Optimisation of surveillance decision

Application to the diced bacon process

N. Commeau [1,2], M. Cornu [3], E. Parent[1]

[2]INRA/AgroParisTech UMR 518, Paris, France

[2]French agency for food, environmental and occupational health & safety/ Laboratory for food safety, Maisons-Alfort, France

[3]Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PPR-ENV, SERIS, LM2E, Cadarache, France

Introduction

- In the food industry, food business operators (FBOs) must ensure that "foodstuffs comply with the relevant microbiological criteria" (European regulation (EC) No.2073/2005, Article 3)
- Sampling plan is a tool to assess microbiological contamination in food
- Two-class attribute sampling plan : n units are sampled in a subpart of the production and analyzed for the presence of a given microorganism

Introduction

- In the food industry, food business operators (FBOs) must ensure that “foodstuffs comply with the relevant microbiological criteria” (European regulation (EC) No.2073/2005, Article 3)
- Sampling plan is a tool to assess microbiological contamination in food
- Two-class attribute sampling plan : n units are sampled in a subpart of the production and analyzed for the presence of a given microorganism
- This kind of sampling-plan estimates the prevalence ψ in a batch : ψ is estimated with x/n
- Sampling-plan can be used to take a decision about the lot
  - if $x \leq c$, a decision is taken
  - if $x > c$, another decision is taken.
- How to choose n and c properly?

Decision and uncertainty

- How to make a decision in an uncertain environment?
- Define $\mathcal{D}$ the set of all possible decisions
- Define Ψ the set of states of nature ψ (also called the parameter)
- Use a criterion called the *loss function* defined for all $(d, \psi) \in \mathcal{D} \times \Psi$ in $\mathbb{R}^+$
- Every decision d taken when the state of nature is ψ has a loss $L(\psi, d)$

Decision rule

- The decision d depends on observation x through a *decision rule* δ defined from the set of observations $\mathcal{X}$ into $\mathcal{D}$: $\delta : x \mapsto d$
- Aim: find the "best" strategy to associate a decision d to observation x, even though ψ is not fully known
- Remark : when $\mathcal{D} = \Psi$, δ is an estimator. Often, $L(\delta(x), \theta) = (\delta(x) - \theta)^2$

Bayesian expected loss and risk function

The Bayesian approach considers the posterior expected cost called the *Bayesian expected loss*:

$$\rho(d|x) = \int_{\Psi} L(\psi, d)\pi(\psi|x)d\psi,$$

Bayesian expected loss and risk function

The Bayesian approach considers the posterior expected cost called the *Bayesian expected loss*:

$$\rho(d|x) = \int_{\Psi} L(\psi, d)\pi(\psi|x)d\psi,$$

For all x, the aim is to find decision d minimizing $\rho(d|x)$. The Bayesian decision rule is : $\delta : \; x \mapsto d = Arg\min \rho(d|x)$.
If, for all $x \in \mathcal{X}$, δ minimizes ρ, it also minimises the *Bayes risk*

$$r(\delta) = \int_{\mathcal{X}} \int_{\Psi} L(\psi, \delta)\pi(\psi|x)f(x)d\psi dx$$

Predictive analysis

If data x have not been observed yet:

- Choose the experimental set-up e which will give x_e
- The best decision rule is the one which minimizes $r(\delta_e) = \int \int L(\delta_e(x), \psi)\pi(\psi)f(x|\psi)dxd\psi$.

Needed specifications

1. Ψ the set of the states of nature ψ and its distribution $\pi(\psi)$
2. The experimental set-up e
3. The set of the observations $\mathcal{X}$ and its distribution $f(x|\psi)$
4. The set of decisions $\mathcal{D}$
5. The loss function L

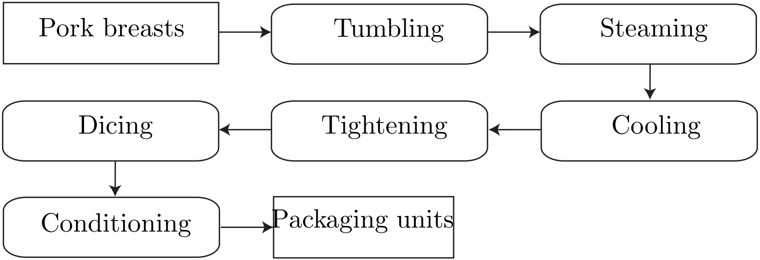
Difficult points: 1, 4 and 5.

Needed specifications

1. Ψ the set of the states of nature ψ and its distribution $\pi(\psi)$
2. The experimental set-up e
3. The set of the observations $\mathcal{X}$ and its distribution $f(x|\psi)$
4. The set of decisions $\mathcal{D}$
5. The loss function L

Difficult points: 1, 4 and 5.

Process


Pork breasts
Tumbling
Steaming
Cooling
Tightening
Dicing
Conditioning
Packaging units

Batch

- Statistic *batch*: repetitive behaviour, same probability distribution
- Usual decisions: accepting or rejecting the batch
- In this application: the production batches are not all analyzed, production batch $\neq$ statistic batch
- Here, the *batch* is a production period of one month

States of nature, experimental set-up, decisions

- ψ : prevalence of a one-month production, $\Psi = [0; 1]$ et $\pi(\psi) = \mathcal{B}eta(a, b)$
  - low if $\psi \in \Psi_0 = [0; \psi_0]$
  - medium if $\psi \in \Psi_1 =]\psi_0; \psi_1]$
  - high if $\psi \in \Psi_2 =]\psi_1; 1]$
- Experimental set-up : number of observations n over a period
- Observations : number of positive results x among n analyses, $\mathcal{X} = [0 : n]$, $f(x|\psi, n) = \mathcal{B}in(n, \psi)$
- Decisions

$$\begin{cases} d_0, & \text{no corrective action needed} \\ d_1, & \text{a slight corrective action is needed} \\ d_2, & \text{a big corrective action is needed} \end{cases}$$

Posterior and marginal distributions

If $\psi \sim \mathcal{B}eta(a;b)$ and $x|\psi \sim \mathcal{B}in(n,\psi)$, then $\pi(\psi|x) = \mathcal{B}eta(a+x;b+n-x)$ and

$$\begin{aligned}[x] &= \frac{\pi(\psi)f(x|\psi)}{\pi(\psi|x)} \\ &= \frac{\Gamma(a+b)\Gamma(a+x)\Gamma(b+n-x)\Gamma(n+1)}{\Gamma(a)\Gamma(b)\Gamma(a+b+n)\Gamma(n-x+1)\Gamma(x+1)},\end{aligned}$$

where $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}dt$.

Remark : X follows the Polya distribution (probability that a urn has $a+x$ white balls after n draws. The urn has initially a white balls and b red balls. At each draw a ball is randomly drawn from the urn and is put back along with a new ball of the same color.

Example of loss function L

Assumption: the decisions and the consequences can be quantified in euros

- Decisions taken by the FBO
  - 0 for d_0
  - C_1 for d_1
  - C_2 for d_2
- Fine to be paid to the client (ex: retailer)
  - 0 if $\psi \in \Psi_0$
  - K_1 if $\psi \in \Psi_1$
  - K_2 if $\psi \in \Psi_2$
- As d_1 and d_2 should reduce the prevalence
  - the fine when d_1 is taken is reduced by $1 - \alpha\%$
  - the fine when d_2 is taken is reduced by $1 - \beta\%$

Exemple of loss function L

Decisions ψ	d_0	d_1	d_2
$\psi \in \Psi_0$	$0 + 0 + cn$	$C_1 + cn$	$C_2 + cn$
$\psi \in \Psi_1$	$0 + K_1 + cn$	$C_1 + \alpha K_1 + cn$	$C_2 + \beta K_1 + cn$
$\psi \in \Psi_2$	$0 + K_2 + cn$	$C_1 + \alpha K_2 + cn$	$C_2 + \beta K_2 + cn$

Table: Values taken by the cost function L depending on the decision d taken by the plant and the prevalence ψ of the lot.

Exemple of loss function L

The Bayesian expected loss is equal to:

$$\begin{aligned}\rho(\delta_n(x) = d|x) &= \int L(\delta_n(x), \psi)\pi(\psi|x,n)d\psi \\ &= cn + 1_{\delta_n(x)=d_0}(K_1P_1 + K_2P_2)f(x) \\ &+ 1_{\delta_n(x)=d_1}(C_1 + (\alpha K_1P_1 + \alpha K_2P_2))f(x) \\ &+ 1_{\delta_n(x)=d_2}(C_2 + (\beta K_1P_1 + \beta K_2P_2))f(x),\end{aligned}$$

where $P_i = \mathbb{P}(\psi \in \Psi_i|x,n)$, $i = 1,2$. Decision d_0 is chosen if $\rho(d_0|x) \leq \rho(d_1|x) \Leftrightarrow x \leq c_1$.

$$\begin{aligned}\delta_n(x) = d_0 &\quad \Leftrightarrow x \leq c_1 \\ \delta_n(x) = d_1 &\quad \Leftrightarrow c_1 < x \leq c_2 \\ \delta_n(x) = d_2 &\quad \Leftrightarrow x > c_2,\end{aligned}$$

Expert

We asked an expert to estimate costs c, C_1, C_2, K_1 and K_2.
The task was the following:

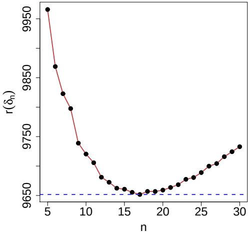
- Define the slight and the big corrections needed to lower the prevalence
- Describe the fines the client charges the plant
- Give a cost to each correction and each kind of fine

c	C_1	C_2	K_1	K_2	α	β	ψ_0	ψ_1	a	b
16	4250	14000	6200	90050	0.3	0.15	0.2	0.6	2	3

Bayes risk and sampling size

Bayesian risk:

$$\begin{aligned} r(\delta_n) &= cn + \sum_{x=0}^{c_1} (K_1 P_1 + K_2 P_2) f(x) \\ &+ \sum_{x=c_1+1}^{c_2} (C_1 + (\alpha K_1 P_1 + \alpha K_2 P_2)) f(x) \\ &+ \sum_{x=c_2+1}^{n} (C_2 + (\beta K_1 P_1 + \beta K_2 P_2)) f(x) \end{aligned}$$



The minimum is reached for $n = 17$, $c_1 = 5$ and $c_2 = 12$.

Conclusion

- Powerful tool to conceptualize the types of decisions taken by a plant
- Elicitation is difficult
- The two most important and difficult issues to be discussed
  - the duration of the period over which it is relevant to consider the prevalence
  - the nature (and costs) of the negative consequences