



黒木玄 Gen Kuroki

@genkuroki

statmodeling.hatenablog.com/en...

からリンクされている

statmodeling.hatenablog.com/en...

も素晴らしいです。

後者の(2)の例を解説します。その話は

slideshare.net/mobile/simizu70...

の36ページ以降の話でもあります。

プロ野球で各選手ごとの年間ホームラン数を集計したときの分布のモデルを作る話です。

続く

2017年05月30日 22:40 · Web · 🔄 0 · ★ 2 · Webで開く



黒木玄 Gen Kuroki @genkuroki

on May 30

まず、1人の選手が打席でホームランを打つことはまれであり、ある一定の確率でホームランを打つと仮定すれば、各選手ごとに年間ホームラン数はPoisson分布になっていると考えることは最初のたたき台としての悪くないでしょう。

しかし、プロ野球の各選手ごとに年間ホームラン数を集計して、ヒストグラムを描くと、Poisson分布とは似ても似つかない形になります。

それは当たり前で、すべての選手の年間ホームラン数の期待値が同じなら、ヒストグラムはPoisson分布のそれに近くなるかもしれませんが、実際には各選手ごとに年間ホームラン数の期待値は違うので、選手の個性をモデルに組み込む必要があります。

続く



黒木玄 Gen Kuroki @genkuroki

on May 30

そこで

slideshare.net/mobile/simizu70...

statmodeling.hatenablog.com/en...

では、各選手ごとにPoisson分布に従う年間ホームラン数の期待値がパラメーター入りのガンマに従うと仮定して、推定を行なっています。

欲しい予測分布はサンプルから漏れている選手の年間ホームラン数の予測分布です。

前者のリンク先では最尤法とベイズ推定法の2つの方法で推定してモデルの整合性を確認しています。単純なPoisson分布のモデルよりも、階層化して選手の個性を考慮したモデルの方が、AICやWAICの指標で測った予測精度が改善されることが確認されています。

続く



黒木玄 Gen Kuroki @genkuroki

on May 30

個性を考慮したモデルは、表に見えて来ない個性のパラメーターで積分してしまえば負の二項分布になります。

そうせずに、あえて階層ベイズのスタイルで推定を行うには、Stanというソフトにモデルの設定を教えてあげて計算してもらう方法があります。

プロ野球選手のサンプルサイズが N 人ならば、個性の分布を記述するガンマ分布のパラメータ ϕ, μ の他に、選手 N 人分のパラメーター r_n ($n = 1, 2, \dots, N$) を使うことを教えてあげ、さらに1人あたりのホームラン数を意味する確率変数 Y を

$$Y \sim \text{poisson}(r),$$

$$r \sim \text{gamma}(\phi, \phi/\mu)$$

とモデル化することも教えてあげます。

詳しい内容は

statmodeling.hatenablog.com/en...

の前半の(2)のStanコードを見て下さい。

続く



黒木玄 Gen Kuroki @genkuroki

on May 31

続き

ここで誤解しやすそうな点を指摘しておきます。

内部で予測分布を計算するためだけに使われる個性を意味するパラメーターたち r_n ($n = 1, \dots, N$) と事前分布を与えておくことが必要なパラメーター ϕ, μ (コードの中では `theta[1,2]`) では立場が違います。

WAICに関する漸近論はサンプル数と無関係にモデルが決まっているような場合にしか使えません。特にモデルのパラメーター数は一定でなければいけない。

そして、サンプルの構成要素の個性を考えるモデルではサンプルサイズに比例した内部計算用のパラメーターを用意することになる。

こういう理由でWAICの漸近論は個性を考慮するモデル(より一般に階層ベイズの典型的場合)には使用できないと誤解する危険性があると思いました。

もちろん、そんなことは皆無なわけです。

続く



黒木玄 Gen Kuroki @genkuroki

on May 31

statmodeling.hatenablog.com/en...

に引用されているbaibai氏の

【グループ数もサンプルサイズも比例して同時に増える】

場合の計算のリクエストと、そこでは引用されていない渡辺澄夫さんの仕事に関する

[twitter.com/ibaibabaibai/statu...](https://twitter.com/ibaibabaibai/status/874111111111111111)

【そもそも漸近論ですけど、階層ベイズとかはパラメータ数もサンプルサイズも同時に増える極限でそういう極限にはならない。】

(リンク先には不快で軽蔑的な発言があるので注意)

を見ると、私による誤解の可能性への注意は現実のものだと思われず。

専門外の話については私も誤解をよくするのですが、社会的に専門家とみなされている人が誤解に基づいて同業者を「宗教系」と呼ぶというのはちょっとすごいと思いました。

世の中には色々な人がいるものです。

mathtod.online powered by [Mastodon](#)