On moderate deviations in Poisson approximation

Aihua Xia (夏爱华)

School of Mathematics and Statistics

The University of Melbourne, VIC 3010

11 July, 2019

(Based on a joint work with Qingwei Liu)

Why?

- Distributional approximation pays little attention to the tail probabilities.
- In statistical inference, the tail probabilities matter!
- The error estimates of distributional approximation are useless because the tail probabilities are often significantly smaller than the error estimates.

What's the moderate deviation?

Petrov (1975), p. 228: let X_i, $1 \le i \le n$, be independent and identically distributed (i.i.d.) random variables with $\mathbb{E}(X_1) = 0$ and $\mathrm{Var}(X_1) = 1$, if for some $t_0 > 0$,

$$\mathbb{E}e^{t_0|X_1|} \le c_0 < \infty,$$

then there exist positive constants c_1 and c_2 depending on c_0 and t_0 such that

$$\frac{\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \ge z\right)}{1 - \Phi(z)} = 1 + O(1)\frac{1+z^3}{\sqrt{n}}, \quad 0 \le z \le c_1 n^{1/6},$$

where $\Phi(z)$ is the distribution function of the standard normal, $|O(1)| \le c_2$.

- c_1 and c_2?
- The range of values of n?

Why do we need Poisson?

- Since the pioneering work Chen (1975), it has been shown that, for the counts of rare events, Poisson distribution and its "relatives" provide a better approximation in terms of stronger metrics.
- BUT for the tail probabilities, we don't need the stronger metrics, can't we use normal?

Example

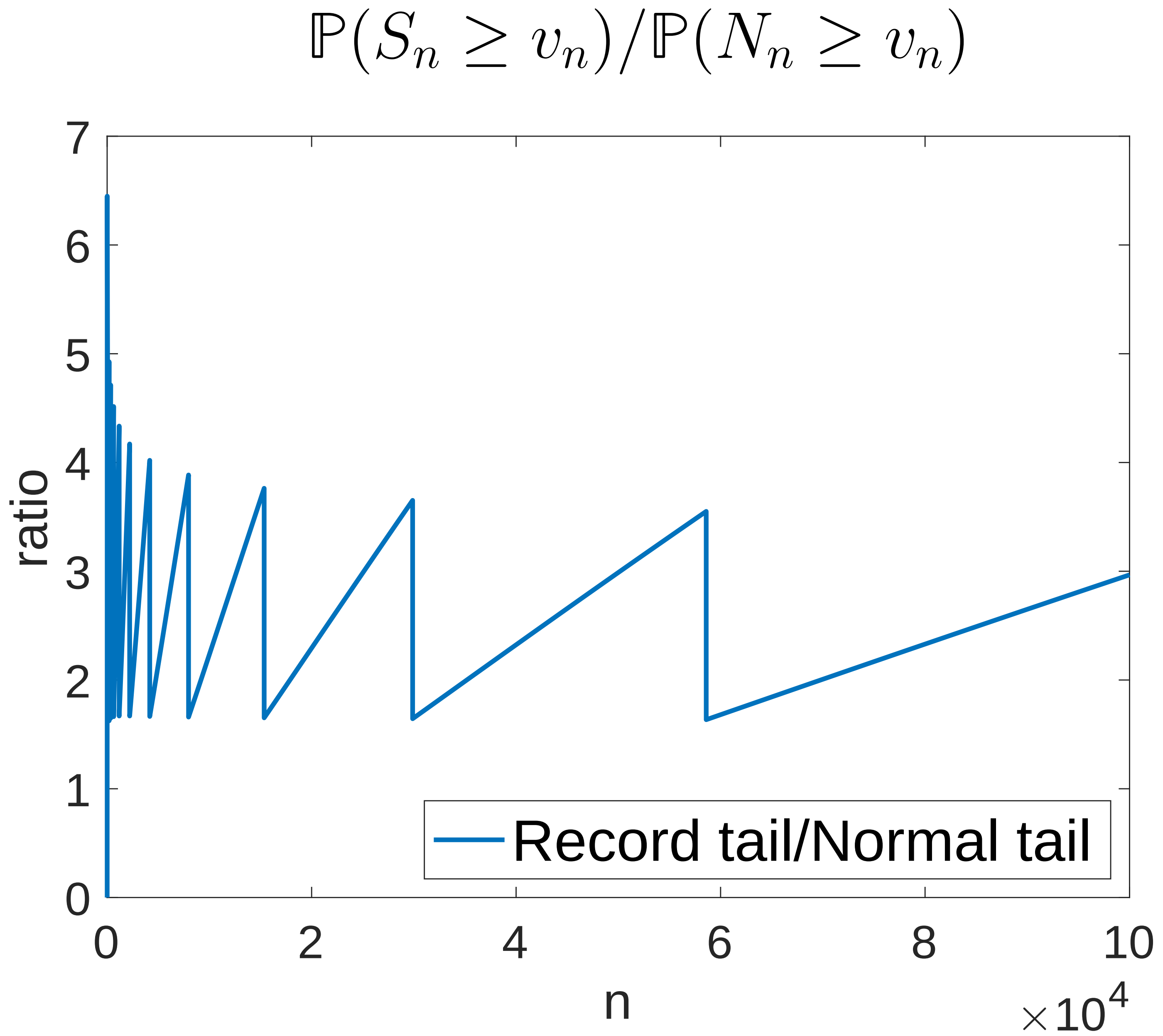
- Let $\{X_i : 1 \leq i \leq n\}$ be iid with a continuous cumulative distribution function, we are interested in the distribution of records in $\{X_i\}$.
- X_1 is always a record: ignore it.
- For $2 \leq i \leq n$, X_i is a record if $X_i > \max_{1 \leq j \leq i-1} X_j$.
- $I_i := \mathbf{1}[X_i > \max_{1 \leq j \leq i-1} X_j]$.
- $S_n := \sum_{i=2}^n I_i$.

Approximations of S_n

- $\mathbb{E} I_i = 1/i$ and $\{I_i : \ 2 \le i \le n\}$ are independent.
- $\lambda_n := \mathbb{E} S_n = \sum_{i=2}^{n} \frac{1}{i}$; $\sigma_n^2 = \mathrm{Var}(S_n) = \sum_{i=2}^{n} \frac{1}{i}\left(1 - \frac{1}{i}\right)$.
- $\lambda_n - \sigma_n \in (0, 1)$.
- Under the Kolmogorov distance, the error of
  - normal approximation is $O(\log^{-1/2} n)$,
  - Poisson approximation is $O(\log^{-1} n)$.

How about the tail probabilities?

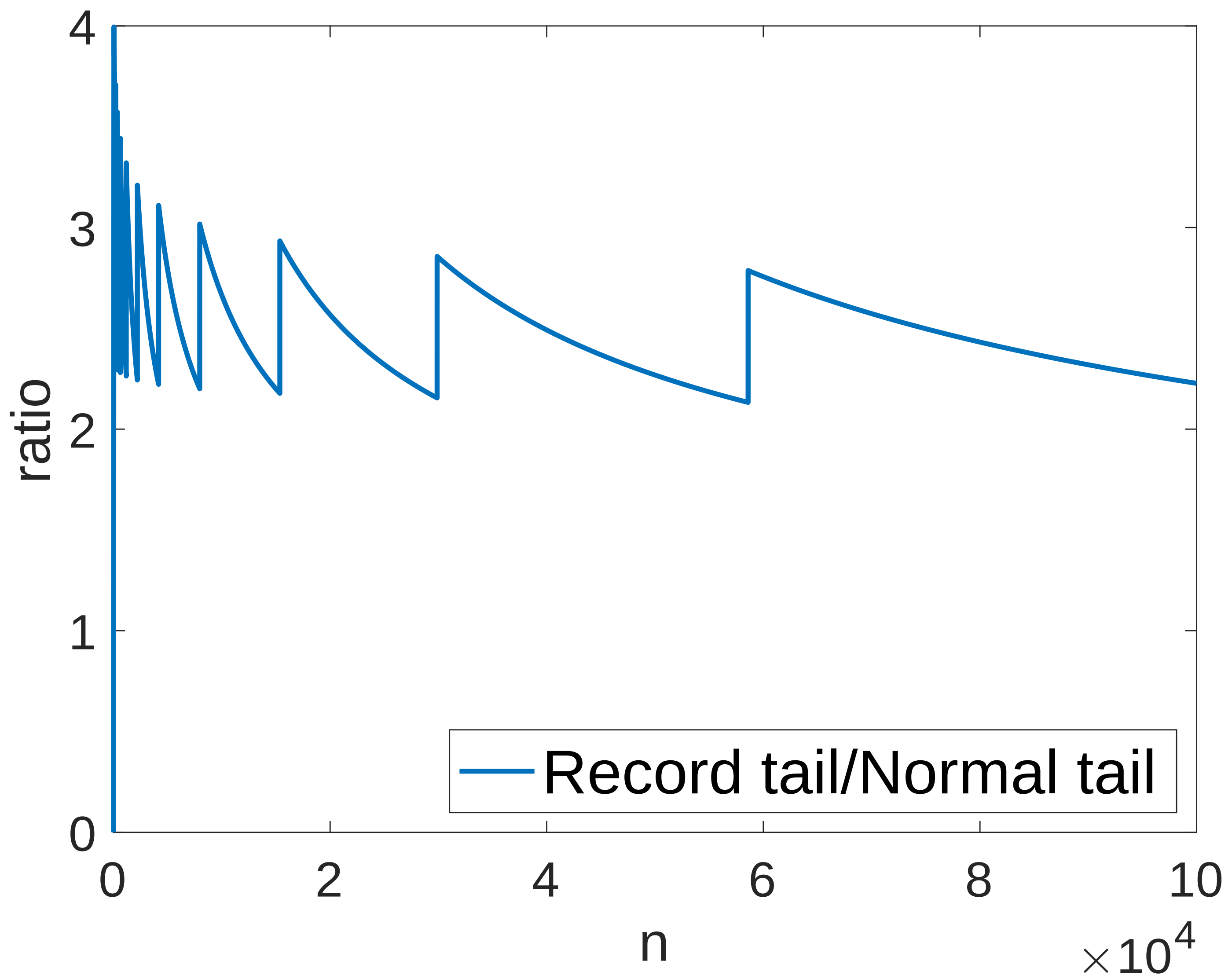
We consider the tail probabilities $\mathbb{P}(S_n \geq v_n)$ with $v_n := \lambda_n + 3 \cdot \sigma_n$ and compare $\mathbb{P}(S_n \geq v_n)$ with moderate deviations based on $\mathrm{Pn}(\lambda_n)$, $\mathrm{Pn}(\sigma_n^2)$, $N_n \sim N(\lambda_n, \sigma_n^2)$.



$\mathbb{P}(S_n \geq v_n)/\mathbb{P}(N_n \geq v_n)$
7
6
5
4
3
2
1
0
ratio
0
2
4
6
8
10
n
$\times 10^4$
Record tail/Normal tail

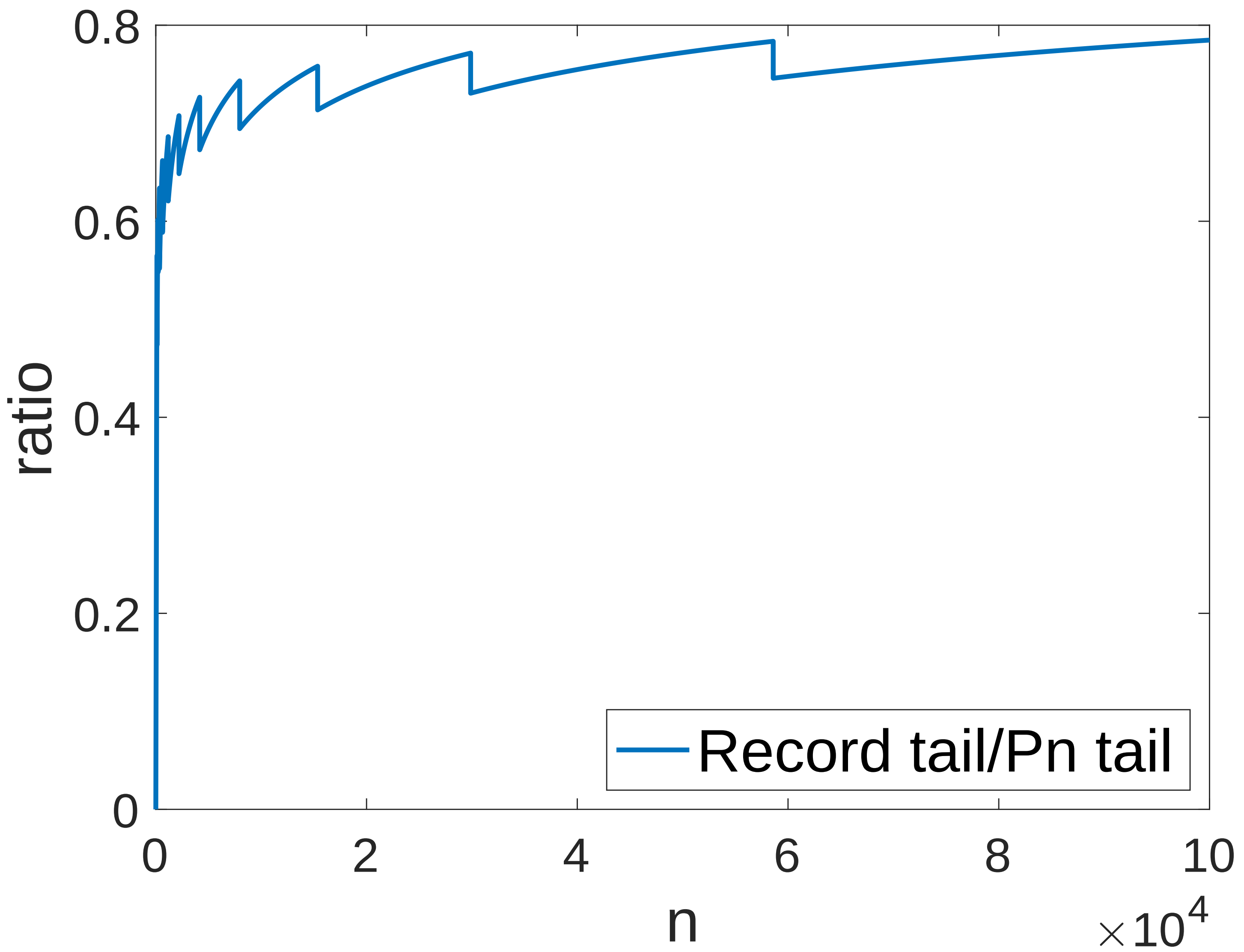
[Slide 7]

$\mathbb{P}(S_n \geq v_n)/\mathbb{P}(N_n \geq v_n)$ with 0.5 correction



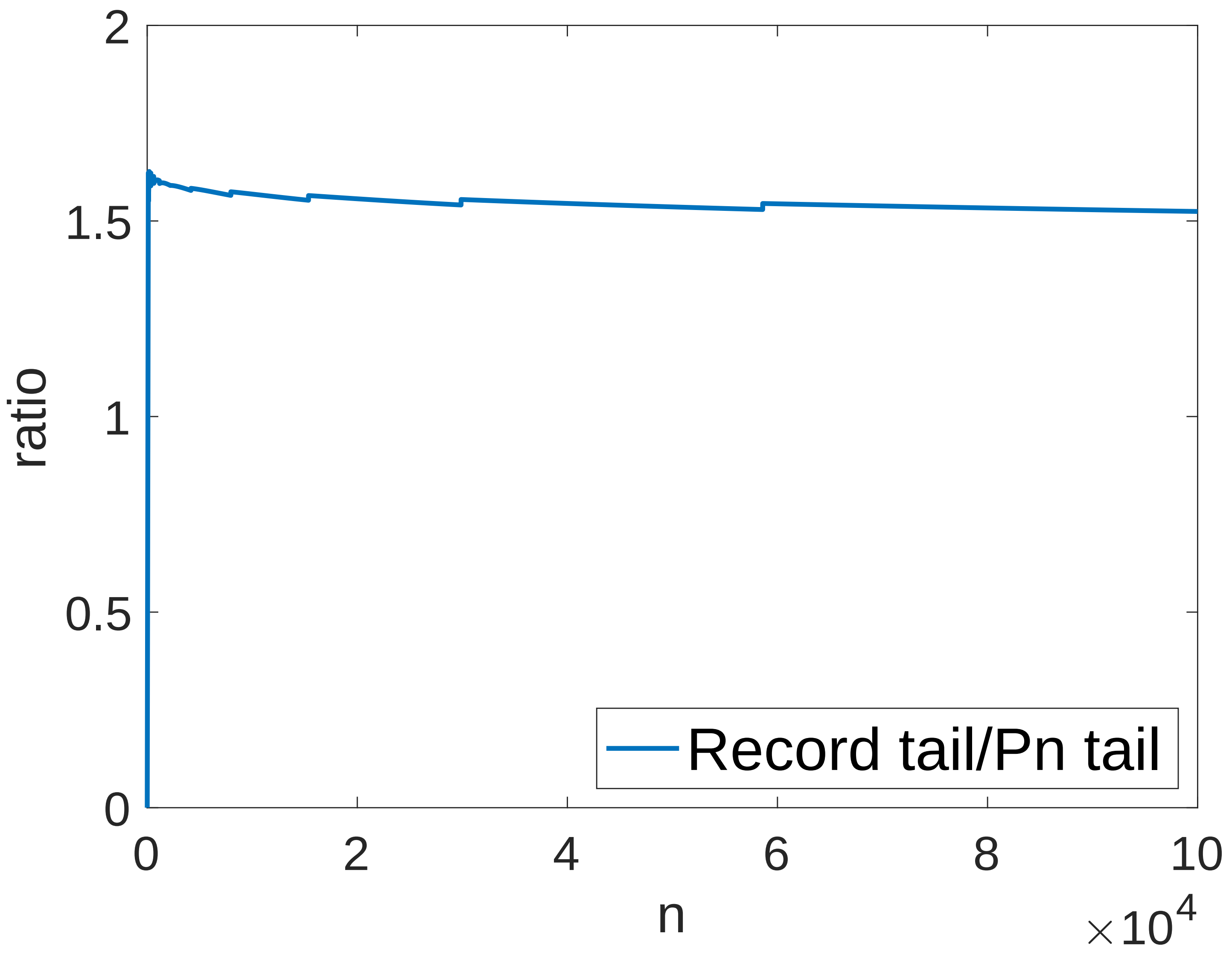
[Slide 8]

$$\mathbb{P}(S_n \geq v_n)/\mathbf{Pn}(\lambda_n)([v_n, \infty))$$



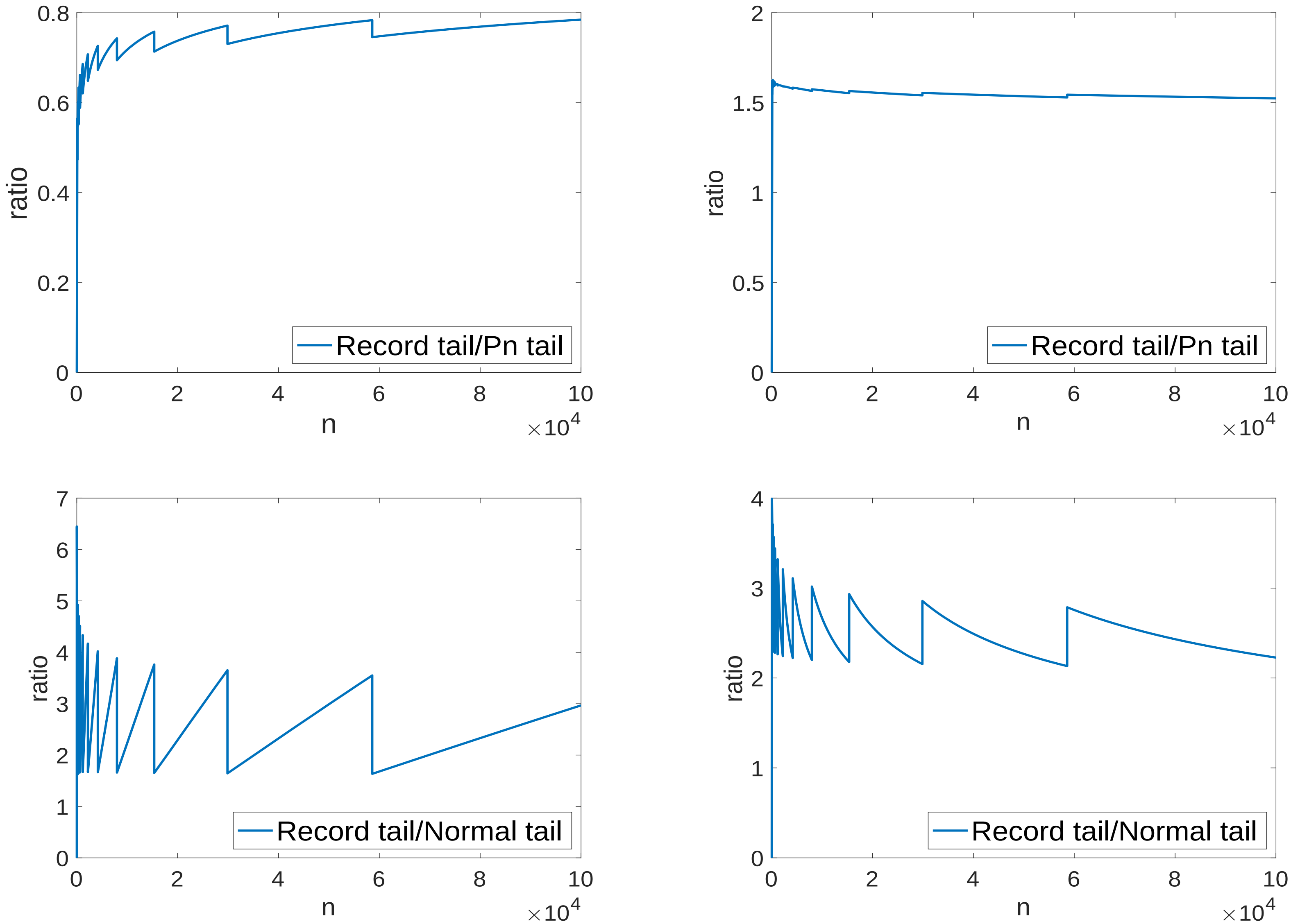
[Slide 9]

$$\mathbb{P}(S_n \geq v_n)/\mathbf{Pn}(\sigma_n^2)([v_n, \infty))$$



[Slide 10]

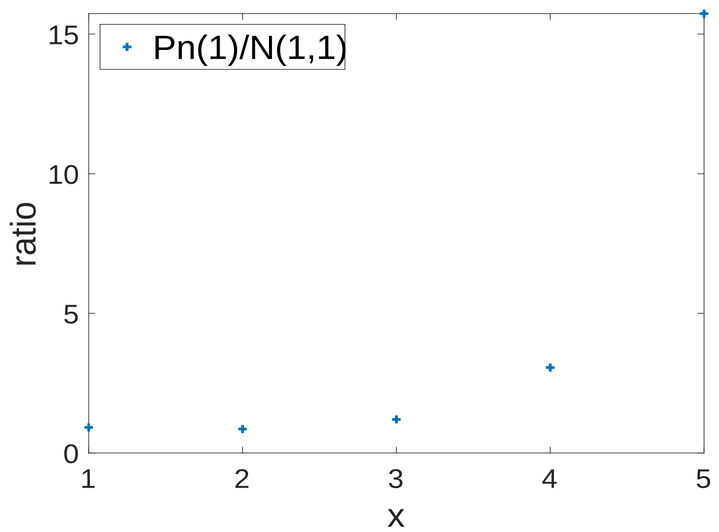
The winner is $\mathbf{Pn}(\lambda_n)$

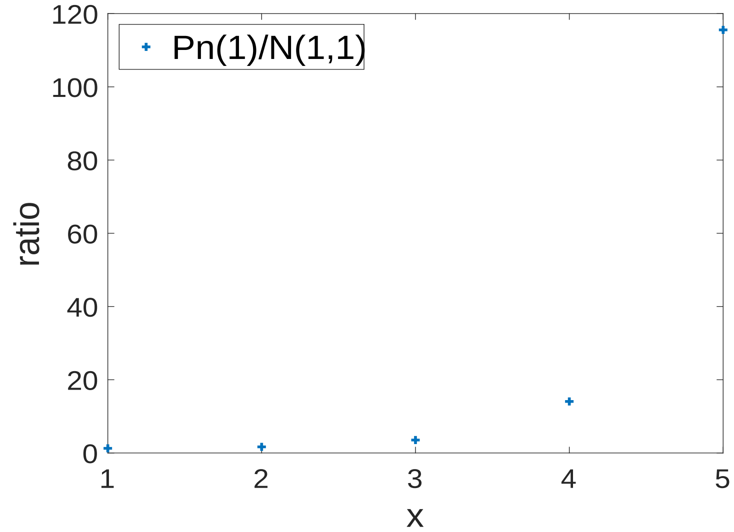


Literature?

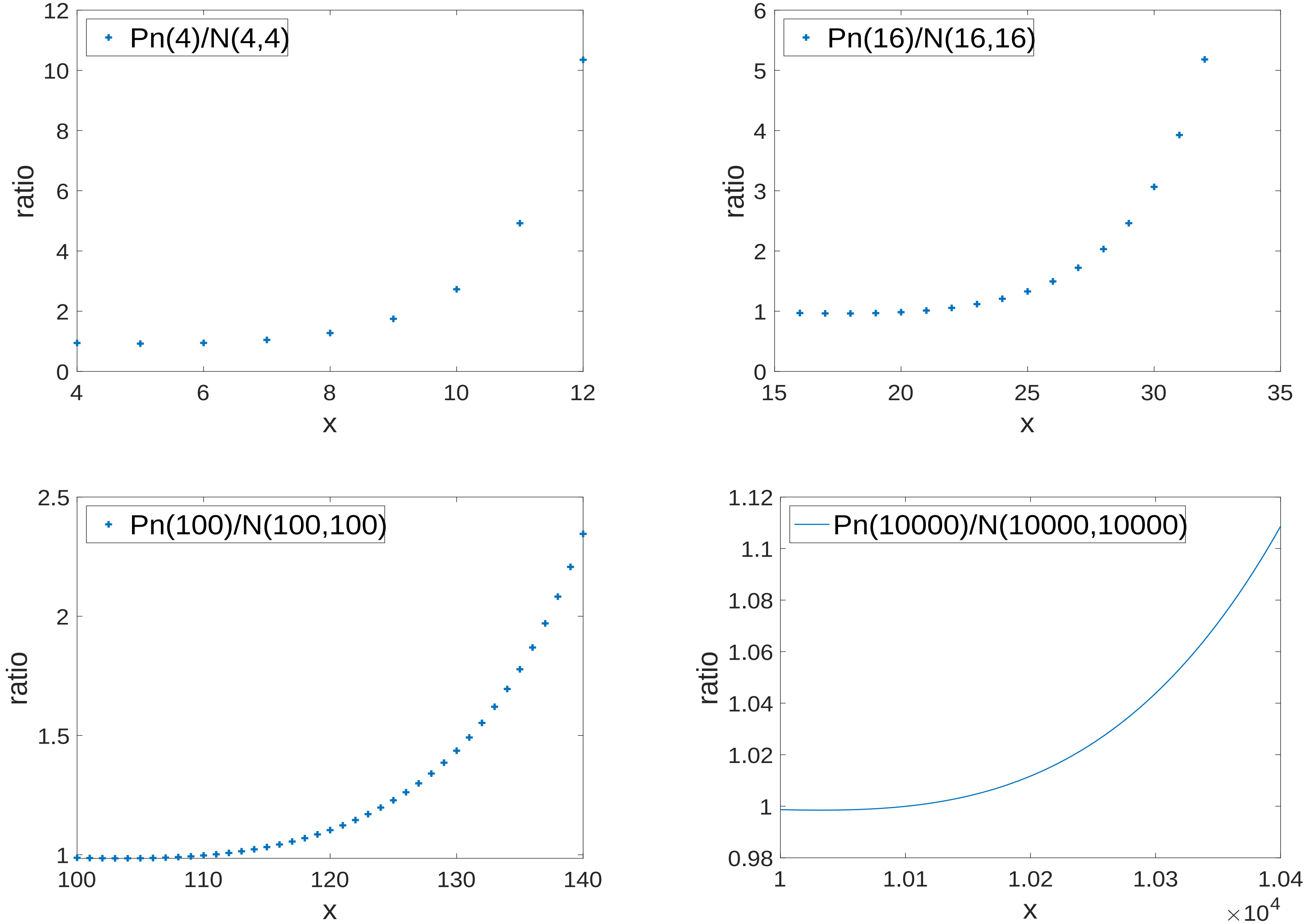
- Chen, L. H. Y. & Choi, K. P. (1992). Some asymptotic and large deviation results in Poisson approximation.
- Barbour, A. D., Chen, L. H. Y. & Choi, K. P. (1995). Poisson approximation for unbounded functions, I: Independent summands.
- Chen, L. H. Y., Fang, X. & Shao, Q.-M. (2013). Moderate deviations in Poisson approximation: a first attempt.
- Čekanavičius, V. & Vellaisamy, P. (2019). On large deviations for sums of discrete m-dependent random variables.

Pn(1) tails vs Normal tails (with and without .5 correction)



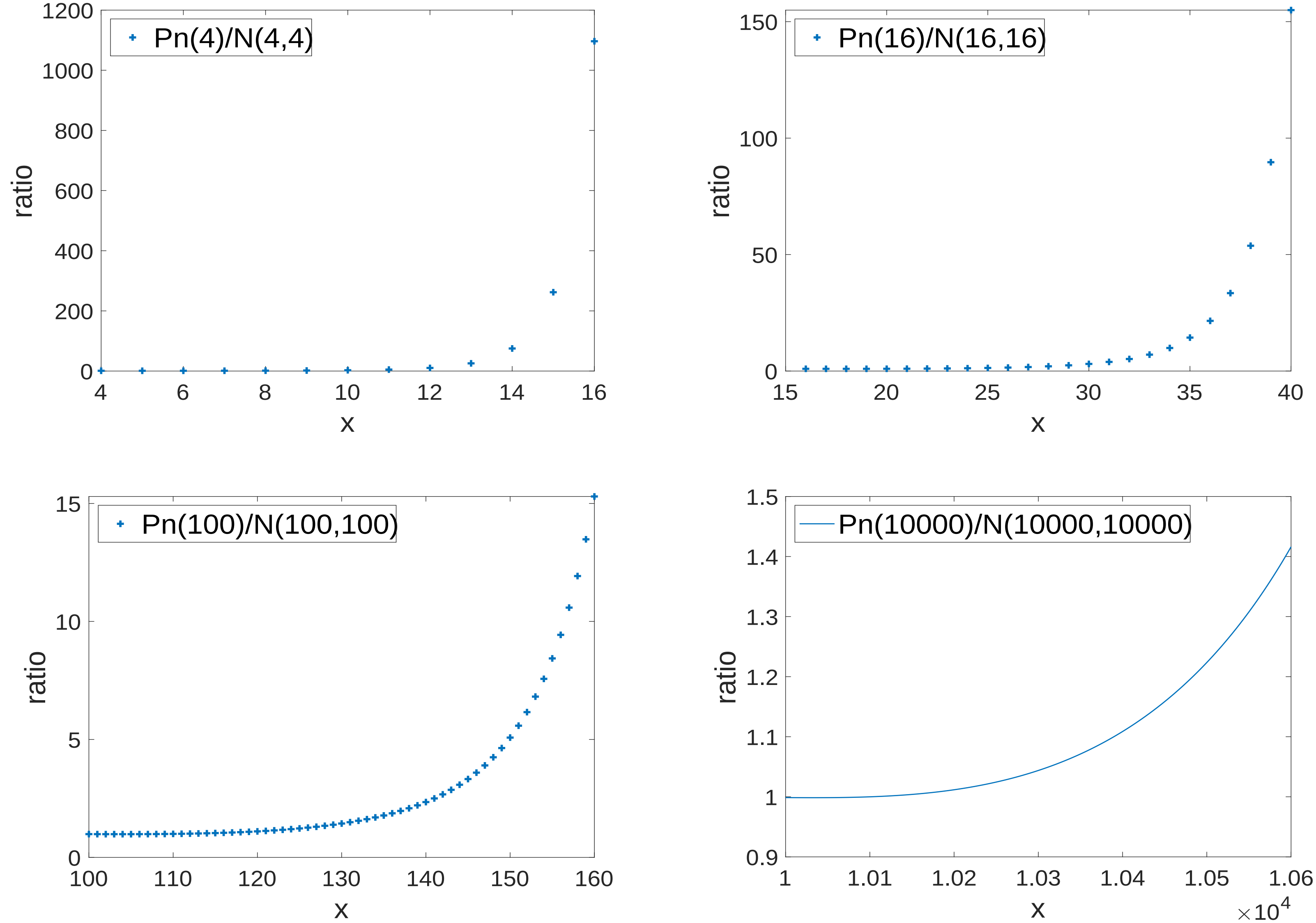


One std to four std away from the mean



[Slide 14]

One std to six std away from the mean



Conclusions

Poisson(λ) vs $N(\lambda, \lambda)$:

- Poisson has a heavier tail than normal tail;
- there is an acceleration of the ratio of the tail probabilities beyond a few standard deviations when λ is not large enough;
- unlike normal, looking at the # of standard deviations away does not work for Pn when λ is not large enough;
- a small change of the value of λ has significant impact on its moderate deviations.

Poisson moderate deviation – Chen, Fang & Shao (2013)

- Let X_i's be Bernoulli rvs with $\mathbb{P}(X_i = 1) = p_i$, $1 \leq i \leq n$.
- $W = \sum_{i=1}^n X_i$.
- The moderate deviation bound is, with $\xi = (k - \lambda)/\sqrt{\lambda}$,

$$\left| \frac{\mathbb{P}(W \geq k)}{\text{Pn}(\mu)([k, \infty))} - 1 \right| \leq C(\max p_i)(1 + \xi^2).$$

- What is C?

Matching: similar

- n: fixed;
- π: a uniform random permutation of $\{1, \ldots, n\}$;
- $X_i = \mathbf{1}_{\{i=\pi(i)\}}$;
- $W = \sum_{i=1}^{n} X_i$: # fixed points in the permutation.
- $\mathbb{P}(\{i = \pi(i)\}) = 1/n$ so $\mu = 1$.
- For $j \neq i$, $\mathbb{E}(X_j | X_i = 1) = 1/(n-1)$, so $\sum_{j \neq i} \mathbb{E}(X_j | X_i = 1) = 1$, $\mathbb{E}(W^2) = 2/n$ and $\mathrm{Var}(W) = 1$.
- For all k with $k^2/n \leq c$,

$$\left| \frac{\mathbb{P}(W \geq k)}{\mathrm{Pn}(1)([k, \infty))} - 1 \right| \leq C \underbrace{k^2/n}_{\leq c}.$$

Our approach

- W: a non-negative integer-valued random variable with mean μ and variance σ^2.
- We consider a Pn approximation of $W - a$ for an $a < \mu$.
- W^s has the size-biased distribution:

$$\mathbb{P}(W = w) = \frac{w\mathbb{P}(W = w)}{\mathbb{E}(W)}, \quad w \geq 0.$$

- Let $\lambda = \mu - a$, $Y \sim \text{Pn}(\lambda)$. Then for fixed integer k with $x := \frac{k-\lambda}{\sqrt{\lambda}} \geq 1$, we have

$$\left|\frac{\mathbb{P}(W - a \geq k)}{\mathbb{P}(Y \geq k)} - 1\right|$$
$$\leq 3\lambda^{-1} x e^{x^2+1} \{\mu\mathbb{E}|W + 1 - W^s| + |\mu - \lambda|\}$$
$$+\mathbb{P}(W - a < -1).$$

When $a = 0$

The bound is reduced to

$$\left|\frac{\mathbb{P}(W \geq k)}{\mathbb{P}(Y \geq k)} - 1\right| \leq 3xe^{x^2+1}\mathbb{E}|W + 1 - W^s|.$$

vs Chen, Fang & Shao (2013):

$$\left|\frac{\mathbb{P}(W \geq k)}{\mathbb{P}(Y \geq k)} - 1\right| \leq C(\delta_1 + \delta_2\lambda)(1 + x^2).$$

- Our bound is easy to compute.
- It contains no unspecified constants.

Can we do better?

- $\mathrm{Pn}(\lambda)$ is the stationary distribution of the birth-death process with birth rate λ and unit per capita death rate.
- The solution of

$$\begin{aligned} \mathcal{A}g(j) &:= \lambda[g(j+1) - g(j)] + j[g(j-1) - g(j)] \\ &= h(j) - \mathrm{Pn}(\lambda)\{h\}, \quad j \geq 0, \end{aligned}$$

is

$$g(j) = -\int_0^\infty \{\mathbb{E}[h(Z_j(t))] - \mathrm{Pn}(\lambda)(h)\}dt.$$

- For $h = \mathbf{1}_{[k,\infty)}$,

$$\begin{aligned} g(i) - g(i-1) &= -\int_0^\infty \mathbb{E}\left[h(Z_i(t)) - h(Z_{i-1}(t))\right] dt \\ &= -\mathbb{E}\int_0^\infty e^{-t}\mathbf{1}_{\{Z_{i-1}(t)=k-1\}} dt. \end{aligned}$$

- Using coupling and the birth-death Markov process, $\|\Delta g\| = O(\lambda^{-1/2})\mathrm{Pn}(\lambda)\{h\}$, $\|\Delta^2 g\| = O(\lambda^{-1})\mathrm{Pn}(\lambda)\{h\}$.

- Using these estimates, together with some dependence structure, we can work on $W - a$ being approximated by $\mathrm{Pn}(\lambda)$ through Stein's equation:

$$\begin{aligned} &\mathbb{P}(W - a \geq k) - \mathrm{Pn}(\lambda)([k,\infty)) \\ &= \mathbb{E}[h(W-a) - \mathrm{Pn}(\lambda)\{h\}] \\ &\approx \mathbb{E}\mathcal{A}g(W-a). \end{aligned}$$

- Some technical manipulation, we then get bounds in terms of $\|\Delta g\|$, $\|\Delta^2 g\|$ and different characteristics of W.
- Because $\mathrm{Pn}(\lambda)([k,\infty))$ is in the bounds of $\|\Delta g\|$ and $\|\Delta^2 g\|$, divide both sides by $\mathrm{Pn}(\lambda)([k,\infty))$.

Example

- $\{X_i,\ 1 \le i \le n\}$: independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i \in (0,1)$, $W = \sum_{i=1}^n X_i$.
- $\lambda = \mathbb{E}W - a > 0$, $\sigma^2 = \mathrm{Var}(W)$, $Y \sim \mathrm{Pn}(\lambda)$ and $x := \frac{k-\lambda}{\sqrt{\lambda}} \ge 1$,

$$\begin{aligned}
&\left|\frac{\mathbb{P}(W - a \ge k)}{\mathbb{P}(Y \ge k)} - 1\right| \\
\le\ & \left[4xe^{x^2+1} + 1\right] \text{ something like } \max_i p_i/\sigma \\
&+3|\lambda - \sigma^2| x\lambda^{-1}e^{x^2+1} + \exp\left\{-\frac{(\mu - a + 2)^2}{2\mu}\right\}.
\end{aligned}$$

Matching problem

For a fixed n, let π be a uniform random permutation of $\{1, \ldots, n\}$, $W = \sum_{i=1}^{n} \mathbf{1}_{\{i=\pi(i)\}}$ be the number of fixed points in the permutation, then

$$\left| \frac{\mathbb{P}(W \geq k)}{\mathrm{Pn}(1)([k, \infty))} - 1 \right| \leq \frac{6}{n} x e^{x^2+1},$$

where $x := k - 1 \geq 1$.

Take home messages

- For the counts of rare events, the tail probabilities can be approximated by the moderate deviations of Pn with twists of the parameters.
- The robustness of the tail behaviour of the Pn for large λ has not been incorporated into the bound.
- We conjecture that bound can be sharpened by a factor possibly as much as 1/3.
- We don't have any idea about the lower bound.

Thank you!