On the Frequency Distribution of Non-Independent Random Values

Thomas Holenstein Renato Renner

Department of Computer Science
Swiss Federal Institute of Technology (ETH), Zurich
CH-8092 Zurich, Switzerland
{thomahol,renner}@inf.ethz.ch

November 2003

Abstract

Let $\mathbf{Z} = (Z_1, \ldots, Z_n)$ be an n-tuple of random variables and let $\mathcal{P}$ be a convex set containing all conditional probability distributions $P_{Z_i|Z_1=z_1 \cdots Z_{i-1}=z_{i-1}}$. We show that the frequency distribution of the elements in the n-tuple $\mathbf{Z}$ is contained in an ε-environment of $\mathcal{P}$, except with small probability.

1 Preliminaries

Definition 1.1. The *variational distance* between two probability distributions P and Q over the same range $\mathcal{Z}$ is defined by

$$\delta(P, Q) := \frac{1}{2} \sum_{z \in \mathcal{Z}} |P(z) - Q(z)| \ .$$

Definition 1.2. Let $\varepsilon \geq 0$. The ε-environment $\mathcal{B}_\varepsilon(\mathcal{P})$ of a set of probability distributions $\mathcal{P}$ with range $\mathcal{Z}$ is the set of probability distributions P' with range $\mathcal{Z}$ such that $\delta(P, P') \leq \varepsilon$ for some $P \in \mathcal{P}$.

Definition 1.3. Let $\mathbf{z} := (z_1, \ldots, z_n)$ be an n-tuple of elements of a set $\mathcal{Z}$. The *frequency* $Q_\mathbf{z}$ of $\mathbf{z}$ is the function from $\mathcal{Z}$ to $[0, 1]$ defined by

$$Q_\mathbf{z}(z) := \frac{|\{i : z_i = z\}|}{n} \ ,$$

for $z \in \mathcal{Z}$.

It is easy to see that the frequency $Q_\mathbf{z}$ is a probability distribution on $\mathcal{Z}$.

Definition 1.4. A *martingale* is an $n+1$-tuple $(Z_0, \ldots, Z_n)$ of random variables on $\mathbb{R}$ such that,

$$E[Z_i | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] = z_{i-1}$$

for any $i \in \{1, \ldots, n\}$ and $z_1, \ldots, z_{n-1} \in \mathcal{Z}$ (if $\text{Prob}[Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] > 0$).

Theorem 1.5 (Azuma's Inequality). *Let* $(Z_0, \ldots, Z_n)$ *be a martingale with* $Z_0 = 0$ *and*

$$|Z_i - Z_{i-1}| \leq 1$$

for all $i \in \{1, \ldots, n\}$*. Then, for any* $\mu > 0$*,*

$$\mathrm{Prob}[Z_n > \mu\sqrt{n}] < e^{-\mu^2/2} \ .$$

Proof. See, e.g., p. 95 of [1]. □

2 Main Theorem and Proof

Theorem 2.1. *Let* $\mathbf{Z} = (Z_1, \ldots, Z_n)$ *be an* n*-tuple of random variables with alphabet* $\mathcal{Z}$ *and let* $\mathcal{P}$ *be a convex set of probability distributions on* $\mathcal{Z}$ *such that*

$$P_{Z_i|Z_1=z_1,\ldots,Z_{i-1}=z_{i-1}} \in \mathcal{P}$$

for all $i \in \{1, \ldots, n\}$ *and* $z_1, \ldots, z_{i-1} \in \mathcal{Z}$ *(if* $\mathrm{Prob}[Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] > 0$*). Then, for any* $\varepsilon \geq 0$*,*

$$\mathrm{Prob}[Q_{\mathbf{Z}} \in \mathcal{B}_\varepsilon(\mathcal{P})] > 1 - 2^{|\mathcal{Z}|} e^{-n\varepsilon^2/2} \ .$$

Proof. Let $\mathcal{A}$ be any fixed nonempty proper subset of $\mathcal{Z}$. For $i \in \{1, \ldots, n\}$, let B_i be the binary random variable which takes the value 1 if $Z_i \in \mathcal{A}$ and 0 otherwise, let p_i be the function on $\mathcal{Z}^{i-1}$ defined by

$$p_i(z_1, \ldots, z_{i-1}) := \mathrm{Prob}[B_i = 1 | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] \ ,$$

and let

$$D_i := B_i - p_i(Z_1, \ldots, Z_{i-1}) \ .$$

For $i \in \{0, \ldots, n\}$, let

$$S_i := \sum_{j=1}^{i} D_i \ .$$

We show that the $n+1$-tuple $(S_0, \ldots, S_n)$ is a martingale, i.e.,

$$E[S_i | S_0 = s_0, \ldots, S_{i-1} = s_{i-1}] = s_{i-1} \qquad (1)$$

for all $i \in \{1, \ldots, n\}$ and $s_0, \ldots, s_{n-1}$. Since B_i is binary, we have

$$E[B_i | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] = p_i(z_1, \ldots, z_{i-1}) \ ,$$

and, by the definition of D_i,

$$E[D_i | Z_1 = z_1, \ldots, Z_{i-1} = z_{i-1}] = 0 \ .$$

The values of $(D_1, \ldots, D_n)$ are uniquely determined by the values of $(Z_1, \ldots, Z_n)$. We thus have

$$E[D_i | D_1 = d_1, \ldots, D_{i-1} = d_{i-1}] = 0$$

and, using the definition of S_i,

$$E[S_i|D_1 = d_1, \ldots, D_{i-1} = d_{i-1}] = E[\sum_{j=1}^{i} D_i|D_1 = d_1, \ldots, D_{i-1} = d_{i-1}] = \sum_{j=1}^{i-1} d_j \ .$$

which implies (1). Since $S_0 = 0$ and $|S_i - S_{i-1}| = |D_i| \leq 1$ for all $i \in \{1, \ldots, n\}$, we can apply Theorem 1.5 leading to

$$\mathrm{Prob}[S_n > \mu\sqrt{n}] < e^{-\mu^2/2} \ . \tag{2}$$

For any $\mathcal{A} \subseteq \mathcal{Z}$, let $s_\mathcal{A}$ be the function

$$s_\mathcal{A} : \quad P \longmapsto \sum_{z \in \mathcal{A}} P(z) \ ,$$

defined on the set of probability distributions on $\mathcal{Z}$. For any $\mathbf{z} = (z_1, \ldots, z_n)$, let $P_\mathbf{z}$ be the probability distribution on $\mathcal{Z}$ defined by

$$P_\mathbf{z}(z) := \frac{1}{n} \sum_{j=1}^{n} \mathrm{Prob}[Z_j = z | Z_1 = z_1, \ldots, Z_{j-1} = z_{j-1}] \tag{3}$$

for $z \in \mathcal{Z}$. Since, by the definition of $\mathbf{B} = (B_1, \ldots, B_n)$,

$$s_\mathcal{A}(Q_\mathbf{Z}) = Q_\mathbf{B}(1) = \frac{1}{n} \sum_{j=1}^{n} B_j$$

and, by the definition of p_j and $P_\mathbf{z}$,

$$s_\mathcal{A}(P_\mathbf{Z}) = \frac{1}{n} \sum_{j=1}^{n} p_j(Z_1, \ldots, Z_{j-1})$$

we find

$$s_\mathcal{A}(Q_\mathbf{Z}) - s_\mathcal{A}(P_\mathbf{Z}) = \frac{1}{n} \sum_{j=1}^{n} \big(B_j - p_j(Z_1, \ldots, Z_{j-1})\big) = \frac{1}{n} \sum_{j=1}^{n} D_j = \frac{1}{n} S_n \ .$$

It follows from (2) that

$$\mathrm{Prob}[s_\mathcal{A}(Q_\mathbf{Z}) - s_\mathcal{A}(P_\mathbf{Z}) > \frac{\mu}{\sqrt{n}}] < e^{-\mu^2/2} \ . \tag{4}$$

The variational distance between $Q_\mathbf{Z}$ and $P_\mathbf{Z}$ can be written as

$$\delta(Q_\mathbf{Z}, P_\mathbf{Z}) = \max_{\mathcal{A} \subset \mathcal{Z}} \big(s_\mathcal{A}(Q_\mathbf{Z}) - s_\mathcal{A}(P_\mathbf{Z})\big)$$

where the maximum ranges over all non-empty proper subsets $\mathcal{A}$ of $\mathcal{Z}$. Applying the union bound for all $2^{|\mathcal{Z}|} - 2$ subsets $\mathcal{A}$, we conclude from (4)

$$\mathrm{Prob}[\delta(Q_\mathbf{Z}, P_\mathbf{Z}) > \frac{\mu}{\sqrt{n}}] < (2^{|\mathcal{Z}|} - 2) e^{-\mu^2/2} < 2^{|\mathcal{Z}|} e^{-\mu^2/2} \ .$$

The assertion then follows from the observation that the probability distribution $P_\mathbf{z}$, as defined in (3), is a convex combination of distributions from the set $\mathcal{P}$, i.e., since $\mathcal{P}$ is convex, $P_\mathbf{z} \in \mathcal{P}$, for any $\mathbf{z}$, and, consequently, $P_\mathbf{Z} \in \mathcal{P}$. $\square$

References

[1] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, second edition, 2000.