Applications of Statistics and Probability in Civil Engineering – Kanda, Takada & Furuta (eds)

The extremal index of a dependent stationary pulse load process

B. Bhattacharya
Dept. of Civil Engineering, Indian Institute of Technology, Kharagpur, WB, India

ABSTRACT: Observing a load process above high thresholds, modeling it as a pulse process with random occurrence times and magnitudes, and extrapolating life-time maximum or design loads from the data is a common task in structural reliability analyses. In this paper, we consider a stationary live load sequence that arrive according to a dependent point process and allow for a weakened mixing-type dependence in the load pulse magnitudes that asymptotically decreases to zero with increasing separation in the sequence. The scale of fluctuation of the loading process is used to identify clusters of exceedances above high thresholds which in turn is used to estimate the extremal index of the process. The pulse arrival instants are modeled as a Cox process governed by a stationary lognormal intensity. An illustrative example utilizes in-service peak strain data from ambient traffic collected on a high volume highway bridge, and analyzes the asymptotic behavior of the maximum load.

1 INTRODUCTION

A general formulation of the time-dependent reliability problem corresponding to an "overload" type limit state is:

$$\mathrm{Rel}(t) = P\begin{bmatrix} R(\tau) - D - L(\tau) - W(\tau) - S(\tau) - \ldots > 0 \\ \text{for all } \tau \in [0,t] \end{bmatrix} \quad (1)$$

R is the resistance of the structural component, possibly time dependent due to aging effects. L, W, S etc. represent time-varying live, wind, snow loads and so on, respectively. In almost every relevant load case, the life-time maximum of one of these load processes must be evaluated (ASCE, 2002). The dead load (D) is generally assumed not to vary with time. Simplification is possible if $R(\tau)$ is replaced by a representative resistance, R_e, that is independent of time. A typical example can be formulated as:

$$R_e - D - L_{\max,t} = 0 \quad (2)$$

where $L_{\max,t}$ is the maximum live load on the structure during (0, t]. Live loads often have discontinuous sample functions with jumps occurring at random instants of time, τ_i, with random magnitudes, L_i. If the duration of the individual live load is small compared to the reference time (or if a filtering with a high threshold is employed in data collection), the live load process may be idealized as a pulse process with random occurrence times. The maximum of the observed live load magnitudes is then simply,

$$L_{\max,t} = \max\left\{L_1, L_2, \ldots, L_{N_t}\right\} \quad (3)$$

N_t is the random number of loading events during the interval [0, t].

The simplest model for predicting the life-time maximum distribution from an observed sequence of size n (n finite) is to assume that the magnitudes, L_i, are mutually independent and are identically distributed (the so-called "i.i.d." assumption) and independent of the occurrence times as well. In many practical instances, each member of the observed sequence $L_1, L_2, \ldots, L_n$ actually represents the "block maximum" to facilitate computations; the cost, however, is the wastage of a large number of potentially useful data points.

Powerful generalizations can be made when the sample size, n, of the i.i.d. sequence approaches infinity. Under very general conditions that are satisfied by most parent distributions, the maximum, $\hat{L}_{\max,t}$, of the i.i.d. sequence approaches one of the three classical extreme value distributions, H_c (elaborated later). Also, even if the occurrence times did not originally constitute a renewal process, they do become rarer on the time axis and approach the Poisson distribution as the said threshold becomes higher.

Nevertheless, the i.i.d. assumption cannot be always justified in maximum live load modeling. There may be significant dependence in the loading sequence (both in regard to load magnitudes and occurrence times) that would make the results from classical i.i.d. analyses conservative at best and erroneous in general. This paper presents a methodology that accounts for more generalized loading sequences including dependence in the occurrence point process and its associated marks. The formulation will be restricted to stationary and aperiodic sequences while the methodology will be demonstrated on a strain response data collected under ambient traffic on a highway bridge.

2 THE MAXIMUM LIVE LOAD FROM A STATIONARY DEPENDENT SEQUENCE

Let F denote the common marginal distribution of the strictly stationary and dependent sequence, $\{L_n\}$, and let M_n denote the maximum of the sequence. The dependence structure of $\{L_n\}$ is such that Leadbetter's (Leadbetter et al., 1983) Conditions $D(u_n)$ and $D'(u_n)$ are satisfied.

Condition $D(u_n)$ is a type of distributional mixing (a much weakened form of strong mixing) and ensures that subsequences of $\{L_n\}$ becomes asymptotically independent with increasing separation between them. Let $F_{i_1}, \ldots, L_{i_n}(x)$ denote the joint distribution function of $L_{i_1}, \ldots, L_{i_n}$ evaluated at the common point x. Condition $D(u_n)$ is said to hold for some given real sequence $\{u_n\}$, if for any integers $1 \le i_1 < \cdots < i_p < j_1 < \cdots < j_{p'} \le n$ for which $j_1 - i_p \ge l$, we have

$$\left| F_{i_1,\cdots,i_p,j_1,\cdots,j_{p'}}(u_n) - F_{i_1,\cdots,i_p}(u_n) F_{j_1,\cdots,j_{p'}}(u_n) \right| \le \alpha_{n,l} \tag{4}$$

and $\alpha_{n,l_n} \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

The second, Condition $D'(u_n)$, limits the possibility of clustering of exceedances from the sequence above a high threshold and is said to hold for the strictly stationary sequence $\{L_n\}$ if for some given real sequence $\{u_n\}$,

$$\limsup_{n\to 0} n \sum_{j=2}^{[n/k]} P\left\{L_1 > u_n, L_j > u_n\right\} = 0 \text{ as } k \to \infty \tag{5}$$

where [ ] denotes the integer part.

The importance of Conditions $D(u_n)$ and $D'(u_n)$ comes from the following two very appealing properties (Leadbetter et al., 1983):

1. The asymptotic distribution of M_n has only three possible forms, namely, the three classical extreme value distribution types. The rate of convergence, however, is slower than in the i.i.d. case, and can be quantified using the *extremal index*, θ, of the sequence, defined below.
2. The point process constituting the instants when the sequence $\{L_n\}$ exceeds the threshold u_n converges in distribution to a Poisson process as n increases. Consequent of the asymptotic Poisson behavior, the maxima in disjoint time intervals become asymptotically independent as well.

2.1 *The extremal index*

The extremal index of a sequence can be interpreted as the reciprocal of the mean limiting cluster size above high thresholds, and is given by (Leadbetter et al., 1983):

$$\lim_{n\to\infty} P\left\{M_n \le u_n(\tau)\right\} = \exp(-\theta\tau)\ ,\ \tau > 0,\ 0 \le \theta \le 1 \tag{6}$$

if for some $\tau > 0$ there exists sequence $\{u_n(\tau)\}$ such that

$$\tau = \lim_{n\to\infty} n\left(1 - F(u_n(\tau))\right) \tag{7}$$

The extremal index, which is a number between 0 and 1, measures the strength of the dependence in the sequence $\{L_n\}$. Heuristically, $\theta = 0$ corresponds to an infinitely long memory sequence, $0 < \theta < 1$ corresponds to a short memory sequence, and $\theta = 1$ corresponds to a memoryless sequence (Hsing, 1993).

It is convenient at this point to formally introduce the "associated" i.i.d. sequence $\{\hat{L}_n\}$ (we have alluded to it earlier) that has the same marginal (or "parent") distribution, F. Let $\hat{M}_n$ be the maximum of the i.i.d. sequence $\{\hat{L}_n\}$. Under suitable regularity conditions that are satisfied by most common parent distributions, F, the distribution of the maximum $\hat{M}_n$, when suitably normalized by constants $\{a_n\}$ and $\{b_n\}$, converges to one of the three classical types, H_c (Galambos, 1987; Castillo, 1998):

$$P\left[\hat{M}_n \le a_n x + b_n\right] \to H_c(z) = \exp\left[-(1 + cz)^{-1/c}\right] \tag{8}$$

where $1 + cz > 0$, $z = (x - \varepsilon)/\delta$ in which ε and $\delta > 0$ are appropriate location and scale parameters of the distribution. Eq. (8) represents the generalized extreme value distribution, in which the parameter c determines the type of the distribution: It is of (i) Type I (the Gumbel type) if $c = 0$, (ii) Type II (the Frechet type) if $c > 0$, and (iii) Type III (the Weibull type) if $c < 0$.

If Eq. (8) holds, the distribution of M_n also converges, with the same set of normalizing constants $\{a_n\}$ and $\{b_n\}$ (or with one set altered) as above, to the type of $(H_c)^\theta$, where the exponent $\theta > 0$ is the extremal index of $\{L_n\}$:

$$P\left[M_n \le a_n x + b_n\right] \to [H_c(z)]^\theta \tag{9}$$

The converse is also true. Clearly, H_c and $(H_c)^\theta$ are of the same type for any given value of c. (Two distribution G and H are of the same *type* if there are constants a and $b > 0$ such that $G(x) = H(a + bx)$). The significance of this result is that the distribution of the maximum M_n of a stationary dependent sequence, provided it converges (which can be guaranteed by Conditions $D(u_n)$ and $D'(u_n)$), may be estimated, at least in the right tail, simply with the help of the marginal distribution F and the extremal index θ of the underlying process, as:

$$P\left[M_n \le u_n\right] \approx F^{n\theta}(u_n) \tag{10}$$

for sufficiently high u_n and large but finite n.

Eq. (10) is significant also as it highlights the degree of conservatism that may be introduced by the common and sometimes indiscriminate engineering practice of assuming a sequence to be i.i.d. when estimating the distribution of its maximum .(Bhattacharya, 2007).

It is therefore clear that, the distribution of the maximum of a stationary dependent sequence may conveniently be estimated in two steps: first by obtaining the distribution of the maximum of the associated i.i.d. sequence, and second, by estimating the extremal index of the parent process. We take up the latter problem first; the estimation of the associated i.i.d. sequence will be taken up subsequently.

2.2 *Estimating* θ

There are two acceptable methods for estimating θ : the *blocks* method and the *runs* method. Under suitable conditions, both estimators are consistent for the extremal index, but we choose the runs method because it usually has the smaller bias of the two (Smith & Weissman, 1994; Weissman & Novak, 1998). Considering the extremal index as threshold dependent and estimating it at various values of the threshold has also been suggested (Tawn, 1990; Smith & Weissman, 1994), but we do not adopt that approach in this paper.

Defining $M_{p,q} = \max\{L_p, \ldots, L_q\}$, the runs estimator of the extremal index, $\hat{\theta}_R$, is given by:

$$\hat{\theta}_R(x; r, n) = \frac{\sum_{i=1}^{n} \mathbb{I}_{B,i}(L_i > x \ge M_{i+1,i+r-1})}{\sum_{i=1}^{n} \mathbb{I}_{A,i}(L_i > x)}, \quad r \ge 2 \tag{11}$$

in which $\mathbb{I}_{A,i}$ $(\cdot)$ and $\mathbb{I}_{B,i}$ $(\cdot)$ are indicator functions verifying the truth of the respective condition in parentheses. The estimate is basically the reciprocal of the average cluster size above high thresholds (x) in which two consecutive exceedances are part of the same cluster if they are less than r observations apart (i.e., a *run* of observations below x of length r or greater are deemed to separate two adjacent clusters). Eq. (11) is based on the property of dependent stationary sequences that, for some appropriately chosen r (Hsing, 1993),

$$P\left[M_{2,r} \le x \middle| L_1 > x\right] = \theta + R\left(q(x)\right) \tag{12}$$

such that $R(q(x)) \to 0$ as $q(x) = 1 - F(x) \to 0$ (i.e., as $x \to x_0$ where $x_0 = \sup\{x : F(x)\,(x) < \}$).

The quality of the estimate in Eq. (11) depends on R which, however, is not known a priori. Nevertheless, it has been suggested(Hsing, 1993) that R has a simple log-linear form for a large class of processes:

$$R(q) \sim \beta_1 q^{\beta_2}, \quad \beta_1 \ne 0,\ \beta_2 > 0, \quad \text{as } q \to 0 \tag{13}$$

This relation, in conjunction with the set of estimates obtained using Eq. (11) for a range of x, may be used to obtain a minimum squared error estimate for θ in Eq. (12).

Note that $\hat{\theta}_R$ depends also on the run length, r, a parameter that must be chosen with care. Attempts have been made to provide optimal estimates of $r$based on minimum absolute bias considerations (Smith & Weissman, 1994). r must be short enough so as not to group relatively independent observations in the same cluster, at the same time long enough to reflect the dependence structure of the underlying physical nature of the sequence, e.g., the average storm length in case of wave data. We propose Vanmarcke's, {1983 #71} scale of fluctuation, τ_c, as an estimate of the run length, r:

$$\hat{r} \approx \tau_c = \lim_{T \to \infty} T\gamma(T) = 2\int_0^T \left(1 - \frac{\tau}{T}\right)\rho(\tau)\, d\tau \tag{14}$$

where $r \ge 2$, (r integer). $\gamma(T)$ is the variance function of a stationary process $X(t)$ and is defined as the ratio of the variance of the local average over a window of length T and the variance, σ^2, of the process $X(t)$.

2.3 *Modeling the random load occurrence times*

The distribution of the maximum load provides only the first part to the solution to the time-dependent problem. The statistical description of the number of occurrences during the given time interval is required to complete the picture.

The Poisson model i.e., a renewal process with exponential inter-arrival times or, equivalently, a process with independent increments and a constant rate of occurrence – is analytically attractive, but may prove too simplistic for most loading processes. Nevertheless, the pure Poisson process can be used as the building block for a large variety of processes showing clustering, dependence, non-stationarity etc. Clustering phenomena can be accounted for by the

Neymann-Scott and the Bartlett-Lewis processes (Cox & Isham, 2000). A Polya process, which is a non-stationary version of the pure birth process, can also be used to model clustering (Wen, 1990).

A more versatile generalization of the pure Poisson process occurs if the rate, $\Lambda(t) \geq 0$, itself is considered to be a random process yielding what is known as a doubly stochastic Poisson process (or Cox process) (Cox & Isham, 2000). The mean measure of the point process in the interval $[0,t]$ is a random variable and is given by:

$$M_t = \int_0^t \Lambda(\tau) d\tau \tag{15}$$

Then, conditioned on $M_t = m_t$ (where m_t is any positive number for given t), the point process $N(t)$ becomes a (generally non-homogeneous) Poisson process, i.e., the counts are distributed according to:

$$P\left[N(t) = x \mid M_t = m_t\right] = e^{-m_t} \frac{(m_t)^x}{x!} \tag{16}$$

We choose the Cox process to model the stochastic arrival rate for the load pulses first for its versatility, and then due to the asymptotic Poisson nature of the filtered point process above high thresholds that is consistent with the conditional Poisson characteristic of the Cox process. There is a rich collection of results pertaining to modeling the random rate function, $\Lambda(t)$; the two most common of which are the Markov modulated Poisson process (Fischer & Meier-Hellstern, 1992) and the exponentiated Gaussian process (i.e. the lognormal process) (Moller et al., 1998).

We can now look at the maximum, $\hat{L}_{\max,t}$, of the associated i.i.d. peak strain sequence $\hat{L}_1, \hat{L}_2, \ldots, \hat{L}_{N_t}$. for fixed $N_t = n$ and $M_t = m_t$, the distribution of $\hat{L}_{\max,t}$ is,

$$P\left[\hat{L}_{\max,t} \leq x \middle| N_t = n, M_t = m_t\right] = \left(F(x)\right)^n \tag{17}$$

Removing the conditioning on $N_t = n$ first, we obtain:

$$P\left[\hat{L}_{\max,t} \leq x \mid M_t = m_t\right] = \exp\left[-m_t \left\{1 - F(x)\right\}\right] \tag{18}$$

Further, if the distribution of the mean measure, M_t, is known, the unconditional distribution of $\hat{L}_{\max,t}$ can be given by:

$$P\left[\hat{L}_{\max,t} \leq x\right] = \int_0^\infty \exp\left[-m_t \left\{1 - F(x)\right\}\right] f_{M_t}(m_t) dm_t \tag{19}$$

2.4 *Inclusion of sampling-related uncertainties*

The maximum live load is estimated based on observed data. Hence, it is important that uncertainties due to sampling be accounted for. For any given l, the true value of F is unknown (see, e.g., (Galambos, 1993)), hence we can describe it as a random variable P with (prior) probability density function f_P'. The unknown P is estimated from the sample as:

$$\hat{p}(l) \equiv \hat{F}(l) = \frac{1}{n+1} \sum_{k=1}^{n} \mathbb{I}(L_k \leq l) \tag{20}$$

It can be shown (Bhattacharya, 2007) that the posterior density of P is $f_P''(p) = (c'/C) p^{(k+1)\theta} (1-p)^{\bar{k}\theta}(1)$, $0 \leq p \leq 1$, which is clearly of the Beta type, allowing us to write:

$$f_P''(x; \alpha_1, \alpha_2) = \begin{cases} \frac{1}{\mathrm{B}(\alpha_1, \alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}, 0 \leq x \leq 1 \\ 0 \qquad , \text{ elsewhere} \end{cases} \tag{21}$$

The two parameters of the distribution are:

$$\alpha_1 = ([(n+1)\hat{p}] + 1)\theta + 1, \ \alpha_2 = (n - [(n+1)\hat{p}])\theta + 1 \tag{22}$$

where n is the number of observations, and the estimate $\hat{p}$ is given by Eq. (20). The mean and variance of this distribution are, respectively, $\alpha_1/(\alpha_1 + \alpha_2)$ and $\alpha_1\alpha_2/(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)$. Hence, the updated mean of P is very close to the estimate $\hat{p}$ regardless of the value of θ ; its variance, however, is inversely proportional to the extremal index.

In light of the above formulation, Eq. (18) can now be interpreted as the conditional distribution of the maximum of the associated i.i.d. sequence during an interval of length t given fixed values of the parent distribution. Assuming the sampling uncertainty to be independent of the rate process, the unconditional distribution of the maximum of the associated i.i.d. sequence, $\hat{L}_{\max,t}$, is:

$$P\left[\hat{L}_{\max,t} \leq x\right] = \int_0^1 \int_0^\infty \exp\left[-m_t \left\{1 - P(x)\right\}\right] f_{M_t}(m_t) f_P''(p) \, dm_t \, dp \tag{23}$$

which may be estimated numerically using Monte-Carlo simulations. Finally, using the extremal index, the unconditional distribution of the maximum of the original sequence $\{L_n\}$ can be given by:

$$F_{L_{\max,t}}(x) = P\left[L_{\max,t} \leq x\right] = \left\{P\left[\hat{L}_{\max,t} \leq x\right]\right\}^\theta \tag{24}$$

3 A NUMERICAL EXAMPLE INVOLVING BRIDGE IN-SERVICE DATA

We now demonstrate the proposed methodology to estimate the maximum live load on a highway bridge

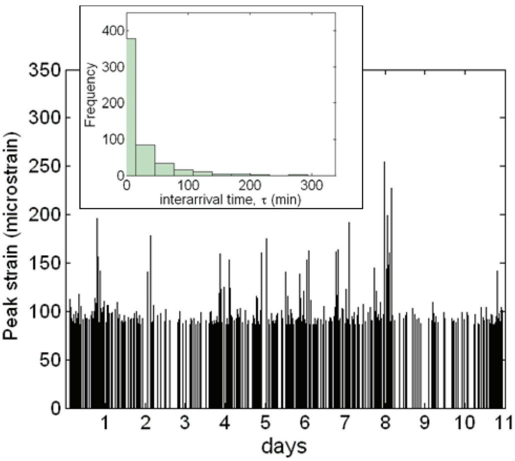

Figure 1. Time-line of loading events spanning 11 days in August 1998 (Inset: Distribution of interarrival times).

using data collected under ambient traffic. The in-service strain monitoring system is analogous to a weigh-in-motion system, and measures peak live-load bridge strains due to site-specific traffic over extended periods of time (Shenton III et al., 2000). The system continuously digitizes an analog signal at 1,200 Hz, and waits for a pre-specified strain threshold to be exceeded. When this threshold is exceeded, the system records, among others, the time at which the event took place and the peak strain during the event.

It should be highlighted that this data acquisition system looks at structural response under ambient unrestricted traffic and thus naturally includes all possible single and multiple presence truck loading cases.

The bridge selected for instrumentation and data acquisition was Bridge 1-791 which is a 3-span continuous, slab-on-steel girder structure carrying two lanes of Interstate-95 over Darley Road in Delaware. In-service strain data were recorded at midspan of the critical girder of the approach span (beneath the right travel lane) during an approximately 11-day period in August 1998 (Figure 1). The trigger level was set at 85 $\mu\varepsilon$ so that only 533 rather large loading events were recorded during the 11 day period.

The data was filtered by raising the threshold, u (Eqs (4) – (7)), and the interarrival times were found to become increasingly Exponential in nature (Bhattacharya, 2007) thus corroborating item 2 in Section 2 above.

3.1 *The run length and the extremal index*

The extremal index of the loading sequence, $\{L_n\}$, is estimated using Eq. (11). As detailed in (Bhattacharya, 2007), the estimate $\hat{\theta}$ shows a decreasing trend with increasing r for any particular value of u, which is consistent with the property that the extremal index is the reciprocal of the average number of exceedances per cluster, and with increasing run length the number of cluster goes down.

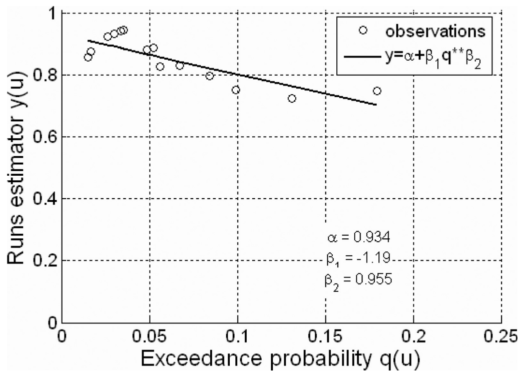

Figure 2. Minimum squared error estimate of θ (for r = 2).

Recall that r is an integer equal to or greater than 2. Three different series are used for estimating the run length: (i) the peak strain series, (ii) the hourly maximum series, and (iii) the two-hourly maximum series (Bhattacharya, 2007). For the first series, $\hat{\tau}_c$, converges between 1 and 3 as T grows large. We then look at the scale of fluctuation in the hourly maximum and two-hourly maximum loads. For both these sequences, the scale of fluctuation is found to converge to around 1 hour ($\sim 1 \times 1$ hr and $\sim 0.5 \times 2$ hr, respectively). Coupled with the observation that the average number of loading events is around 2 per hour (the mean interarrival time being about 29.3 min.), we adopt a value of 2 for r in this analysis.

The runs estimator $\hat{\theta}_R(u; r=2)$ (Eq. (11)) is now calculated as a function of the exceedance probability, $q(u) \equiv 1 - p(u)$, corresponding to u running from 100 to 175 in increments of 5. The exceedance probability $q(u) \equiv 1 - p(u)$ is estimated from Eq. (20). We apply Eqs (12) and (13) to obtain the minimum squared error estimate of the extremal index:

$$y(u) = \alpha + \beta_1 q(u)^{\beta_2} \tag{25}$$

where $y(u) \equiv \hat{\theta}_R(u; r=2)$ corresponds to the estimate, and $\alpha \equiv \theta$ corresponds to the true value (cf. Eq. (12)). The minimum squared error fit to the data according to Eq. (25) are shown in Figure 2, yielding a value of the extremal index as $\theta = 0.93$. This value of θ close to 1 indicates that the load sequence is almost independent, a likely consequence of the rather high trigger of 85 microstrain set for the in-service recording device.

3.2 *Cox process with lognormal arrival rate*

For the Cox process model for the load occurrence process, we assume a simple stationary model that gives non-negative smooth sample functions, namely, a lognormal process given by:

$$\Lambda(t) = \exp[\mu + \sigma z(t)] \tag{26}$$

where μ and σ are constants, and $z(t)$ is a zero-mean unit-variance stationary Gaussian process with auto-correlation function $\rho(\tau)$. The stationary mean and coefficient of variation (c.o.v.) of $\Lambda(t)$ are, respectively, $\mu_\Lambda = exp[\mu + \sigma^2/2]$ and $\delta_\Lambda = \sqrt{exp(\sigma^2) - 1}$. We further assume that the autocorrelation function (ACF) of $z(t)$ is exponentially decaying with correlation length, τ_0, i.e., $\rho(\tau) = exp(-|\tau|/\tau_0)$. The ACFs of $\Lambda(t)$ and $z(t)$ are not equal, being related through $\rho(\tau) = \ln(1 + \rho_\Lambda(\tau)\delta_\Lambda^2)/\ln(1 + \delta_\Lambda^2)$ (der Kiureghian & Liu, 1985), although, the difference becomes negligible for $\delta_\Lambda < 0.3$. The random mean measure, M_t (Eq. (15)), then has its first two moments as:

$$E[M_t] = t\exp(\mu + \sigma^2/2) \tag{27}$$

$$\text{var}[M_t] = 2\sigma_\Lambda^2 \left\{t\tau_0 + \tau_0^3(e^{-t/\tau_0} - 1)\right\} \tag{28}$$

where σ_Λ^2 is the stationary variance of the process Λ.

We estimate the unknown parameters μ, τ_0 and σ as follows: First, estimate $E[M_t]$ for different values of t, minimize the error with Eq. (27) and hence obtain least square estimate for μ_Λ. Then, estimate var $[M_t]$ for different values of t and minimize the error with Eq. (28); hence obtain least square estimates for τ_0 and σ_Λ. Use the estimates of σ_Λ and μ_Λ in turn to obtain μ and σ. Non-linear least square analyses of the data yielded the following estimated parameters: $\mu_\Lambda = 1.99$/hr, $\sigma_\Lambda = 1.50$/hr and $\tau_0 = 19.40$ hrs. Thus the estimated lognormal parameters are (Eq. (26)): $\mu = 0.53$ and $\sigma = 0.56$ when t is expressed in hours.

3.3 *Distribution of the maximum of the associated i.i.d. sequence*

The distribution of the maximum load, $\hat{L}_{\max,t}$, during the interval $(0, t]$ of the associated i.i.d. sequence is estimated next. We select the time interval $t = 1$ day. Point estimates of the empirical CDF, $\hat{p}$, of the load sequence $\{L_n\}$ at nine different strain values (l) are obtained. A Bayesian updating of the CDF is performed (re. Eq. (21) with $\theta = 1$). The c.o.v. of P is found to become smaller and smaller as one moves along the upper tail – a result of the reasonably large sample size. 10,000 Monte Carlo simulations were used in estimating Eq. (23) for each value of l to obtain the unconditional CDF of the daily maximum, $\hat{L}_{\max,1d}$, of the associated i.i.d. sequence (details are provided in (Bhattacharya, 2007)).

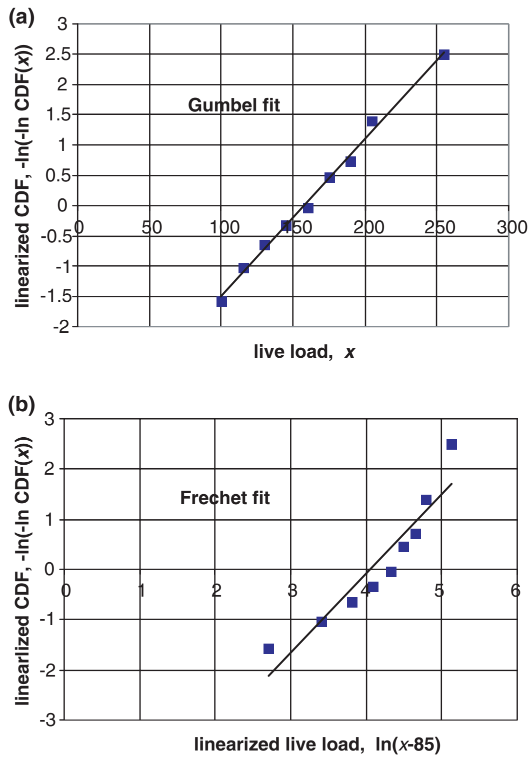

Figure 3. (a) Gumbel and (b) Frechet probability fit of daily maximum of the associated i.i.d. live-load sequence.

The maximum from an i.i.d. sequence approaches one of the three classical extreme value distributions for largest values. The Gumbel fit was better than the Frechet one in the present case (Figure 3) and was adopted for $\hat{L}_{\max,1d}$ in this paper:

$$F_{\hat{L}_{\max,1d}}(x) = \exp\left[-\exp\left(-\hat{\alpha}_{1d}(x - \hat{u}_{1d})\right)\right] \tag{29}$$

where $\hat{\alpha}$ and $\hat{u}$ are the scale and mode, respectively, of the maximum of the associated i.i.d. sequence. The Weibull distribution for maxima was not tried here since it is limited on the right which appeared to be an unreasonable restriction, although this property of the Weibull distribution can be attractive in situations where geometric or any other constraint puts a well-defined upper limit on the vehicular load on the bridge. Figure 3 also gives the best fit straight line through the data from which the parameters can be estimated as $\hat{\alpha}_{1d} = 0.0260$ and $\hat{u}_{1d} = 157.4$ microstrain.

3.4 *Maximum live load for various time intervals*

Recall that the distribution of the maximum of the associated i.i.d. sequence and the actual dependent

Table 1. Maximum live load statistics for different time intervals and effect of dependence in the parent loading process.

	Statistics of maximum live load, $L_{\max,t}$							
	$\theta = 1$		$\theta = 0.93$		$\theta = 0.75$		$\theta = 0.50$	
Time, t	Mean	c.o.v.	Mean	c.o.v.	Mean	c.o.v.	Mean	c.o.v.
1 day	178.0	0.275	177.0	0.279	168.7	0.293	153.3	0.322
1 year	407.7	0.121	404.1	0.122	395.9	0.125	380.1	0.130
2 years	434.5	0.114	430.8	0.115	422.6	0.117	406.7	0.121
10 years	496.6	0.0997	492.8	0.100	484.6	0.102	468.6	0.105
50 years	558.7	0.0886	554.7	0.0890	546.6	0.0904	530.5	0.0930
75 years	574.4	0.0862	570.3	0.0866	562.2	0.0879	546.1	0.0903

stationary sequence are of the same type (cf. Eqs. (8), (9) and (24)) differing only in terms of the shape parameter by the factor θ. For the Gumbel family which is closed under maximization, this leads to an unchanged α (hence an unchanged variance) and a mode (and mean) shifted to the left by an amount $(1/\alpha)\ln(1/\theta), 0 < \theta \leq 1$. The closer θ is to 1, the more modest is this shift. Hence, the Gumbel parameters of the actual maximum, $L_{max,rd}$, corresponding to r days ($r =$ integer) can be given by:

$$\alpha_{rd} = \hat{\alpha}_{1d} \equiv \alpha$$
$$u_{rd} = \hat{u}_{1d} - \frac{1}{\alpha}\ln\left(\frac{1}{\theta}\right) + \frac{1}{\alpha}\ln(r) \tag{30}$$

The mean and c.o.v. of the maximum live load for various time intervals, t, up to 75 years and for various values of θ are listed in Table 1. As expected, with increasing t, the maximum live load distribution becomes narrower and shifts to the right. The case of $\theta = 1$ signifying an i.i.d. assumption, is tabulated first. For the purpose of comparison, the consequence of lower values of θ, signifying increasingly greater dependence in the parent process, is also demonstrated in the Table. The extent of conservatism in the i.i.d. assumption is evident. In general, including the effect of dependence in the parent loading process decreases the mean and increases the c.o.v. of the maximum load. As may be expected, the influence of dependence in the parent loading process on the maximum diminishes with increasing t.

4 CONCLUSION

The methodology outline here allows the use of in-service data to obtain a realistic probabilistic model of extreme live loads. A considerable part of the methodology involved accounting for possible dependence in the parent loading process which was formalized as weak mixing type. Dependence in the arrival rate process as well as in the associated load magnitudes was considered. Inclusion of dependence eliminates the unnecessary conservatism introduced by the potentially unrealistic i.i.d. assumption.

The methodology was used to predict maximum live load on a highway bridge. The asymptotic behavior of extremes from the sample with increasing thresholds was investigated. The dependence was characterized using the extremal index and the marginal distribution of the parent process. The scale of fluctuation of the loading process was used to identify clusters of exceedances above high thresholds. A Bayesian updating, derived from the distribution of order statistics in a dependent stationary series, was performed on the sample distribution function. The parameters of the random arrival rate process modeled as a Cox process with stationary lognormal intensity were determined.

REFERENCES

ASCE (2002). Minimum Design Loads for Buildings and Other Structures. Reston, VA, ASCE.

Bhattacharya, B. (2007). The Extremal Index and the Maximum of a Dependent Stationary Pulse Load Process Observed above a High Threshold. *Structural Safety* in press.

Castillo, E. (1998). *Extreme Value Theory in Engineering*. New York, N.Y., Academic Press.

Cox, D. R. and V. Isham (2000). *Point Processes*. Boca Raton, Chapman & Hall/CRC.

der Kiureghian, A. and P.-L. Liu (1985). Structural reliability under incomplete probability information. *Journal of Engineering Mechanics, ASCE* 112(1): 85–104.

Fischer, W. and K. Meier-Hellstern (1992). The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation* 19: 149–171.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Malabar, FL, Krieger.

Galambos, J. (1993). The development of the mathematical theory of extremes in the past half century. *Theory of Probability and its Applications* 39(2): 234–248.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *The Annals of Statistics* 21(4): 2043–2071.

Leadbetter, M.R., Lindgren G. and Rootzen H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. New York, Springer-Verlag.

Moller, J., Syversveen A.R. and Waagepetersen R.P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* 25(3): 451–482.

Shenton III, H.W., Chajes M.J. and Holloway E.S. (2000). A system for monitoring live load strain in bridges. *CD Proceeding of the Structural Materials Technology IV: An NDT Conference*, Atlantic City, NJ, Lancaster Pennsylvania.

Smith, R. L. and Weissman I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society. Series B (Methodological)* 56(3): 515–528.

Tawn, J. (1990). *Discussion* on the paper “Models of exceedances over high threshold” by Davison and Smith. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(3): 428–429.

Vanmarcke, E. (1983). *Random Fields: Analysis and Synthesis*. Cambridge, MA, MIT Press.

Weissman, I. and Novak S.Y. (1998). On blocks and runs estimators of the extremal index. *Journal of Statistical Planning and Inference* 66: 281–288.

Wen, Y.-K. (1990). *Structural Load Modeling and Combination for Performance and Safety Evaluation*. Amsterdam, Elsevier.