Nestranný odhad

Parametr θ

- Máme statistický (výběrový) soubor, který je realizací náhodného výběru X_1, X_2, X_3,…, X_n z pravděpodobnostní distribuce, která je kompletně stanovena jedním nebo více parametry – **modelové parametry**.
- Statistická veličina, která nás zajímá, odpovídá určité vlastnosti modelové distribuce, která může být sama popsána modelovými parametry.
- Taková vlastnost modelové distribuce se nazývá **parametr θ**.
- Např. v Poiss(λ) rozdělení je modelovým parametrem λ. Parametrem zájmu může být třeba samotné λ nebo třeba pravděpodobnost, že jev nenastane $e^{-\lambda}$.
- Každý parametr θ závisí jenom na statistickém souboru.

Odhad

ESTIMATE. An *estimate* is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e., t is some function of the dataset only:

$$t = h(x_1, x_2, \ldots, x_n).$$

- Popis odhadu je spíše formální, ale myšlenka spočívá ve skutečnost, že funkce *t* spočítaná ze statistického souboru mi dá nějakou představu o parametru θ reálné distribuce.
- Několik odhadů jsme už poznali – viz tabulka v přednášce 6 na str. 14 – jsou to tedy různé číselné hodnoty, množiny čísel nebo samotné křivky.
- Např.:
  - λ je střední hodnota modelové distribuce, podle zákona velkých čísel je výběrový průměr $\overline{x_n}$ přirozeným odhadem pro λ.
  - pro pravděpodobnost, že náhodná proměnná s rozdělením Poiss(λ) bude nabývat nulové hodnoty může být přirozeným odhadem četnost nul ve statistickém souboru $\dfrac{\text{number of } x_i \text{ equal to zero}}{n}$ nebo odhad exp(- $\overline{x_n}$)

Odhad

- Z předchozího plyne, že můžeme vymyslit několik odhadů pro daný parametr θ.
- Kdy je jeden odhad lepší než jiný?
- Existuje nejlepší možný odhad?
- Odpověď musí být negativní, protože nemůžeme říct nic jistého o různých odhadech, protože sami jsou spočítány z náhodného statistického souboru.
- Jediné co můžeme říci je, s jakou pravděpodobností jsou jednotlivé odhady vzdáleny od parametru θ.
- **Odhadová funkce** je vlastně metoda jak počítat odhady. Je to vlastně speciální případ výběrové charakteristiky.
- Odhad je číslo, vypočítané ze statistického souboru.

ESTIMATOR. Let $t = h(x_1, x_2, \ldots, x_n)$ be an estimate based on the dataset $x_1, x_2, \ldots, x_n$. Then t is a realization of the random variable

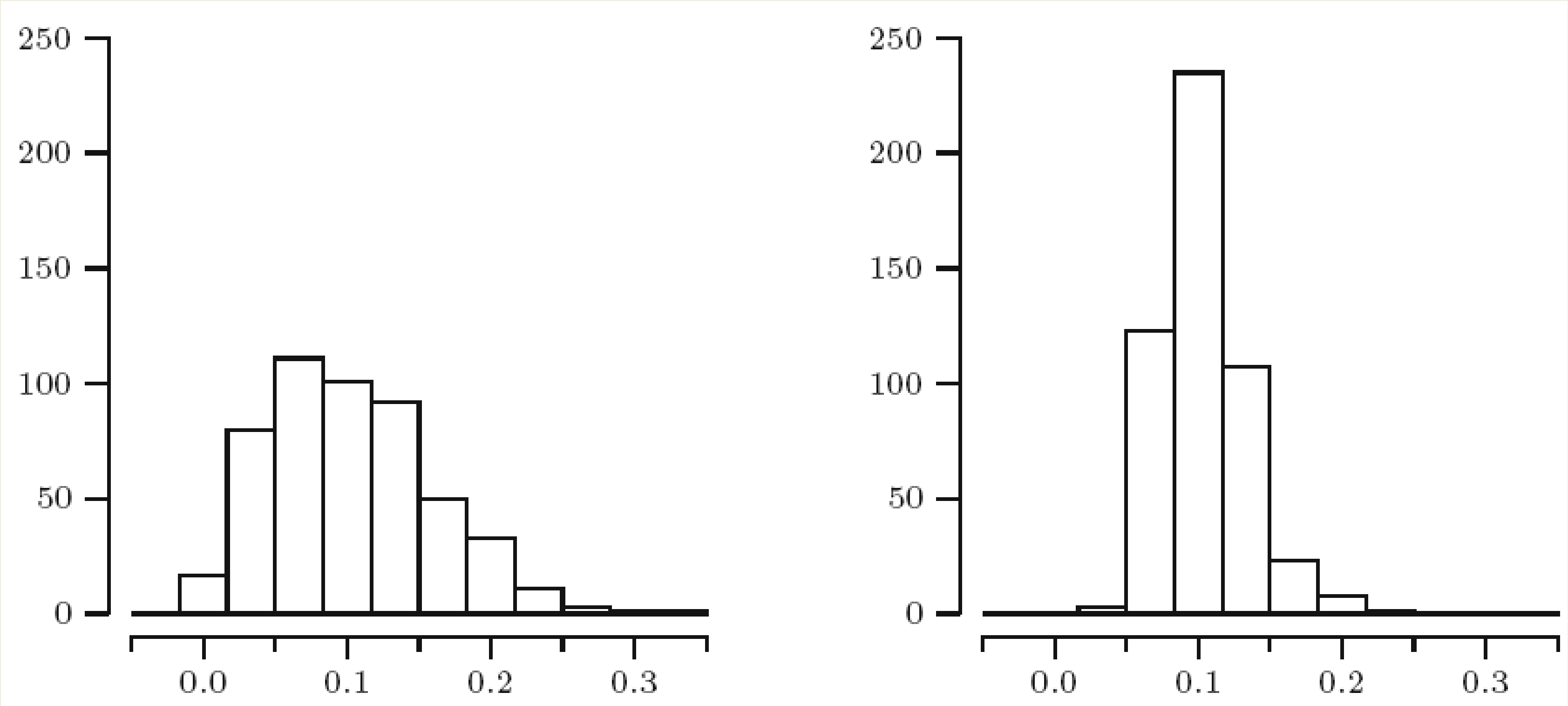
$$T = h(X_1, X_2, \ldots, X_n).$$

The random variable T is called an *estimator*.

Chování odhadových funkcí

- Mějme Poiss(μ) pravděpodobnostní rozdělení a naměříme 30 realizací náhodného výběru z *F*.
- Chceme odhadnout pravděpodobnost p_0, že náhodná proměnná x = 0.
- Zvolíme dvě odhadové funkce *S* a *T* $S = \dfrac{\text{number of } X_i \text{ equal to zero}}{n} \quad \text{and} \quad T = \mathrm{e}^{-\bar{X}_n}$
- *S* může nabývat jen hodnot: 0, 1/30, 2/30, 3/30,…, 1
- *T* může nabývat hodnot: 1, $e^{-1/30}$, $e^{-2/30}$, $e^{-3/30}$,…
- Je zřejmé, že *S* a *T* nemohou dát pro 30 měření stejnou hodnotu p_0.
- Situaci můžeme nasimulovat v počítači pro μ = ln 10 a tedy p_0 = 0,1. 500 krát zopakujeme náhodné vybrání 30 hodnot z Poiss(μ) a máme tedy 500 hodnot pro každou *S* a *T* a vyneseme četnosti hodnot do histogramu.
- Obě odhadové funkce se pohybují kolem správné hodnoty p_0 = 0,1, kterou mají odhadovat.

Chování odhadových funkcí



Výběrová distribuce

- Tedy hodnoty odhadové funkce *S* fluktuují kolem 0,1. Je tedy žádoucí, aby střední hodnota *S* byla rovna 0,1.
- Navíc, chceme aby to platilo pro jakoukoli hodnotu p_0, tedy $E[S] = p_0$, pro $0 < p_0 < 1$.
- Abychom to ověřili potřebujeme znát pravděpodobnostní distribuci odhadové funkce *S*.
- Odhadové funkce jsou konstruovány z náhodného výběru mluvíme o **výběrové distribuci**.

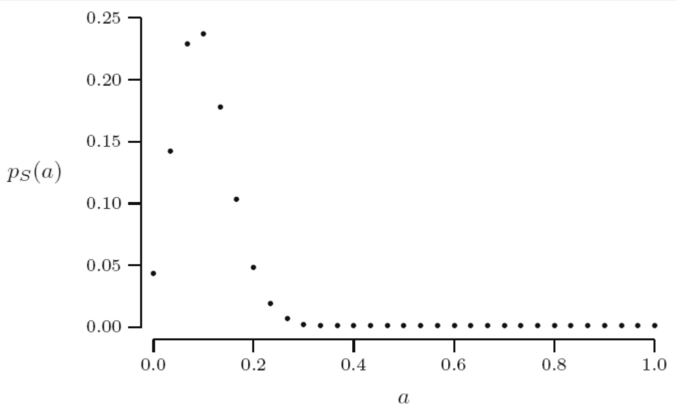
THE SAMPLING DISTRIBUTION. Let $T = h(X_1, X_2, \ldots, X_n)$ be an estimator based on a random sample $X_1, X_2, \ldots, X_n$. The probability distribution of T is called the *sampling distribution* of T.

Výběrová distribuce

- Jak najít konkrétní výběrovou distribuci?
- Nechť $S = Y/n$, kde Y je počet X_i rovných nule a tedy Y je rovno počtu úspěchu v n nezávislých pokusech s pravděpodobností úspěchu p_0.
- Tedy Y musí mít $Bin(n, p_0)$ distribuci a pak $S = Bin(n, p_0)/n$, s diskrétní náhodnou proměnnou k/n.

$$p_S\left(\frac{k}{n}\right) = \mathrm{P}\left(S = \frac{k}{n}\right) = \mathrm{P}(Y = k) = \binom{n}{k} p_0^k (1 - p_0)^{n-k}$$

- Pravděpodobnostní funkce p_S(a) pro $n = 30$ a $p_0 = 0{,}1$.



- Střední hodnota S bude:

$$\mathrm{E}[S] = \frac{\mathrm{E}[Y]}{n} = \frac{np_0}{n} = p_0$$

Výběrová distribuce a nestrannost

- Tedy odhadová funkce S pro p_0 má vlastnosti, že $E[S] = p_0$.
- To odráží fakt, že S nemá systematickou tendenci produkovat odhady, které jsou větší než p_0 nebo menší než p_0. To je žádoucí vlastnost odhadové funkce!!! A taková odhadová funkce se nazývá jako **nestranná**.

DEFINITION. An estimator T is called an *unbiased* estimator for the parameter θ, if

$$\mathrm{E}[T] = \theta$$

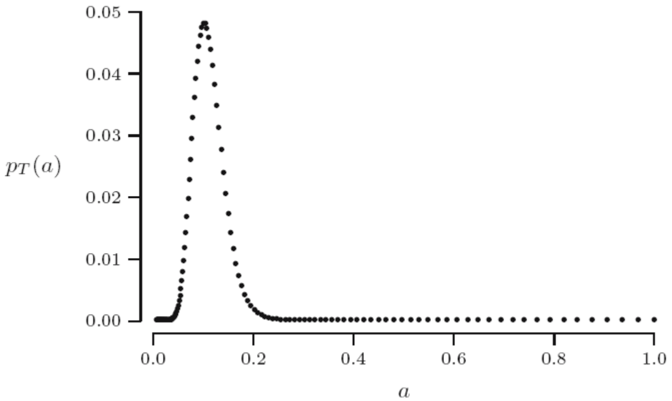
irrespective of the value of θ. The difference $\mathrm{E}[T] - \theta$ is called the *bias* of T; if this difference is nonzero, then T is called *biased*.

Výběrová distribuce a nestrannost

$T = \mathrm{e}^{-\bar{X}_n}.$

- Teď stejnou proceduru provedeme i pro odhadovou funkci *T*.
- Můžeme ji přepsat do tvaru $T = \mathrm{e}^{-Z/n},$ kde $Z = X_1 + X_2 + \cdots + X_n.$
- Náhodná proměnná *Z* je součtem *n* nezávislých *Poiss*(*μ*) náhodných proměnných a má distribuci *Poiss*(*nμ*).
- Tedy *T* je diskrétní náhodná proměnná nabývající hodnot $\mathrm{e}^{-k/n}$ s pravděpodobnostní funkcí:

$$p_T\left(\mathrm{e}^{-k/n}\right) = \mathrm{P}\left(T = \mathrm{e}^{-k/n}\right) = \mathrm{P}(Z = k) = \frac{\mathrm{e}^{-n\mu}(n\mu)^k}{k!}.$$



- Pro *n* = 30 a p_0 = 0,1 je pravděpodobnostní funkce v grafu. Mohlo by se zdát, že *T* je opět nestranná odhadová funkce, ale není to pravda – důkaz Jensenova nerovnost.

Výběrová distribuce a nestrannost

- Funkce exp(-x) je konvexní, musí tedy platit:

$$\mathrm{E}\left[T\right] = \mathrm{E}\left[\mathrm{e}^{-\bar{X}_n}\right] > \mathrm{e}^{-\mathrm{E}\left[\bar{X}_n\right]}$$

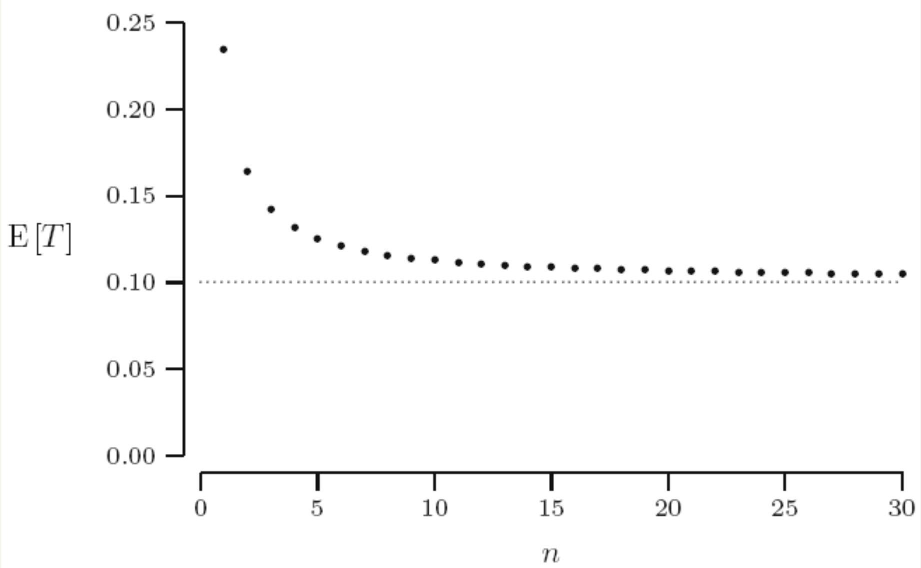
- Ze zákona velkých čísel plyne, že $\mathrm{E}[\bar{X}_n] = \mu$, protože μ je střední hodnota Poiss(μ).
- Pak dostaneme: $\mathrm{E}\left[T\right] > \mathrm{e}^{-\mu} = p_0$
- To znamená, že *T* je pozitivně stranné pro p_0.
- Spočítáme E[T] přesně: $\mathrm{E}\left[T\right] = \mathrm{E}\left[\mathrm{e}^{-\bar{X}_n}\right] = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})}$
- Protože $n(1 - \mathrm{e}^{-1/n}) \to 1$ pro $n \to \infty$, pak

$$\mathrm{E}\left[T\right] = \mathrm{e}^{-n\mu(1-\mathrm{e}^{-1/n})} \to \mathrm{e}^{-\mu} = p_0$$

- Vidíme, že strannost s rostoucím *n* klesá k nule.

Výběrová distribuce a nestrannost

- V grafu je střední hodnota T pro n = 30 a μ = ln 10 a p_0 = 0,1.



- Navíc platí, že výběrový průměr a výběrový rozptyl jsou nestranné odhadové funkce pro střední hodnotu *Poiss*(μ). Tato vlastnosti odhadových funkcí $\bar{X}_n$ a S_n^2 je navíc univerzální pro jakékoliv pravděpodobnostní rozdělení.

Nestranná odhadová funkce pro střední hodnotu a rozptyl

- Na statistickém souboru nás většinou zajímá střední hodnota a rozptyl modelové distribuce.

UNBIASED ESTIMATORS FOR EXPECTATION AND VARIANCE. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with finite expectation μ and finite variance σ^2. Then

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is an *unbiased estimator for* μ and

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

is an *unbiased estimator for* σ^2.

Nestranná odhadová funkce pro střední hodnotu a rozptyl

- Tvrzení v definici neříká nic jiného než, že $E[\bar{X}_n] = \mu$ a $E[S_n^2] = \sigma^2$.

$$\mathrm{E}\left[X_i - \bar{X}_n\right] = \mathrm{E}\left[X_i\right] - \mathrm{E}\left[\bar{X}_n\right] = 0 \qquad \mathrm{E}\left[Y\right] = 0$$

$$\mathrm{E}\left[S_n^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[(X_i - \bar{X}_n)^2\right]$$

$$\mathrm{Var}(Y) = \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 = \mathrm{E}\left[Y^2\right]$$

$$\mathrm{E}\left[(X_i - \bar{X}_n)^2\right] = \mathrm{Var}(X_i - \bar{X}_n)$$

$$X_i - \bar{X}_n = \frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i}X_j.$$

$$\begin{aligned}
\mathrm{Var}(X_i - \bar{X}_n) &= \mathrm{Var}\left(\frac{n-1}{n}X_i - \frac{1}{n}\sum_{j\neq i}X_j\right) \\
&= \frac{(n-1)^2}{n^2}\mathrm{Var}(X_i) + \frac{1}{n^2}\sum_{j\neq i}\mathrm{Var}(X_j) \\
&= \left[\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right]\sigma^2 = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[S_n^2\right] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[(X_i - \bar{X}_n)^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{Var}(X_i - \bar{X}_n) = \frac{1}{n-1}\cdot n \cdot \frac{n-1}{n}\sigma^2 = \sigma^2
\end{aligned}$$

Přenos nestrannosti

- Jaká bude odhadová funkce pro směrodatnou odchylku σ? Bude to funkce S_n?
- Podle Jensen nerovnosti to pravda nebude.

$$\sigma^2 = \mathrm{E}\left[S_n^2\right] > \left(\mathrm{E}\left[S_n\right]\right)^2$$

- Z toho plyne, že: $\mathrm{E}[S_n] < \sigma$
- Obecná vlastnost: nestrannost nějaké odhadové funkce se vždy nepřenáší.
- Je-li T nestranná odhadová funkce parametru θ, potom odhadová funkce $g(T)$ nemusí být nestranná odhadová funkce parametru $g(\theta)$.

Přenos nestrannosti

- Existuje speciální příklad nestranné odhadové funkce, kdy její nestrannost se přenese na novou odhadovou funkci, která vznikne lineární transformací.
- Nechť T je nestranná odhadová funkce pro parametr θ a platí, že $\mathrm{E}[T] = \theta$.
- Potom transformace: $g(T) = aT + b$ je nestranný odhad pro parametr $a\theta + b$.

$$\mathrm{E}\left[aT + b\right] = a\mathrm{E}\left[T\right] + b = a\theta + b$$

Střední kvadratická chyba a porovnání odhadových funkcí

Srovnání odhadových funkcí

- Nestrannost je zásadní vlastnost odhadových funkcí.
- Pokud existuje více nestranných odhadových funkcí pro daný parametr modelové distribuce, tak jak vybrat tu nejvhodnější?
- Přirozených parametrem výběru pro nestranné odhadové funkce bude rozptyl výběrové distribuce.

Odhadová funkce N

- Úkolem je odhadnout celkový počet vyrobených automobilů *N*, pokud náš statistický soubor obsahuje *n* výrobních čísel náhodně vybraných vozů.
- Označme vybraná sériová čísla $x_1, x_2, x_3, \ldots, x_\mathrm{n}$ jako realizaci náhodných proměnných $X_1, X_2, X_3, \ldots, X_\mathrm{n}$ reprezentující *n* výběrů bez vracení se stejnou pravděpodobností z množiny 1, 2, 3, …, N.
- $X_1, X_2, X_3, \ldots, X_\mathrm{n}$ není náhodný výběr, protože náhodné proměnné jsou vzájemně závislé.
- Zkonstruujeme dvě nestranné odhadové funkce T_1 a T_2.

Odhadová funkce N – výběrový průměr

- První bude založena na výběrovém průměru:

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- Druhá bude založena na výběrovém maximu:

$$M_n = \max\{X_1, X_2, \ldots, X_n\}$$

- Jak zkonstruovat nestrannou odhadovou funkci pro *N* na základě výběrového průměru?
- Spočítáme střední hodnotu $\overline{X_n}$; pravidlo součtu středních hodnot platí i pro závislé náhodné proměnné: $$\mathrm{E}\left[\bar{X}_n\right] = \frac{\mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]}{n}$$

Odhadová funkce N – výběrový průměr

- Snadno nahlédneme, že marginální distribuce pro každé X_i je stejná: $\mathrm{P}(X_i = k) = \frac{1}{N} \quad \text{for } k = 1, 2, \ldots, N.$
- Tedy střední hodnota každého X_i je:

$$\begin{aligned} \mathrm{E}[X_i] &= 1 \cdot \frac{1}{N} + 2 \cdot \frac{1}{N} + \cdots + N \cdot \frac{1}{N} = \frac{1 + 2 + \cdots + N}{N} \\ &= \frac{\frac{1}{2} N(N+1)}{N} = \frac{N+1}{2}. \end{aligned}$$

- Potom: $\mathrm{E}\left[\bar{X}_n\right] = \frac{\mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]}{n} = \frac{N+1}{2}.$
- Protože střední hodnota T_1 se musí rovnat hledanému parametru N, pak: $T_1 = 2\bar{X}_n - 1$ je nestranná odhadová funkce pro N,

protože: $\mathrm{E}[T_1] = \mathrm{E}\left[2\bar{X}_n - 1\right] = 2\mathrm{E}\left[\bar{X}_n\right] - 1 = 2 \cdot \frac{N+1}{2} - 1 = N.$

Odhadová funkce N – výběrové maximum

- Spočítáme si střední hodnotu náhodné proměnné M_n. Potřebujeme najít její pravděpodobnostní distribuci – jaká je pravděpodobnost, že $M_n = k$?
- Počet způsobů jak vybrat n čísel bez opakování z N prvkové množiny je $\binom{N}{n}$ a každá kombinace má pravděpodobnost $1/\binom{N}{n}$.
- Aby se $M_n = k$, tak musíme mít jeden výběr rovný k a ostatních n-1 výběrů z čísel 1, 2, 3, …, k-1. Uděláme to $\binom{k-1}{n-1}$ způsoby pro $k = n, n+1, \ldots, N$.
- Potom pro pravděpodobnost, že $M_n = k$ platí:

Odhadová funkce N – výběrové maximum

$$\mathrm{P}(M_n = k) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}} = \frac{(k-1)!}{(k-n)!(n-1)!} \cdot \frac{(N-n)!\,n!}{N!}$$
$$= n \cdot \frac{(k-1)!}{(k-n)!} \frac{(N-n)!}{N!}.$$

- A střední hodnota bude:

$$\mathrm{E}[M_n] = \sum_{k=n}^{N} k\mathrm{P}(M_n = k) = \sum_{k=n}^{N} k \cdot n \cdot \frac{(k-1)!}{(k-n)!} \frac{(N-n)!}{N!}$$
$$= \sum_{k=n}^{N} n \cdot \frac{k!}{(k-n)!} \frac{(N-n)!}{N!}$$
$$= n \cdot \frac{(N-n)!}{N!} \sum_{k=n}^{N} \frac{k!}{(k-n)!}.$$

- Jak spočítat poslední sumu? Použijeme trik.

Odhadová funkce N – výběrové maximum

- Musí platit: $$1 = \sum_{j=n}^{N} \mathrm{P}(M_n = j) = \sum_{j=n}^{N} n \cdot \frac{(j-1)!}{(j-n)!} \frac{(N-n)!}{N!},$$
- Z toho hned plyne následující rovnost, platící pro libovolné *N* a *n* $\leq$ N:

$$\sum_{j=n}^{N} \frac{(j-1)!}{(j-n)!} = \frac{N!}{n\,(N-n)!}.$$

- Zaměňme *N* za *N*+1 a *n* za *n*+1:

$$\sum_{j=n+1}^{N+1} \frac{(j-1)!}{(j-n-1)!} = \frac{(N+1)!}{(n+1)(N-n)!}.$$

- Nahraďme *j* – 1 = *k*: $$\sum_{k=n}^{N} \frac{k!}{(k-n)!} = \frac{(N+1)!}{(n+1)(N-n)!}.$$

Odhadová funkce N – výběrové maximum

- Teď můžeme dopočítat E[M_n]:

$$\mathrm{E}\left[M_n\right] = n \cdot \frac{(N-n)!}{N!} \sum_{k=n}^{N} \frac{k!}{(k-n)!}$$
$$= n \cdot \frac{(N-n)!}{N!} \cdot \frac{(N+1)!}{(n+1)(N-n)!} = n \cdot \frac{N+1}{n+1}.$$

- Protože střední hodnota T_2 se musí rovnat hledanému parametru N, pak: $T_2 = \frac{n+1}{n} M_n - 1$ je nestranná odhadová funkce pro N:

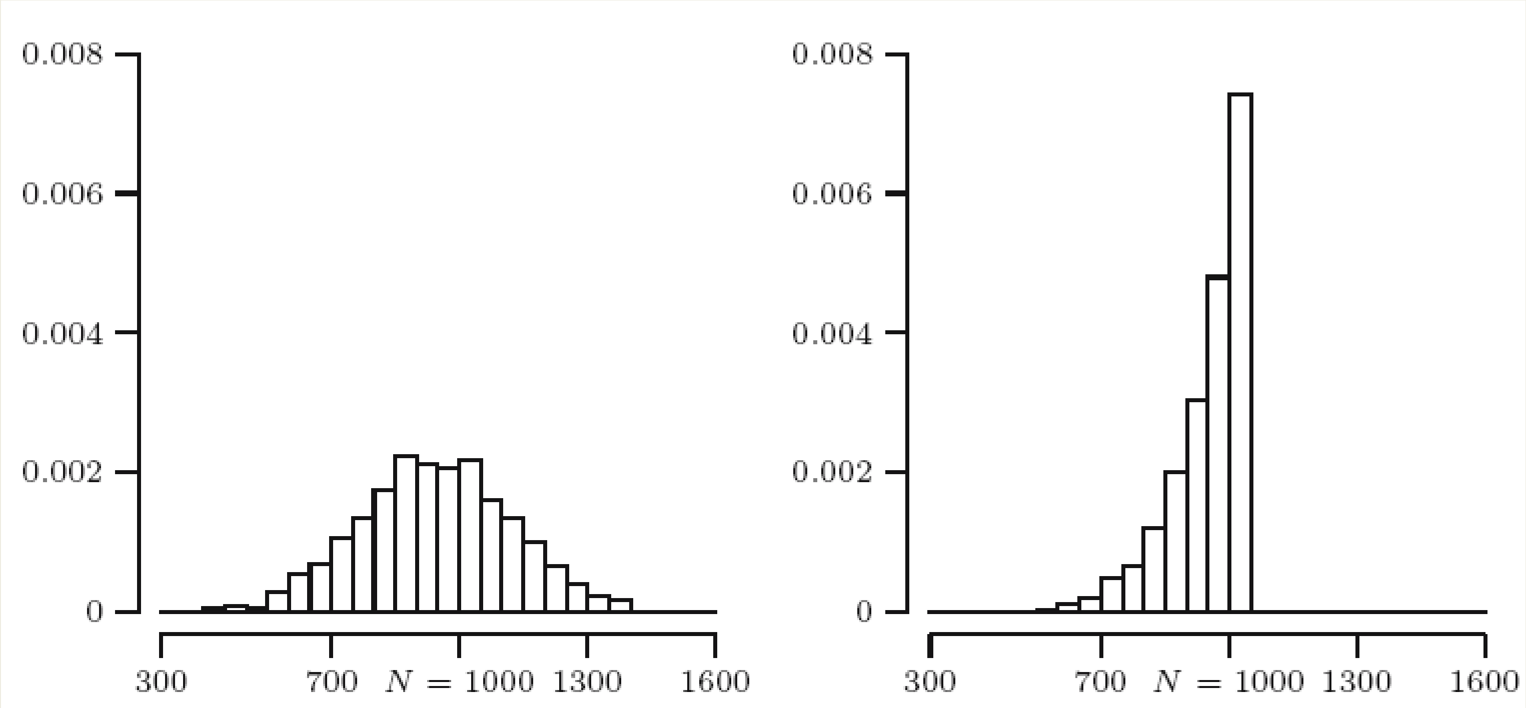
$$\mathrm{E}\left[T_2\right] = \mathrm{E}\left[\frac{n+1}{n} M_n - 1\right] = \frac{n+1}{n} \mathrm{E}\left[M_n\right] - 1 = \frac{n+1}{n} \cdot \frac{n(N+1)}{n+1} - 1 = N.$$

Odhadová funkce N

- Dostali jsme tedy dvě nestranné odhadové funkce pro odhad parametru *N.*
- Která z nich je lepší? Určíme to z toho, jak se T_1 a T_2 mění kolem hodnoty *N*.
- Udělejme simulaci: $N = 1000$, $n = 10$, tedy vybereme bez opakování 10 čísel z 1, 2, 3, …, 1000 a spočítáme hodnoty T_1 a T_2. Celé to 2000-krát zopakujeme a těchto 2000 hodnot vyneseme do histogramu pro každou odhadovou funkci.

Odhadová funkce N

- Vlevo histogram simulované odhadové funkce T_1 a vpravo T_2.



Odhadová funkce N

- Protože histogramy reprezentují pravděpodobnostní funkci, tak vidíme, že distribuce obou odhadových funkcí jsou zcela odlišné.
- Rozptyl T_2 kolem hodnoty N je menší než rozptyl T_1.
- Tedy T_2 odhaduje parametr N účinněji, protože odhady jsou více koncentrovány kolem N v porovnání s T_1.
- Tedy rozptyl odhadové funkce určuje její účinnost.

Odhadová funkce N - rozptyl

EFFICIENCY. Let T_1 and T_2 be two unbiased estimators for the same parameter θ. Then estimator T_2 is called *more efficient* than estimator T_1 if $\mathrm{Var}(T_2) < \mathrm{Var}(T_1)$, irrespective of the value of θ.

- Spočítejme rozptyly odhadových funkcí T_1 a T_2.

$$\mathrm{Var}(T_1) = \mathrm{Var}\big(2\bar{X}_n - 1\big) = 4\mathrm{Var}\big(\bar{X}_n\big)$$

- Protože X_i mají všechny stejnou pravděpodobnostní distribuci, tak i páry (X_i, X_j) pro $i \neq j$ mají stejnou distribuci.
- Potom pro rozptyl součtu náhodných proměnných platí: $\mathrm{Var}(X_1 + \cdots + X_n) = n\mathrm{Var}(X_1) + n(n-1)\mathrm{Cov}(X_1, X_2)$
- Dá se ukázat, že: $\mathrm{Var}(X_1) = \frac{1}{12}(N-1)(N+1), \quad \mathrm{Cov}(X_1, X_2) = -\frac{1}{12}(N+1)$

Odhadová funkce N - rozptyl

- Potom pro rozptyl T_1 máme:

$$\begin{aligned}\mathrm{Var}(T_1) &= 4\mathrm{Var}\left(\bar{X}_n\right) = \frac{4}{n^2}\mathrm{Var}(X_1 + \cdots + X_n) \\ &= \frac{4}{n^2}\left[n \cdot \frac{1}{12}(N-1)(N+1) - n(n-1) \cdot \frac{1}{12}(N+1)\right] \\ &= \frac{1}{3n}(N+1)[N-1-(n-1)] \\ &= \frac{(N+1)(N-n)}{3n}.\end{aligned}$$

- Výpočet rozptylu odhadové funkce T_2 je složitější. Dá se ukázat, že platí: $\mathrm{Var}(M_n) = \dfrac{n(N+1)(N-n)}{(n+2)(n+1)^2}$
- Pozn.: použije se podobného triku jako u výpočtu $E[M_n]$.

Odhadová funkce N - rozptyl

- Potom rozptyl T_2 bude:

$$\mathrm{Var}(T_2) = \mathrm{Var}\left(\frac{n+1}{n}M_n - 1\right) = \frac{(n+1)^2}{n^2}\mathrm{Var}(M_n) = \frac{(N+1)(N-n)}{n(n+2)}$$

- Vidíme, že $\mathrm{Var}(T_2) < \mathrm{Var}(T_1)$ pro všechny N a $n \geq 2$. Pro $n = 1$ jsou obě odhadové funkce rovny X_1.
- Poměr $\mathrm{Var}(T_1)/\mathrm{Var}(T_2)$ se nazývá jako relativní účinnost odhadové funkce T_2 s ohledem na odhadovou funkci T_1.
- V našem případě: $\dfrac{\mathrm{Var}(T_1)}{\mathrm{Var}(T_2)} = \dfrac{(N+1)(N-n)}{3n} \cdot \dfrac{n(n+2)}{(N+1)(N-n)} = \dfrac{n+2}{3}.$
- Tedy je vhodné preferovat odhadovou funkci T_2 pro odhad parametru N před T_1.

Střední kvadratická chyba

- I když je nestrannost důležitá vlastnost odhadové funkce, je třeba účinnost odhadové funkce nějak kvantifikovat i bez znalosti toho, zdali je odhadová funkce nestranná nebo není.
- Je nutné stanovit jak se „rozšiřuje" odhadová funkce kolem hledaného parametru θ.

DEFINITION. Let T be an estimator for a parameter θ. The *mean squared error* of T is the number $\mathrm{MSE}(T) = \mathrm{E}\left[(T-\theta)^2\right]$.

- Střední kvadratická chyba (MSE) odhadové funkce je zobecňující parametr popisující účinnost odhadové funkce.

Střední kvadratická chyba

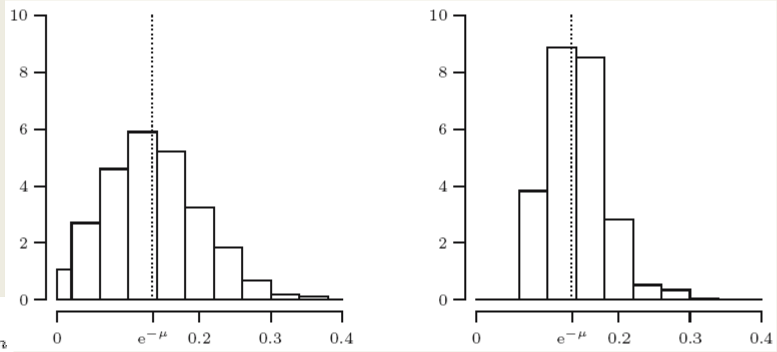
- Tedy odhadová funkce T_1 je účinnější než odhadová funkce T_2 pokud $\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2)$.
- Definiční vztah lze přepsat na:

$$\begin{aligned}\mathrm{MSE}(T) &= \mathrm{E}\left[(T-\theta)^2\right] \\ &= \mathrm{E}\left[(T-\mathrm{E}[T]+\mathrm{E}[T]-\theta)^2\right] \\ &= \mathrm{E}\left[(T-\mathrm{E}[T])^2\right] + 2\mathrm{E}[T-\mathrm{E}[T]]\left(\mathrm{E}[T]-\theta\right) + \left(\mathrm{E}[T]-\theta\right)^2 \\ &= \mathrm{Var}(T) + \left(\mathrm{E}[T]-\theta\right)^2.\end{aligned}$$

- Tedy MSE je součet rozptylu odhadové funkce a její strannosti. Pro nestranné odhadové funkce je MSE rovno $\mathrm{Var}(T)$.

Střední kvadratická chyba

- Může nastat situace, kdy stranná odhadová funkce s malým rozptylem může dávat lepší výsledek než nestranná odhadová funkce s velkým rozptylem.
- Př. hledáme, jaká je hodnota Poiss(μ) pro hodnotu $X = 0$.
- Můžeme nalézt dvě odhadové funkce: $S = \dfrac{\text{number of } X_i \text{ equal to zero}}{n}$ $\quad T = \mathrm{e}^{-\bar{X}_n}$



- Nasimulujeme 1000 opakování náhodného výběru 25 prvků z Poiss(μ) distribuce s μ = 2 – viz histogramy: vlevo odhadová funkce *S*, vpravo *T*.
- Vidíme, že stranná odhadová funkce *T* je blíže k hledanému parametru $e^{-\mu} = e^{-2} = 0{,}1353$ než nestranná funkce *S*. Preferovaný výběr *T* je podporován skutečností, že MSE (*T*) je menší než MSE(*S*).

Maximální věrohodnost

Maximální věrohodnost

- Už umíme zkonstruovat odhadovou funkci pro parametry distribucí, kterým odpovídá přirozený výběrový parametr (E[X] vs. $\overline{X_n}$).
- Co když taková parametrová analogie neexistuje?
- Musím nalézt univerzální princip konstrukce odhadových funkcí pro libovolný parametr.
- K tomu slouží tzv. **metoda maximální věrohodnosti**.

Maximální věrohodnost - příklad

- Ilustrujme si princip maximální věrohodnosti na příkladu.
- Mějme dva balíky po 10 000 stejných elektronických součástek. V jednom balíku je 50% vadných součástek a v druhém balíku je 10% vadných.
- Bohužel neumíme balíky rozeznat. Který balík si máme koupit?
- Otevřu jeden balík a náhodně vyberu deset součástek, které otestuji na vadnost. Zjistím, že jedna je vadná.
- **Závěr**: vyberu si tento balík.
- V balíku s 50% defektních součástek je **více pravděpodobné**, že v 10-ti kusovém výběru se objeví více vadných součástek, zatímco u druhého balíku můžeme jednu vadnou součástku očekávat s větší pravděpodobností.

Maximální věrohodnost

- Tedy vyberu si ten balík, kde je nejvíce pravděpodobné, že jen jedna součástka je vadná
- Toto je základní myšlenka metody maximální věrohodnosti:

> THE MAXIMUM LIKELIHOOD PRINCIPLE. Given a dataset, choose the parameter(s) of interest in such a way that the data are most likely.

- **Důkaz**: nechť náhodná proměnná $R_i = 1$ v případě, že i-tá součástka je vadná a $R_i = 0$ v případě, že je funkční pro $i = 1, 2, 3, \ldots, 10$.
- Tedy $R_1, R_2, R_3, \ldots, R_{10}$ je 10 nezávislých proměnných s distribucí *Ber*(p), kde p je pravděpodobnost, že vybraná součástka je vadná.

$$\mathrm{P}(R_1 = 0, R_2 = 1, R_3 = 0, \ldots, R_{10} = 0) = p(1-p)^9.$$

Maximální věrohodnost

- Pro balík s 10% vadných součástek platí:

$$\mathrm{P}(R_1 = 0, R_2 = 1, R_3 = 0, \ldots, R_{10} = 0) = \frac{1}{10}\left(\frac{9}{10}\right)^9 = 0.039$$

- Pro balík s 50% vadných součástek platí:

$$\mathrm{P}(R_1 = 0, R_2 = 1, R_3 = 0, \ldots, R_{10} = 0) = \frac{1}{2}\left(\frac{1}{2}\right)^9 = 0.00098$$

- Tedy pravděpodobnost, že bude ve výběru právě jedna vadná součástka je asi 40 krát větší pro balík s 10% vadných součástek.

Věrohodnostní funkce

- Mějme statistický soubor prvků x_1, x_2,…, x_n modelovaný jako realizaci náhodného výběru z pravděpodobnostní distribuce charakterizované parametrem θ.
- Pravděpodobnostní funkce diskrétní náhodné proměnné je funkcí θ: $p_\theta(x)$.
- Hustota pravděpodobnosti spojité náhodné proměnné je funkcí θ: $f_\theta(x)$.
- Mějme příklad s diskrétní náhodnou proměnnou.
- Potom metoda maximální věrohodnosti nám říká, že parametr θ odhadneme takovým **číslem**, pro které je funkce $L(\theta)$ maximální.

$$L(\theta) = \mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = p_\theta(x_1) \cdots p_\theta(x_n)$$

Věrohodnostní funkce

- Takové číslo nazýváme jako **maximální věrohodný odhad** parametru θ.
- Funkci $L(\theta)$ nazýváme jako **věrohodnostní funkci**.
- Pro spojitou náhodnou proměnnou je nutné $L(\theta)$ definovat jiným způsobem, protože by se $L(\theta) = 0$.
- Mějme X a $f_\theta(x)$ a malé $\varepsilon>0$. Vybereme takové θ, že pravděpodobnost

  $$\mathrm{P}(x_1 - \varepsilon \leq X_1 \leq x_1 + \varepsilon, \ldots, x_n - \varepsilon \leq X_n \leq x_n + \varepsilon)$$

  je maximální.
- Protože X_i jsou nezávislé, musí platit:

Věrohodnostní funkce

$$\begin{aligned}
&\mathrm{P}(x_1-\varepsilon \le X_1 \le x_1+\varepsilon, \ldots, x_n-\varepsilon \le X_n \le x_n+\varepsilon) \\
&\quad = \mathrm{P}(x_1-\varepsilon \le X_1 \le x_1+\varepsilon) \cdots \mathrm{P}(x_n-\varepsilon \le X_n \le x_n+\varepsilon) \\
&\quad \approx f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n) (2\varepsilon)^n,
\end{aligned}$$

- kdy jsme využili známého faktu, že:

$$\mathrm{P}(x_i-\varepsilon \le X_i \le x_i+\varepsilon) = \int_{x_i-\varepsilon}^{x_i+\varepsilon} f_\theta(x)\,\mathrm{d}x \approx 2\varepsilon f_\theta(x_i)$$

- Tedy pravděpodobnost bude maximální pokud bude maximální funkce: $f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n)$
- Tedy věrohodnostní funkce pro spojitou náhodnou proměnnou bude definována:

$$L(\theta) = f_\theta(x_1) f_\theta(x_2) \cdots f_\theta(x_n)$$

- Můžeme tedy definovat **maximální věrohodný odhad**:

Věrohodnostní funkce

MAXIMUM LIKELIHOOD ESTIMATES. The *maximum likelihood estimate* of θ is the value $t = h(x_1, x_2, \ldots, x_n)$ that maximizes the likelihood function $L(\theta)$. The corresponding random variable

$$T = h(X_1, X_2, \ldots, X_n)$$

is called the *maximum likelihood estimator* for θ.

- Př.: Mějme statistický soubor prvků $x_1, x_2, \ldots, x_n$ modelovaný jako realizaci náhodného výběru z exponenciální pravděpodobnostní distribuce *Exp*(λ) s hustotou pravděpodobnosti $f_\lambda(x) = 0$ pro $x<0$ a $f_\lambda(x) = \lambda e^{-\lambda x}$ pro $x \geq 0$.
- Potom věrohodnostní funkce bude dána:

Věrohodnostní funkce

$$\begin{aligned} L(\lambda) &= f_\lambda(x_1) f_\lambda(x_2) \cdots f_\lambda(x_n) \\ &= \lambda \mathrm{e}^{-\lambda x_1} \cdot \lambda \mathrm{e}^{-\lambda x_2} \cdots \lambda \mathrm{e}^{-\lambda x_n} \\ &= \lambda^n \cdot \mathrm{e}^{-\lambda (x_1 + x_2 + \cdots + x_n)}. \end{aligned}$$

- Získat maximální věrohodný odhad parametru λ znamená nalézt maximum funkce *L*(λ).
- Funkce má maximum v místě, kde první derivace je nulová:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda} L(\lambda) &= n \lambda^{n-1} \mathrm{e}^{-\lambda \sum_{i=1}^n x_i} - \lambda^n \left(\sum_{i=1}^n x_i \right) \mathrm{e}^{-\lambda \sum_{i=1}^n x_i} \\ &= n \left(\lambda^{n-1} \mathrm{e}^{-\lambda \sum_{i=1}^n x_i} \left(1 - \frac{\lambda}{n} \sum_{i=1}^n x_i \right) \right). \end{aligned}$$

Věrohodnostní funkce

- Derivace d($L(\lambda)$)/dλ = 0 pokud 1 - $\lambda\bar{x}_{\mathrm{n}}$ = 0.
- Z toho plyne, že: $\lambda = 1/\bar{x}_{\mathrm{n}}$.
- Tedy věrohodnostní odhadová funkce pro parametr λ je funkce $1/\bar{X}_{\mathrm{n}}$.
- Z definice je vidět, že věrohodnostní funkce $L(\theta)$ je určena součinem pravděpodobnostních funkcí.
- Maximum funkce lze ve většině případů stanovit z její první derivace. Ale derivace součinu funkcí je většinou velmi pracná, protože hledaný parametr je obsažen v každém členu.
- Řešením je logaritmování funkce $L(\theta)$.

Logaritmus věrohodnostní funkce

- Logaritmus součinu funkcí je roven součtu logaritmů jednotlivých funkcí.
- Tedy místo derivování součinu funkcí budeme derivovat součet logaritmů těchto funkcí.
- Definice: $l(\theta) = \ln(L(\theta))$.
- Protože logaritmus je rostoucí funkce, tak funkce $l(\theta)$ i $L(\theta)$ nabývají maxima pro stejný parametr θ.
- Tedy $L(\theta)$ je maximální tehdy a jenom tehdy, když $l(\theta)$ je maximální.