Unsorted notes

Gustavo Pereira

September 23, 2019

Here I store some random notes that I may or may note talk about during recitations.

1 Lectures 1 & 2

- <u>Finite additivity</u>

  Let's define some notation. I can define the following for any indexed collection of sets A_i:

  $$A_1 + A_2 := A_1 \cup A_2$$

  or, more generally

  $$\sum_i A_i := \bigcup_i A_i$$

  whenever the collection A_i is pairwise disjoint.

  The idea of assuming additivity – without any further qualification – is that set-function $\mathbf{P}$ satisfies some form of linearity, that is

  $$\mathbf{P}\left(\sum A_i\right) = \sum_i \mathbf{P}\left(A_i\right)$$

  It turns out that the set of indices over which this assumption is made is consequential.

  We call $\mathbf{P}$ *finitely additive* if the above is required to hold for all finite sets of indices. Similarly, if the relationship holds for countably many indices, $\mathbf{P}$ is called *countably additive*.

  Let's investigate an example of finitely, but not countably, additive measure. Here, we are working with a triple $(X, \mathcal{A}, \mathbf{P})$. $\mathcal{A}$ is an *algebra* of sets. Very similar to the usual $\sigma -$ algebra couterpart, but we don't require the assumptions of closedness under unions and intersections to hold for infinitely many set, only finitely many.

  We will work with the following algebra, which is not a σ-algebra. Let X be the set of

all natural numbers, $\mathbf{N}$. Define also

$$\mathcal{A} = \{A \subset \mathbf{N} : A \text{ is finite or } A^c \text{ is finite}\}$$

Example of sets in $\mathcal{A}$: $\{1, 2, 3\}$ and $\{5001, 5002, \ldots\}$. Example of a sets *not* in $\mathcal{A}$: the set of all odd/even/prime numbers.[1]

It's not hard to see that this is satisfies: $\emptyset \in \mathcal{A}$ (since $\emptyset$ is finite) and closedness under intersections/unions. The reason why $\mathcal{A}$ is not a σ-algebra is that each $A_i = \{1, 3, \ldots, 2i + 1\}$ is in $\mathcal{A}$, but its infinite union, the set of all odd numbers, is not.

Now consider the probability measure: $\mathbf{P} : \mathcal{A} \to [0, 1]$:

$$\mathbf{P}(A) = \begin{cases} 1 & \text{if } A \text{ is infinite} \\ 0 & \text{otherwise} \end{cases}$$

Thus, for example, $\mathbf{P}(1, 2, 3) = 0$ and $P(\{1023, 1024, \ldots\}) = 1$.

Such $\mathbf{P}$ trivially satisfies $\mathbf{P}(A + A') = \mathbf{P}(A) + \mathbf{P}(A')$ because the finite union of finite sets is finite.

This probability measure is interesting because it provides a counter-example to continuity when $\mathbf{P}$ is only finitely, but not countably, additive.

For example, it holds that $\{1, 2, \ldots, n\} \uparrow \mathbf{N}$, but

$$1 = \mathbf{P}(\mathbf{N}) = \mathbf{P}\left(\bigcup_n \{1, 2, \ldots, n\}\right) \neq \lim_n \mathbf{P}\left(\{1, 2, \ldots, n\}\right) = 0$$

Moreover, $\{n + 1, n + 2, \ldots\} \downarrow \emptyset$, but

$$0 = \mathbf{P}(\emptyset) = \mathbf{P}\left(\bigcap_n \{n + 1, n + 2, \ldots\}\right) \neq \lim_n \mathbf{P}\left(\{n + 1, n + 2, \ldots\}\right) = 1$$

The CDF of the random variable $X : \mathbf{N} \to \mathbf{N}$, $X(n) = n$ according to $\mathbf{P}$ will satisfy:

$$F_X(k) = \mathbf{P}\{n : X(n) \leq k\} = 0$$

for all n, so $\lim F_X(k) = 0$ for $k \to \infty$.

[1] The sets whose complement is finite are called co-finite sets.

2 Best linear predictor, matrix version

Let $M(n,k)$ denote the linear space of all matrix of dimension $n \times k$.

Suppose we have random vectors $(\mathbf{y}(\omega), \mathbf{z}(\omega))'$. We know additionally that $\mathbf{y} \in M(n,1)$ and $\mathbf{z} \in M(k,1)$ and these vectors have finite mean and variance. Denote their mean by

$$\begin{bmatrix} \mu_y \\ \mu_z \end{bmatrix}$$

and their variance matrix by

$$\begin{bmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

We define the **best linear predictor** of $\mathbf{y}$ given $\mathbf{z}$ as the random variable $\mathbf{w}$ such that

$$\mathbf{w}^* = \alpha^* + \beta^*(\mathbf{z} - \mu_z)$$

where $\alpha^* \in M(n,1)$ and $\beta^* \in M(n,k)$ solve the minimzation problem

$$\min_{\alpha,\beta} \mathbf{E}\left[\|\mathbf{y} - \alpha - \beta(\mathbf{z} - \mu_z)\|^2\right]$$

You can solve it either by using calculus – which can be cumbersome if you're not used to matrix derivatives – or by noting that the minimand is a squared norm generated by the inner product

$$\langle \mathbf{y}, \mathbf{w} \rangle := \mathbf{E}[\mathbf{w}'\mathbf{y}]$$

of all vectors of the type $\mathbf{y} - \mathbf{w}$ where $\mathbf{w} = \alpha + \beta(\mathbf{z} - \mu_z)$ for some α, β.

Let $\epsilon := \mathbf{y} - \mathbf{w}^*$ denote the residual of the minimization problem. Then ϵ must be orthogonal (by Hilbert's projection theorem) to every $\mathbf{w} = \alpha + \beta(\mathbf{z} - \mu_z)$.

Taking $\beta = 0$, we see that $\mathbf{w}^*$ must satisfy

$$0 = \langle \mathbf{y} - \mathbf{w}^*, \alpha \rangle = \mathbf{E}\left[\alpha'\mathbf{y}\right] - \mathbf{E}\left[\alpha'\mathbf{w}^*\right]$$

for all vectors $\alpha \in M(n,1)$. Taking these to be the elements of the canonical basis, we conclude that

$$\mathbf{w}^* = \mu_y$$

Now take $\alpha = 0$. The orthogonality condition now implies that for any $\beta \in M(n,k)$,

$$0 = \langle \mathbf{y} - \beta^*(\mathbf{z} - \mu_z), \beta(\mathbf{z} - \mu_z) \rangle = \mathbf{E}\left[(\mathbf{z} - \mu_z)'\beta' y\right] - \mathbf{E}\left[(\mathbf{z} - \mu_z)'\beta'\beta^*(\mathbf{z} - \mu_z)\right]$$

Use the properties of the trace – namely, that it's linear and that matrix multiplication commutes inside it – and of the expectation operator to conclude that

$$\operatorname{tr}\left(\beta'\mathbf{E}\left[\mathbf{y}(\mathbf{z}-\mu_z)'\right]\right) = \operatorname{tr}\left(\beta'\beta^*\mathbf{E}\left[(\mathbf{z}-\mu_z)\,(\mathbf{z}-\mu_z)'\right]\right)$$

note that $\mathbf{E}[\mathbf{y}(\mathbf{z}-\mu_z)'] = \Sigma_{yz}$ and $\mathbf{E}[(\mathbf{z}-\mu_z)(\mathbf{z}-\mu_z)'] = \Sigma_{zz}$. The equation above then implies that

$$\operatorname{tr}(\beta'\Sigma_{yz}) = \operatorname{tr}(\beta'\beta^*\Sigma_{zz})$$

should hold for all matrices $\beta \in M(n,k)$. That implies,[2]

$$\Sigma_{yz} = \beta^*\Sigma_{zz}$$

which in turn yields $\beta^* = \Sigma_{yz}\Sigma_{zz}^{-1}$ whenever Σ_{zz} has an inverse. In that case, the BLP is

$$\mathbf{w}^* = \mu_y + \Sigma_{yz}\Sigma_{zz}^{-1}(\mathbf{z}-\mu_z)$$

2.1 Appendix: the Trace operator

- let $A(i,j)$ denote the entry (i,j) of any matrix
- Let A be a $m \times n$ matrix. The trace is defined as

  $$\operatorname{tr}A = \sum_{i=1}^{\min\{m,n\}} A(i,i)$$

  in other words, it's just the sum of elements in the main diagonal.
- Some properties of the trace:
  1. $\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B)$ whenever A and B have similar dimensions
  2. $\operatorname{tr}(kA) = k\operatorname{tr}(A)$ for all scalars k
  3. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ whenever dimensions are such that both multiplications make sense

  Curiosity: any operation $\tilde{\operatorname{tr}}$ that satisfies the properties above is equal to tr (modulo multiplication by a constant)
- The trace and expectation operators commute:

  $$\operatorname{tr}(\mathbf{E}A) = \mathbf{E}(\operatorname{tr}A)$$

[2] See the appendix on the trace operator for details.

- Suppose $A \in M(m,n)$ and you want to select element (i,j) from it. Note that

$$A(i,j) = e_i' A \varepsilon_j = tr(e_i' A \varepsilon_j) = tr(\varepsilon_j e_i' A)$$

  where e_i is the i-th element in the canonical basis of R^m and ε_j is the j-th element of the canonical basis of R^n.

  Hence for any (i,j), letting $B = \varepsilon_j e_i' \in M(n,m)$ we have

$$A(i,j) = \mathrm{tr}(BA)$$

- This implies that if A and $\tilde{A}$ are fixed $m \times n$ matrices, and

$$\mathrm{tr}(BA) = tr(B\tilde{A})$$

  holds for every $B \in M(n,m)$, then

$$A = \tilde{A}$$