LARS Information Note 050571
Purdue University

Statistical Model for Aircraft Data Acquisition

A. G. Wacker and P. H. Swain

Problem: Given N_1 flightlines to be covered in Q days. Assume on any given day that any flightline is "flyable" with probability p (independent of all other flightlines). However, due to limited resources, at most N_2 flightlines can be covered in a day.

What is the expected number of flightlines covered in Q days?

Solution: Let

F_ℓ = random variable designating the number of flightlines covered by the end of day ℓ. F_ℓ can take on values 0, 1, 2,..., N_1 except that it can never exceed $N_2\ell$.

T_ℓ = random variable designating the number of flightlines covered on day ℓ. T_ℓ can take on values 0, 1, 2,..., N_2.

We have the following recursive relation:

$$F_0 = 0$$

$$F_\ell = F_{\ell-1} + T_\ell, \quad \ell = 1, 2, \ldots, Q \qquad (1)$$

We can write:

$$P(F_\ell = x) = P(F_{\ell-1} + T_\ell = x) \qquad 0 \leqq x \leqq \min [N_1, N_2\ell]$$

$$= 0 \qquad \text{otherwise} \qquad (2)$$

Note that:

$$P(F_{\ell-1} + T_\ell = x) = P(F_{\ell-1} = x, T_\ell = 0) + P(F_{\ell-1} = x-1, T_\ell = 1)$$

$$+ \ldots + P(F_{\ell-1} = 0, T_\ell = x)$$

$$= \sum_{i=0}^{x} P(T_\ell = i, F_{\ell-1} = x - i)$$

The right hand side represents all mutually exclusive ways of realizing the event on the left hand side. In terms of conditional probabilities, this can be written as:

$$P(F_{\ell-1} + T_{\ell} = x) = \sum_{i=0}^{x} P(T_{\ell} = i | F_{\ell-1} = x - i) \, P(F_{\ell-1} = x - i)$$

Thus, equation (2) becomes:

$$P(F_{\ell} = x) = \sum_{i=0}^{x} P(T_{\ell} = i | F_{\ell-1} = x - i) \, P(F_{\ell-1} = x - i), \qquad 0 \leq x \leq \min [N_1, N_2 \ell] \qquad (3)$$

$$= 0 \text{ otherwise}$$

(note that $P(T_{\ell} = i | F_{\ell-1} = x - i) = 0$ for $i > N_2$).

Equation (3) can be used recursively to obtain the distribution of F_{ℓ} for an arbitrary day. To do this we need the distribution of F_0 and the conditional distribution of T_{ℓ} given $F_{\ell-1}$, $\ell = 1, 2, \ldots, Q$. The distribution of F_0 is:

$$P(F_0 = x - i) = 1 \qquad x = i \qquad (4)$$

$$= 0 \qquad \text{otherwise}$$

To get the conditional distribution of T_{ℓ} given $F_{\ell-1}$, let $A(\ell, j)$ be a random variable designating the number of flightlines flyable on day ℓ, given that j flightlines have already been flown prior to day ℓ (i.e. $F_{\ell-1} = j$). $A(\ell, j)$ is binomially distributed:

$$P[A(\ell,j) = i] = \binom{N_1 - j}{i} p^i (1-p)^{N-j-i} \qquad i = 0, 1, \ldots N_1 - j$$

$$= 0 \qquad \text{otherwise}$$

(note $N_1 - j$ is the number of flightlines remaining to be flown) since $A(\ell, j)$ is essentially the number of successes in a sequence of N_1-j Bernoulli trials. The number of flightlines <u>flown</u> will be equal to $A(\ell,j)$ if $A(\ell,j) < N_2$ and equal to N_2 if $A(\ell,j) \geq N_2$. Therefore T_ℓ has the distribution:

<u>for $N_1 - j < N_2$</u>

$$P(T_\ell = i | F_{\ell-1} = j) = \binom{N_1-j}{i} p^i (1-p)^{N_1-j-i} \qquad i = 0, 1, \ldots N_1-j$$

$$= 0 \qquad \text{otherwise}$$

<u>for $N_1 - j \geq N_2$</u>

$$P(T_\ell = i | F_{\ell-1} = j) = \binom{N_1-j}{i} p^i (1-p)^{N_1-j-i} \qquad i = 0,1,2,\ldots,N_2-1$$

$$= \sum_{k=N_2}^{N_1-j} \binom{N_1-j}{k} p^k (1-p)^{N_1-j-k} \qquad i = N_2$$

$$= 0 \qquad \text{otherwise}$$

Now we have all that is needed to use equation (3) to obtain the distribution of F_ℓ for any day ℓ. Then the expected number of flightlines flown by the end of day Q is given by:

$$E[F_Q] = \sum_{i=1}^{Q} iP(F_Q = i)$$

For example, if

N_1 = 36 flightlines to be flown

N_2 = 9 flightlines can be flown in a day

Q = 14 days to fly the flightlines

p = 0.3 probability of a flightline "flyable" any given day

then the expected number of flightlines covered in 14 days is 35.72. Furthermore, 36 flightlines can be covered with probability .76, at least 35 flightlines with probability .97.

Discussion: The model is somewhat simplified. For example, it assumes a time-invariant probability that a flightline is flyable on any day. Also, it assumes that given any distribution of flyable flightlines, up to N_2 of them can always be flown. A more realistic situation would be to make N_2 a function of the spatial distribution of the flightlines.

In any case, this model gives something of the flavor of the aircraft-data-acquisition problem, in which weather causes the "flyability" of a flightline to be a probabilistic matter.