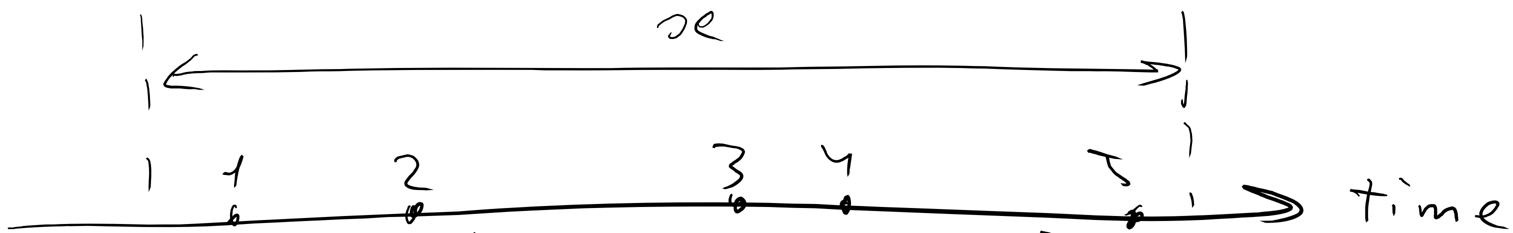
Continuous Probability Distributions

Exponential, Erlang, Gamma

Poisson process



discrete events happen at rate λ

Expected number of events in time x is λx

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

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

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

$p = \lambda \Delta x$

$L = x / \Delta x$

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

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

$\downarrow$ $\quad p \sim \Delta x \to 0, \; L \sim \frac{1}{\Delta x} \to \infty$

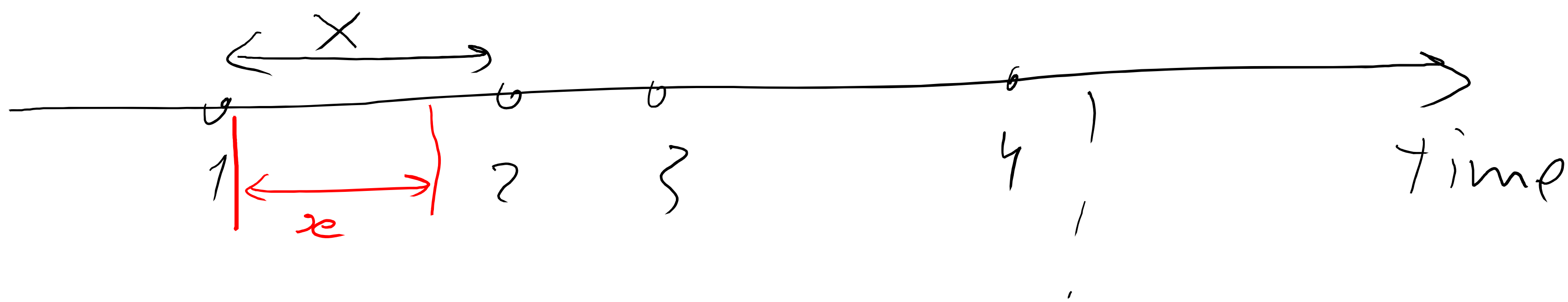
Poisson

Poisson (constant rate) processes

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

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

- Why Discrete Poisson Distribution?
  - Divide time into many tiny intervals of length $x_0 << 1/\lambda$
  - The probability of success (protein production) per internal is small: $p=\lambda\cdot\Delta x << 1$,
  - The number of intervals is large: $L=x/\Delta x >> 1$
  - Mean is constant: $E(N_x)=p\cdot L=(\lambda\Delta x)\cdot(x/\Delta x)=\lambda\cdot x$
  - $P(N_x=n)=L!/n!(L-n)!\ p^n\ (1-p)^{L-n}$
  - In the limit $p\rightarrow 0$, $L\rightarrow\infty$: Binomial distribution → Poisson



$$\text{CCDF}: \text{Prob}(X > x) = \text{Prob}(N_x = 0) = \frac{(\lambda x)^0}{0!} e^{-\lambda x} = e^{-\lambda x}$$

$$\text{PDF} = -\frac{d}{dx}\text{CDF}: \quad \boxed{f(x) = \lambda e^{-\lambda x}}$$

What is the distribution of the interval X between CONSEQUITIVE EVENTS of a constant rate process?

- X is a continuous random variable
- CCDF: $Prob(X>x) = Prob(N_X=0)=exp(-\lambda x)$.
  - Remember: $P_N(N_X=n)=exp(-\lambda x)\ (\lambda x)^n/n!$
- PDF: $f(x)=-d\ CCDF(x)/dx = \lambda exp(-\lambda x)$
- We started with a discrete Poisson distribution where time x was a parameter and N_x – discrete random variable
- We ended up with a continuous exponential distribution where time X between events was a continuous random variable

Exponential Mean & Variance

If the random variable X has an exponential distribution with parameter λ,

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

Note that, for the:

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

Exponential Distribution is a continuous generalization of what discrete probability distribution?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers

Biochemical Reaction Time

- The time x (in minutes) until an enzyme successfully catalyzes a biochemical reaction is approximated by this CDF:

$$F\left(x\right)=1-e^{-x/1.4}\ \text{ for } 0\le x$$

- What is the PDF?

$$f\left(x\right)=\frac{dF\left(x\right)}{dx}=\frac{d}{dx}[1-e^{-x/1.4}]=e^{-x/1.4}/1.4\ \text{ for } 0\le \text{x}$$

- What proportion of reactions is complete within 0.5 minutes?

$$P\left(X<0.5\right)=F\left(0.5\right)=1-e^{-0.5/1.4}=1-0.7=0.3$$

The reaction product is "overdue": no product has been generated in the past 3 minutes.
What is the probability that a product will appear in the next 0.5 minutes?

$$F(x) = 1 - e^{-x/1.4}$$

$$F(0.5) \approx 0.3$$

$$F(3.5) \approx 0.92$$

A. 0.92
B. 0.3
C. 0.62
D. 0.99
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 __the only__ memoryless distribution

Can other random variables be memoryless?

No!

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

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

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

for any t & s

Let $t = \Delta s$ - very small; $F(s) = P(X>s)$

$$F(s+\Delta s) = F(\Delta s)\cdot F(s)$$

$$F(0)=1 \Rightarrow F(\Delta s) \approx 1 - \lambda \Delta s$$

$$F(s+\Delta s) = (1-\lambda\Delta s)\,F(s)$$

$$\frac{F(s+\Delta s)-F(s)}{\Delta s} = -\lambda F(s)$$

$$\frac{dF(s)}{ds} = -\lambda F(s)$$

$$F(s) = \exp(-\lambda s)$$

$$PDF(s) = -\frac{dF}{ds} = \lambda \exp(-\lambda s)$$

Thus, any ~~memoryless~~ continuos memoryless r.v. is exponential

Discrete — geometric

Exponential Distribution in Reliability

- The reliability of electronic components is often modeled by the exponential distribution. A chip might have mean time to failure of 40,000 operating hours.

- The memoryless property implies that the component does not wear out – the probability of failure in the next hour is constant, regardless of the component age.

- The reliability of mechanical components do have a memory – the probability of failure in the next hour increases as the component ages.

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 r^{th} event, i.e., a sum of r exponentially distributed variables.
- The exponential, as well as Erlang distributions, is based on the constant rate Poisson process.

Erlang Distribution

Generalizing from the constant rate
Poisson → Exponential :

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

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

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

Example 4-23: Medical Device Failure

The failures of medical devices can be modeled as a Poisson process. Assume that units that fail are repaired immediately and the mean number of failures per hour is 0.0001. Let *X* denote the time until 4 failures occur. What is the probability that *X* exceed 40,000 hours ~=4.5 years?

Let the random variable *N* denote the number of failures in 40,000 hours. The time until 4 failures occur exceeds 40,000 hours *iff* the number of failures in 40,000 hours is ≤ 3.

$$P\left(X > 40,000\right) = P\left(N \leq 3\right)$$

$$E\left(N\right) = 40,000\left(0.0001\right) = 4 \text{ failures in 40,000 hours}$$

$$P\left(N \leq 3\right) = \sum_{k=0}^{3} \frac{e^{-4} 4^k}{k!} = 0.433$$

Erlang Distribution is a continuous generalization of what discrete probability distribution?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Get your i-clickers

Erlang distribution

$$f(x) = \frac{\lambda^r x^{r-1} \exp(-\lambda x)}{(r-1)!}$$

Can be generalized for any $r > 0$

Q: What to use instead of $(r-1)!$?

Gamma Function

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

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

Properties of the gamma function

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

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

$$\Gamma(1) = 0! = 1$$

$$\Gamma(1/2) = \pi^{1/2} = 1.77$$

$$\Gamma(3/2) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = 0.89$$

interesting facts

$$\left(\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}$$

Gamma Distribution

The random variable *X* with a probability density function:

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

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

Gamma Distribution Graphs

- The r and λ parameters are often called the "shape" and "scale"
- Different parameter combinations change the distribution.
- The distribution becomes progressively more symmetric as *r* increases.
- Matlab uses 1/ λ as a "scale" parameter.

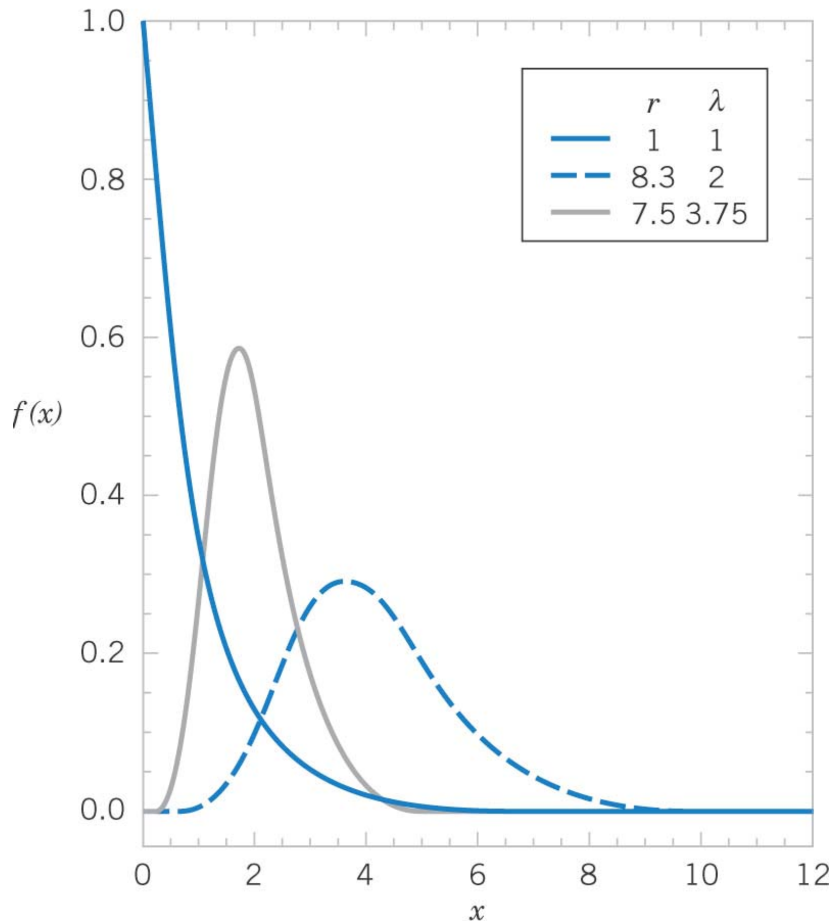


Figure 4-25 Gamma probability density functions for selected values of λ and *r*.

Mean & Variance of the Erlang and Gamma

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

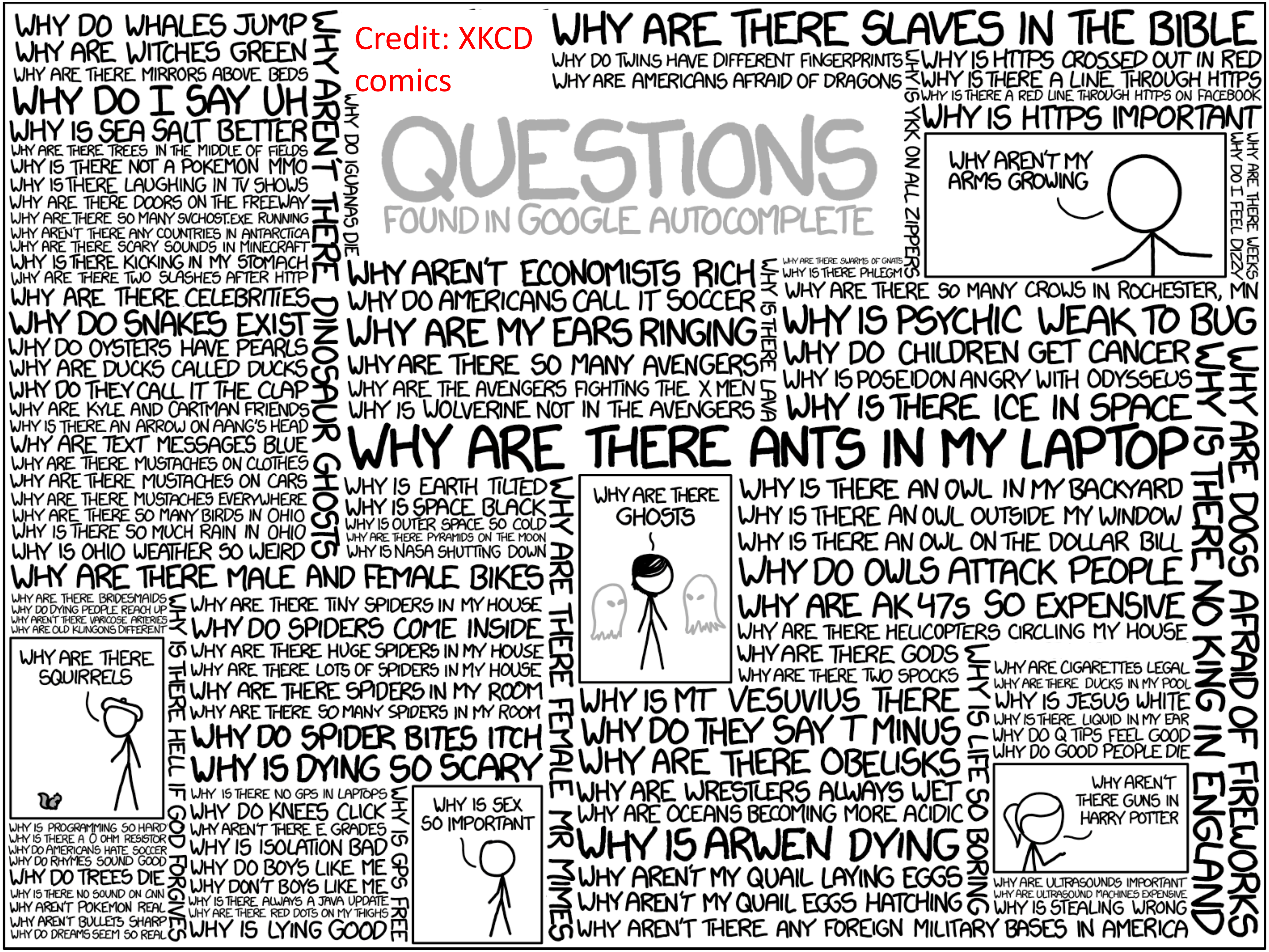
  $\mu = E(X) = r / \lambda$ and $\sigma^2 = V(X) = r / \lambda^2$ (4-19)

- Generalization of exponential results:
  $\mu = E(X) = 1 / \lambda$ and $\sigma^2 = V(X) = 1 / \lambda^2$ or
  Negative binomial results:
  $\mu = E(X) = r / p$ and $\sigma^2 = V(X) = r(1-p) / p^2$

Matlab exercise:

- Generate a sample of 100,000 random numbers drawn from an exponential distribution with rate lambda=0.1.
  Hint: read the help page for random('Exponential'…)
- Calculate mean and standard deviation of the sample and compare to predictions 1/lambda
- Generate PDF and CCDF of the sample and plot them both on a semilogarithmic scale (y-axis)
- After done with exponential modify for Gamma distribution with lambda=0.1, r=4.5

- Stats=??; lambda=??;
- r2=random('Exponential', ??, Stats,1);
- disp([mean(r2),??]);
- disp([std(r2),??]);
- %%
- step=0.1; [a,b]=hist(r2,0:step:max(r2));
- pdf_e=a./sum(a).?? step;
- figure; subplot(1,2,1); semilogy(b,pdf_e,'ko-');
- %%
- X=0:0.01:100;
- for m=1:length(X);
-     ccdf_e(m)=sum(r2 ?? X(m))./Stats;
- end;
- subplot(1,2,2); semilogy(X,ccdf_e,'ko-');



Credit: XKCD comics