Interactions, Partitions, Cumulants

based on B. Streitberg, Lancaster Interactions Revisited, *Ann. Stat.*, 1990

Dino Sejdinovic

Gatsby Unit, UCL

May 17, 2013

An Interaction Measure

- A random vector $(X_1, \ldots, X_n) \sim F$ taking values in the product space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $F \in \mathbb{M}^1 = \mathbb{M}^1(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$
- Write $F_{i_1 i_2 \ldots i_{n'}}$ for the marginal $F_{X_{i_1} X_{i_2} \ldots X_{i_{n'}}}$ of the subvector $(X_{i_1}, \ldots, X_{i_{n'}})$.

An Interaction Measure

- A random vector $(X_1, \ldots, X_n) \sim F$ taking values in the product space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $F \in \mathbb{M}^1 = \mathbb{M}^1(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$
- Write $F_{i_1 i_2 \ldots i_{n'}}$ for the marginal $F_{X_{i_1} X_{i_2} \ldots X_{i_{n'}}}$ of the subvector $(X_{i_1}, \ldots, X_{i_{n'}})$.

Definition (Bahadur (1961); Lancaster (1969))

Interaction measure is a signed measure ΔF that vanishes whenever F can be factorised in a non-trivial way as a product of its (possibly multivariate) marginal distributions.

An Interaction Measure

- A random vector $(X_1, \ldots, X_n) \sim F$ taking values in the product space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$; $F \in \mathbb{M}^1 = \mathbb{M}^1(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$
- Write $F_{i_1 i_2 \ldots i_{n'}}$ for the marginal $F_{X_{i_1} X_{i_2} \ldots X_{i_{n'}}}$ of the subvector $(X_{i_1}, \ldots, X_{i_{n'}})$.

Definition (Bahadur (1961); Lancaster (1969))

Interaction measure is a signed measure ΔF that vanishes whenever F can be factorised in a non-trivial way as a product of its (possibly multivariate) marginal distributions.

$$n = 2: \qquad \Delta F = F_{12} - F_1 F_2.$$

$n = 3$

- Possible non-trivial factorizations:

$n = 3$

- Possible non-trivial factorizations:
  - $F_{123} = F_1 F_2 F_3$ (**total independence**)

$n = 3$

- Possible non-trivial factorizations:
    - $F_{123} = F_1 F_2 F_3$ (**total independence**)
    - $F_{123} = F_{12} F_3$, i.e., $X_3 \perp\!\!\!\perp (X_1, X_2)$ and X_1, X_2 dependent
    - $F_{123} = F_{13} F_2$, i.e., $X_2 \perp\!\!\!\perp (X_1, X_3)$ and X_1, X_3 dependent
    - $F_{123} = F_{23} F_1$, i.e., $X_1 \perp\!\!\!\perp (X_2, X_3)$ and X_2, X_3 dependent

$n = 3$

- Possible non-trivial factorizations:
  - $F_{123} = F_1 F_2 F_3$ (**total independence**)
  - $F_{123} = F_{12} F_3$, i.e., $X_3 \perp\!\!\!\perp (X_1, X_2)$ and X_1, X_2 dependent
  - $F_{123} = F_{13} F_2$, i.e., $X_2 \perp\!\!\!\perp (X_1, X_3)$ and X_1, X_3 dependent
  - $F_{123} = F_{23} F_1$, i.e., $X_1 \perp\!\!\!\perp (X_2, X_3)$ and X_2, X_3 dependent
- pairwise independence ($X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp X_3$ and $X_1 \perp\!\!\!\perp X_3$) does not imply any of the above
  - $X_1, X_2 \overset{i.i.d.}{\sim} Bern(1/2)$, $X_3 = X_1 \veebar X_2$.

Lancaster interaction

- *Lancaster interaction measure* (Lancaster, 1969) is a formal product:

$$\Delta_L F = \prod_{i=1}^{n} \left(F_i^* - F_i\right),$$

where $\prod_{j=1}^{n'} F_{i_j}^*$ is understood as the joint measure $F_{i_1 i_2 \ldots i_{n'}}$ of the subvector $\left(X_{i_1}, \ldots, X_{i_{n'}}\right)$.

Lancaster interaction

- *Lancaster interaction measure* (Lancaster, 1969) is a formal product:

$$\Delta_L F = \prod_{i=1}^{n} \left(F_i^* - F_i\right),$$

where $\prod_{j=1}^{n'} F_{i_j}^*$ is understood as the joint measure $F_{i_1 i_2 \ldots i_{n'}}$ of the subvector $\left(X_{i_1}, \ldots, X_{i_{n'}}\right)$.

- $n = 3$:

$$\Delta_L F = F_{123} - F_{12}F_3 - F_{23}F_1 - F_{13}F_2 + 2F_1F_2F_3.$$

  - Any factorization possible $\Longrightarrow$ $\Delta_L F = 0$

Lancaster interaction

- *Lancaster interaction measure* (Lancaster, 1969) is a formal product:

$$\Delta_L F = \prod_{i=1}^{n} \left(F_i^* - F_i \right),$$

where $\prod_{j=1}^{n'} F_{i_j}^*$ is understood as the joint measure $F_{i_1 i_2 \ldots i_{n'}}$ of the subvector $\left(X_{i_1}, \ldots, X_{i_{n'}} \right)$.

- $n = 3$:

$$\Delta_L F = F_{123} - F_{12}F_3 - F_{23}F_1 - F_{13}F_2 + 2F_1F_2F_3.$$

  - Any factorization possible $\implies \Delta_L F = 0$
  - Example: $X_3 \perp\!\!\!\perp (X_1, X_2) \implies X_3 \perp\!\!\!\perp X_1$, $X_3 \perp\!\!\!\perp X_2$, i.e., $F_{123} = F_{12}F_3$ $\implies F_{23} = F_2F_3$, $F_{13} = F_1F_3$

Lancaster interaction (2)

- $n \geq 4$: Lancaster interaction fails to capture all factorizations: does not necessarily vanish for $(X_1, X_2) \perp\!\!\!\perp (X_3, X_4)$

$$\begin{aligned}\Delta_L F &= (F_1^* - F_1)(F_2^* - F_2)(F_3^* - F_3)(F_4^* - F_4), \\ &= (F_{12} - F_1 F_2)(F_{34} - F_3 F_4).\end{aligned}$$

Lancaster interaction (2)

- $n \geq 4$: Lancaster interaction fails to capture all factorizations: does not necessarily vanish for $(X_1, X_2) \perp\!\!\!\perp (X_3, X_4)$

$$\begin{aligned} \Delta_L F &= (F_1^* - F_1)(F_2^* - F_2)(F_3^* - F_3)(F_4^* - F_4), \\ &= (F_{12} - F_1 F_2)(F_{34} - F_3 F_4). \end{aligned}$$

- Interaction measure valid for all n was constructed by Streitberg (1990):

$$\Delta F = \sum_{\pi} (-1)^{|\pi|-1} (|\pi| - 1)! J_\pi F$$

Partitions, partial order, operations

- Write $[n] = \{1, \ldots, n\}$.
- Partition π of $[n]$ is a set of non-empty pairwise disjoint subsets (blocks) of $[n]$, the union of which is equal to $[n]$.
- **Example**: $\{\{1,3\},\{2\},\{4\}\}$, shorthand $13|2|4$, is a partition of $[4]$

Partitions, partial order, operations

- Write $[n] = \{1, \ldots, n\}$.
- Partition π of $[n]$ is a set of non-empty pairwise disjoint subsets (blocks) of $[n]$, the union of which is equal to $[n]$.
- **Example**: $\{\{1,3\},\{2\},\{4\}\}$, shorthand $13|2|4$, is a partition of $[4]$
- Alternatively, an equivalence relation R_π: $(i,j) \in R_\pi \Leftrightarrow i,j$ belong to the same block of π
- **Example**: $R_{13|2|4} = \{(1,1),(2,2),(3,3),(4,4),(1,3),(3,1)\}$

Partitions, partial order, operations

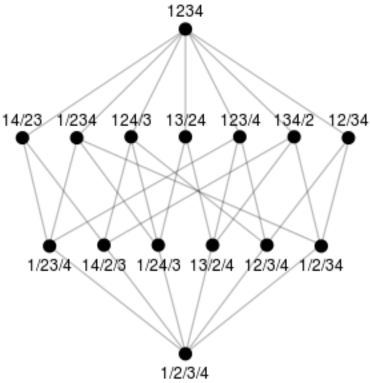
- Write $[n] = \{1, \dots, n\}$.
- Partition π of $[n]$ is a set of non-empty pairwise disjoint subsets (blocks) of $[n]$, the union of which is equal to $[n]$.
- **Example**: $\{\{1,3\},\{2\},\{4\}\}$, shorthand $13|2|4$, is a partition of $[4]$
- Alternatively, an equivalence relation R_π: $(i,j) \in R_\pi \Leftrightarrow i,j$ belong to the same block of π
- **Example**: $R_{13|2|4} = \{(1,1),(2,2),(3,3),(4,4),(1,3),(3,1)\}$
- Partial order: $\tau \leq \pi$ (τ *finer* than π) iff $R_\tau \subseteq R_\pi$
- **meet**: $\tau \wedge \pi \leftrightarrow R_\tau \cap R_\pi$, **join**: $\tau \vee \pi \leftrightarrow \overline{(R_\tau \cup R_\pi)}$

Partitions, partial order, operations

- Write $[n] = \{1, \ldots, n\}$.
- Partition π of $[n]$ is a set of non-empty pairwise disjoint subsets (blocks) of $[n]$, the union of which is equal to $[n]$.
- **Example**: $\{\{1,3\},\{2\},\{4\}\}$, shorthand $13|2|4$, is a partition of $[4]$
- Alternatively, an equivalence relation R_π: $(i,j) \in R_\pi \Leftrightarrow i,j$ belong to the same block of π
- **Example**: $R_{13|2|4} = \{(1,1),(2,2),(3,3),(4,4),(1,3),(3,1)\}$
- Partial order: $\tau \leq \pi$ (τ *finer* than π) iff $R_\tau \subseteq R_\pi$
- **meet**: $\tau \wedge \pi \leftrightarrow R_\tau \cap R_\pi$, **join**: $\tau \vee \pi \leftrightarrow \overline{(R_\tau \cup R_\pi)}$
- **meet refines:** $\pi \wedge \gamma \leq \pi$, **join coarsens:** $\pi \vee \gamma \geq \pi$

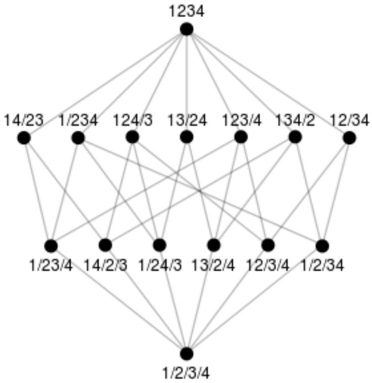
Partitions and interactions (2)

- The set of all partitions: $\mathbf{P}(n)$ forms a lattice



Partitions and interactions (2)

- The set of all partitions: $\mathbf{P}(n)$ forms a lattice



- The finest partition $\mathbf{0} := 1|2|\cdots|n$. The coarsest partition: $\mathbf{1} := 12\cdots n$

Partition operator

- Given a partition $\pi = \pi_1|\pi_2|\ldots|\pi_r$, the associated partition operator on $\mathbb{M}^1$ is given by $J_\pi : F \mapsto F_\pi$, with $F_\pi = \prod_{j=1}^r F_{\pi_j}$, where F_{π_j} is the marginal distribution of the subvector $(X_i\ :\ i \in \pi_j)$.
- **Example**: $J_{13|2|4}F = F_{13}F_2F_4$, $J_{\mathbf{1}}F = F$

Proposition

$$J_\tau \circ J_\pi = J_{\tau\wedge\pi}$$

Interaction operator

Definition

Let $\tau \in \mathbf{P}(n)$. A measure $F \in \mathbb{M}^1$ is said to be τ-decomposable if there exists $\gamma < \tau$ such that $J_\gamma F = F$. A τ-interaction operator is a linear combination

$$\Delta_\tau = \sum_\pi a(\pi, \tau) J_\pi$$

that vanishes for all τ-decomposable measures.
(decomposable$\equiv\mathbf{1}$-decomposable)

- Original definition recovered for $\tau = \mathbf{1}$.
- F isτ-decomposable iff $J_\tau F = F$ and at least one of its marginals F_{τ_j} is itself decomposable.

A property of coefficients

- Δ_τ is a valid τ-interaction operator iff $\Delta_\tau J_\gamma F = 0 \ \forall F, \forall \gamma < \tau$. Now,

$$\begin{aligned} 0 &\equiv \Delta_\tau J_\gamma \\ &= \sum_\pi a(\pi,\tau) J_\pi J_\gamma \\ &= \sum_\pi a(\pi,\tau) J_{\pi \wedge \gamma} \\ &= \textstyle\sum_{\sigma \le \gamma} \left[\displaystyle\sum_{\pi : \pi \wedge \gamma = \sigma} a(\pi,\tau) \right] J_\sigma \end{aligned}$$

A property of coefficients

- Δ_τ is a valid τ-interaction operator iff $\Delta_\tau J_\gamma F = 0 \ \forall F, \ \forall \gamma < \tau$. Now,

$$
\begin{aligned}
0 &\equiv \Delta_\tau J_\gamma \\
&= \sum_\pi a(\pi,\tau) J_\pi J_\gamma \\
&= \sum_\pi a(\pi,\tau) J_{\pi\wedge\gamma} \\
&= \sum_{\sigma\leq\gamma} \left[\sum_{\pi\,:\,\pi\wedge\gamma=\sigma} a(\pi,\tau) \right] J_\sigma
\end{aligned}
$$

- Thus, $\sum_{\pi\,:\,\pi\wedge\gamma=\sigma} a(\pi,\tau) = 0$, for all $\sigma \leq \gamma < \tau$.

Incidence algebra on a lattice

- A bit of notation: *Iverson bracket* $\{\mathcal{P}\} = \begin{cases} 1, & \mathcal{P} \text{ true} \\ 0, & \mathcal{P} \text{ false} \end{cases}$

Incidence algebra on a lattice

- A bit of notation: *Iverson bracket* $\{\mathcal{P}\} = \begin{cases} 1, & \mathcal{P} \text{ true} \\ 0, & \mathcal{P} \text{ false} \end{cases}$
- **Incidence algebra** I_S on the lattice S: set of all maps $g : S \times S \to \mathbb{R}$, s.t. $g(\pi, \tau) \neq 0$ only if $\pi \leq \tau$
- **Convolution** in I_S: $(g * h)(\pi, \tau) = \sum_\gamma g(\pi, \gamma) h(\gamma, \tau)$
- Special elements in I_S: identity $\delta(\pi, \tau) = \{\pi = \tau\}$, and the zeta function $\zeta(\pi, \tau) = \{\pi \leq \tau\}$.
- The **Möbius function** μ is the inverse of the zeta function, i.e., $\mu * \zeta = \zeta * \mu = \delta$.

Main Lemma

Lemma

There exists a unique $a \in I_S$ *such that* $a(\pi, \pi) = 1$ *for all* $\pi \in S$ *and*

$$\sum_{\pi} \{\pi \wedge \gamma = \sigma\} \, a(\pi, \tau) \quad = \quad 0,$$

for all $\tau \in S$ *and all* $\sigma \leq \gamma < \tau$. *Moreover,* $a = \mu$ *is the Möbius function in* I_S.

Proof of the Main Lemma (1)

By definition,
$\delta(\pi,\tau) = \sum_\gamma \zeta(\pi,\gamma)\mu(\gamma,\tau) = \sum_\gamma \{\pi \leq \gamma \leq \tau\}\mu(\pi,\gamma)\zeta(\gamma,\tau)$. Therefore,

$$\begin{aligned} 1 = \delta(\pi,\pi) &= \sum_\gamma \{\pi \leq \gamma \leq \pi\}\mu(\pi,\gamma)\zeta(\gamma,\pi) \\ &= \mu(\pi,\pi)\zeta(\pi,\pi) \\ &= \mu(\pi,\pi). \end{aligned}$$

Proof of the Main Lemma (2)

Let $\tau \in S$ and $\gamma < \tau$. Then,

$$\begin{aligned}\sum_{\pi} \{\pi \wedge \gamma = \gamma\} \mu(\pi, \tau) &= \sum_{\pi} \{\gamma \leq \pi\} \mu(\pi, \tau) \\ &= \sum_{\pi} \zeta(\gamma, \pi)\mu(\pi, \tau) \\ &= \delta(\gamma, \tau) = 0,\end{aligned}$$

which shows the proof for $\sigma = \gamma$.

Proof of the Main Lemma (2)

- Property shown for the case $\sigma = \gamma < \tau$.

Proof of the Main Lemma (2)

- Property shown for the case $\sigma = \gamma < \tau$.

Let σ, such that $\sigma < \gamma < \tau$, be *a maximal element* for which the property has not been shown. Then:

$$\begin{aligned}
\sum_{\pi} \{\pi \wedge \gamma = \sigma\}\, \mu(\pi, \tau) &= \sum_{\pi} \{\pi \wedge \gamma \geq \sigma\}\, \mu(\pi, \tau) \\
&\quad - \sum_{\sigma' > \sigma} \sum_{\pi} \{\pi \wedge \gamma = \sigma'\}\, \mu(\pi, \tau) \\
&= \sum_{\pi} \{\pi \wedge \gamma \geq \sigma\}\, \mu(\pi, \tau) \\
&= \sum_{\pi} \{\pi \geq \sigma\}\, \mu(\pi, \tau) \\
&= \sum_{\pi} \zeta(\sigma, \pi) \mu(\pi, \tau) = \delta(\sigma, \tau) = 0.
\end{aligned}$$

Proof of the Main Lemma (2)

- Property shown for the case $\sigma = \gamma < \tau$.

Let σ, such that $\sigma < \gamma < \tau$, be *a maximal element* for which the property has not been shown. Then:

$$\begin{aligned}
\sum_{\pi} \{\pi \wedge \gamma = \sigma\}\, \mu(\pi,\tau) &= \sum_{\pi} \{\pi \wedge \gamma \geq \sigma\}\, \mu(\pi,\tau) \\
&\quad - \sum_{\sigma' > \sigma} \sum_{\pi} \{\pi \wedge \gamma = \sigma'\}\, \mu(\pi,\tau) \\
&= \sum_{\pi} \{\pi \wedge \gamma \geq \sigma\}\, \mu(\pi,\tau) \\
&= \sum_{\pi} \{\pi \geq \sigma\}\, \mu(\pi,\tau) \\
&= \sum_{\pi} \zeta(\sigma,\pi)\mu(\pi,\tau) = \delta(\sigma,\tau) = 0.
\end{aligned}$$

Induction $\Rightarrow$ ♣

Together

Theorem

Up to multiplicative constant, the τ-interaction operator is given by:

$$\Delta_\tau = \sum_\pi \mu(\pi, \tau) J_\pi$$

- In particular,

$$\begin{aligned} \Delta &= \sum_\pi \mu(\pi, \mathbf{1}) J_\pi \\ &= \sum_\pi (-1)^{|\pi|-1} (|\pi| - 1)! J_\pi \end{aligned}$$

Joint cumulants

- The joint cumulant $\kappa(X_1, \ldots, X_n)$ are defined by the cumulant generating function$g(t_1, \ldots, t_n) = \log \mathbb{E}\left[\exp\left(\sum_{i=1}^n t_i X_i\right)\right]$

Joint cumulants

- The joint cumulant $\kappa(X_1, \ldots, X_n)$ are defined by the cumulant generating functiong$(t_1, \ldots, t_n) = \log \mathbb{E}\left[\exp\left(\sum_{i=1}^n t_i X_i\right)\right]$

1. Symmetry: $\kappa(X_1, \ldots, X_n) = \kappa(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ for all permutations σ
2. Multilinearity:
   $\kappa(\alpha X_1 + \beta Y_1, \ldots, X_n) = \alpha\kappa(X_1, \ldots, X_n) + \beta\kappa(Y_1, \ldots, X_n)$
3. Moment property: $\kappa(\mathbf{X}) = \kappa(\mathbf{Y})$ iff $\mathbf{X}$ and $\mathbf{Y}$ have identical moments up to order n.

Joint cumulants

- The joint cumulant $\kappa(X_1,\ldots,X_n)$ are defined by the cumulant generating function$g(t_1,\ldots,t_n) = \log \mathbb{E}\left[\exp\left(\sum_{i=1}^n t_i X_i\right)\right]$

1. Symmetry: $\kappa(X_1,\ldots,X_n) = \kappa(X_{\sigma(1)},\ldots,X_{\sigma(n)})$ for all permutations σ
2. Multilinearity:
   $\kappa(\alpha X_1 + \beta Y_1,\ldots,X_n) = \alpha\kappa(X_1,\ldots,X_n) + \beta\kappa(Y_1,\ldots,X_n)$
3. Moment property: $\kappa(\mathbf{X}) = \kappa(\mathbf{Y})$ iff $\mathbf{X}$ and $\mathbf{Y}$ have identical moments up to order n.

Fact

$\kappa(X_1,\ldots,X_n) = \int x_1 \cdots x_n d\Delta F$

Corollary

F is decomposable $\Rightarrow \kappa(X_1,\ldots,X_n) = 0.$

Summary

- **Recall**: the second and third cumulant are exactly the same as the second and third central moments

Summary

- **Recall**: the second and third cumulant are exactly the same as the second and third central moments
  - the higher cumulants are neither moments nor central moments, but some other polynomials of the moments

Summary

- **Recall**: the second and third cumulant are exactly the same as the second and third central moments
  - the higher cumulants are neither moments nor central moments, but some other polynomials of the moments
- Lancaster definition: **joint central moments**: $\int x_1 \cdots x_n d\Delta_L F$

Summary

- **Recall**: the second and third cumulant are exactly the same as the second and third central moments
  - the higher cumulants are neither moments nor central moments, but some other polynomials of the moments
- Lancaster definition: **joint central moments**: $\int x_1 \cdots x_n d\Delta_L F$
- Streitberg's correction: **joint cumulants**: $\int x_1 \cdots x_n d\Delta F$

What does this have to do with kernels?

- Three-variable interaction has a simple kernel statistic (joint "kernel cumulant" $\equiv$ joint "kernel central moment"):

$$\left\| \int \left[k_1(\cdot, x_1) \otimes k_2(\cdot, x_2) \otimes k_3(\cdot, x_3) \right] d\widehat{\Delta_L F} \right\|^2_{\mathcal{H}_{k_1 \otimes k_2 \otimes k_3}}$$
$$= \left\| \int \left[\tilde{k}_1(\cdot, x_1) \otimes \tilde{k}_2(\cdot, x_2) \otimes \tilde{k}_3(\cdot, x_3) \right] d\widehat{F} \right\|^2_{\mathcal{H}_{k_1 \otimes k_2 \otimes k_3}}$$
$$= \frac{1}{n^2} \left(\tilde{K}_1 \circ \tilde{K}_2 \circ \tilde{K}_3 \right)_{++}$$

What does this have to do with kernels?

- Three-variable interaction has a simple kernel statistic (joint "kernel cumulant" $\equiv$ joint "kernel central moment"):

$$\left\| \int \left[k_1(\cdot, x_1) \otimes k_2(\cdot, x_2) \otimes k_3(\cdot, x_3)\right] d\widehat{\Delta_L F} \right\|^2_{\mathcal{H}_{k_1 \otimes k_2 \otimes k_3}}$$

$$= \left\| \int \left[\tilde{k}_1(\cdot, x_1) \otimes \tilde{k}_2(\cdot, x_2) \otimes \tilde{k}_3(\cdot, x_3)\right] d\widehat{F} \right\|^2_{\mathcal{H}_{k_1 \otimes k_2 \otimes k_3}}$$

$$= \frac{1}{n^2}\left(\tilde{K}_1 \circ \tilde{K}_2 \circ \tilde{K}_3\right)_{++}$$

- But for $n \geq 4$, the statistic involves the sum over all partitions of order n and different combinations of the centering of the kernel matrices within each summand.