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?

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

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
<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>( \frac{b + a}{2} )</td>
<td>( \frac{(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>( \frac{(1 - p)}{p^2} )</td>
</tr>
<tr>
<td>Negative binomial</td>
<td>( \frac{(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
  - $C = \frac{N}{V}$ – the concentration of proteins in a cell of volume $V$
  - Percentage $\frac{D}{L} \times 100\%$ 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 \( x \)

- Probability Mass Function for the # of mismatches in 4-mers

\[
\begin{array}{c|c}
X & P(X) \\
\hline
0 & 0.6561 \\
1 & 0.2916 \\
2 & 0.0486 \\
3 & 0.0036 \\
4 & 0.0001 \\
\end{array}
\]

\[\sum_x P(X=x) = 1.0000\]
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 $b$. 
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. $P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx = \text{area under } f(x) \, dx \text{ from } a \text{ to } b$
EXAMPLE Suppose that $X$ is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} 
C(4x - 2x^2) & 0 < x < 2 \\
0 & \text{otherwise}
\end{cases}$$

(a) What is the value of $C$?
(b) Find $P\{X > 1\}$. 

EXAMPLE Suppose that $X$ is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} \frac{C(4x - 2x^2)}{2} & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the value of $C$?
(b) Find $P\{X > 1\}$.

\[\int_0^2 \frac{C(4x - 2x^2)}{2} \, dx = 1\]

\[C \cdot \left(4 \cdot \frac{2}{2} - 2 \cdot \frac{2^3}{3}\right) = 1\]

\[\frac{3}{8} \cdot \frac{16}{6} = \frac{8}{3}\]

(b) $P\{X > 1\} = \int_1^2 f(x) \, dx = \frac{1}{2}$ by symmetry
Histogram approximates PDF

A **histogram** is 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 creates an **area** which represents the **probability of X to be within the base**. When base length is narrow, the histogram approximates $f(x)$ (PDF): **height of each rectangle = its area/length of its base**.

![Figure 4-3 Histogram approximates a probability density function.](image)
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).

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

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

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) \, dx$$

(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

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

Compare to discrete

$$f(x) = \frac{1}{(b-a+1)}$$

**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. $\frac{1}{4}$  
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

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
Matlab exercise:

• generate 100,000 random numbers drawn from uniform distribution between 3 and 7
• plot histogram approximating its PDF
• calculate mean, standard deviation and variance
Matlab template: Uniform PDF

- Stats=
- \( r2 = \text{mean} + \text{var}. \times \text{rand} \text{(Stats}, 1) \);
- \text{disp(mean}(r2));
- \text{disp(var}(r2));
- \text{disp(std}(r2));
- step=0.1;
- \[ a, b \] = \text{hist}(r2,0:step:8);
- pdf_e = a./sum(a). \times \text{step} \text{ (or / step);}
- \text{figure; plot}(b, pdf_e, 'ko-');
Matlab exercise: Uniform PDF

- Stats=100000;
- r2=3+4.*rand(Stats,1);
- disp(mean(r2));
- disp(var(r2));
- disp(std(r2));
- step=0.1;
- [a,b]=hist(r2,0:step:8);
- pdf_e=a./sum(a)./step;
- figure; plot(b,pdf_e,'ko-');