Thinning and the Law of Small Numbers

Peter Harremoës

Abstract: The idea of thinning is used to characterize the Poisson distribution and to give the Law of Small Numbers formulations via IID sequences. Both weak and strong convergence is proved.

MSC 2000: 60B10, 94A17

Key words: information divergence, poisson distribution, thinning

1 Introduction

Approximation by Poisson distributions is a well studied subject and the most complete presentation can be found in [1]. Later the connection between information theory has been established in [5] and [8]. For most values of the parameters the best bounds on total variation between a binomial distribution and a Poisson distribution with the same mean have been proved by ideas from information theory via Pinskers inequality [3], [4] and [6]. Here we shall see that the idea of thinning can be used to formulate Law of Small Numbers in a way that resembles the iid formulation of the Central Limit Theorem. Convergence in the Central Limit Theorem has been established in the strong sense that information divergence converges to zero [2] and [7]. The main result of this paper is a similar theorem with convergence to the Poisson distribution.

2 Thinning

Let $P$ denote a distribution on $\mathbb{N}_0$. The $\alpha$-thinning of $P$ is the distribution $T_\alpha (P)$ given by

$$T_\alpha (P) (k) = \sum_{l=k}^{\infty} P(l) \binom{l}{k} \alpha^k (1 - \alpha)^{l-k}.$$ 

If $X_1, X_2, X_3, \ldots$ are independent identically distributed Bernoulli random variables with success probability $\alpha$ and $Y$ has distribution $P$ independent of $X_1, X_2, \ldots$ then the distribution of

$$\sum_{n=1}^{Y} X_n$$

The author wants to thank Oliver Johnson and Yiannis Kontioannis for stimulating discussions and Christophe Vignat and EPFL, Lausanne for organizing the workshop where these ideas were born. The author also thanks the Danish Natural Science Research Council for financial support.
has distribution \( T_\alpha (P) \). Obviously the thinning of an independent sum of random variables is the convolution of thinnings. The factorial moments of the \( \alpha \)-thinning is easy to calculate

\[
E \left( \left( \sum_{n=1}^{Y} X_n \right)_{[k]} \right) = E \left( E \left( \left( \sum_{n=1}^{Y} X_n \right) \mid Y \right) \right) = E (\alpha^k Y_{[k]}) = \alpha^k E (Y_{[k]}).
\]

The thinning conserves the set of Bernoulli sums. To see this we just remark that the \( \alpha \)-thinning of a Bernoulli random variable with success probability \( p \) is a Bernoulli random variable with success probability \( \alpha p \).

**Example 2.1.** Thinning conserves the Poisson distributions.

\[
T_\alpha (\text{Po}(\lambda)) (k) = \sum_{l=k}^{\infty} \text{Po}(\lambda, l) \binom{l}{k} \alpha^k (1 - \alpha)^{l-k} = \sum_{l=k}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda} \alpha^k \binom{l}{k} (1 - \alpha)^{l-k} = \frac{e^{-\lambda}}{k!} \alpha^k \lambda^k \sum_{l=0}^{\infty} \frac{\lambda^{l-k}}{(l-k)!} (1 - \alpha)^{l-k} = \frac{e^{-\lambda}}{k!} \alpha^k \lambda^k \sum_{l=0}^{\infty} \left( \frac{\lambda (1 - \alpha)}{l - k} \right)^l (l - k)! = \frac{e^{-\alpha \lambda}}{k!} \alpha^k \lambda^k e^{\lambda (1 - \alpha)} = \frac{e^{-\alpha \lambda}}{k!} \alpha^k \lambda^k = \text{Po}(\alpha \lambda, k).
\]

For a distribution \( P \) on \( \mathbb{N}_0 \) the \( \alpha \)-thinning operation gives a distribution on \( \mathbb{N}_0 \). We note that \( \mathbb{N}_0 \subset \mathbb{N}_0/n \subset \mathbb{R} \) and that the thinning also can be made on the larger set of grid points \( \mathbb{N}_0/n \). If \( X \) has distribution \( P \) then the distribution of \( T_\alpha (P) \) converges weakly to the distribution of \( \alpha X \) according to the law of large numbers. Similarly, if a positive random variable \( X \) with continuous distribution is quantized according to a uniform quantization and then thinned then the thinned distribution approximately equals the distribution of \( \alpha X \). In this sense thinning is a discrete analog of scaling of a continuous random variable.

**Proposition 2.2.** In the class of ultra-log concave distributions the thinning operation \( T_\alpha, \alpha \in [0; 1] \) is injective.
Proof. An ultra log concave distribution is uniquely determined by its (factorial) moments because ultra log concave distributions satisfies a Cramér condition. The thinning operation just scales the moments so if we know the factorial moments of the thinned distribution we also know the factorial moments of the original distribution.

3 The Law of Small Numbers

By use of the thinning operation it is possible to formulate an iid version of the Law of Small Numbers (Poisson’s Law).

Theorem 3.1 (weak version). Let $P$ be a distribution on $\mathbb{N}_0$ with mean $\lambda$. Then $T_{1/n}(P^{*n})$ converges pointwise to $P_0(\lambda)$ for $n \to \infty$.

Proof. First we note that $T_{1/n}(P^{*n}) = (T_{1/n}(P))^{*n}$. For $\alpha = 1/n$ we have the inequalities

$$T_{1/n}(P)(0) = \sum_{l=0}^{\infty} P(l) (1 - \alpha)^l \geq (1 - \alpha)^\lambda$$

and

$$T_{1/n}(P)(1) = \sum_{l=1}^{\infty} P(l) l \alpha (1 - \alpha)^{l-1}$$

and $T_{1/n}(P)(j) \geq 0$ for $j \geq 2$. Thus

$$(T_{1/n}(P))^{*n}(j) \geq \binom{n}{j} \left( \sum_{l=1}^{\infty} P(l) l \alpha (1 - \alpha)^{l-1} \right)^j \left( (1 - \alpha)^\lambda \right)^{n-j}$$

$$= \frac{n[j]}{n^j \cdot j!} \left( \sum_{l=1}^{\infty} P(l) l \left( 1 - \frac{1}{n} \right)^{l-1} \right)^j \left( 1 - \frac{1}{n} \right)^{(n-j)\lambda}.$$

For a fixed value of $j$ and $n$ tending to infinity we have

$$\frac{n[j]}{n^j \cdot j!} \to \frac{1}{j!}$$

and

$$\left( 1 - \frac{1}{n} \right)^{(n-j)\lambda} \to e^{-\lambda}$$

and by Lebesgue’s Theorem on monoton convergence

$$\sum_{l=1}^{\infty} P(l) l \left( 1 - \frac{1}{n} \right)^{l-1} \to \lambda.$$
Thus
\[ \lim_{n \to \infty} \inf \left( T_{1/n} (P) \right)^* (j) \geq P_0 (\lambda, j). \]

The sequence \((T_{1/n} (P))^*\) is a sequence of discrete probability measures and \(P_0 (\lambda)\) is a discrete probability measure so for any value of \(j\) \((T_{1/n} (P))^* (j)\) converges to \(P_0 (\lambda, j)\) for \(n\) tending to infinity.

According to Sheffe’s Lemma pointwise convergence implies convergence in total variation. An even stronger kind of convergence can be obtained. By \(D (P \parallel Q)\) we shall denote the information divergence from \(P\) to \(Q\) defined by

\[ D (P \parallel Q) = \sum_j P (j) \log \frac{P (j)}{Q (j)} \]

where \(P\) and \(Q\) are discrete probability measures.

**Theorem 3.2 (strong version).** Let \(P\) be a distribution on \(N_0\) with mean \(\lambda\) and finite second moment. Then

\[ T_{1/n} (P^*) \xrightarrow{L} P_0 (\lambda) \text{ for } n \to \infty. \]

**Proof.** Let \(X_1, X_2, \ldots\) denote a sequence of iid random variables with distribution \(P\). Then

\[ D \left( T_{1/n} (P^*) \parallel P_0 (\lambda) \right) = D \left( \sum_{k=0}^{\infty} P^* (k) \ Bi (k, 1/n) \parallel P_0 (\lambda) \right) \]

\[ \leq \sum_{k=0}^{\infty} P^* (k) \cdot D (Bi (k, 1/n) \parallel P_0 (\lambda)). \]

Now, we use that the Poisson distributions belong to an exponential family and the elementary bound \(D (Bi (l, p) \parallel P_0 (lp)) \leq lp^2\) and get

\[ D \left( Bi \left( k, \frac{1}{n} \right) \parallel P_0 (\lambda) \right) = D \left( Bi \left( k, \frac{1}{n} \right) \parallel P_0 \left( \frac{k}{n} \right) \right) + D \left( P_0 \left( \frac{k}{n} \right) \parallel P_0 (\lambda) \right) \]

\[ \leq \frac{k}{n^2} + \sum_{j=0}^{\infty} P_0 \left( \frac{k}{n}, j \right) \log \frac{\frac{k}{n}^j \exp \left( -\frac{k}{n} \right)}{\frac{k}{n}^2 \exp (-\frac{k}{n})} \]

\[ = \frac{k}{n^2} + \lambda \left( \frac{k}{n \lambda} \log \frac{k}{n \lambda} + 1 - \frac{k}{n \lambda} \right) \]

\[ \leq \frac{k}{n^2} + \lambda \left( \frac{k}{n \lambda} - 1 \right)^2, \]
where we have used the elementary inequality $x \log x + 1 - x \leq x (x - 1) + 1 - x = (x - 1)^2$. Hence,

$$D \left( \frac{T_{1/n}}{n} (P^*n) \| Po (\lambda) \right) \leq \sum_{k=0}^{\infty} P^*n (k) \cdot \left( \frac{k}{n^2} + \lambda \left( \frac{k}{n\lambda} - 1 \right)^2 \right)$$

$$= \frac{n\lambda}{n^2} + \frac{1}{\lambda n^2} \sum_{k=0}^{\infty} P^*n (k) \cdot (k - n\lambda)^2$$

$$= \frac{\lambda}{n} + \frac{1}{\lambda n^2} \cdot n Var(X)$$

$$= \frac{\lambda + Var(X)}{n}.$$

The last term obviously converges to zero for $n$ tending to infinity. \hfill \square

For distributions with finite support we can give an alternative proof. The condition on the support is a primitive way to ensure that all sums converge and can easily be relaxed.

**Alternative proof.** For this proof to work we need the additional condition that $D (P\parallel Po (\lambda)) < \infty$. According to the data processing inequality

$$D (P_1 * P_2 * \ldots * P_n \| Po (\lambda/n) * Po (\lambda/n) * \ldots * Po (\lambda/n)) \leq \sum_{i=1}^{n} D (P_i \| Po (\lambda/n)).$$

Hence it is sufficient to show that $n \cdot D \left( \frac{T_{1/n}}{P^*} (P) \| Po (\lambda/n) \right)$ converges to zero for $n$ tending to infinity. We put $\alpha = 1/n$ and have to prove that

$$\frac{D (T_{\alpha} (P) \| Po (\alpha\lambda)) - 0}{\alpha - 0}$$

converges to zero for $\alpha$ tending to zero. By the mean value theorem it is sufficient to show that

$$\frac{\partial}{\partial \alpha} D (T_{\alpha} (P) \| Po (\alpha\lambda))$$

converges to zero for $\alpha$ tending to zero. Now

$$D (T_{\alpha} (P) \| Po (\alpha\lambda))$$

$$= \sum_{k=0}^{\infty} \sum_{l=k}^{\infty} P (l) \left( \frac{l}{k} \right) \alpha^k (1 - \alpha)^{l-k} \log \frac{\sum_{l=k}^{\infty} P (l) \left( \frac{l}{k} \right) \alpha^k (1 - \alpha)^{l-k}}{\alpha^k \lambda^k \exp (-\alpha \lambda)}$$

$$= \alpha \lambda + \sum_{k=0}^{\infty} \left( \sum_{l=k}^{\infty} P (l) \left( \frac{l}{k} \right) \alpha^k (1 - \alpha)^{l-k} \log \left( \sum_{l=k}^{\infty} P (l) \frac{l}{k} \alpha^k (1 - \alpha)^{l-k} \right) \right).$$
The derivative is
\[
\frac{\partial}{\partial \alpha} D(T_\alpha(P) \| P_\lambda) = \lambda + \sum_{k=1}^{\infty} \left( \sum_{l=k}^{\infty} P(l) \binom{l}{k} \alpha^{k-1} (1 - \alpha)^{l-k} \right) \log \left( \sum_{l=k}^{\infty} P(l) \frac{l[l]}{\lambda^l} (1 - \alpha)^{l-k} \right) - \sum_{k=0}^{\infty} \left( \sum_{l=k+1}^{\infty} P(l) \binom{l}{k} \alpha^k (l-k) (1 - \alpha)^{l-k-1} \right) \log \left( \sum_{l=k}^{\infty} P(l) \frac{l[l]}{\lambda^l} (1 - \alpha)^{l-k} \right)
\]
For \(\alpha\) tending to zero this expression tends to
\[
\lambda + \left( \sum_{l=1}^{\infty} P(l) \binom{l}{1} \cdot 1 \right) \log \left( \sum_{l=1}^{\infty} P(l) \frac{l[l]}{\lambda^l} \right) - \left( \sum_{l=0}^{\infty} P(l) \cdot l \right) \log \left( \sum_{l=0}^{\infty} P(l) \right) = 0.
\]

4 Characterizations of the Poisson distribution

We shall finish by some results characterizing the Poisson distribution.

Proposition 4.1. Let \(X\) denote a discrete random variable and let \(Y\) denote an independent Poisson random variable. Assume that
\[
T_\alpha(X) + T_\beta(Y) \sim X
\]
where \(\alpha, \beta \in ]0; 1[\). Then \(X\) has a Poisson distribution.
Proof. First we observe that $\alpha E(X) + \beta E(Y) = E(X)$ so that $E(Y) > 0$ and

$$\beta = (1 - \alpha) \frac{E(X)}{E(Y)}.$$ 

Without loss of generality we may assume that $E(X) = E(Y)$ and $\beta = 1 - \alpha$. Put $P_\alpha(X) = T_\alpha(X) + T_{1-\alpha}(Y)$. Then $P_\alpha(X) \sim X$ and $P_\alpha^n(X) \sim X$. Now,

$$X \sim P_\alpha^n(X) = P_\alpha^n(X) = T_\alpha^n(X) + T_{1-\alpha^n}(Y).$$

The right hand side converges to a Poisson distribution so that $X$ must be Poisson distributed.

Proposition 4.2. If $P$ is a ultra-log concave distribution such that for all $\alpha \in ]0; 1[$ there exists a ultra-log concave distribution $Q_\alpha$ such that $P = T_\alpha(Q_\alpha)$ then $P$ is a Poisson distribution.

Proof. Assume that $P$ is a ultra-log concave distribution such that for all $\alpha \in ]0; 1[$ there exists a ultra-log concave distribution $Q_\alpha$ such that $P = T_\alpha(Q)$. Let $\lambda$ and $V$ denote the first two factorial moments of $P$. Then

$$D(T_{1/n}(Q_{1/n}) \parallel P_\alpha(\lambda)) \leq \frac{\lambda + \frac{V - \lambda^2 + \lambda}{\lambda}}{n} \leq \frac{\lambda + 1}{n},$$

which is proved as in the first proof of Theorem 3.2. Now $T_{1/n}(Q_{1/n}) = P$ for all $n$ so we get $D(P \parallel P_\alpha(\lambda)) = 0$ and $P = P_\alpha(\lambda)$.

References


Peter Harremoës: University of Copenhagen, Department of Mathematics, Universitetsparken 5, Copenhagen O, DK-2100, Denmark, *moes@math.ku.dk*