Neuroinformatics Lecture (L8)

Prof. Dr. Gordon Pipa

Institute of Cognitive Science University of Osnabrück

Summary: **Bernoulli Distribution**

Distribution: $p(x|\mu) = \mu^x(1-\mu)^{1-x}$
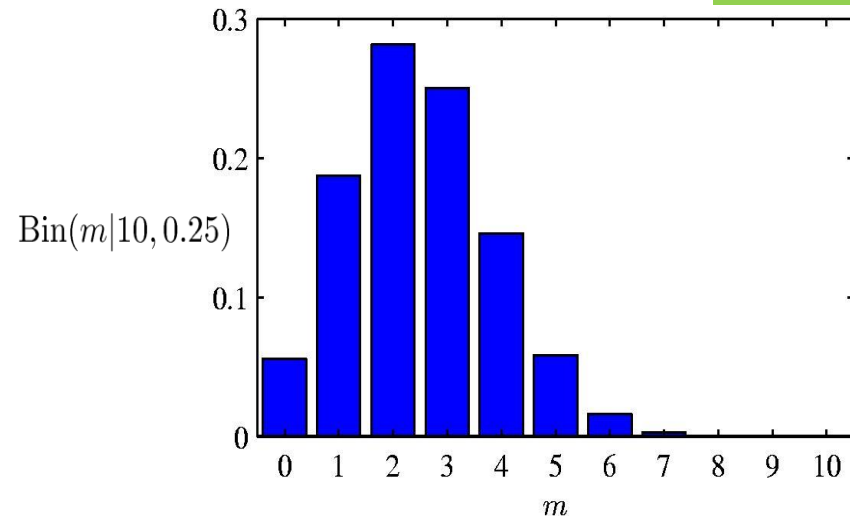
Expected value: $E[x] = \mu$

Variance: $var[x] = \mu(1-\mu)$

Summary: **Binomial Distribution**
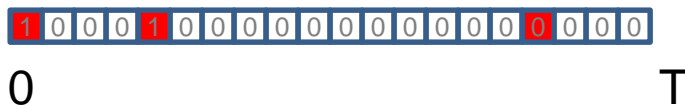
$$\text{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$$

$\text{Bin}(m|10, 0.25)$

$$\mathbb{E}[m] \equiv \sum_{m=0}^{N} m\,\text{Bin}(m|N,\mu) \quad = \quad N\mu$$

$$\text{var}[m] \equiv \sum_{m=0}^{N} (m - \mathbb{E}[m])^2 \,\text{Bin}(m|N,\mu) \quad = \quad N\mu(1-\mu)$$

So far:
$$p(k|\mu, N) = \frac{(N - k + 1) \cdot \ldots \cdot (N - 1) \cdot N}{k!} \cdot \frac{\lambda^k}{N^k} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k}$$

Now we assume that we use this model to describe a spike train using very small temporal bins to describe the spike train via a binary process.

0   T

We now reduce the bin size to a arbitrary small value which is reducing the probability per bin (limes of p going to zero).

At the same time the number of bin must grow if we want to cover the same temporal range between 0 and T (limes of N going to infinity).

Moreover, we know that, irrespective of the bin size and the number of bins, the expected value of events in the time interval from 0 to T stays the same.

We achieve this by making N large, μ small and keeping the product of μ and N constant at the same time.

So far: $p(k|\mu, N) = \dfrac{(N - k + 1) \cdot ... \cdot (N - 1) \cdot N}{k!} \cdot \dfrac{\lambda^k}{N^k} \left(1 - \dfrac{\lambda}{N}\right)^N \left(1 - \dfrac{\lambda}{N}\right)^{-k}$

$$\lim_{\substack{N \to \infty \\ p \to 0 \\ np=const=\lambda}} \frac{\lambda^k}{k!} = \frac{\lambda^k}{k!}$$

Independent of N and p, given λ is constant

$$\lim_{N \to \infty} \frac{N(N-1)\cdot ... \cdot (N-k+1)}{N^k} = 1$$

If N grows, it becomes ~ N to the power of k

$$\lim_{\substack{N \to \infty \\ np=const=\lambda}} \left(1 - \frac{\lambda}{N}\right)^N = e^{-\lambda}$$

known limit (see textbook)

$$\lim_{\substack{N \to \infty \\ np=const=\lambda}} \left(1 - \frac{\lambda}{N}\right)^{-k} = 1$$

λ over N becomes zero for constant k

R

Poisson Distribution: probability of k events given that λ events are expected

$$p(k|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

With k=0,1,2, …. (discrete)   and λ >0 (parameter is continuous)

E(k)   = λ
var(k)   = λ

Example: p(k) spikes in an interval between 0 and T, if λ spikes are expected

Y = poisspdf(X,lambda)

computes the Poisson pdf at each of the values in X using the corresponding parameters in lambda. The parameters in lambda must all be positive.

Discrete Distributions

- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution

Continuous Distributions

- Normal (Gauss)
- Gamma Distribution
- chi^2 Distribution

When he was eight, he figured out how to add up all the numbers from 1 to 100.

$X = 1 + 2 + ... + 100$

$= (1+100) + (2+99) + ... + (50+51)$

$= 50 * 101$

$= 5050$

1777-1855

He completed *Disquisitiones Arithmeticae*, his magnum opus, in 1798 at the age of 21. This work was fundamental in consolidating number theory.

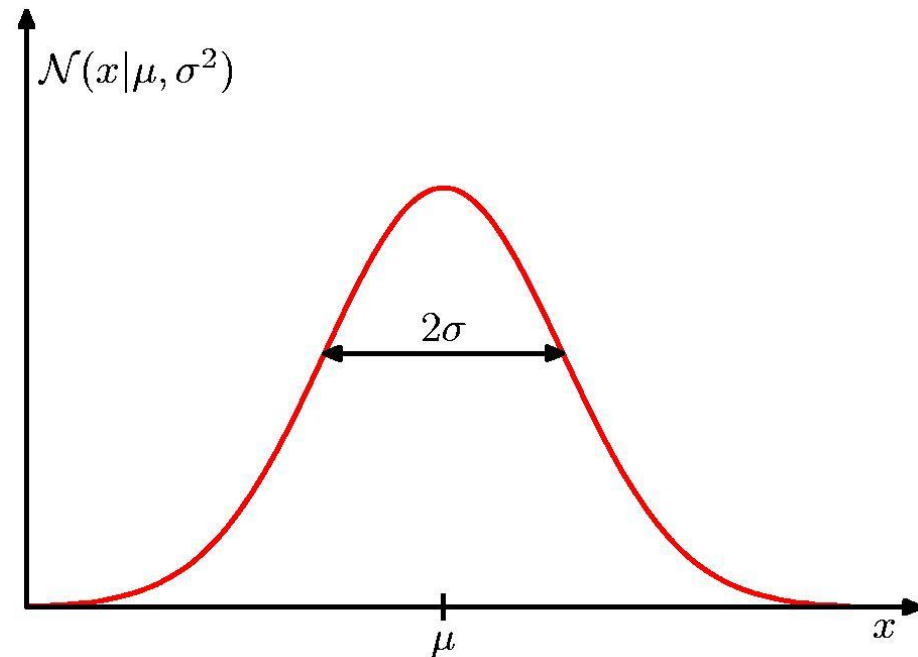See https://de.wikipedia.org/wiki/Datei:10_DM_Serie4_Rueckseite.jpg

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

With x continuous; also parameters μ, σ are continuous

With

$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

Y = normpdf(X,mu,sigma)

computes the pdf at each of the values in X using the normal distribution with mean mu and standard deviation sigma. The parameters in sigma must be positive.

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \, dx = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \, dx = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2 \qquad\qquad \mathrm{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right]$$

$$if \ z = \frac{x-\mu}{\sigma} \ z \ is \ standard \ normal \ distributed \ \ z \sim N(0,1)$$

Let $\{X_1, X_2, ..., X_n\}$ be a random sample of size n with expected values μ and variances σ^2 (independent and identically distributed random variables)

Suppose we are interested in the sample average of these random variables:

$$S_N = \frac{1}{N} \sum_{n=1}^{N} X_n$$

Then the central limit theorem asserts that for large N's, the distribution of S_n is approximately normal with

distribution of S_n ~ Normal with

mean: μ

variance: $\dfrac{1}{N}\sigma^2$

The true strength of the theorem is that S_n approaches normality regardless of the shapes of the distributions of individual X_i's.

Suppose we have two independent random variables X and Y.

To get the the probaility distribtion of p(Z=X+Y), we use a convolution. That is the integral of the joint probability of p(x,y), given that x+y=z (line integral)

$$p(z) = \int_{-\infty}^{\infty} p(x, y \mid x + y = z) dx dy$$

$$= \int_{-\infty}^{\infty} p(x) p(z - x) dx$$

For a discrete distribution, we get:

$$p(z) = \sum p(x) p(z - x)$$

$$p(z) = \sum p(x)p(z-x)$$ With x+y=z   and x,y natural positive numbers

Example:

z=0:   since z=0 → *x=0 & y=0:*

p(z=0)=p(0-0)p(0)

          x=0 & y=0

z=1:   since z=1 x can be 0 or 1

p(z=1)=p(1-0)p(0) + p(1-1)p(1)

z=2:   since z=2 x can be 0 or 1 or 2

p(z=2)=p(2-0)p(0) + p(2-1)p(1) + p(2-2)p(2)

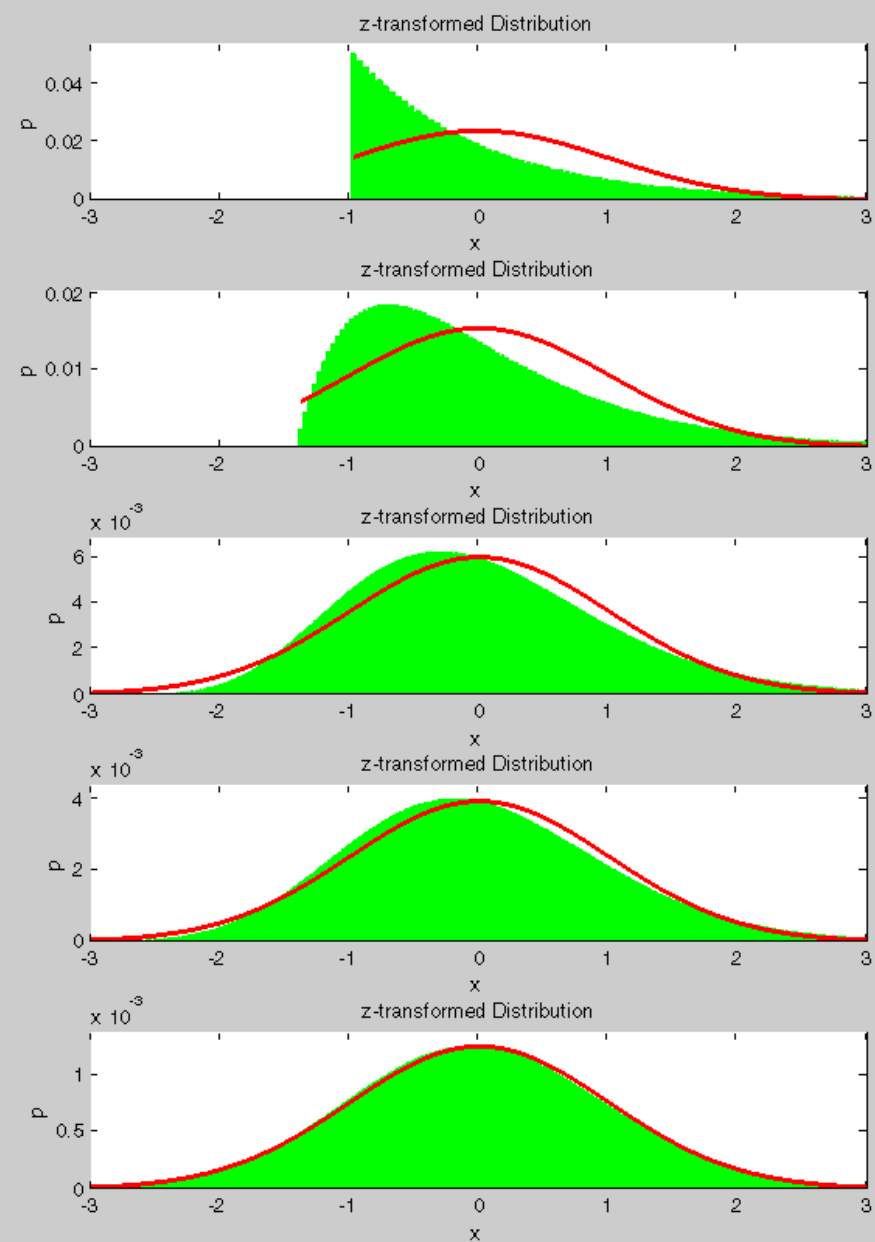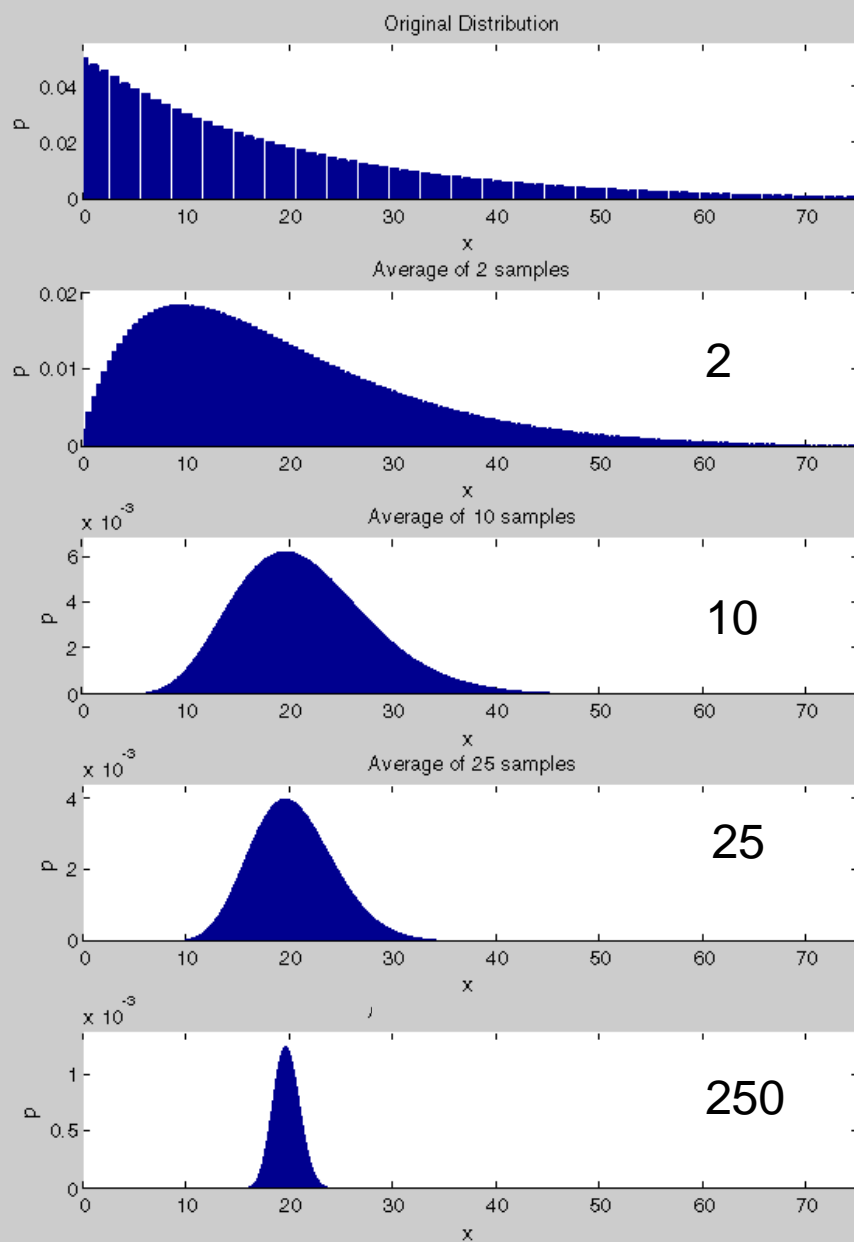Suppose we we apply this itteratively for N random variables X_n.

$Z_2 = X_1 + X_2$

$$p(Z_2) = \int_{-\infty}^{\infty} p^{x_1}(X_1) p^{x_2}(Z_2 - X_1) dX_1$$

$Z_3 = X_1 + X_2 + X_3 = Z_2 + X_3$

$$p(Z_3) = \int_{-\infty}^{\infty} p^{x_3}(X_3) p^{z_2}(Z_3 - X_3) dX_3$$

$Z_4 = X_1 + X_2 + X_3 + X_4 = Z_3 + X_4$

$$p(Z_4) = \int_{-\infty}^{\infty} p^{x_4}(X_4) p^{z_3}(Z_4 - X_4) dX_4$$

```
x= 0:500;
y = gampdf(x,1,20);
```

Gives us a Gamma Distribution for the values at x with gampdf(x,α,β) . Here this is an exponential.

$$p(x) = c \cdot x^{\alpha-1} \, e^{-\frac{x}{\beta}} = c \cdot x^0 \, e^{-\frac{x}{\beta}}$$

Plots the distribution in an grid of 5x2 plots.

```
subplot(5,2,1)
bar(x,y)
set(gca,'Xlim',[0 75])
set(gca,'Ylim',[0 max(y)*1.1])
xlabel('x')
ylabel('p')
title('Original Distribution')
```

```matlab
subplot(5,2,2)
p_vec           = y./sum(y)
E_x             = sum(x.*p_vec)
E_x2            = sum((x.^2).*p_vec)
variance        = E_x2-E_x.^2;
standard_dev    = sqrt(variance);
z_transformed   = (x-E_x)./standard_dev;
bar(z_transformed,y,'green')
set(gca,'Xlim',[-3 3])
set(gca,'Ylim',[0 max(p_vec)*1.1])
z_mormal        = normpdf(z_transformed,0,1);
z_mormal        = z_mormal./sum(z_mormal);
hold on
plot(z_transformed,z_mormal,'r','linewidth',2);
xlabel('x')
ylabel('p')
title('z-transformed Distribution')
```

z-tranforming here

$$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mu)^2\right]$$
$$= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$$

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + (\mathrm{E}[X])^2\right]$$
$$= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + (\mathrm{E}[X])^2$$
$$= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2$$
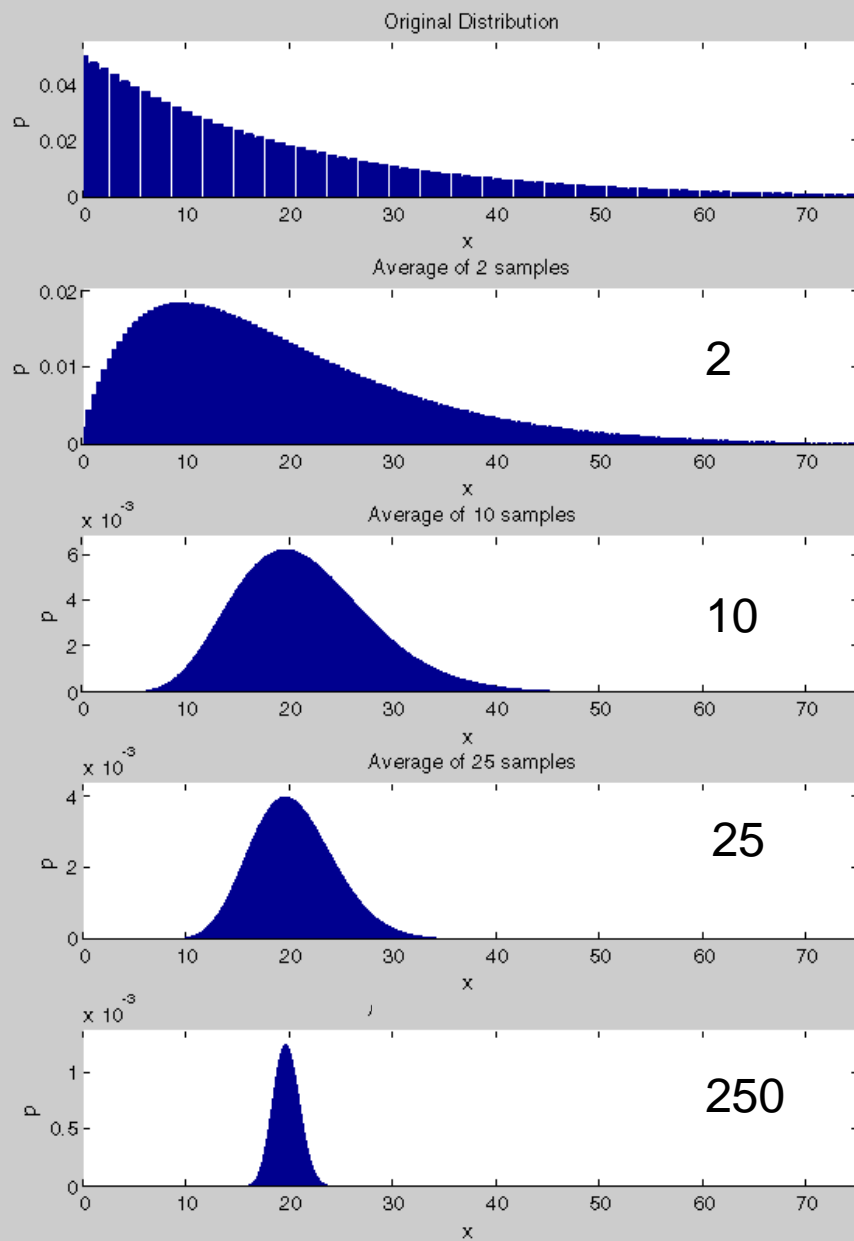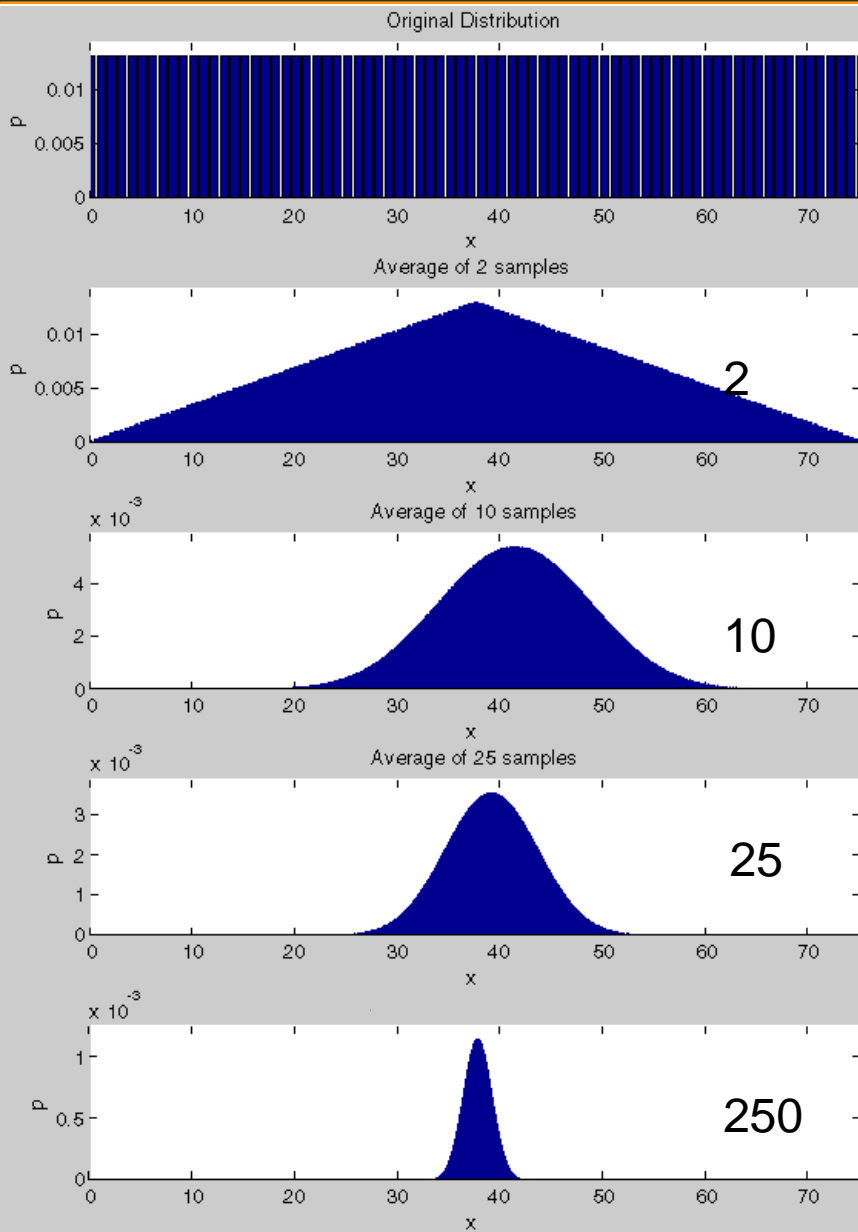
Discrete Distributions

- Bernoulli Distribution

- Binomial Distribution

- Poisson Distribution

Continuous Distributions

- Normal (Gauss)

- Gamma Distribution

- chi^2 Distribution

Gamma Distribution

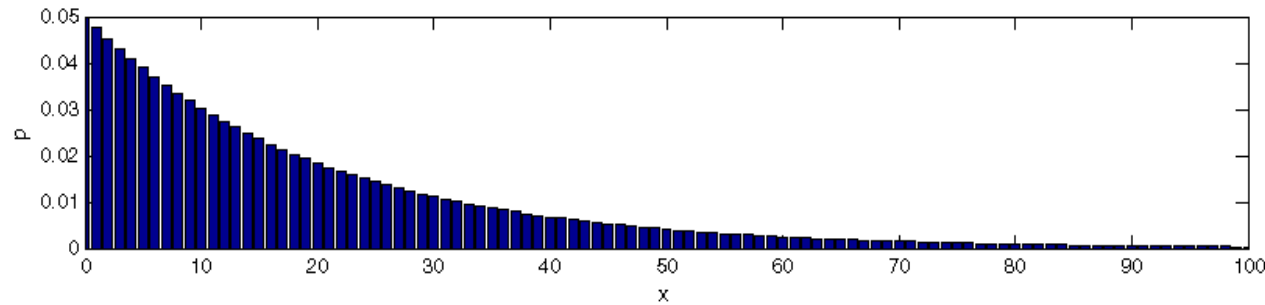$$p(x) = c \cdot x^{\alpha-1}\, e^{-\frac{x}{\beta}}$$

with x,α,β > 0

c is used to normalize such that

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

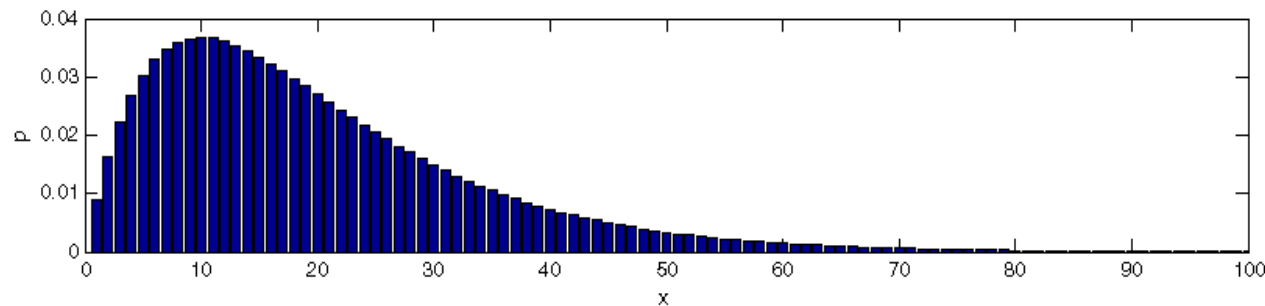Y = gampdf(X,α,β)

computes the gamma pdf at each of the values in X using the corresponding parameters in α and β. The parameters in α and β must all be positive, and the values in X must lie on the interval [0 ∞

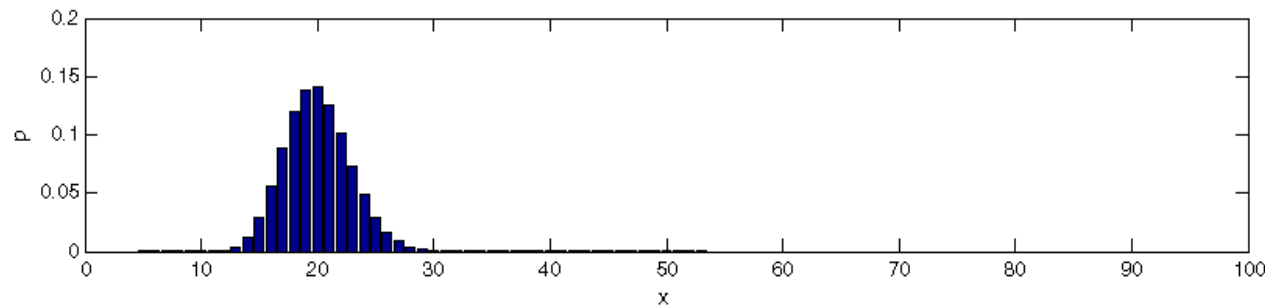# All have the same expected value


```
subplot(3,1,1)
y = gampdf(x,1,20);
bar(x,y)
set(gca,'Xlim',[0 100])
xlabel('x')
ylabel('p')
```

```
subplot(3,1,2)
y = gampdf(x,2,20/2);
bar(x,y)
set(gca,'Xlim',[0 100])
xlabel('x')
ylabel('p')
```

```
subplot(3,1,3)
y = gampdf(x,50,20/50);
bar(x,y)
set(gca,'Xlim',[0 100])
xlabel('x')
ylabel('p')
```

Gamma Distribution

$$p(x) = c \cdot x^{\alpha-1} \, e^{-\frac{x}{\beta}}$$

with x,α,β > 0,   c is chosen such

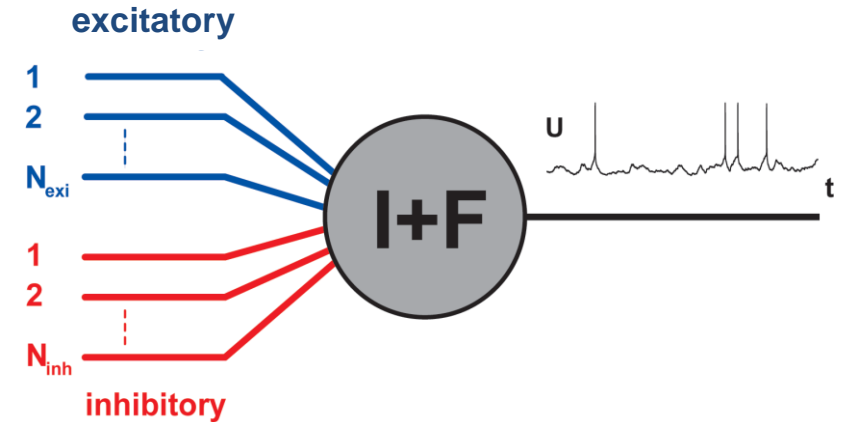$$\int_{-\infty}^{\infty} p(x)dx = 1$$

$$E(x) = \alpha \cdot \beta$$
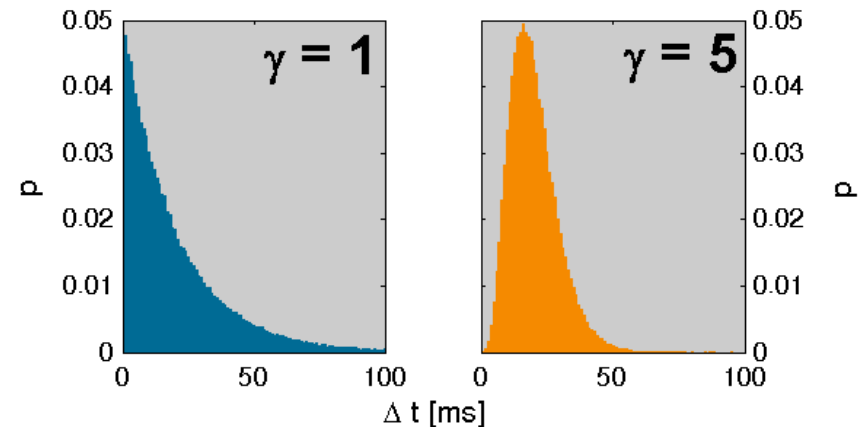
$$var(x) = \alpha \cdot \beta^2$$

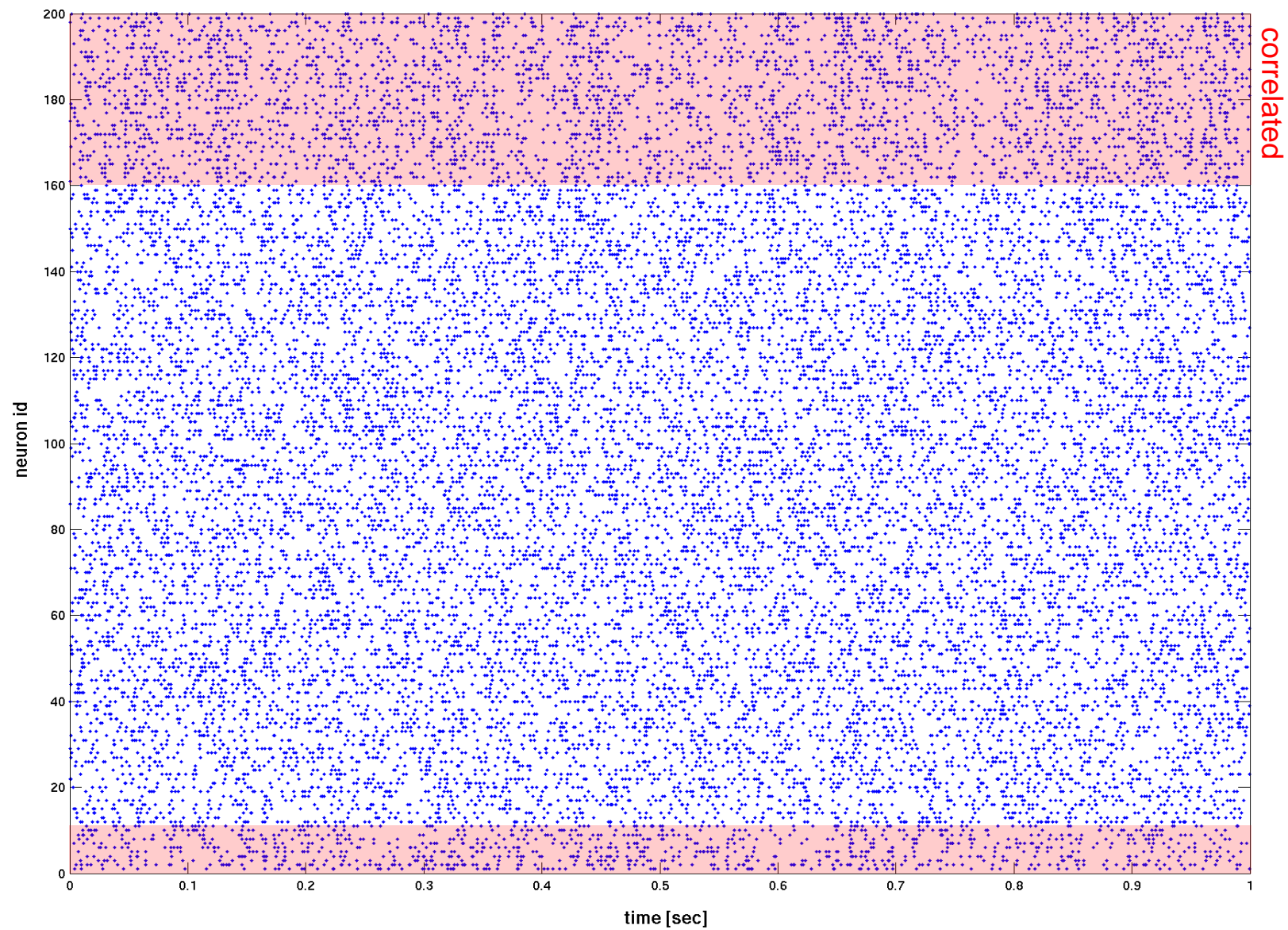**γ-spike trains and a leaky I+F neuron**

- Presynaptic spike trains :

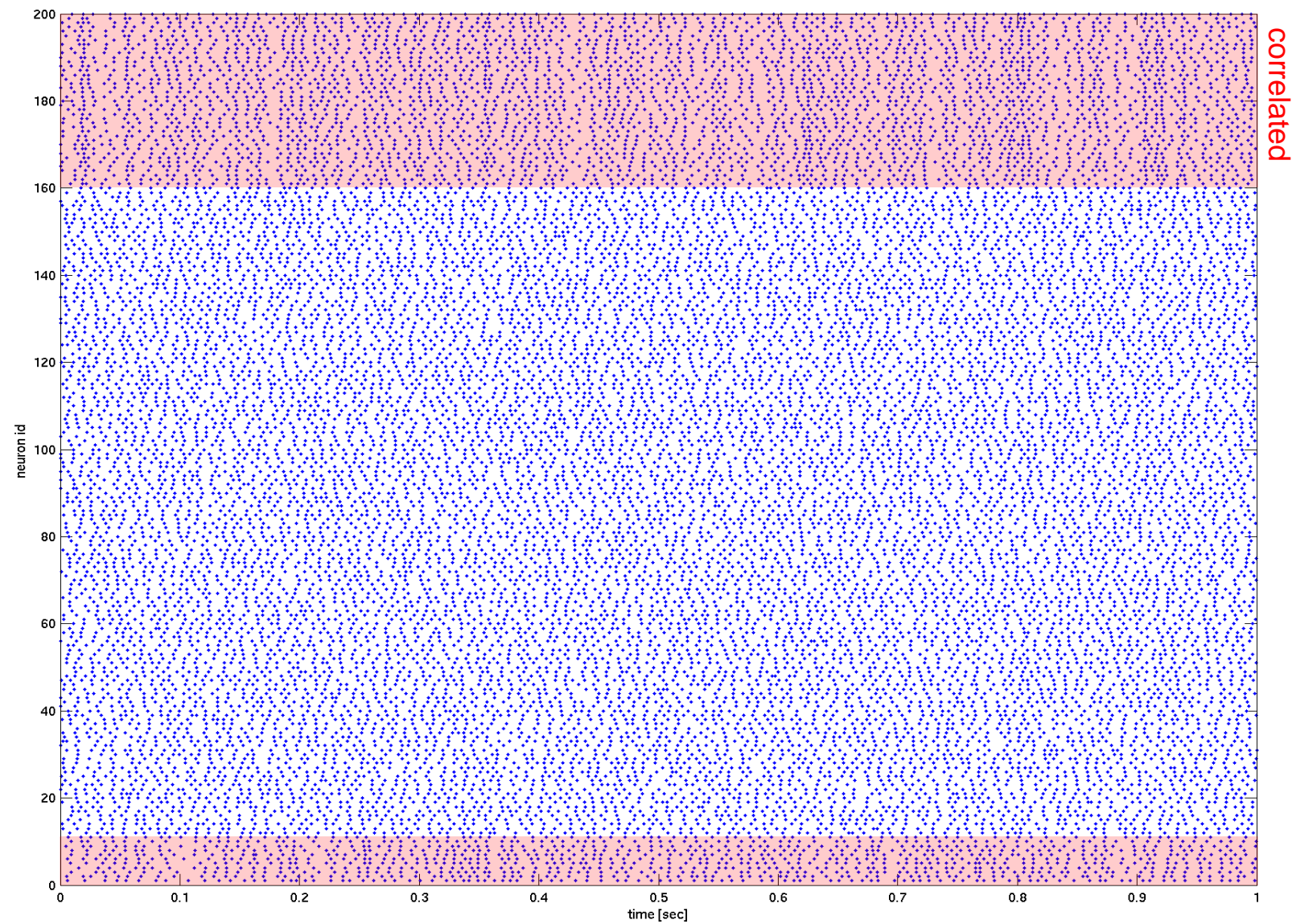 γ-distributed inter-spike

  interval with a shape-factor (γ)

$$p(\tau) = (\gamma \cdot \lambda)^{\gamma} \cdot \tau^{\gamma-1} \cdot \frac{e^{-\gamma \cdot \lambda \tau}}{(\gamma-1)!}$$

Poisson Process

Gamma Process (γ=10)

Discrete Distributions

- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution

Continious Distributions

- Normal (Gauss)
- Gamma Distribution
- chi^2 Distribution

The chi^2 distribution is a special case of the gamma distribution.

Very important for chi^2 goodness of fit test (later in the lecture)

Gamma Distribution

$$p(x) = c \cdot x^{\alpha-1} \, e^{-\frac{x}{\beta}}$$ with x,α,β > 0

c is a norm

chi^2 Distribution

$$p(x) = c \cdot x^{\alpha-1} \, e^{-\frac{x}{\beta}}$$ with α=k/2 , β =2

c is a norm

The chi^2 distribution is a special case of the gamma distribution.

Very important for chi^2 goodness of it test (later in the lecture)

chi^2 distribution:

$$p(x) = c \cdot x^{\alpha-1} e^{-\frac{x}{\beta}}$$ with α=k/2 , β =2

c is a norm

chi^2 distributed with k degrees of freedom

$$p(x) = c \cdot x^{k/2-1} e^{-\frac{x}{2}}$$

Pearson's chi-squared test

The **Pearson's chi-squared goodness of fit** test establishes whether or not an observed frequency distribution differs from a theoretical distribution.

The chi-squared distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables.

The first step in the chi-squared test is to calculate the chi-squared statistic:

Calculating the test-statistic

The value of the test-statistic is

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where

X^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.
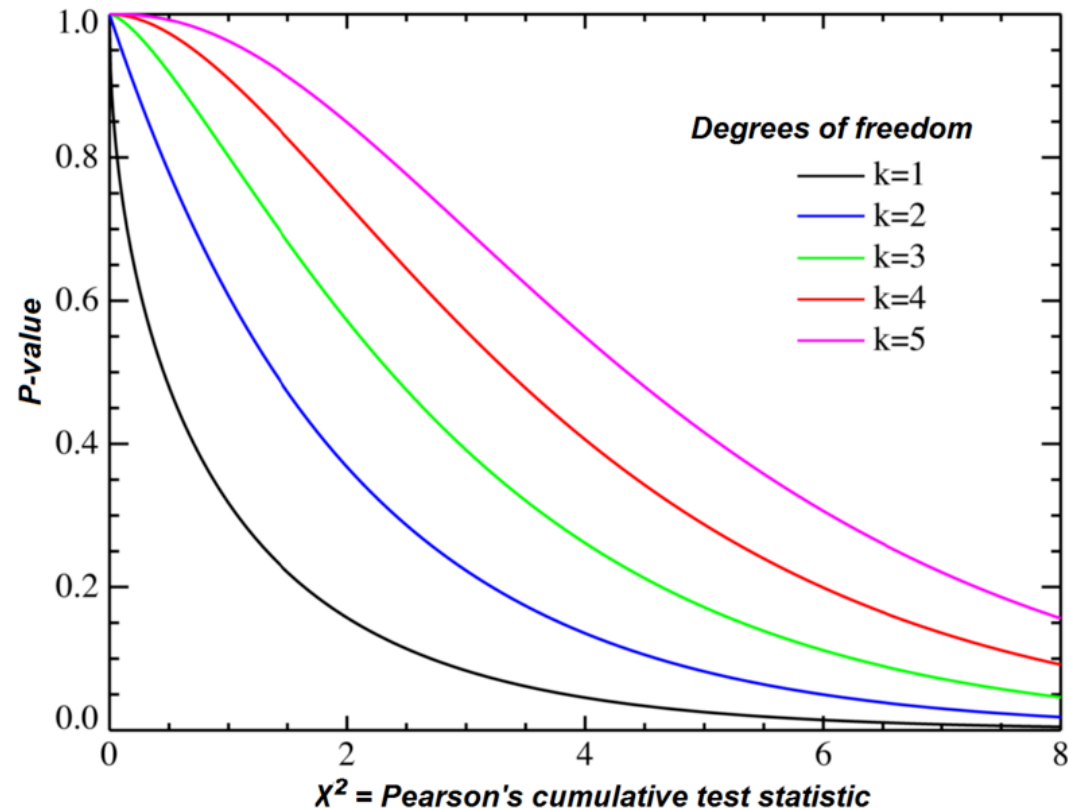
O_i = an observed frequency;

E_i = an expected (theoretical) frequency, asserted by the null hypothesis;

n = the number of cells in the table.

The chi-squared statistic can then be used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution.

The number of degrees of freedom is equal to the number of cells n, minus the reduction in degrees of freedom, p.
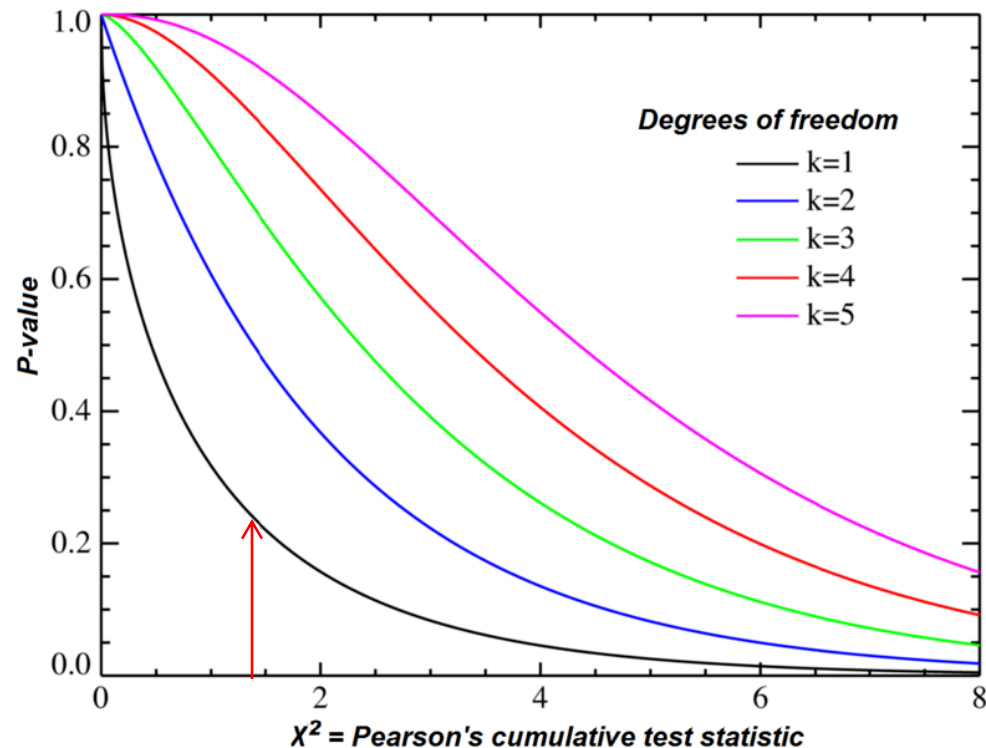

c.f. wikipedia http://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test

For example, to test the hypothesis that a random sample of 100 people has been drawn from a population in which men and women are equal in frequency, the observed number of men and women would be compared to the theoretical frequencies of 50 men and 50 women.

Suppose there were 44 men in the sample and 56 women.

$X^2 = $ Pearson's cumulative test statistic

$$X^2 = \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = 1.44$$

c.f. wikipedia http://en.wikipedia.org/wiki/Pearson%27s_chi-squared_test

The degree of freedom is ONE, since if the male count is known, the female count is determined, and vice-versa.

p = chi2cdf(X2,degree) computes the the p value for a x2 value and a given degree of freedom

We have a sample of 100 people
We have four categories:
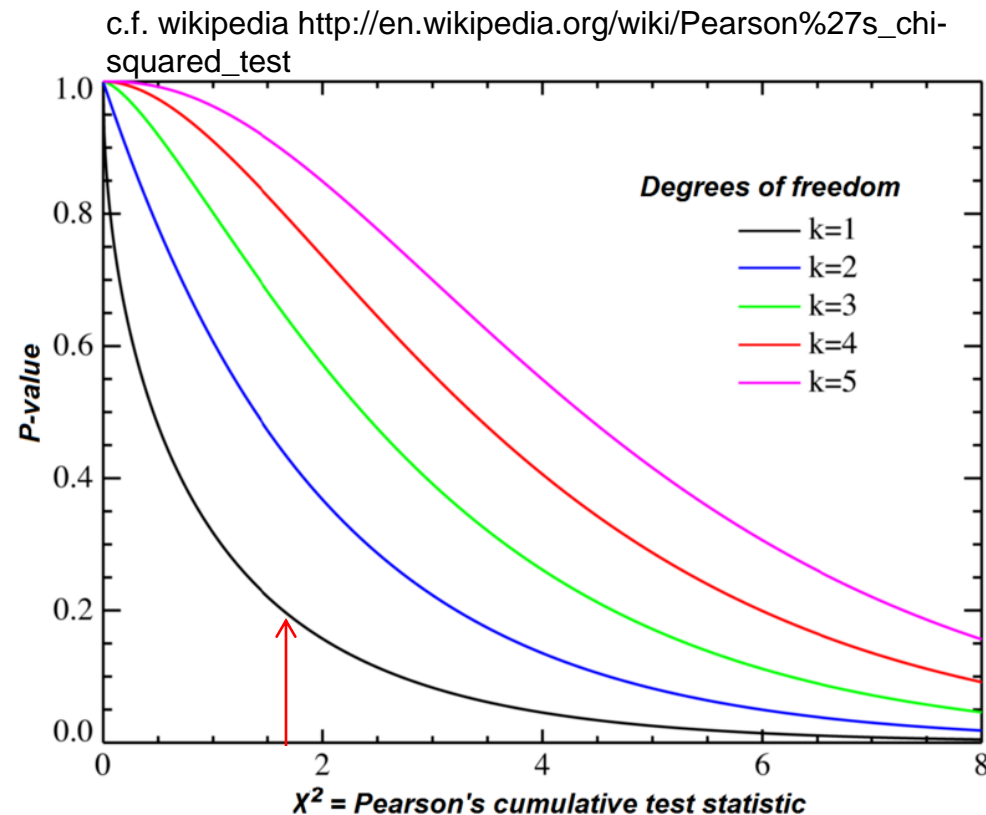
Suppose:

1: Men & born in Osnabrück (10)
2: Men & NOT born in OS (34)
3: Women & born in Osnabrück (10)
4: Women & NOT born in OS (46)

Hypothesis: We p=0.25 for each category .

$$\chi^2 = \frac{(10-25)^2}{25} + \frac{(34-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(46-25)^2}{25}$$

$$= 9 + 3.24 + 9 + 17.64 = 38.88$$

→ p<<0.01

p = chi2cdf(X2,degree) computes the the p value for a x2 value and a given degree of freedom

Q

1. Suppose you observe in two conditions a subjects success in 45, and 55 trials, form a 100. Is that rate different from p=0.5 ?

Degrees of freedom
- k=1
- k=2
- k=3
- k=4
- k=5

P-value

X^2 = Pearson's cumulative test statistic

Calculating the test-statistic

The value of the test-statistic is

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where

X^2 = Pearson's cumulative test statistic, which asymptotically approaches a χ^2 distribution.

O_i = an observed frequency;

E_i = an expected (theoretical) frequency, asserted by the null hypothesis;

n = the number of cells in the table.

Timer (5min): Start — Stop

1. Suppose you observe in two conditions a subjection success in 45, and 55 trials, form a 100. Is that rate different from p=0.5 ?

Calculating the test-statistic

The value of the test-statistic is
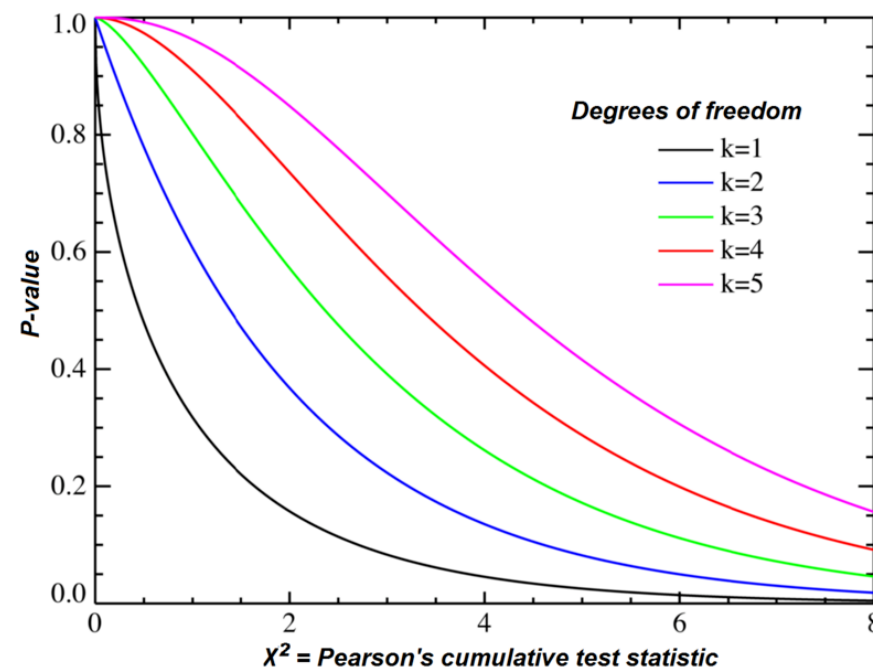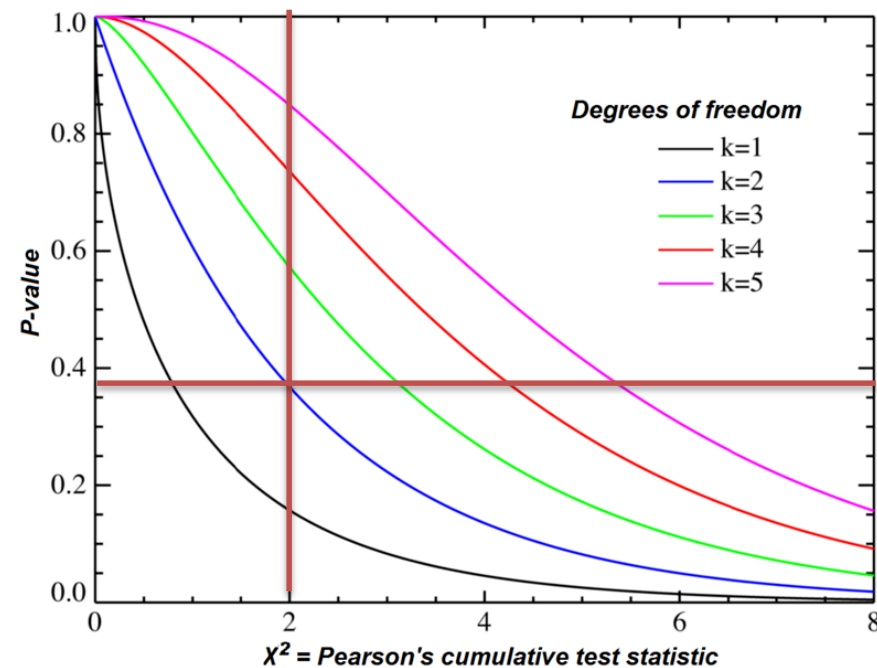
$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

success   failure

$$\chi^2 = \frac{(45-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(45-50)^2}{50} = \frac{4\cdot25}{50} = 2$$

a+b=100   c+d=100   → Degree of freedom is 2

The 4W question session

What is this it all about ?

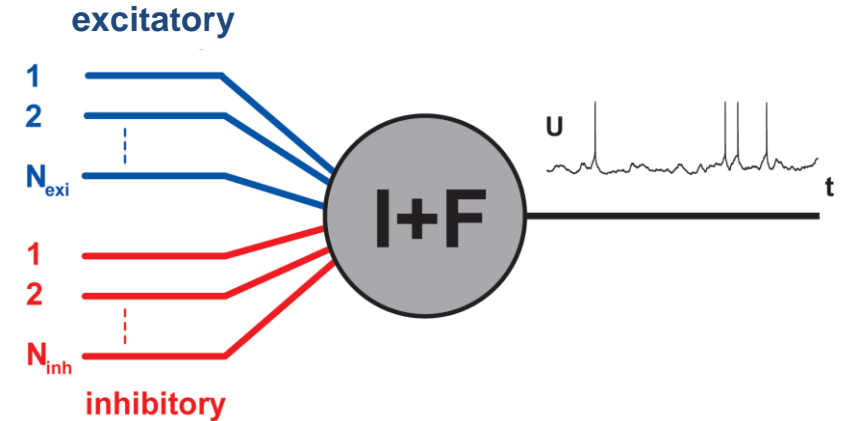**W**hat can **I** use it for ?
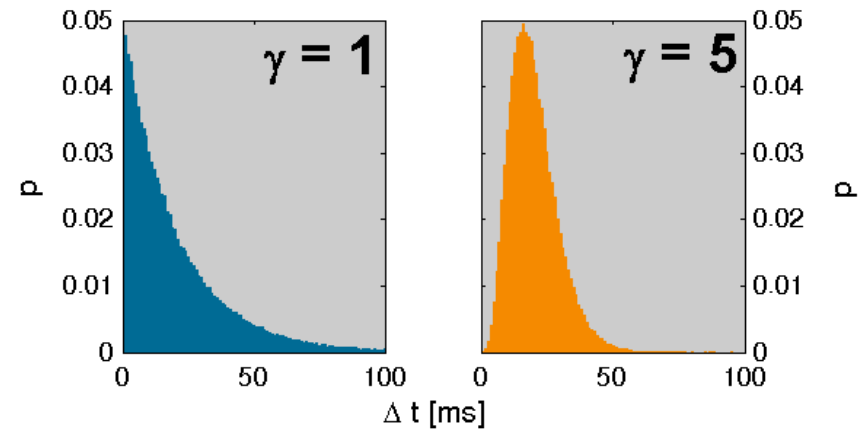
Why should **I** learn it ?

What are potential Bachelor and Master thesis topics ?
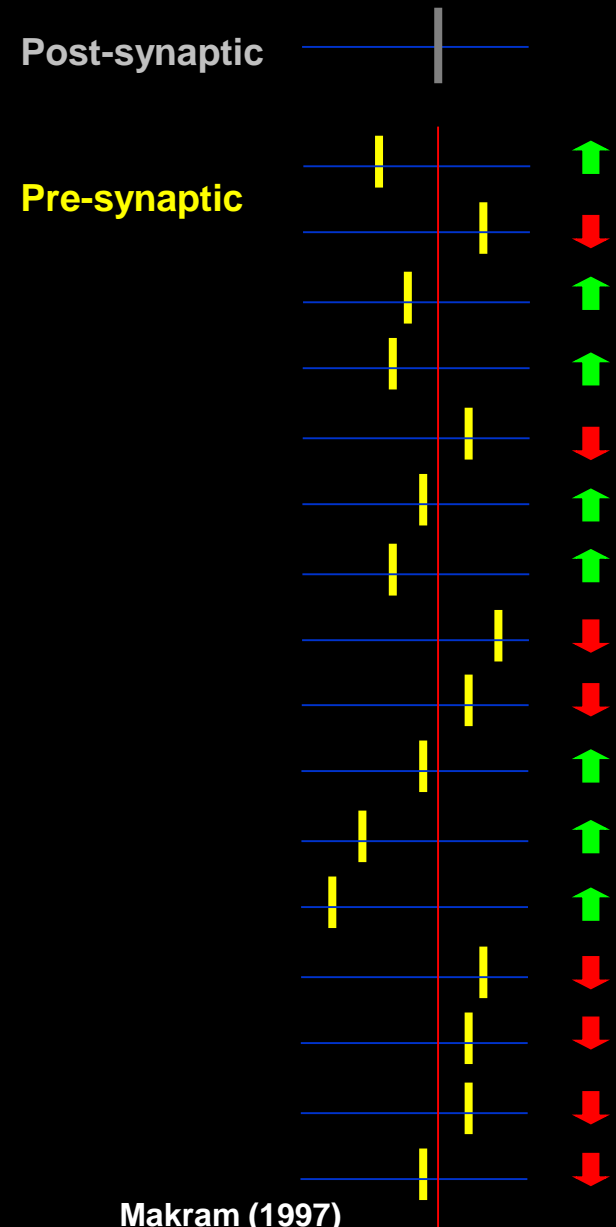
γ-spike trains and a leaky I+F neuron
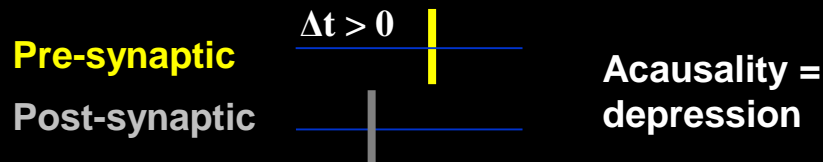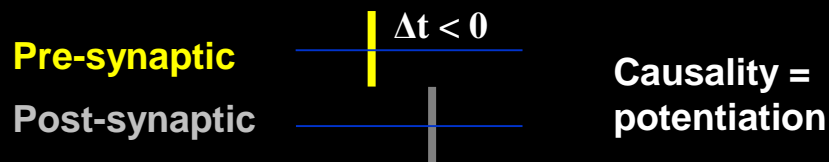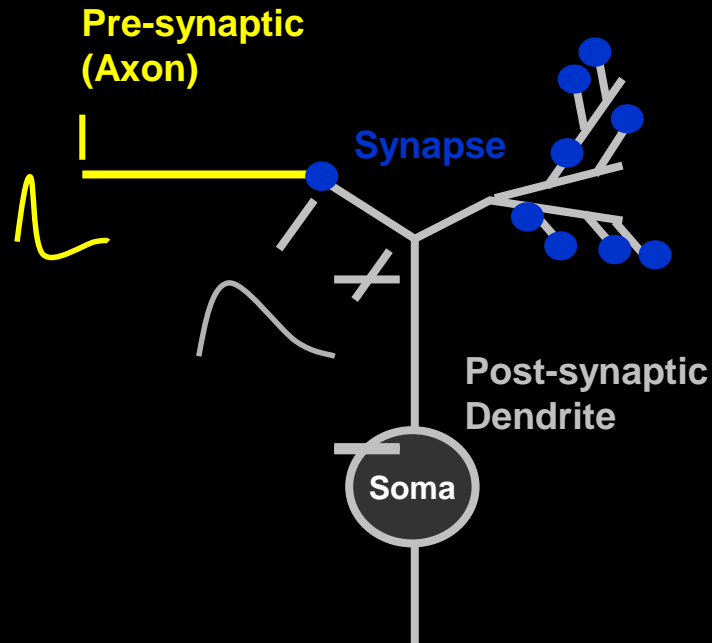
- Presynaptic spike trains :

  γ-distributed inter-spike
  interval with a shape-factor (γ)

$$p(\tau) = (\gamma \cdot \lambda)^{\gamma} \cdot \tau^{\gamma-1} \cdot \frac{e^{-\gamma \cdot \lambda \tau}}{(\gamma - 1)!}$$

Pre-synaptic (Axon)

Synapse

Post-synaptic Dendrite

Soma

Post-synaptic

Pre-synaptic

$\Delta t < 0$

Pre-synaptic

Post-synaptic

Causality = potentiation

$\Delta t > 0$

Pre-synaptic

Post-synaptic

Acausality = depression

Makram (1997)
Abbott (2001)

Spike train auto-structure impacts post-synaptic firing and timing-based plasticity

Bertram Scheller[1][†], Marta Castellano[2,3,4][†], Raul Vicente[3,4] and Gordon Pipa[2,3,4]*

[1] Clinic for Anesthesia, Intensive Care Medicine and Pain Therapy, Johann Wolfgang Goethe University, Frankfurt am Main, Germany
[2] Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany
[3] Department of Neurophysiology, Max-Planck-Institute for Brain Research, Frankfurt am Main, Germany
[4] Frankfurt Institute for Advanced Studies, Johann Wolfgang Goethe University, Frankfurt am Main, Germany

Cortical neurons are typically driven by several thousand synapses. The precise spatiotemporal pattern formed by these inputs can modulate the response of a post-synaptic cell. In this work, we explore how the temporal structure of pre-synaptic inhibitory and excitatory inputs impact the post-synaptic firing of a conductance-based integrate and fire neuron. Both the excitatory and inhibitory input was modeled by renewal *gamma* processes with varying shape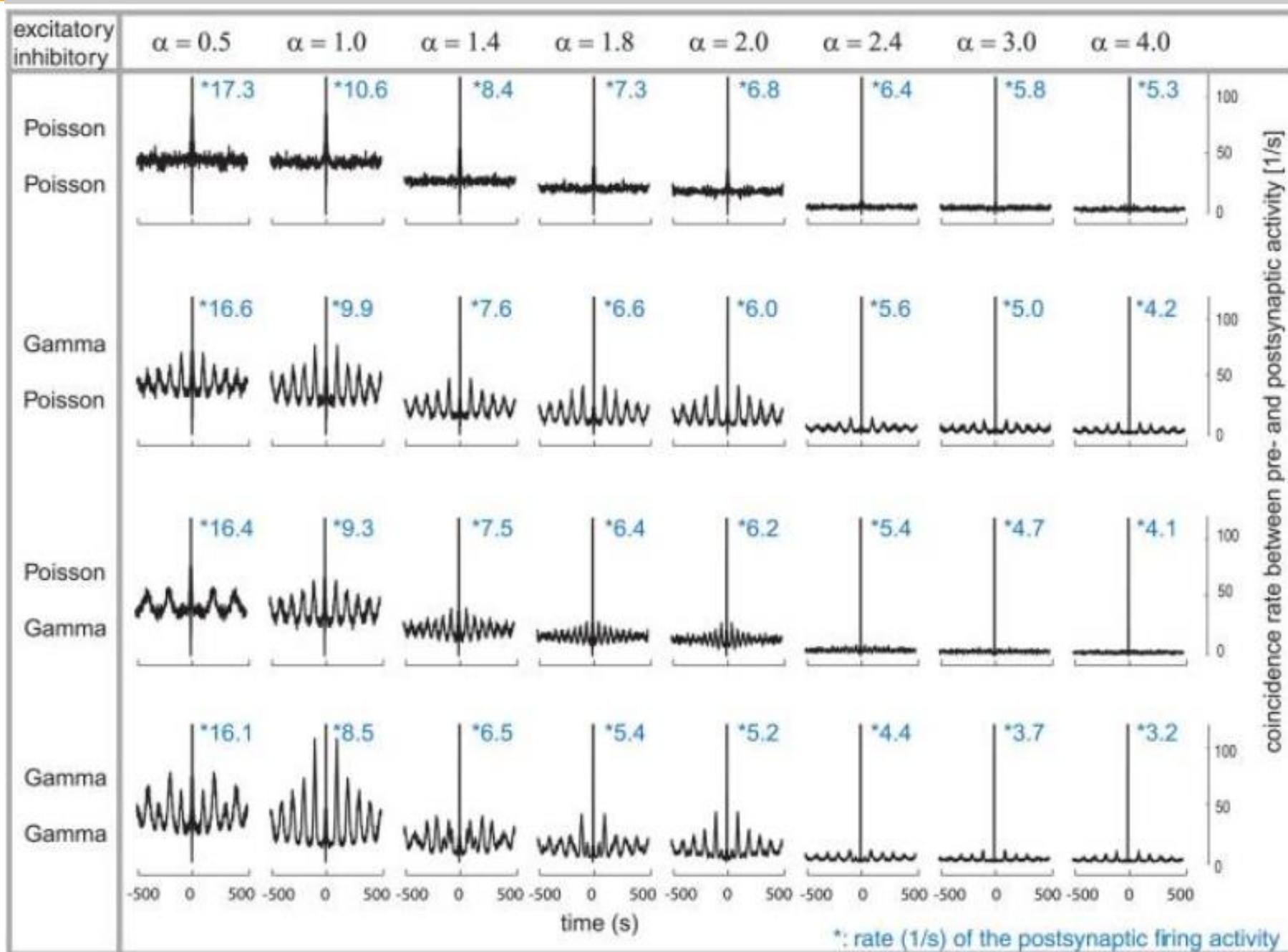 factors for modeling regular and temporally random *Poisson* activity. We demonstrate that the temporal structure of mutually independent inputs affects the post-synaptic firing, while the strength of the effect depends on the firing rates of both the excitatory and inhibitory inputs. In a second step, we explore the effect of temporal structure of mutually independent inputs on a simple version of Hebbian learning, i.e., hard bound spike-timing-dependent plasticity. We explore both the equilibrium weight distribution and the speed of the transient weight dynamics for different mutually independent *gamma* processes. We find that both the equilibrium distribution of the synaptic weights and the speed of synaptic changes are modulated by the temporal structure of the input. Finally, we highlight that the sensitivity of both the post-synaptic firing as well as the spike-timing-dependent plasticity on the auto-structure of the input of a neuron could be used to modulate the learning rate of synaptic modification.

Keywords: spike train, auto-structure, STDP, temporal correlations, integrate and fire, non-*Poissonian*
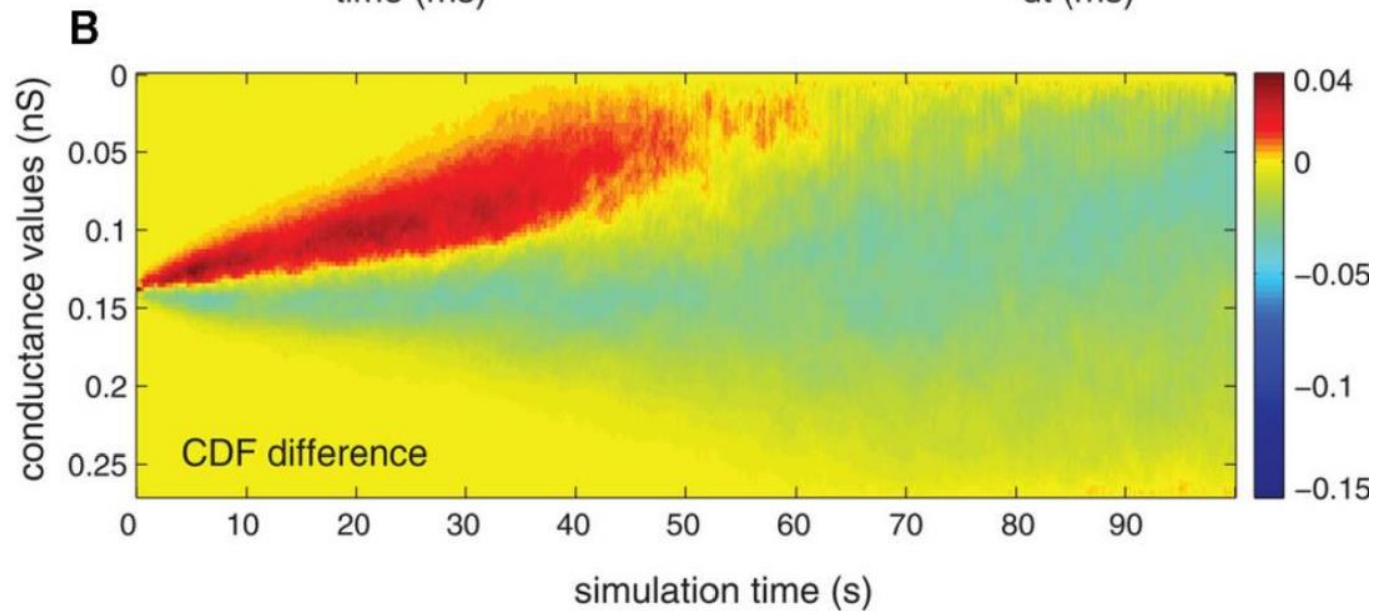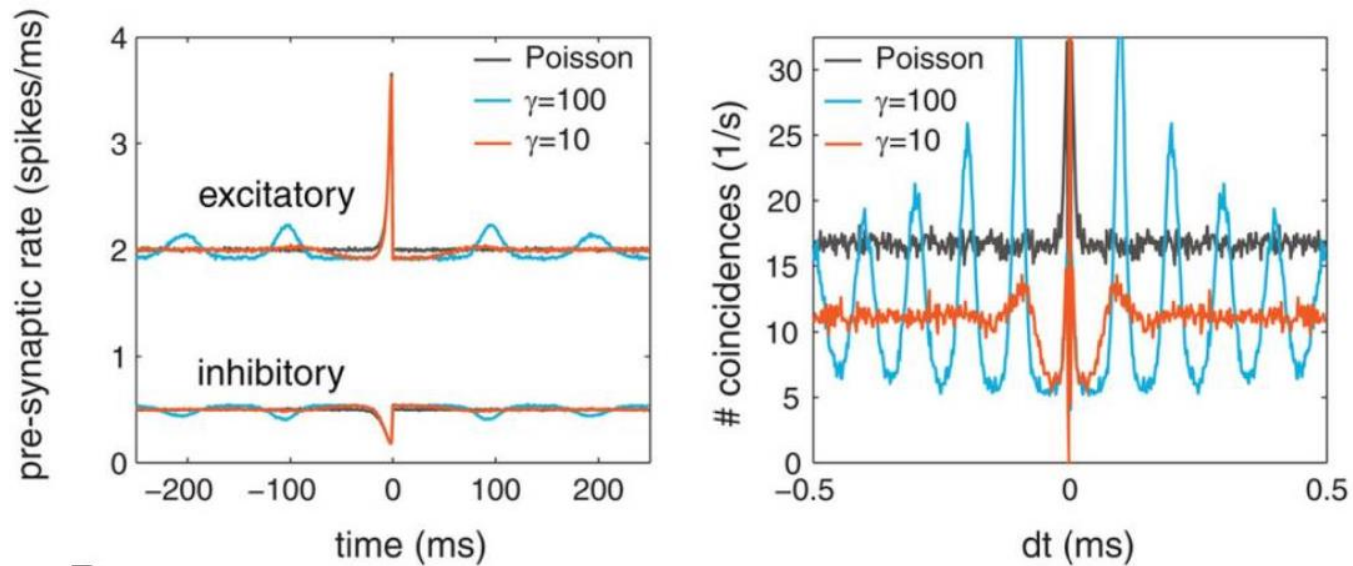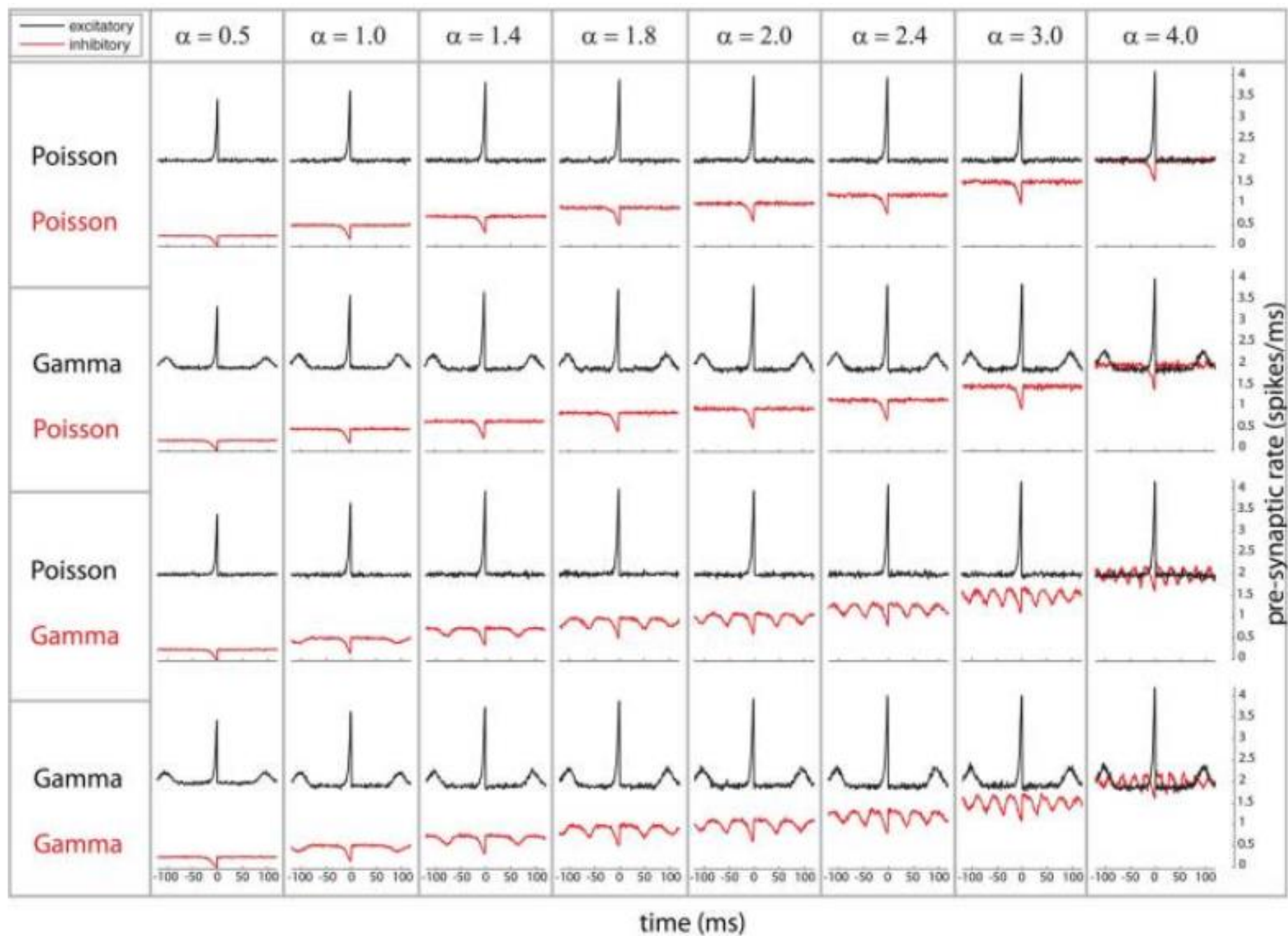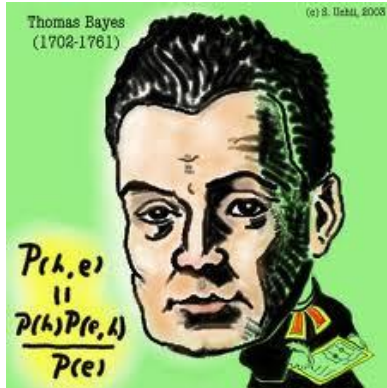
B

CDF difference

excitatory inhibitory	$\alpha = 0.5$	$\alpha = 1.0$	$\alpha = 1.4$	$\alpha = 1.8$	$\alpha = 2.0$	$\alpha = 2.4$	$\alpha = 3.0$	$\alpha = 4.0$
Poisson Poisson	*17.3	*10.6	*8.4	*7.3	*6.8	*6.4	*5.8	*5.3
Gamma Poisson	*16.6	*9.9	*7.6	*6.6	*6.0	*5.6	*5.0	*4.2
Poisson Gamma	*16.4	*9.3	*7.5	*6.4	*6.2	*5.4	*4.7	*4.1
Gamma Gamma	*16.1	*8.5	*6.5	*5.4	*5.2	*4.4	*3.7	*3.2

time (s)

coincidence rate between pre- and postsynaptic activity [1/s]

*: rate (1/s) of the postsynaptic firing activity

Thomas Bayes
(1702-1761)

$P(h,e) = \dfrac{P(h)P(e,h)}{P(e)}$

Special Issue: Probabilistic models of cognition

Bayesian decision theory in sensorimotor control

Konrad P. Körding[1] and Daniel M. Wolpert[2]

[1]Brain and Cognitive Sciences, Massachusetts Institute of Technology, Building NE46-4053, Cambridge, Massachusetts, 02139, USA
[2]Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK

(b)

Likelihood

Posterior

Prior

Decision theory quantifies how people should choose in the context of a given utility function and some partial knowledge of the world. The expected utility is defined as:

$$E[Utility] \equiv \sum_{\substack{possible \\ outcomes}} p(outcome|action)U(outcome)$$

where $p(outcome|action)$ is the probability of an outcome given an action and $U(outcome)$ is the utility associated with this outcome.
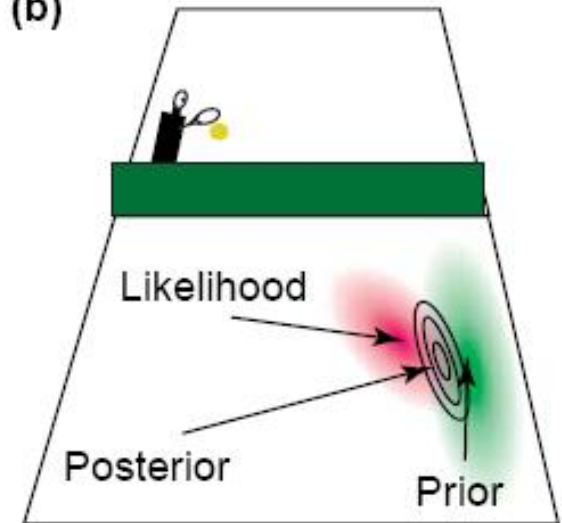
Outcome defined by:

- Prior of ball position
- Likelihood of ball position
- Prior of success in making a point dependent on the position

Is the brain really Bayesian ?

Experiment :

- A subject needs to point to a target.
- There are three types of noise leading uncertainty

(a) Sensor noise   Motor noise

Visual

Proprioceptive

Box 2. Bayesian statistics

When we have a Gaussian prior distribution $p(x)$ and we have a noisy observation o of the position that leads to a Gaussian likelihood (red curve, Figure I) $p(o|x)$ it is possible to use Bayes rule to calculate the posterior distribution (yellow curve, Figure I; how probable is each value given both the observation and the prior knowledge):

$$p(x|o) = p(o|x)\frac{p(x)}{p(o)}$$

This equation assigns a probability to every possible location. If we assume that the prior distribution $p(x)$ is a symmetric one dimensional Gaussian with variance σ_p^2 and mean $\hat{\mu}$ and that the likelihood $p(o|x)$ is also a symmetric one dimensional Gaussian with variance σ_o^2 and mean o, it is possible to compute the posterior that is then also Gaussian in an analytical way. The optimal estimate $\hat{x}$, that is the maximum of the posterior is:

$$\hat{x} = \alpha o + (1-\alpha)\hat{\mu}$$

where

$$\alpha = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2}$$

Moreover we can calculate the width of the posterior as $\sigma^2 = \alpha\sigma_o$. The parameter α is always less than 1. This Bayesian approach leads to a better estimate of possible outcomes than any estimate that is only based on the sensory input.
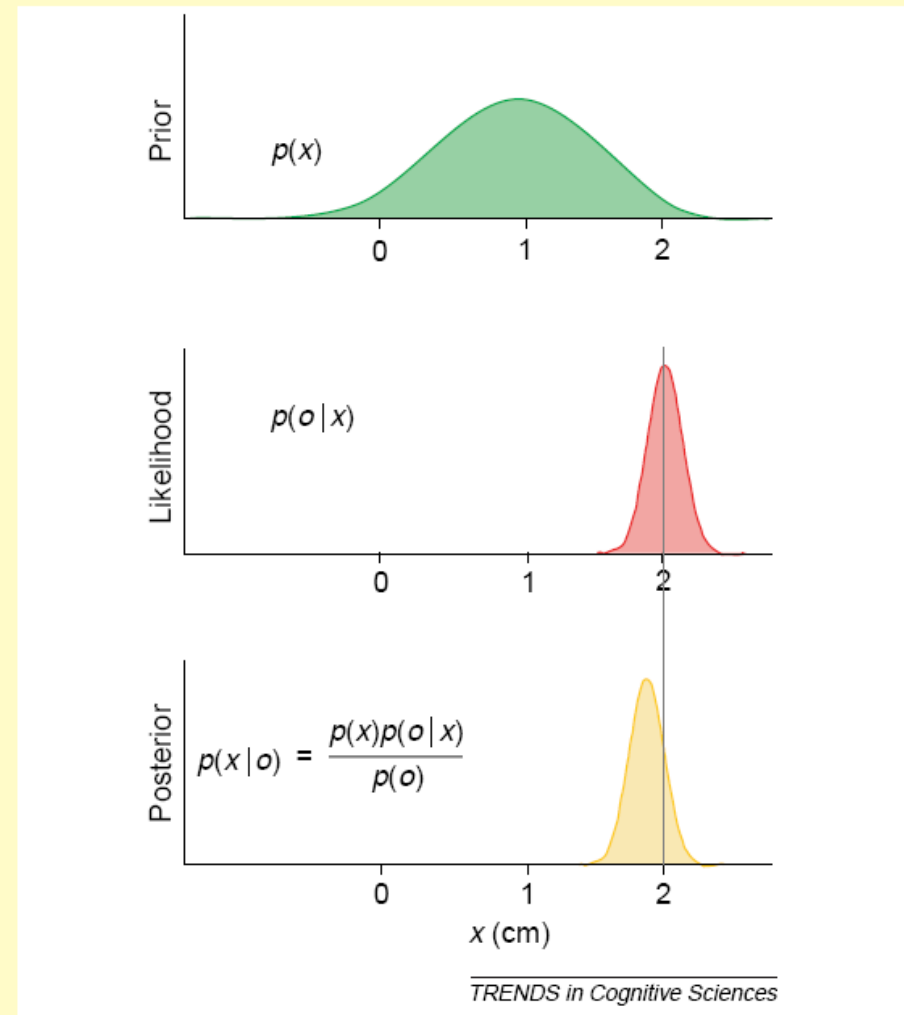

TRENDS in Cognitive Sciences

Figure I. Bayesian integration. The green curve represents the prior and red curve represents the likelihood. The yellow curve represents the posterior, the result from combining prior and likelihood.
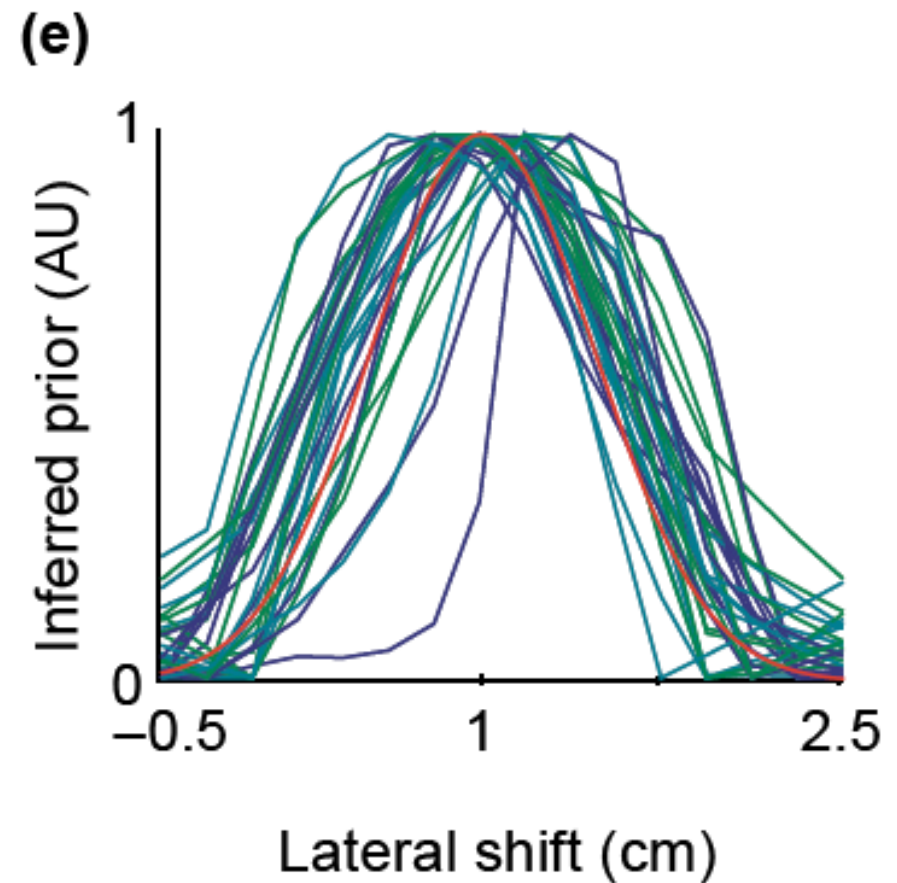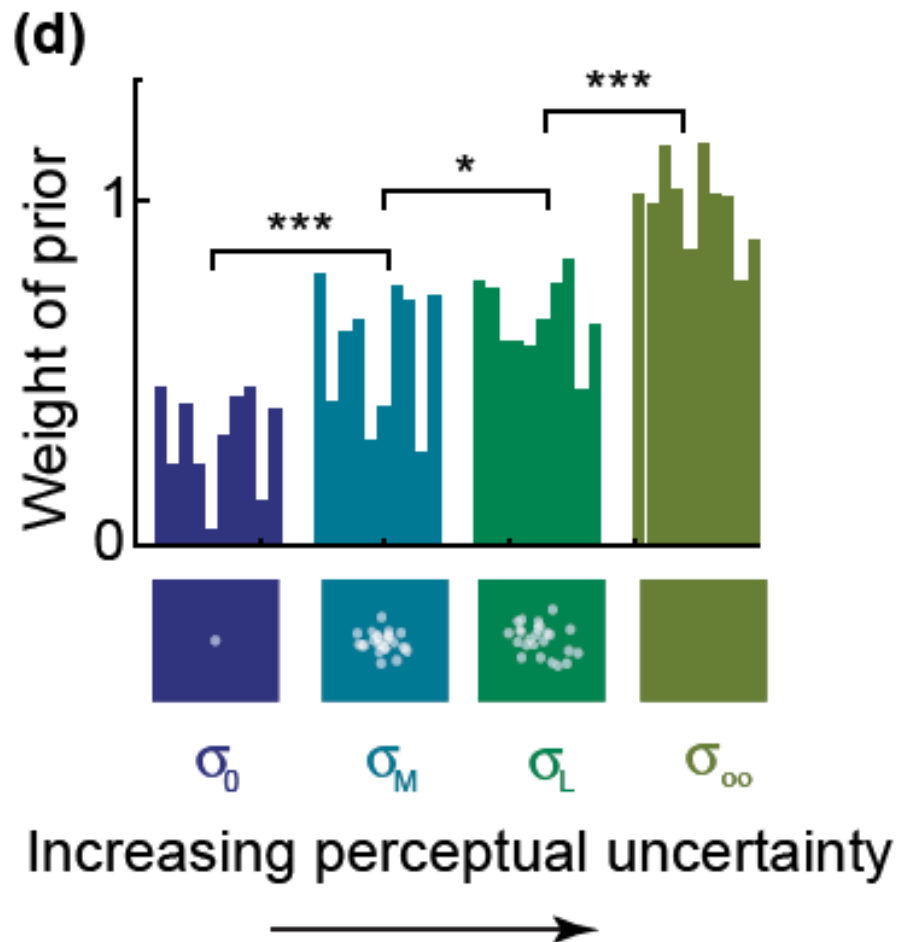
Is the brain really Bayesian ?

Experiment :

- A subject needs to point to a target.
- There are three types of noise leading uncertainty

Test:

- Can the subject learn the uncertainty of the indiviudal sources to estimate the width of the likelihood?

- Can the subject learn a prior?

- Can it combine the prior and likelihood such that it makes a decision that is maximizing the posterior?

(d) Weight of prior vs increasing perceptual uncertainty (σ_0, σ_M, σ_L, σ_{∞}) with significance markers ***, *, ***

(e) Inferred prior (AU) vs Lateral shift (cm)

Yes we can, actually for this example it is Baysian optimal!