# CHAPTER IV

# THE MACRO-LEVEL PARTICIPATION MODELING

After modeling the *micro* PL-site choice decision, the next step is to model the determinants of the total number of trips a licensed angler takes during a season.. It is theoretically possible to model jointly the discrete product-line/site choices and the total participation decision; however, the data and computational requirements for the correct treatment of the corner- solutions implied by zero trips of certain categories makes an integrated utility-theoretic model practically infeasible. Essentially, researchers appear to face a trade-off: they either implement a "utility-theoretic" framework that does not properly model the statistics of the corner solutions; or they model the *micro* and *macro* decisions in separate models that may address the corner-solution problem but do not form an integrated utility-theoretic framework.

In this chapter, we first discuss variants of the former approach, in which total participation is modeled as the sum of independent participation decisions made on each choice occasion throughout the season. We then summarize the Bockstael, Hanemann and Strand (1986) critique of this approach and their alternative proposal to model directly the corner solution. Finally, we develop our own model, which is in the spirit of the second approach.

Due to severe data limitations at the total participation level, our model is substantialiy different from the standard treatment in the literature. We do not know the total number of season trips: our *macro-level* information is limited to the duration between trips, and this variable is censored because we only observe the duration from last trip to the survey return date, not to the subsequent trip. By incorporating a key result from stochastic renewal theory in our modeling, we are able to estimate the determinants of the between-trip durations with a stochastic renewal model and then to derive the total number of trips in a season from the duration model. To accommodate the different trip durations, we develop a competing risks model; to allow for variations in site quality throughout the open-water season: we incorporate time-varying covariates in the model.

**Participation as the Sum of Independent Trip Decisions**

To integrate the participation decision with the PL-site decision in one framework: the utility level $u_0$ associated with not taking a trip on the current choice occasion has to be specified. Individuals are hypothesized to determine whether to take a trip by comparing $u_0$ with the expected maximum utility of taking a trip. A trip will consequently be taken if and only if

$$I^* = max\{u_{m,j};\ \forall(m,j)\} > u_0.$$

For empirical estimation, the relationship between the random element $\epsilon_0$ associated with the no-trip utility $u_0$ and the other random terms $\epsilon_{(l,j)}$ has to be specified. Bockstael et. al. (1986) derive the repeated NMNL model by extending the generalized extreme value (GEV) distribution (III.8) employed for modeling of PL-site choices in the previous chapter to the joint distribution

$$F(\epsilon_0;\ \epsilon_{(m,j)};\ \forall(m,j)) = \exp\left(-e^{-\epsilon_0} - \left[\sum_m \left(\sum_j \exp(-\epsilon_{(m,j)}/\theta)\right)^{\theta/\sigma}\right]^{\sigma}\right). \qquad \text{(IV.18)}$$

The parameter $\theta$ is still the common index of correlation of the random terms *(m, j)* for sites under PL $m$.[1] The participation decision is illustrated in figure IV.1

The probability that an individual will take a trip in period *t* with the given GEV distribution (IV.18) can be shown to be

$$\pi_*^t = \frac{\exp(\sigma I^*)}{\exp(u_0) + \exp(\sigma I^*)}, \qquad \text{(IV. 19)}$$

while the probability of no participation is

$$\pi_0^t = 1 - \pi_*^t = \frac{\exp(u_0)}{\exp(u_0) + \exp(\sigma I^*)}. \qquad \text{(IV.20)}$$

Because the micro-level decisions regarding the trips are nested within the participate/ no-participation decision. the participation choice is not characterized by the IIA restriction. With this model. the *micro* PL-site choices and *macro* participation decision can be estimated simultaneously if the participation and PL-site choice data are available for all periods. However: the framework is one of *repeated* choices, where the decision on any choice occasion is independent of the choices on all other choice occasions

The Alaska fisheries study by Carson: Hanemann, Gum and Mitchell (1987) is the only one to our knowledge that estimates a repeated nested logit model with complete trip information throughout a season [2] In their model, a sport fishing angler can take up to a maximum of three trips in a single week. Let $v_{it}^k$ denote the utility an individual *i*. can receive from taking *k* trips during week *t*. Then the participation probability for having *m* (= 0, 1,2,3) fishing trips is

$$\pi_{it}^m = \frac{\exp(v_{it}^m)}{\sum_{k=0}^3 \exp(v_{it}^k)}.$$

They then maximize the likelihood function

$$\mathcal{L} = \prod_i \prod_t \pi_{it}^{k*},$$

---

[1] When $\sigma = 1$. the no-trip option is treated as just another alternative, and the model degenerates to a standard MNL (i.e., it is not nested over the participate/no-participate decision.)

[2] Stating that fishing opportunities in Alaska change dramatically over a season. Carson et. al. incorporate weekly choices in the mode! and allow the covariates to vary from week to week.

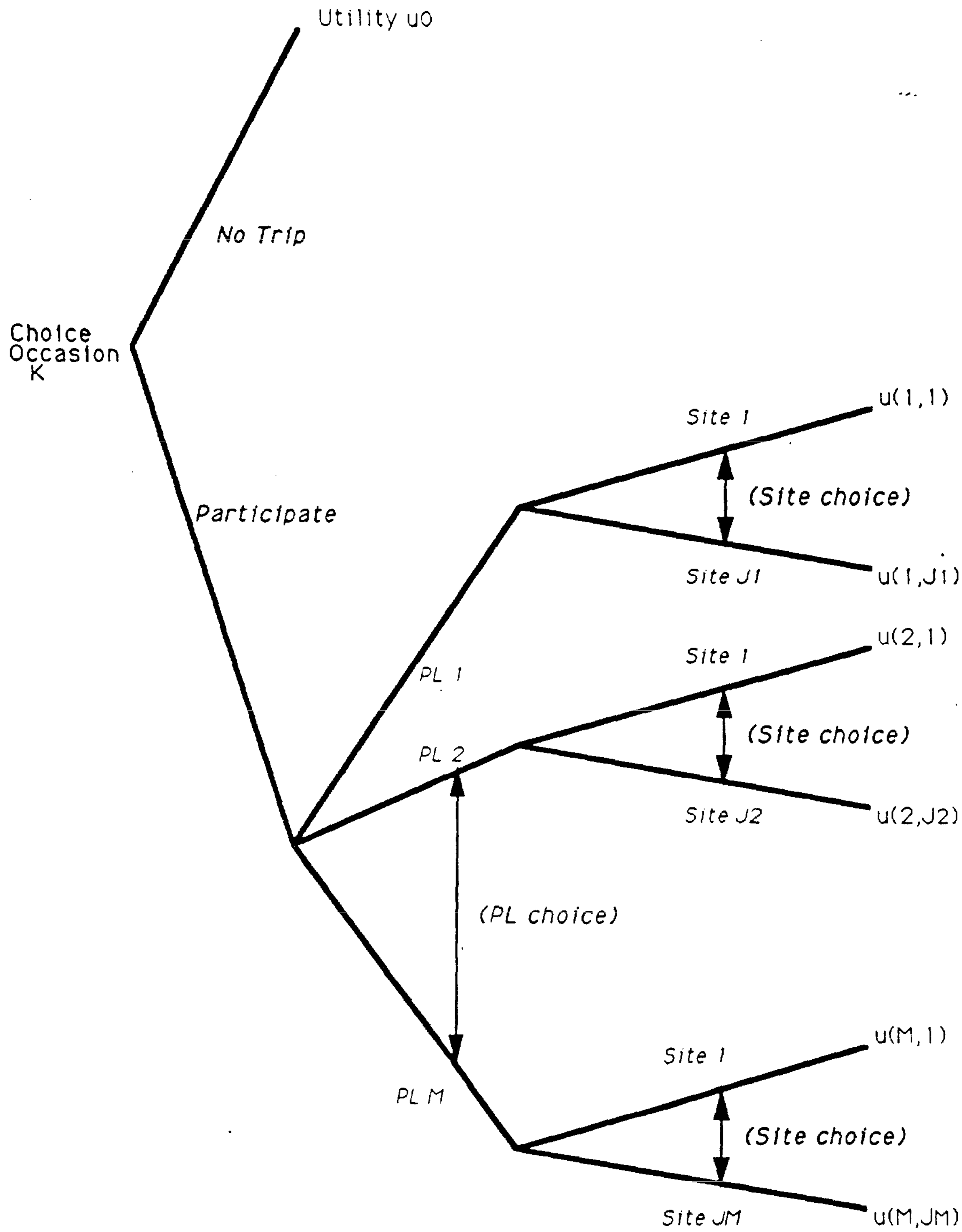


Figure IV.1: The choice occasion participation decision where $k^*$ is the number of trips taken by $i$ in week $t$. Because the mean number of trips during a week taken by those with more than two trips was 3.63, the expected number of seasonal trips is calculated as

$$Q_i = \sum_t \left[\pi_{it}^1 + 2\pi_{it}^2 + 3.63\,\pi_{it}^3\right].$$

More typically, researchers know the total number of trips in a season: but only have detailed trip information about one trip. The recent paper by Morey et al. (1991) provides a good example of a model designed for such data.[3] With the available site choice information $J$. the probability density for the micro decision can be formulated as

$$f(J) = \prod_i \pi_i^*$$

where $\pi_i^*$ is the probability that $i$ would choose his or her actual destination $J$, in the MNL setup. The total number of trips $K$ is then used to derive the combinatorial participation probability density

$$g(K) = \prod_i \left\{ \left[\frac{T!}{K_i!\,(T-K_i)!}\right] (\pi_i^0)^{T-K_i} (1-\pi_i^0)^{K_i} \right\}$$

where $\pi_i^0$ is the probability of $i$ not taking a trip during the period that the site decision is known, $K_i$ is the number of total trips taken by $i$, and $T$ is the number of choice occasions in the whole season. Morey et al. maximize the complete likelihood function

$$\mathcal{L}(J, K) = f(J) \cdot g(K).$$

The expected total number of trips an individual $i$ would take when there are $T$ choice periods in a year is simply $[T \cdot (1 - \pi_i^0)]$.

---

[3] Their model allows for different distributions for the participation and site choice decisions. However, as they note, it is neither a repeated standard MNL nor a repeated nested MNL because it does not incorporate a stochastic component in the indirect utility function conditional upon no participation.

**The Critique and an Alternative Proposal**

Bockstael et al. (1986, pp. 185-86 and 1987, p. 13) critique. the class of participation models highlighted above; on the grounds that they characterize total participation simply as a sum of independent decisions on each choice occasion. In particular, they criticize the models because the occurrence of a season with no trips happens merely by accident: the probability of no participation throughout the season is simply the product of the probabilities of no participation on each choice occasion.

When individuals no longer choose "interior solutions" to the utility maximization problem. then the well-behaved, continuous properties of neoclassical demand theory no longer hold. One must instead model the probability statements with Kuhn-Tucker conditions. The problem with the models developed above is that they do not incorporate the discontinuity of the indirect utility functions as individuals switch among different consumption regimes.

A switching regressions model is appropriate to capture statistically the different regimes. Unfortunately the dimensionality of the problem is generally one less than the number of commodities not consumed. Given the level of detail in the random utility models and the many expected corner solutions for most individuals, it appears practically infeasible to integrate over the number of cumulative distribution functions that would be required with either the direct or indirect Kuhn-Tucker conditions. Bockstael et. al. conclude that "without attempting to estimate the corner solutions, there appears to be no consistent way to link independent discrete choice decisions and a macro decision for total trips with a common underlying utility maximization framework" (1986, p. 186).)

They propose an alternative method in which the expected number of trips to all sites over the season T may be interpreted as:

$$E[T] = E[T|T > 0]Pr[T > 0]$$

where the second term on the right-hand side is the probability that the individual engages in any recreation during the season. The equation can be estimated with Tobit, Cragg or Heckman selection procedures. In this method the decision to ever-participate is estimated directly, allowing the researcher to characterize the role of factors such as poor health, adverse financial conditions, or unusually heavy working loads.

**The Stochastic Renewal Approach**

A major data problem we confront in modeling total trip participation is that we do not know the total number of recreational fishing trips. Therefore, we cannot employ the conventional estimating approaches discussed above! in which the dependent variable is the total number of trips. To accommodate our special data needs, we have developed an alternative framework for modeling the decision about total participation.

As noted above, our information about trips is limited to the duration between trips, and this variable is censored: we only observe the duration from last trip to the survey return date, not to the subsequent trip. Consequently, we estimate a duration mode! of the period between trips, from which we then calculate the expected number of trips. We draw upon a key result in stochastic renewal theory to adapt the duration model to handle the right-censored data.

We incorporate time-varying covariates in the duration model. In addition, we know the length of the most recent trip taken by an angler, which allows us to estimate anglers' demand for trips of different durations. To include this information: we develop a competing risks framework in which individuals may end their spell of no-trips by choosing any one of three trip-lengths (day; weekend, 2-4 days; or vacation, 5+ days). Finally we have some individuals in the sample who took no trips during the period about which they were questioned. We develop procedures to model this right- and left-censored duration data.

The development of the full model requires an extended discussion below due to the many features that have been incorporated. To start, we outline the basic stochastic renewal model, in which the number of trips taken during a period of time is a renewal process. We develop the participation model first for the special case of an exponentially-distributed duration variable and Poisson-distributed trip counts, because the intuition of the model is more accessible with the simpler formulas of the special case. In the next sections of the chapter, we extend the exponential-Poisson model to accommodate: right-censored inter-trip duration data; time-varying covariates: competing risks; and right- and left-censored trip durations.

We then develop the model using the Weibull distribution for inter-trip durations, in order to relax the special assumptions of the exponential-Poisson case. The subsequent four sections follow a similar pattern to the discussion of the exponential model.

## The Stochastic Renewal Process

We assume that the number of trips taken during a period of time is a *renewal process:* in which the between-trip durations are independently and identically distributed. Let $T$ be the random variable of independent time spells between successive trips[4] taken by individual *i.* Denote the probability density function (PDF) of $T$ by

$$f(t) \equiv \lim_{\delta \to 0-} \frac{\text{Prob}\{t \leq T < t + \delta\}}{\delta}$$

and the cumulative density function (CDF) by

$$F(t) \equiv \text{Prob}\{T < t\} = \int_0^t f(s)\, ds.$$

[4] We ignore the spell of a trip. There are two possible interpretations. First. trips are assumed to be instantaneous events for modeling convenience. Second, when an angler decides to begin a trip on a certain day, he/she decides simultaneous!y not to have another trip during the duration, of the trip.

The *survival* function *S(t):* which yields the probability that the duration *T* will be longer than *t,* is defined as

$$S(t) \equiv \text{Prob}\{T \geq t\} = 1 - F(t) = \int_t^{\infty} f(s)\, ds$$

Hence $S(0) = 1$ and $S(\infty) = 0$, while $F(0) = 0$ and $F(\infty) = 1$. Another conceptually useful function, the *hazard rate* function, is defined as

$$h(t) \equiv \lim_{\delta \to 0+} \frac{\text{Prob}\{t \leq T < t + \delta \mid t \leq T\}}{\delta} = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt},$$

which measures the conditional probability of taking the next trip at time *t,* given that no trip has been taken before *t.* A model with a constant hazard rate is said to be *duration independent.*

These functions will be used below to derive the maximum likelihood estimator of *T.* Note that the PDF and hazard rate are just two different ways of describing the same probability distribution. Given the PDF, the hazard rate function can be uniquely determined, and vice versa.

Let us first look at the special case of the Poisson- Exponential distribution. As Kiefer (1988, p. 652) points out, the exponential distribution is simple to work with and to interpret. However. it may be too restrictive in that no duration dependency is allowed. More flexible distributions, such as Weibull, [5] will be considered next.

## The Exponential Distribution

Suppose the time spell *T,* between successively taken trips *k* and *(k* + 1) by individual *i* follows the exponential distribution with parameter $\lambda_i > 0$. All durations are independently distributed. The PDF for $T_i (\geq 0)$ is then

$$f_i(t) = \lambda_i e^{-\lambda_i t}$$

[5] The exponential distribution. is a special case of the Weibull distribution. Thus we can conduct a nested model test to check the appropriateness of using the exponential distribution.

and the corresponding CDF is

$$F_i(t) = 1 - e^{-\lambda_i t}.$$

Thus, the hazard rate is

$$h(t) = \frac{f_i(t)}{S_i(t)} = \lambda_i.$$

Since it is constant for an individual *i* at any time $t > 0$; it is called the *memoryless* property which is unique to the exponential distribution. We assume $\lambda_i = e^{\beta X_i} > 0$, that is, the parameter $\lambda_i$ is a log-linear function of $X_i$, which consists of both personal and site variables. $\beta$ is assumed to be identical across al! individuals.

Given observations of the *completed* durations $t$ for each individual *i* in the data, the log likelihood function *LL* can be formed as follows

$$\begin{aligned} LL &= \sum_i \log f_i(t_i) \\ &= \sum_i \log(\lambda_i e^{-\lambda_i t_i}) \\ &= \sum_i [\log \lambda_i - \lambda_i t_i] \\ &= \sum_i [\beta X_i - t_i e^{\beta X_i}]. \end{aligned}$$

The maximum likelihood estimates $\hat{\beta}$ can then be obtained by maximizing the log likelihood function *LL* with respect to $\beta$ and setting the first derivatives to zero. This gives us

$$\left.\frac{\partial LL}{\partial \beta}\right|_{\beta=\hat{\beta}} = \sum_i \left[X_i - t_i X_i e^{\hat{\beta} X_i}\right] = \sum_i \left[X_i (1 - t_i e^{\hat{\beta} X_i})\right] \equiv 0.$$

Note that the expected duration $E[T_i]$ for the exponential distribution $f_i(t)$ given above is just

$$E[T_i] \equiv \int_0^\infty f_i(t)\, t\, dt = \frac{1}{\lambda_i} = \frac{1}{e^{\beta X_i}}.$$

Our goal, however, is the counting process $N_i(S)$, which records the number of trip occurrences in a time period *S*. In this case, the counting process $N_i(S)$ corresponding to the exponentially distributed between-trip durations is Poisson distributed [6] with

[6] See Ross (1963, pp. 35-36) or Taylor and Karlin (1984, pp. 188-89) for proof.

the discrete PDF:

$$\text{Prob}\{N_i(S) = n\} = \frac{(\lambda_i S)^n e^{-\lambda_i S}}{n!}$$

and expected value

$$E[N_i(S)] = \lambda_i S.$$

The parameter $\lambda_i = e^{\beta X_i}$ (calculated for each individual *i)* has an intuitive interpretation of being the expected number of trips individual *i* will take in one unit of time. Taking days to be the unit of time, the expected number of trips in our Poisson process thus can be readily calculated as the number of days *S* in a fishing season multiplied by $\lambda_i$ for each individual *i*.

To justify the use of the Poisson-Exponential distribution, we have to refer back to the basic postulates of a Poisson process. It has been proved that a counting process $\{N(S), S \geq 0\}$ is Poisson distributed with parameter $\lambda(>0)$ if the following postulates are satisfied: [7]

1. $N(0)=0$. That is, no trip has occured prier to the start of the time interval $S = 0$.

2. The time intervals between trips are stationary and independently distributed.

   A counting process is *stationary* if the distribution of the number of occurrences (in this case, trips) in any interval of time depends only on the length of the time interval. It is *independent* if the number of occurrences taken in disjoint time intervals are independent.

3. $P\{N(s)=1\} = \lambda s + o(s)$ as $s \rightarrow 0$ .

   This posits that the probability of having exactly one trip in a very short time interval *s* is proportional to the length of the interval. The function *o(s)* is defined to have the property that

   $$\lim_{s \rightarrow 0} \frac{o(s)}{s} = 0.$$

4. $P\{N(s) \geq 2\} = o(s)$ as $s \rightarrow 0$.

   This posits that the probability of having at least two trips in a very short time period *s* is very small and can be ignored.

---

[7] See Ross (1983, pp. 32-34) or Taylor and Karlin (1984, pp. 181-184) for proof.

These are reasonable assumptions to make regarding fishing-trip behavior. The advantage of using the Poisson-Exponential pair is that no extra work is necessary to calculate expected total trips $E[N(S)]$.

## Data Limitation

A severe limitation with our data is that we do not observe any completed spell $T_i$. What we have are only the date the last trip began $(L_i)$ and the date the questionnaire was returned $(R_i)$. Consider $R_i$ to be a random censoring point which truncated the spell in question before it was completed.[8] Let the unobserved date of the next trip taken by $i$ (after the questionnaire return date $R_i$) be $V_i$, which would have been the endpoint of the sampled duration if it had not been terminated prematurely.

We can then define the following three random variables

$$\begin{aligned} \textit{Age:} \quad & A_i \equiv R_i - L_i \\ \text{Residual life:} \quad & Y_i \equiv V_i - R_i \\ \text{Life of sampled observation:} \quad & B_i \equiv V_i - L_i. \end{aligned}$$

Of these three variables, only age is observed. See figure (IV.2) to illustrate the relationship among the three variables. This is illustrated in figure (IV.2).

It is well known in the stochastic processes literature that the expected length of an inspected duration $B_i$ is greater than that of *a population* duration $T_i$, due to the greater likelihood of sampling longer intervals. This is called *length-biased sampling*. To distinguish between the sampled interval $B_i$ and the population duration $T_i$, the latter will be called *normal life* in the discussion below, following the convention in the stochastic processes literature.

For the exponential duration case, it can further be shown that (1) both $A_i$ and

---

[8] The censoring mechanism should be independent of the last trip date $L_i$.

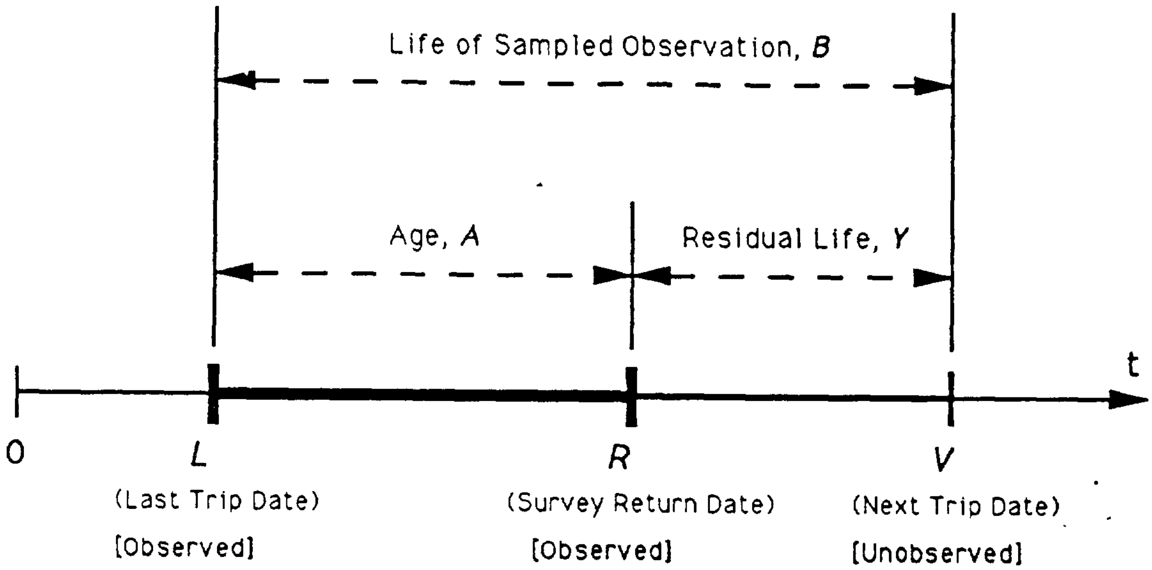


Figure IV.2: The truncated between-trip duration

$Y_i$ have the same distribution as $T_i$ if sampling occurs after the renewal process has been ongoing for a long time. and (2) the length of the sampled interval containing the sampling point $R_i$ is expected to be twice that of a normal life interval $T_i$, known as the famous *inspection paradox.*[9] Therefore, in the limit,

$$E[A_i] = E[Y_i] = E[T_i] = \frac{1}{\lambda_i}.$$

The solution to our limited data problem will then involve the following steps:

1. We can first estimate the parameters of the age $(A_i)$ distribution using the available age data.

2. Since age $A_i$ and normal life $T_i$ have the same distribution, the parameters obtained for $A_i$ are exactly those for $T_i$.

[9] See Taylor and Karlin (1984), pp. 282-84.

3. The seasonal total trips can then be calculated using the estimate $\hat{\lambda}_i$.

The procedure outlined above is not limited to the exponential case. We can always derive the distribution of age *A* from any given distribution function for the regular life *T*. Therefore it is genrally the case that parameters in the normal life distribution can be estimated with age data/distribution, if normal life data/distributions are not available.

Time-Varying Covariates

So far we have assumed that each individual *i* has a constant exponential parameter $\lambda_i$ across time. Since conditions at recreation sites (part of the $X_i$ vector) often vary during a season. both $X_i$ and $\lambda_i$ should be generalized to be time indexed. The time-varying elements in $X_i$ are called *time-varying covariates*. The probability of an individual taking trips at different times will hence depend on the time-dependent explanatory variables $X_i(t)$.

In the following discussion. a mere statement of interval *t* presupposes implicitly a starting point of time 0. The notation $t_s^r$ will be used when necessary to indicate that the duration *t* runs from time *s* to time $r$, instead of from 0 to *t*. The endpoints are important now since the parameters $\lambda_{it}$ are time dependent.

In the case of time-varying parameters, the CDF of $T_i (\geq 0)$ becomes

$$F_i(t) = 1 - \exp\left(-\int_0^t \lambda_i(s)\, ds\right)$$

and the PDF becomes

$$f_i(t) = \lambda_i(t) \exp\left(-\int_0^t \lambda_i(s)\, ds\right)$$

with $\lambda_i(t) > 0$ at any moment of time *t*. The distribution functions still necessarily have the properties that $F_i(0) = 0$ and $F_i(\infty) = 1$. The instantaneous hazard rate $h_i(t) = \lambda_i(t)$ depends solely on the value of parameter $\lambda_i$ at time $t$.

The probability that an individual $i$ has a completed duration $T_i$ greater than $t$ and less than $r\,(t \leq r)$ is then

$$\text{Prob}\{t \leq T_i \leq r\} = \exp\left(-\int_0^t \lambda_i(s)\,ds\right) - \exp\left(-\int_0^r \lambda_i(s)\,ds\right) \geq 0.$$

The corresponding Poisson counting process can be shown to have the distribution

$$\text{Prob}\{N_i(S) = n\} = \frac{\left[\int_0^S \lambda_i(u)\,du\right]^n \exp\left[-\int_0^S \lambda_i(u)\,du\right]}{n!} \equiv \frac{(\lambda_i^*)^n e^{-\lambda_i^*}}{n!}$$

where $\lambda_i^* \equiv \int_0^S \lambda_i(u)\,du$. The proof is only a generalization of that for the time-independent version. The expected number of trips taken by $i$ during a season from day 0 to day $S$ is then

$$E[N_i(S)] = \int_0^S \lambda_i(u)\,du = \lambda_i^*.$$

Given observations of the *completed* durations $t_i$ (running from $L_i$ to $V_i$) and assuming $\lambda_i(t) = e^{\beta X_i(t)}$ for time $t$, we can construct the log likelihood function $LL$ as follows

$$\begin{aligned} LL &= \sum_i \log f_i(t_{i,L_i}^{V_i}) \\ &= \sum_i \left[\log \lambda_i(V_i) - \int_{L_i}^{V_i} \lambda_i(s)\,ds\right] \\ &= \sum_i \left[\beta X_i(V_i) - \int_{L_i}^{V_i} e^{\beta X_i(s)}\,ds\right]. \end{aligned}$$

The discrete time version of the log likelihood function $LL$ is

$$LL = \sum_i \left[\beta X_i(V_i) - \sum_{s=L_i}^{V_i} e^{\beta X_i(s)}\right].$$

The maximum likelihood estimate $\hat{\beta}$ is then the solution to the equation

$$\frac{\partial LL}{\partial \beta} = \sum_i \left[X_i(V_i) - \sum_{s=L_i}^{V_i} X_i(s)\,e^{\beta X_i(s)}\right] \equiv 0.$$

The expected number of trips of individual $i$ during a season from day 0 to day $S$ is then simply

$$E[N_i(S)] = \sum_{s=0}^{S} \lambda_i(s) = \sum_{s=0}^{S} e^{\beta X_i(s)}$$

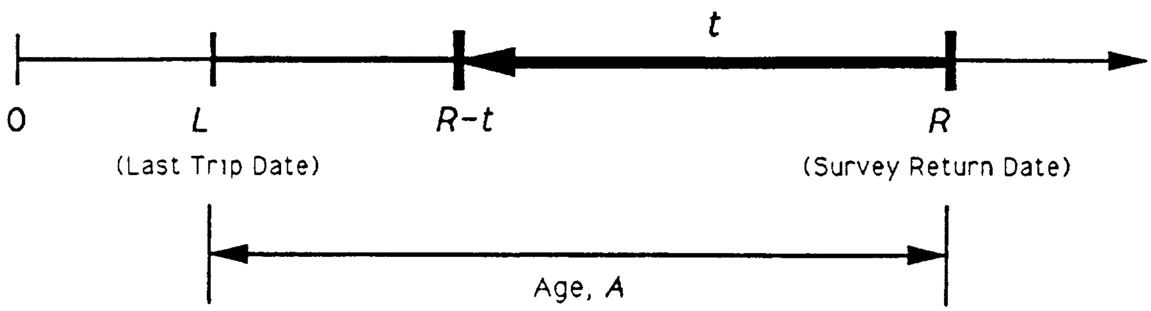


Figure IV.3: Derivation of the age distribution

One more issue wemust address is the estimation of parameters $\beta$ when we only have age data instead of completed durations. This can be done by first deriving the distribution of age $A_i$ as follows.

$$\begin{aligned}\text{Prob}\{A_i \geq t\} &= \text{Prob}\{N(R_i) - N(R_i - t) = 0\} \\ &= \text{Prob}\{N(t|_{R_i-t}^{R_i}) = 0\} \\ &= \exp\left(-\int_{R_i-t}^{R_i} \lambda_i(u)\, du\right)\end{aligned}$$

for $t \leq A_i$. Note that $\text{Prob}\{A_i \geq t\} = 0$ for $t > A_i$. This is shown in figure (IV.3). Basically, we are looking backwards from the given censoring point $R_i$ to find the time the last trip occurred, not looking forwards in search of the next trip date.

Therefore, the CDF of age $A_i$ is

$$F_A(t|_{R_i-t}^{R_i}) = \text{Prob}\{A_i \leq t|_{R_i-t}^{R_i}\} = 1 - \exp\left(-\int_{R_i-t}^{R_i} \lambda_i(u)\, du\right)$$

and the probability density of having an age $A_i = t$ from the date of last trip $L_i$ (= $R_i - t$) to the survey return date $R_i$ is

$$f_A(t \mid {}^{R_i}_{L_i}) = \frac{dF_A(t \mid {}^{R_i}_{L_i})}{dt} = \lambda_i(L_i) \exp\left(-\int_{L_i}^{R_i} \lambda_i(u)\, du\right)$$

The parameters $\beta$ can hence be estimated using age data and the age distribution $f_A(t)$. They are exactly those that appear in the normal life distribution. The total trips can then be calculated as

$$E[N_i(S)] = \sum_{s=0}^{S} e^{\beta X_i(s)}$$

## Competing Risk Participation

To further enrich our participation model, consider the more complicated situation where an individual can take either a day trip, a weekend trip. or a vacation trip. Trips of unequal lengths are considered to represent different. substitute commodities because their utility trade-offs may be different. In other words, long trips are taken for purposes somewhat different from those of short trips. Here we'll think of them as different *types* of events (or risks) that would terminate the durations and index them as $d$ = 1, 2, 3. In the following discussion, individual index $i$ is omitted for notational simplicity.

The single-type exponential specification can now be extended by defining the *type-specific hazard rate* as

$$h_d(t) \equiv \lim_{\delta \to 0+} \frac{\text{Prob}\{t \leq T < t + \delta,\ D = d \mid t \leq T\}}{\delta} = \lambda_d(t) = e^{\beta_d X_i(t)}.$$

This is the probability density of an individual $i$ taking a type $d$ trip immediately after time $t$ conditional on no trip occurrence of any type up to time $t$. The *non-type-specific hazard rate* of individual $i$ taking any type of trip at $t$ is then simply

$$h(t) \equiv \lim_{\delta \to 0-} \frac{\text{Prob}\{t \leq T < t + \delta \mid t \leq T\}}{\delta} = \sum_d \lambda_d(t) = \sum_d e^{\beta_d X_i(t)}$$

since trips of different types cannot be taken simultaneously on a choice occasion, i.e., they are mutually exclusive. It must necessarily follow that $0 \leq h_d(t) \leq 1$ for all types $d$ and $0 \leq h(t) \leq 1$.

The *non-type-specific CDF* of at least one trip of any type up to time $t$ is

$$F(t) = 1 - \exp\left(-\int_0^t h(s)\, ds\right) = 1 - \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right)$$

and the *non-type-specific survival* function of no trip at all from time 0 to time $t$ is

$$S(t) = 1 - F(t) = \exp\left(-\int_0^t h(s)\, ds\right) = \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right).$$

The *type-specific CDF* of having at least one *type d* trip up to time $t$ is

$$F_d(t) = 1 - \exp\left(-\int_0^t h_d(s)\, ds\right) = 1 - \exp\left(-\int_0^t \lambda_d(s)\, ds\right)$$

and the *type-specific survival function* of no *type d* trip from time 0 to time $t$ is

$$S_d(t) = 1 - F_d(t) = \exp\left(-\int_0^t h_d(s)\, ds\right) = \exp\left(-\int_0^t \lambda_d(s)\, ds\right).$$

do not untrue and

The *non-type-specific PDF*

$$f(t) = \frac{dF(t)}{dt} = \left[\sum_d \lambda_d(t)\right] \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right) = h(t)\, S(t)$$

measures the probability of having no trip up to time $t$ and then a trip of any type at time $t$. The *type-specific PDF* below

$$f_d(t) = \lambda_d(t) \exp\left(-\int_0^t \sum_j \lambda_j(s)\, ds\right) = h_d(t)\, S(t)$$

gives us the probability of having no trip before $t$ and then a *type d* trip at $t$. Note that by definition

$$f_d(t) \neq \frac{dF_d(t)}{dt}.$$

As in the previous sections, let $R_i$ and $L_i$ be the observed censoring date and last trip date respectively. Also let $D_i$ be the type of the last trip taken by individual $i$ in our sample data. The likelihood function, incorporating the time-varying covariate results, is

$$L = \prod_i f_{D_i}(t_i | {R_i \atop L_i}) = \prod_i \left[ \lambda_{D_i}(L_i) \exp\left( - \int_{L_i}^{R_i} \sum_d \lambda_d(s)\, ds \right) \right] \tag{IV.21}$$

And the log likelihood function is

$$LL = \sum_i \left[ \log \lambda_{D_i}(L_i) - \int_{L_i}^{R_i} \sum_d \lambda_d(s)\, ds \right]. \tag{IV.22}$$

Alternatively, we can write the likelihood function $L$ as

$$L = \prod_i \left[ \lambda_{D_i}(L_i) \exp\left( - \int_{L_i}^{R_i} \lambda_{D_i}(s)\, ds \right) \prod_{d \neq D_i} \exp\left( - \int_{L_i}^{R_i} \lambda_d(s)\, ds \right) \right].$$

Durations corresponding to all types except the chosen type $D_i$ are regarded as censored at individual *i's* survey return date $R_i$. The parameter $\beta_d$ for $\lambda_{id}(t) = e^{\beta_d X_i(t)}$ can be estimated by maximizing the above likelihood function.

The expected number of type $j$ trips taken by $i$ during a season from day 0 to day $S$ is readily calculated as

$$E[N_{id}(S)] = \sum_{s=0}^{S} e^{\beta_d X_i(s)}$$

The expected number of total trips taken by $i$ is then $\sum_d E[N_{id}(S)]$.

Note that if individuals have homogeneous (i.e.. not time- varying) hazard $\lambda_d$. the log likelihood function in the discrete time context reduces to

$$LL = \sum_i \left[ \log \lambda_{D_i} - \sum_{s=L_i}^{R_i} \sum_d \lambda_d \right].$$

Let $t_i (= R_i - L_i)$ denote the observed age of $i$ and $N_d$ be the number of individuals in the sample whose last trips are type *d*. The MLE of $\lambda_d$ is then

$$\hat{\lambda}_d = \frac{N_d}{\sum_i t_i}.$$

Censored Age Durations

One further complication we address with this model is to accommodate the *censored age duration varible.*

The variable $L_i$ denotes the date individual *i* took the last trip, as reported in the questionnaire returned on date $R_i$. For some individuals *k,* however, $L_k$ is not available, and all we know is that no trip was ever taken from the beginning of sample period *C* (= April 1, 1983) up to the questionnaire return date $R_k$. This gives us the *left censored* age data. Recognizing that age duration is essentially right-censored trip duration! the data for these individuals can alternatively be itnerpreted as fight- and left-censored trip durations.

The fact that the age duration has 'survived' the period from *C* to $R_k$ suggests that we augment the likelihood function (IV.21) with

$$L_0 = \prod_k S_k(t_k{}_C^{R_k}) = \prod_k \exp\left(-\int_C^{R_k} h(s)\,ds\right)$$

to include the non-participants for whom we only have censored age. Therefore the complete log likelihood function is

$$LL = \sum_{i\in P_1}\left[\log\lambda_{D_i}(L_i) - \int_{L_i}^{R_i}\sum_d \lambda_d(s)\,ds\right] - \sum_{k\in P_0}\left[\int_C^{R_k}\sum_d \lambda_d(s)\,ds\right]$$

where $P_1$ is the sample of participating people, and $P_0$ is the set of non-participants.

## Using Less Restrictive Distributions

Estimation can also be performed using other more flexible functional forms (e.g., Weibull, Log-Logistic. or Box-Cox hazards) for the distribution of between-trip durations if the Poisson-Exponential dual appears too restrictive. Note that the exponential distribution has only one parameter $\lambda$, and its mean is equal to its standard deviation $\mathrm{E}(T) = \sqrt{\mathrm{Var}(T)} = 1/\lambda$. Therefore the mean and variance cannot be adjusted separately. As pointed out by Kiefer (1988), the exponential is unlikely to be an adequate description of the data if the sample contains both very long and short durations.

Let $f(t)$ and $F(t)$ be the common PDF and CDF of the independently distributed between-trip intervals $T$ for all individuals. The mean interval between successive trips is then

$$\mu = E[T] = \int_0^\infty f(t)\,t\,dt.$$

The estimation procedure for any $f(t)$ is basically the same as that described in the previous section. To my knowledge, however, no computer package can yet handle the full time-varying competing risk age duration model, though partial estimation can indeed be carried out by some existing commercial programs.[10] In the following sections: the more general Weibull distribution will be employed to illustrate the use of other distribution functions and to test the exponential duration assumption.

<u>The Weibull Distribution</u>

Now assume that the between-trip time intervals are all independent and Weibull-distributed with two parameters: a shape parameter $\gamma(>0)$ and a scale parameter $\lambda(>0)$. The distribution functions are[11]

$$\begin{aligned} PDF: \quad f(t) &= \lambda\gamma(\lambda t)^{\gamma-1}\exp(-(\lambda t)^{\gamma}) \\ CDF: \quad F(t) &= 1-\exp(-(\lambda t)^{\gamma}) \\ Survival: \quad S(t) &= \exp(-(\lambda t)^{\gamma}) \\ Hazard: \quad h(t) &= \lambda\gamma(\lambda t)^{\gamma-1} \end{aligned}$$

The shape parameter determines the shape of the hazard function $h(t)$. When

---

[10] For instance, Limdep (1989, chapters 27 and 28) can only handle Cox's proportional hazards model without competing risks, or basic Weibull with neither time-varying covariates nor competing risks.

[11] For a brief discussion, on the Weibull distribution., see Lee (1980), pp. 162-67.

$\gamma > 1$, the hazard rate $h(t)$ increases with $t$: the case of *positive duration dependence.* When $\gamma < 1$, the hazard rate $h(t)$ declines with $t$, the case of *negative duration dependence.* When $\gamma = 1$, the model reduces to the exponential case and we have a constant hazard regardless of the value of $t$. Therefore, the appropriateness of employing the exponential distribution can be empirically tested by formulating a test of the hypothesis $H_0 : \gamma = 1$ [12]

The expected length of between-trip intervals is

$$\mu = \frac{\Gamma(1 + \frac{1}{\gamma})}{\lambda}$$

and the variance is

$$\sigma^2 = \frac{1}{\lambda^2}\left[\Gamma(1 + \frac{2}{\gamma}) - \Gamma^2(1 + \frac{1}{\gamma})\right]$$

where $\Gamma$ is the gamma function defined as[13]

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du.$$

Note that when $\gamma = 1$, we have $\mu = \Gamma(2)/\lambda = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$, exactly the exponential case.

The three tasks we need to perform to modify the basic Weibull distribution for our estimation problem are

- the derivation of the age distribution,
- the inclusion of the time-varying covariates, and
- the development of the competing risk model.

---

[12] The Weibull hazard is monotonic. Other generalizations that embed Weibull as a special case are Log-Logistic and Box-Cox hazards, for example. Both hazards allow non-monotonic behavior. See Lancaster (1990), chapter 3.

[13] $\Gamma(x)$ is simply $(x-1)!$ when $x$ is a non- negative integer.

The Time-Varying Weibull Age Distribution

To derive the time-varying version of the Weibull distribution, we assume that the scale of the hazard rate is time dependent, i.e., $\lambda(t) > 0$ for all *t*. However, the shape of the hazard function, determined by the value of $\gamma$, is preset and not time-varying. This maintains $\gamma$ as a constant. For the estimation of the Weibull model, we further posit that

$$\begin{aligned} \gamma &= e^{\alpha X_1} > 0 \\ \lambda(t) &= e^{\beta X_2(t)} > 0 \end{aligned}$$

where the explanatory variables $X_1$ are constant through time while $X_2(t)$ vary with time. Conceptually $X_1$ consists of variables that determine the shape of the hazard function, and $X_2(t)$ contains the variables that affect the trip-taking probabilities at *t*. There may possibly be overlapping between $X_1$ and $X_2$ since $X_2$ can also have components that do not vary with time. If $\alpha X_1 > 0$ (or equivalently, $\gamma > 1$), an individual is said to have *positive* duration dependence. If $\alpha X_1 = 0$ (or $\gamma = 1$), there is *no* duration dependence. Otherwise, *negative* duration dependence exists.

By modifying the basic Weibull distribution functions, we can derive the time-varying Weibull probability system as follows:

$$\begin{aligned} \textit{Hazard}: \quad h(t) &= \lambda(t)\gamma \left[\int_0^t \lambda(s)\,ds\right]^{\gamma-1} \\ \textit{Survival}: \quad S(t) &= \exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) \\ \textit{CDF}: \quad F(t) &= 1-\exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) = 1 - S(t) \\ \textit{PDF}: \quad f(t) &= \lambda(t)\gamma \left[\int_0^t \lambda(s)\,ds\right]^{\gamma-1} \exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) \end{aligned}$$

It is straightforward to verify that $F(0) = 0$ and $F(\infty) = 1$ and

$$f(t) = \frac{dF(t)}{dt} = h(t)\,S(t),$$

and hence the above equations constitute a consistent distribution definition.

Following the notation of $(t \mid_{L=R-t}^{R})$ used for age duration in previous sections, the age distribution can be derived:

$$
\begin{aligned}
S_A(t \mid_L^R) &= \exp\left[-\left(\int_L^R \lambda(s)\, ds\right)^{\gamma}\right] \\
F_A(t \mid_L^R) &= 1 - \exp\left[-\left(\int_L^R \lambda(s)\, ds\right)^{\gamma}\right] = 1 - S_A(t \mid_L^R) \\
f_A(t \mid_L^R) &= \frac{dF_A(t \mid_L^R)}{dt} = \lambda(L)\gamma \left[\int_L^R \lambda(s)\, ds\right]^{\gamma-1} S_A(t \mid_L^R)
\end{aligned}
$$

## The Competing Risk Weibull Model

For the three types of trips (day, weekend and vacation, indexed by $j$ = 1, 2, 3, respectively) that an individual may take, we assume that

$$\lambda_j(t) = e^{\beta_j X_2(t)} > 0$$

We further assume that the shape parameter $\gamma$ is constant and identical for all types of trips.[14]

The type-specific hazard rate, under the assumption of inter-type dependence is

$$h_j(t) = \lambda_j(t)\gamma \left[\int_0^t \sum_l \lambda_l(s)\, ds\right]^{\gamma-1} \tag{IV.23}$$

The non-type-specific hazard rate, the sum of the type- specific hazards by definition, is then simply

$$h(t) = \sum_j h_j(t) = \left[\sum_j \lambda_j(t)\right] \gamma \left[\int_0^t \sum_j \lambda_j(s)\, ds\right]^{\gamma-1} \tag{IV.24}$$

These hazard functions imply that

[14] There are two reasons for the different treatments. Firstly, we see no reason why different types of trips should have different duration dependencies. Secondly, and more importantly, we need to keep the model under a controllable degree of complexity.

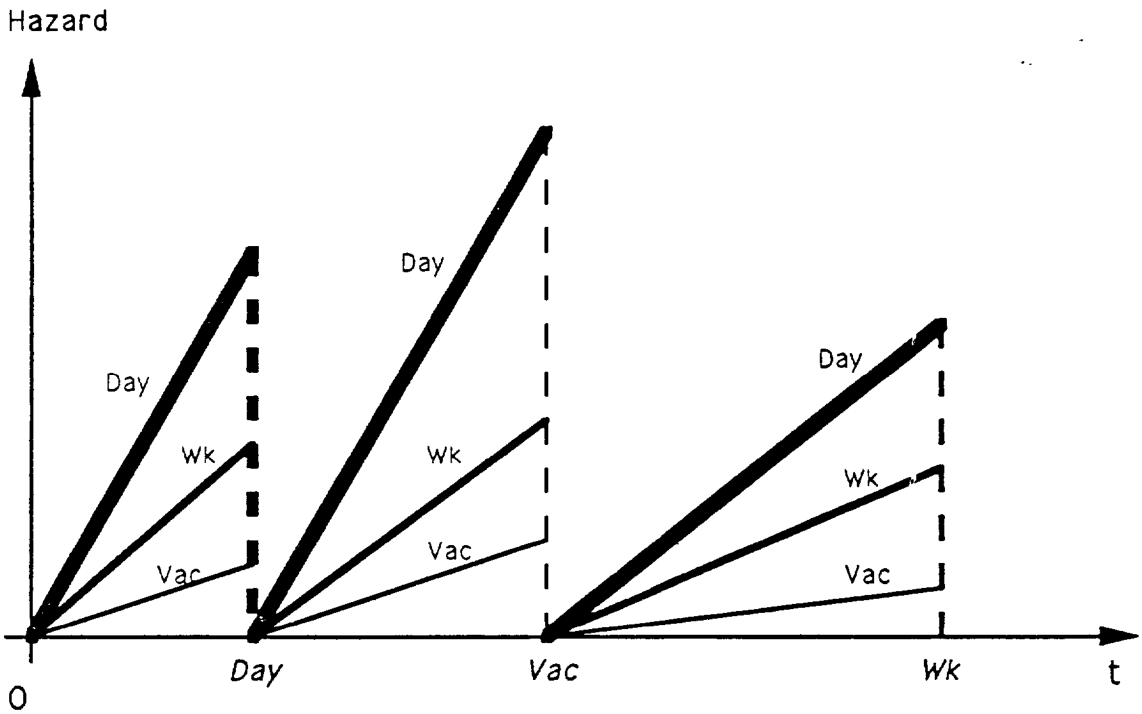


Figure IV.4: Hazard rate with inter-type dependence

- The probability density of taking a type $J$ trip at $t$, conditional on no trip up to $t$, depends not only on the history of parameter $\lambda_J(t)$ before time $t$, but also on the history of parameters of all other types.

- The hazard ratio of having different types of trips at time $t$ is not affected by the parameter values in the past. This can be seen by noting that proportinoaltiy holds as follows:

$$h_1(t) : h_2(t) : h_3(t) = \lambda_1(t) : \lambda_2(t) : \lambda_3(t).$$

Figure (IV.4) shows the hazard rate behaviors of different trip types for the case of positive duration dependence. Note that the hazards of all types become zero with each trip occurrence (i.e., $t$ = 0) whatever its type.[15]

[15] In the case of negative duration dependence, all the hazards become infinity the moment after a trip is taken.

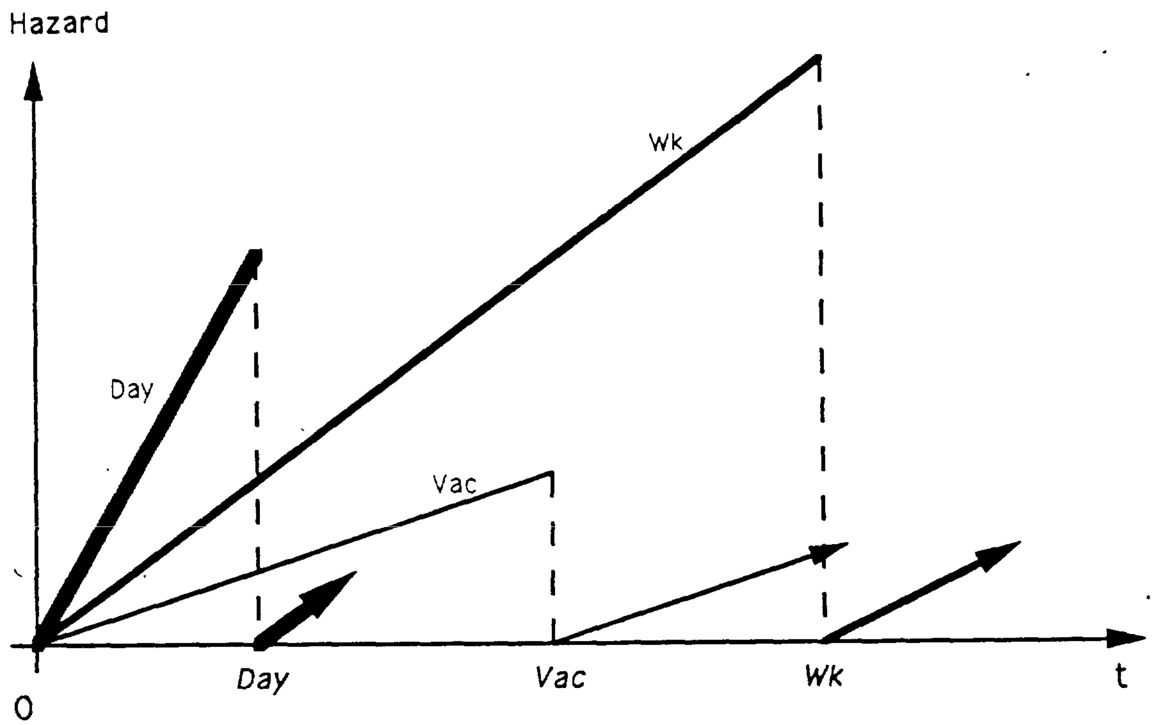


Figure IV.5: Hazard rate without inter-type dependence

The hazard rate based on inter-type dependence discussed above can be contrasted with a hazard rate without inter- type dependence illustrated in figure (IV.5)

$$h_j(t_j) = \lambda_j(t_j)\gamma \left[\int_0^{t_j} \lambda_j(s)\, ds\right]^{\gamma-1}$$

and

$$h(t) = \sum_j h_j(t_j) = \sum_j \left(\lambda_j(t_j)\gamma \left[\int_0^{t_j} \lambda_j(s)\, ds\right]^{\gamma-1}\right).$$

Note that the between-trip duration $t$ is indexed by the trip type $j$ since it is now type-specific. When there is no inter-type dependence among different hazards, the hazard rate of one trip type is not affected by the occurrences of trips of other types. Therefore, the hazard rate of one trip type continues to increase until a trip of its own type is taken, at that time it drops to zero while hazards of other trip types keep increasing.

In our analysis we assume that there is inter-type dependence and employ the hazard rates defined in (IV.23). It is not difficult to verify that the distribution functions corresponding to the hazards (IV.23) and (IV.24) are

$$\begin{aligned} S(t) &= \exp\left(-\left[\int_0^t \sum_j \lambda_j(s)\,ds\right]^\gamma\right) \\ F(t) &= 1-\exp\left(-\left[\int_0^t \sum_j \lambda_j(s)\,ds\right]^\gamma\right) = 1-S(t) \\ f(t) &= \frac{dF(t)}{dt} = h(t)\,S(t) \\ f_j(t) &= h_j(t)\,S(t) \end{aligned}$$

## Estimating the Weibull Model

Let $f^A_{J_i}(t_i|^{R_i}_{L_i})$ denote the probability density of individual $i$ taking the most recent trip of type $J_i$ and having an observed age from $L_i$ to $R_i$. The likelihood function for the sample $P_1$ for whom we have the last trip data is then

$$\begin{aligned} L &= \prod_{i\in P_1} f^A_{J_i}(t_i|^{R_i}_{L_i}) \\ &= \prod_{i\in P_1}\left\{\lambda_{J_i}(L_i)\,\gamma\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]^{\gamma-1}\exp\left(-\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]^\gamma\right)\right\} \end{aligned}$$

For the non-participant sample $P_0$, we know only that no trip was taken from $C$, the beginning of sample period. up to the questionnaire return date $R_k$. The likelihood for this sample is

$$\begin{aligned} L &= \prod_{k\in P_0} S(t_k|^{R_k}_{C}) \\ &= \prod_{k\in P_0} \exp\left(-\left[\int_C^{R_k}\sum_k \lambda_k(s)\,ds\right]^\gamma\right) \end{aligned}$$

Combining the participants and the non-participants, the complete log likelihood function is

$$LL = \sum_{i\in P_1}\left\{\log\lambda_{J_i}(L_i) - \log\gamma + (\gamma-1)\,\log\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]\right.$$

$$
\begin{aligned}
& - \left[ \int_{L_i}^{R_i} \sum_j \lambda_j(s)\, ds \right]^{\gamma} \Bigg\} \\
& - \sum_{k \in P_0} \left[ \int_{C}^{R_k} \sum_k \lambda_k(s)\, ds \right]^{\gamma} \\
= \quad & \sum_{i \in P_1} \Bigg\{ \beta_{J_i} X_2(L_i) + \alpha X_1 + (e^{\alpha X_1} - 1) \ \log \left[ \int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \right] \\
& - \left[ \int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \right]^{e^{\alpha X_1}} \Bigg\} \\
& - \sum_{k \in P_0} \left[ \int_{C}^{R_k} \sum_k e^{\beta_j X_2(s)}\, ds \right]^{e^{\alpha X_1}} \\
= \quad & \sum_{i \in P_1} \left[ \beta_{J_i} X_2(L_i) - \alpha X_1 + (e^{\alpha X_1} - 1) \ \log S(\beta) - S(\beta)^{e^{\alpha X_1}} \right] \\
& - \sum_{k \in P_0} N(\beta)^{e^{\alpha X_1}}
\end{aligned}
$$

where

$$
\begin{aligned}
S(\beta) \ & \equiv \ \int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \quad \text{for a participant } i \text{ in } P_1) \\
N(\beta) \ & \equiv \ \int_{C}^{R_k} \sum_j e^{\beta_j X_2(s)}\, ds \quad \text{(for a non-participant } k \text{ in } P_0).
\end{aligned}
$$

The first derivatives with respect to the parameters $\alpha$ and $\beta$ are consequently

$$
\begin{aligned}
\frac{\partial LL}{\partial \alpha} \ = \ & \sum_{i \in P_1} \left[ X_1 - (X_1 e^{\alpha X_1}) \log S(\beta) - S(\beta)^{e^{\alpha X_1}} \ \log S(\beta)\, e^{\alpha X_1} X_1 \right] \\
& - \sum_{k \in P_0} \left\{ N(\beta)^{e^{\alpha X_1}} \ \log N(\beta)\, e^{\alpha X_1} X_1 \right\} \\
= \ & \sum_{i \in P_1} \left\{ X_1 \left[ 1 - e^{\alpha X_1} \ \log S(\beta) \, (1 - S(\beta)^{e^{\alpha X_1}}) \right] \right\} \\
& - \sum_{k \in P_0} \left\{ N(\beta)^{e^{\alpha X_1}} \ \log N(\beta)\, e^{\alpha X_1} X_1 \right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial LL}{\partial \beta_j} \ = \ & \sum_{i \in P_1} \left[ Q(j, J_i) + (e^{\alpha X_1} - 1) \frac{S_j(\beta_j)}{S(\beta)} - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1} - 1} S_j(\beta_j) \right] \\
& - \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1} - 1} N_j(\beta_j) \right\} \\
= \ & \sum_{i \in P_1} \left\{ Q(j, J_i) + \frac{S_j(\beta_j)}{S(\beta)} \left[ (e^{\alpha X_1} - 1) - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}} \right] \right\}
\end{aligned}
$$

$$- \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1}-1} N_j(\beta_j) \right\}$$

where

$$\begin{aligned} S_j(\beta_j) &\equiv \frac{\partial S(\beta)}{\partial \beta_j} = \int_{L_i}^{R_i} X_2(s)\, e^{\beta_j X_2(s)}\, ds \quad \text{for a participant } i \text{ in } P_1) \\ N_j(\beta_j) &\equiv \frac{\partial N(\beta)}{\partial \beta_j} = \int_{C}^{R_k} X_2(s)\, e^{\beta_j X_2(s)}\, ds \quad \text{(for a non-participant } k \text{ in } P_0) \end{aligned}$$

and

$$Q(j, J_i) = \begin{cases} X_2(L_i) & \text{if } j = J_i \\ 0 & \text{otherwise.} \end{cases}$$

And the second derivatives (the Hessian matrix) are

$$\begin{aligned} \frac{\partial^2 LL}{\partial \alpha^2} &= \sum_{i \in P_1} \Big\{ X_1^2 e^{\alpha X_1} \log S(\beta) \\ &\quad - S(\beta)^{e^{\alpha X_1}} \left[ \log S(\beta)\, e^{\alpha X_1} X_1 \right]^2 \\ &\quad - S(\beta)^{e^{\alpha X_1}} \log S(\beta)\, e^{\alpha X_1} X_1^2 \Big\} \\ &\quad - \sum_{k \in P_0} \Big\{ N(\beta)^{e^{\alpha X_1}} \left[ \log N(\beta)\, e^{\alpha X_1} X_1 \right]^2 \\ &\quad + N(\beta)^{e^{\alpha X_1}} \log N(\beta)\, e^{\alpha X_1} X_1^2 \Big\} \\ &= \sum_{i \in P_1} \left\{ X_1^2 e^{\alpha X_1} \log S(\beta) \left[ 1 - S(\beta)^{e^{\alpha X_1}} \left( \log S(\beta)\, e^{\alpha X_1} + 1 \right) \right] \right\} \\ &\quad - \sum_{k \in P_0} \left\{ X_1^2 e^{\alpha X_1} \log N(\beta)\, N(\beta)^{e^{\alpha X_1}} \left[ \log N(\beta)\, e^{\alpha X_1} - 1 \right] \right\} \\ \frac{\partial^2 LL}{\partial \alpha\, \partial \beta_j} &= \sum_{i \in P_1} \Big[ (X_1 e^{\alpha X_1}) \frac{S_j(\beta_j)}{S(\beta)} \\ &\quad - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}-1} S_j(\beta_j) \log S(\beta)\, e^{\alpha X_1} X_1 \\ &\quad - S(\beta)^{e^{\alpha X_1}} \frac{S_j(\beta_j)}{S(\beta)} e^{\alpha X_1} X_1 \Big] \\ &\quad - \sum_{k \in P_0} \Big[ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1}-1} N_j(\beta_j) \log N(\beta)\, e^{\alpha X_1} X_1 \\ &\quad - N(\beta)^{e^{\alpha X_1}} \frac{N_j(\beta_j)}{N(\beta)} e^{\alpha X_1} X_1 \Big] \\ &= \sum_{i \in P_1} \left\{ X_1 e^{\alpha X_1} \frac{S_j(\beta_j)}{S(\beta)} \left[ 1 - S(\beta)^{e^{\alpha X_1}} \left( \log S(\beta)\, e^{\alpha X_1} + 1 \right) \right] \right\} \end{aligned}$$

$$
- \sum_{k \in P_0} \left\{ X_1 \, e^{\alpha X_1} \, \frac{N_j(\beta_j)}{N(\beta)} \, N(\beta)^{e^{\alpha X_1}} \left( \log N(\beta) \, e^{\alpha X_1} + 1 \right) \right\}
$$

$$
\begin{aligned}
\frac{\partial^2 LL}{\partial \beta_j^2} &= \sum_{i \in P_1} \left\{ \frac{S_{jj}(\beta_j)\, S(\beta) - S_j(\beta_j)^2}{S(\beta)^2} \left[ \left( e^{\alpha X_1} - 1 \right) - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}} \right] \right. \\
&\quad \left. - \left( e^{\alpha X_1} \right)^2 \frac{S_j(\beta_j)}{S(\beta)} S(\beta)^{e^{\alpha X_1} - 1} S_j(\beta_j) \right\} \\
&\quad - \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1}} \left[ \left( e^{\alpha X_1} - 1 \right) \frac{N_j(\beta_j)^2}{N(\beta)^2} + \frac{N_{jj}(\beta_j)}{N(\beta)} \right] \right\} \\
&= \sum_{i \in P_1} \left\{ \left( e^{\alpha X_1} - 1 \right) \frac{S_{jj}(\beta_j)\, S(\beta) - S_j(\beta_j)^2}{S(\beta)^2} \right. \\
&\quad \left. - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}} \left[ \left( e^{\alpha X_1} - 1 \right) \frac{S_j(\beta_j)^2}{S(\beta)^2} + \frac{S_{jj}(\beta_j)}{S(\beta)} \right] \right\} \\
&\quad - \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1}} \left[ \left( e^{\alpha X_1} - 1 \right) \frac{N_j(\beta_j)^2}{N(\beta)^2} + \frac{N_{jj}(\beta_j)}{N(\beta)} \right] \right\}
\end{aligned}
$$

and, for $j \neq m$,

$$
\begin{aligned}
\frac{\partial^2 LL}{\partial \beta_j \partial \beta_m} &= \sum_{i \in P_1} \left\{ (e^{\alpha X_1} - 1) \frac{-S_j(\beta_j) S_m(\beta_m)}{S(\beta)^2} \right. \\
&\quad \left. - e^{\alpha X_1} (e^{\alpha X_1} - 1) \, S(\beta)^{e^{\alpha X_1} - 2} S_j(\beta_j) \, S_m(\beta_m) \right\} \\
&\quad - \sum_{k \in P_0} \left[ e^{\alpha X_1} \, (e^{\alpha X_1} - 1) \, N(\beta)^{e^{\alpha X_1} - 2} N_j(\beta_j) \, N_m(\beta_m) \right] \\
&= \sum_{i \in P_1} \left\{ -(e^{\alpha X_1} - 1) \frac{S_j(\beta_j) S_m(\beta_m)}{S(\beta)^2} \left[ 1 + e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}} \right] \right\} \\
&\quad - \sum_{k \in P_0} \left[ e^{\alpha X_1} \, (e^{\alpha X_1} - 1) \, N(\beta)^{e^{\alpha X_1} - 2} N_j(\beta_j) \, N_m(\beta_m) \right]
\end{aligned}
$$

where

$$
\begin{aligned}
S_{jj}(\beta_j) &\equiv \frac{\partial^2 S(\beta)}{\partial \beta_j^2} = \int_{L_i}^{R_i} X_2(s)^2 e^{\beta_j X_2(s)} \, ds \quad \text{(for a participant } i \text{ in } P_1) \\
N_{jj}(\beta_j) &\equiv \frac{\partial^2 N(\beta)}{\partial \beta_j^2} = \int_{C}^{R_k} X_2(s)^2 e^{\beta_j X_2(s)} \, ds \quad \text{(for a non-participant } k \text{ in } P_0).
\end{aligned}
$$

The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by the Newton-Raphson, algorithm. They are consistent: asymptotically efficient and asymptotically normal. The variance- covariance matrix $(-E[\nabla^2 LL]^{-1})$ can be approximated by

$$
-[\nabla^2 LL]^{-1}_{\alpha = \hat{\alpha},\, \beta = \hat{\beta}}
$$

If the null hypothesis $H_0 : \alpha = 0$ is rejected by the likelihood ratio test, the use of the exponential distribution cannot be justified, and we have to calculate the expected total seasonal trips for the Weibull model.[16]

## Renewal Counting Process for Weibull

If the use of the exponential is rejected in favor of the Weibull, the distribution of the corresponding renewal counting process $N(s), s \geq 0$, has to be determined to calculate the expected number of trips in season.

Define the random variable $W_n$ as the sum of $n$ consecutive between-trip durations. i.e.,

$$W_n = T_1 + T_2 + \cdots + T_n, \; n \geq 1.$$

The distribution of $W_n$ is then the *convolution* of the CDFs of the $n$ durations $T_1, T_2, \ldots, T_n$.

**Definition IV.1** If two independent random variables $X$ and $Y$ have the distribution functions $F_X$ and $F_Y$, respectively. then the distribution of their sum $Z = X + Y$ is the convolution of $F_X$ and $F_Y$, defined as

$$F_Z(z) = \int_{-\infty}^{\infty} F_X(z-u)\,dF_Y(u) = \int_{-\infty}^{\infty} F_Y(z-u)\,dF_X(u). \quad \blacksquare$$

Since all between-trip durations $T_k$ have the. common CDF $F(s) = \text{Prob}\{T \leq s\}$, the CDF of $W_n$ is the n-fold convolution of $F(s)$ with itself, denoted by $F_n$. Hence.

consecutive

$$F_n(s) = \text{Prob}\{W_n \leq s\} = \text{Prob}\{N(s) \geq n\}, \; n \geq 1$$

Given $F_1(s) \equiv F(s)$, any $F_n(s)$ for $n \geq 2$ can be calculated using the recursive formula

$$F_n(s) \;\; = \;\; \int_0^{\infty} F_{n-1}(s-u)\,dF(u)$$

---

[16] An interesting intermediate case is where the intercept is nonzero, but the slope coefficients are zero.

$$
\begin{aligned}
&= \int_0^\infty F_{n-1}(s-u)\, f(u)\, du \\
&= \int_0^s F_{n-1}(s-u)\, f(u)\, du
\end{aligned}
$$

The probability of having exactly $n$ trips in a time period $S$ is

$$
\begin{aligned}
\text{Prob}\{N(S)=n\} &= \text{Prob}\{W_n \le S\} - \text{Prob}\{W_{n+1} \le S\} \\
&= F_n(S) - F_{n+1}(S).
\end{aligned}
$$

The expected number of trips in the time interval $S$ is then

$$
\begin{aligned}
m(S) &= E[N(S)] \\
&= \sum_{n=1}^{\infty} [n \cdot \text{Prob}\{N(S)=n\}] \\
&= \sum_{n=1}^{\infty} \text{Prob}\{N(S) \ge n\} \\
&= \sum_{n=1}^{\infty} F_n(S).
\end{aligned}
$$

While a closed-form formula for $F_n$ is difficult to obtain. the values of $F_n(S)$ can be approximated by numerical solutions and the calculation of $m(S)$ can be carried out as follows for the discrete time case. First the values of $F_1(s) = F(s)$ for $s$ = *1,2. . .*, $S$ are calculated and stored. The values of $F_2(s)$ for $s$ = 1,2, . . . , $S$ can then be computed using the values of $F_1$ now available.

$$
F_2(s) = \sum_{u=0}^{s} [F_1(s-u)\, f(u)]
$$

This goes on until, for some $N$, $F_N(S)$ is small enough to be safely ignored. Note that $f(0)$ = 0 and $F_n(0)$ = 0. $\forall n$. The expected number of seasonal trips is thus approximately

$$
m(S) = \sum_{n=1}^{N} F_n(S) = \sum_{n=1}^{N} \sum_{u=0}^{S} [F_{n-1}(S-u)\, f(u)].
$$

The function $m(S)$, called the *renewal function,* will always converge since it is finite for a finite $S$, as proved in Ross (1983, p. 57). Note that the simple formula

$m(S) = S/\mu$ does not generally hold for other distributions besides the Poisson-Exponential case, where it occurs due to the special memoryless property of the exponential distribution. Though it is true that both respectively.

$$\frac{N(S)}{S} \rightarrow \frac{1}{\mu} \text{ as } S \rightarrow \infty$$

and

$$\frac{m(S)}{S} \rightarrow \frac{1}{\mu} \text{ as } S \rightarrow \infty,$$

it is a mistake to use $(S/\mu)$ as the expected value of $N(S)$ when $S$ is not large enough. In our study, $S$ is the length of a fishing season, the time period during which we count the trips. and is (sibstantially) less than infinity. Hence the extra work of calculating $m(S)$ has to be done if a less restrictive distribution like the Weibull is preferred.