Příklad 2.4. Rozdělení výběrového rozptylu a výběrové směrodatné odchylky
Na základ simulační studie ověřte, že pokud náhodná veličina $X \sim N(\mu, \sigma^2)$, potom pro výběrový rozptyl S_n^2 a výběrovou směrodatnou odchylku S_n platí

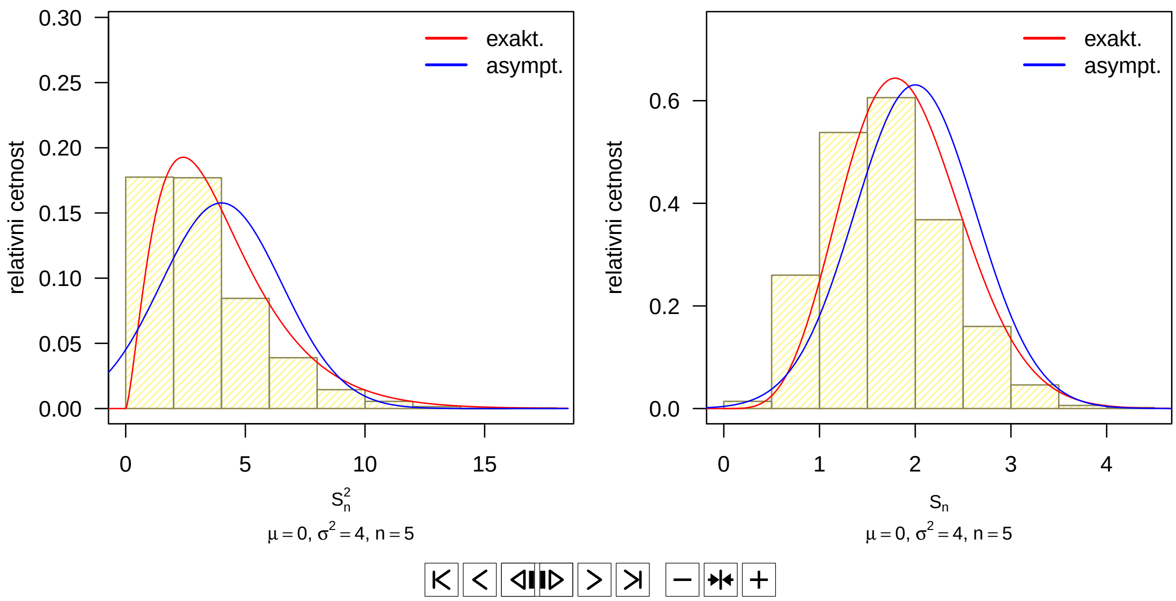
(a) $S_n^2 \sim \Gamma(\frac{n}{2}, \frac{2\sigma^2}{n})$ exaktně;

(b) $S_n^2 \sim N(\sigma^2, \frac{2\sigma^4}{n})$ asymptoticky;

(c) $S_n \sim \Gamma_G(\sqrt{\frac{2\sigma^2}{n}}, 2, \frac{n}{2})$ exaktně;

(d) $S_n \sim N(\sigma, \frac{\sigma^2}{2n})$ asymptoticky.

Vygenerujte $M = 1000$ náhodných výběrů z normálního rozdělení $N(\mu, \sigma^2)$ o rozsahu n, kde $\mu = 0$ a $\sigma^2 = 4$. Pro každý náhodný výběr vypočítejte výběrový rozptyl $S_{n_i}^2$, $i = 1, \ldots, M$. Rozptyly vykreslete pomocí histogramu a superponujte je křivkami hustoty exaktního i asymptotického rozdělení statistiky S_n^2. Dále pro každý náhodný výběr vypočítejte výběrovou směrodatnou odchylku S_{n_i}, $i = 1, \ldots, M$. Odchylky zaneste vykreslete do histogramu a superponujte je křivkami hustoty asymptotického a exaktního rozdělení statistiky S_n. Vytvořte animaci zobrazující konvergenci asymptotického rozdělení obou statistik k exaktnímu rozdělení při zvětšujícím se rozsahu náhodných výběrů n. Hodnoty rozsahu n volte 5, 10, 50, 100, 500, 1000.



Obrázek 4: Exaktní a asymptotické rozdělení (a) vyběrového rozptylu S_n^2; (b) vyběrové směrodatné odchylky S_n

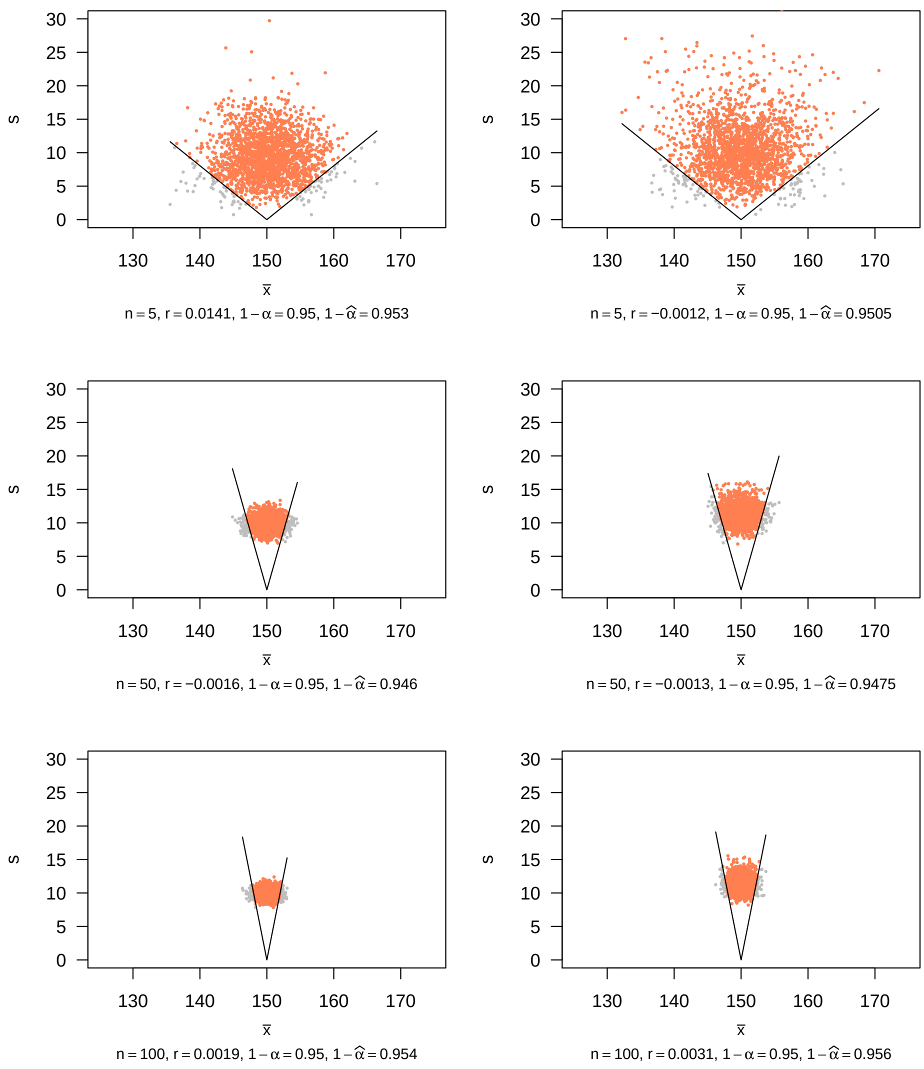
Teorie k příkladu 2.5

- **nominální pravděpodobnost pokrytí** (spolehlivost) ... teoretická pravděpodobnost pokrytí (spolehlivost) $1 - \alpha$ (jaká by teoreticky měla být)
- **aktuální pravděpodobnost pokrytí** (spolehlivost) ... skutečná pravděpodobnost pokrytí (spolehlivost) $1 - \widehat{\alpha}$ (jaká ve skutečnosti je)
- *Příklad: Předokládáme, že nominální spolehlivost námi zvoleného intervalu spolehlivosti (IS) (např. o μ když σ^2 známe) je $1-\alpha = 0.95$. Naměříme (nagenerujeme) data, a na základě nich vypočítáme aktuální spolehlivost $1 - \widehat{\alpha}$. Pokud dodržíme předpoklady (naměřená (nagenerovaná) data budou pocházet z normálního rozdělení, všechny se stejným rozptylem σ^2), bude aktuální spolehlivost $1 - \widehat{\alpha} \doteq 0.95$. Nicméně, pokud předpoklady (ať už vědomě či nevědomě) porušíme, může se aktuální spolehlivost $1 - \widehat{\alpha}$ od nominální spolehlivosti $1 - \alpha$ výrazně lišit.*

Příklad 2.5. Nezávislost μ a σ^2; pravděpodobnost pokrytí
Nechť $X \sim N(\mu, \sigma^2)$, kde $\mu = 150$ a $\sigma^2 = 10^2$. Pomocí simulační studie ověřte nezávislost výběrového průměru $\bar{X}$ a výběrové směrodatné odchylky $S = S_{n-1}$.

Nakreslete rozptylový graf $(\bar{x}_m, s_m)$, $m = 1, 2, \ldots, M$, kde $M = 2000$. Barvu bodů zvolte šedou. Vypočítejte hodnotu Pearsonova korelačního koeficientu $r_{\bar{X},S}$. Červenou barvou dále vyznačte v grafu takové body $(\bar{x}_m, s_m)$, pro které platí $\mu \in IS_m = (dh_m; hh_m) = \left(\bar{x}_m - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}}; \bar{x}_m - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}\right)$. Navíc vykreslete hranice definované body $(\bar{x}_m, s_m)$, jež splňují vztah $dh_m = hh_m = \mu$. Vypočítejte pravděpodobnost pokrytí 95 % DIS pro μ jako podíl $\frac{\sum_m I(\mu \in IS_m)}{M}$. Zvolte (a) $n = 5$, (b) $n = 50$ a (c) $n = 100$.

Simulaci proveďte také za předpokladu, že data pochází ze smíšeného rozdělení $X \sim [pN(\mu, \sigma^2) + (1-p)N(\mu, \sigma_2^2)]$, kde $p = 0.9$, $\mu = 150$, $\sigma^2 = 10^2$ a $\sigma_2^2 = 20^2$.



Obrázek 5: Nezávislost výběrového průměru $\bar{X}$ a výběrové směrodatné odchylky S_{n-1}

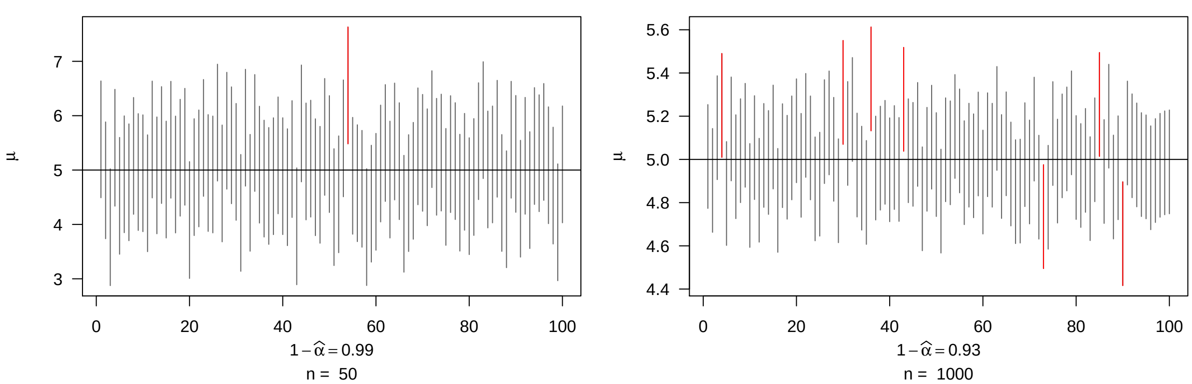
Teorie k příkladu 2.6

- **konzervativní IS** ... interval spolehlivosti, jehož aktuální (skutečná) pravděpodobnost pokrytí $1-\widehat{\alpha}$ je větší než nominální pravděpodobnost pokrytí $1-\alpha$, se nazývá *konzervativní IS* (IS obsahuje θ s větší pravděpodobností, než bychom předpokládali).
- **liberální IS** ... interval spolehlivosti, jehož aktuální (skutečná) pravděpodobnost pokrytí $1-\widehat{\alpha}$ je menší než nominální pravděpodobnost pokrytí $1-\alpha$, se nazývá *liberální IS* (IS obsahuje θ s menší pravděpodobností, než bychom předpokládali).

Příklad 2.6. Odhad koeficientu spolehlivosti $1-\alpha$ Waldova empirického DIS pro parametr μ normálního rozdělení při známém rozptylu σ^2
Nechť $X \sim N(5, 15)$. Pomocí simulační studie ($M = 100$) stanovte Monte Carlo (MC) odhad koeficientu spolehlivosti (pravděpodobnosti pokrytí) 95 % Waldova exaktního empirického DIS pro parametr μ normálního rozdělení při známém rozptylu σ^2.

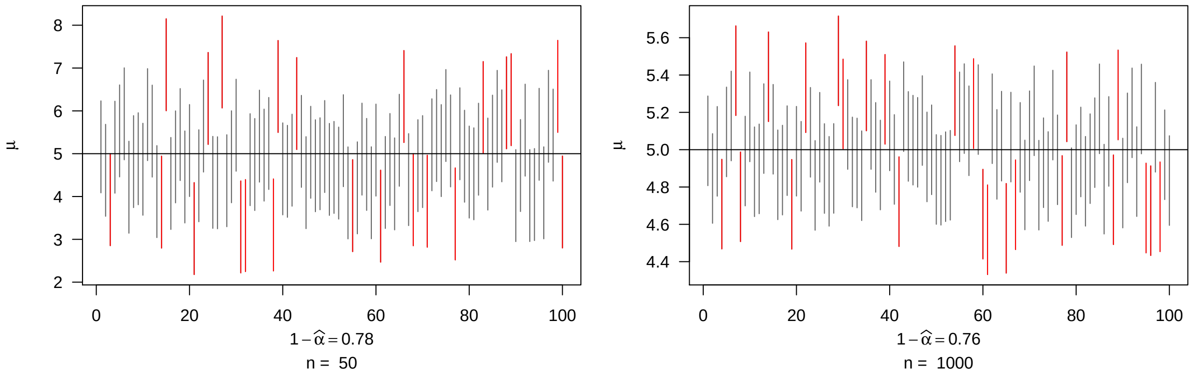
Vygenerujte $M = 100$ náhodných výběrů z normálního rozdělení $N(5, 10)$ a na základě každého náhodného výběru vypočítejte 95 % DIS pro parametr μ. Všechny DIS vykreslete do jednoho grafu jako svislé šedé úsečky. Červenou barvou dále vyznačte v grafu ty DIS, které nepokrývají střední hodnotu $\mu = 5$ a černou barvou vyznačte horizontální referenční čáru v bodě μ. Dále vypočítejte aktuální pravděpodobnost pokrytí 95 % DIS pro μ jako podíl $\frac{\sum_m I(\mu \in IS_m)}{M}$ a porovnejte ji s nominální pravděpodobností pokrytí $1-\alpha$. Rozsah náhodných výběrů volte (a) $n = 50$; (b) $n = 1000$.

Simulaci proveďte také za předpokladu, že data pochází ze smíšeného rozdělení $X \sim pN(5, 15) + (1-p)N(5, 15^2)$, kde $p = 0.9$, resp. ze smíšeného rozdělení $X \sim pN(5, 15) + (1-p)N(3, 15)$, kde $p = 0.9$.

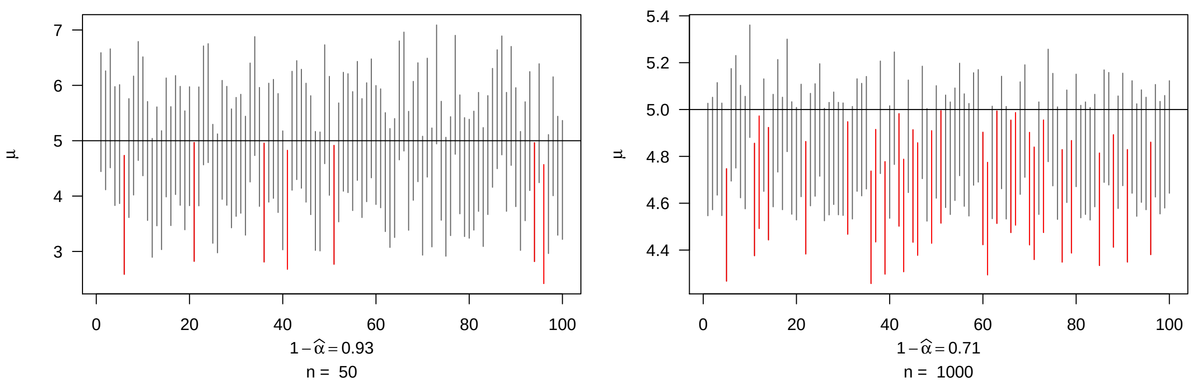


Obrázek 6: Pravděpodobnost pokrytí 95 % Waldových exaktních empirických DIS pro parametr μ normálního rozdělení při známém rozptylu σ^2

Zopakujeme-li několikrát simulační studii, vidíme, že pochází-li náhodné výběry z normálního rozdělení, pohybuje se aktuální pravděpodobnost pokrytí okolo nominální pravděpodobnosti pokrytí $1-\alpha$. V případě, že náhodné výběry pochází ze směsi dvou normálních rozdělení, které se liší pouze v rozptylech, dochází s rostoucím rozdílem mezi oběma rozptyly k postupnému snižování aktuální pravděpodobnosti pokrytí, a tedy ke zvýšení aktuální hladiny významnosti. Aby se však tento trend projevil, museli bychom hodnotu rozptylu σ_2^2 pozměnit opravdu výrazně (viz např. $\sigma_1^2 = 15$ vs $\sigma_2^2 = 15^2$). Podobná situace nastává v případě, že náhodné výběry pochází ze směsi dvou normálních rozdělení, které se liší střední hodnotou. Zde dochází k postupnému snižování aktuální pravděpodobnosti pokrytí, a tedy ke zvyšování aktuální hladiny významnosti s rostoucím rozsahem náhodných výběrů n.



Obrázek 7: Pravděpodobnost pokrytí 95 % Waldových exaktních empirických DIS pro parametr μ normálního rozdělení při známém rozptylu σ^2 za předpokladu, že náhodný výběr pochází ze smíšeného rozdělení $pN(5, 15) + (1-p)N(5, 15^2)$, $p = 0.9$



Obrázek 8: Pravděpodobnost pokrytí 95 % Waldových exaktních empirických DIS pro parametr μ normálního rozdělení při známém rozptylu σ^2 za předpokladu, že náhodný výběr pochází ze smíšeného rozdělení $pN(5, 15) + (1-p)N(3, 15)$, $p = 0.9$

Tabulka 1: Pravděpodobnost pokrytí 95 % Waldových empirických DIS pro parametr μ normálního rozdělení při neznámém rozptylu σ^2

	$1-\alpha$	$1-\widehat{\alpha}$	dh_α	hh_α
normálni r.: $n = 50$	0.9500	0.9900	0.9705	1.0095
normálni r.: $n = 1000$	0.9500	0.9300	0.8800	0.9800
smíšené r. s rozdílnými σ_1^2 a σ_2^2: $n = 50$	0.9500	0.7800	0.6988	0.8612
smíšené r. s rozdílnými σ_1^2 a σ_2^2: $n = 1000$	0.9500	0.7600	0.6763	0.8437
smíšené r. s rozdílnými μ_1 a μ_2: $n = 50$	0.9500	0.9300	0.8800	0.9800
smíšené r. s rozdílnými μ_1 a μ_2: $n = 1000$	0.9500	0.7100	0.6211	0.7989