UNIVERSAL PREDICTION OVER LARGE ALPHABETS

Narayana Santhanam[1] *and Venkat Anantharam*[2]

[1]University of Hawaii, Manoa
2540 Dole St., Honolulu, HI, nsanthan@hawaii.edu
[2] University of California, Berkeley
261 Cory Hall, Berkeley, CA, ananth@eecs.berkeley.edu

ABSTRACT

Insurance transfers losses associated with risks to the insurer for a price, the *premium*. Considering a natural probabilistic framework for the insurance problem, we derive a necessary and sufficient condition on loss models such that the insurer remains solvent despite the losses taken on. In particular, there need not be any upper bound on the loss—rather it is the structure of the model space that decides insurability.

Insurance is a way of managing losses associated with risks—for example, floods, network outages, and earthquakes—primarily by transfering risk to another entity—the insurer, for a price, the *premium*. The insurer attempts to break even by balancing the possible loss that may be suffered by a few (risk) with the guaranteed payments of many (premium).

In 1903, Filip Lundberg [1] defined and formulated this scenario in its natural probabilistic setting as part of his thesis. In particular, Lundberg formulated a collective risk problem pooling together the risk of all the insured. There is an underlying risk model—a probability measure on loss sequences. Typically, the model itself is unknown, but can be imagined to belong to a known class of risk models. Suppose the insurance company sets some premium to be paid by the insured regularly—say, once at the beginning of every time interval. The losses incured by the insured will be of uncertain size in every time interval, governed according to the unknown underlying risk model. For a given class of risk models, how should the premiums be set so that the insurer compensates all losses in full, yet remains solvent?

Related to the insurance problem is the *pricing* problem that several researchers [2, 3] have considered for the Internet—these adopt, among other techniques, game theoretic principles to tackle the problem. A different approach, including that of Lundberg [1] involves studying the loss parametrically, using, for example, Poisson processes as the class of risk models. A more comprehensive theory of risk modeling has evolved [4] which incorporates several model classes for the loss other than Poisson processes, and which also includes some fat tailed distribution classes.

The later approach is very reminiscent of work in probability estimation, universal compression and prediction. Lately, there has been a lot of focus on choosing model classes for new applications such as language modeling, text compression, clustering and classification. Researchers have come up with new classes of models, *e.g.* [5, 6], as well as theoretical and practical approaches that balance the complexity of the model classes with their description power [7]. In particular, one would like to use a model class that is as general as possible, and is yet tractable.

This focus in compression literature is very pertinent to a new slew of scenarios for risk management. In settings like network outages, it is not clear what should constitute a reasonable risk model in the absence of usable information about what might cause the outages. If we are going to model these risks, how does one choose a class that is as general as possible, yet, one on which the insurer can set premiums to remain solvent?

A preliminary question is, then, what are necessary and sufficient conditions for a class of measures on infinite loss sequences to be *insurable*? In this paper, we provide a partial answer. If losses can be modelled as *i.i.d.* samples from a set $\mathcal{P}$ of distributions we determine a necessary and sufficient condition on $\mathcal{P}$ for insurability.

We adopt the collective risk approach, namely, we abstract the problem without loss of generality to include just two players in the insurance game—the insured and the insurer. We denote the sequence of losses by $\{X_i\}_{i\geq 1}$, and we assume that $X_i \in \mathbb{N}$ for all $i \geq 1$, where $\mathbb{N}$ denotes the set of natural numbers, $\{0, 1, 2, , \ldots, \}$. $\mathcal{P}^\infty$ is a collection of measures on infinite length loss sequences. In this paper, we deal with only *i.i.d.* measures. Consequently, we denote by $\mathcal{P}$ the set of distributions on $\mathbb{N}$ obtained as single letter marginals of $\mathcal{P}^\infty$.

Let $\mathbb{N}^*$ be the collection of all finite length strings of natural numbers. The insurer's *scheme* Φ is a mapping from $\mathbb{N}^* \to \mathbb{R}^+$, and is interpreted as the premium demanded by the insurer from the insured after a loss sequence is observed. The insurer can observe the loss for a time prior to entering the insurance game. However, we require the insurer enters the game with probability 1 no matter what loss models are in force, and the insurer cannot quit once entered.

We adopt another abstraction without loss of generality: at any stage if the insurer is surprised by a loss bigger than the premium charged in *that* round, the insurer goes bankrupt. To see why this simplification does not involve any loss of generality, imagine the sequence of premiums set in the paper to represent the cummulative premium thus far.

To eliminate trivial schemes that do not enter the game at all, we require that for all $p \in \mathcal{P}$, the insurer enters the game with probability 1.

A class $\mathcal{P}^\infty$ of measures is insurable if $\forall\ \eta > 0$, there exists a premium scheme Φ such that $\forall\ p \in \mathcal{P}^\infty$, $p(\Phi \text{ goes bankrupt }) < \eta$ and if, in addition, for all $p \in \mathcal{P}^\infty$, $\lim_{n\to\infty} p(\{X^n : \Phi(X^n) < \infty\}) = 1$.

In Section 2, we consider an example each of insurable and non-insurable classes.

1. RESULTS

We model the loss at each time by numbers in $\mathbb{N} = \{0, 1, \ldots\}$. A loss distribution is a distribution over $\mathbb{N}$, and let $\mathcal{P}$ be a set of loss distributions. $\mathcal{P}^\infty$ is the collection of *i.i.d.* measures over infinite sequences from $\mathbb{N}$ such that the set of marginals over $\mathbb{N}$ they induce is $\mathcal{P}$. We call $\mathcal{P}$ the set of *single letter marginals* of $\mathcal{P}^\infty$. Each $p \in \mathcal{P}$ is assumed to have finite support, and the *span* of $p \in \mathcal{P}$ is the highest number which has probability > 0 under p.

An insurer's *scheme* Φ is a mapping from $\mathbb{N}^* \to \mathbb{R}^+$, and is interpreted as the premium demanded by the insurer from the insured after a loss sequence is observed. For convenience, we assume $\Phi(x^n) = \infty$ on every sequence x^n of losses on which Φ has not entered.

Note however that the supremum over all distributions $p \in \mathcal{P}$ of the span of p need not bounded. Thus, we do not assume an upper bound on the possible loss.

The crux of insurability is this: we would like close distributions to be similar in their span. We first define what distributions are close, followed by what distributions have "similar" span. We will then specify the necessary and sufficient conditions for insurability.

1.1. Close distributions

Insurability of $\mathcal{P}^\infty$ depends on the neighborhoods of the probability distributions among its single letter marginals $\mathcal{P}$. The relevant "distance" between distributions in $\mathcal{P}$ that decides the neighborhood is

$$\mathcal{J}(p,q) = D\left(p\|\frac{p+q}{2}\right) + D\left(q\|\frac{p+q}{2}\right).$$

1.2. Cummulative distribution functions

In this paper, we phrase the notion of similarity in span in terms of the cummulative distribution function. Note that we are dealing with distributions over a discrete (countable) support, so a few non-standard definitions related to the cummulative distribution functions need to be clarified.

For our purposes cummulative distribution function of any distribution p is a function from $\mathbb{R} \to [0,1]$, and will be denoted by F_p. We obtain F_p by first defining F_p on points in the support of p and the point at infinity. We define F_p for all other points by linearly interpolating between the values in the support of p.

Let $F_p^{-1}(1)$ be the smallest number y such that $F_p(y) = 1$, and let $F_p^{-1}(x) = 0$ for all $0 \le x < F_p(0)$. Note that for $0 \le x \le 1$, $F_p^{-1}(x)$ is now uniquely defined.

Two technical observations are in order since we are dealing discrete distributions. Consider a distribution p with support $\mathcal{A} \subset \mathbb{N}$. For $\delta > 0$, let (T for tail)

$$T_\delta = \{y \in \mathcal{A} : y \ge F^{-1}(1-\delta)\},$$

and let (H for head)

$$H_\delta = \{y \in \mathcal{A} : y \le 2F^{-1}(1-\delta/2)\}.$$

It is easy to see that

$$p(T_\delta) > \delta \text{ and } p(H_\delta) > 1-\delta.$$

Suppose, for some δ, $F_p^{-1}(1-\delta) > 0$ and the premium is set to $F^{-1}(1-\delta)$, the probability under p of the loss exceeding the premium is $\ge \delta$. If the premium is set to $2F_p^{-1}(1-\delta/2)$, the probability that the loss exceeds the premium is $\le \delta$. We will use these observations in the proofs to follow.

1.3. Necessary and sufficient conditions for insurability

Existence of close distributions with very different spans is what kills insurability. A scheme could be "deceived" by some process $p \in \mathcal{P}^\infty$ into setting low premiums, while a close enough distribution lurks with a high loss. The conditions for insurability of $\mathcal{P}^\infty$ are phrased in terms of its single letter marginals $\mathcal{P}$.

Formally, a distribution p in $\mathcal{P}$ is *deceptive* if $\forall$ neighborhoods $\epsilon > 0$, $\exists\ \delta > 0$ so that no matter what function $f : \mathbb{R} \to \mathbb{R}$ is chosen, $\exists$ a bad distribution $q \in \mathcal{P}$ such that

$$\mathcal{J}(p,q) \le \epsilon$$

and

$$F_q^{-1}(1-\delta) > f(F_p^{-1}(1-\delta)),$$

2. EXAMPLES

The set $\mathcal{N}^\infty$ is the class of *i.i.d.* processes whose single letter marginals have finite moment. Namely, $\forall p \in \mathcal{N}^\infty$, $E_p X_1 < \infty$.

Theorem 1. $\mathcal{N}^\infty$ is not insurable.

Proof Note that the loss measure that puts probability 1 on the all-0 zero sequences exists in $\mathcal{N}^\infty$. Since we consider only schemes that enter with probability 1 no matter what $p \in \mathcal{N}^\infty$ is in force, every insurer must therefore enter after seeing a finite number of zeros.

Fix any scheme. Denote the premiums charged at time i by $\Phi(X^i)$. Suppose the scheme enters the game after seeing N losses of size 0. To show that $\mathcal{N}^\infty$ is not insurable, we show that $\exists \eta > 0$ such that for all schemes Φ, $\exists p \in \mathcal{N}^\infty$ such that

$$p(\ \Phi \text{ goes bankrupt }) \ge \eta.$$

Fix some $\delta = 1 - \eta$. Let ϵ be small enough that

$$(1-\epsilon)^N > 1 - \delta/2,$$

and let M be a number large enough that

$$(1-\epsilon)^M < \delta/2.$$

Note that since $1 - \delta/2 \geq \delta/2$, $N < M$.

Let L be greater than any of premiums charged by Φ for the sequences $0^N, 0^{N+1}, \ldots 0^M$. Let $p \in \mathcal{N}^\infty$ satisfy, for all i,

$$p(X_i) = \begin{cases} 1-\epsilon & \text{if } X_i = 0 \\ \epsilon & \text{if } X_i = L. \end{cases}$$

For the process p, the insurer is bankrupted on all sequences that contain loss L in between the N'th and M'th step. The sequences in question have probabilities (under p)

$$(1-\epsilon)^N\epsilon, (1-\epsilon)^{N+1}\epsilon, \ldots, (1-\epsilon)^{N+M-1}$$

and they also form a prefix free set. Therefore, summing up the geometric series and using the assumptions on ϵ above,

$$p(\ \Phi \text{ is bankrupted }) \geq 1 - \delta/2 - \delta/2 = \eta. \qquad \square$$

One can verify that every distribution in $\mathcal{N}^\infty$ is deceptive.

A monotone distribution on numbers satisfies for all i, probability of $i \geq$ probability of $i+1$. Let $\mathcal{M}^\infty$ be the set of all monotone *i.i.d.* loss processes with finite support. It will follow from Section 3 that

Theorem 2. $\mathcal{M}^\infty$ is not insurable. □

The above results mean that while insurability seems related to weak compressibility [8], it is not identical.

Consider $\mathcal{U}$, the collection of all uniform distributions over a finite support of form $\{m, \ldots, M\}$, with m and M being arbitrary. Let the losses be sampled *i.i.d.* from one of the distributions in $\mathcal{U}$—call these processes $\mathcal{U}^\infty$.

Theorem 3. $\mathcal{U}^\infty$ is insurable.

Proof If the threshold probability of ruin is η, set the premiums Φ as follows. For all sequences $\overline{x}$ with length $\leq \log\frac{1}{\eta} + 1$, $\Phi(\overline{x}) = \infty$. For all sequences longer than $\log\frac{1}{\eta} + 1$, the premium is twice the largest loss observed thus far. It is easy to see this scheme is bankrupted with probability $\leq \eta$. □

3. NECESSARY AND SUFFICIENT CONDITION FOR INSURABILITY

Note that according to the conventions adopted with defining cummulative distribution functions in Section 1.2, if for a sequence x, $F_q^{-1}(1-\delta) > \Phi(x)$, the scheme Φ will be bankrupted with probability $\geq \delta$ in the next step.

$\mathcal{P}^\infty$ is a set of *i.i.d.* measures over infinite sequences from $\mathbb{N}$, and let $\mathcal{P}$ denote the collection of their single letter marginals.

Theorem 4. $\mathcal{P}^\infty$ is insurable iff no $p \in \mathcal{P}$ is deceptive. □

4. REFERENCES

[1] K. Englund and A. Martin-Löf. *Statisticians of the Centuries*, chapter on Ernst Filip Oskar Lundberg, pages 308–311. New York: Springer, 2001.

[2] S. Shakkottai and R. Srikant. Economics of network pricing with multiple ISPs. *IEEE/ACM Transactions on Networking*, pages 1233–1245, Dec 2006.

[3] A. M. Odlyzko. Paris metro pricing for the internet. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 140–147, 1999.

[4] H. Cramer. Historical review of Filip Lundberg's work on risk theory. *Skandinavisk Aktuarietidskrift (Suppl.)*, 52:6–12, 1969. Reprinted in The Collected Works of Harald Cramér edited by Anders Martin-Löf, 2 volumes Springer 1994.

[5] F. M. J. Willems, Y. M. Shtarkov, and Tj. J. Tjalkens. The context-tree weighting method: basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, 1995.

[6] A. Orlitsky, N.P. Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Transactions on Information Theory*, 50(7):1469—1481, July 2004.

[7] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743—2760, oct 1998.

[8] J.C. Kieffer. A unified approach to weak universal source coding. *IEEE Transactions on Information Theory*, 24(6):674—682, nov 1978.