

# A Bayesian Approach to Generating Scenario Parameters

Eric Kemp-Benedict, 15 February 2007 (Last revision 24 March 2007)

## Introduction

When developing quantitative scenarios, it is necessary to choose parameter values to fit a particular story line, such as growth rates and resource-use efficiencies. Typically, the choice is made with an eye to historical experience as well as to the story behind the scenario, which might lead to a departure from historical patterns. This note offers one approach to making such choices by using Bayesian statistical reasoning to systematically combine historical data with qualitative assumptions regarding the scenario.

## Bayes' Rule

Bayes' Rule is applicable when the probability of an event occurring or (as in this case) the probability that a parameter takes on a particular value  $\theta$  needs to be updated in light of new information  $S$ .

Mathematically, Bayes' Rule states that

$$p(\theta|S) \propto p(\theta) p(S|\theta) . \quad (1)$$

That is, the probability that the parameter takes the value  $\theta$  in light of the new information provided by  $S$  (called the *posterior distribution*) is proportional to the probability that it takes on the value  $\theta$  without the new information (called the *prior distribution*), multiplied by the probability that the information  $S$  would appear given that the parameter is equal to  $\theta$ . The final factor is somewhat challenging to conceptualize, and can be easier to think of in terms of the statistical likelihood function, which is defined as

$$L(\theta|S) = p(S|\theta) . \quad (2)$$

which states that the likelihood that the parameter value is  $\theta$ , given the information  $S$ , is equal to the probability that the information provided by  $S$  will be present, given that the parameter value is  $\theta$ . For purposes of elicitation, values for  $L(\theta|S)$  can be estimated by asking, "In scenario  $S$  how much more or less frequently will the parameter take the value of  $\theta$  compared to the reference case?" Combining Equations (1) and (2) gives the form for Bayes' Rule that will be used in the rest of this note.

$$p(\theta|S) \propto p(\theta) L(\theta|S) . \quad (3)$$

## Bayes' Rule and Choosing Scenario Parameters

Suppose that a value  $\theta$  for a parameter  $\hat{\theta}$  must be chosen for scenario  $i$  out of  $N$  scenarios  $S_1, \dots, S_N$ . It is common to ask experts directly what the parameter might be in a particular scenario. However, Bayes' Rule suggests a different approach. Instead of asking directly, ask first what the probability that the parameter might take on a particular value would be without any information about which scenario is being examined. A reasonable answer to this question is that  $\theta$  would be distributed according to a historically experienced prior distribution  $p(\theta)$ .

The historical distribution might be based on the history of the system being studied, or it might be based on the history of a reference system. For example, if a set of scenario variants of a common high income growth scenario are being studied for a country that has experienced low growth in the past, then the historical distribution might be that of a country or region that has experienced high growth in the past, rather than the country being studied.

In order to apply Bayes' rule, the likelihood  $L(\theta|S_i)$  must be generated for each scenario  $S_i$ . This is done through expert elicitation.

## A Practical Implementation

In order to make the task of specifying the likelihood function more manageable, the prior distribution  $p(\theta)$  is replaced with a coarse discretization based on quantiles of the prior distribution. In the approach presented in this note, a substitute prior distribution  $p^*(\theta)$  is constructed, where

$$\begin{aligned}
 p^*(\theta = q_{0.025}) &= 5\% \\
 p^*(\theta = q_{0.150}) &= 20\% \\
 p^*(\theta = q_{0.500}) &= 50\% , \\
 p^*(\theta = q_{0.850}) &= 20\% \\
 p^*(\theta = q_{0.975}) &= 5\%
 \end{aligned}
 \tag{4}$$

in which  $q_k$  designates the  $k^{\text{th}}$  quantile of the distribution  $p(\theta)$ . For the purpose of elicitation, each of the values in Equation (4) is given a qualitative label:

- $q_{0.025}$  is "Very low"
- $q_{0.150}$  is "Low"
- $q_{0.500}$  is "Moderate"
- $q_{0.850}$  is "High"
- $q_{0.975}$  is "Very high"

An example is given in Illustration 1, using data on 15-year average growth rates in GDP per capita for countries in Africa from 1975 through 2002.

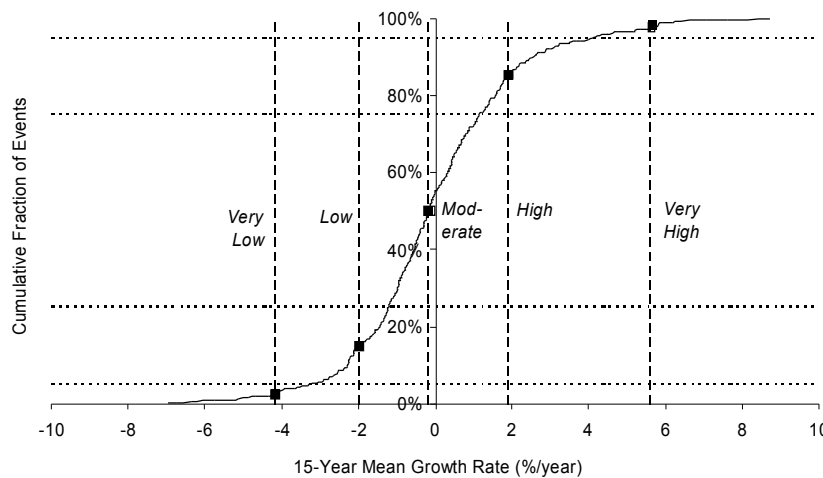


Illustration 1: Distribution and quantiles for growth rates for countries in Africa

The expert is then asked to assign qualitative values for the likelihood function, expressed as a relative frequency of occurrence compared to the baseline. Weights are assigned using a decibel scale, in which neighboring values differ by a factor of 10. The five values used are:

- “Much less frequent”: Likelihood  $\propto 1$
- “Less frequent”: Likelihood  $\propto 10$
- “About the same”: Likelihood  $\propto 100$
- “More frequent”: Likelihood  $\propto 1,000$
- “Much more frequent”: Likelihood  $\propto 10,000$

By combining the weights for different likelihoods with the probability distribution in Equation (4), the distribution of values of  $\hat{\theta}$  in scenario  $S_i$  can be estimated, and statistics calculated.

## Example

As an example, consider choosing an income growth rate for a scenario looking 15 years into the future. For this note this example was implemented using data from the GEO data portal.<sup>1</sup> Running averages of annual income growth rates for all countries were calculated over 15 year intervals using a moving 15-year window. Averages were calculated for all countries in the data set with continuous data from 1975 to 2002. This resulted in a data set with 1,339 data points. Descriptive statistics for this data set are:

- Mean: 1.15%
- Median: 1.20%
- Standard deviation: 2.23%
- Minimum: -6.95%
- Maximum: 8.90%
- 5<sup>th</sup> Percentile: -2.49%
- 95<sup>th</sup> Percentile: 5.20%

An expert might then be asked to assign likelihoods for different income growth rates for, say, a “high growth” scenario, by being asked, for example, “How much more or less frequent will growth rates of this occur in a scenario consistent with the scenario narrative compared to historical experience?” One possible set of likelihoods is:

- Very low: Somewhat less frequent
- Low: About the same
- Moderate: About the same
- High: Somewhat more frequent
- Very high: Somewhat more frequent

Constructing the posterior distribution for the growth rate on the basis of the historically-derived prior distribution and the likelihoods gives a distribution with both a mean of 3.02% and median rate of 2.95%. One of these (for example, the median) could then be used as the growth rate in the scenario.

---

1 <http://geodata.grid.unep.ch/>

## Tcl/Tk Script Implementing the Method

The approach described in this note is somewhat computationally intensive, but is also formulaic, which makes it ideal for implementing in a computer program. This section contains a script written in the Tcl/Tk language that implements the scenario parameter generation approach described in this note. A screenshot is shown in Illustration 2, showing the example given in the previous section. In Microsoft Windows, the script may be run by saving it to a file, downloading and installing the free “batteries included” distribution of Tcl/Tk from ActiveState,<sup>2</sup> and then double-clicking on the file. A reference data set is copied and pasted into the window, the values assigned to the likelihood function are set, and the program is run. It is expected that this will be done by a modeler after the expert consultation has taken place.

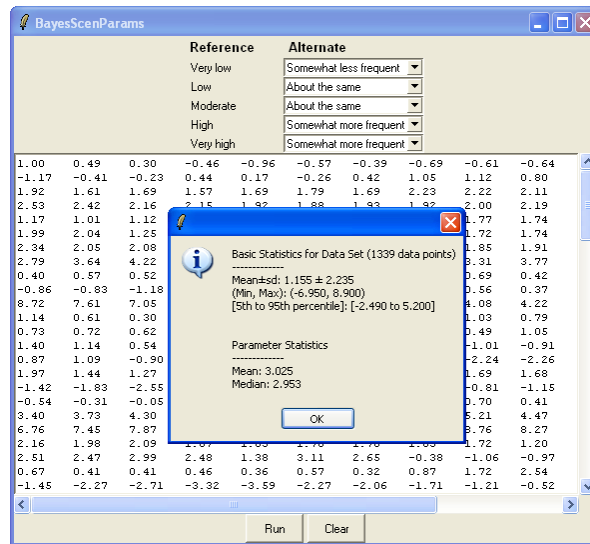


Illustration 2: Screenshot of scenario parameter generator

```
# Copyright 2007 Eric Kemp-Benedict
# This code may be freely copied, modified, and re-used.
package require math::statistics
package require BWidget

namespace eval bayesparams {
    # Data is a text window with the data in it
    variable data
    variable ranges {0.025 0.150 0.500 0.850 0.975}
    variable coverage {0.05 0.20 0.50 0.20 0.05}
    set likelihoods {
        "Much less frequent"
        "Somewhat less frequent"
        "About the same"
        "Somewhat more frequent"
        "Much more frequent"
    }
    set levels {
        "Very low"
        "Low"
        "Moderate"
        "High"
        "Very high"
    }
    # wt = wtbase ^ index of likelihoods
    variable wtbase 10

    proc getdata {} {
        variable data
```

2 Available through a free download from <http://www.activestate.com/Products/ActiveTcl/>.

```

    set numlist {}
    foreach item [split [$data get 1.0 end] " \t\n"] {
        if [string is double -strict $item] {
            lappend numlist $item
        }
    }
    return $numlist
}

proc clear {} {
    variable data

    $data delete 1.0 end
}

proc confinterval {conf} {
    set numlist [getdata]
    if {[llength $numlist] == 0} {
        return {}
    }
    set retval {}
    lappend retval [::math::statistics::quantiles $numlist [expr {1.0 - $conf}]]
    lappend retval [::math::statistics::quantiles $numlist $conf]

    return $retval
}

proc getstats {} {
    set numlist [getdata]
    if {[llength $numlist] == 0} {
        return {}
    }
    return [::math::statistics::basic-stats $numlist]
}

proc getquants {} {
    variable ranges

    set numlist [getdata]
    if {[llength $numlist] == 0} {
        return {}
    }
    set retval {}
    foreach cutoff $bayesparams::ranges {
        lappend retval [::math::statistics::quantiles $numlist $cutoff]
    }
    return $retval
}

proc dataframe {w} {
    variable data

    ScrolledWindow $w
    set data [text $w.t -wrap none -font {"Courier New" 8}]
    $w setwidget $data

    return $data
}

}

## Use library

proc run {} {
    set quants [bayesparams::getquants]
    if {[llength $quants] == 0} {
        tk messageBox -icon info -message "No data have been entered."
        return
    }
    set rawprob {}
    foreach w [wininfo children .f] c $bayesparams::coverage {
        set choice [$w.cb getvalue]
        set wt [expr {pow($bayesparams::wtbase, $choice)}]
        # This applies Bayes rule to update "coverage" given weights
        # "rawprob" doesn't sum to 1
    }
}

```

```

        lappend rawprob [expr {$wt * $c}]
    }
    set sum_pq 0
    set sum_p 0
    foreach q $quants p $rawprob {
        set sum_pq [expr {$sum_pq + $p * $q}]
        set sum_p [expr {$sum_p + $p}]
    }
    set partsum_p 0
    set prevps_p 0
    set prevquant [lindex $quants 0]
    foreach q $quants p $rawprob {
        # Go to midpoint
        set partsum_p [expr {$partsum_p + 0.5 * $p}]
        if {2 * $partsum_p >= $sum_p} {
            set median [expr {$prevquant + ($q - $prevquant) * (0.5 * $sum_p - $prevps_p) / \
                ($partsum_p - $prevps_p)}]
            break
        }
        set prevps_p $partsum_p
        # Move past midpoint before going to next
        set partsum_p [expr {$partsum_p + 0.5 * $p}]
        set prevquant $q
    }
    set stats [bayesparams::getstats]
    set msg_meansd [format "\nMean\u00B1sd: %.3f \u00B1 %.3f" [lindex $stats 0] [lindex $stats 4]]
    set msg_minmax [format "\n(Min, Max): (%.3f, %.3f)" [lindex $stats 1] [lindex $stats 2]]
    set msg "Basic Statistics for Data Set ([lindex $stats 3] data points)"
    set msg "$msg\n-----"
    set msg "$msg$msg_meansd"
    set msg "$msg$msg_minmax"
    set ci [bayesparams::confinterval 0.95]
    set msg_ci [format "\n[5th to 95th percentile]: \[%.3f to %.3f\]" [lindex $ci 0] \
        [lindex $ci 1]]
    set msg "$msg$msg_ci"
    set msg "$msg\n\nParameter Statistics"
    set msg "$msg\n-----"
    set msg_meanmed [format "\nMean: %.3f\nMedian: %.3f" [expr {$sum_pq / $sum_p}] $median]
    set msg "$msg$msg_meanmed"
    tk_messageBox -message $msg
}

frame .t
pack .t -side top
label .t.a -text "Reference" -width 15 -font "helvetica 10 bold"
pack .t.a -side left
label .t.b -text "Alternate" -width 20 -font "helvetica 10 bold" \
    -anchor w
pack .t.b

frame .f
pack .f -side top
foreach lvl $bayesparams::levels {
    set frm [frame .f.[string tolower [join [split $lvl] {}]]]
    pack $frm -side top
    label $frm.l -text $lvl -width 15 -anchor w
    pack $frm.l -side left
    ComboBox $frm.cb -values $bayesparams::likelihoods
    pack $frm.cb
    $frm.cb setvalue @2
}

frame .buttonframe
pack .buttonframe -side bottom

set df [bayesparams::dataframe .data]
pack .data -side top -expand yes -fill both

button .buttonframe.run -text "Run" -width 8 -pady 4 -command run
pack .buttonframe.run -side left -padx 2
button .buttonframe.clear -text "Clear" -width 8 -pady 4 -command bayesparams::clear
pack .buttonframe.clear -padx 2

focus -force $df

```