Dirichlet Process Poisson Mixture Models in a Conjugate Framework

STAT 447 Final Project

Caden Hewlett

2024-04-12

Introduction

```
# TODO
```

We will begin with a brief literature review, in which probability measures are discussed and Dirichlet Processes are introduced from a theoretical standpoint. In the Data Analysis section, we briefly recap the source data aggregation and inspect the plot as well as some summary statistics. In the Methodology section, we outline the stick-breaking process used by the algorithm and formally define both the Bayesian model and hyper-parameter choices for the DPMM. In the Results section, we produce the *a posteriori* results, including the table of posterior means of the rate parameters and the plot of the posterior distribution. We then provide a brief interpretation of these results.

Literature Review

Before we discuss Dirichlet Processes, it is crucial to establish a groundwork in probability measure theory. We will briefly revisit the concepts of σ-algebra and probability measures. In the following, we present generalized definitions which are discussed in rigor works such as (Billingsley 2012) and (Rudin 1986).

Measure Theory

Let $\mathbb{X}$ be a well-defined sample space. A σ-algebra $\mathcal{F} \subseteq P(\mathbb{X})$ is a set satisfying the following:

1. The entire sample space $\mathbb{X}$ is in $\mathcal{F}$ and $\emptyset$ is in $\mathcal{F}$. This is referred to in the literature as "non-emptiness and universality."
2. For all sets $A \in \mathcal{F}$, the complement $A^c \in \mathcal{F}$. This property is referred to as "closure under complementation."
3. For any countable collection of sets $\{A_i\}_{i \in I}$, where I is a countable index set, if $\forall i \in I, A_i \in \mathcal{F}$ then $\bigcup_{i \in I} A_i \in \mathcal{F}$. This is referred to as "closure under countable unions."

For the sake of this work, we are more interested in *probability measures*, which are built on σ-algebra. A probability measure $\mu : \mathcal{F} \mapsto [0, 1]$ satisfies the following familiar axioms of probability:

1. All mappings are non-negative: $\forall A \in \mathcal{F},\ \mu(A) \geq 0$. This is referred to as "non-negativity."
2. $\mu(\mathbb{X}) = 1$, and $\mu(\emptyset) = 0$.
3. For any countable set $\{A_i\} \subseteq \mathcal{F}$ where $\forall i \neq j, A_i \cap A_j = \emptyset$, we have that $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$, this is referred to as "countable additivity."

Before we move on to Dirichlet Processes, we acknowledge that the above definitions might be abstract for those new to measure theory. For clarity, the Appendix includes examples to illustrate σ-algebra properties and verify a probability measure μ on a simple finite set.

Dirichlett Process: A Distribution Over Measures

Now, we introduce the concept of the Dirichlet Process (DP), a distribution across probability measures. We adopt the description used in (Hannah 2011), where the DP is specified by its base measure $\mathbb{G}_0$ (commonly a distribution) and the concentration parameter, α. A random sample from a Dirichlet Process is a probability measure over a σ-algebra from sample space $\mathbb{X}$, which is often assumed to be countable. The resulting probability measure follows the structure of the base measure $\mathbb{G}_0$ with a degree of deviation controlled by α. A Dirichlet Process can also be thought of conceptually of as a Dirichlet Distribution whose support is an infinite-simplex, rather than a discretely-defined k-simplex. (Ferguson 1973)

The α parameter indicates less confidence in the base measure, yielding greater dispersion and a higher number of clusters of measures about $\mathbb{G}_0$. Conversely, a lower α indicates more confidence in $\mathbb{G}_0$ resulting in fewer, larger clusters. The finite approximation of this property is explored in more detail in the Methods section.

Data Analysis

Aggregation code framework modified from (StackOverflow 2023)

```
# TODO: discuss data itself
#       discuss aggregation and latent categorization (and events where this could take place irl)
#       discuss 2014 imputation process (by nearest neighbours)
# What we effectively have is discrete weekly counts, with latent categorization (damage severity) that
```

Methods

We will use a finite-approximation of a Dirichlet Process centered about a gamma base measure using the stick-breaking methodology. In addition, we will utilize the gamma-poisson conjugacy to implicitly define the mixing distribution, which will be utilized be used as the `MixingDistribution` object in the MCMC DP sampler implemented in the `dirichletprocess` package (Markwick 2023) to produce a posterior distribution.

In this section, we outline the stick-breaking process used by the algorithm and formally define both the Bayesian model and hyper-parameter choices for the DPMM.

(Sethuraman 1994) and (Teh 2006)

```
# TODO: First, Discuss finite-approximation via GEM
# TODO: Then discuss implementation via custom Mixing Distribution : https://en.wikipedia.org/wiki/Mixt
# https://dm13450.github.io/2018/02/21/Custom-Distributions-Conjugate.html
# TODO: noting that when deciding on a prior-likelihood model, the characterization
# of the system for a well-functioning mixture distribution (to be used in the DPMM)
# requires the following four characteristics, which we derive from the gamma-poisson conjugacy
# it's important to recall that while this conjugate system allows us to well-define the parameters
# for the mixing distribution as required for a Dirichlett Process sampler, the gamma distribution
# in the prior-likelihood pair is the base measure, not the prior. we can think of these as the
# parameters to the mixing process. (i.e. requires a Mixing Distribution Object `mdObject` which is
# implicitly defined by the following four features: (4 functions))

# TODO: when setup is done, declare model equations
```

Results

```
# Total Runtime of Script:  753.68 seconds
# TODO: generation from the Gibbs sampling procedure detailed in [@Neal2000]
# TODO: report posterior distrubtion (of mean of lambda with CrI error bars) from posterior frame
# TODO: report posterior weights and lambda values (much more useful, actually shows high
# weight on three separate lambdas )
```

Using ((CPD) 2024)

```
# TODO: Fit to Chicago Traffic model. Discuss posterior distribution curve.

# we've actually ended up creating latent categorization by
# crash severity, which is exactly what we are looking for
# we can think of this as "evaluated weekly crash reports"
# assuming that more severe crashes will be processed later
# or, more likely, the number of evaluated crash reports
# declines in groups over time as severity increases
# since more severe reports take longer to evaluate
```

Conclusion and Further Work

```
# TODO: Discuss Chicago Traffic Model and Implementation
```

Appendix

Acknowledgements: Special thanks to Prof. Lasantha Premarathna for sparking my interest in non-parametric statistics!

Miscellaneous information such as knowledge on sets, power sets, subsets and countability from (Demirbas and Rechnitzer 2023).

Example of a σ-Algebra

For additional clarity, we provide an example of a σ-algebra $\mathcal{F}$ to demonstrate the properties mentioned in the literature review.

Let's consider the finitely countable and simple set $\mathbb{X} = \{a, b, c\}$.

Directly, $P(\mathbb{X}) = \big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\big\}$ is the power set of $\mathbb{X}$, where we note $\{a, b, c\} = \mathbb{X}$. Consider $\mathcal{F} = \big\{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\big\} \subseteq P(\mathbb{X})$.

We will apply the properties discussed in the literature review section to prove that $\mathcal{F}$ is a σ-algebra.

To verify Property 1 (universality and non-emptiness), we note that we can also write $\mathcal{F}$ as $\big\{\emptyset, \{a\}, \{b, c\}, \mathbb{X}\big\}$. From this definition, it is direct to see that $\mathbb{X} \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$, verifying universality and non-emptiness.

To verify Property 2 (closure under complementation), we note that $A^{c^c} = A$. Hence, if we show $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ this is equivalent to showing that $A \in \mathcal{F} \iff A^c \in \mathcal{F}$. Since $\mathcal{F}$ is finitely countable, we can consider a case-wise basis for verification. Firstly, we have $A = \emptyset$. By definition, $A^c = \mathbb{X}$. We see that $\mathbb{X} \in \mathcal{F}$. As mentioned before, this implies that the case where $A = \mathbb{X}$ also holds. Now, we can proceed to verify the case where $A = \{a\}$. We note that $\{a\}^c = \{b, c\}$, and that $\{b, c\} \in \mathcal{F}$. Therefore, this case holds and thus so does the case where $A = \{b, c\}$ by biconditionality. Hence, we can conclude that $\mathcal{F}$ is closed under complementation.

To verify Property 3 (closure under countable unions), we first consider the concrete case where $\{A_i\}_{i=1}^{2} = \big\{\{a\}, \{b, c\}\big\}$ where $A_1 = \{a\}$ and $A_2 = \{b, c\}$. We note that $A_1, A_2 \in \mathcal{F}$, so we would expect that $A_1 \cup A_2 \in \mathcal{F}$. Directly, $A_1 \cup A_2 = \{a\} \cup \{b, c\} = \{a, b, c\} = \mathbb{X}$, and we see that $\mathbb{X} \in \mathcal{F}$. For the remaining cases, $\forall A \in \mathcal{F}, A \cup \emptyset = A$ by fundamental set properties, and directly $A \in \mathcal{F}$ by construction. Similarly, $\forall A \in \mathcal{F}, A \cup \mathbb{X} = \mathbb{X}$ and we know that $\mathbb{X} \in \mathcal{F}$. Hence, for all cases, $\mathcal{F}$ is closed under countable unions.

From all of these properties, we can conclude that $\mathcal{F}$ is a σ-algebra, as required $\square$.

Example of a Probability Measure

Using the σ-algebra $\mathcal{F} = \big\{\emptyset, \{a\}, \{b, c\}, \{a, b, c\}\big\}$ we will define $\mu : \mathcal{F} \mapsto [0, 1]$ and prove that μ is a probability measure using the properties discussed in the literature review.

We will define a concrete example as follows, and show it is a probability measure on $\mathcal{F}$.

$$\mu(A) = \begin{cases} 2/3, & |A| = 1 \\ 1/3, & |A| = 2 \\ 1, & |A| = 3 \\ 0, & \text{otherwise} \end{cases}$$

Let's evaluate the properties discussed in the literature review to verify that μ is in fact a probability measure.

First, we evaluate Property 1 (non-negativity.) In effect, we wish to verify that $\forall A \in \mathcal{F}, \mu(A) \geq 0$. We can easily evaluate the universal in a case-wise basis.

a. $\emptyset \in \mathcal{F}$ and $\mu(\emptyset) = 0 \geq 0$.
b. $\{a\} \in \mathcal{F}$ and $\mu(\{a\}) = 2/3 \geq 0$.
c. $\{b, c\} \in \mathcal{F}$ and $\mu(\{b, c\}) = 1/3 \geq 0$.
d. $\mathbb{X} \in \mathcal{F}$ and $\mu(\mathbb{X}) = 1 \geq 0$.

Hence, $\forall A \in \mathcal{F}, \mu(A) \geq 0$, verifying the non-negativity clause. Further, we see that $\forall A \in \mathcal{F}, 0 \leq \mu(A) \leq 1$.

In addition, we can utilize the evaluations above to verify Property 2. Directly, we see that $\mu(\emptyset) = 0$ and $\mu(\mathbb{X}) = 1$, verifying Property 2 that a value of 1 is assigned to the sample space $\mathbb{X}$.

Finally, we verify Property 3 (countable additivity.) Again, because the σ-algebra is finitely countable, we verify all pairwise disjoint intersections on a case-wise basis.

a. First, we consider the set of pairwise disjoint sets $\{\emptyset, A\}$ for $A \in \mathcal{F}$. We see that $\forall A \in \mathcal{F}, \emptyset \cap A = \emptyset$ and $\emptyset \cup A = A$. Using, $\mu(\emptyset) = 0$, we observe that $\mu\Big(\bigcup_i A_i\Big) = \mu(A) = \mu(A) + \mu(\emptyset) = \sum_{i=1} \mu(A_i)$, as required.
b. Similarly, we consider $\mathbb{X} \in \mathcal{F}$, noting that $\forall A \in \mathcal{F}, \mathbb{X} \cap A = A$ is non-disjoint, so the case holds vacuously by falsity of the antecedent. A similar argument can be applied for $\{A, A\}, A \in \mathcal{F}$ which is evidently a non-disjoint pair.
c. The other pairwise disjoint set in $\mathcal{F}$ is $\{A_i\} = \big\{\{a\}, \{b, c\}\big\}$, since $\{a\} \cap \{b, c\} = \emptyset$. Hence, this pair should be countably additive. We can see directly that $\bigcup_i A_i = \{a\} \cup \{b, c\} = \mathbb{X}$, Hence, $\mu\Big(\bigcup_i A_i\Big) = \mu(\mathbb{X}) = 1$, so we anticipate $\sum_i \mu(A_i) = 1$. To verify, we see that $\sum_i \mu(A_i) = \mu(\{a\}) + \mu(\{b, c\}) = 2/3 + 1/3 = 1$, so countable additivity holds.

From all of these properties, we can conclude that μ is a probability measure on σ-algebra $\mathcal{F}$, as required $\square$.

Results Code

```r
library(dirichletprocess)

# set seed for reproducibility
set.seed(447)
# start the clock
start_time = proc.time()
M = 10000
RUN = FALSE # TRUE if running sampler

###########################
### Mixing Distribution ###
###########################

# define the framework conjugate mixture model
poisMd <- MixingDistribution(
  distribution = "poisson",
  priorParameters = c(1, 1),
  conjugate = "conjugate"
)

# Part 1: Poisson Likelihood
Likelihood.poisson <- function(mdobj, x, theta){
  return(as.numeric(dpois(x, theta[[1]])))
}
# Part 2: Gamma Prior : Base Measure
PriorDraw.poisson <- function(mdobj, n){
  draws = rgamma(n, mdobj$priorParameters[1], mdobj$priorParameters[2])
  theta = list(array(draws, dim=c(1,1,n)))
  return(theta)
}
# Part 3: Posterior Draw by Conjugacy
PosteriorDraw.poisson <- function(mdobj, x, n=1){
  priorParameters = mdobj$priorParameters
  theta = rgamma(n, priorParameters[1] + sum(x),
                    priorParameters[2] + nrow(x))
  return(list(array(theta, dim=c(1,1,n))))
}

# Part 4: Predictive Distribution by Marginalization
Predictive.poisson <- function(mdobj, x){
  priorParameters = mdobj$priorParameters
  pred = numeric(length(x))
  for(i in seq_along(x)){
    alphaPost = priorParameters[1] + x[i]
    betaPost = priorParameters[2] + 1
    pred[i] = (priorParameters[2] ^ priorParameters[1]) / gamma(priorParameters[1])
    pred[i] = pred[i] * gamma(alphaPost) / (betaPost^alphaPost)
    pred[i] = pred[i] * (1 / prod(factorial(x[i])))
  }
  return(pred)
}
```

```r
############################
### D.P. Gibbs Sampling ###
############################

# read in cleaned data frame
df = read.csv("final_project/cleaned_crash_data.csv")
# monthly crash count, in 100s of crashes
y = ( round((df$crash_count)/100) )

# create DP Poisson Mixture Model from mix dist. defined earlier
dirp = DirichletProcessCreate(y, poisMd)

if(RUN){
  # initialize and fit DPMM via Gibbs
  dirp = Initialise(dirp)
  dirp = Fit(dirp, M)

  # compute, posterior frame: sampling from the posterior
  cat("Generating Posterior Frame...")
  # include 95% and 99% Credible Intervals
  postf = PosteriorFrame(dirp, 0:22, 1000, ci_size = c(0.1, 0.01))

  # save to avoid repeat simulation
  saveRDS(postf, file = "final_project/posterior_sampleframe.RDS")
  saveRDS(dirp, file =  "final_project/posterior_results.RDS")
}
# report runtime
total_time = proc.time() - start_time
cat("Total Runtime of Script: ", total_time['elapsed'], "seconds\n")
```

$$
\begin{aligned}
\{\pi_k\}_{k=1}^{K} &\sim \text{GEM}(\alpha),\ \text{stick-breaking process} \\
\{\theta_k\}_{k=1}^{K} &\sim \mathbb{G}_0,\ \text{base measure} \\
z_i &\sim \text{categorical}(\{1,2,\ldots,K\},\{\pi_k\}),\ \text{cluster assignment } i \in [1,N] \\
y_i \mid z_i, \{\theta_k\}_1^K &\sim \text{F}(x_i\,;\,\theta_{z_i}),\ \text{likelihood given cluster}
\end{aligned}
$$

Infinite Dirichlet Process Mixture Model

$$
\begin{aligned}
\{\pi_k\}_{k=1}^{\infty} &\sim \text{GEM}(\alpha) \\
\{\langle\alpha_k,\beta_k\rangle\}_{k=1}^{\infty} &\sim \mathbb{G}_0 \\
y_i \mid \{\pi_k\}, \{\langle\alpha_k,\beta_k\rangle\} &\sim \sum_{k=1}^{\infty} \pi_k \text{beta}(y_i \mid \langle\alpha_k,\beta_k\rangle)
\end{aligned}
$$

Where the MCMC implementation is facilitated by the `dirichletprocess` package, created by (Markwick 2023).

Finite-approximated infinite mixture DPMM, similar

$$
\begin{aligned}
\{\pi_k\}_{k=1}^{K} &\sim \text{GEM}(\alpha),\ \text{stick-breaking process} \\
\{\theta_k\}_{k=1}^{K} &\sim \mathbb{G}_0,\ \text{base measure} \\
z_i &\sim \text{categorical}(\{1,2,\ldots,K\},\{\pi_k\}),\ \text{cluster assignment } i \in [1,N] \\
y_i \mid z_i, \{\theta_k\}_1^K &\sim \text{F}(x_i\,;\,\theta_{z_i}),\ \text{likelihood given cluster}
\end{aligned}
$$

$$
\begin{aligned}
\{\pi_k\}_{k=1}^{\infty} &\sim \mathrm{GEM}(\alpha), \text{ stick-breaking process} \\
\{\theta_k\}_{k=1}^{\infty} &\sim \mathbb{G}_0, \text{ base measure,} \\
F &= \prod_{i=1}^{N}\left(\sum_{k=1}^{\infty}\pi_k N(x_i \mid \theta_k)\right), \text{ likelihood from normal kernel}
\end{aligned}
$$

Sources

(Jones 2021)

Billingsley, Patrick. 2012. *Probability and Measure, Anniversary Edition.* Wiley.

(CPD), Chicago Police Department. 2024. "Chicago Traffic Crashes - Chicago Police Dept." Kaggle. https://doi.org/10.34740/KAGGLE/DSV/7339559.

Demirbas, Seckin, and Andrew Rechnitzer. 2023. "An Introduction to Mathematical Proof : MATH 220." Free web and pdf textbook. https://personal.math.ubc.ca/~PLP/.

Ferguson, Thomas. 1973. "Bayesian Analysis of Some Nonparametric Problems." *Annals of Statistics* 1 (2): 209–30. https://doi.org/10.1214/aos/1176342360.

Hannah, Lauren A. 2011. "Dirichlet Process Mixtures of Generalized Linear Models." *Journal of Machine Learning Research* 12: 1923–53. https://www.jmlr.org/papers/volume12/hannah11a/hannah11a.pdf.

Jones, Andy. 2021. "Dirichlet Process Mixture Models." Online. https://andrewcharlesjones.github.io/journal/dpmm.html.

Markwick, Dean. 2023. *Dirichletprocess: An 'r' Package for Dirichlet Process Mixture Models.* https://dm13450.github.io/dirichletprocess/.

Rudin, Walter. 1986. *Real and Complex Analysis.* 3rd ed. McGraw-Hill.

Sethuraman, Jayaram. 1994. "A Constructive Definition of Dirichlet Priors." *Statistica Sinica.*

StackOverflow. 2023. "How to Use Custom Functions in Mutate (Dplyr)." Stack Overflow. https://stackoverflow.com/questions/44730774/how-to-use-custom-functions-in-mutate-dplyr.

Teh, Yee Whye. 2006. "Dirichlet Process." Course Notes for Gatsby Computational Neuroscience Unit Tutorial. https://mlg.eng.cam.ac.uk/zoubin/tut06/ywt.pdf.