

یک الگوی چندمتغیری بیزی برای شبکه‌های ژنی

عبداللله صفری^۱، علی شریفی^۲، حمید پزشک^۱، پیمان نیکچی^۳، سیدامیر مرعشی^۴، چنگیز اصلاح چی^۵

^۱ دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران

^۲ موسسه بیوشیمی و بیوفیزیک، دانشگاه تهران

^۳ پژوهشکده علوم زیستی، پژوهشگاه دانش‌های بنیادی

^۴ گروه بیوتکنولوژی، پردیس علوم، دانشگاه تهران

^۵ دانشکده ریاضی، دانشگاه شهید بهشتی

تاریخ دریافت: ۱۳۹۱/۲/۱۳ تاریخ آخرین بازننگری: ۱۳۹۱/۷/۲۸

چکیده: در روش‌های استنباط موجود برای شبکه‌های ژنی، معمولاً فقط از اطلاعات نمونه استفاده می‌شود و کمتر از اطلاعات گذشته یا موازی استفاده شده است. در این مقاله با استفاده از اطلاعات موجود به استنباط بیزی شبکه‌های ژنی پرداخته می‌شود. با به‌کارگیری توزیع پیشین مناسب و توجه به عدم استقلال پارامترهای موجود در شبکه‌های ژنی تعیین‌الگویی مورد نظر است که بتوان نتایج بهتری نسبت به اطلاعات گذشته و اطلاعات موجود در نمونه‌ها به دست آورد. به‌علاوه ابرپارامترها به دو روش برآورد می‌شوند و نتایج آن در یک مطالعه شبیه‌سازی مورد بررسی قرار گرفته، معایب و مزایای آن‌ها بیان خواهند شد.

واژه‌های کلیدی: شبکه‌های ژنی، استنباط بیزی، مدل‌های چندمتغیری، نمونه‌گیر گیبز.

آدرس الکترونیک مسئول مقاله: حمید پزشک، pezeskh@khayam.ut.ac.ir

کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲E۱۵

اگر در حوزه ژنوم تنها به حیات سلولی نگاه شود، با حجم فراوانی از ابزارها، روش‌ها، مدل‌ها و مقاله‌ها روبرو می‌شویم. یکی از پرطرفدارترین رویکردها، مدل مبتنی بر شبکه است. در این نوع مدل‌ها، اجزای مختلف یک سلول متناظر با راس‌های یک گراف یا نقاط یک شبکه هستند و انواع برهم‌کنش‌های بین آن‌ها، یال‌های گراف یا پیوندهای بین نقاط شبکه را به وجود می‌آورند (اشلیت و برزما، ۲۰۰۷). بررسی شبکه‌های ژنی از این لحاظ قابل اهمیت است که ابزاری مهم برای یافتن میزان تأثیرگذاری پروتئین‌ها، یافتن پروتئین‌های مرتبط کار ساده‌ای نیست. پژوهش‌های فراوانی برای حل این مشکل انجام شده ولی تقریباً در تمامی آن‌ها، محققان نسبت به نتایج تحقیقات قبلی بی‌تفاوت بوده‌اند.

روش‌های گوناگونی برای به‌کارگیری مدل‌های پویای شبکه‌های ژنی به‌کار گرفته شده‌اند. از جمله می‌توان به مدل‌های بولی (کافمن، ۱۹۶۹)، شبکه‌های پتری (پایینی و همکاران، ۲۰۰۳)، مدل‌های ترکیبی^۱ معادلات تفاضلی و دیفرانسیل (داسو و همکاران، ۲۰۰۰)، مدل‌های خطی با حالات محدود^۲ (FSLM)، شبکه‌های تصادفی^۳ (مک ادماس و ارکین، ۱۹۹۷) و مهندسی معکوس شبکه‌های ژنی (فریدمن، ۲۰۰۴) اشاره کرد.

در این مقاله با استفاده از مدل‌های سلسله مراتبی بیزی، به بازسازی شبکه‌های ژنی پرداخته می‌شود. هدف از این بررسی‌ها، یافتن جفت پروتئین‌هایی است که نسبت به یکدیگر برهم‌کنش دارند. در تحلیل نتایج آزمایش‌هایی که تاکنون انجام شده‌اند، از اطلاعات پژوهش‌های قبلی، که در دسترس هستند، هیچ استفاده‌ای نشده است. در حقیقت تحلیل آزمایش پروفایل بیان ژنی^۴ نسبت به نتایج به‌دست آمده توسط دیگران در آزمایش‌های قبلی، با بی‌تفاوتی انجام شده است. به‌وضوح اگر بتوان روش مناسبی برای به‌کارگیری اطلاعات قبلی در آزمایش‌های بعدی پیدا نمود،

۱ Hybrid

۲ Finite State Linear Models

۳ Stochastic Networks

۴ Gene expression profile

انتظار می‌رود نتایجی بهتر و با دقت بیشتری فراهم شود. برای ترکیب نتایج از پیش تعیین شده توسط دیگران با نتایج آزمایش‌های انجام شده در زمان حال، می‌توان از استنباط بیزی استفاده کرد. برای این کار در بخش دوم ابتدا یک توزیع پیشین مناسب در نظر گرفته می‌شود. سپس به دو روش متفاوت پارامترهای توزیع پیشین برآورد می‌شوند. در بخش سوم با روش‌های شبیه‌سازی مونت کارلوی زنجیر مارکوفی و نمونه‌گیر گیبز نتایج حاصل از روش‌های ارائه شده مورد تحلیل حساسیت قرار می‌گیرند. در بخش چهارم به بحث و نتیجه‌گیری پرداخته می‌شود.

۲ استنباط بیزی ماتریس کواریانس

اعداد موجود در شبکه‌های بیان ژنی، بیان‌کننده میزان ارتباط ژن‌ها هستند. بنابراین ساختار شبکه بیان ژنی به صورت ساختار ماتریس کواریانس است که می‌توان با کمک استنباط بیزی برای یک ماتریس کواریانس به تحلیل آن‌ها پرداخت. فرض کنید ماتریس کواریانس جامعه‌ای از توزیع نرمال به صورت $Y \sim N(0, \Sigma)$ باشد. هرچند استفاده از توزیع نرمال برای مدل‌بندی شبکه‌های ژنی و یافتن توزیع‌های پیشین مناسب تقریباً امری ضروری است، اما این توزیع از واقعیت توزیع شبکه‌های ژنی هم فاصله زیادی ندارد. برای انجام این استنباط، ابتدا باید یک توزیع پیشین برای ماتریس کواریانس Σ تعیین کرد. برای این کار روش‌های زیادی وجود دارد. یکی از مهم‌ترین آن‌ها تعیین یک توزیع پیشین مزدوج است (هاف، ۲۰۰۹). چون درایه‌های ماتریس کواریانس از هم مستقل نیستند، شبیه‌سازی از آن‌ها به طور مجزا نتایج قابل قبولی را به دست نمی‌دهد (پاپاسپیلوپولوس و رابرتس، ۲۰۰۸). در این مورد می‌توان از توزیع وارون ویشارت به عنوان توزیع پیشین مزدوج برای ماتریس کواریانس استفاده کرد (ژانگ، ۲۰۰۹). روش دیگری که در این مورد کاربرد زیادی دارد، استفاده از توزیع نرمال چندمتغیره است. درست است که این نوع توزیع پیشین مزدوج نیست، اما تحت شرایطی به‌طور مجانبی مزدوج است (لئونارد و هسو، ۱۹۹۲).

۱.۲ توزیع وارون ویشارت

در این زیربخش فرض می‌شود ماتریس کوواریانس Σ از توزیع پیشین وارون ویشارت $IW_p(G, \nu)$ ، پیروی کند (گلمن و همکاران، ۲۰۰۳). که در آن G ماتریس مقیاس همیشه مثبت، p بعد ماتریس کوواریانس و $\nu < 2p$ درجه آزادی است. در این صورت

$$p(\Sigma|G, \nu) = \frac{c_0 |G|^{\frac{\nu-p-1}{2}}}{|\Sigma|^{\frac{\nu}{2}}} \exp\left\{-\frac{1}{2} \text{tr}(\Sigma^{-1}G)\right\}$$

که در آن c_0 مقدار ثابت نرمال کننده تابع چگالی است. مد توزیع وارون ویشارت به صورت $Mode(\Sigma|G, \nu) = \frac{G}{\nu}$ است (هاگان، ۲۰۱۰). با توجه به چن (۱۹۷۹)، پارامترهای جدیدی برای توزیع وارون ویشارت به صورت

$$\Sigma|G, \nu \sim IW_p(\nu\Omega, \nu) \quad (1)$$

در نظر گرفته می‌شود، که در آن ν درجه آزادی و Ω نشان دهنده اطلاعات ساختاری تعیین شده درباره Σ است به طوری که $Mode(\Sigma|\Omega, \nu) = \frac{\nu\Omega}{\nu} = \Omega$ ویشارت برای ماتریس کوواریانس جامعه‌ای از توزیع نرمال چندمتغیره، پیشین مزدوج است (پرس، ۲۰۰۵). فرض کنید N مشاهده $Y = \{y_1, \dots, y_N\}$ از توزیع $N(0, \Sigma)$ در اختیار باشند به طوری که Σ از توزیع پیشین مزدوج وارون ویشارت $IW_p(\nu\Omega, \nu)$ پیروی کند. در این صورت توزیع پسین Σ نیز وارون ویشارت به صورت $IW_p(\nu^*\Omega^*, \nu^*)$ خواهد بود، که در آن

$$\nu^* = \nu + N, \quad \Omega^* = \left(\frac{N}{N + \nu}\right)S + \left(\frac{\nu}{N + \nu}\right)\Omega \quad (2)$$

توجه شود که ماتریس Ω^* از میانگین وزنی ابرپارامتر پیشین Ω و ماتریس کوواریانس نمونه‌ای S به دست می‌آید که مد توزیع وارون ویشارت پسین نیز می‌باشد.

حال باید ابرپارامترهای توزیع وارون ویشارت پیشین را نیز محاسبه کرد. همان‌طور که قبلاً اشاره شد، Ω^* از ترکیب Ω و S به دست می‌آید. نحوه ترکیب آن‌ها

توسط ν ، درجه آزادی توزیع پیشین، تعیین می‌شود. برای تابع چگالی (۱) لازم است شرط $\nu > 2p$ برقرار باشد ولی در عمل دلیلی برای برقرار بودن این شرط وجود ندارد. به عبارت دیگر، ν نشان دهنده میزان اطمینانی است که برای ابرپارامتر Ω منظور می‌شود و می‌توان آن را به صورت تجربی از مجموعه اعداد حقیقی نامنفی به دست آورد. وقتی ν برابر با صفر در نظر گرفته می‌شود، Ω^* دقیقاً ماتریس کوواریانس نمونه‌ای می‌شود. سوال اصلی، تعیین ابرپارامتر Ω است که نشان دهنده اطلاعات ساختاری مشخص کننده ماتریس کوواریانس است. در حقیقت مشکل اساسی پیدا کردن ساختار کوواریانس Ω از اطلاعات پیشین است. برآورد Ω با روش ترکیب اطلاعات پیشین و اطلاعات نمونه یا روش گشتاوری انجام می‌شود.

۲.۲ ترکیب اطلاعات پیشین و اطلاعات نمونه

ژانگ (۲۰۰۹)، از یک متغیر فاصله‌ای برای یافتن ساختار کوواریانس استفاده کرد، ولی برای غلبه بر برخی مشکلات مانند بدون مقیاس کردن متغیر فاصله‌ای و نزدیک کردن آن به خواص کوواریانس، می‌توان از اطلاعات نمونه‌ای هم کمک گرفت. گلمن و همکاران (۲۰۰۳) نشان دادند یک ماتریس کوواریانس را می‌توان به ماتریس‌های همبستگی و انحراف معیار به صورت $\Omega = VCV$ تجزیه کرد، که در آن V یک ماتریس قطری $p \times p$ از انحراف معیارها است. چون Ω در توزیع وارون و یشارت باید همیشه مثبت باشد، C یک ماتریس همیشه مثبت است. شهود این نوع تجزیه این است که درایه‌های قطری ماتریس V حاوی اطلاعاتی درباره انحراف معیارها بوده و C نمایش دهنده ساختار همبستگی بین متغیرها است. ماتریس قطری V می‌تواند توسط داده‌ها برآورد شود. با توجه به ساختار C ، برای برآورد درایه‌های آن از رابطه $C_{ij} = \exp\left(\frac{d_{ij}}{\alpha}\right)$ استفاده می‌شود (یاگ لام، ۱۹۸۷) که در آن d_{ij} اطلاعات پیشین درباره متغیرهای i ام و j ام است. در مورد شبکه‌های ژنی d_{ij} میزان بیان ژنی خواهد بود که از نتایج بررسی‌های قبلی قابل استخراج است و $\alpha > 0$ یک پارامتر است و برآورد آن از رابطه $\rho_{median} = \exp\left(-\frac{d_{median}}{\alpha}\right)$ محاسبه می‌شود، که در آن ρ_{median} میانه درایه‌های ماتریس همبستگی نمونه‌ای و d_{median}

میانه درایه‌های ماتریس اطلاعات پیشین هستند.

برای برآورد Ω با توجه به شرایط خاص آن و تعداد پارامترهای آن از روش گشتاوری استفاده می‌شود. اگر Σ از توزیع وارون ویشارت (۱) پیروی کند، در این صورت

$$E(\Sigma|\Omega, \nu) = \frac{\nu\Omega}{\nu - p - 1} \quad (۳)$$

بنابراین به شرط مشخص بودن ν ، به سادگی می‌توان Ω را از رابطه $\hat{\Omega} = \frac{(\nu-p-1)D}{\nu}$ محاسبه کرد که در آن D ماتریس اطلاعات پیشین است. برآورد ν مشابه قبل به صورت تجربی تعیین می‌شود.

۳ مطالعه شبیه‌سازی

اگر مدل بندای بیزی معرفی شده مناسب باشد، نتایجی که از توزیع پسین برای مهندسی معکوس ساخت شبکه ژنی به دست می‌آیند باید بهتر از نتایجی باشند که به تنهایی از داده‌ها یا به تنهایی از اطلاعات پیشین به دست می‌آیند. در این بخش با استفاده از روش‌های شبیه‌سازی مدل بیزی بیان شده مورد بررسی قرار می‌گیرد. علاوه بر این دو روش برآورد ابرپارامترهای توزیع پیشین ارائه شده نیز با هم مقایسه می‌شوند.

برای اینکه بتوان نتایج به دست آمده از هر روش را با هم مقایسه نمود، ابتدا ماتریس کوواریانس واقعی به طور تصادفی شبیه‌سازی می‌شود. سپس به طور تصادفی مقداری نوفه به آن اضافه می‌شود. حال دو ماتریس جدید به دست آورده می‌شود یکی به عنوان ماتریس کوواریانس نمونه‌ای و دیگری به عنوان ماتریس اطلاعات پیشین به دست آورده می‌شود. از آنجا که معمولاً اطلاعات پیشین در بررسی‌های برهم‌کنش پروتئینی به صورت ماتریسی با مولفه‌های -1 ، 0 و 1 است،

درایه‌های ماتریس اطلاعات پیشین به صورت

$$d'_{ij} = \begin{cases} -1 & d_{ij} < d_1 \\ 0 & d_1 \leq d_{ij} \leq d_2 \\ 1 & d_{ij} \geq d_2 \end{cases} \quad (4)$$

تبدیل می‌شوند، که در آن d_1 چارک اول و d_2 چارک سوم درایه‌های ماتریس اطلاعات پیشین هستند.

الگوریتم شبیه‌سازی:

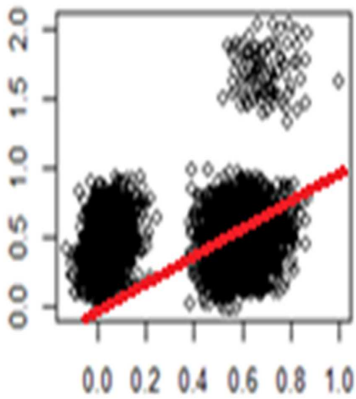
- (۱) تولید یک ماتریس کوواریانس تصادفی به عنوان ماتریس کوواریانس واقعی (R)
- (۲) تولید ماتریس کوواریانس نمونه‌ای با اضافه کردن یک نوفه سفید (از توزیع نرمال صفر و واریانس σ^2) به همه درایه‌های ماتریس کوواریانس واقعی به صورت تصادفی و کاملاً مستقل به عنوان خطای مشاهدات (S)
- (۳) تبدیل ماتریس کوواریانس واقعی به ماتریسی با درایه‌های منفی یک، صفر و یک با استفاده از رابطه (۴) به عنوان ماتریس شبکه واقعی
- (۴) تبدیل درایه‌های صفر به منفی یک، یک یا به عکس با احتمال P از ماتریس شبکه واقعی به عنوان خطای اطلاعات پیشین (D) برای به‌دست آوردن ماتریس شبکه به عنوان اطلاعات پیشین
- (۵) برآورد ماتریس مقیاس توزیع و ارون ویشارت پیشین از دو روش ارائه شده در بخش قبل (Ω_1 و Ω_2)
- (۶) در نظر گرفتن مقداری اولیه برای درجه آزادی توزیع و ارون ویشارت پیشین (ν)
- (۷) به‌دست آوردن پارامترهای (ν^* و Ω^*) توزیع و ارون ویشارت پسین از رابطه (۲)
- (۸) به‌دست آوردن برآوردی برای ماتریس کوواریانس واقعی از توزیع و ارون ویشارت پسین با استفاده از رابطه (۳) با جایگزینی ν^* و Ω^* به جای ν و Ω
- (۹) محاسبه میانگین توان دوم خطا برای برآوردهای ماتریس کوواریانس واقعی از دو روش ارائه شده
- (۱۰) تغییر درجه آزادی توزیع و ارون ویشارت پیشین و بازگشت به گام ۷

(۱۱) تکرار گام‌های ۷ تا ۱۰ برای رسیدن به برآوردی با کمترین میانگین توان دوم خطا برای یکی از دو روش ارائه شده

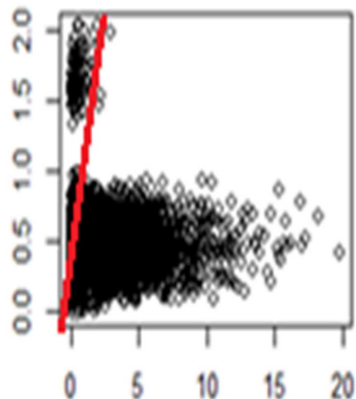
با انجام الگوریتم فوق توسط نرم افزار R و با ۱۰۰۰ تکرار شبیه‌سازی و محاسبه متوسط میانگین توان دوم خطاهای هر تکرار نتایج زیر حاصل شده‌اند.

جدول ۱ میانگین توان دوم خطا برای برآوردهای مختلف با $p = 0.4$ و $\sigma^2 = 0.5$ است. برای محاسبه این اعداد، توان دوم تفاضل درایه‌های هر ماتریس برآورد را از ماتریس واقعی، نظیر به نظیر، محاسبه کرده و میانگین آن‌ها به دست آورده می‌شوند. اعداد نشان داده شده در جدول، متوسط مقادیر میانگین توان دوم خطا در ۱۰۰۰ بار تکرار می‌باشند. همان‌طور که در جدول ۱ ملاحظه می‌شود با افزایش ν ، میانگین توان دوم خطا برای برآورد گشتاوری کمتر می‌شود ولی در مورد برآورد ترکیبی چنین نیست. به نظر می‌رسد که کمتر بودن ν برآورد بهتری از روش ترکیبی به دست می‌دهد. به طور خاص به ازای $\nu = 120$ علاوه بر برآورد ترکیبی، روش گشتاوری از اطلاعات پیشین و داده‌ها هر یک به تنهایی نیز بهتر عمل کرده است. یعنی مدل‌بندی بیزی می‌تواند با انتخاب مناسب پارامترهای پیشین در این زمینه باعث بهبود نتایج شود. قابل ذکر است که میانگین توان دوم خطا برای ν های بیشتر از ۱۲۰ نیز محاسبه شد که روندی صعودی داشت. این نشان‌دهنده بهینه بودن $\nu = 120$ می‌باشد. برای اینکه نحوه فاصله درایه‌های هر یک از برآوردها نسبت به ماتریس کوواریانس واقعی دیده شود، می‌توان درایه‌های هر یک از ماتریس‌های برآورد را در برابر درایه‌های ماتریس کوواریانس واقعی رسم کرد (شکل ۱). به وضوح برآوردی بهتر است که نقاط در نمودار پراکنش^۵ مربوطه به نیمساز ناحیه اول نزدیک‌تر باشند. همان‌طور که در شکل مشخص است، برآوردگر گشتاوری عملکرد بهتری نسبت به بقیه برآوردها داشته است و البته ماتریس کوواریانس نمونه‌ای هم به دلیل وجود خطای بسیار کوچک فاصله زیادی از ماتریس واقعی ندارد. شایان ذکر است، فواصلی که بین نقاط در هر شکل مشاهده می‌شود به دلیل تبدیل ماتریس‌های کوواریانس شبیه‌سازی شده به ماتریس همیشه مثبت است که برای تولید نمونه از توزیع نرمال چندمتغیره امری ضروری است. در

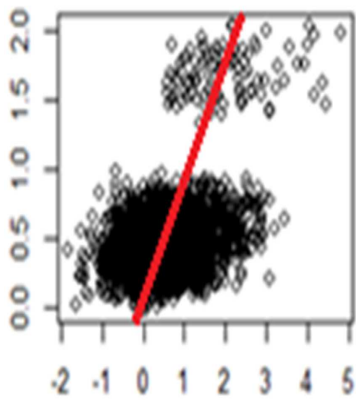
^۵ Scatter plot



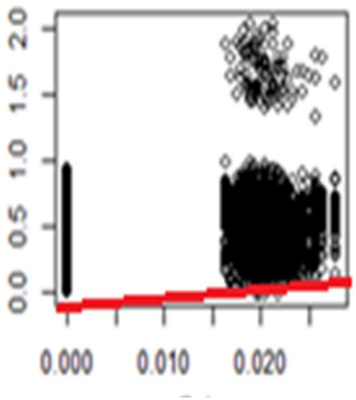
(ب)



(الف)



(د)



(ج)

شکل ۱: درایه‌های ماتریس‌های برآورد در برابر درایه‌های ماتریس واقعی ($\sigma^2 = 0/5$) نمودار (الف) نسبت به Ω_1 ، نمودار (ب) نسبت به Ω_2 ، نمودار (ج) نسبت به پیشینی و نمودار (د) نسبت به داده‌ها رسم شده‌اند.

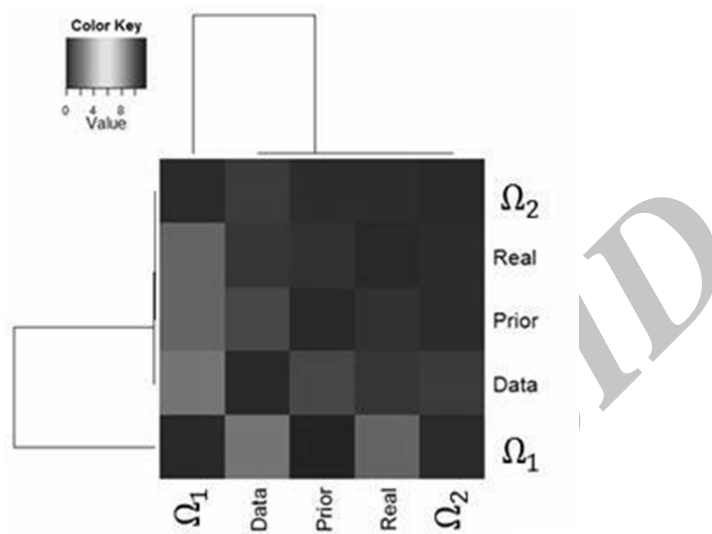
جدول ۱: میانگین توان دوم خطا برای برآوردها به‌ازای مقادیر مختلف درجه آزادی توزیع وارون ویشارت پیشین

پسین		داده	پیشین	ν	σ^2
Ω_2	Ω_1				
۰/۱۴۰۵	۸/۸۳۵۶	۰/۴۷۰۲	۰/۲۸۵۵	۱۲۰	
۰/۳۵۷۷	۸/۴۴۷۰	۰/۴۶۸۸	۰/۲۸۵۱	۹۰	
۰/۷۹۵۷	۷/۶۷۵۹	۰/۴۶۸۴	۰/۲۸۶۱	۶۰	
۱/۱۷۷۵	۶/۲۱۷۸	۰/۴۷۴۷	۰/۲۸۳۸	۳۰	۰/۵
۰/۸۳۱۲	۲/۹۹۲۱	۰/۴۶۸۴	۰/۲۸۴۴	۱۰	
۰/۴۰۲۰	۰/۵۴۷۴	۰/۴۵۸۲	۰/۲۸۶۵	۱	
۰/۴۶۴۳	۰/۴۶۴۳	۰/۴۶۴۳	۰/۲۸۵۱	۰	
۰/۸۷۱۷	۳۴۳۹۴۴/۶	۱۴۸/۳۹۶۶	۰/۲۸۵۵	۱۲۰	
۱/۵۰۴۷	۳۱۸۷۴۲/۵	۱۴۷/۳۲۲۵	۰/۲۸۴۹	۹۰	
۳/۱۵۰۶	۲۹۱۳۳۴/۱	۱۴۷/۰۱۱۰	۰/۲۸۵۶	۶۰	
۹/۱۹۷۷	۲۳۰۹۰۷/۲	۱۴۹/۰۶۳۰	۰/۲۸۳۹	۳۰	۵
۱۲۱/۳۲۱۲	۱۰۴۲۲۰/۸	۱۵۰/۶۲۸۲	۰/۲۸۶۱	۱۰	
۱۲۱/۶۲۳۵	۳۶۵۱/۱۵۵۵	۱۴۷/۹۷۵۰	۰/۲۸۵۸	۱	
۱۵۰/۲۸۸۰	۱۵۰/۲۸۸۰	۱۵۰/۲۸۸۰	۰/۲۸۴۶	۰	

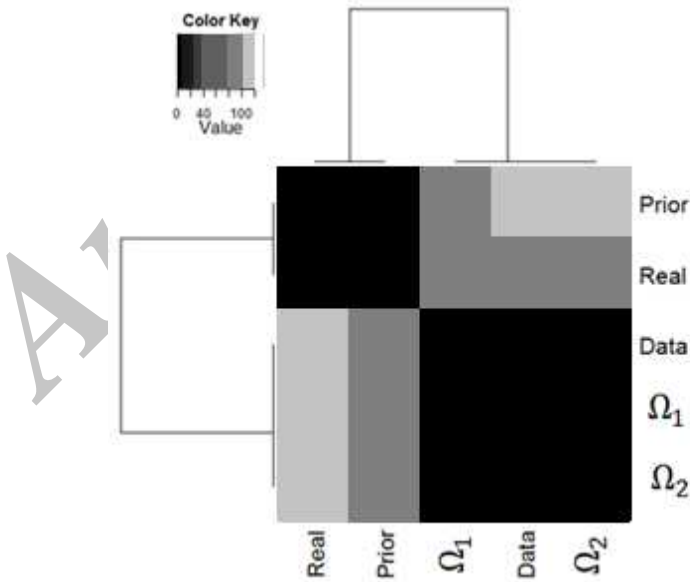
جدول ۱ و شکل ۱ تفاوت درایه‌های هر ماتریس برآورد از درایه‌های ماتریس واقعی، برای بررسی نزدیکی برآوردها به مقدار واقعی مورد توجه قرار داده شده است. حال آن‌که می‌توان این نزدیکی را با در نظر گرفتن روند تغییرات درایه‌ها در هر ماتریس نسبت به ماتریس واقعی، نیز بررسی نمود. برای انجام این کار از نقشه گرمایی^۶ که بر مبنای ماتریس همبستگی است، می‌توان استفاده کرد (شکل ۲). همان‌طور که ملاحظه می‌شود، روش برآورد گشتاوری نسبت به بقیه روش‌ها به ماتریس کوواریانس واقعی نزدیک‌تر است که نتایج جدول ۱ و شکل ۱ را نیز تأیید می‌کند. حال برای بررسی تأثیر مقدار خطای اطلاعات پیشین و خطای مشاهدات، تأثیر هر یک به طور جداگانه روی نتایج بررسی می‌شوند.

جدول ۱ میانگین توان دوم خطا را برای روش‌های مختلف و مقادیر ۰/۵ و ۵ برای σ^2 نمایش می‌دهد. همان‌طور که انتظار می‌رفت، افزایش خطای داده‌های

^۶ Heatmap



شکل ۲: نقشه گرمایی اختلاف برآوردهای مختلف از ماتریس کوواریانس واقعی ($\sigma^2 = 0.5$)



شکل ۳: نقشه گرمایی اختلاف برآوردهای مختلف از ماتریس کوواریانس واقعی ($\sigma^2 = 5$)

مشاهده شده، باعث افزایش میانگین مربع خطای برآورد ماتریس کوواریانس واقعی بر اساس داده‌ها شده است. در مورد برآوردهای دو روش بیزی، خطای هر دو مورد زیاد شده است. باز هم مانند حالت قبل، با افزایش میانگین توان دوم خطا برای روش گشتاوری کاهش و برای روش ترکیبی افزایش می‌یابد ولی تغییرات به اندازه قبل محسوس نیست.

هجوته ابه، کروطن امه، دوشم سره رابود $120 = \nu$ ت ماحری اربی یامرگ هشقذ رگا
ت ماعلاط ط سوت س نایراووکس برتامه درو اربی یکیدزند، ت فری م رماظتنا ۱ ل ولدج به
دوشی م هدیدی عقاوس نایراووکس یرتامه بن پیشیپ

۴ بحث و نتیجه گیری

در این مقاله به یکی از مسائل مهم حوزه‌های سلولولی، مولکولی محاسبات و بیوانفورماتیک یعنی تحلیل شبکه‌های برهم کنش ژنی پرداخته شد. دیدگاهی نظری برای مدل‌سازی بیزی مناسب این گونه شبکه‌ها ارائه شد. در حال حاضر تیم‌های تحقیقاتی مختلفی مشغول تهیه بانک‌های اطلاعاتی داده‌های واقعی ژنی هستند که انتظار می‌رود در آینده‌ای نه چندان دور، مدل ارائه شده روی داده‌های واقعی به کار گرفته شود. در این کار مبانی نظری و نحوه پیاده سازی مدل پیشنهادی مورد بحث قرار گرفت. ملاحظه شد که با افزایش خطای داده‌های مشاهده شده، داده‌هایی به دست می‌آیند که دقت کمتر و در نتیجه ارزش کمتری دارند و اطلاعات پیشین به مراتب دارای اطلاعات بیشتر توأم با دقت بیشتر هستند. بنابراین ترکیب این اطلاعات با اطلاعات داده‌های مشاهده باعث کم شدن دقت و در نتیجه کم شدن ارزش برآورد حاصل می‌شود. حال اگر داده‌ها و اطلاعات پیشین قابل اطمینانی در اختیار باشند، می‌توان از ترکیب آن‌ها به نتایج دقیق‌تری دست یافت. بنابراین برای کسب نتیجه دقیق‌تر بهتر است ابتدا به بررسی دقت داده‌های در اختیار و نیز دقت اطلاعات پیشین پرداخته شود. سپس با توجه به اینکه کدام یک از اطلاعات قابل اطمینان‌تر هستند، به ترکیب اطلاعات پرداخته می‌شود.

۵ تقدیر و تشکر

نویسندگان از داوران محترم مجله کمال تشکر را دارند. از معاونت پژوهشی دانشگاه تهران سپاسگزاری می شود. بخشی از هزینه‌های این پژوهش توسط گروه تحقیقاتی بیوانفورماتیک پژوهشگاه دانش‌های بنیادی تامین شده است ۱-۰-۱۳۸۹-CS

۶ مراجع

- Chen, C. (1979), Bayesian Inference for a Normal Dispersion Matrix and its Application to Stochastic Multiple Regression Analysis, *Journal of Royal Statistical Society, B*, **41**, 235-248.
- Dassow, G., Meir, E., Munro, E. M. and Odell, G. M. (2000), The Segment Polarity Network is a Ribust Developmental Module, *Nature*, **406**, 188-192.
- Friedman, N. (2004), Inferring Cellular Networks Using Probabilistic Graphical Models, *Science*, **303**, 799-805.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003), *Bayesian Data Analysis*, Chapman and Hall CRC.
- Hagan, A. (2010), *Kendalls Advanced Theory of Statistics*, V. 2B, Arnold.
- Hoff, D. (2009), *A First Course in Bayesian Statistical Methods*, Springer Press, New York.
- Kauffman, S. (1969), Homeostasis and Differentiation in Random Genetic Control Networks, *Nature*, **224**, 177-178.
- Leonard, T. and Hsu, S. J. (1992), Bayesian Inference for a Covariance Matrix, *The Annals of Statistics*, **20**, 669-1669.

McAdams, H. and Arkin, A. (1997), Stochastic Echanisms in Gene Expression, *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 814-819.

Papaspiliopoulos, O. and Roberts, G. O. (2008), Stability of the Gibbs sampler for Bayesian Hierarchial Models, *The Annals of Statistics*, **36**, 95-117.

Pinney, J. W., Westhead, D. R. and McConkey, G. A. (2003), Petri Net Representations in Systems Biology, *Biochemical Society Transactions*, **31** 1513-1515.

Press, S. J. (2005), *Applied Multivariate analysis*, Dover Publications Inc.

Schlitt, T. and Brazma, A. (2007), Current Approaches to Gene Regulatory Network Modelling, *BMC Bioinformatics*, **8**, S9-S9.

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions*, Springer-Verlag, New York.

Zhang, Y. (2009), Smart PCA, *Proceeding of the Twenty-First International Joint Conference on Artificial Intelligence, (IJCAI-09)*.