# Inference via randomized test statistics

Nikita Puchkin
(joint work with V. Ulyanov, arXiv:2112.06583)

HSE University and IITP RAS, Moscow

June 24, 2022

# Hypothesis testing

A statistician observes a sample $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n \in \mathbb{R}^r$ of i.i.d. random vectors with multinomial distribution $\mathrm{Mult}(1, \boldsymbol{p}^\circ)$ or the aggregated data $\boldsymbol{Y} = \boldsymbol{\eta}_1 + \dots + \boldsymbol{\eta}_n \sim \mathrm{Mult}(n, \boldsymbol{p}^\circ)$

$$\mathbb{P}\left(\boldsymbol{\eta}_1 = e_j\right) = p_j^\circ, \quad j \in \{1, \dots, r\}$$

# Hypothesis testing

A statistician observes a sample $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n \in \mathbb{R}^r$ of i.i.d. random vectors with multinomial distribution $\mathrm{Mult}(1, \boldsymbol{p}^\circ)$ or the aggregated data $\boldsymbol{Y} = \boldsymbol{\eta}_1 + \dots + \boldsymbol{\eta}_n \sim \mathrm{Mult}(n, \boldsymbol{p}^\circ)$

$$\mathbb{P}\left(\boldsymbol{\eta}_1 = e_j\right) = p_j^\circ, \quad j \in \{1, \dots, r\}$$

A simple hypothesis: $H_0 : \boldsymbol{p}^\circ = \boldsymbol{p}$

# Hypothesis testing

A statistician observes a sample $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n \in \mathbb{R}^r$ of i.i.d. random vectors with multinomial distribution $\mathrm{Mult}(1, \boldsymbol{p}^\circ)$ or the aggregated data $\boldsymbol{Y} = \boldsymbol{\eta}_1 + \dots + \boldsymbol{\eta}_n \sim \mathrm{Mult}(n, \boldsymbol{p}^\circ)$

$$\mathbb{P}\left(\boldsymbol{\eta}_1 = e_j\right) = p_j^\circ, \quad j \in \{1, \dots, r\}$$

A simple hypothesis: $H_0 : \boldsymbol{p}^\circ = \boldsymbol{p}$

Common approach: use Pearson's statistic

$$T_P = \sum_{j=1}^{r} \frac{(Y_j - np_j)^2}{np_j}$$

# Hypothesis testing

A statistician observes a sample $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n \in \mathbb{R}^r$ of i.i.d. random vectors with multinomial distribution $\mathrm{Mult}(1, \boldsymbol{p}^\circ)$ or the aggregated data $\boldsymbol{Y} = \boldsymbol{\eta}_1 + \dots + \boldsymbol{\eta}_n \sim \mathrm{Mult}(n, \boldsymbol{p}^\circ)$

$$\mathbb{P}\left(\boldsymbol{\eta}_1 = e_j\right) = p_j^\circ, \quad j \in \{1, \dots, r\}$$

A simple hypothesis: $H_0 : \boldsymbol{p}^\circ = \boldsymbol{p}$

Common approach: use Pearson's statistic

$$T_P = \sum_{j=1}^{r} \frac{(Y_j - np_j)^2}{np_j} \quad \xrightarrow{d} \chi^2(r-1), \quad n \to \infty$$

and reject $H_0$ if and only if $T_P$ exceeds $\chi^2_{1-\alpha}(r-1)$

# Hypothesis testing

Other possible choices: power divergence test statistics [Cressie and Read, 1984]:

$$T_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{r} Y_j \left[ \left( \frac{Y_j}{np_j} \right)^\lambda - 1 \right]$$

# Hypothesis testing

Other possible choices: power divergence test statistics [Cressie and Read, 1984]:

$$T_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{r} Y_j \left[ \left( \frac{Y_j}{np_j} \right)^\lambda - 1 \right]$$

When $\lambda = -1$ or $\lambda = 0$, the expression should be understood as a passage to limit

# Hypothesis testing

Other possible choices: power divergence test statistics [Cressie and Read, 1984]:

$$T_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{r} Y_j \left[ \left( \frac{Y_j}{np_j} \right)^\lambda - 1 \right]$$

When $\lambda = -1$ or $\lambda = 0$, the expression should be understood as a passage to limit

Theorem:

$$T_\lambda \xrightarrow{d} \chi^2(r-1), \quad n \to \infty$$

# Hypothesis testing

Other possible choices: power divergence test statistics [Cressie and Read, 1984]:

$$T_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{r} Y_j \left[ \left( \frac{Y_j}{np_j} \right)^\lambda - 1 \right]$$

When $\lambda = -1$ or $\lambda = 0$, the expression should be understood as a passage to limit

Theorem:

$$T_\lambda \xrightarrow{d} \chi^2(r-1), \quad n \to \infty$$

**Question 1**: what is the rate of convergence?

Other possible choices: power divergence test statistics [Cressie and Read, 1984]:

$$T_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{r} Y_j \left[ \left( \frac{Y_j}{np_j} \right)^\lambda - 1 \right]$$

When $\lambda = -1$ or $\lambda = 0$, the expression should be understood as a passage to limit

Theorem:

$$T_\lambda \xrightarrow{d} \chi^2(r-1), \quad n \to \infty$$

**Question 1**: what is the rate of convergence?

**Question 2**: can we propose something better?

# Rates of convergence

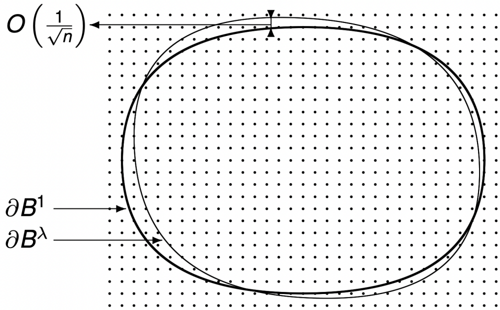
Kolmogorov distance:

$$d_K(\mathrm{Law}(\xi), \mathrm{Law}(\eta)) = \sup_{t \in \mathbb{R}} |\mathbb{P}(\xi > t) - \mathbb{P}(\eta > t)|.$$

[Yarnold, 1972, Ulyanov and Zubov, 2009]:

$$d_K\left(\mathrm{Law}(T_\lambda), \chi^2(r-1)\right) \lesssim \sup_{t>0} \frac{|N^\lambda(t) - n^{(r-1)/2} V^\lambda(t)|}{n^{(r-1)/2}} + \frac{1}{n}$$

- $B^\lambda(t) = \{y : T_\lambda(y) \leqslant t\}$
- $N^\lambda(t)$ — # of lattice points inside $B^\lambda(t)$
- $V^\lambda(t)$ — volume of $B^\lambda(t)$



# Rates of convergence

Kolmogorov distance:

$$d_K(\mathrm{Law}(\xi), \mathrm{Law}(\eta)) = \sup_{t \in \mathbb{R}} \left| \mathbb{P}(\xi > t) - \mathbb{P}(\eta > t) \right|.$$

Pearson's statistic [Götze and Ulyanov, 2003]:

$$d_K\left(\mathrm{Law}(T_P), \chi^2(r-1)\right) = \begin{cases} O\left(n^{-1+1/r}\right), & \text{if } 2 \leqslant r \leqslant 5, \\ O\left(n^{-1}\right), & \text{if } r \geqslant 6. \end{cases}$$

# Rates of convergence

Kolmogorov distance:

$$d_K(\mathrm{Law}(\xi), \mathrm{Law}(\eta)) = \sup_{t \in \mathbb{R}} |\mathbb{P}(\xi > t) - \mathbb{P}(\eta > t)| .$$

Pearson's statistic [Götze and Ulyanov, 2003]:

$$d_K\left(\mathrm{Law}(T_P), \chi^2(r-1)\right) = \begin{cases} O\left(n^{-1+1/r}\right), & \text{if } 2 \leqslant r \leqslant 5, \\ O\left(n^{-1}\right), & \text{if } r \geqslant 6. \end{cases}$$

**Remark**: the $O(1/\sqrt{n})$ rate of convergence cannot be improved when $r = 2$

# Why the rate $O(1/\sqrt{n})$ cannot be improved when $r = 2$?

Consider the case $p_1 = p_2 = 1/2$, $n$ is even:

$$
\begin{aligned}
T_p &= \sum_{j=1}^{2} \frac{(Y_j - p_j)^2}{np_j} \\
&= \frac{(Y_1 - n/2)^2}{n/2} + \frac{(n - Y_1 - n/2)^2}{n/2} \\
&= \frac{4(Y_1 - n/2)^2}{n}.
\end{aligned}
$$

# Why the rate $O(1/\sqrt{n})$ cannot be improved when $r = 2$?

Since $Y_1 \sim \mathrm{Binom}(n, 1/2)$, we have

$$\mathbb{P}\left(Y_1 = \frac{n}{2}\right) = \binom{n}{n/2} 4^{-n} = \frac{n!}{(n/2)!(n/2)!} \cdot 4^{-n}$$
$$\sim \frac{\sqrt{2\pi n}(n/e)^n}{\pi n (n/2e)^n} \cdot 4^{-n} = \sqrt{\frac{2}{\pi n}}$$

Hence, $\mathbb{P}(T_P = 0) \sim (\pi n/2)^{-1/2}$ while $\mathbb{P}(\chi^2(1) \leqslant 0) = 0$

# Rates of convergence

Kolmogorov distance:

$$d_K(\mathrm{Law}\,(\xi)\,,\mathrm{Law}\,(\eta)) = \sup_{t\in\mathbb{R}} \left|\mathbb{P}\,(\xi > t) - \mathbb{P}\,(\eta > t)\right|.$$

Power divergence test statistics [Assylbekov, 2010, Assylbekov et al., 2011]:

$$d_K\left(\mathrm{Law}\,(T_\lambda)\,,\chi^2(r-1)\right) = \begin{cases} O\left(n^{-50/73}(\log n)^{315/146}\right), & \text{if } r = 3, \\ O\left(n^{-1+6/(7r-3)}\right), & \text{if } 4 \leqslant r \leqslant 8, \\ O\left(n^{-1+5/(6r-4)}\right), & \text{if } r \geqslant 9 \end{cases}$$

# Rates of convergence

On an event with high probability

$$T_\lambda = \sum_{j=1}^{r} \frac{(Y_j - np_j)^2}{np_j} + \frac{\lambda - 1}{3} \underbrace{\sum_{j=1}^{r} \frac{(Y_j - np_j)^3}{n^2 p_j^2}}_{O(1/\sqrt{n})} + O\left(\frac{(\log n)^4}{n}\right)$$

The first term converges to $\chi^2(r-1)$ with the rate $O(1/n)$, provided that $r \geqslant 6$

# Rates of convergence

Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex non-negative function such that

$$\phi(1) = \phi'(1) = 0, \quad \phi''(1) > 0$$

Phi-divergence test statistic:

$$T_\phi = \frac{2n}{\phi''(1)} \sum_{j=1}^{r} p_j \phi\left(\frac{Y_j}{np_j}\right)$$

# Rates of convergence

Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a convex non-negative function such that

$$\phi(1) = \phi'(1) = 0, \quad \phi''(1) > 0$$

Phi-divergence test statistic:

$$T_\phi = \frac{2n}{\phi''(1)} \sum_{j=1}^{r} p_j \phi\left(\frac{Y_j}{np_j}\right)$$

**Remark**: power divergence statistics belong to the family of phi-divergence statistics

# Berry-Esseen theorem

- $\xi_1, \dots, \xi_n$ – i.i.d. centered random variables, $\mathbb{E}\xi_1^2 = 1$, $\mathbb{E}|\xi_1|^3 < \infty$

# Berry-Esseen theorem

- $\xi_1, \ldots, \xi_n$ – i.i.d. centered random variables, $\mathbb{E}\xi_1^2 = 1$, $\mathbb{E}|\xi_1|^3 < \infty$
- It holds that

$$\sup_{\substack{a,b\in\mathbb{R},\\ a<b}} \left| \mathbb{P}\left( a \leqslant \sum_{i=1}^n \theta_i \xi_i \leqslant b \right) - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} \mathrm{d}x \right| \leqslant \frac{\mathbb{E}|\xi_1|^3}{\sqrt{n}}$$

# Berry-Esseen theorem

- $\xi_1, \dots, \xi_n$ – i.i.d. centered random variables, $\mathbb{E}\xi_1^2 = 1$, $\mathbb{E}|\xi_1|^3 < \infty$
- It holds that

$$\sup_{\substack{a,b \in \mathbb{R}, \\ a<b}} \left| \mathbb{P}\left( a \leqslant \sum_{i=1}^n \theta_i \xi_i \leqslant b \right) - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} \mathrm{d}x \right| \leqslant \frac{\mathbb{E}|\xi_1|^3}{\sqrt{n}}$$

Using the Berry-Esseen theorem, one can prove that

$$d_K\left(\mathrm{Law}\left(T_\phi\right), \chi^2(r-1)\right) = O(n^{-1/2})$$

under mild assumptions on $\phi$

# CLT for weighted sums

- $\xi_1, \dots, \xi_n$ – i.i.d. centered random variables with the unit variance and a finite fourth moment

# CLT for weighted sums

- $\xi_1, \dots, \xi_n$ – i.i.d. centered random variables with the unit variance and a finite fourth moment
- $\boldsymbol{\theta} \sim \mathcal{U}(\mathcal{S}^{n-1})$ – $n$-dimensional vector of coefficients

# CLT for weighted sums

- $\xi_1, \dots, \xi_n$ – i.i.d. centered random variables with the unit variance and a finite fourth moment
- $\boldsymbol{\theta} \sim \mathcal{U}(\mathcal{S}^{n-1})$ – $n$-dimensional vector of coefficients
- CLT for weighted sums [Klartag and Sodin, 2012]: for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, it holds that

$$\sup_{\substack{a,b \in \mathbb{R}, \\ a<b}} \left| \mathbb{P}\left( a \leqslant \sum_{i=1}^n \theta_i \xi_i \leqslant b \,\middle|\, \boldsymbol{\theta} \right) - \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-x^2/2} \mathrm{d}x \right| \leqslant \frac{C_{KS} \mathbb{E}\xi_1^4 \log^2(1/\delta)}{n},$$

where $C_{KS}$ is an absolute constant

# Multivariate CLT for weighted sums

**Theorem ([Ayvazyan and Ulyanov, 2022])**

*Let $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ be i.i.d. random vectors in $\mathbb{R}^d$, $\mathbb{E}\boldsymbol{\xi}_1 = \mathbf{0}$, $\mathbb{E}\boldsymbol{\xi}_1\boldsymbol{\xi}_1^\top = I_d$, $\mathbb{E}\|\boldsymbol{\xi}_1\|^4 < \infty$. Denote the family of convex Borel sets in $\mathbb{R}^d$ by $\mathfrak{B}$ and let $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, I_d)$ be the standard Gaussian random vector in $\mathbb{R}^d$. Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$ over $\boldsymbol{\theta} \sim \mathcal{U}(\mathcal{S}^{n-1})$, it holds that*

$$\sup_{B \in \mathfrak{B}} \left| \mathbb{P}\left( \sum_{i=1}^n \theta_i \boldsymbol{\xi}_i \in B \,\middle|\, \boldsymbol{\theta} \right) - \mathbb{P}(\boldsymbol{\eta} \in B) \right| \leqslant \frac{C_d \mathbb{E}\|\boldsymbol{\xi}_1\|^4 \log^2(1/\delta)}{n}, \tag{1}$$

*where the constant $C_d$ depends on $d$ only.*

# Randomized phi-divergence test statistic

- $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n \in \mathbb{R}^r$ are i.i.d. random vectors $\mathrm{Mult}(1, \boldsymbol{p})$
- Define a weighted sum

$$\boldsymbol{X}^{\boldsymbol{\theta}} = \sum_{i=1}^{n} \theta_i (\boldsymbol{\eta}_i - \boldsymbol{p})$$

- Randomized phi-divergence test statistic:

$$\mathcal{T}_\phi = \frac{2n}{\phi''(1)} \sum_{j=1}^{r} p_j \phi \left( 1 + \frac{X_j^{\boldsymbol{\theta}}}{\sqrt{n p_j}} \right)$$

# Randomized phi-divergence test statistic

**Assumption**

The function $\phi$ is three times differentiable at 1, $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$, and the third derivative $\phi'''$ is $\mathsf{L}$-Lipschitz on $[1 - \Delta, 1 + \Delta]$ for some $\Delta > 0$.

# Randomized phi-divergence test statistic

**Assumption**

The function $\phi$ is three times differentiable at 1, $\phi(1) = \phi'(1) = 0$, $\phi''(1) > 0$, and the third derivative $\phi'''$ is L-Lipschitz on $[1 - \Delta, 1 + \Delta]$ for some $\Delta > 0$.

**Assumption**

The sample size $n$ is sufficiently large, that is,

$$5|\phi'''(1)| \left(p_j(1 - p_j) + \log n\right) \leqslant 4\phi''(1)\sqrt{n} p_j \quad \text{for all } j \in \{1, \dots, r\},$$

$$5 \log n \leqslant 2 p_{\min} \Delta \sqrt{n}, \quad \text{and} \quad 16r^3 + 16r^2 \log n \leqslant n p_{\min}, \quad \text{where } p_{\min} = \min_{1 \leqslant j \leqslant r} p_j$$

# Randomized phi-divergence test statistic

## Theorem (P. and Ulyanov)

*Let the aforementioned assumptions be satisfied and let* $p_{\min} > 0$*. Then, for any* $\delta \in (0, 1)$*, with probability at least* $1 - \delta$ *(over* $\boldsymbol{\theta}$ *uniformly distributed on the unit sphere* $\mathcal{S}^{n-1}$*), it holds that*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left( \mathcal{T}_\phi > t \,\middle|\, \boldsymbol{\theta} \right) - \mathbb{P}\left( Z > t \right) \right|$$

$$\lesssim \left( \frac{\phi'''(1)}{\phi''(1)} \right)^2 \frac{r^{3/2} + (\log n)^{3/2}}{n p_{\min}} + \frac{C_{r-1} r \log^2(1/\delta)}{n p_{\min}} + \frac{\mathsf{L}\sqrt{r}(\log n)^4}{\phi''(1) n p_{\min}^3}.$$

*Here* $Z \sim \chi^2(r-1)$ *and* $C_{r-1}$ *is the same constant as in (1).*

# Key ideas

**Step 1.** Taylor's expansion:

$$\mathcal{T}_\phi = \underbrace{\sum_{j=1}^{r} \frac{(X_j^{\boldsymbol{\theta}})^2}{p_j} + \frac{\phi'''(1)}{3\phi''(1)} \sum_{j=1}^{r} \frac{(X_j^{\boldsymbol{\theta}})^3}{\sqrt{n p_j^2}}}_{\mathcal{Q}(\boldsymbol{X}^{\boldsymbol{\theta}})}$$

$$+ \underbrace{\frac{1}{\phi''(1)} \sum_{j=1}^{r} \int\limits_0^1 \left[\phi'''\left(1 + \frac{v X_j^{\boldsymbol{\theta}}}{\sqrt{n p_j}}\right) - \phi'''(1)\right] \frac{(X_j^{\boldsymbol{\theta}})^3}{\sqrt{n p_j^2}} (1-v)^2 \mathrm{d}v}_{\mathcal{R}}\,.$$

# Key ideas

**Step 2.** Concentration of measure: there is an event $E_1$, $\mathbb{P}(E_1 \,|\, \boldsymbol{\theta}) \geqslant 1 - 2r/n$, such that, conditionally on $\boldsymbol{\theta}$,

$$|\mathcal{R}| \leqslant \Psi_n = \frac{2\mathsf{L}}{\phi''(1)n} + \frac{2\mathsf{L}r(\log n)^4}{\phi''(1)np_{\min}^3} \qquad \text{on } E_1$$

# Key ideas

**Step 3.** CLT for weighted sums [Ayvazyan and Ulyanov, 2022]:

$$d_K\left(\mathrm{Law}\left(\mathcal{Q}(\boldsymbol{X}^{\boldsymbol{\theta}})\right), \mathrm{Law}\left(\mathcal{Q}(\widetilde{\boldsymbol{X}})\right)\right) \leqslant \frac{2r}{n} + \frac{C_{r-1} r \log^2(1/\delta)}{n p_{\min}},$$

where $\widetilde{\boldsymbol{X}} \sim \mathcal{N}(0, \Sigma)$ and

$$\Sigma = \mathrm{diag}(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^\top$$

# Key ideas

**Step 4.** Key lemma:

$$d_K\left(\mathrm{Law}\left(\mathcal{Q}(\widetilde{\boldsymbol{X}})\right), \chi^2(r-1)\right) \lesssim \left(\frac{\phi'''(1)}{\phi''(1)}\right)^2 \frac{r^{3/2} + (\log n)^{3/2}}{n p_{\min}}$$

# Key ideas

$$d_K\left(\mathrm{Law}\left(\mathcal{Q}(\widetilde{\boldsymbol{X}})\right), \chi^2(r-1)\right) = \sup_{t\in\mathbb{R}} \left|\mathbb{P}(\mathcal{Q}(\widetilde{\boldsymbol{X}}) > t) - \mathbb{P}(\chi^2(r-1) > t)\right|$$

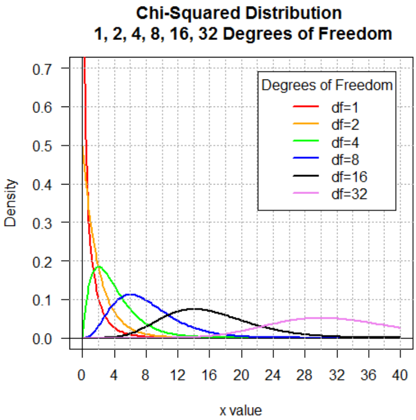
Simple case: $t > 8(r-1) + 8\log n$ or $t < 0$

$$\mathbb{P}(\chi^2(r-1) > t) \leqslant 1/n$$

Large deviations:

$$\mathbb{P}(\mathcal{Q}(\widetilde{\boldsymbol{X}}) > t) \lesssim \left(\frac{\phi'''(1)}{\phi''(1)}\right)^2 \frac{r}{np_{\min}}$$

$$\mathbb{P}(\mathcal{Q}(\widetilde{\boldsymbol{X}}) \leqslant 0) \lesssim \left(\frac{\phi'''(1)}{\phi''(1)}\right)^2 \frac{1}{np_{\min}}$$



# Key ideas

Let $\rho = \|D^{-1/2}\widetilde{\boldsymbol{X}}\|$ and $\boldsymbol{\tau} = D^{-1/2}\widetilde{\boldsymbol{X}}/\rho$, where $D = \operatorname{diag}(\boldsymbol{p})$, and

$$\mathcal{S} = \frac{\phi'''(1)}{3\phi''(1)} \sum_{j=1}^{r} \frac{\tau_j^3 \sqrt{p_{\min}}}{\sqrt{p_j}}$$

Then

- $\rho^2 \sim \chi^2(r-1)$
- $\rho$ and $\mathcal{S}$ are independent

$$\mathcal{Q}(\widetilde{\boldsymbol{X}}) = \sum_{j=1}^{r} \frac{\widetilde{X}_j^2}{p_j} + \frac{\phi'''(1)}{3\phi''(1)} \sum_{j=1}^{r} \frac{\widetilde{X}_j^3}{\sqrt{n} p_j^2}$$
$$= \rho^2 + \frac{\rho^3 \mathcal{S}}{\sqrt{n p_{\min}}}$$

# Key ideas

For any $t \in (0, 8(r-1) + 8\log n)$, let a random variable $\sigma_t$ be a root of the equation

$$t = \sigma_t^2 + \frac{\sigma_t^3 \mathcal{S}}{\sqrt{n p_{\min}}}$$

from the interval $(\sqrt{0.5t}, \sqrt{2t})$

$$\sup_{0 < t \leqslant 8(r-1) + 8\log n} \left| \mathbb{P}\left( \rho^2 + \frac{\rho^3 \mathcal{S}}{\sqrt{n p_{\min}}} > t \right) - \mathbb{P}\left(\rho^2 > t\right) \right|$$

$$= \sup_{0 < t \leqslant 8(r-1) + 8\log n} \left| \mathbb{E}\mathbb{P}\left( \rho^2 + \frac{\rho^3 \mathcal{S}}{\sqrt{n p_{\min}}} > t \,\middle|\, \mathcal{S} \right) - \mathbb{P}\left(\rho^2 > t\right) \right|$$

$$= \sup_{0 < t \leqslant 8(r-1) + 8\log n} \left| \mathbb{E}\mathbb{P}\left( \rho^2 + \frac{\rho^3 \mathcal{S}}{\sqrt{n p_{\min}}} > \sigma_t^2 + \frac{\sigma_t^3 \mathcal{S}}{\sqrt{n p_{\min}}} \,\middle|\, \mathcal{S} \right) - \mathbb{E}\mathbb{P}\left( \rho^2 > \sigma_t^2 + \frac{\sigma_t^3 \mathcal{S}}{\sqrt{n p_{\min}}} \,\middle|\, \mathcal{S} \right) \right|$$

# Key ideas

**Step 5.** Gaussian anti-concentration [Götze et al., 2019]: if $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, then

$$\sup_{t>0} \mathbb{P}\left(t < \|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < t + \varepsilon\right) \lesssim \varkappa_\Sigma \, \varepsilon$$

# Key ideas

**Step 5.** Gaussian anti-concentration [Götze et al., 2019]:

$$
\begin{aligned}
d_K\left(\mathrm{Law}\left(\mathcal{T}_\phi\right), \chi^2(r-1)\right) &= d_K\left(\mathrm{Law}\left(\mathcal{Q}(\boldsymbol{X}^{\boldsymbol{\theta}}) + \mathcal{R}\right), \chi^2(r-1)\right) \\
&\approx d_K\left(\mathrm{Law}\left(\mathcal{Q}(\widetilde{\boldsymbol{X}}) \pm \Psi_n\right), \chi^2(r-1)\right) \\
&\lesssim \left(\frac{\phi'''(1)}{\phi''(1)}\right)^2 \frac{r^{3/2} + (\log n)^{3/2}}{n p_{\min}} \\
&\quad + \frac{C_{r-1} r \log^2(1/\delta)}{n p_{\min}} + \frac{\mathsf{L}\sqrt{r}(\log n)^4}{\phi''(1) n p_{\min}^3}
\end{aligned}
$$

# References


Assylbekov, Z. (2010).
Convergence Rate of Multinomial Goodness-of-fit Statistics to Chi-square Distribution.
*Hiroshima Mathematical Journal*, 40(1):115 – 131.

Assylbekov, Z., Zubov, V., and Ulyanov, V. (2011).
On Approximating Some Statistics of Goodness-of-fit Tests in the Case of Three-dimensional Discrete Data.
*Siberian Mathematical Journal*, 52:571–584.

# References


Ayvazyan, S. and Ulyanov, V. (2022).
A Multivariate CLT for “Typical” Weighted Sums with Rate of Convergence of Order O(1/n).
In: Foundations of Modern Statistics - Festschrift in Honor of Vladimir Spokoiny. Springer Proceedings in Mathematics & Statistics, Springer, Cham. (in print, available at arxiv.org/abs/2112.05815).

Cressie, N. and Read, T. (1984).
Multinomial Goodness-of-Fit Tests.
*Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):440–464.

# References

Götze, F., Naumov, A., Spokoiny, V., and Ulyanov, V. (2019).
Large Ball Probabilities, Gaussian Comparison and Anti-concentration.
*Bernoulli*, 25(4A):2538 – 2563.

Götze, F. and Ulyanov, V. (2003).
Asymptotic Distribution of $\chi^2$-type Statistics.
*Preprint 03-033, Research group Spectral analysis, asymptotic distributions and stochastic dynamics, Bielefeld Univ., Bielefeld.*

Klartag, B. and Sodin, S. (2012).
Variations on the Berry-Esseen Theorem.
*Theory of Probability & Its Applications*, 56(3):403–419.

# References


Ulyanov, V. and Zubov, V. (2009).
Refinement on the Convergence of One Family of Goodness-of-fit Statistics to Chi-squared Distribution.
*Hiroshima Mathematical Journal*, 39(1):133 – 161.

Yarnold, J. (1972).
Asymptotic Approximations for the Probability that a Sum of Lattice Random Vectors Lies in a Convex Set.
*The Annals of Mathematical Statistics*, 43(5):1566 – 1580.

Thank you for attention!

Thank you for attention!

Any questions?