Agnostic insurance tasks and their relation to compression

Narayana Santhanam, *Member, IEEE*   Venkat Anantharam *Fellow, IEEE*

***Abstract*— We consider the following *insurance* problem. Our task is to predict finite upper bounds on unseen samples of an unknown distribution p over the set of natural numbers, using only observations generated *i.i.d.* from p. While p is unknown, it belongs to a known collection $\mathcal{P}$ of possible models. To emphasize, the support of the unknown distribution p is unbounded, and the game proceeds for an *infinitely* long time. If the said upper bounds are accurate over the infinite time window with probability arbitrarily close to 1, we say $\mathcal{P}$ is insurable.**

Insurability of $\mathcal{P}$ is characterized by a condition on the set of models, that is both necessary and sufficient. We examine connections between the insurance problem on the one hand, and weak and strong universal compression on the other. We show that if $\mathcal{P}$ can be strongly compressed, it can be insured as well. However, the connection with weak compression is more subtle. We show by constructing appropriate classes of distributions that neither weak compression nor insurability implies the other.

Keywords: insurance, ℓ_1 topology, non-parametric approaches, prediction, universal compression.

Insurance is a means of managing risk by transfering potential losses to an *insurer*, for a price, the *premium*. The insurer attempts to break even by balancing the possible loss that may be suffered by a few with the guaranteed premiums of many.

It is common practice among insurers to limit payments to a predetermined ceiling, even if the loss suffered by the insured exceeds the ceiling. In both the insurance industry and the legal regulatory framework surrounding it, this is assumed to be common sense. However, as we will show, it is not always necessary to impose such ceilings. Moreover, in scenarios such as reinsurance, a ceiling on compensation is not only undesirable, but also limits the very utility of the business.

A second problem arises in several modern settings for which some sort of insurance is desirable, but no viable scheme exists. For example, insuring against network outages or attacks against future smart grids, where the cascade effect of outages or attacks could be catastrophic. In these settings, it is not even 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 systematic, theoretical, as opposed to empirical, study of insurance goes back to 1903 when Filip Lundberg [1] defined a natural probabilistic setting as part of his thesis. In particular, Lundberg formulated a collective risk problem pooling together the risk of all the insured. Typically, these approaches involve studying the loss parametrically, using, for example, compound Poisson processes as the class of risk models. A more comprehensive theory of risk modeling has evolved [2] which incorporates several model classes for the loss other than Poisson processes, and which also includes some fat tailed distribution classes.

In order to address problems in more modern setups, we deviate here from classical literature on insurance in two crucial aspects. First, we allow the loss to be unbounded, and second, we take a non-parametric approach borrowing on a universal compression framework. To clarify, unbounded loss is not just cosmetic—we do not impose any other restrictions such as bounded entropy of distributions, or bounded moments, or any assumptions that effectively leave us with compact model spaces.

We use a probability measure on loss sequences to model the loss. The *model*, *i.e.,* the probability measure, is unknown but assumed to belong to a known class $\mathcal{P}$ of risk models. As mentioned before, we assume no ceiling on the loss, requiring the insurer to compensate the insured in full. For a given class of probabilistic risk models, how should the premiums be set so that the insurer compensates all losses *in full*, yet remains solvent? If such schemes are possible, the model class is said to be *insurable*.

The crux of insurability is this: we would like close distributions to have comparable percentiles. In Section I, we define what distributions are close, followed by what distributions have "similar" percentiles. Examples of insurable and non-insurable classes of distributions can be found in [3], where necessary and sufficient conditions are derived for $\mathcal{P}$ to be insurable if, in addition, all distributions $\mathcal{P}$ have finite (but not necessarily uniformly bounded) spans. In Sections II and III, we will prove the necessary and sufficient condition for insurability.

The similarity of the framework we adopt for predic-

N. Santhanam is with the Dept of EE, University of Hawaii, Honolulu, HI 96822. nsanthan@hawaii.edu

V. Anantharam is with the Dept of EECS, University of California, Berkeley, CA 94720. ananth@eecs.berkeley.edu.

tion to recent universal compression or Bayesian non-parametric statistics literature is not incidental. The results also reveal potentially interesting connections with the notions of weak universality. In future work, we aim to characterize the hierarchy of this prediction problem relative to well known characterizations of compression.

Related to the insurance problem is the *pricing* problem that several researchers [4], [5] have considered for the Internet—these adopt, among other techniques, game theoretic principles to tackle the problem.

I. Conditions characterizing insurability

We model the loss at each time by numbers in $\mathbb{N} = \{0, 1, \dots\}$. 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 of symbols 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 $\mathcal{P}^\infty$.

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 in $\mathbb{N}^*$ is observed. Note that Φ is supposed to work for *all* models in $\mathcal{P}$ and has no information on the underlying distribution other than through the samples from the distribution.

The insurer can observe the loss for a time prior to entering the insurance game. However, we require the scheme enters the game with probability 1 no matter what loss model $p \in \mathcal{P}$ is in force. The insurer has to keep setting finite premiums from the point it enters. For convenience, we assume $\Phi(x^n) = \infty$ on every sequence x^n of losses on which Φ has not entered.

We adopt an apparent simplification that involves no 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. As mentioned before, one then sees the function Φ to represent the sum of total built up past reserves of the insurer as well as the premium to be set for the next round.

Definition 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 $p \in \mathcal{P}^\infty$,

$$p(\{X^n : \lim_{n\to\infty} \min_{1\le j\le n} \Phi(X^j) < \infty\}) = 1. \qquad \square$$

In Theorems 1 and 4, we determine a condition on $\mathcal{P}$ that is both necessary and sufficient for insurability.

A. 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

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

B. Cumulative distribution functions

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

For our purposes cumulative 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 $0 \le x < F_p(0)$. If p has infinite support then $F_p^{-1}(1) = \infty$. 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)

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

and let (H for head)

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

It is easy to see that

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

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.

C. 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

$f(\delta) \in \mathbb{R}$ is chosen, $\exists$ a (bad) distribution $q \in \mathcal{P}$ such that

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

and

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

In the above definition, $f(\delta)$ is simply an arbitrary number. However, it is useful to think of this number as the evaluation of a function $f : (0,1) \rightarrow \mathbb{R}$ at δ, particularly when thinking of the contrapositive of the definition as below. Equivalently, a distribution p in $\mathcal{P}$ is not *deceptive* if $\exists$ neighborhood $\epsilon_p > 0$, such that $\forall$ $\delta > 0$, $\exists$ $f(\delta) \in \mathbb{R}$, such that all distributions $q \in \mathcal{P}$ with

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

satisfy

$$F_q^{-1}(1-\delta) \leq f(\delta).$$

In Sections II and III, we show that a collection $\mathcal{P}^\infty$ of *i.i.d.* processes is insurable iff there are no deceptive distributions among its single letter marginals $\mathcal{P}$.

II. Necessary condition for insurability

Note that according to the conventions adopted with defining cumulative distribution functions in Section I-B, 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 1. If $\mathcal{P}^\infty$ is insurable, then no $p \in \mathcal{P}$ is deceptive.

Proof In what follows, we will denote by p (or q) both a measure $\in \mathcal{P}^\infty$ as well as the single letter marginal distribution $\in \mathcal{P}$. The context will clarify which of the two is meant if need be. We prove the contrapositive of the theorem: if some $p \in \mathcal{P}$ is deceptive, then $\mathcal{P}^\infty$ is not insurable.

Pick $0 \leq \alpha < h^{-1}(\frac{1}{2})$ and fix $0 < \eta < (1 - 2h(\alpha))\big(1 - \frac{1}{e}\big)$, the bounds chosen in order to satisfy the technical requirements of the proof of Lemma 5. Suppose $p \in \mathcal{P}$ is deceptive. We prove that $\mathcal{P}^\infty$ is not insurable by finding for each scheme Φ, a process q such that

$$q(\text{ } \Phi \text{ goes bankrupt }) \geq \eta.$$

Consider any Φ that enters on p with probability 1. For all n, let

$$R_n = \{x^n : \Phi(x^n) < \infty\}$$

be the set of sequences of length n on which Φ has entered and let N be a number such that

$$p(R_N) > 1 - \alpha/2. \tag{1}$$

For any sequence x^n, let $A(x^n)$ be the set of symbols that appear in it. Recall that the head of distribution p, $H_{p,\gamma}$, was defined in Section I-B to be the set $\{y \in \mathcal{A} : y \leq 2F_p^{-1}(1-\gamma/2)\}$, where $\mathcal{A}$ is the support of p. Furthermore, define for all $\gamma > 0$

$$R_{p,\gamma,n} = \{x^n \in R_n : A(x^n) \subseteq H_{p,\gamma}\}.$$

Set $\epsilon = \frac{1}{N^2}$. Since p is deceptive, there exists $\delta > 0$ such that for all $f(\delta) \in \mathbb{R}$, there exists a distribution $q' \in \mathcal{P}$ satisfying both

$$\mathcal{J}(p, q') < \epsilon = \frac{1}{N^2} \text{ and } F_{q'}^{-1}(1-\delta) > f(\delta).$$

While the number $f(\delta)$ can be arbitrary above, we focus on a specific number dependent only on Φ. To define this number, first pick $k \geq 2$ large enough that

$$(1-\delta^k)^{N+1/\delta} \geq 1 - \alpha/2. \tag{2}$$

Now, for all $0 < \delta' < 1$, let

$$f(\delta') \stackrel{\text{def}}{=} \max_{\substack{x^i \in R_{p,\delta^k,i} \\ N \leq i \leq N + \lceil \frac{1}{\delta'} \rceil}} \Phi(x^i).$$

A word about this parameter k, since it is not immediately apparent why this should be defined. We will effectively ignore the δ^k tail of the distribution p, and focus only on strings in $R_{p,\delta^k,i}$, $N \leq i \leq N + \frac{1}{\delta}$. The advantage of doing so is technical—we will be able to handle p and q as though they were distributions with finite span. Furthermore, note that for $N \leq i < N + \frac{1}{\delta}$, $p(R_{p,\delta^k,i}) \geq 1 - \alpha$ from (1) and (2).

Let $q \in \mathcal{P}$ simultaneously satisfy

$$\mathcal{J}(p, q) < \epsilon = \frac{1}{N^2} \text{ and } F_q^{-1}(1-\delta) > f(\delta).$$

Applying Lemma 5 to distributions over length-n sequences induced by the measures $p, q \in \mathcal{P}^\infty$ corresponding to the distributions above,

$$q(R_{p,\delta^k,N}) \geq 1 - \frac{2}{N} - 2h(\alpha),$$

namely, Φ has entered with probability (under q) at least $1 - \frac{2}{N} - 2h(\alpha)$ for length N sequences. Since the insurer cannot quit once it has entered, the scheme has entered with probability (under q) at least $1 - \frac{2}{N} - 2h(\alpha)$ for all n length sequences where $n > N$. Namely for all $n \geq N$,

$$q(R_{p,\delta^k,n}) \geq 1 - \frac{2}{N} - 2h(\alpha),$$

For convenience, let $M = \lceil \frac{1}{\delta} \rceil$. Let the distribution q be in force. We have set things up so that Φ is bankrupted whenever any element in the δ-tail of q follows any sequence in $R_{p,\delta^k,i}$, where $N \le i \le N+M-1$. To see this, note that

$$F_q^{-1}(1-\delta) \ge f(\delta) = \max_{\substack{X^i \in R_{p,\delta^k,i} \\ N \le i \le N + \lceil \frac{1}{\delta} \rceil}} \Phi(X^i). \tag{3}$$

Equivalently, conditioned on any sequence in $R_{p,\delta^k,i}$ with i between N and $N+M-1$, the scheme Φ fails with probability (under q) at least δ in the following step.

A sequence on which Φ has entered, but such that Φ has not been bankrupted on any of the sequence's prefixes is a *surviving* sequence.

Consider a surviving sequence $\overline{x} \in R_{p,\delta^k,N}$ in the support of p at level N. Given $\overline{x}$, let the conditional probability that Φ is bankrupted in the following step be δ_N. From (3), as mentioned before, we have $\delta_N \ge \delta$.

Now, given $\overline{x} \in R_{p,\delta^k,N}$, the conditional probability that Φ is bankrupted in at most two further steps is,

$$\delta_N + (1-\delta_N)\delta_{N+1} \ge \delta + (1-\delta)\delta,$$

where δ_{N+1} is interpreted as the weighted average (over surviving length-$(N+1)$ suffixes of $\overline{x}$) of the probability that Φ goes bankrupt in step $N+2$. In particular, note that the inequality above holds because $\delta_{N+1} \ge \delta$ thanks to Equation (3).

Similarly, given a sequence $\overline{x} \in R_{p,\delta^k,N}$, the probability that Φ is bankrupted on suffixes of $\overline{x}$ with length between N and $N+M$ is

$$\delta_N + (1-\delta_N)\delta_{N+1} + \ldots + (1-\delta_N)\delta_{N+M} \prod_{i=N+1}^{N+M} (1-\delta_i)$$

for some $\delta_N, \delta_{N+1}, \ldots, \delta_{N+M}$, all of which are $\ge \delta$.

Let q_1 be the probability of all survivors in $R_{p,\delta^k,N}$, and q_2 be the probability of all sequences in $R_{p,\delta^k,N}$ where Φ has already been bankrupted. Therefore $q_1 + q_2 = q(R_{p,\delta^k,N})$.

Therefore Φ is bankrupted with probability

$$\begin{aligned} &\ge q_2 + q_1 \left(\delta_N + \ldots + \delta_{N+M} \prod_{i=N}^{N+M} (1-\delta_i) \right) \\ &\ge \left(1 - \frac{1}{N} - h(\alpha)\right)\left(1 - (1-\delta)^{\lceil 1/\delta \rceil}\right), \end{aligned} \tag{4}$$

where $\overline{\delta}$ stands for $1-\delta$, and the second inequality follows as in [3]. □

III. Sufficient condition for insurability

The necessary condition in Section II is also sufficient for insurability. As per conventions adopted in Section I-B, recall that a distribution q is bankrupted with probability $< \delta$ if the premium set is at least $2F^{-1}(1-\delta/2)$. In a slight abuse of notation, we will represent types by probability distributions where convenient. Namely, a string 123111 has a length-6 type $(4/6, 1/6, 1/6)$.

A. Topology of $\mathcal{P}$ with ℓ_1 metric

We begin by showing that $\mathcal{P}$ with ℓ_1 topology is Lindelöf. To do so, we obtain a countable basis (see [6] for definitions) $\mathcal{B}$ for $\mathcal{P}$ with ℓ_1 topology.

Let the collection of all finite subsets of $\mathbb{N}$ be denoted by $\#2^{\mathbb{N}}$. Let $\mathcal{Q}$ be the set of all distributions over sets in $\#2^{\mathbb{N}}$, such that every probability value is rational. $\mathcal{Q}$ is countable, since $\#2^{\mathbb{N}}$ is countable, and given any set $S \in \#2^{\mathbb{N}}$, the set of distributions in $\mathcal{Q}$ over S is countable.

Let $\mathcal{C}$ be the collection of all distributions over $\mathbb{N}$, let $B_q(\epsilon) = \{p \in \mathcal{C} : |p-q| \le \epsilon\}$, and let

$$\mathcal{B} = \{B_q(\epsilon) \cap \mathcal{P} : q \in \mathcal{Q} \text{ and } \epsilon \text{ rational } \}. \tag{5}$$

Lemma 2. $\mathcal{B}$ is a basis for $\mathcal{P}$ with ℓ_1 topology. □

As a corollary, we obtain that [6]

Corollary 3. $\mathcal{P}$ with the ℓ_1 topology is Lindelöf. □

B. Sufficient condition

We have the machinery required to prove that if no $p \in \mathcal{P}$ is deceptive, then the class of distributions is insurable.

Theorem 4. If no $p \in \mathcal{P}$ is deceptive, then $\mathcal{P}^\infty$ is insurable.

Proof The proof is constructive. For any $0 < \eta < 1$, we obtain a scheme Φ such that for all $p \in \mathcal{P}^\infty$, $p(\Phi \text{ goes bankrupt }) < \eta$.

Since no $p \in \mathcal{P}$ is deceptive, it follows that for all $p \in \mathcal{P}$, $\exists \epsilon_p > 0$ such that for all $\delta > 0$, $\exists f_p(\delta) \in \mathbb{R}$ so that all q with $\mathcal{J}(p,q) < \epsilon_p$ satisfy

$$F_q^{-1}(1-\delta) < f_p(\delta).$$

We say that ϵ_p is the *reach* of p. For $p \in \mathcal{P}$, define

$$B_p = \{p' \in \mathcal{P} : \mathcal{J}(p,p') \le \epsilon_p\},$$

which will play the role of the set of distributions which will not be bankrupted by setting premiums assuming p is in force. Furthermore, for $p \in \mathcal{P}$, let

$$I_p = \left\{ q : |p-q|_1 < \frac{{\epsilon_p}^2 (\ln 2)^2}{16} \right\}.$$

For large n, I_p will play the role of the set of length-n types on which the proposed scheme Φ will have entered, in order to ensure that Φ enters with probability 1 on strings generated by p. Note that if ϵ_p is small enough, $I_p \cap \mathcal{P} \subset B_p$.

Since no $p \in \mathcal{P}$ is deceptive, the space $\mathcal{P}$ of distributions can be covered by the open sets I_p. From Corollary 3, $\mathcal{P}$ is Lindelöf under the ℓ_1 topology. Thus, there is a countable set $\mathcal{Q}$, such that $\mathcal{P}$ is covered by the collection of relatively open sets $\mathcal{I}_{\mathcal{Q}}$, where

$$\mathcal{I}_{\mathcal{Q}} \stackrel{\text{def}}{=} \{I_q \cap \mathcal{P} : q \in \mathcal{Q}\}.$$

The set $\mathcal{I}_Q$ is countable. We index it by $\iota : \mathcal{I}_Q \to \mathbb{N}$.

We now describe the scheme Φ.

a) Preliminaries: Consider a length-n sequence x on which Φ has not entered thus far. Let the type of the sequence be q, and let

$$\mathcal{P}'_q = \{p' \in \mathcal{I}_Q : q \in I_{p'}\}$$

be the set of distributions in $\mathcal{P}$ which potentially *capture* q.

Note that q in general does not belong to $\mathcal{P}$, so we need further refinements to the set $\mathcal{P}'$.

If $\mathcal{P}'_q \neq \emptyset$, we will refine the set of distributions that could capture q further to $\mathcal{P}_q \subset \mathcal{P}'_q$. This is to ensure that distributions in $\mathcal{P}'_q$ do not prematurely capture a type. First we require (6) below to hold, to ensure that if any distribution $p' \in \mathcal{P}_q$ captures q generated by a distribution p out of its reach, then the probability of q under p is not too large. In addition, we impose (7) as well to resolve a technical issue since q need not, in general, belong to $\mathcal{P}$.

For $p' \in \mathcal{P}'_q$, let the reach of p' be $\epsilon_{_{p'}}$, and define

$$D_{_{p'}} \stackrel{\text{def}}{=} \frac{\epsilon^4_{_{p'}}(\ln 2)^3}{512}.$$

This quantity will lower bound the distance of the type q in question from the distribution p in force if p happens to be out of the reach of p'. Specifically, we place $p' \in \mathcal{P}_q$, if n satisfies

$$\exp\left(-nD_{_{p'}}/18\right) \leq \frac{\eta}{2C(p')\iota(p')^2n^2}\frac{36}{\pi^4}, \qquad (6)$$

and

$$F_q^{-1}(1 - \sqrt{D_{_{p'}}}/3) \leq \log C(p'), \qquad (7)$$

where $C(p')$ is

$$C(p') = 2^{\left(\sup_{r\in B_{p'}} F_r^{-1}(1-\sqrt{D_{_{p'}}}/3)\right)}.$$

Note that $C(p')$ is finite since p' is not deceptive. Comparison with Lemma 8 will give a hint as to why the equations above look the way they do.

b) Description of Φ: If $\mathcal{P}_q = \emptyset$, the scheme does not enter yet. If $\mathcal{P}_q \neq \emptyset$, let p_q denote the distribution in $\mathcal{P}_q$ with the smallest index.

All suffixes of the sequence are said to be *trapped* by p_q—namely, premiums will be based on p_q. The premium assigned for a length-m sequence trapped by p_q is

$$2f_{p_q}\left(\frac{6\eta/2}{2\pi^2n^2}\right).$$

c) Φ enters with probability 1: First, we verify that the scheme enters with probability 1, no matter what distribution $p \in \mathcal{P}$ is in force. Every distribution $p \in \mathcal{P}$ is covered by at least one of the sets in $\mathcal{I}_Q$. There is thus a neighborhood $\epsilon > 0$ around p,

$$I(p, \epsilon) = \{q : |p - q| < \epsilon\}$$

such that $I(p, \epsilon) \subseteq Q$ where $Q \in \mathcal{I}_Q$ with the set with the smallest index among all sets in $\mathcal{I}_Q$ that contain $I(p, \epsilon)$. Let p' be the distribution which defines the set Q in $\mathcal{I}_Q$.

With probability 1, the type of sequences generated by p lies within $I(p, \epsilon)$ [7] (see also Lemma 8 for an alternate proof). Now (6) will hold for all types in $I(p, \epsilon)$, if we make n large enough—since $C(p')$ and $\iota(p')$ do not change with n and the right side diminishes to zero polymonimally with n, while the left diminishes exponentially to zero. Lastly, (7) will also hold almost surely, since $F_q^{-1}(1 - \sqrt{D_{_{p'}}}/3) \to F_p^{-1}(1 - \sqrt{D_{_{p'}}}/3)$ with probability 1. which in turn is $\leq \log C(p')$.

Thus the scheme enters with probability 1.

d) Probability of bankruptcy $\leq \eta$: We now analyze the scheme. Consider any $p \in \mathcal{P}$. Among sequences on which Φ has entered, we will distinguish between those that are in *good* traps and those in *bad* traps. If x is trapped by p' such that $p \in B_{p'}$, p' is a good trap. Conversely, if $p \notin B_{p'}$, p' is a bad trap.

(Good traps) Suppose a length-n sequence x^n is in a good trap, namely, it is trapped by a distribution p' such that $p \in B_{p'}$. Recall that the premium assigned is

$$2f_{p'}\left(\frac{6\eta}{2\pi^2n^2}\right) \geq 2F_p^{-1}\left(1 - \frac{6\eta}{2\pi^2n^2}\right),$$

where the inequality follows because p' is not deceptive, and p is within the reach of p'. Therefore, the scheme is bankrupted with probability at most $\delta' = 6\eta/2\pi^2n^2$ in the next step. Therefore, sequences in good traps contribute at most $\eta/2$ to the probability of bankruptcy.

(Bad traps) We will show that the probability with which sequences generated by p fall into bad traps $\leq \eta/2$. Pessimistically, the conditional probability of bankruptcy given a sequence falls into a bad trap is 1. Thus the contribution to bankruptcy by sequences in bad traps is at most $\eta/2$.

Let q be any length-n type trapped by $\tilde{p}$ with reach $\tilde{\epsilon}$ such that $p \notin B_{\tilde{p}}$, we obtain from Lemma 7 that $\mathcal{J}(p, q) \geq \frac{\tilde{\epsilon}^2 \ln 2}{16}$. Hence, for all q trapped by $\tilde{p}$,

$$\frac{1}{2\ln 2}|p - q|_1^2 \geq \mathcal{J}^2(p, q)\frac{\ln 2}{2} = \frac{\tilde{\epsilon}^4(\ln 2)^3}{512} = D^2_{_{\tilde{p}}}$$

Thus, for $p \in \mathcal{P}^\infty$, the probability the type of a length

n sequence q falls is trapped by a bad $\tilde{p}$

$$\leq p\left(|q-p|^2 \geq D_{\tilde{p}} \text{ and } F_q^{-1}(1-\frac{\sqrt{D_{\tilde{p}}}}{3}) \leq \log C(\tilde{p})\right)$$
$$\leq (C(\tilde{p})-2)\exp\left(-\frac{nD_{\tilde{p}}}{18}\right)$$
$$\leq \frac{\eta(C(\tilde{p})-2)}{2C(\tilde{p})\iota(\tilde{p})^2 n^2}\frac{36}{\pi^4}$$
$$\leq \frac{\eta}{2\iota(\tilde{p})^2 n^2}\frac{36}{\pi^4},$$

where the inequalities follow from Lemma 8 and from (6) and (7). Therefore, the probability of sequences falling into bad traps

$$\leq \sum_{n\geq 1}\sum_{p'\in\mathcal{I}_{\mathcal{Q}}} \frac{\eta}{2\iota(p')^2 n^2}\frac{36}{\pi^4} \leq \eta/2$$

since $\sum_{p'\in\mathcal{I}_{\mathcal{Q}}} \frac{1}{\iota(p')^2} \leq \sum_{n\geq 1}\frac{1}{n^2} = \frac{\pi^2}{6}$. The theorem follows. □

IV. Appendix

Proofs of Lemma 5, 6, and 7 can be found in [3].

Lemma 5. For distributions p and q over some support $\mathcal{A}$, with $\mathcal{J}(p,q) \leq \epsilon$. For some $S \subset \mathcal{A}$ and $\alpha < 1-\ln 2 = .30685$, if $p(S) \geq 1-\alpha$, then

$$q(S) \geq 1 - 2\epsilon - 2h(\alpha).$$ □

Lemma 6. For any two distributions p, q,

$$\frac{1}{4\ln 2}|p-q|_1^2 \leq \mathcal{J}(p,q) \leq \frac{1}{\ln 2}|p-q|_1$$

and for any three distributions p, q, r,

$$\mathcal{J}(p,q)+\mathcal{J}(q,r) \geq \mathcal{J}^2(p,r)\frac{\ln 2}{8}.$$ □

Lemma 7. Suppose a type q is trapped by p_0 with reach ϵ_0. For all $p \in \mathcal{P}$ with $\mathcal{J}(p,p_0) \geq \epsilon_0$,

$$\mathcal{J}(p,q) \geq \frac{\epsilon_0^2 \ln 2}{16}.$$ □

Lemma 8. Let p be any distribution over $\mathbb{N}$, and let $\delta > 0$ and k be any number ≥ 2. Let X_1^n be a sequence generated *i.i.d.* p. Then

$$p(|q(X^n)-p| > \delta \text{ and } F_q^{-1}(1-\delta/3) \leq k)$$
$$\leq (2^k-2)\exp\left(-\frac{n\delta^2}{18}\right).$$

Proof There is a similar lemma in [8]. The difference from [8] is that the right side of the inequality above does *not* depend on p, and this property is crucial for its use here.

The starting point is the following result. Suppose p' has a finite support L. Then from [9], if we consider length n sequences,

$$p'(|q(X^n)-p'| \leq \delta) \geq 1-(2^L-2)\exp\left(-\frac{n\delta^2}{2}\right). \quad (8)$$

Since $k \geq 2$, consider the distributions p' and q' with support $A = \{1,\ldots,k-1\}\cup\{-1\}$, obtained as

$$p'(i) = \begin{cases} p(i) & i < k \\ \sum_{j=k}^{\infty} p(j) & i = -1 \end{cases},$$

and similary for q'. We apply (8) to obtain the Lemma.

From (8),

$$p'(|p'-q'| > \delta/3) \leq (2^k-2)\exp\left(-\frac{n\delta^2}{18}\right).$$

We will show that if $F_q^{-1}(1-\delta/3) \leq k$ and $|p-q| > \delta$ then $q'(-1) \leq \delta/3$ and $|p'-q'| > \delta/3$. Thus, we will have

$$p(|q(X^n)-p| > \delta \text{ and } F_q^{-1}(1-\delta/3) \leq k)$$
$$\leq p'(|p'-q'| > \delta/3 \text{ and } q'(-1) \leq \delta/3)$$
$$\leq (2^k-2)\exp\left(-\frac{n\delta^2}{18}\right).$$

Finally, as in [8],

$$|p-q| - \sum_{l=1}^{k-1}|p(l)-q(l)|$$
$$\leq \sum_{j=k}^{\infty}(p(j)-q(j)) + 2\sum_{j=k}^{\infty}q(j)$$
$$\leq |p'(-1)-q'(-1)| + 2\delta/3.$$

Since $p(l) = p'(l)$ and $q(l) = q'(l)$ for all $l = 1,\ldots,k-1$, we have $|p'-q'| \geq |p-q| - 2\delta/3$. If $|p-q| \geq \delta$ in addition, $|p'-q'| \geq \delta/3$. □

Acknowledgments

N. Santhanam was supported by NSF Grants CCF-1065632, CCF-1018984 and EECS-1029081. V. Anantharam was supported by the ARO MURI grant W911NF- 08-1-0233, Tools for the Analysis and Design of Complex Multi-Scale Networks, the NSF grant CNS-0910702, the NSF Science & Technology Center grant CCF-0939370, Science of Information, Marvell Semiconductor Inc., and the U.C. Discovery program.

References

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

[2] 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.

[3] N. Santhanam and V. Anantharam. What risks lead to ruin. In *Allerton Conference on Computing, Communication and Control*, pages 844 – 850, Oct 2010.

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

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

[6] J. Munkres. *Topology*. Prentice Hall, 2000.

[7] K.L. Chung. A note on the ergodic theorem of information theory. *Annals of Mathematical Statistics*, 32:612—614, 1961.

[8] S. Ho and R. Yeung. On information divergence measures and joint typicality. *IEEE Transactions on Information Theory*, 56(12):5893–5905, 2010.

[9] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. Weinberger. Universal discrete denoising: Known channel. *IEEE Transactions on Information Theory*, 51(1):5—28, Jan 2005.