Measuring partial exchangeability with reproducing kernel Hilbert spaces

Hugo Lavenant

Bocconi University

5th Italian Meeting on Probability and Mathematical Statistics

Palermo (Italy), June 9, 2026

Joint work with:

Luiss

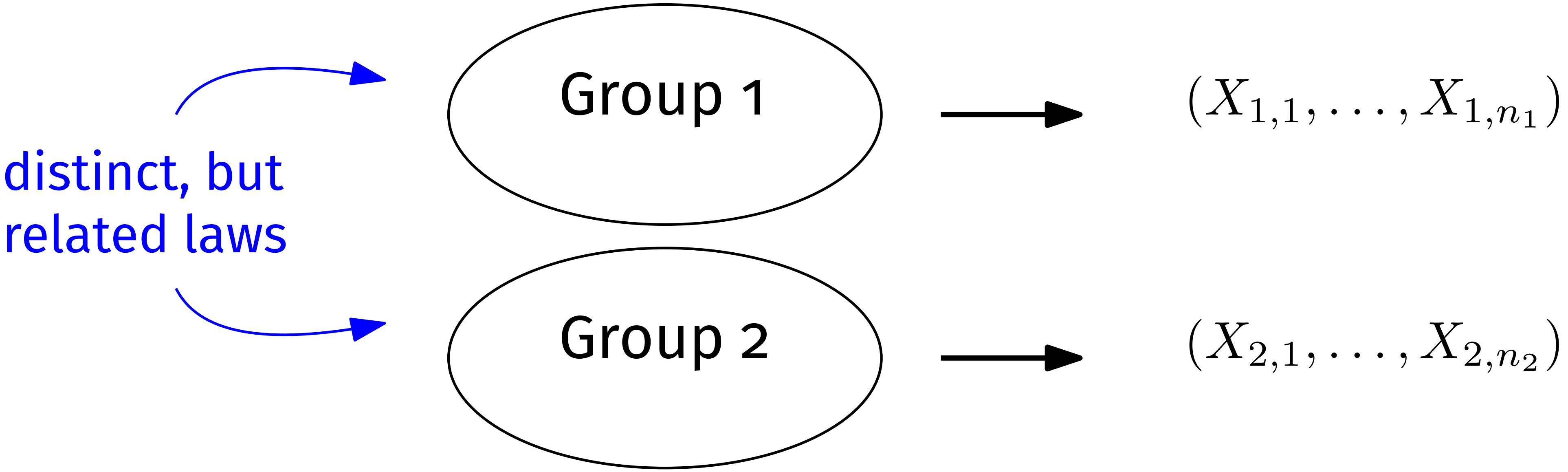
Marta Catalano

Bocconi

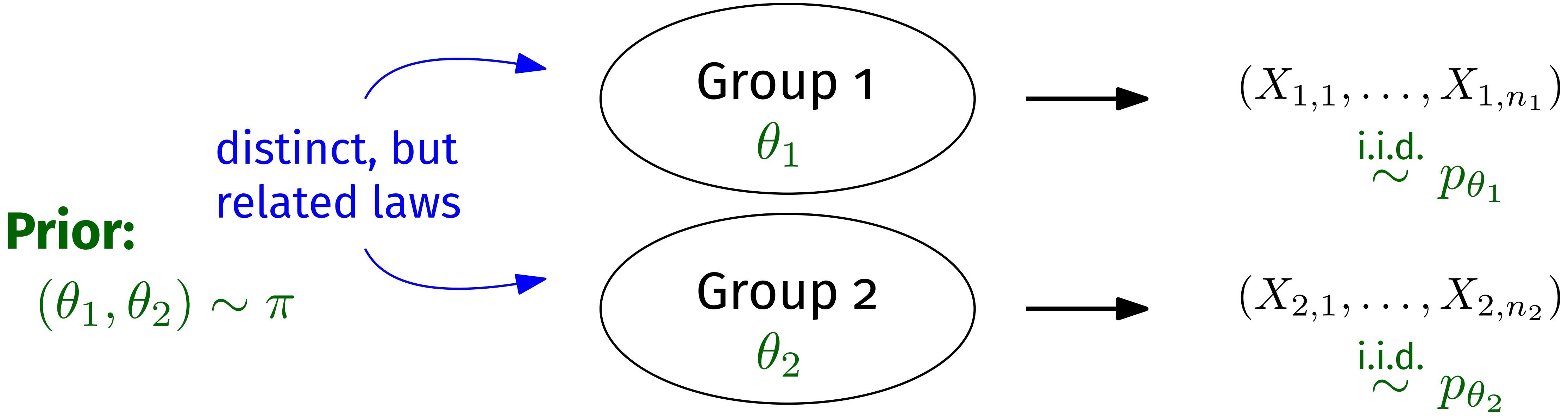
Francesco Mascari

[Arxiv 2509.20221]

Bayesian multi-level modelling



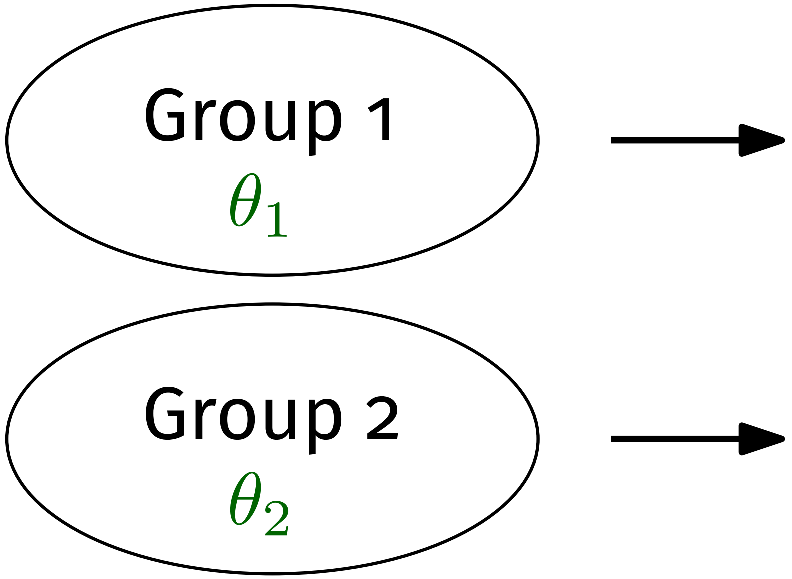
Bayesian multi-level modelling



Bayesian inference allows for borrowing of information if θ_1, θ_2 dependent.

Goal: quantifying the amount of **dependence** between groups a priori and a posteriori.

Bayesian multi-level modelling



$(X_{1,1}, \ldots, X_{1,n_1}) \overset{\text{i.i.d.}}{\sim} p_{\theta_1}$

$(X_{2,1}, \ldots, X_{2,n_2}) \overset{\text{i.i.d.}}{\sim} p_{\theta_2}$

Prior:

$(\theta_1, \theta_2) \sim \pi$

$I = 0$ — $I = 1$

$I \in (0, 1)$

Independence

$\theta_1 \perp \theta_2$

Analyze the two groups separately.

Partial exchangeability: intermediate case

Full exchangeability

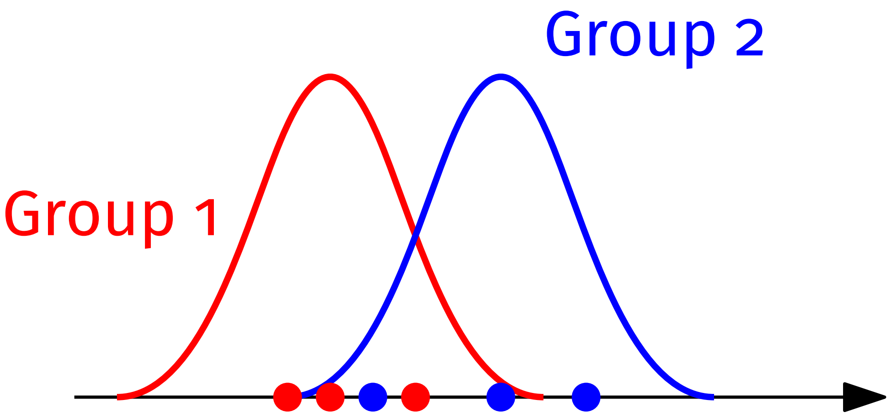
$\theta_1 = \theta_2$ a.s.

There is only one group!

Example: Gaussian case

$$(\theta_1, \theta_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$X_{i,1}, X_{i,2}, \ldots, X_{i,n_i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta_i, 1)$$

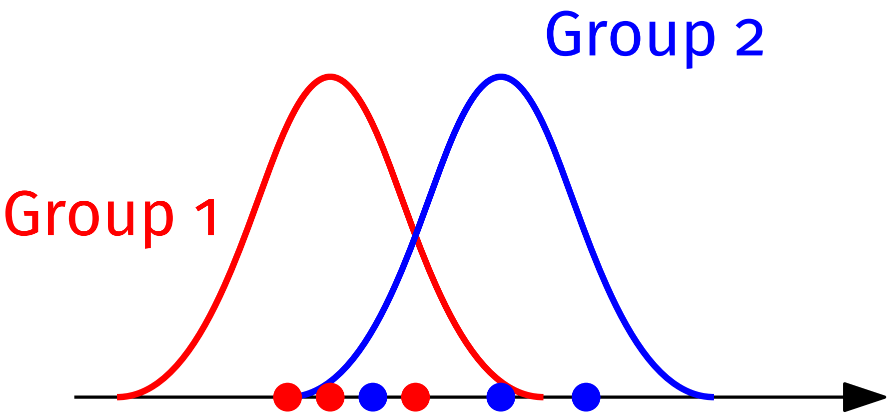


Measure dependence with **linear correlation** $\mathrm{Corr}(\theta_1, \theta_2) = \rho$.

Example: Gaussian case

$$(\theta_1, \theta_2) \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$X_{i,1}, X_{i,2}, \ldots, X_{i,n_i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta_i, 1)$$



Measure dependence with **linear correlation** $\mathrm{Corr}(\theta_1, \theta_2) = \rho$.

Computation. After observing n_1, n_2 data points, the correlation of θ_1, θ_2 a posteriori is:

$$\mathrm{Corr}(\theta_1, \theta_2 | \mathbf{X}_{n_1,n_2}) = \frac{\rho}{\sqrt{1 + n_2(1-\rho^2)}\sqrt{1 + n_2(1-\rho^2)}} \sim \frac{1}{\sqrt{n_1 n_2}} \frac{\rho}{1-\rho^2}.$$

Going nonparametric

$\mathrm{Corr}(\theta_1, \theta_2)$ may depend on the parametrization – and does not apply to nonparametric cases.

Going nonparametric

$\mathrm{Corr}(\theta_1, \theta_2)$ may depend on the parametrization – and does not apply to nonparametric cases.

Note: $(p_{\theta_1}, p_{\theta_2})$ with $(\theta_1, \theta_2) \sim \pi$ is a pair of random measures on $\mathbb{X}$.

Goal: For $(\tilde{P}_1, \tilde{P}_2)$ a pair of random measures over $\mathbb{X}$, build $I(\tilde{P}_1, \tilde{P}_2)$ s.t. :

- $I = 0$ if $\tilde{P}_1 \perp \tilde{P}_2$, and $I = 1$ if $\tilde{P}_1 = \tilde{P}_2$ a.s.
- I is analytically tractable, also for posterior distributions.
- I can be estimated from samples.

Previous proposals

Setwise correlation:

$$I = \mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)).$$

- Can be computed for several models.
- In most priors, does not depend on A.

[Rodríguez et al, 2008], [Leisen et al, 2013], [Griffin and Leisen, 2017], [Camerlenghi et al, 2019], [Beraha et al, 2019], [Ascolani et al, 2023], [Denti et al, 2023], [Lijoi et al, 2023], [Horiguchi et al, 2024], [Colombi et al, 2025]

Previous proposals

Setwise correlation:

$$I = \mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)).$$

- Can be computed for several models.
- In most priors, does not depend on A.

[Rodríguez et al, 2008], [Leisen et al, 2013], [Griffin and Leisen, 2017], [Camerlenghi et al, 2019], [Beraha et al, 2019], [Ascolani et al, 2023], [Denti et al, 2023], [Lijoi et al, 2023], [Horiguchi et al, 2024], [Colombi et al, 2025]

May depend greatly on A on toy examples.

$\tilde{P}_i = W\delta_{x_i} + WP \quad W \in [0,1]$, unif r.v.

There exist A, B such that:

$$\mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = 1$$
$$\mathrm{Corr}(\tilde{P}_1(B), \tilde{P}_2(B)) = -1$$

Previous proposals

Setwise correlation:

$$I = \mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)).$$

- Can be computed for several models.
- In most priors, does not depend on A.

[Rodríguez et al, 2008], [Leisen et al, 2013], [Griffin and Leisen, 2017], [Camerlenghi et al, 2019], [Beraha et al, 2019], [Ascolani et al, 2023], [Denti et al, 2023], [Lijoi et al, 2023], [Horiguchi et al, 2024], [Colombi et al, 2025]

Index based on Wasserstein distance

- Only for Completely Random Vectors.
- Characterize independence and exchangeability.

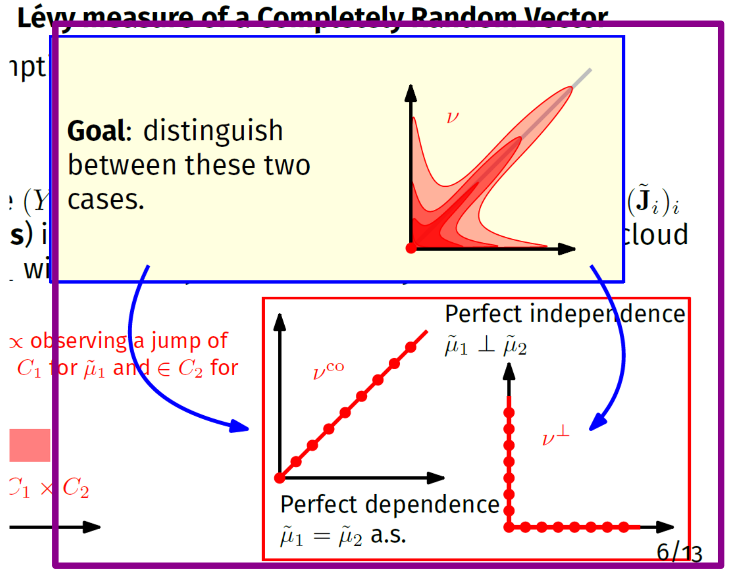
[Catalano et al, 2021, 2024]

May depend greatly on A on toy examples.

$$\tilde{P}_i = W\delta_{x_i} + WP \qquad W \in [0,1], \text{ unif r.v.}$$

There exist A, B such that:

$$\mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)) = 1$$
$$\mathrm{Corr}(\tilde{P}_1(B), \tilde{P}_2(B)) = -1$$



Our proposal: kernel correlation

$k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ measurable, bounded, symmetric and positive definite.

- Gaussian: $k(x,y) = \exp(-|x-y|^2/2\sigma^2)$.
- Linear: $k(x,y) = xy$.

$\forall (a_i)_i, (x_i)_i, \quad \sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0.$

Define:

$$\mathrm{Cov}_k(\tilde{P}_1, \tilde{P}_2) = \mathbb{E}\left[\iint k(x,y)\mathrm{d}\tilde{P}_1(x)\mathrm{d}\tilde{P}_2(y)\right] - \iint k(x,y)\mathrm{d}\mathbb{E}[\tilde{P}_1](x)\mathrm{d}\mathbb{E}[\tilde{P}_2](y),$$

$$\mathrm{Var}_k(\tilde{P}) = \mathrm{Cov}_k(\tilde{P}, \tilde{P}).$$

Our proposal: kernel correlation

$k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ measurable, bounded, symmetric and positive definite.

- Gaussian: $k(x,y) = \exp(-|x-y|^2/2\sigma^2)$.
- Linear: $k(x,y) = xy$.

$\forall (a_i)_i, (x_i)_i, \quad \sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0.$

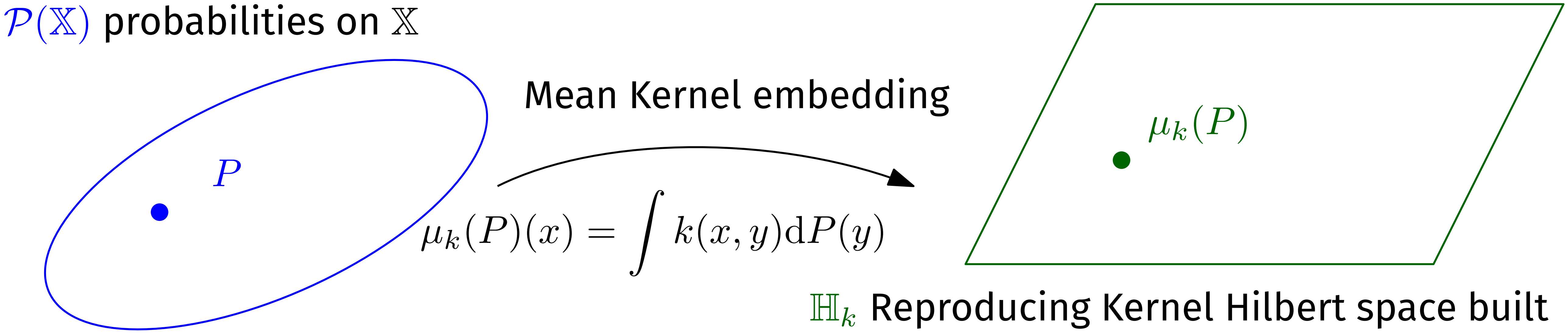
Define:

$$\mathrm{Cov}_k(\tilde{P}_1, \tilde{P}_2) = \mathbb{E}\left[\iint k(x,y)\mathrm{d}\tilde{P}_1(x)\mathrm{d}\tilde{P}_2(y)\right] - \iint k(x,y)\mathrm{d}\mathbb{E}[\tilde{P}_1](x)\mathrm{d}\mathbb{E}[\tilde{P}_2](y),$$

$$\mathrm{Var}_k(\tilde{P}) = \mathrm{Cov}_k(\tilde{P}, \tilde{P}).$$

The index: $$I = \mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \frac{\mathrm{Cov}_k(\tilde{P}_1, \tilde{P}_2)}{\sqrt{\mathrm{Var}_k(\tilde{P}_1)}\sqrt{\mathrm{Var}_k(\tilde{P}_2)}} \in [-1, 1].$$

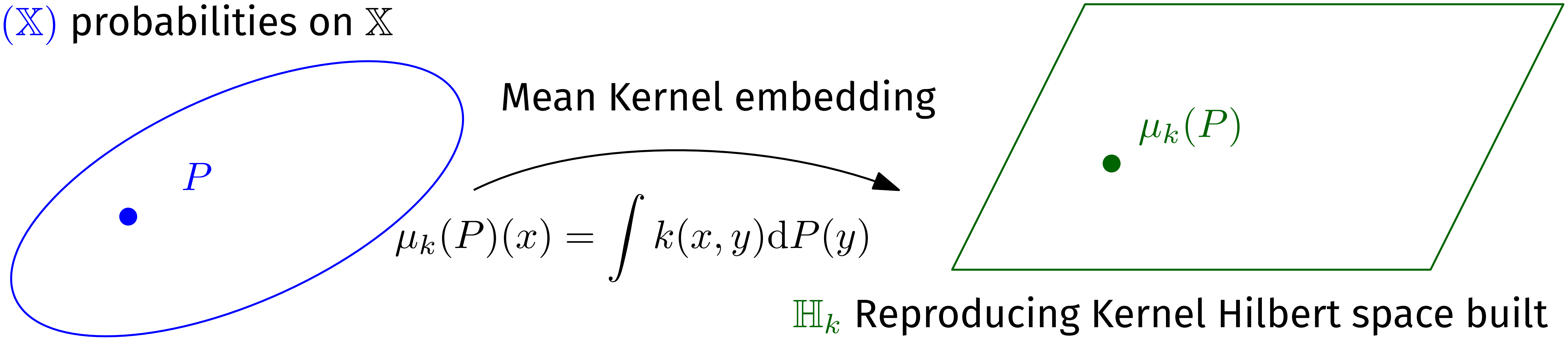
Interpretation: correlation in a Hilbert space



$\mathbb{H}_k$ Reproducing Kernel Hilbert space built on kernel k, subspace of functions $\mathbb{X} \to \mathbb{R}$.

[Berlinet &Thomas-Agnan, 2004]

Interpretation: correlation in a Hilbert space



$\mathbb{H}_k$ Reproducing Kernel Hilbert space built on kernel k, subspace of functions $\mathbb{X} \to \mathbb{R}$.

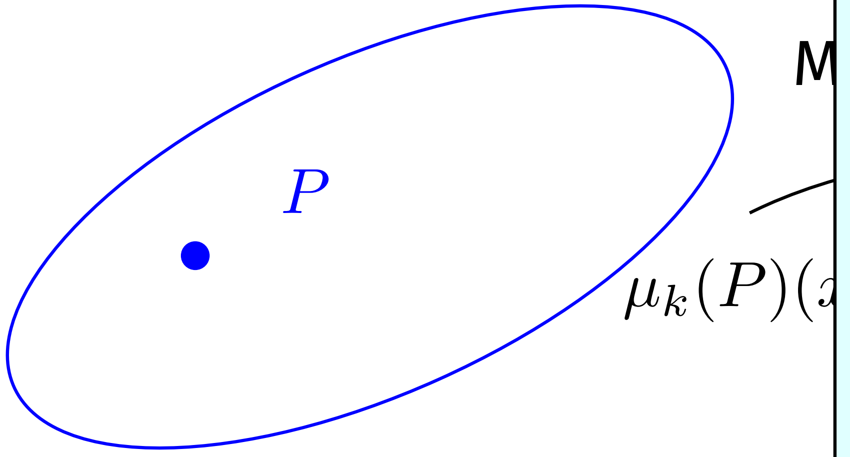
Rewriting:

$$\mathrm{Cov}_k(\tilde{P}_1, \tilde{P}_2) = \mathbb{E}\left[\langle \mu_k(\tilde{P}_1) - \mathbb{E}[\mu_k(\tilde{P}_1)], \mu_k(\tilde{P}_2) - \mathbb{E}[\mu_k(\tilde{P}_2)]\rangle_{\mathbb{H}_k}\right].$$

[Berlinet &Thomas-Agnan, 2004]

Interpretation: correlation in a Hilbert space

$\mathcal{P}(\mathbb{X})$ probabilities on $\mathbb{X}$



Thursday 14:30 – 16:00
Session 15A with **Francesco Mascari** for other choice of embeddings and summaries of the cross-covariance matrix.

Rewriting:

$$\mathrm{Cov}_k(\tilde{P}_1, \tilde{P}_2) = \mathbb{E}\left[\langle \mu_k(\tilde{P}_1) - \mathbb{E}[\mu_k(\tilde{P}_1)], \mu_k(\tilde{P}_2) - \mathbb{E}[\mu_k(\tilde{P}_2)]\rangle_{\mathbb{H}_k}\right].$$

[Berlinet &Thomas-Agnan, 2004]

Estimation from samples

Only law of **two samples** per group: $(X_{1,1}^{(t)}, X_{1,2}^{(t)}, X_{2,1}^{(t)}, X_{2,2}^{(t)})$ for $t = 1, \ldots, M$ i.i.d. samples with law

$$(\tilde{P}_1, \tilde{P}_2)$$

$$(X_{1,1}, X_{1,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_1 \qquad (X_{2,1}, X_{2,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_2$$

Estimation from samples

Only law of **two samples** per group:
$(X_{1,1}^{(t)}, X_{1,2}^{(t)}, X_{2,1}^{(t)}, X_{2,2}^{(t)})$ for
$t = 1, \dots, M$ i.i.d. samples with law

$$(\tilde{P}_1, \tilde{P}_2)$$

$$(X_{1,1}, X_{1,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_1 \qquad (X_{2,1}, X_{2,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_2$$

Unbiased & asymptotically normal estimators:

$$\widehat{\mathrm{Cov}}_k(\tilde{P}_1, \tilde{P}_2) = \frac{1}{M-1}\sum_{t=1}^{M} k\Big(X_{1,1}^{(t)}, X_{2,1}^{(t)}\Big) - \frac{1}{(M-1)M}\sum_{t=1}^{M}\sum_{s=1}^{M} k\Big(X_{1,1}^{(t)}, X_{2,1}^{(s)}\Big)$$

$$\widehat{\mathrm{Var}}_k(\tilde{P}_i) = \frac{1}{M-1}\sum_{t=1}^{M} k\Big(X_{i,1}^{(t)}, X_{i,2}^{(t)}\Big) - \frac{1}{(M-1)M}\sum_{t=1}^{M}\sum_{s=1}^{M} k\Big(X_{i,1}^{(t)}, X_{i,2}^{(s)}\Big)$$

Estimation from samples

Only law of **two samples** per group: $(X_{1,1}^{(t)}, X_{1,2}^{(t)}, X_{2,1}^{(t)}, X_{2,2}^{(t)})$ for $t = 1, \ldots, M$ i.i.d. samples with law

$$(\tilde{P}_1, \tilde{P}_2)$$

$$(X_{1,1}, X_{1,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_1 \qquad (X_{2,1}, X_{2,2}) \overset{\text{i.i.d.}}{\sim} \tilde{P}_2$$

Intergroup

Unbiased & asymptotically normal estimators:

$$\widehat{\mathrm{Cov}}_k(\tilde{P}_1, \tilde{P}_2) = \frac{1}{M-1} \sum_{t=1}^{M} k\left(X_{1,1}^{(t)}, X_{2,1}^{(t)}\right) - \frac{1}{(M-1)M} \sum_{t=1}^{M} \sum_{s=1}^{M} k\left(X_{1,1}^{(t)}, X_{2,1}^{(s)}\right)$$

$$\widehat{\mathrm{Var}}_k(\tilde{P}_i) = \frac{1}{M-1} \sum_{t=1}^{M} k\left(X_{i,1}^{(t)}, X_{i,2}^{(t)}\right) - \frac{1}{(M-1)M} \sum_{t=1}^{M} \sum_{s=1}^{M} k\left(X_{i,1}^{(t)}, X_{i,2}^{(s)}\right)$$

Intragroup

Kernel correlation generalizes setwise correlation

If $A \subseteq \mathbb{X}$ and $k(x,y) = 1_{(x,y)\in A^2}$ then

$$\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)).$$

But for this kernel the mean kernel embedding is **not** injective, contrary to the Gaussian kernel.

Kernel correlation generalizes setwise correlation

If $A \subseteq \mathbb{X}$ and $k(x, y) = 1_{(x,y) \in A^2}$ then

$$\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \mathrm{Corr}(\tilde{P}_1(A), \tilde{P}_2(A)).$$

But for this kernel the mean kernel embedding is **not** injective, contrary to the Gaussian kernel.

Common atoms, dependent weights

Result. If $(\tilde{P}_1, \tilde{P}_2)$ is a **multivariate species sampling process** (mSSP) then the setwise correlation does not depend on A, and the kernel correlation does not depend on k.

First part proved in [Franzolini et al, 2025]

mSSP:

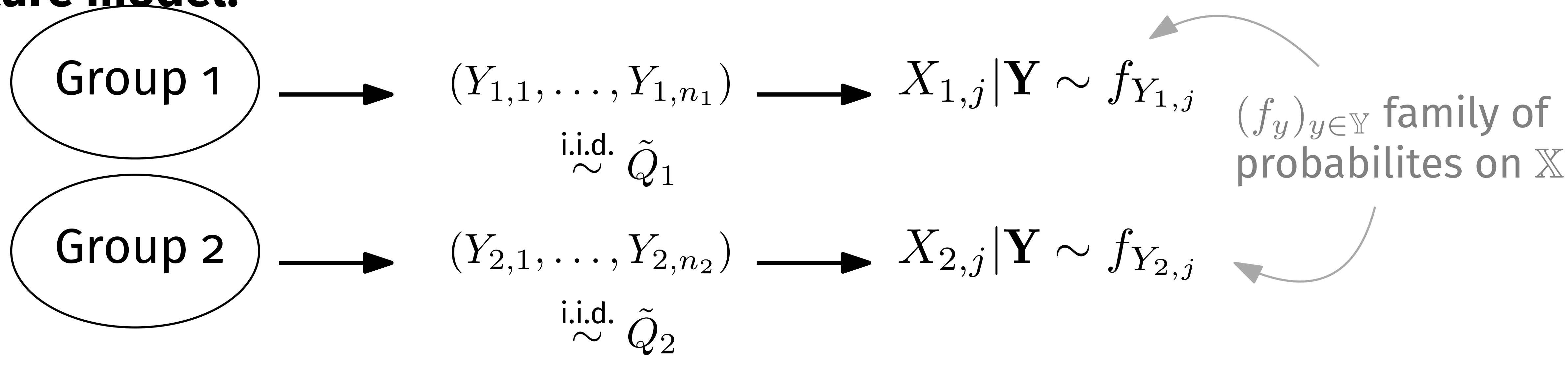
$$\tilde{P}_i = \sum_{k=1}^{+\infty} s_{i,k} \delta_{x_k}$$

$$(s_{1,k}, s_{2,k})_{k \geq 1} \perp (x_k)_{k \geq 1}$$

$$(x_k)_{k \geq 1} \overset{\text{i.i.d.}}{\sim} P_0$$

Behavior under mixtures

Mixture model:



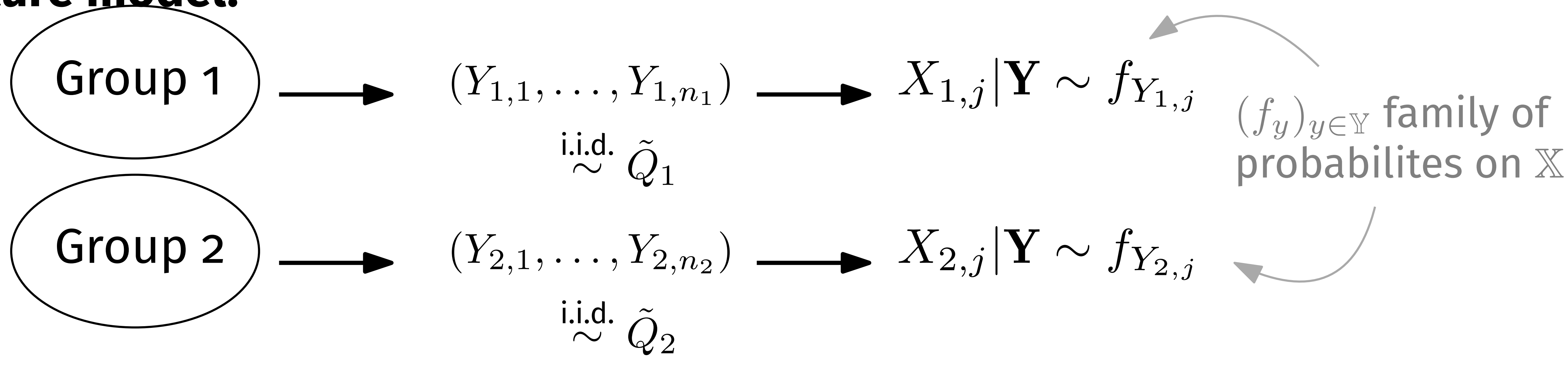
Equivalently:

$$(\tilde{P}_1, \tilde{P}_2) = \iint (f_{y_1}, f_{y_2}) \, \mathrm{d}\tilde{Q}_1(y_1) \mathrm{d}\tilde{Q}_1(y_2).$$

Parametric model with $y \leftrightarrow \theta$: we choose $(\tilde{Q}_1, \tilde{Q}_2) = (\delta_{y_1}, \delta_{y_2})$ with $(y_1, y_2) \sim \pi$.

Behavior under mixtures

Mixture model:



Equivalently:

$$(\tilde{P}_1, \tilde{P}_2) = \iint (f_{y_1}, f_{y_2})\, \mathrm{d}\tilde{Q}_1(y_1)\mathrm{d}\tilde{Q}_1(y_2).$$

Parametric model with $y \leftrightarrow \theta$: we choose $(\tilde{Q}_1, \tilde{Q}_2) = (\delta_{y_1}, \delta_{y_2})$ with $(y_1, y_2) \sim \pi$.

Theorem: $\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \mathrm{Corr}_{k_f}(\tilde{Q}_1, \tilde{Q}_2)$ with

$$k_f(y_1, y_2) = \iint_{\mathbb{X} \times \mathbb{X}} k(x_1, x_2) f_{y_1}(\mathrm{d}x_1) f_{y_1}(\mathrm{d}x_2).$$

Detecting exchangeability

Easy results:

If $\tilde{P}_1 \perp \tilde{P}_2$ then $I = 0$.

If $\tilde{P}_1 = \tilde{P}_2$ a.s. then $I = 1$.

Detecting exchangeability

Easy results:

If $\tilde{P}_1 \perp \tilde{P}_2$ then $I = 0$.

If $\tilde{P}_1 = \tilde{P}_2$ a.s. then $I = 1$.

Weaker conditions: the fixed atoms of $\tilde{P}_1$, $\tilde{P}_2$ can have arbitrary small mass.

Conversely, assume $\tilde{P}_1, \tilde{P}_2$ are a.s. discrete and $\mathbb{E}[\tilde{P}_1]$, $\mathbb{E}[\tilde{P}_2]$ atomless. If $\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = 1$ for a c_0-universal kernel k then

$$\tilde{P}_1 = \tilde{P}_2 \text{ a.s.}$$

μ_k is injective on signed measures: e.g. Gaussian, Laplace, but not setwise.

The hierarchical Dirichlet process

Model:

$$\tilde{P}_0 \sim \mathrm{DP}(c_0, P_0),$$

$$\tilde{P}_1, \tilde{P}_2 | \tilde{P}_0 \overset{\text{i.i.d.}}{\sim} \mathrm{DP}(c, \tilde{P}_0).$$

$\tilde{P} \sim \mathrm{DP}(c, P)$ if $\tilde{P} = \sum_{k=1}^{+\infty} s_k \delta_{x_k}$

$s_k = v_k \prod_{j=1}^{k-1}(1 - v_j)$ with $v_j \overset{\text{i.i.d.}}{\sim} \mathrm{Beta}(1, c)$.

$(x_k)_{k \geq 1} \overset{\text{i.i.d.}}{\sim} P_0$

[Teh et al, 2006], [Camerlenghi et al, 2019]

The hierarchical Dirichlet process

Model: $\tilde{P}_0 \sim \mathrm{DP}(c_0, P_0),$

$$\tilde{P}_1, \tilde{P}_2 | \tilde{P}_0 \overset{\text{i.i.d.}}{\sim} \mathrm{DP}(c, \tilde{P}_0).$$

A priori:

$$\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \frac{1+c}{1+c+c_0}.$$

$\tilde{P} \sim \mathrm{DP}(c, P)$ if $\tilde{P} = \sum_{k=1}^{+\infty} s_k \delta_{x_k}$

$s_k = v_k \prod_{j=1}^{k-1}(1-v_j)$ with $v_j \overset{\text{i.i.d.}}{\sim} \mathrm{Beta}(1, c).$

$(x_k)_{k \geq 1} \overset{\text{i.i.d.}}{\sim} P_0$

[Teh et al, 2006], [Camerlenghi et al, 2019]

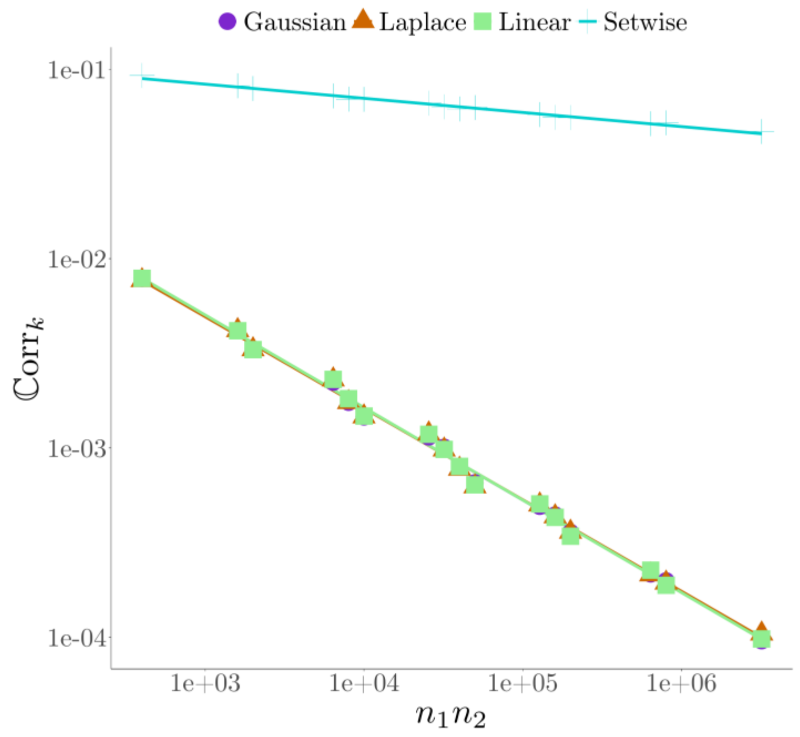
The hierarchical Dirichlet process

Model:
$$\tilde{P}_0 \sim \mathrm{DP}(c_0, P_0),$$
$$\tilde{P}_1, \tilde{P}_2 | \tilde{P}_0 \overset{\text{i.i.d.}}{\sim} \mathrm{DP}(c, \tilde{P}_0).$$

A priori:
$$\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2) = \frac{1+c}{1+c+c_0}.$$

$(x_j)_{j\geq 1}$ non degenerate if $\mu_k(\hat{P}_n)$ with $\hat{P}_n = n^{-1}\sum_{j=1}^{n} \delta_{x_j}$ does not converge to the image of a Dirac mass in $\mathbb{H}_k$.

A posteriori: for a sequence of data $\mathbf{X}_{n_1,n_2}$ **non degenerate** with respect to the kernel k,
$$\mathrm{Corr}_k(\tilde{P}_1, \tilde{P}_2 | \mathbf{X}_{n_1,n_2}) = \mathcal{O}\left(\frac{1}{\sqrt{n_1 n_2}}\right).$$



Thank you for your attention

The 5th Italian Meeting on Probability and Mathematical Statistics