Statistics 240 Lecture Notes

P.B. Stark www.stat.berkeley.edu/∼stark/index.html

Revised 4 November 2010

ROUGH DRAFT!

Please report errors to: stark [AT] stat [DOT] berkeley [DOT] edu

1 Part 1: Mathematical preliminaries.

References:

- Clark, A., 1984. *Elements of Abstract Algebra*, Dover.
- Hardy, G., J.E. Littlewood, and G. Pólya, 1952. *Inequalities, 2nd ed.*, Cambridge University Press, Cambridge.
- Le Cam, L., 1986. *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, NY.
- Luenberger, D.G., 1997. *Optimization by Vector Space Methods*, Wiley.
- Lugosi, G., 2006. *Concentration-of-Measure Inequalities*.
- Rudin, W., 1991. *Functional Analysis*, McGraw-Hill.
- Stark, P.B., 2010. *SticiGui*, www.stat.berkeley.edu/users/stark/SticiGui.

1.1 Sets

A set is a collection of objects, called members or elements of the set, without regard for their order. $a \in A$, pronounced "a is an element of A," "a is in A," or "a is a member of A" means that a is an element of the set A. This is the same as writing $A \ni a$, which is pronounced "A contains a." If a is not an element of A, we write $a \notin A$. Sets may be described explicitly by listing their contents, or implicitly by specifying a property that all elements of the set share, or a condition that they satisfy. The contents of sets are enclosed in curly braces: $\{\}$. Examples:

- $A = \{a, b, c, d\}$: the set containing the four elements a, b, c, and d.
- $\emptyset = \{\}$: the empty set, the set that contains no elements.
- $\mathbf{Z} \equiv \{\ldots, -2, -1, 0, 1, 2, \ldots\}$: the integers.
- $\mathbb{N} \equiv \{1, 2, 3, \ldots\}$: the natural (counting) numbers.
- $\mathbb{R} \equiv (-\infty, \infty)$: the real numbers.
- $\mathbb{R}^+ \equiv [-\infty, \infty]$: the extended real numbers.
- $\mathbf{C} \equiv \{a + bi : a, b \in \mathbb{R}\}$, where $i = \sqrt{-1}$: the complex numbers.
- $\mathbf{Q} \equiv \{a/b : a, b \in \mathbf{Z}\}$: the rational numbers.

B is a *subset* of A, written $B \subset A$ or $A \supset B$, if every element of the set B is also an element of the set A. Thus $\mathbb{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbb{R} \subset \mathbf{C}$. The empty set $\emptyset$ is a subset of every set. If $A \subset B$ and $B \subset A$, A and B are the same set, and we write $A = B$. If B is not a subset of A, we write $B \not\subset A$ or $A \not\supset B$. B is a *proper subset* of A if $B \subset A$ but $A \not\subset B$.

The *complement* of A (with respect to the universe $\mathcal{X}$), written A^c or A', is the set of all objects under consideration ($\mathcal{X}$) that are not elements of A. That is, $A^c \equiv \{a \in \mathcal{X} : a \notin A\}$.

The *intersection* of A and B, written $A \cap B$ or AB, is the set of all objects that are elements of both A and B:

$$A \cap B \equiv \{a : a \in A \text{ and } a \in B\}. \tag{1}$$

If $A \cap B = \emptyset$, we say A and B are *disjoint* or *mutually exclusive*.

The *union* of A and B, written $A \cup B$, is the set of all objects that are elements of A or of B (or both):

$$A \cup B \equiv \{a : a \in A \text{ or } a \in B \text{ or both}\}. \tag{2}$$

The *difference* of A and B, $A \setminus B$, pronounced "A minus B," is the set of all elements of A that are not elements of B:

$$A \setminus B \equiv \{a \in A : a \notin B\} = A \cap B^c. \tag{3}$$

Intervals are special subsets of $\mathbb{R}$:

$$\begin{aligned} [a, b] &\equiv \{x \in \mathbb{R} : a \leq x \leq b\} \\ (a, b] &\equiv \{x \in \mathbb{R} : a < x \leq b\} \\ [a, b) &\equiv \{x \in \mathbb{R} : a \leq x < b\} \\ (a, b) &\equiv \{x \in \mathbb{R} : a < x < b\}. \end{aligned}$$

Sometimes we have a collection of sets, indexed by elements of another set: $\{A_\beta : \beta \in B\}$. Then B is called an *index set*. If B is a subset of the integers $\mathbf{Z}$, usually we write A_i or A_j, etc., rather than A_β. If $B = \mathbb{N}$, we usually write $\{A_j\}_{j=1}^{\infty}$ rather than $\{A_\beta : \beta \in \mathbb{N}\}$.

$$\bigcap_{\beta \in B} A_\beta \equiv \{a : a \in A_\beta \ \ \forall \beta \in B\}. \tag{4}$$

($\forall$ means "for all.") If $B = \{1, 2, \ldots, n\}$, we usually write $\bigcap_{j=1}^{n} A_j$ rather than $\bigcap_{j \in \{1,2,\ldots,n\}} A_j$. The notation $\bigcup_{\beta \in B} A_\beta$ and $\bigcup_{j=1}^{n} A_j$ are defined analogously.

A collection of sets $\{A_\beta : \beta \in B\}$ is *pairwise disjoint* if $A_\beta \cap A_{\beta'} = \emptyset$ whenever $\beta \neq \beta'$. The collection $\{A_\beta : \beta \in B\}$ *exhausts* or *covers* the set A if $A \subset \bigcup_{\beta \in B} A_\beta$. The collection $\{A_\beta : \beta \in B\}$ is a *partition* of the set A if $A = \cup_{\beta \in B} A_\beta$ and the sets $\{A_\beta : \beta \in B\}$ are pairwise disjoint. If $\{A_\beta : \beta \in B\}$ are pairwise disjoint and exhaust A, then $\{A_\beta \cap A : \beta \in B\}$ is a partition of A.

A set is *countable* if its elements can be put in one-to-one correspondence with a subset of $\mathbb{N}$. A set is *finite* if its elements can be put in one-to-one correspondence with $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$. If a set is not finite, it is infinite. $\mathbb{N}$, $\mathbf{Z}$, and $\mathbf{Q}$ are infinite but countable; $\mathbb{R}$ is infinite and uncountable.

The notation $\#A$, pronounced "the cardinality of A" is the size of the set A. If A is finite, $\#A$ is the number of elements in A. If A is not finite but A is countable (if its elements can be put in one-to-one correspondence with the elements of $\mathbb{N}$), then $\#A = \aleph_0$ (aleph-null).

The *power set* of a set A is the set of all subsets of the set A. For example, the power set of $\{a, b, c\}$ is

$$\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}. \tag{5}$$

If A is a finite set, the cardinality of the power set of A is $2^{\#A}$. This can be seen as follows: suppose $\#A = n$ is finite. Consider the elements of A to be written in some canonical order. We can specify an element of the power set by an n-digit binary number. The first digit is 1 if the first element of A is in the subset, and 0 otherwise. The second digit is 1 if the second element of A is in the subset, and 0 otherwise, etc. There are 2^n n-digit binary numbers, so there are 2^n subsets. The cardinality of the power set of $\mathbb{N}$ is not $\aleph_0$.

If A is a finite set, B is a countable set and $\{A_j : \beta \in B\}$ is a partition of A, then

$$\#A = \sum_{\beta \in B} \#A_\beta. \tag{6}$$

1.2 Cartesian Products

The notation $(s_1, s_2, \ldots, s_n) \equiv (s_j)_{j=1}^n$ denotes an *ordered n-tuple* consisting of s_1 in the first position, s_2 in the second position, etc. The parentheses are used instead of curly braces to distinguish n-tuples from sets: $(s_j)_{j=1}^n \neq \{s_j\}_{j=1}^n$. The kth *component* of the n-tuple $s = (s_j)_{j=1}^n$, is s_k, $k = 1, 2, \ldots, n$. Two n-tuples are equal if their components are equal. That is, $(s_j)_{j=1}^n = (t_j)_{j=1}^n$ means that $s_j = t_j$ for $j = 1, \ldots, n$. In particular, $(s, t) \neq (t, s)$ unless $s = t$. In contrast, $\{s, t\} = \{t, s\}$ always.

The *Cartesian product of S and T* is $S \bigotimes T \equiv \{(s, t) : s \in S \text{ and } t \in T\}$. Unless $S = T$, $S \bigotimes T \neq T \bigotimes S$. $\mathbb{R}^n$ is the Cartesian product of $\mathbb{R}$ with itself, n times; its elements are n-tuples of real numbers. If s is the n-tuple $(s_1, s_2, \ldots, s_n) = (s_j)_{j=1}^n$,

Let A be a finite set with $\#A = n$. A *permutation* of (the elements of) A is an element s of $\bigotimes_{j=1}^n A = A^n$ whose components are distinct elements of A. That is, $s = (s_j)_{j=1}^n \in A^n$ is a permutation of A if $\#\{s_j\}_{j=1}^n = n$. There are $n! = n(n-1)\cdots 1$ permutations of a set with n elements: there are n choices for the first component of the permutation, $n - 1$ choices for the second (whatever the first might be), $n - 2$ for the third (whatever the first two might be), etc. This is an illustration of the *fundamental rule of counting*: in a sequence of n choices, if there are m_1 possibilites for the first choice, m_2 possibilities for the second choice (no matter which was chosen in the first place), m_3 possibilities for the third choice (no matter which were chosen in the first two places), and so on, then there are $m_1 m_2 \cdots m_n = \prod_{j=1}^n m_j$ possible sequences of choices in all.

The number of permutations of n things taken k at a time, ${}_nP_k$, is the number of ways there are of selecting k of n things, then permuting those k things. There are n choices for the object that will be in the first place in the permutation, $n-1$ for the second place (regardless of which is first), etc., and $n-k+1$ choices for the item that will be in the kth place. By the fundamental rule of counting, it follows that ${}_nP_k = n(n-1)\cdots(n-k+1) = n!/(n-k)!$.

The number of subsets of size k that can be formed from n objects is

$$ {}_nC_k = \binom{n}{k} = {}_nP_k/k! = n(n-1)\cdots(n-k+1)/k! = \frac{n!}{k!(n-k)!}. \tag{7} $$

Because the power set of a set with n elements can be partitioned as

$$ \cup_{k=0}^n \{\text{all subsets of size } k\}, \tag{8} $$

it follows that

$$ \sum_{j=0}^n {}_nC_k = 2^n. \tag{9} $$

1.3 Mappings and Functions

Functions are subsets of Cartesian products. We write $f : \mathcal{X} \rightarrow \mathcal{Y}$, pronounced "$f$ maps $\mathcal{X}$ into $\mathcal{Y}$" or "f is a function with domain $\mathcal{X}$ and co-domain $\mathcal{Y}$" if $f \subset \mathcal{X} \bigotimes \mathcal{Y}$ such that for each $x \in \mathcal{X}$, $\exists 1 y \in \mathcal{Y}$ such that $(x, y) \in f$. (The notation $\exists 1 y$ means that there exists exactly one value of y.) The set $\mathcal{X}$ is called the *domain* of f and $\mathcal{Y}$ is called the co-domain of f. If the $\mathcal{X}$-component of an element of f is x, we denote the $\mathcal{Y}$-component of that element of f by fx or $f(x)$, so that $(x, fx) \in f$; we write $f : x \mapsto y = f(x)$. The functions f and g are equal if they are the same subset of $\mathcal{X} \bigotimes \mathcal{Y}$, which means that they have the same domain $\mathcal{X}$, and $fx = gx \; \forall x \in \mathcal{X}$.

Let $A \subset \mathcal{X}$. The *image* of A under f is

$$ fA = f(A) \equiv \{y \in \mathcal{Y} : (x, y) \in f \text{ for some } x \in A\}. $$

More colloquially, we would write this as

$$ fA = \{y \in \mathcal{Y} : f(x) = y \text{ for some } x \in A\}. $$

If $f\mathcal{X}$ is a proper subset of $\mathcal{Y}$, f is *into*. If $f\mathcal{X} = \mathcal{Y}$, f is *onto*. For $B \subset \mathcal{Y}$, the *inverse image of B under f* or *pre-image of B under f* is

$$f^{-1}B \equiv \{x \in \mathcal{X} : fx \in B\}.$$

Similarly, $f^{-1}y \equiv \{x \in \mathcal{X} : fx = y\}$ If $\forall y \in \mathcal{Y}$, $\#\{f^{-1}y\} \le 1$, f is *one-to-one* (1:1). If f is one-to-one and onto, i.e., if $\forall y \in \mathcal{Y}$, $\#\{f^{-1}y\} = 1$, f is a *bijection*.

Exercise 1

1. *Does* $f^{-1}(fA) = A$?
2. *Does* $f(f^{-1}B) = B$?
3. *Does* $f^{-1}(C \cap D) = f^{-1}C \cap f^{-1}D$?
4. *Does* $f(C \cap D) = fC \cap fD$?
5. *Does* $f(C \cup D) = fC \cup fD$?

1.4 Groups

Definition 1 *A* group *is an ordered pair* $(\mathcal{G}, \times)$, *where* $\mathcal{G}$ *is a collection of objects (the elements of the group) and* $\times$ *is a mapping from* $\mathcal{G} \bigotimes \mathcal{G}$ *onto* $\mathcal{G}$,

$$\begin{aligned} \times : \quad \mathcal{G} \bigotimes \mathcal{G} & \to \mathcal{G} \\ (a, b) & \mapsto a \times b, \end{aligned}$$

satisfying the following axioms:

1. $\exists e \in \mathcal{G}$ *s.t.* $\forall a \in \mathcal{G}$, $e \times a = a$. *The element* e *is called the* identity.
2. *For each* $a \in \mathcal{G}$, $\exists a^{-1} \in \mathcal{G}$ *s.t.* $a^{-1} \times a = e$. *(Every element has an inverse.)*
3. *If* $a, b, c \in \mathcal{G}$, *then* $a \times (b \times c) = (a \times b) \times c$. *(The group operation is associative.)*

If, in addition, for every $a, b \in \mathcal{G}$, $a \times b = b \times a$ *(if the group operation commutes), we say that* $(\mathcal{G}, \times)$ *is an* Abelian group *or* commutative group.

Examples of groups include the real numbers together with ordinary addition, $(\mathbb{R}, +)$; the real numbers other than zero together with ordinary multiplication, $(\mathbb{R} \setminus \{0\}, *)$; the rational numbers together with ordinary addition, $(\mathbf{Q}, +)$; and the integers 0 to $p-1$, p prime, together with addition modulo p, $(\{0, 1, \ldots, p-1\}, +)$.

Exercise 2 *1. Show that $\forall a \in \mathcal{G}$, $a \times a^{-1} = e$. (The inverse from the left is also the inverse from the right; equivalently, $(a^{-1})^{-1} = a$.)*

2. Show that $\forall a \in \mathcal{G}$, $ae = a$. (The identity from the left is also the identity from the right.)

1.5 Fields

Definition 2 *An ordered triple $(\mathcal{F}, \times, +)$ is a* field *if $\mathcal{F}$ is a collection of objects and $\times$ and $+$ are operations on $\mathcal{F} \times \mathcal{F}$ such that*

1. *$\mathcal{F}$ is an Abelian group under the operation $+$, with identity 0.*
2. *$\mathcal{F} \setminus \{0\}$ is an Abelian group under the operation $\times$, with identity 1.*
3. *$\times$ is distributive over $+$. I.e., for any a, b, $c \in \mathcal{F}$ $a \times (b + c) = a \times b + a \times c$ and $(a + b) \times c = a \times c + b \times c$.*

The additive inverse of a is denoted $-a$; the multiplicative inverse of a is $a^{-1} = 1/a$.

Examples: $(\mathbb{R}, \times, +)$, where $\times$ is ordinary (real) multiplication and $+$ is ordinary (real) addition. The complex numbers $\mathbf{C}$, with complex multiplication and addition.

These (and the extended reals) are the only fields we will use.

1.6 Arithmetic with ∞

We shall use the following conventions:

- $0 \cdot \infty = \infty \cdot 0 = 0$
- $x + \infty = \infty + x = \infty$, $x \in \mathbb{R}$
- $x \cdot \infty = \infty \cdot x = \infty$, $x > 0$

With these conventions, $([-\infty, \infty], \times, +)$ is a field.

1.7 Linear Vector Spaces

Definition 3 *A linear vector space is an ordered quadruple* $((\mathcal{F}, \times, +_1), \mathcal{X}, \cdot, +_2)$ *where* $(\mathcal{F}, \times, +_1)$ *is a field,* $\mathcal{X}$ *is a set of objects (the vectors), and* $\cdot$ *is an operation on* $\mathcal{F} \bigotimes \mathcal{X}$ *and* $+_2$ *is an operation on* $\mathcal{X} \bigotimes \mathcal{X}$ *such that:*

1. $(\mathcal{X}, +_2)$ *is an Abelian group with identity 0 (the zero vector)*

2. $\cdot : \mathcal{F} \times \mathcal{X} \to \mathcal{X}$*;* $(\alpha, x) \mapsto \alpha \cdot x$ *such that:*

   (a) *If* 1 *is the multiplicative identity on* $\mathcal{F}$*,* $1 \cdot x = x \; \forall x \in \mathcal{X}$*.*

   (b) $\alpha \cdot (x +_2 y) = \alpha \cdot x +_2 \alpha \cdot y \; \forall \alpha \in \mathcal{F}$*,* $x, y \in \mathcal{X}$*. (distribution)*

   (c) $\alpha \cdot (\beta \cdot x) = (\alpha \times \beta) \cdot x \; \forall \alpha, \beta \in \mathcal{F}, x \in \mathcal{X}$*. (association)*

   (d) $(\alpha +_1 \beta) \cdot x = \alpha \cdot x +_2 \beta \cdot x, \; \forall \alpha, \beta \in \mathcal{F}$*,* $x \in \mathcal{X}$*. (distribution)*

We rarely distinguish notationally between $+_1$ and $+_2$, between $\times$ and $\cdot$, or between the additive identity 0 of the field $\mathcal{F}$ and the identity 0 of the Abelian group $\mathcal{X}$. Sometimes, the multiplication symbols are omitted; e.g., $\alpha \cdot x = \alpha x$. Usually, one just calls $\mathcal{X}$ a linear vector space (LVS), omitting mention of the other elements of the quadruple. For our purposes, $\mathcal{F}$ is almost always $\mathbb{R}$. In that case, $\mathcal{X}$ is called a real linear vector space (RLVS).

Definition 4 *A* functional *on a linear vector space is a mapping from the vectors* $\mathcal{X}$ *to the field* $\mathcal{F}$ *of the linear vector space.*

Definition 5 *A* linear combination *of* $\{x_j\}_{j=1}^n \subset \mathcal{X}$*,* $\mathcal{X}$ *a linear vector space, is a vector* $x = \sum_{j=1}^n \alpha_j x_j$*, where* $\{\alpha_j\}_{j=1}^n \subset \mathcal{F}$ *A set* $\{x_\alpha : \alpha \in A\} \subset \mathcal{X}$ *is* linearly dependent *if there exist constants* $\{\beta_\alpha : \alpha \in A\} \subset \mathcal{F}$*, not all equal to zero, such that* $\sum_{\alpha \in A} \beta_\alpha x_\alpha = 0$*. A set is* linearly independent *if it is not linearly dependent.*

Definition 6 *A* subspace *of a linear vector space* $\mathcal{X}$ *is a subset of* $\mathcal{X}$ *that is a linear vector space with the same field* $\mathcal{F}$ *and operations* $+_1, \cdot, +_2, \times$ *as* $\mathcal{X}$*.*

Definition 7 *Let* $\mathcal{X}$ *be a linear vector space,* $A, B \subset \mathcal{X}$*,* $x \in x$*,* $\alpha \in \mathcal{F}$*.*

- $\alpha A \equiv \{\alpha a : a \in A\}$

- $-A \equiv \{-1 \cdot a : a \in A\}$
- $x + A \equiv \{x + a : a \in A\} = A + x$
- $x - A \equiv \{x - a : a \in A\}$
- $A + B \equiv \{a + b : a \in A, b \in B\}$

Exercise 3 *1. Does $x - A = -(A - x)$?*

2. Does $A + A = 2A$?

3. When does $A - A = 2A$?

If you cannot find necessary conditions, give sufficient ones.

Definition 8 *A mapping M from a linear vector space $\mathcal{X}$ into a linear vector space $\mathcal{Y}$ with the same field $\mathcal{F}$ is* linear *iff*

$$M(\alpha x + \beta y) = \alpha M x + \beta M y, \ \ \forall \alpha, \beta \in \mathcal{F}, \ \ x, y \in \mathcal{X}.$$

Definition 9 *Let $\mathcal{X}$ be a real linear vector space. A set $C \subset \mathcal{X}$ is* convex *iff $\alpha C + (1 - \alpha) C \subset C$ $\forall \alpha \in [0, 1]$. A set $B \subset \mathcal{X}$ is* balanced *iff $\alpha B \subset B$ $\forall \alpha$ with $|\alpha| \leq 1$.*

Note that this requires us to define $|\cdot|$ on the field $\mathcal{F}$. For $\mathcal{F} = \mathbb{R}$, let $|\cdot|$ be absolute value; for $\mathcal{F} = \mathbf{C}$, let $|\cdot|$ be the modulus.

Exercise 4 *Show that if C, $D \subset \mathcal{X}$ are convex, then*

1. αC is convex, $\alpha \in \mathbb{R}$

2. $C \cap D$ is convex.

Definition 10 *A linear combination $\sum_j \beta_j x_j$ of elements $\{x_j\}$ of a linear vector space is a* convex combination *if*

1. $\{\beta_j\} \subset \mathbb{R}$,

2. $\beta_j \geq 0$, $\forall j$, and

3. $\sum_j \beta_j = 1$.

Definition 11 *The* convex hull *of a set* $A \subset \mathcal{X}$ *is the intersection of all convex sets that contain* A*. Equivalently, it is the set of all convex combinations of elements of* A*. If* C *is convex, a point* $x \in C$ *is an* extreme point *of* C *if* x *cannot be written as a convex combination of a subset of* C *unless that subset contains* x*. A* polytope *is the convex hull of a finite collection of points.*

Definition 12 *A set* $\{x_\alpha : \alpha \in A\}$ *is a* basis *for a linear vector space* $\mathcal{X}$ *if every* $x \in \mathcal{X}$ *has a* **unique** *representation* $x = \sum_{\alpha \in A} \beta_\alpha x_\alpha$ *with* $\{\beta_\alpha : \alpha \in A\} \subset \mathcal{F}$*. If* $\mathcal{X}$ *has a basis with* n *elements,* $n \in \mathbb{N}$*,* $\mathcal{X}$ *is* finite-dimensional *and the dimension of* $\mathcal{X}$, $\dim(\mathcal{X})$*, is* n*. If* $\mathcal{X}$ *is not finite-dimensional, it is* infinite-dimensional.

Exercise 5 *Show that if* $\mathcal{Y}$ *is a subspace of* $\mathcal{X}$ *and* $\dim(\mathcal{X}) = n$*, then* $\dim(\mathcal{Y}) \le n$*.*

Definition 13 *A* Hamel basis *for a linear vector space* $\mathcal{X}$ *is a maximal linearly independent subset of* $\mathcal{X}$*.*

1.8 Normed linear vector spaces

Norms, completeness.

1.9 Metric spaces

1.10 Topological spaces

Neighborhoods, completeness. Topologies induced by metrics. Cauchy sequences.

1.11 Duals of linear vector spaces

Linear functionals, duality, norms of linear functionals.

1.12 Banach Spaces

Normed linear vector spaces that are complete w.r.t. topology induced by norm.

1.13 Hilbert Spaces

Complete inner product spaces. Self-dual. Riesz Hilbert space representation theorem. Isometric isomorphism between $\mathcal{H}^*$, normed dual of $\mathcal{H}$, and $\mathcal{H}$ itself. That is, every bounded linear functional on $\mathcal{H}$ can be written as the inner product with an element of $\mathcal{H}$; every element of $\mathcal{H}$ defines a bounded linear functional on $\mathcal{H}$; norm of the linear functional and norm of the element are equal.

1.14 Reproducing Kernel Hilbert Spaces

Hilbert space $\mathcal{H}$ of functions on some domain $\mathcal{D}$. Point evaluator (the linear functional that evaluates an element of $\mathcal{H}$ at an arbitrary point $x \in \mathcal{D}$) is a bounded linear functional. By Riesz representation theorem, point evaluator is the inner product with an element of $\mathcal{H}$. Elements of $\mathcal{H}$ are functions on $\mathcal{D}$. Denote by $K_x(y)$ the element of $\mathcal{H}$ corresponding to evaluation at the point $x \in \mathcal{D}$. Kernel is $K_x(y)$, viewed as a function of $(x, y) \in \mathcal{D} \times \mathcal{D}$.

Examples: finite-dimensional spaces of functions are reproducing Kernel Hilbert spaces. Suppose $\mathcal{H}$ is an n-dimensional set of functions. Let $\{f_j(y)\}_{j=1}^n$ be an orthonormal basis for $\mathcal{H}$, so that any element f of $\mathcal{H}$ can be written $f = \sum_{j=1}^n \alpha_j f_j(y)$ for some set of real numbers $\{\alpha_j\}_{j=1}^n$. Since $\{f_j(y)\}_{j=1}^n$ are orthonormal, $\alpha_j = \langle f, f_j \rangle$. Hence

$$\begin{aligned} f(x) &= \sum_{j=1}^n \langle f, f_j \rangle f_j(x) \\ &= \langle f, \sum_{j=1}^n f_j(x) f_j \rangle. \end{aligned} \tag{10}$$

Thus $K_x(y) = \sum_{j=1}^n f_j(x) f_j(y)$.

1.15 Partial order and convexity

Definition 14 *A relation* $\leq$ *is a* partial order *on a set* $\mathcal{X}$ *if for all* $x, y, z \in \mathcal{X}$*,*

1. $x \leq x, \forall x \in \mathcal{X}$
2. *if* $x \leq y$ *and* $y \leq x$*, then* $x = y$
3. *if* $x \leq y$ *and* $y \leq z$ *then* $x \leq z$

If, in addition, for any x*,* $y \in \mathcal{X}$*, either* $x \leq y$ *or* $y \leq x$ *(or both), then* $\leq$ *is an* order*.*

The usual $\leq$ is an order on $\mathbb{R}$. Set inclusion gives an order among the power set (set of all subsets) of a given set: $x \leq y$ if $x \subset y$. One can think of orders as subsets of $\mathcal{X} \otimes \mathcal{X}$ or as mappings from $\mathcal{X} \otimes \mathcal{X} \to \{0, 1\}$. Henceforth, we take $\mathbb{R}$ to be ordered by $\leq$.

Definition 15 *A set $C \subset \mathcal{X}$, $\mathcal{X}$ a linear vector space, is a* cone *with vertex* 0 *iff* $\alpha C \subset C$ $\forall \alpha \geq 0$. *A set $C \subset \mathcal{X}$, $\mathcal{X}$ a linear vector space, is a* cone with vertex p *if* $C = p + C_0$*, where* C_0 *is a cone with vertex* 0.

Definition 16 *Let $\mathcal{X}$ be a LVS and let $P \subset \mathcal{X}$ be a convex cone with vertex 0. For any $x, y \in \mathcal{X}$, we write $x \leq y$ (w.r.t. P) if $y - x \in P$. The cone P is called the* positive cone *in $\mathcal{X}$. $N \equiv -P$ is the* negative cone*. We write $x \geq y$ if $y - x \in N$ (equivalently, if $x - y \in P$).*

Examples. In $\mathbb{R}$, $[0, \infty)$ is the usual positive cone. In $\mathbb{R}^n$, the positive orthant (n-tuples whose components are all non-negative) is the usual positive cone. The set of non-negative functions and the set of monotone functions can form the positive cones in some function spaces.

Note: the relation $\leq$ defined above is almost–but not quite–a partial order on the linear vector space $\mathcal{X}$: it does not satisfy the second axiom. If P satisfies ($x \in P$ and $-x \in P$) implies $x = 0$, then the relation $\leq$ is a partial order.

Exercise 6 *For x, $y \in \mathbb{R}^n$, define*

$$R(x, y) = \begin{cases} \textit{true}, & \max_{j=1}^{n}(y_j - x_j) \geq 0 \\ \textit{false}, & \textit{otherwise}. \end{cases} \tag{11}$$

Does $R(\cdot, \cdot)$ define a partial order on $\mathbb{R}^n$? Why or why not?

Definition 17 *Let $\mathcal{X}$ and $\mathcal{Y}$ be linear vector spaces, let P be the positive cone on $\mathcal{Y}$, and let $T : \mathcal{X} \to \mathcal{Y}$ have domain D. T is* convex *if*

1. *D is a convex subset of $\mathcal{X}$*

2. *$T(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha T x_1 + (1 - \alpha) T x_2$, $\forall x_1, x_2 \in D$, and all $\alpha \in [0, 1]$.*

Note that convexity depends on the definition of the positive cone P in $\mathcal{Y}$! Convexity and related (such as pseudoconvexity and quasiconvexity), play a crucial role in optimization theory. (A mapping T is *quasiconvex* if its domain is convex and $T(\alpha x_1 + (1 - \alpha)x_2) \leq \max(T x_1, T x_2)$, $\forall x_1, x_2 \in D$, and all $\alpha \in [0, 1]$.)

Definition 18 *Let T be a subset of a linear vector space $\mathcal{X}$. The* cone generated by T *or* star of T *is the set*

$$C(T) \equiv \{\alpha t : \alpha \in [0, \infty), t \in T\} \subset \mathcal{X}. \tag{12}$$

1.16 General-purpose Inequalities

1.16.1 The Arithmetic-Geometric Mean Inequality

1.16.2 Rearrangement Inequalities

Two functions, three functions.

1.16.3 The Triangle Inequality and Generalizations

1.16.4 The Cauchy-Schwartz Inequality

If $\mathcal{H}$ is a Hilbert space and $x, y \in \mathcal{H}$, then

$$|(x, y)| \le \|x\|\|y\|. \tag{13}$$

1.16.5 Parseval's Theorem

1.16.6 The Projection Theorem

1.17 Probability Inequalities

This follows [**?**] rather closely.

1.17.1 A Helpful Identity

If X is a nonnegative real-valued random variable,

$$\mathbb{E}X = \int_0^\infty \Pr\{X \ge x\} dx \tag{14}$$

1.17.2 Jensen's Inequality

If ϕ is a convex function from $\mathcal{X}$ to $\Re$, then $\phi(\mathbb{E}X) \le \mathbb{E}\phi(X)$.

1.17.3 Markov's, Chebychev's, and related inequalities

From 13,

$$\mathbb{E}X \geq \int_0^t \Pr\{X \geq x\} dx \geq t \Pr\{X \geq t\} \tag{15}$$

so

$$\Pr\{X \geq t\} \leq \frac{\mathbb{E}X}{t}. \tag{16}$$

Moreover, for any strictly monotonic function f and nonnegative X,

$$\Pr\{X \geq t\} = \Pr\{f(X) \geq f(t)\} \leq \frac{\mathbb{E}f(X)}{f(t)}. \tag{17}$$

In particular, for any real-valued X and real $q > 0$, $|X - \mathbb{E}X|$ is a nonnegative random variable and $f(x) = x^q$ is strictly monotonic, so

$$\Pr\{|X - \mathbb{E}X| \geq t\} = \Pr\{|X - \mathbb{E}X|^q \geq t^q\} \leq \frac{\mathbb{E}|X - \mathbb{E}X|^q}{t^q}. \tag{18}$$

Chebychev's inequality is a special case:

$$\Pr\{|X - \mathbb{E}X|^2 \geq t^2\} \leq \frac{\mathbb{E}|X - \mathbb{E}X|^2}{t^2} = \frac{\mathbf{Var}X}{t^2}. \tag{19}$$

1.17.4 Chernoff bounds

Apply 16 with $f(x) = e^{sx}$ for positive s:

$$\Pr\{X \geq t\} = \Pr\{e^{sX} \geq e^{st}\} \leq \frac{\mathbb{E}e^{sX}}{e^{st}} \tag{20}$$

for all s. For particular X, can optimize over s.

1.17.5 Hoeffding's Inequality

Let $\{X_j\}_{j=1}^n$ be independent, and define $S_n \equiv \sum_{j=1}^n X_j$. Then, applying 19 gives

$$\Pr\{S_n - \mathbb{E}S_n \geq t\} \leq e^{-st}\mathbb{E}e^{sS_n} = e^{-st} \prod_{j=1}^n e^{s(X_j - EX_j)}. \tag{21}$$

Hoeffding bounds the moment generating function for a bounded random variable X: If $a \leq X \leq b$ with probability 1, then

$$\mathbb{E}e^{sX} \leq e^{s^2(b-a)^2/8}, \tag{22}$$

from which follows *Hoeffding's tail bound*.

If $\{X_j\}_{j=1}^n$ are independent and $\Pr\{a_j \le X_j \le b_j\} = 1$, then

$$\Pr\{S_n - \mathbb{E}S_n \ge t\} \le e^{-2t^2/\sum_{j=1}^n (b_j - a_j)^2} \tag{23}$$

and

$$\Pr\{S_n - \mathbb{E}S_n \le -t\} \le e^{-2t^2/\sum_{j=1}^n (b_j - a_j)^2} \tag{24}$$

1.17.6 Hoeffding's Other Inequality

Suppose f is a convex, real function and $\mathcal{X}$ is a finite set. Let $\{X_j\}_{j=1}^n$ be a simple random sample from $\mathcal{X}$ and let $\{Y_j\}_{j=1}^n$ be an iid uniform random sample (with replacement) from $\mathcal{X}$. Then

$$\mathbb{E}f\left(\sum_{j=1}^n X_j\right) \le \mathbb{E}f\left(\sum_{j=1}^n Y_j\right). \tag{25}$$

1.17.7 Bernstein's Inequality

Suppose $\{X_j\}_{j=1}^n$ are independent with $\mathbb{E}X_j = 0$ for all j, $\Pr\{|X_j| \le c\} = 1$, $\sigma_j^2 = \mathbb{E}X_j^2$ and $\sigma = \frac{1}{n}\sum_{j=1}^n \sigma_j^2$. Then for any $\epsilon > 0$,

$$\Pr\{S_n/n > \epsilon\} \le e^{-n\epsilon^2/2(\sigma^2 + c\epsilon/3)}. \tag{26}$$

1.17.8 The Probability Transform

1.17.9 The Massart-Dvoretsky-Kiefer-Wolfowitz Inequality

1.17.10 Transformed order statistics

1.17.11 Martingales; the Optional Stopping Theorem

1.17.12 The Neyman-Pearson Lemma

1.17.13 Vapnik-Cervonenkis Classes

1.17.14 Convergence of Empirical Processes

1.18 Optimization

1.18.1 Convex optimization

1.18.2 Nonsmooth optimization

1.18.3 Numerical Programming: Linear, Quadratic, and Integer

1.18.4 Combinatorial Problems

1.18.5 The Knapsack Problem