Final: STAT 570

Philip Pham

December 9, 2018

Consider the failure time data in Table 1.

1. We describe a simple model for these data. Let p $(0 < p < 1)$ denote the weekly failure probability, i.e., the probability of failure during any week, and T the random variable describing the week at which failure occurred. Then T may be modeled as a geometric random variable:

$$\mathbb{P}\left(T = t \mid p\right) = \begin{cases} p\left(1-p\right)^{t-1}, & t = 1, 2, \ldots; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

   Let Y_t represent the number of components that fail in week t, $t = 1, 2, \ldots, N$, and Y_{N+1} the number of components that have not failed by week N.

   (a) Show that the likelihood function is

$$L\left(p\right) = \left[\left(1-p\right)^N\right]^{Y_{N+1}} \prod_{t=1}^N \left[p\left(1-p\right)^{t-1}\right]^{Y_t}. \tag{2}$$

   Solution: An individual component's failure week has distribution Geometric (p). The probability that a single component fails in week t is the probability that it survived $t-1$ weeks and failed on week t, which is $p\left(1-p\right)^{t-1}$ from Equation 1. There are Y_t such components, which gives us the factors for $t = 1, 2, \ldots, N$.

   The probability that a component fails at a later date is

$$\left(1-p\right)^N \sum_{k=1}^\infty p\left(1-p\right)^{k-1} = \left(1-p\right)^N \frac{p}{1-\left(1-p\right)} = \left(1-p\right)^N,$$

   which gives us the remaining factor. There are Y_{N+1} remaining components, so

$$L\left(p\right) = \left\{\prod_{t=1}^N \left[p\left(1-p\right)^{t-1}\right]^{Y_t}\right\} \times \left[\left(1-p\right)^N\right]^{Y_{N+1}}.$$

   (b) Find an expression for the MLE $\hat{p}$.

Solution: The score function is

$$
\begin{aligned}
S(p) &= \frac{\partial}{\partial p} \log L(p) \\
&= \frac{\partial}{\partial p} \left[N Y_{N+1} \log(1-p) + \sum_{t=1}^{N} Y_t \left(\log p + (t-1) \log(1-p) \right) \right] \\
&= -\frac{N Y_{N+1}}{1-p} + \sum_{t=1}^{N} Y_t \left(\frac{1}{p} - \frac{t-1}{1-p} \right) = -\frac{N Y_{N+1}}{1-p} + \sum_{t=1}^{N} Y_t \frac{1-pt}{p(1-p)}. \quad (3)
\end{aligned}
$$

Solving for $S(\hat{p}) = 0$, we find the MLE:

$$
\hat{p} \left(N Y_{N+1} + \sum_{t=1}^{N} t Y_t \right) = \sum_{t=1}^{N} Y_t \implies \boxed{\hat{p} = \frac{\sum_{t=1}^{N} Y_t}{N Y_{N+1} + \sum_{t=1}^{N} t Y_t}.} \quad (4)
$$

(c) Find the form of the observed information and hence the asymptotic variance of the maximum likelihood estimate (MLE).

Solution: Using Equation 3, the expected observed information is

$$
\begin{aligned}
I(p) &= \mathbb{E}\left[-\frac{\partial}{\partial p} S(p) \mid p \right] \\
&= \frac{N \mathbb{E}\left[Y_{N+1} \mid p \right]}{(1-p)^2} + \sum_{t=1}^{N} \mathbb{E}\left[Y_t \mid p \right] \left(\frac{1}{p^2} + \frac{t-1}{(1-p)^2} \right) \\
&= n \frac{N (1-p)^N}{(1-p)^2} + np \sum_{t=1}^{N} (1-p)^{t-1} \left(\frac{1}{p^2} + \frac{t-1}{(1-p)^2} \right) \\
&= n \left[\frac{(1-p)^N}{(1-p)^2} + \frac{1-(1-p)^N}{p^2} + \frac{(1-p)-(1-p)^N}{p(1-p)^2} \right] \\
&= \boxed{n \frac{1-(1-p)^N}{p^2 (1-p)},} \quad (5)
\end{aligned}
$$

where $n = Y_{N+1} + \sum_{t=1}^{N} Y_t$.
From Equation 5, the asymptotic variance of $\hat{p}$ is

$$
\operatorname{var}(\hat{p}) \approx \widehat{\operatorname{var}}(\hat{p}) = I(\hat{p})^{-1} = \frac{1}{n} \times \frac{\hat{p}^2 (1-\hat{p})}{1-(1-\hat{p})^N} \quad (6)
$$

by asymptotic normality of the MLE.

(d) For the data in Table 1, calculate the MLE, $\hat{p}$. the variance of $\hat{p}$, and an asymptotic 95% confidence interval for p.

Solution: The MLE can be calculated with Equation 4 to be $\boxed{\hat{p} = 0.354717.}$
The variance can be found with Equation 6 to be $\boxed{\widehat{\operatorname{var}}(\hat{p}) = 0.00016828.}$
If Φ is the cumulative distribution function for a standard normal, we can use asymptotic normality to find the 95% confidence interval as

$$
\left[\hat{p} + \Phi^{-1}(0.025) \sqrt{\widehat{\operatorname{var}}(\hat{p})}, \hat{p} + \Phi^{-1}(0.975) \sqrt{\widehat{\operatorname{var}}(\hat{p})} \right] = \boxed{[0.32929, 0.38014].}
$$

(e) We now consider a Bayesian analysis. The conjugate prior for p is a beta distribution, $\text{Beta}\,(a, b)$. State the form of the posterior with this choice. Give the form of the posterior mean and write as a weighted combination of the MLE and the prior mean.

Solution: By Bayes' rule, we know the posterior density is proportional to the likelihood times the prior. From Equation 2, we'll have

$$\begin{aligned} L(p) \times \left[p^{a-1}(1-p)^{b-1}\right] &= p^{a-1}(1-p)^{b+NY_{N+1}-1} \prod_{t=1}^{N} \left[p(1-p)^{t-1}\right]^{Y_t} \\ &= p^{a+\sum_{t=1}^{N} Y_t - 1}(1-p)^{b+\sum_{t=1}^{N}(t-1)Y_t + NY_{N+1}-1}, \end{aligned}$$

whose form we recognize as the integrand of beta function, so the posterior also has beta distribution, that is,

$$\begin{aligned} p \mid Y_1, Y_2, \ldots, Y_{N+1} &\sim \text{Beta}\left(a + \sum_{t=1}^{N} Y_t, b + \sum_{t=1}^{N}(t-1)Y_t + NY_{N+1}\right) \\ &= \frac{\Gamma(a'+b')}{\Gamma(a')\Gamma(b')} p^{a'-1}(1-p)^{b'-1}, \end{aligned} \tag{7}$$

where $a' = a + \sum_{t=1}^{N} Y_t$ and $b' = b + \sum_{t=1}^{N}(t-1)Y_t + NY_{N+1}$.
The posterior mean takes the form

$$\begin{aligned} \mathbb{E}\left[p \mid Y_1, Y_2, \ldots, Y_{N+1}\right] &= \frac{a'}{a'+b'} \\ &= \frac{a + \sum_{t=1}^{N} Y_t}{a + b + \sum_{t=1}^{N} tY_t + NY_{N+1}}. \end{aligned} \tag{8}$$

We have that the prior mean is $p_{\text{prior}} = \frac{a}{a+b}$. Equation 8 can be rewritten as

$$\boxed{\frac{(a+b)\,p_{\text{prior}} + \left(\sum_{t=}^{N} tY_t + NY_{N+1}\right)\hat{p}}{a + b + \sum_{t=}^{N} tY_t + NY_{N+1}},} \tag{9}$$

so the posterior mean is a convex combination of the prior mean and MLE.

(f) Suppose we wish to fix the parameters of the prior, a and b, so that the mean is μ and the prior standard deviation is σ. Obtain expressions for a and b in terms of μ and σ^2.

Solution: It is well known that the mean and variance of the $\text{Beta}\,(a, b)$ distribution are $\frac{a}{a+b}$ and $\frac{ab}{(a+b)^2(a+b+1)}$, respectively.
Solving equations

$$\begin{aligned} \frac{a}{a+b} &= \mu \\ \frac{ab}{(a+b)^2(a+b+1)} &= \sigma^2, \end{aligned}$$

we find that

$$a = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right] \tag{10}$$

$$b = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]. \tag{11}$$

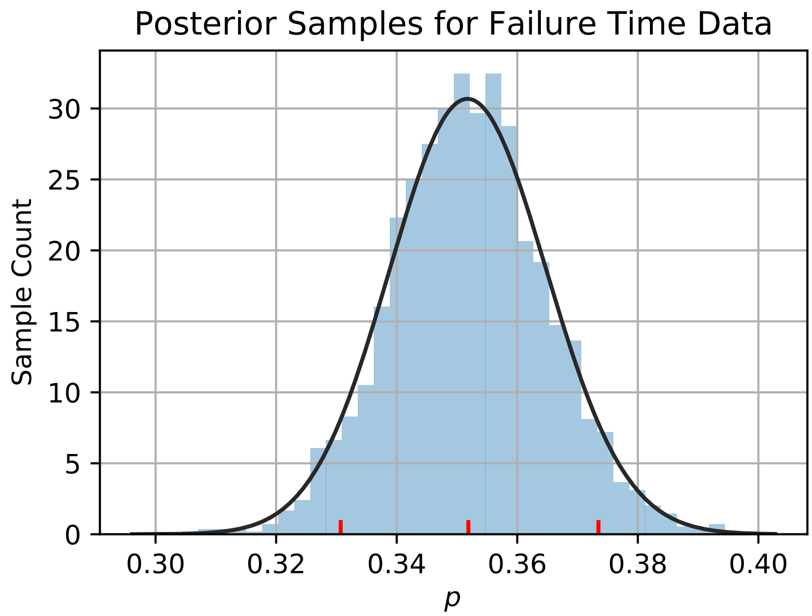

Figure 1: 2,048 samples drawn from the posterior in Equation 13. The red ticks denote the 5%, 50% and 95% quantiles.

(g) For the data in Table 1, assume we wish to have a beta prior with $\mu = 0.2$ and $\sigma = 0.08$. State the posterior for the prior corresponding to this choice and evaluate the posterior mean. Simulate samples from the posterior distribution. Provide a histogram representation of the posterior distribution and calculate the 5%, 50% and 95% points of the posterior distribution.

Solution: Apply Equations 10 and 11 with $\mu = 0.2$ and $\sigma = 0.08$, we find the prior:

$$p \sim \text{Beta}\left(4.8, 19.2\right). \tag{12}$$

Using Equation 7, we have find the posterior:

$$p \sim \text{Beta}\left(474.8, 874.2\right). \tag{13}$$

A histogram of samples drawn from the distribution in Equation 13 can be found in Figure 1. The 5%, 50%, and 95% posterior quantiles are 0.33070873, 0.35189124, and 0.37346975, respectively.

2. (a) A more complex likelihood for these data would assume that the i-th component had their own probability p_i, with the p_i's arising from a distribution $\pi\left(p\right)$. Show that

$$\mathbb{P}\left(T = t\right) = \mathbb{E}\left[\left(1-p\right)^{t-1}\right] - E[(1-p)^t], \tag{14}$$

and

$$\mathbb{P}\left(T > t\right) = \mathbb{E}\left[\left(1-p\right)^{t}\right]. \tag{15}$$

Solution: First let us find the survival function in 15.

$$\begin{aligned}
\mathbb{P}\left(T > t\right) &= \int_0^1 \mathbb{P}\left(T > t \mid p\right) \pi\left(p\right)\, \mathrm{d}p = \int_0^1 \left[\sum_{s=t+1}^{\infty} p(1-p)^{s-1}\right] \pi\left(p\right)\, \mathrm{d}p \\
&= \int_0^1 \left[p \sum_{s=0}^{\infty} (1-p)^s\right] (1-p)^t \pi\left(p\right)\, \mathrm{d}p \\
&= \int_0^1 \left[p \times \frac{1}{1-(1-p)}\right] (1-p)^t \pi\left(p\right)\, \mathrm{d}p = \int_0^1 (1-p)^t \pi\left(p\right)\, \mathrm{d}p \\
&= \mathbb{E}\left[(1-p)^t\right],
\end{aligned}$$

which proves Equation 15.
The probability mass function in Equation 14 follows:

$$\mathbb{P}\left(T = t\right) = \mathbb{P}\left(T > t-1\right) - \mathbb{P}\left(T > t\right) = \mathbb{E}\left[(1-p)^{t-1}\right] - \mathbb{E}\left[(1-p)^t\right].$$

(b) Obtain expressions for $\mathbb{P}\left(T = t \mid \alpha, \beta\right)$ and $\mathbb{P}\left(T > t \mid \alpha, \beta\right)$ with $\pi\left(\cdot\right)$ taken as the beta distribution, $\mathrm{Beta}\left(\alpha, \beta\right)$.

Solution: These follow from Equations 14 and 15.

$$\begin{aligned}
\mathbb{P}\left(T > t\right) &= \mathbb{E}\left[(1-p)^t\right] = \sum_{s=t}^{\infty} \mathbb{E}\left[p\left(1-p\right)^s\right] \qquad (16)\\
&= \sum_{s=t}^{\infty} \int_0^p p\left(1-p\right)^s \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} p^{\alpha-1}\left(1-p\right)^{\beta-1}\, \mathrm{d}p \\
&= \sum_{s=t}^{\infty} \int_0^p \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} p^{\alpha+1-1}\left(1-p\right)^{\beta+s-1}\, \mathrm{d}p \\
&= \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} \sum_{s=t}^{\infty} \frac{\Gamma\left(\alpha+1\right)\Gamma\left(\beta+s\right)}{\Gamma\left(\alpha+\beta+s+1\right)} = \alpha \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\beta\right)} \sum_{s=t}^{\infty} \frac{\Gamma\left(\beta+s\right)}{\Gamma\left(\alpha+\beta+s+1\right)} \\
&= 1 - \alpha \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\beta\right)} \sum_{s=0}^{t-1} \frac{\Gamma\left(\beta+s\right)}{\Gamma\left(\alpha+\beta+s+1\right)} \\
&= 1 - \frac{1}{B\left(\alpha,\beta\right)} \sum_{s=0}^{t-1} B\left(\alpha+1, \beta+s\right) = \frac{B\left(\alpha, \beta+t\right)}{B\left(\alpha,\beta\right)} = \frac{\Gamma\left(\alpha+\beta\right)\Gamma\left(\beta+t\right)}{\Gamma\left(\beta\right)\Gamma\left(\alpha+\beta+t\right)}
\end{aligned}$$

where B is the beta function, and we know $\mathbb{P}\left(T > 0\right) = 1$.
Plugging Equation 16 into Equation 14, one obtains

$$\mathbb{P}\left(T = t\right) = \frac{B\left(\alpha+1, \beta+t-1\right)}{B\left(\alpha,\beta\right)} = \alpha \frac{\Gamma\left(\alpha+\beta\right)\Gamma\left(\beta+t-1\right)}{\Gamma\left(\beta\right)\Gamma\left(\alpha+\beta+t\right)} \qquad (17)$$

for $t \in \mathbb{N}$.

(c) Using the previous part, write down the likelihood function $L\left(\alpha, \beta\right)$ corresponding to data $\{Y_t\}_{t=1}^{N+1}$.

Solution: Our model for T is different, so we can substitute Equations 17 and 16 into Equation 2: we'll have $\mathbb{P}\left(T = t\right)$ in place of $p\left(1-p\right)^{t-1}$ and $\mathbb{P}\left(T > N\right)$

in place of $(1-p)^N$.

$$L(\alpha,\beta) = [\mathbb{P}(T>N)]^{Y_{N+1}} \prod_{t=1}^{N} [\mathbb{P}(T=t)]^{Y_t} \tag{18}$$
$$= \left[\frac{B(\alpha,\beta+N)}{B(\alpha,\beta)}\right]^{Y_{N+1}} \prod_{t=1}^{N} \left[\frac{B(\alpha+1,\beta+t-1)}{B(\alpha,\beta)}\right]^{Y_t}.$$

(d) Find the MLEs $\hat{\alpha}$ and $\hat{\beta}$ for the data of Table 1.

Solution: From Equation 18, we can consider the log-likelihood function:

$$l(\alpha,\beta) = \log L(\alpha,\beta) \tag{19}$$
$$= -n\log B(\alpha,\beta) + Y_{N+1}\log B(\alpha,\beta+N) + \sum_{t=1}^{N} Y_t \log B(\alpha+1,\beta+t-1).$$

The score function is

$$S(\alpha,\beta) = \nabla l(\alpha,\beta) = \begin{pmatrix} S_\alpha(\alpha,\beta) \\ S_\beta(\alpha,\beta) \end{pmatrix} \tag{20}$$
$$S_\alpha(\alpha,\beta) = -n[\psi(\alpha)-\psi(\alpha+\beta)] + Y_{N+1}[\psi(\alpha)-\psi(\alpha+\beta+N)] + \sum_{t=1}^{N} Y_t[\psi(\alpha+1)-\psi(\alpha+\beta+t)],$$
$$S_\beta(\alpha,\beta) = -n[\psi(\beta)-\psi(\alpha+\beta)] + Y_{N+1}[\psi(\beta+N)-\psi(\alpha+\beta+N)] + \sum_{t=1}^{N} Y_t[\psi(\beta+t-1)-\psi(\alpha+\beta+t)],$$

p where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function.
Numerically solving Equation 20 for $S\left(\hat{\alpha},\hat{\beta}\right) = \mathbf{0}$, I obtain $\boxed{\hat{\alpha} = 1.413336}$ and $\boxed{\hat{\beta} = 1.38001102}$ for the MLEs.

3. (a) Show that the likelihood in Equation 2 can be written as a product of binomial distributions.

Solution: We can model the data as taking N draws from a binomial distribution. Following each draw, we discard the failures and make another draw if $t<N$:

$$L(p) = \prod_{t=1}^{N} \left[\binom{n-\sum_{s=1}^{t-1} Y_s}{Y_t} p^{Y_t}(1-p)^{n-\sum_{s=1}^{t} Y_s}\right] \tag{21}$$
$$= \prod_{t=1}^{N} \left[\binom{\sum_{s=t}^{N+1} Y_s}{Y_t} p^{Y_t}(1-p)^{\sum_{s=t+1}^{N+1} Y_s}\right],$$

which is equivlaent to Equation 2 up to a constant of proportionality,
In Equation 21, we have a product of binomial probability mass functions, where $Y_t \mid Y_1,\ldots,Y_{t-1} \sim \text{Binomial}\left(n-\sum_{s=1}^{t-1} Y_s, p\right)$.

(b) Fit the binomial model, and show that the estimate of the probability is identical to that under the previous MLE analysis. Obtain a 95% asymptotic confidence interval for p.

Solution: Since Equation 21 only differs from Equation 2 by a constant of proporitionality, the score function is also Equation 3. Thus, the MLE is same $\hat{p} = 0.354717$.

The observed information will also be the same. To calculate the expected observed information, we can use the law of total expectation and strong induction. For the base case $\mathbb{E}[Y_1] = np$. In general, $\mathbb{E}[Y_t] = np(1-p)^{t-1}$ for $t = 1, 2, \ldots, N$. For $t > 1$, we have

$$\begin{aligned}
\mathbb{E}[Y_t] &= \mathbb{E}[\mathbb{E}[Y_t \mid Y_1, \ldots, Y_{t-1}]] = \mathbb{E}\left[p\left(n - \sum_{s=1}^{t-1} Y_s\right)\right] \\
&= p\left(n - \sum_{s=1}^{t-1} \mathbb{E}[Y_s]\right) = p\left(n - \sum_{s=1}^{t-1} np(1-p)^{s-1}\right) \\
&= np\left(1 - p\sum_{s=0}^{t-2}(1-p)^r\right) = np\left(1 - p\frac{1-(1-p)^{t-1}}{p}\right) \\
&= np(1-p)^{t-1}, \qquad (22)
\end{aligned}$$

which is same as it was under the geometric model.

For Y_{N+1}, we have

$$\begin{aligned}
\mathbb{E}[Y_{N+1}] &= \mathbb{E}[\mathbb{E}[Y_{N+1} \mid Y_1, Y_2, \ldots, Y_N]] = \mathbb{E}\left[n - \sum_{t=1}^{N} Y_t\right] \\
&= n - \sum_{t=1}^{N} \mathbb{E}[Y_t] = n - np\sum_{t=1}^{N}(1-p)^{t-1} \\
&= n - np\frac{1-(1-p)^N}{p} = (1-p)^N, \qquad (23)
\end{aligned}$$

which is also the same as under the geometric model. Therefore, the expected observed information is the same as Equation 5.

Then, the asymptotic 95% confidence interval for p is also $[0.32929, 0.38014]$.

(c) Obtain Pearson residuals and comment on the fit of the model, using any plots you feel are appropriate.

Solution:

(d) Fit a binomial model you feel is appropriate.

Solution:

Time (weeks), i	Failures, y_i	Temperature, x_i
1	210	24.0
2	108	26.0
3	58	24.0
4	40	26.0
5	17	25.0
6	10	22.0
7	7	23.0
8	6	20.0
9	5	21.0
10	4	18.0
11	2	17.0
12	3	20.0
> 12	15	

Table 1: Time until failure for $n = 485$ components, along with average weekly temperature.