# Analýza dat v neurologii

## IX. Poissonovo rozdělení

L. Dušek, T. Pavlík, J. Koptíková
Institut biostatistiky a analýz
Masarykova univerzita, Brno

doc. RNDr. Ladislav Dušek, Dr.
Institut biostatistiky a analýz
Masarykova univerzita, Brno
e-mail: dusek@cba.muni.cz

V minulém díle našeho seriálu jsme se zabývali binomickým rozdělením sloužícím pro popis četnosti výskytu náhodného jevu v *n* nezávislých pokusech. Sledovaný jev má přitom stále stejnou pravděpodobnost (v modelu ji vyjadřuje parametr $\pi$, v odhadu *p*), že nastane. Dále předpokládáme konečný počet opakování experimentu *(n)*, ve kterých sledovaný jev může nastat s četností 0 až *n*. Poissonovo rozdělení je velmi podobné binomickému, neboť rovněž sleduje výskyt jevu v konečném počtu experimentů nebo měření. Binomické rozdělení přechází v Poissonovo při rostoucím $n$ ($n \to \infty$) a klesajícím $p$ ($p \to 0$). Tato aproximace na Poissonovo rozdělení je funkční již při $n > 30$ a $p < 0{,}1$. Proto se o Poissonově rozdělení říká, že modeluje výskyt málo četných nebo vzácných jevů. To ale není zcela přesné, neboť Poissonovo rozdělení lze použít i pro modelování výskytu relativně četných jevů, samozřejmě jsou-li naplněny předpoklady pro jeho aplikaci. U jevů vzácných ale představuje nenahraditelný nástroj.

Strohá definice říká, že náhodná veličina má Poissonovo rozdělení, pokud vyjadřuje počet výskytů málo pravděpodobných jevů v určitém časovém nebo objemovém intervalu. Při praktické výuce lékařů nebo biologů jsme zjistili, že v této definici trochu zaniká přidaná hodnota celého modelu, a proto si zde dovolíme jeho účel poněkud více popsat. U binomického rozdělení (viz VIII. díl našeho seriálu) realizujeme *n* nezávislých pokusů, např. hodů mincí. Základní koncept počítá s tím, že u každého jednotlivého hodu umíme posoudit, zda nastal sledovaný jev. Je lhostejné, zda hodů je 10 nebo 10 milionů, princip zůstává stejný, roste pouze pracnost měření. Začneme-li takto sledovat vzácný jev, opět se princip nemění, jen roste požadavek na nutný počet opakování experimentu, abychom daný jev vůbec uviděli. Nicméně nadále u každé jednotlivé „experimentální jednotky" (hod mincí, měřený objekt, situace) umíme posoudit „individuálně", zda jev nastal nebo ne.

Kvalitativně zcela nová situace nastane ve chvíli, kdy se skutečně naplní $n \to \infty$ a počet jednotlivých experimentů začne být individuálně neměřitelný. Jedinou možností, jak v takové situaci uspět při sledování výskytu jevu, je práce s jinak definovanými širšími experimentálními jednotkami, tedy např. s časovými intervaly, plochou, objemovými intervaly nebo jinak účelově nastavenými celky, ve kterých provádíme sledování (výrobní směna, příjem v nemocnici, sezona…). V takové situaci binomické rozdělení již použít vůbec nelze. Poissonovo rozdělení je ale právě na tyto experimenty nastaveno a sleduje výskyt jevu v časových a prostorových experimentálních jednotkách. Základním cílem je zjistit průměrný výskyt jevu na jednu experimentální jednotku. Jako příklady uvádíme (obr. 1):

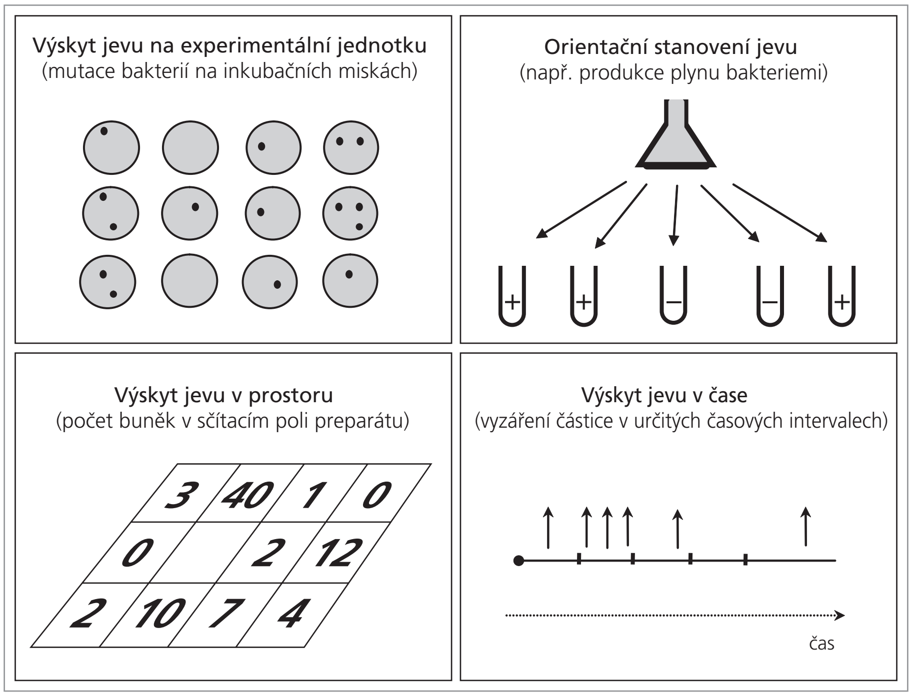
- mutace bakterií měřené klasickým výsevem kolonií na Petriho misku (sledován je průměrný výskyt na 1 misku)
- počet krvinek v poli mikroskopu, kde experimentální jednotkou je buď definovaná plocha políčka, nebo objem hodnocené jamky
- počet žížal vyskytujících se na určité ploše pole (měřeno opakovaným průzkumem vybrané plochy, např. 1 $m^2$)
- počet pooperačních komplikací během určitého časového intervalu po výkonu
- počet částic, které vyzáří zářič za danou časovou jednotku.

Všechny výše uvedené příklady reprezentují stejnou experimentální situaci a postup:

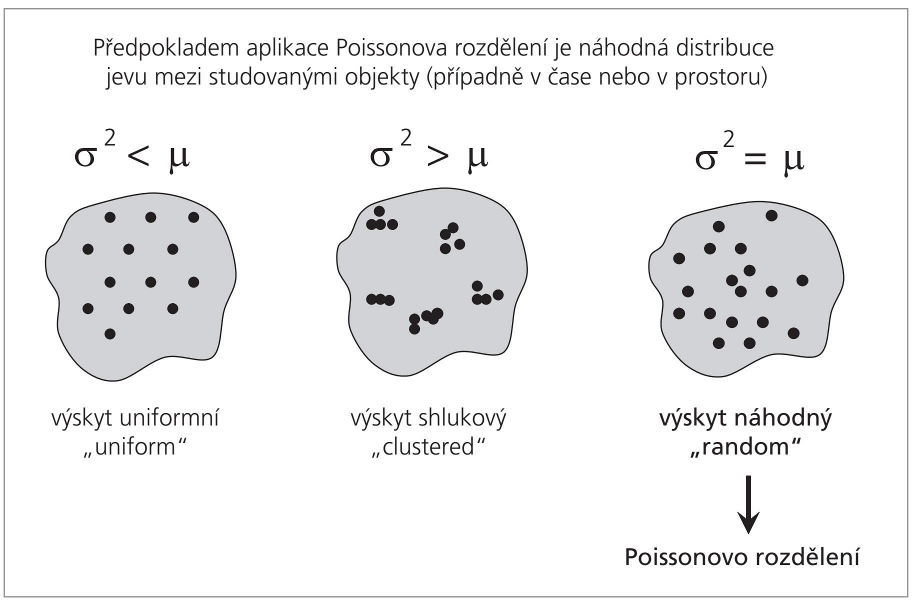
1. Definujeme experimentální jednotku (časový úsek, plocha), kterou následně opakovaně používáme k měření.
2. Za stále stejných podmínek sledujeme výskyt daného jevu v definovaných experimentálních jednotkách (tedy již ne sledování jednotlivých hodů mincí nebo jednotlivých bakterií, sledujeme větší interval nebo časový úsek).
3. Ze zaznamenaných počtů jevů v experimentálních jednotkách počítáme střední výskyt jevu na 1 experimentální jednotku. Jelikož ta má stanovenou měřitelnou velikost, lze střední výskyt jevu přepočítat na větší celky nebo jinak zobecnit.

Poissonovo rozdělení slouží k odhadu středního výskytu jevu a následně k modelování pravděpodobnosti, že jev se vyskytne v nějak určeném počtu (např. právě jednou, více než 5krát apod). Tak jako i jiná rozdělení pravděpodobnosti, má Poissonovo rozdělení nutné předpoklady pro svou aplikaci:

- výskyt jevu je zcela náhodný (tedy náhodný v čase nebo prostoru podle typu situace)
- výskyt jevu v konkrétní experimentální jednotce nijak nezávisí na tom, co se stalo v jiných jednotkách
- není možné, aby 2 nebo více jevů nastaly současně, přesně ve stejném místě prostoru nebo ve stejném časovém okamžiku




Obr. 1. Příklady praktického využití Poissonova rozdělení.




Obr. 2. Poissonovo rozdělení je model sledující náhodný výskyt jevu.

- pro každý dílčí časový okamžik, prostorovou jednotku nebo jinak definovanou jednotku je pravděpodobnost výskytu jevu stejná.

Poissonovo rozdělení je tedy modelem pro výskyt jevů, které se náhodně vyskytují v čase nebo prostoru s neměnnou pravděpodobností. Počítáme-li např. mutované kolonie na Petriho misce, předpokládáme, že pravděpodobnost mutace každé jednotlivé bakterie v průběhu experimentu je stejná a výskyt mutací je zcela náhodný. Obr. 2 ukazuje alternativy k náhodnému výskytu, tedy výskyt rovnoměrný nebo shlukový, což jsou situace, na které musí být použit jiný stochastický model.

Jak již bylo řečeno, cílem hodnocení je odhad střední hodnoty počtu jevů (událostí) na danou jednotku. V Poissonově rozdělení střední hodnotu počtu jevů X označujeme jako λ. Parametr λ je jediným parametrem pravděpodobnostní funkce P (X), která je popsána a graficky znázorněna na obr. 3. Je zajímavé, že proměnná X s Poissonovým rozdělením má střední hodnotu rovnu rozptylu: $E(X) = Var(X) = \mu = \sigma^2 = \lambda$ (obr. 2).

Obr. 3 jasně dokládá, že čím větší je λ, tím více se tvar Poissonova rozdělení blíží rozdělení normálnímu. Tímto potvrzujeme, že Poissonovo rozdělení může být využito i pro modelování výskytu častých jevů. Pokud k Poissonovu rozdělení vede aproximace binomického rozdělení při $n \to \infty$ a $p \to 0$, pak hodnota $\lambda = np$. Z toho je opět patrný číselný význam parametru λ, který vyjadřuje střední počet jevů v dané experimentální situaci.

Hlavní pragmatickou hodnotu Poissonova rozložení asi nejlépe přiblíží konkrétní příklad. Aplikujme Poissonovo rozdělení na sledování určité komplikace u pacienta v čase (náhodná veličina X) s tím, že základním časovým intervalem pro měření je 1 h. Během opakovaných měření v tomto intervalu jsme zjistili, že střední výskyt sledované komplikace je $\lambda = 1$. Jsou-li splněny všechny předpoklady Poissonova rozdělení, můžeme využít jeho pravděpodobnostní funkci k modelovým výpočtům, které umožní odhadnout pravděpodobnost různých potenciálně rizikových situací. Uveďme jako příklad 2 takové otázky, přičemž pro výpočet využijeme vztah uvedený na obr. 3:

1. Jaká je pravděpodobnost, že veličina X nabude hodnoty x = 1?

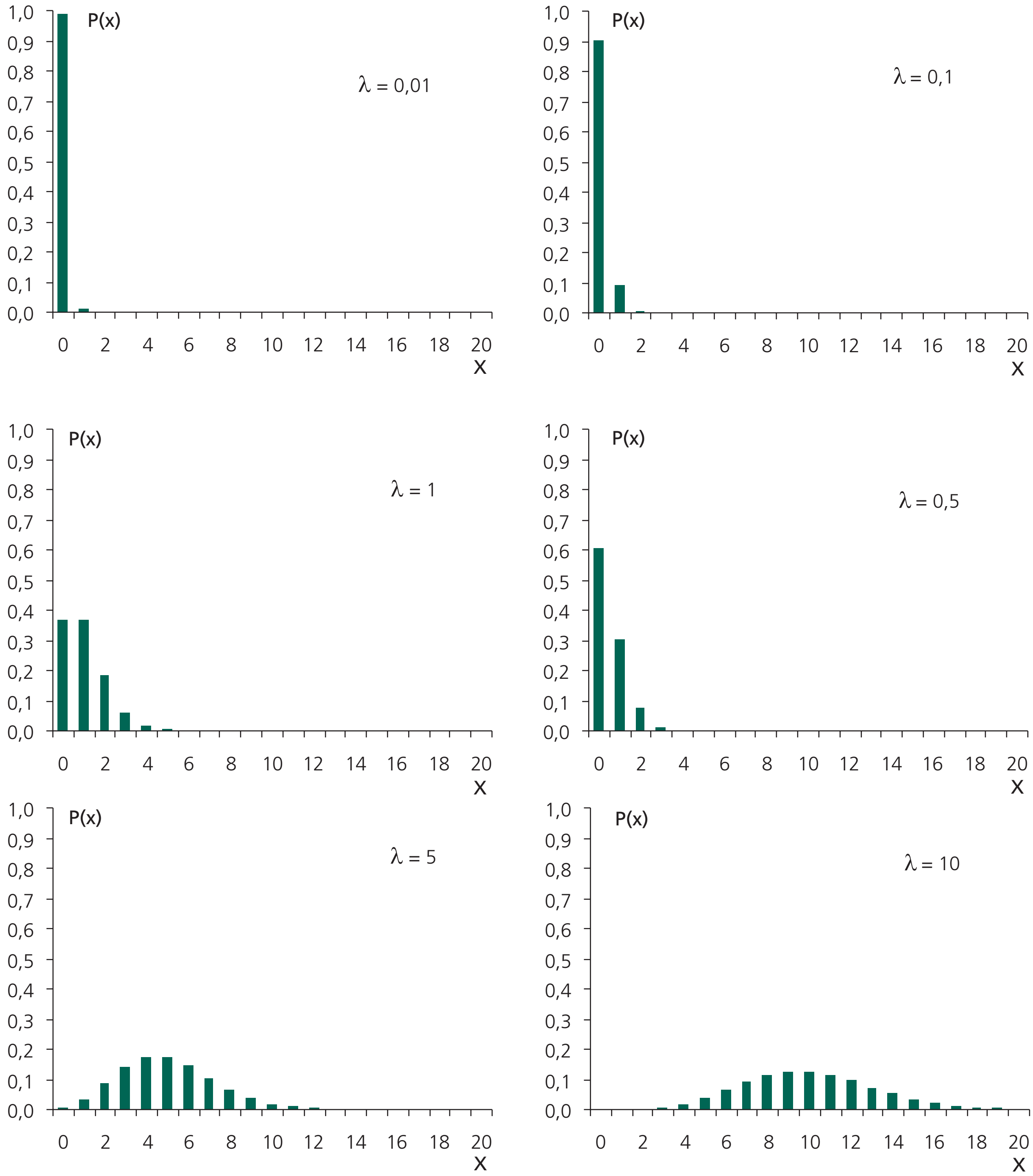
$P(x = 1) = (\lambda^1/1!)/e^{\lambda} = 0{,}3679$

2. Jaká je pravděpodobnost, že veličina X nabude hodnoty alespoň 2, tedy $x \geq 2$?

Pravděpodobnost $P(x \geq 2)$ je rovna součtu pravděpodobností
$P(x = 2) + P(x = 3) + P(x = 4) + \dots$,
lépe ji ale spočítáme jako
$P(x \geq 2) = 1 - P(x = 0) - P(x = 1)$:
$P(x = 0) = (\lambda^0/0!)/e^{\lambda} = 0{,}3679$
$P(x = 1) = 0{,}3679$ (viz bod 1)
potom
$P(x \geq 2) = 1 - 0{,}3679 - 0{,}3679 = 0{,}2642$
Příklad ukazuje, že pro
$\lambda = 1$ je $P(x = 0) = P(x = 1)$.

$$\mathrm{P}(x) = e^{-\lambda} \times \frac{\lambda^x}{x!}$$

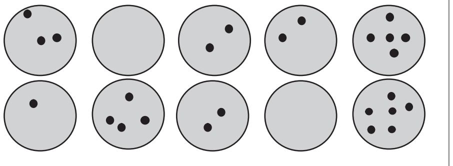
kde x je sledovaná náhodná veličina (počet výskytů jevu v experimentální jednotce – časový interval, plocha, objemová jednotka) a λ je parametr rozdělení vyjadřující střední výskyt jevu na 1 experimentální jednotku




Obr. 3. Poissonovo rozdělení jako model.

S parametrem λ samozřejmě není nutné pracovat pouze jako s bodovým odhadem. Bodový odhad můžeme doplnit i intervalem spolehlivosti tak, jak jsme jej dříve dokumentovali například pro odhad aritmetického průměru. Příklad takového výpočtu intervalu spolehlivosti ukazuje obr. 4.

Příklad: Pokus sledující výskyt nějak změněných kolonií bakterií na 10 Petriho miskách. Sledovaný jev nastal v experimentu celkem 25krát. Bodový odhad parametru λ (průměrný výskyt jevu na experimentální jednotku, tedy misku) je λ = 2,5.



$\bar{x} \approx \lambda = 25/10 = 2{,}5$

95% interval spolehlivosti pro λ:

$$\boxed{\bar{x} - Z_{1-\alpha/2} \times \sqrt{\frac{\bar{x}}{n}} \leq \lambda \leq \bar{x} + Z_{1-\alpha/2} \times \sqrt{\frac{\bar{x}}{n}}}$$

$$2{,}5 - 1{,}96 \times \sqrt{0{,}25} \leq \lambda \leq 2{,}5 + 1{,}96 \times \sqrt{0{,}25}$$

$$1{,}52 \leq \lambda \leq 3{,}48$$

**Obr. 4. Výpočet intervalu spolehlivosti pro parametr λ Poissonova rozdělení.**

Závěrečná poznámka musí patřit samotnému názvu Poissonova rozdělení, neboť ten není tak „anonymní“ jako u rozdělení binomického. Rozdělení je pojmenováno po významném francouzském matematikovi a fyzikovi Poissonovi. Siméon-Denis Poisson žil v letech 1781 až 1840 a za svůj život toho stihnul opravdu hodně. Byl profesorem na École Polytechnique v Paříži, členem Francouzské akademie věd a Petrohradské akademie věd. Zabýval se matematickou analýzou, řešením diferenciálních rovnic i teorií pravděpodobnosti. Zavedl pojem „zákon velkých čísel“. Kromě toho řešil pohybové rovnice v mechanice, významně přispěl k poznání magnetických a elektrických jevů a zabýval se teorií pružnosti a tepla. Nezbývá než konstatovat, že Poissonovo rozdělení má více než důstojný původ svého názvu.