スパイク時系列のヒストグラム作成における最適区間幅決定のレシピ
A Recipe for Making a Time Histogram with an Optimal Bin Width from Spike Sequences

島崎　秀昭 (PY), 篠本　滋
Hideaki Shimazaki(PY) and Shigeru Shinomoto
京都大学理学研究科　物理学・宇宙物理学専攻
hideaki@ton.scphys.kyoto-u.ac.jp

Abstract— Time histograms are widely used to quantify the rate of discharges of neuronal spike trains. Nevertheless ad hoc bin width of the histogram is often used in literatures. We give a recipe to determine an optimal bin width of a time histogram of spike sequences. The optimal bin width diverges for a small number of trials, and scales differently for a large number of trials, depending on the smoothness of the underlying rate. We thus provide the method to estimate the minimum number of spike sequences to construct a meaningful time histogram as well as the exponent of the scaling relation.

Keywords— **Peristimulus Time Histogram, Poisson Process, Mean Integrated Squared Error**

1　序論

電気神経生理学の動物実験では感覚刺激・行動・注意等と神経細胞の発火頻度（レート）の相関関係がよく調べられる. 広く使われているレート推定の手法に, 同一刺激下で行われた複数回の試行のスパイク時系列を適当な時間幅をもつ区間に分割し, その中でのイベント生成率（発火率）を棒グラフとして表す Peristimulus Time Histogram (PSTH) がある. PSTH の形状は分割する区間の時間幅に依存するにもかかわらず, 区間幅は多くの場合研究者により恣意的に与えられている. 小山・篠本はヒストグラムによる時間依存 Poisson 過程のレート推定において最適な区間幅の理論値を導出し, 最適区間幅が発散する場合があることを示した [2]. 従って実験データから最適区間幅を導出する手順や有限区間幅を得るのに必要な最小試行数の推定方法などが求められる.

あまり知られていないが Rudemo は確率分布推定の枠組みで最適な区間幅を求める式を示している [1]. ここではこの先行研究と異なるコスト関数を導入し, スパイク統計から最適区間幅を決定する簡便な公式を導出する. またレートが一般の定常確率過程の場合の最適区間幅の理論値のスケーリング則を導出し, 転移点近傍での振る舞いを調べた. この理論の応用としてコスト関数を外挿することで, データからヒストグラム作成に必要な最小試行数を求める手法, および背後のレート過程がなめらかな過程か否かを推定する手法を提案する.

2　方法

2.1　スパイク時系列からのヒストグラムの作成

長さ T の時間依存 Poisson 過程のレート（強度過程）を λ_t ($t \in [0, T]$) とする. 区間 $[0, \Delta]$ の棒ヒストグラムの真の高さは

$$\Lambda = \frac{1}{\Delta}\int_0^{\Delta} \lambda_t dt. \tag{1}$$

で与えられる. この区間内の n 回の試行数の総スパイク数 k は次の Poisson 分布で与えられる.

$$p(k|n\Lambda\Delta) = \frac{(n\Lambda\Delta)^k}{k!} e^{-n\Lambda\Delta} \tag{2}$$

従って Λ の不偏推定量である $\hat{\Lambda} = k/(n\Delta)$ がデータから求められるヒストグラムの高さである.

2.2　平均積分二乗誤差及びコスト関数の導入

スパイク時系列のレート λ_t とヒストグラムの当てはまりの良さは平均積分二乗誤差 (Mean Integrated Squared Error, MISE) で評価する. 十分長い定常なスパイク時系列が与えられた場合 MISE は次式で与えられる.

$$\mathcal{E} = \left\langle E_\Lambda \left[\frac{1}{\Delta}\int_0^{\Delta} (\,\lambda_t - \hat{\Lambda}\,)^2 dt \right] \right\rangle \tag{3}$$

ここで $\langle\cdot\rangle$ はレート過程 λ_t の経路によるアンサンブル平均を意味し, $E_\Lambda[\cdot]$ はレート λ_t の区間幅 Δ 内での時間平均が Λ である場合のスパイク数の条件付き確率分布 (式 2) による平均操作を表す.

MISE を区間幅 Δ 内でのレートのゆらぎとスパイク生成のゆらぎに分割し, さらに区間幅 Δ の選択に依らない項を除いたコスト関数を導入することができる.

$$\begin{aligned} C(\Delta) &\equiv \mathcal{E} - \left\langle (\lambda_t - \langle\Lambda\rangle)^2 \right\rangle \\ &= \left\langle E_\Lambda \left[(\hat{\Lambda} - \Lambda)^2 \right] \right\rangle - \left\langle (\Lambda - \langle\Lambda\rangle)^2 \right\rangle \end{aligned} \tag{4}$$

式 4 は Λ の分散を第 2 項に含んでいるので, 観測量のみからなる式に書き直し次式を得る.

$$C(\Delta) = 2\left\langle E_\Lambda \left[(\hat{\Lambda} - \Lambda)^2 \right] \right\rangle - \left\langle E_\Lambda \left[(\hat{\Lambda} - \langle\Lambda\rangle)^2 \right] \right\rangle \tag{5}$$

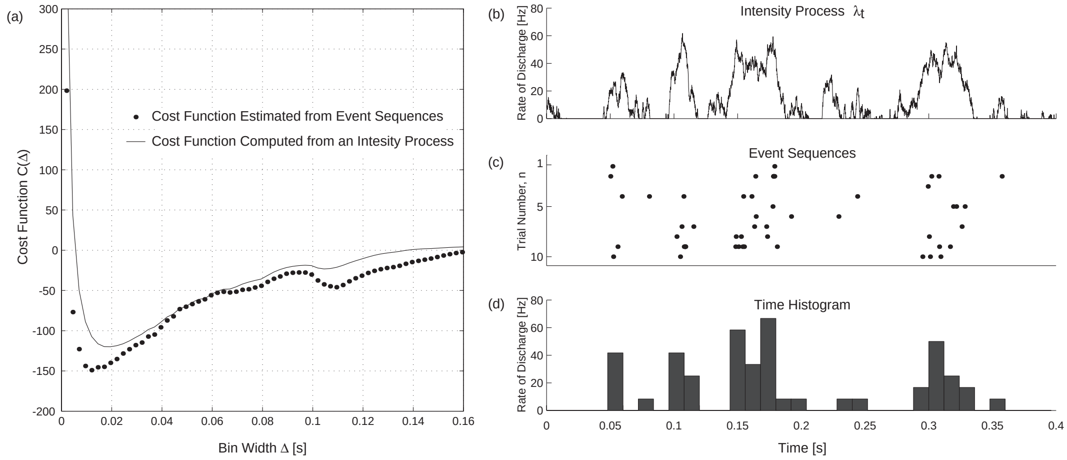

図 1: スパイク時系列からの最適ヒストグラムの作成. (a) 黒丸: 式 6 を用いてスパイク時系列から推定したコスト関数 (各時間幅 Δ 内での移動平均をとった). 実線: 背後のレートを用いて式 7 の右辺 から計算したコスト関数. (b) 時間依存 Poisson 過程に使用したレート (強度過程). (c) スパイク時系列 (時間依存 Poisson 過程, 試行数 $n = 10$, 総スパイク数 51 個). (d) スパイク時系列から推定された最適幅からなる時間ヒストグラム.

ヒストグラムの最適区間幅決定の手順を以下にまとめ, 図 1-a に (i)-(iv) の手順から得られたコスト関数を示す.

最適区間幅決定のレシピ

(i) 観測時間 T, 試行回数 n のスパイク時系列を区間幅 Δ で $N = \lfloor T/\Delta \rfloor$ 個に分割する. i 番目の区間に入るスパイクの数を数え k_i とする.
(ii) スパイク数の標本平均 $\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i$, 及び~~不偏分散~~ $s^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(k_i - \bar{k}\right)^2$ を計算する.
(iii) スパイク統計量 $\bar{k}$, s^2 からコスト関数 (式 5),

$$C(\Delta) = \frac{2\bar{k} - s^2}{(n\Delta)^2} \tag{6}$$

を計算する.
(iv) 異なる Δ に対して (i) から (iii) を繰り返しコスト関数の最小値を与える Δ を探す.

3 議論と発展

3.1 最適区間幅の理論値のスケーリング則と発散

式 4 の第一項に Cramér-Rao の不等式を適用することで, コスト関数の下限が Λ の統計量で与えられる.

$$C(\Delta) \geq \frac{\langle \Lambda \rangle}{n\Delta} - \left\langle (\Lambda - \langle \Lambda \rangle)^2 \right\rangle \tag{7}$$

右辺の極値を考えることで, レートが平均 μ, 相関関数 $\phi(t)$ なる定常確率過程について最適幅の解析解が求まる.

n が十分大きい場合, $\phi(t)$ の原点付近での展開式を用いて式 7 右辺の極値を与える Δ を求める. $\phi(t)$ が原点で Cusp 型となるときは漸近値 $\phi'(0+)$ を用いて最適幅は $\Delta^* \sim \sqrt{-3\mu/\phi'(0+)\,n}$ で与えられる. $\phi(t)$ が原点でなめらかなときは対称性から $\phi'(0) = 0$ であり, 最適幅は $\Delta^* \sim \left(-6\mu/\phi''(0+)\,n\right)^{1/3}$ となる.

n が小さい転移点付近では $\langle(\Lambda - \langle\Lambda\rangle)^2\rangle \simeq \mu/n_c(1/\Delta) - u(1/\Delta)^2 + O\left((1/\Delta)^3\right)$ と展開する (n_c, u は定数). このとき臨界点は n_c でありランダウの 2 次相転移の理論が適用できる. $n > n_c$ では最適幅の振る舞いは $\Delta^* \sim n n_c/(n - n_c)$ で表される.

3.2 データからのスケーリング指数の同定とヒストグラム作成に必要な最小試行回数の推定

n 回の試行数のスパイク統計 $\bar{k}$, s^2 を用いて, 試行数が m 回の場合の最適幅を, 外挿したコスト関数

$$C_m(\Delta) = \left(\frac{1}{m} + \frac{1}{n}\right)\frac{\bar{k}}{n\Delta^2} - \frac{s^2}{(n\Delta)^2} \tag{8}$$

から推定することができる. これによりデータから次の 2 つの数値を推定できる. (i) スケーリング指数: 最適幅の試行数に関するスケーリング指数を調べることができる. 指数が $-1/2$ のときは背後のレートは微分不可能であり, $-1/3$ のときには微分の存在するなめらかな確率過程であると推定される. (ii) ヒストグラム作成に必要な最小試行回数: 転移点で発散やとびを示す指標を用いることで, 転移点を与える n_c を推定することができる. これにより実験者は少ない試行数からヒストグラム作成に最低限必要な試行数を予測することができる.

参考文献

[1] Rudemo, M. (1982) "Empirical Choice of Histograms and Kernel Density Estimators" Scandinavian Journal of Statistics, **9**: 65-78

[2] Koyama, S. and Shinomoto, S. (2004) "Histogram bin width selection for time-dependent Poisson processes." Journal of Physics A-Mathematical and General, **37**: 7255-7265.