \tag{1.1}$$

The second main theorem of probability theory is the Central limit theorem. This theorem describes the universal behavior of re-scaled averages. According to the Central limit theorem, one has to scale up the difference $\left(\frac{1}{n} S_n - \mu\right)$ by $\sqrt{n}$ to obtain the non-trivial limiting behavior

$$\mathbb{P}\left[\frac{1}{n} S_n \geq \mu + \frac{c}{\sqrt{n}}\right] \approx 1 - \Phi\left(\frac{c}{\sigma}\right),$$

where $\Phi(x)$ is the standard normal cumulative distribution function evaluated at x.
We see that $\frac{1}{n}S_n$ typically fluctuates around μ with a distance of order $\frac{1}{\sqrt{n}}$. Such a deviation is called a normal fluctuation.
The Law of large numbers and the Central limit theorem combined imply that the probability of all deviations of an order greater than $\frac{1}{\sqrt{n}}$ will converge to 0, as $n \to \infty$. In this thesis, we restrict ourselves to large deviations i.e. deviations of order 1, as in (1.1). We will in particular study how fast the probability of a large deviation will converge to 0, as $n \to \infty$.
In Section 2 we will see that when the moment generating function is finite, we typically have for large values of n

$$\mathbb{P}\left[\frac{1}{n}S_n \geq a\right] \approx e^{-nI(a)}, \tag{1.2}$$

where I is a strictly convex function with the property that $I\left(\mathbb{E}\left[X\right]\right) = 0$. Here $I(a)$ quantifies the exponential rate at which the tails of the distribution of S_n decay. When the moment generating function is infinite, this probability typically converges slower.

Recall the insurance problem. Suppose the payoffs $X_1, X_2, ...$ corresponding to the claims are i.i.d. random variables and the the periodically paid premium is equal to $a > \mathbb{E}\left[X\right]$. Furthermore, suppose your reserve fund only consists of the premiums paid. Then $S_n := X_1 + ... + X_n$ is the total value of the claims. Now bankruptcy occurs when the average amount of a claim $\frac{1}{n}S_n$ is higher than the periodically paid premium a and therefore the probability of bankruptcy is given by $\mathbb{P}\left(S_n > na\right)$. Now the Law of large numbers implies that the probability of bankruptcy will be equal to 0, as $n \to \infty$. However, there can just be finitely many claims and thus the probability of bankruptcy $\mathbb{P}\left(\frac{1}{n}S_n > a\right)$ will never be equal to zero. So to identify the risk you take corresponding to a premium value a, you are interested in the rate at which this probability decays as n will be very large, and therefore this risk is quantified by (1.2). Moreover, to avoid bankruptcy, you should know if a possible bankruptcy is likely to be caused by a single extremely high claim or by extremely many claims. This can be studied using the distribution of a claim conditional on a bankruptcy i.e. the distribution $\mathbb{P}\left(X_1 | \frac{1}{n}S_n > a\right)$.

In this thesis, we will see that the the moment generating function of the random variable has a huge impact on the limiting behaviour of the large deviation of its sum. In the Sections 2 up to 4, we will study the case where the moment generating function is finite. Cramér's theorem in Section 2 will state how the rate function I, which quantifies the rate of exponential decay, can be derived. For the proof of this theorem we will introduce the tilted measure $\mathbb{Q}_\mu$. Under the measure $\mathbb{Q}_\mu$, a large deviation becomes a typical event. Section 3 describes the large deviation in terms of the empirical measure. Large deviations of the payoff can be derived from large deviations of the type of claims, because the payoff is a function of the claims. If now the payoff corresponding to the claims changes you don't have to make an analysis of your claims all over again, but instead you just have to change the payoff function.
In Section 4 we will analyse the probability $\mathbb{P}\left(X_1 | \frac{1}{n}S_n > a\right)$. We will see that the distribution function of this conditional probability is given by the tilted measure $\mathbb{Q}_\mu$ which turned the large deviation to typical event.
In Section 5 we will look at the case where the moment generating function is infinite. In this case the decay of $\mathbb{P}\left(\frac{1}{n}S_n \geq a\right)$ is not exponential, but polynomial. Therefore the decay is slower than the case where the moment generating function is finite. Further-

more, we will see that if the moment generating function is infinite, a large deviation is most likely to be caused by one single random variable.

2 Rate function

We will show that under certain conditions of the moment generating function of X, that for $n \to \infty$ we have

$$\mathbb{P}\left[\frac{1}{n}S_n \geq a\right] \approx e^{-nI(a)}, \tag{2.1}$$

where the strictly convex function $I(a)$ is called the rate function. Here $I(a)$ quantifies the rate at which the tails of the distribution of S_n decay. The rate function $I(a)$ has the property that $I\left(\mathbb{E}\left[X\right]\right) = 0$, because the Law of large numbers implies that $\mathbb{P}\left[\frac{1}{n}S_n \geq a\right] = 1$. Another property of $I(a)$ is that it increases as a goes further away from $\mathbb{E}\left[X\right]$, this is because then Law of large numbers implies that $\mathbb{P}\left[\frac{1}{n}S_n \geq a\right]$ will be lower.

We will see that the rate functions differ for different probability distributions, even if they have the same mean and variance. If for example $X_i \sim Ber\left(\frac{1}{2}\right)$ and $Y_i \sim N\left(\frac{1}{2}, \frac{1}{4}\right)$, then $\mathbb{E}\left[X\right] = \mathbb{E}\left[Y\right]$ and $Var(X) = Var(Y)$. However, the probability

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq 1.5\right) = 0,$$

and therefore $I(1.5) = \infty$, while the probability

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i \geq 1.5\right) = \int_{1.5}^{\infty} \frac{1}{\sqrt{\frac{1}{2}n\pi}} e^{\frac{1}{2}\left(\frac{x-\frac{n}{2}}{\sqrt{\frac{n}{4}}}\right)^2} dx > 0,$$

and therefore $I(1.5) < \infty$.
In Section 2.2, Proposition 2.10 states the rate functions corresponding to some common probability distributions.

2.1 Cramér's theorem

Cramér's theorem [2] gives an expression for the rate function $I(a)$ of (2.1). The condition of this theorem is that the moment generating function of X is finite. We will see in Proposition 2.9 that for the exponential distribution, where the moment generating function is not finite everywhere, Cramér's theorem is also valid.

Theorem 2.1 (Cramér's Theorem)**.** *Let (X_i) be i.i.d. $\mathbb{R}$-valued random variables satisfying*

$$M_X(t) = \mathbb{E}e^{tX_1} < \infty \qquad \forall t \in \mathbb{R}.$$

Let $S_n = \sum_{i=1}^{n} X_i$. Then, for all $a > \mathbb{E}X_1$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[S_n \geq na\right] = -I(a),$$

where

$$I(a) = \sup_{\mu\in\mathbb{R}} \left[a\mu - \varphi(\mu)\right],$$

with

$$\varphi(\mu) = \log \mathbb{E}e^{\mu X}.$$

Proof. The proof of the theorem consists of two parts, the first part proves $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \geq na] \leq -I(a)$ and the second part proves $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \geq na] \geq -I(a)$. We will start with the proof of part 1.

Let $\lambda > 0$. We use Markov's inequality to derive

$$\begin{aligned}\mathbb{P}[S_n \geq na] &= \mathbb{P}\left[e^{\lambda S_n} \geq e^{\lambda na}\right] \\ &\leq \frac{\mathbb{E}\left[e^{\lambda S_n}\right]}{e^{\lambda na}} \\ &= \frac{\mathbb{E}\left[\prod_{i=1}^n e^{\lambda X_i}\right]}{e^{\lambda na}}.\end{aligned}$$

We know (X_i) are independent, therefore

$$\begin{aligned}\mathbb{P}[S_n \geq na] &\leq \prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] e^{-\lambda na} \\ &= \mathbb{E}\left[e^{\lambda X_1}\right]^n e^{-\lambda na} \\ &= e^{n\left(\frac{1}{n}\log \mathbb{E}\left[e^{\lambda X_1}\right]^n - \lambda a\right)} \\ &= e^{-n\left(\lambda a - \log \mathbb{E}\left[e^{\lambda X_1}\right]\right)}.\end{aligned}$$

Subsequently, optimizing over λ results in

$$\begin{aligned}\frac{1}{n}\log \mathbb{P}[S_n \geq na] &\leq \inf_{\lambda>0} -\left(\lambda a - \log \mathbb{E}\left[e^{\lambda X_1}\right]\right) \\ &\leq -\sup_{\lambda>0}\left(\lambda a - \log \mathbb{E}\left[e^{\lambda X_1}\right]\right).\end{aligned}$$

We are left to prove that

$$\sup_{\lambda>0}[\lambda a - \varphi(\lambda)] = \sup_{\lambda\in\mathbb{R}}[\lambda a - \varphi(\lambda)].$$

This is done in Lemma 2.8 where we analyse $I(a)$ and its properties. Hence, we can conclude

$$\frac{1}{n}\log \mathbb{P}[S_n \geq na] \leq -I(a). \tag{2.2}$$

□

In order to prove part 2 of the proof of Theorem 2.1, we will introduce a probability measure $\mathbb{Q}_\mu$, called the tilted measure. Later on, we will see that this μ is chosen such that $\mathbb{E}_{\mathbb{Q}_\mu}[X] = a$. That implies that $\lim_{n\to\infty} \frac{1}{n} S_n = a$. Therefore the tilted distribution $\mathbb{Q}_\mu$ will shift the corresponding random variables, such that the rare event $\left\{\frac{1}{n}S_n \geq a\right\}$ becomes a typical event.

Definition 2.2. *Let $\mathbb{P}$ be a probability measure on a measurable space. Define the tilted measure $\mathbb{Q}_\mu$ by*

$$\frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(x) = \frac{e^{x\mu}}{\mathbb{E}_{\mathbb{P}}\left[e^{\mu X}\right]},$$

where μ is chosen such that $\mu a - \log \mathbb{E}\left[e^{\mu X}\right] = I(a)$.

The following lemma shows that $Var_{\mathbb{Q}_\mu}[X] \in (0,\infty)$.

Lemma 2.3. *If* $Var_{\mathbb{P}}[X] > 0$ *then*

$$Var_{\mathbb{Q}_\mu}[X] \in (0,\infty)$$

Proof. We know $Var_{\mathbb{P}}[X] > 0$, so X is not a constant on the probability measure $\mathbb{P}$. Therefore there exist an event A and and an event B such that $\mathbb{P}[X \in A], \mathbb{P}[X \in B] > 0$. We know $e^{x\mu}$ is a strictly positive function and thus

$$\frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(x) = \frac{e^{x\mu}}{\mathbb{E}_{\mathbb{P}}[e^{\mu X}]} > 0.$$

This implies

$$\begin{aligned}\mathbb{Q}_\mu[X \in A] &\geq \left\{\min_{x\in A} \frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(x)\right\} \mathbb{P}[X \in A] > 0,\\ \mathbb{Q}_\mu[X \in B] &\geq \left\{\min_{x\in B} \frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(x)\right\} \mathbb{P}[X \in B] > 0.\end{aligned}$$

So X is not constant under the probability measure $\mathbb{Q}$, hence $Var_{\mathbb{Q}}[X] > 0$.
Yet, we are left to show that $Var_{\mathbb{Q}}[X] < \infty$. Therefore, we use that the moment generating function with respect to the probability measure $\mathbb{Q}$ is finite.
Let $t \in \mathbb{R}$, we know $\mu \neq 0$ so

$$\mathbb{E}_{\mathbb{Q}_\mu}\left[e^{tX}\right] = \frac{\mathbb{E}_{\mathbb{P}[e^{\mu X}e^{tX}]}}{\mathbb{E}_{\mathbb{P}}[e^{\mu X}]} = \frac{M_X(t+\mu)}{M_X(\mu)} < \infty. \tag{2.3}$$

We know $e^x \geq x^2$, if $x \geq 0$ and $e^{-x} \geq x^2$, if $x \leq 0$. Therefore, using (2.3) with $t = 1$ and $t = -1$ we derive

$$Var_{\mathbb{Q}_\mu} = \mathbb{E}_{\mathbb{Q}_\mu}\left[X^2\right] - \mathbb{E}_{\mathbb{Q}_\mu}[X]^2 \leq \max\left(\mathbb{E}_{\mathbb{Q}_\mu}\left[e^X\right], \mathbb{E}_{\mathbb{Q}_\mu}\left[e^{-X}\right]\right) < \infty.$$

□

Lemma 2.4. *Let*

$$\varphi(\mu) = \log \mathbb{E}\left[e^{\mu X}\right],$$

where

$$\mathbb{E}\left[e^{\mu X}\right] < \infty \quad \forall \mu \in \mathbb{R}.$$

Then $\varphi(\mu)$ *is differentiable.*

Proof. The moment generating function $\mathbb{E}\left[e^{\mu X}\right]$ is differentiable on B^o, where $B = \{\mu : M_X(\mu) < \infty\}$. We supposed $\mathbb{E}\left[e^{\mu X}\right] < \infty$, hence $\mathbb{E}\left[e^{\mu X}\right]$ is differentiable on $\mu \in \mathbb{R}$. Furthermore, $\log y$ is differentiable for $y > 0$ and $\mathbb{E}\left[e^{\mu X}\right] > 0$ for all $\mu \in \mathbb{R}$. Hence, $\varphi(\mu)$ is a composition of differentiable functions and thus is differentiable on $\mu \in \mathbb{R}$. □

Lemma 2.5.

$$\mathbb{E}_{\mathbb{Q}_\mu}[X] = a$$

Proof. To deduce $\mathbb{E}_{\mathbb{Q}_\mu}[X] = a$, we will study the supremum on $\mu \in \mathbb{R}$ of

$$\mu a - \varphi(\mu). \tag{2.4}$$

By Lemma 2.4 we know φ is differentiable. Setting the derivative of (2.4) equal to zero, results in

$$a = \varphi^{'}(\mu). \tag{2.5}$$

Differentiating $\varphi(\mu)$, gives

$$\begin{aligned}
\varphi'(\mu) &= \frac{\partial}{\partial \mu} \log \mathbb{E}\left[e^{\mu X}\right] \\
&= \frac{1}{\mathbb{E}\left[e^{\mu X}\right]} * \frac{\partial}{\partial \mu} \mathbb{E}\left[e^{\mu X}\right] \\
&= \frac{\mathbb{E}\left[X e^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]} \qquad (2.6) \\
&= \mathbb{E}_{\mathbb{Q}_\mu}[X]. \qquad (2.7)
\end{aligned}$$

We apply the second derivative test to check whether $a = \mathbb{E}_{\mathbb{Q}_\mu}[X]$ indeed implies a supremum on (2.4) so that $I(a) = \mu a - \varphi(\mu)$. We know $\varphi'(\mu)$ is a quotient of two differentiable functions, a finite moment generating function and the derivative of a finite moment generating function, and hence $\varphi''(\mu)$ exists. Differentiating (2.4) twice with respect to μ gives

$$\frac{\partial^2}{\partial \mu^2}\left(\mu a - \varphi(\mu)\right) = -\varphi^{''}(\mu).$$

Differentiating $\varphi(\mu)$ twice, results in

$$\begin{aligned}
\varphi''(\mu) &= \frac{\partial}{\partial \mu}\left(\frac{\mathbb{E}\left[X e^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]}\right) \\
&= \frac{\mathbb{E}\left[X^2 e^{\mu X}\right]\mathbb{E}\left[e^{\mu X}\right] - \mathbb{E}\left[X e^{\mu X}\right]\mathbb{E}\left[X e^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]^2} \\
&= \frac{\mathbb{E}\left[X^2 e^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]} - \left(\frac{\mathbb{E}\left[X e^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]}\right)^2 \\
&= \mathbb{E}\left[X^2 \frac{e^{\mu X}}{\mathbb{E}\left[e^{\mu X}\right]}\right] - \mathbb{E}\left[X \frac{e^{\mu X}}{\mathbb{E}\left[e^{\mu X}\right]}\right]^2 \\
&= \mathbb{E}_{\mathbb{Q}}\left[X^2\right] - \mathbb{E}_{\mathbb{Q}}[X]^2 \\
&= Var_{\mathbb{Q}}[X]. \qquad (2.8)
\end{aligned}$$

From Lemma 2.3 we know $Var_{\mathbb{Q}}[X] > 0$, so $-\varphi''(\mu) < 0$, and thus $a = \mathbb{E}_{\mathbb{Q}_\mu}[X]$ indeed implies a supremum on (2.4). Hence, we can conclude

$$\mathbb{E}_{\mathbb{Q}_\mu}[X] = a. \tag{2.9}$$

□

The following Lemma will show that $\mu(a)$ is continuously differentiable in an open neighborhood of a. We will use this fact in the proof of Lemma 2.7 where we show I is a strictly convex function.

Lemma 2.6. *Let*

$$I(x) = \sup_{\hat{\mu}\in\mathbb{R}} [x\hat{\mu} - \varphi(\hat{\mu})] = x\mu(x) - \varphi(\mu(x)).$$

Then μ is a continuously differentiable function of x for x in an open neighborhood of a.

Proof. Let $J(x,y) = \varphi^{'}(y) - x$. Let μ^* be such that

$$I(a) = \mu^* a - \varphi(\mu^*).$$

Then we know by (2.5) that $\varphi^{'}(\mu^*) = a$. Therefore $J(a,\mu^*) = \varphi^{'}(\mu^*) - a = 0$. Furthermore, using Lemma 2.3 and (2.8) we derive

$$\frac{\partial J}{\partial y}(a,\mu^*) = \varphi^{''}(\mu^*) > 0.$$

Then we know by the implicit function theorem that there exists an open neighborhood V of a, and an open neighborhood W of μ^*, and a continuously differentiable function $\mu : V \to W$ such that

(i) $J(x,\mu(x)) = 0$ for all $x \in V$,

(ii) $J(x,y) \neq 0$ for all $(x,y) \in V$xW with $y \neq \mu(x)$,

(iii) $\dfrac{\partial \mu}{\partial x}(x)$ exists for all $x \in V$.

So for every $x \in V$ the function $\mu(x)$ is such that $\varphi^{'}(\mu(x)) = x$. Therefore, for every $x \in V$ there is a $\mu(x) \in W$ such that

$$I(x) = \mu(x)x - \varphi(\mu(x)).$$

Furthermore, $\mu(x)$ is a continuously differentiable in V. □

Lemma 2.7. *I is strictly convex.*

Proof. Fix a_0. We prove $I^{''}(a_0) > 0$. According to Lemma 2.6 there exists a continuously differentiable function $a \mapsto \mu(a)$ in an open neighborhood of a_0 such that

$$I(a) = \sup_{\hat{\mu}\in\mathbb{R}} [x\hat{\mu} - \varphi(\hat{\mu})] = \mu(a)a - \varphi(\mu(a)). \tag{2.10}$$

Now, differentiating both the left-hand side and the right-hand side of (2.10) with respect to a results in

$$I'(a) = a\frac{d\mu}{da}(a) + \mu(a) - \varphi'(\mu)\frac{d\mu}{da}(a).$$

Subsequently, using (2.5) we get $I'(a_0) = \mu(a_0)$, and thus

$$I''(a_0) = \frac{d\mu}{da}(a_0). \tag{2.11}$$

Differentiating (2.5) with respect to a gives

$$\varphi''(\mu(a_0))\frac{d\mu}{da}(a_0) = 1. \tag{2.12}$$

Hence

$$\frac{d\mu}{da}(a_0) = \frac{1}{\varphi''(\mu(a_0))}. \tag{2.13}$$

Subsequently, (2.11) combined with (2.13) gives

$$I''(a_0) = \frac{d\mu}{da}(a_0) = \frac{1}{\varphi''(\mu(a_0))} = \frac{1}{Var_{\mathbb{Q}}[X]}.$$

Furthermore, we know from Lemma 2.3 that $Var_{\mathbb{Q}}[X] \in (0, \infty)$, so $I''(a_0) > 0$, hence I is strictly convex. $\square$

With the obtained results, we can now prove the remaining part of the proof of Theorem 2.1.

Part 2 of proof of theorem 2.1. Using the Radon-Nikodym derivative from Definition 2.2, we derive

$$\begin{aligned}
\mathbb{P}[S_n \geq na] &= \mathbb{E}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \geq a\right\}}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \geq a\right\}} \frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(X_1, .., X_n) \left(\frac{d\mathbb{Q}_\mu}{d\mathbb{P}}\right)^{-1}(X_1, .., X_n)\right] \\
&= \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \geq a\right\}} \frac{\mathbb{E}\left[e^{\mu X}\right]^n}{e^{S_n\mu}}\right] \\
&= \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \geq a\right\}} e^{\log\left(\mathbb{E}\left[e^{\mu X}\right]^n\right) - S_n\mu}\right] \\
&= \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \geq a\right\}} e^{n\left(\log\left(\mathbb{E}\left[e^{\mu X}\right]\right) - \frac{1}{n}S_n\mu\right)}\right].
\end{aligned}$$

As we want to derive an expression of the lower bound, we need an upper bound on $\frac{1}{n}S_n$. Using the result that $\frac{1}{n}S_n \to a$ under $\mathbb{Q}_\mu$, we can restrict $[a, \infty)$ to $[a, a+\varepsilon)$, for an arbitrary $\varepsilon > 0$, without losing too much information about the upper bound of $\frac{1}{n}S_n$.

$$\begin{aligned}
\mathbb{P}[S_n \geq na] &\geq \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \in [a, a+\varepsilon]\right\}} e^{n\left(\log\left(\mathbb{E}\left[e^{\mu X}\right]\right) - \frac{1}{n}S_n\mu\right)}\right] \\
&\geq \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \in [a, a+\varepsilon]\right\}}\right] e^{n\left(\log\left(\mathbb{E}\left[e^{\mu X}\right]\right) - (a+\varepsilon)\mu\right)} \\
&= \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \in [a, a+\varepsilon]\right\}}\right] e^{-n\left(a\mu - \log\left(\mathbb{E}\left[e^{\mu X}\right]\right)\right)} e^{-n\mu\varepsilon} \\
&\geq \mathbb{E}_{\mathbb{Q}_\mu}\left[\mathbb{1}_{\left\{\frac{1}{n}S_n \in [a, a+\varepsilon]\right\}}\right] e^{-nI(a)} e^{-n\mu\varepsilon} \\
&= \mathbb{Q}_\mu\left[\frac{1}{n}S_n \in [a, a+\varepsilon]\right] e^{-nI(a)} e^{-n\mu\varepsilon}.
\end{aligned} \tag{2.14}$$

We will apply the Central Limit Theorem to prove that the left term of (2.14) tends to $\frac{1}{2}$, as $n \to \infty$. We know from Lemma 2.3 that $\sigma^2 := Var_{\mathbb{Q}_\mu} \in (0, \infty)$. And from (2.9) that $\mathbb{E}_{\mathbb{Q}_\mu}[X] = a$. So by the Central Limit Theorem

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{Q}_\mu\left[\frac{1}{n}S_n \in [a, a+\varepsilon]\right] &= \lim_{n\to\infty} \left(\mathbb{Q}_\mu\left[S_n \leq n(a+\varepsilon)\right] - \mathbb{Q}_\mu\left[S_n \geq na\right]\right) \\
&= \lim_{n\to\infty} \mathbb{Q}_\mu\left[\frac{S_n - na}{\sigma\sqrt{n}} \leq \frac{an + n\varepsilon - an}{\sigma\sqrt{n}}\right] - \left(1 - \lim_{n\to\infty} \mathbb{Q}_\mu\left[S_n \leq na\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{n\to\infty} \mathbb{Q}_\mu\left[\frac{S_n - na}{\sigma\sqrt{n}} \leq \frac{n\varepsilon}{\sigma\sqrt{n}}\right] + \lim_{n\to\infty} \mathbb{Q}_\mu\left[\frac{S_n - na}{\sigma\sqrt{n}} \leq \frac{na - na}{\sigma\sqrt{n}}\right] - 1 \\
&= \lim_{n\to\infty} \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma}\right) + \Phi(0) - 1 \\
&= 1 + \frac{1}{2} - 1 = \frac{1}{2}.
\end{aligned}
$$

This results in combined with (2.14) implies

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log\left(\mathbb{P}\left[S_n \geq na\right]\right) &\geq \lim_{n\to\infty} \frac{1}{n}\log\left(\mathbb{Q}_\mu\left[\frac{1}{n}S_n \in [a, a+\varepsilon]\right] e^{-nI(a)} e^{-n\mu\varepsilon}\right) \\
&= \lim_{n\to\infty}\left(\frac{\log\left(\mathbb{Q}_\mu\left[\frac{1}{n}S_n \in [a, a+\varepsilon]\right]\right)}{n}\right) - I(a) - \mu\varepsilon \\
&= \lim_{n\to\infty}\left(\frac{\log\left(\mathbb{Q}_\mu\left[\frac{1}{n}S_n \in [a, a+\varepsilon]\right]\right)}{n}\right) - I(a) - \mu\varepsilon \\
&= \lim_{n\to\infty}\left(\frac{\log\left(\frac{1}{2}\right)}{n}\right) - I(a) - \mu\varepsilon \\
&= -I(a) - \mu\varepsilon.
\end{aligned}
$$

We chose $\varepsilon > 0$ arbitrarily. Hence

$$
\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left[S_n \geq na\right] \geq -I(a). \tag{2.15}
$$

If we combine (2.2) and (2.15) we can conclude that

$$
\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left[S_n \geq na\right] = -I(a).
$$

□

Lemma 2.8.

$$
\sup_{\lambda\in\mathbb{R}}\left[\lambda a - \varphi(\lambda)\right] = \sup_{\lambda>0}\left[\lambda a - \varphi(\lambda)\right]
$$

Proof. From (2.5) we know

$$
\varphi'(\lambda(a)) = a. \tag{2.16}
$$

Furthermore, (2.7) combined with the fact that $a > \mathbb{E}[X]$ implies

$$
\varphi'(0) = \mathbb{E}[X] < a. \tag{2.17}
$$

Using (2.6) we derive

$$
\varphi'(0) = \mathbb{E}[X]. \tag{2.18}
$$

Hence (2.16) combined with (2.17) and (2.18) results in

$$
\varphi'(\lambda(a)) > \varphi'(0). \tag{2.19}
$$

From Lemma 2.3 we know $Var_{\mathbb{Q}_\mu}[X] > 0$. Subsequently, using (2.8) we derive $\varphi''(\lambda) > 0$. Hence $\varphi'(\lambda)$ is a strictly increasing function of λ. Therefore, (2.19) implies $\lambda(a) > 0$. □

One of the conditions in Cramér's theorem is that the moment generating function $\mathbb{E}e^{\mu X} < \infty$ for all $\mu \in \mathbb{R}$. The moment generating function of the exponential distribution is given by

$$M_X(\mu) = \mathbb{E}e^{\mu X} = \begin{cases} \frac{\lambda}{\lambda-\mu}, & \text{if } \mu < \lambda, \\ \infty, & \text{otherwise.} \end{cases}$$

Therefore, the exponential distribution does not meet the condition in Cramér's theorem. However, in the next proposition we will see that Cramér's theorem still holds for the exponential distribution.

Proposition 2.9. *Theorem 2.1 also holds for the exponential distribution for $a > 0$.*

Proof. In the first part of the proof of Theorem 2.1 we derive the upper bound

$$\frac{1}{n} \log \mathbb{P}\left[S_n \geq na\right] \leq -I(a) = -\sup_{\mu \in \mathbb{R}} \left[a\mu - \log \mathbb{E}e^{\mu X}\right].$$

This upper bound is independent of whether $\mathbb{E}\left[e^{\mu X}\right]$ is finite or not and therefore will still hold.

For the proof of the lower bound we used the tilted measure $\mathbb{Q}_\mu$. Therefore we have to show that this measure exists in the case of the exponential distribution. Let $a > 0$. If we suppose that $\mu \geq \lambda$, then $\mathbb{E}\left[e^{\mu X}\right] = \infty$. This implies

$$a\mu - \log \mathbb{E}\left[e^{\mu X}\right] = -\infty.$$

But $a * 0 - \log \mathbb{E}\left[e^{0 * X}\right] = 0$. Hence

$$\sup_{\mu \in \mathbb{R}} \left[\mu a - \varphi(\mu)\right] = \sup_{\mu < \lambda} \left[\mu a - \varphi(\mu)\right].$$

We know that $\mathbb{E}\left[e^{\mu X}\right]$ is finite for $\mu < \lambda$, and therefore using (2.5) we derive

$$\varphi^{'}(\mu) = \frac{\partial}{\partial \mu} \log\left(\frac{\lambda}{\lambda - \mu}\right) = \frac{1}{\lambda - \mu} = a.$$

Note that $\frac{1}{\lambda-\mu}$ covers the set $(0, \infty)$ for $\mu < \lambda$. Therefore for every $a > 0$ there exists a μ such that $\varphi^{'}(\mu) = a$. Hence, we know the measure $\mathbb{Q}_\mu$ exists for every $a > 0$ and therefore the proof of part 2 of the proof of Cramér's theorem still holds. $\square$

2.2 Rate functions of common probability distributions

From Theorem 2.1 we know that for large n

$$\mathbb{P}\left[\frac{1}{n} S_n \geq a\right] \approx e^{-nI(a)}.$$

where the rate function $I(a) = \sup_{\mu \in \mathbb{R}} \left[a\mu - \varphi(\mu)\right]$.

In Proposition 2.10, the rate functions $I(a)$ corresponding to some common probability distributions are given.

Proposition 2.10. *Let the rate function $I(a) = \sup_{\mu\in\mathbb{R}}[\mu a - \varphi(\mu)]$. The table below contains probability distributions with their corresponding rate functions.*

Distribution	**Rate function I(a)**
Constant(c)	$\begin{cases} 0, \text{ if } a = c \\ \infty, \text{ otherwise} \end{cases}$
Bernoulli(p)	$\begin{cases} a\log\left(\frac{a}{p}\right) + (1-a)\log\left(\frac{1-a}{1-p}\right), \text{ if } a \in [0,1] \\ \infty, \text{ otherwise} \end{cases}$
Binomial(n,p)	$\begin{cases} a\log\left(\frac{a}{p}\right) + (n-a)\log\left(\frac{n-a}{1-p}\right), \text{ if } a \in [0,n] \\ \infty, \text{ otherwise} \end{cases}$
Poisson(λ)	$\begin{cases} a\log\left(\frac{a}{\lambda}\right) - a + \lambda, \text{ if } a > 0 \\ \lambda, \text{ if } a = 0 \\ \infty, \text{ if } a < 0 \end{cases}$
Exponential(λ)	$\begin{cases} a\lambda - 1 - \log(a\lambda), \text{ if } a > 0 \\ \infty, \text{ if } a \leq 0 \end{cases}$
Normal(λ, σ^2)	$\frac{1}{2}\frac{(a-\lambda)^2}{\sigma^2}$, for $a \in \mathbb{R}$

Proof. In the proofs of the rate functions of the distributions, we use the same method for every distribution except for the constant random variable. We first use (2.5) to compute μ from a, subsequently we derive an expression for μ, and after that we use μ to derive an expression for $I(a)$.
(i) Let $X = c \in \mathbb{R}$.
Then $M_X(\mu) = \mathbb{E}\left[e^{\mu X}\right] = e^{c\mu}$.
Therefore

$$I(a) = \sup_{\mu\in\mathbb{R}}[\mu a - \log(e^{c\mu})] = \sup_{\mu\in\mathbb{R}}[\mu(a-c)] = \begin{cases} 0 \text{ if } a = c, \\ \infty \text{ otherwise.} \end{cases}$$

(ii) Let $X \sim$ Bernoulli(p) for $p \in [0,1]$.
We know $M_X(\mu) = \mathbb{E}\left[e^{\mu X}\right] = q + pe^{\mu}$.
If $a < 0$. Then

$$\lim_{\mu\to-\infty}[\mu a - \log(q + pe^{\mu})] = \infty.$$

Hence, $I(a) = \infty$ for $a > 0$.
If $a > 0$. Then

$$\lim_{\mu\to\infty}[\mu a - \log(q + pe^{\mu})] = \infty.$$

Hence, $I(a) = \infty$ for $a < 0$.
If $a \in [0,1]$. Then we can express a as

$$a = \frac{\partial}{\partial\mu}(\log(q + pe^{\mu})) = \frac{pe^{\mu}}{q + pe^{\mu}}.$$

Subsequently, the expression for μ follows by

$$aq + ape^{\mu} = pe^{\mu}$$

$$
\begin{aligned}
pe^{\mu}(1-a) &= aq \\
e^{\mu} &= \frac{a}{1-a}\frac{q}{p} \\
\mu &= \log\left(\frac{a}{p}\right) - \log\left(\frac{1-a}{q}\right).
\end{aligned} \tag{2.20}
$$

Therefore

$$
\begin{aligned}
I(a) &= a\left(\log\left(\frac{a}{p}\right) - \log\left(\frac{1-a}{q}\right)\right) - \log\left(q + p\frac{a}{1-a}\frac{q}{p}\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{1-a}{q}\right) - \log\left(q\left(1+\frac{a}{1-a}\right)\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{1-a}{q}\right) - \log\left(q\left(\frac{1-a}{1-a}+\frac{a}{1-a}\right)\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{1-a}{q}\right) - \log\left(\frac{q}{1-a}\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{1-a}{q}\right) + \log\left(\frac{1-a}{q}\right) \\
&= a\log\left(\frac{a}{p}\right) + (1-a)\log\left(\frac{1-a}{q}\right) \\
&= a\log\left(\frac{a}{p}\right) + (1-a)\log\left(\frac{1-a}{1-p}\right).
\end{aligned}
$$

(iii) Let $X \sim \text{Binomial}(n,p)$ for $n \in \mathbb{N}$ and $p \in [0,1]$.
We know $M_X(\mu) = \mathbb{E}\left[e^{\mu X}\right] = (q+pe^{\mu})^n$.
If $a < 0$. Then

$$
\lim_{\mu\to-\infty}\left[\mu a - \log\left[(q+pe^{\mu})^n\right]\right] = \lim_{\mu\to-\infty}\left[\mu a - n\log\left(q+pe^{\mu}\right)\right] = \infty.
$$

Hence, $I(a) = \infty$ for $a < 0$.
If $a > n$. Then

$$
\lim_{\mu\to\infty}\left[a\mu - \log\left[(q+pe^{\mu})^n\right]\right] = \lim_{\mu\to\infty}\left[a\mu - n\log\left(q+pe^{\mu}\right)\right] = \infty.
$$

Therefore, $I(a) = \infty$ for $a > n$.
If $a \in [0,n]$. Then we can express a as

$$
\begin{aligned}
a &= \frac{\partial}{\partial\mu}\varphi(\mu) \\
&= \frac{\partial}{\partial\mu}\left[\log\left(q+pe^{\mu}\right)^n\right] \\
&= n\frac{\partial}{\partial\mu}\left[\log\left(q+pe^{\mu}\right)\right] \\
&= \frac{npe^{\mu}}{q+pe^{\mu}}.
\end{aligned}
$$

Subsequently, the expression for μ follows by

$$
aq + ape^{\mu} = npe^{\mu}
$$

$$\begin{aligned}
pe^{\mu}(n-a) &= aq \\
e^{\mu} &= \frac{a}{n-a}\frac{q}{p} \\
\mu &= \log\left(\frac{a}{p}\right) - \log\left(\frac{n-a}{q}\right).
\end{aligned} \tag{2.21}$$

Therefore

$$\begin{aligned}
I(a) &= a\left(\log\left(\frac{a}{p}\right) - \log\left(\frac{n-a}{q}\right)\right) - \log\left[\left(q + p\frac{a}{n-a}\frac{q}{p}\right)^n\right] \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{n-a}{q}\right) - n\log\left(q\left(1 + \frac{a}{n-a}\right)\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{n-a}{q}\right) - n\log\left(q\left(\frac{n-a}{n-a} + \frac{a}{n-a}\right)\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{n-a}{q}\right) - n\log\left(\frac{qn}{n-a}\right) \\
&= a\log\left(\frac{a}{p}\right) - a\log\left(\frac{n-a}{q}\right) + n\log\left(\frac{n-a}{nq}\right) \\
&= a\log\left(\frac{a}{p}\right) + (n-a)\log\left(\frac{n-a}{nq}\right) \\
&= a\log\left(\frac{a}{p}\right) + (n-a)\log\left(\frac{n-a}{1-p}\right).
\end{aligned}$$

(iv) Let $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$.
We know $M_X(\mu) = \mathbb{E}\left[e^{\mu X}\right] = e^{\lambda(e^{\mu}-1)}$.
So $\varphi(\mu) = \log M_X(\mu) = \lambda e^{\mu} - \lambda$.
If $a < 0$. Then

$$\lim_{\mu \to -\infty} [a\mu - \lambda e^{\mu} + \lambda] = \infty.$$

Therefore, $I(a) = \infty$ for $a < 0$. If $a = 0$. Then

$$\sup_{\mu \in \mathbb{R}} [\lambda - \lambda e^{\mu}] = \lambda.$$

Hence, $I(a) = \lambda$ for $a = 0$.
If $a > 0$. Then we can express a as

$$a = \frac{\partial}{\partial \mu}\varphi(\mu) = \frac{\partial}{\partial \mu}\left(e^{\lambda e^{\mu} - \lambda}\right) = \lambda e^{\mu}.$$

So

$$\mu = \log\left(\frac{a}{\lambda}\right). \tag{2.22}$$

Hence

$$\begin{aligned}
I(a) &= \sup_{\mu \in \mathbb{R}} [a\mu - \log \varphi(\mu)] \\
&= \mu a - \lambda e^{\mu} + \lambda \\
&= \log\left(\frac{a}{\lambda}\right) a - \lambda e^{\log\left(\frac{a}{\lambda}\right)} + \lambda
\end{aligned}$$

$$= a \log\left(\frac{a}{\lambda}\right) - a + \lambda.$$

(v) Let $X \sim \text{Exponential}(\lambda)$ for $\lambda > 0$.
According to Proposition 2.9 we can use Cramér's theorem to derive the rate function of the exponential distribution.

$$M_X(\mu) = \mathbb{E}e^{\mu X} = \begin{cases} \frac{\lambda}{\lambda-\mu}, & \text{if } \mu < \lambda, \\ \infty, & \text{otherwise.} \end{cases}$$

Let $a \leq 0$. Then

$$\lim_{\mu \to -\infty} \left[a\mu - \log\left(\frac{\lambda}{\lambda-\mu}\right)\right] = \lim_{\mu \to -\infty} \left[a\mu - \log(\lambda) + \log(\lambda-\mu)\right] = \infty.$$

Therefore, $I(a) = \infty$ for $a \leq 0$.
If $a > 0$. Then we can express a as

$$\begin{aligned} a &= \frac{\partial}{\partial\mu}\varphi(\mu) \\ &= \frac{\partial}{\partial\mu}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)\right) \\ &= \frac{1}{\frac{\lambda}{\lambda-\mu}} * \frac{\lambda}{(\lambda-\mu)^2} \\ &= \frac{1}{\lambda-\mu}. \end{aligned}$$

So

$$\mu = \frac{a\lambda - 1}{a}. \tag{2.23}$$

Hence

$$\begin{aligned} I(a) &= a * \frac{a\lambda-1}{a} - \log\left(\frac{\lambda}{\lambda - \frac{a\lambda-1}{a}}\right) \\ &= a\lambda - 1 - \log\left(\frac{\lambda}{\lambda - \frac{a\lambda-1}{a}}\right) \\ &= a\lambda - 1 - \log\left(\frac{\lambda}{\frac{1}{a}}\right) \\ &= a\lambda - 1 - \log(a\lambda). \end{aligned}$$

(iv) Let $X \sim \text{Normal}(\lambda, \sigma^2)$ for $\lambda \in \mathbb{R}$ and $\sigma^2 > 0$. Now

$$\begin{aligned} M_X(\mu) &= \mathbb{E}\left[e^{\mu X}\right] \\ &= \int_{-\infty}^{\infty} e^{\mu x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\lambda)^2}{\sigma^2}} dx. \end{aligned}$$

Apply integration by substitution, let $z = \frac{x-\mu}{\sigma}$, then $x = z\sigma + \lambda$, $\left|\frac{dx}{dz}\right| = \sigma$.

$$M_X(\mu) = e^{\lambda\mu} \int_{-\infty}^{\infty} e^{z\sigma\mu} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{1}{2}z^2} \sigma dz$$

$$
\begin{aligned}
&= e^{\lambda\mu} \int_{-\infty}^{\infty} e^{z\sigma\mu} \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}z^2} dz \\
&= e^{\lambda\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + z\sigma\mu - \frac{1}{2}\sigma^2\mu} e^{\frac{1}{2}\sigma^2\mu^2} dz \\
&= e^{\lambda\mu + \frac{1}{2}\sigma^2\mu^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\sigma\mu)^2} dz.
\end{aligned}
$$

The part inside the integral is the density function of the $N(\sigma\mu, 1)$-distribution. Therefore the integral is equal to 1. Hence $M_X(\mu) = e^{\lambda\mu + \frac{1}{2}\sigma^2\mu^2}$.
Now we can express a as

$$
a = \frac{\partial}{\partial\mu}\varphi(\mu) = \frac{\partial}{\partial\mu}\left(\lambda\mu - \frac{\sigma^2\mu^2}{2}\right) = \lambda + \mu\sigma^2. \tag{2.24}
$$

So

$$
\mu = \frac{a-\lambda}{\sigma^2}. \tag{2.25}
$$

Hence, the rate function is given by

$$
I(a) = a\frac{a-\lambda}{\sigma^2} - \lambda\frac{a-\lambda}{\sigma^2} - \frac{\sigma^2}{2}\left(\frac{a-\lambda}{\sigma^2}\right)^2 = \frac{1}{2}\frac{(a-\lambda)^2}{\sigma^2}.
$$

$\square$

3 Large deviation of the empirical measure

Recall the insurance model from the introduction. Note that there is a clear distinction between the claim and the payoff. The claim is the event for which money is claimed by a client of the insurer, while the payoff is a function of the claim which displays the corresponding amount of money the insurer has to pay for a claim.
If we study Section 2 in the context of insurances, we can derive a rate function that quantifies at which rate the probability of bankruptcy decays, as $n \to \infty$.
However, if due to reinsurance or other circumstances the payoff changes and thus the underlying probabilistic structure of the payoff changes, we have to make a new analysis of our payoff in order to find the corresponding rate function. Note that only the underlying probabilistic structure of the payoff changes, while the underlying probabilistic structure of the claims does not change.
To save us the work of analyzing the probabilistic structure of the payoff again, we can describe our problem as a large deviation of the empirical measure instead of a large deviation of the empirical average. The empirical measure is the relative frequency at which an event occurs. With the use of the empirical measure we can describe the probability of bankruptcy as a large deviation of the type of claims corresponding to a large payoff, instead of a large deviation of the payoffs itself. If now the payoff changes, one can simply choose a different function to describe the payoff corresponding to the claims, while using the same model for the claims as before.
In Section 2 we studied the large deviation of the average of a sequence of random variables. However, in this section we study the large deviation of the empirical measure of a sequence of random variables.

Suppose there are r different type of claims. Let $Y_1, ..., Y_n$ be the claims an insurer receives. Then we can restrict ourselves to the following conditions

$$Y_i \in S = \{1, ..., r\} \subset \mathbb{N}, \tag{3.1a}$$

$$Y_1, Y_2, ... \text{ are i.i.d. with marginal law } \rho = (\rho_s)_{s \in S}, \tag{3.1b}$$

$$\rho_s > 0 \quad \forall s \in S. \tag{3.1c}$$

Now $Y_i = s \in S$ can be interpreted as the i-th claim has type s. Furthermore, let $f : S \to \mathbb{R}$ be the function which maps the type of claim s to the payoff $f(s)$ and thus

$$X_i = f(Y_i), \text{ for } i = 1, ..., n.$$

Then the total payoff is defined as

$$S_n := \sum_{i=1}^{n} X_i.$$

Thus instead of analysing the payoffs X_i, we will analyse the claims Y_i. Later we will derive info over X_i from Y_i.
Note that Y_i take values in the finite set S. This is more restrictive than the conditions we saw in Section 2, because now continuous distribution functions will be excluded. However, the results we will obtain can be extended so that it can also be used for continuous distributions. In that case, some partial sums should be written as integrals.

Definition 3.1 (Empirical measure). *Let $Y_1, Y_2, ...$ be a sequence of independent identically distributed random variables on S. Them the empirical measure L_n is defined as*

$$L_n(s) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i=s\}} \quad \textit{for } s \in S.$$

Let $\mathfrak{M}_1(S)$ be the set of probabilities measures on S. $\mathfrak{M}_1(S)$ is given by

$$\mathfrak{M}_1(S) := \left\{ \nu = (\nu_1, ..., \nu_r) \in [0,1]^r : \sum_{s=1}^{r} \nu_s = 1 \right\}.$$

So clearly $L_n \in \mathfrak{M}_1(S)$.

Definition 3.2 (Total variation distance). *For $\mu, \nu \in \mathfrak{M}_1(S)$, the total variation distance is given by*

$$d(\mu, \nu) = \frac{1}{2}\sum_{s=1}^{r} |\mu_s - \nu_s|.$$

The total variation distance measures the distance between two probability measures in $\mathfrak{M}_1(S)$.
To derive an asymptotic relation for the large deviation of the empirical measure, we need to introduce the concept relative entropy. The relative entropy between two probability distributions is a measure of the distance between them.

Definition 3.3 (Relative entropy). *The relative entropy of a probability vector ν with respect to another probability vector μ is*

$$H(\nu|\rho) := \sum_{s=1}^{r} \nu_s \log \frac{\nu_s}{\rho_s}.$$

3.1 Sanov's Theorem

As in the case of the empirical average, the empirical measure will converge to its average, as $n \to \infty$. Namely, for each s the Law of large numbers implies

$$\lim_{n\to\infty} L_n(s) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i=s\}} \xrightarrow{a.s.} \rho_s,$$

for all $s \in S$ as $n \to \infty$. Therefore

$$d\left(L_n, \rho\right) \xrightarrow{a.s.} 0. \quad \text{as } n \to \infty.$$

A large deviation of the empirical measure can be measured by the distance between the empirical measure L_n and the original measure ρ.
Sanov's theorem describes large deviations of the empirical measure in terms of relative entropy. In the proof of Theorem 3.4 we follow Theorem 2.2 in [2].

Theorem 3.4 (Sanov's Theorem). *Let L_n as in definition 3.1 and $H(\nu|\rho)$ as in definition 3.3. Then, for all $a > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(L_n \in B_a^c(\rho)\right) = - \inf_{\nu \in B_a^c(\rho)} I_\rho(\nu), \tag{3.2}$$

where $B_a(\rho) = \{\nu \in \mathfrak{M}_1(S) : d(\nu, \rho) \leq a\}$, $B_a^c = \mathfrak{M}_1(S) \backslash B_a(\rho)$ and

$$I_\rho(\nu) = H(\nu|\rho). \tag{3.3}$$

Proof. Let

$$K_n = \left\{ k = (k_1, ..., k_r) \in \mathbb{N}_0^r : \sum_{s=1}^{r} k_s = n \right\}$$

and note that $\frac{1}{n} K_n \subset \mathfrak{M}_1(S)$ for all $n \in \mathbb{N}$. Then L_n has the multinomial distribution

$$\mathbb{P}\left(L_n(s) = \frac{k_s}{n} \forall s \right) = n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!}, \qquad k \in K_n. \tag{3.4}$$

For $k \in K_n$, let $\nu_n(k) = \frac{1}{n} k \in \mathfrak{M}_1(S)$. Let

$$Q_n(a) = \max_{k \in K_n : \nu_n(k) \in B_a^c(\rho)} \left(n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!} \right). \tag{3.5}$$

Then clearly,

$$Q_n(a) \le \mathbb{P}\left(L_n \in B_a^c(\rho) \right) \le |K_n| Q_n(a). \tag{3.6}$$

We first study one of the terms in (3.5). Stirling's formula is given by

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e} \right)^n .$$

This implies

$$\log n! = n \log n - n + \mathcal{O}(\log n). \tag{3.7}$$

Subsequently, using (3.7) and $\sum_{s=1}^{r} \frac{k_s}{n} = 1$ we derive

$$\begin{aligned}
\frac{1}{n} \log \left(n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!} \right) &= \frac{1}{n} \log n! + \frac{1}{n} \log \left(\prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!} \right) \\
&= \frac{1}{n} \left(n \log n - n + \mathcal{O}(\log n) \right) + \frac{1}{n} \sum_{s=1}^{r} \left(\log\left(\rho_s^{k_s}\right) - \log(k_s!) \right) \\
&= \log n - 1 + \mathcal{O}\left(\frac{\log n}{n} \right) + \frac{1}{n} \sum_{s=1}^{r} \left[k_s \log(\rho_s) - \left(k_s \log k_s - k_s + \mathcal{O}(\log k_s) \right) \right] \\
&= \log n - 1 + \mathcal{O}\left(\frac{\log n}{n} \right) + \sum_{s=1}^{r} \left[\frac{k_s}{n} (\log \rho_s - \log k_s) + \frac{1}{n} \mathcal{O}\left(\log k_s \right) \right] + \sum_{s=1}^{r} \frac{k_s}{n} \\
&= \sum_{s=1}^{r} \frac{k_s}{n} \log n + \mathcal{O}\left(\frac{\log n}{n} \right) + \sum_{s=1}^{r} \left[\frac{k_s}{n} (\log \rho_s - \log k_s) + \frac{1}{n} \mathcal{O}(\log k_s) \right] \\
&= \sum_{s=1}^{r} \frac{k_s}{n} \left(\log \rho_s - \log \frac{k_s}{n} \right) + \mathcal{O}\left(\frac{\log n}{n} \right) + \sum_{s=1}^{r} \frac{1}{n} \mathcal{O}(\log k_s) \\
&= -\sum_{s=1}^{r} \nu_n(k) \log \left(\frac{\nu_n(k)}{\rho_s} \right) + \mathcal{O}\left(\frac{\log n}{n} \right) + \frac{1}{n} \sum_{s=1}^{r} \mathcal{O}(\log k_s) \\
&= -I_\rho(\nu_n(k)) + \mathcal{O}\left(\frac{\log n}{n} \right) + \frac{1}{n} \sum_{s=1}^{r} \mathcal{O}(\log k_s) \\
&= -I_\rho(\nu_n(k)) + o(1).
\end{aligned} \tag{3.8}$$

Applying (3.5) to (3.8) results in

$$\begin{aligned}\frac{1}{n}\log Q_n(a) &= \max_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} \{-I_\rho(\nu_n(k)) + o(1)\} \\ &= -\min_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} I_\rho(\nu_n(k)) + o(1). \end{aligned} \tag{3.9}$$

We will now prove that

$$\lim_{n\to\infty} \min_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} I_\rho(\nu_n(k)) = \inf_{\nu\in B_a^c(\rho)} I_\rho(\nu). \tag{3.10}$$

We know each $\nu_n(k) \in \mathfrak{M}_1(S)$ and we know $v \mapsto I_\rho(\nu)$ is continuous as it is a composition of continuous functions. Hence

$$\begin{aligned} &(i) \bigcup_{n\in\mathbb{N}} \{\nu_n(k) : k \in K_n\} \text{ is dense in } \mathfrak{M}_1(S), \\ &(ii)\, v \mapsto I_\rho(\nu) \text{ is continuous on } \mathfrak{M}_1(S). \end{aligned}$$

(i) implies that for every $\nu \in \mathfrak{M}_1(S)$ there exists a sequence $(K_n)_{n\in\mathbb{N}}$, with $k_n \in K_n$ for all n, such that

$$\lim_{n\to\infty} d(\nu_n(k_n), v) = 0.$$

Subsequently, (ii) implies

$$\lim_{n\to\infty} I_\rho(\nu_n(k_n)) = I_\rho(\nu). \tag{3.11}$$

$B_a^c(\rho)$ is an open set, so

$$\limsup_{n\to\infty} \min_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} I_\rho(\nu_n(k)) \le \lim_{n\to\infty} I_\rho(\nu_n(k)) = I_\rho(\nu). \tag{3.12}$$

Optimizing over $\nu \in B_a^c(\rho)$ results in

$$\limsup_{n\to\infty} \min_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} I_\rho(\nu_n(k)) \le \inf_{\nu\in B_a^c(\rho)} I_\rho(\nu). \tag{3.13}$$

We know

$$\limsup_{n\to\infty} \min_{k\in K_n:\nu_n(k)\in B_a^c(\rho)} I_\rho(\nu_n(k)) \ge \inf_{\nu\in B_a^c(\rho)} I_\rho(\nu). \tag{3.14}$$

Therefore, (3.13) together with (3.14) gives (3.10). Now using (3.9) and (3.10) we derive

$$\lim_{n\to\infty} \frac{1}{n}\log Q_n(a) = -\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu). \tag{3.15}$$

Every component of the vector K_n belongs to the set $\{\frac{0}{n}, \frac{1}{n}, ..., \frac{n}{n}\}$. The cardinality of this set is $(n+1)$. This vector is specified by at most r of such quantities. So$\mid K_n \mid\le (n+1)^r$. We know $\frac{1}{n}\log((n+1)^r) \to 0$, as $n \to \infty$. Hence $\frac{1}{n}\log \mid K_n \mid\to 0$. Therefore using (3.15) we derive

$$\lim_{n\to\infty} \frac{1}{n}\log(\mid K_n \mid Q_n(a)) = \lim_{n\to\infty} \frac{1}{n}\log \mid K_n \mid + \lim_{n\to\infty} \frac{1}{n}\log Q_n(a) = -\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu).$$

Subsequently, using (3.6), we can conclude that

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}\left(L_n \in B_a^c(\rho)\right) = -\inf_{\nu\in B_a^c(\rho)} I_\rho(\nu).$$

□

3.2 The coincidence of Cramér's theorem and Sanov's theorem

A large deviation of the empirical measure is more general than a large deviation of the empirical average. Therefore, a large deviation of the empirical measure can imply a large deviation of the empirical average. We have the relation

$$\frac{1}{n}S_n = \sum_{s=1}^{r} sL_n(s). \tag{3.16}$$

Theorem 2.1 described the rate of decay for large deviations of the empirical average $\frac{1}{n}S_n$, while Theorem 3.4 did the same for large deviations of the empirical measure $L_n(s)$. This implies that there exists a link between the 2 theorems. First we will study Theorem 3.5 which indicates the link between (3.16) and Theorem 3.4. After that, Theorem 3.6 will link the rate function of corresponding to the empirical average to the rate function corresponding to the empirical measure. In the proof of Theorem 3.5 we follow Theorem 2.15 from [2].

Theorem 3.5. *For $\nu \in \mathfrak{M}_1(S)$, let $m_\nu = \sum_s s\nu_s$. Then, for all $a > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}S_n \in B_a^c(m_\rho)\right) = -\inf_{z\in B_a^c(m_\rho)} \hat{I}(z),$$

where $B_a(m_\rho) = \{z \in \mathbb{R} :\mid z - m_\rho \mid\leq a\}$ and

$$\hat{I}(z) = \inf_{\nu\in\mathfrak{M}_1(S):m_\nu=z} I_\rho(\nu), \tag{3.17}$$

where $I_\rho(\nu)$ is as in (3.3).

Proof. Note that

$$\left\{\frac{1}{n}S_n \in B_a^c(m_\rho)\right\} \iff \left\{L_n \in \hat{B}_a^c(\rho)\right\},$$

where

$$\hat{B}_a(\rho) = \{\nu \in \mathfrak{M}_1(S) :\mid m_\nu - m_\rho \mid\leq a\}.$$

We know $v \mapsto m_\rho$ is continuous. Therefore, using that the image of an open set under a continuous function is again an open set, we derive that $B_a^c(\rho)$ is an open subset of $\mathfrak{M}_1(S)$.

In the proof of Theorem 3.4 we only use that $B_a^c(\rho)$ is open to prove (3.12). Hence, Theorem 3.4 also holds when $B_a^c(\rho)$ is replaced by an arbitrary open set of $\mathfrak{M}_1(S)$. Therefore, using (3.16) and Theorem 3.4 with $B_a^c(\rho)$ replaced by $\hat{B}_a^c(\rho)$, we derive

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}S_n \in B_a^c(m_\rho)\right) = -\inf_{\nu\in\hat{B}_a^c(\rho)} I_\rho(\nu).$$

Now the claim follows from the fact that

$$\inf_{z\in\hat{B}_a^c(\rho)} I_\rho(\nu) = \inf_{z\in B_a^c(m_\rho)} \inf_{\nu\in\mathfrak{M}_1(S):m_\nu=z} I_\rho(\nu).$$

$\square$

Theorem 3.6 shows that $I(z)$ of Theorem 2.1 and $\hat{I}(z)$ of Theorem 3.5 coincide. In the proof of this theorem we use the method of Lagrange multipliers.

Theorem 3.6. *Let $I(z)$ as in Theorem 2.1 and $\hat{I}(z)$ as in Theorem 3.5. Then*

$$I(z) = \hat{I}(z).$$

Proof. From (3.17) we know, we want to compute

$$\inf_{\nu\in\mathfrak{M}_1(\Gamma):m_\nu=z} I_\rho(\nu) = \inf_{\nu\in\mathfrak{M}_1(\Gamma):m_\nu=z} \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right).$$

We know $\nu \log\left(\frac{\nu}{\rho}\right)$ is convex in the set $(0,\infty)$, hence the extremum we will find on $(0,\infty)$ using the method of Lagrange multipliers will be an infimum.

The two conditions that must be satisfied are

$$(i) \sum_s \nu_s = 1, \quad \text{because } \nu \in \mathfrak{M}_1(S),$$
$$(ii) \sum_s s\nu_s = z, \quad \text{because } m_\nu = z.$$

Hence, the optimization problem is given by

$$\begin{aligned} &\text{Minimize} && \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right), \\ &\text{subject to:} && G_1(\nu_s) = \sum_s \nu_s - 1 = 0, \\ &&& G_2(\nu_s) = \sum_s s\nu_s - z = 0. \end{aligned}$$

The Lagrangian becomes

$$\begin{aligned} \mathcal{L} &= \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right) - \mu_1 G_1 - \mu_2 G_2 \\ &= \sum_s \left[\nu_s \log\left(\frac{\nu_s}{\rho_s}\right)\right] - \mu_1\left(\sum_s \nu_s - 1\right) - \mu_2\left(\sum_s s\nu_s - z\right) \\ &= \sum_s \left[\nu_s \log\left(\frac{\nu_s}{\rho_s}\right) - \mu_1\nu_s - \mu_2 s\nu_s\right] - \mu_1 - \mu_2 z. \end{aligned}$$

Differentiating the Lagrangian with respect to ν_s gives

$$\begin{aligned} \frac{\partial}{\partial \nu_s}\mathcal{L} &= \sum_s \left[\log\left(\frac{\nu_s}{\rho_s}\right) + 1 - \mu_1 - \mu_2 s\right] \\ &= \sum_s \left[\log(\nu_s) - \log(\rho_s) + 1 - \mu_1 - \mu_2 s\right] = 0. \end{aligned}$$

Therefore

$$\sum_s \log(\nu_s) = \sum_s \left[\log(\rho_s) + \mu_1 + \mu_2 s - 1\right]$$

$$\begin{aligned}\prod_s \nu_s &= \prod_s \rho_s e^{\mu_1+\mu_2 s-1}\\ \nu_s &= \rho_s e^{\mu_1+\mu_2 s-1}.\end{aligned} \tag{3.18}$$

We know ν_s is a probability measure, and thus $\sum_s \nu_s = 1$. This gives

$$\sum_s \rho_s e^{\mu_1+\mu_2 s-1} = 1.$$

Therefore

$$\sum_s \rho_s e^{\mu_2 s} = e^{1-\mu_1}. \tag{3.19}$$

Subsequently, an expression for μ_1 follows by (3.19) which implies

$$\mu_1 = 1 - \log\left(\sum_s \rho_s e^{\mu_2 s}\right). \tag{3.20}$$

Now (3.18) applied to (3.20) results in

$$\begin{aligned}\nu_s &= \rho_s e^{\mu_1+\mu_2 s-1}\\ &= \rho_s e^{1-\log\left(\sum_s \rho_s e^{\mu_2 s}\right)+\mu_2 s-1}\\ &= \frac{\rho_s e^{\mu_2 s}}{\sum_s \rho_s e^{\mu_2 s}}.\end{aligned} \tag{3.21}$$

Subsequently, using (3.21) with $\mu = \mu_2$, we derive

$$\begin{aligned}\hat{I}(z) &= \inf_{\nu\in\mathfrak{M}_1(S):m_\nu=z} I_\rho(\nu)\\ &= \sum_s \nu_s \log\left(\frac{\nu_s}{\rho_s}\right)\\ &= \sum_s \nu_s \log\left(\frac{e^{\mu s}}{\sum_s \rho_s e^{\mu s}}\right)\\ &= \sum_s \nu_s \mu s - \sum_s \left[\nu_s \log\left(\sum_s \rho_s e^{\mu s}\right)\right]\\ &= \mu \sum_s s\nu_s - \log\left(\sum_s \rho_s e^{\mu s}\right)\sum_s \nu_s\\ &= \mu \sum_s \left[s\frac{\rho_s e^{\mu s}}{\sum_s \rho_s e^{\mu s}}\right] - \log\left(\sum_s \rho_s e^{\mu s}\right).\end{aligned}$$

Replacing the sums by expectation gives

$$\hat{I}(z) = \mu\frac{\mathbb{E}\left[Xe^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]} - \log\left(\mathbb{E}\left[e^{\mu X}\right]\right). \tag{3.22}$$

We know $I(z)$ takes a supremum if the derivative with respect to μ equals zero. We know that

$$\frac{\partial}{\partial\mu}\left(z\mu - \log\mathbb{E}\left[e^{\mu X}\right]\right) = 0,$$

gives

$$z = \frac{\mathbb{E}\left[Xe^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]}.$$

So

$$\begin{aligned} I(z) &= \sup_{\mu \in \mathbb{R}} \left[z\mu - \log \mathbb{E}\left[e^{\mu X}\right]\right] \\ &= \mu \frac{\mathbb{E}\left[Xe^{\mu X}\right]}{\mathbb{E}\left[e^{\mu X}\right]} - \log\left(\mathbb{E}\left[e^{\mu X}\right]\right). \end{aligned} \tag{3.23}$$

Hence, by (3.22) and (3.23) we can conclude

$$I(z) = \hat{I}(z).$$

□

4 Distribution of a random variable conditional on a large deviation

An insurer wants to avoid bankruptcy and thus wants to know what kind of claims will likely lead to bankruptcy. This can be studied by the distribution of the claim conditional on a bankruptcy.
Therefore, we want to derive an expression for the conditional probability of the claim Y_1 when the total payoff $S_n \in A$ for large n, where A is a subset of $\mathbb{R}$. We denote this conditional probability as

$$\rho_n^*(s) := \mathbb{P}_\rho \left(Y_1 = s \mid \frac{1}{n} S_n \in A \right), \quad \text{for } s = 1, ..., r.$$

In Section 2 we studied the case where $\frac{1}{n} S_n \geq a$. However, $\frac{1}{n} S_n \in A$ is a more general case. Because we can interpret A as $A = [a, \infty)$. Then

$$\rho_n^*(s) = \mathbb{P}_\rho \left(Y_1 = s \mid \frac{1}{n} S_n \geq a \right).$$

This describes the probability of the random variable Y constrained to a large deviation of the sum S_n.
We will prove in this section that this conditional probability is the tilted measure $\mathbb{Q}_\mu$ from Definition 2.2.

4.1 Gibbs's Principle

Gibbs's principle gives an expression for the limit points of ρ_n^* as $n \to \infty$ in terms of entropy, which we introduced in Section 3. Before we can state Gibbs's principle we will rewrite $\rho_n^*(s)$ in terms of the empirical measure $L_n(s)$.
Let $f = (f_1, f_2, ..., f_r)$, where $f_s = f(s)$. We can write $\frac{1}{n} S_n$ as

$$\begin{aligned}
\frac{1}{n} S_n &= \frac{1}{n} \sum_{i=1}^{n} X_i \\
&= \frac{1}{n} \sum_{i=1}^{n} f(Y_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} f_1 \mathbb{1}_{\{Y_i = 1\}} + f_2 \mathbb{1}_{\{Y_i = 2\}} + ... + f_r \mathbb{1}_{\{Y_i = r\}} \\
&= f_1 \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i = 1\}} + f_2 \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i = 2\}} + ... + f_r \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i = r\}} \\
&= \langle f, L_n \rangle,
\end{aligned}$$

where $\langle f, \pi \rangle$ denotes the integral of f against the measure π.
Now

$$\begin{aligned}
\rho_n^*(s) &= \mathbb{P}_\rho \left(Y_1 = s | \frac{1}{n} S_n \in A \right) \\
&= \mathbb{E} \left[L_n(s) | \frac{1}{n} S_n \in A \right]
\end{aligned}$$

$$= \mathbb{E}\left[L_n(s) | \langle f, L_n \rangle \in A\right].$$

Hence, with $\Gamma := \{\nu : \langle f, \nu \rangle \in A\}$ we derive

$$\rho_n^* = \mathbb{E}\left[L_n | L_n \in \Gamma\right].$$

Let

$$I_\Gamma := \inf_{\nu \in \Gamma^o} H(\nu|\rho) = \inf_{\nu \in \bar{\Gamma}} H(\nu|\rho). \tag{4.1}$$

The following theorem is from Theorem 3.3.3 in [1].

Theorem 4.1 (Gibbs's principle)**.** *Let*

$$\mathcal{M} := \left\{\nu \in \bar{\Gamma} : H(\nu|\rho) = I_\Gamma\right\}.$$

Then
(a) All the limit points of $\{\rho_n^*\}$ belong to $\overline{co}\,(\mathcal{M})$, the closure of the convex hull of $\mathcal{M}$.
(b) When Γ is a convex set of non-empty interior, the set $\mathcal{M}$ consists of a single point to which ρ_n^* converge as $n \to \infty$

Note that Gibbs's principle states that

$$\lim_{n \to \infty} \rho_n^* = \hat{\nu},$$

where $\hat{\nu}$ is such that

$$H(\hat{\nu}|\rho) = \inf_{\nu \in \bar{\Gamma}} H(\nu|\rho).$$

Therefore, $\mathbb{P}_\rho\left(Y_1 = s | \frac{1}{n} S_n \in A\right)$ converges to the probability measure which has the lowest relative entropy corresponding to the original measure ρ. In section 4.2 we will see that this is the tilted measure $\mathbb{Q}_\mu$ from Definition 2.2.

Proof. (a) The statement : All the limit points of $\{\rho_n^*\}$ belong to $\overline{co}\,(\mathcal{M})$, is equivalent with

$$d\left(\rho_n^*, \overline{co}(\mathcal{M})\right) \to 0,$$

where the total variational distance d is defined in Definition 3.2. Let $\mathcal{M}^\delta := \{\nu : d\,(\nu, \mathcal{M}) < \delta\}$. Since d is a convex function on $M_1\,(\Sigma) \times M_1\,(\Sigma)$, each point in $co\left(\mathcal{M}^\delta\right)$ is within variational distance δ of some point in $co\,(\mathcal{M})$. $\delta > 0$ can be arbitrarily small, and thus it suffices to prove

$$d\left(\rho_n^*, co(\mathcal{M}^\delta)\right) \to 0. \tag{4.2}$$

We will first derive a lower bound of the distance between ρ_n^* and an arbitrary subset $U \subset \mathfrak{M}_1(S)$. Since $\mathbb{E}\left[L_n | L_n \in U \cap \Gamma\right]$ belongs to co(U), while $\rho_n^* = \mathbb{E}\left[L_n | L_n \in \Gamma\right]$, it follows that

$$\begin{aligned} d\left(\rho_n^*, co(U)\right) &= \inf_{\nu \in co(U)} d\left(\mathbb{E}\left[L_n | L_n \in \Gamma\right], co(U)\right) \\ &\leq d\left(\mathbb{E}\left[L_n | L_n \in \Gamma\right], \mathbb{E}\left[L_n | L_n \in U \cap \Gamma\right]\right). \end{aligned} \tag{4.3}$$

By the law of total expectation, we know

$$\mathbb{E}\left[L_n | L_n \in \Gamma\right] = \mathbb{E}\left[L_n | L_n \in \Gamma \cap U\right] \mathbb{P}\left(L_n \in U | L_n \in \Gamma\right) + \mathbb{E}\left[L_n | L_n \in \Gamma \cap U^c\right] \mathbb{P}\left(L_n \in U^c | L_n \in \Gamma\right).$$

Using this, we can derive

$$
\begin{aligned}
&\mathbb{E}\left[L_n|L_n\in\Gamma\right]-\mathbb{E}\left[L_n|L_n\in U\cap\Gamma\right]=\\
&\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\mathbb{P}\left(L_n\in U|L_n\in\Gamma\right)+\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right]\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)-\mathbb{E}\left[L_n|L_n\in U\cap\Gamma\right]\\
&=\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\left(1-\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)\right)-\mathbb{E}\left[L_n|L_n\in U\cap\Gamma\right]\\
&+\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right]\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)\\
&=\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)\left\{\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right]-\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\right\}. \qquad (4.4)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&d\left(\mathbb{E}\left[L_n|L_n\in\Gamma\right],\mathbb{E}\left[L_n|L_n\in U\cap\Gamma\right]\right)=\\
&\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)d\left(\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right],\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\right). \qquad (4.5)
\end{aligned}
$$

Now (4.3) combined with (4.5) results in

$$
\begin{aligned}
d\left(\rho_n^*,co(U)\right)&\leq\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)\left\{\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right]-\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\right\}\\
&=\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right)d\left(\mathbb{E}\left[L_n|L_n\in\Gamma\cap U^c\right],\mathbb{E}\left[L_n|L_n\in\Gamma\cap U\right]\right).
\end{aligned}
$$

Subsequently, using that $L_n\in\mathfrak{M}_1(S)$ and $d(\mu,\nu)\leq 1$ for $\mu,\nu\in\mathfrak{M}_1(S)$, we derive

$$
d\left(\rho_n^*,co(U)\right)\leq\mathbb{P}\left(L_n\in U^c|L_n\in\Gamma\right). \qquad (4.6)
$$

We know $\mathcal{M}^\delta\in\mathfrak{M}_1(S)$, so (4.6) applied to $U=\mathcal{M}^\delta$ gives

$$
d\left(\rho_n^*,co(\mathcal{M}^\delta)\right)\leq\mathbb{P}\left(L_n\in\left(\mathcal{M}^\delta\right)^c|L_n\in\Gamma\right). \qquad (4.7)
$$

Note that

$$
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\rho\left(L_n\in\left(\mathcal{M}^\delta\right)^c|L_n\in\Gamma\right)<0, \qquad (4.8)
$$

implies

$$
\lim_{n\to\infty}\mathbb{P}\left(L_n\in\left(\mathcal{M}^\delta\right)^c|L_n\in\Gamma\right)=0.
$$

Therefore, if the statement (4.8) is true, then (4.7) implies (4.2) and the claim will follow. Hence, it suffices to prove that the statement (4.8) is correct in order to complete the proof.
From Theorem 3.4 and (4.1) we know

$$
I_\Gamma=-\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu\left(L_n\in\Gamma\right). \qquad (4.9)
$$

Using (3.14) we derive

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\rho\left(L_n\in\left(\mathcal{M}^\delta\right)^c\cap\Gamma\right)&\leq\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\rho\left(L_n\in\Gamma\right) \qquad (4.10)\\
&\leq-\inf_{\nu\in\Gamma}H(\nu|\rho)\\
&\leq-\inf_{\nu\in\left(\mathcal{M}^\delta\right)^c\cap\Gamma}H(\nu|\rho)\\
&\leq-\inf_{\nu\in\left(\mathcal{M}^\delta\right)^c\cap\bar{\Gamma}}H(\nu|\rho). \qquad (4.11)
\end{aligned}
$$

Due to the strictly inequality condition on the distance, $\mathcal{M}^\delta$ are clearly open sets.Therefore $\left(\mathcal{M}^\delta\right)^c$ are closed sets. An intersection of closed sets is closed, so $\left(\mathcal{M}^\delta\right)^c \cap \bar{\Gamma}$ is closed. Moreover, $\left(\mathcal{M}^\delta\right)^c \cap \bar{\Gamma}$ is clearly bounded, hence $\left(\mathcal{M}^\delta\right)^c \cap \bar{\Gamma}$ is compact.
A lower semi-continuous function on a compact set always contains its infimum and $\mathcal{M}$ is the set of the minimizers v of $H(\nu|\mu)$. Hence, the set $\left(\mathcal{M}^\delta\right)^c$ does not contain the minimizers of $H(\nu|\mu)$. Thus for some $\tilde{\nu} \in \left(\mathcal{M}^\delta\right)^c \cap \bar{\Gamma}$,

$$\inf_{\nu \in \left(\mathcal{M}^\delta\right)^c \cap \bar{\Gamma}} H(\nu|\rho) = H(\tilde{\nu}|\rho) > \inf_{\nu \in \Gamma} H(\nu|\rho) = I_\Gamma. \tag{4.12}$$

If we combine (4.11) and (4.12) we get

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\rho \left(L_n \in \left(\mathcal{M}^\delta\right)^c \cap \Gamma\right) < -I_\Gamma. \tag{4.13}$$

If we apply the definition of conditional probability and combine (4.9) and (4.13), we derive

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\rho \left(L_n \in \left(\mathcal{M}^\delta\right)^c | L_n \in \Gamma\right) &= \limsup_{n\to\infty} \frac{1}{n} \log \left(\frac{\mathbb{P}_\rho \left(L_n \in \left(\mathcal{M}^\delta\right)^c \cap \Gamma\right)}{\mathbb{P}_\rho \left(L_n \in \Gamma\right)} \right) \\
&= \limsup_{n\to\infty} \left\{ \frac{1}{n} \log \mathbb{P}_\rho \left(L_n \in \left(\mathcal{M}^\delta\right)^c \cap \Gamma\right) - \frac{1}{n} \log \mathbb{P}_\rho \left(L_n \in \Gamma\right) \right\} \\
&< -I_\Gamma + I_\Gamma = 0.
\end{aligned}$$

Now the claim follows.

(b) Note that $H(\nu|\rho)$ is a sum of strictly convex functions and therefore is strictly convex. Hence, $H(\nu|\rho)$ has a unique minimum. Therefore, there exists one unique measure $\hat{\nu} \in \hat{\Gamma}$ such that

$$H(\hat{\nu}|\rho) = \inf_{\nu \in \hat{\Gamma}} H(\nu|\rho).$$

Hence, $\mathcal{M}$ consists of a single point and is therefore a closed set. This implies $\mathcal{M} = \hat{\nu}$. Subsequently using (a) we can conclude that $\rho_n^* \to \hat{\nu}$, as $n \to \infty$. $\square$

4.2 Expression of the conditional distribution given a large deviation of the sum

Note that, by the definition of ρ_n^*, Gibbs's principle states

$$\lim_{n\to\infty} \mathbb{P}_\rho \left(Y_1 = s | \frac{1}{n} S_n \in A \right) = \hat{\nu},$$

where $\hat{\nu}$ is defined such that

$$H(\hat{\nu}|\rho) = \inf_{\nu \in \bar{\Gamma}} H(\nu|\rho).$$

In the proof of Theorem 3.6, we used Lagrange multipliers to derive this minimum. From (3.21) and Theorem 3.6 we know this $\hat{\nu}$ is given by

$$\hat{\nu} = \frac{\rho_s e^{\mu s}}{\sum_s \rho_s e^{\mu s}} = \frac{\rho(s) e^{\mu s}}{\mathbb{E}\left[e^{\mu X}\right]},$$

where μ is chosen such that

$$\mu a - \log \mathbb{E}\left[e^{\mu X}\right] = \sup_{\hat{\mu}\in\mathbb{R}} \left\{\hat{\mu} a - \log \mathbb{E}\left[e^{\hat{\mu} X}\right]\right\}.$$

Hence, as mentioned earlier in this section, the conditional distribution we are looking for, is the tilted measure $\mathbb{Q}_\mu$ from definition 2.2. Thus

$$\lim_{n\to\infty} \mathbb{P}_\rho \left(Y_1 = s \mid \frac{1}{n} S_n \in A\right) = \mathbb{Q}_\mu(s),$$

where $\mathbb{Q}_\mu(s)$ is defined in Definition 2.2.

4.3 distributions conditional on rare events for several probability distributions

From section 4.2 we know that the conditional distribution of a single random variable, constrained to a large deviation of the sum, is given by the tilted measure $\mathbb{Q}_\mu$. The Radon-Nikodym derivative of the conditional probability distribution is given by Definition 2.2. In Proposition 4.2 we will derive an expression for this tilted measure $\mathbb{Q}_\mu$ for some common probability distributions.

Proposition 4.2. *The probability distribution under the tilted measure $\mathbb{Q}_\mu$ defined as in Definition 2.2 are given by*

Original distribution	**Tilted distribution**	**Conditional distribution**
Bernoulli(p)	Bernoulli$\left(\frac{pe^\mu}{1-p+pe^\mu}\right)$	Bernoulli(a)
Binomial(n,p)	Binomial$\left(n, \frac{pe^\mu}{1-p+pe^\mu}\right)$	Binomial$\left(n, \frac{a}{n}\right)$
Poisson(λ)	Poisson(λe^μ)	Poisson(a)
Exponential(λ)	Exponential$(\lambda - \mu)$	Exponential$\left(\frac{1}{a}\right)$
Normal(λ, σ^2)	Normal$(\lambda + \mu\sigma^2, \sigma^2)$	Normal(a, σ^2)

Proof. (i) Let $X \sim$ Bernoulli(p), with $p \geq 0$. Then the tilted distribution at $X = 0$ is given by

$$\begin{aligned}\mathbb{Q}_\mu (X = 0) &= \mathbb{E}_\mathbb{P}\left[\mathbb{1}_{\{X=0\}} \frac{d\mathbb{Q}_\mu}{d\mathbb{P}}(0)\right] \\ &= \mathbb{E}_\mathbb{P}\left[\mathbb{1}_{\{X=0\}} \frac{e^0}{\mathbb{E}_\mathbb{P}\left[e^{\mu X}\right]}\right] \\ &= \mathbb{E}_\mathbb{P}\left[\mathbb{1}_{\{X=0\}} \frac{1}{q + pe^\mu}\right].\end{aligned}$$

Subsequently, using $\mathbb{P}(X = 0) = 1 - p$ and the moment generating function $M_X(t) = 1 - p + pe^t$, results in

$$\begin{aligned}\mathbb{Q}_\mu (X = 0) &= (1-p)\frac{1}{1-p+pe^\mu} \\ &= \frac{1-p}{1-p+pe^\mu}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1-p+pe^{\mu}-pe^{\mu}}{1-p+pe^{\mu}} \\
&= 1 - \frac{pe^{\mu}}{1-p+pe^{\mu}}.
\end{aligned}
$$

Similarly, the tilted distribution at $X = 1$ is given by

$$
\begin{aligned}
\mathbb{Q}_{\mu}(X=1) &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{X=1\}}\frac{d\mathbb{Q}_{\mu}}{d\mathbb{P}}(1)\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{X=1\}}\frac{e^{\mu}}{\mathbb{E}_{\mathbb{P}}\left[e^{\mu X}\right]}\right] \\
&= p\frac{e^{\mu}}{1-p+pe^{\mu}} \\
&= \frac{pe^{\mu}}{1-p+pe^{\mu}}.
\end{aligned}
$$

Hence, $\mathbb{Q}_{\mu}$ has the Bernoulli$\left(\frac{pe^{\mu}}{1-p+pe^{\mu}}\right)$-distribution.

From (2.20) we know

$$
\mu = \log\left(\frac{a}{p}\right) - \log\left(\frac{1-a}{q}\right).
$$

Therefore

$$
\begin{aligned}
\frac{pe^{\mu}}{1-p+pe^{\mu}} &= \frac{pe^{\log\left(\frac{a}{p}\right)-\log\left(\frac{1-a}{q}\right)}}{q+pe^{\log\left(\frac{a}{p}\right)-\log\left(\frac{1-a}{q}\right)}} \\
&= \frac{a\frac{q}{1-a}}{q+a\frac{q}{1-a}} \\
&= \frac{a}{1-a+a} = a.
\end{aligned}
$$

Hence, $\mathbb{Q}_{\mu}$ has the Bernoulli(a)-distribution.
(ii) Let $X \sim$ Binomial(n, p), with $n, p \geq 0$. Then the tilted distribution at $X = k$ is given by

$$
\mathbb{Q}_{\mu}(X=k) = \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{X=k\}}\frac{d\mathbb{Q}_{\mu}}{d\mathbb{P}}(X=k)\right].
$$

Subsequently, using that $\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$ and the moment generating function $M_X(t) = (1-p+pe^t)^n$, we derive

$$
\begin{aligned}
\mathbb{Q}_{\mu}(X=k) &= \binom{n}{k}p^k(1-p)^{n-k}\frac{e^{k\mu}}{\mathbb{E}\left[e^{\mu X}\right]} \\
&= \binom{n}{k}p^k(1-p)^{n-k}\frac{e^{k\mu}}{(1-p+pe^{\mu})^n} \\
&= \binom{n}{k}p^k(1-p)^{n-k}\left(e^{\mu}\right)^k\frac{1}{(1-p+pe^{\mu})^k}\frac{1}{(1-p+pe^{\mu})^{n-k}} \\
&= \binom{n}{k}\left(\frac{pe^{\mu}}{1-p+pe^{\mu}}\right)^k\left(\frac{1-p}{1-p+pe^{\mu}}\right)^{n-k}
\end{aligned}
$$

$$
\begin{aligned}
&= \binom{n}{k}\left(\frac{pe^{\mu}}{1-p+pe^{\mu}}\right)^{k}\left(\frac{1-p+pe^{\mu}-pe^{\mu}}{1-p+pe^{\mu}}\right)^{n-k}\\
&= \binom{n}{k}\left(\frac{pe^{\mu}}{1-p+pe^{\mu}}\right)^{k}\left(1-\frac{pe^{\mu}}{1-p+pe^{\mu}}\right)^{n-k}.
\end{aligned}
$$

Therefore, $\mathbb{Q}_\mu$ has the Binomial$\left(n, \frac{pe^{\mu}}{1-p+pe^{\mu}}\right)$-distribution.
From (2.21) we know

$$\mu = \log\left(\frac{a}{p}\right) - \log\left(\frac{n-a}{q}\right).$$

This implies

$$
\begin{aligned}
\frac{pe^{\mu}}{1-p+pe^{\mu}} &= \frac{pe^{\log\left(\frac{a}{p}\right)-\log\left(\frac{n-a}{q}\right)}}{1-p+pe^{\log\left(\frac{a}{p}\right)-\log\left(\frac{n-a}{q}\right)}}\\
&= \frac{a\frac{q}{n-a}}{q+a\frac{q}{n-a}}\\
&= \frac{a}{n-a+a} = \frac{a}{n}.
\end{aligned}
$$

Hence, $\mathbb{Q}_\mu$ has the Binomial$\left(n, \frac{a}{n}\right)$-distribution.
(iii) Let $X \sim$ Poisson(λ), with $\lambda \in \mathbb{R}^+$. Then the tilted distribution at $X = k$ is given by

$$\mathbb{Q}_\mu\left(X=k\right) = \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{X=k\}}\frac{d\mathbb{Q}_\mu}{d\mathbb{P}}\left(X=k\right)\right].$$

Subsequently, using that $\mathbb{P}\left(X=k\right) = \frac{\lambda^k}{k!}e^{-\lambda}$ and that the moment generating function $M_X(t) = e^{\lambda\left(e^t-1\right)}$, we derive

$$
\begin{aligned}
\mathbb{Q}_\mu\left(X=k\right) &= \frac{\lambda^k}{k!}e^{-\lambda}\frac{e^{k\mu}}{\mathbb{E}\left[e^{\mu X}\right]}\\
&= \frac{\lambda^k}{k!}e^{-\lambda}e^{k\mu}e^{-\lambda(e^{\mu}-1)}\\
&= \frac{\lambda^k}{k!}e^{k\mu-\lambda e^{\mu}}\\
&= \frac{\left(\lambda e^{\mu}\right)^k}{k!}e^{-\lambda e^{-\mu}}.
\end{aligned}
$$

Hence, $\mathbb{Q}_\mu$ has the Poisson(λe^{μ})-distribution.
From (2.22) we know

$$\mu = \log\left(\frac{a}{\lambda}\right).$$

This implies

$$\lambda e^{\mu} = \lambda\frac{a}{\lambda} = a.$$

Hence, $\mathbb{Q}_\mu$ has the Poisson(a)-distribution.
(iv) Let $X \sim$ Exponential(λ), with $\lambda > 0$. The the tilted distribution function is given by

$$\mathbb{Q}_\mu\left(X \le a\right) = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\{X\le a\}}\frac{d\mathbb{Q}}{d\mathbb{P}}\right]$$

$$= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\{X\leq a\}}\frac{e^{x\mu}}{\mathbb{E}\left[e^{\mu X}\right]}\right].$$

Subsequently, using that $\lambda e^{-\lambda x}$ is the density function of the exponential distribution and that the moment generating function is given by $M_X(t) = \frac{\lambda}{\lambda - t}$, we derive

$$\begin{aligned}
\mathbb{Q}_\mu\left(X \leq a\right) &= \int_0^a \frac{e^{x\mu}}{\left(\frac{\lambda}{\lambda-\mu}\right)}\lambda e^{-\lambda x}dx \\
&= (\lambda - \mu)\int_0^a e^{x(\mu-\lambda)}dx \\
&= (\lambda - \mu)\left[\frac{1}{(\mu-\lambda)}e^{x(\mu-\lambda)}\right]_0^a \\
&= -\left[e^{-x(\lambda-\mu)}\right]_0^a \\
&= 1 - e^{-(\lambda-\mu)a}.
\end{aligned}$$

Therefore, $\mathbb{Q}_\mu$ has the Exponential($\lambda - \mu$)-distribution. From (2.23) we know

$$\mu = \frac{a\lambda - 1}{a}.$$

This implies

$$\lambda - \mu = \lambda - \frac{a\lambda - 1}{a} = \frac{1}{a}.$$

Hence, $\mathbb{Q}_\mu$ has the Exponential$\left(\frac{1}{a}\right)$-distribution.

(v) Let $X \sim$ Normal(λ, σ^2), with $\lambda \in \mathbb{R}$ and $\sigma^2 > 0$. The the tilted distribution function is given by

$$\begin{aligned}
\mathbb{Q}_\mu\left(X \leq a\right) &= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\{X\leq a\}}\frac{d\mathbb{Q}}{d\mathbb{P}}\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_{\{X\leq a\}}\frac{e^{x\mu}}{\mathbb{E}\left[e^{\mu X}\right]}\right].
\end{aligned}$$

Subsequently, using that $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2}$ is the density function of the normal distribution and that the moment generating function is given by $M_X(t) = e^{-\lambda\mu - \frac{\sigma^2\mu^2}{2}}$, we derive

$$\begin{aligned}
\mathbb{Q}_\mu\left(X \leq a\right) &= \int_{-\infty}^a e^{x\mu}e^{-\lambda\mu-\frac{\sigma^2\mu^2}{2}}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2}dx \\
&= e^{-\lambda\mu-\frac{\sigma^2\mu^2}{2}}\int_{-\infty}^a e^{x\mu}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2}dx \\
&= e^{-\lambda\mu-\frac{\sigma^2\mu^2}{2}}\int_{-\infty}^a e^{x\mu-\frac{1}{2\sigma^2}\left(x^2-2\lambda x+\lambda^2\right)}\frac{1}{\sqrt{2\pi\sigma^2}}dx \\
&= e^{-\lambda\mu-\frac{\sigma^2\mu^2}{2}}\int_{-\infty}^a e^{-\frac{1}{2\sigma^2}\left(x^2-2x\left(\lambda+\mu\sigma^2\right)+\lambda^2\right)}\frac{1}{\sqrt{2\pi\sigma^2}}dx \\
&= e^{-\lambda\mu-\frac{\sigma^2\mu^2}{2}}\int_{-\infty}^a e^{-\frac{1}{2\sigma^2}\left(x^2-2x\left(\lambda+\mu\sigma^2\right)+\left(\lambda+\mu\sigma^2\right)^2\right)+\frac{1}{2\sigma^2}\left(2\lambda\mu\sigma^2+\mu^2\sigma^2\right)}\frac{1}{\sqrt{2\pi\sigma^2}}dx
\end{aligned}$$

$$= e^{-\lambda\mu - \frac{\sigma^2\mu^2}{2} + \lambda\mu + \frac{\sigma^2\mu^2}{2}} \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{\left(x-\left(\lambda+\mu\sigma^2\right)\right)^2}{\sigma^2}} dx$$
$$= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{\left(x-\left(\lambda+\mu\sigma^2\right)\right)^2}{\sigma^2}} dx.$$

Therefore, $\mathbb{Q}_\mu$ has the Normal$(\lambda + \mu\sigma^2, \sigma^2)$-distribution. From (2.25) we know

$$\mu = \frac{a - \lambda}{\sigma^2}.$$

This implies

$$\lambda + \mu\sigma^2 = \lambda + \frac{a - \lambda}{\sigma^2}\sigma^2 = a.$$

Hence, $\mathbb{Q}_\mu$ has the Normal(a, σ^2)-distribution. □

Note that the Conditional distribution functions in Proposition 4.2 all have expected value a and finite positive variance as stated in Lemma 2.5 and Lemma 2.3 respectively. The expected value a can intuitively be interpreted as that when a large deviations occurs, it will probably occur in the most common way.

5 Large deviation of random variables with an infinite moment generating function

In this thesis we always supposed that the moment generating function of the distribution of the random variable was finite. This section is about the distribution of a random variable, conditional on a large deviation of the sum, where its corresponding moment generating function is not finite.
In Proposition 2.9 we saw that Cramér's theorem was still valid despite the fact that the moment generating function was not finite everywhere, but was finite in a neighborhood of 0. Because then the expectation and variance are finite. However, for many distributions with infinite moment generating functions, the theorems we applied are invalid. A well-known distribution where this is the case is the Pareto distribution.

5.1 Large deviation of the Pareto distribution

If we study part 1 of the proof of Cramér's theorem we see that the upper bound is

$$\frac{1}{n}\log \mathbb{P}\left[S_n \geq na\right] \leq -I(a) = -\sup_{\mu\in\mathbb{R}}\left[a\mu - \log \mathbb{E}e^{\mu X}\right] = 0,$$

for the Pareto distribution. Therefore part 1 of the proof of Cramér's theorem still holds. However, part 2 of the proof of Cramér's theorem is not valid for the Pareto distribution. In order to prove part 2 of the theorem we used the tilted measure $\mathbb{Q}_\mu$ and the Central limit theorem. For the Pareto distribution, the moment generating function is not finite and therefore the tilted measure $\mathbb{Q}_\mu$ is not defined. Therefore, we can not use Cramér's theorem to describe a large deviation for the Pareto distribution.

The Pareto distribution belongs to the class of regularly varying distributions with tail

$$\overline{F}(x) = x^{-\alpha}L(x), \quad x > 0,$$

where $L(x)$ is a slowly varying function. A function $L : (0.\infty) \to (0, \infty)$ is slowly varying if for all $a > 0$,

$$\lim_{x\to\infty}\frac{L(ax)}{L(x)} = 1.$$

Theorem 5.1 gives an expression of the limiting behavior of S_n for large values of a for the Pareto distribution. It states that the probability of a large deviation of the maximum value and the probability of a large deviation of the sum are asymptotically the same. This is different from what we have seen in Section 4. In the following theorem we use page 38 from [4].

Theorem 5.1. *Let $\overline{F}(x) = x^{-\alpha}L(x)$ for $\alpha \geq 0$ and $L(x)$ a slowly varying function. Let $X_1, X_2, ..., X_n$ be independent identically distribution random variables with distribution function F. Define $M_n := \max\left(X_1, X_2, ..., X_n\right)$ and $S_n := \sum_{i=1}^{n} X_i$. Then*

$$\mathbb{P}\left(M_n > b\right) \sim \mathbb{P}\left(S_n > b\right), \ \text{as } b \to \infty. \tag{5.1}$$

In (5.1), the notation

$$\mathbb{P}\left(M_n > b\right) \sim \mathbb{P}\left(S_n > b\right),$$

stands for

$$\lim_{n\to\infty}\frac{\mathbb{P}(M_n > b)}{\mathbb{P}(S_n > b)} = 1.$$

Proof. From the definition of S_n we know

$$\begin{aligned}\mathbb{P}(S_n > b) &= \mathbb{P}\left(\sum_{i=1}^{n} X_i > b\right)\\ &= 1 - \mathbb{P}\left(\sum_{i=1}^{n} X_i \le b\right).\end{aligned}$$

Subsequently, for $F^{n*}(x) := \mathbb{P}\left(\sum_{i=1}^{n} X_i \le x\right)$ we derive

$$\begin{aligned}\mathbb{P}(S_n > b) &= 1 - F^{n*}(b)\\ &= \overline{F^{n*}(b)}.\end{aligned}$$

To calculate $\mathbb{P}(M_n > b)$, we use the fact that the event $M_n \le b$ implies $X_i \le b$ for every $i = 1, ..., n$. This gives

$$\begin{aligned}\mathbb{P}(M_n > b) &= 1 - \mathbb{P}(M_n \le b)\\ &= 1 - \mathbb{P}(X_1 \le b, ..., X_n \le b)\\ &= 1 - \prod_{i=1}^{n}\mathbb{P}(X_i < b)\\ &= 1 - F^n(b)\\ &= \sum_{k=0}^{n-1} F^k(b) - \sum_{k=1}^{n} F^k(b)\\ &= (1 - F(b))\left(\sum_{k=0}^{n-1} F^k(b)\right)\\ &= \overline{F(b)}\sum_{k=0}^{n-1} F^k(b).\end{aligned}$$

Subsequently, $\lim_{b\to\infty} F(b) = 1$ implies

$$\lim_{b\to\infty}\left(\overline{F(b)}\sum_{k=0}^{n-1} F^k(b)\right) = \lim_{b\to\infty}\left(\overline{F(b)}n\right).$$

Hence for $b \to \infty$,

$$\begin{aligned}\mathbb{P}(M_n > b) &\sim n\overline{F(b)},\ \text{ as } b \to \infty, \qquad (5.2)\\ \mathbb{P}(S_n > b) &= \overline{F^{n*}(b)},\ \text{ as } b \to \infty.\end{aligned}$$

Due to Corollary 1.3.2 from [4] it follows that

$$\mathbb{P}(M_n > b) \sim \mathbb{P}(S_n > b),\ \text{ as } b \to \infty.$$

□

Note that Theorem 5.1 only describes the large deviation for large values of b, whereas Theorem 2.1 describes the large deviation for every value of b. Theorem 5.1 can be interpreted as follows. When b is large, then the tail of the largest value in a sample determines the tail of the sum of this sample.
We know that the variance of the Pareto distribution with shape parameter $\alpha > 0$ is equal to ∞. This means that there is definitely a probability that for a sequence of random variables, one variable is much higher than the others. According to Theorem 5.1, the largest variable determines the tail of the sum of independent random variables. So for large sample size n, the probability that the sum of independent Pareto distributed random variables is exceptionally large is equal to n times the probability that the first sample is exceptionally large.

Theorem 5.2 gives an expression for a large deviation of the sum of Pareto distributed random variables. Theorem 5.2 is Theorem 1.9 from [3].

Theorem 5.2. *Let $X_1, X_2, ..., X_n$ be identically distributed and suppose that $\overline{F}(x) = x^{-\alpha}L(x)$, where $L(x)$ is a slowly varying function and $\alpha > 2$. If, in addition, $\mathbb{E}[X] = 0$, $\sigma^2 = 1$ and $\mathbb{E}|X_1|^{2+\delta} < \infty$, then*

$$\mathbb{P}\left(S_n \geq b\right) = \left(1 - \Phi\left(\frac{b}{\sqrt{n}}\right)\right)\left(1 + o(1)\right) + n\left(1 - F(b)\right)\left(1 + o(1)\right), \tag{5.3}$$

for $n \to \infty$ and $b \geq \sqrt{n}$.

Note that the Pareto distribution can not have the property $\mathbb{E}[X] = 0$. However, we can shift by some constant such that these conditions will be obtained. The most important condition of this theorem is that the tails of the distribution are heavy.
For $b = na$ and n large, Theorem 5.2 implies

$$\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) \approx \left(1 - \Phi\left(\sqrt{n}a\right)\right)\left(1 + o(1)\right) + n\left(1 - F(na)\right)\left(1 + o(1)\right). \tag{5.4}$$

From Section 2 we know that for a finite moment generating function we have

$$\mathbb{P}\left[\frac{1}{n}S_n \geq a\right] \approx e^{-nI(a)}.$$

Therefore we see that $\mathbb{P}\left(\frac{1}{n}S_n \geq a\right)$ decays exponentially when the moment generating function is finite.
Note that the right-hand side of (5.4) for the Pareto distribution with $F(x) = x^{-\alpha}$ can be written as

$$n\left(1 - F(na)\right)\left(1 + o(1)\right) \approx n(na)^{-\alpha} = n^{1-\alpha}a^{-\alpha}.$$

This is not an exponential decay. So clearly, the probability of a rare event decays slower for the Pareto distribution than for a distribution with a finite moment generating function.

From the proof in Nagaev [3] we know that (5.3) can be split up in two equations. When b relatively is large the behavior of the tails dominate and therefore

$$\mathbb{P}\left(S_n \geq b\right) = n\left(1 - F(b)\right)\left(1 + o(1)\right). \tag{5.5}$$

When b is relatively small the probability will behave like the Central limit theorem, therefore

$$\mathbb{P}\left(S_n \geq b\right) = \left(1 - \Phi\left(\frac{b}{\sqrt{n}}\right)\right)(1 + o(1)), \tag{5.6}$$

when b is relatively small.
Using (5.2) we know that as $n \to \infty$ and $b \to \infty$, (5.5) becomes

$$\mathbb{P}\left(S_n \geq b\right) = \mathbb{P}\left(M_n \geq b\right). \tag{5.7}$$

Note that (5.7) coincides with Theorem 5.1. Therefore, for sufficiently large values of b, the sum S_n exceeds b essentially because one of the X_i's assumes a value exceeding b. The large deviation of the sum is thus caused by a large deviation of one of the values X_i.
However, (5.6) can be interpreted as follows. Now, the probability that the sum S_n exceeds the value b is normally distributed and therefore the probability that one individual X_i exceeds b is very small compared to the probability that S_n exceeds b.
Hence, the cause of a large deviation, is different for large values of a and relatively small values of b.

6 Discussion

In Section 2, 3 and 4 we assumed that the moment generating function of the random variable X was finite. Cramér's theorem stated that the rate function $I(a)$ such that

$$\mathbb{P}\left[\frac{1}{n}S_n \geq a\right] \approx e^{-nI(a)},$$

is given by $\sup_{\mu\in\mathbb{R}}\left[a\mu - \log \mathbb{E}e^{\mu X}\right]$. In order to prove Cramér's theorem we used the tilted measure $\mathbb{Q}_\mu$ as defined in Definition 2.2.
We studied the case of large deviations in a more general context in Section 3, because we studied the large deviations of the empirical measure instead of large deviations of the empirical average. Now it will be much easier for an insurer to make a small adjustment to his model when the payoff structure of the claims change, instead of analyzing the claims all over again. Sanov's theorem stated

$$\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(L_n \in B_a^c(\rho)\right) = -\inf_{\nu\in B_a^c(\rho)} H(\nu|\rho),$$

and therefore described the exponential decay of a large deviation of the empirical measure in terms of relative entropy In section 3.2 we saw that Cramér's theorem can be derived from Sanov's theorem.
In Section 4 we saw Gibbs's Principle stated that the distribution function constrained to a large deviation was the probability measure that caused the lowest entropy. In Section 4.2 we saw this measure is given by

$$\lim_{n\to\infty}\mathbb{P}_\rho\left(Y_1 = s \mid \frac{1}{n}S_n \in A\right) = \mathbb{Q}_\mu(s),$$

where $\mathbb{Q}_\mu(s)$ is the tilted measure we already used in the proof of the lower bound of Cramér's Theorem.
In Section 5 we studied the cases where the moment generating function does not exist. Theorem 5.1 stated that for a sequences random variables from the Pareto distribution

$$\mathbb{P}\left(M_n > b\right) \sim \mathbb{P}\left(S_n > b\right), \text{ as } b \to \infty,$$

where $M_n = \max\left(X_1, X_2, ..., X_n\right)$. Therefore a large deviation for a large value of b is caused by a large deviation of a single random variable. Theorem 5.2 implies that for large values of n the probability

$$\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) \approx \left(1 - \Phi\left(\sqrt{n}a\right)\right)(1 + o(1)) + n\left(1 - F(na)\right)(1 + o(1)).$$

Therefore, in contrast to the earlier sections, $\mathbb{P}\left(\frac{1}{n}S_n \geq a\right)$ does not decay exponentially for the Pareto distribution, but decays polynomially as

$$\mathbb{P}\left(\frac{1}{n}S_n \geq a\right) \approx n^{1-\alpha}a^{-\alpha},$$

and therefore decays slower.
We saw that if a is relatively large, the large deviation is caused by one single variable. While in the case that a is relatively small, the deviation is described by the Central limit

theorem and therefore the deviation is likely to be caused by multiple variables.

In this thesis we always assumed the random variables $X_1, X_2, ..., X_n$ were independent, however in the case of insurances, this is often not the case. When, for instance, a village is burned down by a forest fire, a household contents insurance company will have to pay much higher claims then they usually do. Therefore, further research of large deviations in the field of mathematical insurance should also study the deviations in the case the random variables are not independent. However, this wouldn't mean that our results are bad. Although claims, of for example a car insurance company, are not dependent, they are certainly not strongly correlated and therefore theorems like the Law of large numbers and the Central Limit Theorem can still be used to make a good approximation.
Furthermore, the theorems we stated gave expression of probabilities as $n \to \infty$. However, in the real world the number of claims will always be finite, therefore these theorems are maybe less powerful than they initially seemed. Therefore, further research can be done on large deviations for a finite number of random variables.
Further studies can also investigate the number of claims instead of the value of the claims. In that case, the number of claims could be determined by a discrete probability distribution like for example the Poisson distribution.

References

[1] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2009. ISBN: 978-3-642-03311-7. URL: https://books.google.nl/books?id=iT9JRlGPx5gC.

[2] F. den Hollander. *Large Deviations*. Fields Institute monographs. American Mathematical Society, 2008. ISBN: 978-0-8218-4435-9. URL: https://books.google.nl/books?id=arxAjD_yl4oC.

[3] A V Nagaev S. V. *Large Deviations of sums of independent random variables*. en. Novosibirsk: Mathematical Institute.

[4] T. Mikosch P. Embrechts C. Klüppelberg. *Modelling Extremal Events for Insurance and Finance*. en. Springer Berlin Heidelberg, Feb. 1997.