黒木玄 **Gen Kuroki** @genkuroki Sunday at 8:45am
大数の法則、中心極限定理、Kullback-Leibler情報量とSanovの定理が確率論に関する基本的な教養だろうという話
mathtod.online/@genkuroki/2104...

二項分布などの正規分布での近似はPoisson分布の正規分布による近似を経由すると楽になる話
mathtod.online/@genkuroki/2137...

Lindebergによるモーメント母函数も特性函数も使わずにテイラーの定理だけで中心極限定理を証明する方法の紹介
mathtod.online/@genkuroki/2178...

は連続する記事のつもりです。

私はずっとKL情報量が分布の違いを測る自然な指標になっていることの理解の重要性を強調して来ました。

続く

黒木玄 **Gen Kuroki** @genkuroki Sunday at 9:07am
多項分布の中心極限定理の計算をPoisson分布に関する中心極限定理を経由して計算すると楽できれいな計算が可能になるのですが、さらに本質的な計算の仕方は、多項分布からKL情報量を出す計算をPoisson分布経由で行い、その計算の帰結として、多項分布の中心極限定理を導くことです。

中心極限定理を中心だと思って計算するのではなく、KL情報量の方がより根源的な量だと思って計算した方が計算がきれいにまとまることが多いと思う。

KL情報量のビルディングブロックは函数 $x\log x$ で中心極限定理のビルディングブロックは $y^2/2$ なのですが、それらの関係は

$$x = 1 + y/\sqrt{n}$$

のとき

$$nx\log x = \sqrt{n}\,y + \frac{y^2}{2} + o(1)$$

です。$\sqrt{n}\,y$ の項が「和」で消える仕組みで多項分布の中心極限定理が導かれる。

黒木玄 **Gen Kuroki** @genkuroki Sunday at 9:28am
一般の中心極限定理をKL情報量の立場から見直す仕事もされています。

X_k は独立同分布確率変数列だとします。(同じルーレットを回して得られる乱数列のこと。) 簡単のため、各々、平均は0で分散は1だとし、

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

とおきます。Z_n の平均と分散も0と1になります。標準正規分布はこの変換の不動点になっています。

Z_n の確率密度函数を $p_n(x)$ と書き、標準正規分布の確率密度函数を $q(x) = e^{-x^2/2}/\sqrt{2\pi}$ と書き、それらのKL情報量を

$$D_n = D(p_n \| q) = \int p_n(x) \log \frac{p_n(x)}{q(x)}\, dx$$

と書き、これが有限の値になると仮定しておきます。

中心極限定理は Z_n の分布が標準正規分布に「近付く」という主張です。

ここまで来れば、誰でも、標準正規分布と Z_n の分布の違いを測るKL情報量 D_n は単調に 0 に収束しているだろうと予想できるでしょう！

続く

黒木玄 **Gen Kuroki** @genkuroki Sunday at 9:44am

続き

情報量の -1 倍がエントロピーなので、KL情報量の単調減少は相対エントロピーの単調増加と同値です。

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

の標準正規分布に対する相対エントロピーは単調増加しながら、0 に収束するだろうと予想される。

その方向への最初の決定的結果は、「次元の呪い」のニューラルネットワークによる克服の論文でも有名なBarronさんによるKL情報量 D_{2^n} は単調減少しながら 0 に収束するという結果です。倍々で増やして行けば単調減少する。
statistics.stanford.edu/sites/...
stat.yale.edu/~arb4/publicatio...
より最近の論文には次がある:
arxiv.org/abs/math/0111020

続く

黒木玄 **Gen Kuroki** @genkuroki Sunday at 9:57am

倍々で行かなくてもKL情報量 D_n 自身が単調減少していることの証明は次の論文にあります。
ams.org/journals/jams/2004-17-...

要するに、中心極限定理の文脈では、適切な設定のもとで、エントロピーの単調増加が証明されているということです。

注意: D_n は以下のように表される:

$$
\begin{aligned}
D_n &= \int p_n(x)\log p_n(x)\,dx \\
&\quad + \int p_n(x)\left(\frac{x^2}{2} + \log\sqrt{2\pi}\right)dx \\
&= \int p_n(x)\log p_n(x)\,dx \\
&\quad + \frac{1}{2} + \log\sqrt{2\pi}.
\end{aligned}
$$

この文脈では第1項の -1 倍をエントロピーと呼びたい(「情報量」ではなく)。

黒木玄 **Gen Kuroki**
@genkuroki

mathtod.online/@genkuroki/2178…
で簡単に紹介した中心極限定理のTaylorの定理しか使わない証明法には

Lindeberg exchange method
Lindeberg replacement trick

のような名前が付けられています。

関連の解説へのリンク

Tao, Terence.
DIscrite random matrices and universality
slide
terrytao.files.wordpress.com/2…

Random matrices: The Universality phenomenon for Wigner ensembles
Terence Tao, Van Vu
arxiv.org/abs/1202.0068

2017年06月04日 23:18 · Web · 0 · 1 · Webで開く

mathtod.online powered by Mastodon