Applied Statistics I

Liang Zhang

Department of Mathematics, University of Utah

June 26, 2008
Poisson Distribution

In some sense, the Poisson distribution can be recognized as the limit of a binomial experiment.

**Proposition**
Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that $np$ approaches a value $\lambda > 0$. Then $b(x; n, p) \to p(x; \lambda)$.

This tells us in any binomial experiment in which $n$ is large and $p$ is small, $b(x; n, p) \approx p(x; \lambda)$, where $\lambda = np$.

As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$. 
Poisson Distribution

In some sense, the Poisson distribution can be recognized as the limit of a binomial experiment.

Proposition

Suppose that in the binomial pmf \( b(x; n, p) \), we let \( n \to \infty \) and \( p \to 0 \) in such a way that \( np \) approaches a value \( \lambda > 0 \). Then \( b(x; n, p) \to p(x; \lambda) \).
In some sense, the **Poisson distribution** can be recognized as the limit of a binomial experiment.

**Proposition**

*Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that $np$ approaches a value $\lambda > 0$. Then $b(x; n, p) \to p(x; \lambda)$.  

This tells us in any binomial experiment in which $n$ is large and $p$ is small, $b(x; n, p) \approx p(x; \lambda)$, where $\lambda = np$.  

In some sense, the Poisson distribution can be recognized as the limit of a binomial experiment.

**Proposition**

*Suppose that in the binomial pmf \( b(x; n, p) \), we let \( n \to \infty \) and \( p \to 0 \) in such a way that \( np \) approaches a value \( \lambda > 0 \). Then \( b(x; n, p) \to p(x; \lambda) \).*

This tells us in any binomial experiment in which \( n \) is large and \( p \) is small, \( b(x; n, p) \approx p(x; \lambda) \), where \( \lambda = np \).

As a rule of thumb, this approximation can safely be applied if \( n > 50 \) and \( np < 5 \).
Example 3.40:
If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors?

Let S denote a page containing at least one error, F denote an error-free page and X denote the number of pages containing at least one error. Then X is a binomial rv, and

$$P(X = 1) = b(1; 400, 0.005) \approx p(1; 400 \cdot 0.005) = p(1; 2) = e^{-2}(2)1! = 0.270671$$
Example 3.40:
If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors?
Example 3.40:
If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors?

Let \( S \) denote a page containing at least one error, \( F \) denote an error-free page and \( X \) denote the number of pages containing at least one error. Then \( X \) is a binomial rv, and

\[
P(X = 1) = b(1; 400, 0.005) \approx p(1; 400 \cdot 0.005) = p(1; 2) = \frac{e^{-2}(2)}{1!} = 0.270671
\]
Poisson Distribution

\[ b(x; n, p) \rightarrow p(x; \lambda) \text{ as } n \rightarrow \infty \text{ and } p \rightarrow 0 \text{ with } np \rightarrow \lambda. \]

\[ b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \]

\[ \lim_{n \rightarrow \infty} \frac{n!}{x!(n-x)!} p^x = \lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-x+1)}{x!} p^x = \lambda^x x! \lim_{n \rightarrow \infty} \left(1 - \frac{p}{n}\right)^{n-x} = e^{-\lambda} \]
Poisson Distribution

A proof for $b(x; n, p) \to p(x; \lambda)$ as $n \to \infty$ and $p \to 0$ with $np \to \lambda$.

\[
b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}
\]

\[
\lim_{n \to \infty} \frac{n!}{x!(n-x)!} p^x = \lim_{n \to \infty} \frac{n(n-1) \cdots (n-x+1)}{x!} p^x
\]

\[
= \lim_{n \to \infty} \frac{(np)[(n-1)p] \cdots [(n-x+1)p]}{x!}
\]

\[
= \lambda^x
\]

\[
= \frac{x!}{x!}
\]

\[
\lim_{n \to \infty} (1-p)^{n-x} = \lim_{n \to \infty} \left(1 - \frac{np}{n}\right)^{n-x}
\]

\[
= \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^{n-x}
\]

\[
= e^{-\lambda}
\]
Probability Density Functions

Recall that a random variable \( X \) is continuous if

1. possible values of \( X \) comprise either a single interval on the number line (for some \( A < B \), any number \( x \) between \( A \) and \( B \) is a possible value)
2. \( P(X = c) = 0 \) for any number \( c \) that is a possible value of \( X \).

Examples:

1. \( X = \) the temperature in one day. \( X \) can be any value between \( L \) and \( H \), where \( L \) represents the lowest temperature and \( H \) represents the highest temperature.

2. Example 4.3: \( X = \) the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences. Is \( X \) really a continuous rv? No. The point is that there are customers lucky enough to have no wait whatsoever before climbing into the barber’s chair, which means \( P(X = 0) > 0 \). Only conditioned on no chairs being empty, the waiting time will be continuous.
Recall that a random variable $X$ is **continuous** if

1. possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;
2. $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

**Examples:**

1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.
2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences. Is $X$ really a continuous rv? No. The point is that there are customers lucky enough to have no wait whatsoever before climbing into the barber’s chair, which means $P(X = 0) > 0$. Only conditioned on no chairs being empty, the waiting time will be continuous.
Probability Density Functions

Recall that a random variable $X$ is continuous if

1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;

2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

Examples:

1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.

2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences.

Is $X$ really a continuous rv? No. The point is that there are customers lucky enough to have no wait whatsoever before climbing into the barber's chair, which means $P(X = 0) > 0$. Only conditioned on no chairs being empty, the waiting time will be continuous.
Probability Density Functions

Recall that a random variable $X$ is **continuous** if

1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;

2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$. 

Examples:
1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.

2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences. Is $X$ really a continuous rv? No. The point is that there are customers lucky enough to have no wait whatsoever before climbing into the barber’s chair, which means $P(X = 0) > 0$. Only conditioned on no chairs being empty, the waiting time will be continuous.
Probability Density Functions

Recall that a random variable $X$ is **continuous** if

1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;

2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

**Examples:**

1. $X =$ the temperature in one day. \(X\) can be any value between \(L\) and \(H\), where \(L\) represents the lowest temperature and \(H\) represents the highest temperature.
Recall that a random variable $X$ is continuous if
1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;
2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

**Examples:**
1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.
2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences.
Probability Density Functions

Recall that a random variable $X$ is continuous if
1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;
2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

Examples:
1. X = the temperature in one day. X can be any value between L and H, where L represents the lowest temperature and H represents the highest temperature.
2. Example 4.3: X = the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences. Is $X$ really a continuous rv?
Recall that a random variable $X$ is continuous if
1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;
2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

Examples:
1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.
2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences. Is $X$ really a continuous rv? No.
Recall that a random variable $X$ is continuous if
1). possible values of $X$ comprise either a single interval on the number line (for some $A < B$, any number $x$ between $A$ and $B$ is a possible value) or a union of disjoint intervals;
2). $P(X = c) = 0$ for any number $c$ that is a possible value of $X$.

Examples:
1. $X =$ the temperature in one day. $X$ can be any value between $L$ and $H$, where $L$ represents the lowest temperature and $H$ represents the highest temperature.
2. Example 4.3: $X =$ the amount of time a randomly selected customer spends waiting for a haircut before his/her haircut commences.

Is $X$ really a continuous rv? No.
The point is that there are customers lucky enough to have no wait whatsoever before climbing into the barber’s chair, which means $P(X = 0) > 0$. Only conditioned on no chairs being empty, the waiting time will be continuous.
Let's consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what's the probability for the temperature between $70^\circ$ and $80^\circ$?

Ultimately, we want to know the probability distribution for $X$.

One way to do that is to record the temperature from time to time and then plot the histogram. However, when you plot the histogram, it's up to you to choose the bin size.

But if we make the bin size finer and finer (meanwhile we need more and more data), the histogram will become a smooth curve which will represent the probability distribution for $X$. 
Let’s consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what’s the probability for the temperature between 70° and 80°?
Let’s consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what’s the probability for the temperature between $70^\circ$ and $80^\circ$? Ultimately, we want to know the probability distribution for $X$. 
Let’s consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what’s the probability for the temperature between 70° and 80°? Ultimately, we want to know the probability distribution for $X$. One way to do that is to record the temperature from time to time and then plot the histogram.
Probability Density Functions

Let’s consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what’s the probability for the temperature between $70^\circ$ and $80^\circ$?

Ultimately, we want to know the probability distribution for $X$.

One way to do that is to record the temperature from time to time and then plot the histogram.

However, when you plot the histogram, it’s up to you to choose the bin size.
Let’s consider the temperature example again. We want to know the probability that the temperature is in any given interval. For example, what’s the probability for the temperature between $70^\circ$ and $80^\circ$? Ultimately, we want to know the probability distribution for $X$. One way to do that is to record the temperature from time to time and then plot the histogram. However, when you plot the histogram, it’s up to you to choose the bin size. But if we make the bin size finer and finer (meanwhile we need more and more data), the histogram will become a smooth curve which will represent the probability distribution for $X$. 
Probability Density Functions
Probability Density Functions

(a) bin size 2

(b) bin size 0.8

(c) limit case
Definition
Let $X$ be a continuous rv. Then a probability distribution or probability density function (pdf) of $X$ is a function $f(x)$ such that for any two numbers $a$ and $b$ with $a \leq b$,

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

That is, the probability that $X$ takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function. The graph of $f(x)$ is often referred to as the density curve.
Probability Density Functions

Definition
Let $X$ be a continuous rv. Then a **probability distribution** or **probability density function** (pdf) of $X$ is a function $f(x)$ such that for any two numbers $a$ and $b$ with $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

That is, the probability that $X$ takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function. The graph of $f(x)$ is often referred to as the *density curve*. 
Figure: $P(60 \leq X \leq 70)$
Remark:

For \( f(x) \) to be a legitimate pdf, it must satisfy the following two conditions:

1. \( f(x) \geq 0 \) for all \( x \);
2. \( \int_{-\infty}^{\infty} f(x) \, dx = \text{area under the entire graph of } f(x) = 1. \)
Remark:
For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:
Remark:
For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:
1. $f(x) \geq 0$ for all $x$;
Remark:
For \( f(x) \) to be a legitimate pdf, it must satisfy the following two conditions:
1. \( f(x) \geq 0 \) for all \( x \);
2. \( \int_{-\infty}^{\infty} f(x) \, dx = \text{area under the entire graph of } f(x) = 1. \)
Example:
A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is
\[
f(x) = \begin{cases} 
\frac{1}{24} & 0 \leq x \leq 24 \\
0 & \text{otherwise}
\end{cases}
\]
The density curve for $X$ is showed below:
Example:
A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is known as

$$f(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$
Example:
A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is known as

$$f(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

The density curve for $X$ is showed below:
Probability Density Functions

Example: (continued)

A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is known as

$$f(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

If we want to know the probability that the clock will stop between 2:00pm and 2:45pm, then

$$P(14 \leq X \leq 14.75) = \int_{14}^{14.75} \frac{1}{24} \, dx = \frac{1}{24} \cdot 0.75 = \frac{1}{32}$$
Example: (continued)

A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is known as

$$f(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

If we want to know the probability that the clock will stop between 2:00pm and 2:45pm, then

$$P(14 \leq X \leq 14.75) = \int_{14}^{14.75} \frac{1}{24} \, dx = \frac{1}{24} \cdot 0.75 = \frac{1}{32}$$
Example: (continued)
A clock stops at random at any time during the day. Let $X$ be the time (hours plus fractions of hours) at which the clock stops. The pdf for $X$ is known as

$$f(x) = \begin{cases} \frac{1}{24} & 0 \leq x \leq 24 \\ 0 & \text{otherwise} \end{cases}$$

If we want to know the probability that the clock will stop between 2:00pm and 2:45pm, then

$$P(14 \leq X \leq 14.75) = \int_{14}^{14.75} \frac{1}{24} \, dx = \frac{1}{24} |_{14}^{14.75} = \frac{1}{32}$$
A continuous rv $X$ is said to have a uniform distribution on the interval $[A, B]$, if the pdf of $X$ is

$$f(x; A, B) = \begin{cases} 
1/A - 1/B & A \leq x \leq B \\
0 & \text{otherwise}
\end{cases}$$

The graph of any uniform pdf looks like the graph in the previous example:
Probability Density Functions

Definition

A continuous rv $X$ is said to have a **uniform distribution** on the interval $[A, B]$, if the pdf of $X$ is

$$f(x; A, B) = \begin{cases} \frac{1}{A-B} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$
Probability Density Functions

Definition
A continuous rv $X$ is said to have a **uniform distribution** on the interval $[A, B]$, if the pdf of $X$ is

$$f(x; A, B) = \begin{cases} \frac{1}{A-B} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

The graph of any uniform pdf looks like the graph in the previous example:
Comparisons between continuous rv and discrete rv:

For discrete rv $Y$, each possible value is assigned positive probability;

For continuous rv $X$, the probability for any single possible value is 0!

$$P(X = c) = \int_{c}^{c} f(x) \, dx = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x) \, dx = 0$$

Since $P(X = c) = 0$ for continuous rv $X$ and $P(Y = c') > 0$, we have

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b)$$

while

$$P(a' \leq Y \leq b'), P(a' < Y < b'), P(a' < Y \leq b') \text{ and } P(a' \leq Y < b')$$

are different.
Comparisons between continuous rv and discrete rv:

For discrete rv $Y$, each possible value is assigned positive probability;
For continuous rv $X$, the probability for any single possible value is 0!

$$P(X = c) = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x) \, dx$$

Since $P(X = c) = 0$ for continuous rv $X$ and $P(Y = c') > 0$, we have

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b)$$

while $P(a' \leq Y \leq b')$, $P(a' < Y < b')$, $P(a' < Y \leq b')$ and $P(a' \leq Y < b')$ are different.
Comparisons between continuous rv and discrete rv:
For discrete rv $Y$, each possible value is assigned positive probability;
For continuous rv $X$, the probability for any single possible value is 0!
Comparisons between continuous rv and discrete rv:

For discrete rv $Y$, each possible value is assigned positive probability;

For continuous rv $X$, the probability for any single possible value is 0!

\[
P(X = c) = \int_{c}^{c} f(x)dx = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x)dx = 0
\]
Comparisons between continuous rv and discrete rv:
For discrete rv $Y$, each possible value is assigned positive probability;
For continuous rv $X$, the probability for any single possible value is 0!

$$P(X = c) = \int_{c}^{c} f(x)\,dx = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x)\,dx = 0$$

Since $P(X = c) = 0$ for continuous rv $X$ and $P(Y = c') > 0$, 
Comparisons between continuous rv and discrete rv:
For discrete rv $Y$, each possible value is assigned positive probability;
For continuous rv $X$, the probability for any single possible value is 0!

$$
P(X = c) = \int_{c}^{c} f(x) \, dx = \lim_{\epsilon \to 0} \int_{c-\epsilon}^{c+\epsilon} f(x) \, dx = 0
$$

Since $P(X = c) = 0$ for continuous rv $X$ and $P(Y = c') > 0$, we have

$$
P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b)
$$

while $P(a' \leq Y \leq b')$, $P(a' < Y < b')$, $P(a' < Y \leq b')$ and $P(a' \leq Y < b')$ are different.
Cumulative Distribution Functions

Definition

The cumulative distribution function $F(x)$ for a continuous rv $X$ is defined for every number $x$ by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) \, dy$$

For each $x$, $F(x)$ is the area under the density curve to the left of $x$. 

Liang Zhang (UofU)  
Applied Statistics I  
June 26, 2008  15 / 25
Cumulative Distribution Functions

Definition

The **cumulative distribution function** \( F(x) \) for a continuous rv \( X \) is defined for every number \( x \) by

\[
F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) \, dy
\]

For each \( x \), \( F(x) \) is the area under the density curve to the left of \( x \).
Cumulative Distribution Functions

Definition

The **cumulative distribution function** \( F(x) \) for a continuous rv \( X \) is defined for every number \( x \) by

\[
F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y) \, dy
\]

For each \( x \), \( F(x) \) is the area under the density curve to the left of \( x \).
Cumulative Distribution Functions

Example 4.6

Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} 
1 & \text{if } A \leq x \leq B \\
0 & \text{otherwise}
\end{cases}$$

Then the cdf for $X$ is calculated as following:

For $x < A$, $F(x) = 0$;

for $A \leq x < B$, we have

$$F(x) = \int_{-\infty}^{x} f(y) \, dy = \int_{A}^{x} \frac{1}{B-A} \, dy = \frac{x - A}{B - A}$$

for $x \geq B$, $F(x) = 1$.

Therefore the entire cdf for $X$ is

$$F(x) = \begin{cases} 
0 & x < A \\
\frac{x - A}{B - A} & A \leq x < B \\
1 & x \geq B
\end{cases}$$
Example 4.6
Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

Then the cdf for $X$ is calculated as following:
Example 4.6
Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

Then the cdf for $X$ is calculated as following:
For $x < A$, $F(x) = 0$;
Example 4.6
Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} 
\frac{1}{B-A} & A \leq x \leq B \\
0 & \text{otherwise}
\end{cases}$$

Then the cdf for $X$ is calculated as following:
For $x < A$, $F(x) = 0$; for $A \leq x < B$, we have

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{A}^{x} \frac{1}{B-A} dy = \frac{1}{B-A} \cdot y \bigg|_{y=A}^{x} = \frac{x-A}{B-A};$$

then for $x \geq B$, $F(x) = 1$.
Cumulative Distribution Functions

Example 4.6
Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

Then the cdf for $X$ is calculated as following:
For $x < A$, $F(x) = 0$; for $A \leq x < B$, we have

$$F(x) = \int_{-\infty}^{x} f(y) dy = \int_{A}^{x} \frac{1}{B-A} dy = \frac{1}{B-A} \cdot y \bigg|_{y=A}^{y=x} = \frac{x-A}{B-A};$$

for $x \geq B$, $F(x) = 1$. 
Example 4.6
Let $X$, the thickness of a certain metal sheet, have a uniform distribution on $[A, B]$. The pdf for $X$ is

$$f(x) = \begin{cases} 
\frac{1}{B-A} & A \leq x \leq B \\
0 & \text{otherwise}
\end{cases}$$

Then the cdf for $X$ is calculated as following:
For $x < A$, $F(x) = 0$; for $A \leq x < B$, we have

$$F(x) = \int_{-\infty}^{x} f(y) dy = \int_{A}^{x} \frac{1}{B-A} dy = \frac{1}{B-A} \cdot y \mid_{y=A}^{x} = \frac{x - A}{B - A};$$

for $x \geq B$, $F(x) = 1$.

Therefore the entire cdf for $X$ is

$$F(x) = \begin{cases} 
0 & x < A \\
\frac{x - A}{B - A} & A \leq x < B \\
1 & x \geq B
\end{cases}$$
Cumulative Distribution Functions

Proposition

Let $X$ be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number $a$, $P(X > a) = 1 - F(a)$ and for any two numbers $a$ and $b$ with $a < b$, $P(a \leq X \leq b) = F(b) - F(a)$. 

Liang Zhang (UofU)
Applied Statistics I
June 26, 2008 17 / 25
Proposition

Let $X$ be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number $a$,

$$P(X > a) = 1 - F(a)$$

and for any two numbers $a$ and $b$ with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a).$$
Proposition

Let $X$ be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number $a$,

$$P(X > a) = 1 - F(a)$$

and for any two numbers $a$ and $b$ with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a).$$
Cumulative Distribution Functions

Example (Problem 15)

Let $X$ denote the amount of space occupied by an article placed in a 1-ft $^3$ packing container. The pdf of $X$ is

$$f(x) = \begin{cases} 
90x - 8 & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases}$$

Then what is $P(X \leq 0.5)$ and $P(0.25 < X \leq 0.5)$?
Example (Problem 15)
Let $X$ denote the amount of space occupied by an article placed in a 1-ft$^3$ packing container. The pdf of $X$ is

$$f(x) = \begin{cases} 90x^8(1 - x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then what is $P(X \leq 0.5)$ and $P(0.25 < X \leq 0.5)$?
Proposition

If $X$ is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every $x$ at which the derivative $F'(x)$ exists, $F'(x) = f(x)$.

e.g. for the previous example, we know the cdf for $X$ is

$$F(x) = \begin{cases} 
0 & x \leq 0 \\
10 & 9 < x \leq 10 \\
1 & x > 10
\end{cases}$$

Then the derivative of $F(x)$ exists on $(-\infty, \infty)$ and we get

$$F'(x) = \begin{cases} 
90 & 0 < x < 1 \\
0 & -\infty < x \leq 0 \text{ or } 1 \leq x < \infty
\end{cases}$$

which is just the pdf of $X$. 
Proposition

If $X$ is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every $x$ at which the derivative $F'(x)$ exists, $F'(x) = f(x)$. 

For example, we know the cdf for $X$ is

$$F(x) = \begin{cases} 
0 & x \leq 0 \\
10 - 9x & 0 < x < 1 \\
1 & x \geq 1 
\end{cases}$$

Then the derivative of $F(x)$ exists on $(-\infty, \infty)$ and we get

$$F'(x) = \begin{cases} 
90x - 90 & 0 < x < 1 \\
0 & -\infty < x \leq 0 \text{ or } 1 \leq x < \infty
\end{cases}$$

which is just the pdf of $X$. 
Proposition

If $X$ is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every $x$ at which the derivative $F'(x)$ exists, $F'(x) = f(x)$.

e.g. for the previous example, we know the cdf for $X$ is

$$F(x) = \begin{cases} 
0 & x \leq 0 \\
10x^9 - 9x^{10} & 0 < x < 1 \\
1 & x \geq 1
\end{cases}$$

Then the derivative of $F(x)$ exists on $(-\infty, \infty)$ and we get

$$F'(x) = 90x^8 - 90x^9$$

for $0 < x < 1$ and $F'(x) = 0$ for $-\infty < x \leq 0$ and $1 \leq x < \infty$, which is just the pdf of $X$. 
Cumulative Distribution Functions

Definition
The expected value or mean valued of a continuous rv \( X \) with pdf \( f(x) \) is

\[
\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx
\]

Definition
The variance of a continuous random variable \( X \) with pdf \( f(x) \) and mean value \( \mu \) is

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

The standard deviation (SD) of \( X \) is \( \sigma_X = \sqrt{V(X)} \).
Cumulative Distribution Functions

Definition

The **expected value** or **mean valued** of a continuous rv $X$ with pdf $f(x)$ is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$
Cumulative Distribution Functions

Definition

The expected value or mean valued of a continuous rv $X$ with pdf $f(x)$ is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

Definition

The variance of a continuous random variable $X$ with pdf $f(x)$ and mean value $\mu$ is

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

The standard deviation (SD) of $X$ is $\sigma_X = \sqrt{V(X)}$. 
Cumulative Distribution Functions

Proposition

\[ V(X) = E(X^2) - [E(X)]^2 \]

e.g. for the previous example, the pdf of \( X \) is given as

\[
f(x) = \begin{cases} 
90x^8(1-x) & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases}
\]

Then the expected value of \( X \) is

\[
E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{0}^{1} x \cdot 90x^8(1-x) \, dx
\]

\[
= 90 \left[ \frac{x^9}{9} - \frac{x^{10}}{10} \right]_{x=0}^{x=1} = \frac{9}{11}
\]
Cumulative Distribution Functions

Proposition

\[ V(X) = E(X^2) - [E(X)]^2 \]
Cumulative Distribution Functions

Proposition

\[ V(X) = E(X^2) - [E(X)]^2 \]

e.g. for the previous example, the pdf of \( X \) is given as

\[ f(x) = \begin{cases} 
90x^8(1 - x) & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases} \]
Cumulative Distribution Functions

Proposition

\[ V(X) = E(X^2) - [E(X)]^2 \]

e.g. for the previous example, the pdf of \( X \) is given as

\[ f(x) = \begin{cases} 
90x^8(1-x) & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases} \]

Then the expected value of \( X \) is

\[
E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx = \int_{0}^{1} x \cdot 90x^8(1-x) \, dx
\]

\[
= 90 \int_{0}^{1} (x^9 - x^{10}) \, dx = 90\left( \frac{1}{10}x^{10} - \frac{1}{11}x^{11} \right) \bigg|_{x=0}^{x=1} = \frac{9}{11}
\]
Cumulative Distribution Functions

Example continued: the pdf for $X$ is 

$$f(x) = \begin{cases} 
90x^8(1-x) & 0 < x < 1 \\
0 & \text{otherwise} 
\end{cases}$$

The variance of $X$ is 

$$V(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x) \, dx - \left[ \int_{-\infty}^{\infty} x \cdot f(x) \, dx \right]^2$$

$$= \int_{0}^{1} x^2 \cdot 90x^8(1-x) \, dx - \left[ \int_{0}^{1} x \cdot 90x^8(1-x) \, dx \right]^2$$

$$= 90 \int_{0}^{1} \left( x^{10} - x^{11} \right) \, dx - \left( \frac{9}{11} \right)^2$$

$$= 15 \frac{22}{242} - \frac{81}{121} = \frac{3}{242}.$$
Cumulative Distribution Functions

Example continued: the pdf for $X$ is

$$f(x) = \begin{cases} 
90x^8(1 - x) & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases}$$
Example continued: the pdf for X is

\[ f(x) = \begin{cases} 
90x^8(1-x) & 0 < x < 1 \\
0 & \text{otherwise}
\end{cases} \]

The variance of X is

\[
V(X) = E(X^2) - [E(X)]^2 = \int^\infty_{-\infty} x^2 \cdot f(x)dx - [\int^\infty_{-\infty} x \cdot f(x)dx]^2
\]

\[
= \int^1_0 x^2 \cdot 90x^8(1-x)dx - [\int^1_0 x \cdot 90x^8(1-x)dx]^2
\]

\[
= 90 \int^1_0 (x^{10} - x^{11})dx - (\frac{9}{11})^2
\]

\[
= 90\left(\frac{1}{11}x^{11} - \frac{1}{12}x^{12}\right)\bigg|_{x=0}^{x=1} - (\frac{9}{11})^2
\]

\[
= \frac{15}{22} - \frac{81}{121} = \frac{3}{242}
\]
Cumulative Distribution Functions

Definition

Let $p$ be a number between 0 and 1. The \((100p)\)th percentile of the distribution of a continuous rv $X$, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) \, dy$$

In words, the \((100p)\)th percentile $\eta(p)$ is the $X$ value such that there are $100p\%$ of $X$ values below $\eta(p)$.

Graphically, $\eta(p)$ is the value on the measurement axis such that $100p\%$ of the area under the graph of $f(x)$ lies to the left of $\eta(p)$ and $100(1-p)\%$ lies to the right.
Cumulative Distribution Functions

Definition

Let $p$ be a number between 0 and 1. The $(100p)$th percentile of the distribution of a continuous rv $X$, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) \, dy$$

In words, the $(100p)$th percentile $\eta(p)$ is the $X$ value such that there are $100p\%$ of $X$ values below $\eta(p)$.

Graphically, $\eta(p)$ is the value on the measurement axis such that $100p\%$ of the area under the graph of $f(x)$ lies to the left of $\eta(p)$ and $100(1-p)\%$ lies to the right.
Definition

Let $p$ be a number between 0 and 1. The \textbf{(100$p$)th percentile} of the distribution of a continuous rv $X$, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) \, dy$$

In words, the \textbf{(100$p$)th percentile} $\eta(p)$ is the $X$ value such that there are $100p\%$ $X$ values below $\eta(p)$. 

Cumulative Distribution Functions

Definition

Let $p$ be a number between 0 and 1. The \textbf{(100$p$)th percentile} of the distribution of a continuous rv $X$, denoted by $\eta(p)$, is defined by

$$ p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y) \, dy $$

In words, the (100$p$)th percentile $\eta(p)$ is the $X$ value such that there are $100p\%$ $X$ values below $\eta(p)$.

Graphically, $\eta(p)$ is the value on the measurement axis such that $100p\%$ of the area under the graph of $f(x)$ lies to the left of $\eta(p)$ and $100(1 - p)\%$ lies to the right.
Cumulative Distribution Functions
Cumulative Distribution Functions

\[ p = F(n(p)) \]
Cumulative Distribution Functions

Definition

The median of a continuous distribution, denoted by \( \tilde{\mu} \), is the 50th percentile, so \( \tilde{\mu} \) satisfies \( 0.5 = F(\tilde{\mu}) \). That is, half the area under the density curve is to the left of \( \tilde{\mu} \) and half is to the right of \( \tilde{\mu} \).

e.g. for the continuous rv \( X \) with cdf \( F(x) = \begin{cases} 0 & x \leq 0 \\ 10 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases} \)

the 100 \( p \)th percentile is calculated as following:

\[
p = F(\eta(p)) = 10 \eta(p) - 9 \eta(p)\]

Therefore, the 75th percentile is \( \eta(0.75) \approx 0.9036 \) and the median is \( \eta(0.5) \approx 0.8377 \).
**Definition**

The **median** of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $0.5 = F(\tilde{\mu})$. That is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$. 

**Example**

For the continuous rv $X$ with cdf $F(x) = \begin{cases} 0 & x \leq 0 \\ 0.9 & x < 1 \\ 1 & x \geq 1 \end{cases}$

the $100\text{th}$ percentile is calculated as following:

$$p = F(\eta(p)) = 0.9 - 0.9 \eta(p)$$

Therefore, the $75$th percentile is $\eta(0.75) \approx 0.9036$ and the median is $\eta(0.5) \approx 0.8377$. 
Cumulative Distribution Functions

Definition

The median of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $0.5 = F(\tilde{\mu})$. That is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.

e.g. for the continuous rv $X$ with cdf

\[
F(x) = \begin{cases} 
0 & x \leq 0 \\
10x^9 - 9x^{10} & 0 < x < 1 \\
1 & x \geq 1
\end{cases}
\]

the 100th percentile is calculated as following:

\[
p = F(\eta(p)) = 10\eta(p)^9 - 9\eta(p)^{10}
\]
Definition

The **median** of a continuous distribution, denoted by \( \tilde{\mu} \), is the 50th percentile, so \( \tilde{\mu} \) satisfies \( 0.5 = F(\tilde{\mu}) \). That is, half the area under the density curve is to the left of \( \tilde{\mu} \) and half is to the right of \( \tilde{\mu} \).

e.g. for the continuous rv \( X \) with cdf

\[
F(x) = \begin{cases} 
0 & x \leq 0 \\
10x^9 - 9x^{10} & 0 < x < 1 \\
1 & x \geq 1
\end{cases}
\]

the **100th percentile** is calculated as following:

\[ p = F(\eta(p)) = 10\eta(p)^9 - 9\eta(p)^{10} \]

Therefore, the 75th percentile is \( \eta(0.75) \approx 0.9036 \) and the median is \( \eta(0.5) \approx 0.8377 \).