Phase transitions in the estimation of event rate: a path integral analysis

2007 J. Phys. A: Math. Theor. 40 F383

(http://iopscience.iop.org/1751-8121/40/20/F01)

IOP Publishing

Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **40** (2007) F383–F390 doi:10.1088/1751-8113/40/20/F01

FAST TRACK COMMUNICATION

Phase transitions in the estimation of event rate: a path integral analysis

S Koyama[1], T Shimokawa and S Shinomoto

Department of Physics, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

E-mail: shinomoto@scphys.kyoto-u.ac.jp

Received 16 November 2006, in final form 9 April 2007
Published 30 April 2007
Online at stacks.iop.org/JPhysA/40/F383

Abstract

We try to capture the instantaneous rate of event occurrence as a function of time. A jagged or smooth rate function can be arbitrarily chosen, but it is nevertheless possible to select a plausible smoothness according to the principle of maximum likelihood. Upon optimization, the smoothness may diverge, indicating that the data are insufficient to uncover the time dependence of the underlying rate. By evaluating the likelihood function through a marginalization path integral, we found not only first-order but also second-order phase transitions leading to the divergence of the optimized smoothness.

PACS numbers: 02.50.Fz, 02.50.Ey, 02.50.Tt, 05.70.Fh

1. Introduction

Given a sequence of events, such as economic events, seismic activity or neuronal spikes, we describe the rate as a function of time [1–5]. The smoothness of the rate function can be chosen arbitrarily depending on one's assumption or prejudice about the smoothness of the underlying rate. The assumption for the smoothness can be incorporated in the Bayesian prior probability of the underlying rate. It is nevertheless possible to select the very assumption by maximizing the likelihood function, which is the probability of obtaining data given an assumption about the underlying rate. This probability of events can be obtained through 'marginalization', the integration of a joint probability over the other arguments [6–10]. In the present case, the marginalization is carried out through functional integration over potential rate processes. Bialek *et al* pointed out that the marginalization can be carried out through a corresponding path integral, a technique developed by Richard Feynman in the fields of quantum mechanics and statistical mechanics [11].

[1] Present address: Department of Statistics and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

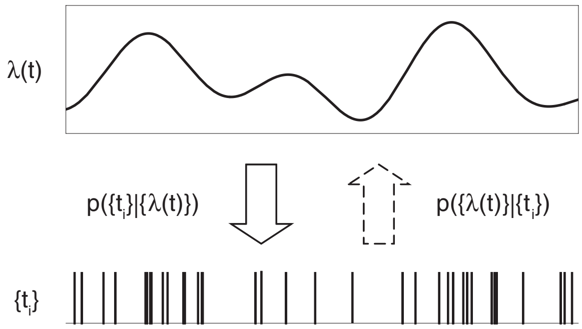

Figure 1. Events $\{t_i\}$ are derived from the underlying rate $\lambda(t)$ according to the conditional probability $p(\{t_i\} \mid \{\lambda(t)\})$. The underlying rate is inferred from the data with the inverse probability $p(\{\lambda(t)\} \mid \{t_i\})$.

In the present paper, we formulated a method of evaluating the likelihood function for the time-dependent Poisson process using a path integral for the marginalization, and applied the formulation to two rate processes that are modulated sinusoidally and with the Ornstein–Uhlenbeck process, respectively. We observed in these applications that the time scale of the rate-estimation kernel diverges for events derived from a modestly fluctuating rate. The phase transition describing the divergence can be either first order (discontinuous) or second order (continuous), depending on the characteristics of the rate fluctuation.

2. The Bayes methods

Let us start with a time-dependent Poisson process in which events are derived from a given underlying rate $\lambda(t)$. In this process, the probability for events to occur at $\{t_i\} \equiv \{t_1, t_2, \ldots, t_n\}$ in the period of $t \in [0, T]$ is given by

$$p\left(\{t_i\} \mid \{\lambda(t)\}\right) = \left[\prod_{i=1}^{n} \lambda(t_i)\right] \exp\left(-\int_0^T \lambda(t)\,\mathrm{d}t\right), \tag{1}$$

where the exponential term is the survivor function that represents the probability that no events occur in the inter-event intervals [12, 13].

Next, consider inverting the arguments of this conditional probability so that the unknown underlying rate is inferred from the events received (see figure 1). This 'inverse probability' can be obtained using the Bayes formula

$$p_\gamma(\{\lambda(t)\} \mid \{t_i\}) = \frac{p(\{t_i\} \mid \{\lambda(t)\})\, p_\gamma(\{\lambda(t)\})}{p_\gamma(\{t_i\})}. \tag{2}$$

In this paper, we introduce a prior distribution of $\lambda(t)$ that incorporates the tendency of the estimated rate to be relatively flat:

$$p_\gamma(\{\lambda(t)\}) = \frac{1}{Z(\gamma)} \exp\left[-\frac{1}{2\gamma^2}\int_0^T \left(\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t}\right)^2 \mathrm{d}t\right], \tag{3}$$

where γ is a hyperparameter representing the degree of fluctuation and $Z(\gamma)$ is a normalization constant. We can consider a higher order derivative of $\lambda(t)$, but the qualitative features of the results remain unchanged.

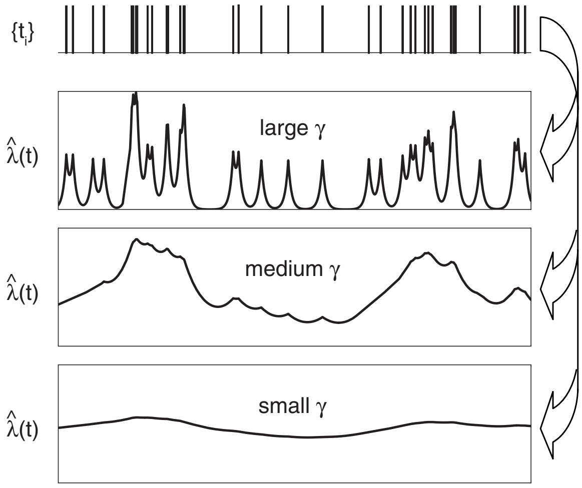

Figure 2. The MAP estimate $\hat{\lambda}(t)$ for a given sequence of events $\{t_i\}$ depends on the hyperparameter of the prior, γ.

The probability of having events $p_\gamma(\{t_i\})$ in equation (2) can be obtained by the 'marginalization' of the joint probability,

$$p_\gamma(\{t_i\}) = \int p_\gamma(\{t_i\}, \{\lambda(t)\}) d\{\lambda(t)\}$$
$$= \int p(\{t_i\} \mid \{\lambda(t)\}) p_\gamma(\{\lambda(t)\}) d\{\lambda(t)\}, \tag{4}$$

where $d\{\lambda(t)\}$ represents a functional integration over all possible paths of $\lambda(t)$. This probability, called the 'marginal likelihood function' or the 'evidence', represents a likelihood function with respect to the hyperparameter. The method of selecting a hyperparameter by maximizing the marginal likelihood function is called the 'empirical Bayes' [6–10].

2.1. *The path integral method*

The marginal likelihood function (4) can be represented in the form of a path integral,

$$p_\gamma(\{t_i\}) = \frac{1}{Z(\gamma)} \int \exp\left[-\int_0^T L(\dot{\lambda}, \lambda, t)\, \mathrm{d}t\right] d\{\lambda(t)\}, \tag{5}$$

where $L(\dot{\lambda}, \lambda, t)$ is a 'Lagrangian' of the form

$$L(\dot{\lambda}, \lambda, t) = \frac{1}{2\gamma^2}\dot{\lambda}^2 + \lambda - \sum_{i=1}^{n} \delta(t - t_i) \log \lambda. \tag{6}$$

The maximum *a posteriori* (MAP) estimate of the rate, $\hat{\lambda}(t)$, corresponds to the 'classical path' obtained from the Euler–Lagrange equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\lambda}}\right) - \frac{\partial L}{\partial \lambda} = 0. \tag{7}$$

Note that the MAP estimate $\hat{\lambda}(t)$ depends critically on the hyperparameter γ (see figure 2).

The marginal likelihood function is obtained by performing the integration over all paths of $\lambda(t)$ around the 'classical path', $\hat{\lambda}(t)$. By approximating the action integral to a range quadratic in the deviation from the classical path, the integral can be performed analytically as [14–17]

$$p_\gamma(\{t_i\}) = R \exp\left[-\int_0^T L(\dot{\hat{\lambda}}, \hat{\lambda}, t)\, \mathrm{d}t\right], \tag{8}$$

where R represents the 'quantum' contribution of the quadratic deviation to the path integral:

$$R = \sqrt{\frac{\frac{\partial^2 L}{\partial \dot{\lambda}^2}}{2\pi T}} \left[\frac{\det\left(\frac{\partial^2 L}{\partial \dot{\lambda}^2}\partial_t^2 + \frac{\partial^2 L}{\partial \lambda \partial \dot{\lambda}}\partial_t - \frac{\partial^2 L}{\partial \lambda^2}\right)}{\det\left(\frac{\partial^2 L}{\partial \dot{\lambda}^2}\partial_t^2\right)}\right]^{-\frac{1}{2}}$$

$$= \sqrt{\frac{\frac{\partial^2 L}{\partial \dot{\lambda}^2}}{2\pi T}} \left[\frac{f_1(T,0)}{f_2(T,0)}\right]^{-\frac{1}{2}}. \tag{9}$$

We obtained the determinant using the Gelfand–Yaglom method. f_1 and f_2, respectively, satisfy

$$\left(\frac{\partial^2 L}{\partial \dot{\lambda}^2}\partial_t^2 + \frac{\partial^2 L}{\partial \lambda \partial \dot{\lambda}}\partial_t - \frac{\partial^2 L}{\partial \lambda^2}\right) f_1(t,0) = 0, \qquad f_1(0,0) = 0, \qquad \left.\frac{\mathrm{d}f_1(t,0)}{\mathrm{d}t}\right|_{t=0} = 1,$$

$$\frac{\partial^2 L}{\partial \dot{\lambda}^2}\partial_t^2 f_2(t,0) = 0, \qquad f_2(0,0) = 0, \qquad \left.\frac{\mathrm{d}f_2}{\mathrm{d}t}(t,0)\right|_{t=0} = 1.$$

2.2. *Evaluation of the marginal likelihood*

The maximization of the marginal likelihood function $p_\gamma(\{t_i\})$ with respect to a particular sequence of events can be carried out numerically by using the expectation and maximization (EM) method [18–20]. Here, we formulate a method for computing a marginal likelihood function averaged over the ensemble of event sequences derived from a given underlying rate,

$$\lambda(t) = \mu + \sigma f(t/\tau), \tag{10}$$

where μ is the mean rate and $\sigma f(t/\tau)$ represents a rate fluctuation characterized by the amplitude σ and a time scale τ. In each realization, the occurrence of events fluctuates around the underlying rate, and $\sum_{i=1}^n \delta(t - t_i)$ can be represented as a stochastic process [21],

$$\lambda(t) + \sqrt{\lambda(t)}\xi(t), \tag{11}$$

where $\xi(t)$ is a white noise characterized by the ensemble averages $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$.

As the rate is non-negative, the path integral defined by equation (5) should be carried out in the half space, $\lambda \geqslant 0$. Under the condition that the rate fluctuation is small, $\sigma/\mu \ll 1$, however, the orbits passing through $\lambda < 0$ do not practically contribute to the integral.

Utilizing the same condition $\sigma/\mu \ll 1$, the Lagrangian can be approximated to the range quadratic in the deviation $x(t) \equiv \lambda(t) - \mu$ as

$$L = \frac{1}{2\gamma^2}\dot{x}^2 + x - \{\mu + \sigma f(t/\tau) + \sqrt{\mu + \sigma f(t/\tau)}\xi(t)\} \log\left(1 + \frac{x}{\mu}\right) \tag{12}$$

$$\approx \frac{1}{2\gamma^2}\dot{x}^2 - \frac{\sigma f(t/\tau) + \sqrt{\mu}\xi(t)}{\mu}x + \frac{1}{2\mu}x^2, \tag{13}$$

where we have ignored the term of $O((\sigma/\mu)^{3/2})$. The solution of the Euler–Lagrange equation (7), representing the 'classical path' for this Lagrangian, is given by

$$\hat{x}(t) = \frac{\gamma}{2\sqrt{\mu}} \int_0^T \mathrm{e}^{-\frac{\gamma}{\sqrt{\mu}}|t-s|}(\sigma f(s/\tau) + \sqrt{\mu}\xi(s))\,\mathrm{d}s. \tag{14}$$

The 'classical action' $\int_0^T L(\dot{\hat{x}}, \hat{x}, t)\,\mathrm{d}t$ and the 'quantum effect' R are obtained analytically. Utilizing the Euler–Lagrange equation, the classical action can be rewritten as

$$\int_0^T L\,\mathrm{d}t = \int_0^T \left\{ \frac{1}{2\gamma^2}\frac{\mathrm{d}}{\mathrm{d}t}(\dot{\hat{x}}\hat{x}) - \frac{\sigma f(t/\tau) + \sqrt{\mu}\xi(t)}{2\mu}\hat{x} \right\} \mathrm{d}t. \tag{15}$$

The first term in the rhs representing the end-point effect is negligible compared to the second term whose contribution is of the order of $T \gg 1$. The classical action can be obtained explicitly by inserting the solution (equation (14)) into the second term of equation (15), and assuming the stationarity of the event-rate fluctuation, as

$$-\frac{\gamma}{4\sqrt{\mu}} \int_0^T \int_0^T \mathrm{e}^{-\frac{\gamma}{\sqrt{\mu}}|t-s|} \left\{ \langle \xi(s)\xi(t) \rangle + \frac{\sigma^2}{\mu} \langle f(s/\tau) f(t/\tau) \rangle \right\} \mathrm{d}s\,\mathrm{d}t \tag{16}$$

$$\simeq -\frac{\gamma}{4\sqrt{\mu}} T \left\{ 1 + 2\frac{\sigma^2\tau}{\mu} \int_0^\infty \phi(u)\,\mathrm{e}^{-\frac{\gamma\tau}{\sqrt{\mu}}u}\,\mathrm{d}u \right\}, \tag{17}$$

where $\phi(u) \equiv \langle f(u+u')f(u') \rangle$ is the correlation function of the rate fluctuation.

The 'quantum' contribution is obtained from equation (9), using

$$f_1(t,0) = -\frac{\sqrt{\mu}}{2\gamma}\mathrm{e}^{-\frac{\gamma}{\sqrt{\mu}}} + \frac{\sqrt{\mu}}{2\gamma}\mathrm{e}^{\frac{\gamma}{\sqrt{\mu}}}, \qquad f_2(t,0) = t, \tag{18}$$

as

$$R = (\pi\gamma\sqrt{\mu})^{-\frac{1}{2}} \exp\left(-\frac{\gamma}{2\sqrt{\mu}} T \right). \tag{19}$$

Summing the classical action and quantum effect, the 'free energy' or the negative log marginal likelihood function is written explicitly as

$$F(\gamma) \equiv -\frac{1}{T}\ln p_\gamma(\{t_i\}) = -\frac{1}{T}\left(\log R - \int_0^T L\,\mathrm{d}t \right)$$
$$= \frac{\gamma}{4\sqrt{\mu}} \left(1 - 2\beta \int_0^\infty \phi(u)\,\mathrm{e}^{-\frac{\gamma\tau}{\sqrt{\mu}}u}\,\mathrm{d}u \right), \tag{20}$$

where $\beta \equiv \sigma^2\tau/\mu$ represents the degree of fluctuation in the underlying rate. The hyperparameter γ characterizing the fluctuation in the estimated rate can be selected so that the marginal likelihood function $p_\gamma(\{t_i\})$ is maximized, or the free energy is minimized,

$$\hat{\gamma} = \arg\min_\gamma F(\gamma). \tag{21}$$

3. Results

3.1. Sinusoidally regulated Poisson process

We applied this formula to two time-dependent Poisson processes. The first application is the sinusoidally modulated process:

$$\lambda(t) = \mu + \sigma \sin t/\tau. \tag{22}$$

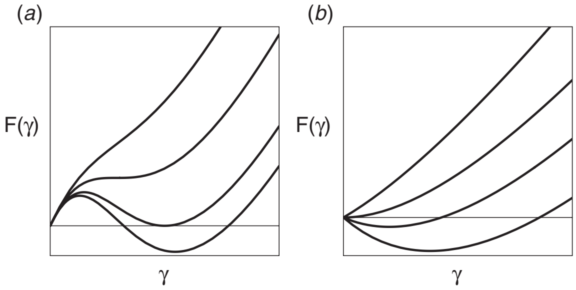

Figure 3. Free energy functions $F(\gamma)$ for two rate processes. (*a*) The first-order phase transition exhibited by the sinusoidally modulated Poisson process. (*b*) The second-order phase transition exhibited by the doubly stochastic Poisson process.

The free energy for this case is

$$F(\gamma) = \frac{\gamma}{4\sqrt{\mu}} - \frac{\beta}{4}\frac{\gamma^2\tau}{\mu + \gamma^2\tau^2}. \tag{23}$$

A vanishing hyperparameter $\gamma = 0$ indicates a constant rate. This free energy defined for $\gamma \geqslant 0$ always has a minimum at $\gamma = 0$, as $\mathrm{d}F(\gamma)/\mathrm{d}\gamma|_{\gamma=0} > 0$ (see figure 3(*a*)). As the amplitude of rate fluctuation σ is increased from zero to $\beta > (4/3)^{3/2}$, another local minimum appears at a finite γ. As the rate fluctuation σ is increased further to $\beta > 2$, the minimum at a finite γ becomes lower than the minimum at $\gamma = 0$, implying that event sequence should be interpreted as being derived from a fluctuating rate. In the large fluctuation extreme $\beta \gg 1$, the optimized time scale obeys the scaling relation

$$\sqrt{\mu}/\hat{\gamma} \sim \sigma^{-\frac{2}{3}}\tau^{\frac{2}{3}}\mu^{\frac{1}{3}} = \tau\beta^{-\frac{1}{3}}. \tag{24}$$

This is consistent with the result found by Bialek *et al* [11].

3.2. Doubly stochastic Poisson process

The second example application is the doubly stochastic process in which the rate obeys the Ornstein–Uhlenbeck process,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = -\frac{\lambda - \mu}{\tau} + \sigma\sqrt{\frac{2}{\tau}}\xi(t), \tag{25}$$

where $\xi(t)$ is a Gaussian white noise. Here, the free energy is

$$F(\gamma) = \frac{\gamma}{4\sqrt{\mu}} - \frac{\beta}{2}\frac{\gamma}{\sqrt{\mu} + \gamma\tau}. \tag{26}$$

It is notable that this free energy never has multiple minima (see figure 3(*b*)). The minimum of this function defined for $\gamma \geqslant 0$ stays at $\gamma = 0$ until β exceeds 1/2. The optimized time scale obeys the scaling relation

$$\sqrt{\mu}/\hat{\gamma} \sim \sigma^{-1}\tau^{\frac{1}{2}}\mu^{\frac{1}{2}} = \tau\beta^{-\frac{1}{2}}, \tag{27}$$

in the large fluctuation $\beta \gg 1$ extreme.

The different phase transitions in these two rate processes are due to the different functional forms of the second term in the free energy (20) which is identical to the Laplace transform of the rate-fluctuation correlation. In the sinusoidally modulated process, the Laplace transform has a peak at a finite frequency and the phase transition is of the first order (discontinuous). In the Ornstein–Uhlenbeck process, the Laplace transform has no peak and the transition is of the second order (continuous).

4. Summary

We carried out the exact marginalization with the path integral method, and optimized the time scale of the rate-estimation kernel by maximizing the likelihood function. It was observed that the time scale of the rate-estimation kernel diverges for the data derived from a moderately fluctuating underlying rate. This implies that under such conditions, any rate estimated with a finite time scale kernel is likely to capture a spurious rate. We also revealed that the time scale diverges either discontinuously or continuously, or in other words, the phase transition is of either the first or the second order.

In the evaluation of a path integral, the Lagrangian is approximated to the quadratic form under the condition of $\sigma/\mu \ll 1$, for which exact solutions can be obtained. Even in this limit, the dimensionless parameter $\beta = \sigma^2\tau/\mu$ can be made finite by choosing large τ. Thus, there is certainly a range of parameters in which the conclusions about the type of phase transitions are rigorous. In addition, we carried out numerical marginalization of the raw joint probability distribution by means of the EM method and evaluated the free energy for a finite value of σ/μ, and confirmed that the numerical results are very close to the analytical results.

Note that the inability to estimate the rate is not necessarily due to the Bayesian method applied to the kernel estimation. We have also formulated a method of optimizing a time histogram under the principle of minimizing the mean integrated square error (MISE), and found that the optimized bin size may diverge [22, 23]. The bin size of the time histogram plays a similar role to the time scale of the rate-estimation kernel in determining the smoothness of the rate. The parametric condition for the divergence of the optimal bin size and that of the time scale of the rate-estimation kernel studied here are very similar (they are identical for the doubly stochastic process; $\beta_c = 1/2$). The asymptotic characteristics (24) and (27) are respectively the same as those for the optimal bin size determined with the MISE criteria. It would be interesting to investigate the relationship of these apparently independent principles.

Acknowledgments

We thank Kensuke Arai for his suggestions for improving the manuscript. This study is supported in part by Grants-in-Aid for Scientific Research to SS from the Ministry of Education, Culture, Sports, Science and Technology of Japan (16300068, 18020015) and the 21st Century COE 'Center for Diversity and Universality in Physics'. SK is supported by the Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists.

References

[1] Snyder D L 1975 *Random Point Processes* (New York: Wiley)
[2] Rieke F, Warland D, de Ruyter van Steveninck R and Bialek W 1996 *Spikes: Exploring the Neural Code* (Cambridge, MA: MIT Press)
[3] Dayan P and Abbott L F 2001 *Theoretical Neuroscience* (Cambridge, MA: MIT Press)
[4] Ventura V, Carta R, Kass R E, Gettner S N and Olson C R 2002 *Biostatistics* **3** 1
[5] Kass R E, Ventura V and Cai C 2003 *Network* **14** 5
[6] Good I J 1965 *The Estimation of Probabilities: An Essay on Modern Bayesian Methods* (Cambridge, MA: MIT Press)

[7] Akaike H 1980 *Bayesian Statistics* ed J M Bernardo, M H DeGroot, D V Lindley and A F M Smith (Valencia: University Press) p 143
[8] MacKay D J C 1992 *Neural Comput.* **4** 415
[9] Bruce A D and Saad D 1994 *J. Phys. A: Math. Gen.* **27** 3355
[10] Carlin B P and Louis T A 2000 *Bayes and Empirical Bayes Methods for Data Analysis* 2nd edn (New York: Chapman and Hall)
[11] Bialek W, Callan C G and Strong S P 1996 *Phys. Rev. Lett.* **77** 4693
[12] Cox D R and Lewis P A W 1966 *The Statistical Analysis of Series of Events* (New York: Wiley)
[13] Daley D J and Vere-Jones D 2003 *An Introduction to the Theory of Point Processes* vol 1, 2nd edn (New York: Springer)
[14] Feynman R P and Hibbs A R 1965 *Quantum Mechanics and Path Integrals* (New York: McGraw-Hill)
[15] Coleman S 1975 *Aspects of Symmetry* (Cambridge: Cambridge University Press)
[16] Schulman L S 1981 *Techniques and Applications of Path Integration* (New York: Wiley)
[17] Kleinert H 2004 *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, and Financial Markets* (Singapore: World Scientific)
[18] Dempster A P, Laird N M and Rubin D B 1977 *J. R. Stat. Soc.* B **39** 1
[19] Smith A C and Brown E N 2003 *Neural Comput.* **15** 965
[20] Koyama S and Shinomoto S 2005 *J. Phys. A: Math. Gen.* **38** L531
[21] Tuckwell H C 1988 *Introduction to Theoretical Neurobiology* (Cambridge: Cambridge University Press)
[22] Koyama S and Shinomoto S 2004 *J. Phys. A: Math. Gen.* **37** 7255
[23] Shimazaki H and Shinomoto S 2007 *Neural Comput.* at press