Probability Density Function (PDF)

Density functions, in contrast to mass functions, distribute probability continuously along an interval

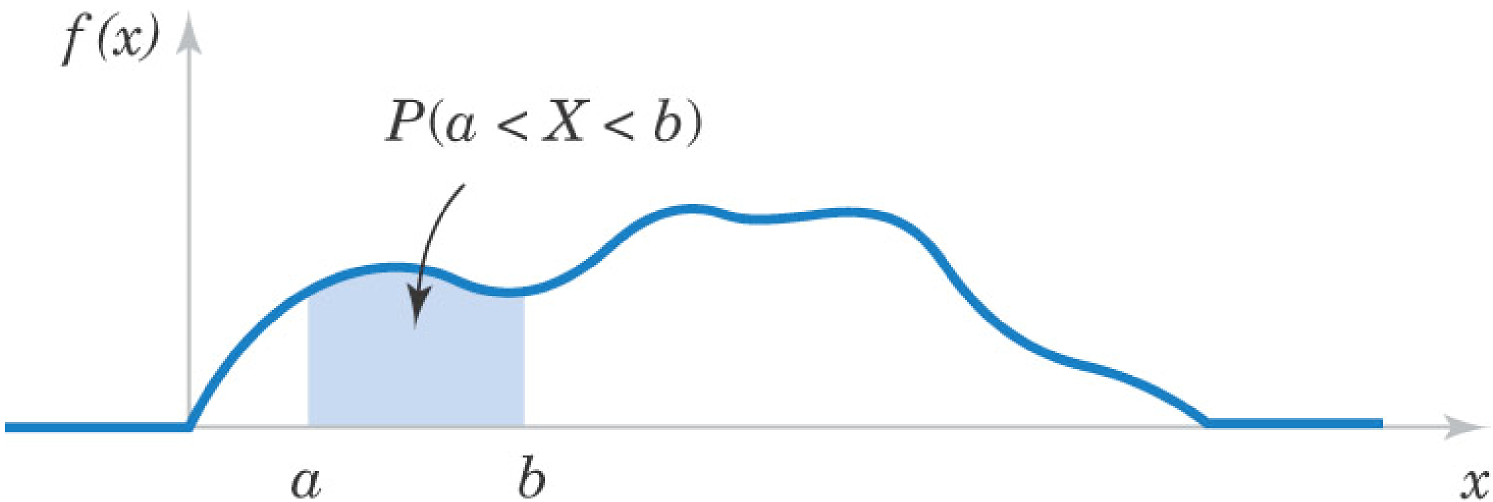


Figure 4-2 Probability is determined from the area under *f*(*x*) from a to b.

X is a continuous random variable with a uniform distribution between 0 and 5.

What is Probability(X=2)?

A. 1/6
B. 1/5
C. 0
D. Infinity
E. I have no idea

Get your i-clickers

X is a continuous random variable with a uniform distribution between 0 and 5.

What is Probability(X=2)?

A. 1/6
B. 1/5
C. 0
D. Infinity
E. I have no idea

Get your i-clickers

X is a continuous random variable with a uniform distribution between 0 and 5.

What is Probability(X<=2)?

A. 3/6
B. 3/5
C. 0
D. 2/5
E. I have no idea

Get your i-clickers

X is a continuous random variable with a uniform distribution between 0 and 5.

What is Probability(X<=2)?

A. 3/6
B. 3/5
C. 0
D. 2/5
E. I have no idea

Get your i-clickers

X is a **discrete** random variable with a uniform distribution between 0 and 5.

What is Probability(X<=2)?

A. 3/6
B. 3/5
C. 0
D. 2/5
E. I have no idea

Get your i-clickers

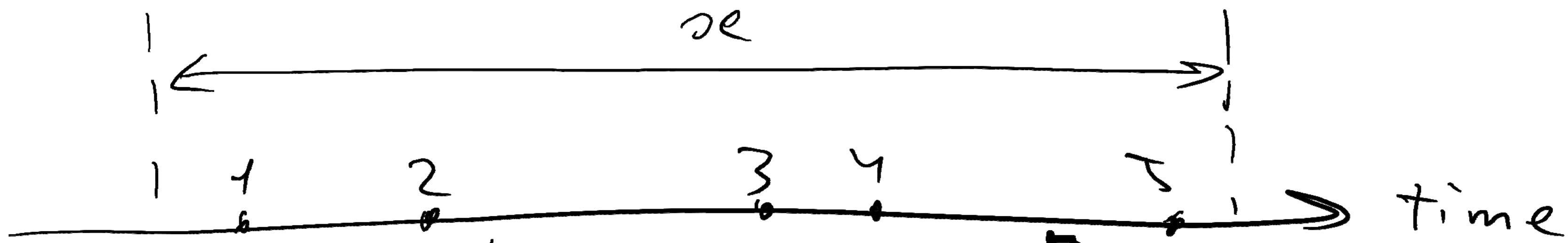
X is a continuous random variable with a uniform distribution between 0 and 5.

What is Probability(X<=2)?

A. 3/6
B. 3/5
C. 0
D. 2/5
E. I have no idea

Get your i-clickers

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

$p = \Gamma \Delta x$

$L = x/\Delta x$

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

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

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

Poisson

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 $\Delta x \ll 1/r$
  - The probability of success (protein production) per internal is small: $p\_success=r\Delta x \ll 1$,
  - The number of intervals is large: $n= x/\Delta x \gg 1$
  - Mean is constant: $r=E(N_x)=p\_success \cdot n= r\Delta x \cdot x/\Delta x = r \cdot x$
  - In the limit $\Delta x \ll x$, 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:

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

Closely related to the discrete geometric distribution

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

To summarize constant rate processes:

r – rate per unit of time / length 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,

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

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?

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

- 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"

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

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

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

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

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

$$= \frac{\exp(-r(t+s))}{\exp(-rs)} = \exp(-rt) =$$

$$= P(X>t)$$

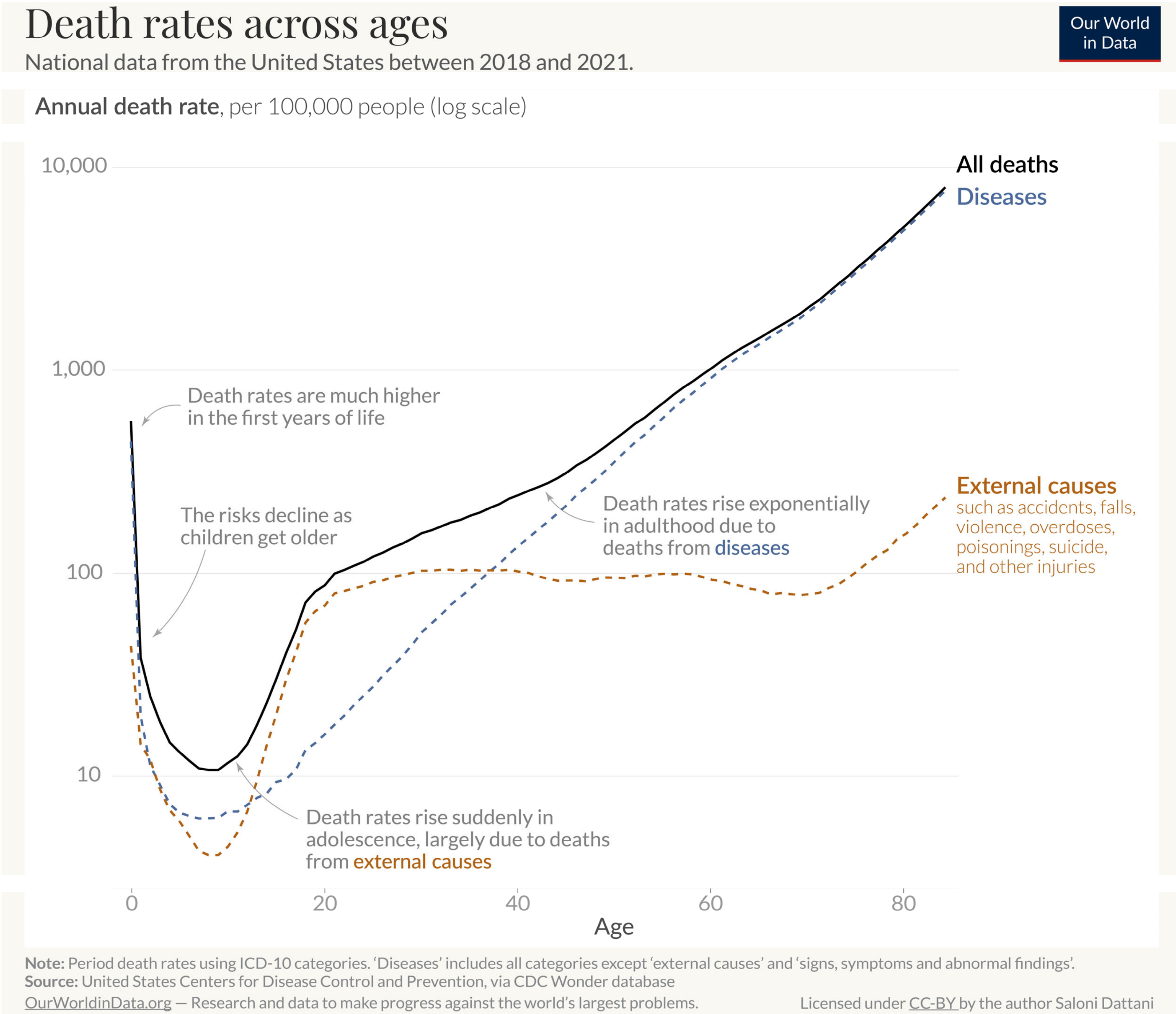
Exponential is <u>the only</u> memoryless distribution

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-');**

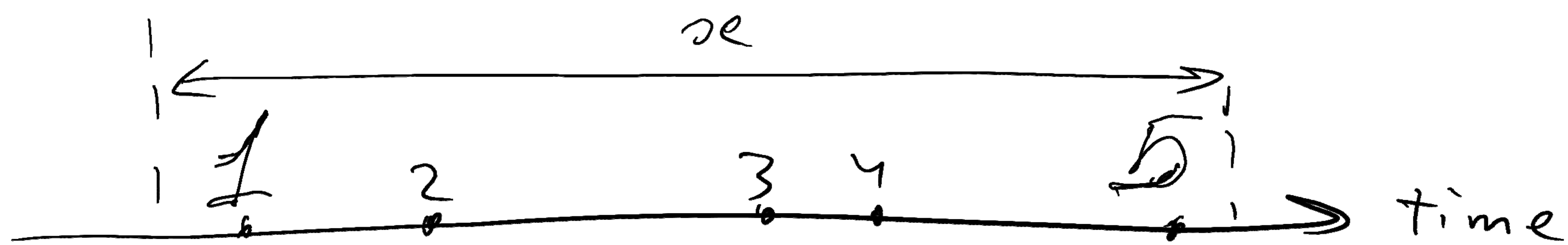


https://ourworldindata.org/how-do-the-risks-of-death-change-as-people-age

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 independently from each other at constant rate $= r$: $E[N_x] = rx$

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:

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

Gamma Distribution

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

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

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.

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

Properties of the gamma function

$$\Gamma(1) = 1$$

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

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

$$\Gamma\left(\frac{1}{2}\right) = \pi^{\frac{1}{2}} = 1.77 = \left(-\frac{1}{2}\right)! \qquad \text{interesting fact}$$