Modeling in Probability and Statistics

Andre A. Napoli

Scenario 1: The Commo Officer

Suppose you are the communications officer for an armored battalion. Your battalion TOC must maintain communications with the brigade TOC at all times when in a combat environment. In preparation for a deployment that will last a month, your battalion XO wants you to tell him the number of radios the unit should bring in order to have a 99% reliable link from battalion to brigade. You learn that one of these radios typically fails every 300 hours.

Problem 1: Model this situation as a standby redundant system (as shown in Figure 1) and determine the number of radios required to achieve 99% reliability.

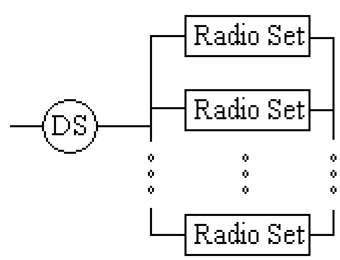

Figure 1. Battalion Communications System

Solution 1:

We assume that a standby redundant system models radio failures as a Poisson process. There are three assumptions required of such a process – stationarity, rarity and independence.

- Stationarity – the expected number of radio failures in any subinterval is proportional to the length of the subinterval. A consequence of this is that the radio failure rate stays constant throughout the entire interval.
- Rarity – the probability that one radio fails in a very small time interval is very small, and the probability that two fail in the same interval is even smaller, essentially zero.
- Independence – the number of radio failures in one time interval is independent of the number in any other non-overlapping subinterval.

Are these assumptions realistic? They might oversimplify an otherwise complex system, but should suffice for a rough estimate.

A sketch of this system is presented in Figure 1. The "DS" node represents a decision switch – which essentially "replaces" components as they fail.

Let X be a random variable representing the number of radios working out of n radios in the system. We know that under certain conditions X ~ Poisson (l) where l is the mean number of failures in a time period, t. In this problem, X ~ Poisson ($l = 3.6$).

Then, an expression representing the reliability of the system is

$$R(720) = P(X \leq n-1) = \sum_{x=0}^{n-1} \frac{e^{-l} l^{x}}{x!} \geq 0.99 \, .$$

Note the lower bound on the system reliability of 0.99. The smallest value of n for which this cumulative Poisson is satisfied with mean 3.6 is n = 9. So, we conclude that the battalion should deploy with nine radios.

DISCUSSION: We assume that once a radio fails, it does so in a catastrophic fashion – we cannot use it again. Is this realistic? Not at all. In practice, broken radios are repaired at some point and returned to the system. Does this invalidate the model? Maybe not. If we wanted our model to include repair we'd have to update it every time a repaired radio returned to the system.

What happens if the radios in the system do not operate and fail independently of each other? There are models that capture this dependence – and they are more complicated. The bottom line is that this model is a conservative one – it represents the worst-case scenario. That is, if a radio breaks, it can never be repaired.

Scenario 2: The Movement Officer

Suppose you are a mechanized infantry company executive officer involved in the planning of a battalion operation. This particular operation requires an extended mounted movement to an assembly area. You are tasked to plan a unit refuel stop during this movement. A consideration is the mean cruising range of your M2A3 Bradley fleet. You decide to investigate this mean range and plan a refuel at approximately three-fourths of this distance. You randomly select 10 of your battalion's M2A3's and measure their cruising range. The average of your sample is 46.04 km and the sample standard deviation is 7.966 km.

Problem 1: Construct a 95% confidence interval for the true population mean M2A3 cruising range.

Solution 1:

The first step in constructing such a confidence interval is to determine if the underlying distribution of the cruising range data is normally distributed. A normal probability plot of the data provides us with a tool to verify this assumption. A normal plot of the ten range data is shown in Figure 1. Essentially, if the data is linear, then it is reasonable to claim that the data comes from a normal distribution.

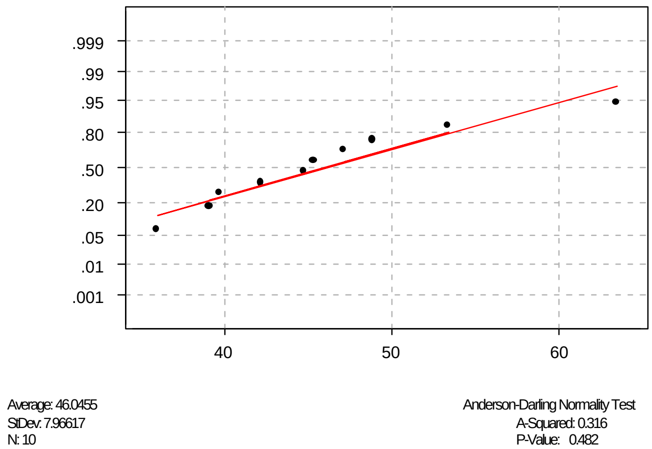

Figure 1. Normal Probability Plot for Range Data

Notice that the data appears to be linear. Therefore, we conclude that the cruising range distances are normally distributed. Since we do not know the population variance and the sample size is small (less than 30) we will construct a *t*-interval.

A 95% confidence interval for the true population mean cruising range m is of the form

$$\left(\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}\right) \tag{1}$$

where $\bar{x}$ is the sample mean, *s* is the sample standard deviation, *n* is the sample size and $t_{\alpha/2}$ is the *t* critical value with $n-1$ degrees of freedom and upper tail probability $\alpha/2$. This critical value can be regarded as a penalty function for small sample sizes. Crudely put, *t* critical values get larger with decreasing sample size, which has the effect of enlarging interval estimates. Conversely, a larger sample size results with a narrower, more precise interval (assuming all else remains the same).

Substituting the appropriate critical value and sample statistics into equation (1) yields the following 95% confidence interval:

$$\left(46.04 \pm 2.228\frac{7.966}{\sqrt{10}}\right) = (40.43, 51.65)$$

How do we interpret this interval estimate? Since this is a 95% confidence interval, roughly 95% of intervals constructed in a similar manner would contain the true value of m. In other words, we are 95% confident that the true value of m lies between 40.43 and 51.65 km.

What should you do with this information? Recall that the refuel point is to be located at three-fourths of the mean cruising range. A conservative approach would be to use the lower bound on m , or 40.43 km. This yields a planning distance of approximately 30 km. Now, you should find terrain suitable for a refueling operation at this distance.

Now let's look at the assumptions required of this interval estimate. There are really only two major assumptions – normality and that the sample was taken in a random manner. What happens if the data were non-normal? If it were possible to obtain a sample size greater than 30, then a large-sample confidence interval could be constructed that takes advantage of the Central Limit Theorem. If such a large sample were impossible, then one would resort to non-parametric methods.

The notion of a random sample cannot be stressed enough. If the M2A3's were selected in a non-random fashion, this confidence interval would be inappropriate to use due to the violation of independence assumptions.