An Analysis of Approaches to Presence-Only Data

William Fithian and Trevor Hastie
Department of Statistics
Stanford University

Eco-Stats, UNSW Sydney, July 2013
Species Distribution Modeling

Question: where may a given species be found?
Species Distribution Modeling

Question: where may a given species be found?

Motivations:
- Plan wildlife management actions
- Monitor endangered or invasive species
- Scientific understanding
- etc.
Species Distribution Modeling

Question: where may a given species be found?

Motivations:

• Plan wildlife management actions
• Monitor endangered or invasive species
• Scientific understanding
• etc.

What geographic features predict greater abundance?
Presence-Absence / Count Data

Scientists visit patch of land
Presence-Absence / Count Data

Scientists visit patch of land

Record whether any specimens encountered / how many
Presence-Absence / Count Data

Scientists visit patch of land

Record whether any specimens encountered / how many

Relatively high quality data
Presence-Absence / Count Data

Scientists visit patch of land

Record whether any specimens encountered / how many

Relatively high quality data

Expensive, difficult for rare or elusive species
Presence-Only Data

Motorist spies koala
Presence-Only Data

Motorist spies koala

Calls museum excitedly
Presence-Only Data

Motorist spies koala

Calls museum excitedly

Museum records location
Presence-Only Data

Motorist spies koala

Calls museum excitedly

Museum records location

Lower quality data
Presence-Only Data

Motorist spies koala

Calls museum excitedly

Museum records location

Lower quality data

More of it exists
Koala Sightings

Figure 1. Koala records (courtesy of New South Wales National Parks & Wildlife Service) and the road network on part of the New South Wales north coast.

Taken from Margules and Austen (1994)
Notation

\[ n_1 \text{ presence observations, } n_0 \text{ random background locations} \]
Notation

$n_1$ presence observations, $n_0$ random background locations

Geographic coordinates $z_i \in D \subseteq \mathbb{R}^2$, $i = 1, \ldots, n_0 + n_1$
Notation

\( n_1 \) presence observations, \( n_0 \) random background locations

Geographic coordinates \( z_i \in \mathcal{D} \subseteq \mathbb{R}^2, i = 1, \ldots, n_0 + n_1 \)

Features \( x_i = x(z_i) \) measured via geographic info systems (rainfall, temp., elevation, ...)

Notation

$n_1$ presence observations, $n_0$ random background locations

Geographic coordinates $z_i \in D \subseteq \mathbb{R}^2$, $i = 1, \ldots, n_0 + n_1$

Features $x_i = x(z_i)$ measured via geographic info systems (rainfall, temp., elevation, ...)

$y_i = 1$ for presence, 0 for background
Notation

$n_1$ presence observations, $n_0$ random background locations

Geographic coordinates $z_i \in \mathcal{D} \subseteq \mathbb{R}^2$, $i = 1, \ldots, n_0 + n_1$

Features $x_i = x(z_i)$ measured via geographic info systems (rainfall, temp., elevation, ...)

$y_i = 1$ for presence, 0 for background

Outline

1. Inhomogeneous Poisson Process Model

2. Maxent

3. Logistic Regression
Outline

1 Inhomogeneous Poisson Process Model

2 Maxent

3 Logistic Regression
Inhomogeneous Poisson Process

Intensity function

\[ \lambda(z) : \mathcal{D} \rightarrow [0, \infty) \]
Inhomogeneous Poisson Process

Intensity function

\[ \lambda(z) : \mathcal{D} \to [0, \infty) \]

\[ \Lambda(A) = \int_A \lambda(z) \, dz \]

Assume \( \Lambda(\mathcal{D}) < \infty \).
Inhomogeneous Poisson Process

Intensity function

\[ \lambda(z) : \mathcal{D} \rightarrow [0, \infty) \]

\[ \Lambda(A) = \int_A \lambda(z) \, dz \]

Assume \( \Lambda(\mathcal{D}) < \infty \).

\[ p_\lambda(z) = \frac{\lambda(z)}{\Lambda(\mathcal{D})} \]
Inhomogeneous Poisson Process

Intensity function

\[ \lambda(z) : \mathcal{D} \rightarrow [0, \infty) \]

\[ \Lambda(A) = \int_A \lambda(z) \, dz \]

Assume \( \Lambda(\mathcal{D}) < \infty \).

\[ p_\lambda(z) = \lambda(z) / \Lambda(\mathcal{D}) \]

**Definition 1:** random sample of random size

\[ n_1 \sim \text{Poisson}(\Lambda(\mathcal{D})) \]

\[ z_i \overset{\text{i.i.d.}}{\sim} p_\lambda \quad (y_i = 1) \]
Inhomogeneous Poisson Process

Intensity function

\[ \lambda(z) : \mathcal{D} \to [0, \infty) \]

\[ \Lambda(A) = \int_A \lambda(z) \, dz \]

Assume \( \Lambda(\mathcal{D}) < \infty \).

\[ p_{\lambda}(z) = \frac{\lambda(z)}{\Lambda(\mathcal{D})} \]

**Definition 1:** random sample of random size

\[ n_1 \sim \text{Poisson}(\Lambda(\mathcal{D})) \]

\[ z_i \text{ i.i.d.} \sim p_{\lambda} \quad (y_i = 1) \]

**Definition 2:** continuous version of discrete poisson model

\[ N(A) = \#\{z_i \in A : y_i = 1\} \]

\[ \sim \text{Poisson}(\Lambda(A)) \]

\( N(A_i) \) independent across disjoint \( A_i \)
Warton & Shepherd (2010) propose log-linear IPP for presence-only data
Warton & Shepherd (2010) propose log-linear IPP for presence-only data

\[ \lambda(z) = e^{\alpha + \beta' x(z)} \]
Warton & Shepherd (2010) propose log-linear IPP for presence-only data

\[ \lambda(z) = e^{\alpha + \beta' x(z)} \]

\[ p\lambda(z) = \frac{e^{\beta' x(z)}}{\int_D e^{\beta' x(z)} \, dz} \]
Warton & Shepherd (2010) propose log-linear IPP for presence-only data

\[
\lambda(z) = e^{\alpha + \beta' x(z)}
\]

\[
p_\lambda(z) = \frac{e^{\beta' x(z)}}{\int_{D} e^{\beta' x(z)} \, dz}
\]

\(\beta\) determines \(p_\lambda\)
Warton & Shepherd (2010) propose log-linear IPP for presence-only data

\[ \lambda(z) = e^{\alpha + \beta' x(z)} \]

\[ p_\lambda(z) = \frac{e^{\beta' x(z)}}{\int_D e^{\beta' x(z)} \, dz} \]

\( \beta \) determines \( p_\lambda \)

\( \alpha \) determines \( \Lambda(D) \)
Identifiability and Observer Bias

Occurrence process of scientific interest
Identifiability and Observer Bias

Occurrence process of scientific interest

Presence-only data reflect rate of sightings
Identifiability and Observer Bias

Occurrence process of scientific interest

Presence-only data reflect rate of *sightings*

Observation process is thinned occurrence process

\[ \lambda_{\text{obs}}(z) = \lambda_{\text{occ}}(z)s(z) \]
\[ e^{\alpha + \beta' x(z)} = e^{\tilde{\alpha} + \tilde{\beta}' x(z)} e^{\gamma + \delta' x(z)} \]
Identifiability and Observer Bias

Occurrence process of scientific interest

Presence-only data reflect rate of *sightings*

Observation process is thinned occurrence process

\[
\lambda_{\text{obs}}(z) = \lambda_{\text{occ}}(z)s(z)
\]

\[
e^{\alpha + \beta' x(z)} = e^{\tilde{\alpha} + \tilde{\beta}' x(z)} e^{\gamma + \delta' x(z)}
\]

Options:

1. Assume \( s \) is constant (optimistic)
Identifiability and Observer Bias

Occurrence process of scientific interest

Presence-only data reflect rate of *sightings*

Observation process is thinned occurrence process

\[ \lambda_{\text{obs}}(z) = \lambda_{\text{occ}}(z) s(z) \]
\[ e^{\alpha+\beta'x(z)} = e^{\tilde{\alpha}+\tilde{\beta}'x(z)} e^{\gamma+\delta'x(z)} \]

Options:

1. Assume $s$ is constant (optimistic)
2. Assume $s$ and $\lambda_{\text{occ}}$ depend on different features
Identifiability and Observer Bias

Occurrence process of scientific interest

Presence-only data reflect rate of sightings

Observation process is thinned occurrence process

\[ \lambda_{\text{obs}}(z) = \lambda_{\text{occ}}(z) s(z) \]

\[ e^{\alpha + \beta' x(z)} = e^{\tilde{\alpha} + \tilde{\beta}' x(z)} e^{\gamma + \delta' x(z)} \]

Options:

1. Assume \( s \) is constant (optimistic)
2. Assume \( s \) and \( \lambda_{\text{occ}} \) depend on different features

Either way, \( \tilde{\alpha} \) unidentifiable (\( \alpha = \gamma + \tilde{\alpha} \))
Occurrence Probability versus Occurrence Rate

Probability of what event?
Occurrence Probability versus Occurrence Rate

Probability of what event?

Of seeing one (or more) member of species, in a quadrat of size $A$, if observed for time $T$?
Occurrence Probability versus Occurrence Rate

Probability of what event?

Of seeing one (or more) member of species, in a quadrat of size $A$, if observed for time $T$?

Occurrence rate measures the expected number of species (seen) *per unit area*, if observed for time $T$. 
Occurrence Probability versus Occurrence Rate

Probability of what event?

Of seeing one (or more) member of species, in a quadrat of size $A$, if observed for time $T$?

Occurrence rate measures the expected number of species (seen) *per unit area*, if observed for time $T$.

The IPP occurrence rate uses one less unit, and the PO sampling process seems more aligned with that assumed by an IPP.
Maximum Likelihood for IPP

Log-likelihood

\[ \ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_D e^{\alpha + \beta' x(z)} \, dz \]
Maximum Likelihood for IPP

Log-likelihood

\[ \ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_D e^{\alpha + \beta' x(z)} \, dz \]

Score equation for \( \alpha \):

\[ n_1 = \Lambda(D) \]
Maximum Likelihood for IPP

Log-likelihood

\[
\ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_{\mathcal{D}} e^{\alpha + \beta' x(z)} \, dz
\]

Score equation for \(\alpha\):

\[
n_1 = \Lambda(\mathcal{D})
\]

For \(\beta\):

\[
\frac{1}{n_1} \sum_{y_i=1} x_i = \mathbb{E}_{p_\lambda} x(z)
\]
Maximum Likelihood for IPP

Log-likelihood

\[ \ell(\alpha, \beta) = \sum_{y_i=1} (\alpha + \beta' x_i) - \int_D e^{\alpha + \beta' x(z)} \, dz \]

Score equation for \( \alpha \):

\[ n_1 = \Lambda(D) \]

For \( \beta \):

\[ \frac{1}{n_1} \sum_{y_i=1} x_i = \mathbb{E}_{p_x} x(z) \]

Interpretation:

1. Choose \( \hat{\beta} \) to match means of features \( x(z) \)
Maximum Likelihood for IPP

Log-likelihood

\[ \ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_{\mathcal{D}} e^{\alpha + \beta' x(z)} \, dz \]

Score equation for \( \alpha \):

\[ n_1 = \Lambda(\mathcal{D}) \]

For \( \beta \):

\[ \frac{1}{n_1} \sum_{y_i=1} x_i = \mathbb{E}_{p_{\lambda}} x(z) \]

Interpretation:

1. Choose \( \hat{\beta} \) to match means of features \( x(z) \)
2. Choose \( \hat{\alpha} \) so \( \Lambda(\mathcal{D}) = n_1 \)
Maximum Likelihood for IPP

Log-likelihood

$$\ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_D e^{\alpha + \beta' x(z)} \, dz$$

Score equation for $\alpha$:

$$n_1 = \Lambda(D)$$

For $\beta$:

$$\frac{1}{n_1} \sum_{y_i=1} x_i = \mathbb{E}_{p_{\lambda}} x(z)$$

Interpretation:

1. Choose $\hat{\beta}$ to match means of features $x(z)$
2. Choose $\hat{\alpha}$ so $\Lambda(D) = n_1$

1. Estimate density.
Maximum Likelihood for IPP

Log-likelihood

\[ \ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \int_D e^{\alpha + \beta' x(z)} \, dz \]

Score equation for \( \alpha \):

\[ n_1 = \Lambda(\mathcal{D}) \]

For \( \beta \):

\[ \frac{1}{n_1} \sum_{y_i=1} x_i = \mathbb{E}_{p_\lambda} x(z) \]

Interpretation:

1. Choose \( \hat{\beta} \) to match means of features \( x(z) \)
2. Choose \( \hat{\alpha} \) so \( \Lambda(\mathcal{D}) = n_1 \)

1. Estimate density. 2. Multiply by \( n_1 \).
Numerical Approximation of IPP Likelihood

In practice, can’t evaluate integrals analytically
Numerical Approximation of IPP Likelihood

In practice, can’t evaluate integrals analytically

Use background points for numerical approximation

$$\ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \frac{|D|}{n_0} \sum_{y_\ell=0} e^{\alpha + \beta' x_\ell}$$
Numerical Approximation of IPP Likelihood

In practice, can’t evaluate integrals analytically

Use background points for numerical approximation

\[ \ell(\alpha, \beta) = \sum_{y_i=1} \alpha + \beta' x_i - \frac{|D|}{n_0} \sum_{y_\ell=0} e^{\alpha + \beta' x_\ell} \]

Same interpretation of score equations
Outline

1. Inhomogeneous Poisson Process Model
2. Maxent
3. Logistic Regression
Maxent

Phillips et al. (2004) model presence points as $z_i \overset{\text{i.i.d.}}{\sim} p(z)$.
Maxent

Phillips et al. (2004) model presence points as $z_i \overset{i.i.d.}{\sim} p(z)$

Maximize $H(p) = - \int p(z) \log p(z) \, dz$ subject to

$$\frac{1}{n_1} \sum_{y_i=1} x(z_i) = \mathbb{E}_p x(z)$$

*Maximum Entropy* makes $p(z)$ as uniform as possible.
Maxent

Phillips et al. (2004) model presence points as \( z_i \overset{\text{i.i.d.}}{\sim} p(z) \)

Maximize \( H(p) = - \int p(z) \log p(z) \, dz \) subject to

\[
\frac{1}{n_1} \sum_{y_i=1} x(z_i) = \mathbb{E}_p x(z)
\]

*Maximum Entropy* makes \( p(z) \) as uniform as possible.

Authors show solution has parametric form:

\[
z_i \overset{\text{i.i.d.}}{\sim} p(z) = \frac{e^{\beta' x(z)}}{\int e^{\beta' x(u)} \, du}
\]
Maxent

Phillips et al. (2004) model presence points as $z_i \sim \mathcal{N}(p(z))$

Maximize $H(p) = -\int p(z) \log p(z) \, dz$ subject to

$$\frac{1}{n_1} \sum_{y_i=1}^{n_1} x(z_i) = \mathbb{E}_p x(z)$$

**Maximum Entropy** makes $p(z)$ as uniform as possible.

Authors show solution has parametric form:

$$z_i \sim \mathcal{N}(p(z)) = \frac{e^{\beta' x(z)}}{\int e^{\beta' x(u)} \, du}$$

Aarts et al. (2012), FH (2013):

Exactly same form as conditional IPP, with same estimating equations! Hence same slopes $\hat{\beta}$ as IPP
Maxent in Practice

- Uses a sample of background points
Maxent in Practice

- Uses a sample of background points
  
  *IPP can too, as in our numerical MLE of IPP.*
Maxent in Practice

- Uses a sample of background points
  *IPP can too, as in our numerical MLE of IPP.*
- Uses derived features (polynomials in $x(z)$, hinge functions, etc).
Maxent in Practice

- Uses a sample of background points
  *IPP can too, as in our numerical MLE of IPP.*
- Uses derived features (polynomials in $x(z)$, hinge functions, etc).
  *Can do the same with IPP*
Maxent in Practice

- Uses a sample of background points
  *IPP can too, as in our numerical MLE of IPP.*
- Uses derived features (polynomials in $x(z)$, hinge functions, etc).
  *Can do the same with IPP*
- Imposes $\ell_1$ (lasso) bounds on parameters
Maxent in Practice

- Uses a sample of background points
  *IPP can too, as in our numerical MLE of IPP.*
- Uses derived features (polynomials in $x(z)$, hinge functions, etc).
  *Can do the same with IPP*
- Imposes $\ell_1$ (lasso) bounds on parameters
  *Can do the same with IPP*
Maxent in Practice

- Uses a sample of background points
  *IPP can too, as in our numerical MLE of IPP.*
- Uses derived features (polynomials in $x(z)$, hinge functions, etc).
  *Can do the same with IPP*
- Imposes $\ell_1$ (lasso) bounds on parameters
  *Can do the same with IPP*

Paradigm: enrich linear model via basis expansions, and then regularize coefficients to control variance inflation
Geographic versus Environmental models

IPP and Maxent model the density of the locations $z$ for presence sites:

$$p_\lambda(z) \propto e^{\beta' x(z)}.$$
Geographic versus Environmental models

IPP and Maxent model the density of the locations $z$ for presence sites:

$$p_\lambda(z) \propto e^{\beta' x(z)}.$$ 

Can represent this as a model for the density of features $x = x(z)$ for presence sites:

$$f_1(x) \propto h(x)e^{\beta' x(z)}$$

where $h(x)$ is the marginal distribution of the environmental features (over the whole domain).
Geographic versus Environmental models

IPP and Maxent model the density of the locations \( z \) for presence sites:

\[
p_\lambda(z) \propto e^{\beta' x(z)}.
\]

Can represent this as a model for the density of features \( x = x(z) \) for presence sites:

\[
f_1(x) \propto h(x) e^{\beta' x(z)}
\]

where \( h(x) \) is the marginal distribution of the environmental features (over the whole domain).

Since by Bayes rule

\[
\Pr(\text{Presence at } z| x(z) = x) = \frac{f_1(x) \pi_1}{h(x)}
\]

where \( \pi_1 \) is the overall prevalence, \( e^{\beta' x(z)} \) measures the presence probability up to a constant \( \pi_1 \) (Elith et al, 2011).
Outline

1. Inhomogeneous Poisson Process Model
2. Maxent
3. Logistic Regression
“Naive” Logistic Regression

Presence-only modeling as classification
“Naive” Logistic Regression

Presence-only modeling as classification

Treat $x_i$ as fixed, presence/background $y_i$ as random, and assume:

$$y_i | x_i \sim \text{Bernoulli} \left( \frac{e^{\eta + \beta' x_i}}{1 + e^{\eta + \beta' x_i}} \right)$$
Presence-only modeling as classification

Treat $x_i$ as fixed, presence/background $y_i$ as random, and assume:

$$y_i | x_i \sim \text{Bernoulli} \left( \frac{e^{\eta+\beta'x_i}}{1 + e^{\eta+\beta'x_i}} \right)$$

Flexible modeling framework: GAM, MARS, boosting, LASSO, etc.
Case-Control Sampling

Back to IPP Model $\lambda(z) = e^{\alpha + \beta'x(z)}$ and $p_\lambda(z) \propto e^{\beta'x(z)}$.

Consider mixture of $n_1$ presence samples, and $n_0$ uniform background samples. Using Bayes rule, can show that

$$\mathbb{P}(y_i = 1|z_i) = \frac{e^{\eta + \beta'x_i}}{1 + e^{\eta + \beta'x_i}}$$

where $\eta$ is a constant that depends on $n_1$, $n_2$, $|D|$, $\alpha$ and more.
Case-Control Sampling

Back to IPP Model $\lambda(z) = e^{\alpha + \beta'x(z)}$ and $p_\lambda(z) \propto e^{\beta'x(z)}$.

Consider mixture of $n_1$ presence samples, and $n_0$ uniform background samples. Using Bayes rule, can show that

$$P(y_i = 1|z_i) = \frac{e^{\eta + \beta'x_i}}{1 + e^{\eta + \beta'x_i}}$$

where $\eta$ is a constant that depends on $n_1$, $n_2$, $|D|$, $\alpha$ and more.
Case-Control Sampling

Back to IPP Model $\lambda(z) = e^{\alpha + \beta'x(z)}$ and $p_\lambda(z) \propto e^{\beta'x(z)}$.

Consider mixture of $n_1$ presence samples, and $n_0$ uniform background samples. Using Bayes rule, can show that

$$P(y_i = 1|z_i) = \frac{e^{\eta + \beta'x_i}}{1 + e^{\eta + \beta'x_i}}$$

where $\eta$ is a constant that depends on $n_1$, $n_2$, $|D|$, $\alpha$ and more.

“Case-control” sampling design
Case-Control Sampling

Back to IPP Model $\lambda(z) = e^{\alpha + \beta'x(z)}$ and $p_\lambda(z) \propto e^{\beta'x(z)}$.

Consider mixture of $n_1$ presence samples, and $n_0$ uniform background samples. Using Bayes rule, can show that

$$P(y_i = 1 | z_i) = \frac{e^{\eta + \beta'x_i}}{1 + e^{\eta + \beta'x_i}}$$

where $\eta$ is a constant that depends on $n_1, n_2, |D|, \alpha$ and more.

“Case-control” sampling design

Logistic regression as density estimation
Logistic Regression vs IPP

If linear IPP model is correct (!!), then both are estimating same $\beta$, but get different $\hat{\beta}$.
Logistic Regression vs IPP

If linear IPP model is correct (!!), then both are estimating same $\beta$, but get different $\hat{\beta}$

Warton & Shepherd (2010) show $\hat{\beta}_{LR} \rightarrow \hat{\beta}_{IPP}$ as $n_0 \rightarrow \infty$ with $n_1$ fixed
Logistic Regression vs IPP

If linear IPP model is correct (!!), then both are estimating same $\beta$, but get different $\hat{\beta}$

Warton & Shepherd (2010) show $\hat{\beta}_{LR} \rightarrow \hat{\beta}_{IPP}$ as $n_0 \rightarrow \infty$ with $n_1$ fixed

Not true if $n_0, n_1 \rightarrow \infty$ together
Logistic Regression vs IPP

If linear IPP model is correct (!!), then both are estimating same $\beta$, but get different $\hat{\beta}$

Warton & Shepherd (2010) show $\hat{\beta}_{LR} \rightarrow \hat{\beta}_{IPP}$ as $n_0 \rightarrow \infty$ with $n_1$ fixed

Not true if $n_0, n_1 \rightarrow \infty$ together

If linear model an approximation (i.e. as in always!), limiting $\hat{\beta}_{LR}$ depends on limiting ratio $n_1/n_0$
Logistic Regression vs IPP

If linear IPP model is correct (!!), then both are estimating same \( \beta \), but get different \( \hat{\beta} \)

Warton & Shepherd (2010) show \( \hat{\beta}_{LR} \to \hat{\beta}_{IPP} \) as \( n_0 \to \infty \) with \( n_1 \) fixed

Not true if \( n_0, n_1 \to \infty \) together

If linear model an approximation (i.e. as in always!), limiting \( \hat{\beta}_{LR} \) depends on limiting ratio \( n_1/n_0 \)

\( n_1 \) large \( \Rightarrow \) may need very large \( n_0 \)
Logistic Regression vs IPP

Fixed presence sample, \( n_1 = 1000 \). True \( \lambda \) quadratic in \( x \)
Weighted Logistic Regression

Don’t really need \( n_0 \to \infty \)
Weighted Logistic Regression

Don’t really need $n_0 \to \infty$

Weight sample to reflect undersampling of background points
Weighted Logistic Regression

Don’t really need \( n_0 \rightarrow \infty \)

Weight sample to reflect undersampling of background points

\[
    w_i = \begin{cases} 
        W & \text{if } y_i = 0 \\
        1 & \text{if } y_i = 1 
    \end{cases}
\]
Weighted Logistic Regression

Don’t really need $n_0 \to \infty$

Weight sample to reflect undersampling of background points

$$w_i = \begin{cases} W & y_i = 0 \\ 1 & y_i = 1 \end{cases}$$

As $W \to \infty$, $\hat{\beta}_{WLR} \to \hat{\beta}_{IPP}$
Weighted Logistic Regression

Don’t really need \( n_0 \to \infty \)

Weight sample to reflect undersampling of background points

\[
 w_i = \begin{cases} 
 W & y_i = 0 \\
 1 & y_i = 1 
\end{cases}
\]

As \( W \to \infty \), \( \hat{\beta}_{WLR} \to \hat{\beta}_{IPP} \)

Weighted logistic regression = numerical IPP = numerical Maxent
Weighted Logistic Regression

Don’t really need $n_0 \to \infty$

Weight sample to reflect undersampling of background points

$$w_i = \begin{cases} W & y_i = 0 \\ 1 & y_i = 1 \end{cases}$$

As $W \to \infty$, $\hat{\beta}_{WLR} \to \hat{\beta}_{IPP}$

Weighted logistic regression = numerical IPP = numerical Maxent

Implication: can fit IPP model via weighted logistic regression
Weighted Logistic Regression

Don’t really need $n_0 \to \infty$

Weight sample to reflect undersampling of background points

$$w_i = \begin{cases} W & y_i = 0 \\ 1 & y_i = 1 \end{cases}$$

As $W \to \infty$, $\hat{\beta}_{WLR} \to \hat{\beta}_{IPP}$

Weighted logistic regression = numerical IPP = numerical Maxent

Implication: can fit IPP model via weighted logistic regression / weighted poisson glm
Weighted vs Unweighted Logistic Regression

Weighted LR converges faster to large-$n_0$ limit.

Weighted and Unweighted Estimates for Logistic Regression
Conclusions

MaxEnt and logistic regression can be derived from IPP, same $\beta$. 

```r
boosted.ipp <- gbm(y ~ ., family="bernoulli",
data=banksia, weights=1000^(1-y))
lasso.ipp <- glmnet(x[,1:100],y, family="binomial",
data=banksia, weights=1000^(1-y))
```
Conclusions

MaxEnt and logistic regression can be derived from IPP, same $\beta$

$\hat{\beta}$ for IPP / MaxEnt may be fit by weighted logistic regression
MaxEnt and logistic regression can be derived from IPP, same \( \hat{\beta} \)

\( \hat{\beta} \) for IPP / MaxEnt may be fit by weighted logistic regression/
GAM / Boosted Trees / MARS / Group LASSO / ...
MaxEnt and logistic regression can be derived from IPP, same $\hat{\beta}$ for IPP / MaxEnt may be fit by weighted logistic regression/ GAM / Boosted Trees / MARS / Group LASSO / ...

boosted.ipp <- gbm(y ~ ., family="bernoulli", 
                   data=banksia, weights=1000^(1-y))

lasso.ipp <- glmnet(x[,1:100],y, family="binomial", 
                   data=banksia, weights=1000^(1-y))
Thanks