Pravděpodobnostní prostor $(\Omega, \mathcal{A}, P)$ = (množina všech náhodných elem. jevů, σ-algebra,pravděpodobnost)
Základní vlastnosti:

1. $0 \leq P(A) \leq 1$
2. $P(A^c) = 1 - P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $A \subset B \to P(A) \leq P(B)$
5. $A \subset B \to P(B - A) = P(B) - P(A)$

Podmíněná pravděpodobnost - jev A za podmínky jevu B (např. pst, že padne 6 za podmínky, že padlo sudé číslo), chová se jako nepodmíněná pst.
$P(A|B) = \frac{P(A \cap B)}{P(B)} \to$ průnik lze hledat např. skrz tabulku či obrázek (případně u nezávislých jevů skrz vzorec níže)
Věta o násobení psti (řetězové pravidlo) - $P(\bigcap_{i=1}^{n}) = P(A_1)P(A_2|A_1)P(A3|A_1 \cap A_2)...P(A_n|A_1 \cap ... \cap A_{n-1})$
Věta o úplné psti - známe $P(B|A_i) \to P(B) = \sum_{i=0}^{\infty} P(A_i)P(B|A_i)$, např. pst, že nám během roku praskne nějaký z X typů žárovek
Bayesova věta - známe $P(B|A_i) \to P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}$; např. jaká je pst, že člověk s pozitivním testem je opravdu nakažený, když máme danou spolehlivost - ideálně řešit skrz tabulku
Nezávislost jevů - dva jevy jsou nezávislé, pokud platí $P(A \cap B) = P(A)P(B)$, např. při prvním hodu padne panna a při druhém orel. Pro více jevů je potřeba pro totální nezávislost dokázat nezávislost všech n-tic pro n$\geq$ 2. Pokud pro každou dvojici jevů platí, že jsou jevy nezávislé, pak jsou obecně po dvou nezávislé.

Náhodná veličina X - zobrazení $X : \Omega \to \mathbb{R}$, které přiřazuje jevům náhodná čísla (lze na to pohlížet jako na náhodné číslo); operace s náhodnými veličinami vrací náhodné veličiny
Distribuční funkce - $F(x) = P(X \leq x)$, např. jaká je šance, že budeme na bus čekat méně jak x minut. Je neklesající a zprava spojitá v každém bodě. Platí, že $\lim_{x \to -\infty} F(x) = 0$ a $\lim_{x \to \infty} F(x) = 1$.
Tabulka základních informací o diskrétních, spojitých a směsích n. v.:

Diskrétní	Spojitá	Směs $Mix_c(D, S)$
F(x) je skokovitá, velikost skoků odpovídá jejich pravděpodobnostem, $F(x) = \sum_i P(X = x_i)$	F(x) je spojitá a její derivace je f(x), tj. $F(x) = \int_{-\infty}^{x} f(t)dt$	F(x) je směs dílčích distribučních funkcí F_D (s "váhou"c) a F_S, tj. $F(x) = cF_D(x) + (1 - c)F_S(x)$
	f(x) - hustota pravděpodobnosti, $\int_{-\infty}^{\infty} f(x)dx = 1$	
$\mathbb{E}X = \sum_{i=1}^{\infty} x_i \cdot P(X = x_i)$	$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx$	$\mathbb{E}X = c\mathbb{E}_D + (1 - c)\mathbb{E}S$
$\mathbb{E}X^2 = \sum_{i=1}^{\infty} x_i^2 \cdot P(X = x_i)$	$\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 f(x)dx$	

Základní vlastnosti střední hodnoty $\mathbb{E}X$:

1. a je konst. $\to \mathbb{E}a = a$
2. $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$
3. $X_1 \leq X \leq X_2 \to \mathbb{E}X_1 \leq \mathbb{E}X \leq \mathbb{E}X_2$

Variance - česky rozptyl, $varX = \mathbb{E}(X - \mathbb{E}X)^2$, **Kovariance** - $cov(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$
Základní vlastnosti variance $varX$ **a kovariance** $cov(X, Y)$:

1. X je n.v. $\to varX = \mathbb{E}X^2 - \mathbb{E}^2X = cov(X, X)$
2. a je konst. $\to var(a) = 0$
3. X je n.v., a je reálné číslo $\to var(aX) = a^2 varX$
4. X,Y $\to var(X + Y) = varX + varY + 2cov(X, Y)$
5. X,Y jsou n.v. $\to cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$

Čebyševova nerovnost - X je n.v. a pro každé $\mathcal{E} > 0$ platí, že $P(|X = \mathbb{E}X| \geq \mathcal{E}) \leq \frac{varX}{\mathcal{E}^2}$; např. odhadněte, že při 120 hodech padne 10-15 šestek (tohle jsou zrovna prý dost blbá čísla, ale princip snad jde poznat :D)

Alt(p) - model pro "úspěch/neúspěch"; např. šestka padne v hodu s pravděpodobností p
Binom(n,p) - model pro "počet úspěchů v n nezávislých situacích"; např. počet šestek, které padají s pravděpodobností p, v n hodech
Po(λ) - model pro "počet vzájemně nezávislých událostí v intervalu"; např. počet prasklých žárovek v průběhu měsíce; λ vztahujeme k našemu časovému intervalu
Ge(p) - model pro "počet neúspěchů před prvním úspěchem"; např. počet hodů kostkou než padne šestka
HypGe(N,K,n) - model pro "vybíráme z hromady N předmětů, ze kterých má K předmětů specifickou vlastnost, celkem n předmětů"; např. z klobouku, kde je 10 kuliček, z toho 3 černé, vybíráme 6
Ro(a,b) - model pro "dobu čekání na událost, která přichází v pravidelných intervalech"; např. doba, kterou od příchodu budeme čekat na bus, co jezdí každých 10 minut
Exp(λ) - model pro "dobu čekání na události, které jsou navzájem nezávislé"; např. doba, za kterou praskne další žárovka; λ lze označi za intenzitu (prasknou obvykle 3 žárovky za měsíc $\to \lambda = 3$)
Norm(μ, σ^2) - v realitě často se vyskytující rozdělení, převod na normované pomocí $Y = \frac{(X-\mu)}{\sigma}$

Tabulka jednotlivých modelů rozdělení:

P_x	*Hodnoty X*	$P(X=k)$ nebo $f(x), F(x)$	$\mathbb{E}X$	$varX$
Alt(p)	0 nebo 1	$p^k(1-p)^{1-k}$	p	$p(1-p)$
Binom(n,p)	$<0,n>$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Po(λ)	$<0,\infty)$	$\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ
Ge(p)	$<0,\infty)$	$p(1-p)^k$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
HypGe(N,K,n)	<max$(0,n+K-N)$, min(n,K)>	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$n\frac{K}{N}$	$n\frac{K}{N}(1-\frac{K}{N})\frac{N-n}{N-1}$
Ro(a,b)	$<a,b>$	$\frac{1}{b-a}$, $\frac{x-a}{b-a}$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exp(λ)	$(0,\infty)$	$\lambda e^{-\lambda x}$, $1-e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Norm(μ,σ^2)	$(-\infty,\infty)$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, Φ	μ	σ^2

Konvoluce

KAT throwback, ale tak kdyby to náhodou po nás ve zkoušce chtěla, hodí se mít všechno, žejo.

Diskrétní - F,G distribuční fce nezávislých n.v. s pstmi p_n, respektivě q_n. Pak pro H = F*G (distribuční fci Z = X+Y) platí, že:

$H = \sum h_n$, kde $h_n = \sum p_k \cdot q_(n-k)$

Spojité - X,Y jsou nezávislé n.v. s hustotou psti $f(x)$, respektivě $g(x)$. Pro Z = X+Y s hustotou psti $h(z)$ platí, že:

$h(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$

Tldr. pro jednotlivá rozdělení:

Alt(p)$\rightarrow$Binom(2,p)	Binom($n_{1,2}$,p) $\rightarrow$ Binom(n_1+n_2,p)	Po($\lambda_{1,2}$)$\rightarrow$Po($\lambda_1+\lambda_2$)
Ro(a,b/c,d) - moc textu,dopiš ručně	N($\mu_{1,2}$,$\sigma^2_{1,2}$)$\rightarrow$N($\mu_1+\mu_2$,$\sigma_1^2+\sigma_2^2$)	Exp(λ) $\rightarrow$ h(z) = $\lambda^2 ze^{-\lambda z}$

Náhodný vektor- máme n.v. X_1,X_2,...,X_n definované na pstním prostoru. Náhodný vektor (dál jako n.vec.) pak je $\mathbb{X} = \{X_1, X_2, ..., X_n\}^T$

Má sdruženou distribuční funkci, kde $F_{\mathbb{X}}(x)$ říká, že platí $x_k^i \leq x^i$ pro všechny i (kde i = 0,1,...,n). Obdobně existuje sdružená hustota (integrujeme přes všechny x_i).

Marginální rozdělení - rozdělení podvektoru n.vec., tj. "vyščítání podsituací"(např. pokud máme složky XYZ, tak zkoumáme situace, kde X = .., bez ohledu na ostatní složky))

Vlastnosti n.vec.

Vektor středních hodnot - $\mathbb{E}\mathbb{X} = (\mathbb{E}X_1, ..., \mathbb{E}X_n)^T$

Variační matice - Var$\mathbb{X}$ s prvky $cov(X,Y) = \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) \rightarrow$ Var$\mathbb{X}$ = $\begin{bmatrix} varX & cov(X,Y) \\ cov(X,Y) & varY \end{bmatrix}$

Korelační matice - Corr$\mathbb{X}$ s prvky $corr(X,Y) = \frac{cov(X,Y)}{\sqrt{varX}\sqrt{varY}}$; na diagonále má vždycky 1

Nezávislost n.v. - podobné jako u jevů, akorát zkoumáme distribuční fce, tj. $F_{X_{i1}...X_{ir}}(x_{i1}...x_{ir}) = F_{X_{i1}}(x_{i1}) \cdot ... \cdot F_{X_{ir}}(x_{ir})$

Diskrétní - testujeme $P(X_1 = x_1, ..., X_n = x_n) = P(X_1 = x_1) \cdot ... \cdot P(X_n = x_n)$

Spojité - testujeme $f_{\mathbb{X}}(x_1, ..., x_n) = f_{X_1}(x_1) \cdot ... \cdot f_{X_n}(x_n)$

Centrální limitní věta

$Z_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}}$, $n = 1, 2, ...$ potom $\lim_{n\to\infty} F_n(x) = \Phi(x)$ (Tohle je nutné dohledat v tabulce :))

Přehled jednotlivých statistických rozdělení:

Chí-kvadrát rozdělení χ^2_n - pokud máme nezávislé n.v. $X_1, ..., X_n$ s rozdělením N(0,1), pak n.v. $Y = \sum X_i^2$ má rozdělení χ^2_n, kde n je počet stupňů volnosti

Studentovo t-rozdělení t_n - pokud máme n.v. X s rozdělením N(0,1) a na ní nezávislo n.v. Y s rozdělením χ^2_n, pak n.v. $Z = \frac{X}{\sqrt{Y}}\sqrt{n}$ má t_n rozdělení, kde n je počet stupňů volnosti

Fisherovo-Snedecorovo rozdělení $F_{n,m}$ - pokud máme nezdávislé n.v. U a V s rozdělením χ^2_n (χ^2_m), pak n.v. $W = \frac{U/n}{V/m}$ má $F_{n,m}$ roydělení, kde n a m jsou parametry

Výběrový průměr $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **Výběrový rozptyl** $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$
Kvantil - pro kvantil z_β platí $F(z_\beta) = \beta$
Kvartil - kvantily pro $\beta = \frac{i}{4}$; dle hodnoty i máme 1., 2. a 3. kvartil (2. kvartil se označuje za medián)
Modus - nejčastěji zastoupená hodnota ve vzorku

Empirická distribuční funkce - podobně jako klasická distribuční fce, pouze místo n.v. máme realizaci náhodného výběru.
$F_{emp}(x) = \frac{\#\{x_i, x_i < x\}}{n}$, # je počet prvků
Když ji zakreslujeme, tak na osu x umisťujeme data a na osu y hodnotu F_{emp}. Opět platí, že je vždy zprava spojitá.
Histogram - aneb jeden ze způsobů jak získat hrubý odhad o např. rozdělení dat
Jak zanést data do histogramu? Rozdělíme si je do intervalů (např. prvky od 0 do 1 tvoří jeden interval), které zakreslíme jako sloupce (šířka = interval, výška = počet prvků).
Je nutné provádět normalizaci, aby byly všechny nakreslené sloupce stejně široké! Např. skrz $w = \frac{max_x - min_x}{k}$, kde k je počet námi chtených intervalů.

Bodové odhady
Pokud platí $\mathbb{E}\hat{\theta}(X_1, X_2, ..., X_n) = \theta$ nazýváme odhad nestranný.
Pokud platí $\lim_{n\to\infty} var\hat{\theta}(X_1, X_2, ..., X_n) = 0$ nazýváme odhad konzistentní (měla by tady platit i asymptotická nestrannost).
Metoda momentů odhadujeme parametry $\theta_1 ... \theta_k$, vytvoříme soustavu k rovnic dle předpisu $\mathbb{E}X_1^i = m_i, m_i = \frac{1}{n}\sum_{j=1}^{n} x_j^i$
Metoda max. věrohodnosti $\prod_{i=1}^{n} P_{\hat{\theta}}(X_1 = x_i) = max \prod_{i=1}^{n} P_\theta(X_1 = x_1)$, často je lepší počítat s logaritmem tohoto vztahu

Interval spolehlivosti
Jak název napovídá, máme intervalový odhad pro nějaký parametr (nejčastěji μ nebo σ^2)
Potřebujeme vědět, na jaké hladině α testovat - koukáme do tabulky kvantilů N(0,1) rozdělení (tab. č. 1) nebo t-rozdělení
Pro odhad střední hodnoty založený na CLV využijeme následující:
$(\bar{X} - \mu_{1-\alpha/2}\frac{S_n}{\sqrt{n}} < \mu < \bar{X} + \mu_{1-\alpha/2}\frac{S_n}{\sqrt{n}})$, tj. μ leží v otevřeném intervalu těchto dvou dopočítaých hodnot
Pro normální rozdělení, kde neznáme hodnotu μ a naopak známe σ^2:
$(\bar{X} - \mu_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + \mu_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$
Pokud neznáme ani μ, ani σ^2, využijeme následující:
$(\bar{X} - t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}} < \mu < \bar{X} + t_{1-\alpha/2,n-1}\frac{S_n}{\sqrt{n}})$ pro odhad μ $(\frac{(n-1)S_n^2}{\chi^2_{1-\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S_n^2}{\chi^2_{\alpha/2,n-1}})$ je odhad pro σ^2
Občas lze za S_n v odhadu založeném na CLV elegantně dosadit dle rozdělení výběru, např. pro Po(λ)$\to$ $\sqrt{\bar{X}_n}$ nebo Alt(p)$\to\sqrt{\bar{X}_n(1-\bar{X}_n)}$

Jak obecně testovat hypotézy:
1. Charakterizuj soubor - Jak velký máme vzorek? Jaký je výběrový průměr či rozptyl?
2. Model n.v. - Jaké má rozdělení, pokud ho vůbec známe? Jaké má dané rozdělení parametry? Potřebujeme zohlednit nějaké vazby? (Jak vypadá teoretická četnost?)
3. Hypotéza - Jak vypadá H_0 a H_A? Máme správně nastavené (ne)rovnosti?
4. Výběr testovací statistiky - Která je ta správná? Jak vypadá vzorec? Potřebujeme aplikovat CLV (ano, pokud neznáme rozdělení $\to$ asi spíš řešit přes intervalový odhad)?
5. Porovnání na testovací hladině - Na jaké hladině testujeme? Jakému kvantilu odpovídá? Máme zamítací pásmo jen z jedné strany, nebo z obou? Spadá naše hodnota z body 4. do pásma zamítání?
6. Výsledek - (ne)zamítáme H_0 ve prospěch H_A

t-test střední hodnoty normálního rozdělení:
Pro normální rozdělení, pokud není známe (není normální), POUŽIJ "CLV"(nebo na to hoď intervalový odhad :D)!
Potřebujeme výběrový průměr, směrodatnou odchylku a počet vzorků
$T_0 = \frac{\bar{X}_n - \mu_0}{S_n}\cdot\sqrt{n} \to \mu_0$ je dané z H_0; porovnáváme s $t_{1-\alpha/2;n-1}$

Párový test:
Sledujeme párové znaky (např. dioprie na obou očích), označíme Y a Z
Testujeme, zda se rozdíl středních hodnot rovná μ_0
Položíme $X_n = Y_n - Z_n$... a pokud to má normální rozdělení, aplikujeme test výše (pokud ne, tak máme asi smůlu)

Testování shody rozptylů:
Máme dva nezávislé výbery X,Y s normálním rozdělením s neznámým nenulovým rozptylem, vytvoříme výběrový rozptyl S_x^2

a S_y^2
$F_0 = \frac{S_X^2}{S_Y^2}$; porovnáváme s $F_{\alpha/2;m-1;n-1}$ (nebo $1-\alpha/2$), kde m je u X a n u Y

Test dobré shody:
Pro χ^2 rozdělení, zobecňuje binomické rozdělení. Bacha, n NENÍ k, n j počet vzorků od každého z k typů!
Testujeme H_0, že platí $P(X = x_i) = p_i$, kde $p_i = P(X = x_i)$
Potřebujeme teoretickou četnost (np_i)
$\chi_0^2 = \sum \frac{(X_i - np_i)^2}{np_i}$; zkoumáme, zda $\chi_0^2 \leq \chi^2_{1-\alpha;k-1}$, kde k je počet (typů) vzorku
Test nezávislosti:
Dva náhodné výběry Y a Z, každý z nich má n prvků a nabývá hodnot 1-r (respektive 1-c)
Testujeme, zda jsou X a Y nezávislé (naše H_0)
Vytvoříme si fešnou kontingenční tabulku s celkem r sloupci a c řádky, do kterých doplníme četnosti dle zadání a součty řádků/sloupců
Dopočítáme teoretické četnosti $n_{teor.} = \frac{\sum_i n_{ij} \cdot \sum_j n_{ij}}{n}$
$\chi_0^2 = \sum_{i,j} \frac{(n_{i,j} - n_{teor.i,j})^2}{n_{teor.i,j}}$; zkoumáme, zda $\chi_0^2 < \chi_{1-\alpha,(r-1)(c-1)}$

Random poznámky - χ^2 vždy porovnáváme jen z jedné strany! A taky je fajn koukat do správného sloupečku v tabulce, protože 0.95 není to samé jako 0.995...

Náhodné procesy - Rodina n.v. x_t, t $\in$ T definovaná v pravděpodobnostním prostoru; dělíme dle typu stavu (spojité/diskrétní) a času (spojitý - T = [a,b]; diskrétní - T = $\mathbb{N}$)
Markovská vlastnost - $P(X_{n+1} = j | x_n = i_n, x_{n-1} = i_{n-1}, ..., x_0 = i_0) = P(x_{n+1} = j | x_n = i)$, tj. nezáleží zde na minulosti, zajímá nás pouze poslední známý stav
Homogenní řetězec - nezávisí na n, ale pouze na stavech i a j, mezi kterými přecházíme. Tj. např. $P(x_3 = j | x_2 = i)$ je to samé jako $P(x_9 = j | x_8 = i)$
Matice pravděpodobnosti přechodů - hodnota na pozici i, j odpovídá pravděpodobnosti, že ze stavu i v následujícím kroce přejdeme do stavu j. Tuto matici lze mocnit, čímž zjistíme psti přechodu po n krocích.
Stacionární řešení - značíme π; odpovídá hodnotě psti přechodů, na které se ustálí jednotlivé stavy po ∞ krocích
(A)periodický stav - periodický, pokud perioda je větší než 1; jinak aperiodický
Perioda - "po kolika krocích se můžeme vrátit do původního stavu?"$\rightarrow$ největší společný dělitel tohohle pro všechny stavy
Dosažitelný stav - říkáme, že stav j je dosažitelný ze stavu i, pokud platí, že $p_{ij}^(n) > 0$
Uzavřená množina - pro uzavřenou množinu platí, že žádný stav mimo ni z ní není dosažitelný (tj. jakmile se dostaneme do uzavřené množiny, už ji neopustíme)
Trvalý stav - po nekonečně krocích je právděpodobnost, že se v něm můžeme nacházet, rovna 1 (např. stav, ze kterého nevede přechod do všech "sousedů")
Přechodný stav - po nekoneřně krocích je nenulová pravděpodobnost, že se v něm můžeme nacházet (ale není rovna 1); co není trvalé, je přechodné
Absorbční stav - podmnožina trvalých, jednobodová, "cesty vedou do absorbčního stavu, ale z něj už ne"$\rightarrow$ matematicky jde o jednobodovou uzavřenou množinu
Nulový stav - střední doba návratu do stavu je rovna ∞; jinak jde o stav nenulový
Random poznámky - periodické stavy NEKONVERGUJÍ k stac. řešení!

Jak tohle všechno aplikovat v praxi?
1. Podíváme se na schéma nebo slovní úlohu, ze které vycházíme
2. Vytvoříme si matici přechodů - řádky odpovídají aktuálnímu stavu, sloupce nadcházejícímu. Zapíšeme do ní jednotlivé psti (pokud pst není daná, je rovna 0)
3. Vypíšeme si jednotlivé složky stacionárního řešení - vznikne nám soustava lineárních rovnic. Např. $\pi_1 = P_{1,1} \cdot \pi_1 + P_{2,1} \cdot \pi_2 + P_{3,1} \cdot \pi_3$
4. Aplikujeme trik (součet všech složek stacionárního řešení se rovná 1)
5. Vyřešíme soustavu rovnic danou bodem 3. a 4. a jednotlivé hodnoty zapíšeme do sloupců matice přechodů k příslušným stavům
6. Vypočteme cokoliv dalšího, co po nás zlatíčko Helisová bude ve zkoušce chtít :)
7. A když má rozdělení více komponent? Aplikujeme krok 3.-5. pro každou komponentu a napíšeme stac. řešení v podobě "a$\cdot\pi_a$+b$\cdot\pi_b$..."
8. Kolik konstant a,b... potřebujeme spočítat? Stačí nám vždy dopočítat jen N-1 konstant, poslední dopočteme díku triku z bodu 4. 9. A jak je spočítat? Aplikujeme Beckův trik - "S jakou pstí se ze startovního políčka dostaneme do cílové komponenty?"