Výpočet exaktných pravdepodobnostných rozdelení vybraných testov mnohorozmernej štatistickej analýzy[1]

A Note on Computing the Exact Distribution of Selected Multivariate Test Criteria

Viktor WITKOVSKÝ

Ústav merania SAV, Bratislava
witkovsky@savba.sk

ROBUST 2018

20. letní škola JČMF, Rybník, ČR, 21.-25. január 2018

[1] Práca bola finančne podporená projektom Agentúry na podporu výskumu a vývoja APVV-15-0295 a projektami VEGA 2/0054/18 a VEGA 2/0011/16.

Úvod

- Aplikácia metód štatistickej inferencie často vedie k neštandardným pravdepodobnostným rozdeleniam odhadov a testovacích štatistík.
- Typickým príkladom sú exaktné (*small sample*) rozdelenia testovacích štatistík mnohorozmernej štatistickej analýzy:
- Príklad:
  Ak $E \sim W_p(n, \Sigma)$ a $H \sim W_p(q, \Sigma)$ označujú nezávislé Wishartové matice, potom

  $$T_1 = \frac{|E|}{|\Sigma|} \sim Q_1 \times \cdots \times Q_p, \quad \text{and} \quad \Lambda = \frac{|E|}{|E+H|} \sim B_1 \times \cdots \times B_p,$$

  kde $Q_i \sim \chi^2_{n-p+i}$ a $B_i \sim Beta(\frac{n-p+i}{2}, \frac{q}{2})$ sú nezávislé náhodné premenné, $i = 1, ..., p$.
- Rozdelenie mnohých testovacích štatistík (za platnosti nulovej hypotézy) možno štrukturálne vyjadriť ako súčin alebo ako lineárnu kombináciu nezávislých náhodných premenných.
- Často platí, že charakteristická funkcia (CF) pravdepodobnostného rozdelenia testovacej štatistiky (za platnosti nulovej alebo alternatívnej hypotézy) je známa.

Úvod

- Cieľom tohto príspevku je spropagovať a povzbudiť ďalší výskum v oblasti metód a algoritmov založených na numerickej inverzii characteristickej funkcie.
- Tu uvedené príklady a vypočty sú ilustrované pomocou vytvoreného MATLAB toolboxu CharFunTool (The Characteristic Functions Toolbox), ktorý je v procese vývoja a prístupný na stránke GitHub:

  `https://github.com/witkovsky/CharFunTool/`

- Aplikovateľnosť metódy založenej na numerickom invertovaní charakteristickej funkcie budeme ilustrovať určením príslušných charakteristických funkcií a výpočtom exaktného pravdepodobnostného rozdelenia (nulové resp. alternatívne rozdelenie) niektorých testovacích kritérií používaných v mnohorozmernej štatistickej analýze:
  - Bartlettov test homogenity rozptylov nezávislých normálnych rozdelení,
  - Wilksovo Λ-rozdelenie a jeho aplikácie pri testovaní hypotéz v mnohorozmernej štatistickej analýze.

Bartlettov test homogenity rozptylov

- Bartlett (1937) navrhol modifikovanú LRT testovaciu štatistiku na test hypotézy o rovností rozptylov k normálnych populácií a jej približné rozdelenie.

- Bartlett navrhol korekciu LRT (multiplikatívny korekčný faktor), ktorá významne urýchľuje konvergenciu k asymptotickému χ^2_{k-1} rozdeleniu:

$$\chi^2 = \frac{\nu \log(S_p^2) - \sum_{i=1}^k \nu_i \log(S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{\nu_i} - \frac{1}{\nu} \right)} \overset{\text{approx}}{\sim} \chi^2_{k-1},$$

kde $S_i^2 = \frac{1}{\nu_i} \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)^2$, $S_p^2 = \frac{1}{\nu} \sum_{i=1}^k \nu_i S_i^2$, $\nu_i = n_i - 1$, $\nu = \sum_{i=1}^k \nu_i$.

- Exaktné rozdelenie Bartlettovej testovacej štatistiky (jeho analytické vyjadrenie) je neznáme. Bolo študované (okrem iných) v prácach Glaser (1976, 1976b) a Chao & Glaser (1978).

- Chao a Glaser dokázali, že exaktné rozdelenie Bartlettovej testovacej štatistiky za platnosti nulovej hypotézy (o rovnosti rozptylov) je v priamom vzťahu s rozdelením podielu (váženého) geometrického a aritmetického priemeru nezávislých gamma rozdelených náhodných premenných.

Rozdelenie podielu geometrického a aritmetického priemeru

- Nech R_w označuje podiel váženého geometrického priemeru a aritmetického priemeru nezávislých náhodných gamma náhodných premenných,

$$R_w = \frac{G_w}{A} = \frac{\prod_{l=1}^{k} X_l^{w_l}}{\frac{1}{k}\sum_{l=1}^{k} X_l}, \quad \text{pričom} \quad X_l \sim Gamma(\alpha_l, \beta)$$

kde α_l označuje parametre tvaru (shape) a β je spoločný parameter škály (scale).

- Chao a Glaser (1978) určili r-tý moment náhodnej premennej R_w

$$E\left(R_w^r\right) = k^r \frac{\Gamma\left(\sum_{l=1}^{k} \alpha_l\right)}{\Gamma\left(\sum_{l=1}^{k} \alpha_l + r\right)} \prod_{l=1}^{k} \frac{\Gamma\left(\alpha_l + r w_l\right)}{\Gamma\left(\alpha_l\right)}.$$

- Vo všeobecnosti, pre ľubovoľnú log-transformovanú nezápornú náhodnú premennú, napr. $Y = \log(X)$, platí, že jej CF možno vyjadriť pomocou formálneho výrazu pre r-tý moment premennej X substitúciou (ak existuje),

$$E\left(X^r\right) = E\left(e^{r\log(X)}\right) \Rightarrow E\left(e^{it\log(X)}\right) = \mathrm{cf}_{\log(X)}(t) = \mathrm{cf}_Y(t).$$

- Odtiaľ teda, CF log-transformovanej premennej $W = log(R_w)$ je

$$\mathrm{cf}_W(t) = k^{it} \frac{\Gamma\left(\sum_{l=1}^{k} \alpha_l\right)}{\Gamma\left(\sum_{l=1}^{k} \alpha_l + it\right)} \prod_{l=1}^{k} \frac{\Gamma\left(\alpha_l + i w_l t\right)}{\Gamma\left(\alpha_l\right)}.$$

Charakteristická funkcia Bartlettovej testovacej štatistiky

- Označme $X_l = \nu_l S_l^2 \overset{H_0}{\sim} \sigma^2 \chi^2_{\nu_l} \equiv \mathrm{Gamma}\left(\frac{\nu_l}{2}, \frac{1}{2\sigma^2}\right)$ fpre$l = 1, \ldots, k$. Bartlettova χ^2 štatistika je v priamom vzťahu s log-transformovaným podielom $W = log(R_w)$,

$$\chi^2 = \frac{c}{b} - \frac{\nu}{b} \log(R_w),$$

kde

  - $b = 1 + \frac{1}{3(k-1)} \left(\sum_{l=1}^{k} \frac{1}{\nu_l} - \frac{1}{\nu}\right)$ označuje Bartlettov korekčný faktor,
  - $c = \nu \log(c_w) = \nu \log(\frac{k}{\nu}) + \sum_{l=1}^{k} \nu_l \log(\nu_l)$ je konštanta závislá od váh (veľkosti nezávislých výberov).

- Odtiaľ CF Bartlettovej χ^2 testovacej štatistiky za platnosti nulovej hypotézy je

$$\mathrm{cf}_{\chi^2}(t) = e^{\mathrm{i}\frac{c}{b}t} k^{-\mathrm{i}\frac{\nu}{b}t} \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2} - \mathrm{i}\frac{\nu}{b}t\right)} \prod_{l=1}^{k} \frac{\Gamma\left(\frac{\nu_l}{2} - \mathrm{i}\frac{\nu_l}{b}t\right)}{\Gamma\left(\frac{\nu_l}{2}\right)}.$$

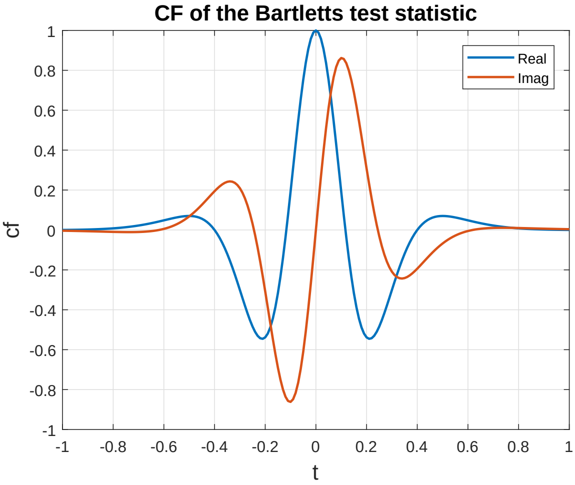
- Exaktné nulové rozdelenie (PDF/CDF) resp. kvantily rozdlenia Bartlettovej χ^2 testovacej štatistiky možno spočítať numericky z uvedenej charakteristickej pomocou vhodného algoritmu pre numerickú inverziu CF.

Numerický príklad: Exaktné rozdelenie Bartlettovej testovacej štatistiky

```
% Computing the exact distribution of the Bartlett's test statistic
  k             = 15;                                    % number of normal populations
  nu_l          = [1 1 1 1 1 2 2 2 2 2 3 3 3 3 3];       % sample degrees of freedom
  nu            = sum(nu_l);                             % total degrees of freedom
  alpha{1}      = nu_l/2;                                % alpha_l parameters
  weight{1}     = alpha{1}/sum(alpha{1});                % weights_
  c             = nu * log(k * prod(weight{1}.^weight{1})); % coefficient c
  b             = 1 + 1/(3*(k-1))*(sum(1./nu_l) - 1/nu);  % Bartlett's correction b
  shift         = c/b;
  coef          = -nu/b;

  % Characteristic function of the Bartlett's test statistic (16)
  cf_logR       = @(t) cf_LogRV_MeansRatioW(t,k,alpha,weight,coef);
  cf            = @(t) exp(1i*t*shift) .* cf_logR(t);

  % Evaluate the distribution function by using the algorithm cf2DistGP
  x             = linspace(0,40);
  prob          = [0.9 0.95 0.99];
  options.xMin = 0;
  result        = cf2DistGP(cf,x,prob,options);
```

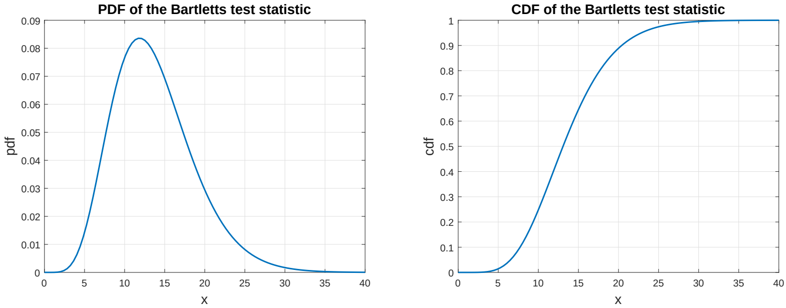
Obr.: MATLABovský kód (s využitím algoritmov v balíčku CharFunTool) na výpočet charakteristickej funkcie a exaktného nulového rozdelenia (PDF/CDF) Bartlettovej testovacej štatistiky pre test homogenity k normálnych populácií so špecifickou voľbou parametrov: $k = 15$ a $\nu_l(n_l - 1) \in \{1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3\}$.

Numerický príklad: Exaktné rozdelenie Bartlettovej testovacej štatistiky



Obr.: Charakteristická funkcia exaktného nulového rozdelenia Bartlettovej testovacej štatistiky pre test homogenity k normálnych populácií so špecifickou voľbou parametrov: $k = 15$ a $\nu_l(n_l - 1) \in \{1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3\}$.

Numerický príklad: Exaktné rozdelenie Bartlettovej testovacej štatistiky



Obr.: Exaktné nulové rozdelenie (PDF/CDF) Bartlettovej testovacej štatistiky pre test homogenity k normálnych populácií so špecifickou voľbou parametrov: $k = 15$ a $\nu_l(n_l - 1) \in \{1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3\}$.

- Pre špecifickú voľbu pravdepodobností 0.9, 0.95 and 0.99 dostávame vypočítané exaktné hodnoty kvantilov: $q_{0.9} = 20.3969$, $q_{0.95} = 22.8508$ a $q_{0.99} = 27.9221$.
- Hodnoty približných kvantilov, vypočítaných z približného asymptotického rozdelenia χ^2_{k-1} pre $k = 15$, sú: $\tilde{q}_{0.9} = 21.0641$, $\tilde{q}_{0.95} = 23.6848$ a $\tilde{q}_{0.99} = 29.1412$.

Charakteristická funkcia Wilksovho Lambda rozdelenia

- The Wilksová testovacia štatistika sa často využíva pre testovanie hypotéz v mnohorozmernej analýze, špeciálne v súvislosti s rôznymi LRT (testami pomerom vierohodností) a mnohorozmernou analýzou rozptylu (MANOVA).
- Wilksova Λ štatistika a jej exaktné nulové rozdelenie je

$$\Lambda = \frac{|E|}{|E+H|} \sim \prod_{j=1}^{p} B_j \equiv \Lambda(p,n,q),$$

kde $E \sim W_p(n,\Sigma)$ a $H \sim W_p(q,\Sigma)$ sú nezávislé matice s Wishartovým rozdelením a $B_j \sim \mathrm{Beta}\left(\frac{n-j+1}{2}, \frac{q}{2}\right)$ sú nezávislé beta rozdelené náhodné premenné.

- Vo všeobecnosti, analytický tvar distribučnej funkcie nie je známy resp. je komplikovaný, obzvlášť pre veľké hodnoty parametrov p a q, preto sú pre aplikácie často používé rôzne aproximácie, často známa aproximácia založená na asymptotických argumentoch:

$$-n\left(1 - \frac{p-q+1}{2n}\right)\log(\Lambda) \overset{n\to\infty}{\sim} \chi^2_{pq},$$

Charakteristická funkcia Wilksovho Lambda rozdelenia

- Exaktné rozdelenie log-transformovanej štatistiky $\lambda = -\log(\Lambda)$ možno s vysokou presnosťou vypočítať numerickou inverziou jej CF.

- Platí,

$$\mathrm{cf}_\lambda(t) = \mathrm{cf}_{\log(\Lambda)}(-t) = \prod_{j=1}^{p} \mathrm{cf}_{\log(B_j)}(-t) = \prod_{j=1}^{p} \frac{\Gamma\left(\frac{n-j+1}{2} - \mathrm{i}t\right)}{\Gamma\left(\frac{n-j+1}{2}\right)} \frac{\Gamma\left(\frac{n+q-j+1}{2}\right)}{\Gamma\left(\frac{n+q-j+1}{2} - \mathrm{i}t\right)},$$

  kde $\mathrm{cf}_{\log(B_j)}(t)$ označuje CF log-transformovanej náhodnej premennej $Y_j = \log(B_j)$, $B_j \sim \mathrm{Beta}\left(\frac{n-j+1}{2}, \frac{q}{2}\right)$, $j = 1, \ldots, p$.

- Výsledná CF bola odvodená na základe znalostí r-tého momentu beta rozdelenia $B \sim \mathrm{Beta}(\alpha, \beta)$:

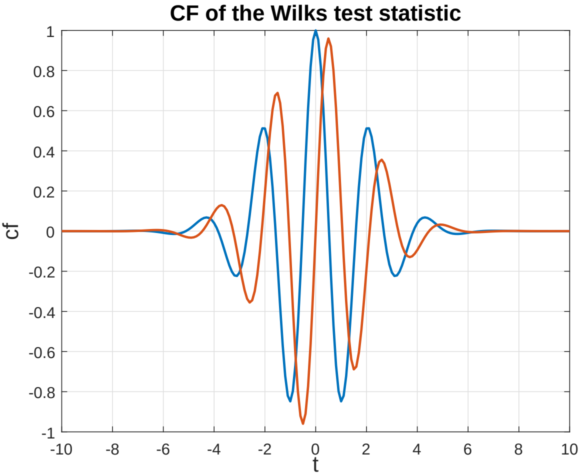
$$E\left(B^r\right) = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + r)},$$

Numerický príklad: Exaktné rozdelenie Wilksovej testovacej štatistiky

```
% Exact distribution of lambda = -log(Lambda) for testing equality of p-dimensional
% mean vectors of q normal populations with common unstructured covariance matrix

p    = 10;        % dimension of the normal populations
q    = 7;         % number of normal populations
n    = 30;        % total number of samples

% Characteristic function of lambda = -log(Lambda) with unstructured covariance matrix
cf   = @(t) cf_LogRV_WilksLambda(t,p,n-q,q-1,-1);

% Evaluate the distribution of -log(Lambda) by using the cf2DistGP
x            = linspace(0,6)';
prob         = [0.9 0.95 0.99];
options.xMin = 0;
result       = cf2DistGP(cf,x,prob,options);
```

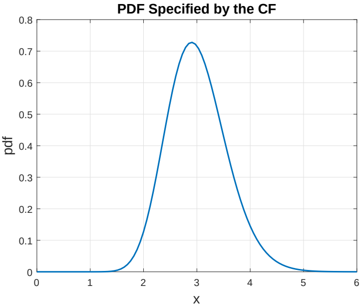
Obr.: MATLABovský kód (s využitím algoritmov balíčka CharFunTool) na výpočet charakteristickej funkcie a exaktného nulového rozdelenia (PDF/CDF) log-transformovanej Wilksovej testovacej štatistiky pre testovanie hypotézy o rovnosti vektorov stredných hodnôt q normálnych populácií so spoločnou neznámou neštrukturovanou kovariančnou maticou, s parametrami $p = 10$ (dimenzia vektorov), $q = 7$ (počet porovnávaných normálnych populácií), $n = 30$ (rozsah náhodných výberov spolu).
V tomto prípade, za platnosti nulovej hypotézy, má Wilksova testovacia štatistika rozdelenie $\Lambda(p, n-q, q-1)$.

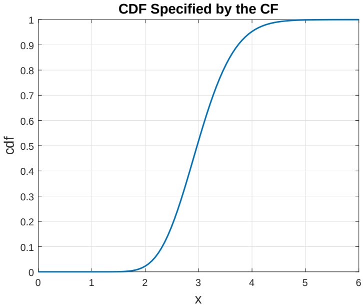
Numerický príklad: Exaktné rozdelenie Wilksovej testovacej štatistiky



Obr.: Charakteristická funkcia exaktného nulového rozdelenia log-transformovanej Wilksovej testovacej štatistiky (s rozdelením $\Lambda(p, n-q, q-1)$) pre testovanie hypotézy o rovnosti vektorov stredných hodnôt s parametrami $p = 10$ (dimenzia), $q = 7$ (počet populácií), $n = 30$ (rozsah výberov spolu).

Numerický príklad: Exaktné rozdelenie Wilksovej testovacej štatistiky





Obr.: Exaktné nulové rozdelenie log-transformovanej Wilksovej testovacej štatistiky (PDF/CDF) pre testovanie hypotézy o rovnosti vektorov stredných hodnôt. V tomto prípade, za platnosti nulovej hypotézy, má Wilksova testovacia štatistika rozdelenie $\Lambda(p, n-q, q-1)$ s parametrami $p = 10$ (dimenzia), $q = 7$ (počet populácií), $n = 30$ (rozsah výberov spolu).

- Pre špecifickú voľbu pravdepodobností 0.9, 0.95 and 0.99 dostávame vypočítané exaktné hodnoty kvantilov: $q_{0.9} = 3.7399$, $q_{0.95} = 3.9797$ a $q_{0.99} = 4.4573$.
- Hodnoty približných kvantilov, vypočítaných z približného asymptotického rozdelenia χ^2_{k-1} pre $k = 15$, sú: $\tilde{q}_{0.9} = 3.6291$, $\tilde{q}_{0.95} = 3.8577$ a $\tilde{q}_{0.99} = 4.3112$.

Exaktné rozdelenia mnohorozmerných testovacích kritérií za alternatívy

- Ako uviedol Mathai (1973), významný posun v oblasti určenia exaktných nenulových (alternatívnych) rozdelení testovacích štatistík v mnohorozmernej štatistickej analýze bol možný vďaka rozvoju špeciálnych funkcií s maticovým argumentom: hypergeometrických funkcií s maticovým argumentom a ďalších zovšeobecnených funkcií ako napr. Meijerových G-funktcií alebo Foxových H-funkcií.
- Zovšeobecnená hypergeometrická funkcia s maticovým argumentom je definovaná:

$$ {}_pF_q\left(a_1,\ldots,a_p; b_1,\ldots,b_q \,|\, X\right) = \sum_{k=1}^{\infty}\sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_p)_\kappa}{k!(b_1)_\kappa \cdots (b_q)_\kappa} C_\kappa(X) $$

  kde
  - $p \geq 0$ a $q \geq 0$ sú celočíselné parametre,
  - X je $n \times n$ symetrická matica s nezápornými vlastnými reálnymi číslami $x_1, x_2, \ldots, x_n$,
  - $\kappa = (\kappa_1, \kappa_2, \ldots)$ reprezentuje možné partície čísla k,
  - $(a)_\kappa$ a $(b)_\kappa$ reprezentujú zovšeobecnené Pochhammerové symboly,
  - $C_\kappa(X)$ je Jackova funkcia — symetrická, homogénna polynomická funkcia stupňa $|\kappa|$ v premenných $x_1, x_2, \ldots, x_n$ of X.
- Viac detailov a možné stratégie pre efektívny numerický výpočet zovšeobecnených hypergeometrických funkcií s maticovým argumentom možno nájsť napr. v práci Koev a Edelman (2006).

Charakteristické funkcie rozdelení testovacích štatistík za alternatívy

- V špeciálnych prípadoch je možné odvodiť explicitný tvar momentov testovacích štatistík za platnosti alternatívy:
- Napríklad r-tý moment Wilksovej zovšeobecnenej variancie $|S|$, kde S je matica s necentrálnym Wishartovým rozdelením s n stupňami voľnosti a maticovými parametrami Σ (kovariančná matica) a Ω (maticový parameter necentrality), $S \sim W_p(n, \Sigma, \Omega)$, je daný vzťahom

$$E\left(|S|^r\right) = \frac{\Gamma_p\left(\frac{n}{2} + r\right)}{\Gamma_p\left(\frac{n}{2}\right)} |2\Sigma|^r \exp\left(-\operatorname{trace}(\Omega)\right) {}_1F_1\left(\frac{n}{2} + r; \frac{n}{2} \,|\, \Omega\right),$$

kde $\Gamma_p(a) = \pi^{\frac{p(p-1)}{2}} \prod_{j=1}^{p} \Gamma\left(a - \frac{j-1}{2}\right)$ označuje mnohorozmernú gama funkciu.

- Odtiaľ dostávame charakteristickú funkciu log-transformovanej náhodnej premennej $W = -\log(|S|)$ je

$$\mathrm{cf}_W(t) = \frac{\Gamma_p\left(\frac{n}{2} - \mathrm{i}t\right)}{\Gamma_p\left(\frac{n}{2}\right)} |2\Sigma|^{-\mathrm{i}t} \exp\left(-\operatorname{trace}(\Omega)\right) {}_1F_1\left(\frac{n}{2} - \mathrm{i}t; \frac{n}{2} \,|\, \Omega\right).$$

- Požadované rozdelenie za alternatívy (PDF/CDF/QF) možno dostatočne presne vypočítať numerickým invertovaním CF.

Charakteristické funkcie rozdelení testovacích štatistík za alternatívy

- V prípade testovania hypotéz o regresných koeficientov je Wilksova Λ testovacia štatistika špecifikovaná maticami H a E. Vo všeobecnosti (za alternatívy) má matica H necentrálne Wishartovo rozdelenie s q stupňami voľnosti, kovariančnou maticou Σ, a maticovým parametrom necentrality $\Omega = \frac{1}{2}MM'\Sigma^{-1}$, kde $M = E(X)$, pričom $H = XX'$, teda $H \sim W_p(q, \Sigma, \Omega)$. Matica E ma centrálne Wishartovo rozdelenie s n stupňami voľnosti a kovariančnou maticou Σ, teda $E \sim W_p(n, \Sigma)$.
- Charakteristická funkcia log-transformovanej štatistiky $\lambda = -\log(\Lambda)$, odvodenej z r-tého momentu štatistiky Λ za platnosti alternatívy, viď. Constantine (1963), je daná vzťahom

$$\mathrm{cf}_\lambda(t) = \frac{\Gamma_p\left(\frac{n}{2} - \mathrm{i}t\right)}{\Gamma_p\left(\frac{n}{2}\right)} \frac{\Gamma_p\left(\frac{n+q}{2}\right)}{\Gamma_p\left(\frac{n+q}{2} - \mathrm{i}t\right)} \, {}_1F_1\left(-\mathrm{i}t; \frac{n+q}{2} - \mathrm{i}t \,|\, -\Omega\right).$$

- Distribúciu log-transformovanej testovacej štatistiky (PDF/CDF/QF) za alternatívy možno vypočítať numerickým invertovaním charakteristickej funkcie, s využitím algoritmov na výpočet zovšeobecnených hypergeometrických funkcií s maticovým argumentom, viď. napr. Koev and Edelman (2006), alebo s využitím ich vhodných aproximácií.
- Efektívny a presný výpočet zovšeobecnených hypergeometrických funkcií s maticovým argumentom zostáva stále veľkou výzvou. Presnosť a efektívnosť výpočtu exaktných rozdelení numerickým invertovaním CF závisí významne od kvality dostupných algoritmov.

Charakteristické funkcie rozdelení testovacích štatistík za alternatívy

- Vo všeobecnosti, pokiaľ sú známe momenty rozdelenia uvažovanej mnohorozmernej testovacej štatistiky za alternatívy, potom je možné priamo odvodiť charakteristickú funkciu log-transformovanej testovacej štatistiky, z ktorej možno následne vypočítať hodnoty jej distribučnej funkcie (PDF/CDF/QF) numerickým invertovaním tejto CF.

- Momenty mnohorozmerných testovacích štatistík boli študované v rozsiahlej štatistickej literatúre. Avšak pre mnohé dôležité testovacie štatistiky v mnohorozmernej analýze sú ich alternatívne rozdelenia resp. ich charakteristické funkcie neznáme, resp. výpočtovo zložité pre praktické použitie.

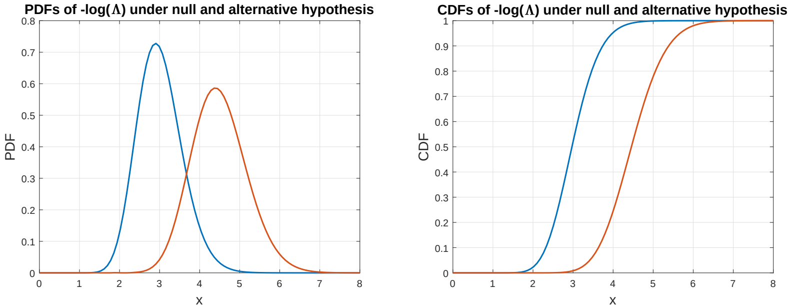
- Uvedené problémy zostávajú otvorené pre ďalší výskum.

Exaktné rozdelenie Wilksovej testovacej štatistiky za alternatívy

```
%  PDF/CDF of log-tranformed Wilks Lambda for testing equality of means with dimension p=10,
%  total sample size n=30, number of populations q=7. CF of the non-null distribution is
%  calculated by using the noncentrality matrix parameter Omega with nonzero eigenvalues
%  [10 5 4 3 2 1] and the algorithm Hypergeom1F1Mat.m with MAX = 30. This could take LONG time!

p      = 10;
n      = 30;
q      = 7;
Omega  = [10 5 4 3 2 1];

% CF under null hypothesis
cf_H0  = @(t) cf_LogRV_WilksLambdaNC(t,p,n-q,q-1,[],-1);
% CF under non-null (alternative) hypothesis specified by eigenvalues of matrix Omega
cf_HA  = @(t) cf_LogRV_WilksLambdaNC(t,p,n-q,q-1,Omega,-1,30);

x      = linspace(0,8);
prob   = [0.9 0.95 0.99];
options.xMin         = 0;
options.SixSigmaRule = 10;

% Evaluate the distribution of -log(Lambda) by using the cf2DistGP
result_H0 = cf2DistGP(cf_H0,x,prob,options);
result_HA = cf2DistGP(cf_HA,x,prob,options);
```

Obr.: MATLABovský kód (s využitím algoritmov balíčka CharFunTool) na výpočet charakteristických funkcií a exaktného nulového rozdelenia (PDF/CDF), za platnosti nulovej hypotézy a za platnosti alternatívy špecifikovanej vlastnými číslami matice Ω, log-transformovanej Wilksovej testovacej štatistiky pre testovanie hypotézy o rovnosti vektorov stredných hodnôt.

Exaktné rozdelenie Wilksovej testovacej štatistiky za alternatívy



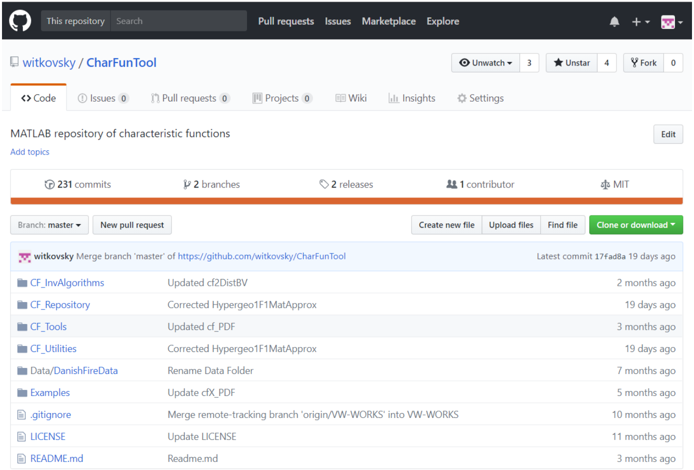
Obr.: Exaktné nulové a alternatívne rozdelenie log-transformovanej Wilksovej testovacej štatistiky (PDF/CDF) pre testovanie hypotézy o rovnosti vektorov stredných hodnôt. V tomto prípade, za platnosti nulovej hypotézy má Wilksova testovacia štatistika rozdelenie $\Lambda(p, n-q, q-1)$ s parametrami $p = 10$ (dimenzia), $q = 7$ (počet populácií), $n = 30$ (rozsah výberov spolu).

- Pre špecifickú voľbu pravdepodobností 0.9, 0.95 and 0.99 dostávame vypočítané exaktné hodnoty kvantilov: $q_{0.9} = 3.7399$, $q_{0.95} = 3.9797$ a $q_{0.99} = 4.4573$.
- Hodnoty kvantilov za platnosti špecifikovanej alternatívy sú: $\tilde{q}_{0.9} = 5.3858$, $\tilde{q}_{0.95} = 5.6692$ a $\tilde{q}_{0.99} = 6.2253$.

Zhrnutie

- Znalosť charakteristickej funkcie poskytuje úplnú charakterizáciu distribúcie rozdelenia testovacej štatistiky. Analytické invertovanie CF (pokiaľ je vôbec možné) často vedie ku komplikovaným a výpočtovo náročným výrazom na určenie hodnôt PDF/CDF resp. kvantilov daného rozdelenia.
- Ako alternatívu navrhujeme využitie metód a algoritmov pre numerické invertovanie charakteristických funkcií.
- Ako sa ukazuje, v mnohých praktických situáciach na presný a efektívny výpočet hodnôt distribučných funkcií často postačuje jednoduchá implementácia Gil-Pelaezovej inverzie CF na CDF pomocou aplikácie jednoduchého lichobežníkového (trapezoidal) integrálneho pravidla. Alternatívne a sofistikované metódy sú potrebné pre špeciálne situácie a pre výpočty vyžadujúce vysokú numerickú presnosť.
- Štandardné štatistické balíky (napr. R, SAS, MATLAB) neposkytujú dostatočné nástroje na výpočet, kombinovanie a numerické invertovanie charakteristických funkcií.
- Ďalší výskum je potrebný pre odvodenie charakteristických funkcií rozdelení testovacích štatistík v mnohorozmernej analýze za platnosti alternatívy, rozvoj metód a algoritmov pre presné a efektívne vyčíslenie (zovšeobecnených) špeciálnych funkcií v komplexnej premennej (resp. s maticovými parametrami), rozvoj metód a algoritmov pre integrovanie vysoko oscilačných komplexných funkcií pre numerické invertovanie charakteristických funkcií.

CharFunTool: The Characteristic Funtion Toolbox



CharFunTool: The Characteristic Functions Toolbox.

Balíček MATLAB algoritmov pre výpočet, kombinovanie a numerické invertovanie charakteristických funkcií: `https://github.com/witkovsky/CharFunTool`.

CharFunTool: The Characteristic Funtion Toolbox

No.	Probability Distribution Name	Notation	MATLAB Call Syntax
1	Arcsine (symmetric)	X ~ ArcsineSymmetric	cfS_Arcsine(t)
2	Beta (symmetric)	X ~ BetaSymmetric(theta)	cfS_Beta(t,theta)
3	Gaussian (symmetric)	X ~ N(0,1)	cfS_Gaussian(t)
4	Rectangular (symmetric)	X ~ RectangularSymmetric	cfS_Rectangular(t)
5	Student's t (symmetric)	X ~ t(df)	cfS_Student(t,df)
6	Trapezoidal (symmetric)	X ~ TrapezoidalSymmetric(lambda)	cfS_Trapezoidal(t,lambda)
7	Triangular (symmetric)	X ~ TriangularSymmetric(lambda)	cfS_Triangular(t)
8	Two Sided Power (symmetric)	X ~ TSPSymmetric(theta)	cfS_TSP(t,theta)
9	Normal	X ~ N(mu,sigma)	cf_Normal(t,mu,sigma,coef,niid)
10	Student's t	X ~ t(df)	cf_Student(t,df,coef,niid)
11	Beta	X ~ Beta(alpha,beta)	cf_Beta(t,alpha,beta,coef,niid)
12	Fisher-Snedecor F	X ~ F(df1,df2)	cf_FisherSnedecor(t,df1,df2,coef,niid,tol)
13	Chi-Square	X ~ ChiSquare(df,ncp)	cf_ChiSquare(t,df,ncp,coef,niid)
14	Inverse Gamma	X ~ InvGamma(alpha,beta)	cf_InverseGamma(t,alpha,beta,coef,niid)
15	Stable	X ~ Stable(alpha,mu,sigma)	cf_Stable(t,alpha,mu,sigma,coef,niid)
16	von-Mises	X ~ vonMises(mu,kappa)	cf_vonMises(t,mu,kappa,coef,niid)
17	Exponential	X ~ Exponential(lambda)	cfX_Exponential(t,lambda)
18	Log-Logistic	X ~ LogLogistic(alpha,beta)	cfX_LogLogistic(t,alpha,beta,tol)
19	Log-Normal	X ~ LogNormal(mu,sigma)	cfX_LogNormal(t,mu,sigma,tol)
20	Pareto	X ~ Pareto(alpha,sigma,type)	cfX_Pareto(t,alpha,sigma,type,tol)
21	Generalized Pareto	X ~ GeneralizedPareto(xi,sigma,theta)	cfX_GeneralizedPareto(t,xi,sigma,theta,tol)
22	Pearson type V	X ~ PearsonV(alpha,beta)	cfX_PearsonV(t,alpha,beta)
23	Pearson type VI	X ~ PearsonVI(alpha,beta)	cfX_PearsonVI(t,alpha,beta)
24	Weibull	X ~ Weibull(alpha,beta)	cfX_Weibull(t,alpha,beta,tol)

CharFunTool: The Characteristic Functions Toolbox.

Triedy distribučných funkcií, zvolená parametrizácia a MATLAB algoritmy na výpočet ich charakteristických funkcií v balíčku CharFunTool.

CharFunTool: The Characteristic Funtion Toolbox

No.	Probability Distribution Name	Notation	MATLAB Call Syntax
25	Log of Beta RV (random variable)	X ~ LogBeta(alpha,beta)	cf_LogRV_Beta(t,alpha,beta,coef,niid)
26	Log of Fisher-Snedecor F RV	X ~ LogFisher(df1,df2)	cf_LogRV_FisherSnedecor(t,df1,df2,coef,niid)
27	Log of Gamma RV	X ~ LogGamma(alpha,beta)	cf_LogRV_Gamma(t,alpha,beta,coef,niid)
28	Log of Chi-Square RV	X ~ LogChiSquare(df)	cf_LogRV_ChiSquare(t,df,coef,niid)
29	Log of Inverse Gamma RV	X ~ LogInvGamma(alpha,beta)	cf_LogRV_InverseGamma(t,alpha,beta,coef,niid)
30	Log of Means Ratio RV	X ~ LogMeansRatio(n,alpha)	cf_LogRV_MeansRatio(t,n,alpha,coef,niid)
31	Log of weighted Means Ratio RV	X ~ LogMeansRatioW(n,alpha,weight)	cf_LogRV_MeansRatioW(t,n,alpha,weight,coef,niid)
32	Log of Wilk's Lambda RV	X ~ LogLambda(p,m,n)	cf_LogRV_WilksLambda(t,p,m,n,coef,niid)
33	Dirac Mixture	X ~ DiracMixture(d,weight)	cfE_DiracMixture(t,d,weight,cfX)
34	Empirical	X ~ Empirical(data)	cfE_Empirical(t,data,cfX)
35	Empirical Ogive	X ~ EmpiricalOgive(bins,freq)	cfE_EmpiricalOgive(t,bins,freq,cfX)
36	Binomial	X ~ Binomial(n,p)	cfN_Binomial(t,n,p,cfX)
37	Delaporte	X ~ Delaporte(a,b,c)	cfN_Delaporte(t,a,b,c,cfX)
38	Generalized Poisson	X ~ GeneralizedPoisson(a,p)	cfN_GeneralizedPoisson(t,a,p,cfX)
39	Geometric	X ~ Geometric(p,type)	cfN_Geometric(t,p,type,cfX)
40	Logarithmic	X ~ Logarithmic(p)	cfN_Logarithmic(t,p,cfX)
41	Negative Binomial	X ~ NegativeBinomial(r,p)	cfN_NegativeBinomial(t,r,p,cfX)
42	Poisson	X ~ Poisson(lambda)	cfN_Poisson(t,lambda,cfX)
43	Polya-Eggenberger	X ~ PolyaEggenberger(a,b,m)	cfN_PolyaEggenberger(t,a,b,m,cfX)
44	Quinkert	X ~ Quinkert(a,b)	cfN_Quinkert(t,a,b,cfX)
45	Waring	X ~ Waring(a,b,r)	cfN_Waring(t,a,b,r,cfX)

CharFunTool: The Characteristic Functions Toolbox.
Triedy distribučných funkcií, zvolená parametrizácia a MATLAB algoritmy na výpočet ich charakteristických funkcií v balíčku CharFunTool.

LITERATÚRA

- M. S. Bartlett, Properties of sufficiency and statistical tests, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 160 (901) (1937) 268–282.
- M. S. Bartlett, A note on the multiplying factors for various χ^2 approximations, Journal of the Royal Statistical Society. Series B (Methodological) 16 (2) (1954) 296–298.
- A. M. Mathai, A review of the different techniques used for deriving the exact distributions of multivariate test criteria, Sankhyā: The Indian Journal of Statistics, Series A (1973) 39–60.
- R. E. Glaser, The ratio of the geometric mean to the arithmetic mean for a random sample from a gamma distribution, Journal of the American Statistical Association 71 (354) (1976) 480–487.
- R. E. Glaser, Exact critical values for Bartlett's test for homogeneity of variances, Journal of the American Statistical Association 71 (354) (1976) 488–490.
- M.-T. Chao, R. E. Glaser, The exact distribution of Bartlett's test statistic for homogeneity of variances with unequal sample sizes, Journal of the American Statistical Association 73 (362) (1978) 422–426.
- R. J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley & Sons, 2009.
- P. Koev, A. Edelman, The efficient evaluation of the hypergeometric function of a matrix argument, Mathematics of Computation 75 (254) (2006) 833–846.
- A. G. Constantine, Some non-central distribution problems in multivariate analysis, The Annals of Mathematical Statistics 34 (4) (1963) 1270–1285.