On the estimability of interval probabilities

Marco Cattaneo

School of Mathematics and Physical Sciences
University of Hull

14 March 2018

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

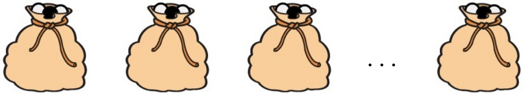
bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n = p$,

can p be learned from data (i.e. consistently estimated)?

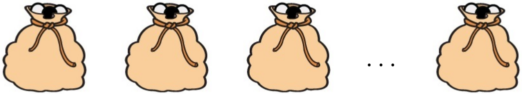
bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n = p$,

can p be learned from data (i.e. consistently estimated)? YES

bags of marbles



bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)?

interpretations of probability

$P(A) \in [0, 1]$: probability of event A

frequentist

subjective

interpretations of probability

$P(A) \in [0, 1]$: probability of event A

frequentist

$P(A) \approx$ relative frequency of occurrence of A in a large number of independent repetitions

subjective

interpretations of probability

$P(A) \in [0,1]$: probability of event A

frequentist

$P(A) \approx$ relative frequency of occurrence of A in a large number of independent repetitions

subjective

$P(A) =$ fair price for a security that pays 1 if A occurs, and 0 otherwise

interpretations of probability

$P(A) \in [0, 1]$: probability of event A

frequentist

$P(A) \approx$ relative frequency of occurrence of A in a large number of independent repetitions

subjective

$P(A) =$ fair price for a security that pays 1 if A occurs, and 0 otherwise

$[\underline{P}(A), \overline{P}(A)] \subseteq [0, 1]$: interval/imprecise probability of event A

interpretations of probability

$P(A) \in [0,1]$: probability of event A

frequentist

$P(A) \approx$ relative frequency of occurrence of A in a large number of independent repetitions

subjective

$P(A) =$ fair price for a security that pays 1 if A occurs, and 0 otherwise

$[\underline{P}(A), \overline{P}(A)] \subseteq [0,1]$: interval/imprecise probability of event A

subjective

$\underline{P}(A), \overline{P}(A) =$ maximum buying price and minimum selling price for a security that pays 1 if A occurs, and 0 otherwise

interpretations of probability

$P(A) \in [0, 1]$: probability of event A

frequentist

$P(A) \approx$ relative frequency of occurrence of A in a large number of independent repetitions

subjective

$P(A) =$ fair price for a security that pays 1 if A occurs, and 0 otherwise

$[\underline{P}(A), \overline{P}(A)] \subseteq [0, 1]$: interval/imprecise probability of event A

frequentist

$\underline{P}(A), \overline{P}(A) \approx$?

subjective

$\underline{P}(A), \overline{P}(A) =$ maximum buying price and minimum selling price for a security that pays 1 if A occurs, and 0 otherwise

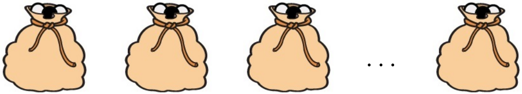
bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)?

bags of marbles

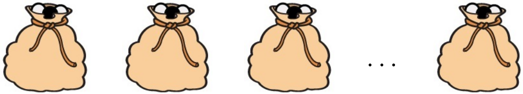
bags of ●/○ marbles:  $\cdots$ $\cdots$

proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)? NO

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\inf p_n, \sup p_n]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)?

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\inf p_n, \sup p_n]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)? NO

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)?

bags of marbles

bags of $\bullet/\circ$ marbles:				$\cdots$		$\cdots$
proportion of $\bullet$:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	$\bullet$	$\circ$	$\circ$	$\cdots$	$\bullet$	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n]$,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated)? NO

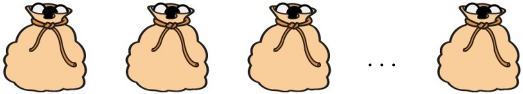
bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup\ p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ?

bags of marbles

bags of $\bullet/\circ$ marbles:				$\cdots$		$\cdots$
proportion of $\bullet$:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	$\bullet$	$\circ$	$\circ$	$\cdots$	$\bullet$	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ? NO (except when δ is 0 or 1)

bags of marbles

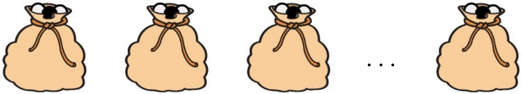
bags of $\bullet/\circ$ marbles:				$\cdots$		$\cdots$
proportion of $\bullet$:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	$\bullet$	$\circ$	$\circ$	$\cdots$	$\bullet$	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ? NO (except when δ is 0 or 1)
- $\mathcal{I}$ is finite?

bags of marbles

bags of $\bullet/\circ$ marbles:  $\cdots$ $\cdots$

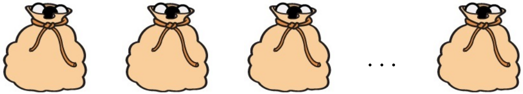
proportion of $\bullet$:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	$\bullet$	$\circ$	$\circ$	$\cdots$	$\bullet$	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ? NO (except when δ is 0 or 1)
- $\mathcal{I}$ is finite? NO (in general)

bags of marbles



bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup\ p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ? NO (except when δ is 0 or 1)
- $\mathcal{I}$ is finite? NO (in general)
- $\mathcal{I}$ consists of pairwise disjoint intervals?

bags of marbles

bags of ●/○ marbles:				$\cdots$		$\cdots$
proportion of ●:	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$
one marble at random:	●	○	○	$\cdots$	●	$\cdots$

assuming that $p_n \in [\underline{p}, \overline{p}] = [\liminf p_n, \limsup p_n] \in \mathcal{I}$,
where $\mathcal{I}$ is a given set of probability intervals,

can $[\underline{p}, \overline{p}]$ be learned from data (i.e. consistently estimated), when:

- $\mathcal{I}$ consists of all intervals of a given length δ? NO (except when δ is 0 or 1)
- $\mathcal{I}$ is finite? NO (in general)
- $\mathcal{I}$ consists of pairwise disjoint intervals? YES

theorem

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $X_n \sim Ber(p_n)$, and let $\mathcal{I}$ be a set of closed (possibly degenerate) subintervals of $[0, 1]$.

Then the following two statements are equivalent:

(i) There are (sequences of) estimators $\underline{\pi}_n, \overline{\pi}_n : \{0, 1\}^n \to [0, 1]$ such that for all $[\underline{p}, \overline{p}] \in \mathcal{I}$ and all sequences $p_n \in [\underline{p}, \overline{p}]$ with $[\liminf p_n,\ \limsup p_n] = [\underline{p}, \overline{p}]$,

$$[\underline{\pi}_n(X_1, \ldots, X_n),\ \overline{\pi}_n(X_1, \ldots, X_n)] \xrightarrow{p} [\underline{p}, \overline{p}].$$

(ii) The elements of $\mathcal{I} \setminus \{[0, 0],\ [1, 1]\}$ are pairwise disjoint.

proof: (ii) $\Rightarrow$ (i)

$$\underline{\pi}_n(X_1,\ldots,X_n) = \begin{cases} 1 & \text{if } X_1 = \cdots = X_n = 1, \\ \inf\left\{\underline{p} : [\underline{p},\overline{p}] \in \mathcal{I},\ \bar{X}_n < \overline{p} + c_n\right\} & \text{otherwise,} \end{cases}$$

$$\overline{\pi}_n(X_1,\ldots,X_n) = \begin{cases} 0 & \text{if } X_1 = \cdots = X_n = 0, \\ \sup\left\{\overline{p} : [\underline{p},\overline{p}] \in \mathcal{I},\ \bar{X}_n > \underline{p} - c_n\right\} & \text{otherwise,} \end{cases}$$

where c_n is any sequence of real numbers such that $\lim c_n = 0$ and $\lim \sqrt{n}\, c_n = +\infty$, while $\inf \varnothing$ and $\sup \varnothing$ can be defined arbitrarily.

proof: (i) $\Rightarrow$ (ii)

Let $[\underline{p}, \overline{p}], [\underline{p}', \overline{p}'] \in \mathcal{I} \setminus \{[0,0],\ [1,1]\}$ be different.

Then there is a sequence of events $A_n \in \sigma(X_1, \ldots, X_n)$ such that

$$\lim P(A_n) = \begin{cases} 1 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

proof: (i) $\Rightarrow$ (ii)

Let $[\underline{p}, \overline{p}], [\underline{p}', \overline{p}'] \in \mathcal{I} \setminus \{[0,0],\ [1,1]\}$ be different.

Then there is a sequence of events $A_n \in \sigma(X_1, \ldots, X_n)$ such that

$$\lim P(A_n) = \begin{cases} 1 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

Now assume that $p_1, p_2, \ldots$ is also a sequence of independent random variables.

The dominated/bounded convergence theorem implies that

$$\lim P(A_n) = \begin{cases} 1 & \text{if a.s. } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if a.s. } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

proof: (i) $\Rightarrow$ (ii)

Let $[\underline{p}, \overline{p}], [\underline{p}', \overline{p}'] \in \mathcal{I} \setminus \{[0,0],\ [1,1]\}$ be different.
Then there is a sequence of events $A_n \in \sigma(X_1, \ldots, X_n)$ such that

$$\lim P(A_n) = \begin{cases} 1 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

Now assume that $p_1, p_2, \ldots$ is also a sequence of independent random variables.
The dominated/bounded convergence theorem implies that

$$\lim P(A_n) = \begin{cases} 1 & \text{if a.s. } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if a.s. } p_n \in [\liminf p_n,\ \limsup\ p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

Furthermore, $X_1, X_2, \ldots$ are independent with $X_n \sim Ber\,(E(p_n))$.

proof: (i) $\Rightarrow$ (ii)

Let $[\underline{p}, \overline{p}], [\underline{p}', \overline{p}'] \in \mathcal{I} \setminus \{[0,0], [1,1]\}$ be different.
Then there is a sequence of events $A_n \in \sigma(X_1, \ldots, X_n)$ such that

$$\lim P(A_n) = \begin{cases} 1 & \text{if } p_n \in [\liminf p_n, \limsup p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if } p_n \in [\liminf p_n, \limsup p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

Now assume that $p_1, p_2, \ldots$ is also a sequence of independent random variables.
The dominated/bounded convergence theorem implies that

$$\lim P(A_n) = \begin{cases} 1 & \text{if a.s. } p_n \in [\liminf p_n, \limsup p_n] = [\underline{p}, \overline{p}], \\ 0 & \text{if a.s. } p_n \in [\liminf p_n, \limsup p_n] = [\underline{p}', \overline{p}']. \end{cases}$$

Furthermore, $X_1, X_2, \ldots$ are independent with $X_n \sim Ber\,(E(p_n))$.

Hence, $\underline{p} \geq \overline{p}'$ or $\underline{p}' \geq \overline{p}$, because otherwise there are two probability distributions for p_n on $\{\underline{p}, \overline{p}\}$ and $\{\underline{p}', \overline{p}'\}$, respectively, with the same expectation and positive probability for both endpoints.

proof: (i) $\Rightarrow$ (ii)

Without loss of generality, assume that $\underline{p} \geq \overline{p}'$, and let $P_{(p'_n)}$ denote the probability distribution of $X_1, X_2, \ldots$ corresponding to the (deterministic) sequence $p_n = p'_n$.

The second Borel–Cantelli lemma and the above results imply that

$$\delta \left(P_{\left(\underline{p} + \frac{\alpha}{n} \, (\overline{p} - \underline{p})\right)}, \, P_{\left(\overline{p}' - \frac{\alpha}{n} \, (\overline{p}' - \underline{p}')\right)} \right) = 1$$

for all $\alpha \in (0, 1)$, where δ is the total variation distance.

proof: (i) $\Rightarrow$ (ii)

Without loss of generality, assume that $\underline{p} \geq \overline{p}'$, and let $P_{(p'_n)}$ denote the probability distribution of $X_1, X_2, \ldots$ corresponding to the (deterministic) sequence $p_n = p'_n$.

The second Borel–Cantelli lemma and the above results imply that

$$\delta\left(P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)},\, P_{\left(\overline{p}'-\frac{\alpha}{n}\,(\overline{p}'-\underline{p}')\right)}\right) = 1$$

for all $\alpha \in (0,1)$, where δ is the total variation distance.

Using Pinsker's inequality (which connects δ to the Kullback–Leibler divergence D_{KL}) and Weierstrass's definition of the gamma function we obtain

$$\delta\left(P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)},\, P_{(\underline{p})}\right) \leq \sqrt{\tfrac{1}{2}\, D_{\mathrm{KL}}\left(P_{(\underline{p})} \,\|\, P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)}\right)} =$$

$$= \sqrt{\underline{p}\, \ln\Gamma\left(\frac{\underline{p}+\alpha\,(\overline{p}-\underline{p})}{\underline{p}}\right) + (1-\underline{p})\, \ln\Gamma\left(\frac{1-\underline{p}-\alpha\,(\overline{p}-\underline{p})}{1-\underline{p}}\right)} \xrightarrow{\alpha\to 0} 0.$$

proof: (i) $\Rightarrow$ (ii)

Without loss of generality, assume that $\underline{p} \geq \overline{p}'$, and let $P_{(p_n')}$ denote the probability distribution of $X_1, X_2, \ldots$ corresponding to the (deterministic) sequence $p_n = p_n'$.

The second Borel–Cantelli lemma and the above results imply that

$$\delta\left(P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)},\, P_{\left(\overline{p}'-\frac{\alpha}{n}\,(\overline{p}'-\underline{p}')\right)}\right) = 1$$

for all $\alpha \in (0,1)$, where δ is the total variation distance.

Using Pinsker's inequality (which connects δ to the Kullback–Leibler divergence D_{KL}) and Weierstrass's definition of the gamma function we obtain

$$\delta\left(P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)},\, P_{(\underline{p})}\right) \leq \sqrt{\tfrac{1}{2}\, D_{\mathrm{KL}}\left(P_{(\underline{p})} \,\|\, P_{\left(\underline{p}+\frac{\alpha}{n}\,(\overline{p}-\underline{p})\right)}\right)} =$$

$$= \sqrt{\underline{p}\, \ln\Gamma\left(\frac{\underline{p}+\alpha\,(\overline{p}-\underline{p})}{\underline{p}}\right) + (1-\underline{p})\, \ln\Gamma\left(\frac{1-\underline{p}-\alpha\,(\overline{p}-\underline{p})}{1-\underline{p}}\right)} \xrightarrow{\alpha\to 0} 0.$$

Analogously, $\delta\left(P_{(\overline{p}')},\, P_{\left(\overline{p}'-\frac{\alpha}{n}\,(\overline{p}'-\underline{p}')\right)}\right) \xrightarrow{\alpha\to 0} 0$, and thus the triangle inequality implies $\underline{p} > \overline{p}'$.