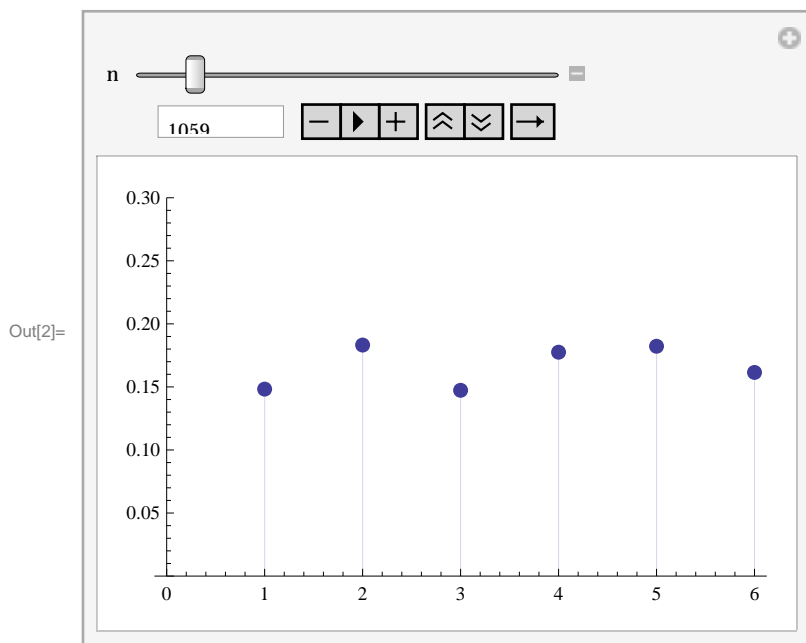


## Beispiele empirischer Verteilungen:

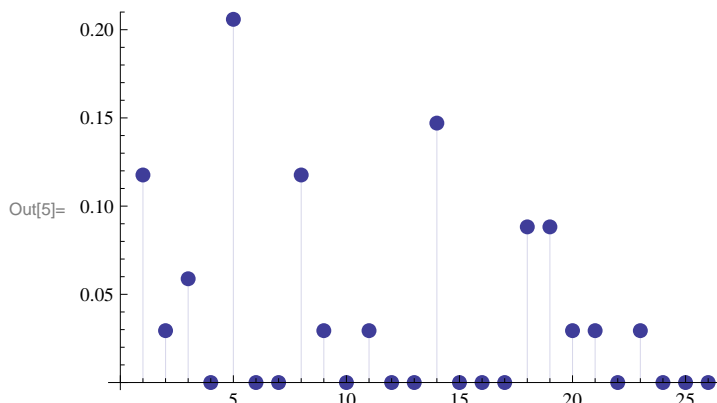
### ■ Empirische Verteilung von n Zufallszahlen aus {1,2,3,4,5,6}

```
In[1]:= x = RandomChoice[{1, 2, 3, 4, 5, 6}, 10 000];  
Manipulate[ListPlot[BinCounts[x[[1 ;; n]], {1, 7, 1}] / n, Filling -> Axis,  
PlotRange -> {0, 0.3}, PlotStyle -> PointSize[Large]], {n, 1, 10 000, 1}]
```



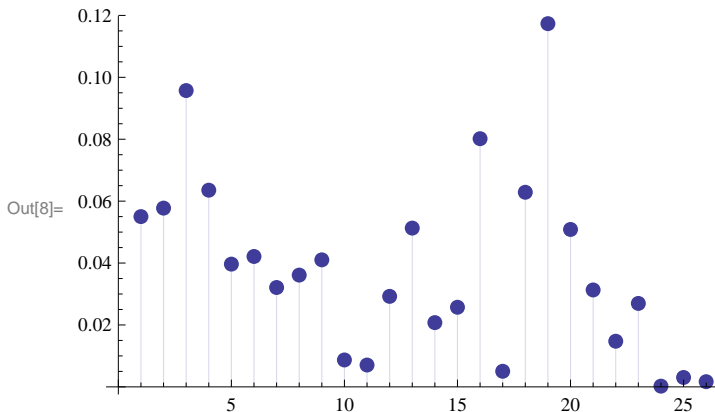
### ■ Empirische Verteilung der Buchstaben "a" bis "z" in dem Wort "Eisenbahnschrankenwaerterhaeuschen"

```
In[3]:= freq =  
StringCount["eisenbahnschrankenwaerterhaeuschen", #] & /@ CharacterRange["a", "z"];  
relfreq = freq / Total[freq];  
ListPlot[relfreq, Filling -> Axis, PlotStyle -> PointSize[Large]]
```



### ■ Empirische Verteilung der Buchstaben "a" bis "z" in einem englischen Wörterbuch

```
In[6]:= freq = Length[DictionaryLookup[# ~~ ____]] & /@ CharacterRange["a", "z"];
relfreq = freq / Total[freq];
ListPlot[relfreq, Filling -> Axis, PlotStyle -> PointSize[Large]]
```



### ■ Benfordsches Gesetz (Häufigkeitsverteilung der Anfangsziffern von Zahlen)

Das Benfordsche Gesetz, auch Newcomb-Benford's Law (NBL) beschreibt eine Gesetzmäßigkeit in der Verteilung der Ziffernstrukturen von Zahlen in empirischen Datensätzen, zum Beispiel ihrer ersten Ziffern. Es lässt sich etwa in Datensätzen über Einwohnerzahlen von Städten, Geldbeträge in der Buchhaltung, Naturkonstanten etc. beobachten. Kurz gefasst besagt es:

Je niedriger der zahlenmäßige Wert einer Ziffernsequenz bestimmter Länge an einer bestimmten Stelle einer Zahl ist, umso wahrscheinlicher ist ihr Auftreten. Für die Anfangsziffern in Zahlen des Zehnersystems gilt zum Beispiel: Zahlen mit der Anfangsziffer 1 treten etwa 6,5-mal so häufig auf wie solche mit der Anfangsziffer 9.“

1881 wurde diese Gesetzmäßigkeit von dem Mathematiker Simon Newcomb entdeckt und im „American Journal of Mathematics“ publiziert. Er soll bemerkt haben, dass in den benutzten Büchern mit Logarithmentafeln, die Seiten mit Tabellen mit Eins als erster Ziffer deutlich schmutziger waren als die anderen Seiten, weil sie offenbar öfter benutzt worden seien. Die Abhandlung Newcombs blieb unbeachtet und war schon in Vergessenheit geraten, als der Physiker Frank Benford (1883–1948) diese Gesetzmäßigkeit wiederentdeckte und darüber 1938 neu publizierte. Seither war diese Gesetzmäßigkeit nach ihm benannt, in neuerer Zeit wird aber durch die Bezeichnung „Newcomb-Benford's Law“ (NBL) dem eigentlichen Urheber wieder Rechnung getragen. Bis vor wenigen Jahren war diese Gesetzmäßigkeit nicht einmal allen Statistikern bekannt. Erst seit der US-amerikanische Mathematiker Theodore Hill versucht hat, die Benford-Verteilung zur Lösung praktischer Probleme nutzbar zu machen, ist ihr Bekanntheitsgrad gewachsen. (aus Wikipedia)

**Benfordsches Gesetz:** Ist  $d$  die erste Ziffer einer Dezimalzahl, so tritt sie nach dem Benfordschen Gesetz in empirischen Datensätzen näherungsweise mit folgenden relativen Häufigkeiten  $p(d)$  auf:

$$p(d) = \log_{10}(1 + 1/d) = \log_{10}(d+1) - \log_{10}(d)$$

In der Demonstration unten werden die relativen Häufigkeiten der Anfangsziffern 1 bis 9 in verschiedenen Datensätzen mit den nach dem Benfordschen Gesetz prognostizierten relativen Häufigkeiten verglichen.

```

In[11]= Manipulate[With[{data = Sort[
  Tally[Cases[First /@ First /@ RealDigits /@ Rest[Cases[Cases[CountryData[#, prop] & /@
    CountryData[], _Real], Except[Missing["NotAvailable"]]], Except[0]]]],
  Grid[{{Text@Grid[{{#[[1]], ToString[N[#[[2]] / Total[Last /@ data] 100, 2]] <> "%"} & /@ data,
    Frame → All, Alignment → Center, ItemSize → {{1, 4}, Automatic}},
  Show[ListPlot[Last /@ data / Total[Last /@ data], Filling → Axis,
    ImageSize → 400, PlotStyle → {PointSize[Large], Thick, Black},
    ImagePadding → 25, PlotRange → {{0, 10}, {0, All}},
    Ticks → {Range[9], Automatic}}, Plot[Log[10, 1 + 1 / x], {x, 1, 9}]]]],
  {{prop, "Area", "property"}, CountryDataNumericProps,
  ControlType → PopupMenu},
  Alignment → Center,
  SynchronousInitialization →
  False, Initialization ⇒
  {CountryDataNumericProps = First /@ Cases[#, CountryData["USA", #]] & /@
    Cases[CountryData[All, "Properties"], Except[{}], {_, _Real}];}]

```

Out[11]=

