Probabilities are calculated as areas under the density curve
• The distribution function supplies cumulative probabilities
• The density function of a continuous random variable $x$ is denoted by $f(x)$
• The distribution function by $F(x)$.
• $P(x_l < X < x_u) = F(x_u) - F(x_l)$

$P(x_l < X < x_u)$ for continuous r.v.'s - 1

$P(x_l < X < x_u)$ for continuous r.v.'s - 2

$P(x_l < X < x_u)$ for continuous r.v.'s - 3

$P(x_l < X < x_u)$ for continuous r.v.'s - 4
• Note that the probability of an event being exact, rather than falling an interval, is zero since, $Pr(X = a) = Pr(a < X < a) = F(a) - F(a) = 0$.

• A discrete random variable had probability mass concentrated at each possible discrete number and the notion of $P(X = a)$ is defined for discrete random variables.

• Where we are specifically interested in $X$ falling in a narrow interval $(a - \delta, a + \delta)$, the probability is given by $Pr ((a - \delta) < X < (a + \delta)) = F(a + \delta) - F(a - \delta)$.

Determination of tail probabilities for discrete random variables required consideration of the events $(X \leq x)$, $(X > x)$, $(X < x)$ and $(X \geq x)$.

Whilst it is necessary to also carefully define the event with continuous random variables, there is not the issue of $X \leq x$ or $X \geq x$ since $P(X = x) = 0$. Hence, $P(X > x) = 1 - P(X < x)$. 
Lecture 11  Expected values, Variances

- $E(X) = \sum_{i=1}^{N} x_i P(X = x_i)$
- $E(g(X)) = \sum_{i=1}^{N} g(x_i) P(X = x_i)$
- $E(X^2) = \sum_{i=1}^{N} x_i^2 P(X = x_i)$

Taking expectations is a process where the random component is averaged out and the expected value of a random variable is a constant, i.e. the expected value is not random.

**Expectations**

- The expected value of a constant is that constant,
  \[ E(aX) = aE(X) \] \hspace{1cm} (11.1)
- The expected value of a constant times a random variable is that constant times the expected value of the random variable,
  \[ E(aX) = aE(X) \] \hspace{1cm} (11.2)

It follows that
\[
E((aX)^2) = E(a^2X^2) = a^2E(X^2)
\]

\[ E() \text{ is additive} \]

\[
E(X_1 + X_2) = E(X_1) + E(X_2) \quad (11.3)
\]

\[
E((X_1 + X_2)^2) = E(X_1^2 + X_2^2 + 2X_1X_2) \quad (11.4)
\]

\[
= E(X_1^2) + E(X_2^2) + 2E(X_1X_2) \quad (11.5)
\]

**Mean**

\[ \mu = E(X). \]

Since we do not measure the population, $\mu$ and $P(X = x)$ are unknown and we rely on the information in the sample.

- The mean of a sample is the average of the observations,
  \[
  \bar{X} = \frac{1}{n} \times \sum_{i=1}^{n} X_i \quad (11.6)
  \]
  \[
  = \sum_{i=1}^{n} \frac{1}{n} X_i
  \]
  \[
  = \sum_{i=1}^{n} w_i X_i \text{ where } w_i = \frac{1}{n} \quad (11.7)
  \]
• The population variance for discrete random variables is defined,
\[ \sigma^2 = \text{var}(X) = E \left( (X - \mu)^2 \right) = \sum_x (X - \mu)^2 P(X = x) \] (11.8)

1

• The sample variance is
\[ s^2 = \text{var}(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \] (11.9)

\[ = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} \right) \] (11.10)


• If \( X_1 \) and \( X_2 \) are independent,
\[ \text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2) \] (11.11)
\[ \text{var}(X_1 - X_2) = \text{var}(X_1) + \text{var}(X_2) \] (11.12)

• The variance of a constant is zero.
\[ \text{var}(b) = E(b^2) - E(b)^2 = 0 = b^2 - (b)^2 \]

• The variance of a constant times a random variable
\[ \text{var}(aX) = a^2 \text{var}(X) \] (11.13)

• A simple extension is
\[ \text{var}(aX + b) = a^2 \text{var}(X) \]

since \( \text{var}(b) = 0. \)

Degrees of Freedom

\(^1\sigma \) is the lowercase of the Greek letter “sigma”.
A sample of size $n$ has $n$ bits of information or more correctly $n$ degrees of freedom.

Calculation of $s^2$ first requires that $\bar{X}$ be extracted from the data, using 1 degree of freedom (df).

Having already used 1 df for the sample mean, there are $(n - 1)$ df remaining for the sample variance.

The degrees of freedom (df) are the number of values that are free to vary after a statistic is computed from a set of values.

If the mean of 10 values is 3, then 9 of the 10 values are free to vary. Once 9 values have been selected, the 10th number is known.

It must be the number such that the sum of all the numbers is $10 \times 3 = 30$.

Thus the degrees of freedom are $10 - 1 = 9$.

Example

Suppose that $X$ is equally likely to take on values 1, 2, 3, 4.

Compute (a) $E(X)$ and (b) $\text{var}(X)$.

Equally likely means that $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = \frac{1}{4}$.

\[
E(X) = \sum_{i=1}^{4} x_i P(X = x_i) \\
= \frac{1}{4} (1 + 2 + 3 + 4) = 2.5 \\
\text{var}(X) = \frac{1}{3} \left( \sum_{i=1}^{4} X_i^2 - \frac{(\sum_{i=1}^{4} X_i)^2}{4} \right) \\
= \frac{1}{3} \left( (1)^2 + (2)^2 + (3)^2 + (4)^2 - \frac{10^2}{4} \right) = \frac{5}{3}
\]

Standard Deviation

The variance is a squared term making it $> 0$.

The statistic which expresses the randomness or spread on the same scale as the data $X$ is the standard deviation

\[
s = \sqrt{\text{sample variance}}
\]

Standard Error

We use the standard deviation (sometimes abreviated to s.d.) to express the randomness associated with a single observation.

If a mean is calculated from a set of data, more information is used and hence it will be known with better precision than for a single observation.

Variance of the mean

For independent observations,

\[
\text{var}(\bar{X}) = \frac{\sigma^2}{n}
\]

1. For each observation, $X_i$, $\text{var}(X_i) = \text{var}(X_j) = \sigma^2$
2. For independent observations,
\[ \text{var}(X_1 + X_2 + \ldots + X_n) = \sigma^2 + \sigma^2 + \ldots = n \times \sigma^2 \]

3.
\[ \text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{var}(X_i) \]
\[ = \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n} \]

The term used to measure the randomness of a mean is **standard error** which is
\[ \text{s.e.} = \sqrt{\text{variance of a mean}} \]
\[ = \sqrt{\frac{\sigma^2}{n}} \]
\[ = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \frac{s}{\sqrt{n}} \]

Note that se(\(\bar{Y}\)) reads as “standard error of \(\bar{Y}\)”.

**Simulations of var(\(\bar{X}\)).**

- Simulate 12 data sets of sample size 100.
- Extract the mean from each
- Calculate the variance of the means

<table>
<thead>
<tr>
<th>sample</th>
<th>(X)</th>
<th>var((X))</th>
<th>var((X))/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.11</td>
<td>4.48</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>3.12</td>
<td>4.29</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>3.21</td>
<td>4.99</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>3.58</td>
<td>6.64</td>
<td>0.07</td>
</tr>
<tr>
<td>5</td>
<td>3.18</td>
<td>4.33</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>3.16</td>
<td>5.77</td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>2.88</td>
<td>4.12</td>
<td>0.04</td>
</tr>
<tr>
<td>8</td>
<td>2.99</td>
<td>6.75</td>
<td>0.07</td>
</tr>
<tr>
<td>9</td>
<td>3.51</td>
<td>7.08</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>2.96</td>
<td>5.29</td>
<td>0.05</td>
</tr>
<tr>
<td>11</td>
<td>2.97</td>
<td>4.42</td>
<td>0.04</td>
</tr>
<tr>
<td>12</td>
<td>2.83</td>
<td>4.73</td>
<td>0.05</td>
</tr>
</tbody>
</table>
# Simulate variance of the mean

```r
options(digits=2)
nsim <- 12; nsamp <- 100
mnx <- varx <- numeric(length=0)
for (i in 1:nsim){
x <- rgamma(shape=2,scale=1.5,n=nsamp)
mnx <- c(mnx,mean(x))
varx <- c(varx,var(x))  }

print( cbind(mnx,varx,varx/nsamp) )
print( var(mnx) )
```

### Summary of results

<table>
<thead>
<tr>
<th>Discrete</th>
<th>equation no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( E(X) )</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( E(X - \mu)^2 )</td>
</tr>
</tbody>
</table>

\[
\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i \\
\hat{\sigma}^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n} \right) \\
\]

\[
E(a) = a \\
E(aX) = aE(X) \\
E(X_1 + X_2) = E(X_1) + E(X_2) \\
E(X_1 - X_2) = E(X_1) - E(X_2) \\
\]

\[
\text{var}(a) = 0 \\
\text{var}(aX) = a^2 \text{var}(X) \\
\text{var}(X_1 \pm X_2) = \text{var}(X_1) + \text{var}(X_2) \quad \text{if } X_1, X_2 \text{ independent} \\
\text{var}(X + a) = \text{var}(X) \\
\text{s.d.}(X) = \sigma \\
\text{var}(\bar{X}) = \frac{\sigma^2}{n} \\
\text{s.e.}(\bar{X}) = \frac{s}{\sqrt{n}} \\
\]

### Example E(aX)

Example 2.3 concerned calculation of body mass index using \( n = 6 \) metric measurements of weight (in kgs) and height (in metres). The conversion factors to their imperial counterparts are:

\[
\begin{align*}
\text{lbs} & \approx 2.2 \times \text{kgs} \\
\text{inches} & \approx 39.37 \times \text{metres}
\end{align*}
\]

If the mean weight in kilograms is 74kgs, what is the mean weight in lbs?

\[
E(\text{lbs}) = E(2.2 \times \text{kgs}) = 2.2E(\text{kgs}) = 2.2 \times 74 = 164
\]

The variance of weight on the kilogram scale is 238. What is the variance on the lbs scale?

\[
\text{var(lbs)} = \text{var}(2.2 \times \text{kgs}) = (2.2)^2 \text{var(\text{kgs})} = 4.84 \times 238 = 1151
\]

What is the standard error of the mean weight in kgs?

\[
\text{s.e.}(\text{kgs}) = \frac{s}{\sqrt{n}} = \frac{\sqrt{238}}{\sqrt{6}} = 6.3
\]
Lecture 12  The normal distribution

• The data-based curve
• Normal density curve which is calculated from the mean and the standard deviation
• Normal curve is a very close approximation to the “data-based curve”.
• If the approximation is satisfactory, calculations from the normal density curve are more straightforward
• The imperfections are of a smaller order of magnitude than the errors themselves and such discrepancies will not impact severely on the interpretation.

\[ \phi_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\} \]

(12.1)

where \( \mu \) specifies the centre of the curve and \( \sigma \) regulates the spread of the curve.

In Rcmdr,
Distributions → Continuous distributions → Normal distribution → Plot normal distribution.

Enter mean, standard deviation, quantiles (which are variables) in the menu. Check Plot density function.

• A general form for a density which depends upon \( \mu \) and \( \sigma \).
• The type of density will apply to a vast number of data sets but each data set will have its own specific value of \( \mu \) and of \( \sigma \).
• The terms \( \mu \) and \( \sigma \) are called parameters and the information about the density is contained in these parameters.
Because the density can be derived entirely from the parameters, it is called a *parametric* density function.

if the pattern of randomness is reliably approximated by a normal distribution, the information on randomness can be condensed into 2 numbers.

For the Millikan data, the sample estimates are:

\[
\bar{X} = 4.781 \quad s = 0.015
\]

\[
\phi(x) = \frac{1}{\sqrt{2\pi} s} \exp \left\{ -\frac{1}{2} \left( \frac{x - 4.781}{0.0151} \right)^2 \right\}
\]

Calculating normal distribution probabilities

\[
\begin{align*}
P(X < \mu - 3\sigma) &= P(X < 4.78 - 3 \times 0.015) = P(X < 4.74) \\
P(X > \mu + 2\sigma) &= P(X > 4.78 + 2 \times 0.015) = P(X > 4.81) \\
P(\mu - \sigma < X < \mu + \sigma) &= P(4.77 < X < 4.80) = P(X < 4.80) - P(X < 4.77)
\end{align*}
\]
\begin{align*}
P_1 & = 0.0014, \quad P_2 = 0.023, \quad P_3 = 0.68. \\
\text{Rcmdr menus:-} \\
\text{Distributions \rightarrow Continuous distributions \rightarrow Normal distribution \rightarrow Normal probabilities.} \\
\text{Enter mean, standard deviation, quantiles (which are variables) in the menu.}
\end{align*}

\textbf{Standard Normal distribution}

Suppose that \( X \sim N(\mu, \sigma^2) \) and

\[
Z = \frac{X - \mu}{\sigma}
\tag{12.2}
\]

Recall that for a variable \( X \) and constants \( a, b \),

\[
E(aX) = aE(X) \\
E(b) = b \\
E(X_1 + X_2) = E(X_1) + E(X_2)
\]

Hence if \( E(X) = \mu, \) \( E(Z) = \frac{E(X) - \mu}{\sigma} = 0. \)

\[
\text{var}(aX) = a^2 \text{var}(X) \\
\text{var}(b) = 0 \\
\text{var}(X + c) = \text{var}(X)
\]

For \( \text{var}(X) = \sigma^2 \), \( \text{var}(Z) = \frac{1}{\sigma^2} \times \text{var}(X) + 0 = 1. \)

Consider the Milikan data where \( \bar{x} = 4.781 \) and \( s = 0.0151. \)

For say \( X = 4.74, \) \( Z = \frac{4.74 - 4.781}{0.015} = -2.67 \)

\begin{verbatim}
> pnorm(q=4.74,mean=4.78,sd=0.015)
[1] 0.0038
> pnorm(q=-2.67,mean=0,sd=1)
[1] 0.0038
> 
\end{verbatim}