

A crash course in
Dirichlet processes
Part 3

Jason Swanson
UCF Probability Seminar
April 6, 2021

REVIEW +

(S, Σ) is a standard Borel space if \exists a metric d on S s.t. (S, d) is complete and separable, and $\Sigma = \mathcal{B}(S)$.

Every standard Borel sp. is Borel isomorphic to one of

- $(\mathbb{R}, \mathcal{B})$
- $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$
- $(\{0, 1, \dots, n-1\}, \mathcal{P}(\{0, 1, \dots, n-1\}))$

(Thm. 8.3.6, Measure Theory, 2nd ed., Cohn 2013)

Standard Borel spaces are important for us.

They give us:

- De Finetti

exchangeable \Leftrightarrow conditionally i.i.d.

- Glivenko-Cantelli-Varadarajan

exchangeable $\Rightarrow \frac{1}{n} \sum_{j=1}^n \delta_{x_j} \xrightarrow{\text{weakly}} \mu$ a.s.,
where $X|_{\mu} \sim \mu^\infty$

- regular conditional distributions

$$E[f(x)|\mathcal{G}] = \int f(x) \mathcal{L}(x|\mathcal{G}; dx)$$

From now on, we fix a metric d so that
 (S, d) is complete, separable, and $\mathcal{S} = \mathcal{B}(S)$.

$M_1(S)$: set of prob. measures on (S, \mathcal{S})

$\mathcal{M}_1(S)$: σ -alg. generated by projections

\hat{d} : Prokhorov metric on $M_1(S)$

(metrizes weak convergence)

$\mathcal{B}(M_1(S))$: Borel σ -alg. on $(M_1(S), \hat{d})$

$M_1(S) \subset \mathcal{B}(M_1(S))$ (pf at end)

Why work with $\mathcal{B}(M_1(S))$?

If (S, δ) is complete and separable, then
 $(M_1(S), \hat{\delta})$ is complete and separable;
hence, $(M_1(S), \mathcal{B}(M_1(S)))$ is a standard
Borel space.

(Thm. 3.1.7, Markov Processes, Ethier & Kurtz 1986)

Dirichlet processes

$\rho \in M_1(S)$ (base measure)

$\kappa > 0$ (stability constant)

$\lambda \sim D(\kappa\rho)$ (Dir. proc. on S with parameter $\kappa\rho$)

$$(\lambda(B_0), \dots, \lambda(B_d)) \sim \text{Dir}(\kappa_p(B_0), \dots, \kappa_p(B_d))$$

partition

↑ Dirichlet distribution

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

(Sethuraman stick breaking construction)

↑ i.i.d. $U_k \sim p$

random positive weights that sum to 1

λ is a random measure, i.e.

$$\lambda: \Omega \rightarrow M_+(S) \text{ is } (\mathcal{F}, \mathcal{M}_+(S))\text{-m'ble}$$

λ is also $(\mathcal{F}, \mathcal{B}(M_+(S)))$ -m'ble. (pf at end)

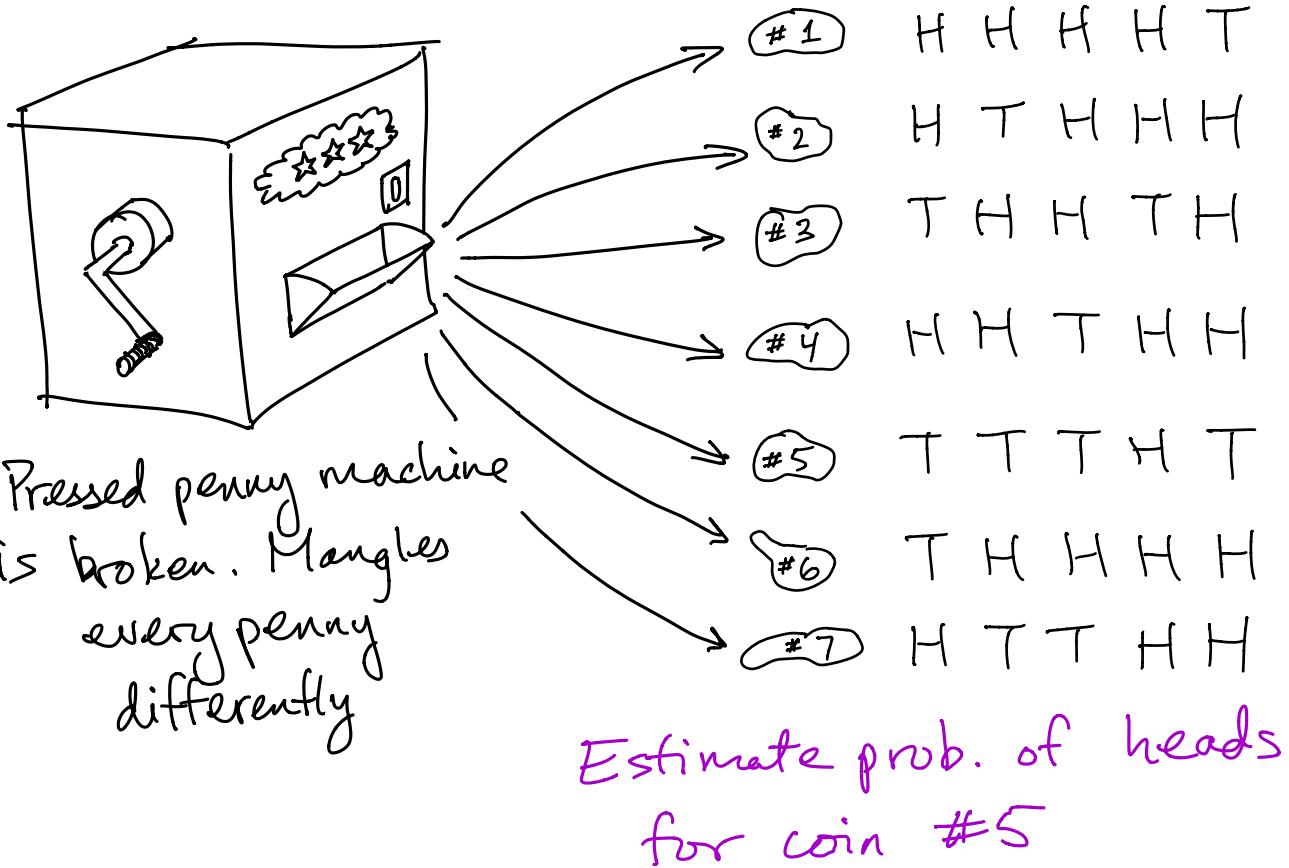
We will consider $\lambda \sim D(\kappa p)$ as an $(M, (S), \mathcal{B}(M, (S)))$ -val. r.v.

Therefore, $D(\kappa p)$ (the law of λ) is a prob. meas. on a standard Borel space.

This is important if we want to construct $w \sim D(\kappa' D(\kappa p))$, which is a Dir. proc. on $M, (S)$.

(We want our Dir. processes to be on standard Borel spaces.)

Sequence of bent coins



#1	H H H H T	#4	H H T H H	#7	H T T H H
#2	H T H H H	#5	T T T H T	← prob. heads for Coin 5?	
#3	T H H T H	#6	T H H H H		

- Machine seems to produce coins weighted toward heads
 - 24 out of 35 flips are heads
 - 6 out of 7 coins have majority H.
- Even if all coins had prob. of heads 0.6, e.g., still 43% chance of at least 1 coin in 7 getting 4 tails in 5 flips.
- Intuitively, Coin 5's flips should matter more than other coins. How to balance?

$\{X_{ij} : i, j \in \mathbb{N}\} : \{0, 1\}$ -val. r.v.s

X_{ij} : result of j^{th} flip of i^{th} coin
(0 = tails, 1 = heads)

$X_i = \{X_{ij} : j \in \mathbb{N}\}$, $X = \{X_i : i \in \mathbb{N}\}$

X_1, X_2, \dots exchangeable

For fixed i , X_{i1}, X_{i2}, \dots exchangeable

θ_i : [0, 1]-val. r.v.

X_{i1}, X_{i2}, \dots cond. i.i.d. given θ_i

with $P(X_{ij} = 1 | \theta_i) = \theta_i$

$$\frac{1}{m} \sum_{j=1}^m X_{ij} \xrightarrow{m \rightarrow \infty} \theta_i \quad \text{a.s.}$$

So $\theta_i = f(X_i)$ for some m'ble f
 (that doesn't depend on i)

$\therefore \theta_1, \theta_2, \dots$ exchangeable

$\varpi : M_1([0,1])$ -val. r.v.

$\theta_1, \theta_2, \dots$ cond. i.i.d. given ϖ

$$P(\theta_i \in A | \varpi) = \varpi(A)$$

Let ϖ be a Dirichlet process.

Must choose a base measure ρ on $[0,1]$

and a stability constant $K > 0$ s.t.

$$\omega \sim D(K\rho).$$

For simplicity, let ρ be uniform.

[A uniform r.v. is a Dirichlet process on $\{0,1\}$.

So $\omega \sim D(KD(\alpha))$.]

$$P(X_{k,m+1} = 1 \mid X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m) = ?$$

($1 \leq k \leq n$)

First issues:

- (i) $P(X_{11}=1, X_{21}=1 \mid \theta_1, \theta_2) \stackrel{?}{=} P(X_{11}=1 \mid \theta_1) P(X_{21}=1 \mid \theta_2)$
- (ii) $P(X_{11}=1 \mid \varpi, \theta_1) \stackrel{?}{=} P(X_{11}=1 \mid \theta_1)$

Not immediately obvious.

$$(i) P(X_{11}=1, X_{21}=1 | \theta_1, \theta_2) \stackrel{?}{=} P(X_{11}=1 | \theta_1) P(X_{21}=1 | \theta_2)$$

$$(ii) P(X_{11}=1 | \bar{\omega}, \theta_1) \stackrel{?}{=} P(X_{11}=1 | \theta_1)$$

Conjecture:

$$P(X_{ij} = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n | \theta_1, \theta_2, \dots)$$

$$= \prod_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} P(X_{ij} = a_{ij} | \theta_i) \text{ (ideas at end)}$$

Implies (i) (condition inside on $\mathcal{F}_\infty^\theta$)

Implies (ii) since

$$\bar{\omega} = \lim \frac{1}{n} \sum_{j=1}^n \delta_{\theta_j} \in \sigma(\theta_1, \theta_2, \dots)$$

With the conjecture established (or assumed),
can prove:

$$\begin{aligned} P(X_{k,m+1} = 1 \mid X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m) \\ = E[\theta_k \mid X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m] \end{aligned}$$

(1 \leq k \leq n)

Pf sketch:

$$A := \{X_{ij} = a_{ij}, 1 \leq i \leq n, 1 \leq j \leq m\}$$

Let $\mu_i(a) = \begin{cases} \theta_i & \text{if } a=1, \\ 1-\theta_i & \text{if } a=0. \end{cases}$

$$\begin{aligned}
 P(X_{k,m+1} = 1, A) &= E[P(X_{k,m+1} = 1, A | \theta_1, \theta_2, \dots)] \\
 &= E\left[\left(\prod_{i=1}^n \prod_{j=1}^m \mu_i(a_{ij})\right) \theta_k\right]
 \end{aligned}$$

OTOF,

$$\begin{aligned}
 E[\theta_k 1_A] &= E\left[E[\theta_k 1_A | \theta_1, \theta_2, \dots]\right] \\
 &= E\left[\left(\prod_{i=1}^n \prod_{j=1}^m \mu_i(a_{ij})\right) \theta_k\right]. \quad \square
 \end{aligned}$$

So it's enough to understand

$$\mathcal{L}(\theta_1, \dots, \theta_n | X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m)$$

Explicit formulas for

$$L(\theta_1, \dots, \theta_n | X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m)$$

are not tractable. (See Antoniak (1974) and
Berry and Christensen (1979); even the case $n=2$
occupies nearly a page in the latter paper.)

How to estimate?

$$L(\theta_k | \theta_1, \dots, \theta_{k-1}, X_{k1}, \dots, X_{km})$$

is manageable.

Regard original problem as a Bayesian
missing data problem.

Approaches:

- (i) Gibbs sampler (Geman & Geman 1984)
- (ii) data augmentation (Tanner & Wong 1987)
- (iii) sequential imputation and importance sampling (Kong, Liu, Wong 1994; Liu 1996)
 - avoids iteration



topic of Part 4, if
there ever is a Part 4

requires
iteration

Thm: $M_1(S) \subset \mathcal{B}(M_1(S))$

Pf sketch:

- Suffices to show $\pi_A(v) = v(A)$ is Borel-m'ble
- $\mathcal{L} = \{A \in \mathcal{S} : \pi_A \text{ is Borel-m'ble}\}$ is a λ -system
- Suffices to show \mathcal{L} contains open sets
- Let $U \subset S$ be open. Choose cont. $f_n \uparrow 1_U$ p.w.
e.g. <https://math.stackexchange.com/a/294845>
- $\varphi_n(v) := \int f_n d\nu$ is cont., \therefore m'ble
- $\varphi_n \rightarrow \pi_A$ p.w. $\Rightarrow \pi_A$ m'ble. □

Thm: If $\lambda \in \mathbb{A}(kp)$, then $\lambda: \Omega \rightarrow M_1(S)$ is $(\mathbb{F}, \mathcal{B}(M_1(S)))$ -m'ble.

Pf sketch:

- $(x, y) \in (0, 1)^n \times S^n \mapsto \frac{\sum_{k=1}^n x_k \delta_{y_k}}{\sum_{k=1}^n x_k} \in M_1(S)$ cont., \therefore m'ble
- $\therefore \frac{\sum_{k=1}^n R_k \delta_{U_k}}{\sum_{k=1}^n R_k}$ is $(\mathbb{F}, \mathcal{B}(M_1(S)))$ -m'ble
- λ is $(\mathbb{F}, \mathcal{B}(M_1(S)))$ -m'ble
(pointwise limit of m'ble functions) □

Conjecture:

$$P(X_{ij} = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n | \theta_1, \theta_2, \dots)$$

$$= \prod_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} P(X_{ij} = a_{ij} | \theta_i)$$

Pf sketch that doesn't work:

- X is a "separately exchangeable" array of r.v.'s (the distribution of $\{X_{ij} : i, j \in \mathbb{N}\}$ is invariant under separate permutations of i and j).

- By a result of Aldous and Hoover,^{*} \exists m'ble $f: [0, 1]^4 \rightarrow \mathbb{R}$ and indep. $U(0, 1)$ r.v.'s $\alpha, \xi_i, \eta_j, \zeta_{ij}$ such that $X_{ij} = f(\alpha, \xi_i, \eta_j, \zeta_{ij})$ a.s.
-

* Proved independently by Aldous and Hoover between 1979 and 1985. Also appears as Cor. 7.23 in Probabilistic Symmetries and Invariance Principles by Kallenberg (2005). The version here is taken from Multivariate Sampling and the Estimation Problem for Exchangeable Arrays, Kallenberg, Journal of Theoretical Probability, Vol. 12, No. 3 (1999)

- $\frac{1}{m} \sum_{j=1}^m X_{ij} \rightarrow \theta_i$ a.s.
 - $\frac{1}{m} \sum_{j=1}^m X_{ij}$
 $= \frac{1}{m} \sum_{j=1}^m f(\alpha, \xi_i, \eta_j, \zeta_{ij}) \rightarrow \int_0^1 \int_0^1 f(\alpha, \xi_i, x, y) dx dy$ a.s.
 - $\therefore \theta_i = \int_0^1 \int_0^1 f(\alpha, \xi_i, x, y) dx dy$ a.s.
- $\Rightarrow \theta_i \in \sigma(\alpha, \xi_i)$
- $P(X_{ij} = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n | \theta_1, \theta_2, \dots)$
- $= E[P(X_{ij} = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n | \alpha, \xi_1, \xi_2, \dots) | \theta_1, \theta_2, \dots]$

- Given $\alpha, \xi_1, \xi_2, \dots$, the X_{ij} 's are indep. \leftarrow NO!

$$\therefore P(X_{ij} = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n | \alpha, \xi_1, \xi_2, \dots)$$

$$= \prod_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} P(X_{ij} = a_{ij} | \alpha, \xi_1, \xi_2, \dots)$$

$$\begin{aligned} \bullet P(X_{ij} = 1 | \alpha, \xi_1, \xi_2, \dots) &= E[X_{ij} | \alpha, \xi_1, \xi_2, \dots] \\ &= E[f(\alpha, \xi_i, \eta_j, \zeta_{ij}) | \alpha, \xi_1, \xi_2, \dots] \\ &= \int_0^1 \int_0^1 f(\alpha, \xi_i, x, y) dx dy = \theta_i = P(X_{ij} = 1 | \theta_i) \end{aligned}$$

$$\text{Thus, } P(X_{ij} = 0 | \alpha, \xi_1, \xi_2, \dots) = 1 - \theta_i = P(X_{ij} = 0 | \theta_i)$$

- Put it all together... □

How to fix?

$X_{1j}, X_{2j}, X_{3j}, \dots$ cond. indep. given $\bar{\omega}$

$$\begin{aligned}\therefore \frac{1}{n} \sum_{i=1}^n X_{ij} &\rightarrow E[X_{ij} | \bar{\omega}] \\&= E[E[X_{ij} | \theta_1, \theta_2, \dots] | \bar{\omega}] \\&= E[E[X_{ij} | \alpha, \xi_1, \xi_2, \dots] | \bar{\omega}] \\&= E[\theta_i | \bar{\omega}] \\&= \int_0^1 x \bar{\omega}(dx)\end{aligned}$$

OTOH,

$X_{ij}, X_{2j}, X_{3j}, \dots$ cond. indep. given α, η_j

$$\therefore \frac{1}{n} \sum_{i=1}^n X_{ij} \rightarrow E[X_{ij} | \alpha, \eta_j]$$

$$= \int_0^1 \int_0^1 f(\alpha, x, \eta_j, y) dx dy$$

$$\therefore \int_0^1 \int_0^1 f(\alpha, x, \eta_j, y) dx dy = \int_0^1 x \omega(dx) \quad \forall j.$$

Perhaps can show it's possible to choose η_j so they don't depend on j . Then they're absorbed into α :

$$X_{ij} = \tilde{f}(\alpha, \xi_i, \zeta_{ij})$$