Constant rate (Poisson) process

Discrete events happen at rate $\Gamma$.
Expected number of events in time $x$ is $\Gamma x$.

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

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

Why Poisson? Divide $x$ into many tiny intervals of length $\Delta x$.

$$p = \frac{\Gamma x}{\Delta x}$$
$$l = \frac{x}{\Delta x}$$

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

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

As $p \sim \Delta x \to 0$, $L \sim \frac{l}{\Delta x} \to \infty$.
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 \( r \cdot x \):
  \[
P_N(N_x=n)=\exp(-r \cdot x)(r \cdot x)^n/n! \quad \text{E}(N_x)= r x
  \]
- 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: \( p_{\text{success}}=r\Delta x << 1 \),
  - The number of intervals is large: \( n= x/\Delta x >> 1 \)
  - Mean is constant: \( r=E(N_x)=p_{\text{success}} \cdot n= r\Delta x \cdot x/\Delta x = r \cdot x \)
  - In the limit \( \Delta x << x \), \( p_{\text{success}} \) is small and \( n \) is large, thus Binomial distribution \( \rightarrow \) 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:

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

Closely related to the discrete geometric distribution

$$f(x) = p(1-p)^{x-1} = p e^{(x-1) \ln(1-p)} \approx pe^{-px} \text{ for small } p$$
What is the interval $X$ between two successes of a constant rate process?

- $X$ is a continuous random variable
- CCDF: $P_X(X>x) = P_{\mathcal{N}}(N_X=0) = \exp(-r \cdot x)$.  
  - Remember: $P_{\mathcal{N}}(N_X=n) = \exp(-r \cdot x) \frac{(r \cdot x)^n}{n!}$
- 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 $r$,

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

Note that, for the:

• Poisson distribution: \text{mean}= \text{variance}
• Exponential distribution: \text{mean} = \text{standard deviation} = \text{variance}^{0.5}
Biochemical Reaction Time

- The time $x$ (in minutes) until an enzyme catalyzes a biochemical reaction and generates a product is approximated by this CCDF:
  
  $$F_>(x) = e^{-2x} \text{ for } 0 \leq x$$

  Here the rate of this process is $r=2 \text{ min}^{-1}$ and $1/r=0.5 \text{ min}$ is the average time between successive products of this enzyme.

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

- What proportion of reactions will not generate another product within 0.5 minutes of the previous product?
  
  $$P(X > 0.5) = F_>(0.5) = e^{-2 \times 0.5} = 0.37$$
We observed our enzyme 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?

\[ F_X(x) = e^{-2x} \]
\[ F_X(0.5) \approx 0.37 \]
\[ F_X(1.5) \approx 0.05 \]
\[ F_X(1.0) \approx 0.13 \]

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 the only memoryless distribution
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 $k^{\text{th}}$ 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 $\lambda$. \( E[N_x] = \lambda x \)

\( X \) follows Erlang distribution

\[ f(X > x) = \sum_{n=1}^{r-1} P(N_x = n) = \sum \frac{(rx)^n}{n!} e^{-rx} \]
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:

\[ f(x) = \frac{r^k x^{k-1} e^{-rx}}{(k-1)!} \quad \text{for } x > 0 \quad \text{and } k = 1, 2, 3, ... \]
Gamma Distribution

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

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

has a gamma random distribution with parameters $r > 0$ and $k > 0$. If $k$ is a positive integer, then $X$ has an Erlang distribution.
$f(x) = \frac{r^k x^{k-1} e^{-rx}}{\Gamma(k)}$, for $x > 0$

$$\int_{0}^{+\infty} f(x) \, dx = 1$$

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.

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

(4-17)

Properties of the gamma function

$$\Gamma(1) = 1$$

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

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

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

Daniel Bernoulli's Gamma
Mean & Variance of the Erlang and Gamma

• If $X$ is an Erlang (or more generally Gamma) random variable with parameters $r$ and $k$, 
  $\mu = E(X) = k/r$ and $\sigma^2 = V(X) = k/r^2$ \hspace{1cm} (4-19)

• Generalization of exponential results:
  $\mu = E(X) = 1/r$ and $\sigma^2 = V(X) = 1/r^2$ \hspace{1cm} 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 \textbf{Exponential distribution} with \( r = 0.1 \)
• Generate a sample of 100,000 variables with \textit{“Harry Potter” Gamma distribution} with \( r = 0.1 \) and \( k = 9 \frac{3}{4} (9.75) \)
• Calculate mean and compare it to \( \frac{1}{r} \) (Exp) and \( \frac{k}{r} \) (Gamma)
• Calculate standard deviation and compare it to \( \frac{1}{r} \) (Exp) and \( \frac{\sqrt{k}}{r} \) (Gamma)
• Plot semilog-y plots of PDFs \textbf{and CCDFs}.
• \textbf{Hint:} read the help page (better yet documentation webpage) for \texttt{random(‘Exponential’...)} and \texttt{random(‘Gamma’...)}: 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-');
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-');