Continuous Probability Distributions

Uniform Distribution
Important terms & concepts for discrete random variables

• 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

**Boldface and underlined** are the same for continuous distributions
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?

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

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><strong>Discrete</strong></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>( \frac{1}{p} )</td>
<td>( \frac{(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>( \frac{r}{p} )</td>
<td>( \frac{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 $x$

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

<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.

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

$P(a < X < b)$
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 \( f(x) \):

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

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

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) \, dx \quad (4-4)$$

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 $\sigma = \sqrt{\sigma^2}$. 

Sec 4-4 Mean & Variance of a Continuous Random Variable
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

\[ f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b \]  

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

Continuous:

• PMF: \( f(x) = 1/(b-a) \)
• Mean and Variance:
  \[ \mu = E(x) = (b+a)/2 \]
  \[ \sigma^2 = V(x) = (b-a)^2/12 \]
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. $\frac{1}{4}$

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
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 $\Delta x$

$p = \lambda \Delta x$

$L = \frac{x}{\Delta x}$

$E(N_x) = \rho L = \lambda x$

$\rho \Delta x \to 0$, $L \sim \frac{1}{\Delta x} \to \infty$ Poisson
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 $r \cdot x$:

$$P_{N}(N_x=n)=\exp(-r\cdot x)(r\cdot x)^n/n! \quad E(N_x)=rx$$

• 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: $\lambda=E(N_x)=p_{success} \cdot n= r\Delta x \cdot x/\Delta x = r \cdot x$
  – In the limit $\Delta x \ll x$, $p_{success}$ is small and $n$ is large, thus Binomial distribution $\rightarrow$ Poisson distribution
Exponential Distribution Definition

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

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

CDF: $P(X < x) = \int_0^x f(u) \, du$

CCDF: $P(X > x) = \int_x^{\infty} f(u) \, du = P(N_x = 0) = \frac{(rx)^n}{n!} e^{-rx}$

$P(N_x = n) = \frac{(rx)^n}{n!} e^{-rx}$

CCDF: $P(x > x) = e^{-rx}$

PDF: $f(x) = -e^{-rx}$
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: $P_{N}(N_X=n)=\exp(-r \cdot x) \ (r \cdot x)^n/n!$
- PDF: $f_X(x)=-dCDF_X(x)/dx = r \cdot \exp(-r \cdot x)$
- We started with a discrete Poisson distribution where time $x$ was a parameter.
- We ended up with a continuous exponential distribution.