توزیع بتای دو متغیره جدید

نجمه السادات موسوی احمدآبادی

دانشگاه پیام نورشیراز

n.moosavi44@gmail.com

چکیده

یک توزیع بتای دو متغیره جدید و مناسب با توانایی بالا که در مقایسه با تمام رقبای قبلی خود بهترین برازش را داشته باشد معرفی می کنیم. نمایشهای متنوع و مختلف از گشتاورهای حاصل ضربی، تابع چگالی احتمال حاشیه‌ای، گشتاورهای حاشیه‌ای، تابع چگالی شرطی و همچنین توابع مولد گشتاور شرطی ارائه شده است. روش حداکثر درست نمایی جهت برآورد آن به کار گرفته شده و در آخر یک مجموعه شش تایی از داده های کاربردی همراه با معیارهای مربوطه، به عنوان مثال توضیح داده شده است.

کلمات کلیدی: برآورد، گرده افشانی. ، شمارش .گشتاورهای حاصل ضربی

۱. مقدمه

توزیع بتا یکی از توزیع های آماری انعطاف پذیر می باشد که در بسیاری از علوم از جمله اقتصاد، عمران، روانشناسی، ادبیات و ... کاربرد دارد. تعمیم توزیع بتا به حالت چند متغیره نیز مطرح شده و کاربردهای وسیعی در بحث استنباط آماری دو متغیره دارد.

برای توزیع های جدید بد ست آمده نیز خواص مختلفی مانند تابع چگالی احتمال، تابع توزیع، گ شتاور، تخمین در ست نمایی ماک سیمم و ... مورد بررسی قرار گرفته است. در بحث تعمیمی توزیع بتای دو متغیره جدید عباراتی برای چگالی های حاشیه ای، امید ریاضی شرطی، توابع مولد گشتاور شرطی و حاصل ضربی بدست آورده شده است. توزیع های بتای دو متغیره زیادی تا کنون مطرح شده اند اما در ادبیات و مقاله های آماری، توزیع های بتای دو متغیری کمی یافت می شود به گونه ای که کارکردن با آن راحت باشد. فصل ۹ مرجع [۹] و فصل چهار مرجع [۳] را ملاحظه کنید. می توان گفت که بهترین توزیع های بتای دو متغیری تا قبل از سال ۲۰۰۰ در این مراجع معرفی شده اند. به نظر می رسد اولین توزیع بتای دو متغیره ، توزیع دریخله باشد. کانور و ماسیمان [۴] اولین کسانی بودند که توزیع دریخله را تعمیم دادند. لیبی و نویک [۵] توزیع بتای چند متغیره را بسط و توسعه دادند. توزیع های بتای دو متغیره ارائه شده توسط جنر [۷] و الکین و لیو [۸] توزیع جدیدی نمی باشد، آنها در حقیقت حالت خاصی از توزیع های بتای چند متغیره لیبی و نویک می باشند. سه تابع توزیع ارائه شده توسط ناداراژه و کتس [۹] مبنای توصیف خاصیت حاصل ضرب متغیرهای تصادفی بتا به شمار می رود. دو تا از این توابع توزیع دارای تابع چگالی احتمال (pdfs) می باشند. سارابیا و کاستیلو [۱۴] چندین توزیع بتای دو متغیره با سه پارامتر اضافه را توسعه دادند. ناگار و اروزکو [۱۵] با استفاده از توابع اپل نوع دوم از تابعهای فوق هندسی گوس، توزیع بتای دو متغیره دیگری با تابع چگالی احتمال را توسعه دادند. توزیع های بتای دو متغیره ارائه شده توسط باسیونی و جنز [۱۷] نیز توزیع جدیدی به شمار نمی رود و ظاهری مشابه توزیعهای ارائه شده توسط ناگار و اروزکو [۱۵] دارد.

توزیع بتای دو متغیره معرفی شده توسط آرنولد [۱۸] جهت معرفی مدلی جهت نمایش همبستگی مثبت یا منفی مجموعه داده ها مورد استفاده قرار می گیرد، اما توابع چگالی احتمال آن به صورت انتگرال سه گانه که به صورت فرم غیر بسته است، می باشد. گاپتا [۲۰] توزیع بتای دو متغیره غیر مرکزی ارائه نمود با تابع چگالی احتمال شامل توابع فوق هندسی. ژاکوب [۲۲] با الحاق توزیع بتای دو متغیره کومر همراه با تابع چگالی

احتمال به صورت سری نامتناهی از توابع، یک توزیع بتای دو متغیره دیگری را معرفی کرد. همچنین الکین و تریکالینوس [۲۳] یک توزیع بتای دو متغیره با تابع چگالی احتمال از توابع اپل نوع اول را بنا نهاد. از توزیع های بتای دو متغیره که دارای توابع چگالی احتمال بوده و در ابتدا از داده ها و نقاط واقع در ناحیه مربع $[0,1] \times [0,1]$ استفاده می کنند، از توزیع های بتای لیبی و نویک [۵] و ناداراژه [۱۲و۱۳] نشات می گیرند. داده های مورد نیاز می تواند از ناحیه ساده $\{(x,y):\ x > 0,\ y > 0\ ,x + y\ < 1\}$ انتخاب و مورد استفاده قرار گیرند مانند توزیع بتای دریخله.

در این مقاله ما قصد داریم یک توزیع بتای دو متغیره جدیدی ارائه کنیم با توابع چگالی احتمال مربوطه. ما نشان می دهیم که شش مجموعه داده حقیقی با این توزیع می تواند از تمام توزیع های بتای دو متغیره شناخته شده تاکنون، برازش بهتری داشته باشد. در ادامه به معرفی و بررسی کامل این توزیع جدید پرداخته می‌شود.

توزیع جدید یک توزیع شش پارامتری بوده با تابع چگالی احتمال توام تعریف شده به صورت زیر:

$$f(x,y) = \frac{c\, x^{\alpha-1}(1-x)^{\gamma-\alpha-1}(1-y)^{\gamma'-\beta-1}}{(1-ux)^{\gamma+\gamma'-\alpha-1}(1-vy)^{\gamma+\gamma'-\beta-1}(1-ux-vy)^{\alpha+\beta-\gamma-\gamma'+1}} \tag{1}$$

برای $0 < x < 1\ , 0 < y < 1\ , 0 < \alpha < \beta\ , 0 < \beta < \gamma'\ , 0 \le u < 1\ \ , 0 \le v < 1\ \ ,\ \ 0 \le u + v\ < 1$

توضیح اینکه مقدار ثابت c به گونه ای تعیین می شود که حاصل انتگرال تابع چگالی احتمال یک باشد. شش پارامتر مجهول عبارتند از $\alpha\,,\beta\,,\gamma\,,\gamma',u\,,v$

c را می‌توان با استفاده از رابطه زیر تعیین کرد:

$$\frac{1}{c} = \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma-\alpha)\Gamma(\gamma'-\beta)}{\Gamma(\alpha)\Gamma(\gamma')}\ .F_4\ (\alpha,\beta;\gamma,\gamma';u(1-v),v(1-u)\ )$$

تابع $F_4\ (\mathrm{a},\mathrm{b}\ ;\mathrm{c},c';\ z,\xi\ )$ تابع اپل نوع چهارم بوده و به صورت زیر بیان می گردد.

$$F_4\ (\mathrm{a},\mathrm{b}\ ;\mathrm{c},c';\ z,\xi\ ) = \sum_{k=0}^{\infty} \quad \sum_{l=0}^{\infty} \frac{(a)_{k+l}(b)_{k+l}\, z^k\, \xi^l}{(c)_k(c')_l\, k!l!}$$

و $(f)_\mathrm{k}$ به صورت زیر تعریف می شود:

$$.\quad (f)_k = f\ (f+1) \ldots (\ f+k-1)$$

حال با میل دادن $x \to 0$ می توان نوشت :

$$f(x,y) \sim c\ x^{\alpha-1}\ \frac{y^{\beta-1}(1-y)^{\gamma'-\beta-1}}{(1-vy)^{\alpha}}$$

حال مشابه حالت قبل این بار با میل دادن $y \to 0$ و گرفتن حد تابع خواهیم داشت:

$$f(x,y) \sim c\ y^{\beta-1}\ \frac{x^{\alpha-1}(1-x)^{\gamma-\alpha-1}}{(1-ux)^{\beta}}$$

اجازه دهید راجع به شکل و نمودار معادله (۱) بحث کوتاهی دا شته با شیم. با م شتق گرفتن از $\log f$ نسبت به متغیرهای x,y می توان چنین بدست آورد:

$$\frac{\partial\ \log \mathrm{f}}{\partial\ \mathrm{x}} = \frac{\alpha-1}{x} - \frac{\gamma-\alpha-1}{1-x} + \frac{u(\gamma+\gamma'-\alpha-1)}{1-ux} + \frac{u(\alpha+\beta-\gamma-\gamma'+1)}{1-ux-vy} \tag{2}$$

مشابها داریم :

$$\frac{\partial\ \log \mathrm{f}}{\partial\ \mathrm{y}} = \frac{\beta-1}{y} - \frac{\gamma\prime-\beta-1}{1-\gamma} + \frac{v(\gamma+\gamma'-\beta-1)}{1-vy} + \frac{v(\alpha+\beta-\gamma-\gamma'+1)}{1-ux-vy} \tag{3}$$

حال با صفر قرار دادن رابطه های (۲) و (۳) و حل همزمان معادلات درجه سه زیر نقاط بحرانی معادله (۱) بدست می آید:

$$u^2(2-\gamma+\beta)x^3 + u(\alpha u - u - \beta u + vy - 4 - \beta + 2\gamma + \gamma' vy - \alpha vy)x^2 + (\gamma vy - 2\alpha u + \beta u + 2u + 2 - 2vy + 2\alpha uvy - \gamma - \gamma uvy - \gamma' uvy)x + (1-\alpha)(vy-1) = 0$$

بنابر این معادله (۱) می تواند با نه نقطه بحرانی معرفی و تعیین شود. شکلهای ۱ و ۲ برخی از حالتهای ممکن معادله (۱) را به تصویر می کشد. ملاحظه می شود که معادله (۱) می تواند متقارن، پادمتقارن یا متمایل به گو شه های ناحیه مربعی با شد و تصویر یا نمایشی از وابستگی قوی و ضعیف آن دو باشد.حال کارهای باقیمانده که می بایست در این مقاله انجام گردد عبارتند از ارائه گشتاورهای حاصل ضربی، توابع چگالی حاشیه ای، گشتاورهای حاشیه ای، توابع چگالی احتمال شرطی و توابع مولد گشتاور شرطی مربوط به معادله (۱) . در ادامه به مباحث اشاره شده پرداخته می شود.لازم به ذکر است که از روش برآورد حداکثر درستنمایی برای ماتریس اطلاع مشاهده شده استفاده شده که در بخش شش در مورد آن بحث خواهد شد.

۲. گشتاورهای حاصل ضربی

قضیه زیر با استفاده از سریهای نامتناهی مضاعف از توابع فوق هندسی گوس، روش محاسبه و بدست آوردن گشتاورهای حاصل ضربی مربوط به معادله (۱) را بیان می کند.

قضیه ۲-۱ : گشتاورهای حاصل ضربی X, Y مربوط به معادله (۱) بوسیله فرمول زیر محاسبه می گردد.

$$E(X^m Y^n) = C.B(\beta+n, \gamma'-\beta).B(\alpha+m, \gamma-\alpha).\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(\beta+m)_{k+l}(\gamma+\gamma'-\beta-1)_l v^{k+1}}{(n+\gamma')_{k+l} K!l!}$$

$$.(\alpha+\beta-\gamma-\gamma'+1)_l \ . \ {}_2F_1(\alpha+m, l+\beta; m+\gamma; u) \tag{4}$$

برای هر عدد حقیقی $m, n > 0$. $B(a,b)$ ، ${}_2F_1(a, b; c; x)$ تابع بتا و تابع فوق هندسی گوس بوده که به ترتیب به صورت زیر تعریف می شوند:

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt \tag{5}$$

و

$${}_2F_1(a, b; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k x^k}{(c)_k k!} \tag{6}$$

اثبات : با توجه به تعریف امید ریاضی می توان نوشت :

$$E(X^m Y^n) = C\int_0^1 \int_0^1 \frac{x^{m+\alpha-1} y^{n+\beta-1} (1-x)^{\gamma-\alpha-1} (1-y)^{\gamma'-\beta-1}}{(1-ux)^{\gamma+\gamma'-\alpha-1}(1-vy)^{\gamma+\gamma'-\beta-1}(1-ux-vy)^{\alpha+\beta-\gamma-\gamma'+1}}\, dy\, dx \tag{7}$$

با کمی دقت به انتگرال داخلی که نسبت به المان y می باشد و عنایت به روابط (۵) و (۶) می توان نوشت:

$$E(X^m Y^n) = C.B(\beta+n, \gamma'-\beta) \int_0^1 \frac{x^{m+\alpha-1}(1-x)^{\gamma-\alpha-1}}{(1-ux)^{\beta}} .F_1(\beta+n, \gamma+\gamma'-\beta-1, \alpha+\beta-\gamma-\gamma'$$

$$+1, n+\gamma'; v, \frac{v}{1-ux})\, dx \tag{8}$$

در اینجا تابع F_1 تابع اپل نوع اول تعریف شده به صورت زیر می باشد:

$$F_1(a, b, b', c; z, \xi) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(a)_{k+l}(b)_k (b')_l z^k \xi^l}{(c)_{k+l} k!l!} \tag{9}$$

با جایگذاری رابطه اخیر در رابطه (۸) و بازنویسی مجدد می توان نوشت:

$$E(X^m\,Y^n) = \ C.B(\beta + n\,,\gamma' - \beta)\ \int_0^1 \frac{x^{m+\alpha-1}\ (1-x)^{\gamma-\alpha-1}}{(1-ux)^{\beta}}.$$

$$\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\beta+n)_{k+l}(\gamma+\gamma'-\beta-1)_k\,(\alpha+\beta-\gamma-\gamma'+1)_l\,v^{k+1}}{(n+\gamma\prime)_{k+l}\,k!l!\,(1-ux)^l}\ \ dx$$

$$=\ C.B(\beta + n\,,\gamma' - \beta).\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\beta+n)_{k+l}(\gamma+\gamma'-\beta-1)_k\,(\alpha+\beta-\gamma-\gamma'+1)_l\,v^{k+1}}{(n+\gamma\prime)_{k+l}\,k!l!}$$

$$.\int_0^1 \frac{x^{m+\alpha-1}\ (1-x)^{\gamma-\alpha-1}}{(1-ux)^{l+\beta}}\ \mathrm{dx} \qquad (10)$$

۳. توابع چگالی احتمال و توابع مولد گشتاور حاشیه ای

قضایای ۳-۱ و ۳-۲ که در زیر آمده است با استفاده از تابع اپل نوع اول و سری نامتناهی مضاعف تابع فوق هندسی گوس به محاسبه توابع چگالی و گشتاورهای حاشیه ای می پردازد.

قضیه ۳-۱: اگر f تابع چگالی احتمال توام X,Y معادله (۱) باشد در این صورت تابع چگالی حاشیه ای X ,Y به صورت زیر بدست می آید:

$$f_{\mathrm{X}}\,(\mathrm{X}) = \ \mathrm{C.B}(\beta\,,\gamma' - \beta)\mathrm{x}^{\alpha}\ \ (1-\mathrm{x})^{\gamma-\alpha-1}\,(1-\mathrm{ux})^{-\beta}$$

$$.\,F_1\,\left(\beta\,,\gamma+\gamma'-\beta-1\,;\,\alpha+\beta-\gamma-\gamma'+1\,,\gamma';\,v\,,\frac{v}{1-ux}\,\right) \qquad (11)$$

و

$$f_{\mathrm{Y}}\,(\mathrm{Y}) = \ \mathrm{C.B}(\alpha\,,\gamma - \alpha)\mathrm{y}^{\beta-1}\,(1-\mathrm{y})^{\gamma\prime-\beta-1}\,(1-\mathrm{vy})^{-\alpha}$$

$$.\,F_1\,\left(\alpha\,,\gamma+\gamma'-\alpha-1\,;\,\alpha+\beta-\gamma-\gamma'+1\,,\gamma\ \ ;\,u\,,\frac{u}{1-vy}\,\right) \qquad (12)$$

وقتی که $0 < y < 1\ \ ,\ \ 0 < x < 1$ و $B(a,b)$ تابع بتا تعریف شده بوسیله رابطه (۵) و F_1 تابع اپل نوع اول تعریف شده توسط معادله (۹) می باشد.

قضیه ۳-۲ : توابع مولد گشتاور حاشیه ای و توابع چگالی احتمال حاشیه ای مربوط به معادلات (۱۱) و (۱۲) به صورت زیر می باشند:

$$E(X^m\,) = \ C.B(\beta\,,\gamma' - \beta)B(\alpha + m\,,\gamma - \alpha)\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\beta)_{k+l}\,(\gamma+\gamma'-\beta-1)_l\ v^{k+l}}{(\gamma\prime)_{k+l}\,k\,!l!}$$

$$.\,(\alpha+\beta-\gamma-\gamma'+1)_l\ \ {}_2F_1\,(\alpha+m\,,l+\beta\,;m+\gamma\,;u\,)$$

و همچنین داریم :

$$E(Y^n\,) = \ C.B(\beta + n\,,\gamma' - \beta)B(\alpha\,,\gamma - \alpha)\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\beta+n)_{k+l}\,(\gamma+\gamma'-\beta-1)_l\ v^{k+l}}{(n+\gamma\prime)_{k+l}\,k\,!l!}$$

$$.\,(\alpha+\beta-\gamma-\gamma'+1)_l\ \ {}_2F_1\,(\alpha\,,l+\beta\,;\gamma\,;u\,)$$

وقتی که $n > 0\,,m > 0$ و $B(a,b)$ تابع بتا تعریف شده بوسیله رابطه (۵) و ${}_2\mathrm{F}_1$ تابع فوق هندسی گوس تعریف شده توسط معادله (۱۶) می باشد.

$$f_X(x) = \ C.B(\beta\,,\gamma' - \beta)\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\beta)_{k+l}\,(\gamma+\gamma'-\beta-1)_k\ \ (\alpha+\beta-\gamma-\gamma'+1)_l\,v^{k+l}}{(\gamma')_{k+l}\,k\,!l!\,(1-ux)^{\beta+l}}\ .x^{\alpha-1}\,(1-x)^{\gamma-\alpha-1}$$

و

$$f_Y(y) = \ C.B(\alpha\,,\gamma - \alpha)\sum_{k=0}^{\infty}\quad \sum_{l=0}^{\infty} \frac{(\alpha)_{k+l}\,(\gamma+\gamma'-\alpha-1)_k\ \ (\alpha+\beta-\gamma-\gamma'+1)_l\,u^{k+l}\,y^{\beta}(1-y)^{\gamma'-\beta-1}}{(\gamma)_{k+l}\,k\,!l!\,(1-vy)^{\alpha+l}}$$

ملاحظه می شود که با قرار دادن $u=0$ (و به نحو مشابه $v=0$) معادله (۱۲) و به نحو مشابه معادله (۱۱) ، تابع بتای لیبی و نویکس حاصل می‌شوند.

۴. توابع چگالی احتمال و توابع مولد گشتاور شرطی

قضایای ۴-۱ و ۴-۲ توابع چگالی احتمال و توابع مولد گشتاور و حاشیه ای مربوط به معادله (۱) را با استفاده از تابع اپل نوع اول، ارائه می‌دهد.

قضیه ۴-۱: اگر f تابع چگالی احتمال توام X,Y مربوط به معادله (۱) باشند، تابع چگالی احتمال X با شرط $Y=y$ به صورت زیر محاسبه می‌گردد.

$$f_{X|Y}(x|y) = \frac{(1-vy)^{1+\alpha+\beta-\gamma-\gamma\prime}}{B(\alpha,\gamma-\alpha)F_1\left(\alpha,\gamma+\gamma'-\alpha-1\,;\alpha+\beta-\gamma-\gamma'+1,\gamma\,;u,\frac{u}{1-vy}\right)} \cdot \frac{x^{\alpha-1}(1-x)^{\gamma-\alpha-1}}{(1-ux)^{\gamma+\gamma'-\alpha-1}\,(1-ux-vy)^{\alpha+\beta-\gamma-\gamma'+1}} \tag{14}$$

وقتی که $0 < x < 1$ و $B(a,b)$ تابع بتا تعریف شده بوسیله رابطه (۵) و F_1 تابع اپل نوع اول تعریف شده توسط معادله (۹) می باشد.

همچنین تابع چگالی احتمال شرطی Y به شرط $X=x$ به صورت زیر می باشد:

$$f_{Y|X}(y|x) = \frac{(1-ux)^{1+\alpha+\beta-\gamma-\gamma\prime}}{B(\beta,\gamma\prime-\beta)F_1\left(\beta,\gamma+\gamma'-\beta-1,\alpha+\beta-\gamma-\gamma'+1\,;\gamma\prime\,;v,\frac{v}{1-ux}\right)} \cdot \frac{y^{\beta-1}(1-y)^{\gamma\prime-\beta-1}}{(1-vy)^{\gamma+\gamma'-\beta-1}\,(1-ux-vy)^{\alpha+\beta-\gamma-\gamma'+1}} \tag{15}$$

برای $0 < y < 1$

قضیه ۴-۲: تابع مولد گشتاور شرطی از توابع چگالی احتمال شرطی معادلات (۱۴) و (۱۵) به صورت زیر محاسبه می‌گردد.

$$E(\mathrm{X}^{\mathrm{m}}|y) = \frac{B(\alpha+m,\gamma-\alpha)F_1\left(\alpha+m,\gamma+\gamma'-\alpha-1\,;\alpha+\beta-\gamma-\gamma'+1,\gamma+m;u,\frac{u}{1-vy}\right)}{B(\alpha,\gamma-\alpha)F_1\left(\alpha,\gamma+\gamma'-\alpha-1\,;\alpha+\beta-\gamma-\gamma'+1,\gamma\,;u,\frac{u}{1-vy}\right)} \tag{16}$$

و به نحو مشابه می توان نوشت :

$$E(\mathrm{Y}^{\mathrm{n}}|x) = \frac{B(\beta+n,\gamma'-\beta)F_1\left(\beta+n,\gamma+\gamma'-\beta-1,\alpha+\beta-\gamma-\gamma'+1\,;\ \gamma\prime+n;v,\frac{v}{1-ux}\right)}{B(\beta,\gamma\prime-\beta)F_1\left(\beta,\gamma+\gamma'-\beta-1,\alpha+\beta-\gamma-\gamma'+1\,;\gamma\prime\,;v,\frac{v}{1-ux}\right)} \tag{17}$$

وقتی که $0 < n > 0\,,m$ و $B(a,b)$ تابع بتا تعریف شده بوسیله رابطه (۵) و F_1 تابع اپل نوع اول تعریف شده توسط معادله (۶) می‌باشد
.

اثبات :

$$E(\mathrm{X}^{\mathrm{m}}|y) = \frac{(1-vy)^{1+\alpha+\beta-\gamma-\gamma\prime}}{B(\alpha,\gamma-\alpha)F_1\left(\alpha,\gamma+\gamma'-\alpha-1,\alpha+\beta-\gamma-\gamma'+1\,;\gamma\,;u,\frac{u}{1-vy}\right)} \cdot \int_0^1 \frac{\mathrm{x}^{\mathrm{m}+\alpha-1}\,(1-\mathrm{x})^{\gamma-\alpha-1}}{(1-\mathrm{ux})^{\gamma+\gamma'-\alpha-1}\,(1-\mathrm{ux}-\mathrm{vy})^{\alpha+\beta-\gamma-\gamma'+1}}\,\mathrm{dx} \tag{18}$$

۵. راه حلهای دیگر جهت توابع چگالی احتمال توام حاشیه ای

ما در اینجا با بازنویسی بعضی از عبارات، تابع چگالی احتمال توام $f(x,y)$ را به صورت دیگری محاسبه می کنیم همراه با تابع چگالی احتمال حاشیه ای X و تابع چگالی احتمال حاشیه ای Y از آن . با کمی دقت می توان دید که :

$$1-ux-vy = (1-ux)(1-vy) - uvxy$$

بنابراین تابع چگالی احتمال توام $f(x,y)$ به صورت زیر بدست می‌آید:

$$f(x,y) = C\ \frac{x^{\alpha-1}y^{\beta-1}(1-x)^{\gamma-\alpha-1}\ (1-y)^{\gamma'-\beta-1}}{(1-ux)^{\beta}\ (1-vy)^{\alpha}}\ .[1-\frac{uvxy}{(1-ux)(1-vy)}]^{-(\alpha+\beta-\gamma-\gamma'+1)}$$

بنابراین ملاحظه می شود که تابع چگالی احتمال توام به صورت ترکیب خطی از دو تابع چگالی احتمال مستقل بتای لیبی و نویک بوده ، همچنین توابع چگالی حاشیه ای X نیز به صورت ترکیب از تابع چگالی بتای لیبی و نویک می باشد. به نحو مشابه توابع چگالی حاشیه ای Y نیز به صورت ترکیب از تابع چگالی بتای لیبی و نویک می باشد. و نهایتا، ثابت C به صورت یک سری نامتناهی شامل ضرب دو تابع فوق هندسی گوس نوشته شد.

۶. برآورد

توزیع پارامتری داده شده در معادله (۱) با پارامترهای $\alpha, \beta, \gamma, \gamma', u, v$ می باشد. در اینجا ما برآورد حداکثر درستنمایی را با محاسبه ماتریس اطلاع مشاهده شده مد نظر قرار می‌دهیم. فرض کنید $(x_1\ ,y_1)\ ...\ (x_n\ ,y_n)$ یک نمونه تصادفی از معادله (۱) باشد. با محاسبه لگاریتم درستنمایی ماکسیمم داریم:

$$\log L\ (\alpha,\beta,\gamma,\gamma',u,v) = n\ \log C + (\alpha-1)\sum_{j=1}^{n}\log x_j + (\beta-1)\sum_{j=1}^{n}\log y_j + (\gamma-\alpha-1)\sum_{j=1}^{n}\log(1-x_j)$$
$$+(\gamma'-\beta-1)\sum_{j=1}^{n}\log(1-y_j) + (\gamma+\gamma'-\alpha-1)\sum_{j=1}^{n}\log(1-ux_j) - (\gamma+\gamma'-\beta-1)\sum_{j=1}^{n}\log(1-vy_j)$$
$$-(\alpha+\beta-\gamma-\gamma'+1)\sum_{j=1}^{n}\log(1-ux_j-vy_j)$$

حال باید نسبت به هریک از پارامترها مشتق گیری انجام گیرد مثلا:

$$\frac{\partial\ \log L}{\partial\alpha} = \frac{n}{C}\frac{\partial C}{\partial\alpha} + \sum_{j=1}^{n}\log x_j - \sum_{j=1}^{n}\log(1-x_j) + \sum_{j=1}^{n}\log(1-ux_j) - \sum_{j=1}^{n}\log(1-ux_j-vy_j)$$

برآورد حداکثر درستنمایی پارامترهای $\alpha, \beta, \gamma, \gamma', u, v$ که به صورت $(\hat{\alpha}\ ,\hat{\beta},\hat{\gamma},\widehat{\gamma'},\hat{u},\hat{v})$ در نظر می گیریم از حل معادلات فوق بدست می آیند. برای حل معادلات فوق از الگوریتم nlm نرم افزار R استفاده شده است. هر چند فانگشن nlm نرم افزار می نیمم تابع را محاسبه می کند اما اگر منفی عبارت را به عنوان تابع ورودی به فانگش بدهیم ماکسیمم را برای ما محاسبه می نماید. با ورودیهای مناسب الگوریتم مطمئن هستیم که این روش همواره همگرا به جواب یگانه و منحصر بفرد می باشد.ماتریس اطلاع مشاهده شده برای $\alpha, \beta, \gamma, \gamma', u, v$ به صورت $(I_{i,j})-$ بوده که به عنوان نمونه بعضی از درایه های آن به صورت زیر می باشد:

$$I_{1,1} = \frac{n}{C}\frac{\partial^2 C}{\partial\alpha^2} - \frac{n}{C^2}(\frac{\partial C}{\partial\alpha})^2\ ,\ I_{1,2} = \frac{n}{C}\frac{\partial^2 C}{\partial\alpha\partial\beta} - \frac{n}{C^2}\left(\frac{\partial C}{\partial\alpha}\right)\left(\frac{\partial C}{\partial\beta}\right)\ ,\ I_{1,3} = \frac{n}{C}\frac{\partial^2 C}{\partial\alpha\partial\gamma} - \frac{n}{C^2}\left(\frac{\partial C}{\partial\alpha}\right)\left(\frac{\partial C}{\partial\gamma}\right)$$

حال این سوال مطرح می شود که برای n های بزرگ چگونه این تخمینها را نگهداری کنیم؟ ما این سوال را با تحقیق و انجام شبیه سازی به در بخش بعد خواهیم دید.

۷. داده های کاربردی

در این بخش توضیح داده می شود که چگونه توزیع بتای دو متغیره جدید تعریف شده با معادله (۱) با تمام توزیع های بتای دو متغیره شناخته شده تا کنون، مقایسه می گردد. برای این منظور از داده های جدول ۱ مربوط به جنگلی از گرده دانه گردو استفاده می کنیم. این داده ها در جدول ۱ آورده شده است. اطلاعات جدول ۱ مربوط به یک نمونه به حجم ۴۲ از چهار نوع دانه به نامهای *Pinus,Abies,Quercus,Alnus* می‌با شد. ضمنا جمع داده های مربوط به هر سطر ۱۰۰ می با شد. توزیع های با تابع چگالی احتمال روی ناحیه $(o,1)\times(0,1)$ که می توانند برازش گردند عبارتند از :

- توزیع بتای دو متغیره لیبی و نویک *[۵]* با تابع چگالی احتمال زیر، با نام اختصاری *'LN'*:

$$f(x,y) = \frac{C_1\ x^{\alpha-1}y^{\beta-1}(1-x)^{\gamma-\alpha-1}(1-y)^{\gamma-\beta-1}}{(1-xy)^{\gamma}}$$

- توزیع بتای دو متغیره گاپتا و ونگ [۶] با تابع چگالی احتمال زیر، با نام اختصاری $'GW'$:

$$f(x,y)=\frac{x^{a-1}y^{c-1}(1-x)^{b-1}(1-y)^{d-1}}{B(a,b)B(c,d)}\{1+\lambda[2I_x(a,b)-1][2I_x\,(c,d)-1]\}$$

$$I_x(a,b)=\int_0^x t^a(1-t)^{b-1}\;/B(a,b)$$

- توزیع بتای دو متغیره ناداراژه [۱۲] باتابع چگالی احتمال زیر، با نام اختصاری 'N2007a':

$$f(x,y)=\frac{C_2x^{\alpha-1}y^{\beta-1}(1-x)^{\gamma-\alpha-1}(1-y)^{\gamma-\beta-1}}{(1-xy\delta)^{\gamma}}$$

- توزیع بتای دو متغیره ناداراژه [۱۳] باتابع چگالی احتمال زیر، با نام اختصاری $'N2007b'$:

- $$f(x,y)=\frac{C_3x^{\beta-1}y^{\beta\prime-1}(1-x)^{\gamma-\beta-1}(1-y)^{\gamma\prime-\beta\prime-1}}{(1-ux-vy)^{\alpha}}$$

- توزیع بتای دو متغیره سارابیا و کاستیلو [۱۴] با تابع چگالی احتمال زیر، با نام اختصاری $'SC'$:

$$f(x,y)=\frac{C_4x^{a-1}y^{b-1}[1-x-y+(1-de)xy]^{c-1}}{[1-(1-d)x]^{a+c}\,[1-(1-e)y]^{b+c}}$$

جهت تشخیص تفاوت توزیع های فوق از معیارها به شرح ذیل بهره گرفته شده است:

۱. معیار اطلاع آکائیک با ضابطه تعریف شده به صورت : $AIC=2k-2\log L\,(\hat{\theta}\,)$

۲. معیار اطلاع بیزی با ضابطه تعریف شده به صورت : $BIC=k\log n\;-2\log\;L(\hat{\theta}\,)$

۳. معیار سازگاری اطلاع آکائیک با ضابطه تعریف شده به صورت: $CAIC=-\,2\log\;L(\hat{\theta}\,)+\;k\;(\log n\;+\;1)$

۴. معیار تصحیح اطلاع آکائیک با ضابطه تعریف شده به صورت: $AICc=AIC+\frac{2\,k\,(k+1)}{n-k-1}$

۵. معیار هنان – کویین با ضابطه تعریف شده به صورت : $HQC=-\,2\log\;L(\hat{\theta}\,)+2\,k\;\log\log n$

لازم به ذکر ا ست که هر کدام از معیارهای فوق که کمتر با شد بهترین برازش را دارد. برای بحث بی شتر و مف صل تر در این خ صوص به بارنهام و اندرسون [۲۴] مراجعه شود.

جدول ۱

Pinus	Abies	Quercus	Alnus	Pinus	Abies	Quercus	Alnus
۷۵	۲	۱۴	۹	۸۴	۱	۱۲	۳
۸۱	۲	۱۳	۴	۸۱	۱	۱۵	۳
۸۹	۳	۱	۷	۸۱	۱	۱۶	۲
۸۴	۵	۷	۴	۷۶	۲	۱۸	۴
۸۱	۳	۱۰	۶	۸۷	۳	۷	۳
۸۶	۱	۸	۵	۹۱	۱	۵	۳
۸۶	۲	۱۱	۱	۹۳	۴	۲	۱
۸۲	۲	۱۰	۶	۸۹	۱	۶	۴
۷۲	۱	۱۶	۱۱	۸۷	۳	۸	۲
۹۳	۴	۲	۱	۹۴	۱	۳	۲
۸۷	۱	۱۱	۱	۸۱	۲	۹	۸
۹۵	۱	۳	۱	۶۹	۷	۱۸	۶
۸۵	۱	۱۲	۲	۸۶	۱	۸	۵
۹۱	۱	۴	۴	۷۴	۵	۱۶	۵
۷۹	۱	۱۹	۱	۸۲	۲	۱۱	۵
۹۵	۲	۱	۲	۸۷	۳	۹	۱
۹۰	۳	۵	۲	۶۸	۳	۲۶	۳
۹۰	۲	۷	۱	۷۷	۳	۱۱	۹
۸۸	۱	۹	۲	۸۶	۲	۷	۵
۸۶	۱	۱۰	۳	۷۹	۱	۱۱	۹
۷۵	۲	۱۴	۹	۷۹	۱	۱۷	۳

جدول ۲		
نوع گرده	داربین و واتسون	گادفری و بروش
Pinus	۰,۳۶۱	۰,۸۴۰
Abies	۰,۶۵۱	۰,۵۷۵
Quercus	۰,۲۳۶	۰,۵۷۰
Alnus	۰,۵۵۵	۰,۷۴۵

جدول ۳			
نوع گرده	هریسون و مک کاب	گلد فلد و کوانت	بروش و پاگان
Pinus	۰,۴۰۷	۰,۲۱۸	۰,۸۹۹
Abies	۰,۰۵۲	۰,۰۵۸	۰,۳۹۱
Quercus	۰,۲۳	۰,۲۰۹	۰,۵۸۴
Alnus	۰,۳۰۵	۰,۱۵۱	۰,۸۲۵

جدول ۴ (مقادیر p – value آزمون)						
نوع گرده	رتبه بارتلس	کوکس و استوارت	علامت	رتبه	نقطه بازگشت	باکس
Pinus	۰,۲۴۰	۰,۰۵۱	۰,۵۳۷	۰,۱۱۲	۰,۵۰۳	۰,۳۵۸
Abies	۰,۸۰۷	۰,۸۱۵	۰,۳۸۰	۰,۱۱۴	۰,۰۵۱	۰,۶۰۱
Quercus	۰,۲۷۶	۰,۱۱۰	۰,۸۳۰	۰,۰۹۱	۰,۱۹۲	۰,۲۰۷
Alnus	۰,۶۳۸	۰,۲۴۸	۱	۰,۸۲۶	۰,۲۳۳	۰,۸۶۹

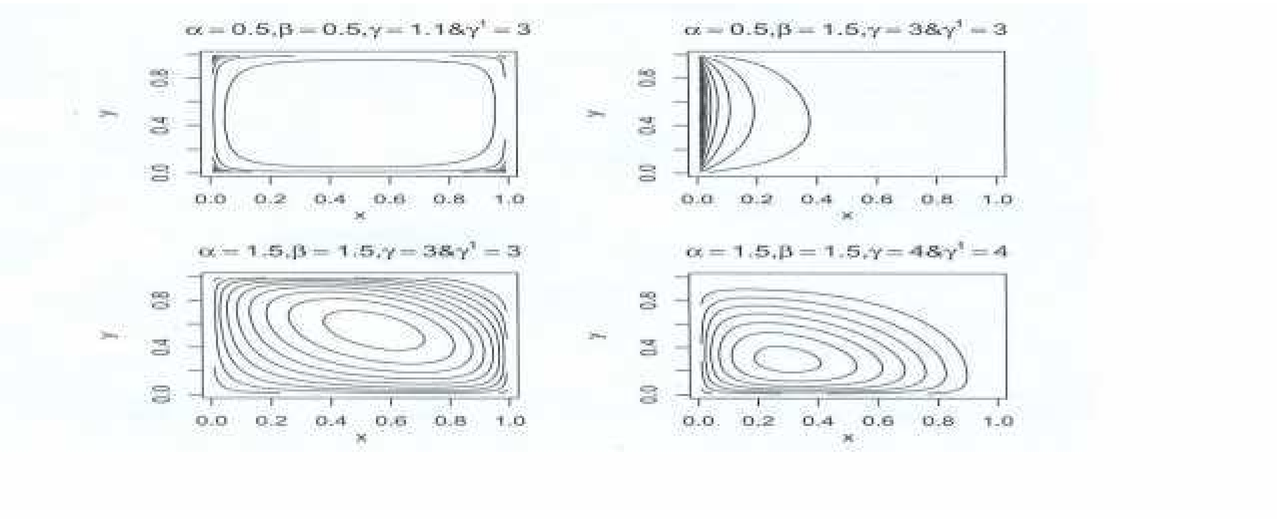
جدول ۵ (اندازه گیری معیارها برای (Pinus , Abies))						
نام توزیع	-log L	AIC	BIC	CAIC	AICc	HQC
LN	-167.458	-328.916	-323.180	-320.180	-328.394	-326.731
N2007a	-168.233	-328.466	-321.515	-317.515	-327.385	-325.918
New	-182.983	-353.967	-342.494	-336.494	-352.013	-349.598
GW	-180.047	-350.094	-340.534	-335.534	-348.73	-346.453
SC	-125.774	-241.547	-231.987	-226.987	-240.184	-237.907
N2007b	3.331×10^{-14}	14	27.384	34.384	16.667	19.097

جدول ۶ (اندازه گیری معیارها برای (Pinus , Quercus))						
نام توزیع	-log L	AIC	BIC	CAIC	AICc	HQC
LN	-166.065	-326.129	-319.299	-316.299	-325.776	-323.410
N2007a	-168.912	-329.824	-322.873	-318.873	-328.743	-327.276
New	-175.057	-349.670	-336.010	-330.010	-348.377	-344.232
GW	-155.356	-300.712	-289.329	-284.329	-299.803	-296.180
SC	-130.149	-250.297	-238.914	-233.914	-249.388	-245.766
N2007b	4.796×10^{-14}	14	29.937	36.937	15.75	20.344

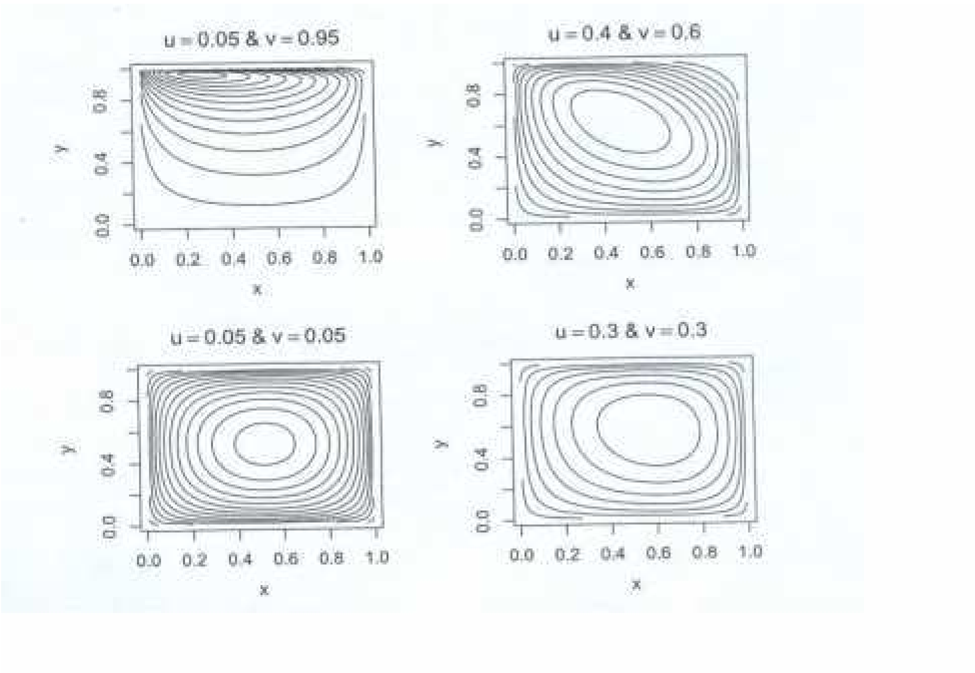
جدول ۷ (اندازه گیری معیارها برای (Pinus , Alnus))						
نام توزیع	-log L	AIC	BIC	CAIC	AICc	HQC
LN	-195.130	-384.259	-377.830	-374.830	-383.853	-381.731
N2007a	-197.233	-386.466	-379.515	-375.515	-385.385	-383.918
New	-209.482	-411.593	-398.734	-392.734	-410.093	-406.536
GW	-170.477	-330.955	-320.239	-315.239	-329.902	-326.740
SC	-137.733	-265.466	-254.751	-249.751	-264.414	-261.252
N2007b	4.197×10^{-14}	14	29.002	36.002	16.036	19.900

داده ها	برآورد تجربی	برآورد برازش شده
	جدول ۸ (ضریب همبستگی برآورد برازش شده با مقیاس اصلی)	
(Pinus , Abies)	-0.253	-0.260
(Pinus , Quercus)	-0.898	-0.872
(Pinus , Alnus)	-0.561	-0.525

داده ها	برآورد تجربی	برآورد برازش شده
	جدول ۹ (ضریب همبستگی برآورد برازش شده با مقیاس تبدیل شده)	
(Pinus , Abies)	0.185	0.188
(Pinus , Quercus)	0.252	0.244
(Pinus , Alnus)	0.296	0.299



شکل ۱ : تابع چگالی احتمال (۱) برای $u = 0.5$, $v = 0.5$



شکل۲ : توابع چگالی احتمال (۱) برای $\alpha = 1.5$, $\beta = 1.5$, $\gamma = 3$, $\gamma' = 3$

۸. نتیجه گیری

ملاحظه گردید که توزیع بتای دو متغیره جدید ارائه شده با معادله (۱) با توجه به معیارهای شش گانه عنوان شده، در مقایسه با سایر توزیع های بتای دو متغیره دارای بهترین برازش ا ست. مقادیر مربوط به معیارهای *AIC, BIC, CAIC, AICc, HQC* محا سبه و در جدول ۵ تا ۷ آورده شـده اسـت. می بینیم که توزیع بتای دو متغیره جدید دارای کمترین مقدار در هریک از معیارهای عنوان شـده برای هر مجموعه داده بوده، و این درحالی است که توزیع دو متغیره ناداراژه دارای بیشترین مقدار می باشد. توزیعی که دارای کمترین معیار باشد مناسب تر است و با توجه به مقادیر معیارهای فوق در می‌یابیم که توزیع دو متغیره جدید از سایر رقبای خود بهتر و تواناتر می باشد و این یک کشف قابل توجه می باشد.

مراجع

[1] Hutchinson TP, Lai CD. The engineering statistician guide to continuous bivariate distributions. Adelaik:Rumsby Scientific Publishing;1991.

[2] Arnol BC, Castillo E,Sarabia JM. Conditional specification of statistical models. New York: Springer; 1999.

[3] Kotz S,Balakrishnan N, Johonson NL. Continuous multivariate distributions, volume 1: models and Applications. 2nd ed. New York: Wiley; 2000.

[4] Connor RJ, Mosimann JE. Concepts of independence for proportions with a generalization of the Dirichlet distribution. J Amer Statist Assoc. 1969;64;194-206.

[5] Libby DL, Novick MR. Multivariate generalized beta-distributions with application to utility assessment.

[6] Gupta AK, Wong CF. On three and five parameter bivariate beta distributions. Metrika. 1985;32:85

[7] Jones MC. Multivariate t and the beta distributions associated with the multivariate F distribution. Metrika 20002;54:215-231.

[8] Olkin I, Liu R. A bivariate beta distribution. Stat Probab Lett. 2003;62:407-412.

[9] Nadarajah S, Kotz S. Some bivariate beta distributions. Statistics. 2005;39:457-466

[10] Nadarajah S, Kotz S. The bivariate F1-beta distribution. Math Rep Acad Sci. R Soc Can. 2005;27:58

[11] Nadarajah S, S. The bivariate F3-beta distribution. Commun. Kor Math Soc. 2006;21:363-374.

[12] Nadarajah S, S. A new bivariate beta distribution wuth application to drought data. Metron 2007LXV

[13] Nadarajah S, S. The bivariate F2 beta distribution. Amer J Math Manage Sci. 2007;27;351-368

[14] Sarabia JM, Castillo E. bivariate distribution based on the generalized three-parameter beta distribution. In:Balakrishnan N, Sarabia JM, Castillo E, editor. USA: Birkhauser Boston; 2006. P.85-110

[15] Nagar DK, Orozco-Castaneda JM. Generalized bivariate beta type 1 distribution. 2009;12:421-429

[16] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974:19:716-723

[17] Hannan EJ, Quinn BG. The determination of the order of an autoregression. J R Statist Soc, B.

[18] Arnold BC, Ng HKT. Flexible bivariate beta distributions. J Multivariate Anal. 2011;102:1194-1202

[19] Bran-Cardona PA, Orozco-Castaneda JM.Nagar DK.

[20] Gupta AK, Orozco-Castaneda JM, Nagar DK. Non – central bivariate beta distribution. Statist Pap. 2011;52:139-152

[21] Durbin J, Watson GS. Testing for serial correlation in least squares regression.

[22] Jacobs R, Bekker A, Human SW. On the bivariate Kummer-beta type IV distribution. Commun Stat-Theory Methods. 2012;41:3339-3354.

[23] Olkin I, Trikalinos TA. Constructions for a bivariate beta distribution. Stat Probab Lett 2015;96:54-60

[24] Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33:261-304.

A new bivariate beta distribution

Najmah Sadat Moosavi ahmadabadi

Abstract. A new bivariate beta distribution capable of providing better than all its competitiors is introduced. Varius representations are derived for its prouct moments, marginal densities, marginal moments, conditional densities and conditional moments. The method of maximum likelihood is used to derive the associated estimation procedure. Applications to six bivariate data sets are illustrated.

Key words: Estimation, pollen, count, product moments