Continuous Probability Distributions

Uniform Distribution
Important Terms & Concepts Learned

• Probability Mass Function (PMF)
• Cumulative Distribution Function (CDF)
• Complementary Cumulative Distribution Function (CCDF)

• Expected value
• Mean
• Variance
• Standard deviation

• Uniform distribution
• Bernoulli distribution/trial
• Binomial distribution
• Poisson distribution
• Geometric distribution
• Negative binomial distribution
Which distribution is this?

\[
\binom{n}{x} p^x (1 - p)^{n-x}
\]

A. Uniform
B. Binomial
C. Geometric
D. Negative Binomial
E. Poisson

Get your i-clickers
Which distribution is this?

\[ \binom{n}{x} p^x (1 - p)^{n-x} \]

A. Uniform  
B. Binomial  
C. Geometric  
D. Negative Binomial  
E. Poisson

Get your i-clickers
Which distribution is this?

\[ \binom{x - 1}{r - 1} (1 - p)^{x-r} p^r \]

A. Uniform  
B. Binomial  
C. Geometric  
D. Negative Binomial  
E. Poisson

Get your i-clickers
Which distribution is this?

\[
\binom{x-1}{r-1} (1 - p)^{x-r} p^r
\]

A. Uniform  
B. Binomial  
C. Geometric  
D. Negative Binomial  
E. Poisson  

Get your i-clickers
Which distribution is this?

\[ \frac{e^{-\lambda} \lambda^x}{x!} \]

A. Uniform  
B. Binomial  
C. Geometric  
D. Negative Binomial  
E. Poisson

Get your i-clickers
Which distribution is this?

\[ e^{-\lambda} \frac{\lambda^x}{x!} \]

A. Uniform
B. Binomial
C. Geometric
D. Negative Binomial
E. Poisson

Get your i-clickers
<table>
<thead>
<tr>
<th>Name</th>
<th>Probability Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform</td>
<td>$\frac{1}{n}, a \leq b$</td>
<td>$(b + a)$ / $2$</td>
<td>$(b - a + 1)^2 - 1$ / $12$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$\binom{n}{x}p^x(1 - p)^{n-x}$, $x = 0, 1, \ldots, n, 0 \leq p \leq 1$</td>
<td>$np$</td>
<td>$np(1 - p)$</td>
</tr>
<tr>
<td>Geometric</td>
<td>$(1 - p)^{x-1}p$, $x = 1, 2, \ldots, 0 \leq p \leq 1$</td>
<td>$1/p$</td>
<td>$(1 - p)/p^2$</td>
</tr>
<tr>
<td>Negative binomial</td>
<td>$\binom{x - 1}{r - 1}(1 - p)^{x-r}p^r$, $x = r, r + 1, r + 2, \ldots, 0 \leq p \leq 1$</td>
<td>$r/p$</td>
<td>$r(1 - p)/p^2$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, 2, \ldots, 0 &lt; \lambda$</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
</tbody>
</table>
Continuous & Discrete Random Variables

• A **discrete random variable** is usually integer number
  – N – the number of proteins in a cell
  – D- number of nucleotides different between two sequences

• A **continuous random variable** is a real number
  – C=N/V – the concentration of proteins in a cell of volume V
  – Percentage D/L*100% of different nucleotides in protein sequences of different lengths L (depending on set of L’s may be discrete but dense)
Probability Mass Function (PMF)

• $X$ – discrete random variable

• Probability Mass Function: $f(x) = P(X=x)$ – the probability that $X$ is exactly equal to $x$

<table>
<thead>
<tr>
<th>$X$</th>
<th>$P(X=x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.6561</td>
</tr>
<tr>
<td>1</td>
<td>0.2916</td>
</tr>
<tr>
<td>2</td>
<td>0.0486</td>
</tr>
<tr>
<td>3</td>
<td>0.0036</td>
</tr>
<tr>
<td>4</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

$\sum_x P(X=x) = 1.0000$
Probability Density Function (PDF)

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

\[ f(x) \]

\[ P(a < X < b) \]

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

For a continuous random variable $X$, a probability density function is a function such that

1. $f(x) \geq 0$ means that the function is always non-negative.
2. $\int_{-\infty}^{\infty} f(x) \, dx = 1$
3. $P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx = \text{area under } f(x) \, dx \text{ from } a \text{ to } b$
Normalized histogram approximates PDF

A histogram is a graphical display of data showing a series of adjacent rectangles. Each rectangle has a base which represents an interval of data values. The height of the rectangle is a number of events in the sample within the base.

When base length is narrow, the histogram could be normalized to approximate PDF (f(x)):

\[ \text{height of each rectangle} = \frac{\text{(# of events within base)}}{\text{(total # of events)/width of its base}}. \]

Normalized histogram approximates a probability density function.
Cumulative Distribution Functions (CDF & CCDF)

The cumulative distribution function (CDF) of a continuous random variable $X$ is,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) \, du \quad \text{for} \quad -\infty < x < \infty \quad (4-3)$$

One can also use the inverse cumulative distribution function or complementary cumulative distribution function (CCDF)

$$F_>(x) = P(X > x) = \int_{x}^{\infty} f(u) \, du \quad \text{for} \quad -\infty < x < \infty$$

Definition of CDF for a continuous variable is the same as for a discrete variable.
Density vs. Cumulative Functions

• The probability density function (PDF) is the derivative of the cumulative distribution function (CDF).

\[ f(x) = \frac{dF(x)}{dx} = -\frac{dF_>(x)}{dx} \]

as long as the derivative exists.
Mean & Variance

Suppose \( X \) is a continuous random variable with probability density function \( f(x) \). The mean or expected value of \( X \), denoted as \( \mu \) or \( E(X) \), is

\[
\mu = E(X) = \int_{-\infty}^{\infty} x f(x) \, dx \quad (4-4)
\]

The variance of \( X \), denoted as \( V(X) \) or \( \sigma^2 \), is

\[
\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx = \int_{-\infty}^{\infty} x^2 f(x) \, dx - \mu^2
\]

The standard deviation of \( X \) is \( \sigma = \sqrt{\sigma^2} \).
Gallery of Useful
Continuous Probability Distributions
Continuous Uniform Distribution

• This is the simplest continuous distribution and analogous to its discrete counterpart.

• A continuous random variable $X$ with probability density function

$$f(x) = 1 / (b-a) \text{ for } a \leq x \leq b \quad (4-6)$$

Compare to discrete

$$f(x) = 1/(b-a+1)$$

Figure 4-8 Continuous uniform PDF
Comparison between Discrete & Continuous Uniform Distributions

Discrete:

• PMF: \( f(x) = \frac{1}{b-a+1} \)
• Mean and Variance:
  \[ \mu = E(x) = \frac{b+a}{2} \]
  \[ \sigma^2 = V(x) = \frac{[(b-a+1)^2-1]}{12} \]

Continuous:

• PMF: \( f(x) = \frac{1}{b-a} \)
• Mean and Variance:
  \[ \mu = E(x) = \frac{b+a}{2} \]
  \[ \sigma^2 = V(x) = \frac{(b-a)^2}{12} \]
X is a continuous random variable with a uniform distribution between 0 and 3. What is \( P(X=1) \)?

A. 1/4  
B. 1/3  
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 3. What is P(X=1)?

A. 1/4  
B. 1/3  
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 3.

What is $P(X<1)$?

A. $\frac{1}{4}$
B. $\frac{1}{3}$
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 3.

What is \( P(X<1) \)?

A. 1/4  
B. 1/3  
C. 0  
D. Infinity  
E. I have no idea
Poisson process

Discrete events happen at rate \( \lambda \).

Expected number of events in time \( x \) is \( \lambda x \).

The actual number of events \( N \) is a Poisson distributed discrete random variable.

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

Why Poisson? Divide \( x \) into many tiny intervals of length \( \Delta x \)

\[
p = \lambda \Delta x
\]

\[
L = \frac{x}{\Delta x}
\]

\[
E(N_x) = pL = \lambda x
\]

Prove \( N \) follows Poisson

\[
\lim_{\Delta x \to 0, \ L \to \infty} \frac{p_n}{x^n} = e^{-\lambda}
\]
Poisson (constant rate) processes

• Let’s assume that proteins are produced by all ribosomes in the cell at a rate \( \lambda \) per second.
• The expected number of proteins produced in \( x \) seconds is \( \lambda x \).
• The actual number of proteins \( N_x \) is a discrete random variable following a Poisson distribution with mean \( \lambda x \):
  \[
P_N(N_x=n) = \exp(-\lambda x)(\lambda x)^n/n! \quad E(N_x) = \lambda x
  \]
• Why Discrete Poisson Distribution?
  – Divide time into many tiny intervals of length \( x_0 \ll 1/\lambda \)
  – The probability of success (protein production) per internal is small: \( p = \lambda \cdot x_0 \ll 1 \)
  – The number of intervals is large: \( L = x/x_0 \gg 1 \)
  – Mean is constant: \( E(N_x) = p \cdot L = \lambda x_0 \cdot x/x_0 = \lambda \cdot x \)
  – \( P(N_x=n) = L!/n!(L-n)! \cdot p^n \cdot (1-p)^{L-n} \)
  – In the limit \( p \to 0 \) we have Binomial distribution \( \to \) Poisson
Exponential Distribution Definition

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

The probability density function of $X$ is:

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \leq x < \infty \quad (4-14)$$

Closely related to the discrete geometric distribution

$$f(x) = p(1-p)^{x-1} \quad (3-9)$$
What is the interval $X$ between two successes of a constant rate process?

- $X$ is a continuous random variable
- CDF: $P_X(X>x) = P_{\mathcal{N}}(N_X=0)=\exp(-\lambda x)$.
  - Remember: $P_{\mathcal{N}}(N_X=n)=\exp(-\lambda x) \frac{(\lambda x)^n}{n!}$
- PDF: $f_X(x) = -\frac{d}{dx} P_X(x) = \lambda \exp(-\lambda x)$
- 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 is \( f(x) = \lambda e^{-\lambda x} \)
Exponential Mean & Variance

If the random variable $X$ has an exponential distribution with parameter $\lambda$, 

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

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:

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

• What is the PDF?

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

• 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

\[ \mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t) \]

\[ \mathbb{P}(X > t + s \mid X > s) = \frac{\mathbb{P}(X > t + s, X > s)}{\mathbb{P}(X > s)} = \frac{\exp(-\lambda (t + s))}{\exp(-\lambda s)} = \exp(-\lambda t) = \]

\[ = \mathbb{P}(X > t) \]

Exponential is the only memoryless distribution
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.
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 $\rightarrow$ 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:

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

for $x > 0$ and $r = 1, 2, 3, ...$
Gamma Distribution

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

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

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 Function

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

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

Properties of the gamma function

$$\Gamma(1) = 1$$
$$\Gamma(r) = (r-1)\Gamma(r-1) \quad \text{recursive property}$$
$$\Gamma(r) = (r-1)! \quad \text{factorial function}$$
$$\Gamma(1/2) = \pi^{1/2} = 1.77 \quad \text{interesting fact}$$
Daniel Bernoulli's Gamma

$\Gamma(x)$
Mean & Variance of the Erlang and Gamma

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

• Generalization of exponential results:
  \[ \mu = E(X) = \frac{1}{\lambda} \]  
  \[ \sigma^2 = V(X) = \frac{1}{\lambda^2} \]  
  or

  Negative binomial results:
  \[ \mu = E(X) = \frac{r}{p} \]  
  \[ \sigma^2 = V(X) = \frac{r(1-p)}{p^2} \]
Matlab exercise:

• Generate a sample of 100,000 variables with exponential distribution with $\lambda = 0.1$
• Calculate mean and standard deviation and compare them to $1/\lambda$
• Plot semilog-y plots of PDF and CCDF.
• Hint: read the help page for `random('Exponential'...)`: it uses something else instead of $\lambda$. What?