Background on Order and Concentration

Andrew Nobel

August, 2023

Maxima and Minima

The Usual Order Relation

Definition: For $a, b \in \mathbb{R}$ write $a \leq b$ if $(b - a) \geq 0$ and $a < b$ if $(b - a) > 0$

Basic Properties

1. If $a \leq b$ and $b \leq a$ then $a = b$
2. If $a \leq b$ then $-b \leq -a$
3. If $a \leq b$ and $c \leq d$ then $a + c \leq b + d$
4. If $0 \leq a \leq b$ and $0 \leq c \leq d$ then $ac \leq bd$

Note: (2)-(4) continue to hold if we replace $\leq$ by $<$

Maxima and Minima

Basic Properties: Let $a_1, \ldots, a_n \in \mathbb{R}$ and $b_1, \ldots, b_n \in \mathbb{R}$ be finite sequences

1. If $a_i \le b_i$ each i, then $\max_i a_i \ \le \ \max_i b_i$ and $\min_i a_i \ \le \ \min_i b_i$
2. If $a \ge a_i$ for each i then $a \ge \max_i a_i$
3. $-\min_i a_i \ = \ \max_i(-a_i)$ and $-\max_i a_i \ = \ \min_i(-a_i)$
4. If $c \ge 0$ and b are constants, $c \max_i a_i + b \ = \ \max_i(c\, a_i + b)$
5. $\max_i(a_i + b_i) \le \max_i a_i + \max_i b_i$
6. $\min_i(a_i + b_i) \ge \min_i a_i + \min_i b_i$
7. $\max_i a_i - \max_i b_i \le \max_i |a_i - b_i|$
8. $\min_i a_i - \min_i b_i \ \le \ \max_i |a_i - b_i|$

Suprema and Infima

Definition: Let $A \subseteq \mathbb{R}$ be bounded. Recall that

1. $\sup(A) =$ least upper bound for A
2. $\inf(A) =$ greatest lower bound for A

Existence of $\sup$ and $\inf$ follows from construction of the real numbers

- By convention $\sup(\emptyset) = -\infty$ and $\inf(\emptyset) = +\infty$
- If $A \subseteq B$ then $\sup(A) \leq \sup(B)$ while $\inf(A) \geq \inf(B)$

Key Fact: Basic properties of minima and maxima extend to infima and suprema over infinite sets

Iterated Minima and Maxima

Fact: Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be any function. Then the following relations hold

$$\sup_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} h(x, y) = \sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{X}} h(x, y)$$

$$\inf_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} h(x, y) = \inf_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} h(x, y)$$

$$\sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} h(x, y) \leq \inf_{y \in \mathcal{Y}} \sup_{x \in \mathcal{X}} h(x, y)$$

Argmax and Argmin

Definition: The *argmax* of a function $f : \mathcal{X} \to \mathbb{R}$ is the set of points $y \in \mathcal{X}$ where f is maximized

$$\begin{aligned}\operatorname*{argmax}_{x \in \mathcal{X}} f(x) &= \{y \in \mathcal{X} : f(y) \geq f(x) \text{ for all } x \in \mathcal{X}\} \\ &= \left\{y \in \mathcal{X} : f(y) = \max_{x \in \mathcal{X}} f(x)\right\}\end{aligned}$$

Similarly, the *argmin* of f is the set of points $y \in \mathcal{X}$ where f is minimized

$$\begin{aligned}\operatorname*{argmin}_{x \in \mathcal{X}} f(x) &= \{y \in \mathcal{X} : f(y) \leq f(x) \text{ for all } x \in \mathcal{X}\} \\ &= \left\{y \in \mathcal{X} : f(y) = \min_{x \in \mathcal{X}} f(x)\right\}\end{aligned}$$

Argmax and Argmin, cont.

Note that $\operatorname{argmax}_{x \in \mathcal{X}} f(x)$ is a subset of $\mathcal{X}$

- $\max_{x \in \mathcal{X}} f(x)$ is the maximum value of $f(x)$ if this exists
- $\operatorname{argmax}_{x \in \mathcal{X}} f(x)$ is the set of arguments x achieving the maximum value
- $\operatorname{argmax}_{x \in \mathcal{X}} f(x)$ is non-empty iff $\max_{x \in \mathcal{X}} f(x)$ defined

Note that $\operatorname{argmin}_{x \in \mathcal{X}} f(x)$ is a subset of $\mathcal{X}$

- $\min_{x \in \mathcal{X}} f(x)$ is the minimum value of $f(x)$ if this exists
- $\operatorname{argmin}_{x \in \mathcal{X}} f(x)$ is the set of arguments x achieving the minimum value
- $\operatorname{argmin}_{x \in \mathcal{X}} f(x)$ is non-empty iff $\min_{x \in \mathcal{X}} f(x)$ defined

Basic Concentration Inequalities

Recall: Elementary Inequalities for Probability

Fact: If A, B are events, the axioms of probability ensure that

1. If $A \subseteq B$ then $P(A) \leq P(B)$
2. $P(A \cup B) \leq P(A) + P(B)$

Examples: Let X, Y be random variables and $a, b > 0$

- $P(|X + Y| \geq a + b) \leq P(|X| \geq a) + P(|Y| \geq b)$
- $P(|XY| \geq a) \leq P(|X| \geq a/b) + P(|Y| \geq b)$

Idea: Show that the event on the left is contained in the union of the events on the right using elementary logic. Then apply the basic properties of probability above.

Concentration Inequalities

Recall: For a random variable X

- $\mathbb{E}X$ tells us about the center of its distribution
- $\mathrm{Var}(X)$ tells us about the spread of its distribution

Concentration Inequalities: Bounds on the probability that a random variable is far from its expectation

$$\mathbb{P}(X \geq \mathbb{E}X + t) \qquad \mathbb{P}(X \leq \mathbb{E}X - t) \qquad \mathbb{P}(|X - \mathbb{E}X| \geq t)$$

- Often $X = U_1 + \cdots + U_n$ sum of independent random variables
- More generally $X =$ function of independent random variables

Markov's and Chebyshev's Inequalities

Markov's inequality: If $X \geq 0$ and $t > 0$ then

$$\mathbb{P}(X \geq t) \;\leq\; \frac{\mathbb{E}X}{t}$$

Chebyshev's Inequality: If $\mathbb{E}X^2 < \infty$ then for each $t > 0$

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \;\leq\; \frac{\mathrm{Var}(X)}{t^2}$$

- Upper bound may be larger than 1 (not useful)
- Upper bound is less than 1 if $t >$ SD(X)

Moment Generating Functions

Recall: The moment generating function (MGF) of a rv X is defined by

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] \quad \text{for } s \in \mathbb{R}$$

Note that $M_X(s) \geq 0$, and that $M_X(s)$ may be $+\infty$.

Fact: if $X_1, \ldots, X_n$ are independent and $M_{X_i}(s)$ are finite in a neighborhood of 0 then $S_n = X_1 + \cdots + X_n$ has MGF

$$M_{S_n}(s) = \prod_{i=1}^{n} M_{X_i}(s)$$

MGFs are a good way to study sums of independent random variables

Chernoff's Bound

Chernoff Bound: For any random variable X and $t \in \mathbb{R}$

$$\mathbb{P}(X \geq t) \;\leq\; \min_{s>0} e^{-st}\, \mathbb{E}e^{sX} \;=\; \min_{s>0} e^{-st}\, M_X(s)$$

Corollary: If MGF of $(X - \mathbb{E}X)$ is at most $M(s)$ for $s \geq 0$, then for $t > 0$

$$\mathbb{P}(X \geq \mathbb{E}X + t) \;\leq\; \inf_{s>0} e^{-st}\, M(s)$$

- Inequalities for left tail $\mathbb{P}(X \leq \mathbb{E}X - t)$ established in same way
- Bound on $\mathbb{P}(|X - \mathbb{E}X| \geq t)$ obtained by adding L/R tail bounds

Hoeffding's MGF Bound and Hoeffding's Inequality

MGF bound: If $X \in [a, b]$ then for every $s \geq 0$

$$\mathbb{E}e^{s(X-\mathbb{E}X)} \leq e^{s^2(b-a)^2/8}$$

Probability Inequality: Let $X_1, \ldots, X_n$ be independent with $a_i \leq X_i \leq b_i$ and let $S_n = X_1 + \cdots + X_n$. For every $t \geq 0$,

$$\mathbb{P}(S_n - \mathbb{E}S_n \geq t) \;\leq\; \exp\left\{ \frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}$$

Also $\mathbb{P}(S_n - \mathbb{E}S_n \leq -t) \leq$ RHS and $\mathbb{P}(|S_n - \mathbb{E}S_n| \geq t) \leq$ 2 RHS

Note: Bound does not use information about variance of the X_is

Example: Bernoulli Random Variables

Let $X_1, \ldots, X_n$ be iid Bern(p). Note that $\mathbb{E}(\sum_{i=1}^n X_i) = np$

Chebyshev: Uses $\mathrm{Var}(X_i) = p(1-p)$. For each $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^n X_i - np \geq t\right) \leq \frac{n\,p(1-p)}{t^2} \leq \frac{n}{4t^2}$$

Hoeffding: Uses $0 \leq X_i \leq 1$. For each $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^n X_i - np \geq t\right) \leq \exp\left\{\frac{-2t^2}{n}\right\}$$

Note: In each case upper bounds are useful only when $t \gtrsim \sqrt{n}$

Bernoulli Example, cont.

Compare bounds of Chebyshev and Hoeffding when $n = 100$

t	Chebyshev	Hoeffding
5	1	.607
10	.250	.135
12	.173	.0561
14	.128	.0198
16	.0977	.0060
20	.0625	.000335

Upshot: Once the bounds kick in, Hoeffding is better

Bounded Difference Inequality

Setting: Let $\mathcal{X}$ be a set, possibly finite

- Function $f : \mathcal{X}^n \to \mathbb{R}$
- $X_1, \ldots, X_n \in \mathcal{X}$ independent, not necessarily identically distributed

Of interest: bounds on the probability that the random variable

$$Z = f(X_1, \ldots, X_n)$$

is far from its mean $\mathbb{E}Z$

Bounded Difference Inequality

Definition: The ith *difference coefficient* c_i of f is the maximum possible change in the value of f if we change the value of the ith coordinate,

$$c_i \;=\; \sup |f(x_1^n) - f(x_1^{i-1}, x_i', x_{i+1}^n)|$$

where the supremum is over all sequences $x_1, \ldots, x_i, x_i', x_{i+1}, \ldots, x_n \in \mathcal{X}$

Theorem (McDiarmid) If $X_1, \ldots, X_n \in \mathcal{X}$ independent, then for every $t \geq 0$

$$\mathbb{P}\left(|f(X_1^n) - \mathbb{E}f(X_1^n)| \geq t\right) \;\leq\; 2\,\exp\left\{\frac{-2t^2}{\sum_{i=1}^n c_i^2}\right\}$$

Moreover, $\mathrm{Var}(f(X_1^n)) \leq \sum_{i=1}^n c_i^2/4$