History of the stick breaking representation for the Dirichlet process

Jayaram Sethuraman
Department of Statistics
Florida State University
Tallahassee, FL 32306

sethu@stat.fsu.edu

February 16, 2017

Summary

- What is the stick breaking construction?
- Stick breaking construction almost in Blackwell and McQueen (1973)
- Stick breaking construction in Ferguson (1973) when combined with McCloskey (1965)
- Sethuraman's stick breaking construction

Nonparametric priors

A nonparametric prior is just a probability distribution on $\mathcal{P}$ the space of all probability measures (say on the real line). Measurable sets in $\mathcal{P}$ are of the form $\{P : P(A) < r\}$. So we should specify the distribution of $(P(A_1), P(A_2), \ldots, P(A_k))$, etc. Ferguson (1973) defined the Dirichlet process $\mathcal{D}(\alpha, \beta)$ to the random probability measure for which

$$(P(A_1), P(A_2), \ldots, P(A_k)) \sim Dirich(\alpha\beta(A_1), \alpha\beta(A_2), \ldots, \alpha\beta(A_k))$$

for all partitions $(A_1, A_2, \ldots, A_k)$ of the real line.

Nonparametric priors

A nonparametric prior can also be defined as the distribution of a random variable P taking values in $\mathcal{P}$.
The stick breaking construction does just this.
Let $\mathbf{V} = (V_1, V_2, \dots)$ be i.i.d. $Beta(1, \alpha)$. Let $p_1 = V_1, p_2 = (1 - v_1)V_2, \dots$. Then $\mathbf{p} = (p_1, p_2, \dots)$ is a random discrete distribution. Let $\mathbf{Z} = (Z_1, z_2, \dots)$ be i.i.d. β and be independent of $\mathbf{V}$. Let

$$P(A) = \sum_1^\infty p_i \delta_{Z_i}(A).$$

This P is a random probability measure and it defines a nonparametric prior. It is the stick breaking representation of the Dirichlet process.

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

This follows from an examination of De Finetti's theorem (1931), Blackwell and MacQueen (1973). See also Hewitt and Savage (1955), Kingman (1978).

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

This follows from an examination of De Finetti's theorem (1931), Blackwell and MacQueen (1973). See also Hewitt and Savage (1955), Kingman (1978).

Let $X_1, X_2, \ldots$ be an infinite sequence of exchangeable (def?) sequence of random variables with a joint distribution Q.

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

This follows from an examination of De Finetti's theorem (1931), Blackwell and MacQueen (1973). See also Hewitt and Savage (1955), Kingman (1978).

Let $X_1, X_2, \ldots$ be an infinite sequence of exchangeable (def?) sequence of random variables with a joint distribution Q.

Then, from De Finetti's theorem

1. The empirical distribution functions $F_n(x) \to F(x)$ with probability 1 for all x.

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

This follows from an examination of De Finetti's theorem (1931), Blackwell and MacQueen (1973). See also Hewitt and Savage (1955), Kingman (1978).

Let $X_1, X_2, \ldots$ be an infinite sequence of exchangeable (def?) sequence of random variables with a joint distribution Q.

Then, from De Finetti's theorem

1. The empirical distribution functions $F_n(x) \to F(x)$ with probability 1 for all x. In fact, $\sup_x |F_n(x) - F(x)| \to 0$ with probability 1.

Nonparametric priors

The class of all nonparametric priors are the same as the class of all exchangeable sequences of random variables!

This follows from an examination of De Finetti's theorem (1931), Blackwell and MacQueen (1973). See also Hewitt and Savage (1955), Kingman (1978).

Let $X_1, X_2, \dots$ be an infinite sequence of exchangeable (def?) sequence of random variables with a joint distribution Q.

Then, from De Finetti's theorem

1. The empirical distribution functions $F_n(x) \to F(x)$ with probability 1 for all x. In fact, $\sup_x |F_n(x) - F(x)| \to 0$ with probability 1.
   (Note that $F(x)$ is a random distribution function.)

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.
3. Given P, $X_1, X_2, \ldots$ are i.i.d. P.

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.
3. Given P, $X_1, X_2, \ldots$ are i.i.d. P.
4. Let us denote the distribution of P under Q by ν^Q. This ν^Q is a nonparametric prior – it is a pm on the space of pm's on R_1.

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.
3. Given P, $X_1, X_2, \ldots$ are i.i.d. P.
4. Let us denote the distribution of P under Q by ν^Q. This ν^Q is a nonparametric prior – it is a pm on the space of pm's on R_1.
5. The class of all nonparametric priors arises in this fashion.

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.
3. Given P, $X_1, X_2, \ldots$ are i.i.d. P.
4. Let us denote the distribution of P under Q by ν^Q. This ν^Q is a nonparametric prior – it is a pm on the space of pm's on R_1.
5. The class of all nonparametric priors arises in this fashion.
6. The distribution of $X_2, X_3, \ldots,$ given X_1 is also exchangeable; denote it by Q_{X_1}.
7. The limit P of the empirical probability measures of $X_1, X_2, \ldots$ is also the limit of the empirical probability measures of $X_2, X_3, \ldots$.

Nonparametric priors

2. The empirical probability measures P_n converge to a random probability measure P weakly with probability 1.
3. Given P, $X_1, X_2, \ldots$ are i.i.d. P.
4. Let us denote the distribution of P under Q by ν^Q. This ν^Q is a nonparametric prior – it is a pm on the space of pm's on R_1.
5. The class of all nonparametric priors arises in this fashion.
6. The distribution of $X_2, X_3, \ldots,$ given X_1 is also exchangeable; denote it by Q_{X_1}.
7. The limit P of the empirical probability measures of $X_1, X_2, \ldots$ is also the limit of the empirical probability measures of $X_2, X_3, \ldots$. Thus the distribution of P given X_1 (the posterior distribution) is the distribution of P under Q_{X_1} and, by mere notation, is $\nu^{Q_{X_1}}$.

Dirichlet prior based on a Pólya urn sequences

The Pólya urn sequence is an example of an infinite exchangeable random variables.

Let β be a pm on R_1 and let $\alpha > 0$. Define the joint distribution $Pol(\alpha, \beta)$ of $X_1, X_2, \ldots$ through

$$X_1 \sim \beta(\cdot), \; X_2 | X_1 \sim \frac{\alpha\beta(\cdot) + \delta_{X_1}(\cdot)}{\alpha + 1}$$

$$X_n | (X_1, \ldots, X_{n-1}) \sim \frac{\alpha\beta(\cdot) + \sum_1^{n-1} \delta_{X_i}(\cdot)}{\alpha + n - 1}, n = 3, 4, \ldots$$

This defines $Pol(\alpha, \beta)$ as an exchangeable probability measure. (It takes just some effort to establish this.)

What about the distribution of $(X_2, X_3, \ldots) | X_1$?

Dirichlet prior based on a Pólya urn sequences

The Pólya urn sequence is an example of an infinite exchangeable random variables.

Let β be a pm on R_1 and let $\alpha > 0$. Define the joint distribution $Pol(\alpha, \beta)$ of $X_1, X_2, \ldots$ through

$$X_1 \sim \beta(\cdot), \; X_2 | X_1 \sim \frac{\alpha\beta(\cdot) + \delta_{X_1}(\cdot)}{\alpha + 1}$$

$$X_n | (X_1, \ldots, X_{n-1}) \sim \frac{\alpha\beta(\cdot) + \sum_1^{n-1} \delta_{X_i}(\cdot)}{\alpha + n - 1}, n = 3, 4, \ldots$$

This defines $Pol(\alpha, \beta)$ as an exchangeable probability measure. (It takes just some effort to establish this.)

What about the distribution of $(X_2, X_3, \ldots) | X_1$? It is $Pol(\alpha + 1, \frac{\alpha\beta + \delta_{X_1}}{\alpha+1})$.

Dirichlet prior based on a Pólya urn sequences

- The nonparametric prior $\nu^{Pol(\alpha,\beta)}$ is the same as the Dirichlet prior $\mathcal{D}(\alpha\beta)$!

Dirichlet prior based on a Pólya urn sequences

- The nonparametric prior $\nu^{Pol(\alpha,\beta)}$ is the same as the Dirichlet prior $\mathcal{D}(\alpha\beta)$!
- That is, the distribution of $(P(A_1), \ldots, P(A_k))$ for any partition $(A_1, \ldots, A_k)$, under $Pol(\alpha, \beta)$, is the finite dimensional Dirichlet $\mathcal{D}(\alpha\beta(A_1), \ldots, \alpha\beta(A_k))$. This is proved in Blackwell and MacQueen (1973).

Dirichlet prior based on a Pólya urn sequences

- The nonparametric prior $\nu^{Pol(\alpha,\beta)}$ is the same as the Dirichlet prior $\mathcal{D}(\alpha\beta)$!
- That is, the distribution of $(P(A_1), \ldots, P(A_k))$ for any partition $(A_1, \ldots, A_k)$, under $Pol(\alpha, \beta)$, is the finite dimensional Dirichlet $\mathcal{D}(\alpha\beta(A_1), \ldots, \alpha\beta(A_k))$. This is proved in Blackwell and MacQueen (1973).

  For any A, $P(A) \sim Beta(\alpha\beta(A), \alpha\beta(A^c))$.

Dirichlet prior based on a Pólya urn sequences

- The nonparametric prior $\nu^{Pol(\alpha,\beta)}$ is the same as the Dirichlet prior $\mathcal{D}(\alpha\beta)$!
- That is, the distribution of $(P(A_1), \ldots, P(A_k))$ for any partition $(A_1, \ldots, A_k)$, under $Pol(\alpha, \beta)$, is the finite dimensional Dirichlet $\mathcal{D}(\alpha\beta(A_1), \ldots, \alpha\beta(A_k))$. This is proved in Blackwell and MacQueen (1973).

  For any A, $P(A) \sim Beta(\alpha\beta(A), \alpha\beta(A^c))$. Can we allow $A = \{X_1\}$ in the above?

Dirichlet prior based on a Pólya urn sequences

- The conditional distribution of $(X_2, X_3, \dots)$ given X_1 is $Pol(\alpha + 1, \frac{\alpha\beta + \delta_{X_1}}{\alpha + 1})$.

Dirichlet prior based on a Pólya urn sequences

- The conditional distribution of $(X_2, X_3, \dots)$ given X_1 is $Pol(\alpha + 1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})$.
- Thus posterior distribution of P given X_1 is $\nu^{Pol(\alpha+1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})}$ which is equal to $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

Dirichlet prior based on a Pólya urn sequences

- The conditional distribution of $(X_2, X_3, \dots)$ given X_1 is $Pol(\alpha + 1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})$.
- Thus posterior distribution of P given X_1 is $\nu^{Pol(\alpha+1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})}$ which is equal to $\mathcal{D}(\alpha\beta + \delta_{X_1})$.
- Though each P_n is a discrete rpm and the limit P in general will be just a rpm.

Dirichlet prior based on a Pólya urn sequences

- The conditional distribution of $(X_2, X_3, \dots)$ given X_1 is $Pol(\alpha + 1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})$.

- Thus posterior distribution of P given X_1 is $\nu^{Pol(\alpha+1, \frac{\alpha\beta+\delta_{X_1}}{\alpha+1})}$ which is equal to $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

- Though each P_n is a discrete rpm and the limit P in general will be just a rpm.

- For the present case of a Pólya urn sequence, Blackwell and MacQueen (1973) show that $P(\{X_1, \dots, X_n\}) \to 1$ with probability 1 and thus P is a discrete rpm. (A little tricky. We will show some details.)

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n}$

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n}$

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\} | X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n} \to 1.$

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n} \to 1$.
This means that P is a discrete random probability measure.

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n} \to 1$.
This means that P is a discrete random probability measure.

From now on, assume that β is non-atomic.

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1,\ldots,X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n} \to 1$.
This means that P is a discrete random probability measure.

From now on, assume that β is non-atomic.

The above conditional distribution of $P(\{X_1\})$ given X_1 becomes $B(1, \alpha)$ which does not depend on X_1

Dirichlet prior based on a Pólya urn sequences

The conditional distribution of P given X_1 is $\mathcal{D}(\alpha\beta + \delta_{X_1})$.

The conditional distribution of $P(\{X_1\})$ given X_1 is

$$B(\alpha\beta(\{X_1\}) + 1, \alpha\beta(R_1 \setminus \{X_1\})).$$

This is tricky. Is $P(\{X_1\})$ measurable to begin with?

The conditional distribution of $P(\{X_1, \ldots, X_n\})$ given $(X_1, \ldots, X_n)$ is $Beta(\alpha\beta(\{X_1, \ldots, X_n\}) + n, \alpha\beta(R_1 \setminus \{X_1, \ldots, X_n\}))$
and
$E(P(\{X_1, \ldots, X_n\}|X_1, \ldots, X_n)) = \frac{\alpha\beta(\{X_1, \ldots, X_n\})+n}{\alpha+n} \geq \frac{n}{\alpha+n} \to 1$.
This means that P is a discrete random probability measure.

From now on, assume that β is non-atomic.

The above conditional distribution of $P(\{X_1\})$ given X_1 becomes $B(1, \alpha)$ which does not depend on X_1 and thus X_1 and $P(\{X_1\})$ are independent.

Dirichlet prior based on a Pólya urn sequences

Let $Y_1, Y_2, \ldots$ be the distinct values among $X_1, X_2, \ldots$ listed in the order of their appearance.

Then $Y_1 = X_1$,

$Y_1, P(\{Y_1\})$ are independent

Dirichlet prior based on a Pólya urn sequences

Let $Y_1, Y_2, \ldots$ be the distinct values among $X_1, X_2, \ldots$ listed in the order of their appearance.

Then $Y_1 = X_1$,

$Y_1, P(\{Y_1\})$ are independent and $Y_1 \sim \beta, P(\{Y_1\}) \sim B(1, \alpha)$.

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \ldots$ and remove all occurrences of X_1 which is the same as Y_1.

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \ldots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1.

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \ldots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1. Its first element is Y_2.

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \ldots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1. Its first element is Y_2.

As before, Y_2 and $\frac{P(\{Y_2\})}{1-P(\{Y_1\})}$ are independent, $Y_2 \sim \beta, \frac{P(\{Y_2\})}{1-P(\{Y_1\})} \sim B(1, \alpha)$.

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \dots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1. Its first element is Y_2.

As before, Y_2 and $\frac{P(\{Y_2\})}{1-P(\{Y_1\})}$ are independent,
$Y_2 \sim \beta, \frac{P(\{Y_2\})}{1-P(\{Y_1\})} \sim B(1, \alpha)$.

Thus $P(\{Y_1\}), \frac{P(\{Y_2\})}{1-P(\{Y_1\})}, \frac{P(\{Y_3\})}{1-P(\{Y_1\})-P(\{Y_2\})}, \dots$ are i.i.d. $B(1, \alpha)$

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \ldots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1. Its first element is Y_2.

As before, Y_2 and $\frac{P(\{Y_2\})}{1-P(\{Y_1\})}$ are independent, $Y_2 \sim \beta$, $\frac{P(\{Y_2\})}{1-P(\{Y_1\})} \sim B(1, \alpha)$.

Thus $P(\{Y_1\}), \frac{P(\{Y_2\})}{1-P(\{Y_1\})}, \frac{P(\{Y_3\})}{1-P(\{Y_1\})-P(\{Y_2\})}, \ldots$ are i.i.d. $B(1, \alpha)$ (i.e. stick breaking, $GEM(\alpha)$)

Dirichlet prior based on a Pólya urn sequences

Consider the sequence $X_2, X_3, \dots$ and remove all occurrences of X_1 which is the same as Y_1. This reduced sequence is the Pólya urn sequence $Pol(\alpha, \beta)$ and independent of Y_1. Its first element is Y_2.

As before, Y_2 and $\frac{P(\{Y_2\})}{1-P(\{Y_1\})}$ are independent, $Y_2 \sim \beta$, $\frac{P(\{Y_2\})}{1-P(\{Y_1\})} \sim B(1, \alpha)$.

Thus $P(\{Y_1\}), \frac{P(\{Y_2\})}{1-P(\{Y_1\})}, \frac{P(\{Y_3\})}{1-P(\{Y_1\})-P(\{Y_2\})}, \dots$ are i.i.d. $B(1, \alpha)$ (i.e. stick breaking, $GEM(\alpha)$)

and all these are independent of $Y_1, Y_2, Y_3 \dots$ which are i.i.d. β.

Dirichlet prior based on a Pólya urn sequences

Since P is discrete and just sits on the set $\{X_1, X_2, \ldots\}$ which is $\{Y_1, Y_2, \ldots\}$,

Dirichlet prior based on a Pólya urn sequences

Since P is discrete and just sits on the set $\{X_1, X_2, \dots\}$ which is $\{Y_1, Y_2, \dots\}$,

and thus $P = \sum_1^\infty P(\{Y_i\})\delta_{Y_1}$.
Put $p_i = P(Y_i), i = 1, 2, \dots$. Then $P = \sum_1^\infty p_i \delta_{Y_i}$; i.e. we have the Sethuraman stick breaking construction of the Dirichlet prior (if β is non-atomic).

Dirichlet prior based on a Pólya urn sequences

Since P is discrete and just sits on the set $\{X_1, X_2, \dots\}$ which is $\{Y_1, Y_2, \dots\}$,

and thus $P = \sum_1^\infty P(\{Y_i\})\delta_{Y_1}$.
Put $p_i = P(Y_i), i = 1, 2, \dots$. Then $P = \sum_1^\infty p_i \delta_{Y_i}$; i.e. we have the Sethuraman stick breaking construction of the Dirichlet prior (if β is non-atomic).

This is how we can turn around the article by Blackwell and MacQueen (1973) to obtain the stick breaking result when β is non-atomic.

Dirichlet prior based on a Pólya urn sequences

Since P is discrete and just sits on the set $\{X_1, X_2, \dots\}$ which is $\{Y_1, Y_2, \dots\}$,

and thus $P = \sum_1^\infty P(\{Y_i\})\delta_{Y_1}$.
Put $p_i = P(Y_i), i = 1, 2, \dots$. Then $P = \sum_1^\infty p_i\delta_{Y_i}$; i.e. we have the Sethuraman stick breaking construction of the Dirichlet prior (if β is non-atomic).

This is how we can turn around the article by Blackwell and MacQueen (1973) to obtain the stick breaking result when β is non-atomic.

However, note that the statement of the stick breaking construction does not assume any properties of β!

ISBP

As an aside, let us examine how $(p_i = P(\{Y_i\}), i = 1, 2, \dots)$ depends on P.

ISBP

As an aside, let us examine how $(p_i = P(\{Y_i\}), i = 1, 2, \dots)$ depends on P.

It is a size biased permutation of the probabilities $(\pi_i, i = 1, 2, \dots)$ of the atoms of the discrete pm P. We can write this as $\mathbf{p} = SBP(\boldsymbol{\pi})$.

ISBP

As an aside, let us examine how $(p_i = P(\{Y_i\}), i = 1, 2, \dots)$ depends on P.

It is a size biased permutation of the probabilities $(\pi_i, i = 1, 2, \dots)$ of the atoms of the discrete pm P. We can write this as $\mathbf{p} = SBP(\boldsymbol{\pi})$.

A further size biased permutation of $\mathbf{q}$ of $\mathbf{p}$ is also a SBP of $\boldsymbol{\pi}$ and thus $\mathbf{p} \sim \mathbf{q}$. This is called the ISBP property of $\mathbf{p}$.

ISBP

As an aside, let us examine how $(p_i = P(\{Y_i\}), i = 1, 2, \dots)$ depends on P.

It is a size biased permutation of the probabilities $(\pi_i, i = 1, 2, \dots)$ of the atoms of the discrete pm P. We can write this as $\mathbf{p} = SBP(\boldsymbol{\pi})$.

A further size biased permutation of $\mathbf{q}$ of $\mathbf{p}$ is also a SBP of $\boldsymbol{\pi}$ and thus $\mathbf{p} \sim \mathbf{q}$. This is called the ISBP property of $\mathbf{p}$.

In particular, if R is an observation from $\mathbf{p}$, then $(p_R, \mathbf{p}^{-R}) \sim \mathbf{p} = (p_1, \mathbf{p}^{-1})$.

ISBP

As an aside, let us examine how $(p_i = P(\{Y_i\}), i = 1, 2, \ldots)$ depends on P.

It is a size biased permutation of the probabilities $(\pi_i, i = 1, 2, \ldots)$ of the atoms of the discrete pm P. We can write this as $\mathbf{p} = SBP(\boldsymbol{\pi})$.

A further size biased permutation of $\mathbf{q}$ of $\mathbf{p}$ is also a SBP of $\boldsymbol{\pi}$ and thus $\mathbf{p} \sim \mathbf{q}$. This is called the ISBP property of $\mathbf{p}$.

In particular, if R is an observation from $\mathbf{p}$, then $(p_R, \mathbf{p}^{-R}) \sim \mathbf{p} = (p_1, \mathbf{p}^{-1})$.

Further more $\frac{\mathbf{p}^{-R}}{1-p_R} \sim \frac{\mathbf{p}^{-1}}{1-p_1} \sim \mathbf{p}$, and so on.

Ferguson (1973)

The full stick breaking construction was available right in Ferguson (1973) paper if we can add McCloskey (1965)!

Ferguson (1973)

The full stick breaking construction was available right in Ferguson (1973) paper if we can add McCloskey (1965)!

Ferguson showed that

$$P^* = \sum_1^\infty \pi_i^* \delta_{Z_i}$$

has distribution $\mathcal{D}(\alpha\beta)$ if $\mathbf{Z}$ are i.i.d. β and $(\pi_1^*, \pi_2^*, \dots)$ are independent of $\mathbf{Z}$ and are the ordered normalized jumps of a Gamma process on $[0, 1]$ with shape parameter α.

Ferguson (1973)

The full stick breaking construction was available right in Ferguson (1973) paper if we can add McCloskey (1965)!

Ferguson showed that

$$P^* = \sum_1^\infty \pi_i^* \delta_{Z_i}$$

has distribution $\mathcal{D}(\alpha\beta)$ if $\mathbf{Z}$ are i.i.d. β and $(\pi_1^*, \pi_2^*, \dots)$ are independent of $\mathbf{Z}$ and are the ordered normalized jumps of a Gamma process on $[0, 1]$ with shape parameter α.

McCloskey (1965) showed that a size biased permutation of $(\pi_1^*, \pi_2^*, \dots)$ is $GEM(\alpha)$ and so we can rewrite

$$P = \sum_1^\infty p_i \delta_{Z_i}$$

and get the stick breaking construction.

Sethuraman construction of Dirichlet priors

Sethuraman (1994)

Sethuraman construction of Dirichlet priors

Let $\alpha > 0$ and let $\beta(\cdot)$ be a pm on $\mathcal{X}$.

We do not assume that β is non-atomic. Restrictions like $\mathcal{X} = R_1$ do not have to made.

Let $V_1, V_2, \ldots,$ be i.i.d. $B(1, \alpha)$ and let $Z_1, Z_2, \ldots$ be independent of $V_1, V_2, \ldots$ and be i.i.d. $\beta(\cdot)$.

Let $p_1 = V_1, p_2 = (1 - V_1)V_2, p_3 = V_3(1 - V_1)(1 - V_2), \ldots.$

Sethuraman construction of Dirichlet priors

The stick breaking construction is

$$P(\cdot) = P(\mathbf{p}, \mathbf{Z})(\cdot) = \sum_{1}^{\infty} p_i \delta_{Z_i}(\cdot)$$

Sethuraman construction of Dirichlet priors

The stick breaking construction is

$$P(\cdot) = P(\mathbf{p}, \mathbf{Z})(\cdot) = \sum_{1}^{\infty} p_i \delta_{Z_i}(\cdot)$$

It is clearly a discrete random probability measure.

Sethuraman construction of Dirichlet priors

The stick breaking construction is

$$P(\cdot) = P(\mathbf{p}, \mathbf{Z})(\cdot) = \sum_{1}^{\infty} p_i \delta_{Z_i}(\cdot)$$

It is clearly a discrete random probability measure.
We have the special identity

$$P = p_1\delta_{Z_1} + (1-p_1)\sum_{2}^{\infty} \frac{p_i}{1-p_1}\delta_{Z_i} = p_1\delta_{Z_1} + (1-p_1)P(\mathbf{p}^{-1}/(1-p_1), \mathbf{Z}^{-1})$$

where $\mathbf{p}^{-1}, \mathbf{Z}^{-1}$ have the obvious meanings.

Sethuraman construction of Dirichlet priors

The stick breaking construction is

$$P(\cdot) = P(\mathbf{p}, \mathbf{Z})(\cdot) = \sum_{1}^{\infty} p_i \delta_{Z_i}(\cdot)$$

It is clearly a discrete random probability measure.
We have the special identity

$$P = p_1\delta_{Z_1} + (1 - p_1)\sum_{2}^{\infty} \frac{p_i}{1 - p_1}\delta_{Z_i} = p_1\delta_{Z_1} + (1 - p_1)P(\mathbf{p}^{-1}/(1 - p_1), \mathbf{Z}^{-1})$$

where $\mathbf{p}^{-1}, \mathbf{Z}^{-1}$ have the obvious meanings.
We could have split the above with index R, (even a random index R) instead of the index 1.

Sethuraman construction of Dirichlet priors

The stick breaking construction is

$$P(\cdot) = P(\mathbf{p}, \mathbf{Z})(\cdot) = \sum_{1}^{\infty} p_i \delta_{Z_i}(\cdot)$$

It is clearly a discrete random probability measure.
We have the special identity

$$P = p_1\delta_{Z_1} + (1 - p_1)\sum_{2}^{\infty} \frac{p_i}{1 - p_1}\delta_{Z_i} = p_1\delta_{Z_1} + (1 - p_1)P(\mathbf{p}^{-1}/(1 - p_1), \mathbf{Z}^{-1})$$

where $\mathbf{p}^{-1}, \mathbf{Z}^{-1}$ have the obvious meanings.
We could have split the above with index R, (even a random index R) instead of the index 1. We will use this identity to prove that the distribution of P is $\mathcal{D}(\alpha\beta)$ and to obtain the posterior distribution.

Sethuraman construction of Dirichlet priors

The special identity shows that

$$P = p_1 \delta_{Z_1} + (1 - p_1) P^*$$

where all the random variables are independent,
$p_1 \sim B(1, \alpha), Z_1 \sim \beta$ and the two rpm's P, P^* have the same distribution.

Sethuraman construction of Dirichlet priors

The special identity shows that

$$P = p_1 \delta_{Z_1} + (1 - p_1) P^*$$

where all the random variables are independent,
$p_1 \sim B(1, \alpha), Z_1 \sim \beta$ and the two rpm's P, P^* have the same distribution.

That is, we have a distributional equation for the distribution of P:

$$P \stackrel{d}{=} p_1 \delta_{Z_1} + (1 - p_1) P.$$

Sethuraman construction of Dirichlet priors

The special identity shows that

$$P = p_1 \delta_{Z_1} + (1 - p_1) P^*$$

where all the random variables are independent,
$p_1 \sim B(1, \alpha), Z_1 \sim \beta$ and the two rpm's P, P^* have the same distribution.

That is, we have a distributional equation for the distribution of P:

$$P \stackrel{d}{=} p_1 \delta_{Z_1} + (1 - p_1) P.$$

In Sethuraman (1994) we show that $\mathcal{D}(\alpha\beta)$ is a solution to this equation, and also that, if there is a solution then it is unique.

Sethuraman construction of Dirichlet priors

What about the posterior distribution?

Sethuraman construction of Dirichlet priors

What about the posterior distribution?

Let R be a random variable such $Q(R = r|\mathbf{p}) = p_r, r = 1, 2, \ldots$ and let $Y = Z_R$. Then

$$\begin{aligned} Q(Y \in A|P) &= Q(Y \in A|(\mathbf{p}, \mathbf{Z})) \\ &= \sum_r Q(Y \in A, R = r|(\mathbf{p}, \mathbf{Z})) \\ &= \sum_r Q(Z_r \in A)p_r = P(A) \end{aligned}$$

Thus Y is a like an observation from P and we need the distribution of P given Y.

Sethuraman construction of Dirichlet priors

The special identity gives

$$P = p_R \delta_Y + (1 - p_R) P(\mathbf{p}^{-R}/(1 - p_R), \mathbf{Z}^{-R}).$$

Sethuraman construction of Dirichlet priors

The special identity gives

$$P = p_R \delta_Y + (1 - p_R) P(\mathbf{p}^{-R}/(1 - p_R), \mathbf{Z}^{-R}).$$

Conditional on (R, Y), the right hand side has distribution

$$p_R \delta_Y + (1 - p_R)\mathcal{D}(\alpha\beta) \text{ which is } \sim (1 - p_1)\delta_Y + (1 - p_1)\mathcal{D}(\alpha\beta)$$

Sethuraman construction of Dirichlet priors

The special identity gives

$$P = p_R \delta_Y + (1 - p_R) P(\mathbf{p}^{-R}/(1 - p_R), \mathbf{Z}^{-R}).$$

Conditional on (R, Y), the right hand side has distribution

$p_R \delta_Y + (1 - p_R) \mathcal{D}(\alpha\beta)$ which is $\sim (1 - p_1)\delta_Y + (1 - p_1)\mathcal{D}(\alpha\beta)$

since p_1 is independent of $\frac{\mathbf{p}^{-1}}{(1-p_1)}$.
This is the same as $\mathcal{D}(\alpha\beta + \delta_Y)$, from standard identities of Dirichlet distributions.

Sethuraman construction of Dirichlet priors

The special identity gives

$$P = p_R \delta_Y + (1 - p_R) P(\mathbf{p}^{-R}/(1 - p_R), \mathbf{Z}^{-R}).$$

Conditional on (R, Y), the right hand side has distribution

$$p_R \delta_Y + (1 - p_R)\mathcal{D}(\alpha\beta) \text{ which is } \sim (1 - p_1)\delta_Y + (1 - p_1)\mathcal{D}(\alpha\beta)$$

since p_1 is independent of $\frac{\mathbf{p}^{-1}}{(1-p_1)}$.
This is the same as $\mathcal{D}(\alpha\beta + \delta_Y)$, from standard identities of Dirichlet distributions.

Thus the distribution of P given Y is $\mathcal{D}(\alpha\beta + \delta_Y)$.

Miconceptions on the stick breaking construction

It is amply clear that Sethuraman (1994) did not impose any conditions on the base measure $\beta(\cdot)$ that it should be non-atomic.

Many papers continue to assert that Sethuraman (1994) assumes that $\beta(\cdot)$ should be non-atomic.

Miconceptions on the stick breaking construction

It is amply clear that Sethuraman (1994) did not impose any conditions on the base measure $\beta(\cdot)$ that it should be non-atomic.

Many papers continue to assert that Sethuraman (1994) assumes that $\beta(\cdot)$ should be non-atomic.

Paisley (2010) says "We use a little-known property of the constructive definition in (Sethuraman, 1994)" following my personal assurance to him that he can use the stick breaking construction to generate the $Beta(a, b)$ distribution.

Miconceptions on the stick breaking construction

It is amply clear that Sethuraman (1994) did not impose any conditions on the base measure $\beta(\cdot)$ that it should be non-atomic.

Many papers continue to assert that Sethuraman (1994) assumes that $\beta(\cdot)$ should be non-atomic.

Paisley (2010) says "We use a little-known property of the constructive definition in (Sethuraman, 1994)" following my personal assurance to him that he can use the stick breaking construction to generate the $Beta(a, b)$ distribution.

Let $Z_1, Z_2, \ldots$ be i.i.d. with $Q(Z_1 = 1) = 1 - Q(Z_1 = 0) = \frac{a}{a+b}$ and $(p_1, p_2, \ldots)$ be $\mathrm{GEM}(a + b)$.

Miconceptions on the stick breaking construction

It is amply clear that Sethuraman (1994) did not impose any conditions on the base measure $\beta(\cdot)$ that it should be non-atomic.

Many papers continue to assert that Sethuraman (1994) assumes that $\beta(\cdot)$ should be non-atomic.

Paisley (2010) says "We use a little-known property of the constructive definition in (Sethuraman, 1994)" following my personal assurance to him that he can use the stick breaking construction to generate the $Beta(a, b)$ distribution.

Let $Z_1, Z_2, \dots$ be i.i.d. with $Q(Z_1 = 1) = 1 - Q(Z_1 = 0) = \frac{a}{a+b}$ and $(p_1, p_2, \dots)$ be $\mathrm{GEM}(a + b)$.

$$P = \sum p_i \; I(Z_1 = 1) \sim Beta(a, b)$$

Miconceptions on the stick breaking construction

Ferguson showed that the support of the $\mathcal{D}(\alpha\beta)$ is the collection of probability measures in $\mathcal{P}$ whose support is contained in the support of β.

If the support of β is R_1 then the support of $\mathcal{D}_{\alpha\beta}$ is $\mathcal{P}$.

Miconceptions on the stick breaking construction

Ferguson showed that the support of the $\mathcal{D}(\alpha\beta)$ is the collection of probability measures in $\mathcal{P}$ whose support is contained in the support of β.

If the support of β is R_1 then the support of $\mathcal{D}_{\alpha\beta}$ is $\mathcal{P}$.

We already saw that $\mathcal{D}(\alpha\beta)$ gives probability 1 to the class of discrete pm's.

Miconceptions on the stick breaking construction

Ferguson showed that the support of the $\mathcal{D}(\alpha\beta)$ is the collection of probability measures in $\mathcal{P}$ whose support is contained in the support of β.

If the support of β is R_1 then the support of $\mathcal{D}_{\alpha\beta}$ is $\mathcal{P}$.

We already saw that $\mathcal{D}(\alpha\beta)$ gives probability 1 to the class of discrete pm's.

$\mathcal{D}(\alpha\beta)$ is not itself a discrete probability measure.