Bayesian Stock Management (Solution proposal)

contact: jean-marc.freyermuth@univ-amu.fr

Applied Statistics, Krakow, 2020

In the company

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

Table 1: Data for a single component

age	nb_comp_obs	nb_comp_fail	nb_comp_new
0	10	0	9
1	15	0	5
2	25	2	11
3	30	3	4
4	20	10	5

To tackle this problem we start by dealing with one age class having in mind the relatively small amount of historical data, the very valuable prior information from mechanics and the need for probabilistic prediction suggest that a Bayesian modelling is a good option.

For a single age class

The simplest first step to approach this problem is to model separately of the 'age' of the component. This problem can be modelled using standard bayesian model for binomial data. It is rather simple and convenient to model the prior knowledge on the probability of having broken component using a beta distribution (conjugate bayesian model). If no information available, there are different ways to specificy a non informative prior; here, we will specify a uniform distribution over the unit interval $\theta \sim \mathcal{B}e\,(1, 1)$. In such model, the posterior distribution of $\theta|y$ is also a beta~:

$$
\begin{aligned}
&y|\theta \sim Bin(n, \theta) \\
&\theta \sim \mathcal{B}e\,(a, b) \\
&\theta|y \sim \mathcal{B}e\,(a' = a + y, b' = n - y + b)
\end{aligned}
$$

Knowing the form of the posterior distribution, we know the posterior mean and variance:

$$
\begin{aligned}
&\mathbb{E}\,[\theta\,|y] = \frac{y + a}{n + a + b} \\
&Var\,[\theta\,|y] = \frac{(y + a)\,(n - y + b)}{(n + a + b)^2\,(a + b + 1)}
\end{aligned}
$$

The beta prior is a conjugate prior with the nice interpretation that the parametrization (a, b) corresponds respectively to the number of success and failures from a previous experiment.

Remark: the posterior mean is a compromise between the prior and MLE mean.

$$\begin{aligned}
\frac{y+a}{n+a+b} &= \quad \lambda\left(\frac{a}{a+b}\right)+(1-\lambda)\,\frac{y}{n} \\
&= \qquad\qquad\qquad\qquad \dots \\
&\Rightarrow \quad \lambda = \frac{a+b}{n+a+b} \in (0,1)
\end{aligned}$$

The posterior summarizes all the information about the paramater. We would like now to predict the number of broken components in a set of new gearboxes. Let us denote as $\tilde{y}$ the number of components (of a given age) in the new set of gearboxes, assume that it follows a binomial distribution. Then we can compute the posterior predictive distribution~:

$$\begin{aligned}
p\left(\tilde{y}\,|y\right) &= \int_0^1 p\left(\tilde{y}\,|\theta\right)p\left(\theta\,|y\right)d\theta \\
&= \frac{n!}{(n-y)!y!}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta^{a+y-1}(1-\theta)^{n-y+b-1}d\theta \\
&= \frac{\Gamma\left(n'+1\right)}{\Gamma\left(n'-\tilde{y}+1\right)\Gamma\left(\tilde{y}+1\right)}\frac{\Gamma\left(a'+b'\right)}{\Gamma\left(a'\right)\Gamma\left(b'\right)}\frac{\Gamma\left(a'+\tilde{y}\right)\Gamma\left(n'-\tilde{y}+b'\right)}{\Gamma\left(a'+b'+n'\right)}
\end{aligned}$$

This is a BetaBinomial with parameters (n', a', b') with expectation and variance:

$$\begin{aligned}
\mathbb{E}\left[\tilde{y}\right] &= \mathbb{E}\left[\mathbb{E}\left[\tilde{y}\,|\theta\right]\right] \\
&= \mathbb{E}\left[n'\theta\right] \\
&= n'\frac{a'}{a'+b'}
\end{aligned}$$

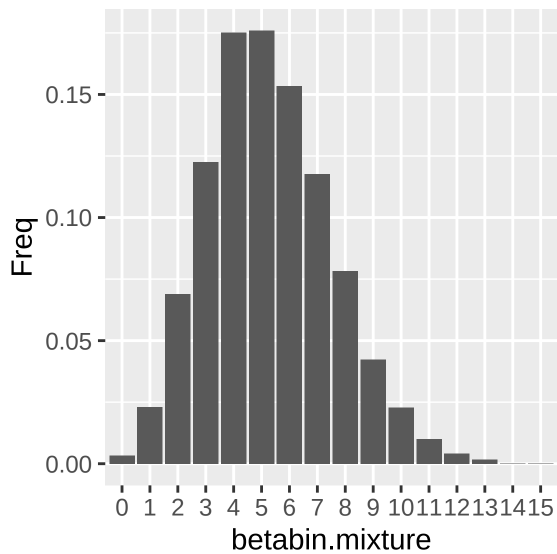
$$\begin{aligned}
\mathrm{Var}\left[\tilde{y}\right] &= \mathrm{Var}\left[\mathbb{E}\left[\tilde{y}\,|\theta\right]\right]+\mathbb{E}\left[\mathrm{Var}\left[\tilde{y}\,|\theta\right]\right] \\
&= \dots \\
&= \frac{n'a'b'\left(n'+a'+b'\right)}{\left(a'+b'\right)^2\left(a'+b'+1\right)}
\end{aligned}$$

including age class

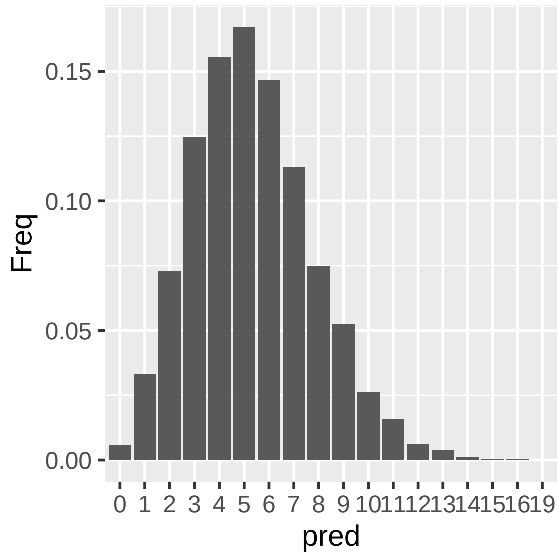
If there would be large number of age classes, a preliminary historical check of correlation between the age class of a component and its probability of failure would be important to look at. Then we could model use a logistic regression model for example. Nevertheless, the prior specification in such model is more involved. This fact and the relatively small number of age categories, lead us to prefer to model independently the different age categories (and therefore specify priors for each age class). At the end, we consider a mixture of posterior predictive distribution with known proportions.

```
M = 10000
alpha1 = alpha2 = alpha3 = alpha4 = alpha5 = 1
beta1 =beta2 = beta3 = beta4 = beta5 = 1

data2 = cbind(data, data.frame(alpha = 1, beta = 1))
```

```r
sim = apply(data2, 1, function(x){rbinom(n = M, size = x[4], rbeta(x[5] + x[3],
                                                           x[2]+ x[6] - x[3], n = M))})

# Mixture - predictive distribution of the broken components
betabin.mixture <- rowSums(sim)
tN <- table( betabin.mixture) / M
r = data.frame(tN)
p<-ggplot(data=r, aes(x=betabin.mixture, y=Freq)) +
  geom_bar(stat="identity")
p
```




```r
# compared without taking into account the age class
n <- 100;
y <- 15;
n.tilde <- 34
alpha.1 <- 1 ; beta.1 <- 1

alpha.post.1 <- y + alpha.1 ; beta.post.1 <- n - y + beta.1
pred <- rbinom(n = M, size = n.tilde, prob = rbeta(n = M, alpha.post.1, beta.post.1))

par(mfrow=c(2,1),mar=c(3,3,1,1),mgp=c(2,0.8,0))
tN <- table( pred) / M
r = data.frame(tN)
p<-ggplot(data=r, aes(x=pred, y=Freq)) +
  geom_bar(stat="identity")
p
```



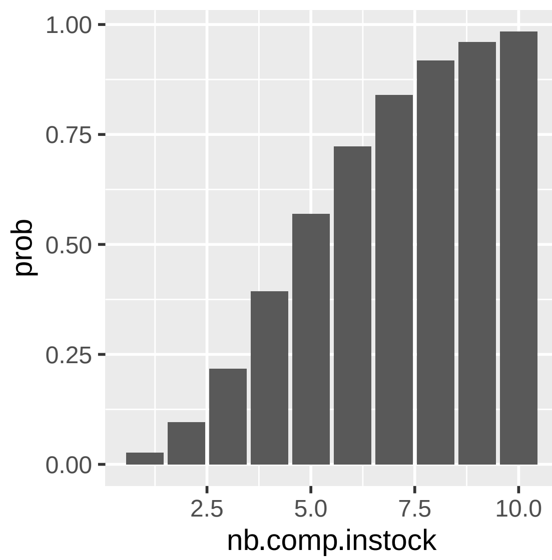

```
# calculation of the 90 percent probability of have tilde(y) broken components
quantile(betabin.mixture, c(0.90) )
```

```
## 90%
##   8
```

```
quantile(pred, c(0.90) )
```

```
## 90%
##   9
```

```
# finally the decision of the number of components to have in stock can be taken accordingly to the pro
nb.comp.instock =1:10
prob = rep(0, length = 10)
for (i in 1:10)
{
prob[i] = 1-sum(betabin.mixture > nb.comp.instock[i]) / M
}

r =data.frame(prob, nb.comp.instock)
p<-ggplot(data=r, aes(x=nb.comp.instock, y=prob)) +
  geom_bar(stat="identity")
p
```



... in the company

Subjective prior specifcation !

Not much of historical data is available, hence, it is a crucial to benefit from the mechanics expertise, i.e., to express their prior beliefs as input in our model. They need a tool to be use in real-life (user friedly interface). It is not convenient to specify the prior parametrisation in terms of Beta parameters. A simple solution consists in expressing the prior belief as an expected probability of failures and associated variability (here think of confidence).

$$\mu_1 = \frac{a}{ab}$$
$$\mu_2 = \frac{ab}{(a+b)^2(a+b+1)}$$

from where we get

$$a = \frac{(1-\mu_1)\mu_1^2 - \mu_1\mu_2}{\mu_2}$$
$$b = \frac{1-\mu_1}{\mu_2}\left[(1-\mu_1)\mu_1 - \mu_2\right]$$

A MUST, make it user-friendly: building a shiny app

To make it friendly one could go to shiny R app and use 2 cursors to give prior specified from mean and variance.

http://web.maths.unsw.edu.au/~lafaye/RShiny/course.nb.html

- server.R
- ui.R (for user interface)

In this problem we need to consider one specific aspect that can be achieved using R-Shiny. We wanna do reactive programming which is a coding style that starts with reactive values that are given by the users and functional programming ahead with reactive-expressions.

In shiny the reactive values are obtained using the input object which is pass to shiny server function.

Appendix

expectation and variance

Consider two random variables (X, Y) with a joint pdf $p(x, y)$ then

$$\mathbb{E}[Y] = \mathbb{E}_X[\mathbb{E}[Y|x]] \mathbb{V}ar[Y] = \mathbb{E}_X[\mathbb{V}ar[Y|x]] + \mathbb{V}ar_X[\mathbb{E}[Y|x]]$$

Proof

$$\begin{aligned}
\mathbb{E}[Y] &= \int_Y yp(y)dy = \int_Y y \int_X p(x,y)dxdy \\
&= \int_X \left[\int_Y yp(y|x)dy\right] p(x)dx \\
&= \mathbb{E}_X[\mathbb{E}[Y|x]]
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}ar[Y] &= \int_Y (y - \mathbb{E}[Y])^2 p(y)dy = \int_X \left[\int_Y (y - \mathbb{E}[Y])^2 p(y|x)dy\right] p(x)dx \\
&= \int_X \left[\int_Y (y - \mathbb{E}[Y|x] + \mathbb{E}[Y|x] - \mathbb{E}[Y])^2 p(y|x)dy\right] p(x)dx \\
&= \int_X \left[\int_Y (y - \mathbb{E}(Y|x))^2 p(y|x)dy\right] p(x)dx + \\
&\int_X (\mathbb{E}(Y|x) - \mathbb{E}(Y))^2 \left[\int_Y p(y|x)dy\right] p(x)dx \\
&= \mathbb{E}_X[\mathbb{V}ar[Y|x]] + \mathbb{V}ar_X[\mathbb{E}[Y|x]]
\end{aligned}$$