بسم الله الرحمن الرحیم

دانشکده علوم ریاضی

گروه آمار

برآورد و آزمون فرضیه‌ی استقلال در توابع مفصل براساس اندازه‌های اطلاع

ارائه شده برای دریافت درجه کارشناسی ارشد

در رشته‌ی آمار ریاضی

استاد راهنما

دکتر محمد امینی

استاد مشاور

دکتر غلامرضا محتشمی برزادران

توسط

سمانه خسروی

آبان ۹۰

تقدیـــــــم به

آستان ملکوتی ثامن الحجج

علـــــی ابـــن موســی الــــرضا(ع)

و

پـدر و مادر عــــزیــزم

قـدردانــــی

سپاس خداوندی را که سخنوران از ستودن او عاجزند و حسابداران از شمارش نعمت‌های او ناتوان و تلاشگران از ادای حق او درمانده‌اند. خدایی‌که افکار ژرف‌اندیش ذات او را درک نمی‌کنند و دست غواصان علم به او نخواهد رسید.[1]

بعد از حمد و ثنای یگانه‌ی بی‌همتا و درود بر آخرین فرستاده‌اش حضرت محمد مصطفی (ص)، وظیفه‌ی خود می‌دانم که از اساتید گرانقدرم تشکر و قدردانی نمایم.

اول از همه، از استاد راهنمای عزیزم جناب آقای **دکتر محمد امینی** تشکر می‌کنم که با صبر، حوصله، دقت نظر و ظرافت کم نظیری در تهیه این پایان‌نامه نقش برجسته‌ای داشتند. هم‌چنین از جناب آقای دکتر **غلامرضا محتشمی برزادران** بسیار سپاسگزارم که راهنمایی‌های ارزنده و مشاوره‌ی ایشان در جهت بهبود مطالب این پایان‌نامه مؤثر بوده است.

هم‌چنین تشکر می‌کنم از،

اساتید داور جناب آقای **دکتر عبدالحمید رضایی رکن آبادی** و جناب آقای **دکتر هادی جباری نوقابی** که داوری این پایان‌نامه را بر عهده داشتند و ما را از نظرات سازنده خود بهره‌مند ساختند.

خانواده عزیزم بخصوص **همسرم** که همواره مشوقی برای کسب علم بوده‌اند.

کارشناس محترم کتابخانه **سرکار خانم صادقی** و **جناب آقای داودنژاد** که در تهیه مقالات کمک نموده‌اند.

مهلا قاسم نژاد دوست بسیار عزیزم که با تشویق و راهنمایی، یاریم رساند.

---

[1] فرازهایی از نهج البلاغه

چکیده

در این پایان‌نامه به معرفی یک رده‌ی کلی از اندازه‌های وابستگی موسوم به اندازه‌های ϕ–واگرا که فاصله‌ی بین توزیع توأم و توزیع استقلال عناصر بردار توأم را اندازه می‌گیرد، می‌پردازیم، سپس ویژگی‌های این اندازه‌ها را بیان و آن‌ها را بر مبنای مفصل معرفی می‌کنیم. سپس معیارهای بهینگی اندازه‌های وابستگی را معرفی می‌نماییم. در ادامه پس از معرفی اصول رنی، برقراری این اصول را روی این اندازه‌ها بررسی می‌کنیم. معرفی برخی از اندازه‌های اطلاع، بدست آوردن این اندازه‌ها بر مبنای مفصل، برآورد اطلاع متقابل به‌کمک چند روش و مقایسه‌ی برخی از روش‌های برآورد، از دیگر مباحثی است که در این پایان‌نامه مطرح می‌شود. در پایان براساس معیارهای بهینگی، اندازه مفصل ϕ–واگرای بهینه برای مفصل نرمال، FGM و برخی از تعمیم‌های FGM تعیین می‌کنیم. برآورد اندازه‌های ϕ–واگرا و آزمون فرضیه‌ی استقلال برای این اندازه‌ها براساس مفصل به‌روش مونت کارلو از نتایج دیگر این پایان‌نامه است.

بسمه تعالي

Graduate Studies Dissertation Information
Ferdowsi University of Mashhad

Title of Dissertation: Estimation and test of independence in copula models using Information measures

Author: Samane Khosravi
Supervisor: Dr.M.Amini
Advisor: Dr.Gh.R.Mohtashami Borzadaran

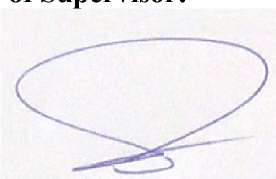
Faculty: Mathematical of Sciences	**Department:** Statistics	**Specialization:** Mathematical Statistics

Approval Date: 9 July 2011	**Defence Date:** 3 November 2011
M.Sc. ● **Ph.D.** ○	**Number of Pages:** 119

Abstract:

In this thesis, we study measures of multivariate dependence based on the phi-divergence of the joint distribution of a random vector and the distribution that corresponds to independence of the components of the vector, the product of the marginals. Properties of these measures are also investigated and we get them via copula. Moreover, we introduce some Information measures and obtain these measures based on copula. Estimating Mutual Information using some method, comparison of method for estimating is other issues to be discussed in this thesis. Finally, based on optimaization's axioms, we determine optimaized phi-divergence measure for normal copula, FGM and some of extensions of FGM. Estimation and test of independence based on copula via Monte Carlo methods is another result this thesis.

Signature of Supervisor:



Date: 9 November 2011

Key Words:

1. Information measures

2. phi-divergance measures

3. Axiomatic framework of Renyi

4. Mutual Information

5. Empirical copula

پیش‌گفتار

در مدت زمان طولانی آماردانان علاقه‌مند به بررسی رابطه‌ی بین توابع توزیع چند متغیره و توابع توزیع حاشیه‌ای آن‌ها بودند. بررسی این مسأله برای حالتی که توابع توزیع حاشیه‌ای یک متغیره می‌باشند، توسط اسکلار(۱۹۵۹) انجام شد. بر این اساس اسکلار کلاس جدیدی از توابع موسوم به تابع مفصل را معرفی نمود. بعد از آن، فریز و والدز (۱۹۹۷) و نلسن (۲۰۰۶) مفصل را به‌عنوان ابزاری که رابطه‌ی بین برآمدهای چند متغیره را نشان می‌دهد، بررسی نمودند. اولین بار سیزر (۱۹۶۳) اندازه‌های ϕ−واگرا را برای تعیین وابستگی بین توزیع توأم و حاصلضرب حاشیه‌ای معرفی نمود، در ادامه میچز و زاگرافوس (۲۰۰۶) ویژگی‌های این اندازه‌ها را مورد بررسی قرار دادند و برای توزیع نرمال اندازه‌ی ϕ−واگرای بهینه را تعیین نمودند و آزمون فرضیه استقلال اندازه‌های ϕ− واگرا را بر مبنای توزیع نرمال انجام دادند. در ادامه محتشمی برزادران و امینی (۲۰۱۰) این اندازه‌ها را بر حسب مفصل بیان کردند. هم چنین ما و سان (۲۰۰۸) اطلاع متقابل را براساس مفصل تجربی برآورد نمودند و در ادامه کلساورینی و ویسنت (۲۰۰۹) و ما و سان (۲۰۱۱) رابطه‌ی بین مفصل و آنتروپی تحقیق نموده، آنتروپی مفصل را معرفی کرده و ثابت کردند آنتروپی مفصل برابر منفی اطلاع کولبک−لایبلر می‌باشد. هدف اصلی این پایان نامه مقایسه‌ی اندازه‌های ϕ−واگرا برای برخی مفصل‌ها و برآورد و آزمون فرضیه‌ی استقلال اندازه‌های ϕ−واگرا بر مبنای مفصل‌ها می‌باشد. در تبیین ضرورت استفاده از اندازه‌های ϕ−واگرا برای آزمون فرضیه‌ی استقلال، می‌توان گفت استقلال برقرار است اگر و تنها اگر اندازه وابستگی ϕ−واگرا برابر صفر باشد.

این پایان‌نامه شامل ۵ فصل می‌باشد که خلاصه مطالب هر فصل به شرح زیر است:

- در فصل ۱، ویژگی‌های اندازه‌های ϕ−واگرا و اندازه‌های اطلاع و مشخصه‌سازی وابستگی تصادفی براساس آنتروپی را بیان می‌کنیم.
- فصل ۲ شامل اندازه‌های ϕ−واگرا و اندازه‌های اطلاع بر حسب مفصل و استفاده از مفصل در

مشخصه‌سازی وابستگی تصادفی براساس آنتروپی می‌باشد.

- برآورد اندازه‌های اطلاع در حالت کلی و برحسب مفصل و برآورد مفصل در مشخصه‌سازی وابستگی تصادفی براساس آنتروپی در فصل ۳ بیان می‌شود.
- در فصل ۴، اصول رنی را روی اندازه‌های ϕ—واگرا بررسی می‌کنیم. در ادامه اندازه‌های ϕ—واگرا برای برخی مفصل‌ها مقایسه می‌شوند.
- در فصل ۵ آزمون فرضیه استقلال اندازه‌های ϕ—واگرا را برمبنای مفصل‌ها، به‌روش مونت کارلو انجام می‌دهیم.

لازم به ذکر است که در این پایان نامه علامت $*$ در ابتدای زیر بخش به معنای جدید بودن کل آن زیر بخش می‌باشد.

سمانه خسروی

آبان ۱۳۹۰

فهرست مندرجات

فهرست جداول

فهرست نمودارها

نمادها

آنتروپی شانون	$Shanon\ entropy$	$H_e(X) \equiv H(X)$
اندازه‌های ϕ−واگرا	$\phi - Divergance$	$D_\phi(f \parallel g) \equiv D_\phi(\tilde{X})$
اطلاع متقابل	$Mutual\ information$	$D_{KL}(f \parallel g) \equiv \delta(\tilde{X}) \equiv MI$
اطلاع متقابل درجه ۲	$Quadratic\ mutual\ information$	$K(X,Y)$
آنتروپی مفصل	$Copula\ entropy$	$H_c(\tilde{X})$
فاصله‌ی هلینجر	$Hellinger\ Distance$	$D_H(f \parallel g)$
فاصله‌ی $\chi^۲$	$\chi^۲ - Divergance$	$D_{\chi^۲}(f \parallel g)$
فاصله‌ی α−واگرا	$\alpha - Divergance$	$D_\alpha(f \parallel g)$
فاصله‌ی جفری	$J - Divergance$	$D_J(f \parallel g)$
نسخه‌ی ترکیبی α−واگرا	$Combination\ of\ version\ of\ \alpha - Divergance$	$D_{C\alpha}(f \parallel g)$
فاصله‌ی هارمونیک	$Harmonic\ Distance$	$D_{Ha}(f \parallel g)$
فاصله‌ی مثلثی	$Triangular\ Discrimination$	$D_\Delta(f \parallel g)$
فاصله‌ی لی و ونگ	$Lei\ and\ Wang\ Divergance$	$D_{LW}(f \parallel g)$
اندازه‌ی کریس و رد	$Cressie\ and\ Read\ Measure$	$D_{CR}(f \parallel g)$
استقلال	$Independence$	$=$
تکین	$Singular$	$\perp$
نزدیک‌ترین همسایه‌ی k	$k - nearest\ neighbour$	knn

فصل ۱

اندازه‌های وابستگی ϕ—واگرا و اندازه‌های اطلاع

۱—۱ مقدمه

۱—۲ اندازه‌های وابستگی ϕ—واگرا

۱—۳ اندازه‌های اطلاع

۱—۴ مشخصه‌سازی وابستگی تصادفی به‌کمک آنتروپی

۱—۵ نتیجه‌گیری

۱—۱ مقدمه

مسأله وابستگی دو (چند) متغیره همواره مورد توجه محققین بوده است. زیرا استقلال در داده‌ها یک ویژگی مطلوب است. در ابتدای این فصل اندازه‌های وابستگی ϕ—واگرا را که فاصله‌ی بین توزیع توأم یک بردار تصادفی و حاصلضرب توزیع‌های حاشیه‌ای را بیان می‌کند، معرفی نموده و ویژگی‌ها و روابط بین آن‌ها را بیان می‌کنیم. در ادامه اندازه‌های اطلاع که محور اصلی تحقیق می‌باشند را معرفی نموده و به معرفی آنتروپی نسبی در حالت چند متغیره می‌پردازیم و در پایان وابستگی تصادفی را برحسب آنتروپی مشخصه‌سازی می‌کنیم. در تدوین مطالب این فصل از منابع جو (۱۹۸۹)، میچز و زاگرافوس (۲۰۰۶) و کوواچ (۲۰۰۶) استفاده شده است.

۱—۲ اندازه‌های وابستگی ϕ—واگرا

اندازه‌های وابستگی ϕ-واگرا اولین بار توسط سیزر(۱۹۶۳) معرفی شد و در ادامه افرادی از قبیل علی و سیلوی (۱۹۶۵)، کنت (۱۹۸۳)، وجدا(۱۹۸۹) و پاردو (۲۰۰۶) پژوهش‌هایی در مورد این اندازه‌ها انجام دادند.

فرض کنید $\tilde{X} = (X_۱, X_۲..., X_n)$ یک بردار تصادفی روی فضای اندازه‌ی حاصلضربی(χ, A, μ) باشد که $\chi = (\chi_۱, \chi_۲, ..., \chi_n)$ ، $\mu = (\mu_۱, \mu_۲, ..., \mu_n)$ و $A = (A_۱, A_۲, ..., A_n)$ و برای $i = ۱, ۲, ..., n$، χ_i : فضای اقلیدسی ، A_i : σ—جبر مجموعه‌های بورلی و μ_i : اندازه لبگ می‌باشد.

همچنین فرض کنید $f = f(\tilde{x})$ چگالی توأم $\tilde{X}$ نسبت به μ بوده و برای $i = ۱, ۲, ..., n$ ، f_i چگالی حاشیه‌ای x_i نسبت به μ_i باشد، آن‌گاه یک رده‌ی کلی از اندازه‌های ϕ—واگرا برای اندازه‌گیری وابستگی توأم به‌صورت زیر معرفی می‌شود:

$$D_\phi(f \parallel g) = \int_\chi g \ \phi(\frac{f}{g})d\mu, \quad \phi \in \Phi^*$$

که در آن g توزیع استقلال نمونه‌ای میان X_1، X_2،...،X_n به‌صورت $g=\prod_{i=1}^{n} f_i$ بوده و Φ^* رده‌ی همه‌ی توابع محدب ϕ روی $[0,\infty)$ است، که ϕ در شرایط زیر صدق می‌کند:

$$0\phi(\frac{0}{0}) = 0 \qquad , \qquad 0\phi(\frac{t}{0}) = t\; lim_{u\to\infty}\frac{\phi(u)}{u}, \tag{۱–۱}$$

با انتخاب شکل‌های مختلفی از ϕ، اندازه‌های واگرایی معروفی بدست می‌آید که نتایج در جدول ۱–۱ بیان شده است:

ردیف	عنوان	$\phi(u)$	اندازه های واگرایی
۱	اطلاع کولبک –لایبلر(D_{KL})	$u\log(u)$	$\int_\chi f\log(\frac{f}{g})d\mu$
۲	فاصله‌ی χ^2 (D_{χ^2})	$\frac{(u-1)^2}{u} or\ u^2-1$	$\int_\chi \frac{(f-g)^2}{f}d\mu$
۳	فاصله‌ی هلینجر(D_H)	$(\sqrt{u}-1)^2$	$\int_\chi(\sqrt{f}-\sqrt{g})^2 d\mu$
۴	فاصله‌ی α– واگرا(D_α)	$\frac{4}{1-\alpha^2}(1-u^{-(\frac{1+\alpha}{2})})$	$\frac{4}{1-\alpha^2}\int_\chi(1-(\frac{g}{f})^{\frac{1+\alpha}{2}})f\ d\mu$
۵	فاصله‌ی جفری(D_J)	$(u-1)\log(u)$	$\int_\chi(f-g)\log(\frac{f}{g})d\mu$
۶	نسخه‌ی ترکیبی α– واگرا$(D_{C\alpha})$	$\frac{4}{\beta^2}(1-u^{\frac{\beta}{2}})^2$	$\frac{4}{\beta^2}\int_\chi(1-(\frac{f}{g})^{\frac{\beta}{2}})^2 g\ d\mu$
۷	فاصله‌ی هارمونیک(D_H)	$\frac{2u}{u+1}-u$	$\int_\chi \frac{2f}{f+g}d\mu$
۸	فاصله‌ی مثلثی(D_Δ)	$\frac{(u-1)^2}{u+1}$	$\int_\chi \frac{(f-g)^2}{f+g}d\mu$
۹	فاصله‌ی لی و ونگ (D_{LW})	$u\log(\frac{2u}{u+1})-\frac{1}{2}(u-1)$	$\int_\chi f\log(\frac{2f}{f+g})d\mu$
۱۰	اندازه‌ی کریس و رد(D_{CR})	$\frac{u^{\lambda+1}-u-\lambda(u-1)}{\lambda(\lambda+1)}$	$\frac{1}{\lambda(\lambda+1)}\int_\chi f((\frac{f}{g})^\lambda-1)d\mu$

جدول ۱–۱. جدول اندازه‌های واگرایی .

۱–۲–۱ ویژگی‌های اندازه‌ی کریس و رد

کریس و رد (۱۹۸۴) اندازه‌ای را معرفی نمودند که ویژگی‌های جالبی دارد. همان‌طور که در جدول ۱–۱ ملاحظه نمودید، این خانواده به‌صورت زیر مشخص می‌شود:

$$D_{CR}(f \parallel g) = \frac{۱}{\lambda(\lambda+۱)} \int_{\chi} f((\frac{f}{g})^{\lambda} - ۱) d\mu$$

مهم‌ترین اندازه‌ی ϕ–واگرا اندازه‌ی کریس و رد می‌باشد که مقادیر $\lambda = -۲, -۱, -\frac{۱}{۲}, ۱$ به‌ترتیب فاصله‌ی کای دو نیمن، کولبک–لایبلر، مربع فاصله‌ی هلینجر و کای دو پیرسن را نتیجه می‌دهد: زیرا اگر $\lambda = -۲, -۱, -\frac{۱}{۲}, ۱$ باشد، آن‌گاه

$$\begin{aligned} \phi_{-۲}(u) &= \frac{u^{-۲+۱} - u + ۲(u-۱)}{-۲(-۲+۱)} \\ &= \frac{۱}{۲}(u + \frac{۱}{u} - ۲), \end{aligned}$$

بنا به قاعده‌ی هوپیتال داریم

$$\begin{aligned} \phi_{-۱}(u) &= lim_{\lambda \to -۱} \frac{u^{\lambda+۱} - u - \lambda(u-۱)}{\lambda(\lambda+۱)} \\ &= lim_{\lambda \to -۱} \frac{\lambda u^{\lambda+۱}(\log u) - (u-۱)}{۲\lambda+۱} \\ &= u \log u - u + ۱, \end{aligned}$$

$$\begin{aligned} \phi_{-\frac{۱}{۲}}(u) &= \frac{u^{-\frac{۱}{۲}+۱} - u + \frac{۱}{۲}(u-۱)}{-\frac{۱}{۲}(-\frac{۱}{۲}+۱)} \\ &= ۴(u^{\frac{۱}{۲}} + \frac{۱}{۲}u - \frac{۱}{۲}) \\ &= ۲(\sqrt{u} - ۱)^{۲}, \end{aligned}$$

$$\begin{aligned} \phi_{۱}(u) &= \frac{۱}{۲}(u^{۲} - u - u + ۱) \\ &= \frac{۱}{۲}(u-۱)^{۲}. \end{aligned}$$

۱–۲–۲ روابط بین اندازه‌های ϕ–واگرا

بروفکوف (۱۹۹۸) و محتشمی برزادران و امینی(۲۰۱۰) روابط بین برخی از اندازه‌های ϕ–واگرا را بررسی نمودند. ما در این بخش به بیان این روابط می‌پردازیم.

$$D_{KL}(f \parallel g) \approx \frac{1}{2} D_{\chi^2}(f \parallel g) \quad (1)$$

$$D_J(f \parallel g) \approx \frac{1}{2}[D_{\chi^2}(f \parallel g) + D_{\chi^2}(g \parallel f)] \quad (2)$$

$$D_{\chi^2}(f \parallel g) \approx 4 D_H(f \parallel g) \quad , \quad D_{\chi^2}(f \parallel g) \geq D_H(f \parallel g) \quad (3)$$

$D_{C\alpha}$ و D_α به‌صورت زیر رابطه دارند:

$$D_{C\alpha}(f \parallel g) = 2(\frac{2}{\beta} - 1)A + 4(1 - \frac{1}{\beta})B \quad (4)$$

که در آن A و B به‌ترتیب $D_\alpha(f \parallel g)$ با $\alpha = 1 - \beta$ و $\alpha = 1 - 2\beta$ می‌باشد.

$$D_{\chi^2}(f \parallel g) = D_{C\alpha}(f \parallel g) \quad , \quad D_{\chi^2}(f \parallel g) = 2 D_\alpha(f \parallel g) \quad (5)$$

(که در آن $\alpha = 3$ در اندازه‌ی α–واگرا و $\beta = 2$ در اندازه‌ی نسخه‌ی ترکیبی α–واگرا)

$$D_H(f \parallel f) = \frac{1}{4} D_{C\alpha}(f \parallel g) \quad , \quad D_H(f \parallel g) = \frac{1}{2} D_\alpha(f \parallel g) \quad (6)$$

(که در آن $\alpha = 0$ در اندازه‌ی α–واگرا و $\beta = 1$ در اندازه‌ی نسخه‌ی ترکیبی α–واگرا)

فاصله هلینجر متقارن است و همه‌ی ویژگی‌های یک متر را دارد هم‌چنین $D_{Ha}(f \parallel g)$ و $D_\Delta(f \parallel g)$ نیز متقارن هستند و

$$D_{LW}(f \parallel g) + D_{LW}(g \parallel f) \leq D_\Delta(f \parallel g) \quad (7)$$

اکنون به بیان اثبات روابط فوق می‌پردازیم.

اثبات (۱):

$$\begin{aligned} D_{KL}(f \parallel g) &= \int_\chi f \log \frac{f}{g} d\mu = -\int_\chi f \log \frac{g}{f} d\mu \\ &= -\int_\chi f \log(1 + (\frac{g}{f} - 1)) d\mu, \end{aligned}$$