Neuroinformatics Lecture (L8)

Prof. Dr. Gordon Pipa

Institute of Cognitive Science University of Osnabrück

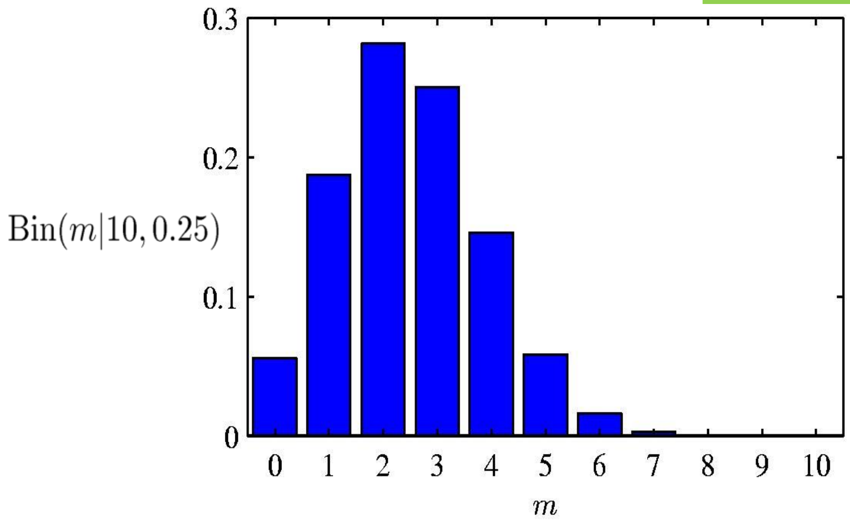
Summary: Bernoulli Distribution

Distribution: $p(x|\mu) = \mu^x(1-\mu)^{1-x}$

Expected value: $E[x] = \mu$

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

Summary: Binomial Distribution



$$\mathrm{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$$

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

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

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.

1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

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

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

Independent of N and p, given λ is constant

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

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

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

1777-1855

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

$$\begin{aligned} X &= 1 + 2 + \ldots + 100 \\ &= (1+100) + (2+99) + \ldots + (50+51) \\ &= 50 * 51 \\ &= 5050 \end{aligned}$$

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

The Gaussian Distribution

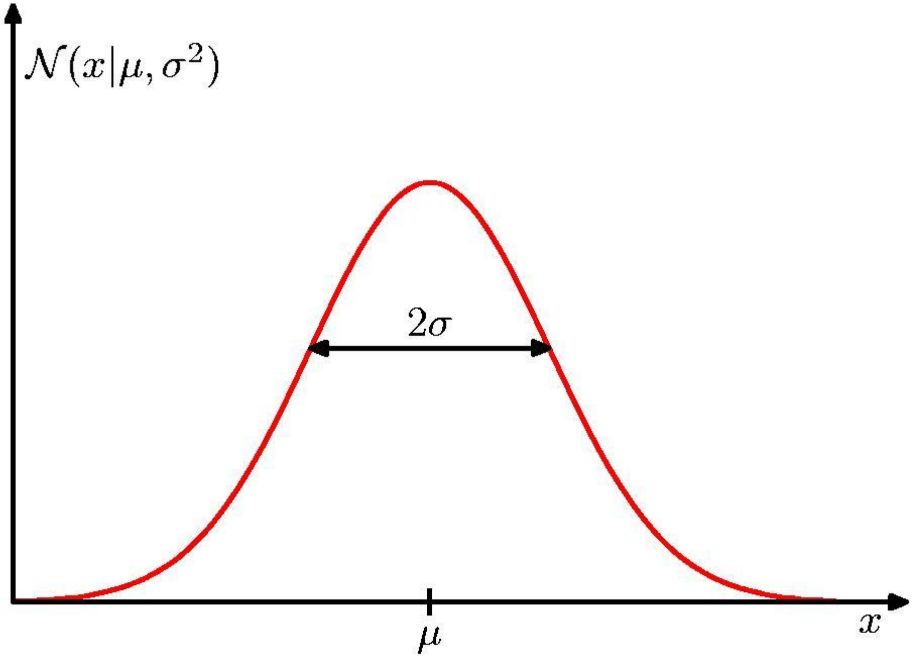
$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

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.

Gaussian Mean and Variance

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

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

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

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

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

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

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

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

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

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

Central Limit Theorem

Let $\{X_1, X_2, \ldots, X_n\}$ 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: $\frac{1}{N}\sigma^2$

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:

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

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

$Z_2 = X_1+X_2$

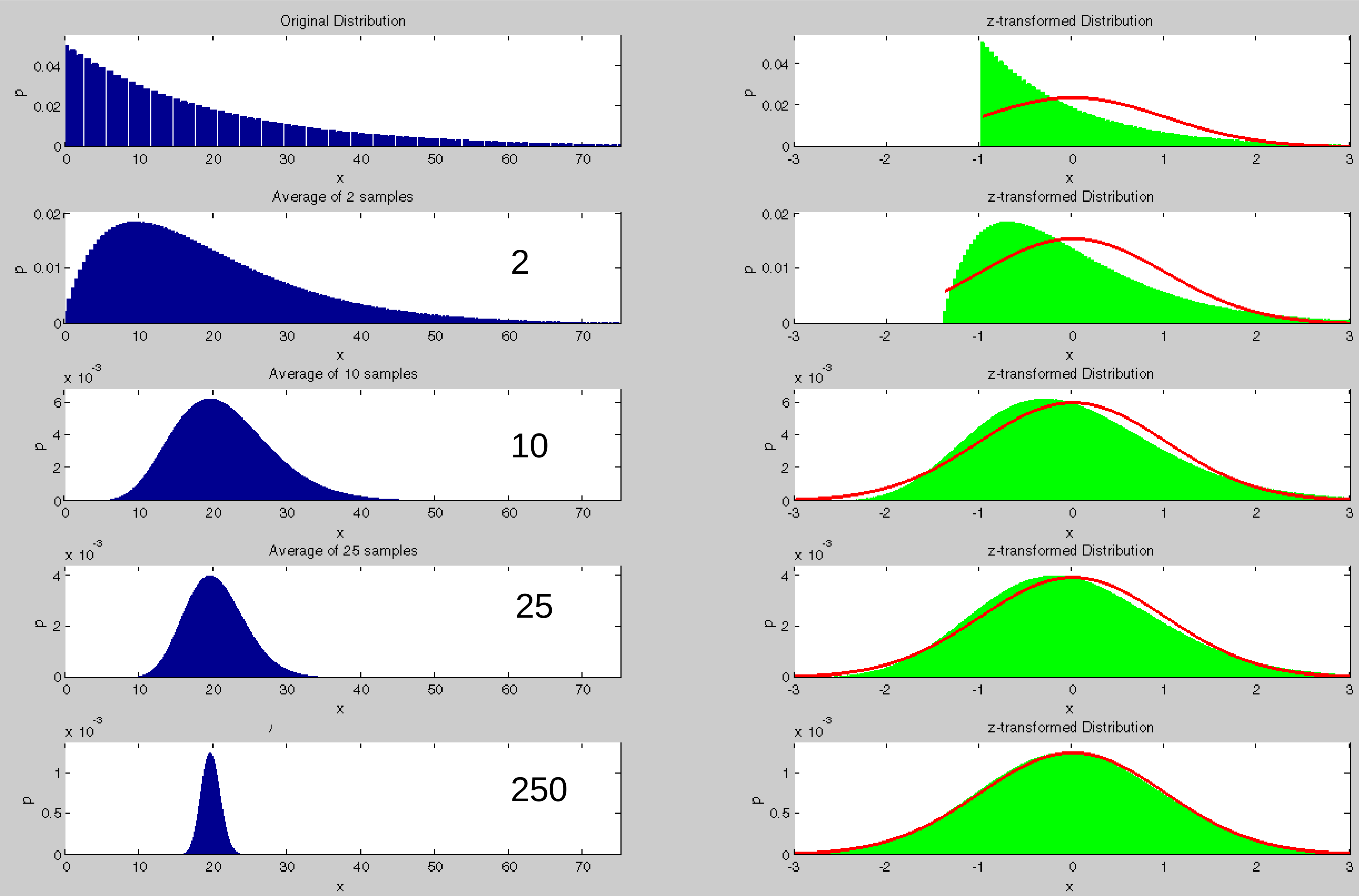
$$p(Z_2) = \int_{-\infty}^{\infty} p^{x_1}(X_1) p^{x_2}(Z_2 - X_1) dX_1$$

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

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

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

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



File convolution_example.m to produce this figures is only online

```
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.

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

```
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')
```

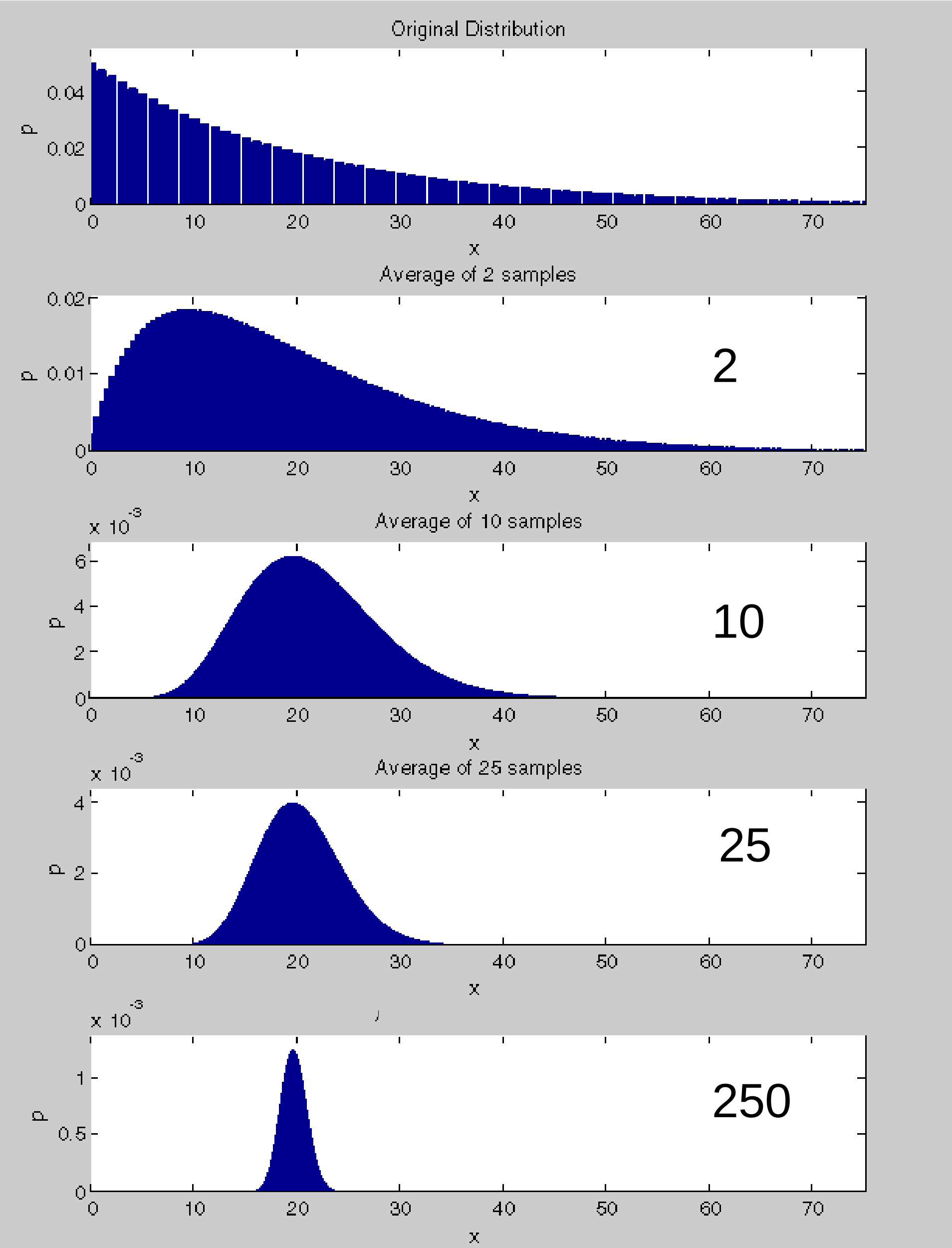
Plots the distribution in an grid of 5x2 plots.

```
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')
```

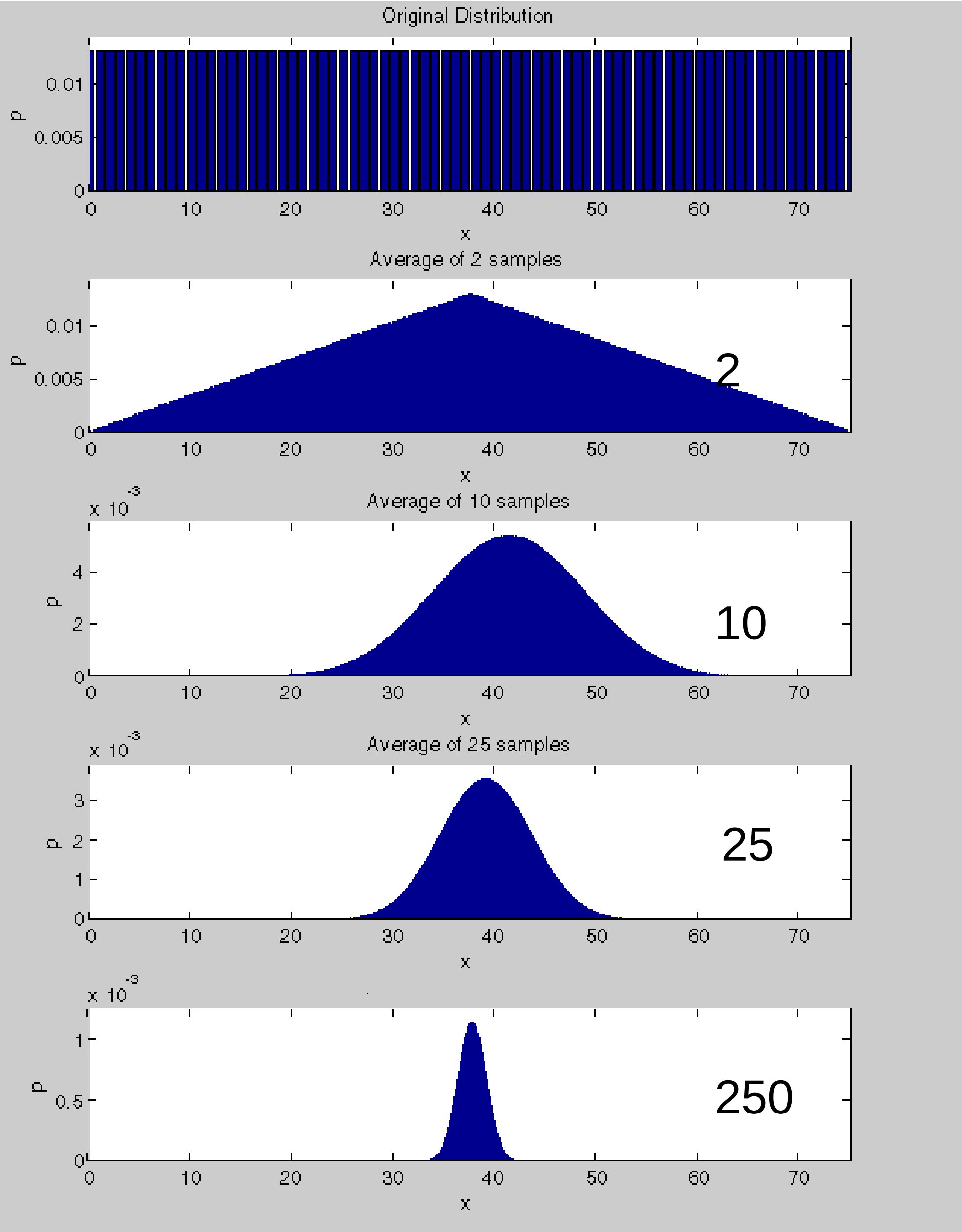
z-tranforming here

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

$$\begin{aligned}\operatorname{Var}(X) &= \operatorname{E}\left[X^2 - 2X\operatorname{E}[X] + (\operatorname{E}[X])^2\right] \\ &= \operatorname{E}\left[X^2\right] - 2\operatorname{E}[X]\operatorname{E}[X] + (\operatorname{E}[X])^2 \\ &= \operatorname{E}\left[X^2\right] - (\operatorname{E}[X])^2\end{aligned}$$



File convolution_example.m to produce this figures is online on studip



File convolution_example.m to produce this figures is online on studip

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

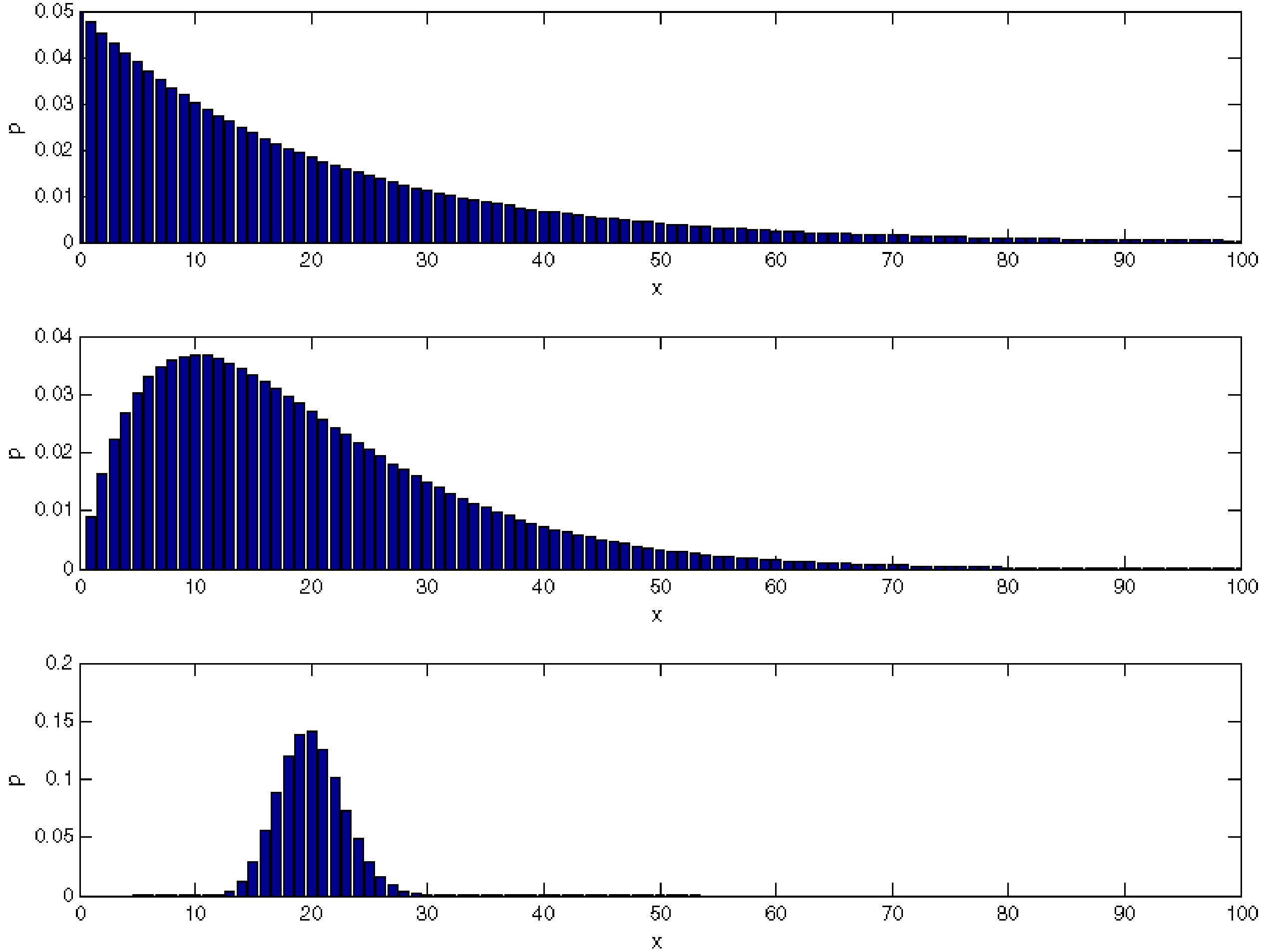
c is used to normalize such that

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

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

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

$$E(x) = \alpha \cdot \beta$$

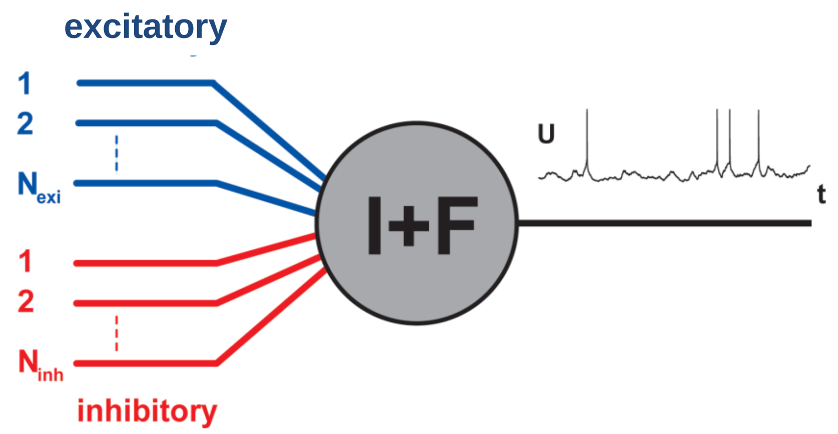
$$var(x) = \alpha \cdot \beta^2$$

γ-spike trains and a leaky I+F neuron

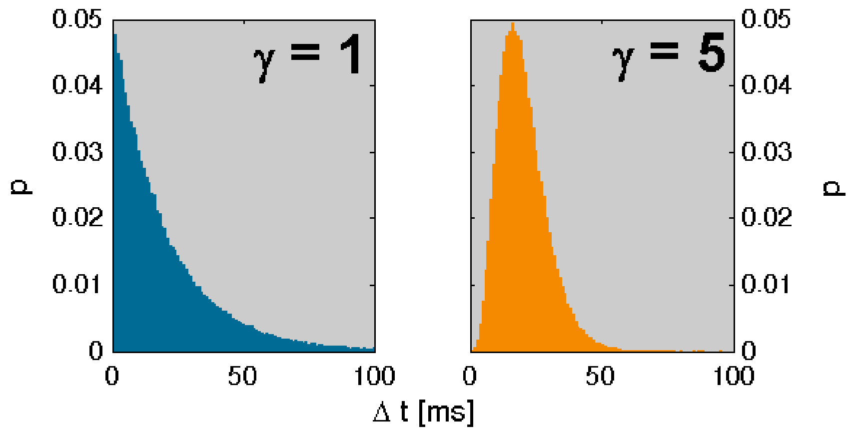
- Presynaptic spike trains :

  γ-distributed inter-spike

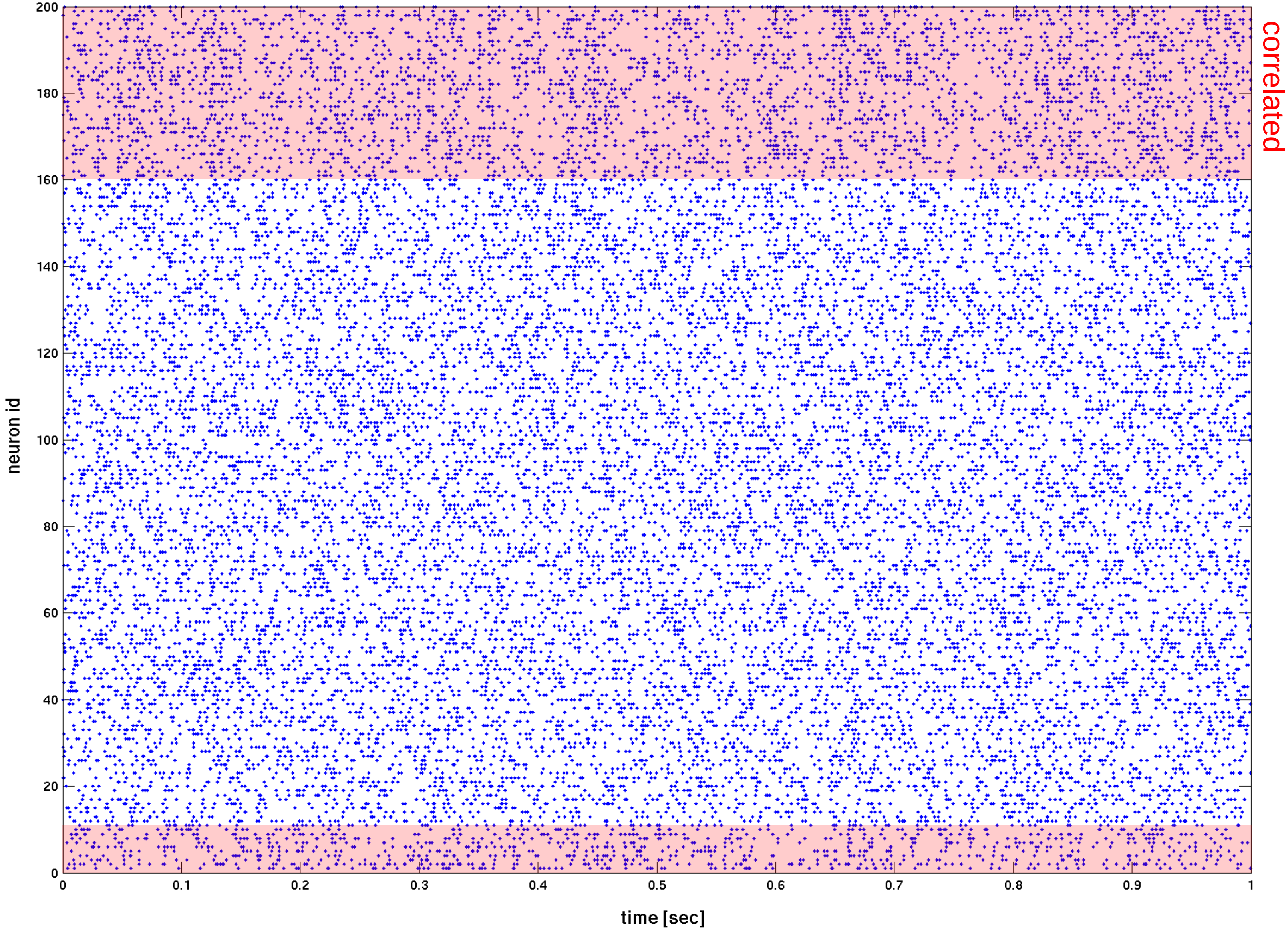
  interval with a shape-factor (γ)



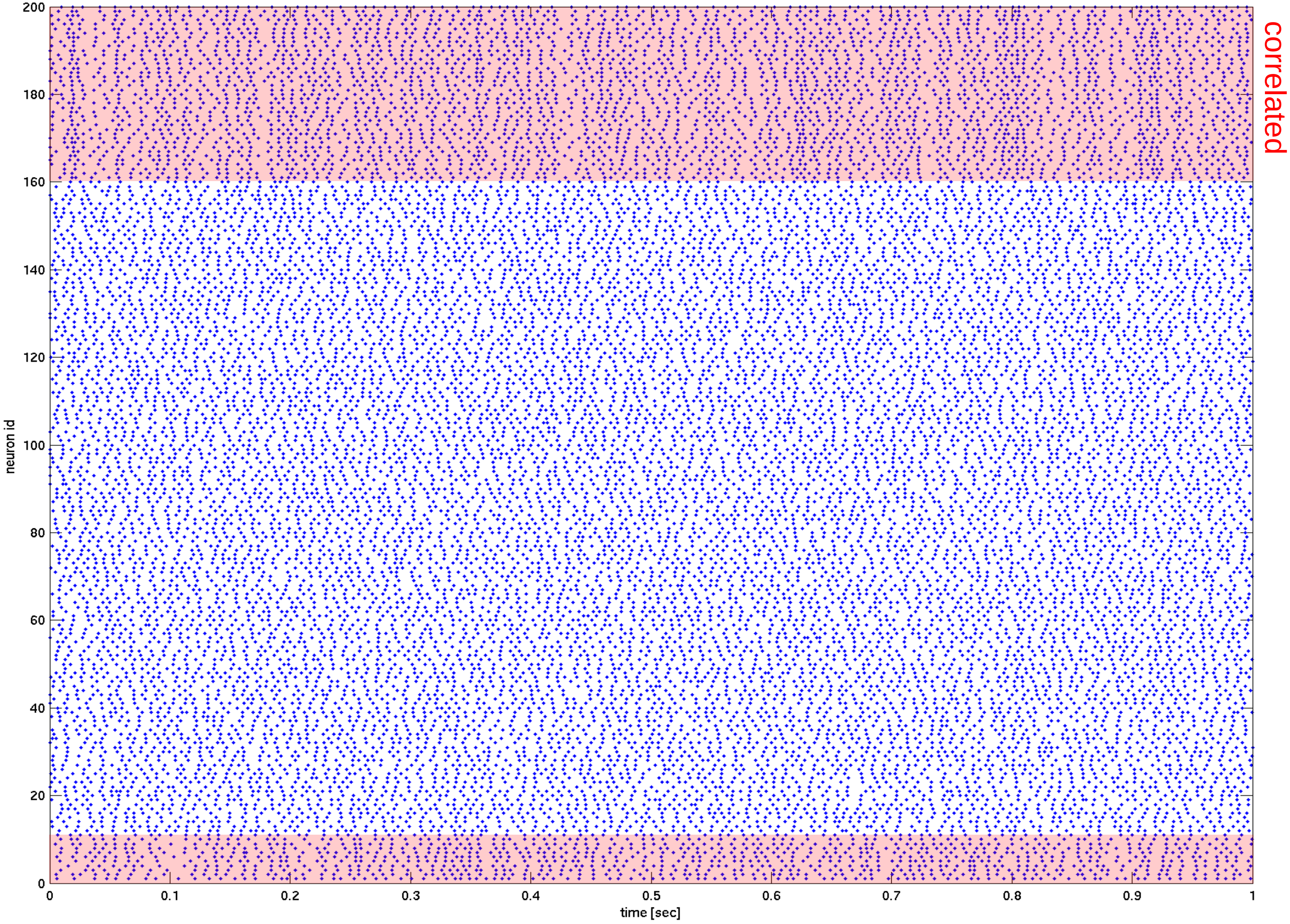
$$p(\tau) = (\gamma \cdot \lambda)^{\gamma} \cdot \tau^{\gamma-1} \cdot \frac{e^{-\gamma \cdot \lambda \tau}}{(\gamma - 1)!}$$



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.

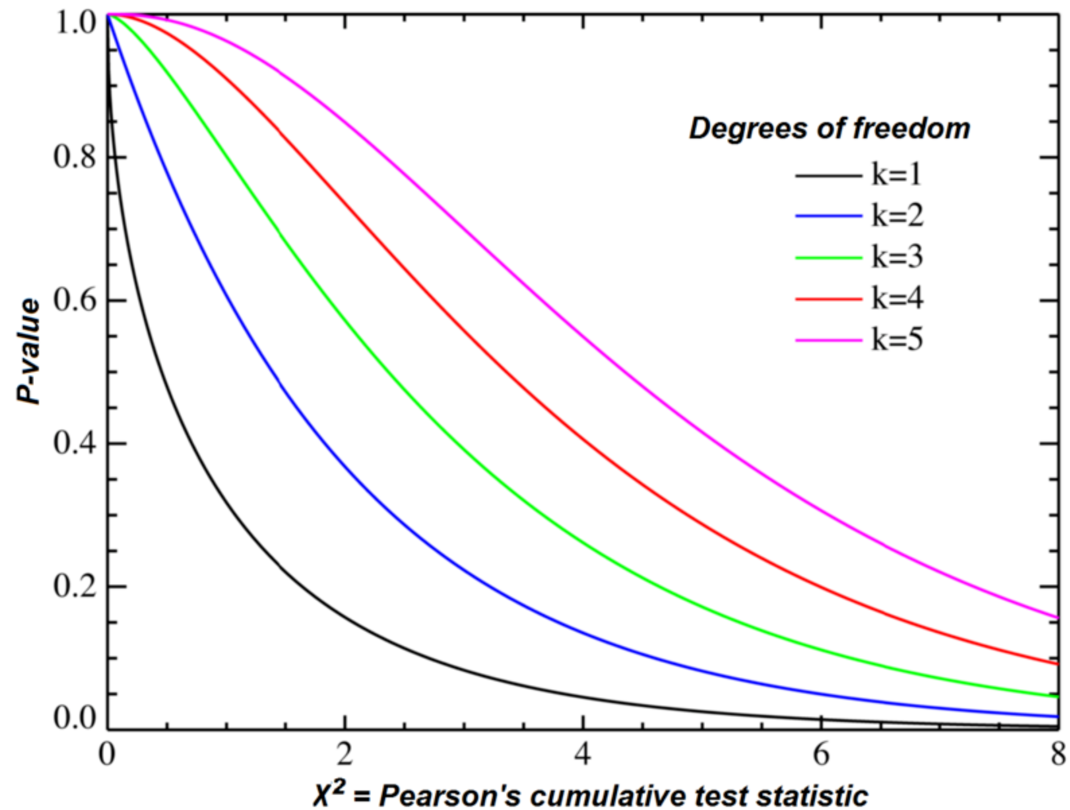
O_i = an observed frequency;

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

n = the number of cells in the table.

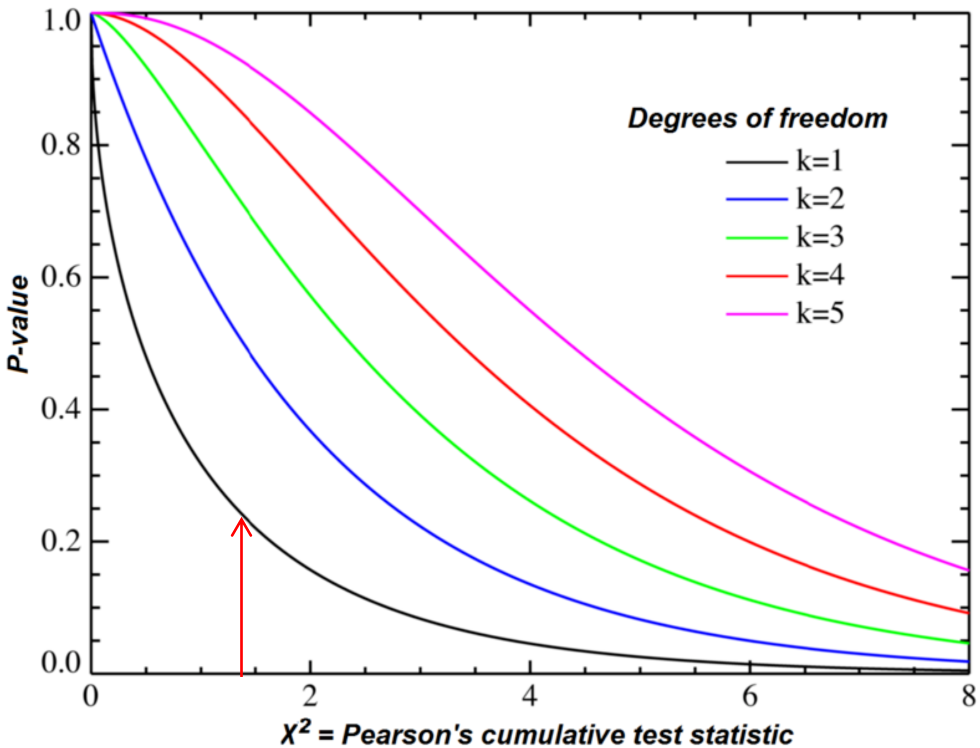
The chi-squared statistic can then be used to calculate a p-value by comparing the value of the statistic to a chi-squared distribution.

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.



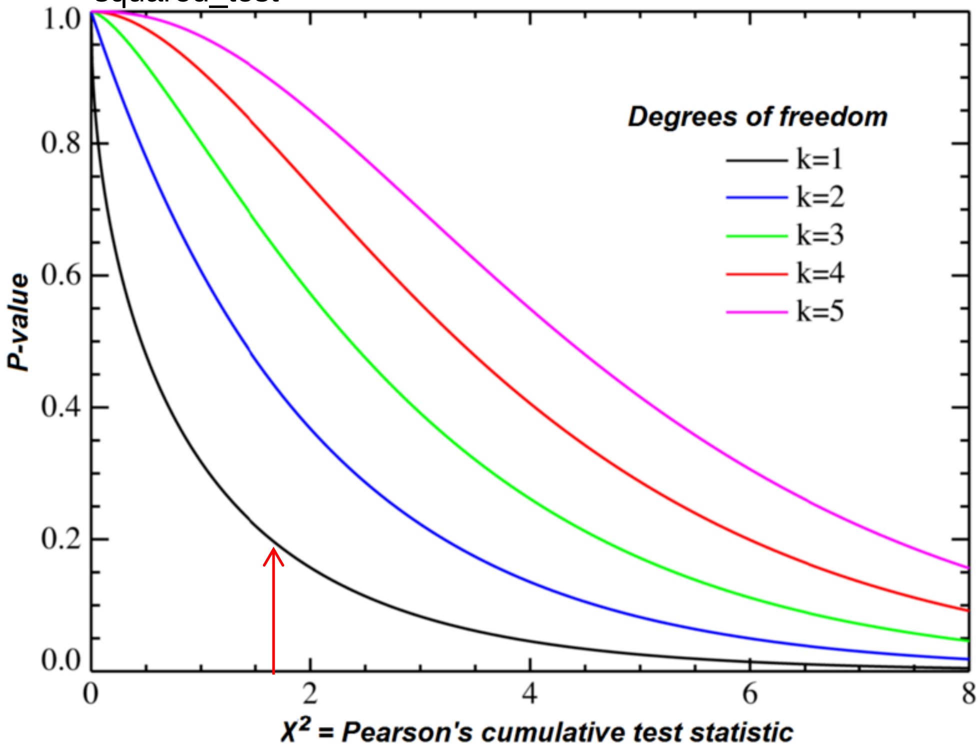
Suppose there were 44 men in the sample and 56 women.

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

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



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)

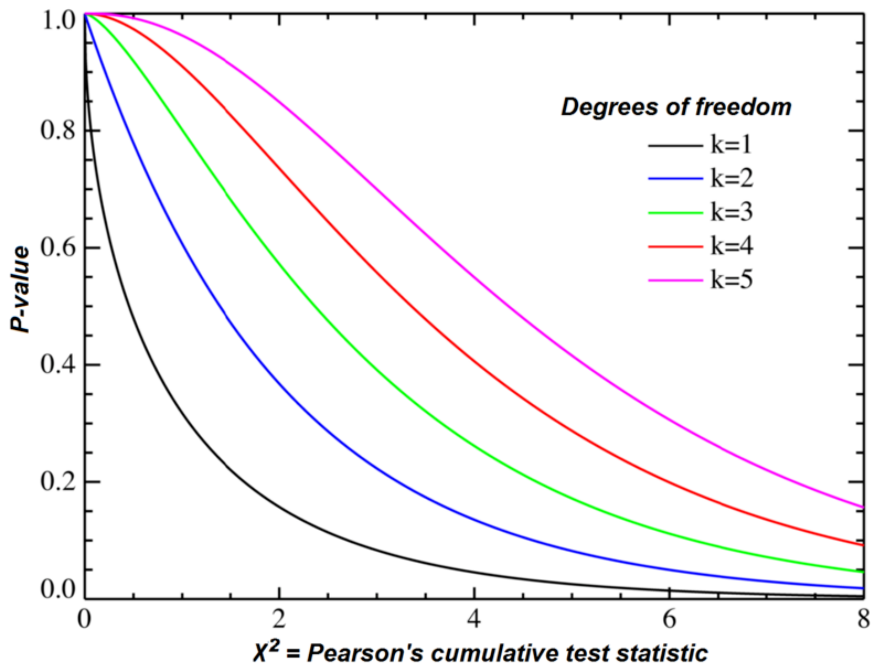
Hypothesis: We p=0.25 for each category .

$$\chi^2 = \frac{(10-25)^2}{25} + \frac{(34-25)^2}{25} + \frac{(10-25)^2}{25} + \frac{(46-25)^2}{25}$$
$$= 9 + 3.24 + 9 + 17.64 = 38.88$$

→ p<<0.01

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

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 ?



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.

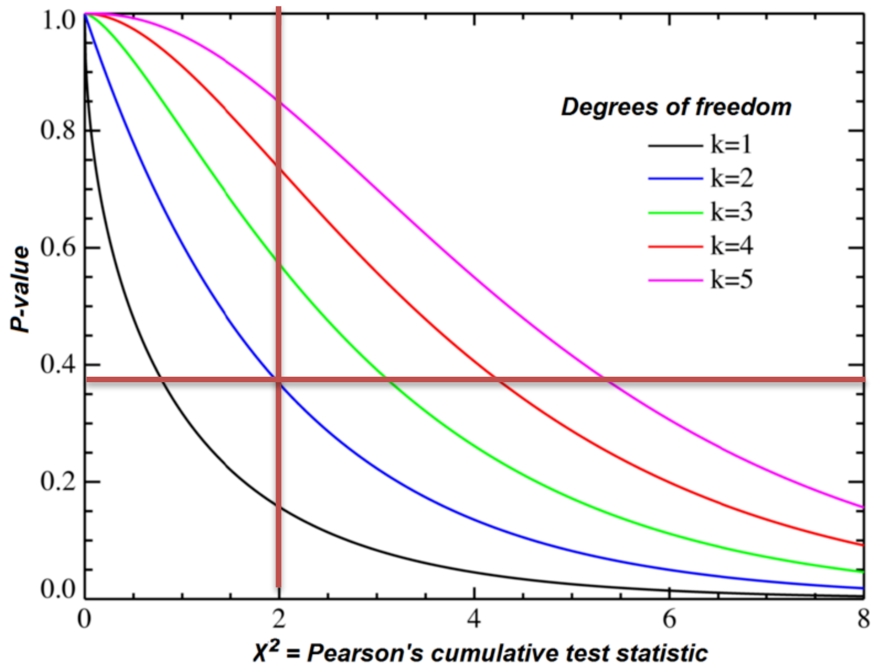
O_i = an observed frequency;

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

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

$$\chi^2 = \underbrace{\frac{(45-50)^2}{50} + \frac{(55-50)^2}{50}}_{\text{success}} + \underbrace{\frac{(55-50)^2}{50} + \frac{(45-50)^2}{50}}_{\text{failure}} = \frac{4 \cdot 25}{50} = 2$$

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

The 4W question session

What is this it all about ?

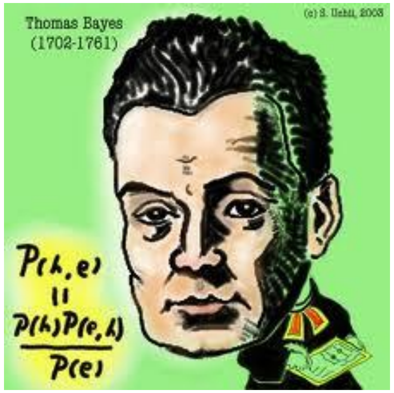
What can I use it for ?

Why should I learn it ?

What are potential Bachelor and Master thesis topics ?

The Brain is Bayesian

4W



Review

TRENDS in Cognitive Sciences Vol.10 No.7 July 2006

Full text provided by www.sciencedirect.com

ELSEVIER

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

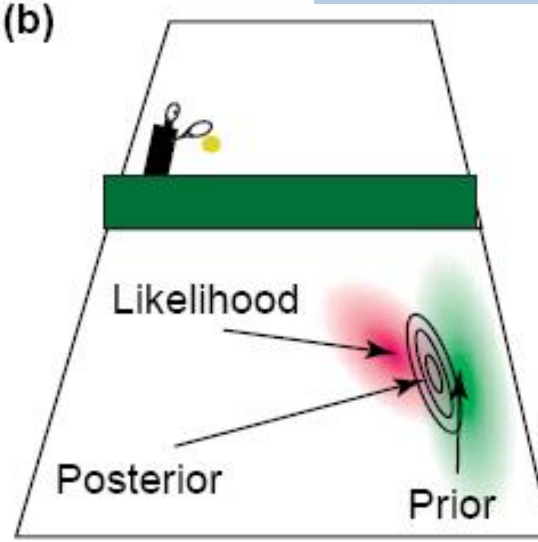
4W

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:

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

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

(b)



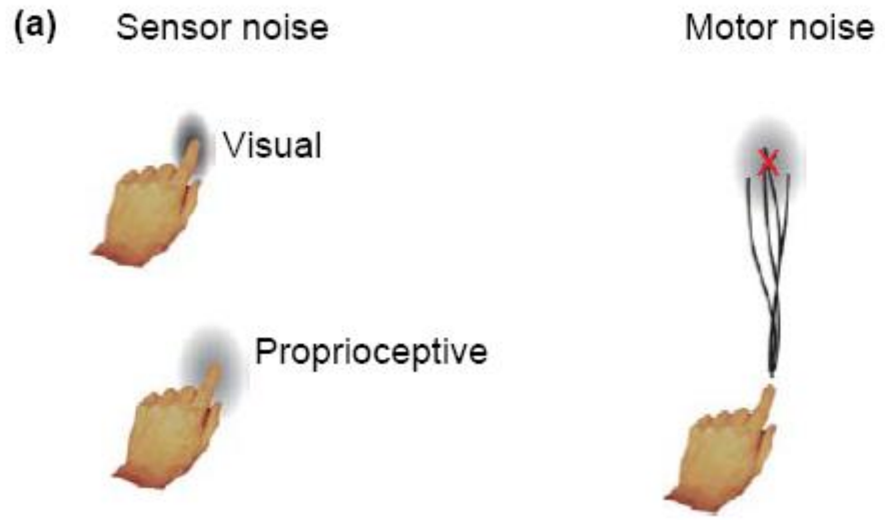
Outcome defined by:

- 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



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.

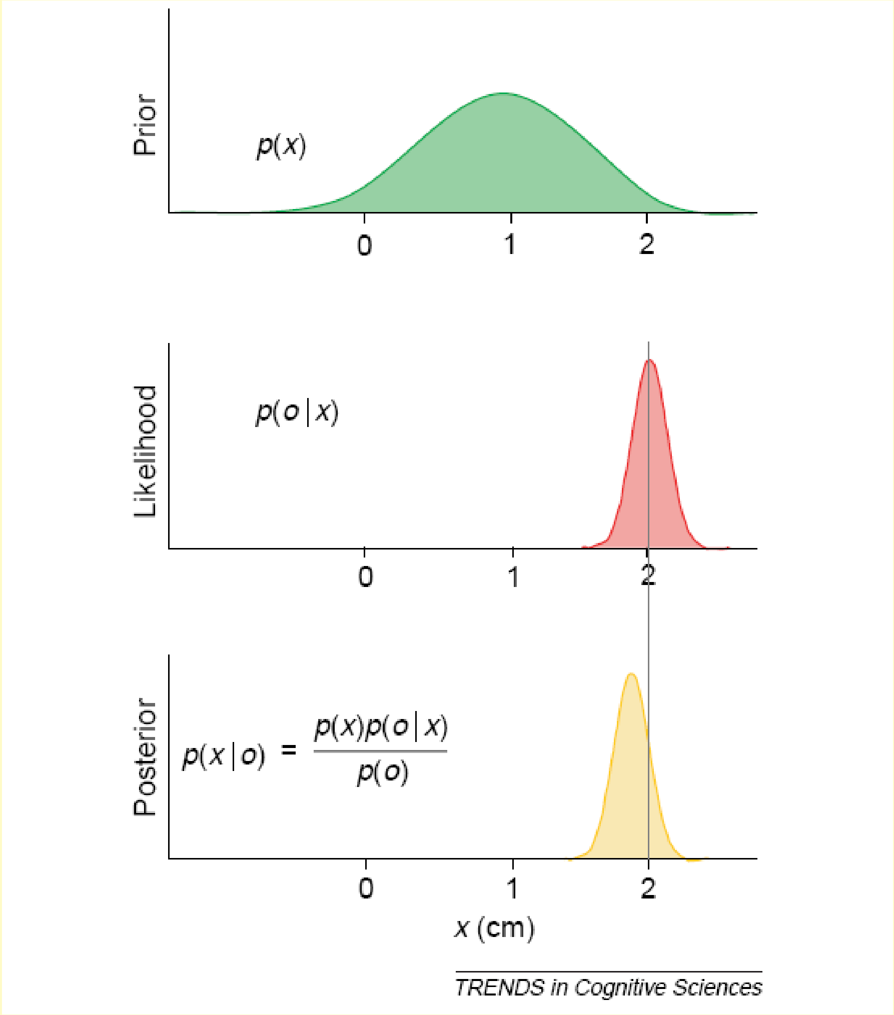


Figure I. Bayesian integration. The green curve represents the prior and red curve represents the likelihood. The yellow curve represents the posterior, the result from combining prior and likelihood.

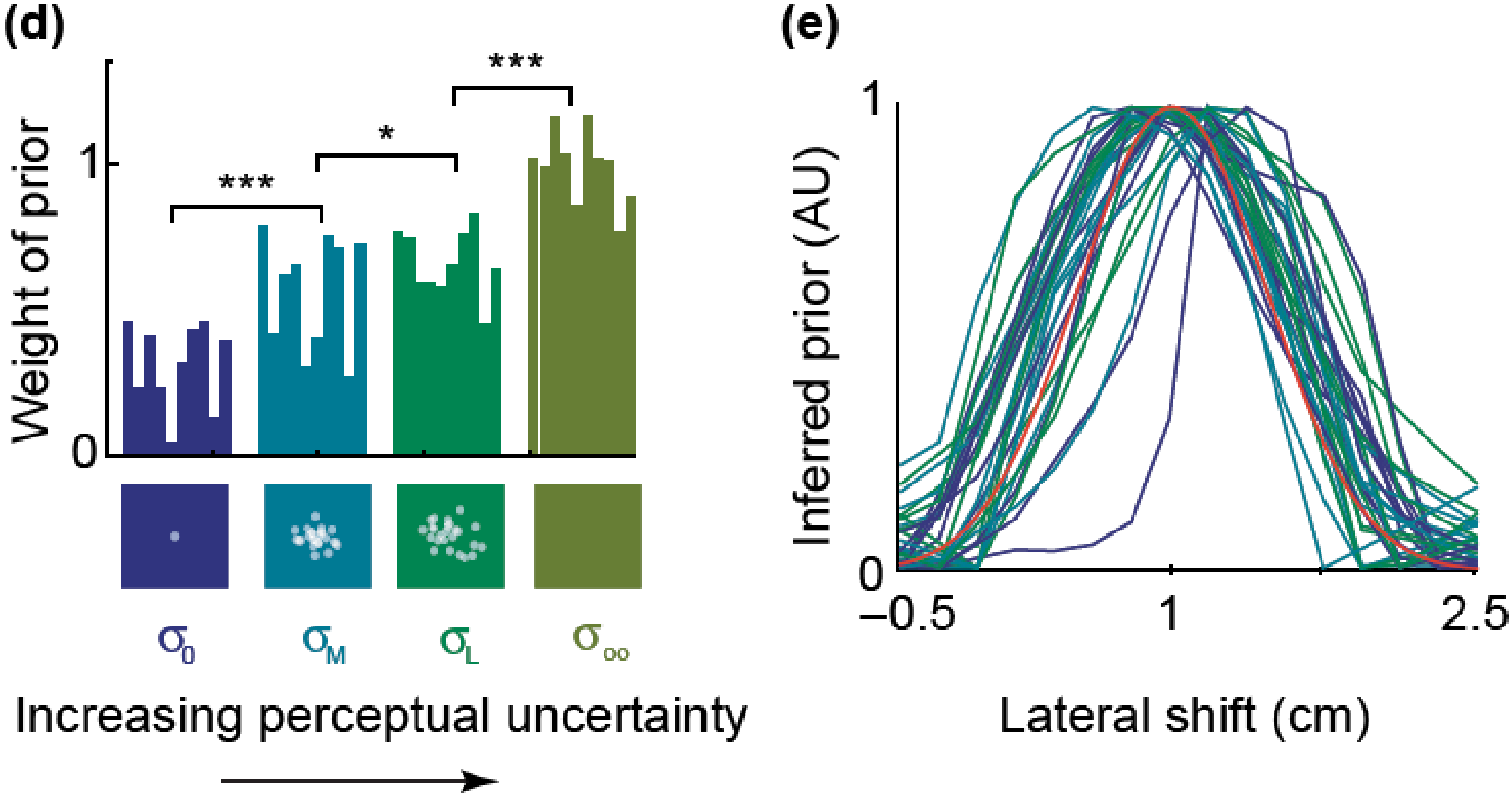
Is the brain really Bayesian ?

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?



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