Self-Validated Computations for the Probabilites of the Central Bivariate Chi-Square Distribution

Kevin Wright
Iowa State University
October, 1997

Ph.D. advisor: William Kennedy

Iowa State Dept of Statistics
50th Anniversary Conference
October, 1997

Outline

- Introduction to Interval Analysis
- Bivariate Chi-Square Distributions
- Interval Bracket-Secant/Bisection
- Numerical results

Motivation

- Error analysis
- Cancellation
- Rounding errors

Solution? Interval analysis

Goal: Guaranteed error bounds

Interval Anaylsis (Moore, 1966, 1979)

A real interval $\mathbf{x}$ is defined $\mathbf{x} = [\underline{x}, \overline{x}]$ where $\underline{x} \in \Re, \overline{x} \in \Re, \underline{x} \leq \overline{x}$.

Let $\mathbf{x} = [\underline{x}, \overline{x}]$ and $\mathbf{y} = [\underline{y}, \overline{y}]$.

Arithmetic operations for intervals are defined:

$\mathbf{x} * \mathbf{y} = \{x * y : x \in \mathbf{x}, y \in \mathbf{y}\}$ for $* \in \{+, -, \times, \div\}$

Equivalently:

$$\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$$

$$\mathbf{x} - \mathbf{y} = [\underline{x} - \overline{y}, \overline{x} - \underline{y}]$$

$$\mathbf{x} \cdot \mathbf{y} = [\text{min}(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}), \text{max}(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y})]$$

$$1/\mathbf{y} = [1/\overline{y}, 1/\underline{y}], \quad 0 \notin \mathbf{y}$$

$$\mathbf{x}/\mathbf{y} = \mathbf{x} * (1/\mathbf{y}), \quad \notin \mathbf{y}$$

Examples:

$$[1,1]+[-2,5]=[-1,6]$$

$$[-2,3]*[1,4]=[-8,12]$$

Note: Subtraction and division are not the inverse of addition and multiplication.

$$[0,1]-[0,1]=[-1,1]$$

$$[1,2]/[1,2]=[0.5,2]$$

Theorem: Rational interval functions are inclusion monotone, i.e. $\mathbf{f}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{y})$ whenever $\mathbf{x} \subseteq \mathbf{y}$.

• IEEE Floating Point Specifications
Round to zero
Round to nearest
Round to $+\infty$
Round to $-\infty$

Compute $[1,1]/[3,3]$:
$[\bigtriangledown(1/3), \bigtriangleup(1/3)] = [0.333, 0.334] = [0.33^4_3]$

• Ensures guaranteed enclosure for rational interval functions.

• C++ BIAS/PROFIL (O. Knuppel, 1993)
Operator overloading : $\mathbf{x} + \mathbf{y}$
Rounding mode control

Dependency difficulties

Reduce the number of occurrences of a given interval in an expression.

- Ex: Suppose $\mathbf{x} = [-1, 2]$.

$\mathbf{x}^2 = \mathbf{x} \cdot \mathbf{x} = [-1, 2] \cdot [-1, 2] = [-2, 4]$

Fix this with an appropriate definition of $\mathbf{x}^2$

$\mathbf{x}^2 = \{x^2 : x \in \mathbf{x}\}$

$\mathbf{x}^2 = [0, 4]$

- Ex: Two extensions of $f(x) = x^2 - x$ are:

$\mathbf{f}_1(\mathbf{x}) = \mathbf{x}^2 - \mathbf{x}$

$\mathbf{f}_2(\mathbf{x}) = (\mathbf{x} - \frac{1}{2})^2 - \frac{1}{4}$

$\mathbf{f}_3(\mathbf{x}) = \mathbf{x}(\mathbf{x} - 1)$

$\mathbf{f}_1([0, 2]) = [-2, 4]$

$\mathbf{f}_2([0, 2]) = [-\frac{1}{4}, 2]$

$\mathbf{f}_3([0, 2]) = [-2, 2]$

The range of f over $[0, 2]$ is $[-\frac{1}{4}, 2]$.

Replacing each occurrence of x by $\mathbf{x}$ is called the *natural interval extension* of $f(x)$.

Bivariate Chi-Square Applications

- Simultaneous inferences for variances
- Simultaneous tests in ANOVA
- Simultaneous tests for goodness of fit
- Distribution of larger of correlated χ^2-variates
- Density of linear combination of independent χ^2-variates

See Gunst & Webster, 1973; Jensen & Howe, 1968

Bivariate Chi-square distribution : Case I
Gunst, 1973

$(Z_{1i}, Z_{2i}), i = 1, \ldots, m$ independent
$Z_{ij} \sim N(0,1)$
$Corr(Z_{1i}, Z_{2i}) = \rho, 1 = 1, \ldots, m$
$Y_i = \sum_{j=1}^{m} Z_{ij}^2, i = 1, 2$

$(Y_1, Y_2) \sim Biv\chi^2(m, m, m)$

Density

$$f(y_1, y_2) = (1-\rho^2)^{m/2} \sum_{j=0}^{\infty} \frac{\Gamma(\frac{m}{2}+j)\rho^{2j}}{j!\Gamma(\frac{m}{2})} \times \frac{(y_1 y_2)^{(m/2)+j-1} \exp[-(y_1+y_2)/2(1-\rho^2)]}{[2^{(m/2)+j}\Gamma(\frac{m}{2}+j)(1-\rho^2)^{(m/2+j)/2}]^2}$$

Bivariate Chi-square distribution : Case I

Distribution

$$P[Y_1 \le d_1, Y_2 \le d_2] = (1-\rho^2)^{m/2} \times \sum_{j=0}^{\infty} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})} \rho^{2j} \gamma(\frac{m}{2}+i, d_1^*) \gamma(\frac{m}{2}+i, d_2^*)$$

$\gamma(\alpha, d) = \int_0^d \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx$

$d_j^* = d_j/(1-\rho_{12}^2)$.

Let P_t be the sum truncated at t

Truncation error (Gunst, 1973)

$$0 \le R_t \le 1 - (1-\rho^2)^{m/2} \sum_{j=0}^{t} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})} \rho^{2j} \quad (1)$$

$\mathbf{p}_t, \mathbf{r}_t$ interval extensions of P_t, R_t

$P[Y_1 \leq d_1, Y_2 \leq d_2] = P_t + R_t$ for all t

$\in [\underline{\mathbf{p}}_t, \, \overline{\mathbf{p}}_t + \overline{\mathbf{r}}_t]$ for all t

•t determined by machine/software precision
• Stopping rules
$\mathbf{r}_{t-1} = \mathbf{r}_t$
$width(\mathbf{r}_t) < \epsilon$

Bivariate Chi-square distribution : Case II

$(Z_{1i}, Z_{2i}), i = 1, \ldots, m$ independent
$Z_{2i}, i = m+1, \ldots, m+n$
$Z_{ij} \sim N(0,1)$
$Corr(Z_{1i}, Z_{2i}) = \rho, 1 = 1, \ldots, m$
$Y_1 = \sum_{j=1}^{m} Z_{1j}^2, Y_2 = \sum_{j=1}^{m+n} Z_{2j}^2$

$(Y_1, Y_2) \sim Biv\chi^2(m, m+n, m)$

Distribution

$$P[Y_1 \leq d_1, Y_2 \leq d_2,] = (1-\rho^2)^{(m+n)/2} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})} \frac{\Gamma(\frac{n}{2}+k)}{k!\Gamma(\frac{n}{2})} \times \rho^{2(j+k)} \gamma(\frac{m}{2}+j, d_1^*) \gamma(\frac{n}{2}+k, d_2^*)$$

Truncation error

$$R_{t_1,t_2} \leq 1 - (1-\rho^2)^{m/2+n/2} \times \sum_{j=0}^{t_1} \sum_{k=0}^{t_2} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})} \frac{\Gamma(\frac{n}{2}+k)}{k!\Gamma(\frac{n}{2})} \rho^{2(j+k)}$$

$P[Y_1 \leq d_1, Y_2 \leq d_2] \in [\underline{\mathbf{p}}_{t_1,t_2}, \overline{\mathbf{p}}_{t_1,t_2} + \overline{\mathbf{r}}_{t_1,t_2}]$ for all pairs (t_1, t_2).

Bivariate Chi-square distribution : Case III

$(Z_{1i}, Z_{2i}), i = 1, \ldots, m$ independent
$Z_{1i}, i = m+1, \ldots, m+n$
$Z_{2i}, i = m+1, \ldots, m+p$
$Z_{ij} \sim N(0,1)$
$Corr(Z_{1i}, Z_{1i}) = \rho, 1 = 1, \ldots, m$
$Y_1 = \sum_{j=1}^{m+n} Z_{1j}^2, Y_2 = \sum_{j=1}^{m+p} Z_{2j}^2$

$(Y_1, Y_2) \sim Biv\chi^2(m+n, m+p, m)$

Distribution

$$P[Y_1 \le d_1, Y_2 \le d_2,] = (1-\rho^2)^{(m+n+p)/2} \times \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})}\frac{\Gamma(\frac{n}{2}+k)}{k!\Gamma(\frac{n}{2})}\frac{\Gamma(\frac{p}{2}+l)}{l!\Gamma(\frac{p}{2})} \times \rho^{2(j+k+l)}\gamma(\frac{m}{2}+\frac{n}{2}+k+j, d_1^*)\gamma(\frac{m}{2}+\frac{p}{2}+j+l, d_2^*)$$

Truncation error

$$R_t \leq 1 - (1 - \rho^2)^{m/2+n/2+p/2} \times \\ \sum_{j=0}^{t_1} \sum_{k=0}^{t_2} \sum_{l=0}^{t_3} \frac{\Gamma(\frac{m}{2}+j)}{j!\Gamma(\frac{m}{2})} \frac{\Gamma(\frac{n}{2}+k)}{k!\Gamma(\frac{n}{2})} \frac{\Gamma(\frac{p}{2}+l)}{l!\Gamma(\frac{p}{2})} \rho^{2(j+k+l)}$$

$P[Y_1 \leq d_1, Y_2 \leq d_2] \in [\underline{\mathbf{p}}_{t_1,t_2,t_3},\ \overline{\mathbf{p}}_{t_1,t_2,t_3} + \overline{\mathbf{r}}_{t_1,t_2,t_3}]$
for all triples (t_1, t_2, t_3)

Critical points of distributions

$P(Y_1 \leq x, Y_2 \leq x) - \alpha = 0$

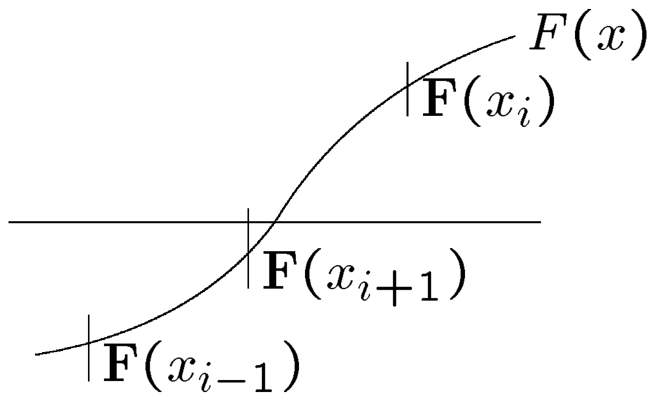
- Intervalized Newton methods
- Interval extension of F, f

Alternative: Derivative free method

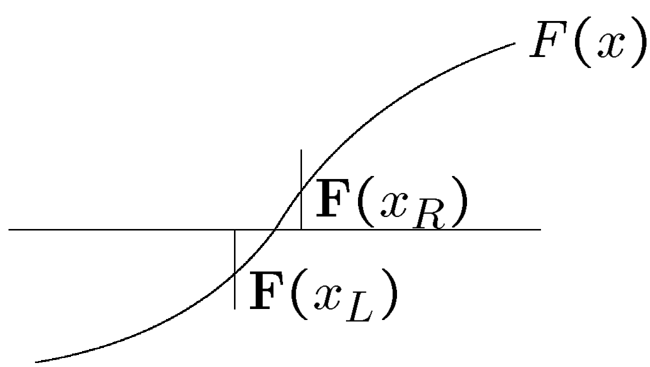
- Bracket-Secant / Illinois Modification

Bracket-Secant stops too soon

- Bisection



Termination of the bracket-secant



Termination of the bisection portion

Upper 0.05 Percentile points of the Bivriate Chi-Square Distribution : Case I

$$(Y_1, Y_2) \sim Biv\chi^2(m, m, m)$$

ρ	m = 2	m = 12
0.1	7.348735242636^{94}_{62}	23.291675614644^{9}_{3}
0.2	7.337736654468^{73}_{52}	23.279893907495^{5}_{0}
0.3	7.318116097295^{33}_{00}	23.257752618706^{8}_{3}
0.4	$7.28777721964^{231}_{197}$	23.22124105410^{20}_{13}
0.5	7.243389878426^{35}_{04}	23.16405309924^{80}_{71}
0.6	7.179739084402^{47}_{01}	23.07634122520^{18}_{07}
0.7	7.088168635581^{73}_{23}	22.94173918779^{26}_{10}
0.8	$6.95217862545^{561}_{462}$	22.72905146810^{28}_{03}
0.9	$6.73002568707^{695}_{497}$	22.35957467992^{6}_{0}

Upper 0.05 Percentile points of the Bivariate
Chi-Square Distribution : Case II
$(Y_1, Y_2) \sim Biv\chi^2(m, m+n, m)$

m	$m+n$	$\rho = 0.2$	$\rho = 0.4$
8	10	19.25562949840^{83}_{57}	$19.2145575272^{118}_{051}$
8	12	21.43987191019^{31}_{01}	$21.4161617406^{313}_{244}$
8	14	$23.8527709314^{125}_{093}$	$23.8412410690^{154}_{086}$
8	16	$26.3613220302^{700}_{665}$	$26.356264614^{8028}_{7942}$
8	18	28.89382780550^{93}_{57}	$28.8917353286^{768}_{657}$

m	$m+n$	$\rho = 0.6$	$\rho = 0.8$
8	10	$19.1173225955^{464}_{268}$	$18.90189596^{30341}_{29503}$
8	12	$21.3604107620^{625}_{460}$	$21.243609643^{9163}_{8388}$
8	14	$23.8140331030^{886}_{654}$	$23.759702671^{6459}_{5236}$
8	16	$26.3442149104^{735}_{419}$	$26.321129600^{3400}_{2338}$
8	18	$28.8866926661^{732}_{480}$	$28.877401302^{6241}_{4647}$

Upper 0.05 Percentile points of the Bivariate Chi-Square Distribution: Case III

$(Y_1, Y_2) \sim Biv\chi^2(m+n, m+p, m)$

m	n	p	$\rho = 0.4$	$\rho = 0.6$
7	1	11	$28.892120984^{7235}_{6487}$	$28.887920316^{7461}_{3450}$
6	2	12	$28.8924903043^{928}_{190}$	$28.88907530^{42914}_{39005}$
5	3	13	$28.892843220^{4233}_{3506}$	$28.890155051^{4641}_{1105}$
4	4	14	$28.893179676^{8562}_{7852}$	$28.891157355^{6196}_{1279}$
3	5	15	$28.893499628^{9161}_{8498}$	$28.89208036^{93269}_{88632}$
2	6	16	$28.893803042^{8767}_{7961}$	$28.892922583^{3886}_{0023}$
1	7	17	$28.894089895^{9045}_{8055}$	$28.89368281^{14038}_{09926}$

Conclusions

- Theoretical error analysis
- Rounding errors
- Cancellation

- Interval analysis
- Guaranteed error bound
- Improved Bivariate Chi-Square values
- Corrections to previous tables

- Further research in statistics

- Univariate Distributions : Wang & Kennedy, 1994