University of Washington

STATISTICS

# Elements of Statistical Methods

# Discrete & Continuous Random Variables

# (Ch 4-5)

Fritz Scholz

Spring Quarter 2010

April 13, 2010

# Discrete Random Variables

The previous discussion of probability spaces and random variables was completely general. The given examples were rather simplistic, yet still important.

We now widen the scope by discussing two general classes of random variables, discrete and continuous ones.

**Definition:** A random variable $X$ is discrete iff $X(S)$, the set of possible values of $X$, i.e., the range of $X$, is countable.

As a complement to the cdf $F(y) = P(X \leq y)$ we also define the probability mass function (pmf) of $X$

**Definition:** For a discrete random variable $X$ the probability mass function (pmf) is the function $f : R \longrightarrow R$ defined by

$$f(x) = P(X = x) = F(x) - F(x-) \quad \text{the jump of } F \text{ at } x$$

# Properties of pmf's

1) $f(x) \geq 0$ for all $x \in R$

2) If $x \notin X(S)$, then $f(x) = 0$.

3) By definition of $X(S)$ we have

$$\sum_{x \in X(S)} f(x) = \sum_{x \in X(S)} P(X = x) = P\left( \bigcup_{x \in X(S)} \{x\} \right) = P(X \in X(S)) = 1$$

$$\begin{aligned} F(y) &= P(X \leq y) = \sum_{x \in X(S): x \leq y} f(x) = \sum_{x \leq y} f(x) \\ P_X(B) &= P(X \in B) = \sum_{x \in B \cap X(S)} f(x) = \sum_{x \in B} f(x) \end{aligned}$$

Technically the last summations might be problematic, but $f(x) = 0$ for all $\notin X(S)$.

Both $F(X)$ and $f(x)$ characterize the distribution of $X$,

i.e., the distribution of probabilities over the various possible $x$ values,

either by the $f(x)$ values or by the jump sizes of $F(x)$.

# Bernoulli Trials

An experiment is called a Bernoulli trial when it can result in only two possible outcomes, e.g., $\mathtt{H}$ or $\mathtt{T}$, success or failure, etc., with respective probabilities $p$ and $1-p$ for some $p \in [0,1]$.

**Definition:** A random variable $X$ is a Bernoulli r.v. when $X(S) = \{0,1\}$.

Usually we identify $X = 1$ with a “success” and $X = 0$ with a “failure”.

**Example** (coin toss): $X(\mathtt{H}) = 1$ and $X(\mathtt{T}) = 0$ with

$$f(0) = P(X=0) = P(\mathtt{T}) = 0.5 \qquad \text{and} \qquad f(1) = P(X=1) = P(\mathtt{H}) = 0.5$$

and $f(x) = 0$ for $x \notin X(S) = \{0,1\}$.

For a coin spin we might have: $f(0) = 0.7$, $f(1) = 0.3$ and $f(x) = 0$ for all other $x$.

# Geometric Distribution

In a sequence of independent Bernoulli trials (with success probability $p$)

let $Y$ count the number of trials prior to the first success.

$Y$ is called a geometric r.v. and its distribution the geometric distribution.

Let $X_1, X_2, X_3, \ldots$ be the Bernoulli r.v.s associated with the Bernoulli trials.

$$\begin{aligned} f(k) = P(Y = k) &= P(X_1 = 0, \ldots, X_k = 0, X_{k+1} = 1) \\ &= P(X_1 = 0) \cdot \ldots \cdot P(X_k = 0) \cdot P(X_{k+1} = 1) \\ &= (1-p)^k p \qquad \text{for} \quad k = 0, 1, 2, \ldots \end{aligned}$$

To indicate that $Y$ has this distribution we write $Y \sim \text{Geometric}(p)$.

Actually, we are dealing with a whole family of such distributions,

one for each value of the parameter $p \in [0, 1]$.

$$F(k) = P(Y \leq k) = 1 - P(Y > k) = 1 - P(Y \geq k+1) = 1 - (1-p)^{k+1}$$

since $\{Y \geq k+1\} \iff \{X_1 = \ldots = X_{k+1} = 0\}$.

# Application of the Geometric Distribution

By 2000, airlines had reported an annoying number of incidents where the shaft of a 777 engine component would break during engine check tests before take-off.

Before engaging in a shaft diameter redesign it was necessary to get some idea of how likely the shaft would break with the current design.

Based on the supposition that this event occurred about once in ten take-offs it was decided to simulate such checks until a shaft break occurred.

50 such simulated checks after repeated engine shutdowns produced no failure. The probability of this to happen when $p = \frac{1}{10}$ is

$$P(Y \geq 50) = (1-p)^{50} = 1 - P(Y \leq 49) = 1 - \texttt{pgeom}(49, .1) = 0.005154$$

What next? Back to the drawing board.

# What Could Have Gone Wrong?

1. The assumed $p = \frac{1}{10}$ was too high.

   The airlines were not counting the successful tests with equal accuracy?

   Check the number of reported incidents against the number of 777 flight.

2. The airline pilots did this procedure not the same way as the Boeing test pilot.

   Check variation of incident rates between airlines.

3. Does the number of flight cycles influence the incident rate?

   Check the number of cycles on the engines with incidence reports.

4. Are our Bernoulli trials not independent?

   Let the propulsion engineers think about it.

5. And probably more.

# Hypergeometric Distribution

In spite of its hyper name, this distribution is modeled by a very simple urn model.

We randomly grab $k$ balls from an urn containing $m$ red balls and $n$ black balls.

The term grab is used to emphasize selection without replacement.

Let $X$ denote the number of red balls that are selected in this grab.

$X$ is a hypergeometric random variable with a hypergeometric distribution.

If we observe $X = x$ we grabbed $x$ red balls and $k - x$ black balls from the urn.

The following restrictions apply: $0 \le x \le \min(m,k)$ and $0 \le k - x \le \min(n,k)$

i.e., $k - \min(n,k) \le x \le \min(m,k)$ or $\max(0, k-n) \le x \le \min(m,k)$

$$X(S) = \{x \in Z : \max(0, k-n) \le x \le \min(m,k)\}$$

# The Hypergeometric pmf

There are $\binom{m+n}{k}$ possible ways to grab $k$ balls from the $m+n$ in the urn.

There are $\binom{m}{x}$ ways to grab $x$ red balls from the $m$ red balls in the urn

and $\binom{n}{k-x}$ ways to grab $k-x$ black balls from the $n$ black balls in the urn.

By the multiplication principle there are $\binom{m}{x}\binom{n}{k-x}$ ways to grab $x$ red balls and $k-x$ black balls from the urn in a grab of $k$ balls. Thus

$$f(x) = P(X = x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}} = \texttt{dhyper(x,m,n,k)}$$

This is again a whole family of pmf's, parametrized by a triple of integers $(m,n,k)$

with $m,n \geq 0$, $m+n \geq 1$ and $0 \leq k \leq m+n$.

To express the hypergeometric nature of a random variable $X$ we write

$X \sim \text{Hypergeometric}(m,n,k)$. In R: $F(x) = P(X \leq x) = \texttt{phyper(x,m,n,k)}$.

# A Hypergeometric Application

In a famous experiment a lady was tested for her claim that she could tell the difference of milk added to a cup of tea (method 1) & tea added to milk (method 2).

`http://en.wikipedia.org/wiki/The_Lady_Tasting_Tea`

8 cups were randomly split into 4 and 4 with either preparation.

The lady succeeded in identifying all cups correctly.

How do urns relate to cups of tea?

Suppose there is nothing to the lady's claim and that she just randomly picks 4 out of the 8 to identify as method 1, or she picks the same 4 cups as method 1 regardless of the a priori randomization of the tea preparation method.

We either rely on our own randomization or the assumed randomization by the lady.

# Tea Tasting Probabilities

If there is nothing to the claim, we can view this as a random selection urn model, picking $k = 4$ cups (balls) randomly from $m+n = 8$, $\{$● ● ● ● ● ● ● ●$\}$ with $m = 4$ prepared by method 1 (●) and $n = 4$ by method 2 (●).

By random selection alone with $X =$ number of correctly identified method 1 cups (red balls)

$$P(\text{total success}) = P(X=4) = \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}} = \frac{1}{70} = \texttt{dhyper}(4,4,4,4) = 0.0143$$

Randomness alone would surprise us. It induces us to accept some claim validity.

What about weaker performance by chance alone?

$$P(X \geq 3) = P(X=4) + P(X=3) = \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}} + \frac{\binom{4}{3}\binom{4}{1}}{\binom{8}{4}} = \frac{17}{70} = .243$$

no longer so unusual.

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \texttt{phyper}(2,4,4,4) = .243$$

# Allowing for More Misses

The experimental setup on the previous slide did not allow any misses.

Suppose we allow for up to $20\%$ misses, would $m = n = 10$ with $X \geq 8$ provide sufficient inducement to believe that there is some claim validity beyond randomness?

$$P(X \geq 8) = \frac{\binom{10}{10}\binom{10}{0}}{\binom{20}{10}} + \frac{\binom{10}{9}\binom{10}{1}}{\binom{20}{10}} + \frac{\binom{10}{8}\binom{10}{2}}{\binom{20}{10}} = \frac{1 + 10^2 + 45^2}{184756} = \frac{2126}{184756}$$

$$= 1 - P(X \leq 7) = 1 - \texttt{phyper}(7, 10, 10, 10) = 0.0115$$

Thus $X \geq 8$ with $m = n = 10$ would provide sufficient inducement to accept partial claim validity. This is a first taste of statistical reasoning, using probability!

# Chevalier de Mére and Blaise Pascal

Around 1650 the Chevalier de Mére posed a famous problem to Blaise Pascal.

How should a gambling pot be fairly divided for an interrupted dice game.

Players $A$ and $B$ each select a different number from $D = \{1, 2, 3, 4, 5, 6\}$.

For each roll of a fair die the player with his chosen number facing up gets a token.

The player who first accumulates 5 tokens wins the pot of $100\$$.

The game got interrupted with $A$ having four tokens and $B$ just having one.

How should the pot be divided?

# The Fair Value of the Pot

We can ignore all rolls that show neither of the two chosen numbers.

Player $B$ wins only when his number shows up four times

in the first four unignored rolls.

$P($4 successes in a row in 4 fair and independent Bernoulli trials$) = 0.5^4 = .0625$.

A fair allocation of the pot would be $0.0625 \cdot \$100 = \$6.25$ for player $B$

and $\$93.75$ for player $A$.

These amounts can be viewed as probability weighted averages of the two possible

gambling outcomes had the game run its course, namely $\$0.00$ and $\$100$:

$\$6.25 = 0.9375 \cdot \$0.0 + 0.0625 \cdot \$100 \quad \& \quad \$93.75 = 0.0625 \cdot \$0.0 + 0.9375 \cdot \$100$

With that starting position $(4,1)$, had players $A$ and $B$ contributed $\$93.75$ and $\$6.25$

to the pot they could expect to win what they put in. The game would be fair.

# Expectation

This fair value of the game is formalized in the notion of the expected value or expectation of the game pay-off $X$.

**Definition:** For a discrete random variable $X$ the expected value $E(X)$, or simply $EX$, is defined as the probability weighted average of the possible values of $X$,i.e.,

$$EX = \sum_{x \in X(S)} x \cdot P(X = x) = \sum_{x \in X(S)} x \cdot f(x) = \sum_{x} x \cdot f(x)$$

The expected value of $X$ is also called the population mean and denoted by $\mu$.

**Example** (Bernoulli r.v.): If $X \sim$ Bernoulli$(p)$ then

$$\mu = EX = \sum_{x \in \{0,1\}} x \cdot P(X = x) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(X = 1) = p$$

# Expectation and Rational Behavior

The expected value EX plays a very important role in many walks of life:

casino, lottery, insurance, credit card companies, merchants, Social Security.

Each time you have to ask what is the fair value of a payout $X$?

Everybody is trying to make money or cover the cost of running the business.

Read the very entertaining text discussion on pp. 96-99.

It discusses psychological aspects of certain expected value propositions.

# Expectation of $\varphi(X)$

If $X$ is a random variable with pmf $f(x)$, so is $Y = \varphi(X)$ with some pmf $g(y)$.

We could find its expectation by first finding the pmf $g(y)$ of $Y$ and then

$$EY = \sum_{y \in Y(S)} y g(y) = \sum_{y} y g(y)$$

But we have a more direct formula that avoids the intermediate step of finding $g(y)$:

$$EY = E\varphi(X) = \sum_{x \in X(S)} \varphi(x) f(x) = \sum_{x} \varphi(x) f(x)$$

The proof is on the next slide.

# $E\varphi(X) = \sum_x \varphi(x) f(x)$

$$\begin{aligned}\sum_x \varphi(x) f(x) &= \sum_y \sum_{x:\varphi(x)=y} \varphi(x) f(x) = \sum_y \sum_{x:\varphi(x)=y} y f(x) = \sum_y y \sum_{x:\varphi(x)=y} f(x) \\ &= \sum_y y P(\varphi(X) = y) = \sum_y y g(y) = EY = E\varphi(X)\end{aligned}$$

In the first $=$ we just sliced up $X(S)$, the range of $X$, into $x$-slices indexed by $y$.

The slice corresponding to $y$ consists of all $x$'s that get mapped into $y$, i.e., $\varphi(x) = y$.

These slices are mutually exclusive. The same $x$ is not mapped into different $y$'s.

We add up within each slice and use the distributive law $a(b+c) = ab + ac$ at $=$

$$\sum_{x:\varphi(x)=y} \varphi(x) f(x) = \sum_{x:\varphi(x)=y} y f(x) = y \sum_{x:\varphi(x)=y} f(x) = y\, P(\varphi(X) = y) = y g(y)$$

followed by the addition of all these $y$-indexed sums over all slices $\sum_y y g(y) = E(Y)$.

# St. Petersburg Paradox

We toss a fair coin. The jackpot starts a $\$1$ and doubles each time a `Tail` is observed. The game terminates as soon as a `Head` shows up and the current jackpot is paid out.

Let $X$ be the number of tosses prior to game termination. $X \sim$ Geometric($p = 0.5$).

The payoff is $Y = \varphi(X) = 2^X$ (in dollars). The pmf of $X$ is $f(x) = 0.5^x, x = 0, 1, \ldots$.

$$E(Y) = E2^X = \sum_{x=0}^{\infty} 2^x \cdot 0.5^x = \sum_{x=0}^{\infty} 1 = \infty$$

Any finite price to play the game is smaller than its expected value.

Should a rational player play it at any price?

Should a casino offer it at a sufficiently high price?

http://en.wikipedia.org/wiki/St._Petersburg_paradox

http://plato.stanford.edu/entries/paradox-stpetersburg/

# Properties of Expectation

In the following, $X$ is always a discrete random variable.

If $X$ takes on a single value $c$, i.e., $P(X = c) = 1$, then $EX = c$. (obvious)

For any constant $c \in R$ we have

$$E[c\varphi(X)] = \sum_{x \in X(S)} c\varphi(x) f(x) = c \sum_{x \in X(S)} \varphi(x) f(x) = cE[\varphi(X)]$$

In particular

$$E[cY] = cEY$$

For any two r.v.'s $X_1 : S \to R$ and $X_2 : S \to R$ (not necessarily independent) we have

$$E[X_1 + X_2] = EX_1 + EX_2 \qquad \text{and} \qquad E[X_1 - X_2] = EX_1 - EX_2$$

provided $EX_1$ and $EX_2$ are finite. See the proof on the next two slides.

# Joint and Marginal Distributions

A function $X : S \longrightarrow R \times R = R^2$ with values $X(s) = (X_1(s), X_2(s))$ is a discrete random vector if its possible value set (range) is countable.

$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ is the joint pmf of the random vector $(X_1, X_2)$

$$\begin{aligned} \sum_{x_1 \in X_1(S)} f(x_1, x_2) &= \sum_{x_1 \in X_1(S)} P(X_1 = x_1, X_2 = x_2) = \sum_{x_1} P(X_1 = x_1, X_2 = x_2) \\ &= P(X_1 \in X_1(S), X_2 = x_2) = P(X_2 = x_2) = f_{X_2}(x_2) \end{aligned}$$

$f_{X_2}(x_2)$ denotes the marginal pmf of $X_2$ alone.

Similarly,

$$f_{X_1}(x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)$$

is the marginal pmf of $X_1$.

# Proof for $E[X_1+X_2]=EX_1+EX_2$

$$
\begin{aligned}
E[X_1+X_2] &= \sum_{(x_1,x_2)} (x_1+x_2) f(x_1,x_2) = \sum_{x_1}\sum_{x_2} [x_1 f(x_1,x_2) + x_2 f(x_1,x_2)] \\
&= \sum_{x_1}\sum_{x_2} x_1 f(x_1,x_2) + \sum_{x_2}\sum_{x_1} x_2 f(x_1,x_2) \\
&= \sum_{x_1} x_1 \left[\sum_{x_2} f(x_1,x_2)\right] + \sum_{x_2} x_2 \left[\sum_{x_1} f(x_1,x_2)\right] \\
&= \sum_{x_1} x_1 f_{X_1}(x_1) + \sum_{x_2} x_2 f_{X_2}(x_2) = EX_1 + EX_2
\end{aligned}
$$

Note: we just make use of the distributive law $a(b+c)=ab+ac$ at $=$

and that the order of summation does not change the sum at $=$.

The proof for $E[X_1-X_2]=EX_1-EX_2$ proceeds along the same lines.

# $\mu$ as Probability Distribution Center

For $\mu = EX$ we have

$$E[X-\mu] = EX - \mu = \mu - \mu = 0$$

i.e., the expected deviation of $X$ from its expectation (or mean) is zero.

$$E[X-\mu] = \sum_{x<\mu}(x-\mu)f(x) + \sum_{x>\mu}(x-\mu)f(x) = 0$$

$$\sum_{x>\mu}|x-\mu|f(x) = \sum_{x<\mu}|x-\mu|f(x) \quad (*)$$

View the probability mass function $f(x)$ as a distribution of physical masses.

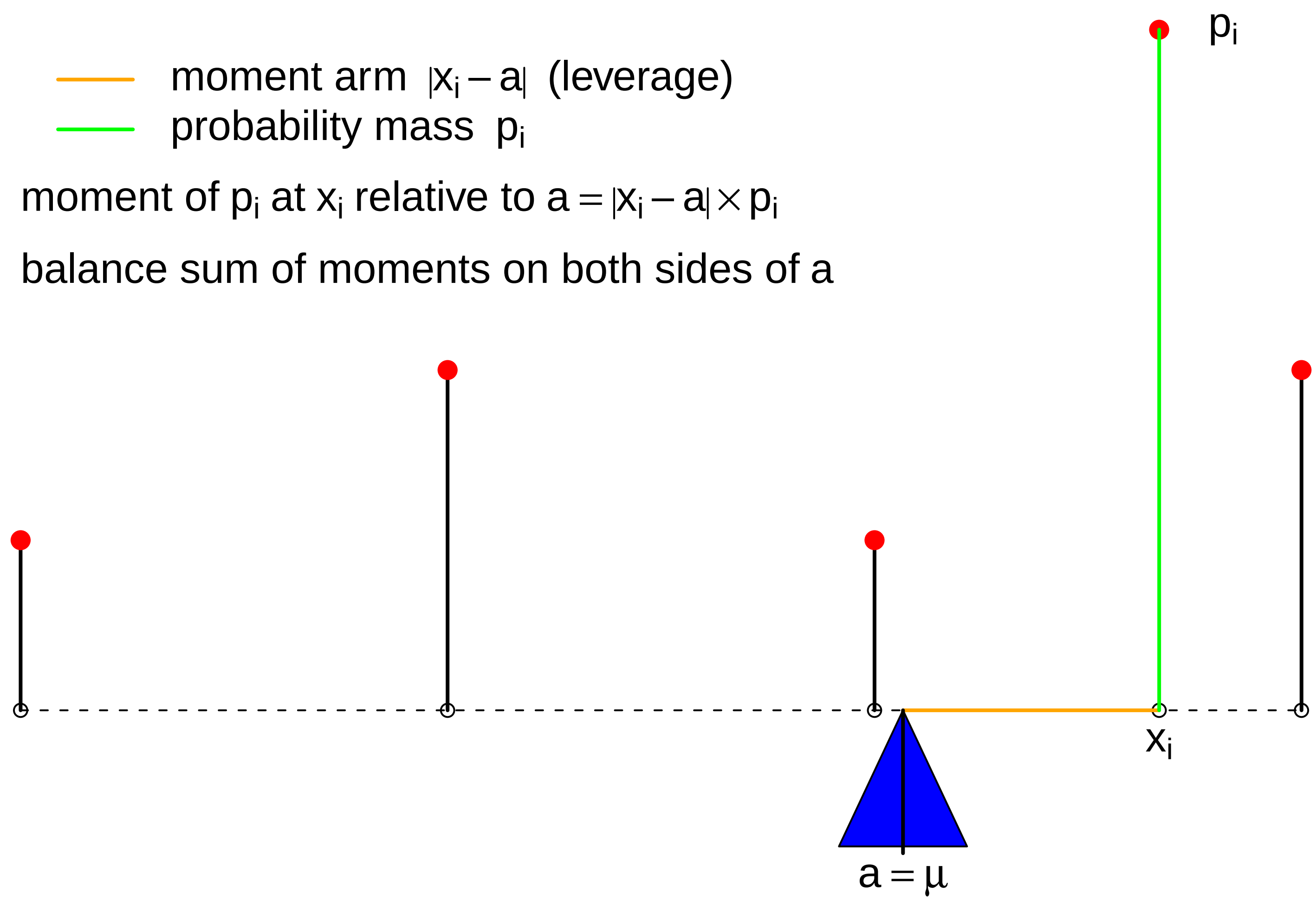
Find its center of "gravity" as that location $a$ for which the sum of the

moments $f(x)|x-a|$ to the right of $a$, balance those on the left of $a$.

$a = \mu$ exactly satisfies this requirement, see $(*)$ and illustration on the next slide.

Thus $\mu$ is viewed as one of the prime characterizations of the distribution center.

# Mean Balancing Act



# The Variance of $X$

For a discrete random variable $X$ with $\mu = EX$ we consider the function

$$\varphi(x) = (x-\mu)^2$$

which expresses the squared deviation of the $X$ values from the mean $\mu$.

$$\mathrm{Var}(X) = \mathrm{Var}\, X = E\varphi(X) = E(X-\mu)^2 = \sum_{x \in X(S)} (x-\mu)^2 f(x)$$

is called the variance of $X$, also called the population variance and denoted by $\sigma^2$.

If $X$ is measured in meters then $\mathrm{Var}\, X$ is measured in units of meters$^2$.

To recover the correct units of measurement from $\mathrm{Var}\, X$ one takes its square root

$$\sqrt{\mathrm{Var}\, X} = \sqrt{\sigma^2} = \sigma$$

and $\sigma$ is called the standard deviation of $X$,

also referred to as population standard deviation.

# The Variance of a Bernoulli Random Variable

If $X \sim \text{Bernoulli}(p)$, then (using $X^2 = X$, since $X = 1$ or $X = 0$ with probability 1)

$$\begin{aligned}
\text{Var}\, X &= E(X-\mu)^2 = E(X-p)^2 = E[X^2 - 2pX + p^2] \\
&= EX^2 + E[-2pX] + E[p^2] = EX^2 - 2pEX + p^2 \\
&= EX - 2p^2 + p^2 = p - p^2 = p(1-p)
\end{aligned}$$

Here we used the addition rule: $E[Y_1 + Y_2 + Y_3] = EY_1 + EY_2 + EY_3$

the expectation of a constant: $Ep^2 = p^2$,

and the constant multiplier rule: $E[-2pX] = -2pEX$.

Note how the variance becomes arbitrarily close to zero as $p \to 0$ or $p \to 1$,

i.e., as $X$ becomes more and more like a constant ($0$ or $1$).

# Alternate Variance Formula

If $X$ is a discrete random variable, then

$$\text{Var } X = EX^2 - (EX)^2 = EX^2 - \mu^2$$

Proof: Using $(a-b)^2 = a^2 - 2ab + c^2$

$$\begin{aligned} \text{Var } X &= E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - E(2\mu X) + E(\mu^2) \\ &= EX^2 - 2\mu EX + \mu^2 = EX^2 - 2\mu^2 + \mu^2 = EX^2 - \mu^2 \end{aligned}$$

# Example Variance Calculation

Suppose we have a discrete r.v. with $X(S) = \{2, 3, 5, 10\}$ with pmf $f(x) = P(X = x) = x/20$. Calculate mean, variance and standard deviation of $X$.

| $x$ | $f(x)$ | $xf(x)$ | $x^2$ | $x^2 f(x)$ |
|---|---|---|---|---|
| 2 | 0.10 | 0.20 | 4 | 0.40 |
| 3 | 0.15 | 0.45 | 9 | 1.35 |
| 5 | 0.25 | 1.25 | 25 | 6.25 |
| 10 | 0.50 | 5.00 | 100 | 50.00 |
| | | 6.90 | | 58.00 |

$EX = 6.90$, $\mathrm{Var}\, X = 58.00 - 6.90^2 = 10.39$, and $\sigma = \sqrt{10.39} = 3.223352$.

# Variance Rules

For any discrete random variable $X$ and constant $c$

$$\text{Var}\,(X+c) = \text{Var}\,X \qquad \text{and} \qquad \text{Var}\,c = 0 \qquad \text{and} \qquad \text{Var}\,(cX) = c^2\text{Var}\,X$$

The first and second (intuitively obvious) follow from

$X+c-E(X+c) = X+c-EX-c = X-EX$ and $\mu = Ec = c$ and $(c-\mu)^2 = 0$,

while

$$\begin{aligned}\text{Var}\,(cX) &= E[cX-E(cX)]^2 = E[cX-cEX]^2 = E[c(X-EX)]^2 \\ &= E[c^2(X-EX)^2] = c^2E(X-EX)^2 = c^2\text{Var}\,X\end{aligned}$$

using expectation properties & $(ab)^2 = a^2b^2$, i.e., invariance of multiplication order.

# $\mathrm{Var}\,(X_1 + X_2)$

If $Y_1$, $Y_2$ are independent discrete random variables, then $E(Y_1 Y_2) = EY_1 \cdot EY_2$.

$$\begin{aligned} E(Y_1 Y_2) &= \sum_{y_1,\,y_2} y_1 y_2 f(y_1, y_2) = \sum_{y_1} \sum_{y_2} y_1 y_2 f_{Y_1}(y_1) f_{Y_2}(y_2) \\ &= \sum_{y_1} y_1 f_{Y_1}(y_1) \sum_{y_2} y_2 f_{Y_2}(y_2) = EY_1 \cdot EY_2 \end{aligned}$$

For independent discrete random variables $X_1$ and $X_2$ we have

$$\mathrm{Var}\,(X_1 + X_2) = \mathrm{Var}\,X_1 + \mathrm{Var}\,X_2$$

With $\mu_1 = EX_1$, $\mu_2 = EX_2$, $Y_1 = X_1 - \mu_1$ and $Y_2 = X_2 - \mu_2$ with $EY_1 = EY_2 = 0$

$$\begin{aligned} \Longrightarrow \quad \mathrm{Var}\,(X_1 + X_2) &= \mathrm{Var}\,(X_1 + X_2 - \mu_1 - \mu_2) = \mathrm{Var}\,(Y_1 + Y_2) = E(Y_1 + Y_2)^2 \\ &= E[Y_1^2 + 2Y_1 Y_2 + Y_2^2] = EY_1^2 + 2E[Y_1 Y_2] + EY_2^2 \\ &= \mathrm{Var}\,Y_1 + EY_1 \cdot EY_2 + \mathrm{Var}\,Y_2 = \mathrm{Var}\,Y_1 + \mathrm{Var}\,Y_2 \\ &= \mathrm{Var}\,X_1 + \mathrm{Var}\,X_2 \end{aligned}$$

# Binomial Random Variables

Let $X_1, \ldots, X_n$ be independent with $X_i \sim \text{Bernoulli}(p)$. Then

$$Y = X_1 + \ldots + X_n = \sum_{i=1}^{n} X_i$$

is called a binomial random variable and we write $Y \sim \text{Binomial}(n; p)$.

$Y = \#$ of successes in $n$ independent Bernoulli trials with success probability $p$.

$$EY = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} p = np$$

$$\text{Var}\, Y = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\, X_i = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

# Binomial Distribution

The binomial random variable $Y$ has pmf

$$f(k) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n$$

We can get the sum $Y = k$ if and only if there are exactly $k$ ones among the sequence $X_1, \ldots, X_n$ and thus exactly $n-k$ zeros. For any such sequence, in particular for $X_1 = \ldots = X_k = 1$ and $X_{k+1} = \ldots = X_n = 0$, we get

$$P(X_1 = \ldots = X_k = 1, X_{k+1} = \ldots = X_n = 0)$$

$$= \overbrace{P(X_1 = 1) \cdot \ldots \cdot P(X_k = 1)}^{k} \cdot \overbrace{P(X_{k+1} = 0) \cdot \ldots \cdot P(X_n = 0)}^{n-k}$$

$$= \underbrace{p \cdot \ldots \cdot p}_{k} \cdot \underbrace{(1-p) \cdot \ldots \cdot (1-p)}_{n-k} = p^k (1-p)^{n-k}$$

For all other such sequences we get $p^k(1-p)^{n-k}$ (different order of $p$ and $1-p$).

There are $\binom{n}{k}$ such sequences with exactly $k$ ones $\Longrightarrow$ above pmf.

# Binomial CDF & R Calculations

As a consequence the binomial cdf is

$$F(k) = P(Y \le k) = \sum_{j=0}^{k} P(Y = j) = \sum_{j=0}^{k} f(j) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

Except for small $n$ the calculation of the pmf and the cdf is very tedious.

Tables used to facilitate such calculations by look-up and interpolation.

R gives for $0 \le k \le n$ (integer), $p \in [0,1]$ and $\gamma = \texttt{gamma} \in [0,1]$

$$f(k) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} = \texttt{dbinom(k,n,p)}$$

and

$$F(k) = P(Y \le k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} = \texttt{pbinom(k,n,p)}$$

$$F^{-1}(\gamma) = \min\{k : \text{ such that } F(k) \ge \gamma\} = \texttt{qbinom(gamma,n,p)}$$

# Example Calculations

When we roll a fair die 600 times we would expect about 100 sixes. Why?

If $Y$ denotes the number of sixes in $n = 600$ rolls, then $Y \sim$ Binomial$(600; 1/6)$.

Thus $EY = np = 100$. Can we be more specific concerning: “about 100 sixes?”

$$f(100) = P(Y = 100) = \texttt{dbinom}(100, 600, 1/6) = 0.04366432$$

not a strong endorsement of our expectation.

$$\begin{aligned} P(90 \le Y \le 110) &= P(Y \le 110) - P(Y \le 89) = F(110) - F(89) \\ &= \texttt{pbinom}(110, 600, 1/6) - \texttt{pbinom}(89, 600, 1/6) = 0.750125 \end{aligned}$$

somewhat reassuring. For more assurance we need to widen the range $[90, 110]$.

$$P(80 \le Y \le 120) = \texttt{pbinom}(120, 600, 1/6) - \texttt{pbinom}(79, 600, 1/6) = 0.9754287$$

with still a $2.5\%$ chance to see a result outside that wide range.

Such chance calculations are at the heart of statistics.

Don’t jump to unwarranted conclusions based on “unexpected’ results.

# Overbooking (critique asumptions)

An airline (say LUFTLOCH) knows from past experience that $10\%$ of its booked business passengers don't show. Thus it sells more seats than are available.

Assume the plane has 180 seats.

How likely is it that too many seat are sold if LUFTLOCH sells 192 seats.

How likely is it that more than 10 seats stay empty?

Let $Y$ be the number of claimed seats. As a first stab assume that each seat claim constitutes an independent Bernoulli trial with claim probability $p = .9$.

We have $n = 192$ trials. The two desired probabilities are

$$P(Y \geq 181) = 1 - P(Y \leq 180) = 1 - \texttt{pbinom}(180, 192, .9) = 0.0255$$

$$P(Y \leq 169) = \texttt{pbinom}(169, 192, .9) = 0.2100696$$

Such probabilities, together with costs of overbooked and empty seats can be used to optimize seat selling strategies (minimize expected losses/maximize profits).

# Continuous Random Variables

Suppose a random experiment could result in any number $X \in [0,1]$,

and suppose we consider all these numbers as equally likely, i.e., $P(X = x) = p$.

$X$ could represent the random fraction of arrival within a bus schedule interval of 30 minutes. $X = 0.6$ means that the arrival was at minute 18 of the interval $[0,30]$.

What is the probability of the event $A = \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4} \ldots\}$?

According to our rules we would get for $p > 0$ the nonsensical

$$P(X \in A) = P(X = 1/2) + P(X = 1/3) + P(X = 1/4) + \ldots = p + p + p + \ldots = \infty$$

Thus we can only assign $p = 0$ in our probability model and we get $P(A) = 0$ for any countable event. The fact that $P(X = x) = 0$ for any $x \in [0,1]$ does not mean that these values $x$ are impossible.

How could we get $P(0.2 \le X \le 0.3)$? $A = [0.2, 0.3]$ has uncountably many points.

# Attempt at a Problem Resolution

Consider the two intervals $(0,0.5)$ and $(0.5,1)$.

Equally likely choices within $[0,1]$ should make these intervals equally probable.

Since $P(X=x)=0$, the intervals $[0,0.5)$ and $[0.5,1]$ should be equally probable.

Since their disjoint union is the full set of possible values for $X$ we conclude that

$$P(X\in[0,0.5))=P(X\in[0.5,1])=\frac{1}{2} \quad \text{since only that way we get} \quad \frac{1}{2}+\frac{1}{2}=1$$

Similarly we can argue

$$P(X\in[0.2,0.3])=\frac{1}{10}=0.3-0.2=0.1$$

$[0,1]$ can be decomposed into 10 adjacent intervals that should all be equally likely.

Going further, the same principle should give for any rational endpoints $a\le b$

$$P(X\in[a,b])=b-a$$

and from there it is just a small technical step ($\Longleftarrow$ countable additivity of $P$) that shows that the same should hold for any $a,b\in[0,1]$ with $a\le b$.

# Do We Have a Resolution?

We started with the intuitve notion of equally likely outcomes $X \in [0,1]$ and what interval probabilities should be if we had a sample space $S$, with a set $\mathcal{C}$ of events and a probability measure $P(A)$ for all events $A \in \mathcal{C}$.  Do we have $(S, \mathcal{C}, P)$?

Take $S = [0,1]$, and let $\mathcal{C}$ be the collection of Borel sets in $[0,1]$,
i.e., the smallest sigma field containing the intervals $[0,a]$ for $0 \le a \le 1$.
To each such interval assign the probability $P([0,a]) = a$.

It can be shown (Caratheodory extension theorem) that this specification is enough to uniquely define a probability measure over all the Borel sets in $\mathcal{C}$,
with the property that $P([a,b]) = b - a$ for all $0 \le a \le b \le 1$.

In this context our previous random variable simply is

$$X : S = [0,1] \longrightarrow R \qquad \text{with} \qquad X(s) = s$$

# The Uniform Distribution over $[0, 1]$

The r.v. $X(s) = s$ defined w.r.t. $(S, \mathcal{C}, P)$ constructed on the previous slide is said to have a continuous uniform distribution on the interval $[0, 1]$, we write $X \sim$ Uniform$[0, 1]$.

Its cdf is easily derived as

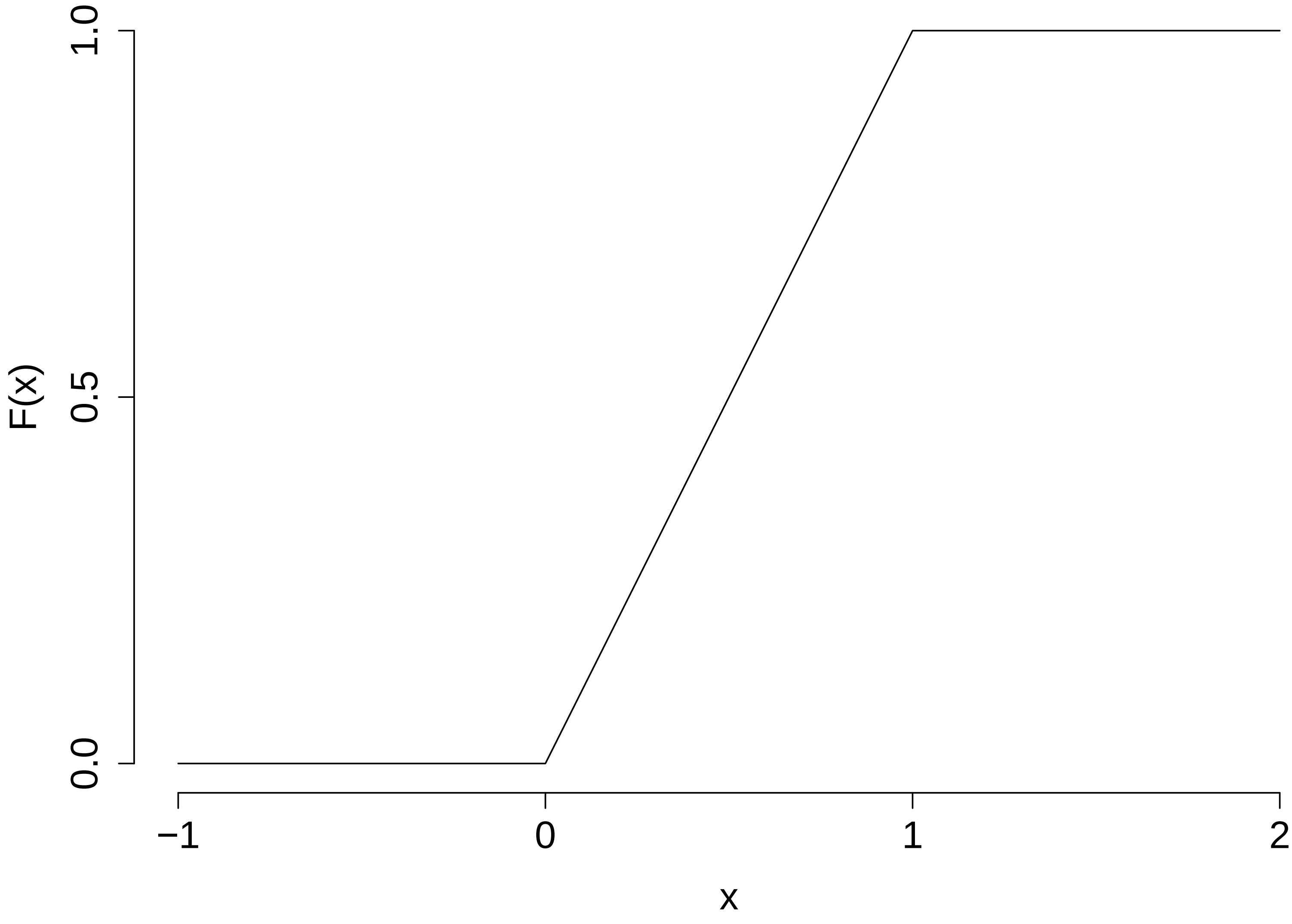
$$
\begin{aligned}
F(x) &= P(X \le x) \\
&= \begin{cases} P(X \in (-\infty, x]) = P(\emptyset) = 0 & \text{for } x < 0 \\ P(X \in (-\infty, 0)) + P(X \in [0, x]) = 0 + (x - 0) = x & \text{for } 0 \le x \le 1 \\ P(X \in (-\infty, 0)) + P(X \in [0, 1]) + P(X \in (1, x]) = 0 + 1 + 0 = 1 & \text{for } x > 1 \end{cases}
\end{aligned}
$$

Its plot is shown on the next slide.

What about a pmf? Previously that was defined as $P(X = x)$, but since that is zero for any $x$ it is not useful.

We need a function $f(x)$ that is useful in calculating interval probabilities.

# CDF of Uniform $[0, 1]$



# Probability Density Function (PDF) for Uniform $[0,1]$

Consider the following function illustrated on the next slide

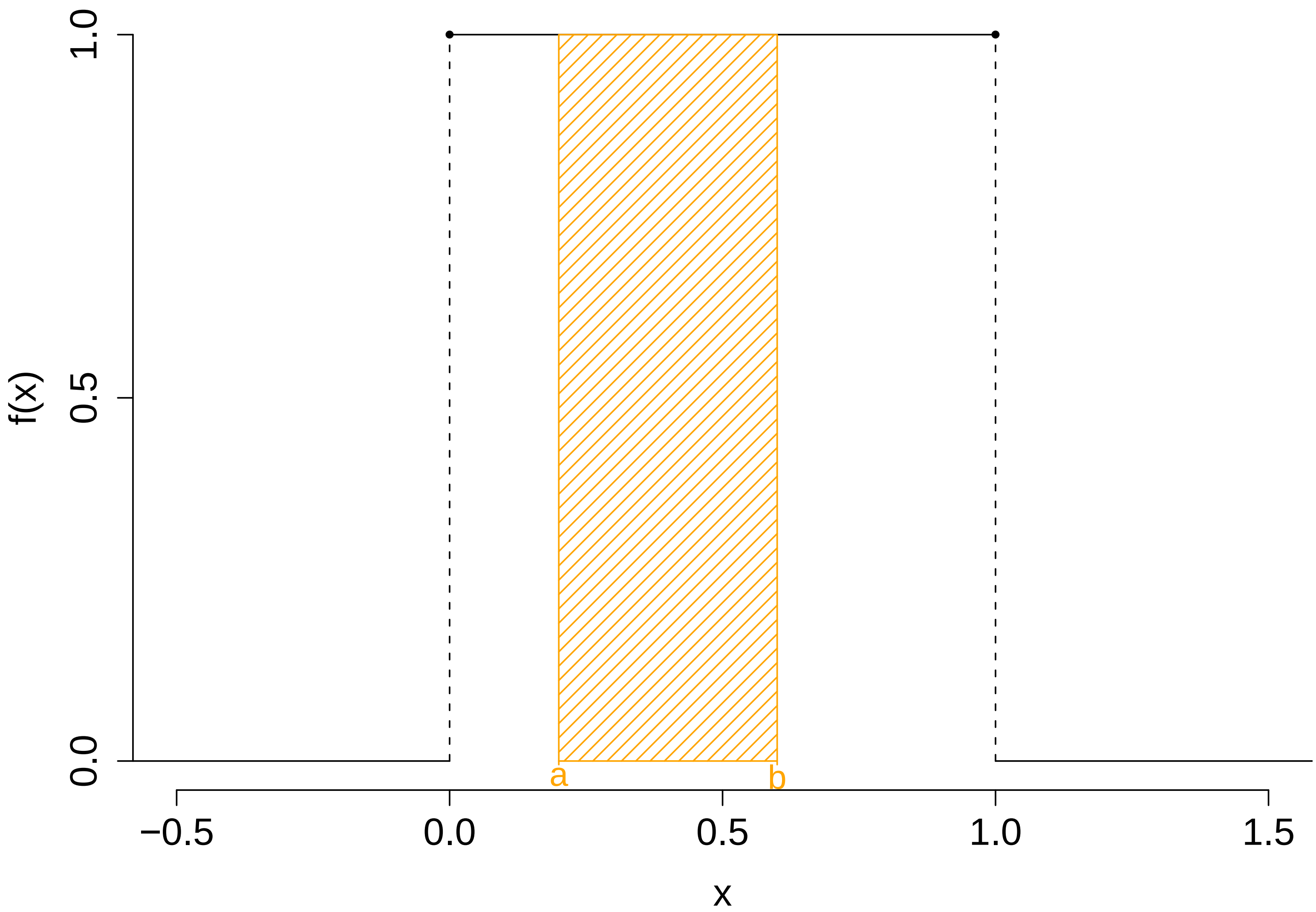
$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \in [0,1] \\ 0 & \text{for } x > 1 \end{cases}$$

Note that $f(x) \equiv 1$ for all $x \in [0,1]$. Thus the rectangle area under $f$ over any interval $[a,b]$ with $0 \leq a \leq b \leq 1$ is just $(b-a) \times 1 = b-a$ (see shaded area in the illustration)

This area is also denoted by $\text{Area}_{[a,b]}(f)$.

This is exactly the probability assigned to such an interval by our Uniform$[0,1]$ random variable: $P(X \in [a,b]) = b-a$.

# Density of Uniform$[0, 1]$



# General Continuous Distributions

We generalize the Uniform$[0,1]$ example as follows:

**Definition:** A probability density function (pdf) $f(x)$ is any function $f: R \longrightarrow R$ such that

1. $f(x) \geq 0$ for all $x \in R$
2. $\text{Area}_{(-\infty,\infty)}(f) = \int_{-\infty}^{\infty} f(x)\, dx = 1$

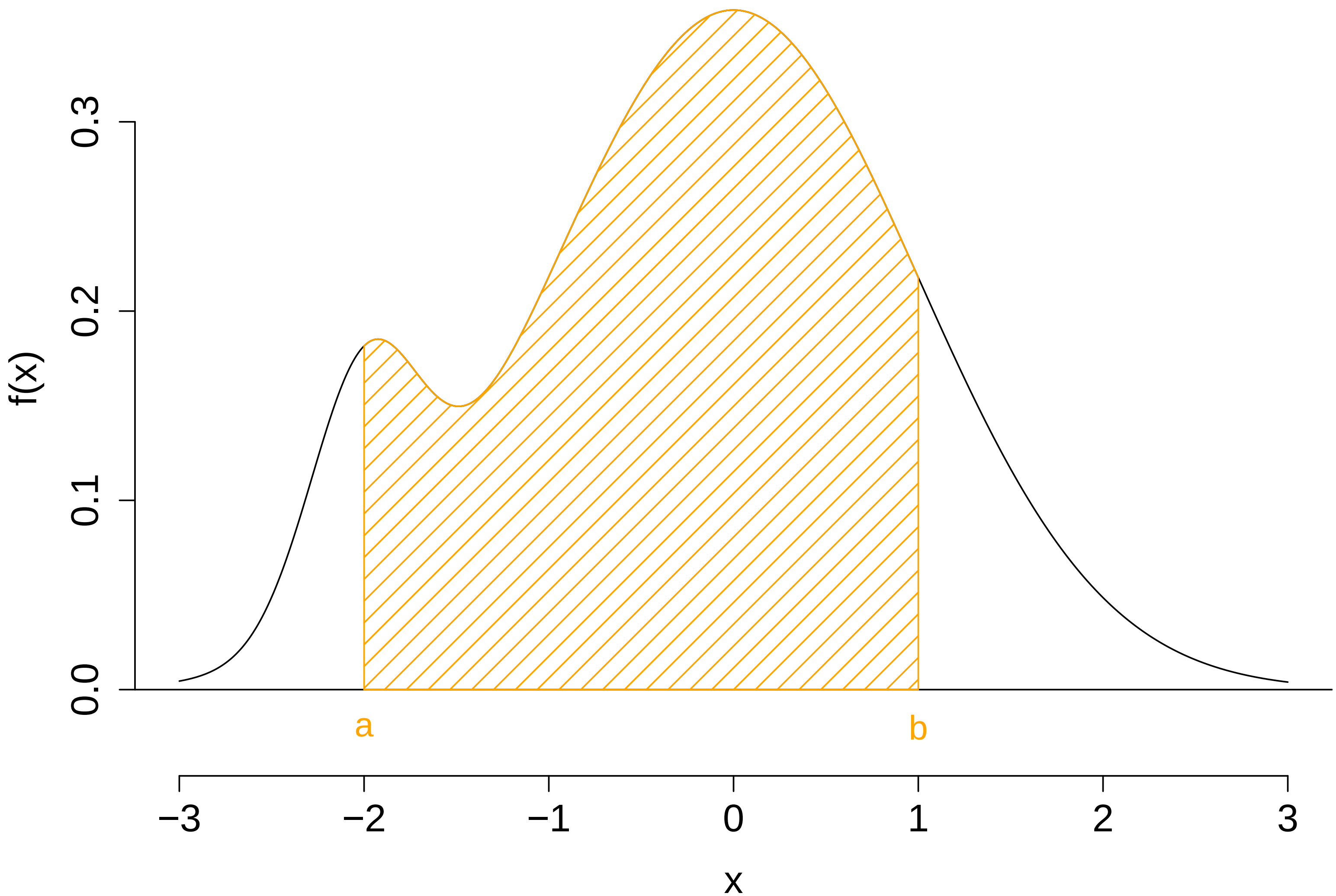
**Definition:** A random variable $X$ is continuous if there is a probability density function $f(x)$ such that for any $a \leq b$ we have

$$P(X \in [a,b]) = \text{Area}_{[a,b]}(f) = \int_a^b f(x)\, dx$$

Such a continuous random variable has cdf

$$F(y) = P(X \leq y) = P(X \in (-\infty, y]) = \text{Area}_{(-\infty,y]}(f) = \int_{-\infty}^{y} f(x)\, dx$$

# General Density (PDF)



# Comments on $\text{Area}_{(-\infty,y]}(f) = \int_{-\infty}^{y} f(x)\,dx$

For those who have had calculus, the above area symbolism is not new.

Calculus provides techniques for calculating such areas for certain functions $f$.

For less tractable functions $f$ numerical approximations will have to suffice,
making use of the fact that $\int_a^b f(x)dx$ stands for $\int$ummation from $a$ to $b$
of many narrow rectangular slivers of height $f(x)$ and base $dx$: $f(x) \cdot dx$.

We will not use calculus. Either use R to calculate areas for certain functions
or use simple geometry (rectangular or triangular areas).

A rectangle with sides $A$ and $B$ has area $A \cdot B$.
A triangle with base $A$ and height $B$ has area $A \cdot B/2$.

# Expectation in the Continuous Case

The expectation of a continuous random variable $X$ with pdf $f(x)$ is defined as

$$\mu = EX = \int_{-\infty}^{\infty} x f(x) dx = \text{Area}_{(-\infty,\infty)}(f(x)x) = \text{Area}_{(-\infty,\infty)}(g)$$

where $g(x) = xf(x)$. We assume that this area exists and is finite.

Since $g(x) = xf(x)$ is typically no longer $\geq 0$ we need to count area under positive portions of $xf(x)$ as positive and areas under negative portions as negative.

Another way of putting this is as follows:

$$EX = \text{Area}_{(0,\infty)}(f(x)x) - \text{Area}_{(-\infty,0)}(f(x)|x|)$$

If $g : R \longrightarrow R$ is a function, then $Y = g(X)$ is a random variable and it can be shown that

$$EY = Eg(X) = \int_{-\infty}^{\infty} g(x) f(x) dx = \text{Area}_{(-\infty,\infty)}(g(x)f(x))$$

again assuming that this area exists and is finite.

# The Discrete & Continuous Case Analogy

discrete case $\quad EX = \sum_x x f(x)$

continuous case $\quad EX = \int_{-\infty}^{\infty} x f(x) dx \approx \sum_x x \cdot (f(x)dx)$

where $f(x)dx =$ area of the narrow rectangle at $x$ with height $f(x)$ and base $dx$.

This narrow rectangle area $=$ the probability of observing $X \in x \pm dx/2$.

Thus in both cases we deal with probability weighted averages of $x$ values.

discrete case $\quad Eg(X) = \sum_x g(x) f(x)$

continuous case $\quad Eg(X) = \int_{-\infty}^{\infty} g(x) f(x) dx \approx \sum_x g(x) \cdot (f(x)dx)$

again both are probability weighted averages of $g(x)$ values.

# The Variance in the Continuous Case

For $g(x) = (x - \mu)^2$ we obtain the variance of the continuous r.v. $X$

$$\text{Var}\, X = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

as the probability weighted average of the squared deviations of the $x$'s from $\mu$.

Again, $\sigma = \sqrt{\text{Var}\, X}$ denotes the standard deviation of $X$.

We will not dwell so much on computing $\mu$ and $\sigma$ for continuous r.v.'s $X$, because we bypass calculus techniques in this course.

The analogy of probability averaged quantities in discrete and continuous case is all that matters.

# A Further Comment on the Continuous Case

It could be argued that most if not all observed random phenomena are intrinsically discrete in nature.

From that point of view the introduction of continuous r.v.'s is just a mathematical artifact that allows a more elegant treatment using calculus ideas.

It provides more elegant notational expressions.

It also avoids the choice of the fine grid of measurements that is most appropriate in any given situation.

# Example: Engine Shutdown

Assume that a particular engine on a two engine airplane, when the pilot is forced to shut it down, will do so equally likely at any point in time during its 8 hour flight.

Given that there is such a shutdown, what is the chance that it will have to be shut down within half an hour of either take-off or landing?

Intuitively, that conditional chance should be $1/8 = 0.125$.

Formally: Let $X =$ time of shutdown in the 8 hour interval. Take as density

$$f(x) = \begin{cases} 0 & \text{for } x \in (-\infty, 0) \\ \frac{1}{8} & \text{for } x \in [0, 8] \\ 0 & \text{for } x \in (8, \infty) \end{cases} \qquad \text{total area under } f \text{ is } 8 \cdot \frac{1}{8} = 1$$

$$P(X \in [0, 0.5] \cup [7.5, 8]) = \text{area under } f \text{ over } [0, 0.5] \cup [7.5, 8] = \frac{1}{2}\frac{1}{8} + \frac{1}{2}\frac{1}{8} = \frac{1}{8}$$

# Example: Engine Shutdown (continued)

Given that both engines are shut down during a given flight (rare), what is the chance that both events happen within half an hour of takeoff or landing?

Assuming independence of the shutdown times $X_1$ and $X_2$

$$\begin{aligned} P(\max(X_1,X_2) \le 0.5 \quad & \cup \quad \min(X_1,X_2) \ge 7.5) \\ & = \quad P(\max(X_1,X_2) \le 0.5) + P(\min(X_1,X_2) \ge 7.5) \\ & = \quad P(X_1 \le 0.5) \cdot P(X_1 \le 0.5) + P(X_1 \ge 7.5) \cdot P(X_2 \ge 7.5) \\ & = \quad \frac{1}{16} \cdot \frac{1}{16} + \frac{1}{16} \cdot \frac{1}{16} = 0.0078125 \end{aligned}$$

On a 737 an engine shutdown occurs about 3 times in 1000 flight hours.

The chance of a shutdown in an 8 hour flight is $8 \cdot 3/1000 = 0.024$,
the chance of two shutdowns is $0.024^2 = 0.000576$, the chance of two shutdowns within half an hour of takeoff of landing is $0.000576 \cdot 0.0078125 = 4.5 \cdot 10^{-6}$.

# Example: Sprinkler System Failure

A warehouse has a sprinkler system. Given that it fails to activate, it may incur such a permanently failed state equally likely at any time $X_2$ during a 12 month period.

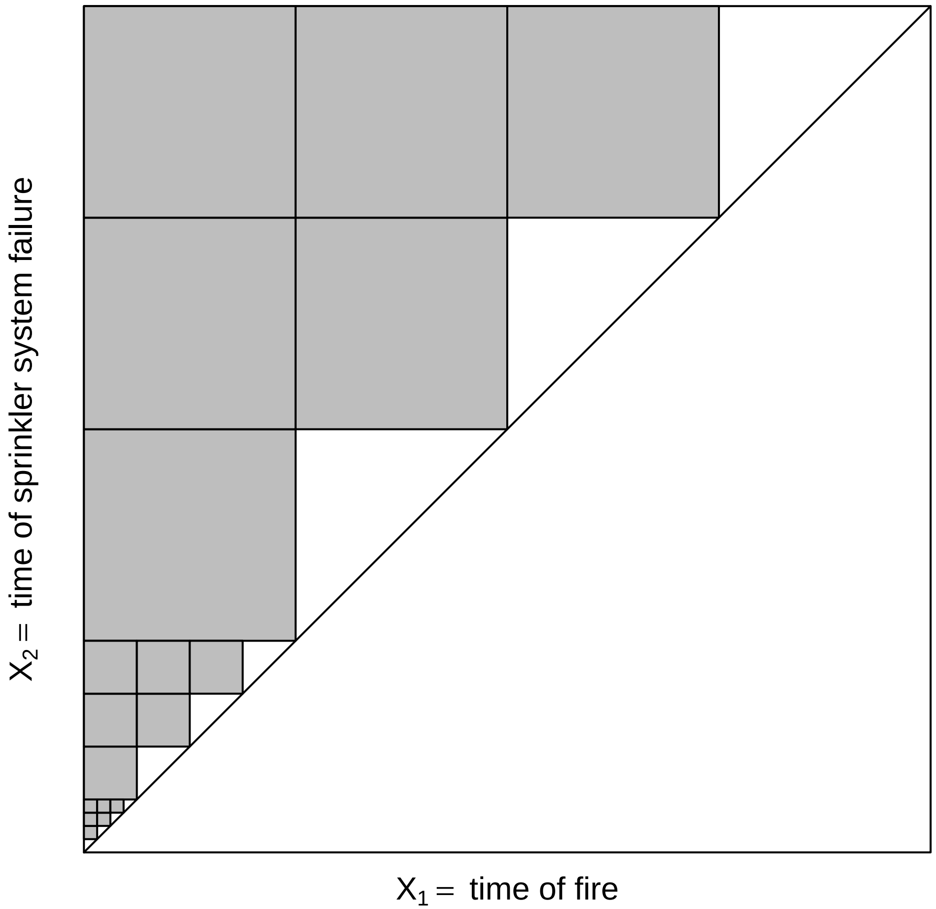
Given that there is a fire during these 12 months, the time of fire can occur equally likely at any time point $X_1$ in that interval (single fire!).

Given that we have a sprinkler system failure and a fire during that year, what is the chance that the sprinkler system will fail before the fire, i.e., will be useless?

Assuming reasonably that the occurrence times $X_1$ and $X_2$ are independent, it seems intuitive that $P(X_1 < X_2) = 0.5$. Rigorous treatment $\Longrightarrow$ next slide.

The same kind of problem arises with a computer hard drive and a backup drive, or with a flight control system and its backup system. (latent failures)

# Example: Sprinkler System Failure (continued)




independence of $X_1$ & $X_2$

$$\Longrightarrow \quad P(X_1 \in [a,b], X_2 \in [c,d])$$
$$= P(X_1 \in [a,b]) \cdot P(X_2 \in [c,d])$$
$$= (b-a) \cdot (d-c) = \text{rectangle area}$$

The upper triangle represents the region where $X_1 < X_2$, i.e., the fire occurs before sprinkler failure.

This triangle region can be approximated by the disjoint union of countably many squares.

Thus $P(X_1 < X_2) = \text{triangle area} = 1/2$.

# Example: Sprinkler System Failure (continued)

Given: there is a one fire and one sprinkler system failure during the year.

Policy: the system is inspected after 6 months and fixed, if found in a failed state. Failure in 0-6 months implies no failure in 6-12 months, due to the "Given".

Given a fire and a failure, the chance that they occur in different halves of the year are $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ for fire in first half and failure in the second half and the same for the other way around. In both cases we are safe, because of the fix in the second case.

The chance that both occur in the first half and in the order $X_1 < X_2$ is $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$, and the same when both occur in the second half in that order.

The chance of staying safely covered is $\frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{3}{4}$.

$\Longrightarrow$ Maintenance checks are beneficial. If the check is done at any other time than at 6 months, the chance of staying safely covered is between $\frac{1}{2}$ and $\frac{3}{4}$ (exercise).

# Example: Waiting for the Bus

Suppose I arrive at a random point in time during the $30$ minute gap between buses. It takes me to another bus stop where I catch a transfer. Again, assume that my arrival there is at a random point of the 30 minute gap between transfer buses.

What is the chance that I waste less than 15 minutes waiting for my buses?

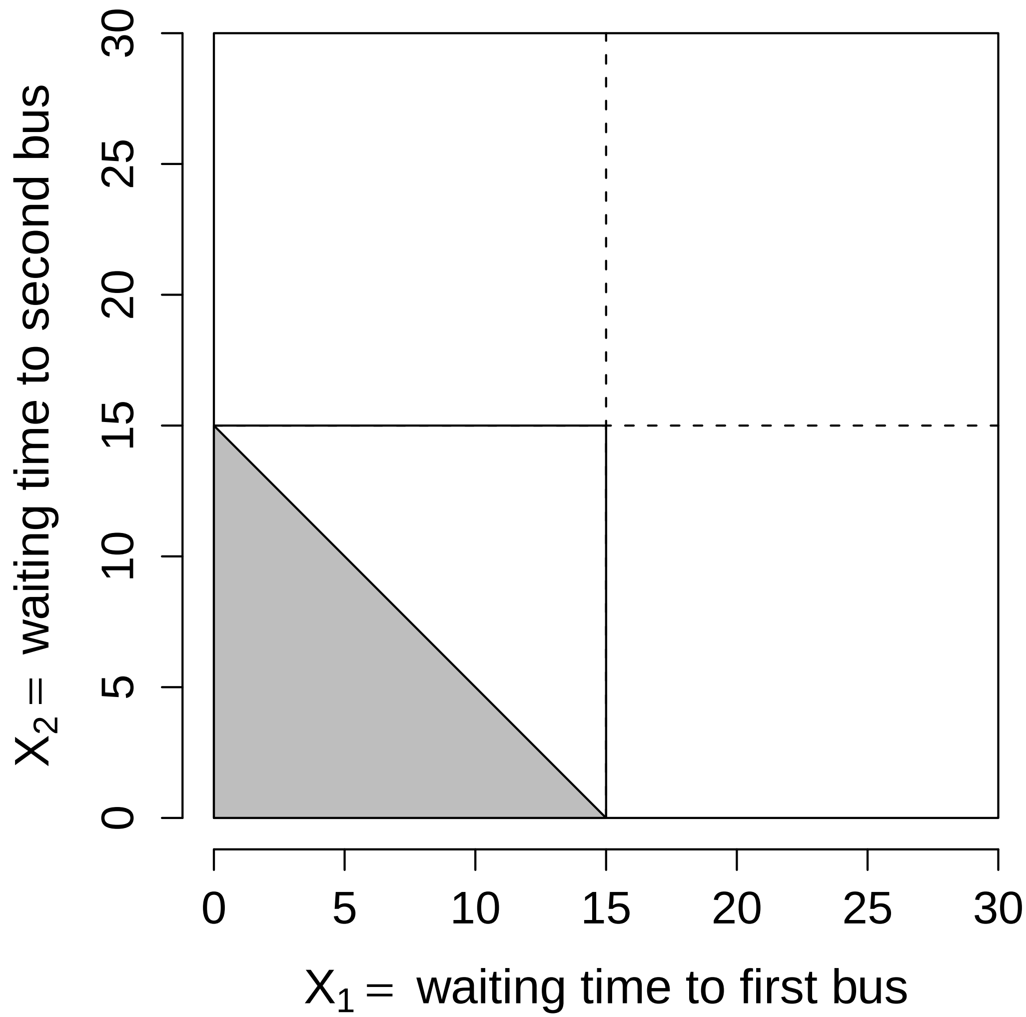
Assume that my waiting times $X_1$ and $X_2$ are independent and $X_i \sim$Uniform$[0,30]$.

$$P(X_1 + X_2 \leq 15) = \frac{1}{2} \cdot P(X_1 \leq 15 \cap X_2 \leq 15) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

See illustration on next slide.

Modification of different schedule gaps, e.g., 20 minutes for the first bus and 30 minutes for the second bus, should be a simple matter of geometry.

# Bus Waiting Time Illustration



# The Normal Distribution

The most important (continuous) distribution in probability and statistics is the normal distribution, also called the Gaussian distribution.

**Definition:** A continuous random variable $X$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, i.e., $X \sim \mathcal{N}(\mu, \sigma^2)$ or $X \sim \text{Normal}(\mu, \sigma^2)$ iff its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad \text{for all } x \in R$$

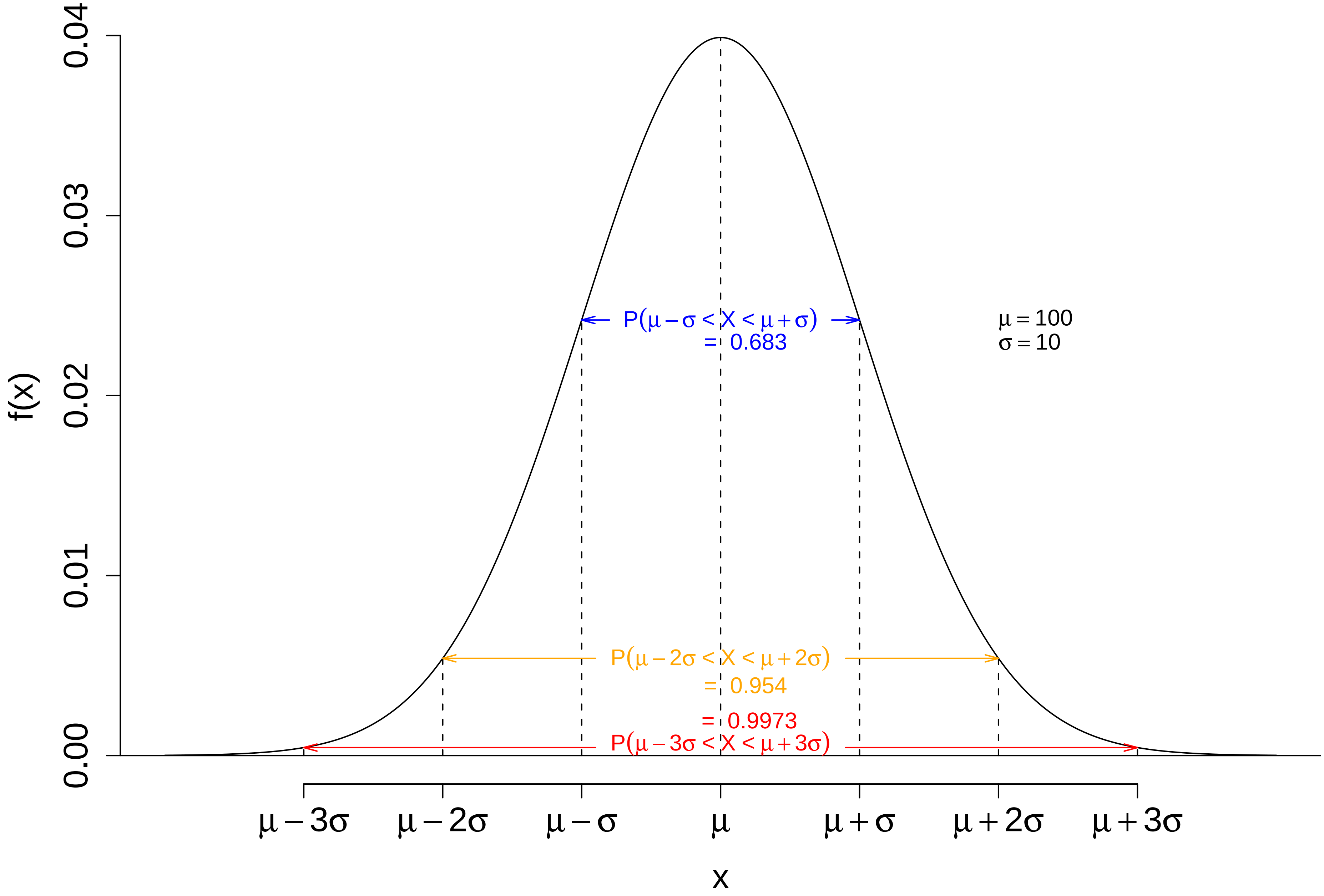
1) $f(x) > 0$ for all $x \in R$. Thus for any $a < b$

$$P(X \in (a,b)) = \text{Area}_{(a,b)}(f) > 0$$

2) $f$ is symmetric around $\mu$, i.e., $f(\mu - x) = f(\mu + x)$ for all $x \in R$.

3) $f$ decreases rapidly as $|x - \mu|$ increases (light tails).

# Normal Density



# Normal Family & Standard Normal Distribution

We have a whole family of normal distributions, indexed by $(\mu, \sigma)$, with $\mu \in R$, $\sigma > 0$.

$\mu = 0$ and $\sigma = 1$ gives us $\mathcal{N}(0,1)$, the standard normal distribution.

**Theorem:** $X \overset{1}{\sim} \mathcal{N}(\mu, \sigma^2) \sim f_X(x) \quad \Longleftrightarrow \quad Z = (X-\mu)/\sigma \overset{2}{\sim} \mathcal{N}(0,1) \sim f_Z(z)$

$Z = (X-\mu)/\sigma$ is called the standardization of $X$, or conversion to standard units.

**Proof:** For very small $\Delta$

$$\begin{array}{ccc} P\left(\mu+\sigma z-\frac{\sigma\Delta}{2} \le X \le \mu+\sigma z+\frac{\sigma\Delta}{2}\right) & = & P\left(z-\frac{\Delta}{2} \le \frac{X-\mu}{\sigma} \le z+\frac{\Delta}{2}\right) \\ 1 \updownarrow \approx & & \approx \updownarrow 2 \\ f_X(\mu+\sigma z)\sigma\Delta & = & f_Z(z)\Delta \end{array}$$

where the probability and density equalities hold by definition.

The loop is closed to arbitrarily close approximation by assuming 1 or 2, q.e.d.

# Using pnorm in R

Introductory statistics texts always used to have a table for $P(Z \le z) = \Phi(z)$.

Now we simply use the R function `pnorm`. For example, for $X \sim \mathcal{N}(1,4)$

$$\begin{aligned} P(X \le 3) &= P\left(\frac{X-\mu}{\sigma} \le \frac{3-\mu}{\sigma}\right) = P\left(Z \le \frac{3-1}{2}\right) \\ &= P(Z \le 1) = \Phi(1) = \texttt{pnorm}(1) = 0.8413447 \\ \text{or} &= \texttt{pnorm}(3, \texttt{mean} = 1, \texttt{sd} = 2) = \texttt{pnorm}(3,1,2) = 0.8413447 \end{aligned}$$

The second usage of `pnorm` uses no standardization, but $\texttt{sd} = \sigma$ is needed.

Standardization is such a fundamental concept, that we emphasize using it.

**Example:** $X \sim \mathcal{N}(4,16)$ find $P(X^2 \ge 36)$

$$\begin{aligned} P(X^2 \ge 36) &= P(X \le -6 \cup X \ge 6) = P(X \le -6) + P(X \ge 6) \\ &= P\left(\frac{X-\mu}{\sigma} \le \frac{-6-\mu}{\sigma}\right) + P\left(\frac{X-\mu}{\sigma} \ge \frac{6-\mu}{\sigma}\right) \\ &= \Phi((-6-4)/4) + (1-\Phi((6-4)/4)) = \Phi(-2.5) + (1-\Phi(0.5)) \\ &= \Phi(-2.5) + \Phi(-0.5) = \texttt{pnorm}(-2.5) + \texttt{pnorm}(-0.5) = 0.3147472 \end{aligned}$$

# Two Important Properties

On the previous slide we used the following identity

$$1-\Phi(z)=\Phi(-z) \quad \text{, i.e.,} \quad \text{Area}_{[z,\infty)}(f_Z)=\text{Area}_{(-\infty,-z]}(f_Z)$$

simply because of the symmetry of the standard normal density around $0$.

If $X_i \sim \mathcal{N}(\mu_i,\sigma_i^2),\ i=1,\ldots,n$, are independent normal random variables, then

$$X_1+\ldots+X_n \sim \mathcal{N}(\mu_1+\ldots+\mu_n,\sigma_1^2+\ldots+\sigma_n^2)$$

While the text states this just for $n=2$,the case $n>2$ follows immediately from that by repeated application, e.g.,

$$X_1+X_2+X_3+X_4 = \overbrace{X_1+\underbrace{(X_2+\overbrace{[X_3+X_4]}^{\text{sum of 2}})}_{\text{sum of 2}}}^{\text{sum of 2}} \quad \text{etc.}$$

# Normal Sampling Distributions

Several distributions arise in the context of sampling from a normal distribution.

These distributions are not so relevant in describing data distributions.

However, they play an important role in describing the random behavior of various quantities (statistics) calculated from normal samples.

For that reason they are called sampling distributions.

We simply give the operational definitions of these distributions and show how to compute respective probabilities in R.

# The Chi-Squared Distributions

When $Z_1, \ldots, Z_n$ are independent standard normal random variables then the continuous random variable

$$Y = Z_1^2 + \ldots + Z_n^2$$

is said to have chi-squared distribution with $n$ degrees of freedom. Note that $Y \geq 0$. We also write $Y \sim \chi^2(n)$.

Since $E(Z_i^2) = \text{Var } Z_i + (EZ_i)^2 = \text{Var } Z_i = 1$ is follows

$$EY = E(Z_1^2 + \ldots + Z_n^2) = EZ_1^2 + \ldots + EZ_n^2 = n$$

One can also show that $\text{Var } Z_i^2 = 2$ so that

$$\text{Var } Y = \text{Var } (Z_1^2 + \ldots + Z_n^2) = \text{Var } Z_1^2 + \ldots + \text{Var } Z_n^2 = 2n$$

The cdf and pdf of $Y$ are given in R by

$$P(Y \leq y) = \texttt{pchisq}(\texttt{y},\texttt{n}) \qquad \text{and} \qquad f_Y(y) = \texttt{dchisq}(\texttt{y},\texttt{n})$$

# Properties of the Chi-Squared Distribution

While generally there is no explicit formula for the cdf and we have to use `pchisq`, for $Y \sim \chi^2(2)$ we have $P(Y \leq y) = 1 - \exp(-y/2)$.

When $Y_1 \sim \chi^2(n_1)$ and $Y_2 \sim \chi^2(n_2)$ are independent chi-squared random variables, it follows that $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

The proof follows immediately from our definition. Let $Z_i, i = 1, \ldots, Z_{n_1}$ and $Z'_1, \ldots, Z'_{n_2}$ denote independent standard normal random variable.

$$\begin{array}{ccccccc} Y_1 & \sim & \chi^2(n_1) & \Longleftrightarrow & Y_1 & = & Z_1^2 + \ldots + Z_{n_1}^2 \\ Y_2 & \sim & \chi^2(n_2) & \Longleftrightarrow & Y_2 & = & (Z'_1)^2 + \ldots + (Z'_{n_2})^2 \\ Y_1 + Y_2 & \sim & \chi^2(n_1 + n_2) & \Longleftrightarrow & Y_1 + Y_2 & = & Z_1^2 + \ldots + Z_{n_1}^2 + (Z'_1)^2 + \ldots + (Z'_{n_2})^2 \end{array}$$

# A Chi-Squared Distribution Application

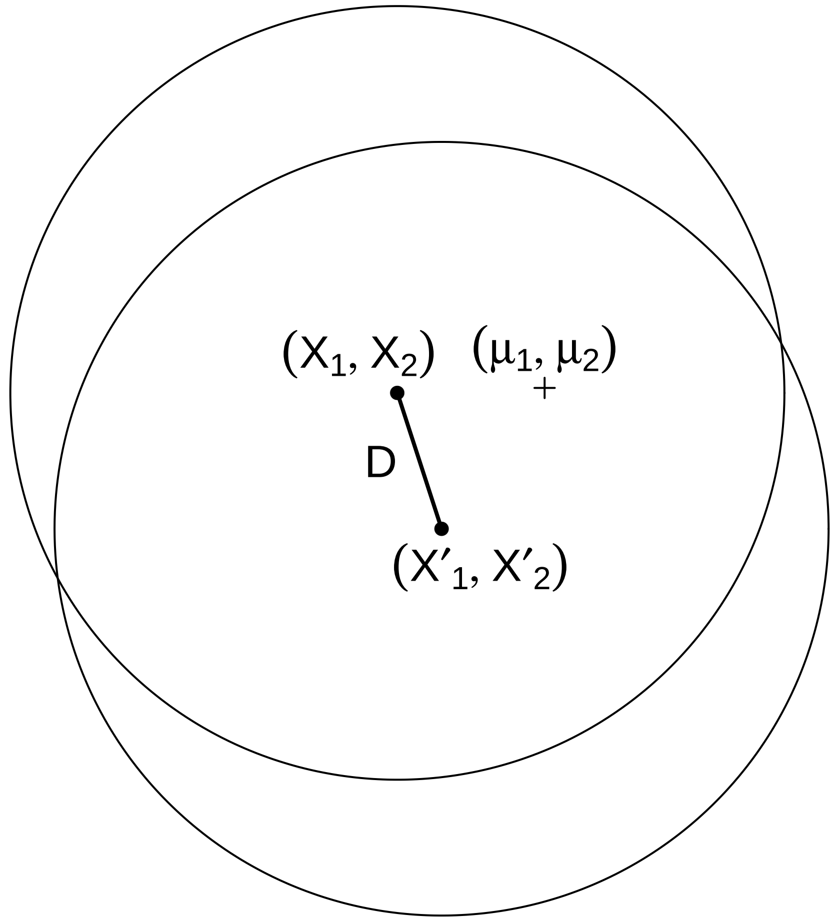
A numerically controlled (NC) machine drills holes in an airplane fuselage panel.
Such holes should match up well with holes of other parts (other panels, stringers) so that riveting the parts together causes no problems.
It is important to understand the capabilities of this process.

In an absolute coordinate system (as used by the NC drill) the target hole has center $(\mu_1, \mu_2) \in R^2$.

The actually drilled center on part 1 is $(X_1, X_2)$, while on part 2 it is $(X_1', X_2')$.

Assume that $X_1, X_1' \sim \mathcal{N}(\mu_1, \sigma^2)$ and $X_2, X_2' \sim \mathcal{N}(\mu_2, \sigma^2)$ are independent.
The respective aiming errors in the perpendicular directions of the coordinate system are $Y_i = X_i - \mu_i$, $Y_i' = X_i' - \mu_i, i = 1, 2$.
$\sigma$ expresses the aiming capability of the NC drill, say it is $\sigma = .01$ inch.
What is the chance that the drilled centers are at most $.05 < \frac{1}{16}$ inches apart, when the parts are aligned on their common nominal center $(\mu_1, \mu_2)$?

# Hole Centers



# Solution

The distance between the drilled hole centers is

$$\begin{aligned} D &= \sqrt{(X_1 - X_1')^2 + (X_2 - X_2')^2} = \sqrt{(X_1 - \mu_1 - (X_1' - \mu_1))^2 + (X_2 - \mu_2 - (X_2' - \mu_2))^2} \\ &= \sqrt{(Y_1 - Y_1')^2 + (Y_2 - Y_2')^2} = \sigma\sqrt{2}\sqrt{\frac{(Y_1 - Y_1')^2}{2\sigma^2} + \frac{(Y_2 - Y_2')^2}{2\sigma^2}} = \sigma\sqrt{2}\sqrt{Z_1^2 + Z_2^2} \end{aligned}$$

$$Y_i - Y_i' \sim \mathcal{N}(0, 2\sigma^2) \Rightarrow Z_i = (Y_i - Y_i')/(\sigma\sqrt{2}) \sim \mathcal{N}(0,1) \Rightarrow V = Z_1^2 + Z_2^2 \sim \chi^2(2).$$

For $d = .05$ and $\sigma = .01$ we get

$$P(D \le d) = P(D^2 \le d^2) = P\left(V \le \frac{d^2}{2\sigma^2}\right) = \texttt{pchisq}(12.5, 2) = 0.9980695$$

We drill 20 holes on both parts, aiming at $(\mu_{1i}, \mu_{2i}), i = 1, \ldots, 20$, respectively.
What is the chance that the maximal hole center distance is at most $.05$?

Assuming independent aiming errors at all hole locations

$$\begin{aligned} P(\max(D_1, \ldots, D_{20}) \le d) &= P(D_1 \le d, \ldots, D_{20} \le d) = P(D_1 \le d) \cdot \ldots \cdot P(D_{20} \le d) \\ &= 0.9980695^{20} = 0.96209 \end{aligned}$$

# Student's $t$ Distribution

When $Z \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(\nu)$ are independent random variables, then

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

is said to have a Student's $t$ distribution with $\nu$ degrees of freedom.

We denote this distribution by $t(\nu)$ and write $T \sim t(\nu)$.

$T$ and $-T$ have the same distribution, since $Z$ and $-Z$ are $\sim \mathcal{N}(0,1)$,

i.e., the $t$ distribution is symmetric around zero.

For large $\nu$ (say $\nu \geq 40$) $t(\nu) \approx \mathcal{N}(0,1)$.

R lets us evaluate the cdf $F(x) = P(T \leq x)$ and pdf $f(x)$ of $t(k)$ via

$$F(x) = \texttt{pt}(\texttt{x},\texttt{k}) \qquad \text{and} \qquad f(x) = \texttt{dt}(\texttt{x},\texttt{k})$$

for example: $\texttt{pt}(2,5) = 0.9490303$

# The $F$ Distribution

Let $Y_1 \sim \chi^2(\nu_1)$ and $Y_2 \sim \chi^2(\nu_2)$ be independent chi-squared r.v.'s, then

$$F = \frac{Y_1/\nu_1}{Y_2/\nu_2}$$

has an $F$ distribution with $\nu_1$ and $\nu_2$ degrees of freedom and we write $F \sim F(\nu_1, \nu_2)$.

Note that $$F = \frac{Y_1/\nu_1}{Y_2/\nu_2} \sim F(\nu_1, \nu_2) \implies \frac{1}{F} = \frac{Y_2/\nu_2}{Y_1/\nu_1} \sim F(\nu_2, \nu_1)$$

Also, with $Z \sim \mathcal{N}(0,1)$ and $Y \sim \chi^2(\nu)$ we have $Z^2 \sim \chi^2(1)$ and thus

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu) \implies T^2 = \frac{Z^2/1}{Y/\nu} \sim F(1, \nu)$$

R lets us evaluate the cdf $F(x) = P(F \leq x)$ and pdf $f(x)$ of $F(k1, k2)$ via

$$F(x) = \mathtt{pf}(\mathtt{x}, \mathtt{k1}, \mathtt{k2}) \qquad \text{and} \qquad f(x) = \mathtt{df}(\mathtt{x}, \mathtt{k1}, \mathtt{k2})$$

If $F \sim F(2, 27)$ then $P(F \geq 2.5) = 1 - P(F \leq 2.5) = 1 - \mathtt{pf}(2.5, 2, 27) = 0.1008988$