Constructing & disintegrating p-ty measures. Couplings & transports

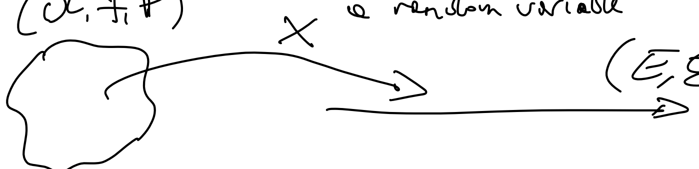
A generic p-ty space will be denoted $(\Omega, \mathcal{F}, \mathbb{P})$. Here $\mathcal{F}$ is a σ-algebra and is not completed, unless specified.

Most of the time, I will work with complete separable metric spaces (Polish spaces). There will always be endowed with their Borel σ-algebra & typically denoted $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ etc. Note that $\mathcal{B}(\mathcal{X})$ is countably generated & (taking finite intersections of complements of the generating open sets $= \mathcal{C}$) countably determined (i.e. μ, ν on $\mathcal{B}(\mathcal{X})$ then $\mu = \nu$ iff $\mu = \nu$ on $\mathcal{C}$).

It follows that $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\mathcal{Y}) = \mathcal{B}(\mathcal{X} \times \mathcal{Y})$.

Pushforward (image) measure

$(\Omega, \mathcal{F}, \mathbb{P})$ $\qquad X$ a random variable $\qquad$ i.e., $X^{-1}(A) \in \mathcal{F} \ \forall_{A \in \mathcal{E}}$



$(E, \mathcal{E})$ measurable space

then X naturally pushes $\mathbb{P}$ onto a p-ty measure on E often called the distribution of X, denoted $\mathcal{L}(X) = \mathbb{P} \circ X^{-1} =: X_{\#}\mathbb{P}$. $\qquad \mathcal{L}(X)(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$, $A \in \mathcal{E}$.

Marginals are the pushforward measures of projections on a product space.

Say $(\mathcal{X} \times \mathcal{Y}, \pi)$ is a p-ty space. Let $\mathrm{proj}_x : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$, $(x,y) \to x$ be the projection on the first coordinate (& likewise proj_y). Then

$\mu := \mathrm{proj}_{x\#}\pi$ is called the first marginal of π & $\nu := \mathrm{proj}_{y\#}\pi$ is the 2nd one.

This is equivalent to saying

(i) $\mu(A) = \pi(A \times \mathcal{Y})$ $\quad \forall \ A \in \mathcal{B}(\mathcal{X})$

$\nu(B) = \pi(\mathcal{X} \times B)$ $\quad B \in \mathcal{B}(\mathcal{Y})$

(ii) $$\int_{\mathcal{X}\times\mathcal{Y}} (\varphi(x)+\psi(y))\,d\pi(dx,dy) = \int_{\mathcal{X}} \varphi\,d\mu + \int_{\mathcal{Y}} \psi\,d\nu$$ $\forall$ measurable ≥ 0 functions φ on $\mathcal{X}$ & ψ on $\mathcal{Y}$

|| this really is

$$\int_{\mathcal{X}\times\mathcal{Y}} (\tilde{\varphi}+\tilde{\psi})\,d\pi, \text{ for } \tilde{\varphi}(x,y)=\varphi(x),\ \tilde{\psi}(x,y)=\psi(y)$$

Fubini - constructing measures

Def. A p-ty kernel on $\mathcal{X}\times\mathcal{Y}$ is a mapping $\theta : \mathcal{X}\times\mathcal{B}(\mathcal{Y}) \to [0,1]$ such that

- for each $x\in\mathcal{X}$, $\theta(x,\cdot)$ is a p-ty measure on $\mathcal{B}(\mathcal{Y})$
- for each $B\in\mathcal{B}(\mathcal{Y})$, $\theta(\cdot,B)$ is measurable

Thm (Fubini) Let $(\Omega,\mathcal{F}) = (\mathcal{X}\times\mathcal{Y}, \mathcal{B}(\mathcal{X})\times\mathcal{B}(\mathcal{Y}))$, μ be a p-ty measure on $\mathcal{X}$ & θ a p-ty kernel. Then there exists a unique p-ty measure π on $(\Omega,\mathcal{F})$ s.t.

$$\pi(A\times B) = \int_A \mu(dx)\,\theta(x,B), \quad \forall A\in\mathcal{B}(\mathcal{X}),\ B\in\mathcal{B}(\mathcal{Y})$$

For a non-negative r.v. f on Ω, the function $x\mapsto \int_{\mathcal{Y}} f(x,y)\,\theta(x,dy)$ is measurable &

$$\int f\,d\pi = \int_{\mathcal{X}} \mu(dx) \int_{\mathcal{Y}} f(x,y)\,\theta(x,dy)$$

Rk Note that $\pi(A\times\mathcal{Y}) = \int_A \mu(dx)\,\underbrace{\theta(x,\mathcal{Y})}_{=1} = \mu(A)$ so $\mathrm{proj}_{\mathcal{X}\#}\pi = \mu$.

On the other hand, $\nu := \mathrm{proj}_{\mathcal{Y}\#}\pi$ is given by $\nu(B) = \pi(\mathcal{X}\times B) = \int_{\mathcal{X}} \mu(dx)\,\theta(x,B)$.

Rk A special case is when $\theta(x,\cdot) = \nu(\cdot)$ is a fixed p-ty measure on $\mathcal{Y}$. π is then called the product measure & denoted $\mu\otimes\nu$.

Couplings & Transport

Def (Coupling) Let $(\mathcal{X},\mu), (\mathcal{Y},\nu)$ be two p-ty measures. A coupling of μ, ν is a p-ty space $(\Omega,\mathcal{F},\mathbb{P})$ with two r.variables $X:\Omega\to\mathcal{X}$ & $Y:\Omega\to\mathcal{Y}$ s.t. $X_\#\mathbb{P} = \mu$ & $Y_\#\mathbb{P} = \nu$.

Rk We sometimes refer to $\pi = (X,Y)_\#\mathbb{P}$ (which is a p-ty measure on $\mathcal{X}\times\mathcal{Y}$) as the coupling.

The set of such measures is denoted $\Pi(\mu,\nu) = \{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}) : \mathrm{proj}_{\mathcal{X}\#}\pi = \mu,\ \mathrm{proj}_{\mathcal{Y}\#}\pi = \nu\}$

Rk If no further requirements are given, $(\mathcal{X}\times\mathcal{Y}, \mu\otimes\nu)$ with $X=\mathrm{proj}_{\mathcal{X}}$, $Y=\mathrm{proj}_{\mathcal{Y}}$ gives the independent coupling. So couplings always exists, i.e. $\Pi(\mu,\nu)\neq\emptyset$.

Def A coupling (X,Y) is said to be deterministic (or Monge) if $Y=T(X)$ for some measurable function $T:\mathcal{X}\to\mathcal{Y}$.

This is equivalent to saying that $Y=T(X)$ and $T_\#\mu=\nu$.

Or that the distribution of (X,Y) is concentrated on the graph of a measurable function T

Or that T provides a change of variables:
$$\int_{\mathcal{Y}}\varphi(y)\,\nu(dy)=\int_{\mathcal{X}}\varphi(T(x))\,\mu(dx)\quad\forall\,\varphi\geq 0$$

Or that $\mathcal{L}((X,Y))=\pi=(\mathrm{Id},T)_\#\mu$.

Such a map T is called a (Monge) transport.

Rk Given $\pi=\mathcal{L}((X,Y))$, such T above is unique μ-a.s. While $\Pi(\mu,\nu)\neq\emptyset$, a transport map may fail to exist (e.g. $\mu=\delta_{\{0\}}$, $\nu=\frac{1}{2}(\delta_{-1}+\delta_1)$).

Examples: (The increasing rearrangement on $\mathbb{R}$)

For a p-ty measure μ on $\mathbb{R}$, let $F_\mu(x)=\mu((-\infty,x])$ & F_μ^{-1} be its right-cont. inverse $F_\mu^{-1}(t)=\inf\{x\in\mathbb{R}: F_\mu(x)>t\}$. If $X\sim\mathrm{Unif}[0,1]$ then $Y=F_\mu^{-1}(X)\sim\mu$

i.e. $(\mathrm{Id},F_\mu^{-1})_\#\mathrm{Leb}|_{[0,1]}\in\Pi(\mathrm{Unif}[0,1],\mu)$ & F_μ^{-1} is a transport map.

If μ has no atoms then the reverse works: $X\sim\mu$ gives $Y=F_\mu(X)\sim\mathrm{Unif}[0,1]$ & hence
$$\left(X,\,F_\nu^{-1}(F_\mu(X))\right)\text{ is a coupling of }(\mu,\nu).$$

- <u>Optimal coupling/transport</u>

Consider a cost function $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ & look for
$$\inf_{T:\,T_\#\mu=\nu}\int_{\mathcal{X}}c(x,T(x))\,\mu(dx)\quad\text{or}\quad\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{Y}}c(x,y)\,d\pi=:\mathcal{P}(\mu,\nu)=\mathcal{P}_c(\mu,\nu).$$

transport maps / plans

e.g. $\|\mu-\nu\|_{TV}=\sup_{A\in\mathcal{B}(\mathcal{X})}|\mu(A)-\nu(A)|=\inf_{\pi\in\Pi(\mu,\nu)}\int 1\!\!1_{x\neq y}\,d\pi$ [proof to follow when we work on duality]

Lemma If μ is atomless then deterministic (transport) couplings are dense in all couplings.

Proof $\to$ Suppose X & Y are compact. Divide them into small sets $K_{i,n}$, $L_{i,n}$
$\mathrm{diam}(K_{i,n}) \leq \frac{1}{2n} \Rightarrow K_{i,n} \times L_{j,n}$ are small ($\leq 1/n$) & partition $X \times Y$.

If $\pi_n(K_{i,n} \times L_{j,n}) = \pi(K_{i,n} \times L_{j,n})$ $\forall_{i,j}$ $\forall_n$ for some $\pi \in \mathcal{P}(X \times Y)$

then $\pi_n \xrightarrow{w} \pi$ since $\left| \int f \, d\pi_n - \int f \, d\pi \right| \leq \sum_{i,j} \left| \int f \, (d\pi_n - d\pi) \right| \leq \omega_f(1/n)$
$\to 0$

where ω_f is the modulus of cont. of f which is our cont. bdd test function.

Then given $\pi \in \Pi(\mu, \nu)$ consider $\pi_i := \pi \big|_{K_{i,n} \times Y} \cdot \frac{1}{\mu(K_{i,n})}$

& $T_i^n = F^{-1}_{\nu_i^n} \circ F_{\mu_i}$, (μ_i^n, ν_i^n) marginals of π_i

Then $\tilde{\pi}_i^n := (\mathrm{Id}, T_i^n)_{\#} \mu_i$ is a Monge coupling which agrees with π on each $K_{i,n} \times L_{j,n}$, i.e..

& by putting $\tilde{\pi}_i^n$ together (over i) we get a Monge couplings

$\tilde{\pi}_n \longrightarrow \pi$.

$\to$ General Polish X & Y, see A. Pratelli PhD '03.

Disecting measures (disintegration)

We saw above how to use μ & a kernel Θ to obtain a p-ty measure on $\mathcal{X}\times\mathcal{Y}$. We now want to reverse this procedure. This is called disintegration in analysis & regular conditioning in p-ty.

Consider $(\Omega,\mathcal{F},\mathbb{P})$ & a sub-σ-algebra $\mathcal{G}$. We know that conditional expectations exist so that $\forall_{A\in\mathcal{F}}$ $\mathbb{E}[\mathbb{1}_A \mid \mathcal{G}]$ is a $\mathcal{G}$-measurable r.v. It is defined $\mathbb{P}$-a.s., i.e. outside of some $\mathbb{P}$-null set... which may depend on A. As $A\in\mathcal{F}$ varies I can easily be left with no $\omega\in\Omega$ on which my objects are jointly identified.

Def $\Theta:\Omega\times\mathcal{F}\to[0,1]$ is called a regular conditional p-ty for $\mathcal{F}$ given $\mathcal{G}$, if

- $\forall_{A\in\mathcal{F}}$, $\Theta(\cdot,A):\Omega\to[0,1]$ is $\mathcal{G}$-measurable; $\circ$ $\forall_{\omega\in\Omega}$ $\Theta(\omega,\cdot)$ is a p-ty measure on $(\Omega,\mathcal{F})$
- $\forall_{A\in\mathcal{F}}$, $\Theta(\cdot,A)=\mathbb{E}[\mathbb{1}_A\mid\mathcal{G}]$ $\mathbb{P}$-a.s.

Thm If Ω is a complete separable metric space & $\mathcal{F}=\mathcal{B}(\Omega)$, then a regular cond. p-ty for $\mathcal{F}$ given $\mathcal{G}$ exists & is unique. Furthermore, if $\mathcal{H}\subseteq\mathcal{G}$ is a countably determined σ-algebra then $\exists\, N\in\mathcal{G}$, $\mathbb{P}(N)=0$,

$$\Theta(\omega,A)=\mathbb{1}_A(\omega)\quad,\ \forall_{A\in\mathcal{H}}\ ,\ \forall_{\omega\in\Omega\setminus N}.$$

$\Rightarrow$ If X is a $\mathcal{G}$-m. r.v. taking values in $(\mathcal{X},\mathcal{B}(\mathcal{X}))$, $\mathcal{H}=\sigma(X)$ then

$$\Theta(\omega,\{\omega'\in\Omega: X(\omega)=X(\omega')\})=1\quad \mathbb{P}\text{-a.s.}$$

Rk This is often stated for $\mathcal{G}=\sigma(\xi)$ & then "$\Theta(x,A)=\mathbb{E}[\mathbb{1}_A\mid\xi=x]$" & $\Theta: S\times\mathcal{F}\to[0,1]$ for $\xi:\Omega\to S$.

/See KS (chp 5) & Parthasarathy ('67).

Let us apply this to the particular case of $(\Omega,\mathcal{F})=(\mathcal{X}\times\mathcal{Y},\mathcal{B}(\mathcal{X})\otimes\mathcal{B}(\mathcal{Y}))$. Let $\pi=\mathbb{P}$ & $\mathcal{G}=\mathcal{H}=\sigma(\mathrm{proj}_{\mathcal{X}})$. Note that $\Theta(\cdot,A)$ being $\mathcal{G}$-measurable means it is a function composition $\Theta(\cdot,A)=f(\mathrm{proj}_{\mathcal{X}}(\omega))$, i.e., is $\mathcal{B}(\mathcal{X})$-measurable.

By restricting $\Theta(x,\cdot)$ to $\mathcal{B}(\mathcal{Y})$ we obtain a p-ty kernel (still denoted Θ) s.t.

$$\Theta(x,B)=\mathbb{E}_\pi[\mathbb{1}_B\mid\mathcal{G}](x)\quad d\pi\text{-}x \text{ a.e.}\quad,\text{or}$$

$$\forall_{A\in\mathcal{G}}\quad \int_{A\times\Omega}\Theta(x,B)\,d\pi=\int_{A\times\Omega}\mathbb{1}_B\,d\pi=\int_{A\times B}d\pi$$

$$\int_A\Theta(x,B)\,d\mu \quad \text{for } \mu:=\mathrm{proj}_{\mathcal{X}}\#\pi$$

$\mu: \mathcal{B}(\mathcal{X}) \to [0,1]$ & $\Theta: \mathcal{X} \times \mathcal{B}(\mathcal{Y}) \mapsto [0,1]$ uniquely determine π on $\mathcal{B}(\mathcal{X}\times\mathcal{Y})$.

We may write $\pi = \mu \otimes \Theta$ & say this is the disintegration of π along its first marginal.

Example. (Martingale couplings) We say that a coupling (X,Y) of (μ,ν), p-ties on $\mathbb{R}$, is a martingale coupling if $\mathbb{E}[Y \mid \sigma(X)] = X$ a.s. & we write $\mathcal{L}((X,Y)) \in \mathcal{M}(\mu,\nu)$.

Let $\pi = \mathcal{L}((X,Y))$. Note that $\mathbb{E}[Y \mid \sigma(X)](x) = \int_{\mathcal{Y}} y\,\Theta(x,dy)$ $\mu(dx)$-a.e.

Indeed, by measurability of Θ, $x \mapsto \int y\,\Theta(x,dy)$ is Borel-measurable & $\forall A \in \mathcal{B}(\mathbb{R})$

$$\mathbb{E}[\mathbb{1}_{X\in A}\, Y] = \iint \mathbb{1}_A(x)\, y \, d\pi = \int_A \mu(dx) \int_{\mathbb{R}} y\,\Theta(x,dy) \quad \text{as required.}$$

So the coupling is a martingale one iff $\int_{\mathbb{R}} y\,\Theta(x,dy) = x$ $\mu(dx)$-a.e., i.e. the kernel Θ is barycentre-preserving.

Example (Knothe-Rosenblatt rearrangement in $\mathbb{R}^n$). Let $\mu,\nu \in \mathcal{P}(\mathbb{R}^n)$ with $\mu \ll Leb$.

1) Let $\mu_1 = proj_1 \# \mu$ & $y_1 = T_1(x_1)$ for $T_1 = F_{\nu_1}^{-1} \circ F_{\mu_1}$. This gives a coupling of (μ_1,ν_1).

2) Let $\mu_{12} = proj_{12} \# \mu$ & $\mu_{12} = \mu_1 \otimes \Theta_1$, $\nu_{12} = \nu_1 \otimes \eta_1$. Let $y_2 = T_2(x_2; x_1)$ where $T_2 = F^{-1}_{\eta_1(y_1,\cdot)} \circ F_{\Theta_1(x_1,\cdot)}$

i.e. for x_1 (& y_1) fixed, we transport the cond distribution $\mathcal{L}(x_2 \mid x_1) \rightsquigarrow \mathcal{L}(y_2 \mid y_1)$.

3) Let $\mu_{13} = \mu_{12} \otimes \Theta_2$, $\nu_{13} = \nu_{12} \otimes \eta_2$ ... etc. $\rightsquigarrow$ map T.

Note that Jacobian matrix of this change of variables T is upper triangular with positive entries on the diagonal.

Example (Gluing) Let $(\mathcal{X}_i, \mu_i)$ $i=1,2,3$ be Polish p-ty spaces & $\pi_{12} \in \Pi(\mu_1,\mu_2)$ & $\pi_{23} \in \Pi(\mu_2,\mu_3)$.

Then there exists $\pi \in \Pi(\mu_1,\mu_2,\mu_3)$ with $proj_{12} \# \pi = \pi_{12}$ & $proj_{23} \# \pi = \pi_{23}$.

Proof (sketch). Disintegrate $\pi_{12} = \Theta_{12} \otimes \mu_2$, $\pi_{23} = \mu_2 \otimes \Theta_{23}$ & glue along ~~those~~ common marginal:

$$\pi(dx_1, dx_2, dx_3) = \Theta_{12}(x_2, dx_1)\, \mu_2(dx_2)\, \Theta_{23}(x_2, dx_3).$$

$\boxtimes$

Lemma $\pi \in \Pi(\mu, \nu) \iff \exists\, T : \mathcal{X} \times [0,1] \to \mathcal{Y}$ s.t.

$(X, T(X,U)) \sim \pi$ & $T(X,U) \sim \nu$

where $(X,U) \sim \mu \otimes \mathrm{Unif}[0,1]$.

"couplings $\equiv$ randomisations of maps"

Remarks on the literature / sources:

As advertised, these notes follow & borrow from

- Villani '03 "Topics in Optimal Transportation"
- Villani '09 "Optimal transport. Old and New."
- Santambrogio '15 "Optimal Transport for Applied Mathematicians"

these are all wonderful books!