



ELEMENTS OF STATISTICAL INFERENCE (LECTURE 2A)

BALANCED EXCITATION AND INHIBITION IN INTEGRATE-AND-FIRE MODELS (LECTURE 2B)

Rob Kass

Department of Statistics
Machine Learning Department
Center for the Neural Basis of Cognition
Carnegie Mellon University

PLAN FOR LECTURES

1. Motivation: Neural coding
2. Statistical concepts
Shadlen and Newsome (1998), section 1
3. Motivation: More advanced methods
Statistical thinking
4. Point process analysis via generalized regression

OUTLINE OF LECTURE 2

- 2a. Quick review of common probability distributions
- 2b. Maximum Likelihood (ML) estimation
- 2c. Quick review of LLN and CLT
- 2d. Properties of ML estimators
- 2e. Point process formalization
- 2g. Shadlen and Newsome (1998), section 1

Important distributions:

Binomial (and Bernoulli)

Poisson

Normal

Exponential

Gamma

Beta

Inverse Gaussian

Important distributions:

Binomial (and Bernoulli)

Poisson

Normal

Exponential

Gamma

Beta

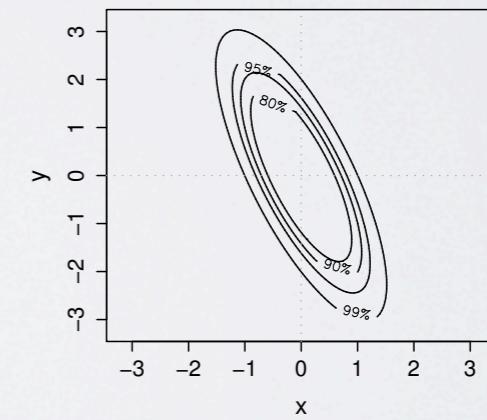
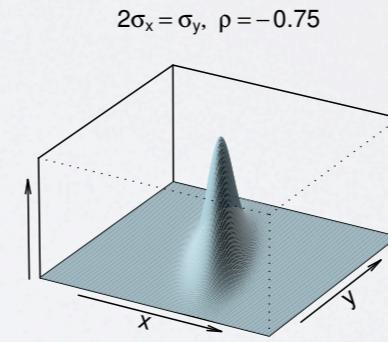
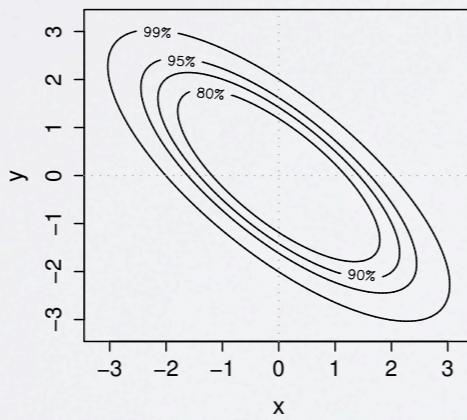
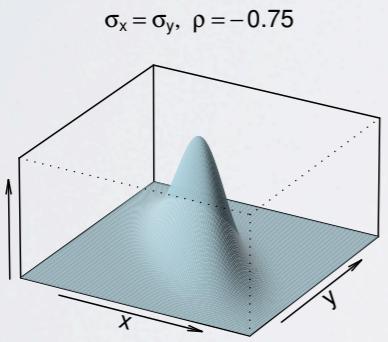
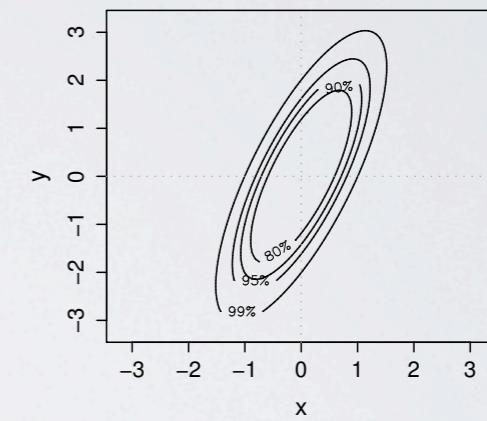
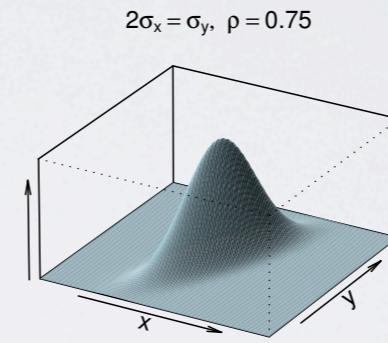
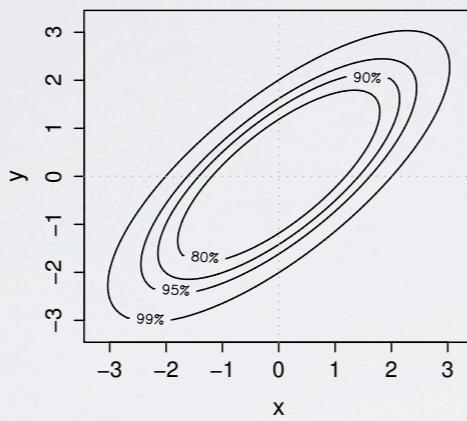
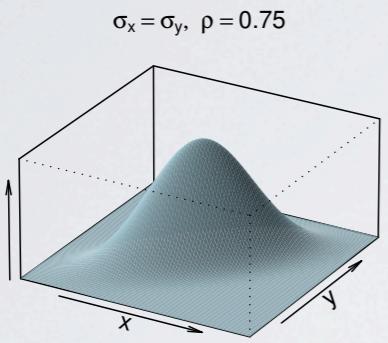
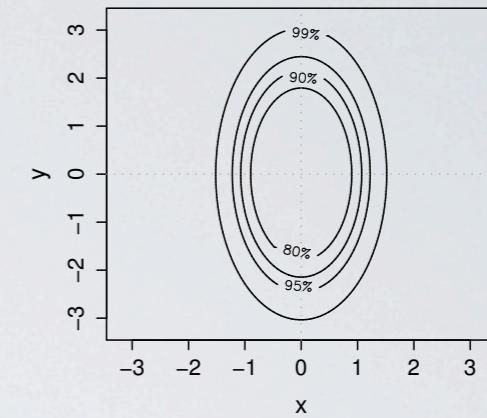
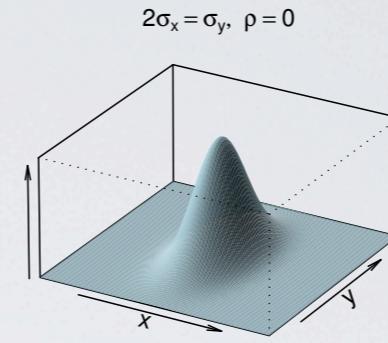
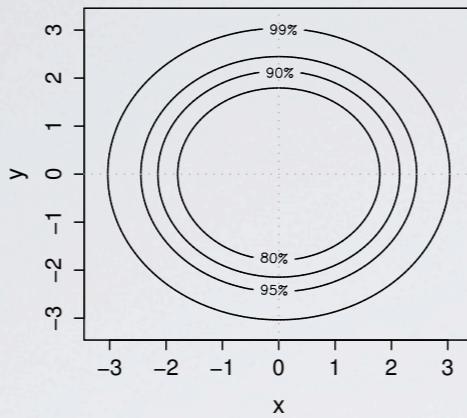
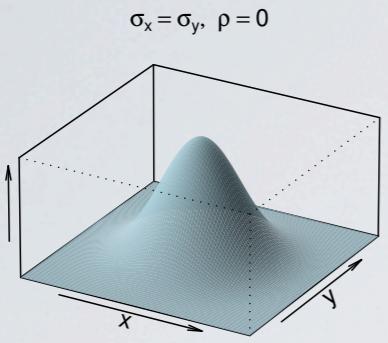
Inverse Gaussian

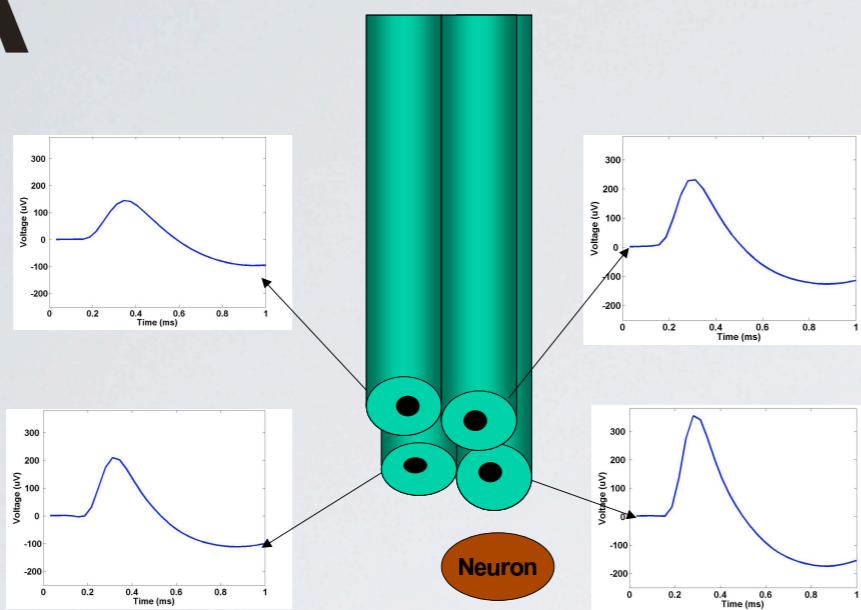
Note: all are *exponential families*

Example 1.2.1 Blindsight in patient P.S. Marshall and Halligan (1988, *Nature*, 336: 766–767) reported an interesting neuropsychological finding from a particular patient, identified as P.S. This patient was a 49 year-old woman who had suffered damage to her right parietal cortex that reduced her capacity to process visual information coming from the left side of her visual space. For example, she would frequently read words incorrectly by omitting left-most letters (“smile” became “mile”) and when asked to copy simple line drawings, she accurately drew the right-hand side of the figures but omitted the left-hand side without any conscious awareness of her error. To show that she could actually see what was on the left but was simply neglecting it—a phenomenon known as *blindsight*—the examiners presented P.S. with a pair of cards showing identical green line drawings of a house, except that on one of the cards bright red flames were depicted on the left side of the house. They presented to P.S. both cards, one above the other (the one placed above being selected at random), and asked her to choose which house she would prefer to live in. She thought this was silly “because they’re the same” but when forced to make a response chose the non-burning house on 14 out of 17 trials. This would seem to indicate that she did, in fact, see the left side of the drawings but was unable to fully process the information. But how convincing is it that she chose the non-burning house on 14 out of 17 trials? Might she have been guessing?

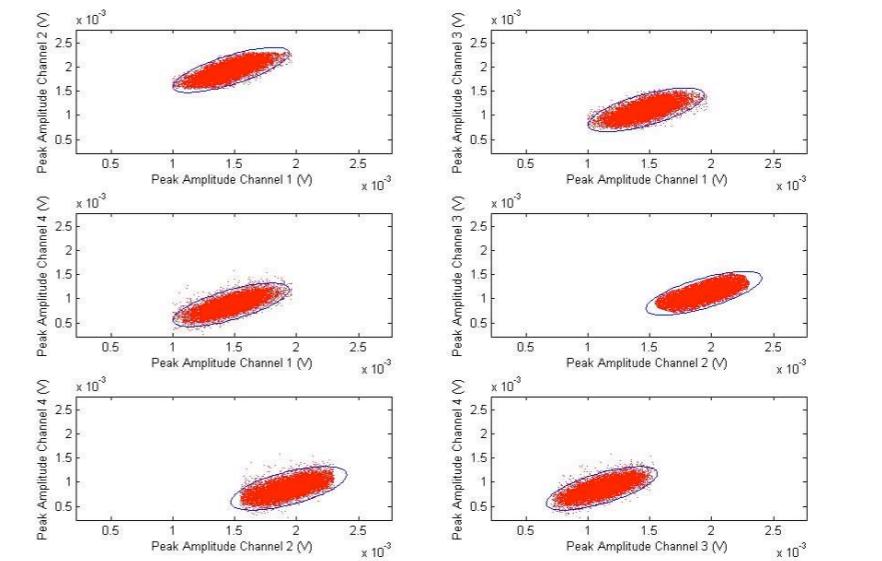
In thinking about binomial assumption for a random variable X one generally ponders whether it is reasonable to think of X as a sum of Bernoulli trials with the independence and homogeneity assumptions. Similarly, in the Poisson case, one typically asks whether the count variable X could be considered a sum of Bernoulli trials for small p (and large n). The first requirement is that

Neuronal spike counts are sometimes assumed to be Poisson-distributed. Let us consider the underlying assumptions in this case. First, if measurements are made on a single neuron to a resolution of 1 millisecond or less, it is the case that a sequence of dichotomous firing events will be observed: in any given time bin (e.g., any given millisecond) the neuron either will or will not have an action potential, and it can not have two. But are these events independent? The membrane of a neuron undergoes changes that alter its propensity to fire immediately after it has fired. In particular, there is a refractory period during which it can not fire again. This clearly violates the assumption of independence. In addition, there may be a build-up of ions, or activity in the local neural network, that makes a neuron more likely to fire if it has fired recently in the past (it may be “bursting”). This again would be a violation of independence. In many experiments such violations of independence produce markedly non-Poisson count distributions and turn out to have a substantial effect, but in others the effects are relatively minor and may be ignored. We indicated that, in the case of vesicle release of neurotransmitters, the homogeneity assumption is not needed in order to apply the Poisson approximation. The same is true for neuronal spike counts: the spike probabilities can vary across time and still lead to Poisson-distributed counts. The key assumption, requiring thought, is independence. On the other hand, the question of whether it is safe to assume Poisson variation remains an empirical matter, subject to statistical examination. As in nearly all statistical situations, judgment of the accuracy of the modeling assumptions—here, the accuracy of the Poisson distribution in describing spike count variation—will depend on the analysis to be performed.

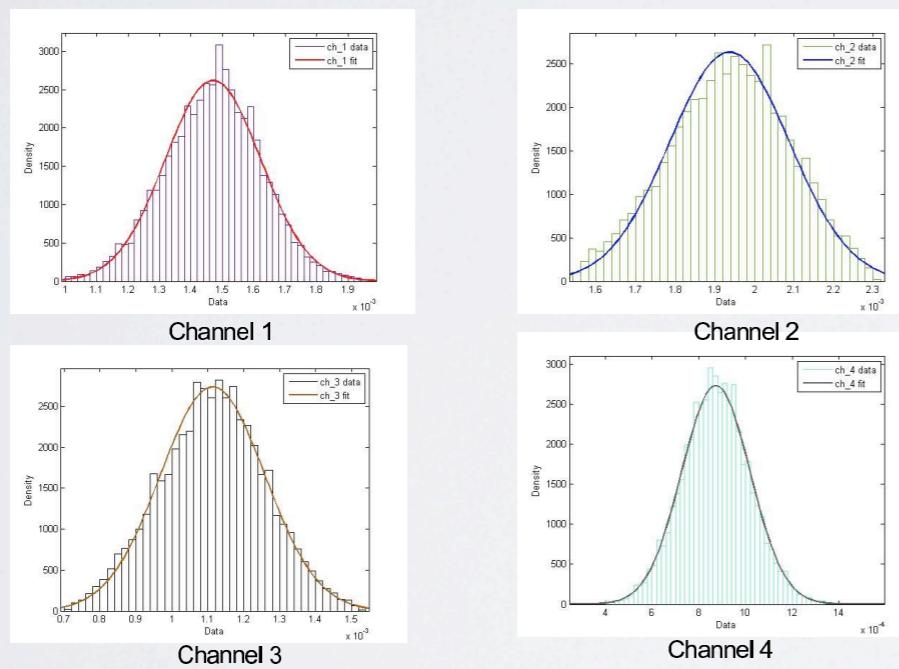


A**B**

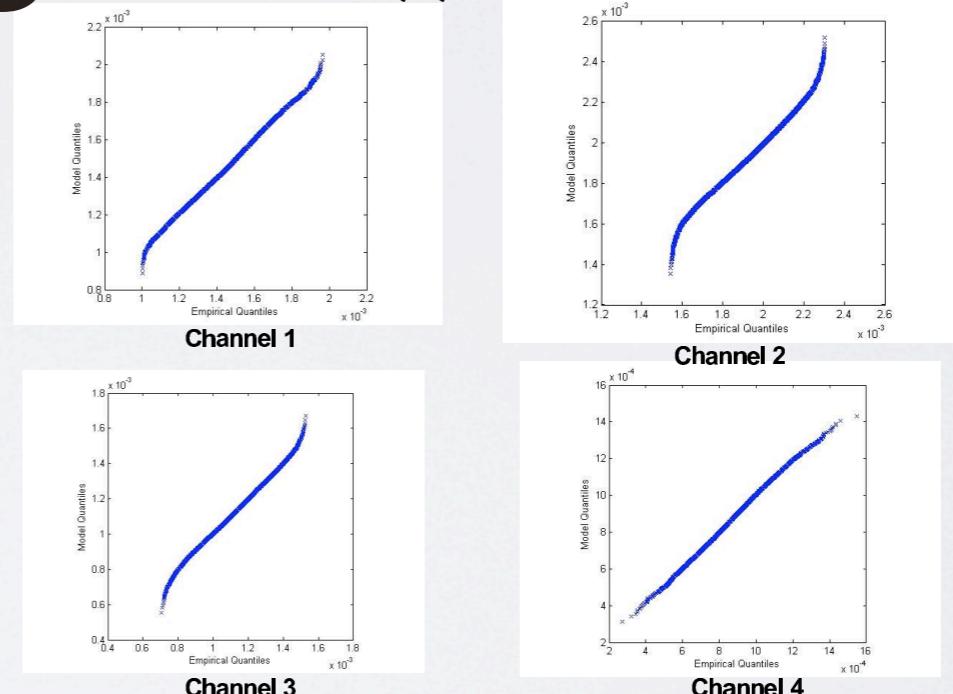
Six Bivariate Plots of Tetrode Channel Recordings With 95% Probability Contour

**C**

Histograms of Spike Events By Channel

**D**

Q-Q Plots



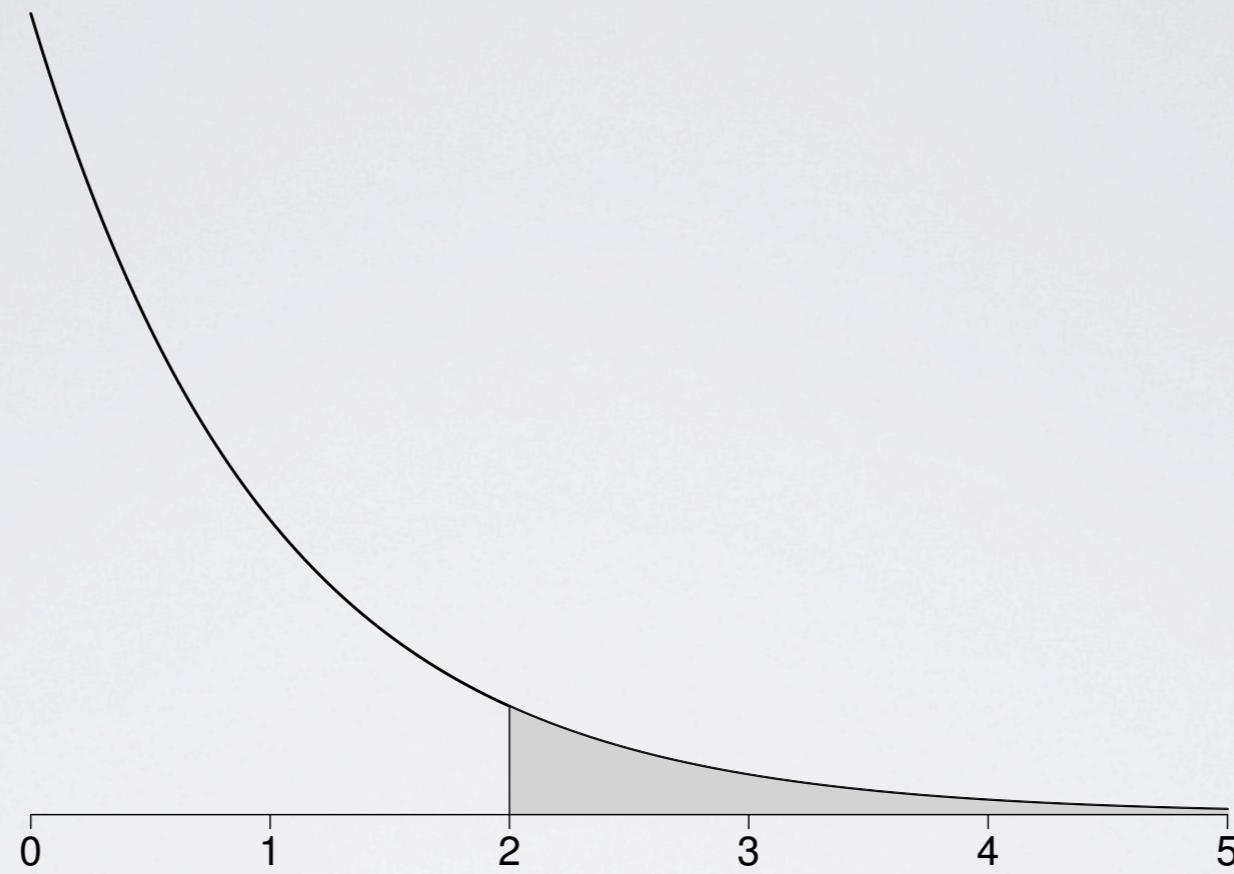
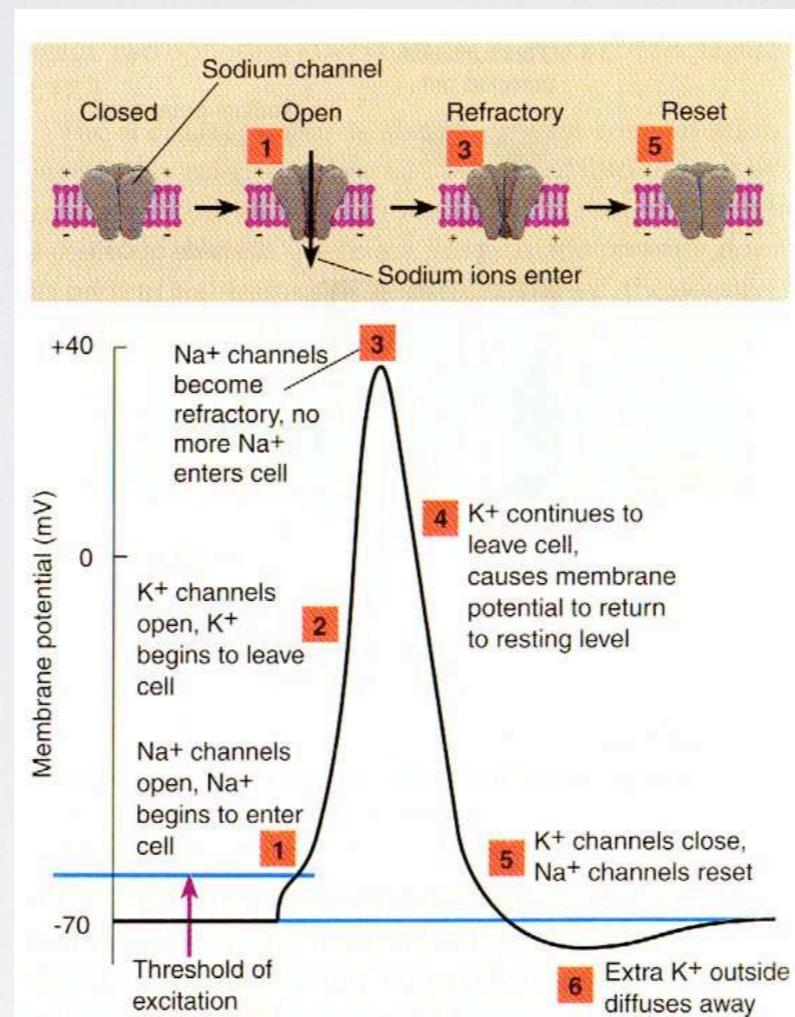


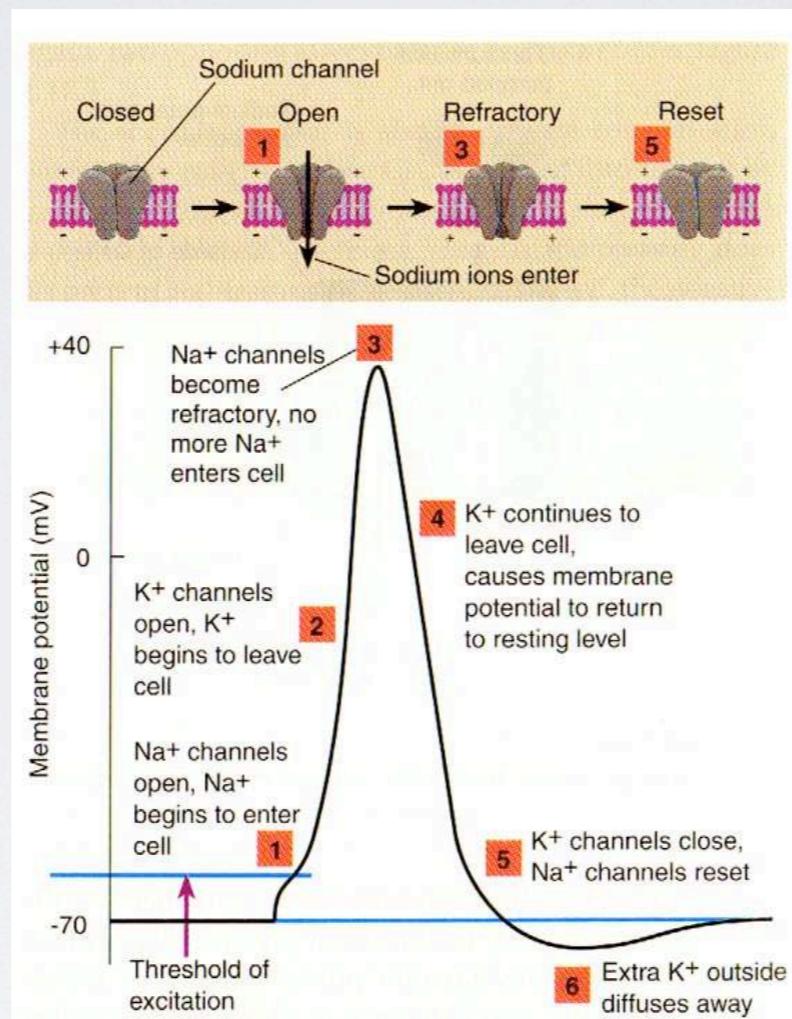
Figure 3.5: The pdf of a random variable X having an exponential distribution with $\lambda = 1$. The shaded area under the pdf gives $P(X > 2)$.

Quick aside on Hodgkin-Huxley model

Hodgkin-Huxley Model (1)

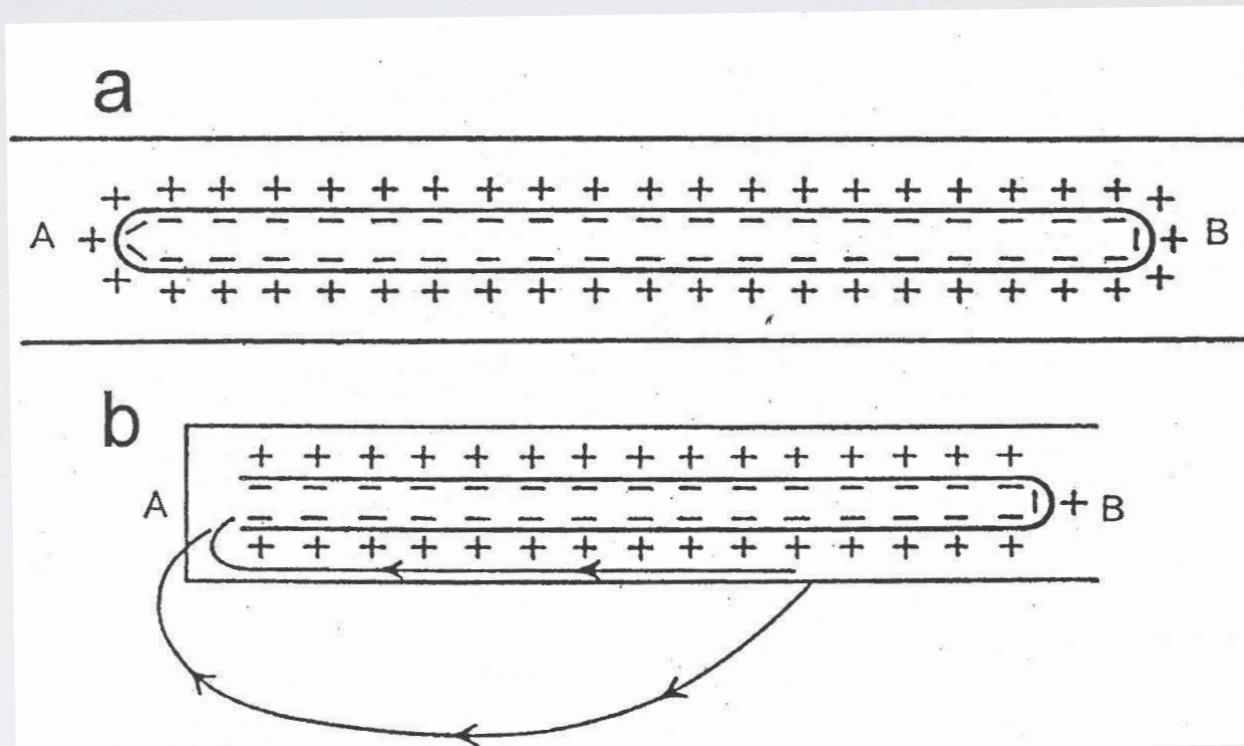


Hodgkin-Huxley Model (1)



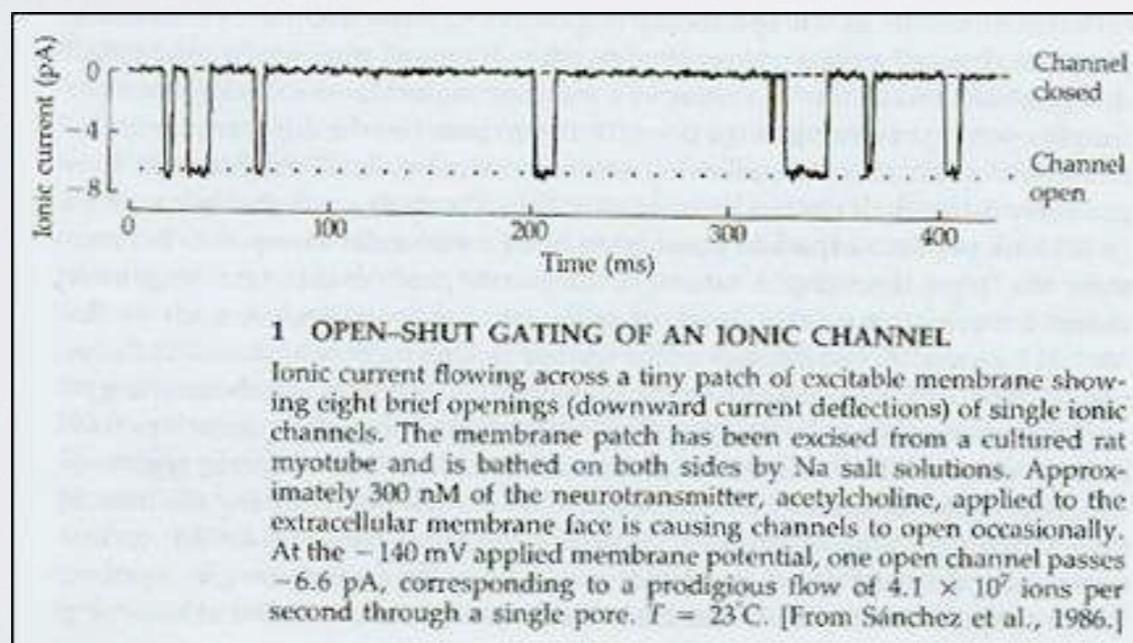
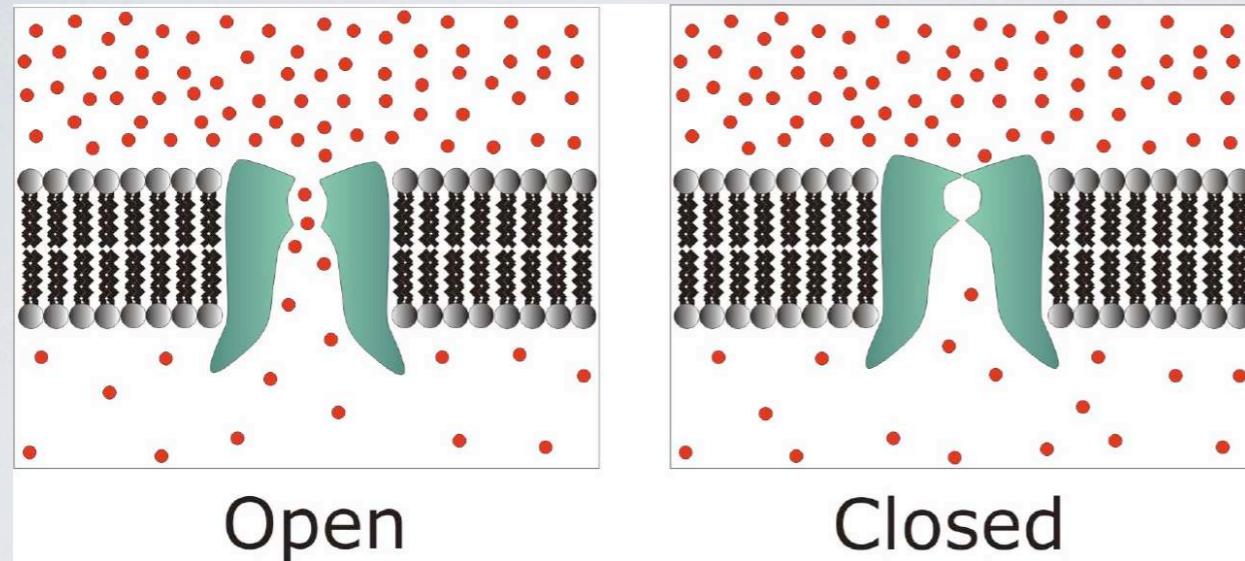
Current conception is based on more than 100 years of investigation.

The “membrane hypothesis.”

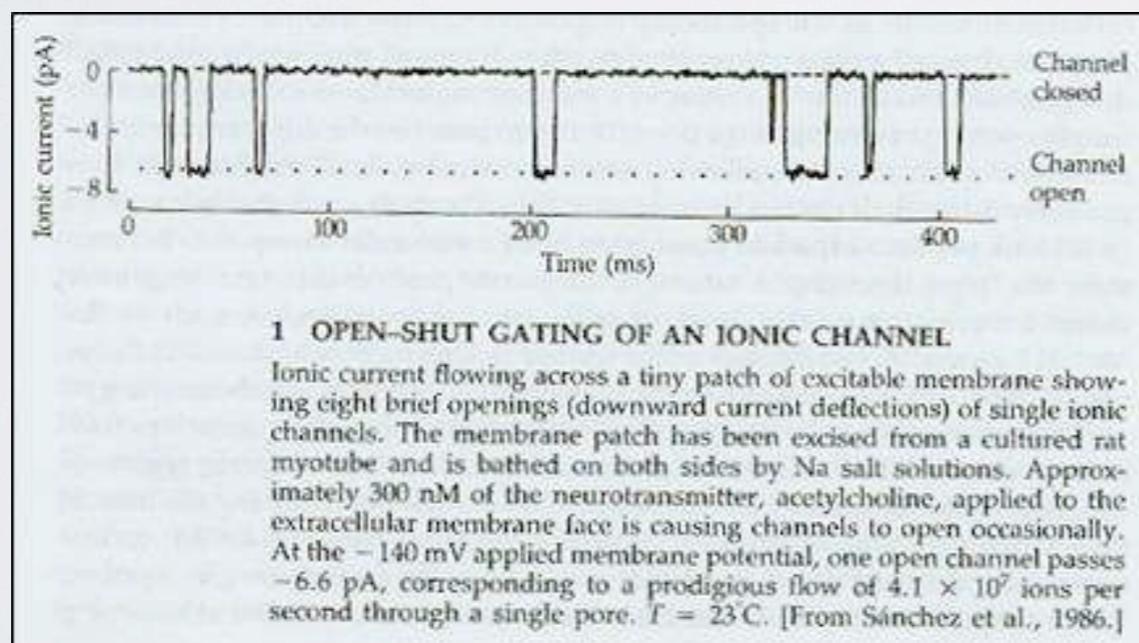
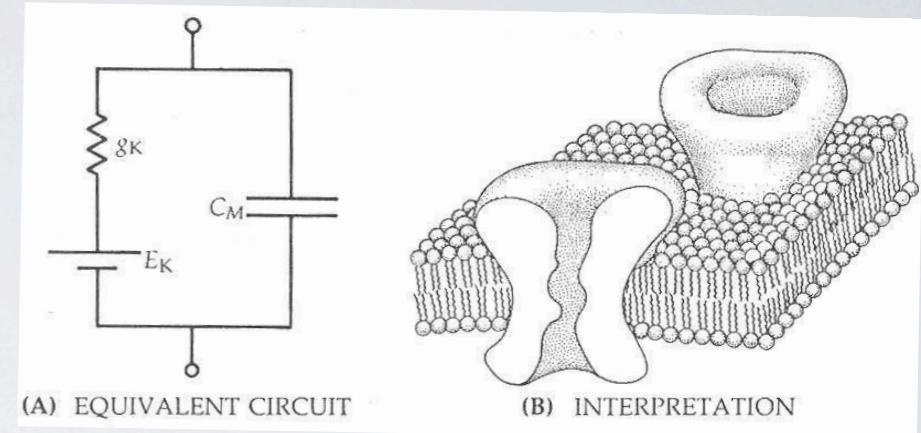
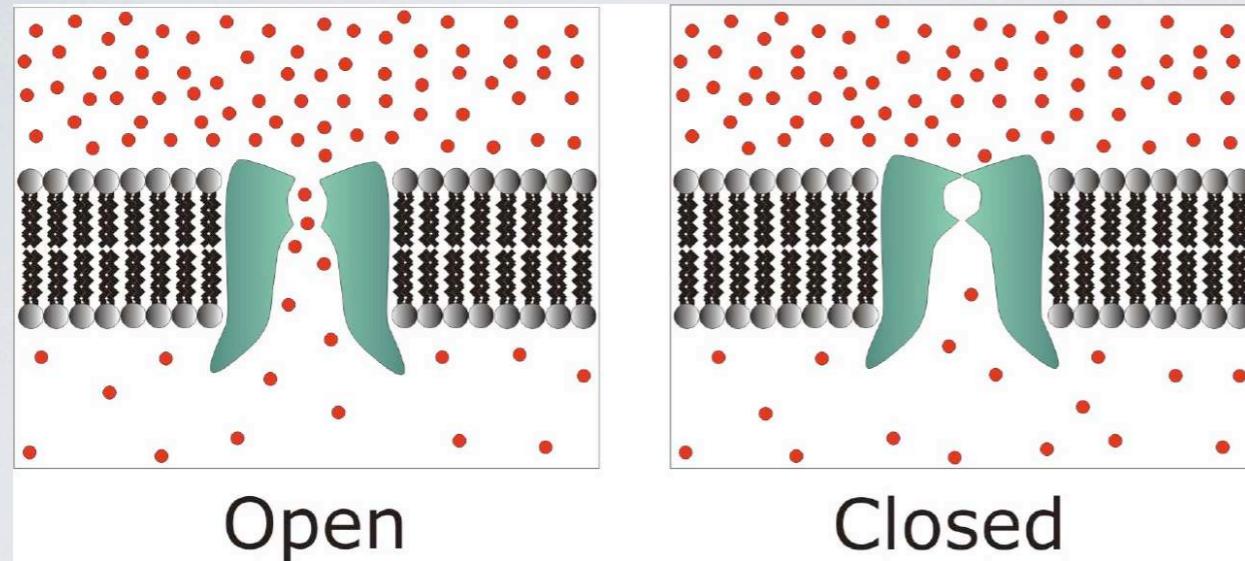


Membranes can be electrically excitable (Bernstein, 1902; based on Nernst, 1888).

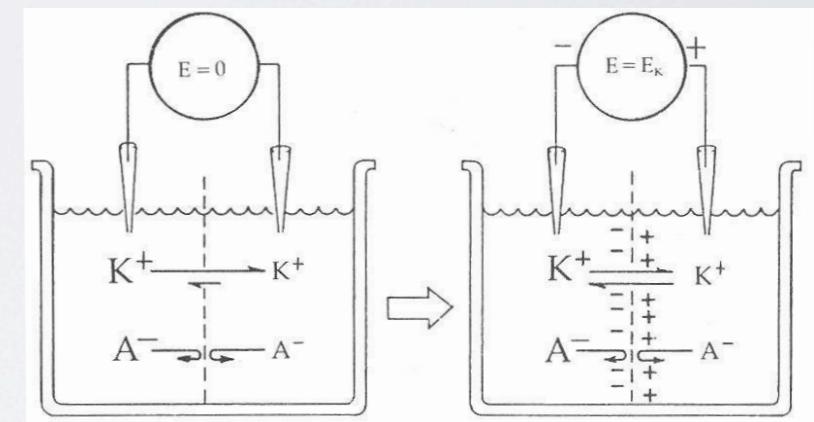
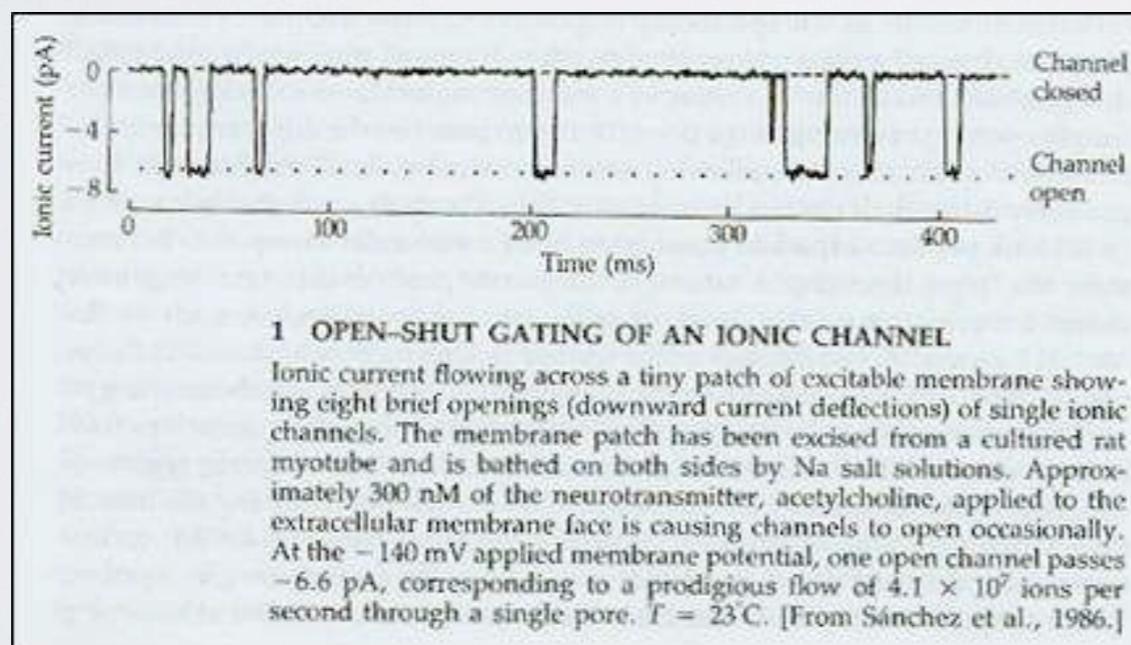
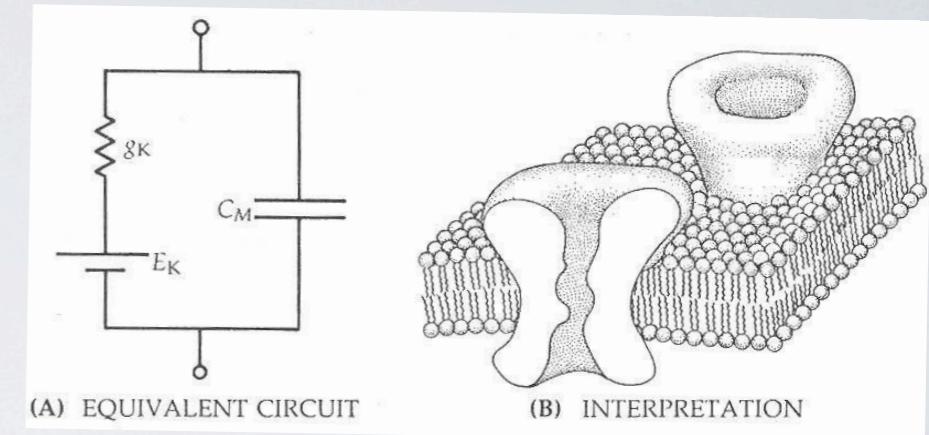
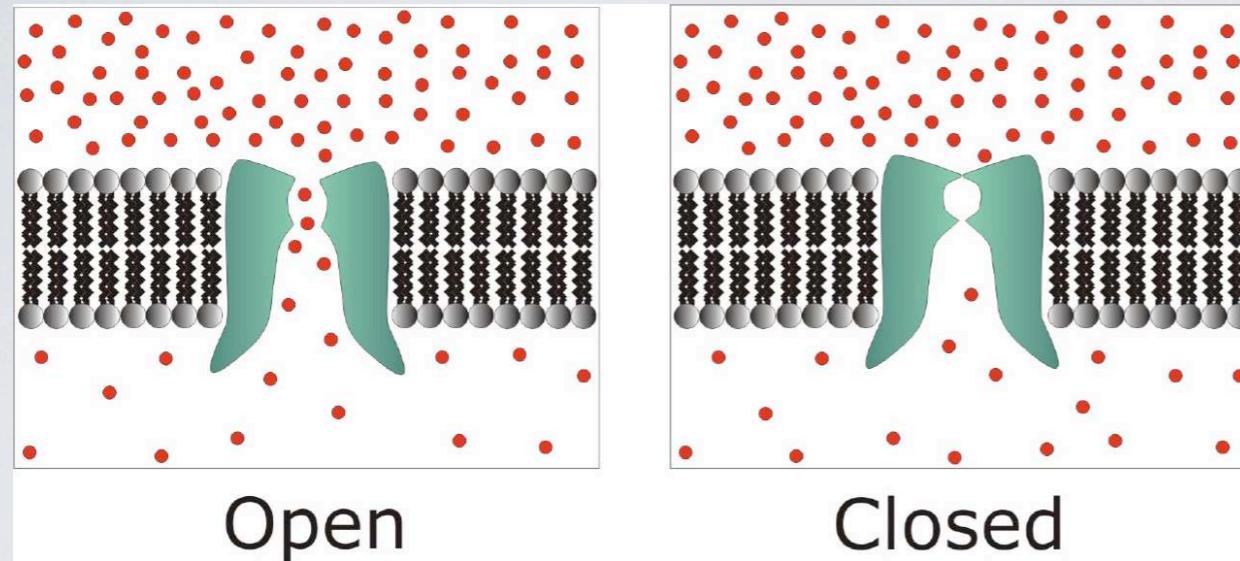
Ion channels gate the flow of ions across membranes (Cole and Curtis, 1939).



Ion channels gate the flow of ions across membranes (Cole and Curtis, 1939).



Ion channels gate the flow of ions across membranes (Cole and Curtis, 1939).



$$\begin{aligned}\text{membrane current} &= \text{capacitative current} + \text{ionic current} \\ &= C \frac{dV}{dt} + g(t)V(t)\end{aligned}$$

The diagram illustrates the decomposition of membrane current. At the top, the equation $C \frac{dV}{dt} + g(t)V(t)$ is shown, where the first term $C \frac{dV}{dt}$ is highlighted with a blue arrow pointing upwards, and the second term $g(t)V(t)$ is also highlighted with a blue arrow pointing upwards. Below this, two equations are provided: $C \frac{dV}{dt} = I(t)$ and $I(t) = g(t)V(t)$. To the right of the second term, a bracket groups it with the term $g(t)V(t)$, with the label '(actually several ionic currents)' positioned above the bracket.

$$\begin{aligned}C \frac{dV}{dt} &= I(t) \\ I(t) &= g(t)V(t)\end{aligned}$$

(actually several
ionic currents)

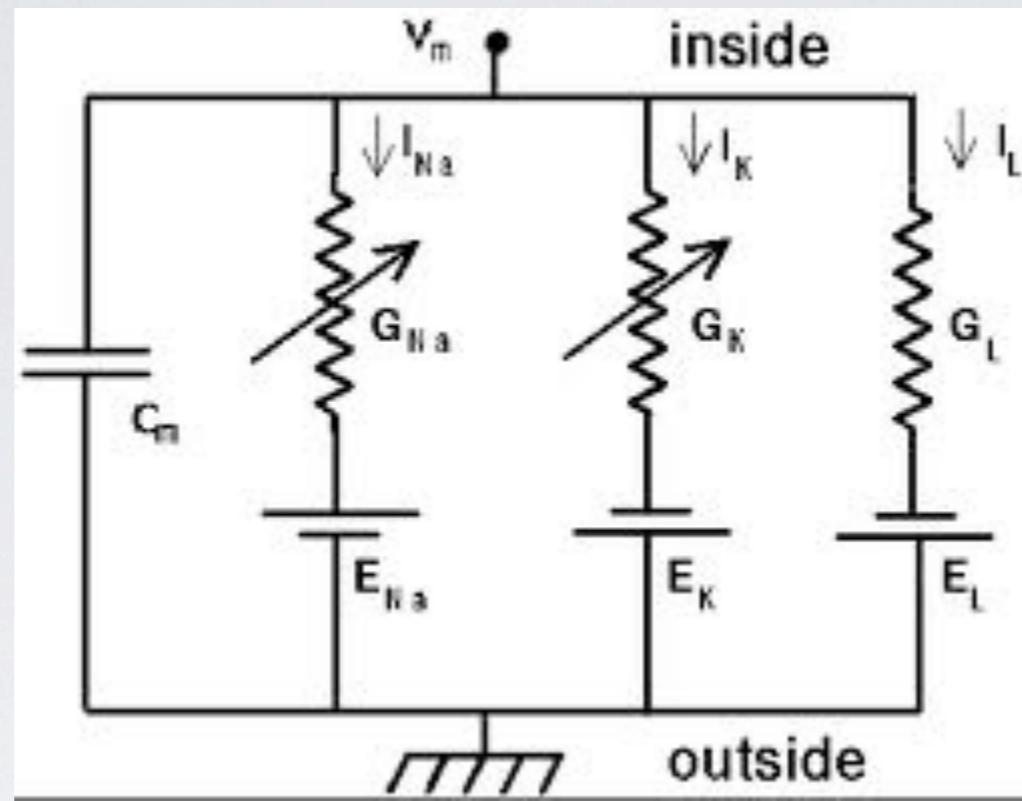
In 1952 Hodgkin and Huxley combined

membrane current = capacitative current + ionic currents

with quantitative description of

- potassium channel kinetics
- sodium channel kinetics

Hodgkin and Huxley Model (2)



$$\text{membrane current} = C \frac{dV}{dt} + I_K + I_{Na} + I_{leak}$$

$$\left. \begin{array}{l} dn = (1 - n(t))\alpha_n dt - n(t)\beta_n dt \\ dm = (1 - m(t))\alpha_m dt - m(t)\beta_m dt \\ dh = (1 - h(t))\alpha_h dt - h(t)\beta_h dt \end{array} \right\} \text{gating eqns}$$

$$\left. \begin{array}{l} I_K = \bar{g}_K n^4 (V - V_k) \\ I_{Na} = \bar{g}_{Na} m^3 h (V - V_{Na}) \\ I_{leak} = g_{leak} (V - V_{leak}) \\ \text{membrane current} = C \frac{dV}{dt} + I_K + I_{Na} + I_{leak} \end{array} \right\} \text{H-H model}$$

150-year-old problem of “animal electricity” solved; correct predictions of

- conductances (shown above);
- form of action potential (obtained by laborious solution of equations; 8 hours per 5 milliseconds), including “undershoot”;
- change in action potentials with varying concentrations of sodium;
- number of sodium ions involved in inward flux;
- speed of action potential propagation;
- voltage curves for sodium and potassium separately.

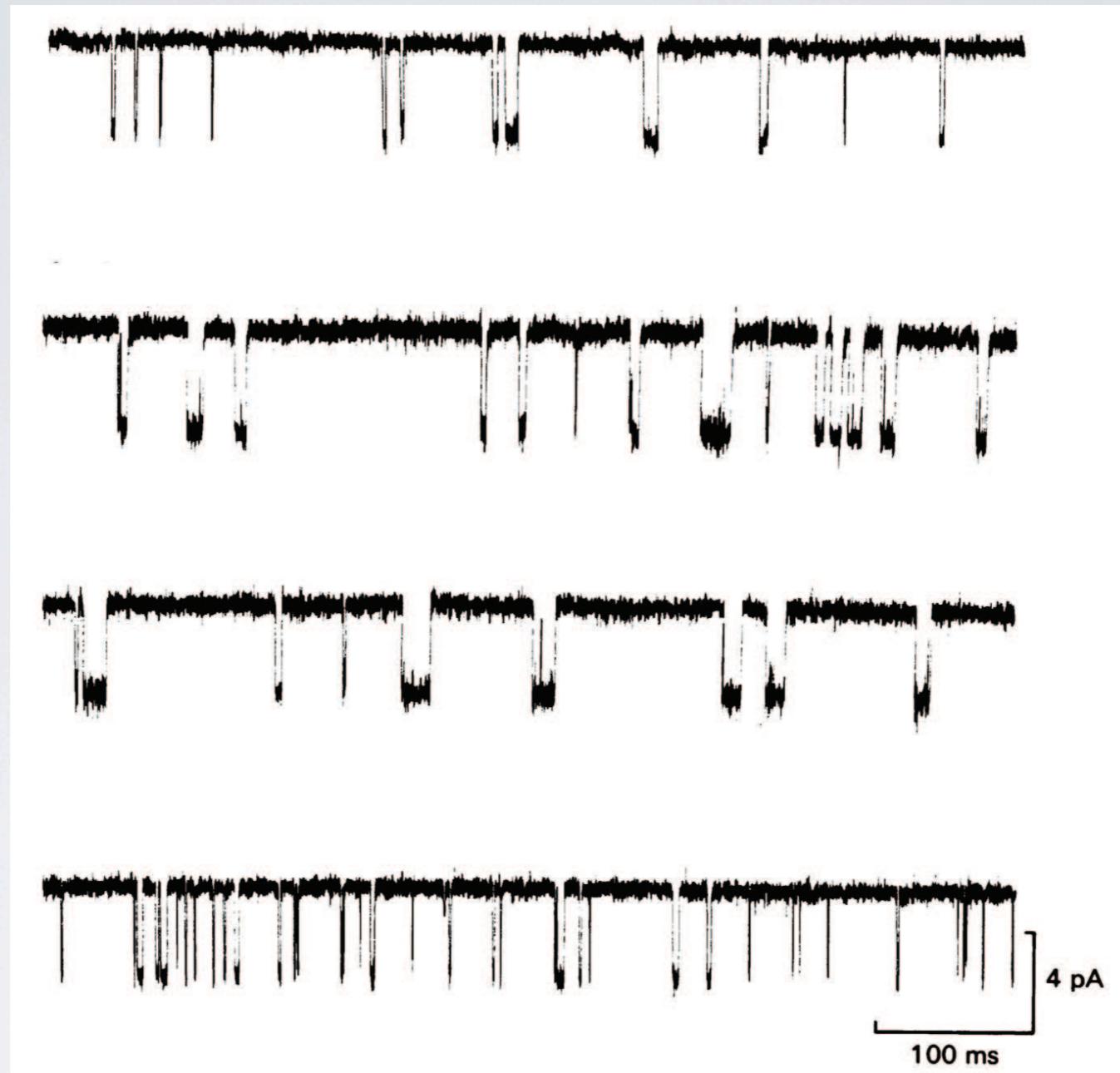


Figure 3.6: Current recordings from individual ion channels in the presence of acetylcholine-type agonists. The records show the opening (higher current levels) and closing (lower current levels), with the timing of opening and closing being stochastic. Modified from Colquhoun and Sakmann (1985).

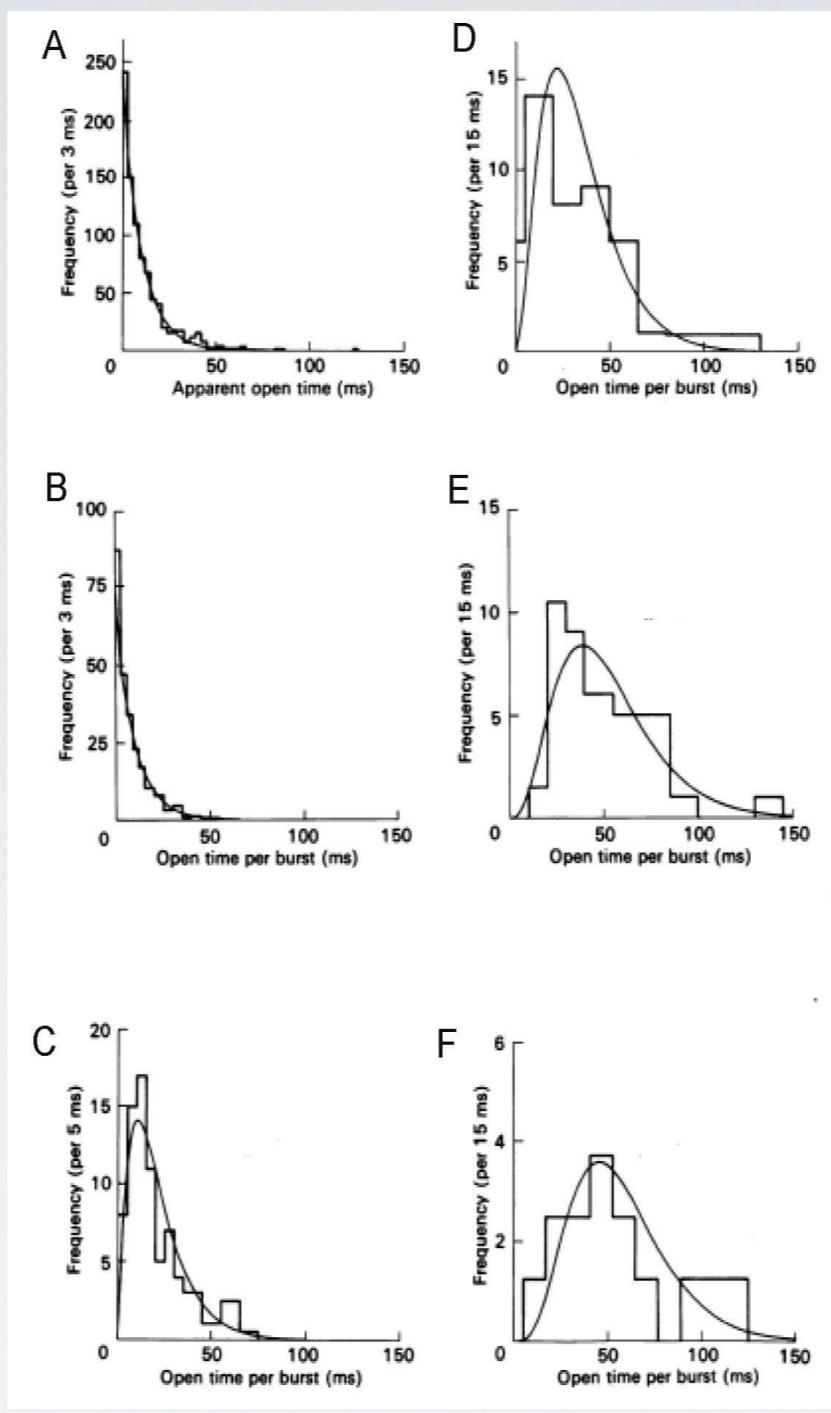


Figure 3.7: Duration of channel openings. Panel A depicts the distribution of burst durations for a particular agonist. Panel B displays the distribution of bursts for which there was only 1 opening, with an exponential pdf overlaid. This illustrates the good fit of the exponential distribution to the durations of ion channel opening. Panels C displays the distributions of bursts for which there were 2 apparent openings, with a gamma pdf, with shape parameter 2, overlaid. Panel C again indicates good agreement. Panels D-F show similar results, for bursts with 3-5 openings. Modified from Colquhoun and Sakmann (1985).

5.4.2 The exponential distribution is used to describe waiting times without memory.

We defined the exponential distribution in Equation (3.7), using it to illustrate calculations based on the pdf, and we showed how it may be applied to ion channel activation durations in Example 3.2.2. The exponential distribution is very special⁶ because of its “memoryless” property. To understand this, let X be the length of time an ion channel is open, and let us consider the probability that the channel will remain open for the next time interval of length h . For example, h might be 5 milliseconds. How do we write this? If we begin the moment the channel opens, i.e., at $x = 0$, the next interval of length h is $(0, h)$ and we want $P(X > h)$. On the other hand, if we begin at time $x = t$, for some positive t , such as 25 milliseconds, the interval in question is $(t, t + h)$ and we are asking for a *conditional* probability: if the channel is open at time t we must have $X > t$, so we are asking for $P(X > t + h | X > t)$. We say that the channel opening duration is memoryless if

$$P(X > t + h | X > t) = P(X > h) \quad (5.8)$$

for all $t > 0$ and $h > 0$. That is, if $t = 25$ milliseconds, the channel does not “remember” that it has been open for 25 milliseconds already; it still has the same probability of remaining open for the next 5 milliseconds that it had when it first opened. And this is true regardless of the time t we pick. The exponential distributions are the *only* distributions that satisfy Equation (5.8).

Contrast this memorylessness with, say, a uniform distribution on the interval $[0, 10]$, measured in milliseconds. According to this uniform distribution, the event (e.g., the closing of the channel) must occur within 10 milliseconds and initially every 5 millisecond interval has the same probability. In particular, the probability the event will occur in the first 5 milliseconds, i.e., in the interval $[0, 5]$, is the same as the probability it will occur in the last 5 milliseconds, in $[5, 10]$. Both probabilities are equal to $\frac{1}{2}$. However, if at time $t = 5$ milliseconds the event has not yet occurred then we are *certain* it will occur in the next half second $[5, 10]$, i.e., this probability is 1, which is

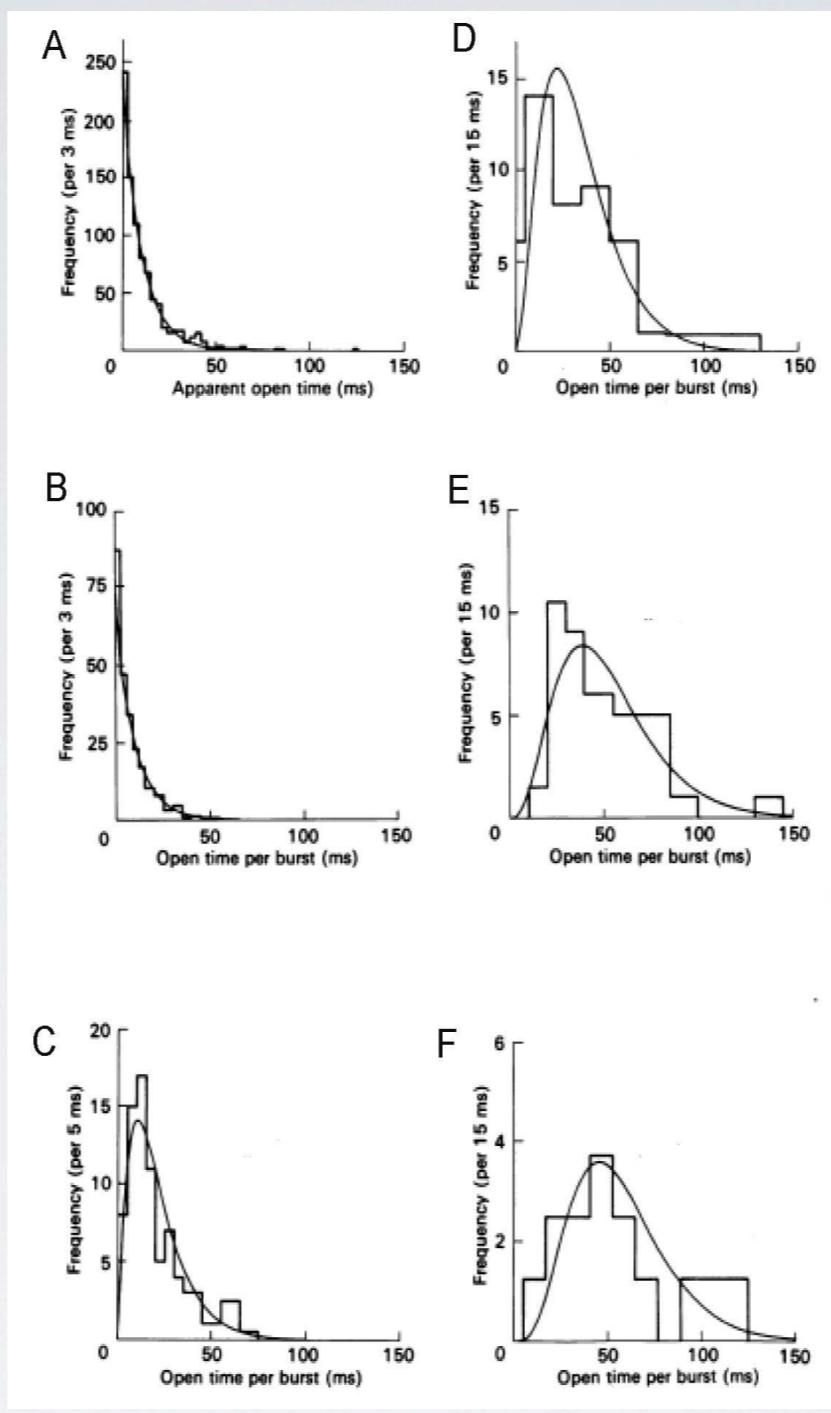
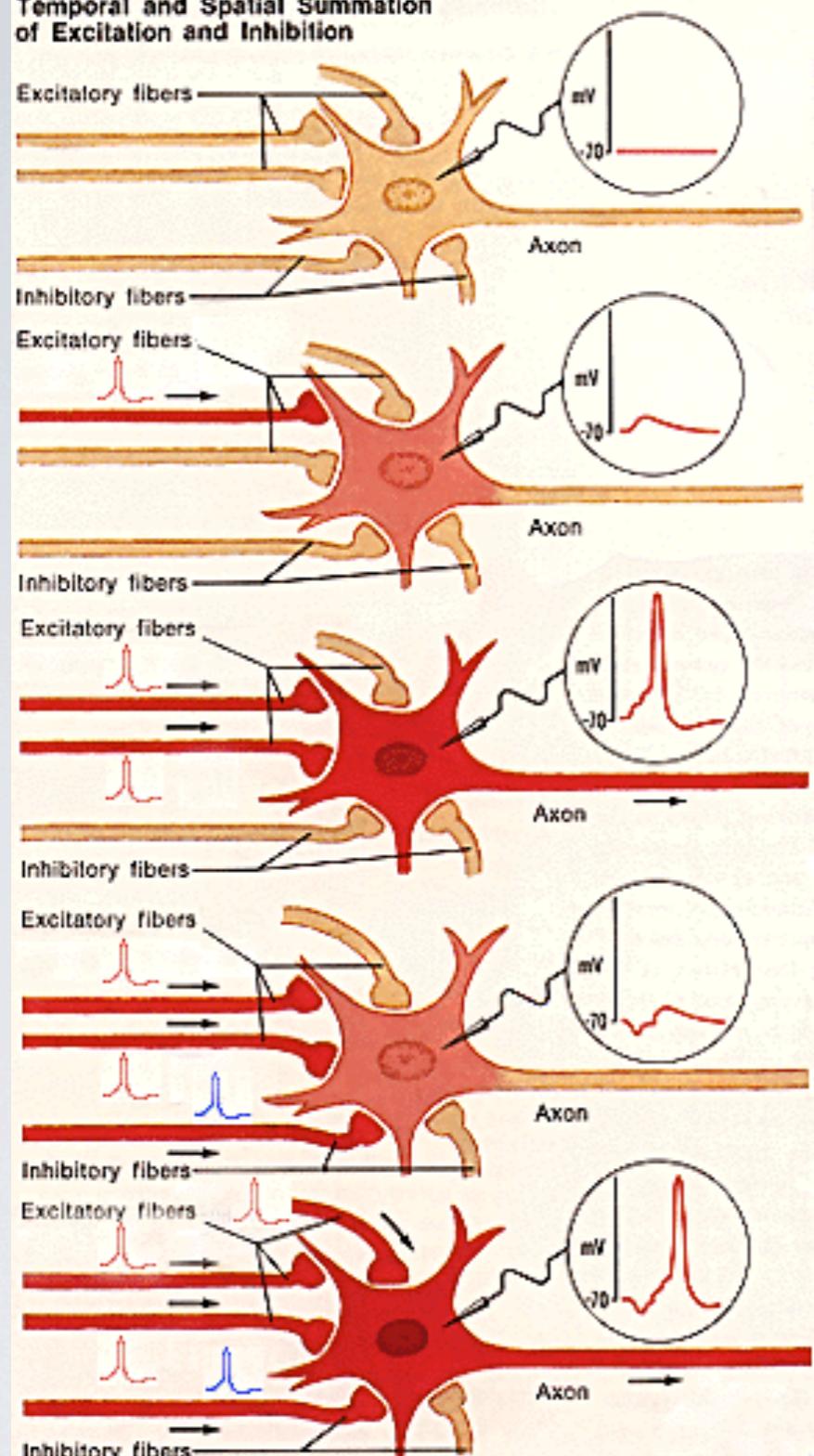
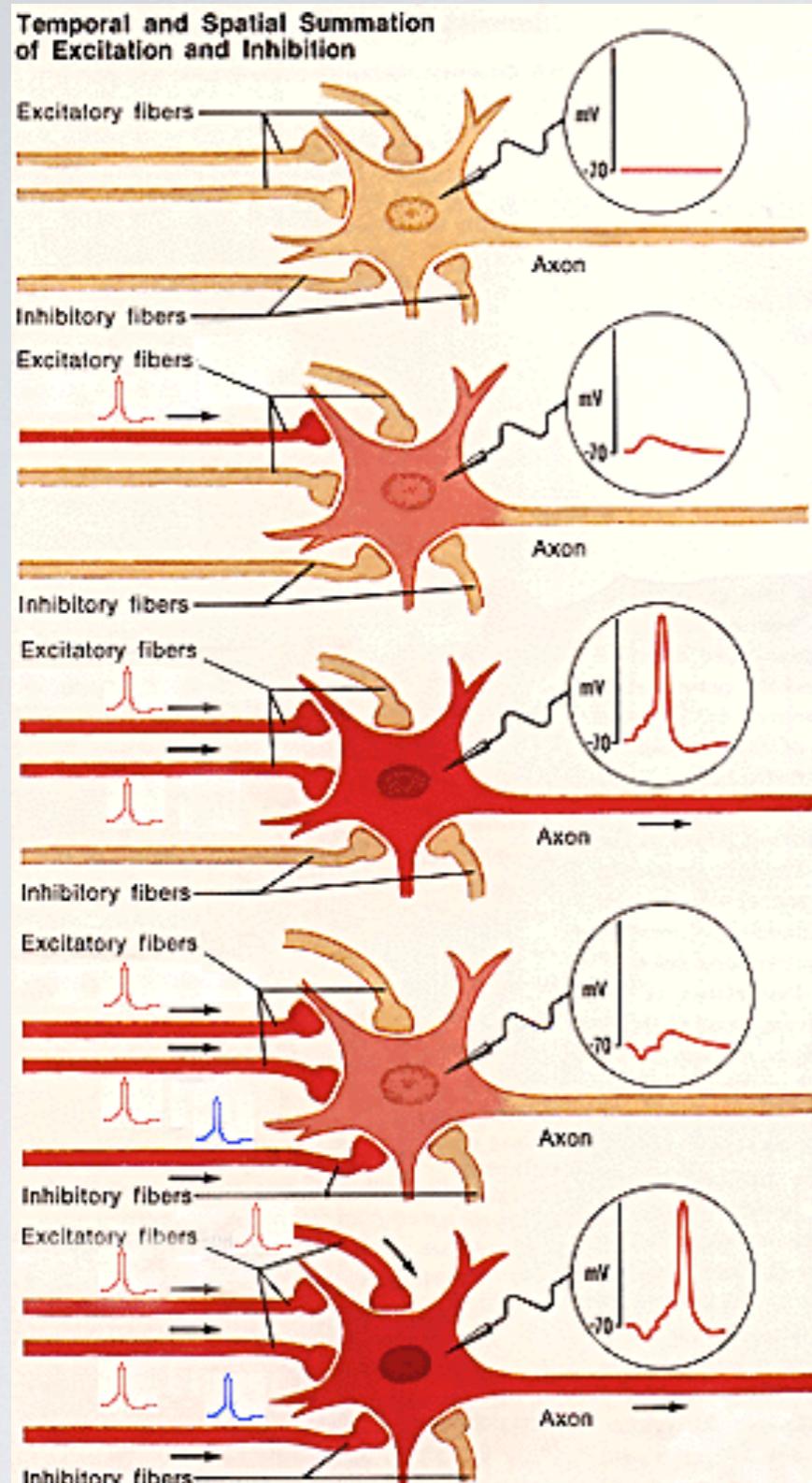


Figure 3.7: Duration of channel openings. Panel A depicts the distribution of burst durations for a particular agonist. Panel B displays the distribution of bursts for which there was only 1 opening, with an exponential pdf overlaid. This illustrates the good fit of the exponential distribution to the durations of ion channel opening. Panels C displays the distributions of bursts for which there were 2 apparent openings, with a gamma pdf, with shape parameter 2, overlaid. Panel C again indicates good agreement. Panels D-F show similar results, for bursts with 3-5 openings. Modified from Colquhoun and Sakmann (1985).

**Temporal and Spatial Summation
of Excitation and Inhibition**





integrate-and-fire
neuron

$$C \frac{dV}{dt} = I(t)$$

leaky stochastic
integrate-and-fire neuron

$$dV(t) = (-g(t)V(t) + I(t)) dt + \sigma dW_t$$

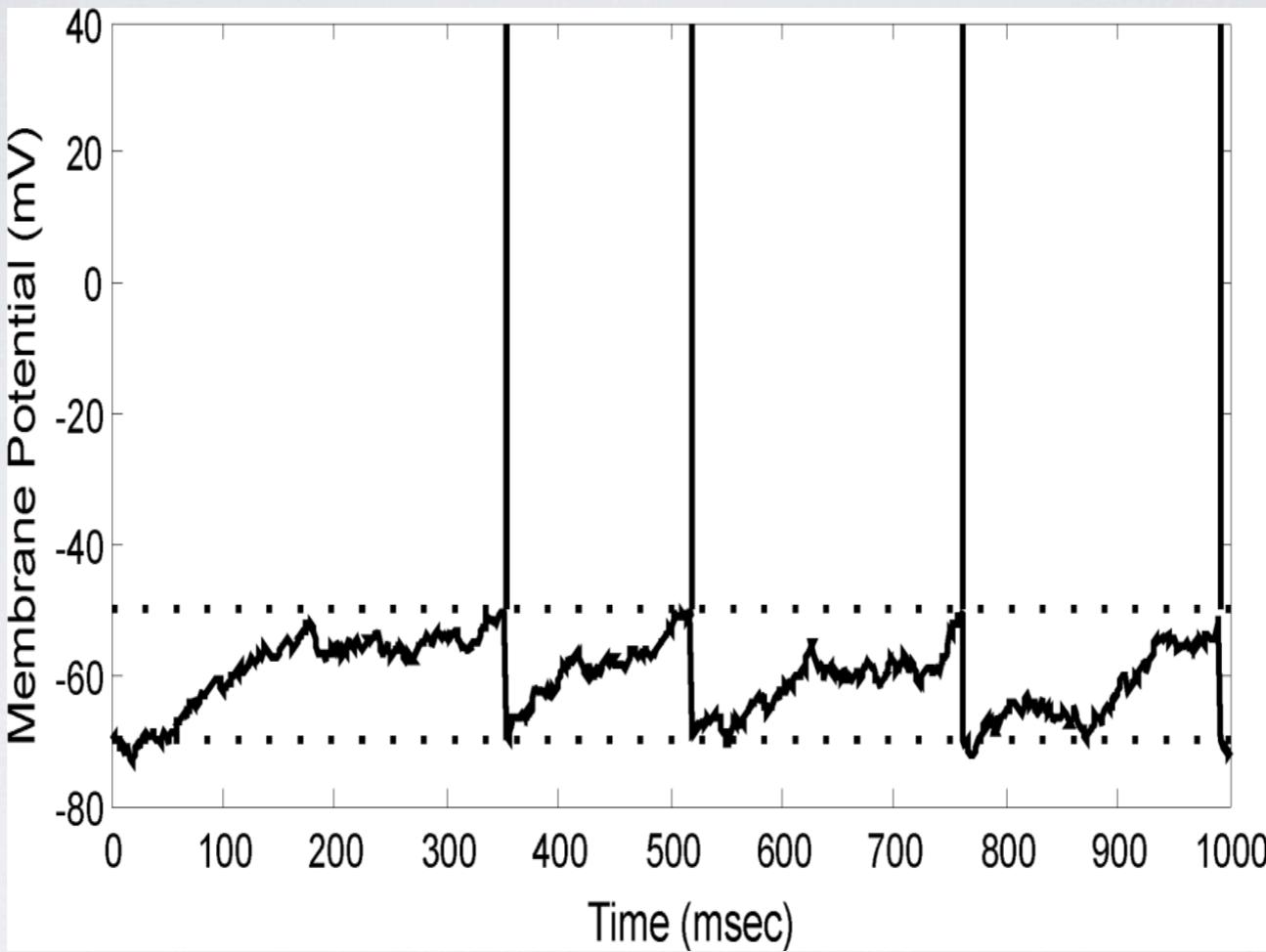


Figure 5.5: Example of an integrate-and-fire neuron. At each time step there is either an EPSP or an IPSP, with probabilities p and $1 - p$. For $p > 1 - p$ this creates a stochastic upward “drift” of the voltage (as the inputs are summed or “integrated”) until it crosses the threshold and the neuron fires. The neuron then resets to its baseline voltage. The resulting ISI distribution is approximately inverse Gaussian.

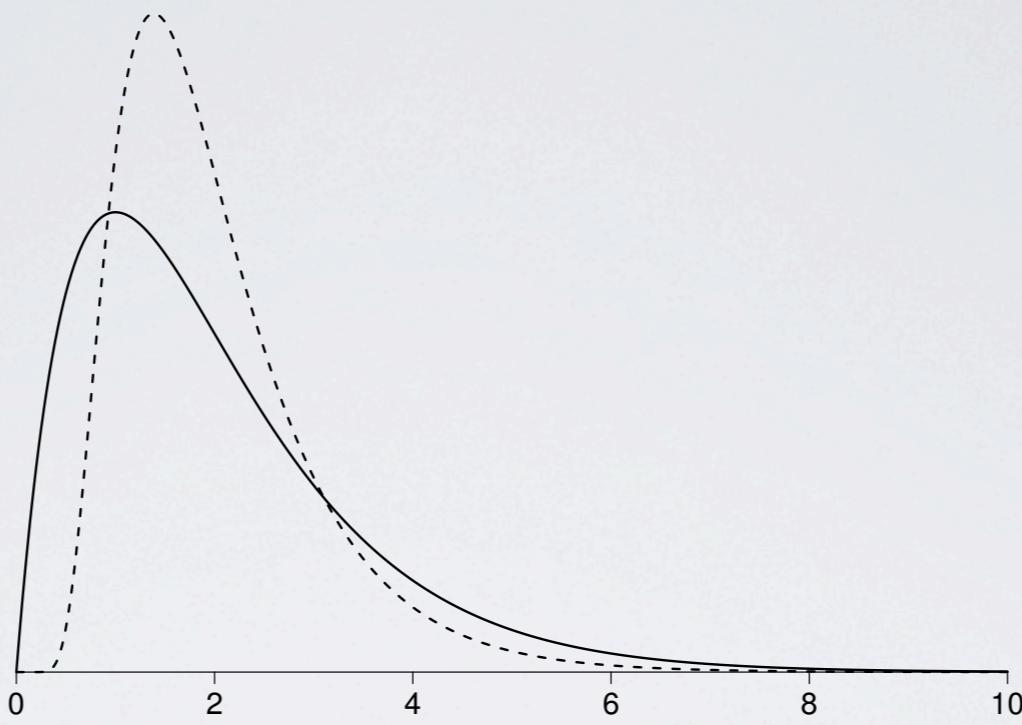


Figure 5.6: Inverse Gaussian pdf plotted together with a $\text{Gamma}(2, 1)$ pdf. The Inverse Gaussian (blue) has the same mean and variance as the Gamma. Note its convexity near 0.

OUTLINE OF LECTURE 2

- 2a. Quick review of common probability distributions
- 2b. Maximum Likelihood (ML) estimation ←
- 2c. Quick review of LLN and CLT
- 2d. Properties of ML estimators
- 2e. Point process formalization
- 2g. Shadlen and Newsome (1998), section 1

Maximum Likelihood (ML) Estimation

To introduce maximum likelihood estimation, let us begin by framing the estimation problem concretely, using the binomial, and let us write the binomial pdf in the form

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

which was previously denoted by $f(x) = P(X = x)$, with p replacing θ . Here the notation $f(x|\theta)$ is used to imply that we are examining the pdf of X given the value of θ . The binomial pdf describes the probabilities to be attached to varying possible values $X = x$ for a given fixed value of θ . That is, once we plug in a value of θ we have completely determined the pdf for all values of x . The problem of estimation, however, attempts to find a sensible guess at θ given that $X = x$ has been observed. It thus reverses the situation: instead of assuming a value for θ and finding values of x , we must assume a value of $X = x$ and come up with a value of θ . In this sense, it is involves an *inverse* or *inductive* form of reasoning. The method of maximum likelihood chooses the value $\hat{\theta}$ of θ that assigns to the observed data x the highest possible probability:

$$f(x|\hat{\theta}) = \max_{\theta} f(x|\theta).$$

In the binomial problem we will, below, show that $\hat{\theta} = x/n$. In other words, maximum likelihood estimates the theoretical proportion (or propensity) θ by the observed proportion x/n .

MLE satisfies

$$f(x|\hat{\theta}) = \max_{\theta} f(x|\theta).$$

Likelihood function:

$$L(\theta) \propto f(x|\theta)$$

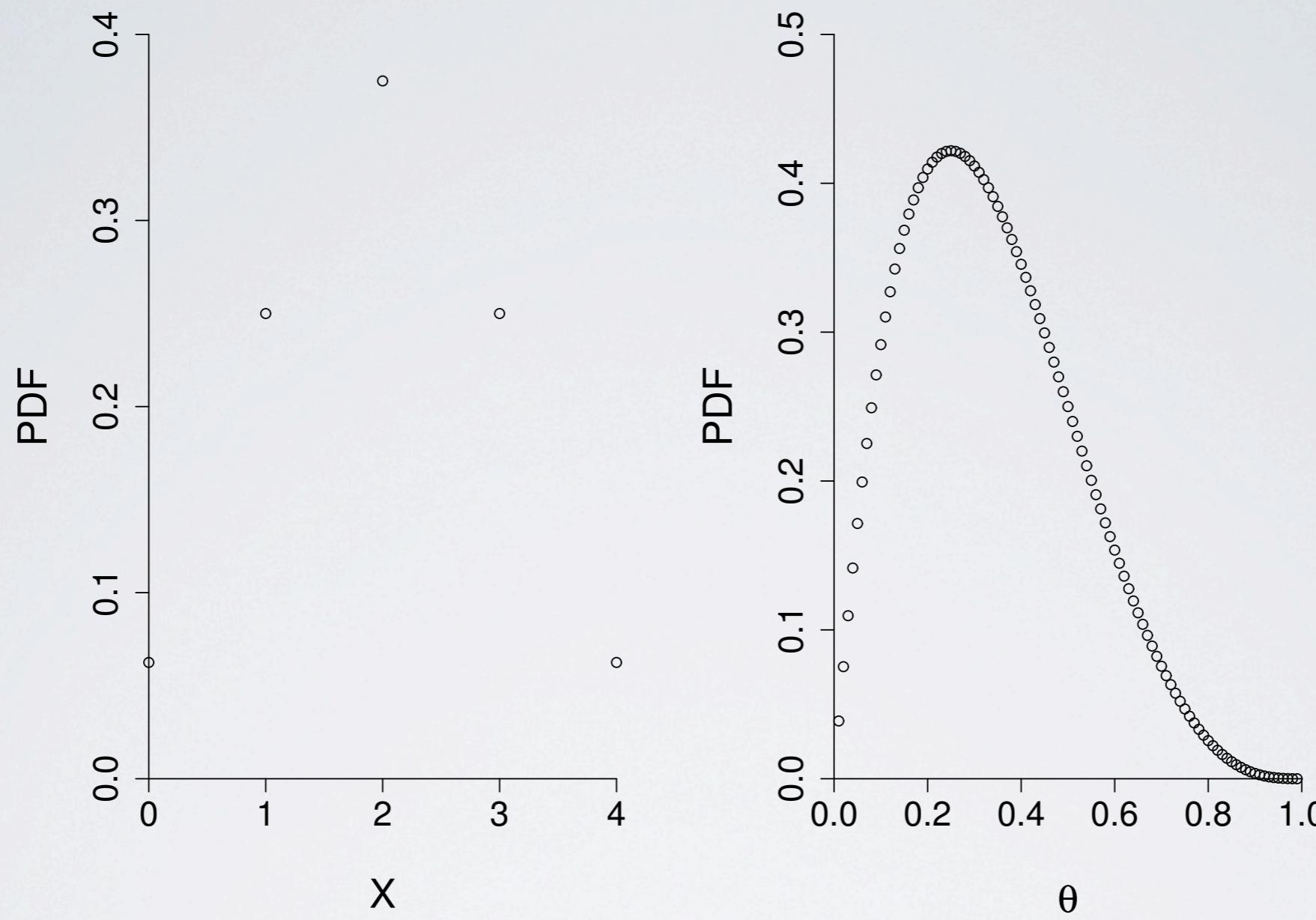


Figure 7.2: Comparison of pdf $f(x|\theta)$ when viewed as a function of x with θ fixed at $\theta = .5$ (on left) or of θ with x fixed at $x = 1$ (on right). On the right-hand side, the pdf is evaluated for 99 equally-spaced values of θ from .01 to .99.

$$0 = \ell'(\theta)$$

$$\ell'(\theta) = \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_m} \end{pmatrix}$$

$$I_{OBS}(\hat{\theta}) = -\ell''(\hat{\theta}).$$

$$\sqrt{I_{OBS}(\hat{\theta})}(\hat{\theta} - \theta) \xrightarrow{D} N(0, 1). \quad (8.18)$$

Result For large samples, under certain general conditions, the MLE $\hat{\theta}$ satisfies (8.18), so that its standard error is given by

$$SE = \frac{1}{\sqrt{-\ell''(\hat{\theta})}} \quad (8.19)$$

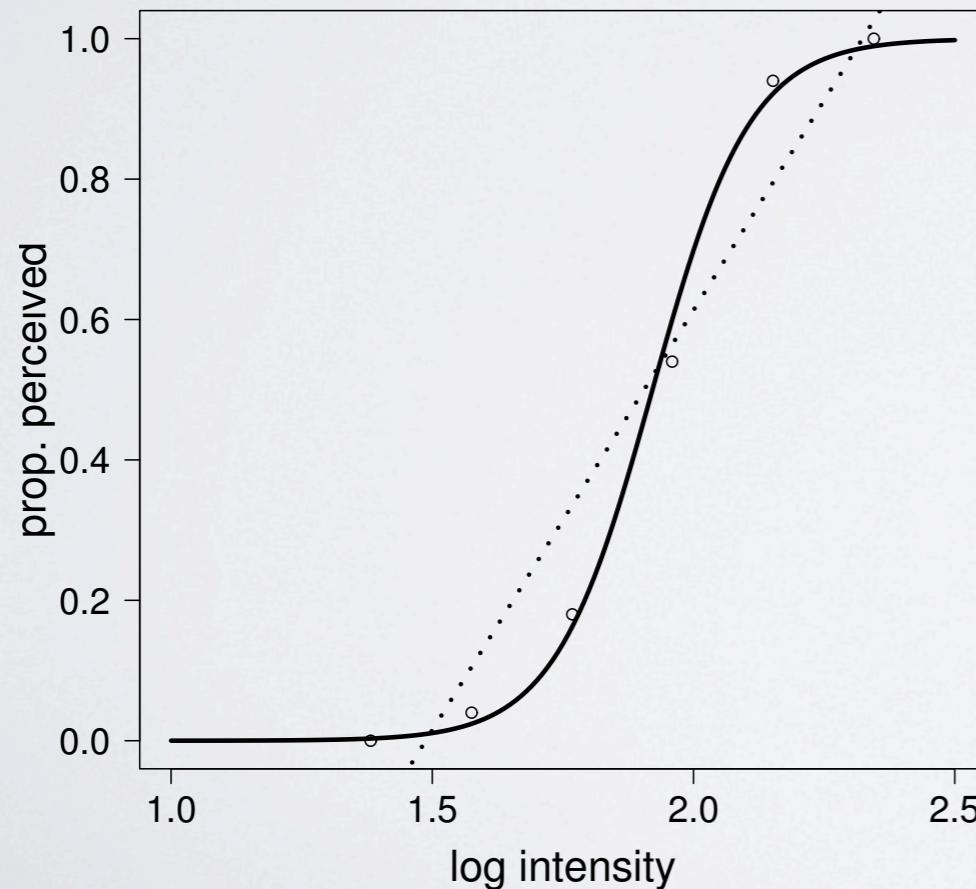
and an approximate 95% CI for θ is given by $(\hat{\theta} - 2SE, \hat{\theta} + 2SE)$.

$$\ell''(\theta) = \begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta_1^2} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_m} \\ \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ell}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_m} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_m} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_m} & \cdots & \frac{\partial^2 \ell}{\partial \theta_m^2} \end{pmatrix}.$$

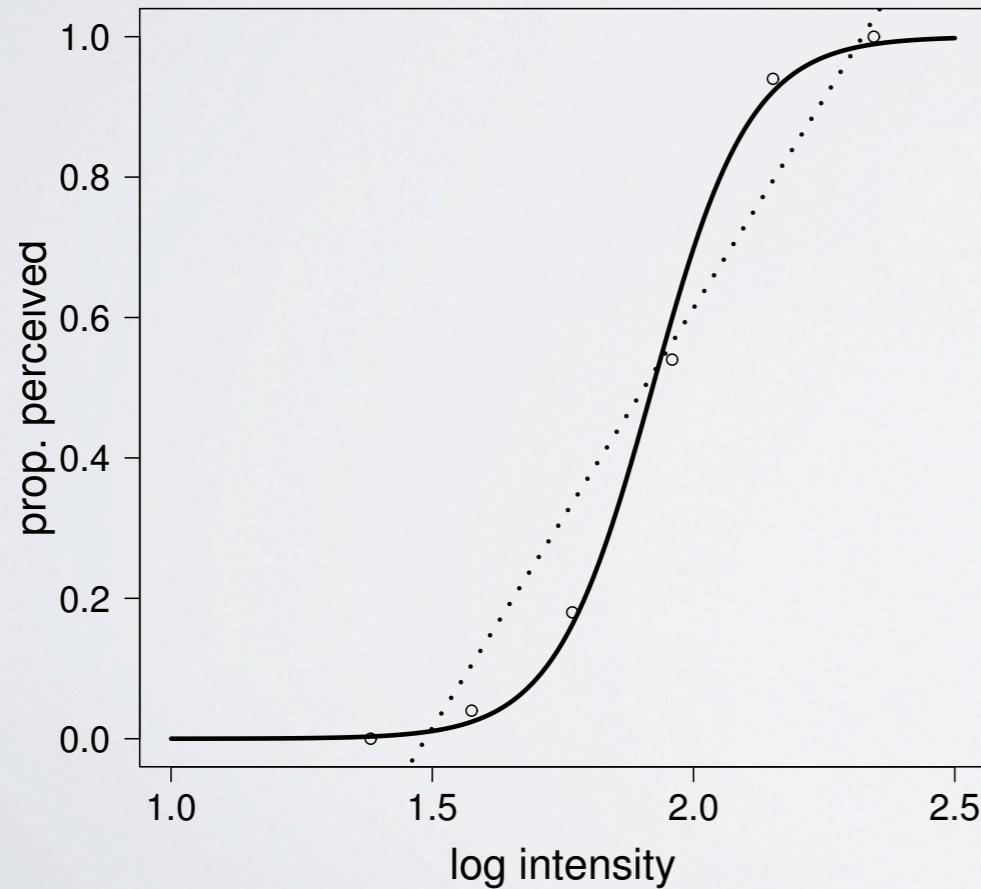
Result For large samples, under certain general conditions, the MLE $\hat{\theta}$ of the m -dimensional parameter θ is distributed approximately as an m -dimensional multivariate normal random vector with variance matrix

$$\hat{\Sigma} = -\ell''(\hat{\theta})^{-1}. \quad (8.23)$$

Example 8.4.1 Human perception of dim light Hecht *et al* (1942, *J. General Physiology*) investigated the threshold for visual perception by exposing human observers to very weak flashes of light in a darkened room. The response variable was an indication of whether or not light was observed (“yes” or “no”), and the explanatory variable was the intensity of the light (in units of average number of light quanta per flash). Several different intensities were used, and for each the experiment was repeated many times. (Similar experiments were performed by Hartline *et al* (1947) on the horseshoe crab *Limulus*, where the response was the proportion of trials on which an isolated optic nerve fiber fired.) The results for one series of trials in one subject are plotted in Figure 8.6.



Example 8.4.1 Human perception of dim light Hecht *et al* (1942, *J. General Physiology*) investigated the threshold for visual perception by exposing human observers to very weak flashes of light in a darkened room. The response variable was an indication of whether or not light was observed (“yes” or “no”), and the explanatory variable was the intensity of the light (in units of average number of light quanta per flash). Several different intensities were used, and for each the experiment was repeated many times. (Similar experiments were performed by Hartline *et al* (1947) on the horseshoe crab *Limulus*, where the response was the proportion of trials on which an isolated optic nerve fiber fired.) The results for one series of trials in one subject are plotted in Figure 8.6.



$$Y_i \sim B(n_i, p_i)$$

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}.$$

In simple cases, ML estimation gives intuitive estimates

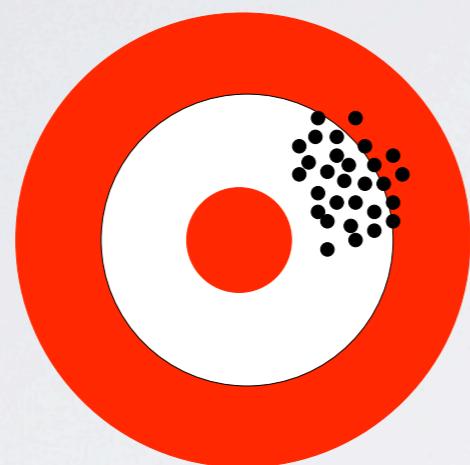
Illustration: Binomial MLE.

$$\hat{\theta} = \frac{x}{n}.$$

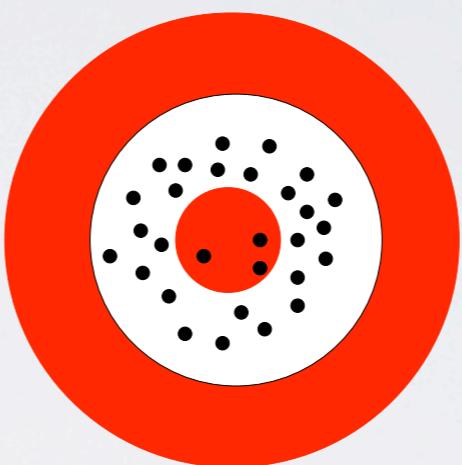
Illustration: Normal MLE. ;

$$; \hat{\theta} = \bar{x}.$$

$$MSE(T) = \text{Bias}(T)^2 + \text{Variance}(T).$$



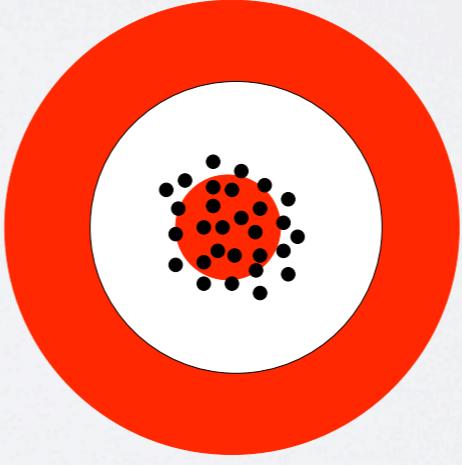
high bias
low variance



low bias
high variance



high bias
high variance



low bias
low variance

8.3.1 In large samples, ML estimation is optimal.

8.3.1 In large samples, ML estimation is optimal.

note: in large samples, estimators are typically
normally distributed

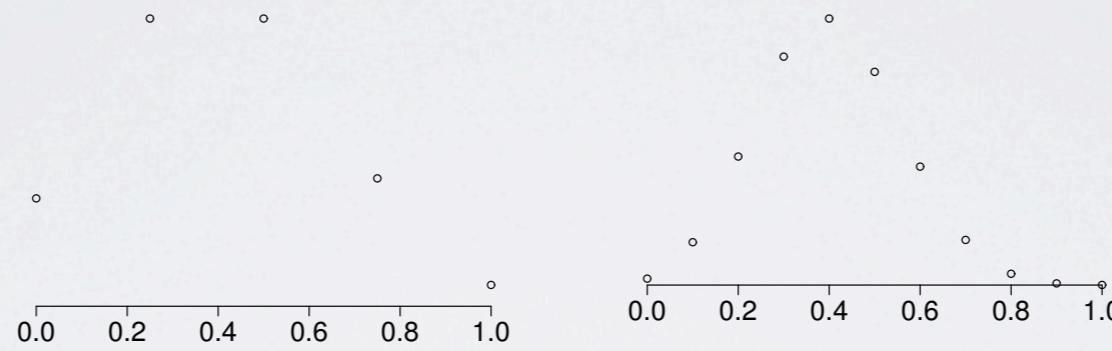
(follows essentially from Central Limit Theorem)

6.1.1 The standard deviation of the sample mean decreases as $1/\sqrt{n}$.

6.2.1 As the sample size n increases, the sample mean converges to the theoretical mean.

6.3.1 For large n , the sample mean is approximately normally distributed.

$n = 4$,



$n = 10, 25$ and 100

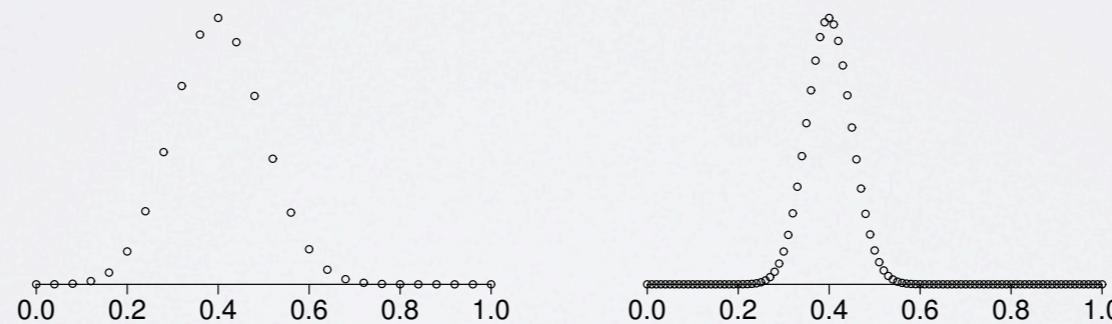


Figure 6.1: The pdf of the binomial mean \bar{X} when $p = .4$ for four different values of n . As n increases the distribution becomes concentrated ($\sigma_{\bar{X}}$ becomes small), with the center of the distribution getting close to $\mu_X = .4$ (the LLN). In addition, the distribution becomes approximately normal (the CLT).

- 8.2.1 In large samples, an estimator should be very likely to be close to its estimand.
- 8.2.2 In large samples, the precision with which a parameter may be estimated is bounded by Fisher information.

$$T \stackrel{\sim}{\sim} N(\theta, \sigma_T^2).$$

Result Roughly speaking, for large samples, under certain general conditions, the precision $1/\sigma_T^2$ of a consistent and asymptotically normal estimator T can not exceed the Fisher information bound $I(\theta)$.

- 8.2.1 In large samples, an estimator should be very likely to be close to its estimand.
- 8.2.2 In large samples, the precision with which a parameter may be estimated is bounded by Fisher information.

$$T \stackrel{\sim}{\sim} N(\theta, \sigma_T^2).$$

Result Roughly speaking, for large samples, under certain general conditions, the precision $1/\sigma_T^2$ of a consistent and asymptotically normal estimator T can not exceed the Fisher information bound $I(\theta)$.

- 8.2.3 Estimators that minimize large-sample variance are called efficient.

8.3.1 In large samples, ML estimation is optimal.

8.3.4 In large samples, ML estimation is approximately Bayesian.

8.3.1 In large samples, ML estimation is optimal.

8.3.4 In large samples, ML estimation is approximately Bayesian.

e.g., binomial case

Let us apply this to the problem of estimating the binomial parameter p . In this section we replace p by θ , so we suppose $X \sim B(n, \theta)$. To apply (7.17) we take $U = \theta$ and $V = X$ to get

$$f_{\theta|x}(\theta|x) = \frac{f_{X|\theta}(x|\theta)f_\theta(\theta)}{\int f_{X|\theta}(x|\theta)f_\theta(\theta)d\theta}.$$

Let us apply this to the problem of estimating the binomial parameter p . In this section we replace p by θ , so we suppose $X \sim B(n, \theta)$. To apply (7.17) we take $U = \theta$ and $V = X$ to get

$$f_{\theta|x}(\theta|x) = \frac{f_{X|\theta}(x|\theta)f_\theta(\theta)}{\int f_{X|\theta}(x|\theta)f_\theta(\theta)d\theta}.$$

In order to do computations we must assign a specific probability distribution as the prior distribution. Assuming we know very little about the value of θ *a priori*, a natural choice is to use the uniform distribution, $\theta \sim U(0, 1)$, i.e., $f_\theta(\theta) = 1$. With this prior pdf we obtain

$$f(\theta|x) = \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x} \cdot 1}{\int \binom{n}{x}\theta^x(1-\theta)^{n-x} \cdot 1 d\theta}$$

which reduces to

$$f(\theta|x) = \frac{\theta^x(1-\theta)^{n-x}}{\int \theta^x(1-\theta)^{n-x} d\theta}. \quad (7.18)$$

This formula is a special case of a *Beta* distribution introduced briefly in Chapter 5: in general, the $Beta(\alpha, \beta)$ density is

$$f(w) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} w^{\alpha-1} (1-w)^{\beta-1}. \quad (7.19)$$

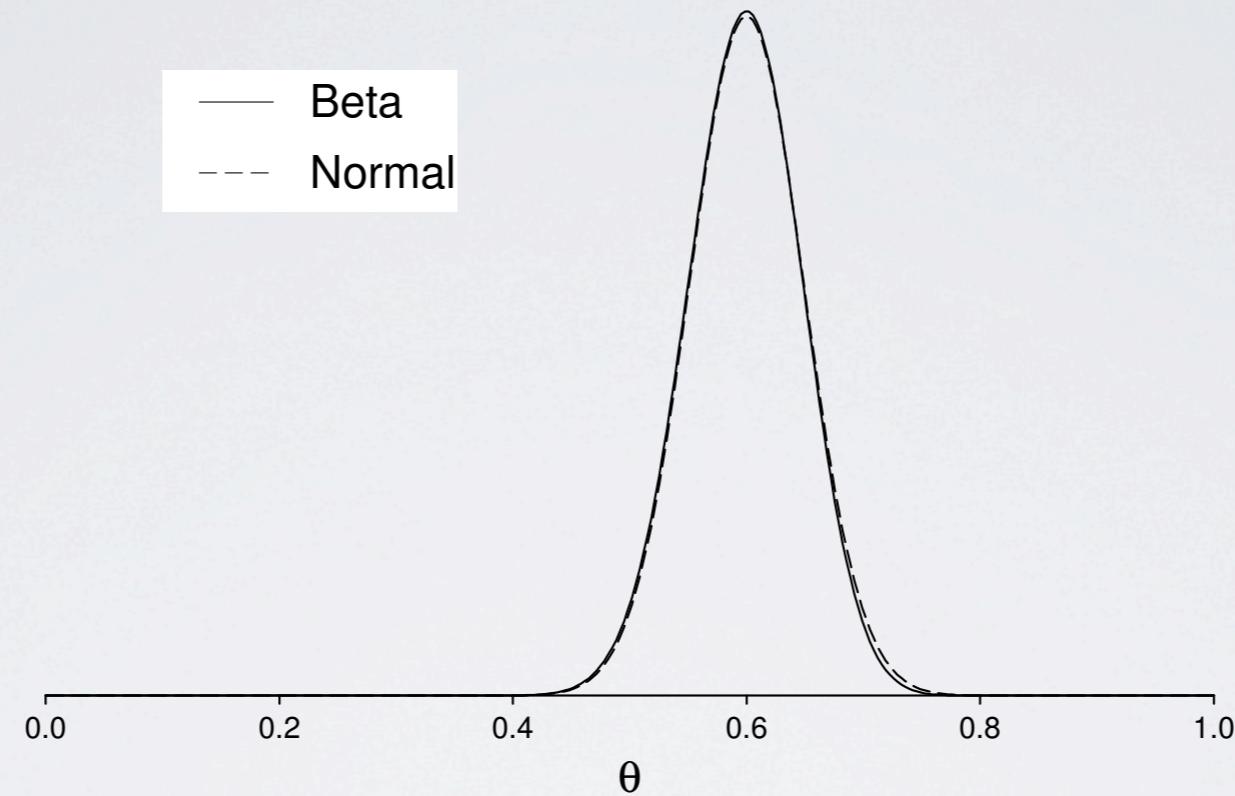


Figure 8.5: Normal approximation $N(.6, (.049)^2)$ to beta posterior $Beta(61, 41)$.

Recap of big idea:

Once we have a statistical model that depends on some parameter vector, we can write down a likelihood function and then find the MLE to estimate the parameter vector and get uncertainty.

OUTLINE OF LECTURE 2

- 2a. Quick review of common probability distributions
- 2b. Maximum Likelihood (ML) estimation
- 2c. Quick review of LLN and CLT
- 2d. Properties of ML estimators
- 2e. Point process formalization 
- 2g. Shadlen and Newsome (1998), section 1

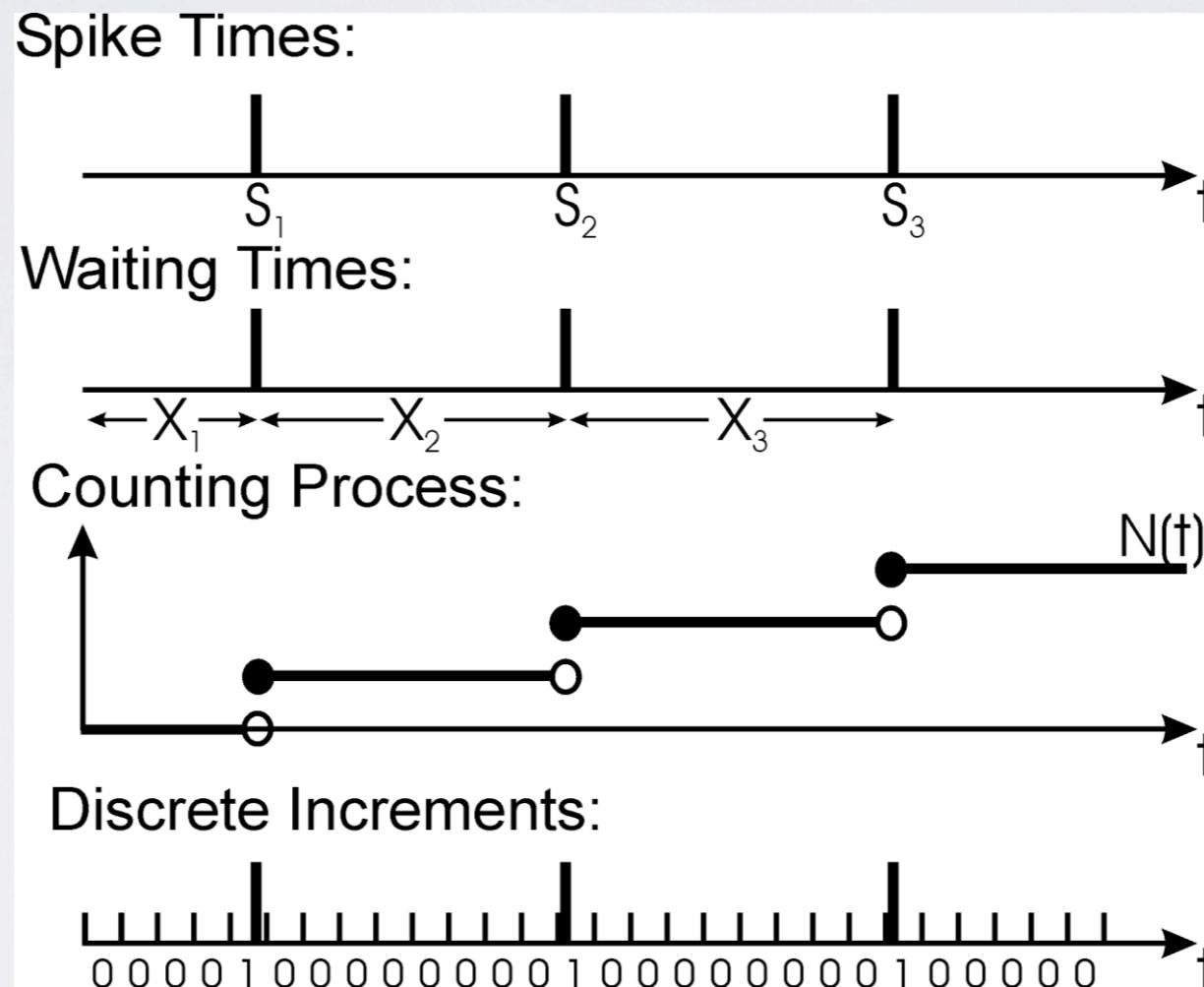
Spike Trains as Point Processes

Main ideas:

1. spike time variability similar to point process;
2. statistical model of instantaneous firing rate provides (a) description and (b) scientific inference via statistical inference.
3. *The ability to go back and forth between continuous time, where neuroscience and statistical theory reside, and discrete time, where measurements are made and data are analyzed, is central to statistical modeling of spike trains.*

19.1.1 A point process may be specified in terms of event times, inter-event intervals, or event counts.

19.1.2 A point process may be considered, approximately, to be a binary time series.



19.2.1 Poisson processes are point processes for which event probabilities do not depend on occurrence or timing of past events.

Definition: A homogeneous Poisson process with intensity λ is a point process satisfying the following conditions:

1. For any interval, $(t, t + \Delta t]$, $\Delta N_{(t,t+\Delta t]} \sim P(\mu)$ with $\mu = \lambda \Delta t$.
2. For any non-overlapping intervals, $(t_1, t_2]$ and $(t_3, t_4]$, $\Delta N_{(t_1,t_2]}$ and $\Delta N_{(t_3,t_4]}$ are independent.

19.2.1 Poisson processes are point processes for which event probabilities do not depend on occurrence or timing of past events.

Definition: A homogeneous Poisson process with intensity λ is a point process satisfying the following conditions:

1. For any interval, $(t, t + \Delta t]$, $\Delta N_{(t,t+\Delta t]} \sim P(\mu)$ with $\mu = \lambda \Delta t$.
2. For any non-overlapping intervals, $(t_1, t_2]$ and $(t_3, t_4]$, $\Delta N_{(t_1,t_2]}$ and $\Delta N_{(t_3,t_4]}$ are independent.

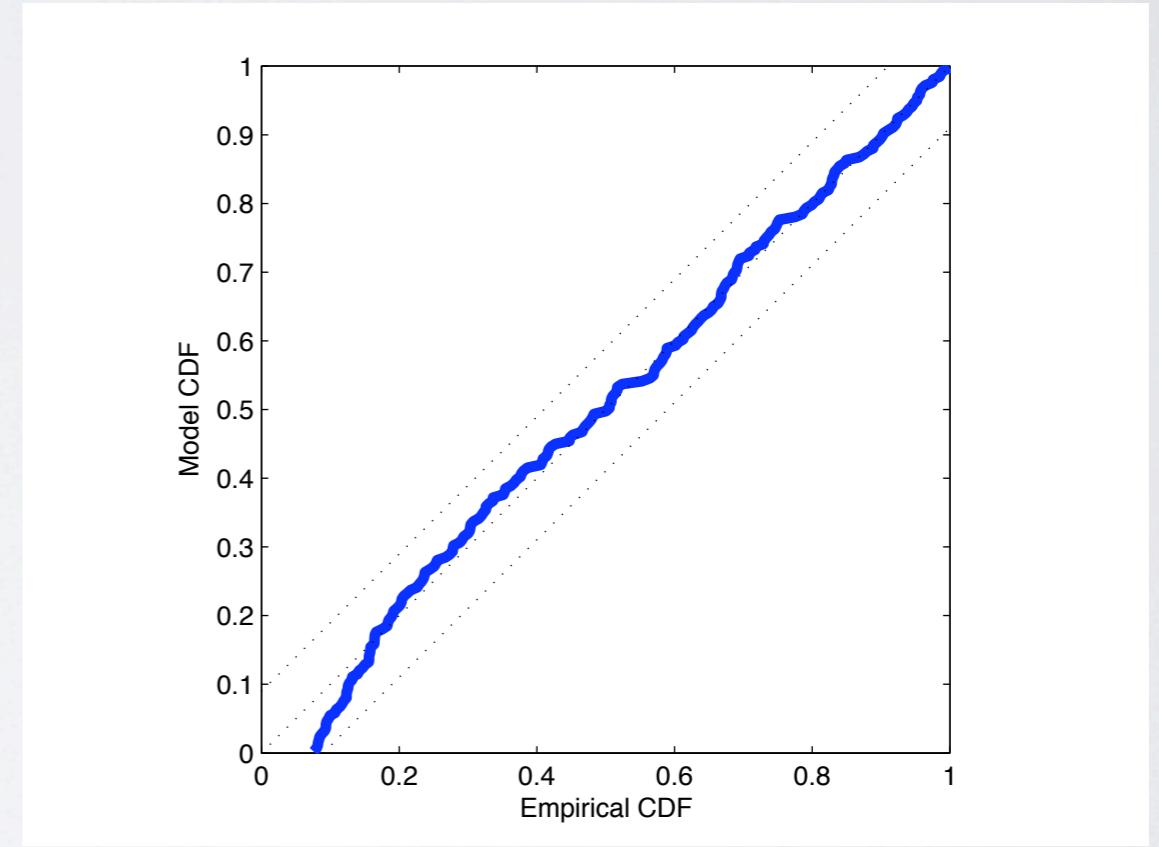
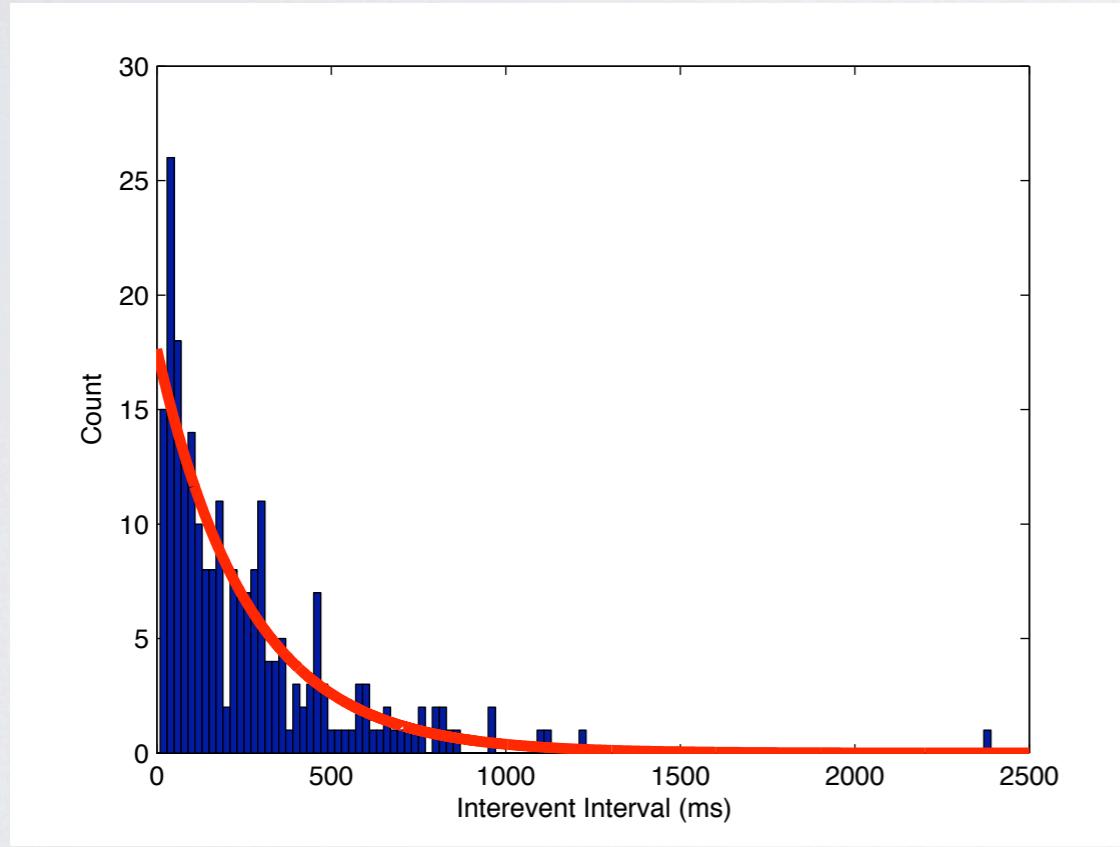
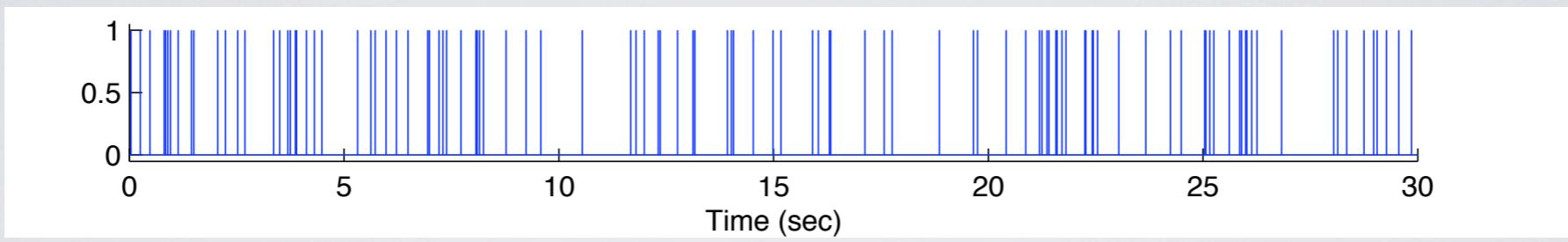
discrete time approximation:

1. For the i th time bin $(i\Delta t, (i + 1)\Delta t]$, $\Delta N_{(i\Delta t,(i+1)\Delta t)} \sim Bernoulli(p)$.
2. For any two distinct time bins, $(i\Delta t, (i + 1)\Delta t]$ and $(j\Delta t, (j + 1)\Delta t]$, $\Delta N_{(i\Delta t,(i+1)\Delta t)}$ and $\Delta N_{(j\Delta t,(j+1)\Delta t)}$ are independent.

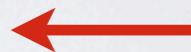
Theorem: A point process is a homogeneous Poisson process with intensity λ if and only if its inter-event waiting times are i.i.d. $Exp(\lambda)$.

exponential random variables are memoryless

Result: Suppose we observe $N(T) = n$ events from a homogeneous Poisson process on an interval $(0, T]$. Then the distribution of the event times is the same as that of a sample of size n from a uniform distribution on $(0, T]$.

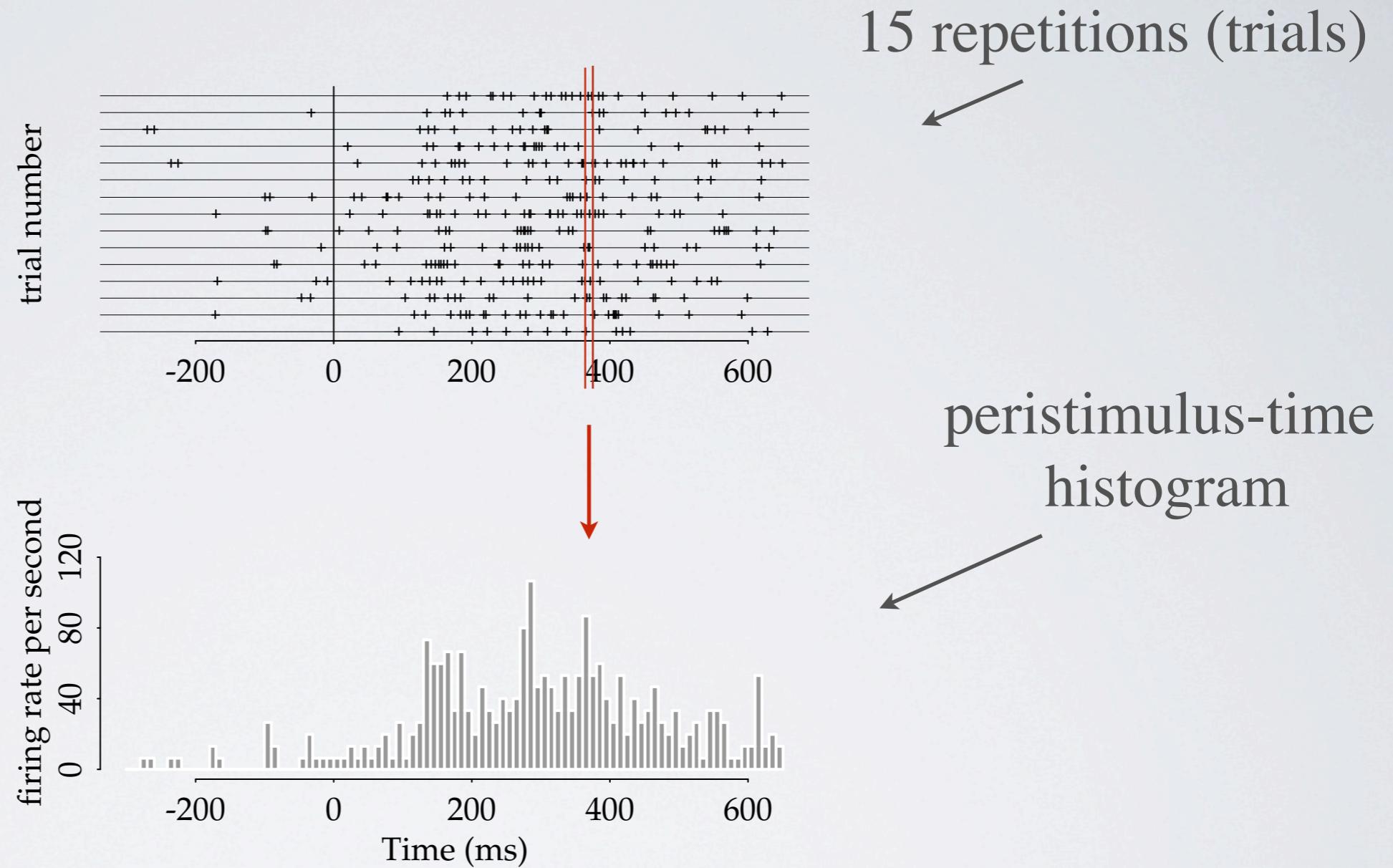


OUTLINE OF LECTURE 2

- 2a. Quick review of common probability distributions
- 2b. Maximum Likelihood (ML) estimation
- 2c. Quick review of LLN and CLT
- 2d. Properties of ML estimators
- 2e. Point process formalization
- 2g. Shadlen and Newsome (1998), section 1 

Shadlen and Newsome

Why do neurons fire irregularly?



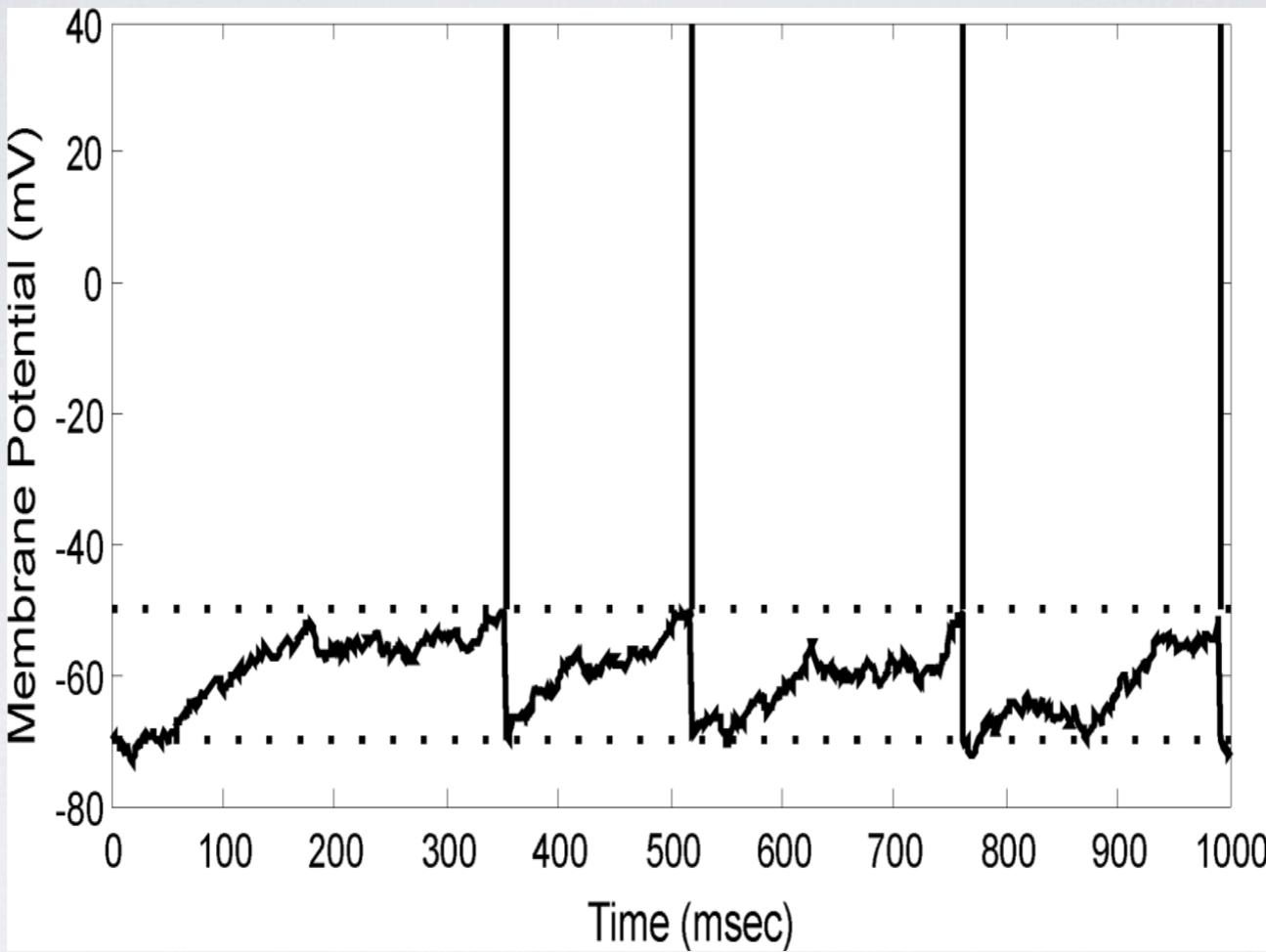


Figure 5.5: Example of an integrate-and-fire neuron. At each time step there is either an EPSP or an IPSP, with probabilities p and $1 - p$. For $p > 1 - p$ this creates a stochastic upward “drift” of the voltage (as the inputs are summed or “integrated”) until it crosses the threshold and the neuron fires. The neuron then resets to its baseline voltage. The resulting ISI distribution is approximately inverse Gaussian.

Shadlen and Newsome

cortical neurons

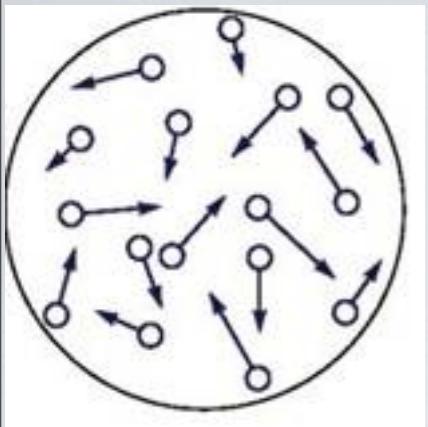
rely on large numbers of inputs

respond with large range of firing rates

Table 1. Properties of statistical homogeneity for cortical neurons

Response property	Approximate value
Dynamic range of response	0–200 spikes/sec
Distribution of interspike intervals	Approximately exponential
Spike count variance	Variance \sim 1–1.5 times the mean count
Spike rate modulation	Expected rate can vary in \sim 1 ISI, 5–10 msec

moving dots

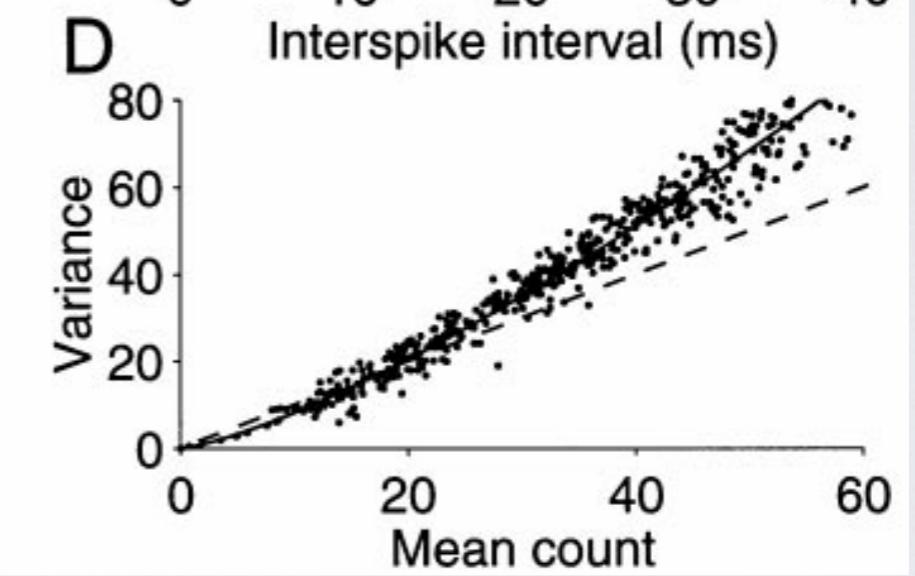
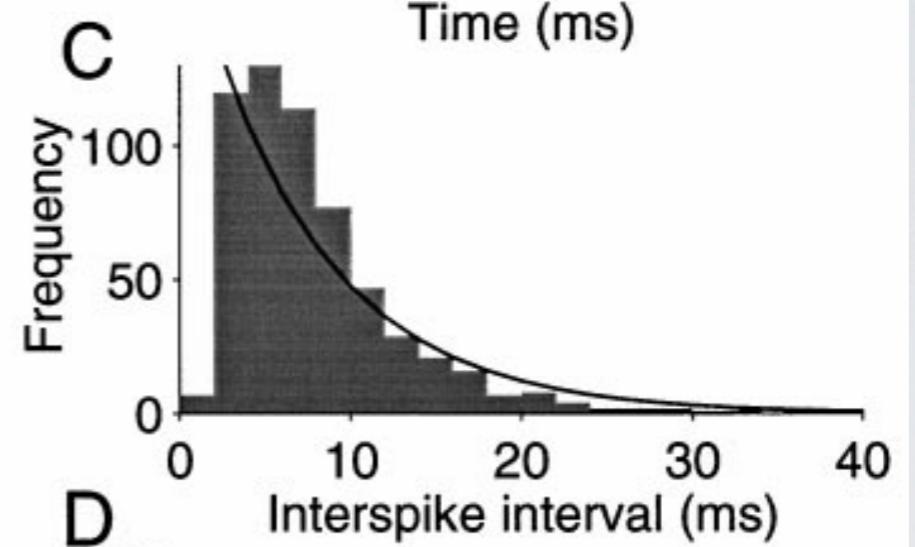
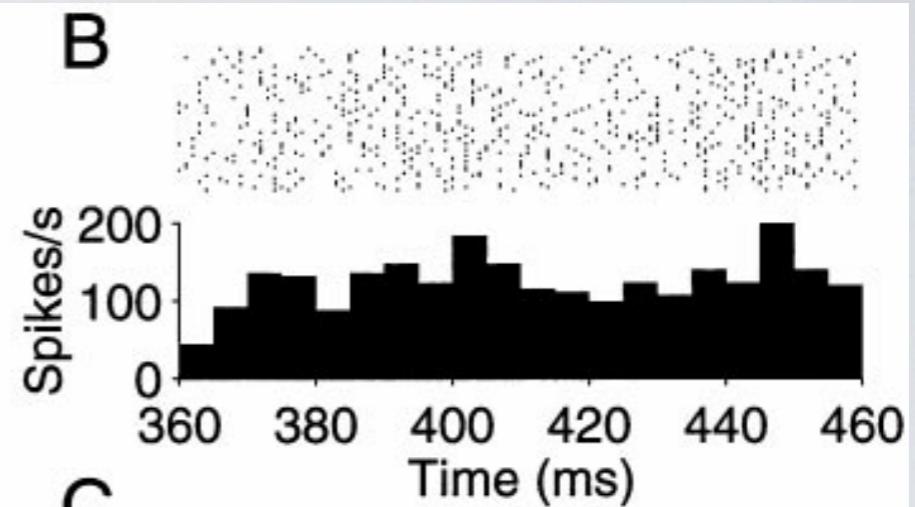
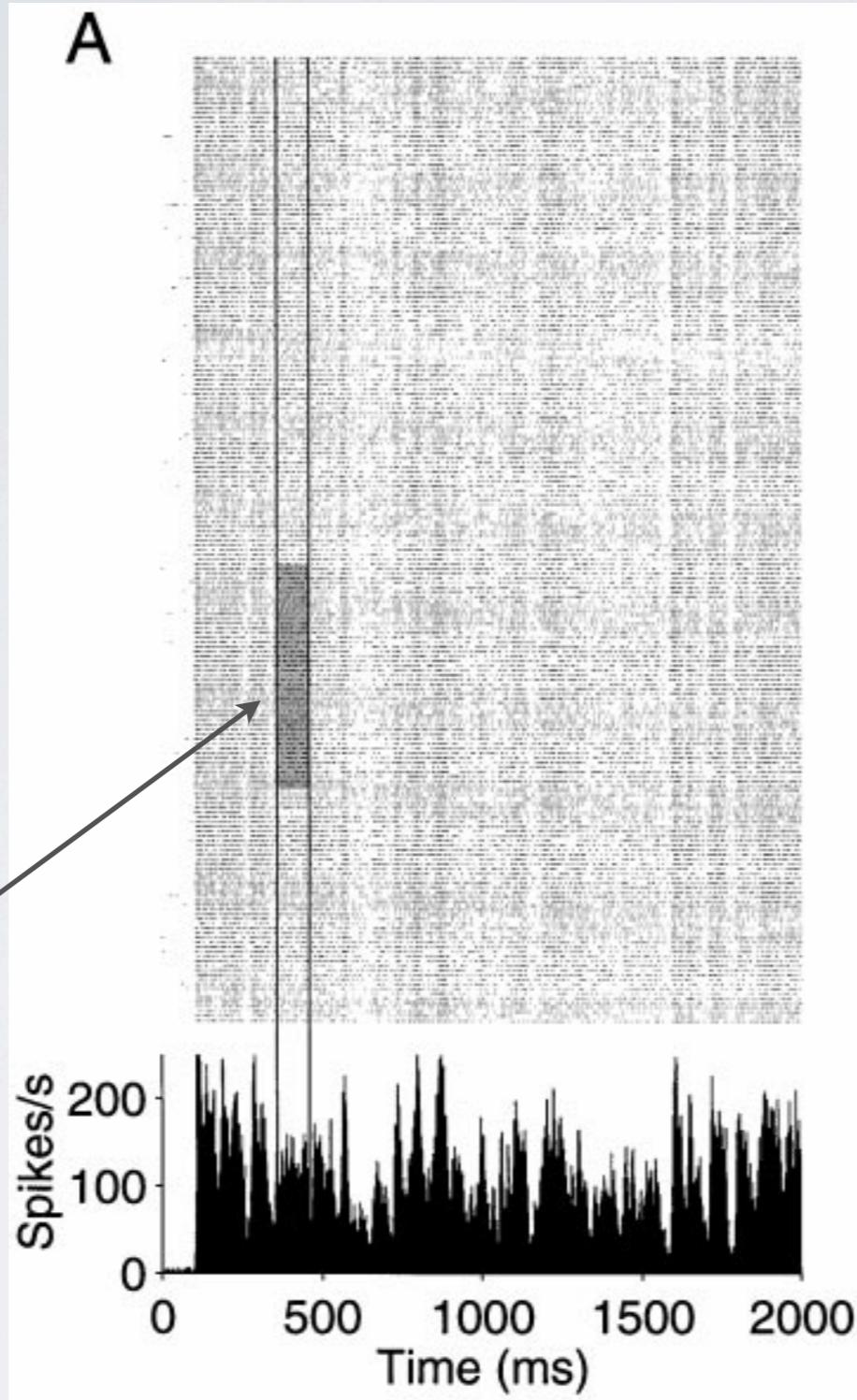


MT neuron

210 trials

15 - 220 Hz

50 trials



How can cortical high-input regime
produce a varying firing rate?

solution: balanced excitation and inhibition

How can cortical high-input regime
produce a varying firing rate?

solution: balanced excitation and inhibition

“the price of a reasonable dynamic range is noise”

Shadlen and Newsome

cortical neurons

rely on large numbers of inputs

EPSP 3-10% of $V_{thresh} - V_{rest}$

\Rightarrow AP when 10-40 spikes within 10-20 msec

Shadlen and Newsome

cortical neurons

rely on large numbers of inputs

EPSP 3-10% of $V_{\text{thresh}} - V_{\text{rest}}$

\Rightarrow AP when 10-40 spikes within 10-20 msec

with 100 inputs at 100 Hz, would expect
10 spikes per ms, 40 spikes per 4 ms,
i.e., at least 250 Hz response

model:

inputs have exponential ISIs

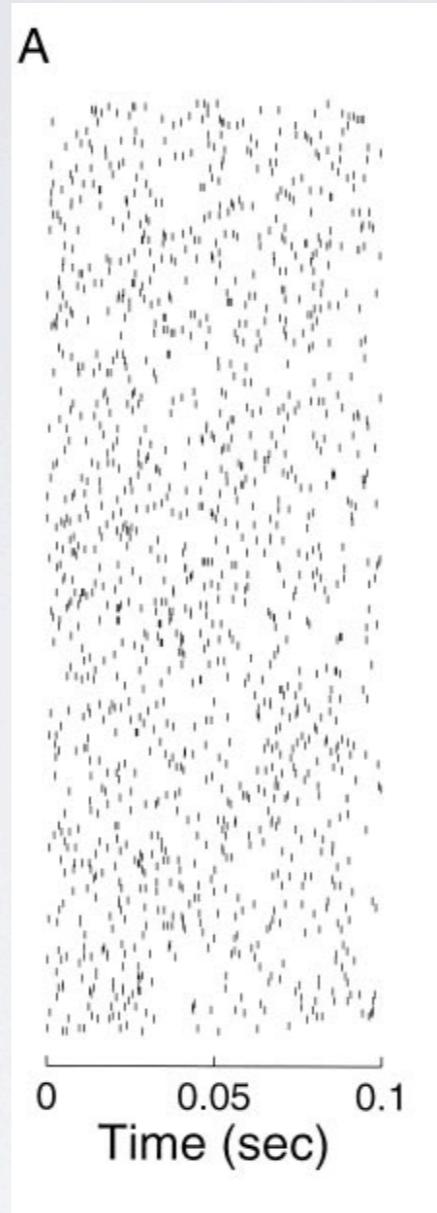
each EPSP +1 step, IPSP -1 step

exponential decay; time constant tau

lower barrier Vrest -1

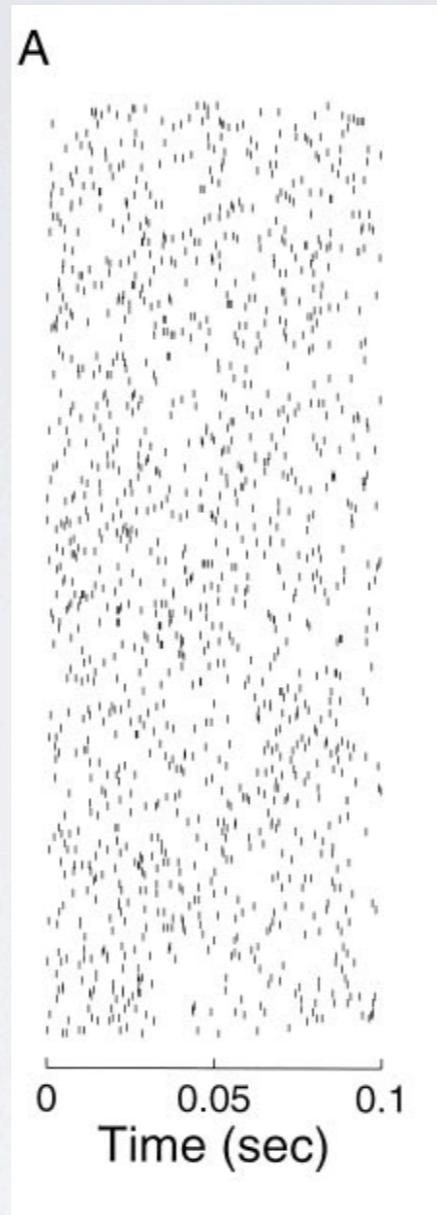
Vthresh = # steps

300 excitatory inputs
each at 50 Hz
exponential ISIs

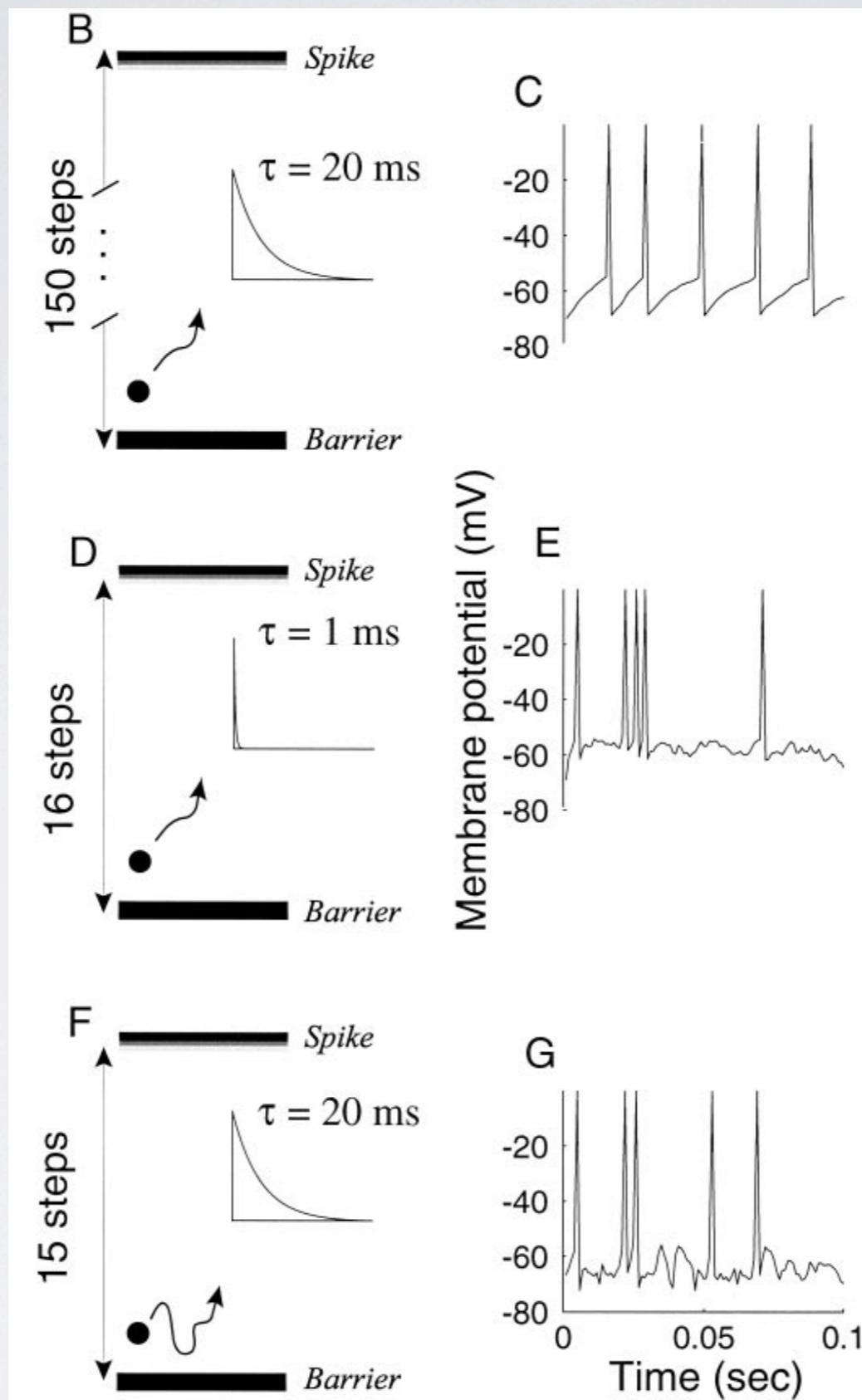


300 excitatory inputs
each at 50 Hz
exponential ISIs

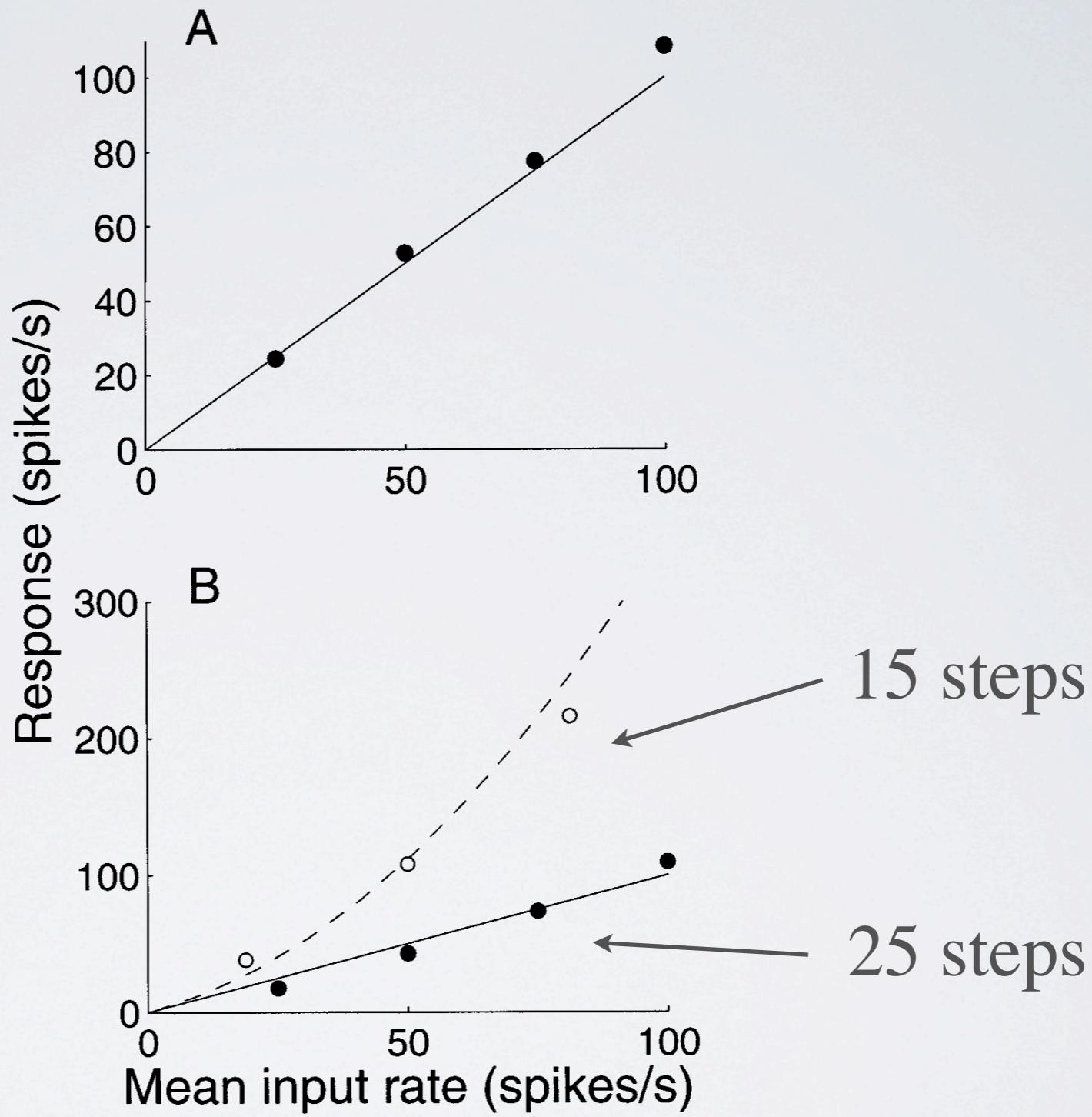
Want: output 50 Hz



B: high threshold
D: small tau
F: balanced input



A: 300 inputs
B: 600 inputs



Conclusions: balanced excitation and inhibition

- (1) allow reproduction of firing rate,
with a few hundred inputs;
- (2) produce irregular ISIs and, when varying firing
rate is considered, these are realistic.

