1 Nějaké ty úvodní definice

X je náhodná veličina, $[X < b]$ je rychlé označení pro všechny prvky Ω, že $X(\omega) < b$.[1]

1.1 Očekávaná střední hodnota

Očekávaná střední hodnota neboli $\mathsf{E}X$ je jinak definovaná u diskrétních a u spojitých rozdělení - u obou se ale dá vnímat jako jakýsi *průměr*, jak už název napovídá.

U diskrétních znamená součet řady (pokud konverguje, což nemusí)

$$\sum_{i=1}^{\infty} x_i \mathsf{P}[X = x_i],$$

kde x_i jsou všechny možné hodnoty X.

U spojitých je očekávaná střední hodnota podobná - je

$$\int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x,$$

kde f_X je hustota (viz o něco níž)[5].

Platí pěkně $\mathsf{E}(a + bX) = a + b\mathsf{E}X$, $\mathsf{E}(X + Y) = \mathsf{E}X + \mathsf{E}Y$.

Občas se hodí uvědomit si, že $\mathsf{E}(g(X, Y)) = \sum_{i=1}^{\infty} \sum_{i=1}^{\infty} (x_i y_j) \mathsf{P}[X = x_i, Y = y_j]$. Tedy, pokud jsou nezávislé, platí třeba $\mathsf{E}\,(XY) = (\mathsf{E}X)\,(\mathsf{E}Y)$.

1.2 Rozptyl

Máme $\mathsf{E}X$, ale to nám stejně neřiká tolik, chceme vědět, o kolik se v průměru od $\mathsf{E}X$ lišíme. Stačí nám udělat $\mathsf{E}(X - \mathsf{E}X)$?

Nestačí! Vyšlo by nám to totiž stejně vždycky 0! Proto definujeme rozptyl $\mathsf{var}X$ jako

$$\mathsf{var}X = \mathsf{E}\,(X - \mathsf{E}X)^2\,.$$

Bacha na uzávorkování - neni to $(\mathsf{E}\,(X - \mathsf{E}X))^2$, to by byla zase nula, ale $\mathsf{E}\left((X - \mathsf{E}X)^2\right)$.

Platí třeba

$$\mathsf{var}X = \mathsf{E}X^2 - (\mathsf{E}X)^2$$

nebo taky $\mathsf{var}(a + bX) = b^2 \mathsf{var}X$.

1.3 Borelovská σ algebra

Tohle je taková spíš pomocná definice, kterou pak už nikde nepoužiju, ale mám jí tu, aby to dávalo smysl. Trochu.

Borelovská σ algebra je nejmenší algebra, co obsahuje mj. i otevřené intervaly na $\mathbb{R}$, a platí tam nějaká pravidla, co se mi sem nechce psát. Důležité je, že členy téhle algebry jsou *množiny na* $\mathbb{R}$.

1.4 Rozdělení pravděpodobnosti

Rozdělení pravděpodobnosti je funkce *z* $b\sigma a$ *do* $\mathbb{R}$, definovaná jako

$$\mathsf{P}_X(B) = \mathsf{P}\left(X^{-1}(B)\right),$$

kde druhé P je pravděpodobnost na „původním“ Ω. Vyjadřuje tedy *pravděpodobnost, že* X *padne do nějaké množiny* - často intervalu. Tohle rozdělení ale tak často nekonstruujeme přímo, stačí nám distribuční funkce.

[1] Ty hranaté závorky jsou tu víceméně jenom proto, že je to „čistčí“ a že je to v učebnici, ze které opisuju[2] , nic víc.

[2] Karel Zvára a Josef Štěpán - pravděpodobnost a matematická statistika[3]

[3] `Btw, kdo jste četl House of Leaves?`[4]

[4] `Nikdo? Škoda :/`

[5] Je to asi definice kruhem nebo tak něco, ale o nic jako o koherentní výklad se tu nesnažim, žejo

1.5 Distribuční funkce

Distribuční funkce $F_X(x)$, někdy také jenom $F(x)$, je funkce $z\ \mathbb{R}$ *do* $\mathbb{R}$, definována

$$F_X(x) = \mathsf{P}[X < x].$$

Znamená pravděpodobnost, že *náhodná veličina je ostře menší než nějaké* x. (ty závorky jsou opravdu hranaté, viz 0. odstavec)

1.6 Hustota

Máme tedy distribuční funkci, co pro každé x dá pravděpodobnost, že $X < x$. Občas ale chceme vidět „rozložení pravděpodobnosti intervalů“ tak nějak „příměji“ - k tomu slouží hustota.

Náhodná veličina X má hustotu f_X, pokud

$$F_X(x) = \int_{-\infty}^{x} f_X(T)\mathrm{d}t.$$

To znamená, že požadavek *plocha pod funkcí nad intervalem = pravděpodobnost, že* X *je v intervalu* není vlastně důsledkem nějaké komplikované definice, ale je v podstatě[6] přímo *definicí*.

Všechna spojitá rozdělení hustotu mají; naopak žádná diskrétní nemají.

1.7 Kvantilová funkce

Kvantilová funkce je jakoby inverzní k distribuční funkci. Je definována jako

$$F^{-1}(u) = \inf\left\{x \in \mathbb{R} : F(x) \geq u\right\},$$

samozřejmě pro $u \in (0, 1)$.

Znamená to tedy *nejmenší* (a někdy i jedinou) *hodnotu* h *takovu, že* $\mathsf{P}[X \in (-\infty, h)]$ *je stejná nebo větší než daná pravděpodobnost.*

Občas se používá termín *kritická hodnota*, ale já přesně nechápu, co to znamená.

1.8 Kovariance

$$\mathsf{cov}(X, Y) = \mathsf{E}(X - \mathsf{E}X)(Y - \mathsf{E}Y) = \mathsf{E}(XY) - (\mathsf{E}X)(\mathsf{E}Y)$$

Kovariance určitým způsobem určuje „závislost“ dvou náhodných veličin. Platí, že $\mathsf{cov}(X, X) = \mathsf{var}X$, naopak pro nezávislé X, Y je $\mathsf{cov}(X, Y) = 0$.

1.9 Korelace

Korelace je normovaná kovariance. Přesněji:

$$\rho_{X,Y} = \frac{\mathsf{cov}(X, Y)}{\sqrt{\mathsf{var}X\mathsf{var}Y}}$$

Platí takové hezké vlastnosti - $|\rho_{X,Y}| \leq 1$, $\rho_{X,X} = 1$ a nakonec pro nezávislé X a Y platí $\rho_{X,Y} = 0$.

2 Diskrétní rozdělení

Stačí nám reálná čísla $x_1, x_2, \ldots$ a k nim $\mathsf{P}[X = x_i]$ takové, že $\sum_{i=1}^{\infty} \mathsf{P}[X = x_i] = 1$. Pak distribuční funkce $F(x) = \sum_{i:x_i<x} \mathsf{P}[X = x_i]$, tj. součet všech menších.

2.1 Binomické rozdělení

Máme n nezávislých pokusů, zdar s pravděpodobností p, nezdar tedy $1 - p$. X je definováno jako počet zdarů. Potom platí, že

$$\mathsf{P}[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Zapisujeme jako $X \sim \mathsf{bi}(n, p)$. Ještě jednou - n je počet pokusů, p pravděpodobnost zdaru. Platí $\mathsf{E}X = np$, $\mathsf{var}X = np(1 - p)$.

[6] Píšu „v podstatě“, protože v definici je jenom interval mínus nekonečno až x, ale ten zbytek je vlastně strašně triviální

2.2 Poisonnovo rozdělení

Mějme binomická rozdělení $X_1, X_2, ...$, nechť $X_n \sim \mathsf{bin}(n, p_n)$ a limita násobku np_n je nějaké λ. Potom existuje rozdělení X, ke kterému se tato rozdělení „blíží“. Přesněji:

$$\forall k \in \mathbb{N} : \lim_{n\to\infty} \mathsf{P}[X_n = k] = \frac{\lambda^k}{k!} e^{-\lambda} = \mathsf{P}[X = k].$$

Bez větších překvapení toto rozdělení nazveme Poissonovo.[7] Napíšeme $X \sim \mathsf{Po}(\lambda)$.

Řečeno jinak - binomické rozdělení $\mathsf{bi}(n, p)$ můžeme při velkém počtu n a malé pravděpodobnosti p aproximovat jako $\mathsf{Po}(np)$.

Platí $\mathsf{E}X = \lambda$, $\mathsf{var}X = \lambda$.

2.3 Geometrické rozdělení

Máme n nezávislých pokusů, zdar s pravděpodobností p, nezdar tedy $1 - p$. X je definováno jako počet nezdarů před prvním zdarem. Potom

$$\mathsf{P}[X = k] = p(1 - p)^k.$$

Tohle rozdělení je fairly easy. Občas se objaví „naopak“ - X je číslo prvního zdaru. Sebecvičení - je to potom $k+1$, nebo $k - 1$?

2.4 Negativně binomické rozdělení

Jde o zobecnění předchozího. Máme n nezávislých pokusů, zdar s pravděpodobností p, nezdar tedy $1 - p$. X je definováno jako počet nezdarů před r zdary. Potom

$$\mathsf{P}[X = k] = \binom{k + r - 1}{r - 1} p^r (1 - p)^k.$$

Píšeme také $X \sim \mathsf{NB}(r, p)$ - ještě jednou, r je počet zdarů, co musí nastat, p je pravděpodobnost zdaru.

2.5 Hypergeometrické rozdělení

Máme N koulí, A z nich je bílých, vytáhneme n předmětů. X je počet bílých vytáhlých. Platí

$$\forall \max(0, A + n - N) \le k \le \min(A, n) : \mathsf{P}[X = k] = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}$$

Nazveme to hypergeometrické rozdělení.

3 Spojitá rozdělení

Stačí je napsat jako vzoreček buď pro hustotu, nebo distribuční funkci.[8] Budu psát obé.

3.1 Rovnoměrné rozdělení

Něco má na nějakém intervalu všude „stejnou“ pravděpodobnost, na zbytku ale nic.

$$f(x) = \begin{cases} 0 & x \le a \\ \frac{1}{b-a} & a < x < b \\ 0 & x \ge b \end{cases}$$

$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$

Taky platí $\mathsf{E}X = \frac{a+b}{2}$, $\mathsf{var}X = \frac{(b-a)^2}{12}$.

[7] Ve skutečnosti je definice „obráceně“, ale mně se to takhle líbí víc.

[8] ačkoliv např. F_X hustotu jednoznačně nedává - ale spojitou dává jenom jednu, mám ten pocit nejistý

3.2 Exponenciální rozdělení

Máme náhodný jev, co se vyskytuje v náhodných okamžicích, výskyty v nepřekrývajících se časových intervalech jsou nezávislé. Víme, že za časovou jednotku se průměrně stane λ-krát. X označíme čas čekání na první výskyt jevu. Platí:

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

Píšeme $X \sim \mathsf{ex}(\lambda)$. Taky platí $\mathsf{E}X = \frac{1}{\lambda}$, $\mathsf{var}X = \frac{1}{\lambda^2}$.

Exponenciální je jediné spojité rozdělení, co takzvaně *nemá paměť'* - znamená to, že, pokud si například představíme X jako dobu dožití produktu, tak v každém okamžiku t platí, že $\mathsf{P}(X > t + n | X > t) = \mathsf{P}(X > n)$[9], tj. pořád mám stejnou pravděpodobnost, že se dalšího dne dožiju.

3.3 Normální rozdělení

Normální rozdělení je „vtipné“ v tom, že distribuční funkci sice má, ale není jí možno vyjádřit jinak, než jako integrál.

Normované normální rozdělení je takové, že hustota, zhusta označovaná $\varphi(x)$, je

$$f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}.$$

Distribuční funkce je definovatelné pouze jako

$$F(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \mathrm{d}t.$$

Nejde ho definovat jinak, ale jde ho tabelovat (tj. některé její hodnoty jsou známy) nebo aproximovat.

Mějme teď čísla $\mu \in \mathbb{R}$, $\sigma > 0$. Náhodnou veličinu Y definujeme jako $Y = \mu + \sigma X$, kde X má normované normální rozdělení. Potom

$$F_Y(x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t,$$

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Píšeme $Y \sim \mathsf{N}(\mu, \sigma^2)$ - normované normální rozdělení je tedy $X \sim \mathsf{N}(0, 1)$.

Pro $X \sim \mathsf{N}(\mu, \sigma^2)$ platí pěkné $\mathsf{E}X = \mu$, $\mathsf{var}X = \sigma^2$.

Platí taky, že kvantil normovaného rozdělení $\Phi^{-1}(0,975) = 1,96$, kvůli symetrii kolem nuly potom tedy platí $\mathsf{P}(|X| < 1,96) = 0,95$ pro $X \sim \mathsf{N}(0, 1)$.

4 Důležité asymptotické věty

4.1 Čebyševova nerovnost

Mějme $\mathsf{var}X < \infty$, nechť $\varepsilon > 0$. Pak platí jednoduché

$$\mathsf{P}\left[|X - \mathsf{E}X| \geq \varepsilon\right] \leq \frac{\mathsf{var}X}{\varepsilon^2}.$$

4.2 iid

V dalších dvou větách se objevují **nezávislé, stejně rozdělené náhodné veličiny** neboli taky **iid**, kterých je ještě k tomu nekonečně mnoho. Jak si to přesně představit?

Mějmě nekonečně hodů kostek, každý je se stejnou pravděpodobností. Pak si můžu jednoduše definovat výsledek 1. hodu jako náhodnou veličinu X_1, výsledek 2. hodu jako X_2 atd. Každý hod je určitě nezávislý, stejně rozdělený, a je jich nekonečně mnoho.

[9] Tady jsem přestal používat tu [hranatou] notaci, co jsem doteď poměrně důsledně dodržoval, ale ona nedává v podmíněných pravděpodobnostech moc smysl. Nějak si to dodefinujte sami :)

4.3 Zákon velkých čísel

Mějme iid $X_1, X_2, ...$ s konečnou střední hodnotou $\mathsf{E}X_n = \mu$. Potom

$$\mathsf{P}\left(\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j = \mu\right\}\right) = 1.$$

Například: Ω jsou všechny možné nekonečné sekvence hodů kostkou (viz výše), $\mathsf{E}X_n$ je vždy $3,5$. Část za dvojtečkou říká, že beru čím dál tím víc hodů kostek, ze kterých dělám průměr. A celá ta věta říká poměrně překvapivě, že *s pravděpodobností 1 budu házet tak, že průměr všech hodů bude* ***přesně*** *3,5*. Nekonečno dokáže divy.

4.4 Centrální limitní věta

Mějme $Y_1, Y_2, ...$ iid, o kterých navíc platí, že $\mathsf{E}Y_i = \mu, \mathsf{var}Y_i = \sigma^2 > 0, \mathsf{E}\left|Y_i\right|^3 < \infty$.

Pro distribuční funkci náhodné veličiny

$$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Y_i - \mu}{\sigma}\right)$$

platí

$$\lim_{n\to\infty} \mathsf{P}\left[S_n < x\right] = \Phi(x),$$

tedy že normovaný průměr se blíží k $\mathsf{N}(0,1)$.

Ten první vzoreček není až tak hrozný, dá se odvodit z postupu normalizace a toho, že S_n je průměr.