Translated Poisson approximation using exchangeable pair couplings

Röllin, A

DOI: https://doi.org/10.1214/105051607000000258

Originally published at:
Röllin, A (2007). Translated Poisson approximation using exchangeable pair couplings. Annals of Applied Probability, 17(5-6):1596-1614.
DOI: https://doi.org/10.1214/105051607000000258

TEXed on February 27, 2008

TRANSLATED POISSON APPROXIMATION USING EXCHANGEABLE PAIR COUPLINGS

By Adrian Röllin*

It is shown that the method of exchangeable pairs introduced by Stein (1986) for normal approximation can effectively be used for translated Poisson approximation. Introducing an additional smoothness condition, one can obtain approximation results in total variation and also in a local limit metric. The result is applied in particular to the anti-voter model on finite graphs as analysed by Rinott and Rotar (1997), obtaining the same rate of convergence, but now for a stronger metric.

1. Introduction

Let W be a random variable with $\mathbb{E}W = \mu$ and $\operatorname{Var} W = \sigma^2$. Stein (1986) introduced a method (which is commonly called the *exchangeable pair approach*) to approximate $W_c := (W-\mu)/\sigma$ by the standard normal distribution; Rinott and Rotar (1997) then generalised the result and successfully applied it to weighted U-statistics and the antivoter model. Their results imply convergence in the Kolmogorov and even in some stronger metrics; however, they do not to provide approximations in the total variation metric or prove local limit like results.

We will consider such results in this paper in the special case, in which the sum W is integer valued, the most common situation being the one where W is a sum of random indicators. As the total variation between W and the normal distribution will always be 1, we will instead use a translated Poisson distribution, matching the first two moments of W as well as possible. Note that we will consider the approximation of the unstandardised variable W and assume that $\sigma^2 \to \infty$, so there is actually no convergence taking place. We will also consider a metric from which local limit approximations can be obtained.

Recall the setting of Stein (1986) and Rinott and Rotar (1997). A pair of random variables (W, W') is called exchangeable, if $\mathscr{L}(W, W') = \mathscr{L}(W', W)$. Assume now that there is a positive number $\lambda < 1$ and a random variable

*Partially supported by Schweizerischer Nationalfondsprojekt 20-107935/1

R such that

$$\mathbb{E}^W(W' - \mu) = (1 - \lambda)(W - \mu) + R, \tag{1.1}$$

holds, where $\mathbb{E}^W$ denotes the conditional expectation with respect to W. Of course, one can always find R to satisfy (1.1), so R must be thought of as being small for the approximation to be successful. Note that (1.1) implies $\mathbb{E}R = 0$.

If the pair (W, W') can be chosen such that condition (1.1) holds and $\mathbb{E}^W(W' - W)^2$ does not fluctuate too much, convergence of W_c to the standard normal distribution will follow in the Kolmogorov metric. As the behaviour of the difference $W' - W$ is mainly responsible for the quality of the approximation, it is an obvious starting point to introduce a smoothness condition, to make sure that the local perturbations of W are not too strong. Recall that Rinott and Rotar (1997) propose to choose W and W' as two successive steps of a reversible Markov chain with stationary distribution $\mathscr{L}(W)$. Then, condition (1.1) states that a particle on $\mathbb{Z}$ obeying the transition rules of such a Markov chain is forced to have a linear drift to the centre. Now $\mathbb{E}^{W=k}(W' - W)^2$ is the average of the squared jump size of the Markov chain if the particle is in k, so that for a good normal approximation, the average jump size of the particle has to be about the same wherever it is. It is now clear that, under these conditions, the particle may still behave irregularly on a local scale, for instance, the particle could still make only jumps of size two and thus stay on the odd or even integers, such that an approximation with a distribution on $\mathbb{Z}$ with span 1 will not be successful in total variation.

Thus, in addition to (1.1), we assume further that

$$W' - W \in \{-1, 0, +1\}, \tag{1.2}$$

and we will see that this seems to be an appropriate condition. Note that under condition (1.2) the corresponding Markov chain does not need to be reversible for (W, W') to be a exchangeable pair; see Lemma 1.1 of Rinott and Rotar (1997).

Condition (1.2) is in sharp contrast to other approaches using Stein's method for the translated Poisson distribution such as Čekanavičius and Vaĭtkus (2001), Röllin (2005) or Barbour and Lindvall (to appear), where an embedded sum of independent random variables within W is used for an explicit smoothing argument; in contrast, the smoothing effect of (1.2) will enter only implicitly into the proof of the main result. Chatterjee et al. (2005) use the same condition to obtain Poisson approximation results with the exchangeable pair approach.

As we are restricted to the integers, we cannot arbitrarily shift a Poisson distribution with a given variance to fit the mean, so some care is needed here. We say that an integer valued random variable Y has a *translated Poisson distribution* with parameters μ and σ^2 and write

$$\mathscr{L}(Y) = \mathrm{TP}(\mu, \sigma^2)$$

if $\mathscr{L}(Y - \mu + \sigma^2 + \gamma) = \mathrm{Po}(\sigma^2 + \gamma)$ where $\gamma = \langle \mu - \sigma^2 \rangle$ and $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x; in particular $\mathrm{TP}(\sigma^2, \sigma^2) = \mathrm{Po}(\sigma^2)$. So, approximating W with $\mathrm{TP}(\mu, \sigma^2)$, we can fit the mean exactly, but note that for the variance we have $\sigma^2 \leqslant \operatorname{Var} Y = \sigma^2 + \gamma \leqslant \sigma^2 + 1$. This will, however, cause no further problems, as the order of error of this mismatch is $\mathrm{O}(\sigma^{-2})$.

Throughout the paper, we shall be concerned with two metrics for probability distributions, the total variation metric d_{TV} and the local limit metric d_{loc}, where, for two probability distributions P and Q given by the point probabilities $\{p_k, k \in \mathbb{Z}\}$ and $\{q_k, k \in \mathbb{Z}\}$ respectively,

$$\begin{aligned} d_{\mathrm{TV}}\bigl(P, Q\bigr) &:= \sup_{A \subset \mathbb{Z}} \bigl|P(A) - Q(A)\bigr| = \frac{1}{2} \sum_{k \in \mathbb{Z}} |p_k - q_k|, \\ d_{\mathrm{loc}}\bigl(P, Q\bigr) &:= \sup_{k \in \mathbb{Z}} |p_k - q_k|. \end{aligned}$$

2. Main results

Theorem 2.1. *Assume that* (W, W') *is an exchangeable pair with values on the integers and which satisfies* (1.1) *and* (1.2). *Then, with* $S = S(W) = \mathbb{P}[W' = W + 1 | W]$ *and* $q_{\max} = \max_{k \in \mathbb{Z}} \mathbb{P}[W = k]$,

$$d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant \frac{\sqrt{\operatorname{Var} S}}{\lambda \sigma^2} + \frac{2\sqrt{\operatorname{Var} R}}{\lambda \sigma} + \frac{2}{\sigma^2}, \tag{2.1}$$

$$\begin{aligned} d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leqslant{}& \frac{2\sqrt{q_{\max} \operatorname{Var} S}}{\lambda \sigma^2} + \frac{2 q_{\max} \sqrt{\operatorname{Var} R}}{\lambda \sigma} \\ &+ \frac{\sqrt{\operatorname{Var} R}}{\lambda \sigma^2} + \frac{2}{\sigma^2}. \end{aligned} \tag{2.2}$$

Before proving Theorem 2.1, we give a short introduction into Stein's method for distributional approximation. The starting point for translated Poisson approximation is the Stein-Chen method for the Poisson distribution as presented in detail by Barbour et al. (1992).

Let W be an integer valued random variable with expectation μ and variance $\sigma^2 > 0$, and let $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \langle \mu - \sigma^2 \rangle$ where $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that, if $Y \sim \mathrm{TP}(\mu, \sigma^2)$, then $Y - s \sim \mathrm{Po}(\sigma^2 + \gamma)$. Let $\mathcal{A}g(j) = (\sigma^2 + \gamma)g(j+1) - jg(j)$ be the usual Stein operator for the Poisson distribution with mean $\sigma^2 + \gamma$, and for $A \subset \mathbb{Z}_+ := \{0, 1, 2, \dots\}$ let $g_A : \mathbb{Z} \to \mathbb{R}$ be the solution of

$$i)\;\; g(j) = 0 \text{ for all } j \leqslant 0, \tag{2.3}$$

$$ii)\;\; \mathcal{A}g(j) = I[j \in A] - \mathrm{Po}(\sigma^2 + \gamma)\{A\} \text{ for all } j \geqslant 0. \tag{2.4}$$

We can thus bound the total variation distance as

$$\begin{aligned} d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) &= d_{\mathrm{TV}}\bigl(\mathscr{L}(W - s), \mathrm{Po}(\sigma^2 + \gamma)\bigr) \\ &= \sup_{B \subset \mathbb{Z}} \bigl|\mathbb{E} I[W - s \in B] - \mathrm{Po}(\sigma^2 + \gamma)\{B\}\bigr| \\ &\leqslant \sup_{A \subset \mathbb{Z}_+} \bigl|\mathbb{E}\mathcal{A}g_A(W - s)\bigr| + \mathbb{P}[W - s < 0], \end{aligned} \tag{2.5}$$

The last terms in (2.5) can be bounded using Chebyshev's inequality as

$$\mathbb{P}[W - s < 0] = \mathbb{P}\bigl[W - \mu < -(\sigma^2 + \gamma)\bigr] \leqslant \mathbb{P}\bigl[|W - \mu| > \sigma^2 + \gamma\bigr] \leqslant \frac{1}{\sigma^2}. \tag{2.6}$$

From (Barbour et al., 1992, Lemma 1.1.1) we have the well-known bounds on the supremum norm of g_A,

$$\|g_A\| \leqslant (\sigma^2 + \gamma)^{-1/2} \leqslant \sigma^{-1}, \quad \|\Delta g_A\| \leqslant (\sigma^2 + \gamma)^{-1} \leqslant \sigma^{-2}, \tag{2.7}$$

where $\Delta g_A(j) := g_A(j+1) - g_A(j)$. If $A = \{k\}$ for some $k \in \mathbb{Z}$ we even have

$$\|g_{\{k\}}\| \leqslant \sigma^{-2}. \tag{2.8}$$

For the proof of the results in the d_{loc} metric, we will also need the following non-standard but simple result.

Lemma 2.2. *Let g_i be the solution of (2.3)–(2.4) for $A = \{i\}$. Then*

$$\sum_k |\Delta g_i(k)| \leqslant 2\sigma^{-2}, \qquad \sum_k \bigl(\Delta g_i(k)\bigr)^2 \leqslant 4\sigma^{-4}. \tag{2.9}$$

Proof. Recall from (Barbour et al., 1992, Proof of Lemma 1.1.1) that $g_i(k)$ is negative and decreasing in $0 \leqslant k \leqslant i$ and positive and decreasing in $k > i$ with the only positive jump in i satisfying

$$|\Delta g_i(i)| \leqslant (\sigma^2 + \gamma)^{-1} \leqslant \sigma^{-2}.$$

From this, it is easy to see that the first bound of (2.9) holds and the second bound is then immediate. □

With $\tilde{g}_A(j) := g_A(j-s)$ we can rewrite the Stein operator $\mathcal{A}$ as

$$\begin{aligned}\mathcal{A}g_A(W-s) &= (\sigma^2+\gamma)g_A(W-s+1) - (W-s)g_A(W-s)\\ &= \sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W) + \gamma\Delta\tilde{g}_A(W).\end{aligned} \tag{2.10}$$

The bounds on $\tilde{g}_A$ are of course the same as on g_A in (2.7)–(2.9). Thus, the last term in (2.10) is easily bounded by

$$\bigl|\mathbb{E}\{\gamma\Delta\tilde{g}_A(W)\}\bigr| \leqslant \gamma\sigma^{-2} \leqslant \sigma^{-2}. \tag{2.11}$$

Inserting (2.10) into (2.5) and invoking the bounds (2.6) and (2.11) we obtain

$$\begin{aligned}&d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr)\\ &\qquad\leqslant \sup_{A\subset\mathbb{Z}_+}\bigl|\mathbb{E}\bigl\{\sigma^2\Delta\tilde{g}_A(W) - (W-\mu)\tilde{g}_A(W)\bigr\}\bigr| + 2\sigma^{-2};\end{aligned} \tag{2.12}$$

the same estimate holds for d_{loc} but with the supremum taken only over the sets $A=\{i\}$ for $i\in\mathbb{Z}_+$.

Proof of Theorem 2.1. We only have to bound the supremum in (2.12). Stein (1986) showed, that, if F satisfies $F(w,w') = -F(w',w)$ for all w and w', exchangeability implies $\mathbb{E}F(W,W') = 0$. Define the random variable $D := W'-W$ and the function $F(w,w') := (w'-w)\bigl(g(w')+g(w)\bigr)$ for $g\equiv\tilde{g}_A$ and note that, from (1.1), $\mathbb{E}^W D = -\lambda(W-\mu)+R$. This yields

$$\begin{aligned}0 &= \mathbb{E}F(W,W') = \mathbb{E}\bigl\{D\bigl(2g(W)+g(W')-g(W)\bigr)\bigr\}\\ &= -2\lambda\mathbb{E}\bigl\{(W-\mu)g(W)\bigr\} + 2\mathbb{E}\bigl\{Rg(W)\bigr\} + \mathbb{E}\bigl\{D(g(W')-g(W))\bigr\}.\end{aligned} \tag{2.13}$$

Note now that, for $D_i := I[D=i]$, $i\in\{-1,+1\}$, we can write

$$D(g(W')-g(W)) = D_{+1}\Delta g(W) + D_{-1}\Delta g(W-1),$$

and further, using exchangeability,

$$\begin{aligned}\mathbb{E}\bigl\{D_{-1}\Delta g(W-1)\bigr\} &= \mathbb{E}\bigl\{I[W'-W=-1]\Delta g(W-1)\bigr\}\\ &= \mathbb{E}\bigl\{I[W-W'=1]\Delta g(W')\bigr\}\\ &= \mathbb{E}\bigl\{D_{+1}\Delta g(W)\bigr\},\end{aligned} \tag{2.14}$$

thus,

$$\mathbb{E}\bigl\{D(g(W')-g(W))\bigr\} = 2\mathbb{E}\bigl\{D_{+1}\Delta g(W)\bigr\}. \tag{2.15}$$

Together with (2.13) this yields

$$\mathbb{E}\bigl\{(W-\mu)g(W)\bigr\} = \frac{\mathbb{E}\bigl\{D_{+1}\Delta g(W)\bigr\}}{\lambda} + \frac{\mathbb{E}\bigl\{Rg(W)\bigr\}}{\lambda}. \tag{2.16}$$

Note now that, by exchangeability, $\mathbb{E}D_{+1} = \mathbb{E}D_{-1}$ and hence that

$$\begin{aligned}\mathbb{E}D_{+1} &= \tfrac{1}{2}\mathbb{E}(W'-W)^2 \\ &= \tfrac{1}{2}\bigl[\mathbb{E}(W'-\mu)^2 - 2\mathbb{E}\bigl\{(W'-\mu)(W-\mu)\bigr\} + \mathbb{E}(W-\mu)^2\bigr] \\ &= \lambda\sigma^2 + \mathbb{E}\bigl\{(W-\mu)R\bigr\} =: \lambda\sigma^2 + a,\end{aligned} \tag{2.17}$$

from (1.1); then use (2.16) to express the expectation in (2.12) as

$$\begin{aligned}&\mathbb{E}\bigl\{(W-\mu)g(W) - \sigma^2\Delta g(W)\bigr\} \\ &= \mathbb{E}\bigl\{(W-\mu)g(W) - (\sigma^2+\lambda^{-1}a)\Delta g(W)\bigr\} + \lambda^{-1}a\mathbb{E}\Delta g(W) \\ &= \mathbb{E}\bigl\{\bigl(D_{+1}\lambda^{-1} - \sigma^2 - \lambda^{-1}a\bigr)\Delta g(W)\bigr\} + \lambda^{-1}\mathbb{E}\bigl\{Rg(W)\bigr\} + \lambda^{-1}a\mathbb{E}\Delta g(W) \\ &=: B_1 + B_2 + B_3.\end{aligned}$$

Now, recall that $S = \mathbb{E}^W D_{+1}$, and thus, with the estimates

$$|B_1| \leqslant \|\Delta g\|\lambda^{-1}\mathbb{E}\bigl|S - \mathbb{E}S\bigr| \leqslant \|\Delta g\|\lambda^{-1}\sqrt{\operatorname{Var} S}, \tag{2.18}$$

$$|B_2| \leqslant \|g\|\lambda^{-1}\mathbb{E}|R| \leqslant \|g\|\lambda^{-1}\sqrt{\operatorname{Var} R}, \tag{2.19}$$

$$|B_3| \leqslant \|\Delta g\|\lambda^{-1}\mathbb{E}\bigl|(W-\mu)R\bigr| \leqslant \|\Delta g\|\lambda^{-1}\sigma\sqrt{\operatorname{Var} R}, \tag{2.20}$$

and the bounds (2.7), (2.1) follows.

To prove (2.2), we also use (2.12), but now we take the supremum only over all subsets $A = \{i\}$ for $i \in \mathbb{Z}$. Writing $g \equiv \tilde{g}_{\{i\}}$ and following the proof as for d_{TV} above, the bound on (2.19) remains and recalling (2.8), the third term in (2.2) follows. We thus need only refine the bounds on B_1 and B_3. Note that by the Cauchy-Schwartz inequality

$$|B_1| \leqslant \lambda^{-1}\sqrt{\operatorname{Var} S}\sqrt{\mathbb{E}(\Delta g(W))^2}.$$

Using Lemma 2.2, the latter expectation can be bounded by

$$\mathbb{E}(\Delta g(W))^2 = \sum_k (\Delta g(k))^2\mathbb{P}[W=k] \leqslant q_{\max}\sum_k(\Delta g(k))^2 \leqslant 4\sigma^{-4}q_{\max} \tag{2.21}$$

which implies the first term in (2.2). Using a similar argument on B_3, we obtain

$$|B_3| \leqslant \lambda^{-1}\sigma\sqrt{\operatorname{Var} R}\,q_{\max}\sum_k|\Delta g(k)|,$$

which, together with Lemma 2.2, yields the second term in (2.2). □

Remark 2.1. Theorem 2.1 is a direct analogue of Theorem 1.2 of Rinott and Rotar (1997). However, the first term in (2.1) is slightly different in quality from Theorem 1.2 of Rinott and Rotar (1997), as can be seen by comparing the result of their Theorem 1.3 for the anti-voter model with estimate (3.7). The additional $2/\sigma^2$ in (2.1) and (2.2) occurs because the Possion distribution cannot take negative values, and because the translation must be integer valued. Depending on the problem at hand, this error term can be further reduced or even be omitted; see estimates (2.6) and (2.11).

Remark 2.2. In some of the applications, instead of $S(W) = \mathbb{P}[W' = W+1|W]$, we will estimate the variance of a random variable $S^* = S^*(X) := \mathbb{P}[W' = W+1|X]$ for some random variable X such that the corresponding σ-algebras satisfy $\sigma(W) \subset \sigma(X)$ and then use the basic fact that $\operatorname{Var} S \leqslant \operatorname{Var} S^*$.

Remark 2.3. As becomes clear from equation (2.16), there is a close connection between the random variable $S = S(W)$ and the so called w-functions as examined for example by Cacoullos et al. (1994) and Cacoullos and Papathanasiou (1997) for the normal and the Poisson distributions. In the case of the standard normal distribution, their problem is as follows: for a given random variable X with $\mathbb{E}X = 0$ and $\operatorname{Var} X = 1$, find a function $w : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{E}\big\{Xf(X)\big\} = \mathbb{E}\big\{w(X)f'(X)\big\} \tag{2.22}$$

holds for a large set of functions f. For the translated Poisson distribution, the corresponding equation is

$$\mathbb{E}\big\{(W-\mu)f(W)\big\} = \mathbb{E}\big\{w(W)\Delta f(W)\big\}, \tag{2.23}$$

and it is indeed satisfied for any W as in Theorem 2.1 if $R = 0$ and if we choose $w(W) = S(W)/\lambda$. Unfortunately, it is often difficult to give an explicit expression for S as a function of W. However, if we allow $w(W)$ in (2.23) to be replaced by a more general random variable, we see from (2.16) that we can use the random variable $S^*(X)/\lambda$ from Remark 2.2 instead. For instance, for the anti-voter model as discussed in the next section, $S^*(X)$ has the nice and explicit representation (3.10).

Using (2.1) with the following corollary one easily obtains a bound for $q_{\max}$.

Corollary 2.3. *For any $\mathbb{Z}$-valued random variable W,*

$$q_{\max} \leqslant d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big) + \frac{1}{2.3\sigma}$$

Proof. Just apply Proposition A.2.7 of Barbour et al. (1992). □

Remark 2.4. Estimate (2.2) in combination with Corollary 2.3 is enough to obtain a local limit theorem in the applications of the next section. Although it can be easily calculated in many circumstances, the example of the Poisson-binomial distribution shows that the bound on d_{loc} need not be optimal; estimate (2.2) is of order $\mathrm{O}(n^{-3/4})$ in the special case of the binomial distribution, in contrast to the true order $\mathrm{O}(n^{-1})$. Under additional assumptions on S however, the bound (2.2) can be used to derive the better d_{loc}-bound, given in the following theorem. This bound is used in the examples 3.2 and 3.3 to obtain the correct order $\mathrm{O}(n^{-1})$ of approximation.

Theorem 2.4. *Assume the conditions of Theorem 2.1; assume in addition that S, as a function of W, can be extended on $\mathbb{R}$ such that it is Lipschitz continuous. Then,*

$$d_{\mathrm{loc}}\big(\mathscr{L}(W),\mathrm{TP}(\mu,\sigma^2)\big) \leqslant \frac{2L_S\big(\sigma^{-3}\mathbb{E}|W-\mu|^3 \vee (d\sigma^{3/2}+1)\big)}{\lambda\sigma^2} + \frac{2L_S q_{\max}}{\lambda\sigma} + \frac{2q_{\max}\sqrt{\mathrm{Var}\,R}}{\lambda\sigma} + \frac{\sqrt{\mathrm{Var}\,R}}{\lambda\sigma^2} + \frac{2}{\sigma^2}. \tag{2.24}$$

where d is the d_{loc}-bound (2.2) *and L_S is the Lipschitz constant of S.*

To prove Theorem 2.4 we need the following Lemma.

Lemma 2.5. *For any μ and σ^2, the bound*

$$\mathrm{TP}(\mu,\sigma^2)\{k\}\,|k-\mu| \leqslant 1$$

holds for all $k \in \mathbb{Z}$.

Proof. Recall from (2.4) that, if $Z \sim \mathrm{TP}(\mu,\sigma^2)$,

$$\mathbb{E}\big\{(Z-\mu)g(Z) - (\sigma^2+\gamma)\Delta g(Z)\big\} = 0 \tag{2.25}$$

for any g for which the expectations exist. With $\pi_k = \mathrm{TP}(\mu,\sigma^2)\{k\}$ and putting $g(\cdot) = I[\cdot = k]$ we obtain from (2.25) the bound

$$\begin{aligned}\pi_k|k-\mu| &\leqslant (\sigma^2+\gamma)|\pi_{k-1}-\pi_k| \\ &\leqslant (\sigma^2+\gamma)\,d_{\mathrm{loc}}\big(\mathrm{TP}(\mu+1,\sigma^2),\mathrm{TP}(\mu,\sigma^2)\big) \\ &= (\sigma^2+\gamma)\,d_{\mathrm{loc}}\big(\mathscr{L}(Y+1),\mathscr{L}(Y)\big).\end{aligned}$$

where $Y \sim \mathrm{Po}(\sigma^2+\gamma)$. The later d_{loc}-distance can easily be bounded using Stein's method for the Poisson distribution, that is, (2.4) in connection with the bound (2.8), which yields $d_{\mathrm{loc}}\big(\mathscr{L}(Y+1),\mathscr{L}(Y)\big) \leqslant (\sigma^2+\gamma)^{-1}$. □

Proof of Theorem 2.4. Follow the proof of Theorem 2.1 for the d_{loc} metric up to the bounds on the B_i. The bounds on $|\mathrm{B}_2|$ and $|B_3|$ remain. Recalling that S is a function defined on all $\mathbb{R}$, write now B_1 as

$$\begin{aligned} B_1 &= \lambda^{-1}\mathbb{E}\big\{(S(W) - \mathbb{E}S(W))\Delta g(W)\big\} \\ &= \lambda^{-1}\mathbb{E}\big\{(S(W) - S(\mu))\Delta g(W)\big\} + \lambda^{-1}\mathbb{E}\big\{(S(\mu) - S(W))\big\}\mathbb{E}\Delta g(W) \\ &=: B_{1,1} + B_{1,2}. \end{aligned}$$

Exploiting Lipschitz continuity of S and recalling (2.9) we obtain with $q_k = \mathbb{P}[W = k]$

$$|B_{1,2}| \leqslant \lambda^{-1}\sigma L_S \sum_k q_k|\Delta g(k)| \leqslant \frac{2L_S q_{\max}}{\lambda\sigma}$$

which is the second term in (2.24). For $B_{1,1}$ we have

$$\begin{aligned} |B_{1,1}| &\leqslant \lambda^{-1}\sum_k q_k|S(k) - S(\mu)||\Delta g(k)| \\ &\leqslant \lambda^{-1}L_S \sum_k q_k|k - \mu||\Delta g(k)|. \end{aligned} \tag{2.26}$$

We now bound $q_k|k-\mu|$. Assume first that $|k-\mu| > \sigma^{3/2}$; then, by Chebyshev's inequality,

$$q_k \leqslant \mathbb{P}[W \geqslant k] \leqslant \frac{\mathbb{E}|W-\mu|^3}{|k-\mu|^3}$$

and thus

$$q_k|k-\mu| \leqslant \sigma^{-3}\mathbb{E}|W-\mu|^3.$$

On the other hand, if $|k-\mu| \leqslant \sigma^{3/2}$, observe that

$$q_k \leqslant d + \mathrm{TP}(\mu, \sigma^2)\{k\}$$

and hence, using Lemma 2.5,

$$q_k|k-\mu| \leqslant d\sigma^{3/2} + 1.$$

Thus, (2.26) can be further bounded to

$$|B_{1,1}| \leqslant \lambda^{-1}L_S\big(\sigma^{-3}\mathbb{E}|W-\mu|^3 \vee (d\sigma^{3/2}+1)\big)\sum_k |\Delta g(k)|$$

and applying again (2.9), this yields the first term in (2.24). □

The following lemma can be used to estimate the second and third moments of W.

Lemma 2.6. *Under the assumptions of Theorem 2.1 and with* $A = \{w : \mathbb{P}[W = w] > 0\}$ *and* $a := \mathbb{E}\{R(W-\mu)\}$,

$$\lambda^{-1}\Big(\inf_{w\in A} S(w) - a\Big) \leqslant \sigma^2 \leqslant \lambda^{-1}\Big(\sup_{w\in A} S(w) - a\Big),$$
$$\mathbb{E}|W-\mu|^3 \leqslant \lambda^{-1}\big(8q_{\max} + 1 + \sigma + \mathbb{E}\{|R|(W-\mu)^2\}\big).$$

Proof. The estimates for the variance are immediate from equality (2.17) and the bounds

$$\inf_{w\in A} S(w) \leqslant \mathbb{E}S(W) \leqslant \sup_{w\in A} S(w).$$

Note now that from equation (2.16),

$$\mathbb{E}\big\{(W-\mu)g(W)\big\} = \lambda^{-1}\mathbb{E}\big\{S(W)\Delta g(W)\big\} + \lambda^{-1}\mathbb{E}\big\{Rg(W)\big\}$$

for all functions g, for which the expectations exist. With $K_\mu(w) = I[w > \mu] - I[w \leqslant \mu]$ and $g(w) = K_\mu(w)(w-\mu)^2$ we thus obtain

$$\begin{aligned}\mathbb{E}|W-\mu|^3 &= \lambda^{-1}\mathbb{E}\big\{S(W)\big[(W-\mu)^2 + 2(W-\mu) + 1\big]\Delta K_\mu(W)\big\}\\ &\quad + \lambda^{-1}\mathbb{E}\big\{S(W)(2(W-\mu)+1)K_\mu(W)\big\}\\ &\quad + \lambda^{-1}\mathbb{E}\big\{R(W-\mu)^2K_\mu(W)\big\} =: B_1' + B_2' + B_3'\end{aligned}$$

Note now, that $|K(w)| = 1$ and

$$\Delta K_\mu(w) = \begin{cases} 2 & \text{if } w = \lfloor\mu\rfloor,\\ 0 & \text{else,}\end{cases}$$

and thus, as $|\lfloor\mu\rfloor - \mu| \leqslant 1$ and $|S(w)| \leqslant 1$,

$$|B_1'| \leqslant 8\lambda^{-1}q_{\max},$$
$$|B_2'| \leqslant \lambda^{-1} + \lambda^{-1}\sigma.$$

The bound on B_3' is immediate. □

3. Applications

In this section we illustrate our results using some examples in in which $W = \sum_{i=1}^n J_i$ for a sequence $J = (J_1, J_2, \dots, J_n)$ of random indicators. Barbour and Xia (1999) and Röllin (2005) considered cases where the J_i have a local dependence structure; in contrast, the examples in this paper (with the exception of the first, standard example) exhibit global dependence.

For latter use we recall the following easy to prove fact.

Lemma 3.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant L_f. Then, for any random variable X,*

$$\operatorname{Var} f(X) \leqslant L_f^2 \operatorname{Var} X.$$

3.1. Poisson-Binomial distribution.

Theorem 3.2. *Let $J = (J_1, \dots, J_n)$ be a sequence of independent random indicators with $\mathbb{E}J_i = p_i$. Then, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 1/n$; we have*

$$S^*(J) := \mathbb{P}[W' = W + 1|J] = \frac{1}{n}\sum_{i=1}^{n}(1 - J_i)p_i, \tag{3.1}$$

$$\operatorname{Var} S(W) \leqslant \operatorname{Var} S^*(J) = n^{-2}\sum_{i=1}^{n} p_i^3(1 - p_i), \tag{3.2}$$

$$q_{\max} \leqslant 0.47\sigma^{-1}, \tag{3.3}$$

where (3.3) *holds if $\sigma^2 = \sum_{i=1}^{n} p_i(1 - p_i) \geqslant 4$. Thus, if $\sigma^2 \asymp n$,*

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) &= \mathrm{O}(n^{-1/2}) \\ d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu, \sigma^2)\big) &= \mathrm{O}(n^{-3/4}). \end{aligned}$$

Remark 3.1. As already mentioned in Remark 2.4, the order of the above d_{loc}-bound is not optimal. However, Röllin (2006) obtains the correct order $\mathrm{O}(n^{-1})$, using a different variant of Stein's method where an explicit smoothing argument is involved.

Remark 3.2. Corollary 2.1 of Čekanavičius and Vaĭtkus (2001) seems to be better in constant than (2.1) in the above example. For instance, for the Binomial distribution, we have

$$d_{\mathrm{TV}}\big(\mathrm{Bi}(n,p), \mathrm{TP}(\mu, \sigma^2)\big) \leqslant C\sqrt{\frac{p}{n(1-p)}} + \frac{2}{np(1-p)},$$

where Čekanavičius and Vaĭtkus (2001) obtain $C = 0.93$ and (2.1) yields $C = 1$.

Proof. We use the standard argument from Stein (1986). Let K be uniformly distributed on $\{1, \dots, n\}$ and let J^* be an independent copy of J. Then it is easy to see that $W' := W - J_K + J_K^*$ will satisfy (1.1) with $R = 0$ and

$\lambda = 1/n$. Further

$$\begin{aligned} S^*(J) &= \mathbb{E}^J I[W' - W = 1] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}^J I[J_i = 0, J_i^* = 1] \\ &= \frac{1}{n}\sum_{i=1}^n (1 - J_i)\mathbb{E}^J J_i^* = \frac{1}{n}\sum_{i=1}^n (1 - J_i) p_i \end{aligned}$$

Thus, $\operatorname{Var} S^* = n^{-2}\sum_{i=1}^n p_i^3(1-p_i)$, which, following Remark 2.2, proves (3.2). From (Čekanavičius and Vaĭtkus, 2001, estimate (2.22)) we obtain

$$q_{\max} \leqslant 0.45\sigma^{-1}(16/15)^{1/2},$$

if $\sigma^2 \geqslant 4$ which proves (3.3). □

3.2. Hypergeometric distribution. Assume that we have N urns and m balls, and that we distribute the balls uniformly into the N urns, in such a way that there is at most one ball per urn. Clearly, the number of balls W in the first n urns has the hypergeometric distribution $\operatorname{Hyp}(m, n, N)$, for which

$$\sigma^2 = \operatorname{Var} W = \frac{nm(N-n)(N-m)}{(N-1)N^2}.$$

Theorem 3.3. *If W has the hypergeometric distribution, Theorem 2.1 can be applied with $R = 0$ and $\lambda = \frac{N}{m(N-m+1)}$; we have*

$$\begin{aligned} S(W) &= \frac{mn - (m+n)W + W^2}{m(N-m+1)}, \\ \operatorname{Var} S(W) &\leqslant \frac{nm(m+n)^2(N-n)(N-m)}{m^2(N-m+1)^2(N-1)N^2}; \end{aligned} \tag{3.4}$$

thus, if $N = N(n) \asymp n$ and $m = m(n) \asymp n$,

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W), \operatorname{TP}(\mu,\sigma^2)\big) &= \mathrm{O}(n^{-1/2}), \\ d_{\mathrm{loc}}\big(\mathscr{L}(W), \operatorname{TP}(\mu,\sigma^2)\big) &= \mathrm{O}(n^{-1}). \end{aligned}$$

Proof. Consider the following construction. Pick uniformly an urn with a ball, and put this ball into any empty urn (including the urn, from which the ball was picked). Denote now by W' the number of balls in the first n urns. Exchangeability of (W, W') is easy to see and condition (1.2) is clearly satisfied. Now, $W' - W = 1$ is the event that a ball is picked from one of the urns $n+1, \ldots, N$ and put into one of the empty urns $1, \ldots, n$, thus

$$S(W) = \mathbb{P}[W' = W + 1 | W] = \frac{m - W}{m} \times \frac{n - W}{N - m + 1}$$

and conversely

$$\mathbb{P}[W' = W - 1|W] = \frac{W}{m} \times \frac{N - n - m + W}{N - m + 1}$$

thus

$$\begin{aligned}&\mathbb{E}^W(W' - W)\\&\quad = \mathbb{E}^W I[W' - W = 1] - \mathbb{E}^W I[W' - W = -1] = \frac{mn - NW}{m(N - m + 1)},\end{aligned}$$

and (1.1) is satisfied with $R = 0$ and $\lambda = \frac{N}{m(N-m+1)}$.

Note now that S, as a function of W, is Lipschitz continuous with constant $L_S = \frac{m+n}{m(N-m+1)}$; thus applying Lemma 3.1 we have

$$\operatorname{Var} S \leqslant \frac{(m+n)\sigma^2}{m^2(N - m + 1)^2}. \qquad \square$$

This is enough to prove the d_{TV}-order and, together with Corollary 2.3, the order $\mathrm{O}(n^{-3/4})$ for the d_{loc}-metric. Now, noting that Lemma 2.6 yields $\mathbb{E}|W - \mu|^3 = \mathrm{O}(n^{3/2})$, we obtain from Theorem 2.4 the desired order $\mathrm{O}(n^{-1})$ for the d_{loc}-metric.

3.3. A parity problem. Let $J_1, \dots, J_n$ be a sequence of independent Be(1/2)-distributed random indicators. Define

$$J_{n+1} := \begin{cases} 1 & \text{if } \sum_{i=1}^n J_i \text{ is odd}, \\ 0 & \text{else}, \end{cases}$$

and $V := \sum_{i=1}^{n+1} J_i$, so V is simply obtained by 'rounding' a $\mathrm{Bi}(n, 1/2)$-distributed random variable to the next even integer. An approximation of V by a translated Poisson distribution will clearly not succeed; however, we may try with $W := \frac{1}{2}V$.

Regard now the following exchangeable pair coupling. Pick two random indices K, L uniformly on $\{1, \dots, n+1\}$ so that almost surely $K \neq L$, and define

$$V' = V + 2 - 2J_K - 2J_L; \tag{3.5}$$

that is, take two summands of V at random, and replace each of them by its complement.

Lemma 3.4. *The pair (V, V') defined as above is an exchangeable pair and $(W, W') := (\frac{1}{2}V, \frac{1}{2}V')$ satisfies (1.1) and (1.2) with $\lambda = 2/(n + 1)$.*

Proof. It is enough to regard the situation on $M = \{0,1\}^n$ because the values $J_1, \dots, J_n$ uniquely determine the random variable J_{n+1}. Note first that construction (3.5) gives rise to a discrete time Markov chain on M, with jumps from $j \in M$ to $j' \in M$, if j' differs from j in exactly one or two coordinates (j' differing in exactly one coordinate corresponds to K or L being equal to $n+1$). Now, as the jump from j to j' happens with the same probability as from j' to j and all the states are connected, it is easy to see that the such defined Markov chain is irreducible and reversible and that the equilibrium distribution assigns equal probability to any $j \in M$, which corresponds to n independent $\mathrm{Be}(1/2)$ random variables. Thus, exchangeability is proved.

Note now that

$$\begin{aligned}\mathbb{E}^J(V'-V) &= 2 - \frac{2}{n(n+1)}\sum_{k=1}^{n+1}\sum_{\substack{l=1\\ l\neq k}}^{n+1}(J_k+J_l)\\ &= 2 - \frac{2}{n(n+1)}2nV = 2 - \frac{4V}{n+1},\end{aligned}$$

thus, we can take $\lambda = 2/(n+1)$. □

Theorem 3.5. *For W defined as above, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 2/(n+1)$; if $n \geqslant 2$, we have $\sigma^2 = (n+1)/16$ and*

$$S(W) = \frac{n(n+1) - (4n-2)W + 4W^2}{n(n+1)}, \quad \operatorname{Var} S(W) \leqslant \frac{(4n-2)^2(n+1)}{16n^2(n+1)^2};$$

thus, as $n \to \infty$,

$$\begin{aligned}d_{\mathrm{TV}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) &= \mathrm{O}(n^{-1/2}),\\ d_{\mathrm{loc}}\bigl(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\bigr) &= \mathrm{O}(n^{-1}).\end{aligned}$$

Proof. Note first that if $n \geqslant 2$ the J_i are uncorrelated and thus

$$\sigma^2 = \mathrm{Var}(V)/4 = (n+1)/16.$$

Now,

$$\begin{aligned}\mathbb{E}^J I[W'-W=1] &= \frac{1}{n(n+1)}\sum_{k=1}^{n+1}\sum_{\substack{l=1\\ l\neq k}}^{n+1}(1-J_k)(1-J_l)\\ &= \frac{n(n+1) - (4n-2)W + 4W^2}{n(n+1)} =: S(W).\end{aligned}$$

Observe that S, as a function of W, is Lipschitz continuous with $L_S = \frac{4n-2}{n(n+1)}$; thus, applying Lemma 3.1,

$$\operatorname{Var} S(W) \leqslant \frac{(4n-2)^2\sigma^2}{n^2(n+1)^2}.$$

This is enough to prove the d_{TV}-order and, together with Corollary 2.3, the order $O(n^{-3/4})$ for d_{loc}. Now, noting that Lemma 2.6 yields $\mathbb{E}|W-\mu| = \mathrm{O}(n^{3/2})$, we obtain from Theorem 2.4 the desired order $\mathrm{O}(n^{-1})$ for the d_{loc}-metric.

□

3.4. Anti-voter model on finite graphs. We closely follow the setup of Rinott and Rotar (1997); see also references therein and Huber and Reinert (2004). Let G be a n-vertex r-regular graph, which is neither bipartite nor an n-cycle. At each vertex i we assume that there is a 'voter' attached, having an opinion $J_i^{(t)}$ which can take the values 0 or 1 in every time point $t \in \mathbb{N}$. Define a Markov chain by the following transition rule. Choose uniformly a random vertex, say i; then, out of the neighbourhood $\mathcal{N}_i$ of i, choose uniformly a random vertex, say j, and let $J_i^{(t+1)}$ be the opposite of $J_j^{(t)}$ and leave the other voters untouched. Assume now that the Markov chain is in its equilibrium and put $W = \sum_{i=1}^n J_i := \sum_{i=1}^n J_i^{(0)}$.

Theorem 3.6. *For the anti-voter model as described above, Theorem 2.1 can be applied with $R = 0$ and $\lambda = 2/n$; we have*

$$S^*(J) := \mathbb{E}^J I[W' - W = 1] = \frac{3rn - 4rW + Q}{4rn}, \tag{3.6}$$

$$\operatorname{Var} S(W) \leqslant \operatorname{Var} S^*(J) \leqslant \frac{16r^2\sigma^2 + \operatorname{Var} Q}{16r^2n^2}, \tag{3.7}$$

where

$$Q = \sum_{i=1}^n \sum_{j\in\mathcal{N}_i} (2J_i - 1)(2J_j - 1);$$

hence, as $n \to \infty$,

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big) = \mathrm{O}\left(\frac{\sqrt{\operatorname{Var} Q}}{r\sigma^2} + \frac{1}{\sigma}\right),$$

$$d_{\mathrm{loc}}\big(\mathscr{L}(W), \mathrm{TP}(\mu,\sigma^2)\big) = \mathrm{O}\left(\frac{(\operatorname{Var} Q)^{3/4}}{r^{3/2}\sigma^3} + \frac{1}{\sigma^{3/2}}\right).$$

Remark 3.3. Note that the bounds for d_{TV} in Theorem 3.6 is very similar to the bound for the weaker Kolmogorov metric d_{K} given in Theorem 1.3 of Rinott and Rotar (1997); they obtain

$$d_{\mathrm{K}}\bigl(\mathscr{L}(W_c),\mathcal{N}(0,1)\bigr) = \mathrm{O}\biggl(\frac{\sqrt{\operatorname{Var} Q}}{r\sigma^2} + \frac{n}{\sigma^3}\biggr), \tag{3.8}$$

where $W_c = (W-\mu)/\sigma$.

Example 3.7. Consider the sequence K_n of complete graphs of size n. Rinott and Rotar (1997) show that $\sigma^2 \asymp n$ and $\operatorname{Var} Q = \mathrm{O}(n^3)$. Thus, from Theorem 2.1, the d_{TV}-distance is of the order $\mathrm{O}(n^{-1/2})$ and the d_{loc}-distance of order $\mathrm{O}(n^{-3/4})$ which proves the LLT. Now,

$$S^*(J) = \frac{n(n-1)-(2n-1)W+W^2}{n(n-1)} = S(W), \tag{3.9}$$

and we can thus take $L_S = \frac{2}{n-1}$. From Lemma 2.6 we obtain $\mathbb{E}|W-\mu|^3 = \mathrm{O}(n^{3/2})$ and therefore Theorem 2.4 yields the order $\mathrm{O}(n^{-1})$ for d_{loc}. Note that the estimates on L_S is obtained only because of the explicit representation (3.9); they are difficult to obtain in general. For further examples of graphs see Rinott and Rotar (1997).

Proof of Theorem 3.6. Define $W' := \sum_{i=1}^n J_i^{(1)}$, and note that (W,W') is an exchangeable pair, satisfying (1.1) and (1.2) with the choices $\lambda = 2/n$ and $R = 0$ (for more details see Rinott and Rotar (1997)). Now, let K be the random index of the vertex that was resampled in the transition from W to W'. As $W' = W - J_K + J_K^{(1)}$,

$$\begin{aligned} S^*(J) &= \mathbb{E}^J I[W'-W=1] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}^J\bigl\{I[J_i=0, J_i^{(1)}=1] \bigm| K=i\bigr\} \\ &= \frac{1}{n}\sum_{i=1}^n (1-J_i)\mathbb{E}^J\bigl\{J_i^{(1)} \bigm| K=i\bigr\} = \frac{1}{n}\sum_{i=1}^n (1-J_i)\Bigl(1-\frac{1}{r}\sum_{j\in\mathcal{N}_i} J_j\Bigr) \end{aligned} \tag{3.10}$$

With $X_i = 2J_i - 1$ and $\tilde{W} = \sum_{i=1}^n X_i$, (3.10) becomes

$$\begin{aligned} S^*(J) &= \frac{1}{4rn}\sum_{i=1}^n (1-X_i)\Big(r - \sum_{j\in\mathcal{N}_i} X_j\Big) \\ &= \frac{1}{4rn}\Big(rn - \sum_{i=1}^n \sum_{j\in\mathcal{N}_i} X_j - r\sum_{i=1}^n X_i + Q\Big) \\ &= \frac{rn - 2r\tilde{W} + Q}{4rn}. \end{aligned}$$

The variance of S^* is thus

$$\operatorname{Var} S^*(J) = \frac{\operatorname{Var}(2r\tilde{W}) + \operatorname{Var} Q - 4r\operatorname{Cov}(\tilde{W}, Q)}{16r^2n^2} = \frac{16r^2\sigma^2 + \operatorname{Var} Q}{16r^2n^2}, \tag{3.11}$$

because $\mathbb{E}\{X_iX_jX_k\} = 0$ for any choice of i, j and k, due to the symmetry of the anti-voter model, and hence $\mathbb{E}\{\tilde{W}Q\} = 0$. □

Acknowledgements

I thank Andrew Barbour for many helpful suggestions and comments.

References

A. D. Barbour and T. Lindvall (to appear). Translated Poisson approximation for Markov chains. *J. Theor. Probab.*.

A. D. Barbour, L. Holst, and S. Janson (1992). *Poisson approximation.* The Clarendon Press Oxford University Press, New York.

A. D. Barbour and A. Xia (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3**, 131–150.

T. Cacoullos and V. Papathanasiou (1997). Characterizations of distributions by generalizations of variance bounds and simple proofs of the CLT. *J. Statist. Plann. Inference* **63**, 157–171.

T. Cacoullos, V. Papathanasiou, and S. A. Utev (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

V. Čekanavičius and P. Vaĭtkus (2001). Centered Poisson approximation by the Stein method. *Lithuanian Math. J.* **41**, 319–329.

S. Chatterjee, P. Diaconis, and E. Meckes (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2**, 64–106.

M. Huber and G. Reinert (2004). The stationary distribution in the antivoter model: exact sampling and approximations. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, 79–93.

Y. Rinott and V. Rotar (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7**, 1080–1105.

A. Röllin (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11**, 1115–1128.

A. Röllin (2006, submitted manuscript). Symmetric binomial approximation for sums of locally dependent random variables.

C. Stein (1986). *Approximate computation of expectations*, IMS Lecture Notes. Institute of Mathematical Statistics, Hayward.

Universität Zürich, Winterthurerstrasse 190, 8052 Zürich, Switzerland
e-mail: adrian.roellin@math.unizh.ch