Midterm (tentatively)  
Thursday 3/14/2019  
9:30am-10:50am  
Room: TBD  

Lectures next week will be given by Prof. Roy Dar
Grader Office Hours
(Location TBD: Everitt Lab/Grainger):

A. Mondays 3-4pm
B. Tuesdays 2-4pm
C. Wednesdays 2:30-4pm
D. Thursdays 1-4pm
E. I don’t plan to come to office hours
\[ E[N_x] = \int_0^x \]  

\[ \Delta x \]

\[ \text{start} \]

\[ \text{end} \]

\[ \lambda = r \Delta x \cdot \frac{n}{\Delta x} \]

\[ = r \cdot n \Delta x \]

\[ \text{Binomial} \]

\[ P(N_x = n) = \frac{(r \cdot x)^n e^{-r \cdot x}}{n!} \]

\[ \text{Poisson} \]
Poisson process AKA Constant rate process

rate = \( r \) \((1/\text{min})\)

\[ P(N = n) = \frac{(rx)^n}{n!} e^{-rx} \]

\[ \lambda = n \cdot p = \frac{nx}{\Delta x} \cdot r \Delta x = r \cdot x \]
Constant rate (AKA Poisson) processes

- Let’s assume that proteins are produced by all 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 $r \cdot x$:
  \[ P_N(N_x=n) = e^{-r \cdot x} \frac{(r \cdot x)^n}{n!} \quad E(N_x) = r \cdot x \]

- 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: $p_{\text{success}} = r \Delta x \ll 1$
  - The number of intervals is large: $n = x/\Delta x \gg 1$
  - Mean is constant: $\lambda = E(N_x) = p_{\text{success}} \cdot n = r \Delta x \cdot x/\Delta x = r \cdot x$
  - In the limit $\Delta x \ll x$, $p_{\text{success}}$ is small and $n$ is large, thus Binomial distribution $\rightarrow$ Poisson distribution
$X \sim \text{time interval between two consecutive events}$

PDF of $X = f(x)$?

CCDF of $X = \text{Prob}(X > x)$

$\text{CCDF}(x) = \text{Prob}(X > x) = \text{Prob}(N_x = 0)$
Exponential Distribution Definition

**Exponential random variable** $X$ describes interval between two successes of a constant rate (Poisson) random process with success rate $p$ 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

$$f(x) = p(1-p)^{x-1} =p e^{(x-1)\ln(1-p)}\approx pe^{-px} \text{ for small } p$$
PDF of $X$ defined as time interval between consecutive events.

CDF:

$$P(X \leq x) = \int_{0}^{x} f(u) du$$

CCDF:

$$P(X > x) = \int_{x}^{\infty} f(u) du = P(N_x = 0) = \frac{(rx)^0}{0!} e^{-rx} = e^{-rx}$$

PDF:

$$f(x) = re^{-rx}$$

$P(N_x = n) = \frac{(rx)^n}{n!} e^{-rx}$
What is the interval X between two successes of a constant rate process?

- X is a continuous random variable
- CDF: \( P_X(X>x) = P_{N}(N_X=0) = \exp(-r \cdot x) \).
  - Remember: \( P_{N}(N_X=n) = \exp(-r \cdot x) \frac{(r \cdot x)^n}{n!} \)
- PDF: \( f_X(x) = -\frac{d}{dx} CDF_X(x) = r \cdot \exp(-r \cdot x) \)
- We started with a discrete Poisson distribution where time x was a parameter
- We ended up with a continuous exponential distribution
To summarize constant rate processes:

\[ \lambda \]  
- rate per unit of length

\[ N(x) \]  
- discrete number of events in time \( x \)

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

Time interval \( X \) between successive events is a continuously distributed random variable.

Its PDF if \( f(x) = e^{-\lambda x} \)
Exponential Mean & Variance

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

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

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 an enzyme successfully catalyzes a biochemical reaction is approximated by this CDF:

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

• What is the PDF?

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

• What proportion of reactions is complete within 0.5 minutes?

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

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

\[ P(X > t + s | X > s) = P(X > t) \]

\[ P(X > t + s | X > s) = \frac{P(X > t + s, X > s)}{P(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t) \]

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 1\textsuperscript{st} 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 Poisson process.
Erlang Distribution

Generalizing from the constant rate Poisson \( \rightarrow \) Exponential:

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

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

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

The random variable \( X \) with a probability density function:

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

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

\[ \int_{0}^{+\infty} f(x) \, dx = 1, \text{ Hence} \]

\[ \Gamma(k) = \int_{0}^{+\infty} r^k x^{k-1} e^{-rx} \, dx = \int_{0}^{+\infty} y^{k-1} e^{-y} \, dy \]

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.

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

Properties of the gamma function

$$
\Gamma(1) = 1
$$

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

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

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

$\Gamma(x)$
BERNOULLI FAMILY

The Bernoulli family

Nicolaus
1623-1708

Jacob
1654-1705

Nicolaus (I)
1687-1759

Daniel
1700-1782

Nicolaus II
1695-1720

Johann (II)
1710-1790

Johann III
1744-1807

Jacob II
1759-1789

Daniel
1700-1782

Johann
1667-1748

Those shown in **bold** above are in our archive.

Run This

Gamma function

Bernoulli Trials
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 $\sigma^2 = V(X) = 1/r^2$ 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 exponential distribution with $r = 0.1$
• Calculate mean and standard deviation and compare them to $1/r$ and $1/r$ respectively.
• Make and print semilog-y plots of PDF and CCDF.
• Hint: read the help page for random(‘Exponential’…): it uses something else instead of $r$. What?
• Repeat these calculations for Gamma distribution with $r = 0.1$ and $k=4.5$ (sum of 4.5 waiting times in a Poisson process akin to Platform 9 ¾ in Harry Potter)