A crash course in Dirichlet processes
Part 2

Jason Swanson
UCF Probability Seminar
Feb 9, 2021

Loose ends

Notation:

(deterministic) measure

- If $\mu = \mathcal{L}(X)$, then

$$\mathcal{L}(X; A) := \mu(A).$$

random measure

- If $\mu = \mathcal{L}(X \mid \mathcal{G})$, then

$$\mathcal{L}(X \mid \mathcal{G}; A) = \mu(A)$$

- $X \sim \mu$ means $\mu = \mathcal{L}(X)$
- $X \mid \mathcal{G} \sim \mu$ means $\mu = \mathcal{L}(X \mid \mathcal{G})$

"Bayesian/frequentist combo"

S: separable metric sp., $\mathcal{S} = \mathcal{B}_S$

$X = \{X_n : n \in \mathbb{N}\}$: S-val. stoch. process

$$\boxed{\begin{array}{l}\textbf{Thm } (\text{De Finetti}^{*} + \text{Glivenko-Cantelli-Varadarajan}^{**}) \\ \text{If } X \text{ is exchangeable, then, with probability one,} \\ \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j} \text{ converges weakly to a random measure } \mu. \\ \text{Moreover, } \mathcal{L}(X|\mu) = \mu^{\infty}.\end{array}}$$

* Thm. 9.16, Foundations of Modern Probability, Kallenberg (1997)

** Thm. 11.4.1, Real Analysis and Probability, Dudley (2004)

III. An asymmetric die

$S = \{0, 1, \dots, d\}$, $\quad \mathcal{S} = \mathcal{P}(S)$

$X = \{X_n : n \in \mathbb{N}\}$: S-val. stoch. proc.

Assume X is exchangeable.

Let μ be the unique random meas. on S such that $\mathcal{L}(X \mid \mu) = \mu^\infty$.

$\mu : \Omega \to M_1(S)$ is $(\mathcal{F}, \mathcal{M}_1(S))$-m'ble

$M_1(S)$ isomorphic to a d-simplex

SIMPLICES

$$\Delta^d = \{ t \in \mathbb{R}^{d+1} : t_j \geq 0, \underbrace{\textstyle\sum_0^d t_j}_{t = (t_0, t_1, \dots, t_d) \in \mathbb{R}^{d+1}} = 1 \}$$

standard d-simplex

σ : surface measure on Δ^d

<u>Notation:</u>
$$\int_{\Delta^d} f(t)\,dt := \int_{\Delta^d} f\,d\sigma$$

Calculating:

For $t = (t_0, t_1, \ldots, t_d)$ and $k \in \{0, 1, \ldots, d\}$,

let $\hat{t}_k = (t_0, \ldots, t_{k-1}, t_{k+1}, \ldots, t_d)$

$$\Delta^d_c = \{ t \in \mathbb{R}^d : t_j \geq 0, \sum_0^{d-1} t_j \leq 1 \}$$

corner d-simplex

$$\boxed{\begin{aligned} &\int_{\Delta^d} f(t)\, dt \\ &= \int_{\Delta^d_c} f\Big(t_0, \ldots, t_{k-1}, 1 - \sum_{j \neq k} t_j, t_{k+1}, \ldots, t_d\Big) \sqrt{1+d}\; d\hat{t}_k \end{aligned}}$$

Expl

$$\int_{\Delta^2} t_0^2 t_1^5 t_2^3 \, dt = \sqrt{3} \int_{\Delta_c^2} t_0^2 t_1^5 (1-t_0-t_1)^3 \, dt_1 \, dt_2$$

Δ_c^2 ← triangle in $\mathbb{R}^2$ with vertices $(0,0), (1,0), (0,1)$

$$= \sqrt{3} \int_0^1 \int_0^{1-t_1} t_0^2 t_1^5 (1-t_0-t_1)^3 \, dt_1 \, dt_2$$

$$= \sqrt{3} \int_0^1 \int_0^{1-t_0} t_0^2 (1-t_0-t_2)^5 t_2^3 \, dt_2 \, dt_0$$

$= \cdots$ etc.

$$\mu : \Omega \to M_1(S), \quad S = \{0, 1, \ldots, d\}$$

Define $\varphi : M_1(S) \longrightarrow \Delta^d$ by

$$\varphi(\nu) = (\nu(\{0\}), \ldots, \nu(\{d\}))$$

φ is a bijection; φ and φ^{-1} both m'ble

$\theta = (\theta_0, \theta_1, \ldots, \theta_d) := \varphi(\mu)$, a Δ^d-val. r.v.

Either μ or θ can be regarded as our "prior"

$$\sigma(\theta) = \sigma(\mu)$$

$\forall n \in \mathbb{N}, \ \forall k \in \{0, \dots, d\},$

$$P(X_n = k \mid \theta) = \theta_k$$

$\mathcal{L}(\theta \mid X_1, \dots, X_n) = ?$ Again, basic calculations ...

$$\mathcal{L}(\theta \mid X_1, \dots, X_n ; dt) \propto \underbrace{t_0^{N_0} t_1^{N_1} \cdots t_d^{N_d}}_{} \mathcal{L}(\theta ; dt),$$

where $N_k = |\{X_j : X_j = k\}|$

density of the <u>Dirichlet distribution</u>, a.k.a. the multivariate beta distribution

DIRICHLET DISTRIBUTION

$\alpha = (\alpha_0, \ldots, \alpha_d) \in (0,\infty)^{d+1}$

The Dirichlet distribution with parameter α, denoted by $\mathrm{Dir}(\alpha)$, is the probability measure on Δ^d proportional to

$$t_0^{\alpha_0 - 1} t_1^{\alpha_1 - 1} \cdots t_d^{\alpha_d - 1} \, dt$$

Relationship to beta distribution

If $\theta = (\theta_0, \ldots, \theta_d) \sim \mathrm{Dir}(\alpha_0, \ldots, \alpha_d)$.

and $k \in \{0, \ldots, d\}$, then $\theta_k \sim \mathrm{Beta}(\alpha_k, \sum_{j \neq k} \alpha_j)$.

$$\mathcal{L}(\theta) = \mathrm{Dir}(\alpha) \Rightarrow \mathcal{L}(\theta; dt) \propto t_0^{\alpha_0 - 1} \cdots t_d^{\alpha_d - 1}\, dt$$

$$\Rightarrow \mathcal{L}(\theta \mid X_1, \ldots, X_n; dt) \propto t_0^{N_0} t_1^{N_1} \cdots t_d^{N_d}\, \mathcal{L}(\theta; dt)$$

$$= t_0^{\alpha_0 + N_0 - 1} \cdots t_d^{\alpha_d + N_d - 1}\, dt$$

where $N_k = |\{X_j : X_j = k\}|$

$$\boxed{\begin{aligned} &\mathcal{L}(\theta) = \mathrm{Dir}(\alpha_0, \ldots, \alpha_d) \\ &\Rightarrow \mathcal{L}(\theta \mid X_1, \ldots, X_n) = \mathrm{Dir}(\alpha_0 + N_0, \ldots, \alpha_d + N_d) \end{aligned}}$$

IV Random object generator

$(S, \mathcal{S})$: standard Borel space

Imagine a machine that produces random objects from S.

$X = \{X_n : n \in \mathbb{N}\}$: S-val. stoch. proc.

Assume X is exchangeable.

Let μ be the unique random meas. on S such that $\mathcal{L}(X|\mu) = \mu^\infty$.

$\mu : \Omega \to M_1(S)$ is $(\mathcal{F}, \mathcal{M}_1(S))$-m'ble

What distribution to put on μ?

μ is a random prob. meas. on $(S, \mathcal{S})$

For each $A \in \mathcal{S}$, $\mu(A)$ is a r.v.

So $\{\mu(A) : A \in \mathcal{S}\}$ is a stoch. proc.

What finite-dimensional distributions do we want?

If $\{B_0, \ldots, B_d\}$ is a m'ble partition of S,

we want the random vector

$$(\mu(B_0), \ldots, \mu(B_d))$$

to have a Dirichlet distribution.

(But with what parameters?)

Definition

Let $(S, \mathcal{S})$ be a m'ble sp. and $\lambda = \{\lambda(A) : A \in \mathcal{S}\}$ a stoch. proc. indexed by $\mathcal{S}$. Let α be a nonzero, finite meas. on $(S, \mathcal{S})$. Then λ is a Dirichlet process on $(S, \mathcal{S})$ with parameter α if, whenever $\{B_0, \ldots, B_d\}$ is a m'ble partition of S, we have

$$\mathcal{L}(\lambda(B_0), \ldots, \lambda(B_d)) = \mathrm{Dir}(\alpha(B_0), \ldots, \alpha(B_d))$$

Think about observing only

$$\tilde{X}_n = \sum_{k=0}^{d} k 1_{B_k}(X_n)$$

First defined by Tom Ferguson (1973).

(i) Do Dirichlet processes exist? (yes)

- could maybe show with appropriate version of Kolmogorov extension theorem
- we will see the explicit "stick-breaking" construction

(ii) Does the parameter α uniquely determine the finite-dimensional distributions, $\{\mathcal{L}(\lambda(A_0),\ldots,\lambda(A_d)) : d \in \mathbb{N},\ A_j \in \mathcal{E}\}$? (yes)

- calculations w/ Dirichlet dist. See Ferguson

(iii) Is every Dirichlet process a random measure? (yes) really?

- will be seen from construction

Sethuraman stick-breaking construction

We will construct a random prob. meas. on $(S, \mathcal{S})$ of the form

$$\lambda = \sum_{k=1}^{\infty} R_k \delta_{U_k}$$

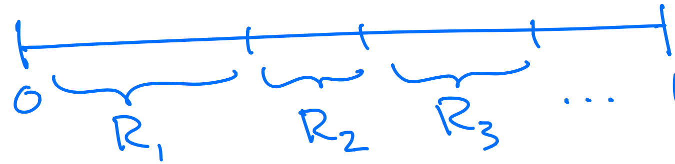
random points in S

random positive weights that sum to 1 (created by "breaking a stick" of length 1)

$U_1, U_2, \ldots$ iid

$\mathcal{L}(U_k) \propto \alpha$

Assume $\alpha(S) = 1$



$$\mathcal{L}(R_1) = \text{Unif}(0,1)$$

$$\mathcal{L}(R_2 \mid R_1) = \text{Unif}(0, 1-R_1)$$

$$\mathcal{L}(R_3 \mid R_1, R_2) = \text{Unif}(0, 1-R_1-R_2)$$

$$\vdots$$

$$\mathcal{L}(R_k \mid R_1, \ldots, R_{k-1}) = \text{Unif}\left(0, 1-\sum_1^{k-1} R_j\right)$$

$$\vdots$$

The R_k's are positive and sum to 1.

Construction w/o conditioning:

$V_1, V_2, \ldots$ iid Unif $(0,1)$

Recursive:

$R_1 = V_1$

$R_2 = (1-R_1)V_2$

$R_3 = (1-R_1-R_2)V_3$

$\vdots$

$R_k = (1-\sum_1^{k-1} R_j)V_k$

$\vdots$

Explicit:

$$\boxed{R_k = V_k \prod_{j=1}^{k-1} (1-V_k)}$$

For general $\alpha(S) \in (0,\infty)$, change Unif $(0,1)$ to Beta $(1, \alpha(S))$

$\lim_{\alpha(S)\to\infty} \lambda = ?$

$$\lambda = \sum_{k=1}^{\infty} R_k \delta_{U_k}$$

- For fixed ω, $\lambda(\omega)$ is a discrete prob. meas.
- For fixed A, $\lambda(A) = \sum_{k=1}^{\infty} R_k 1_A(U_k)$ is a random variable.
- So λ is a probability kernel from Ω to S, i.e. $\lambda : \Omega \to M_1(S)$ is m'ble (λ is a random measure)

Does every Dir. proc. have a mod. that is a random meas.?

$$\boxed{\{\lambda(A) : A \in \mathcal{S}\} \text{ is a Dirichlet process with parameter } \alpha}$$

λ : Dirichlet process on $(S, \mathcal{S})$ with parameter α ← a pos., finite meas. on S

λ : an $M_1(S)$-valued r.v.

$\mathcal{L}(\lambda)$: a prob. meas. on $M_1(S)$

$\mathcal{L}(\lambda) \in M_1(M_1(S))$

Notation: $\underline{\mathcal{D}(\alpha)} := \mathcal{L}(\lambda)$

distribution of a Dir. proc. w/param. α

For $A \in \mathcal{M}_1(S)$, $\mathcal{D}(\alpha; A) := \mathcal{L}(\lambda; A) = P(\lambda \in A)$

m'ble set of prob. measures

These are all synonymous:

- λ is a Dir. proc. w/ param. α
- $\mathcal{L}(\lambda) = \mathcal{D}(\alpha)$
- $\lambda \sim \mathcal{D}(\alpha)$

$\rho = \dfrac{\alpha}{\alpha(S)}$, $\kappa = \alpha(S)$

prob. meas. on S ("base" meas.)

pos. number ("stability")

$\mathcal{D}(\alpha) = \mathcal{D}(\kappa\rho)$

Can think of Dir. proc. as being determined by <u>two</u> parameters

Back to random object generator:

$X_1, X_2, \ldots$ exchangeable S-val. r.v.'s

$X = \{X_n : n \in \mathbb{N}\}$

$\mathcal{L}(X \mid \mu) = \mu^\infty$

Fix pos., finite meas. α on S

Assume $\mathcal{L}(\mu) = \mathcal{D}(\alpha)$

$\mathcal{L}(X_1) = ?$

$$P(X_1 \in A) = E[P(X_1 \in A \mid \mu)]$$

$$= E[\mu(A)]$$

$$\mathcal{L}(\mu(A), \mu(A^c)) = \mathrm{Dir}(\alpha(A), \alpha(A^c))$$

$$\mathcal{L}(\mu(A)) = \mathrm{Beta}(\alpha(A), \alpha(A^c))$$

$$E[\mu(A)] = \frac{\alpha(A)}{\alpha(A)+\alpha(A^c)} = \frac{\alpha(A)}{\alpha(S)} = \rho(A)$$

So $\mathcal{L}(X_1) = \rho$ ← base measure

$$\mathcal{L}(X_{n+1} \mid X_1, \ldots, X_n) = ?$$

$$\mathcal{L}(\mu \mid X_1, \ldots, X_n) = ?$$

$$\mathcal{L}(\mu) = \mathcal{D}(\alpha)$$

$$\Rightarrow \quad \mathcal{L}(\mu \mid X_1, \dots, X_n) = \mathcal{D}\left(\alpha + \sum_1^n \delta_{X_j}\right)$$

$$P(X_{n+1} \in A \mid X_1, \dots, X_n)$$

$$= E\left[P(X_{n+1} \in A \mid \mu, X_1, \dots, X_n) \mid X_1, \dots, X_n \right]$$

$$= E\left[P(X_{n+1} \in A \mid \mu) \mid X_1, \dots, X_n \right]$$

$$= E\left[\mu(A) \mid X_1, \dots, X_n \right]$$

$$= \frac{\left(\alpha + \sum_1^n \delta_{X_j}\right)(A)}{\left(\alpha + \sum_1^n \delta_{X_j}\right)(S)}$$

$$P(X_{n+1} \in A \mid X_1, \ldots, X_n) = \frac{\left(\alpha + \sum_1^n \delta_{x_j}\right)(A)}{\left(\alpha + \sum_1^n \delta_{x_j}\right)(S)}$$

$$= \frac{\kappa \rho(A) + \sum_1^n \delta_{x_j}(A)}{\kappa + n}$$

$$= \underbrace{\frac{\kappa}{\kappa + n}}_{\text{stability}} \underbrace{\rho(A)}_{\text{base meas.}} + \frac{n}{\kappa + n} \cdot \underbrace{\frac{1}{n} \sum_1^n \delta_{x_j}(A)}_{\text{empirical distribution}}$$

- κ small : learn quickly
- κ large : learn slowly

V Sequence of mangled coins

The pressed penny machine is broken.
Every penny comes out different.
Get 10 pennies. Flip each 10 times.

First 9 pennies: 70 heads out of 90 flips.
Seems the machine favors heads-heavy pennies.
10^{th} penny, 1^{st} 9 flips: all tails
Prob. of head on 10^{th} flip?
Propensity of machine VS propensity of penny

Iterated Dirichlet process

(to be continued)