A stock Management problem - Solution proposal Part I

Master DS

Academic year 2024-2025

In the company

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 address this problem, we begin by focusing on a single age class. Considering the relatively limited amount of historical data, the valuable prior knowledge derived from mechanics, and the necessity for probabilistic predictions, Bayesian modeling emerges as a suitable and effective approach.

For a single age class

The simplest initial step to address this problem is to model the 'age' of the component separately. This issue can be effectively analyzed using a standard Bayesian model for Binomial data. A Bayesian approach offers a straightforward and convenient framework to incorporate prior knowledge about the probability of a component being broken, using a Beta distribution as the prior (a conjugate Bayesian model).

If no prior information is available, there are several ways to define a non-informative prior. In this case, we will use a uniform distribution over the unit interval, expressed as $\theta \sim \mathcal{B}e\left(1,1\right)$. Within this model, the posterior distribution of $\theta|y$ is also a Beta distribution:

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

Since we know the posterior distribution, we readily obtain its posterior mean and variance:

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

The Beta prior is a conjugate prior giving insightful interpretation of the prior parameter (a, b) whenever they are positive integers. These parameters can be interpreted as respectively the number of success and failures from a previous experiment.

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

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

Next, we aim to predict the number of broken components in a set of $\tilde{n}$ new gearboxes. Let us denote as $\tilde{y}$ the number of broken components (of a given age) in this new set of gearboxes. We assume it follows a binomial distribution and it is independent from y conditionally on θ, we can calculate the posterior predictive distribution as follows:

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

This is a BetaBinomial distribution with parameters $(\tilde{n}, a', b')$ with expectation and variance:

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

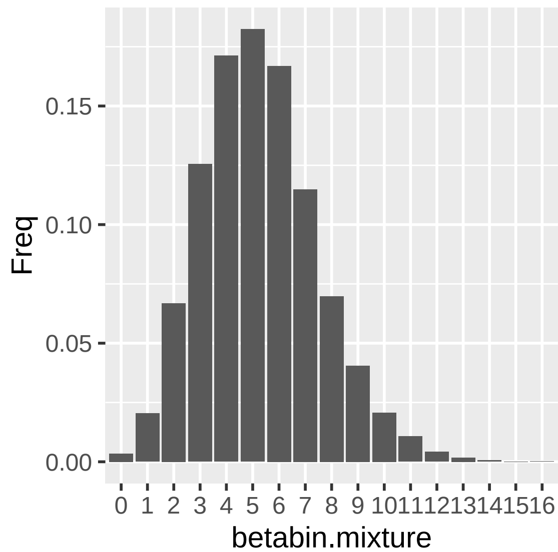
$$
\begin{aligned}
\mathrm{Var}[\tilde{y}] &= \mathrm{Var}[\mathbb{E}[\tilde{y}\,|\theta\,]]+\mathbb{E}[\mathrm{Var}[\tilde{y}\,|\theta\,]] \\
&= \dots \\
&= \frac{\tilde{n}a'b'(\tilde{n}+a'+b')}{(a'+b')^2(a'+b'+1)}.
\end{aligned}
$$

Including age class

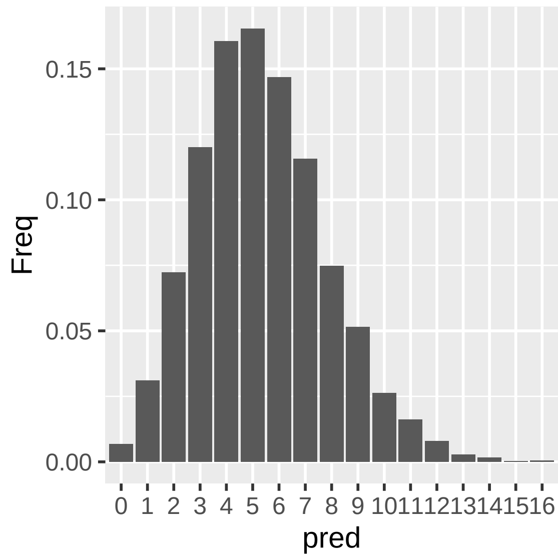
If there were a large number of age classes, it would be important to conduct a preliminary historical analysis to assess the correlation between a component's age class and its probability of failure. Based on the findings, we could, for instance, employ a logistic regression model. However, specifying priors in such models is more complex. Given this complexity and the relatively small number of age categories in our case, we opt to model each age category independently, specifying separate priors for each class. Ultimately, we combine these independent models into a mixture of posterior predictive distributions, weighted by known proportions.

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

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

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
```



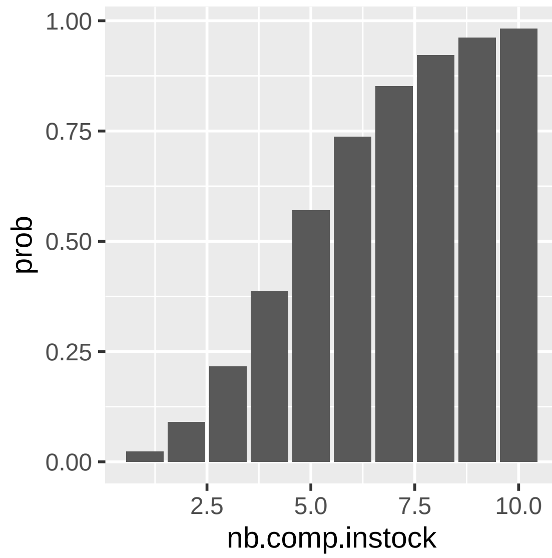

```r
# calculation of the 90 percent probability of have tilde(y) broken components
quantile(betabin.mixture, c(0.90) )
```

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

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

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

```r
# 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
```


prob
1.00
0.75
0.50
0.25
0.00
2.5
5.0
7.5
10.0
nb.comp.instock