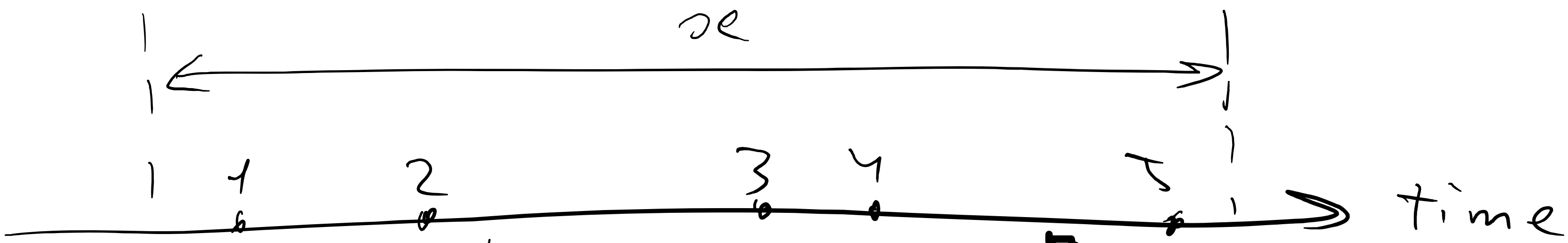
Next two weeks: projects

- Short (20-30 minute) videoclips of lectures will be posted on the website (the last column). Today's material is here: Erlang and Gamma
- For each of the next 4 lectures, there will be a project assignment to be solved in class
- I will not be present in class, but I recommend that you meet in person two times a week during regular hours
- All 4 projects will be submitted as one document from each of the 4 groups
- The projects are worth 20% of the grade, the midterm – 30%, the final exam – 50%

Constant rate (Poisson) process



discrete events happen at rate Γ

Expected number of events in time x is Γx

The actual number of events N_x is a Poisson distributed discrete random variable

$$P(N_x = n) = \frac{(\Gamma x)^n}{n!} e^{-\Gamma x}$$

Why Poisson? Divide x into many tiny intervals of length Δx

$p = \Gamma \Delta x$

$L = x / \Delta x$

$$\mathrm{Prob}(N = n) = \binom{L}{n} p^n (1-p)^{L-n}$$

$p \sim \Delta x \to 0$, $L \sim 1/\Delta x \to \infty$

↓

Poisson

$E(N_x) = pL = \Gamma x$

Constant rate (AKA Poisson) processes

- Let's assume that proteins are produced by ribosomes in the cell at a rate r per second.
- The expected number of proteins produced in x seconds is $r \cdot x$.
- The actual number of proteins N_x is a discrete random variable following a Poisson distribution with mean $r \cdot x$:

$$P_N(N_x = n) = \exp(-r \cdot x)(r \cdot x)^n / n! \qquad E(N_x) = rx$$

- Why Discrete Poisson Distribution?
  - Divide time into many tiny intervals of length $\Delta x << 1/r$
  - The probability of success (protein production) per internal is small: $p\_success = r\Delta x << 1$,
  - The number of intervals is large: $n = x/\Delta x >> 1$
  - Mean is constant: $r = E(N_x) = p\_success \cdot n = r\Delta x \cdot x/\Delta x = r \cdot x$
  - In the limit $\Delta x << x$, p_success is small and n is large, thus Binomial distribution → Poisson distribution

Exponential Distribution Definition

Exponential random variable X describes interval between two successes of a constant rate (Poisson) random process with success rate r per unit interval.

The probability density function of X is:

$$f(x) = re^{-rx} \quad \text{for} \quad 0 \leq x < \infty$$

Closely related to the discrete geometric distribution

$$f(x) = p(1\text{-}p)^{x-1} = \text{p } e^{(x-1)\ln(1-p)} \approx pe^{-px} \quad \text{for small p}$$

To summarize constant rate processes:

r – rate per unit of length (time) 1D

$N(x)$ – disrete number of events in time x

Poisson: $$P(N(x)=n) = \frac{(r \cdot x)^n}{n!} e^{-r \cdot x}$$

Time interval X between successive events is a continuously distributed random variable

Its PDF if $$f(x) = e^{-rx}$$

What is the interval X between two successes of a constant rate process?

- X is a continuous random variable
- CCDF: $P_X(X>x) = P_N(N_X=0)=exp(-r\cdot x)$.
  - Remember: $P_N(N_X=n)=exp(-r\cdot x)\ (r\cdot x)^n/n!$
- PDF: $f_X(x)=-dCCDF_X(x)/dx = r\cdot exp(-r\cdot x)$
- We started with a discrete Poisson distribution where time x was a parameter
- We ended up with a continuous exponential distribution

Exponential Mean & Variance

If the random variable X has an exponential distribution with rate r,

$$\mu = E(X) = \frac{1}{r} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{r^2} \quad (4-15)$$

Note that, for the:

- Poisson distribution: mean= variance
- Exponential distribution: mean = standard deviation = variance$^{0.5}$

Biochemical Reaction Time

- The time x (in minutes) until all enzymes in a cell catalyze a biochemical reaction and generate a product is approximated by this CCDF:

$$F_>(x) = e^{-2x} \text{ for } 0 \leq x$$

Here the rate of this process is r=2 min^{-1} and 1/r=0.5 min is the average time between successive products of these enzymes

- What is the PDF?

$$f(x) = -\frac{dF_>(x)}{dx} = -\frac{d}{dx}e^{-2x} = 2e^{-2x} \text{ for } 0 \leq \mathrm{x}$$

- What proportion of reactions will not generate another product within 0.5 minutes of the previous product?

$$P(X > 0.5) = F_>(0.5) = e^{-2*0.5} = 0.37$$

We observed our cell for 1 minute and no product has been generated: The product is "overdue"

What is the probability that a product will not appear during the next 0.5 minutes?

$$F_{>}(x) = e^{-2x}$$
$$F_{>}(0.5) \approx 0.37$$
$$F_{>}(1.5) \approx 0.05$$
$$F_{>}(1.0) \approx 0.13$$

A. 0.32
B. 0.37
C. 0.08
D. 0.24
E. I have no idea

Get your i-clickers

Memoryless property of the exponential distribution

$$P(X>t+s \mid X>s) = P(X>t)$$

$$P(X>t+s \mid X>s) = \frac{P(X>t+s,\ X>s)}{P(X>s)} =$$

$$= \frac{\exp(-\lambda(t+s))}{\exp(-\lambda s)} = \exp(-\lambda t) =$$

$$= P(X>t)$$

Exponential is <u>the only</u> memoryless distribution

Matlab exercise:

- Generate a sample of 100,000 variables from Exponential distribution with r =0.1
- Calculate mean and compare it to 1/r
- Calculate standard deviation and compare it to 1/r
- Plot semilog-y plots of PDFs **and CCDFs.**
- Hint: read the help page (better yet documentation webpage) for random('Exponential'...) one of their parameters is different than r

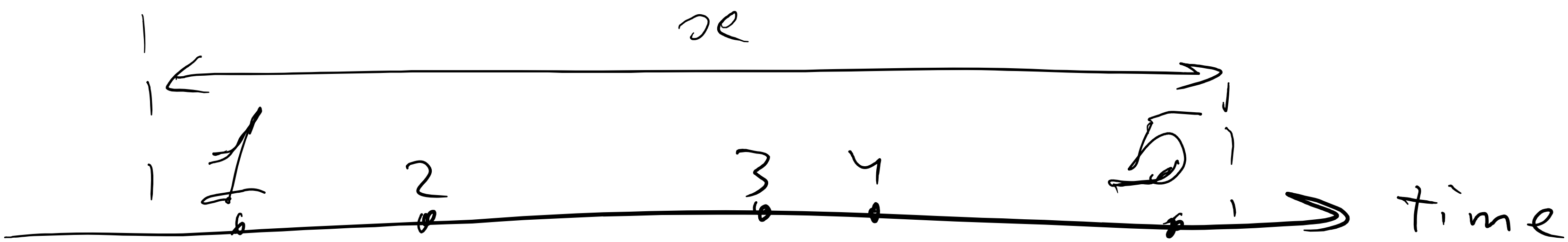
Matlab exercise: Exponential

- **Stats=100000; r=0.1;**
- **r2=random('Exponential', 1./r, Stats,1);**
- **disp([mean(r2),1./r]); disp([std(r2),1./r]);**
- **step=1; [a,b]=hist(r2,0:step:max(r2));**
- **pdf_e=a./sum(a)./step;**
- **subplot(1,2,1); semilogy(b,pdf_e,'rd-’);**
- **x=0:0.01:max(r2);**
- **for m=1:length(x);**
- **ccdf_e(m)=sum(r2>x(m))./Stats;**
- **end;**
- **subplot(1,2,2); semilogy(x,ccdf_e,'ko-');**

Erlang Distribution

- The Erlang distribution is a generalization of the exponential distribution.
- The exponential distribution models the time interval to the 1st event, while the
- Erlang distribution models the time interval to the k^{th} event, i.e., a sum of k exponentially distributed variables.
- The exponential, as well as Erlang distributions, is based on the constant rate (or Poisson) process.

Constant rate (Poisson) process



Events happen <u>independently</u> from each other at <u>constant rate = r</u> : $E[N_x] = rx$

X follows Erlang distribution

$$P(X > x) = \sum_{n=0}^{r-1} P(N_x = n) =$$

$$= \sum \frac{(rx)^n}{n!} e^{-rx}$$

Erlang Distribution

Generalizes the Exponential Distribution:

waiting time until k's events

(constant rate process with rate=r)

$$P(X > x) = \sum_{m=0}^{k-1} \frac{e^{-rx}(rx)^m}{m!} = 1 - F(x)$$

Differentiating $F(x)$ we find that all terms in the sum except the last one cancel each other:

$$f(x) = \frac{r^k x^{k-1} e^{-rx}}{(k-1)!} \quad \text{for } x > 0 \ \text{ and } \ \mathrm{k} = 1, 2, 3, \ldots$$

Gamma Distribution

The random variable X with a probability density function:

$$f(x) = \frac{r^k x^{k-1} e^{-rx}}{\Gamma(k)}, \text{ for } x > 0 \qquad (4\text{-}18)$$

has a gamma random distribution with parameters r > 0 and $k > 0$. If k is a positive integer, then X has an Erlang distribution.

$$f(x) = \frac{r^k x^{k-1} e^{-rx}}{\Gamma(k)}, \text{ for } x > 0$$

$$\int_0^{+\infty} f(x)\,dx = 1, \quad \text{Hence}$$

$$\Gamma(k) = \int_0^{+\infty} r^k x^{k-1} e^{-rx}\,dx = \int_0^{+\infty} y^{k-1} e^{-y}\,dy$$

Comparing with Erlang distribution

for integer k one gets

$$\Gamma(k) = (k-1)!$$

Gamma Function

The gamma function is the generalization of the factorial function for r > 0, not just non-negative integers.

$$\Gamma(k) = \int_0^\infty y^{k-1} e^{-y}\, dy, \quad \text{for } r > 0 \qquad (4\text{-}17)$$

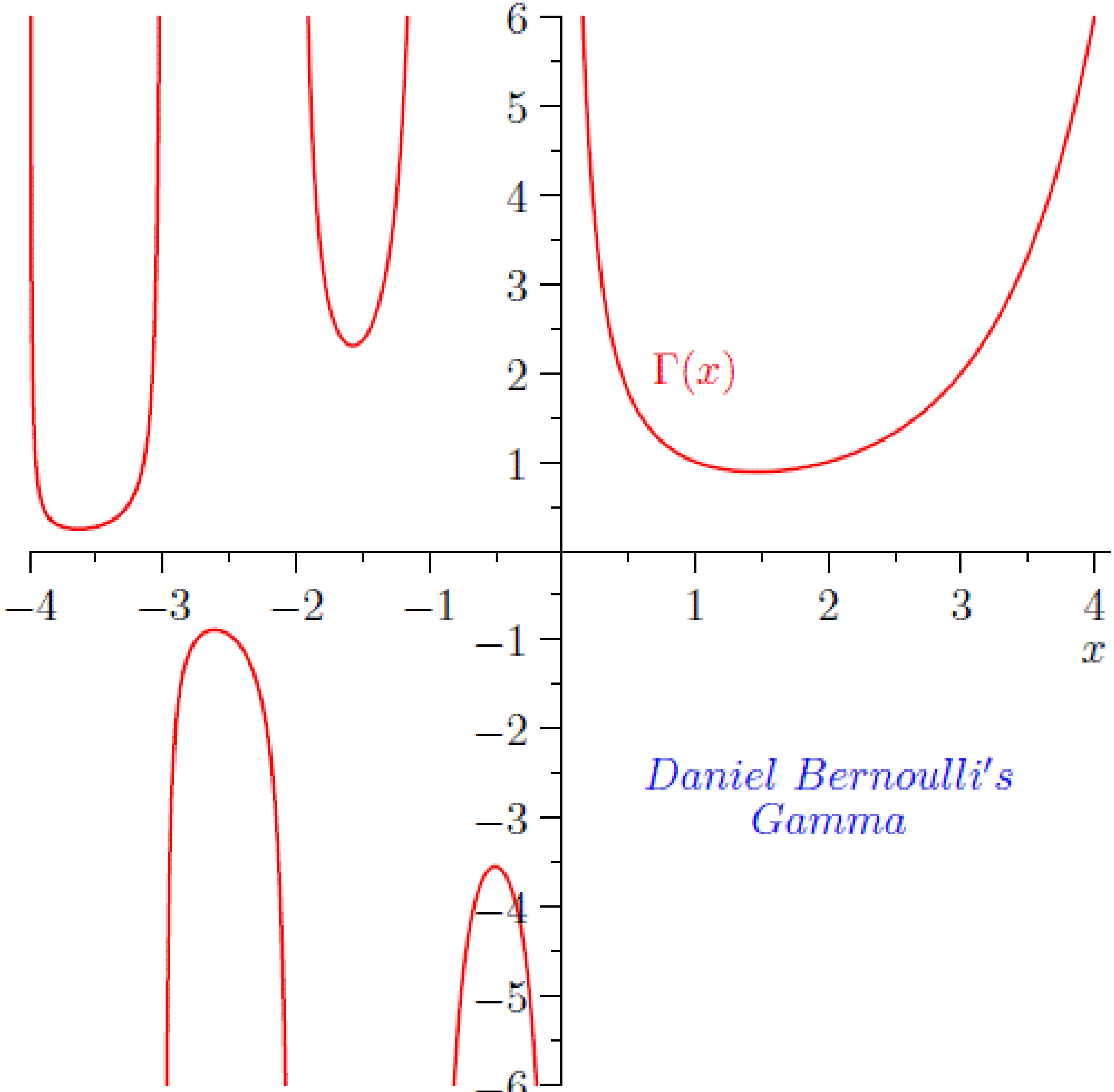
Properties of the gamma function

$$\Gamma(1) = 1$$

$$\Gamma(k) = (k-1)\Gamma(k-1) \quad \text{recursive property}$$

$$\Gamma(k) = (k-1)! \quad \text{factorial function}$$

$$\Gamma(1/2) = \pi^{1/2} = 1.77 \quad \text{interesting fact}$$



Γ(x)
x
Daniel Bernoulli's Gamma

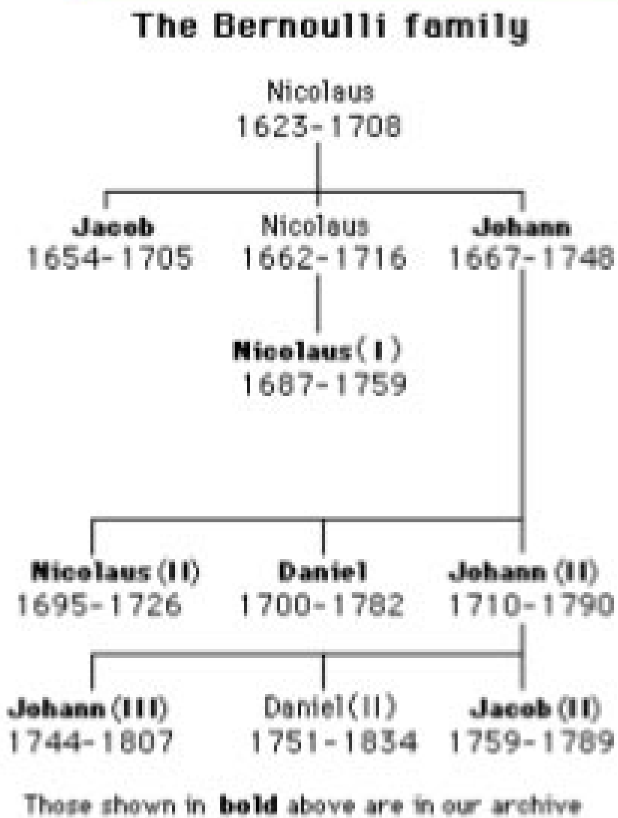
SOLO

Bernoulli trials

BERNOULLI FAMILY

SOLO HERMELIN

http://www.solohermelin.com



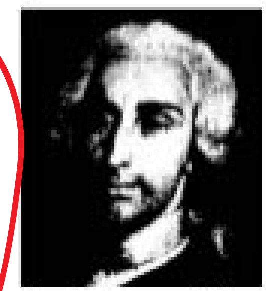
Run This

Jacob 1654-1705

Johann 1667-1748

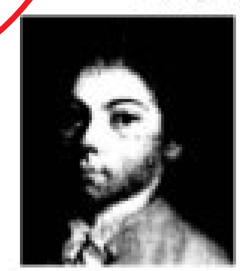
Nicolaus II 1695-1720

Daniel 1700-1782



Johann II 1710-1790

Johann III 1744-1807



Jacob II 1759-1789

Gamma function

Mean & Variance of the Erlang and Gamma

- If X is an Erlang (or more generally Gamma) random variable with parameters r and k,

  $\mu = E(X) = k/\text{r}$ and $\sigma^2 = V(X) = \text{k}/\text{r}^2$ (4-19)

- Generalization of exponential results:

  $\mu = E(X) = 1/\text{r}$ and $\sigma^2 = V(X) = 1/\text{r}^2$ or

  Negative binomial results:

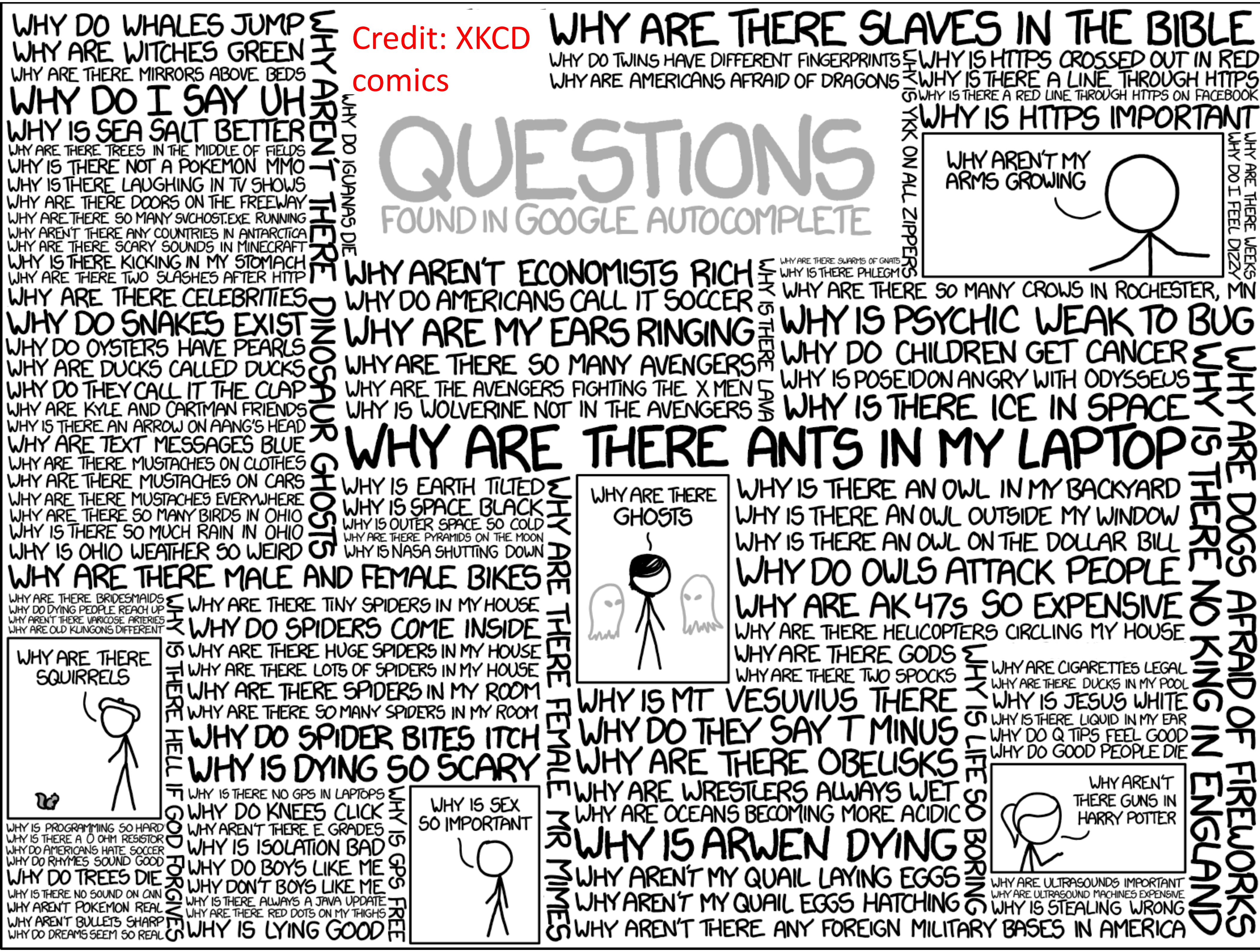
  $\mu = E(X) = k/p$ and $\sigma^2 = V(X) = k(1-p) / p^2$

Matlab exercise:

- Generate a sample of 100,000 variables with “Harry Potter” Gamma distribution with r =0.1 and k=9 ¾ (9.75)
- Calculate mean and compare it to k/r (Gamma)
- Calculate standard deviation and compare it to sqrt(k)/r (Gamma)
- Plot semilog-y plots of PDFs **and CCDFs.**
- Hint: read the help page (better yet documentation webpage) for random(‘Gamma’…): one of their parameters is different than r

Matlab exercise: Gamma

- **Stats=100000; r=0.1; k=9.75;**
- **r2=random('Gamma', k,1./r, Stats,1);**
- **disp([mean(r2),k./r]);**
- **disp([std(r2),sqrt(k)./r]);**
- **step=0.1; [a,b]=hist(r2,0:step:max(r2));**
- **pdf_g=a./sum(a)./step;**
- **figure;**
- **subplot(1,2,1); semilogy(b,pdf_g,'ko-'); hold on;**
- **x=0:0.01:max(r2); clear cdf_g;**
- **for m=1:length(x);**
-     **cdf_g(m)=sum(r2>x(m))./Stats;**
- **end;**
- **subplot(1,2,2); semilogy(x,cdf_g,'rd-');**



Credit: XKCD comics