PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 21, Fasc. 1 (2001), pp. 71–88

LIMITS OF TRUNCATION EXPERIMENTS

BY

FRANK MAROHN (EICHSTÄTT)

Abstract. Given n i.i.d. copies $X_1, \ldots, X_n$ of a random variable X with distribution P_ϑ, $\vartheta \in \Theta \subset \boldsymbol{R}^k$, we are only interested in those observations that fall into some set $D = D(n) \subset \boldsymbol{R}^d$ having but a small probability of occurrence. The truncation set D is assumed to be known and non-random. Denoting the distribution of the truncated random variable $X 1_D(X)$ by $P_{n\vartheta}$ we consider the triangular array of experiments $(\boldsymbol{R}^d, \mathscr{B}^d, (P_{n\vartheta})_{\vartheta\in\Theta})$, $n \in \boldsymbol{N}$, and investigate the asymptotic behavior of the n-fold products $((\boldsymbol{R}^d)^n, (\mathscr{B}^d)^n, (P_{n\vartheta})^n_{\vartheta\in\Theta})$. Under a suitable density expansion, Gaussian shifts as well as Poisson experiments occur in the limit, where the latter case typically occurs when the number of expected observations falling in D is bounded. Finally, we investigate invariance properties of the occurring Poisson limits.

MSC 1991 Subject Classification: 62B15.

Key words and phrases: Statistical experiment, truncation model, Gaussian experiment, local asymptotic normality, Poisson experiment, Poisson process, translation invariance, scale invariance, independent increments.

1. INTRODUCTION

Suppose that $X_1, \ldots, X_n$ are i.i.d. copies of a real-valued random variable X with distribution P_ϑ, where the parameter $\vartheta = (\vartheta_1, \ldots, \vartheta_k)$ belongs to some open set $\Theta \subset \boldsymbol{R}^k$, and that we are only interested in those observations among $X_1, \ldots, X_n$ which fall into a measurable set $D \subset \boldsymbol{R}^d$. Typical examples occur in regression analysis and density estimation, in which case D is located in the center of the underlying distribution (see, e.g., Falk and Reiss [7], Falk and Marohn [6]) or in extreme value analysis, where D is located at the border (see, e.g., Resnick [18], Reiss [17], Falk et al. [4], Falk [2]). In the following the truncation set D is assumed to be known and non-random. We are primarily interested in the cases where $D = D(n)$ has shrinking probability with increasing sample size n. Thus, the present approach is not meant in the sense of survival analysis, where typically the truncation is random and independent of the sample size n (see, e.g., the monograph by Andersen et al. [1]).

The present situation can be described by the truncated random variable $X 1_D(X)$. If P_ϑ has a Lebesgue density f_ϑ, then $X 1_D(X)$ has the distribution

$$P_{n\vartheta} := f_\vartheta 1_D \, d\lambda^d + P_\vartheta(D^c)\,\varepsilon_0,$$

where λ^d denotes the d-dimensional Lebesgue measure, and ε_z the Dirac measure at point $z \in \mathbf{R}^d$. The distribution $P_{n\vartheta}$ has then with respect to $\lambda^d + \varepsilon_0$ the density

$$\frac{dP_{n\vartheta}}{d(\lambda^d + \varepsilon_0)} = f_\vartheta(x)\, 1_D(x)\, 1_{\mathbf{R}^d \setminus \{0\}}(x) + P_\vartheta(D^c)\, 1_{\{0\}}(x).$$

We consider the triangular array of rowwise identical experiments $(\mathbf{R}^d, \mathscr{B}^d, \{P_{n\vartheta}: \vartheta \in \Theta\})$, $n \in N$, which are called *truncation experiments*. Denoting the n-fold product measure $P_{n\vartheta} \otimes \ldots \otimes P_{n\vartheta}$ by $P^n_{n\vartheta}$, we are interested in the asymptotic behavior of the rescaled product experiments

$$E_{n,n} = \left((\mathbf{R}^d)^n, (\mathscr{B}^d)^n, (P^n_{n\vartheta_0 + t\delta_n})_{t \in T_n}\right), \qquad \vartheta_0 \in \Theta,$$

where the set $D = D(n)$ is a rare event, i.e., $P_{\vartheta_0}(D) \xrightarrow{n \to \infty} 0$. Thereby,

$$T_n := \{t \in \mathbf{R}^k : \vartheta_0 + t\delta_n \in \Theta\}$$

and $\delta_n = (\delta_{n1}, \ldots, \delta_{nk})$ is a positive sequence in $\mathbf{R}^k$ usually tending to zero. The multiplication is meant componentwise i.e., $t\delta_n = (t_1 \delta_{n1}, \ldots, t_k \delta_{nk})$. Recall that the rescaling procedure is in general necessary in order to get a non-degenerated limit experiment. For an appealing review of the local approach we refer to the introductory part of Janssen et al. [10].

In recent papers (Falk [3], Marohn [14]) it was shown that under a certain density expansion of D local asymptotic normality (LAN) holds, i.e., the sequence $(E_n)_n$ converges weakly to a Gaussian shift experiment. This was established by direct calculations, i.e., by showing the validity of the LAN expansion of the likelihood ratios.

On the other hand, there exists a well-developed theory of the limit behavior of (rescaled) product experiments (LeCam [11], Strasser [19], Janssen et al. [10], Janssen [8]). It is known that under fairly general conditions every weak accumulation point of product experiments of a triangular array of experiments is a product of a Gaussian experiment and a Poisson experiment. Moreover, under certain conditions, every limit experiment fulfills invariance properties, namely translation invariance and scale invariance. This was the motivation of the present paper. Applying the theory of statistical experiments, we investigate the asymptotic behavior of the product of a truncation experiment. Describing the local structure by a density expansion, it turns out that Gaussian as well as Poisson experiments occur in the limit, where the latter typically occur if the number of expected observations falling in D is bounded. It was shown in Janssen and Marohn [9] that in extreme value analysis Poisson experiments occur if the local structure is described by means of the (infinitely dimensional) extreme value tangent space.

For the remainder of this section, the notation is introduced and some facts concerning statistical experiments are recalled. The reader is referred to LeCam [12], LeCam and Yang [13], Janssen et al. [10], Strasser [21], and Torgersen [22].

Let $E = (\Omega, \mathscr{A}, (P_t)_{t\in T})$ be a statistical experiments, i.e., $(\Omega, \mathscr{A})$ is a measurable space and $(P_t)_{t\in T}$ is a family of probability measures indexed by a set T. Sometimes, we simply write $(P_t)_t$ for E and for binary experiments we use the condensed form (P_s, P_t). Denote by dP_t/dP_s the Radon–Nikodym density of the absolutely continuous part of P_t with respect to P_s. The likelihood process of E with base $s\in T$ is $(dP_t/dP_s)_{t\in T}$, the distribution is taken with respect to P_s. Two statistical experiments $E = (\Omega_1, \mathscr{A}_1, (P_t)_{t\in T})$ and $F = (\Omega_2, \mathscr{A}_2, (Q_t)_{t\in T})$ are called *equivalent* (in the notation, $E \sim F$) if

$$\mathscr{L}\left(\left(\frac{dP_t}{dP_s}\right)_{t\in T}\middle| P_s\right) = \mathscr{L}\left(\left(\frac{dQ_t}{dQ_s}\right)_{t\in T}\middle| Q_s\right), \qquad s\in T.$$

The respective classes are called *experiment-types*.

Let $A(T)$ be the class of all finite non-empty subsets of T. A sequence of experiments $E_n = (\Omega_n, \mathscr{A}_n, (P_{nt})_{t\in T})$, $n\in N$, converges weakly to an experiment $E = (\Omega, \mathscr{A}, (P_t)_{t\in T})$ (in the notation, $E_n \to E$) if for every $\alpha\in A(T)$ and every $s\in\alpha$

$$\mathscr{L}\left(\left(\frac{dP_{nt}}{dP_{ns}}\right)_{t\in\alpha}\middle| P_{ns}\right) \to \mathscr{L}\left(\left(\frac{dP_t}{dP_s}\right)_{t\in\alpha}\middle| P_s\right) \text{ weakly.}$$

Denote by $E_\alpha = (\Omega, \mathscr{A}, (P_t)_{t\in\alpha})$ the restriction of E on α and by

$$S_\alpha := \{(x_t)_{t\in\alpha} \in [0, 1]^\alpha\colon \sum_{t\in\alpha} x_t = 1\}$$

the simplex in $\mathbf{R}^\alpha$. If μ is a σ-finite measure on $\mathscr{A}$ dominating $\{P_t\colon t\in\alpha\}$, then

$$H(E_\alpha)(z) := \int \prod_{t\in\alpha} \left(\frac{dP_t}{d\mu}\right)^{z_t} d\mu, \qquad z\in S_\alpha,$$

is called the *Hellinger transform* of E_α. It is well known that $E \sim F$ if and only if $H(E_\alpha) = H(F_\alpha)$, $\alpha\in T$, and that $E_n \to E$ weakly if and only if $H(E_{n,\alpha}) \to H(E_\alpha)$ pointwise on S_α, $\alpha\in A(T)$. Denote by

$$d_2(P_s, P_t) = \frac{1}{\sqrt{2}}\left(\int\left(\left(\frac{dP_s}{d\mu}\right)^{1/2} - \left(\frac{dP_t}{d\mu}\right)^{1/2}\right)^2 d\mu\right)^{1/2}$$

the Hellinger distance between P_s and P_t. Note the relation

$$d_2(P_s, P_t) = \left(1 - H(E_{\{s,t\}})(1/2, 1/2)\right)^{1/2}.$$

An experiment is a *Gaussian experiment* if it is homogeneous and if one loglikelihood process is a Gaussian process. The most simple Gaussian experiment is the *Gaussian shift* $\left(\mathbf{R}^k, \mathscr{B}^k, (N(\Gamma t, t))_{t\in\mathbf{R}^k}\right)$, where $N(\mu, \Gamma)$ denotes the normal distribution with mean vector zero and covariance matrix Γ on $\boldsymbol{R}^k$. Let $(\Omega, \mathscr{A}, P)$ be a probability space and let $(S, \mathscr{B})$ be a measurable space. A measurable mapping $N: (\Omega, \mathscr{A}, P) \to (M(S, \mathscr{B}), \mathscr{M}(S, \mathscr{B}))$ into the space of point measures on $(S, \mathscr{B})$ equipped with a suitable σ-field $\mathscr{M}(S, \mathscr{B})$ is called a *point process* (for details see, e.g., Resnick [18] and Reiss [17]). Let ν be a finite measure on $(S, \mathscr{B})$ and denote by $\mathrm{Po}(\lambda)$ the Poisson distribution with parameter $\lambda > 0$. A point process N is called a *Poisson process* with intensity measure ν if $N(B)$ is $\mathrm{Po}(\nu(B))$-distributed for $B \in \mathscr{B}$ and $N(B_1), \ldots, N(B_k)$ are independent for every $k \in N$ and pairwise disjoint sets $B_1, \ldots, B_k \in \mathscr{B}$. Let N_1 and N_0 be two Poisson processes with (finite) intensity measures ν_1 and ν_0, respectively, on $(S, \mathscr{B})$. If ν_1 has a ν_0-density h, then

$$\frac{d\mathscr{L}(N_1)}{d\mathscr{L}(N_0)}(\mu) = \Big(\prod_{i=1}^{\mu(S)} h(x_i)\Big) \exp\big(\nu_0(S) - \nu_1(S)\big)$$

if $\mu = \sum_{i=1}^{\mu(S)} \varepsilon_{x_i}$ (for the density formula see, e.g., Reiss [17], Theorem 3.1.1). If N_t is a family of Poisson processes with intensities ν_t, $t \in T$, then $\big(\mathscr{L}(N_t)\big)_{t \in T}$ is a *compound Poisson experiment*. (Here we use the terminology of Milbrodt and Strasser [15]. If the ν_t are σ-finite intensity measures, then $\big(\mathscr{L}(N_t)\big)_{t \in T}$ is called a *Poisson experiment*).

2. RESULTS

We assume the density expansion

(1) $$\left(\frac{f_\vartheta}{f_{\vartheta_0}}\right)^{1/2} = 1 + \langle \vartheta - \vartheta_0, g/2 \rangle + \|\vartheta - \vartheta_0\| \, r_{n\vartheta}$$

on D, $g = (g_1, \ldots, g_k)$, where the remainder term $r_{n\vartheta}$ vanishes asymptotically in a certain sense (see condition (3) below). Condition (1) is a slightly stronger condition compared to that of Falk [3] and Marohn [4]. Despite the fact that there is some similarity of (1) to the concept of L_2-differentiability, there is one crucial difference: The "tangent" g does *not* necessarily satisfy $\int_D g \, dP_{\vartheta_0} = 0$.

In the following we restrict ourselves to the case $\vartheta_0 = 0$. From Marohn [14] we adapt the rescaling rate

(2) $$\delta_{nj} = \frac{1}{\sqrt{n \int_D g_j^2 \, dP_{\vartheta_0}}},$$

and the neglibility condition on the remainder term

(3) $$\frac{\int_D r_{nt\delta_n}^2 \, dP_0}{\int_D g_j^2 \, dP_0} \xrightarrow{n \to \infty} 0, \qquad t \in T_n, \ j = 1, \ldots, k.$$

Moreover, we assume as in Marohn [14] that

$$\text{(4)}\qquad \frac{\int_D g_j\,dP_0}{P_0(D)}\bigg/\sqrt{\frac{\int_D g_j^2\,dP_0}{P_0(D)}} \xrightarrow{n\to\infty} a_j\in[-1,1],\quad j=1,\dots,k,$$

and that the correlations (conditional on D) satisfy

$$\text{(5)}\qquad \frac{\dfrac{1}{P_0(D)}\int_D\left(g_i-\dfrac{\int_D g_i\,dP_0}{P_0(D)}\right)\left(g_j-\dfrac{\int_D g_j\,dP_0}{P_0(D)}\right)dP_0}{\sqrt{\dfrac{\int_D g_i^2\,dP_0}{P_0(D)}-\left(\dfrac{\int_D g_i\,dP_0}{P_0(D)}\right)^2}\sqrt{\dfrac{\int_D g_j^2\,dP_0}{P_0(D)}-\left(\dfrac{\int_D g_j\,dP_0}{P_0(D)}\right)^2}} \xrightarrow{n\to\infty} \varrho_{ij}$$

for $1\leqslant i,j\leqslant k$. The following lemma is crucial. It shows once more that the rescaling rate as defined in (2) is suitable. We do not assume that δ_{nj} tends to zero as $n\to\infty$.

2.1. LEMMA. *For $s,t\in\mathbf{R}^k$ we have*

$$a(s,t):=\lim_{n\to\infty} nd_2^2(P_{ns\delta_n},P_{nt\delta_n})=\tfrac{1}{8}(s-t)'\Gamma(s-t),$$

where Γ denotes the $k\times k$-matrix

$$\Gamma=(\gamma_{ij})=\left(a_ia_j+(1-a_i^2)^{1/2}(1-a_j^2)^{1/2}\varrho_{ij}\right).$$

For $i=j$ we have $\gamma_{ij}=1$. The matrix Γ was introduced in Marohn [14]. It turned out that Γ is a correlation matrix and therefore positive semidefinite. If it is positive definite, then $a(s,t)>0$ iff $s\neq t$. In this case every weak accumulation point of $(E_n)_n$ is homogeneous and the corresponding subsequence is contiguous (see, e.g., Strasser [21], Theorem 61.3). Moreover, the preceding lemma implies that the sequence $(E_n)_n$ is *bounded* and *infinitesimal* in the sense of Definition (5.4) in Milbrodt and Strasser [15], and thus every weak accumulation point is infinitely divisible. If $(P^n_{ns\delta_n},P^n_{nt\delta_n})\to(Q_s,Q_t)$, then

$$\text{(6)}\qquad \lim_{n\to\infty} d_2^2(P^n_{ns\delta_n},P^n_{nt\delta_n})=d_2^2(Q_s,Q_t)=1-\exp\left(-\tfrac{1}{8}(s-t)'\Gamma(s-t)\right)$$

(cf. Milbrodt and Strasser [15], Lemma 5.7 and its proof).

Proof of Lemma 2.1. Applying the Cauchy–Schwarz inequality and the Minkowski inequality, by straightforward calculations as in Marohn [14] we obtain

$$\begin{aligned} nd_2^2(P_{ns\delta_n},P_{nt\delta_n}) &= n\frac{1}{2}\int\left(\left(\frac{dP_{ns\delta_n}}{d(\lambda+\varepsilon_0)}\right)^{1/2}-\left(\frac{dP_{nt\delta_n}}{d(\lambda+\varepsilon_0)}\right)^{1/2}\right)^2 d(\lambda+\varepsilon_0)\\ &= n\frac{1}{2}\int_D\left(\left(\frac{f_{s\delta_n}}{f_0}\right)^{1/2}-\left(\frac{f_{t\delta_n}}{f_0}\right)^{1/2}\right)^2 dP_0+\frac{n}{2}\left(P_{s\delta_n}^{1/2}(D^c)-P_{t\delta_n}^{1/2}(D^c)\right)^2\end{aligned}$$

$$= n\,\frac{1}{8}\int_D \langle (s-t)\,\delta_n,\, g_n\rangle^2\, dP_0 + \frac{n}{2}\left(\frac{P_{s\delta_n}(D)-P_{t\delta_n}(D)}{P^{1/2}_{s\delta_n}(D^c)+P^{1/2}_{t\delta_n}(D^c)}\right)^2 + o(1)$$

$$= \frac{1}{8}(s-t)'\,\Gamma\,(s-t)+o(1)$$

and

$$|P_{s\delta_n}(D)-P_{t\delta_n}(D)| \leqslant \int_D |f_{s\delta_n}-f_{t\delta_n}|\, d\lambda = \int_D |f^{1/2}_{s\delta_n}-f^{1/2}_{t\delta_n}|\,|f^{1/2}_{s\delta_n}+f^{1/2}_{t\delta_n}|\, d\lambda$$

$$\leqslant \Big(\int_D (f^{1/2}_{s\delta_n}-f^{1/2}_{t\delta_n})^2\, d\lambda\Big)^{1/2} \Big(\int_D (f^{1/2}_{s\delta_n}+f^{1/2}_{t\delta_n})^2\, d\lambda\Big)^{1/2}$$

$$\leqslant \left(\int_D \left(\left(\frac{f_{s\delta_n}}{f_0}\right)^{1/2} - \left(\frac{f_{t\delta_n}}{f_0}\right)^{1/2}\right)^2 dP_0\right)^{1/2} (P^{1/2}_{s\delta_n}(D)+P^{1/2}_{t\delta_n})(D)).$$

Note that $P_{t\delta_n}(D)\to 0$ for $n\to\infty$, $t\in \mathbf{R}$, which is a consequence of the expansion

$$\text{(7)}\qquad f_t/f_0 = (1+\langle t,\, g_n/2\rangle + \|t\|\, r_{nt})^2$$
$$= 1+\langle t,\, g_n\rangle + \langle t,\, g_n/2\rangle^2 + 2\,\|t\|\, r_{nt} + \langle t,\, g_n\rangle\, \|t\|\, r_{nt} + \|t\|^2\, r^2_{nt}$$

and the assumptions (3) and (4). ∎

THE GAUSSIAN CASE

In this section we give a different proof of LAN to Falk [3] and Marohn [14], using the theory of statistical experiments. The following theorem gives a sufficient condition for the LAN case of truncation experiments:

2.2. THEOREM. *For* $\varepsilon > 0$ *define the sets* $A_{nj} = \{x\in D\colon\ |g_j(x)| > \varepsilon/\delta_{nj}\}$, $j = 1, \ldots, k$. *If for every* $\varepsilon > 0$

$$\text{(8)}\qquad \frac{\int_{A_{nj}} g_j^2\, dP_0}{\int_D g_j^2\, dP_0} \xrightarrow{n\to\infty} 0, \qquad j = 1, \ldots, k,$$

then

$$E_{n,n} \to \big(\mathbf{R}^k,\, \mathscr{B}^k,\, (N(\Gamma t,\, t))_{t\in \mathbf{R}^k}\big)$$

with Γ *as in Lemma* 2.1.

Condition (8), which is of Lindeberg type, claims that D has to shrink not too fast and that the rescaling rate δ_n has to tend to zero, since (8) implies $nP(D)\to\infty$, as we will see below, and $nP_0(A_{nj})\to 0$. It turned out that condition (8) is crucial in order to derive the LAN expansion of the likelihood ratios; cf. Falk and Liese [5]. Thereby, it was assumed that the number of expected observations falling in D tends to infinity, i.e., $nP_0(D)\to\infty$. But this is already a consequence of condition (8). Since

$$1=\frac{\int_{A_{nj}} g_j^2\,dP_0}{\int_D g_j^2\,dP_0}+\frac{\int_{A_{nj}^c\cap D} g_j^2\,dP_0}{\int_D g_j^2\,dP_0}$$

and the second term becomes arbitrarily small if $nP_0(D)$ is bounded,

$$\frac{\int_{A_{nj}^c\cap D} g_j^2\,dP_0}{\int_D g_j^2\,dP_0}\leqslant\frac{(\varepsilon/\delta_{nj})^2 P_0(D)}{\int_D g_j^2\,dP_0}=\varepsilon\, n\, P_0(D),$$

we conclude that $nP_0(D)\to\infty$ is a necessary condition for (8). This can also be seen from a different point of view: If $nP_0(D)$ is bounded, then Poisson sequences occur as we show in the following. Since only the trivial statistical experiment is Gaussian as well as Poisson and under condition (8) the Gaussian limit is non--trivial, we conclude that $nP_0(D)$ must converge to infinity. Finally, we remark that there is a formal accordance between this result and the well-known approximation of the binomial distribution by the Poisson distribution, if the probability of occurrence is small, and by the normal distribution, else.

Proof of Theorem 2.2. According to Theorem (6.3) and Remark (6.6) in Milbrodt and Strasser [15], $(E_{n,n})_n$ is Gaussian (i.e. every weak accumulation point is a Gaussian experiment) iff for every $t\in T$ and $\varepsilon>0$

$$nP_{n0}\left\{\left|\frac{dP_{nt\delta_n}}{dP_{n0}}-1\right|>\varepsilon\right\}\xrightarrow{n\to\infty}0. \tag{9}$$

Taking into account relation (7), we have

$$\begin{aligned}P_{n0}\{|dP_{nt\delta_n}/dP_{n0}-1|>\varepsilon\} &= P_0\{|f_{t\delta_n}/f_0-1|>\varepsilon\}\quad\text{(for } n \text{ large)}\\ &\leqslant P_0\{|\langle t\delta_n, g_n\rangle|>\varepsilon/5\}+P_0\{\langle t\delta_n, g/2\rangle^2>\varepsilon/5\}+P_0\{2\,\|t\delta_n\|\,|r_{nt\delta_n}|>\varepsilon/5\}\\ &\quad+P_0\{|\langle t\delta_n, g_n\rangle|\,|t\delta_n|\,|r_{nt\delta_n}|>\varepsilon/5\}+P_0\{\|t\delta_n\|^2\, r_{nt\delta_n}^2>\varepsilon/5\}.\end{aligned}$$

Applying the Markov inequality, we get

$$\begin{aligned}nP_0\{|\langle t\delta_n, g_n\rangle|>\varepsilon/5\} &\leqslant \sum_{j=1}^{k} nP_0\left\{|g_j|>\frac{\varepsilon}{5k\,|t_j|\,\delta_{nj}}\right\}\\ &\leqslant \sum_{j=1}^{k} n\left(\frac{5kt_j\delta_{nj}}{\varepsilon}\right)^2\int_{\{|g_j|>\varepsilon/(5k|t_j|\delta_{nj})\}} g_j^2\,dP_0\xrightarrow{n\to\infty}0\end{aligned}$$

by condition (8). The other terms tend also to zero, which is verified by similar arguments, and condition (9) is therefore satisfied. Lemma 2.1 implies now the weak convergence to a Gaussian experiment with covariance

$$\begin{aligned}K(s,t) &= 4\big(a(s,0)+a(t,0)-a(s,t)\big)\\ &= 4\big(\tfrac{1}{8}s'\Gamma s+\tfrac{1}{8}t'\Gamma t-\tfrac{1}{8}(s-t)'\,\Gamma\,(s-t)\big)=s'\Gamma t\end{aligned}$$

(see Milbrodt and Strasser [15], Corollary (6.4) and Remark (6.5)). ∎

Lemma 2.1 shows that the Gaussian experiment is continuous, translation invariant and scale invariant with exponent 2. This holds if and only if the Gaussian experiment is a Gaussian shift. This characterization was established by Strasser [19], Theorem (4.3). That the Gaussian experiment must be a Gaussian shift can also be seen from Lemma (5.1) in Janssen [8] together with relation (6). Finally, Theorem 16.2 in Strasser [20] together with relation (6) show that the Gaussian shift has no independent increments.

We now discuss the rate of convergence of the rescaling sequence $(\delta_n)_n$. First, the following fact is recalled. In the i.i.d. case, LAN can be characterized by the rescaling rate. Strasser [19] showed that LAN holds if and only if the rescaling rate is given by $n^{-1/2}a_n$, where $(a_n)_n$ is a slowly varying sequence, i.e., $\lim_{n\to\infty} a_{nm}/a_n = 1$ for every $m \in N$. In the present situation of non-identically distributed random variables (note that the underlying probability measures depend on $D = D(n)$), this rate is no longer expected. Due to the truncation we lose information about the unknown parameters and, since the rescaling rate is a measure of the increase of precision achieved by enlarging the sample size, the rate must be of a lower rate than $n^{-1/2}$. Typically, the rates are of the order $n^{-1/\alpha}a_n$ with $\alpha > 2$ depending on the choice of D. To illustrate this point, consider, for example, the Gaussian shift $N(\vartheta, 1)$, $\vartheta \in R$, and the set $D = [d_n, \infty)$ with $d_n \to \infty$. Then expansion (1) holds with $g(x) := g_1(x) = x$, condition (4) holds with $a = 1$, and condition (8) is fulfilled whenever d_n tends to infinity not too fast. Denote by Φ the distribution function of the standard normal distribution $N(0, 1)$, and by φ its density. The relation $1-\Phi(x) \sim \varphi(x)/x$, $x \to \infty$, implies

$$nP_0(D) \sim n\varphi(d_n)/d_n \quad \text{and} \quad \delta_n^{-2} \sim nd_n\varphi(d_n).$$

Now, $nP_0(D) \to \infty$ is fulfilled for the slowly varying functions $d_n = \log(n^{1/s})$, $s > 1$, and for these we obtain the rates

$$\delta_n \sim n^{-2^{-1}(1-1/(2s))}a_n, \qquad s > 1,$$

with $(a_n)_n$ slowly varying.

THE POISSON CASE

The following theorem shows that Poisson accumulation points occur if the number of expected observations falling in D is bounded.

2.3. THEOREM. *If* $\limsup_{n\to\infty} nP_0(D) < \infty$, *then every weak accumulation point of* $(E_{n,n})_n$ *is a compound Poisson experiment.*

Proof. According to Theorem (6.9) in Milbrodt and Strasser [15], $(E_{n,n})_n$ is Poisson (i.e. every weak accumulation point is a Poisson experiment) if and only if for every pair $(s, t) \in T^2$

$$\lim_{\varepsilon\to\infty} \limsup_{n\to\infty} n \int_{\{|dP_{nt\delta_n}/dP_{ns\delta_n}-1|<\varepsilon\}} \left(\frac{dP_{nt\delta_n}}{dP_{ns\delta_n}}-1\right)^2 dP_{ns\delta_n} = 0. \tag{10}$$

Now,

$$n \int_{\{|dP_{nt\delta_n}/dP_{ns\delta_n}-1|<\varepsilon\}} \left(\frac{dP_{nt\delta_n}}{dP_{ns\delta_n}}-1\right)^2 dP_{ns\delta_n} \leqslant \varepsilon^2 nP_{s\delta_n}(D)+n\left(\frac{P_{t\delta_n}(D^c)}{P_{s\delta_n}(D^c)}-1\right)^2 P_{s\delta_n}(D^c)$$
$$= \varepsilon^2 nP_{s\delta_n}(D)+n\left(P_{s\delta_n}(D)-P_{t\delta_n}(D)\right)^2/P_{s\delta_n}(D^c).$$

Since the first term is bounded and the second term vanishes as $n \to \infty$ (whether $nP_0(D)$ is bounded or not, see the proof of Lemma 2.1), condition (10) is fulfilled. That every Poisson accumulation point is a compound Poisson experiment follows now from Theorem (6.11) in Milbrodt and Strasser [15], since

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} n\, P_{ns\delta_n}\{|dP_{nt\delta_n}/dP_{ns\delta_n}-1| > \varepsilon\}$$
$$= \lim_{\varepsilon\to 0}\limsup_{n\to\infty} n\, P_{s\delta_n}\{|dP_{t\delta_n}/dP_{s\delta_n}-1| > \varepsilon\} \leqslant \lim_{\varepsilon\to 0}\lim_{n\to\infty} nP_{s\delta_n}(D) < \infty. \blacksquare$$

Below we will see that under suitable conditions weak convergence to a compound Poisson experiment holds (Theorem 2.5).

The Gaussian case as well as the Poisson case occur for the rescaling rate

$$\delta_{nj} = \left(nP_0(D)\right)^{-1/2}\left(\frac{\int_D g_j^2\, dP_0}{P_0(D)}\right)^{-1/2}. \tag{11}$$

The crucial point is that in the LAN case the rescaling rate has to converge to zero, whereas in the Poisson case the rescaling rate may not necessarily converge to zero. Indeed, $(\delta_n)_n$ may be bounded or can even tend to infinity (see the examples given below). Recall again that the underlying probability measures depend on the sample size n through the set D, which becomes smaller for increasing sample sizes.. Hence, it can happen that the statistical information is kept constant or even decreases by enlarging the sample size. Moreover, in the Gaussian case the first factor in (11) becomes relevant, whereas in the Poisson case the second factor in (11) determines the rescaling rate.

But before giving the examples we have to specify the parameter space. Throughout we assume that $nP_0(D) \to \lambda \in (0, \infty)$. Since

$$\lim_{n\to\infty} n\int_D \left(\frac{f_{t\delta_n}}{f_0}\right)^{1/2} dP_0 = \lim_{n\to\infty} n\int_D (1+\langle t\delta_n, g_n/2\rangle+\|t\delta_n\|\, r_{nt\delta_n})\, dP_0$$
$$= \lambda+\frac{\lambda^{1/2}}{2}\sum_{j=1}^{k} a_j t_j,$$

one has to choose $t\in T$ with

$$T = \left\{t\in \mathbf{R}^k\colon\ \lambda+\frac{\lambda^{1/2}}{2}\sum_{j=1}^{k} a_j t_j \geqslant 0\right\}.$$

If $a_j \geqslant 0$ for $j = 1, \ldots, k$, which covers the most interesting examples, then we may choose the cone $T = [0, \infty)^k$. If $a_j = 0$ for $j = 1, \ldots, k$, then $T = \mathbf{R}^k$.

For the following examples LAN was established if $nP_0(D)$ tends to infinity and D shrinks not too fast (cf. Falk [3], Marohn [14]). If $k=1$, we simply write g, a instead of g_1, a_1, etc.

2.4. Examples. (i) (Normal case) Let $P_\vartheta = N(\vartheta, 1)$, $\vartheta \geqslant 0$, be the Gaussian shift. Denote as before by Φ the distribution function of the standard normal distribution $N(0, 1)$ and by φ its density. Choose again $D = D(n) = [d_n, \infty)$ such that $nN(0, 1)(D) = \lambda$, i.e., $d_n = \Phi^{-1}(1-\lambda/n)$. Condition (4) is satisfied with $a=1$. We have

$$\delta_n = \Big(n \int_D g^2 \, dN(0, 1)\Big)^{-1/2} \to \infty$$

since, by integration by parts,

$$\frac{\int_D g^2 \, dN(0, 1)}{N(0, 1)(D)} = \frac{d_n \varphi(d_n) + 1 - \Phi(d_n)}{1-\Phi(d_n)} = d_n^2 \frac{\varphi(d_n)/d_n}{1-\Phi(d_n)} + 1 \to \infty$$

as $n \to \infty$. Note that $\varphi(x)/x \sim 1-\Phi(x)$, $x \to \infty$. It can be shown that $\delta_n = O\big((\log n)^{-1/2}\big)$ (cf. Reiss [16], Example 5.1.4).

(ii) (Pareto) Let $P_\vartheta = H_\beta$, $\beta > 0$, be the Pareto distributions defined by $H_\beta\big((-\infty, x]\big) = 1 - x^{-\beta}$, $x > 1$, which play a central role in extreme value analysis. Fix $\beta_0 > 0$ and let $D = [d_n, \infty)$. Then $g(x) = 1/\beta_0 - \log(x)$ and condition (4) holds again with $a=1$. Now choose D such that $nH_{\beta_0}(D) = nd_n^{-\beta_0} = \lambda$, i.e., $d_n = (n/\lambda)^{1/\beta_0}$. Integration by parts yields

$$\frac{\int_D g^2 \, dH_{\beta_0}}{H_{\beta_0}(D)} = \frac{(1/\beta_0^2 + \log^2 d_n) d_n^{-\beta_0}}{d_n^{-\beta_0}} = \frac{1}{\beta_0^2}\big(1 + \log^2(n/\lambda)\big) \to \infty$$

showing that δ_n converges to zero. As in the previous example, the rescaling sequence $(\delta_n)_n$ is slowly varying.

(iii) (Double exponential) Let P_ϑ, $\vartheta \in \mathbf{R}$, be the translation family of the double exponential distribution with Lebesgue density $f(x) = \exp(-|x|)/2$, $x \in \mathbf{R}$. Choose $D = [-d_n, d_n]$. Then $g(x) = \operatorname{sign}(x)$ and condition (4) holds with $a=0$. We get Poisson limits if $nP_0(D)$ is bounded (see also below) and the rescaling rate $\delta_n = \big(nP_0(D)\big)^{-1/2}$ converges to $\lambda^{-1/2}$ whenever $nP_0(D) \to \lambda$.

(iv) (Pareto shift) Fix some $\beta > 0$ and denote by $P_0 = H_\beta$ the Pareto distribution with shape parameter β (see (ii)). Let $(P_\vartheta)_\vartheta$ be the Pareto shift with location parameter $\vartheta \geqslant 0$. Choose $D = [d_n, \infty)$ with $d_n \to \infty$ such that $nP(D) = \lambda$. Then we have $g(x) = (1+\beta)/x$ and condition (4) holds with

$$a = \big(\beta/(1+\beta)\big)\sqrt{1+2/\beta} < 1.$$

The rescaling rate tends to infinity, since

$$\frac{\int_D g^2 \, dH_\beta}{H_\beta(D)} = \frac{(1+\beta)^2 \beta}{2+\beta} d_n^{-2} \to 0.$$

(v) Denote by P_ϑ, $\vartheta = (\vartheta_1, \vartheta_2)$, the normal distribution with expectation ϑ_1 and variance $(1+\vartheta_2)^{-2}$, $\vartheta \in \Theta = [0, \infty)^2$, i.e. $P_\vartheta = N(\vartheta_1, (1+\vartheta_2)^{-2})$. Choose $D = [-d_n, d_n]$ and let $d_n \to 0$. Then expansion (1) is satisfied with $g_1(x) = x$ and $g_2(x) = 1-x^2$. Integration by parts yields

$$\int_D g_1\, dN(0, 1) = 0, \qquad \int_D g_1^2\, dN(0, 1) = P_0(D) - 2d_n \varphi(d_n)$$

and

$$\int_D g_2\, dN(0, 1) = 2d_n \varphi(d_n), \qquad \int_D g_2^2\, dN(0, 1) = 2(1-d_n)^2 d_n \varphi(d_n).$$

Since

$$\frac{N(0, 1)(D)}{d_n} = \frac{\Phi(d_n) - \Phi(-d_n)}{d_n} = 2\frac{\Phi(d_n) - \Phi(0)}{d_n} \to 2\varphi(0),$$

we see that condition (4) is satisfied with $a_1 = 0$ and $a_2 = 1$, i.e.,

$$\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Moreover, we have

$$\delta_{n1} \sim \lambda^{-1/2} \left(1 - \frac{2\varphi(d_n)}{N(0, 1)(D)/d_n}\right)^{-1/2} \to \infty$$

and, as we have already known,

$$\delta_{n2} \sim \lambda^{-1/2} \left(\frac{2(1-d_n)^2 \varphi(d_n)}{N(0, 1)(D)/d_n}\right)^{-1/2} \to \lambda^{-1/2}.$$

(vi) (Regression model) Let us fix $v_0, \alpha_0, \beta_0 \in \boldsymbol{R}$ and suppose that $X = (V, W) \in \boldsymbol{R}^2$, where V has a continuous density f near v_0 with $f(v_0) > 0$. Assume that under $\vartheta = (\vartheta_1, \vartheta_2) \in \boldsymbol{R}^2$ and for v near v_0 the conditional distribution of W, given $V = v$, is $N(\alpha_0 + \vartheta_1 + (\beta_0 + \vartheta_2)(v - v_0), 1)$, i.e., the regression function

$$E_\vartheta(W \mid V = v) = \alpha_0 + \vartheta_1 + (\beta_0 + \vartheta_2)(v - v_0)$$

is linear in v. Precisely, P_ϑ has the Lebesgue density

$$f_\vartheta(v, w) := f(v)\, q_\vartheta(w \mid v)$$

for v near v_0 and $w \in \boldsymbol{R}$, where

$$q_\vartheta(w \mid v) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\tfrac{1}{2}\left(w - (\alpha_0 + \vartheta_1 + (\beta_0 + \vartheta_2)(v - v_0))\right)^2\right).$$

On $D = D(n) = [v_0 - d_n, v_0 + d_n] \times \mathbf{R}$, where $0 < d_n \xrightarrow{n\to\infty} 0$, expansion (1) holds if n is large with the tangents

$$g_1(v, w) = w - (\alpha_0 + \beta_0(v - v_0)), \quad g_2(v, w) = (v - v_0)(w - (\alpha_0 + \beta_0(v - v_0))).$$

We obtain $P_0(D) \sim 2f(v_0)d_n$ and

$$\int_D g_1\, dP_0 = 0, \quad \int_D g_1^2\, dP_0 = \int_{v_0 - d_n}^{v_0 + d_n} f(v)\, dv \sim 2f(v_0)d_n,$$

$$\int_D g_2\, dP_0 = 0, \quad \int_D g_2^2\, dP_0 = \int_{v_0 - d_n}^{v_0 + d_n} f(v)(v - v_0)^2 dv \sim \tfrac{2}{3} f(v_0)\, d_n^3.$$

Thus, condition (4) holds with $a_1 = a_2 = 0$, i.e.,

$$\Gamma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and we have $\delta_{n1} \sim \lambda^{-1/2}$, $\delta_{n2} \sim 3^{1/2}/(\lambda^{1/2} d_n)$.

In contrast to the Gaussian case, Lemma 2.1 does not imply the weak convergence of a Poisson sequence in general. This is due to the fact that Poisson experiments – in contrast to Gaussian experiments – are not determined by its binary subexperiments. But if the class of Poisson experiments is restricted to the class of Poisson experiments with independent increments, then they are completely described by its binary subexperiments (Strasser [20]). Unfortunately, in our present situation of truncation experiments Poisson limits with independent increments cannot be expected. For example, the Poisson limit for the normal case as well as for the Pareto case of Examples 2.4 (i) and (ii) has no independent increments. To see this crucial point we make use of the following characterization (Strasser [20], Theorem 17.2). The Poisson limit of $(E_n)_n$ has independent increments iff for every $r \in T$ and $\varepsilon > 0$

$$\lim_{n\to\infty} nP_{nr\delta_n}\left\{\left|\frac{dP_{ns\delta_n}}{dP_{nr\delta_n}} - \frac{dP_{nt\delta_n}}{dP_{nr\delta_n}}\right| \geq \varepsilon, \left|\frac{dP_{nu\delta_n}}{dP_{nr\delta_n}} - \frac{dP_{nv\delta_n}}{dP_{nr\delta_n}}\right| \geq \varepsilon\right\} = 0 \tag{12}$$

whenever $s < t < u < v$. We show that condition (12) is violated for the normal case. Recall that D is chosen such that $nP_0(D) = \lambda$. Let $0 < s < t$. For n large we have

$$P_{n0}\left\{\left|\frac{dP_{ns\delta_n}}{dP_{n0}} - \frac{dP_{nt\delta_n}}{dP_{n0}}\right| \geq \varepsilon\right\} = P_0\left\{x \in D\colon \left|\frac{dP_{s\delta_n}}{dP_0} - \frac{dP_{t\delta_n}}{dP_0}\right| \geq \varepsilon\right\}.$$

Now,

$$\begin{aligned} A_n &:= \left\{x \in D\colon \left|\frac{dP_{s\delta_n}}{dP_0} - \frac{dP_{t\delta_n}}{dP_0}\right| \geq \varepsilon\right\} \\ &= \left\{x \geq d_n\colon \left|\exp\left(s\delta_n x - \frac{s^2\delta_n^2}{2}\right) - \exp\left(t\delta_n x - \frac{t^2\delta_n^2}{2}\right)\right| \geq \varepsilon\sqrt{2\pi}\right\}. \end{aligned}$$

Since $\delta_n^{-1} \sim \lambda^{1/2} d_n$, we have

$$|\exp(s\delta_n x - s^2\delta_n^2/2) - \exp(t\delta_n x - t^2\delta_n^2/2)| \to \exp(\lambda^{-1/2} t) - \exp(\lambda^{-1/2} s)$$

for $x = d_n$. Hence $d_n \in A_n$ for n large and ε small, and a monotonicity argument shows that $A_n = [d_n, \infty) = D$. Consequently, condition (12) cannot be satisfied. Similar arguments show that also in the Pareto case condition (12) is not fulfilled.

In the following we establish weak convergence to a Poisson experiment and compute the intensity measures. We consider first binary subexperiments and the case $a_j \in \{-1, 1\}$ for $j = 1, \ldots, k$ (which covers Example 2.4 (i) and (ii)). Let $(P_{n0}^n, P_{nt\delta_n}^n)$ be a binary subexperiment of $E_{n,n}$ and let, as before, $nP_0(D) \to \lambda$. To study the limit behavior of the likelihood ratio it is convenient to use the point process approach. Denote by $N_{n,D}(\cdot) = \sum_{i=1}^n \varepsilon_{X_i}(\cdot \cap D)$ the truncated empirical point process.

It is easily seen that $(P_{n0}^n, P_{nt\delta_n}^n) \sim (\mathscr{L}(N_{n,D} | P_0^n), \mathscr{L}(N_{n,D} | P_{t\delta_n}^n))$ with

$$\frac{d\mathscr{L}(N_{n,D} | P_{t\delta_n}^n)}{d\mathscr{L}(N_{n,D} | P_0^n)}(\mu) = \left(\prod_{i=1}^{\mu(D)} \frac{f_{t\delta_n}(x_i)}{f_0(x_i)}\right)\left(\frac{1-P_{t\delta_n}(D)}{1-P_0(D)}\right)^{n-\mu(D)}$$

if $\mu = \sum_{i=1}^{\mu(D)} \varepsilon_{x_i}$ and $0 \leqslant \mu(D) \leqslant n$ (see, e.g., Reiss [17], Example 3.1.2 for the density formula). Now we make use of the representation

$$\mathscr{L}(N_{n,D} | P_0^n) = \mathscr{L}\left(\sum_{l=1}^{\tau_n} \varepsilon_{Y_l}\right),$$

where $\tau_n, Y_1, Y_2, \ldots$ (defined on some probability space $(\Omega, \mathscr{A}, P)$) are independent, τ_n is binomial $B(n, nP_0(D))$-distributed, and $Y_1, Y_2, \ldots$ are identically distributed according to $P_0(\cdot \cap D)/P_0(D)$; see, e.g., Reiss [17], Theorem 1.4.1. It remains, therefore, to study the asymptotic behavior of

$$L_{n,t} := \left(\prod_{l=1}^{\tau_n} \frac{f_{t\delta_n}(Y_l)}{f_0(Y_l)}\right)\left(\frac{1-P_{t\delta_n}(D)}{1-P_0(D)}\right)^{n-\tau_n}.$$

Denote by τ a Po(λ)-distributed random variable. Then

$$d_1(\mathscr{L}(\tau_n | P), \mathscr{L}(\tau | P)) \to 0 \quad \text{as } n \to \infty$$

by the Scheffé lemma. Since $\tau_n/n \xrightarrow{n\to\infty} 0$ and, by expansion (7),

$$nP_{t\delta_n}(D) = nP_0(D) + n\int_D \langle t\delta_n, g\rangle\, dP_0 + n\int_D \langle t\delta_n g/2\rangle^2\, dP_0 + o(1)$$

$$\xrightarrow{n\to\infty} \lambda + \lambda^{1/2} \sum_{j=1}^k \operatorname{sign}(a_j)\, t_j + \tfrac{1}{4} t' \Gamma t = \left(\lambda^{1/2} + \tfrac{1}{2} \sum_{j=1}^k \operatorname{sign}(a_j)\, t_j\right)^2$$

(recall that by assumption $a_j \in \{-1, 1\}$, $j = 1, \ldots, k$), we conclude that

$$\left(\frac{1-P_{t\delta_n}(D)}{1-P_0(D)}\right)^{n-\tau_n} \xrightarrow{n\to\infty} \exp\left(-\left(\lambda^{1/2} + \tfrac{1}{2} \sum_{j=1}^k \operatorname{sign}(a_j)\, t_j\right)^2 + \lambda\right).$$

Moreover,

$$\frac{f_{t\delta_n}(Y_l)}{f_0(Y_l)} = 1 + \langle t\delta_n, g(Y_l)\rangle + \langle t\delta_n, g(Y_l)/2\rangle^2 + o_P(1)$$

$$\xrightarrow{n\to\infty} 1 + \lambda^{-1/2} \sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j + \frac{1}{4\lambda}\Big(\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\Big)^2$$

$$= \frac{1}{\lambda}\left(\lambda^{1/2} + \frac{1}{2}\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\right)^2$$

in P-probability. Note that $\delta_{nj}\, g_j(Y_l) \to \operatorname{sign}(a_j)\,\lambda^{-1/2}$, which follows from

$$\int \big(\delta_{nj}\, g_j(Y_l) - \operatorname{sign}(a_j)\,\lambda^{-1/2}\big)^2\, dP$$

$$= \delta_{nj}^2 \frac{\int_D g_j^2\, dP_0}{P_0(D)} - 2\operatorname{sign}(a_j)\,\lambda^{-1/2}\,\delta_{nj} \frac{\int_D g_j\, dP_0}{P_0(D)} + \lambda^{-1}$$

$$\xrightarrow{n\to\infty} \lambda^{-1} - 2\lambda^{-1/2}\,\lambda^{-1/2} + \lambda^{-1} = 0.$$

Summarizing the above results we obtain

$$L_{n,t} \to \left(\frac{1}{\lambda}\big(\lambda^{1/2} + \tfrac{1}{2}\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\big)^2\right)^{\tau} \exp\Big(-\big(\lambda^{1/2} + \tfrac{1}{2}\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\big)^2 + \lambda\Big)$$

$$:= L_t := \frac{dQ_t}{dQ_0}$$

in P-probability, i.e., $(P_{n0}^n, P_{nt\delta_n}^n) \xrightarrow{n\to\infty} (Q_0, Q_t)$ with

$$Q_t := \operatorname{Po}\Big(\big(\lambda^{1/2} + \tfrac{1}{2}\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\big)^2\Big).$$

Since

$$d_2^2(\operatorname{Po}(\lambda_1), \operatorname{Po}(\lambda_2)) = 1 - \exp\big(-\tfrac{1}{2}(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2\big),$$

we obtain

$$d_2^2(Q_0, Q_t) = 1 - \exp\Big(-\tfrac{1}{8}\big(\sum_{j=1}^{k} \operatorname{sign}(a_j)\, t_j\big)^2\Big)$$

in accordance with relation (6). For Examples 2.4 (i) and (ii) we have $Q_t = \operatorname{Po}\big((\lambda^{1/2} + t/2)^2\big)$.

Now, let us determine the binary Poisson limit for the general case $a_j \in [-1, 1]$. Let Y be distributed according to $P_0(\cdot \cap D)/P_0(D)$. Since only a finite number of observations are available as the sample size tends to infinity, we have to claim a weak convergence condition. We assume that

$$\delta_n g(Y) = \big(nP_0(D)\big)^{-1/2} \left(\frac{\int_D g^2\, dP_0}{P_0(D)}\right)^{-1/2} g(Y) \to \lambda^{-1/2} Z \tag{13}$$

in distribution with $Z=(Z_1, \ldots, Z_k)$. (For $a_j \in \{-1, 1\}$ we have already known from the above calculations that $\delta_{nj} g_j(Y) \to \operatorname{sign}(a_j)\lambda^{-1/2}$, i.e., $\mathscr{L}(Z_j)=\varepsilon_{\operatorname{sign}(a_j)}$.) Similar calculations as before show that

$$L_{n,t} \to \left(\frac{1}{\lambda}\Big(\lambda+\lambda^{1/2}\sum_{j=1}^{k} t_j Z_j + \tfrac{1}{4}\sum_{1\leqslant i,j\leqslant k} t_i t_j Z_i Z_j\Big)\right)^{\tau} \exp\Big(-\lambda^{1/2}\sum_{j=1}^{k} a_j t_j - \tfrac{1}{4} t'\Gamma t\Big)$$
$$:= L_t := \frac{dQ_t}{dQ_0}.$$

Hence $(P_{n0}^n, P_{nt\delta_n}^n) \xrightarrow{n\to\infty} (Q_0, Q_t)$, where Q_t is the distribution of a Poisson process having the (finite) intensity measure

$$\nu_t(\cdot) = \int (\lambda+\lambda^{1/2}\langle t, z\rangle + \langle t, z\rangle^2/4)\, d\mathscr{L}(Z)(z) \tag{14}$$

(see the density formula for Poisson processes). Note that $E(Z_j)=a_j$ and $E(Z_i Z_j)=\gamma_{ij}$, which imply $\nu_t(R^k)=\lambda+\lambda^{1/2}\sum_{j=1}^{k} a_j t_j + t'\Gamma t/4$. The measure ν_t is absolutely continuous with respect to ν_0. Again, in accordance with relation (6), we have $d^2(Q_0, Q_t) = 1-\exp(-t'\Gamma t/8)$.

The convergence of the finite-dimensional marginal distributions of the loglikelihood process $(\log L_{n,t})_{t\in T}$ to the corresponding marginal distributions of $(\log L_t)_{t\in T}$ is now seen by the Cramér–Wold device and the continuous mapping theorem. Since Q_t is absolutely continuous with respect to Q_0, we have proved the following result:

2.5. THEOREM. *Assume that $nP_0(D)\to\lambda$ and that condition* (13) *is fulfilled. Then $(E_{n,n})_n$ converges weakly to a compound Poisson experiment*

$$\big(M(R^k, \mathscr{B}^k), \mathscr{M}(R^k, \mathscr{B}^k), \mathscr{L}(N_t)_{t\in T}\big),$$

where N_t are Poisson processes with intensity measures (14).

In the situation of Examples 2.4 (i) and (ii) we have $\mathscr{L}(Z)=\varepsilon_1$ and the intensity measures

$$\nu_t = (\lambda^{1/2}+t/2)^2\,\varepsilon_1$$

on the space $S=\{1\}$. In Example 2.4 (iii) with $a=0$ we have $\mathscr{L}(g(Y)) = (\varepsilon_{-1}+\varepsilon_1)/2$, which yields the intensity measures

$$\nu_t = \tfrac{1}{2}(\lambda^{1/2}+t/2)^2\,\varepsilon_1 + \tfrac{1}{2}(\lambda^{1/2}-t/2)^2\,\varepsilon_{-1}$$

on $S=\{-1, 1\}$. In Example 2.4 (iv) we have

$$\left(\frac{\int_D g_j^2\, dP_0}{P_0(D)}\right)^{-1/2} g(Y_i) \sim \left(\frac{(1+\beta)^2\beta}{2+\beta}\right)^{-1/2} \frac{1+\beta}{W} := Z,$$

where W follows the Pareto law H_β. Note that the distribution of $d_n^{-1}Y$ is again H_β, which is a well-known property of the Pareto distribution. Now, let us determine the intensity measure for the scale-location model of Example 2.4 (v). Applying l'Hôspital's rule and elementary calculations we obtain

$$\lim_{n\to\infty} \delta_{n1} d_n = \lambda^{-1/2} \quad \text{and} \quad \delta_{n1} g_1(Y) \to \lambda^{-1/2} U,$$

where U is uniformly distributed on $[-1, 1]$. Consequently, we have $\mathscr{L}(Z_1, Z_2) = \mathscr{L}(U) \otimes \varepsilon_1$. Finally, consider the regression model of Example 2.4 (vi), where Y is distributed according to

$$P_0\big((V, W) \in \cdot \cap D\big) / P_0\big((V, W) \in D\big) \quad \text{with } D = [v_0 - d_n, v_0 + d_n] \times \boldsymbol{R}.$$

Straightforward calculations show that

$$P_0\big((V, W) \in \cdot \cap D\big) / P_0\big((V, W) \in D\big) \to \varepsilon_{v_0} \otimes N(\alpha_0, 1)$$

and

$$P_0\left(\left(\frac{V-v_0}{d_n}, W\right) \in \cdot \cap D\right) \Big/ P_0\left(\left(\frac{V-v_0}{d_n}, W\right) \in D\right) \to \mathscr{L}(U) \otimes N(\alpha_0, 1)$$

weakly with U as before. Hence $\mathscr{L}\big(g_1(Y), g_2(Y)/d_n\big) \to \mathscr{L}(\tilde{Z}, U\tilde{Z})$, where $\tilde{Z}$ is $N(0, 1)$-distributed.

The preceding examples have already shown that the Poisson limit E of a truncation experiment cannot be scale invariant. Otherwise, the logarithm of the Hellinger transform

$$(t_1, \ldots, t_m) \to \log H(E_{\{t_1, \ldots, t_m\}})(z), \qquad (t_1, \ldots, t_m) \in [0, \infty)^m, \ m \in \boldsymbol{N},$$

has to be a homogeneous function (Strasser [19], Lemma (2.6)). Elementary calculations show that for $k = 1$ and $a = 1$ we have

$$\log H(E_{\{s,t\}})(z_s, z_t) = (\lambda^{1/2} + s/2)^{2z_s} (\lambda^{1/2} + t/2)^{2z_t} - z_s (\lambda^{1/2} + s/2)^2 - z_t (\lambda^{1/2} + t/2)^2,$$

which is not homogeneous.

In conclusion, we see that Poisson limits of truncation experiments are not translation invariant and not scale invariant, and they have no independent increments.

We briefly discuss an application in testing theory. Consider the case $k = 1$, $a = 1$, and let $nP_0(D) \to \lambda$. We want to test the hypothesis P^n_{n0} against the alternative $\{P^n_{nt\delta_n}: t > 0\}$ or, equivalently, $\mathscr{L}(N_{n,D} \mid P^n_0)$ against $\{\mathscr{L}(N_{n,D} \mid P^n_{t\delta_n}): t > 0\}$. The limit experiment is given by $\big(\mathrm{Po}\big((\lambda^{1/2} + t/2)^2\big)\big)_{t \geq 0}$, which is a one-dimensional exponential family with strictly monotone likelihood ratios with respect to the identity. Thus, the test

$$\varphi^*(x) = 1_{(c,\infty)}(x) + \gamma 1_{\{c\}}(x), \qquad x \in \boldsymbol{N} \cup \{0\},$$

with c and γ such that $\int \varphi^* \, d\mathrm{Po}(\lambda) = \mathrm{Po}(\lambda)\big((c, \infty)\big) + \gamma \mathrm{Po}(\{c\}) = \alpha$, is optimal for $t = 0$ against $t > 0$ at level α. Hence, based on $N_{n,D}$, and asymptotically optimal test at level α is given by $\varphi^*(N_{n,D})$.

Acknowledgement. I would like to thank Professor H. Strasser for valuable suggestion, which led to the present paper.

REFERENCES

[1] P. K. Andersen, Ø. Borgan, R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, Springer Ser. Statist., New York 1993.

[2] M. Falk, *On testing the extreme value index via the POT-method*, Ann. Statist. 23 (1995), pp. 2013–2035.

[3] M. Falk, *Local asymptotic normality of truncated empirical processes*, Ann. Statist. 26 (1998), pp. 692–718.

[4] M. Falk, J. Hüsler and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, DMV Seminar, Birkhäuser, Basel 1994.

[5] M. Falk and F. Liese, *LAN of thinned empirical processes with an application to fuzzy set density estimation*, Extremes 1 (1998), pp. 323–349.

[6] M. Falk and F. Marohn, *On asymptotically optimal tests for conditional distributions*, Ann. Statist. 21 (1993), pp. 45–60.

[7] M. Falk and R.-D. Reiss, *Statistical inference for conditional curves: Poisson process approach*, Ann. Statist. 20 (1992), pp. 779–796.

[8] A. Janssen, *Limits of translation invariant experiments*, J. Multivariate Anal. 20 (1986), pp. 129–142.

[9] A. Janssen and F. Marohn, *On statistical information of extreme order statistics, local extreme value alternatives, and Poisson point processes*, J. Multivariate Anal. 48 (1994), pp. 1–30.

[10] A. Janssen, H. Milbrodt and H. Strasser, *Infinitely Divisible Statistical Experiments*, Lecture Notes in Statist. 27, Springer, Berlin–Heidelberg 1985.

[11] L. LeCam, *Sur les contraintes imposées par les passages à la limite usuels en statistique*, Bull. Inst. Internat. Statist., Proc. 39th Session, Vienna, Vol. XLV, Book 4 (1973), pp. 169–177.

[12] L. LeCam, *Asymptotic Methods in Statistical Decision Theory*, Springer Ser. Statist., New York 1986.

[13] L. LeCam and G. L. Yang, *Asymptotics in Statistics. Some Basic Concepts*, Springer Ser. Statist., New York 1990.

[14] F. Marohn, *Local asymptotic normality of truncation models*, Statist. Decisions 17 (1999), pp. 237–253.

[15] H. Milbrodt and H. Strasser, *Limits of triangular arrays of experiments*, in: A. Janssen, H. Milbrodt and H. Strasser, *Infinitely Divisible Statistical Experiments*, Lecture Notes in Statist. 27, Springer, Berlin–Heidelberg 1985, pp. 14–54.

[16] R.-D. Reiss, *Approximate Distributions of Order Statistics. With Applications to Nonparametric Statistics*, Springer Ser. Statist., New York 1989.

[17] R.-D. Reiss, *A Course on Point Processes*, Springer Ser. Statist., New York 1993.

[18] S. I. Resnick, *Extreme Values, Regular Variation, and Point Processes*, Springer, New York 1987.

[19] H. Strasser, *Scale invariance of statistical experiments*, Probab. Math. Statist. 5 (1985), pp. 1–20.
[20] H. Strasser, *Statistical experiments with independent increments*, in: A. Janssen, H. Milbrodt and H. Strasser, *Infinitely Divisible Statistical Experiments*, Lecture Notes in Statist. 27, Springer, Berlin–Heidelberg 1985, pp. 124–153.
[21] H. Strasser, *Mathematical Theory of Statistics*, De Gruyter, Berlin–New York 1985.
[22] E. Torgersen, *Comparison of Statistical Experiments*, Cambridge University Press, Cambridge 1991.

Mathematisch-Geographische Fakultät
Katholische Universität Eichstätt
85071, Eichstätt
frank.marohn@ku-eichstaett.de

Received on 21.4.1998;
revised version on 28.8.2000