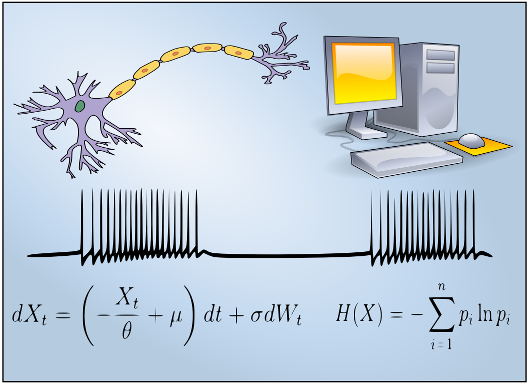
Variability and randomness in neuronal firing patterns

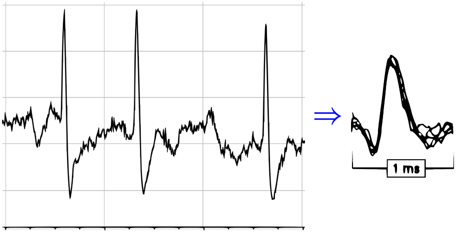
Lubomir Kostal
Institute of Physiology CAS, Prague, Czech Republic

Department of Computational Neuroscience



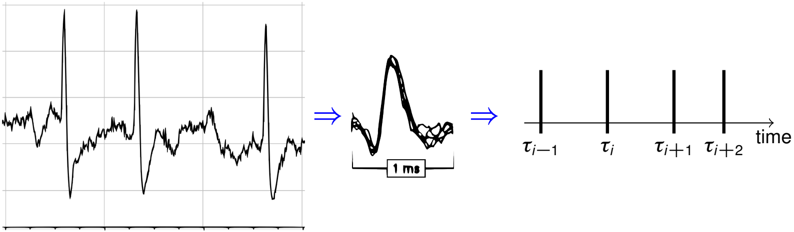
http://comput.biomed.cas.cz

Neuronal signal and “code”



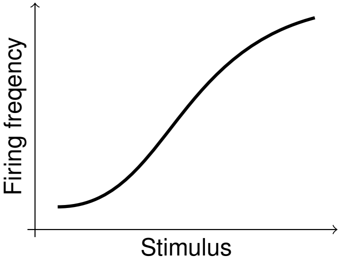
- *Action potential* (AP, spike): activates synaptic *transmission*
- AP shape: *constant* for each individual neuron

Neuronal signal and “code”



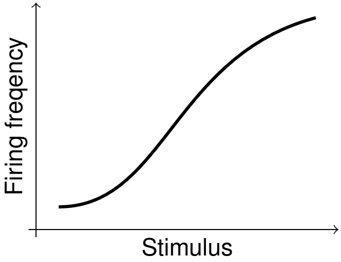
- *Action potential* (AP, spike): activates synaptic *transmission*
- AP shape: *constant* for each individual neuron
- AP is a *point* event in time

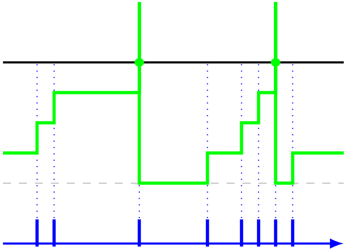
Frequency *vs.* temporal coding



1. Frequency: Adrian (1926), number of APs per unit time (suitably chosen)

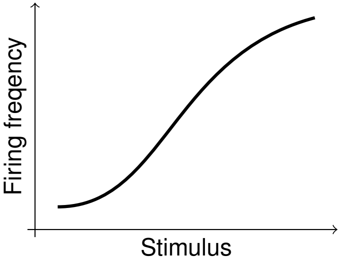
Frequency *vs.* temporal coding

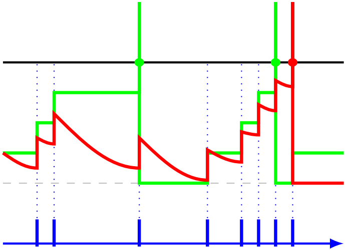




1. Frequency: Adrian (1926), number of APs per unit time (suitably chosen)
2. Temporal: intervals between AP matter (*see:* “leaky” neurons)

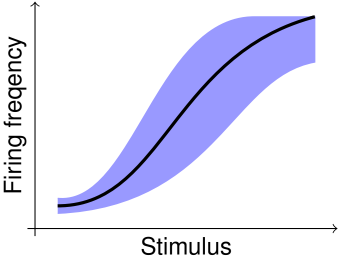
Frequency *vs.* temporal coding

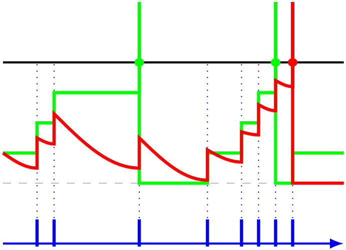




1. Frequency: Adrian (1926), number of APs per unit time (suitably chosen)
2. Temporal: intervals between AP matter (*see:* “leaky” neurons)

Frequency *vs.* temporal coding



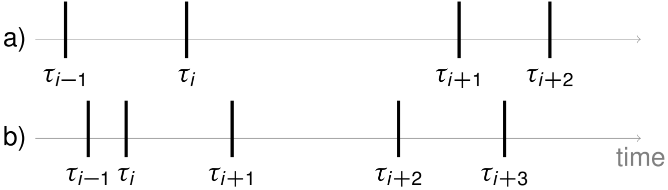


1. Frequency: Adrian (1926), number of APs per unit time (suitably chosen)
2. Temporal: intervals between AP matter (*see:* “leaky” neurons)

- 1. and 2. are *not* mutually exclusive (Perkel & Bullock, 1968), *variability* (Stein et al., *Nat. Rev. Neurosci.* 2005)

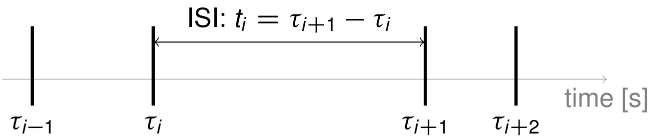
Comparing patterns of neuronal activity

- Spike train: series of APs in time
- “Variability” both within and across trials: unpredictability



- **Methods that compare spike trains are important for characterizing different neuronal coding schemes**
- E.g., *stationary* neural firing: differences beyond the mean firing rate *(frequency coding)* may characterize the *temporal* code

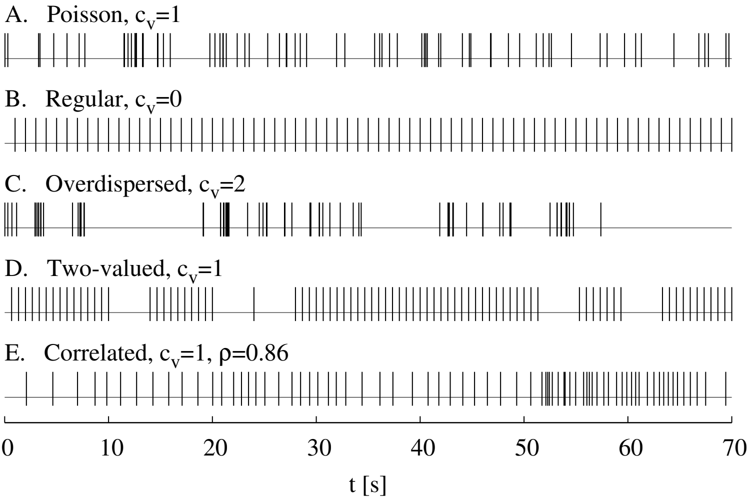
Assumptions



- Neuronal firing under steady-state conditions is often described as a renewal process of *interspike intervals* (ISIs) t_i
- ISIs are then independent realizations of a positive continuous random variable T
- Spike train is fully described by the probability density function (p.d.f.) $f(t)$ (statistical *vs.* biophysical models)
- Extension under *stationarity* conditions: $f(t_1, t_2, \ldots)$

Example spike trains (simulated)

- Different spiking patterns, $\mathrm{E}(T) = 1$, $c_v = \sqrt{\mathrm{Var}(T)} / \mathrm{E}(T)$



Kostal *et al.*, *Eur. J. Neurosci*, 2007

Motivation

- Observation: patterns of neuronal activity can be very different even if $\mathrm{E}(T)$ (mean ISI) is fixed
- How to describe the differences "beyond" $\mathrm{E}(T)$? (Note that: $\#\mathrm{APs}/\Delta = 1/\,\mathrm{E}(T)$)
- Variability (classical): variance or coefficient of variation
  $\mathrm{Var}(T) \propto \mathrm{E}(T)^{\alpha}$ (Koyama, 2015; Koyama & Kobayashi, 2016)
- Shinomoto *et al.*, 2003: *local variability* (Aoki, Takaguchi, Kobayashi and Lambiotte, 2016)

$$L_V = \sum_{i=1}^{n-1} \frac{1}{n-1} \cdot 3 \frac{(t_i - t_{i+1})^2}{(t_i + t_{i+1})^2}$$

Statistical dispersion

- Classical *dispersion measures* in statistics:
  standard deviation, inter-quartile range, mean difference, ...
- Relative statistical dispersion coefficients: normalized to the mean value
  - Variability: $c_v = \sigma/\mu$
- Other global and *intuitive* characteristics?
  - Randomness or predictability: entropy-based
  - Smoothness, modes, ... of the ISI density
- Goal: propose the corresponding *relative dispersion* coefficients

“Variability”, c_v

- Coefficient of variation, c_v

$$c_v = \frac{\sqrt{\mathrm{Var}(T)}}{\mathrm{E}(T)}$$

- $0 \leq c_v < \infty$: no unique c_v-maximizing $f(t)$
- $c_v = 0$ for regular firing
- $c_v = 1$ for $f(t)$ exponential (=Poisson process), but also for other models . . .

- Poisson process is the most random model (by construction), hence c_v *does not* measure randomness

"Randomness", c_h

- Shannon's entropy (discrete r.v. X, $p_i = \Pr(X = x_i)$)

$$H(X) = -\sum_i p_i \log_2 p_i \quad \text{(bit)}$$

- Differential entropy of r.v. $T \sim f(t), t \in \mathcal{T}$:

$$h(T) = -\int_{\mathcal{T}} f(t) \log f(t) \, \mathrm{d}t$$

- $h(T)$ not directly usable (may be negative, ...)
- Propose the entropy-based dispersion as

$$\sigma_h = e^{h(T)}, \quad \sigma_h > 0$$

“Randomness”, c_h

- Inspired by the Asymptotic Equipartition Property:
  Almost any sequence of n realizations of the r.v. T comes from a set A_T in the n-dimensional space of all possible outcomes, and Vol $A_T \approx \exp[nh(T)] = \sigma_h^n$.
- The dispersion coefficient (c_v analogy)

$$c_h = \frac{\sigma_h}{\mathrm{E}(T)} = e^{1 - D_{\mathrm{KL}}\left[f \,\|\, f_{\mathrm{exp}}\right]}, \quad c_h > 0$$

- $\max c_h = e$ iff $f(t)$ is exponential, note $c_v = 1$
- c_h measures the overall “spread” of $f(t)$, “randomness”

$$\text{“Variability''} \neq \text{“Randomness''}$$

Maximum c_h given c_v

- Maximum entropy distribution given $E(T)$ and $E(T^2)$:

$$f_{\max}(t) = \frac{1}{Z}\exp(\lambda_1 t + \lambda_2 t^2)$$

- Euler-Lagrange: $f_{\max}(t) = N(\mu, \sigma^2) / \int_0^\infty N(\mu, \sigma^2)\, dt$:

$$f_{\max}(t) = \sigma\sqrt{\frac{2}{\pi}}\left[1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right]^{-1} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]$$

- Only for $c_v < 1$!
- For $c_v = 1$: $f_{\max}$ is exponential, for $c_v > 1$: unique $f_{\max}$ does not exist ("perturbed" exponential: $c_h \to 1$)

“Smoothness”, c_J

- “Shape” of $f(t)$: translational parameter $\theta \in \mathbb{R}$: $f(t) \to f(t-\theta)$
- Sensitivity: $\mathrm{Var}(\hat{\theta}) \geq 1/J(T)$ where

$$J(T) = \int_0^\infty \left[\frac{\partial \ln f(t)}{\partial t}\right]^2 f(t)\,\mathrm{d}t$$

- Fisher information-based dispersion coefficient, c_J

$$c_J = \frac{1}{\mathrm{E}(T)\sqrt{J(T)}}, \quad c_J > 0$$

- $c_J = 1$ for $f(t)$ exponential
- Any locally steep slope or the presence of modes in the shape of $f(t)$ decreases c_J

Maximal c_J

- Version of *regularity conditions*: $f(t) \in \mathcal{C}^1$ for *all* $t > 0$ and $f(0) = f'(0) = 0$, also $0 < J(T) < \infty$
- Define the *probability amplitude* (real)

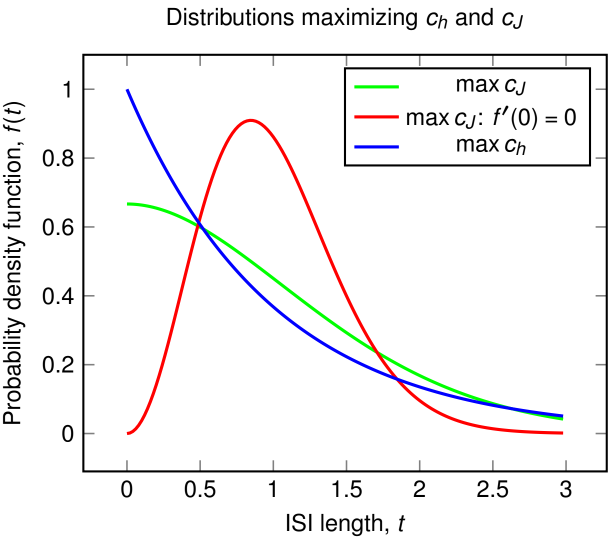
$$u(t) = \sqrt{f(t)} \quad \Rightarrow \quad J(T) = 4 \int_{\mathcal{T}} u'(t)^2 \, \mathrm{d}t$$

- Euler-Lagrange *s.t.* $\int u(t)^2 \, \mathrm{d}t = 1$ and $\int u(t)^2 t \, \mathrm{d}t = \mathrm{E}(T)$

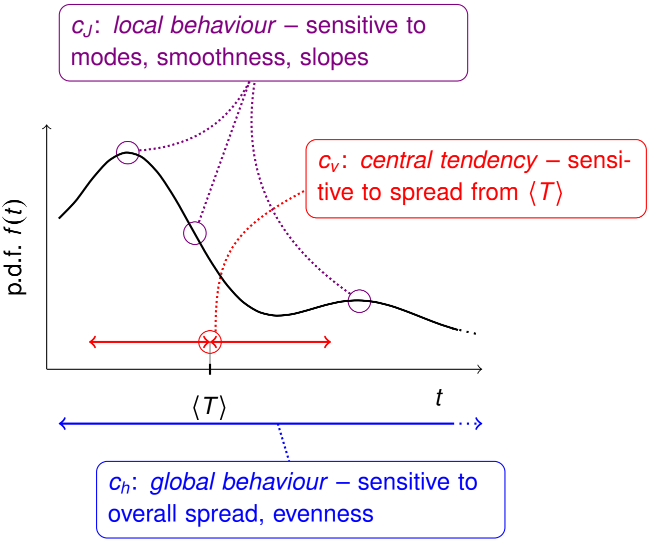
$$u''(t) = (\lambda_1 - \lambda_2 t) u(t) \quad \Rightarrow \quad f(t) \propto \mathrm{Ai}^2 \left(c_1 + c_2 / \mathrm{E}(T) \right)$$

- $\max c_J \doteq 1.26$: $f(t)$ based on the Airy function

Distributions maximizing c_h and c_J given $\mathrm{E}(T) = 1$



Properties of the proposed measures



Statistical ISI models

- Common two-parametric, for convenience $f(t; \mathrm{E}(T) = 1, c_v)$
- Both c_h, c_J can be found analytically

Gamma p.d.f.

- c_J exists for $0 < c_v < 1/\sqrt{2}$ and for $c_v = 1$

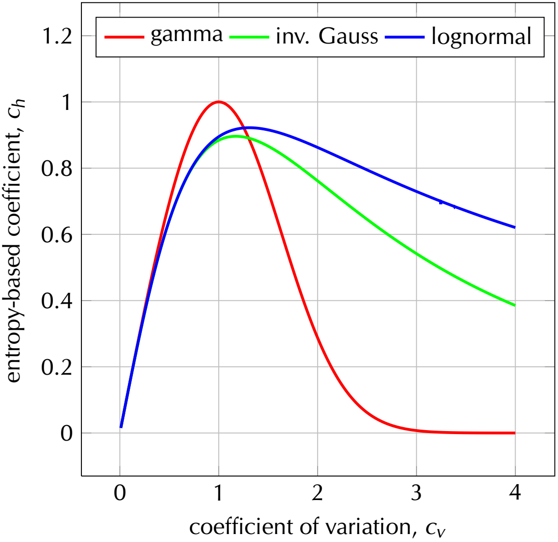
Log-normal p.d.f.

- For $c_v = 1$ it is not exponential ($c_h < 1$)

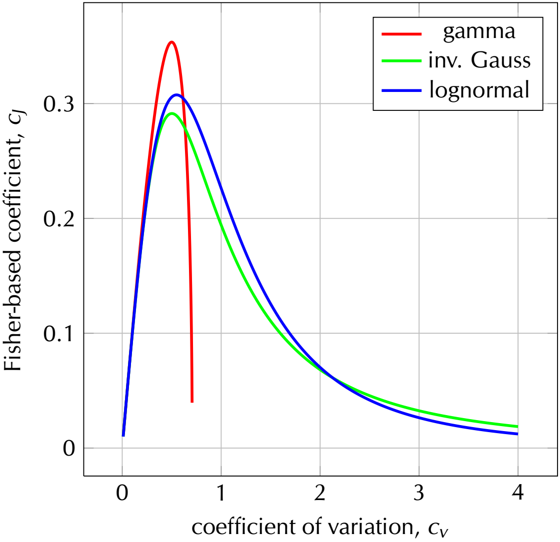
Inverse Gaussian p.d.f.

- "Similar" to log-normal

Dispersion coefficients for some typical ISI models



Dispersion coefficients for some typical ISI models



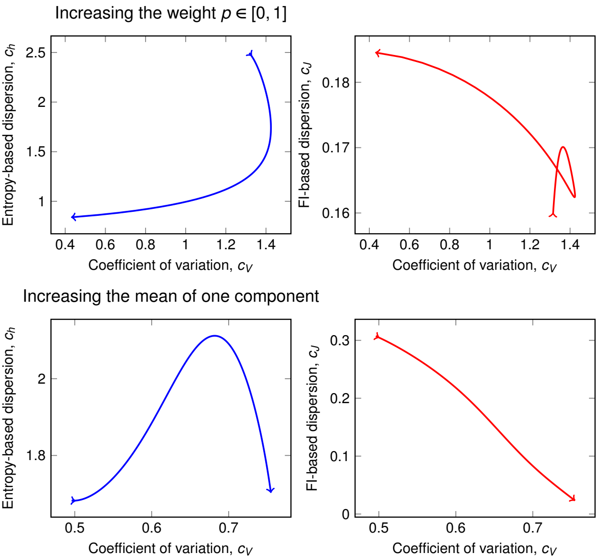
Example: log-normal mixture

- Mixture models: wide applicability, including statistical ISI models
- Consider *mixture of normals*

$$g_m(x) = p\phi(x, m_1, s_1) + (1 - p)\phi(x, m_2, s_2)$$

- $T = \exp X$: log-normal mixture $\approx$ ISI model
- “Variability $\neq$ Randomness”
- “Smoothness $\neq$ Randomness”

Log-normal mixture (sensitivity of c_J to modes)



Non-parametric estimation of c_v, c_h, c_J

- $\hat{c}_v = \hat{\sigma}/\hat{\mu}$ may be problematic (Ditlevsen & Lansky, 2011)

- Estimate c_h *without* $\widehat{f(t)}$: non-parametric binless estimate
- Vasicek estimator given n ranked ISIs $\{t_{[1]} < t_{[2]} < \cdots < t_{[n]}\}$

$$\hat{h} = \frac{1}{n}\sum_{i=1}^{n} \ln\left[\frac{n}{2m}\left(t_{[i+m]} - t_{[i-m]}\right)\right] + \varphi_{\mathrm{bias}}, \quad m \doteq \sqrt{n}$$

$$\varphi_{\mathrm{bias}} = \ln\frac{2m}{n} - \left(1 - \frac{2m}{n}\right)\Psi(2m) + \Psi(n+1) - \frac{2}{n}\sum_{i=1}^{m}\Psi(i+m-1), \quad \Psi(z) = \frac{\mathrm{d}}{\mathrm{d}z}\ln\Gamma(z)$$

- $f(t_1, t_2, \dots)$: Kozachenko-Leonenko estimator

Non-parametric estimation of c_v, c_h, c_J

- *Maximum Penalized Likelihood* (MPL) estimation of $f(t)$
- Likelihood *vs.* roughness penalty (Good & Gaskins, 1971)
- Let $u(t) = \sqrt{f(t)}$ and for the given sample $\{t_1, \ldots, t_n\}$

$$\max_{u(t)} : 2 \sum_{i=1}^{n} \log |u(t_i)| - 4\alpha \int u'(t)^2 \, \mathrm{d}t - \beta \int u''(t)^2 \, \mathrm{d}t$$

Non-parametric estimation of c_v, c_h, c_J

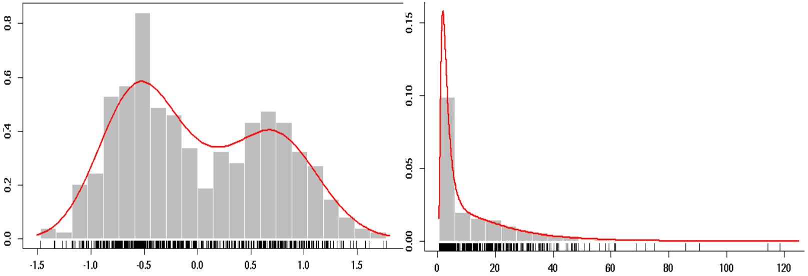
- *Maximum Penalized Likelihood* (MPL) estimation of $f(t)$
- Likelihood *vs.* roughness penalty (Good & Gaskins, 1971)
- Let $u(t) = \sqrt{f(t)}$ and for the given sample $\{t_1, \ldots, t_n\}$

$$\max_{u(t)} : 2 \sum_{i=1}^{n} \log |u(t_i)| - 4\alpha \int u'(t)^2 \, \mathrm{d}t - \beta \int u''(t)^2 \, \mathrm{d}t$$

- Assume Hermitian base for $u \Rightarrow$ nonlinear algebraic eqns. (log-transform of t_i is desirable since $T > 0$)
- α, β “tune” the likelihood/penalty balance, depend on n?

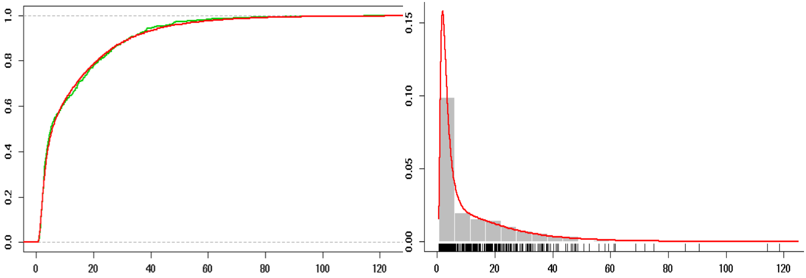
Density estimation (MPL)

- Log-normal mixture ($n = 1000$))
- α, β: "smoothness" regulation

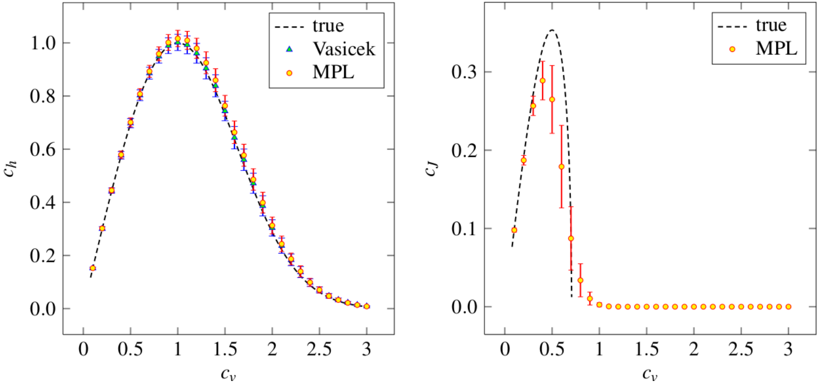


Density estimation (MPL)

- Log-normal mixture ($n = 1000$))
- α, β: “smoothness” regulation



$c_{h,J}$ estimation: gamma ($n = 1000$)

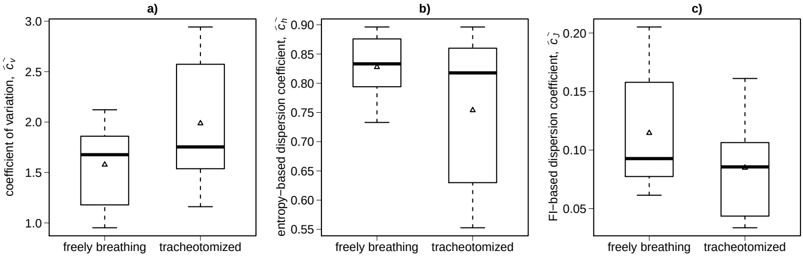


Non-parametric estimation: work in progress

- What if $\beta = 0$? Closed-form solution is known (Laplacian "kernels", Klonias (1982))
- Consistency: fixes asymptotics of $\alpha(n)$
- MLE: arbitrary but convenient, small sample size?

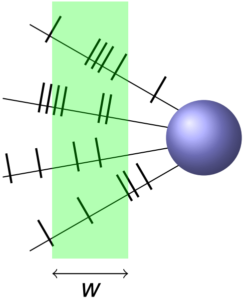
- Different (more robust?) approach:
  - Theorem (Huber, 1974): There exists a *unique* $\hat{f}(t)$ such that its CDF *interpolates* the ECDF based on $\{t_1, \ldots, t_n\}$ and *minimizes* the Fisher information $J(T)$.
  - $J(T)$ is *convex* in $f(t)$
  - Convergence $\hat{c}_J \to c_J$ guaranteed
  - No free parameters? Simplicity of estimation?

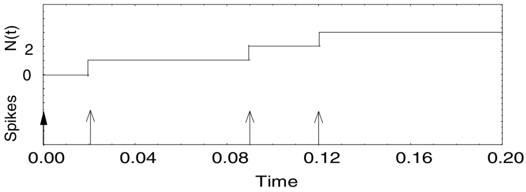
Experimental data application



- ORN of *freely breathing* and *tracheotomized* rats, *spontaneous* single-unit APs recorded
- Note: variability *vs.* randomness

Intervals *vs.* counts: back to frequency coding?





Equally 'good' $\begin{cases} \textit{intervals} : T \sim f(t) \\ \textit{counts} \text{ in a window} : N(w) \end{cases}$

- **Variability, randomness, . . . coefficients for ISIs and counts?**
- $\mathrm{Var}(\cdot) \propto \mathrm{E}(\cdot)^{\alpha}$ for T and $N(w)$ (Koyama and Kobayashi, 2016)

Equilibrium renewal point processes

- Need to specify the *start* of the observation window w
- Equilibrium: the start of w is *random* with respect to APs

$$N(w) \geq n \Leftrightarrow T_0 + T_1 + \cdots + T_{n-1} \leq w$$

- $N(w)$ is stationary, $T_i \sim f(t)$ for $i \geq 1$, however the *time to first spike* is distributed as

$$T_0 \sim f_0(t) = \frac{1 - \int_0^t f(\tilde{t})\,\mathrm{d}\tilde{t}}{\mathrm{E}(T)}$$

- Note that

$$\mathrm{E}\left(N(w)\right) = \frac{w}{\mathrm{E}(T)}$$

Distributions of intervals and counts

- ISI distribution: $T \sim f(t)$
- Let $p_n(w) = \Pr\left(N(w) = n\right)$, where $n \geq 0$ and $w > 0$
- Let $\mathcal{L}\{f\}(s) = \int_0^\infty f(t)e^{-ts}\,\mathrm{d}t$, $s \in \mathbb{C}$ then it holds (Jewell, 1960)

$$p_0(w) = \mathcal{L}^{-1}\left\{\frac{1}{s} - \frac{1 - \mathcal{L}\{f\}(s)}{s^2\,\mathrm{E}(T)}\right\}(w),$$

$$p_n(w) = \mathcal{L}^{-1}\left\{\frac{\left[1 - \mathcal{L}\{f\}(s)\right]^2\left[\mathcal{L}\{f\}(s)\right]^{n-1}}{s^2\,\mathrm{E}(T)}\right\}(w)$$

Variability of counts

- Fano factor: variability with respect to the *Poisson process* $N_P(w)$ with the *same* rate as $N(w)$

$$\mathrm{FF}(w) = \frac{\mathrm{Var}\left(N(w)\right)}{\mathrm{E}\left(N(w)\right)} = \frac{\mathrm{Var}\left(N(w)\right)}{\mathrm{Var}\left(N_P(w)\right)}$$

- Due to the Bernoulli- and CLT-limiting behavior:

$$\lim_{w \to \infty} \mathrm{FF}(w) = c_v^2, \quad \lim_{w \downarrow 0} \mathrm{FF}(w) = 1$$

Randomness of counts

- "Entropy factor" as an analogy to *Fano factor*

$$\mathrm{HF}(w) = \frac{H\big(N(w)\big)}{H\big(N_P(w)\big)}, \quad H\big(N(w)\big) = -\sum_{n=0}^{\infty} p_n(w) \log p_n(w)$$

- Poisson process with intensity λ

$$H\big(N_P(w)\big) = \lambda w[1 - \log(\lambda w)] + e^{-\lambda w} \sum_{n=0}^{\infty} \frac{(\lambda w)^n \log(n!)}{n!}$$

$$\overset{\lambda w \to \infty}{\approx} \frac{1}{2} \log(2\pi e \lambda w) - \frac{1}{12\lambda w} - \frac{1}{24(\lambda w)^2} - \cdots$$

- Limits (Bernoulli p. for $w \downarrow 0$ *vs* CLT for $N(w \to \infty)$)

$$\lim_{w \to \infty} \mathrm{HF}(w) = \lim_{w \downarrow 0} \mathrm{HF}(w) = 1$$

Maximal value of the entropy factor

- Maximum entropy among all $N(w)$ with mean value λw: geometric distribution N_g

$$\Pr\left(N_g(w) = n\right) = \left[1 - \frac{1}{1+\lambda w}\right]^n \frac{1}{1+\lambda w}, \quad n = 0, 1, 2, \ldots$$

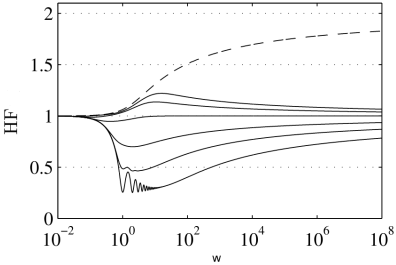
- The entropy is

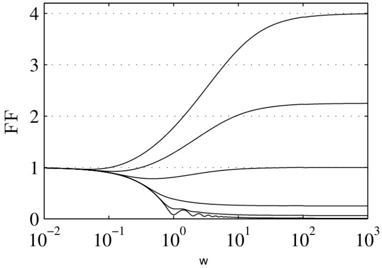
$$H\bigl(N_g(w)\bigr) = (1+\lambda w)\log(1+\lambda w) - \lambda w \log(\lambda w)$$

- However (*c.f.* with the general $\mathrm{HF}(w)$ limit!)

$$\lim_{w\to\infty} \mathrm{HF}_g(w) = \lim_{w\to\infty} \frac{H\bigl(N_g(w)\bigr)}{H\bigl(N_P(w)\bigr)} = 2$$

Entropy factor *vs.* Fano factor for inverse Gaussian





- Parameters: $\mathrm{E}(T) = 1$, $c_v = \{0.1, 0.25, 0.5, 1, 1.5, 2\}$

- Randomness of the Poisson process: ISIs and counts
- *Bursting*: random counts (*c.f.* ISI randomness)
- Information in *temporal* or *frequency* codes?

- Small variability ⇒ low randomness, variable ≠ random
  (Kostal *et al.*, *Eur. J. Neurosci.*, 2007)
- Dispersion-like quantities, compare p.d.f. shapes
  (Kostal *et al.*, *Inform. Sci.*, 2013)
- Parametric × non-parametric estimation of c_h and c_J
  (Kostal and Pokora, *Entropy*, 2012)
- Randomness and variability: counts vs. intervals
  (Rajdl *et al.*, submitted)

- **Collaborators: Petr Lansky, Ondrej Pokora, Kamil Rajdl**