Rachunek prawdopodobieństwa MAT1332
Wydział Matematyki, Matematyka Stosowana
Wykładowca: dr hab. Agnieszka Jurlewicz

Wykład 11: Inne twierdzenia graniczne. Związki asymptotyczne między rozkładami.

Twierdzenie Poissona

Niech S_n oznacza liczbę sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p_n. Jeżeli $p_n \underset{n\to\infty}{\longrightarrow} 0$ tak, że $np_n \underset{n\to\infty}{\longrightarrow} \lambda > 0$, to

$$S_n \xrightarrow[n\to\infty]{d} Y_\lambda,$$

gdzie Y_λ ma rozkład Poissona $\mathcal{P}(\lambda)$.
Inaczej mówiąc, dla dowolnego ustalonego $k \in \mathbb{N}$

$$P(S_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \underset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!} e^{-\lambda} = P(Y_\lambda = k)$$

Dowód:

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!} \cdot \frac{n(n-1)\dots(n-k+1)}{n^k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^n (1-p_n)^{-k} =$$

$$= \frac{1}{k!} \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(\frac{np_n}{1-p_n}\right)^k \left(1 - \frac{np_n}{n}\right)^n \underset{n\to\infty}{\longrightarrow}$$

$$\underset{n\to\infty}{\longrightarrow} \frac{1}{k!} \cdot 1^k \cdot \left(\frac{\lambda}{1}\right)^k e^{-\lambda} = \frac{\lambda^k}{k!} e^{-\lambda}. \square$$

Wniosek: aproksymacja rozkładu Bernoulliego rozkładem Poissona.

Niech S_n oznacza liczbę sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu p, a
Y - zmienną losową o rozkładzie Poissona $\mathcal{P}(\lambda)$ z $\lambda = np$.
Wtedy dla dowolnego zbioru borelowskiego B

$$|P(S_n \in B) - P(Y \in B)| \leqslant \frac{\lambda^2}{n} = np^2.$$

Inaczej mówiąc,

$$P(S_n \in B) \approx P(Y \in B)$$

z błędem, który nie przekracza $\frac{\lambda^2}{n} = np^2$.
W szczególności, dla dowolnego ustalonego $k \in \mathbb{N}$

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} = P(Y = k).$$

(W praktyce przybliżenie powyższe stosuje się, gdy np^2 jest małe,
przy czym $n \geqslant 50$, $p \leqslant 0.1$, $np \leqslant 10$.)

Opracowanie: dr hab. Agnieszka Jurlewicz

Losowanie bez zwracania i losowanie ze zwracaniem:

Załóżmy, że mamy zbiór N elementów, z których M posiada pewną cechę, a pozostałe nie. Losujemy z tego zbioru n elementów. Możemy zrobić to na dwa sposoby:

(a) **ze zwracaniem**
albo
(b) **bez zwracania**.

Oznaczmy przez X ilość elementów posiadających badaną cechę wśród n wylosowanych. Jaki jest rozkład zmiennej losowej X dla obu sposobów losowania?

(a) W przypadku **losowania ze zwracaniem** $X_{(a)}$ to ilość sukcesów w n próbach Bernoulliego z prawdopodobieństwem sukcesu równym $p = \dfrac{M}{N}$ (gdzie sukces jest wtedy, gdy wylosowany element posiada wyróżnioną cechę).
Zatem $X_{(a)}$ ma **rozkład Bernoulliego** $\mathcal{B}\left(n, p = \dfrac{M}{N}\right)$, czyli

$$P(X_{(a)} = k) = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1 - \frac{M}{N}\right)^{n-k}$$

dla $k = 0, 1, \ldots, n$.
Ponadto, $\mathrm{E}X_{(a)} = n\dfrac{M}{N}$, $\mathrm{D}^2X_{(a)} = n\dfrac{M}{N}\left(1 - \dfrac{M}{N}\right)$.

(b) W przypadku **losowania bez zwracania** musi być $n \leqslant N$ i mamy

$$P(X_{(b)} = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

dla $k = 0, 1, \ldots, n$ takich, że $k \leqslant M$ i $n - k \leqslant N - M$.
Zatem $X_{(b)}$ ma tzw. **rozkład hipergeometryczny** z parametrami N, M, n.
Ponadto, $\mathrm{E}X_{(b)} = n\dfrac{M}{N}$, $\mathrm{D}^2X_{(b)} = n\dfrac{M}{N}\left(1 - \dfrac{M}{N}\right)\left(\dfrac{N-n}{N-1}\right)$.

Opracowanie: dr hab. Agnieszka Jurlewicz

FAKT:

Gdy $N \to \infty$ i $M \to \infty$ tak, że $\dfrac{M}{N} \to p$ dla pewnego $0 < p < 1$, to

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \longrightarrow \binom{n}{k} p^k (1-p)^{n-k},$$

czyli

$$X_{(b),N,M,n} \xrightarrow{d} X,$$

gdzie X ma rozkład Bernoulliego $\mathcal{B}(n,p)$.
(Zauważmy, że X ma prawie taki sam rozkład jak $X_{(a)}$, tylko, że $p = \lim \dfrac{M}{N}$ zamiast $\dfrac{M}{N}$.)

DOWÓD:

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{M!}{k!(M-k)!} \cdot \frac{(N-M)!}{(n-k)!(N-M-n+k)!} \cdot \frac{n!(N-n)!}{N!} =$$

$$= \frac{n!}{k!(n-k)!} \cdot \frac{M!}{(M-k)!} \cdot \frac{(N-M)!}{(N-M-n+k)!} \cdot \frac{(N-n)!}{N!} =$$

$$= \binom{n}{k} \frac{(M(M-1)\ldots(M-k+1)) \cdot ((N-M)(N-M-1)\ldots(N-M-n+k+1))}{N(N-1)\ldots(N-n+1)} =$$

(w liczniku w pierwszym iloczynie jest k czynników, w drugim $n-k$;
w mianowniku w iloczynie jest n czynników)

$$= \binom{n}{k} \frac{M}{N} \cdot \frac{M-1}{N-1} \cdot \ldots \cdot \frac{M-k+1}{N-k+1} \cdot \frac{N-M}{N-k} \cdot \frac{N-M-1}{N-k-1} \cdot \ldots \cdot \frac{N-M-n+k+1}{N-n+1} =$$

$$= \binom{n}{k} \left(\frac{M}{N} \cdot \frac{\frac{M}{N} - \frac{1}{N}}{1 - \frac{1}{N}} \cdot \ldots \cdot \frac{\frac{M}{N} - \frac{k-1}{N}}{1 - \frac{k-1}{N}} \right) \cdot \left(\frac{1 - \frac{M}{N}}{1 - \frac{k}{N}} \cdot \frac{1 - \frac{M}{N} - \frac{1}{N}}{1 - \frac{k+1}{N}} \cdot \ldots \cdot \frac{1 - \frac{M}{N} - \frac{n-k-1}{N}}{1 - \frac{n-1}{N}} \right) \longrightarrow$$

$$\longrightarrow \binom{n}{k} p^k (1-p)^{n-k}, \text{ gdy } N \to \infty,\ \frac{M}{N} \to p. \square$$

Opracowanie: dr hab. Agnieszka Jurlewicz

Wniosek (ważny dla teorii statystycznej kontroli jakości):

Przy losowaniu próbki o stosunkowo małej liczności n ze zbioru o wielkiej liczebności N praktycznie jest bez znaczenia, czy losowanie odbywa się ze zwracaniem czy bez, jeśli zagwarantowany jest wybór nieumyślny, przypadkowy. Wówczas bowiem

$$P(X_{(b)} = k) \approx P(X_{(a)} = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

dla $p = \dfrac{M}{N}$. Aby aproksymacja była sensowna wystarczy, że $n < \dfrac{N}{10}$

Przybliżenie rozkładu hipergeometrycznego rozkładem Poissona:

Ze względu na twierdzenie Poissona rozkład hipergeometryczny można przybliżać także rozkładem Poissona.

Mianowicie, gdy $N \to \infty$, $M \to \infty$ tak,że $\dfrac{M}{N} \to 0$, oraz $n \to \infty$ tak,że $n\dfrac{M}{N} \to \lambda > 0$, to

$$P(X_{(b),N,M,n} = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda} = P(Y = k),$$

gdzie Y ma rozkład Poissona $\mathcal{P}(\lambda)$.

Inaczej mówiąc, przy losowaniu bez zwracania dla dostatecznie dużych n i małej proporcji M do N

$$P(X_{(b),N,M,n} = k) \approx \frac{\lambda^k}{k!} e^{-\lambda},$$

gdzie $\lambda = n\dfrac{M}{N}$.

Opracowanie: dr hab. Agnieszka Jurlewicz