

مروری بر کاربردهای خانواده نمایی در استنباط آماری

مهدی شمس^۱

تاریخ دریافت: ۱۳۹۶/۸/۲۴

تاریخ پذیرش: ۱۳۹۷/۱۰/۲۱

چکیده:

در این مقاله پس از معرفی خانواده نمایی و تاریخچه‌ای از کارهای انجام شده توسط محققان در شاخه‌های مختلف آمار، به شرح کاربردهایی از این خانواده در استنباط آماری به‌ویژه در مسئله برآورد، آزمون‌های فرضیه آماری و مفاهیم نظریه اطلاع آماری پرداخته می‌شود.

واژه‌های کلیدی: خانواده نمایی منظم، فضای پارامتر طبیعی، برآورد، آزمون‌های فرضیه آماری، آنتروپی، اطلاع متقابل.

۱ مقدمه

مطالب ارزنده‌ای از این توزیع‌ها مطالعه کرد. به‌عنوان مثال از تحقیقات مهمی که در این زمینه انجام شده می‌توان به کاربرد در تقسیم‌پذیر نامتناهی بودن [۸، ۵۶، ۶۰]، توزیع‌های پایدار چندمتغیره [۴۴]، سیستم‌های دینامیکی تصادفی [۹۹]، میدان‌های تصادفی شرطی [۳]، وابستگی در خانواده‌های نمایی شرطی با استفاده از میدان‌های تصادفی مارکوفی [۶۱]، سازماندهی فرایندهای لوی [۵۷]، توصیف فرایندهای مارکوفی زمان پیوسته روی فضای حالت متناهی که توزیع آنها در هر لحظه متعلق به یک خانواده نمایی یک‌پارامتری است [۱۰۲، ۱۰۳، ۱۰۴]، مجاناً نرمال بودن نسبت لگاریتم درست‌نمایی در یک مدل مارکوفی ارگودیک مانا با تابع چگالی انتقال متعلق به خانواده نمایی شرطی و کاربرد در سری‌های زمانی غیرخطی [۴۸]، توزیع‌های پواسون آمیخته [۳۱]، کاربرد در معادلات دیفرانسیل تصادفی [۹]، بررسی نتایج هندسی در خانواده نمایی [۲۸]، دامنه جاذبه در خانواده‌های نمایی [۶]، مدل‌های خطی تعمیم‌یافته (GLM)^۲ [۱۵، ۶۸]، مدل‌های غیرخطی برای داده‌های رسته‌ای [۷۷]، مدل‌های رگرسیون کانونی [۱۰۱]، تحلیل عاملی [۷۲]، تحلیل مؤلفه اصلی [۷۱، ۱۸]، تحلیل مدل‌های لجستیک [۷]، روش کاهش واریانس برای متغیرهای پاسخ دوحالتی [۹۶]، داده‌های رسته‌ای [۲۹]، داده‌های دودویی چندمتغیره خوشه‌ای [۷۳]، خوشه‌بندی فازی

خانواده نمایی یک رده وسیع و مهم از توزیع‌های احتمال را در بر می‌گیرد که در اکثر شاخه‌های آماری کاربرد دارد. در این مقاله کاربردهایی از این خانواده در استنباط آماری بیان می‌شود. در ابتدا تحقیقاتی که در این زمینه در شاخه‌های مختلف آمار نظیر استنباط آماری، نظریه تصمیم آماری، احتمال، فرایندهای تصادفی، معادلات دیفرانسیل تصادفی، سری‌های زمانی، مدل‌های خطی و غیرخطی، آمار فازی و شاخه‌های دیگر آمار انجام شده است، معرفی می‌شود. سپس در بخش‌های بعدی پس از معرفی خانواده نمایی و شرایط خانواده نمایی منظم، مهم‌ترین کاربردهای این خانواده مهم در استنباط آماری را تشریح می‌کنیم. از مهم‌ترین این موارد که در بخش‌های بعدی بررسی می‌شوند نقش آنها در پیدا کردن مفاهیم آماری نظیر بسندگی، بسندگی مینیمال، کامل بودن، برآوردگرهای ماکسیمم درست‌نمایی، برآوردگرهای بیزی و مجاز و مینیماکس، برآوردگرهای ناریب با کمترین واریانس، نامساوی کرامر-رائو و کاکولوس، بهترین برآوردگر مجاناً نرمال، ماکسیمم کردن آنتروپی یک خانواده نمایی، محاسبه فاصله کولبک-لایبلر بین دو خانواده نمایی و آزمون‌های فرضیه آماری است. در مراجع [۱۰، ۱۵] می‌توان

^۱ هیأت علمی گروه آمار، دانشگاه کاشان، ایران

^۲ Generalized Linear Model

از روش مونته‌کارلو برای برآورد فاصله کولبک-لایبلر [۴۸]، توزیع‌های پیشینی غیرآگاهنده در خانواده‌ی نمایی واریانس درجه دو [۱۹، ۲۰]، نامساوی‌های نسبت پیشگویانه و واریانس توزیع پسینی [۵۳]، مجانباً نرمال بودن توزیع پسینی [۱۶]، انتشار مورد انتظار از دیدگاه بیزی [۸۷]، تعمیم معیار اطلاع بیزی (BIC) [۶]، دیدگاه بیزی ناپارامتری برای مدل‌بندی خوشه‌های تداخل [۴۵]، نرخ همگرایی برآوردهای بیزی تجربی [۶۵، ۶۶]، بیزی تجربی ناپارامتری [۳۸، ۳۹]، بیزی تجربی در آزمون‌های فرضیه با تابع زیان خطی در خانواده‌های نمایی غیر منظم [۹۰]، برآوردگر بیزی و مینیماکس تحت زیان ۱-۰ [۲] و برآورد دنباله‌ای سه مرحله‌ای برای میانگین با تابع زیان توان دوم خطاهای موزون [۹۳]، اشاره کرد.

همان‌طور که اشاره شد، یکی از خانواده‌ی توابع چگالی یا جرم احتمال مهم در آمار خانواده‌ی نمایی هستند که یک رده کلی از اکثر توابع چگالی یا جرم احتمال معروف در آمار را شامل می‌شوند و به خانواده کوپمن-پیتمن-داریموس (به واسطه تحقیقات مستقل این دانشمندان در این زمینه) نیز معروف هستند. یک خانواده توابع چگالی یا توابع جرم احتمال:

$$\{f(x|\theta) : \theta = (\theta_1, \dots, \theta_q) \in \Theta\}$$

را یک خانواده‌ی نمایی q -پارامتری گویند، هرگاه به‌ازای هر $x \in \mathcal{X}$ که $c(\theta) \geq 0$ و $h(x) \geq 0$ داشته باشیم:

$$f(x|\theta) = h(x)c(\theta) \exp \left[\sum_{i=1}^q w_i(\theta)t_i(x) \right]. \quad (1)$$

که توابع c ، h ، t_i و w_i توابع با مقادیر حقیقی هستند. بازپارامتریدن^۷ رابطه (۱) یکتا نیست، به‌عنوان نمونه، w_i می‌تواند در ثابت غیر صفر b ضرب و t_i به آن تقسیم شود. اگر $h(x) = g(x)I_A(x)$ که $g(x) \geq 0$ ، یک بازپارامتریدن دیگر به‌صورت زیر است:

$$f(x|\theta) = \exp \left[\sum_{i=1}^q w_i(\theta)t_i(x) + d(\theta) + s(x) \right] I_A(x),$$

[۳۴، ۴۶]، مدل‌های گرافیکی و استنباط تغییرات [۹۷]، مدل‌های دولایه‌ای نظیر ماشین ترکیب یا ماشین بولتزمن محدود شده و کاربرد آن در بازیابی اطلاعات [۹۸]، کاربرد در یادگیری به‌صورت خودآموز و رده‌بندی متن‌ها و ادراک رباتیک [۵۹]، روش‌های هسته [۱۱]، تکیه‌گاه در نظریه ماتروید جهت‌دار [۸۳]، بسته بودن تحت آماره‌های ترتیبی [۹]، معیاری برای تشخیص متعلق بودن به خانواده‌ی نمایی یک توزیع و مقدار فاصله یک توزیع دلخواه از نمایی بودن [۲۷]، ماتریس خانواده‌ی نمایی تحت محدودیت‌های ساختاری [۳۷]، نرخ خطر برای خانواده‌های نمایی با توان درجه دو [۸۰]، تعمیم معادله اشتاین [۱۴]، کارایی آزمون‌های نسبت درست‌نمایی [۵۴]، آزمون نسبت درست‌نمایی برای آزمون‌های فرضیه با فرضیه مقابل ترتیب مقید بودن [۴۹]، آزمون نیکویی برازش [۳۶]، مجاز بودن آزمون نیکویی برازش برای خانواده‌های نمایی گسسته [۱۷]، کارایی آزمون‌های بهینه در حضور پارامتر مزاحم [۷۰]، آزمون‌های امتیاز شرطی [۷۶]، آزمون‌های امتیاز اجتماع-اشتراک [۹۵]، آزمون‌های امتیاز اصلاح‌شده برای مدل‌های غیرخطی [۳۲]، آزمون همگنی مجانبی برای میانگین [۸۸]، دنباله مجموعه اطمینان محدب [۲۳]، وجود، سازگاری و مجانباً نرمال بودن MLE^۳ [۴، ۷، ۶۹]، خواص مجانبی MLE شرطی [۳۳]، تقریب تابع چگالی MLE در خانواده‌ی نمایی خمیده با استفاده از بسط نقطه‌ی زینی [۴۷]، برآورد هم‌زمان توسعه‌یافته شامل MLE، UMVUE^۴ و MRE^۵ [۳۵]، سازگاری و کارایی مجانبی و کاربرد آن در توزیع نرمال چندمتغیره [۸۵، ۸۶]، شرط وجود و یافتن آماره‌های کمکی در خانواده‌ی نمایی خمیده [۵۸]، بهترین برآوردگر مجانباً نرمال گام به گام [۸۶]، تقریب توسط چندجمله‌ای‌های دریکله [۶۷]، معیاری برای نیرومندی برآوردگرهای نارایب خانواده‌ی نمایی با تابع واریانس درجه دو [۶]، روش‌های دنباله‌ای و ثابت برای برآورد تفاضل میانگین‌های دو جامعه [۹۴]، چگالی‌های پیشینی مزدوج غیردقیق [۲۱]، استنباط با آمیخته‌ای از توزیع‌های پیشینی و استفاده

^۳ Maximum Likelihood Estimator

^۴ Uniformly Minimum Variance Unbiased Estimator

^۵ Minimum Risk Equivariant

^۶ Bayes Information Criterion

^۷ Reparametrization

است و فقط تابعی از یک متغیر که قید خطی داشته باشد، تابع ثابت است.

مثال ۱.۱. اگر $X \sim N(\mu, \sigma^2)$ ، $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ ، این که بازپارامتریدن معمولی به صورت رابطه (۱) یعنی

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] I_{\mathbb{R}}(x)$$

است که در آن $c(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\mu^2}{2\sigma^2}\right)$ ، $h(x) = I_{\mathbb{R}}(x)$ ، $\eta_1 = w_1(\theta) = \frac{-1}{2\sigma^2}$ ، $\eta_2 = w_2(\theta) = \frac{\mu}{\sigma^2}$ ، $t_1(x) = x^2$ ، $t_2(x) = x$ این خانواده یک خانواده نمایی دوپارامتری و به طور کلی تر یک REF است.

به طور مشابه توزیع های گاما، بتا، پواسون، دوجمله ای، پاسکال و هندسی خانواده نمایی هستند ولی توزیع هایی نظیر کوشی و یکنواخت متعلق به خانواده نمایی نیستند زیرا تابع چگالی آنها را نمی توان به صورت رابطه (۱) نوشت. اگر t_i یا η_i در قید خطی صدق کنند، آن گاه می توان تعداد جملات توان در رابطه (۱) را کاهش داد. به عنوان مثال، فرض کنید $\dim \Theta = j$ که در آن $(X_1, \dots, X_j) \sim M_j(n, p_1, \dots, p_j)$ ، در این صورت $\sum_{i=1}^j X_i = n$ و $\sum_{i=1}^j p_i = 1$. اگر $h(x) = n! / \prod_{i=1}^j x_i!$ ،

$$\begin{aligned} f(x_1, \dots, x_j) &= n! \prod_{i=1}^j \frac{p_i^{x_i}}{x_i!} \\ &= \exp \left[n \log p_j + x_j \log \frac{p_1}{p_j} + \dots \right. \\ &\quad \left. + x_{j-1} \log \frac{p_{j-1}}{p_j} \right] h(x) \end{aligned}$$

REF، $(j-1)$ -بعدی است. به طور مشابه اگر μ یک بردار سطری $j \times 1$ و Σ یک ماتریس معین مثبت $j \times j$ باشد، آن گاه بازپارامتریدن معمول $N_j(\mu, \Sigma)$ با $\dim \Theta = j + j^2$ یک REF با $\frac{j(j+1)}{2}$ پارامتر هست. این موارد مثال هایی از REF هستند که $\dim \Theta > \dim \Omega$. خانواده $N(\theta, \theta^2)$ ، خانواده نمایی دوپارامتری غیر منظم است (زیرا $\eta_1 = -\frac{1}{2\theta}$ و $\eta_2 = \frac{1}{\theta}$ و در پی آن نمودار فضای پارامتر طبیعی درجه دو است و شامل یک مستطیل دوبعدی نیست). در این جا $\dim \Omega = 2 < 1 = \dim \Theta$ عموماً

که در آن $s(x) = \log(g(x))$ ، $d(\theta) = \log(c(\theta))$ ، A بستگی به θ ندارد. بنا بر این خانواده $U(\cdot, \theta)$ متعلق به خانواده نمایی نیستند، زیرا تکیه گاه آن به θ بستگی دارد.

بازپارامتریدنی که کاربردهای نظری دارد استفاده از پارامتر طبیعی η است. فضای پارامتر طبیعی Ω یک مجموعه است که انتگرال تابع هسته متناهی باشد، یعنی:

$$\begin{aligned} \Omega &= \left\{ \eta = (\eta_1, \dots, \eta_k) : \frac{1}{c^*(\eta)} \right. \\ &\quad \left. \equiv \int_{-\infty}^{\infty} h(x) \exp \left[\sum_{i=1}^q \eta_i t_i(x) \right] dx < \infty \right\}, \end{aligned}$$

که یک مجموعه محدب است ([۳۰]، ص ۱۲۸). بازپارامتریدن طبیعی برای یک خانواده نمایی عبارت است از

$$f(x|\eta) h(x) c^*(\eta) \exp \left[\sum_{i=1}^q \eta_i t_i(x) \right] \quad (2)$$

که $h(x)$ و $t_i(x)$ همان توابع در رابطه (۱) هستند و $\eta \in \Omega$. مشابه قبل بازپارامتریدن در این جا نیز یکتا نیست.

اکنون به تعریف خانواده نمایی منظم (یا خانواده نمایی پر رتبه) پرداخته می شود. اگر $d_i(y)$ به صورت یکی از عبارت های t_i ، $w_i(\theta)$ یا η_i تعریف شود، یک قید خطی (وابسته خطی) توسط $d_1(y), \dots, d_q(y)$ به صورت $\sum_{i=1}^q a_i d_i(y) = c$ تعریف می شود که a_i ها و c ثابت اند به طوری که همه a_i ها صفر نیستند. اگر تساوی بالا فقط وقتی که $a_1 = \dots = a_k = 0$ برقرار باشد، آن گاه $d_i(y)$ یک قید خطی نیستند (مستقل خطی هستند). در حالتی که یک وابستگی خطی بین t_i ها برقرار باشد، برخی از اعضای فضای پارامتر اضافی هستند و در این حالت خانواده نمایی را بیش کامل گویند و در غیر این صورت آن را مینیمال نامند [۸۷]. شرایط زیر را در نظر بگیرید.

شرط ۱. فرض کنید در بازپارامتریدن طبیعی، η_i و t_i ها هیچ یک قید خطی نباشند.

شرط ۲. فرض کنید Ω شامل یک مستطیل k -بعدی باشد.

اگر شرط ۱ و ۲ برقرار باشد، خانواده نمایی منظم است (REF^۸). برای خانواده نمایی یک پارامتری، مستطیل یک بعدی، فاصله

^۸ Regular Exponential Family

اگر $\dim \Theta < \dim \Omega$ ، خانواده منظم نیست. کسلا و برگر ([۱۲]، ص ۱۱۴) به این نکته اشاره کرده‌اند که

$$\{\eta : \eta = (w_1(\theta), \dots, w_k(\theta)) | \theta \in \Theta\} \subseteq \Omega$$

اما اغلب Ω اکیداً مجموعه بزرگ‌تری است. به‌عنوان نمونه χ_p^2 یک REF نیست، زیرا مجموعه اعداد طبیعی یک زیرمجموعه محدب از اعداد حقیقی نیست، در صورتی که بازپارامتریدن طبیعی $\Gamma(\alpha, \gamma)$ یک REF است. فرض کنید $\dim \Theta = k = \dim \Omega$ و بازپارامتریدن معمولی

$$\Theta = \left\{ \theta \in \mathbb{R}^q : \int f(x|\theta) dx = 1 \right\}$$

در حد امکان بزرگ باشد و قرار دهید:

$$\Theta_\eta = \{ \theta \in \Theta : w_q(\theta), \dots, w_1(\theta) \}.$$

در این صورت فرض کنید فضای پارامتر طبیعی

$$\Omega = \{ (w_1(\theta), \dots, w_q(\theta)) : \theta \in \Theta_\eta \}$$

باشد. به عبارت دیگر برای سادگی تعریف کنید $\eta_i = w_i(\theta)$ به‌ازای هر توزیع عمومی، η یک تابع یک به یک از θ است.

مثال ۲.۱. اگر $X \sim B(n, p)$ ، $\Theta = [0, 1]$ و

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= h(x) c(p) \exp[w(p)t(x)], \end{aligned}$$

برای $x = 0, 1, \dots, n$ به‌طوری که $h(x) = \binom{n}{x} \geq 0$ ، $t(x) = x$ و $w(p) = \log \frac{p}{1-p}$ ، $c(p) = (1-p)^n \geq 0$ از این که $\eta = \log \frac{p}{1-p}$ برای $p = 0, 1$ تعریف نشده است، $\Theta_\eta = (0, 1)$.

خانواده‌ی نمایی یک پارامتری کانونی

$$f(x|\eta) = h(x) \exp \left(\sum_{i=1}^q \eta_i t_i(x) - \Lambda(\eta) \right) \quad (3)$$

را در نظر بگیرید. به راحتی ثابت می‌شود ([۳۰]، ص ۱۳۱):

$$\begin{aligned} E(t_i(X)) &= \frac{\partial}{\partial \eta_i} \Lambda(\eta), \\ Cov(t_i(X), t_j(X)) &= \frac{\partial^2}{\partial \eta_i \partial \eta_j} \Lambda(\eta). \end{aligned}$$

معادله اشتاین روشی ساده برای محاسبه گشتاورهای توزیع خانواده‌ی نمایی ارائه می‌کند ([۶۳]، ص ۳۱). در این روش به‌ازای هر تابع مشتق‌پذیر g که $E|g'(X)| < \infty$ و در حالتی که تکیه‌گاه متغیر تصادفی $\mathbb{R}$ است، داریم:

$$E \left\{ \left[\frac{h'(X)}{h(X)} + \sum_{i=1}^q \eta_i t'_i(X) \right] g(X) \right\} = -E(g'(X)) \quad (4)$$

و در حالتی که تکیه‌گاه متغیر تصادفی (a, b) است شرط برقراری رابطه (۴) این است که داشته باشیم:

$$\lim_{x \rightarrow a \text{ or } b} h(x) \exp \left\{ \sum_{i=1}^q \eta_i t_i(x) \right\} = 0.$$

در حالت توزیع نرمال $N(\mu, \sigma^2)$ رابطه بالا به صورت زیر

$$E\{g(X)(X - \mu)\} = \sigma^2 E(g'(X))$$

آشنای بدل می‌شود که در حالت اخیر با قرار دادن $g(x) = x$ و $E(X) = \mu$ به ترتیب گشتاورهای $E(X) = \mu$ و $E(X^2) = \mu^2 + \sigma^2$ به دست می‌آیند.

در پایان یک مثال نقض از [۱۴] ارائه می‌شود که در خانواده‌ی نمایی یک پارامتری کانونی رابطه (۳)، به صورت زیر

$$f(x|\eta) = h(x) \exp(\eta x - \Lambda(\eta)) I_{(0, \infty)}(x)$$

معادله اشتاین برقرار نیست. برای $n = 2, 3, \dots$ تعریف می‌کنیم:

$$h(x) = \begin{cases} \frac{1}{\gamma} \sin\left(\frac{\pi}{\gamma} x\right) & 0 \leq x \leq \gamma \\ \left(\frac{1}{n} + c_n(x)\right) + \left(\frac{1}{n} - c_n(x)\right) \times \cos(\pi n^\gamma (x - n) - \pi) & n \leq x \leq n + \frac{\gamma}{n^\gamma} \\ \frac{\gamma}{(n+1)^\gamma} & n + \frac{\gamma}{n^\gamma} \leq x < n+1 \end{cases}$$

که در آن

$$c_n(x) = \begin{cases} \frac{1}{n^\gamma} & n \leq x < n + \frac{1}{n^\gamma} \\ \frac{1}{(n+1)^\gamma} & n + \frac{1}{n^\gamma} \leq x < n + \frac{\gamma}{n^\gamma}. \end{cases}$$

اکنون با قرار دادن $g(x) = x$ مشاهده می‌شود، معادله اشتاین برای $\eta = 0$ برقرار نیست. در مقاله [۱۴] در یک قضیه اساسی شرایط نظامی که تحت آن معادله اشتاین برقرار است ذکر می‌شوند.

۲ خانواده‌های نمایی و بسندگی

(ب) یک آماره بسنده مینیمال برای η است، هرگاه $\eta_1, \dots, \eta_k$ قید خطی نباشند.

(ج) یک آماره کامل برای η است، هرگاه Ω شامل یک مستطیل k -بعدی باشد.

قسمت (الف) با استفاده از قضیه تجزیه فشر-نیمن قابل بررسی است [۶۲]، ص ۳۵. برای اثبات قسمت (ب)، با توجه به این که رابطه زیر

$$\frac{f(x|\eta)}{f(y|\eta)} = \frac{\prod_{j=1}^n h(x_j)}{\prod_{j=1}^n h(y_j)} \exp \left[\sum_{i=1}^k \eta_i (T_i(x) - T_i(y)) \right]$$

به η بستگی ندارد اگر و تنها اگر به ازای هر η_i و یک ثابت d ,

$$\sum_{i=1}^q \eta_i [T_i(x) - T_i(y)] = \sum_{i=1}^q \eta_i a_i = d$$

که در آن $T_i(x) = \sum_{j=1}^n t_i(x_j)$ و $a_i = T_i(x) - T_i(y)$ این که η_i ها قید خطی نیستند، داریم $a_i = 0$. بنا بر این $T(X)$ یک آماره بسنده مینیمال است ([۶۲]، ص ۳۹ و [۵۱]، ص ۳). برای اثبات قسمت (ج) به لیمان و رومانو ([۶۳]، ص ۱۱۶) رجوع کنید.

به عنوان نمونه در مثال ۱.۱، $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ یک آماره بسنده کامل برای (μ, σ^2) است و در پی آن توابع یک به یک $(\bar{X}, S^2)$ و $(\bar{X}, S)$ نیز بسنده کامل هستند. در مثال ۱.۲، $\sum_{i=1}^n t(X_i) = \sum_{i=1}^n X_i$ یک آماره بسنده کامل برای p است. شگردهای دیگر برای بررسی بسندگی مینیمال بودن در [۸۵] آمده است.

مثال ۱.۲. اگر $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \gamma^2 \mu^2)$ که $\gamma^2 > 0$ معلوم و $\mu > 0$ است، آن گاه $f(x|\mu)$ یک خانواده نمایی یک پارامتری با $\Theta = (0, \infty)$ (که شامل یک مستطیل یک بعدی است) هست و $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ یک آماره بسنده مینیمال است و از آنجایی که به ازای هر μ عبارت زیر برقرار است:

$$E_\mu \left[\frac{n+\gamma^2}{1+\gamma^2} \sum_{i=1}^n X_i^2 - \left[\sum_{i=1}^n X_i \right]^2 \right] = 0,$$

آماره بسنده مینیمال بالا کامل نیست. این مثال نشان می دهد که مستطیل باید در عوض Θ ، شامل Ω باشد. شایان ذکر است که

در تلخیص داده ها توسط یک آماره ممکن است اطلاعات نهفته در نمونه از بین برود و از این رو آماره ای که برای استنباط به کار برده می شود نباید باعث تلخیص نمونه اولیه به قسمی شود که دیگر برای بررسی مسئله کفایت نکند. علت مطالعه بسندگی در استنباط آماری این است که خلاصه کردن و فشرده کردن داده ها توسط آماره ای مورد علاقه هست که بدون از دست دادن اطلاعات، شامل تمام اطلاعات نمونه پیرامون پارامتر باشد. آماره بسنده مینیمال، حداکثر تلخیص را روی نمونه انجام می دهد، ولی ممکن است شامل اطلاعات اضافی (اطلاعات فرعی یا کمکی) باشد که برای تخمین پارامتر مجهول مفید نیست و در اصطلاح گوییم آماره بسنده هنوز به طور کامل فشرده نشده است و به عبارت دیگر ممکن است یک تابع از آماره بسنده که کمکی است وجود داشته باشد. وجود آماره بسنده کامل تضمین می کند که تابعی از آماره بسنده کامل که کمکی باشد وجود ندارد و این بدین معنی است که این آماره شامل اطلاعات کمکی و غیر لازم در باره پارامتر مجهول نیست. به عبارت عامیانه تر کامل بودن یک آماره بسنده، منجر به بیشترین فشرده گی در داده ها می شود و یا این که آماره بسنده کامل در ایجاد بیشترین فشرده گی در داده ها موفق است.

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از خانواده های نمایی با بازپارامتریدن معمولی داده شده در رابطه (۲) باشند در این صورت تابع چگالی احتمال یا جرم احتمال توأم آن که در ذیل آمده است یک خانواه نمایی k -پارامتری است:

$$f(x_1, \dots, x_n | \eta) = \left(\prod_{j=1}^n h(x_j) \right) [c^*(\eta)]^n \exp \left[\eta_1 \sum_{j=1}^n t_1(x_j) + \dots + \eta_k \sum_{j=1}^n t_k(x_j) \right].$$

در این صورت

$$T(X) = (T_1(X), \dots, T_k(X)) \\ = \left(\sum_{j=1}^n t_1(X_j), \dots, \sum_{j=1}^n t_k(X_j) \right),$$

(الف) یک آماره بسنده برای η است.

برای θ بسنده باشد، آن‌گاه P متعلق به یک خانواده‌ی نمایی است (در [۸۹]، ص ۱۴۴).

در پایان به این نکته اشاره می‌شود که اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x|\theta)$ متعلق به خانواده‌ی نمایی $P = \{P_\theta : \theta \in \Theta\}$ باشد، آن‌گاه توزیع X که روی مجموعه‌ی بول A بریده شده است (یعنی توزیع شرطی X به شرط $X \in A$) نیز متعلق به خانواده‌ی نمایی با تابع چگالی احتمال $f(x|\theta)I_A(x)/P_\theta(A)$ است ([۸۹]، ص ۱۴۴). همچنین به راحتی نشان داده می‌شود که اگر $T(X)$ یک آماره‌ی بسنده‌ی کامل برای خانواده P باشد، آن‌گاه این آماره برای خانواده توزیع بریده شده روی مجموعه‌ی بول A نیز بسنده‌ی کامل است ([۸۹]، ص ۱۴۸).

۳ خانواده‌های نمایی و MLE

روش ماکسیمم درست‌نمایی برای اولین بار توسط ریاضیدانانی چون لاگرانژ، برنولی، اویلر، لاپلاس و گاوس مورد استفاده قرار گرفت ولی سازماندهی اصلی و محبوبیت این روش به واسطه‌ی تحقیقات فیشر است و واژه «ماکسیمم درست‌نمایی» اولین بار توسط این دانشمند مشهور مورد استفاده قرار گرفت. کسلا و برگر ([۱۲]، ص ۳۱۷) یک نتیجه مفید برای یافتن برآورد ماکسیمم درست‌نمایی (MLE) در حالت یک پارامتری ارائه کرده‌اند. برای این منظور از این حقیقت استفاده می‌شود که اگر $K(\theta)$ یک تابع پیوسته روی (a, b) باشد که $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ و این تابع روی این فاصله مشتق‌پذیر و نقطه بحرانی یکتا باشد، در این صورت این نقطه اگر یک ماکسیمم موضعی باشد، ماکسیمم مطلق نیز هست. قابل یادآوری است که اگر MLE یکتا باشد، تابعی از آماره‌ی بسنده‌ی مینیمال است [۶۴، ۷۴]. این حقیقت در خانواده‌های نمایی دارای اهمیت است، زیرا در آنها لگاریتم تابع درست‌نمایی صورت ساده‌ای دارد و به راحتی می‌توان آماره‌ی بسنده‌ی مینیمال را پیدا کرد. به عنوان نمونه در مثال ۳.۲ MLE برای β عبارت است از $\hat{\beta} = T/(n+m)$. در حالت کلی با اختیار کردن یک نمونه n تایی از خانواده‌ی نمایی یک پارامتری

$$f(x|\eta) = h(x)c^*(\eta)\exp(\eta t(x))$$

یک آماره‌ی بسنده‌ی مینیمال k -پارامتری برای یک پارامتر j -بعدی که $j < k$ ، کامل نیست. در این مثال $j = 1 < 2 = k$ ([۲۲]، ص ۳۱).

مثال زیر نشان می‌دهد که هر آماره‌ی بسنده از یک REF لزومی ندارد کامل باشد.

مثال ۲.۲. اگر $X \sim N(0, \sigma^2)$ ، این خانواده یک REF با آماره‌ی بسنده‌ی مینیمال کامل X^2 است. آماره‌ی X بسنده است ولی تابعی از X^2 نیست و بنا بر این X بسنده‌ی مینیمال نیست و طبق قضیه‌ی بهادر [۵] کامل هم نیست.

مثال ۳.۲. اگر $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} E(\beta)$ و $Y_1, \dots, Y_m \stackrel{i.i.d.}{\sim} E(\frac{\beta}{\gamma})$ دو نمونه از هم مستقل باشند، آن‌گاه تابع چگالی توأم $f(x, y|\beta)$ متعلق به REF یک پارامتری با آماره‌ی بسنده‌ی کامل $\sum_{i=1}^n X_i + \gamma \sum_{i=1}^m Y_i$ است.

در حالت کلی لزومی ندارد که توزیع کناری یک خانواده‌ی نمایی، متعلق به خانواده‌ی نمایی باشد ولی در برخی حالات مثل نرمال چندمتغیره و چندجمله‌ای این خصوصیت برقرار است [۸۷]؛ ولی توزیع کناری آماره‌ی بسنده طبق لم زیر متعلق به خانواده‌ی نمایی است.

لم ۴.۲. اگر $X_1, \dots, X_n$ یک نمونه‌ی تصادفی از (۱) باشد، تابع احتمال آماره‌ی بسنده زیر

$$T(\mathbf{X}) = (T_1(\mathbf{X}), \dots, T_q(\mathbf{X})) \\ = \left(\sum_{j=1}^n t_1(X_j), \dots, \sum_{j=1}^n t_q(X_j) \right)$$

به صورت زیر خواهد بود:

$$f_T(t|\theta) = h_*(t)c_*(\theta) \exp \left[\sum_{i=1}^q w_i(\theta)t_i \right],$$

که متعلق به خانواده‌ی نمایی است و همچنین $\sum_{i=1}^q t_i(X_i)$ نیز متعلق به یک خانواده‌ی نمایی است ([۳۰]، ص ۱۲۷).

نشان داده می‌شود اگر $P = \{f(x|\theta) : \theta \in \Theta\}$ یک خانواده از توابع چگالی احتمال باشد که به ازای هر $\theta \in \Theta$ ، $f(x|\theta)$ یک تابع پیوسته در x باشد و برای $X_1, X_2 \stackrel{i.i.d.}{\sim} f(x|\theta)$

در حقیقت یک تابع محدب است. در تحلیل لگاریتم تابع درست‌نمایی ماکسیم شده یعنی

$$\Lambda^*(T) = \sup_{\eta \in \Omega} (\eta T - \Lambda(\eta))$$

را تبدیل فنچل-لژاندر Λ گویند. همچنین در تحلیل محدب داریم $\frac{d}{dt} \Lambda^*(T) = \arg \max_{\eta \in \Omega} (\eta T - \Lambda(\eta))$ در برخی موارد تبدیل فنچل-لژاندر، مشتق‌پذیر نیست و در این حالت $\arg \max$ بالا یک مجموعه به نام زیر دیفرانسیل Λ^* در t نامیده می‌شود. این حالت زمانی اتفاق می‌افتد که بیش از یک MLE وجود داشته باشد که در حالت خانواده‌نمایی یک پارامتری این اتفاق نمی‌افتد. مشاهده می‌شود که Λ^* یک MLE را محاسبه می‌کند، یعنی $\eta(\mu) = \dot{\Lambda}^*(\mu)$ همچنین داریم $\Lambda^*(\mu) = \eta(\mu)\mu - \Lambda(\eta(\mu))$ در پی آن $\dot{\Lambda}(\dot{\Lambda}^*(\mu)) = \dot{\Lambda}(\eta(\mu)) = E_{\eta(\mu)}(t(X)) = \mu$ نتیجه می‌دهد $\dot{\Lambda} \circ \dot{\Lambda}^*$ یک تبدیل همانی است و از این رو $\dot{\Lambda}^{-1} = \dot{\Lambda}^*$.

در حالتی که بخواهیم MLE برای μ به دست آوریم با نوشتن تابع $\tilde{l}(\mu) = l(\dot{\Lambda}^*(\mu))$ داریم

$$\begin{aligned} \frac{d}{d\mu} \tilde{l}(\mu) &= n \frac{\frac{d}{d\eta} (\eta T(\mathbf{X}) - \Lambda(\eta))}{\frac{d\mu}{d\eta}} \\ &= n \frac{T(\mathbf{X}) - \mu}{\text{Var}_{\eta(\mu)}(t(X))} \\ &= n \ddot{\Lambda}^*(\mu) [T(\mathbf{X}) - \mu] \end{aligned}$$

که نشان می‌دهد $\hat{\mu} = T(\mathbf{X})$ در [۶۹] برای یک خانواده‌نمایی به صورت $f(x|\eta) = \exp(\eta x - \Lambda(\eta))$ ، شرط وجود، سازگاری و مجانباً نرمال بودن MLE مورد تحلیل قرار می‌گیرد. در پایان با گرفتن لگاریتم از دو طرف رابطه زیر

$$M_{t(X)}(\theta) = e^{\Lambda(\theta+\eta) - \Lambda(\eta)}$$

و نوشتن بسط تیلور مقدار حاصل یعنی

$$\Lambda(\theta + \eta) - \Lambda(\eta) = \kappa_1 \theta + \kappa_2 \frac{\theta^2}{2} + \kappa_3 \frac{\theta^3}{6} + \dots$$

که در آن $\kappa_i = \Lambda^{(i)}(\eta)$ ها کومولان‌های $t(X)$ هستند، چولگی و برجستگی یک خانواده‌نمایی یک پارامتری به ترتیب به صورت

لگاریتم تابع درست‌نمایی به صورت زیر است:

$$\begin{aligned} l(\eta) &= \sum_{i=1}^n \ln(h(x_i)) + n \ln(c^*(\eta)) + \eta \sum_{i=1}^n t(x_i) \\ &= \sum_{i=1}^n \ln(h(x_i)) - n\Lambda(\eta) + \eta \sum_{i=1}^n t(x_i) \end{aligned}$$

که در آن $\Lambda(\eta) = -\ln(c^*(\eta))$. اگر تابع امتیاز به صورت $\dot{l}(\eta) = \frac{\partial}{\partial \eta} l(\eta)$ باشد، MLE برای η عبارت است از $\hat{\eta} = \arg \max_{\eta \in \Omega} l(\eta)$ که از معادله

$$\dot{l}(\eta) = n (T(\mathbf{x}) - \dot{\Lambda}(\eta)) = 0$$

محاسبه می‌شود و در آن $T(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n t(x_i)$ بنا بر این

$$\begin{aligned} \hat{\eta}(T(\mathbf{x})) &= \dot{\Lambda}^{-1}(T(\mathbf{x})) \\ &= \arg \max_{\eta \in \Omega} (\eta T(\mathbf{x}) - \Lambda(\eta)). \end{aligned}$$

در ابتدا امید ریاضی و واریانس $t(X)$ از روی تابع مولد گشتاور آن متغیر تصادفی پیدا می‌شود:

$$\begin{aligned} M_{t(X)}(\theta) &= E_{\eta}(e^{\theta t(X)}) \\ &= \int_{\Omega} h(x) e^{(\theta+\eta)t(x) - \Lambda(\eta)} dx \\ &= e^{\Lambda(\theta+\eta) - \Lambda(\eta)}. \end{aligned}$$

از طرف دیگر با به ترتیب یک بار و دو بار مشتق‌گیری از طرفین تساوی زیر

$$e^{\Lambda(\eta)} = \int_{\Omega} h(x) e^{\eta t(x)} dx$$

نسبت به η داریم:

$$\begin{aligned} \dot{\Lambda}(\eta) e^{\Lambda(\eta)} &= \int_{\Omega} h(x) t(x) e^{\eta t(x)} dx = e^{\Lambda(\eta)} E_{\eta}(t(X)), \\ (\ddot{\Lambda}(\eta) + \dot{\Lambda}'(\eta)) e^{\Lambda(\eta)} &= \int_{\Omega} h(x) t^2(x) e^{\eta t(x)} dx = e^{\Lambda(\eta)} E_{\eta}(t^2(X)). \end{aligned}$$

بنا بر این:

$$\begin{aligned} E_{\eta}(t(X)) &= \dot{\Lambda}(\eta), \\ \text{Var}_{\eta}(t(X)) &= \ddot{\Lambda}(\eta). \end{aligned}$$

با توجه به قضیه حدی مرکزی می‌توان نتیجه گرفت که $T(\mathbf{x})$ به طور تقریبی دارای توزیع $(\dot{\Lambda}(\eta), \frac{1}{n} \ddot{\Lambda}(\eta))$ است. همچنین به راحتی می‌توان ثابت کرد که $\Lambda(\eta)$ بی‌نهایت بار مشتق‌پذیر و

زیر محاسبه می‌شوند:

$$\gamma(\eta) = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{\Lambda^{(3)}(\eta)}{\Lambda(\eta)^{3/2}},$$

$$\delta(\eta) = \frac{\kappa_4}{\kappa_2^2} = \frac{\Lambda^{(4)}(\eta)}{\Lambda(\eta)^2}.$$

$$P(X=r) = \frac{\gamma(r)}{c(\theta)} \theta^r$$

برابر است با $\gamma(r)h(T)$ [۸۹]، ص ۱۶۵).

برخی از مرجع‌های آماری پیشنهاد می‌کنند که UMVUE توسط تعیین این که یک برآوردگر نارایب W دارای واریانس برابر با کران پایین نامساوی کرامر-رائو (CRLB)^۹ است یا نه پیدا می‌شود. این روش به ندرت جواب می‌دهد زیرا عموماً تساوی زمانی برقرار است که

۱. داده‌ها از یک REF یک پارامتری با آماره بسنده کامل T استخراج شوند.

۲. $W = a + bT$ یک تابع خطی از T باشد.

در حالتی که داده‌ها از یک REF یک پارامتری یا REF، $-q$ پارامتری با $q > 1$ بیابند، عموماً برای توابع غیرخطی از T ، نامساوی کرامر-رائو اکید است. اگر T کامل باشد، $g(T)$ ، UMVUE امید خودش است و تعیین این که T یک آماره بسنده کامل از یک REF یک پارامتری است نسبت به محاسبه $\text{Var}_\theta W$ و CRLB ساده‌تر است. اگر خانواده‌ی نمایی نباشد، CRLB ممکن است یک کران پایین برای واریانس برآوردگر نارایب باشد [۵۲، ۱۰۰]. اگرچه CRLB برای یافتن UMVUE‌ها خیلی مورد استفاده قرار نمی‌گیرد، اما برای یافتن واریانس‌های مجانبی UMVUE‌ها و MLE‌ها مفید است [۵۵، ۸۱]. در حالت کلی تابع امتیاز برای یک تابع وارون‌پذیر η یعنی $h(\eta) = \xi$ برابر است با:

$$\frac{d}{d\xi} l(h^{-1}(\xi)) = \frac{\frac{d}{d\eta} l(\eta)|_{\eta=h^{-1}(\xi)}}{\frac{dh}{d\eta}|_{\eta=h^{-1}(\xi)}}$$

که دارای خاصیت اصلی زیر است:

$$\begin{aligned} E_\eta \left[\frac{d}{d\xi} l(\xi) |_{\xi=\eta} \right] &= n \left[E_\eta (T(\mathbf{X})) - \dot{\Lambda}(\eta) \right] \\ &= n [\mu(\eta) - \mu(\eta)] \\ &= 0. \end{aligned}$$

۴ خانواده‌های نمایی و UMVUE CRLB

در انتخاب یک برآوردگر نقطه‌ای به دنبال برآوردگری هستیم که به ازای تمام مقادیر در فضای پارامتر، تابع مخاطره (متوسط تابع زیان) را کمینه کند، ولی با توجه به بزرگی کلاس برآوردگرها چنین امکانی وجود ندارد. یکی از راه‌ها برای حل این مشکل (پیدا کردن برآوردگرهای بهینه)، محدود کردن کلاس برآوردگرها و پیدا کردن بهترین برآوردگر در کلاس محدود شده است. اگر فقط کلاس برآوردگرهای نارایب در نظر گرفته شود، برآوردگر بهینه با این ساختار را برآوردگر نارایب با کمترین واریانس (UMVUE) نامند. در ابتدا به یافتن یک UMVUE در یک کلاس کلی از خانواده‌ی نمایی گسسته پرداخته می‌شود.

مثال ۱.۴. فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از خانواده‌ی نمایی یک پارامتری با تابع احتمال زیر باشد.

$$P(X_i = x) = \frac{\gamma(x)}{c(\theta)} \theta^x, \quad x = 0, 1, \dots$$

با استفاده از لم ۴.۲، توزیع آماره بسنده $T(\mathbf{X}) = \sum_{i=1}^n X_i$ نیز یک خانواده‌ی نمایی یک پارامتری با تابع احتمال زیر است:

$$P(T = t) = \frac{\gamma_n(t)}{[c(\theta)]^n} \theta^t, \quad t = 0, 1, \dots$$

از این که رابطه‌ی زیر برقرار است:

$$\sum_{t=0}^{\infty} h(t) \gamma_n(t) \theta^t = [c(\theta)]^{n-p} \theta^r,$$

می‌توان نتیجه گرفت که UMVUE برای $\theta^r/[c(\theta)]^p$ (که r و p اعداد صحیح نامنفی هستند) به صورت زیر خواهد بود:

$$h(T) = \begin{cases} 0 & T < r \\ \frac{\gamma_{n-p}(T-r)}{\gamma_n(T)} & T \geq r \end{cases}$$

^۹ Cramer Rao Lower Bound

معقول است و چون در عمل مقدار η مجهول است از برآورد آن استفاده می‌شود یعنی:

$$\widehat{\text{Var}}(\hat{\xi}) \approx \frac{\dot{h}(\hat{\eta})^2}{n\text{Var}_{\hat{\eta}}(t(X))}.$$

برای بررسی بیشتر موضوع در حالت خانواده‌ی نمایی چندمتغیره، تحت شرایط نظم ثابت می‌شود کمترین مقدار ماتریس اطلاع فیشر $I(\theta)$ از دیدگاه نیمه‌معین مثبت بودن توسط خانواده‌ی نمایی زیر

$$f(x|\theta) = h(x) c(\theta) \exp \{Q^t(\theta) T(x)\} I_A(x)$$

که در آن A به پارامتر $\theta \in \Theta \subseteq \mathbb{R}^q$ بستگی ندارد و یک مجموعه‌ی باز است، $h(x), c(\theta) > 0$ و $Q(\theta)$ اولین مشتقات جزئی پیوسته دارد حاصل می‌شود و مقدار مینیمم برابر با ماتریس $M^t(\theta)\Sigma^{-1}M(\theta)$ است که در آن Σ ماتریس کواریانس $T(X) = (T_1(X), \dots, T_k(X))$ و $M(\theta) = \left[\frac{\partial E_{\theta}(T_i(X))}{\partial \theta_j} \right]_{k \times q}$ ماتریس مشتقات جزئی است [۱۰۶]. همچنین ثابت می‌شود دنباله‌ی ماتریس‌های اطلاع فیشر $\{I_n(\theta)\}_{n \in \mathbb{N}}$ صعودی است از این جهت که به‌ازای هر $n_1, n_2 \in \mathbb{N}$ که $n_1 \leq n_2$ ، $I_{n_2}(\theta) - I_{n_1}(\theta) \geq 0$ ، یعنی ماتریس $I_{n_2}(\theta) - I_{n_1}(\theta)$ نیمه‌معین مثبت است [۱۰۶]. در حالت چندمتغیره نیز مشابه حالت یک‌متغیره نامساوی کرامر-رائو به تساوی تبدیل می‌شود اگر و تنها اگر:

الف) $\frac{\partial}{\partial \theta_i} \log f(x|\theta) = \sum_{j=1}^q u_{ij} [T_j(x) - E_{\theta}(T_j(X))]$ که در آن u_{ij} ها ممکن است به θ بستگی داشته باشند ولی به x و $E_{\theta}(T_j(X))$ وابسته نیستند ([۱۰۵]، ص ۱۹۵).

ب) کران پایین نامساوی کرامر رائو زمانی حاصل می‌شود که $f(x|\theta)$ متعلق به خانواده‌ی نمایی q -پارامتری باشد یعنی:

$$f(x|\theta) = h(x) c(\theta) \exp \left\{ \sum_{j=1}^q Q_j(\theta) T_j(x) \right\} I_A(x)$$

که در آن A به پارامتر $\theta = (\theta_1, \dots, \theta_q)$ بستگی ندارد [۱۰۶].

با ترکیب نتایج الف و ب می‌توان کلاس توابع پارامتری و UMVUE های متناظر را از طریق نامساوی کرامر-رائو

اطلاع فیشر در نمونه $X_1, \dots, X_n$ برای $\xi = h(\eta)$ در $\eta \in \Omega$ به صورت $\left[\frac{d}{d\xi} l(h^{-1}(\xi)) \right]$ $\text{Var}_{\eta(\xi)}$ است، یعنی

$$I_{\eta}^{(n)}(h(\eta)) = E_{\eta} \left[\frac{d}{d\xi} l(h^{-1}(\xi)) \Big|_{\xi=h(\eta)} \right]^2 = \frac{I_{\eta}^{(n)}}{h(\eta)^2}$$

در عبارت بالا

$$\begin{aligned} I_{\eta}^{(n)} &= I_{\eta}^{(n)}(\eta) \\ &= nE_{\eta} \left[\left(t(X) - \dot{\Lambda}(\eta) \right)^2 \right] \\ &= n\text{Var}_{\eta}(t(X)) \end{aligned}$$

که در آن $\text{Var}_{\eta}(t(X))$ اطلاع فیشر برای η در یک مشاهده است. با توجه به این که $-\ddot{l}(\eta) = n\text{Var}_{\eta}(t(X))$ ، نتیجه می‌شود که $I_{\eta}^{(n)} = -\ddot{l}(\eta)$ اکنون بر اساس یک متغیر تصادفی متعلق به خانواده‌ی نمایی یک پارامتری زیر

$$f(x|\eta) = h(x) \exp(\eta t(x) - \Lambda(\eta))$$

$\text{Var}_{\eta}(t(X))$ تحت شرایط نظم برابر است با اطلاع فیشر برای پارامتر طبیعی η و یا برابر است با وارون اطلاع فیشر برای پارامتر $E_{\eta}(t(X))$ ([۱۰۹]، ص ۱۷۱). همچنین بر اساس یک نمونه تصادفی n تایی از توزیع بالا، CRLB برای برآوردگر نارایب $\hat{\xi}$ از $\xi = h(\eta)$ برابر است با:

$$\begin{aligned} \text{Var}_{\eta}(\hat{\xi}) &\geq \frac{1}{I_{\eta}^{(n)}(h(\eta))} \\ &= \frac{\dot{h}(\eta)^2}{n\text{Var}_{\eta}(t(X))}. \end{aligned}$$

با به کار گیری این نامساوی برای $\mu = \dot{\Lambda}(\eta)$ داریم:

$$\begin{aligned} \text{Var}_{\eta}(\hat{\mu}) &\geq \frac{\ddot{\Lambda}(\eta)^2}{n\text{Var}_{\eta}(t(X))} \\ &= \frac{\text{Var}_{\eta}(t(X))}{n} \end{aligned}$$

و $\hat{\mu}$ تساوی در CRLB را می‌دهد که این اتفاق فقط برای توابع خطی از μ رخ می‌دهد.

با این که در حالت کلی MLE یعنی $\hat{\xi} = h(\hat{\eta})$ نارایب نیست، با استفاده از تقریب دلتا تساوی تقریبی زیر

$$\text{Var}(\hat{\xi}) \approx \frac{\dot{h}(\eta)^2}{n\text{Var}_{\eta}(t(X))}$$

مثال ۲.۴. در توزیع سه‌جمله‌ای با تابع جرم احتمال زیر

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 | \theta_1, \theta_2, \theta_3) \\ = \frac{n!}{x_1! x_2! x_3!} \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{x_3}, \\ x_1 + x_2 + x_3 = n, \quad \theta_1 + \theta_2 + \theta_3 = 1$$

داریم، $Q_1(\theta) = \frac{\theta_1}{1 - \theta_1 - \theta_2}$ ، $c(\theta) = (1 - \theta_1 - \theta_2)^n$ ، $Q_2(\theta) = \frac{\theta_2}{1 - \theta_1 - \theta_2}$ و $T_1(X) = X_1$ ، $T_2(X) = X_2$ ، $h(x) = \frac{n!}{x_1! x_2! x_3!} I_A(x)$ بنا بر این با توجه به رابطه (۵) داریم:

$$ET_1 = n\theta_1,$$

$$ET_2 = n\theta_2,$$

$$\frac{\partial}{\partial \theta_1} \log P = \frac{n(1 - \theta_2)}{\theta_1(1 - \theta_1 - \theta_2)} \left(\frac{X_1}{n} - \theta_1 \right) + \frac{n}{1 - \theta_1 - \theta_2} \left(\frac{X_2}{n} - \theta_2 \right),$$

$$\frac{\partial}{\partial \theta_2} \log P = \frac{n}{1 - \theta_1 - \theta_2} \left(\frac{X_1}{n} - \theta_1 \right) + \frac{n(1 - \theta_1)}{\theta_2(1 - \theta_1 - \theta_2)} \left(\frac{X_2}{n} - \theta_2 \right),$$

$$I^{-1}(\theta) = \frac{1}{n} \begin{bmatrix} \theta_1(1 - \theta_1) & -\theta_1\theta_2 \\ -\theta_1\theta_2 & \theta_2(1 - \theta_2) \end{bmatrix}.$$

بنا بر این برآوردگر $(\frac{X_1}{n}, \frac{X_2}{n})$ برای $UMVUE$ ، $T^t = (T_1^*, T_2^*) = (\frac{X_1}{n}, \frac{X_2}{n})$ است و

$$VarT_1^* = I_{11}^{-1}(\theta) = \frac{\theta_1(1 - \theta_1)}{n},$$

$$VarT_2^* = I_{22}^{-1}(\theta) = \frac{\theta_2(1 - \theta_2)}{n}.$$

به غیر از نامساوی کرامر-رائو، نامساوی‌های دیگری نیز هستند که کران پایینی برای واریانس برآوردگر ارائه می‌کنند. از جمله معروف‌ترین آنها نامساوی کاکولوس است که پاپاتاناسیو [۷۸] آن را برای خانواده‌ی نمایی تعمیم داد. فرض کنید $(X_1, \dots, X_q)$ یک خانواده‌ی نمایی q -بعدی پیوسته با تابع چگالی زیر باشد:

$$f(x|\theta) = c(\theta) \exp(\theta \cdot x - s(\theta)) I_A(x)$$

که در آن $\theta = (\theta_1, \dots, \theta_q)$ یک بردار پارامتر طبیعی و « $\cdot$ » نشان‌دهنده ضرب اسکالر و مجموعه A یک اجتماع متناهی از مجموعه‌های همبند باز در $\mathbb{R}^q$ و $s(x)$ دارای مشتقات جزئی پیوسته باشد. تحت شرایط نظم مثل متناهی بودن قدر مطلق امید ریاضی [۱۴] نشان داده می‌شود:

$$E[(\nabla s(X) - \theta)g(X)] = E[\nabla g(X)]$$

پیدا کرد. برای سادگی حالت $q = 2$ بررسی می‌شود. اگر $X_1, \dots, X_n$ یک نمونه تصادفی از عبارت زیر باشد:

$$f(x|\theta) = h(x) c(\theta) \exp \left\{ \sum_{j=1}^2 Q_j(\theta) T_j(x) \right\} I_A(x)$$

باشد، برآوردگر $T^t = (T_1, T_2)$ از دیدگاه نامساوی کرامر-رائو یک UMVUE برای $\theta^t = (\theta_1, \theta_2)$ است هرگاه:

$$ET_1 = \theta_1,$$

$$ET_2 = \theta_2,$$

$$VarT_1 = I_{11}^{-1}(\theta) = \frac{I_{22}(\theta)}{\det(I(\theta))},$$

$$VarT_2 = I_{22}^{-1}(\theta) = \frac{I_{11}(\theta)}{\det(I(\theta))}$$

که در آن $I_{ii}(\theta) = E\left(\frac{\partial}{\partial \theta_i} f(x|\theta)\right)^2$ و $I_{ii}^{-1}(\theta)$ امین-دراية قطری ماتریس $I^{-1}(\theta)$ است.

با توجه به نتیجه الف داریم:

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \log \prod_{i=1}^n f(x_i|\theta) &= nQ'_{1\theta_1} \left[\frac{1}{n} \sum_{i=1}^n T_1(x_i) - E_{\theta}(T_1(X)) \right] \\ &\quad + nQ'_{2\theta_1} \left[\frac{1}{n} \sum_{i=1}^n T_2(x_i) - E_{\theta}(T_2(X)) \right], \\ \frac{\partial}{\partial \theta_2} \log \prod_{i=1}^n f(x_i|\theta) &= nQ'_{1\theta_2} \left[\frac{1}{n} \sum_{i=1}^n T_1(x_i) - E_{\theta}(T_1(X)) \right] \\ &\quad + nQ'_{2\theta_2} \left[\frac{1}{n} \sum_{i=1}^n T_2(x_i) - E_{\theta}(T_2(X)) \right], \end{aligned} \quad (۵)$$

که در آن

$$(E_{\theta}(T_1), E_{\theta}(T_2))^t = [Q(\theta)]^{-1} c_*(\theta),$$

با $c_*^t(\theta) = (c'_1, c'_2)$ ، $Q(\theta) = [Q'_{i\theta_j}]_{r \times r} = \left[\frac{\partial}{\partial \theta_j} Q_i(\theta) \right]_{r \times r}$ و $c'_i = \frac{\partial}{\partial \theta_i} \log c(\theta)$ بنا بر این برآوردگر $T^t = (T_1^*, T_2^*)$ با $T_j^* = \frac{1}{n} \sum_{i=1}^n T_j(x_i)$ برای UMVUE $(E_{\theta}(T_1), E_{\theta}(T_2))$ است. (برآوردگر ناریب با واریانس به‌طور یکنواخت مینیمم)های دیگر از توابع پارامتری دیگر تبدیلات خطی از $(E_{\theta}(T_1), E_{\theta}(T_2))$ هستند. به عبارت دیگر از دیدگاه نامساوی کرامر-رائو فقط توابع پارامتری برآوردپذیر به صورت $(k_1 + k_2 E_{\theta}(T_1), k_3 + k_4 E_{\theta}(T_2))$ هستند که UMVUE نظیر آن $(k_1 + k_2 T_1^*, k_3 + k_4 T_2^*)$ است و k_i ها ثابت‌هایی هستند که به θ و x_i ها بستگی ندارند. از این رو برای خانواده‌های غیرنمایی نمی‌توان UMVUE ها را از روش نامساوی کرامر-رائو پیدا کرد.

در حالتی که یک نمونه تصادفی n تایی از خانواده‌ی نمایشی یک پارامتری در اختیار باشد، مجموع فاصله به صورت زیر خواهد بود:

$$D^{(n)}(\eta_1; \eta_2) = n D(\eta_1; \eta_2)$$

که با افزایش نمونه زیاد می‌شود. با بسط تیلور مرتبه اول و دوم داریم:

$$\begin{aligned} \Lambda(\eta_2) &= \Lambda(\eta_1) + (\eta_1 - \eta_2) \dot{\Lambda}(\eta_1) - \frac{1}{2} D(\eta_1; \eta_2) \\ &= \Lambda(\eta_1) + (\eta_1 - \eta_2) \dot{\Lambda}(\eta_1) + \frac{1}{6} (\eta_1 - \eta_2)^2 \ddot{\Lambda}(\eta_1) \\ &\quad + \frac{1}{24} \Lambda^{(3)}(\vartheta) (\eta_1 - \eta_2)^3 \end{aligned}$$

و $\vartheta = \vartheta(\eta_1, \eta_2) \in [\eta_1, \eta_2]$ یک مقدار دلخواه است. در این صورت:

$$\left| D(\eta_1; \eta_2) - I_{\eta_1}(\eta_1 - \eta_2) \right| \leq \frac{1}{6} C(\eta_1; |\eta_1 - \eta_2|) (\eta_1 - \eta_2)^3$$

که در آن

$$\begin{aligned} C(\eta; r) &= \sup_{|\vartheta - \eta| \leq r} \frac{|\ddot{\Lambda}(\vartheta) - \ddot{\Lambda}(\eta)|}{|\vartheta - \eta|} \\ &= \sup_{\vartheta: |\vartheta - \eta| \leq r} \frac{|I_\vartheta - I_\eta|}{|\vartheta - \eta|} \end{aligned}$$

یک ثابت لپ‌شیتز موضعی یا هنگ پیوستگی برای I در η است. اگر متغیر تصادفی X متعلق به خانواده‌ی نمایشی چندپارامتری

$$f(x|\eta) = \exp \left[\sum_{i=1}^k \eta_i t_i(x) - \Lambda(\eta) \right]$$

باشد و بخواهیم آنتروپی زیر را

$$\begin{aligned} H(X) &= E(-\log f(X|\eta)) \\ &= - \int f(x|\eta) \log f(x|\eta) d\mu(x) \end{aligned}$$

تحت شرط

$$E(t_i(X)) = \int f(x|\eta) t_i(x) d\mu(x) = \alpha_i$$

که در آن $i = 1, \dots, k$ ماکسیم کنیم، با استفاده از روش ضرایب لاگرانژ داریم:

$$\begin{aligned} \psi(f, \theta_1, \dots, \theta_k, \lambda) &= \int f(x|\eta) \log f(x|\eta) d\mu(x) \\ &\quad + \sum_{i=1}^k \theta_i \left(\alpha_i - \int f(x|\eta) t_i(x) d\mu(x) \right) \\ &\quad + \theta_0 \left(\int f(x|\eta) d\mu(x) - 1 \right) \\ &\quad - \int f(x|\eta) \lambda(x) d\mu(x). \end{aligned}$$

که تعمیمی بر معادله اشتاین [۹۱، ۹۲] است. همچنین با برقراری شرایط نظم بالا پاپاتاناسیو [۷۸] عبارت زیر را ثابت کرد:

$$\text{Var}(g(\mathbf{X})) \geq E[\nabla g(\mathbf{X})] J^{-1} E[\nabla g(\mathbf{X})]$$

که در آن $J = (s_{ij})$ وارون ماتریس نامفرد $J = (s_{ij})$ با درایه‌های $s_{ij} = E(s_{ij}(\mathbf{X}))$ است که در آن $s_{ij}(\mathbf{X}) = \frac{\partial^2 s(\mathbf{X})}{\partial x_i \partial x_j}$ هست. شایان ذکر است که پاپاتاناسیو [۷۸] نشان داد این مطلب هم صحیح است و با برقراری نامساوی بالا و داشتن شرایط نظم، خانواده‌ی نمایشی به صورت داده شده است. به عنوان نمونه در توزیع نرمال چندمتغیره نامساوی به صورت زیر بدل می‌شود:

$$\text{Var}(g(\mathbf{X})) \geq E[\nabla g(\mathbf{X})] \Sigma E[\nabla g(\mathbf{X})].$$

برای اطلاعات بیشتر در مورد نامساوی‌های مشابه در حالت خانواده‌ی نمایشی گسسته می‌توانند به [۷۸] مراجعه کنند. در انتهای این بخش با انجام محاسبات می‌توان اطلاع متقابل (فاصله کولبک-لایبلر) بین دو خانواده‌ی نمایشی یک پارامتری

$$f(x|\eta_1) = h(x) \exp(\eta_1 t(x) - \Lambda(\eta_1)),$$

و

$$f(x|\eta_2) = h(x) \exp(\eta_2 t(x) - \Lambda(\eta_2)).$$

را به صورت زیر محاسبه کرد:

$$D(\eta_1; \eta_2) = 2 \left[\Lambda(\eta_2) - \Lambda(\eta_1) + (\eta_1 - \eta_2) \dot{\Lambda}(\eta_1) \right] \geq 0.$$

در حقیقت داخل براکت باقیمانده بسط تیلور زیر است:

$$\Lambda(\eta_2) = \Lambda(\eta_1) + (\eta_2 - \eta_1) \dot{\Lambda}(\eta_1) + R(\eta_1; \eta_2);$$

که

$$R(\eta_1; \eta_2) = \frac{1}{6} D(\eta_1; \eta_2) = \frac{1}{6} \ddot{\Lambda}(\vartheta) (\eta_1 - \eta_2)^2$$

و $\vartheta = \vartheta(\eta_1, \eta_2) \in [\eta_1, \eta_2]$ یک مقدار دلخواه است. به راحتی مشاهده می‌شود که

$$\arg \max_{\eta \in \Omega} (\eta T(X) - \Lambda(\eta)) = \arg \min_{\eta \in \Omega} D(\eta(T(X)); \eta)$$

یعنی برآورد ماکسیم درست‌نمایی با برآوردگر حد اقل فاصله کولبک-لایبلر یکسان است.

از حل معادله‌ی زیر

$$\frac{d\psi(f, \theta_*, \dots, \theta_k, \lambda)}{df(x|\eta)} = 1 + \log f(x|\eta) - \sum_{i=1}^k \theta_i t_i(x) + \theta_* - \lambda(x) = 0.$$

مقدار ماکسیمم آنتروپی برابر است با:

$$f(x|\eta) = \exp \left[\sum_{i=1}^k \eta_i t_i(x) - 1 - \theta_* - \lambda(x) \right].$$

با توجه به برقراری شرط $f(x|\eta) > 0$ با اختیار کردن $\lambda(x) = 0$ و همچنین انتخاب $\theta_* = -1 + \Lambda(\eta)$ داریم:

$$f(x|\eta) = \exp \left[\sum_{i=1}^k \eta_i t_i(x) - \Lambda(\eta) \right],$$

که می‌توان نشان داد جواب یکتاست. برای این منظور اگر فرض شود متغیر تصادفی Y متعلق به خانواده‌ی نمایی چندپارامتری و با تابع چگالی g باشد و شرایط قبلی برقرار باشد، آن‌گاه:

$$\begin{aligned} H(Y) &= - \int g \log g d\mu \\ &= - \int g \log \frac{g}{f} d\mu - \int g \log f d\mu \\ &= -D(g; f) - \int g(x|\eta) \left(\sum_{i=1}^k \eta_i t_i(x) - \Lambda(\eta) \right) d\mu(x) \\ &= -D(g; f) - \int f(x|\eta) \left(\sum_{i=1}^k \eta_i t_i(x) - \Lambda(\eta) \right) d\mu(x) \\ &= -D(g; f) - H(X), \end{aligned}$$

که در تساوی ماقبل آخر از رابطه‌ی زیر استفاده شد:

$$\int f(x|\eta) t_i(x) d\mu(x) = \int g(x|\eta) t_i(x) d\mu(x) = \alpha_i$$

با توجه به این که $D(g; f) > 0$ مگر $f = g$ ، نتیجه گرفته می‌شود که جواب یکتا است.

مثال ۳.۴. ماکسیمم آنتروپی روی توزیع‌ها تحت شرط $E(X^2) = 1$ را در حالت‌های زیر محاسبه می‌شود:

الف) با در نظر گرفتن اندازه‌ی شمارشی μ روی مجموعه‌ی $\{-1, 1\}$ به‌طوری که $\mu(\{1\}) = \mu(\{-1\}) = 1$ ، ماکسیمم آنتروپی برای توزیع یکنواخت گسسته $DU\{-1, 1\}$ رخ می‌دهد.

ب) برای اندازه‌ی لبگ μ روی $\mathbb{R}$ یعنی $\mu([a, b]) = b - a$ که $a \leq b$ ، ماکسیمم آنتروپی برای توزیع $N(0, 1)$ اتفاق می‌افتد.

۵ خانواده‌های نمایی و بهترین برآوردگر

یکی از برآوردگرهای بهینه بهترین برآوردگر مجانباً نرمال است که ساختار مجانبی برای نمونه‌های بزرگ را نشان می‌دهد. یک نمونه‌ی تصادفی مثل $X_1, \dots, X_n$ از خانواده‌ی نمایی چندپارامتری در نظر بگیرید که تابع چگالی آن به‌صورت زیر باشد:

$$f(x|\theta) = h(x)c(\theta) \exp \left[\sum_{i=1}^q w_i(\theta) t_i(x) \right].$$

فرض کنید به‌ازای هر $i = 1, \dots, q$ ، $E(t_i(X)) = c_i \theta + d_i$ که c_i ها و d_i ها مقادیر معلوم هستند. همچنین علاوه بر برقراری شرایط نظم برای برقراری نامساوی کرامر-رائو فرضیات زیر نیز برقرار باشد:

۱. به‌ازای هر $\theta \in \Theta$ و $i = 1, \dots, q$ ، $w'_i(\theta)$ در θ پیوسته باشند.

۲. به‌ازای هر $i = 1, \dots, q$ داشته باشیم:

$$\sup_{\theta \in \Theta} \left| w'_i(\theta) / \sum_{j=1}^q c_j w'_j(\theta) \right| < \infty.$$

۳. برای یک $\omega > 0$ ، $E |c_j \theta + d_j - t_j(X)|^{2+\omega} < \infty$.

اگر $\bar{w}_i(\theta)$ یک گسترش $w'_i(\theta)$ روی $\mathbb{R}$ باشد که به‌ازای هر $\theta \in \Theta$ و $i = 1, \dots, q$ داشته باشیم:

$$\sum_{j=1}^q c_j \bar{w}_j(\theta) \neq 0,$$

$$\sup_{\theta \in \Theta} \left| \bar{w}_i(\theta) / \sum_{j=1}^q c_j \bar{w}_j(\theta) \right| < \infty$$

و به‌ازای هر $n \in \mathbb{N}$ تعریف کنیم:

$$Y_n(\theta) = \frac{\sum_{j=1}^q \bar{w}_j(\theta) (c_j \theta + d_j - t_j(X))}{\sum_{j=1}^q c_j \bar{w}_j(\theta)},$$

آن‌گاه در [۸۶] ثابت می‌شود برآوردگر گام به گام $\hat{\theta}_n$ و با قرار دادن

$$\bar{w}_1(x) = \bar{w}_2(x) = \begin{cases} x(1-x^2)^{-2} & |x| < 1 \\ x & |x| \geq 1 \end{cases}$$

برای θ که به صورت بازگشتی به ازای هر $n \in \mathbb{N}$ به صورت زیر تعریف می‌شود:

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \frac{1}{n} Y_n(\hat{\theta}_{n-1})$$

به طوری که $\hat{\theta}_0$ یک مقدار اولیه است. بهترین برآوردگر مجانباً نرمال برای θ است، یعنی

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2),$$

که در آن $\xrightarrow{d}$ به معنی همگرایی در توزیع است و CRLB برای یک برآوردگر نااریب θ برابر است با:

$$\sigma^2 = \frac{-1}{\sum_{j=1}^q c_j w'_{j,j}(\theta)}.$$

ذکر این نکته جالب است که برای خانواده نمایی یک پارامتری ($k=1$)، با قرار دادن $\theta_0 = \theta^*$ ، MLE برای θ بر اساس n مشاهده برابر با $\hat{\theta}_n$ است.

مثال ۱.۵. یک مدل اتورگرسیو را در نظر بگیرید که در آن بردار تصادفی $\mathbf{X} = (X_1, \dots, X_r)^t$ (بردار $1 \times r$) دارای توزیع $N(\mathbf{0}, \Sigma)$ با تابع چگالی زیر باشد:

$$c(\rho) \exp \left[\sum_{i=1}^r w_i(\rho) t_i(\mathbf{x}) \right],$$

که در آن $\Sigma = (\sigma_{ij}) = (\rho^{|i-j|})$ و $|\rho| < 1$ و همچنین:

$$w_1(\rho) = -(2 - 2\rho^2)^{-1},$$

$$w_2(\rho) = -\rho^2(2 - 2\rho^2)^{-1},$$

$$w_3(\rho) = -\rho(1 - \rho^2)^{-1},$$

$$t_1(\mathbf{x}) = \mathbf{x}^t \mathbf{x},$$

$$t_2(\mathbf{x}) = \sum_{i=2}^{r-1} x_i^2,$$

$$t_3(\mathbf{x}) = \sum_{i=1}^{r-1} x_i x_{i+1}.$$

به راحتی مشاهده می‌شود که

$$E(t_1(\mathbf{X})) = r,$$

$$E(t_2(\mathbf{X})) = r - 2,$$

$$E(t_3(\mathbf{X})) = (r - 1)\rho$$

شرایط داده شده برقرار است و $\hat{\theta}_n$ بهترین برآوردگر مجانباً نرمال برای θ است.

برای مشاهده مثال‌های بیشتر در این زمینه و همچنین کاربردهای این مسئله در توزیع نرمال چندمتغیره به [۸۶] مراجعه شود.

۶ خانواده‌های نمایی و برآوردگرهای بیزی، مجاز و مینیماکس

در بخش ۴ یکی از روش‌های پیدا کردن برآوردگرهای بهینه، یعنی محدود کردن کلاس برآوردگرها و پیدا کردن بهترین برآوردگر در کلاس محدود شده بررسی شد. روش دیگر یافتن برآوردگرهای بهینه، مرتب کردن برآوردگرها طبق یک قاعده خاص است که منجر به برآوردگرهای بیزی، مجاز و مینیماکس می‌گردد.

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از خانواده‌های نمایی با تابع چگالی زیر باشد:

$$f(x|\theta) = h(x) \exp(\theta \cdot t(x) - \Lambda(\theta))$$

که در آن $h(x)$ یک تابع نامنفی و $\theta \in \Theta \subseteq \mathbb{R}^q$ باشد. در اینجا $\theta \cdot t(x)$ به معنی ضرب اسکالر هست. این چگالی نسبت به یک اندازه رادون (غالباً اندازه لبگ برای توزیع‌های پیوسته و اندازه شمارشی برای توزیع‌های گسسته) در نظر گرفته می‌شود. دنباله توزیع پیشینی مزدوج $\{\pi_{\alpha,\beta}\}_{\alpha,\beta}$ برای پارامتر طبیعی θ به صورت تابع چگالی‌هایشان نسبت به اندازه لبگ به صورت زیر

$$\pi_{\alpha,\beta}(\theta) \propto \exp\{\theta \cdot \alpha - \beta \Lambda(\theta)\}$$

خاص $\eta = \theta$ داریم:

$$\pi_{\alpha, \beta}^J(\theta) \propto \exp \{ \theta \cdot \alpha + \beta \Lambda(\theta) \} \sqrt{|I_{\theta}(\theta)|}.$$

در حالتی که توزیع پیشینی جفریز است، توزیع پسینی متعلق به خانواده‌ی JCP با $\alpha = T_n(x)$ و $\beta = n$ است. با استفاده از خاصیت ناوردایی اندازه‌ی جفریز یعنی:

$$\sqrt{|I_{\eta}(\eta)|} = \sqrt{|I_{\theta}(\theta(\eta))|} \left| \frac{d\theta(\eta)}{d\eta} \right|$$

نتیجه گرفته می‌شود:

$$\pi_{\alpha, \beta}^J(\eta) = \left| \frac{d\theta(\eta)}{d\eta} \right| \pi_{\alpha, \beta}^J(\theta(\eta))$$

که نشان‌دهنده‌ی یکسان بودن توزیع‌های $\pi_{\alpha, \beta}^J(\theta)$ و $\pi_{\alpha, \beta}^J(\eta)$ یا به عبارت دیگر خاصیت ناوردایی JCP ها تحت بازپارامتریدن هموار است. این حقیقت نشان می‌دهد که شرایط روی α و β که منجر به توزیع‌های پیشینی سره برای $\pi_{\alpha, \beta}^J(\eta)$ می‌شود، نیاز به انتخاب بازپارامتریدن هموار η ندارد. برای خانواده‌ی نمایی طبیعی درجه‌ی دو یک متغیره بیان‌کننده‌ی این است که JCP های متناظر با پارامتر میانگین $m(\theta) = (X_i | \theta)$ معادل با پیشین‌های مزدوج استاندارد رابطه‌ی (۶) هستند [۲۵]. بنا بر این در یک خانواده‌ی نمایی طبیعی درجه‌ی دو بین پیشین‌های مزدوج استاندارد و JCP ها برای پارامتر میانگین، تمایزی نیست.

مثال ۱.۶. توزیع گوسین معکوس $IG(\mu, 1)$ با تابع چگالی

$$f(x | \mu) = \sqrt{\frac{1}{\pi x^3}} \exp \left\{ -\frac{(x-\mu)^2}{\sqrt{\mu} x} \right\}, \quad x > 0$$

نسبت به اندازه‌ی لبگ روی $\mathbb{R}^+$ را در نظر بگیرید. ماتریس اطلاع فیشر برای μ عبارت است از $I_{\mu}(\mu) = \mu^{-3}$ و JCP به صورت زیر خواهد بود:

$$\pi_{\alpha, \beta}^J(\mu) \propto \mu^{-3} \exp \left\{ \frac{-\alpha}{\sqrt{\mu}} + \frac{\beta}{\mu} \right\} I_{(\cdot, \infty)}(\mu).$$

برای $\alpha = 0$ ، $\pi_{\alpha, \beta}^J$ دارای توزیع $IG(\frac{1}{\beta}, -\beta)$ خواهد بود که در حالت $\beta < 0$ یک توزیع پیشینی سره است [۲۵].

در حالت کلی توزیع پسینی متناظر

$$\pi_{\alpha, \beta}^J(\mu | x) = \pi_{\alpha + S, \beta + n}^J(\mu)$$

تعریف می‌شوند که در آن $\alpha \in \mathbb{R}^q$ و $\beta \in \mathbb{R}$ و منجر به توزیع پسینی

$$\pi_{\alpha, \beta}(\theta | x) = \pi_{\alpha + T_n(x), \beta + n}(\theta)$$

خواهد شد که در آن $T_n(x) = \sum_{i=1}^n t(x_i)$. در این حالت توزیع پیشینی و پسینی از یک نوع و در اصطلاح مزدوج هستند. در [۲۴] شرایط لازم و کافی برای سره بودن این توزیع‌های پیشینی به دست آمده است. با در نظر گرفتن پارامتر $\eta = h(\theta)$ که در آن h یک تبدیل یک به یک به طور پیوسته دو بار مشتق پذیر است (و از این به بعد آن را به طور خلاصه بازپارامتریدن هموار نامیم)، توزیع‌های پیشینی مزدوج زیر پیشنهاد می‌شوند:

$$\pi_{\alpha, \beta}(\eta) \propto \left| \frac{d\theta(\eta)}{d\eta} \right| \exp \{ \theta(\eta) \cdot \alpha + \beta \Lambda(\theta(\eta)) \} \quad (۶)$$

$$\pi_{\alpha, \beta}(\eta) \propto \exp \{ \theta(\eta) \cdot \alpha + \beta \Lambda(\theta(\eta)) \} \quad (۷)$$

که تفاوت آنها در ژاکوبی $|d\theta(\eta)/d\eta|$ هست. در حالت کلی این دو خانواده از توزیع‌های پیشینی متفاوت هستند و در برخی حالات برابرند. به عنوان نمونه در [۱۹] نشان داده می‌شود در یک خانواده‌ی نمایی طبیعی با $t(x_i) = x_i$ که $x_i, \theta \in \mathbb{R}$ با پارامتری کردن توسط میانگین توزیع‌ها یعنی $m(\theta) = E(X_i | \theta)$ خانواده توزیع‌های پیشینی تعریف شده در رابطه‌های (۶) و (۷) معادل‌اند اگر و تنها اگر خانواده‌ی نمایی درجه‌ی دو باشد یعنی واریانس توزیع یک چندجمله‌ای درجه ۲ از میانگین باشد. خانواده‌های نمایی طبیعی درجه‌ی دو شامل توزیع‌های دوجمله‌ای، پواسون، دوجمله‌ای منفی، نرمال، گاما، هذلولوی هستند [۷۵]. بنا بر این برای خانواده مزبور می‌توان از نوشتن ژاکوبی صرف نظر کرد.

در [۲۵] خانواده‌ی جدیدی از پیشین‌های مزدوج برای خانواده‌ی نمایی که نسبت به بازپارامتریدن ناوردا هستند معرفی می‌شود. به ازای هر بازپارامتریدن هموار $\eta(\theta)$ ، پیشینی مزدوج جفریز (JCP^۱) متناظر به صورت زیر تعریف می‌شود:

$$\pi_{\alpha, \beta}^J(\eta) \propto \exp \{ \theta(\eta) \cdot \alpha + \beta \Lambda(\theta(\eta)) \} \sqrt{|I_{\eta}(\eta)|},$$

که در آن $I_{\eta}(\eta)$ ماتریس اطلاع فیشر برای η است. در حالت

^۱ Jeffreys Conjugate Prior

تقلیل می‌یابد که در آن $c(x, n)$ یک تابع از $x = (x_1, \dots, x_n)^t$ و $n = \tau^r$ که $\theta = \tau^r$ و $\nu > 0$ تابعی از n و $T(x)$ آماره بسنده کامل برای θ با توزیع $\Gamma(\nu, \theta)$ هست. مدل اخیر در برآورد $\theta = \tau^r$ ناوردای مقیاسی است، هرگاه به‌ازای هر $a, x_i > 0$ $T(ax) = a^r T(x)$ و $c(ax, n) = a^{r\nu-n} c(x, n)$.

برآوردگرهای خطی مجاز برای $\theta = \tau^r$ تحت زیان آنتروپی در خانواده نمایشی کلی (۸) در رابطه [۷۹] به‌دست آمده است. در این‌جا به‌عنوان نمونه در خانواده نمایشی در رابطه (۸) برآوردگر مجاز و مینیماکس برای $\theta = \tau^r$ که $\theta \in [a, +\infty)$ ($a > 0$) محاسبه می‌گردد.

با در نظر گرفتن برآوردگرهای $\hat{\theta}$ که به‌ازای هر $\theta \geq a$ در شرط $P_{\theta}(\hat{\theta} \geq a) = 1$ صدق می‌کنند و تابع زیان ناوردای مقیاسی $L(\theta, \delta) = (\frac{\delta}{\theta} - 1)^2$ و توزیع پیشینی θ با تابع چگالی زیر

$$\pi_m(\theta) = \frac{1}{m} \left(\frac{1}{\theta}\right) \left(\frac{a}{\theta}\right)^{1/m} I_{[a, +\infty)}(\theta)$$

در [۵۰] ثابت می‌شود برآوردگر بیز برای θ به‌صورت زیر خواهد بود:

$$\delta_m(X) = \frac{T(X)}{\nu + 1 + \frac{1}{m}} \left[1 + \frac{\left(\frac{T(X)}{a}\right)^{\nu+1+\frac{1}{m}} e^{-\frac{T(X)}{a}}}{\int_a^{\infty} \frac{T(X)}{a} t^{\nu+1+\frac{1}{m}} e^{-t} dt} \right]$$

و علاوه بر آن مخاطره بیزی این برآوردگر متناهی است و

$$\lim_{m \rightarrow \infty} r_m(\delta_m) = \frac{1}{\nu+1}.$$

همچنین ثابت می‌شود که:

$$\begin{aligned} \delta(X) &= \lim_{m \rightarrow \infty} \delta_m(X) \\ &= \frac{T(X)}{\nu+1} \left[1 + \frac{\left(\frac{T(X)}{a}\right)^{\nu+1} e^{-\frac{T(X)}{a}}}{\int_a^{\infty} \frac{T(X)}{a} t^{\nu+1} e^{-t} dt} \right] \end{aligned}$$

یک برآوردگر مجاز و مینیماکس برای $\theta = \tau^r$ که $\theta \geq a$ هست. برای مقایسه با حالت $\theta \geq 0$ مشاهده می‌شود که در این حالت برآوردگر مینیماکس برای θ ، $T(X)/(\nu+1)$ هست.

مثال ۳.۶. بر اساس نمونه تصادفی $X_1, \dots, X_n$ از توزیع گوسین معکوس $IG(\infty, \lambda)$ با $\theta = \lambda^{-1}$ ($\tau = \lambda, r = -1$) $\nu = n/2$

هست که در آن $S = \sum_{i=1}^n x_i$. در [۲۵] با ارائه چند مثال با داده‌های واقعی برآوردهای بیزی تحت JCP محاسبه می‌شوند و برای کسب مطالب بیشتر به این مقاله مراجعه کنید. برای یک خانواده نمایشی با تابع چگالی زیر

$$f(x|\theta) = c(\theta) \exp[\theta t(x)]$$

که در آن $\theta \in (\theta_-, \theta_+) \subset \mathbb{R}$ مسئله برآورد $\theta = E(t(X))$ را تحت تابع زیان توان دوم خطاها در نظر بگیرید. شرط لازم و کافی برای مجاز بودن $\frac{t(X)+\gamma\lambda}{1+\lambda}$ آن است که

$$\int_{\theta_-}^{\theta_+} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} d\theta = \int_{\theta_-}^{\theta_+} \frac{e^{-\gamma\lambda\theta}}{[c(\theta)]^\lambda} d\theta = \infty$$

که در آن $\theta \in (\theta_-, \theta_+)$. [۸۹]، ص ۲۶۵. همچنین در حالتی که $\theta_- = -\infty$ و $\theta_+ = +\infty$ در برآورد $\theta = E(t(X))$ و تحت تابع زیان $L(\theta, a) = \frac{1}{\text{var}(T)}(a - \theta)^2$ ، با قرار دادن $\lambda = 0$ در قسمت بالا به‌راحتی نتیجه گرفته می‌شود که $t(X)$ یک برآوردگر مجاز و در پی آن از این که مخاطره این برآوردگر ثابت و تابع زیان اکیداً محدب است این نتیجه حاصل می‌شود که $t(X)$ یک برآوردگر مینیماکس یکتا نیز هست ([۸۹]، ص ۲۶۶).

مثال ۲.۶. اگر $X_1, \dots, X_n$ نمونه تصادفی از توزیع $P(\theta)$ باشد، به‌راحتی نشان داده می‌شود که $\bar{X}$ یک برآوردگر مجاز و مینیماکس یکتا تحت تابع زیان $L(\theta, a) = \frac{1}{\theta}(a - \theta)^2$ است ([۸۹]، ص ۲۶۷).

در [۵۰] در یک زیرکلاس از خانواده‌های نمایشی با پارامتر مقیاس کران‌دار و تحت تابع زیان ناوردای مقیاسی برای توانی از پارامتر برآوردگر مجاز و مینیماکس پیدا می‌شود.

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی از یک توزیع با چگالی $\frac{1}{\tau} g(\frac{x}{\tau})$ باشد، که g یک تابع معلوم و τ پارامتر مقیاس مجهول باشد. تابع چگالی توأم نمونه تصادفی $X_1, \dots, X_n$ به‌صورت زیر تعریف می‌شود:

$$f(x|\tau) = \frac{1}{\tau^n} f\left(\frac{x}{\tau}\right)$$

تعریف می‌شود. در برخی مواقع این مدل به‌صورت زیر

$$f(x|\theta) = c(x, n) \theta^{-\nu} e^{-T(x)/\theta} \quad (۸)$$

که $c(x, n) = e^{h(x)+k_1}$ بنا بر این

$$-2a(\mathbf{X})b(\eta) = 2a(\mathbf{X})/\theta \sim \Gamma(k/2, 2)$$

و یا به‌طور معادل $a(\mathbf{X}) \sim \Gamma(k/2, \theta)$

بنا بر این اگر شرط (۱۰) برقرار باشد، خانواده‌ی یک پارامتری (۹) به‌شکل خانواده‌ی نمایی با پارامتر مقیاس (۸) با $\nu = k/2$ برای $T(x) = a(x)$ و $\theta = -1/b(\eta)$ خواهد بود. بنا بر این برای برآورد پارامتر $\theta = -1/b(\eta) \in [a, +\infty)$ می‌توان از ساختار ذکر شده در قسمت قبل استفاده کرد و برآوردگر مجاز و مینیماکس برای $\theta = -1/b(\eta)$ تحت شرایط قبلی (تابع زیان ناوردی مقیاسی داده شده و فضای پارامتر بریده شده از چپ) به‌صورت ذیل خواهد بود [۵۰]:

$$\delta_m(\mathbf{X}) = \frac{a(\mathbf{X})}{\frac{k}{2} + 1} \left[1 + \frac{\left(\frac{a(\mathbf{X})}{a}\right)^{\frac{k}{2}+1} e^{-\frac{a(\mathbf{X})}{a}}}{\int_0^{\frac{a(\mathbf{X})}{a}} t^{\frac{k}{2}+1} e^{-t} dt} \right].$$

مثال ۴.۶. بر اساس نمونه‌ی تصادفی $X_1, \dots, X_n$ از توزیع رابلی $T(x) = \nu = n$, $(\tau = \beta, r = 2)$ $\theta = \beta^2$ با $R(\beta)$ $c(\beta) = -2n \ln \beta$ $b(\beta) = -1/\beta^2$ $a(x) = \frac{1}{2} \sum_{i=1}^n x_i^2$ و $c(x, n) = \prod_{i=1}^n x_i$ تابع چگالی توأم نمونه‌ی تصادفی برای $i = 1, \dots, n$ به‌صورت ذیل خواهد بود:

$$f(x|\beta) = \left(\prod_{i=1}^n x_i\right) \beta^{-2n} e^{-\frac{1}{2\beta^2} \sum_{i=1}^n x_i^2}, x_i > 0.$$

با توجه به برقراری شرط (۱۰) یعنی $k = 2c'(\eta)b(\eta)/b'(\eta) = 2n$ داریم:

$$-2a(\mathbf{X})b(\beta) = \frac{\sum_{i=1}^n X_i^2}{\beta^2} \sim \Gamma(n, 2)$$

و در این حالت برآوردگر مجاز و مینیماکس برای $\theta = \beta^2 \in [a, +\infty)$ برابر است با:

$$\delta(\mathbf{X}) = \frac{\sum_{i=1}^n X_i^2}{2n + 2} \left[1 + \frac{\left(\frac{1}{2a} \sum_{i=1}^n X_i^2\right)^{n+1} e^{-\frac{1}{2a} \sum_{i=1}^n X_i^2}}{\int_0^{\frac{1}{2a} \sum_{i=1}^n X_i^2} t^{n+1} e^{-t} dt} \right].$$

^{۱۱} Uniformly Most Powerful Test

$T(x) = \frac{1}{2} \sum_{i=1}^n \frac{1}{x_i}$ و $c(x, n) = \left(\prod_{i=1}^n 2\pi x_i^2\right)^{-1/2}$ تابع چگالی توأم نمونه‌ی تصادفی برای $i = 1, \dots, n$ به‌صورت ذیل خواهد بود:

$$f(x|\lambda) = \left(\prod_{i=1}^n 2\pi x_i^2\right)^{-1/2} \lambda^{-n/2} e^{-\frac{\lambda}{2} \sum_{i=1}^n \frac{1}{x_i}}, x_i > 0.$$

در این حالت برآوردگر مجاز و مینیماکس برای $\theta = \lambda^{-1} \in [a, +\infty)$ برابر است با:

$$\delta(\mathbf{X}) = \frac{\sum_{i=1}^n \frac{1}{X_i}}{n + 2} \left[1 + \frac{\left(\frac{1}{2a} \sum_{i=1}^n \frac{1}{X_i}\right)^{\frac{n}{2}+1} e^{-\frac{1}{2a} \sum_{i=1}^n \frac{1}{X_i}}}{\int_0^{\frac{1}{2a} \sum_{i=1}^n \frac{1}{X_i}} t^{\frac{n}{2}+1} e^{-t} dt} \right].$$

در ادامه نمونه‌ی تصادفی $X_1, \dots, X_n$ از یک خانواده‌ی نمایی با تابع چگالی توأم نمونه به‌صورت زیر را در نظر بگیرید:

$$f(x|\eta) = e^{a(x)b(\eta)+c(\eta)+h(x)}. \quad (9)$$

در [۸۲] ثابت شد $-2a(\mathbf{X})b(\eta)$ دارای توزیع $\Gamma(\frac{k}{2}, 2)$ است اگر و تنها اگر داشته باشیم:

$$2c'(\eta)b(\eta) = kb'(\eta) \quad (10)$$

که در آن k مثبت و به η بستگی ندارد. در حالت خاص که k یک عدد طبیعی باشد، $-2a(\mathbf{X})b(\eta)$ دارای توزیع χ_k^2 خواهد بود. خانواده‌های نمایی (۹) که در شرط (۱۰) برای $k \in \mathbb{N}$ صدق کنند را توزیع‌های تبدیل‌یافته به‌خی‌دو نامند. توجه کنید با توجه به ساختار تابع گاما a و b مختلف‌العلامه هستند که فرض می‌شود a مثبت و b منفی است. از حل معادله‌ی دیفرانسیل (۱۰)، داریم، $c(\eta) = \frac{k}{2} \ln |b(\eta)| + k_1$ و در پی آن

$$e^{c(\eta)} = \left(-\frac{1}{b(\eta)}\right)^{-\frac{k}{2}} e^{k_1} = \theta^{-\frac{k}{2}} e^{k_1}$$

که در آن $\theta = -1/b(\eta) > 0$. از این رو رابطه (۹) به‌صورت زیر ساده می‌شود:

$$f(x|\eta) = e^{h(x)} e^{c(\eta)} e^{a(x)[-1/b(\eta)]} = c(x, n) \theta^{-\frac{k}{2}} e^{-\frac{a(x)}{\theta}}$$

۷ خانواده‌های نمایی و آزمون‌های فرضیه بهینه

قضیه ۱.۷. اگر $f(x|\theta) = h(x)c(\theta)\exp[\theta x]$ ، به‌ازای هر آزمون دلخواه φ و $\theta_1 \in \Theta$ و $\theta_2 \in \Theta$ که $\theta_1 < \theta_2$ یک آزمون دوطرفه

$$\varphi'(x) = \begin{cases} 1 & x < x_1 \text{ or } x > x_2 \\ \gamma_i & x = x_i, i = 1, 2 \\ 0 & x_1 < x < x_2 \end{cases}$$

وجود دارد که

$$E_{\theta_1}\varphi'(X) = E_{\theta_1}\varphi(X),$$

$$E_{\theta_2}\varphi'(X) = E_{\theta_2}\varphi(X).$$

علاوه بر آن به‌ازای هر آزمون دوطرفه

$$\varphi' E_{\theta}\varphi'(X) - E_{\theta}\varphi(X) \begin{cases} \leq 0 & \theta_1 < \theta < \theta_2 \\ \geq 0 & \theta < \theta_1, \theta > \theta_2. \end{cases}$$

همچنین φ' یک $UMPU$ در اندازه α است به‌طوری که x_1, x_2, γ_1 و γ_2 طوری انتخاب می‌شوند که $E_{\theta_1}\varphi'(X) = E_{\theta_2}\varphi'(X) = \alpha$ ([۳۰]، ص ۲۱۷).

در حالتی که با آزمون‌های دوطرفه $H_0: \theta = \theta_0$ در مقابل $H_1: \theta \neq \theta_0$ روبرو هستیم قضیه زیر می‌تواند راه‌گشا باشد:

قضیه ۲.۷. اگر $f(x|\theta) = h(x)c(\theta)\exp[\theta x]$ ، به‌ازای هر آزمون دلخواه φ و $\theta_0 \in \Theta$ ، یک آزمون دوطرفه

$$\varphi'(x) = \begin{cases} 1 & x < x_1 \text{ or } x > x_2 \\ \gamma_i & x = x_i, i = 1, 2 \\ 0 & x_1 < x < x_2 \end{cases}$$

وجود دارد که $E_{\theta_0}\varphi'(X) = E_{\theta_0}\varphi(X)$ و

$$\frac{d}{d\theta} E_{\theta}\varphi'(X) |_{\theta=\theta_0} = \frac{d}{d\theta} E_{\theta}\varphi(X) |_{\theta=\theta_0}.$$

علاوه بر آن به‌ازای هر آزمون دوطرفه φ' و به‌ازای هر $\theta \in \Theta$

$$E_{\theta}\varphi'(X) \geq E_{\theta}\varphi(X).$$

آزمون‌های فرضیه آماری یکی از شاخه‌های مورد توجه در استنباط آماری به حساب می‌آید و آزمون‌های بهینه مثل پرتوان‌ترین آزمون یکنواخت ($UMPT^{11}$)، پرتوان‌ترین آزمون یکنواخت نااریب ($UMPB^{12}$) و آزمون نسبت درست‌نمایی (LRT^{13}) از معروف‌ترین آزمون‌های بهینه هستند.

در بیکل و دوکسوم ([۸]، ص ۱۹۹) آزمون‌های UMP یک‌طرفه برای خانواده‌های نمایی مورد بررسی قرار گرفته‌اند. برای این منظور فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی با تابع چگالی احتمال یا تابع جرم احتمال از یک خانواده نمایی یک‌پارامتری زیر باشند:

$$f(x|\theta) = h(x)c(\theta)\exp[\omega(\theta)t(x)]$$

که $\omega(\theta)$ اکیداً صعودی است. این خانواده در آماره بسنده کامل $T(X) = \sum_{i=1}^n t(X_i)$ دارای خاصیت MLR^{14} است. در این صورت تابع آزمون $UMPT$ برای آزمون فرضیه آماری $H_0: \theta \leq \theta_0$ در مقابل $H_1: \theta > \theta_0$ برابر است با

$$\varphi(x) = \begin{cases} 1 & T(x) > k \\ \gamma & T(x) = k \\ 0 & T(x) < k \end{cases}$$

$$\alpha = P_{\theta_0}(T > k) + \gamma P_{\theta_0}(T = k)$$

فرض کنید بازپارامتریدن به‌صورت زیر باشد:

$$f(x|\theta) = h(x)c(\theta)\exp[\tilde{\omega}(\theta)\tilde{t}(x)]$$

که در آن $\tilde{\omega}(\theta)$ اکیداً نزولی است. در این صورت مطابق قبل قرار می‌دهیم $\omega(\theta) = -\tilde{\omega}(\theta)$ و $t(x) = -\tilde{t}(x)$

در حالتی که با آزمون‌های دوطرفه $H_0: \theta_1 \leq \theta \leq \theta_2$ در مقابل $H_1: \theta < \theta_1 \text{ or } \theta > \theta_2$ روبرو هستیم قضیه زیر می‌تواند راه‌گشا باشد:

¹¹ Uniformly Most Powerful Unbiased Test

¹² Likelihood Ratio Test

¹³ Monotone Likelihood Ratio

با توجه به ارتباط فاصله‌ی اطمینان و آزمون‌های فرضیه‌ی آماری می‌توان با استفاده از توابع آزمون بالا فواصل اطمینان مناسب برای پارامترهای جامعه پیدا کرد. به عنوان نمونه خانواده‌ی نمایی یک پارامتری زیر در نظر بگیرد:

$$f(x|\theta) = h(x)c(\theta) \exp[\omega(\theta)t(x)],$$

که در آن θ پارامتر با مقادیر حقیقی و تابع $\omega(\theta)$ اکیداً صعودی است. ناحیه پذیرش UMPT در اندازه α برای آزمون فرضیه‌ی آماری $H_0: \theta = \theta_0$ در مقابل $H_1: \theta > \theta_0$ با تابع آزمون زیر

$$\varphi(x) = \begin{cases} 1 & T(x) > k \\ \gamma & T(x) = k \\ 0 & T(x) < k \end{cases}$$

برابر است با:

$$A(\theta_0) = \{x : T(x) \leq k(\theta_0)\}$$

که در آن $k(\theta_0) = k$ به راحتی ثابت می‌شود $k(\theta)$ اکیداً صعودی است. بنا بر این $C(X) = [-\bar{\theta}(X), \infty)$ یک فاصله‌ی اطمینان یک‌طرفه برای θ در سطح معنی‌داری $1 - \alpha$ است که در آن $\bar{\theta}(X) = \inf \{\theta : T(x) \leq k(\theta)\}$ به طور مشابه برای آزمون فرضیه‌ی آماری $H_0: \theta = \theta_0$ در مقابل $H_1: \theta < \theta_0$ ، فاصله‌ی اطمینان یک‌طرفه برای θ در سطح معنی‌داری $1 - \alpha$ به صورت $C(X) = (-\infty, \bar{\theta}(X)]$ است که در آن $\bar{\theta}(X) = \sup \{\theta : T(x) \geq k(\theta)\}$ به طور مشابه متناظر با آزمون‌های دوطرفه نیز می‌توان فاصله‌ی اطمینان پیدا کرد که مطالعه جزئیات و مثال‌های بیشتر به لیمان و رومانو ([۶۳]، صفحه‌های ۱۶۴-۱۶۱) و شائو ([۸۹]، صفحه‌های ۴۸۰-۴۷۷) مراجعه کرد. علاوه بر آن در مقالاتی نظیر [۲۳] نیز می‌توان راجع به فاصله‌های اطمینان در خانواده‌ی نمایی مطالب ارزنده‌ای را پیدا کرد. در [۲۳] ثابت می‌شود که اگر $X_1, \dots, X_n$ یک نمونه تصادفی با تابع چگالی احتمال یا تابع جرم احتمال از یک خانواده‌ی نمایی

$$f(x|\theta) = c(\theta) \exp(\theta \cdot x)$$

باشند و F یک اندازه احتمال روی مجموعه محدب $\Theta \subseteq \mathbb{R}^q$

همچنین φ' یک UMPU در اندازه α است به طوری که x_1, x_2, γ_1 و γ_2 طوری انتخاب می‌شوند که $E_{\theta_0}[\varphi'(X)] = \alpha$ و $E_{\theta_0}(\varphi'(X)) = \alpha E_{\theta_0}(X)$ ([۳۰]، ص ۲۲۰).

در فرگوسن ([۳۰]، صفحات ۲۳۳-۲۲۶) می‌توان مطالب بیشتری در مورد خانواده‌های نمایی چندپارامتری مشاهده کرد. به عنوان نمونه برای خانواده‌ی نمایی چندپارامتری زیر:

$$f(x|\theta, \omega) = \exp[\theta T(x) + \omega U(x) - \Lambda(\theta, \omega)]$$

از این که (T, U) دارای تابع چگالی $\exp[\theta t + \omega u - \Lambda(\theta, \omega)]$ است، نتیجه گرفته می‌شود که $T|U=u$ دارای تابع چگالی $\exp[\theta t]$ هست که متعلق به یک خانواده‌ی نمایی یک پارامتری است. بنا بر این UMPU در اندازه α در برای آزمون فرضیه‌ی آماری $H_0: \theta \leq \theta_0$ در مقابل $H_1: \theta > \theta_0$ برابر است با:

$$\varphi(T, U) = \begin{cases} 1 & T > k(U) \\ \gamma(U) & T = k(U) \\ 0 & T < k(U) \end{cases}$$

که در آن $k(u)$ و $\gamma(u)$ به گونه‌ای محاسبه می‌شوند که به ازای هر u ، $E_{\theta_0}[\varphi(T, U)|U=u] = \alpha$ برای آزمون‌های دوطرفه $H_0: \theta_1 \leq \theta \leq \theta_2$ در مقابل $H_1: \theta < \theta_1 \text{ or } \theta > \theta_2$ ، UMPU در اندازه α برابر است با:

$$\varphi(T, U) = \begin{cases} 1 & T < k_1(U) \text{ or } T > k_2(U) \\ \gamma_i(U) & T = k_i(U), i = 1, 2 \\ 0 & k_1(U) < T < k_2(U) \end{cases}$$

که در آن $k_i(u)$ و $\gamma_i(u)$ به گونه‌ای محاسبه می‌شوند که به ازای هر u داریم:

$$E_{\theta_1}[\varphi(T, U)|U=u] = E_{\theta_2}[\varphi(T, U)|U=u] = \alpha$$

$$(E_{\theta_0}[\varphi(T, U)|U=u] = \alpha E_{\theta_0}[T|U=u])$$

برای مشاهده مثال در این زمینه به مرجع‌های لیمان و رومانو ([۶۳]، صفحه‌های ۱۲۴-۱۱۹) و شائو ([۸۹]، صفحه‌های ۴۰۶-۴۱۷) مراجعه شود.

است. به راحتی می توان نشان داد $h'(\theta) = \text{Var}_{\theta}(X) > 0$. در پی آن h یک تابع اکیداً صعودی و لذا H یکتاست. برای اثبات این که برای $\hat{\theta}_i = H(x_i)$ $i = 1, \dots, q$ تابع $L(x, \theta)$ را ماکسیم می کند از این که $\frac{\partial^2}{\partial \theta_i^2} L(x, \theta) = -\text{Var}_{\theta}(X) < 0$ و $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(x, \theta) = 0$ برای $i \neq j$ نتیجه می گیریم که ماتریس مشتقات جزئی مرتبه دوم یک ماتریس معین نامنفی باشد و از این رو نقطه زیر

$$(\hat{\theta}_1, \dots, \hat{\theta}_q) = (H(x_1), \dots, H(x_q))$$

یک MLE برای $(\theta_1, \dots, \theta_q)$ خواهد بود. بنا بر این:

$$\lambda(x) = \frac{L(x, \hat{\theta}_1, \dots, \hat{\theta}_q)}{L(x, \circ)} = \prod_{i=1}^q \left(\frac{B(H(x_i))}{B(\circ)} \right) \exp \left(\sum_{i=1}^q x_i H(x_i) \right)$$

و در پی آن LRT به صورت زیر خواهد بود:

$$\varphi(x) = \begin{cases} 1 & \prod_{i=1}^q B(H(x_i)) \exp \left(\sum_{i=1}^q x_i H(x_i) \right) > k \\ 0 & o.w. \end{cases} \quad (11)$$

قضیه های زیر می تواند برای اثبات مجاز بودن آزمون ها مورد استفاده قرار گیرند.

قضیه ۳.۷. اگر $f(x|\theta) = h(x)c(\theta)\exp[\theta x]$ و تابع زیان برای $\theta_1 < \theta_2$ در نامساوری زیر صدق کند:

$$L(\theta, a_1) - L(\theta, a_2) \begin{cases} \geq 0 & \theta_1 < \theta < \theta_2 \\ \leq 0 & \theta < \theta_1 \text{ or } \theta > \theta_2 \end{cases}$$

کلاس آزمون های دوطرفه

$$\varphi'(x) = \begin{cases} 1 & x < x_1 \text{ or } x > x_2 \\ \gamma_i & x = x_i, i = 1, 2 \\ 0 & x_1 < x < x_2 \end{cases}$$

اساساً کامل است. اگر علاوه بر شرایط بالا $\theta'_1 < \theta'_2 < \theta'_3$ وجود داشته باشند که

$$L(\theta'_i, a_1) - L(\theta'_i, a_2) \begin{cases} < 0 & i = 1, 3 \\ > 0 & i = 2 \end{cases}$$

هر آزمون دوطرفه

$$\varphi'(x) = \begin{cases} 1 & x < x_1 \text{ or } x > x_2 \\ \gamma_i & x = x_i, i = 1, 2 \\ 0 & x_1 < x < x_2 \end{cases}$$

باشد، به ازای هر $\varepsilon > 0$ و $n \in \mathbb{N}$

$$G_n = \left\{ \lambda \in \Theta : \frac{\int_{\Theta} \prod_{i=1}^n f_{\theta}(X_i) dF(\theta)}{\prod_{i=1}^n f_{\lambda}(X_i)} < \varepsilon \right\}$$

یک دنباله از فاصله های اطمینان محدب در سطح $1 - \varepsilon^{-1}$ است و به ازای هر $\theta_0 \in \Theta$ $\lim_{n \rightarrow \infty} \rho(G_n, \theta_0) = 0$ که در آن $\rho(G, x) = \sup \{|y - x| : y \in G\}$

در [۲۸] ثابت می شود آزمون نسبت درست نمایی (LRT)

برای فرضیه ساده در مقابل مرکب برای پارامترهای یک خانواده نمایی چند پارامتری مجاز است. برای این منظور فرض کنید X_i ها $(i = 1, \dots, q)$ متغیرهای تصادفی مستقل از توزیع زیر باشند:

$$f(x|\theta) = A(x)B(\theta_i) \exp[\theta_i x], \quad x \in X \subseteq \mathbb{R}, \quad \theta_i \in \Theta \subseteq \mathbb{R}$$

تابع چگالی توأم $(X_1, \dots, X_q)$ به صورت قضیه بالا که در آن $A(x) = \prod_{i=1}^q A(x_i)$ ، $B(\theta) = \prod_{i=1}^q B(\theta_i)$ باشند، اگر θ_{i_0} ها نقاط درونی Θ_i باشند برای $i = 1, \dots, q$ می خواهیم LRT برای آزمون فرضیه آماری $H_0: \theta = \theta_{1_0}, \dots, \theta_{q_0}$ در مقابل $H_1: \theta \neq \theta_{1_0}, \dots, \theta_{q_0}$ پیدا کنیم. بدون این که به کلیت مسئله لطمه ای وارد شود با تبدیل مناسب می توان فرض کرد به ازای هر $\theta_{i_0} = 0$ ، $i = 1, \dots, q$ برای آزمون فرضیه آماری به صورت زیر خواهد بود:

$$\varphi(x) = \begin{cases} 1 & \lambda(x) > k \\ 0 & o.w. \end{cases}$$

و

$$\lambda(x) = \frac{L(x, \hat{\theta}_1, \dots, \hat{\theta}_q)}{L(x, \circ)}$$

که در آن $\hat{\theta}_i$ ها MLE برای θ_i ها هستند که در آن $i = 1, \dots, q$ با توجه به این که

$$\ln L(x, \theta) = \sum_{i=1}^q \ln B(\theta_i) + \sum_{i=1}^q \ln A(x_i) + \sum_{i=1}^q \theta_i x_i$$

داریم:

$$\frac{\partial}{\partial \theta_i} \ln L(x, \theta) = \frac{B'(\theta_i)}{B(\theta_i)} + x_i = 0 \Leftrightarrow \theta_i = \theta_i^* = H(x_i)$$

که در آن $H(\cdot)$ وارون تابع

$$h(\theta) = -\frac{B'(\theta_i)}{B(\theta_i)} = E_{\theta}(X)$$

مجاز است) ([۳۰], ص ۲۱۲).

قضیه ۴.۷. فرض کنید X دارای توزیع

$$f(x|\theta) = A(x)B(\theta) \exp\left[\sum_{i=1}^q \theta_i x_i\right]$$

باشد که در آن $\theta \in \Theta \subseteq \mathbb{R}^q$ شرط لازم و کافی برای این که تابع آزمون φ برای آزمون فرضیه آماری $H_0: \theta_1 = \dots = \theta_p$ در مقابل نفی آن که $p \leq q$ مجاز باشد آن است که

$$\varphi(x) = \begin{cases} 1 & S_p \in A(t_P) \\ 0 & o.w. \end{cases}$$

که در آن $S_p = (x_1, \dots, x_q)$ و $t_P = (x_{p+1}, \dots, x_q)$ به ازای هر t_P یک مجموعه محدب بسته در $\mathbb{R}^p$ است [۲۶].

در [۲۸] به کمک قضیه ۴.۷ نشان داده می شود که برای آزمون فرضیه آماری $H_0: \theta = \theta_{10}, \dots, \theta_{q0}$ در مقابل $H_1: \theta \neq \theta_{10}, \dots, \theta_{q0}$ که با استفاده از رابطه (۱۱) محاسبه شد یک آزمون مجاز نیز هست.

در پایان در مورد آزمون نیکویی برازش در خانواده نمایی به طور خلاصه مطالبی بیان می شود [۳۶]. اگر $X_1, \dots, X_n$ نمونه تصادفی با تابع توزیع F_0 باشد، می خواهیم آزمون فرضیه آماری مرکب $F_0 \in F = \{F_\theta(x) : \theta \in \Theta \subseteq \mathbb{R}^q\}$ را انجام دهیم. فرض کنید مقدار واقعی پارامتر θ_0 باشد، یعنی تحت فرضیه H_0 ، تابع توزیع واقعی $F_0(x) = F_{\theta_0}(x)$ است. روش های متعددی برای بررسی این آزمون به کار گرفته شده است. در یکی از این روش ها می توان آماره آزمون را بر حسب تابع توزیع تجربی و توابع چندک نوشت و در پی آن با استفاده از خواص مجانبی، توزیع های حدی را پیدا کرد. اغلب پارامتر مجهول تحت فرضیه H_0 ، توسط دنباله ای از برآوردگرها مثل $\hat{\theta}_n$ برآورد می شود. در این حالت به جای $F_\theta(x)$ از تابع توزیع برآورد شده یعنی $F_{\hat{\theta}_n}(x)$ استفاده می شود. از دیگر روش های پیشنهادی تصادفی کردن و باز نمونه گیری است. در مقاله [۳۶] روش جدیدی برای حالتی که $F_\theta(x)$ به ازای هر $\theta \in \Theta$ در x پیوسته است پیشنهاد شده که در آن از متغیرهای

تصادفی مستقل $U_1 = F_{\theta_0}(X_1), \dots, U_n = F_{\theta_0}(X_n)$ که دارای توزیع $U(0, 1)$ هستند، استفاده می شود. با جایگزینی $\hat{\theta}_n$ به جای پارامتر مجهول θ_0 ، متغیرهای تصادفی $\hat{U}_1 = F_{\hat{\theta}_n}(X_1), \dots, \hat{U}_n = F_{\hat{\theta}_n}(X_n)$ و در پی آن آماره های ترتیبی $U_{(1)}, \dots, U_{(n)}$ و $\hat{U}_{(1)}, \dots, \hat{U}_{(n)}$ که به ترتیب برابر $U_{(1)}, \dots, U_{(n)}$ و $\hat{U}_{(1)}, \dots, \hat{U}_{(n)}$ هستند را در نظر می گیریم. به راحتی ثابت می شود اگر $k(n), n - k(n) \rightarrow \infty$ آن گاه

$$\sqrt{n} \frac{U_{(k(n))} - \frac{k(n)}{n}}{\sqrt{\frac{k(n)}{n} \left(1 - \frac{k(n)}{n}\right)}} \rightarrow 0$$

به طور مجانبی دارای توزیع نرمال استاندارد است.

می توان توزیع حدی

$$Z_n = \sqrt{n} \max_{1 \leq k \leq n} \left| \frac{U_{(k(n))} - \frac{k(n)}{n}}{\sqrt{\frac{k(n)}{n} (n - k(n))}} \right|$$

$$\hat{Z}_n = \sqrt{n} \max_{1 \leq k \leq n} \left| \frac{\hat{U}_{(k(n))} - \frac{k(n)}{n}}{\sqrt{\frac{k(n)}{n} (n - k(n))}} \right|$$

را بررسی کرد. فرض کنید شرایط نظم زیر برقرار باشد:

۱. تابع چگالی متعلق به خانواده نمایی q -پارامتری (۱) باشد، یعنی

$$f(x|\theta) = h(x)c(\theta) \exp\left[\sum_{i=1}^q w_i(\theta)t_i(x)\right], \quad x \in D$$

که در آن D به پارامتر مجهول θ بستگی ندارد.

۲. مشتقات جزئی $c(\theta)$ و $w_i(\theta)$ ها نسبت به θ_j که در آن $j = 1, \dots, q$ تا مرتبه سوم وجود داشته باشند و در x و یک همسایگی حول θ_0 به طور یکنواخت کران دار باشند.

$$\max_{1 \leq i \leq q} \int_{-\infty}^{+\infty} (t_i(x))^4 f(x|\theta_0) dx < \infty.$$

۳. $\sqrt{n} \|\hat{\theta}_n - \theta_0\| = O_P(1)$ که در آن $\|\cdot\|$ نرم ماکسیمم در $\mathbb{R}^q$ است.

در [۳۶] نشان داده می شود Z_n و $\hat{Z}_n$ دارای توزیع مجانبی یکسانی هستند و همچنین تحت فرضیه H_0 و برقراری شرایط نظم بالا به ازای هر x داریم:

$$\lim_{n \rightarrow \infty} P\{a(\log(n)) \hat{Z}_n \leq x + b(\log(n))\} = \exp(-\lambda e^{-x})$$

که در آن

به طوری که

$$\liminf_{n \rightarrow \infty} P \left\{ \frac{1}{\sqrt{n}} \hat{Z}_n \geq \delta \right\} \geq \varepsilon$$

و این حقیقت سازگاری مجانبی آزمون داده شده را اثبات می کند.

$$a(x) = \sqrt{2 \log(x)},$$

$$b(x) = 2 \log(x) + \frac{1}{2} \log(\log(x)) - \frac{1}{2} \log(\pi).$$

با توجه به این حقیقت فرضیه H_0 را برای مقادیر بزرگ $a(\log(n)) \hat{Z}_n - b(\log(n))$ رد می کنیم. اکنون فرض کنید تحت فرضیه مقابل، x و $\delta > 0$ وجود داشته باشد به گونه ای که

$$\inf_{\theta \in \Theta} |F_{\theta}(x_0) - F_{\theta}(x_*)| > \delta$$

که رابطه اخیر برای n بزرگ و $1 + [nF_{\theta}(x_*)]$ منجر به عبارت زیر می شود:

$$P \left\{ \left| \hat{U}_{(k(n))} - U_{(k(n))} \right| > \frac{1}{2} \delta \right\} > 1 - \delta$$

بنا بر این به ازای هر $\varepsilon \in (0, 1)$ ، $\delta = \delta(\varepsilon) > 0$ وجود دارد

۸ نتیجه گیری

در این مقاله مروری، تاریخچه ای از کارهای انجام شده راجع به خانواده نمای که توسط محققان در شاخه های مختلف آمار انجام شده ارائه شد و مهم ترین کاربردهایی از این خانواده که در استنباط آماری از آنها استفاده می شود به تفصیل شرح داده شد. علاقه مندان می توانند از این مقاله به عنوان شروعی برای تحقیقات مرتبط با خانواده نمای بهره گیرند و همچنین کاربردهای این رده مهم در شاخه های دیگر آمار را بررسی کنند.

مراجع

- [1] Abu-Dayyeh, W. and Madan, K. C. (2005). The admissibility of the LRT for the exponential family model, *Applied Mathematics and Computation*, **160**, 303-308.
- [2] Abughalous, M. M. and Bansal, N. K. (1995). On selecting the best natural exponential families with quadratic variance function, *Statistics and Probability Letters*, **25**(4), 341-349.
- [3] Altun, Y., Smola, A. J. and Hofmann, T. (2004). *Exponential Families for Conditional Random Fields*, In Uncertainty in Artificial Intelligence (UAI), 2-9, AUAI Press.
- [4] Atienza, N., Garcia-Heras, J., Munoz-Pichardob, J. M. and Villa, R. (2007). On the consistency of MLE in finite mixture models of exponential Families, *Journal of Statistical Planning and Inference*, **137**, 496-505.
- [5] Bahadur, R. R. (1958). Examples of Inconsistency of Maximum Likelihood Estimators, *Sankhya*, **20**, 207-210.
- [6] Balkema, A. A., Kluppelberg, C. and Resnick, S. I. (2003). Domains of attraction for exponential families, *Stochastic Processes and their Applications*, **107**, 83-103.
- [7] Banerjee, A. (2007). *An Analysis of Logistic Models: Exponential Family Connections and Online Performance*, Siam Data Mining.

- [8] Bar-Lev, S. K. and Bshouty, D. (1992). Natural exponential families and self-decomposability, *Statistics and Probability Letters*, **13**, 147-152.
- [9] Bar-Lev, S. K. and Bshouty, D. (2008). Exponential families are not preserved by the formation of order statistics, *Statistics and Probability Letters*, **78**, 2787-2792.
- [10] Barndorff-Nielsen, O. (2014). *Information and Exponential Families in Statistical Theory*, Wiley, New York.
- [11] Barranco-Chamorro, I. and Moreno-Rebollo, J. L. (2008). Some measures of robustness for unbiased estimators in one-parameter natural exponential families with quadratic variance function, *Journal of Mathematical Analysis and Applications*, **341**, 346-356.
- [12] Berk, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models, *Annals of Mathematics*, **4**, 193-204.
- [13] Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, CA.
- [14] Brigo, D. (2000). On SDEs with marginal laws evolving in finite-dimensional exponential families, *Statistics and Probability Letters*, **49**, 127-134.
- [15] Brown, L. D. (1986). Fundamentals of statistical exponential families with applications in statistical decision theory, *Annals of the Institute of Statistical Mathematics*, Hayward, California.
- [16] Canu, S. and Smola, A. J. (2006). Kernel methods and the exponential family, *Neurocomputing*, **69**(7-9), 714-720.
- [17] Casella, G. and Berger, R. L. (2002). *Statistical Inference, Australia, Pacific Grove, CA*, Thomson Learning.
- [18] Chakrabarti, A. and Ghosh, J. K. (2006). A generalization of BIC for the general exponential family, *Journal of Statistical Planning and Inference*, **136**, 2847-2872.
- [19] Chou, J. P. (1988). An identity for multidimensional continuous exponential families and its applications, *Journal of Multivariate Analysis*, **24**, 129-142.
- [20] Clark, D. R. and Thayer, C. A. (2004). *A Primer on the Exponential Family of Distributions*, CAS 2004, discussion papers 117-148, Casualty Actuarial Society, Arlington, Virginia.
- [21] Clarke, B. and Ghosal, S. (2010). Reference priors for exponential families with increasing dimension, *Electronic Journal of Statistics*, **4**, 737-780.

- [22] Cohen, A. and Sackrowitz, H. B. (1987). Admissibility of goodness of fit tests for discrete exponential families, *Statistics and Probability Letters*, **5**, 1-3, 1987.
- [23] Collins, M., Dasgupta, S. and Schapire, R. (2002). *A Generalization of Principal Component Analysis to the Exponential Family*, Advances in Neural Information Processing Systems, **14**.
- [24] Consonni, G. and Veronese, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *Journal of the American Statistical Association*, **87**, 1123-1127.
- [25] Consonni, G. and Veronese, P. and Gutierrez-Pena. (2004). Reference priors for exponential families with simple quadratic variance function, *Journal of Multivariate Analysis*, **88**, 335-364.
- [26] Coolen, F. P. A. (1993). Imprecise conjugate prior densities for the one-parameter exponential family of distributions, *Statistics and Probability Letters*, **16**, 337-342, North-Holland.
- [27] Cox, D. R. and Hinckley, D. V. (1974). *Theoretical Statistics*, Chapman and Hall, London.
- [28] Csenki, A. (1979). A note on confidence sequences in multiparameter exponential families, *Journal of Multivariate Analysis*, **9**, 337-340.
- [29] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families, *The Annals of Statistics*, **7**, 269-281.
- [30] Druilhet, P. and Pommeret, D. (2012). Invariant conjugate analysis for exponential families, Bayesian Analysis, *International Society for Bayesian Analysis*, **7(4)**, 903-916.
- [31] Eaton, M. L. (1970). A complete class theorem for multidimensional one sided alternatives, *The Annals of Statistics*, **41**, 1884-1888.
- [32] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency), *The Annals of Statistics*, **3(6)**, 1189-1242.
- [33] Efron, B. (1978). The geometry of exponential families, *The Annals of Statistics*, **6(2)**, 362-376.
- [34] Eshima, N. (2004). Canonical exponential models for analysis of association between two sets of variables, *Statistics and Probability Letters*, **66**, 135-144.
- [35] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.
- [36] Ferrari, A., Letac, G. and Tournet, J. Y. (2007). Exponential families of mixed Poisson distributions, *Journal of Multivariate Analysis*, **98**, 1283-1292.

- [37] Ferrari, S. L. P. and Cordeiro, G. M. (1996). Corrected score tests for exponential family nonlinear models, *Statistics and Probability Letters*, **26**(1), 7-12.
- [38] Fujisawa, H. (2003). Asymptotic properties of conditional maximum likelihood estimator in a certain exponential model, *Journal of Multivariate Analysis*, **86**, 126-142.
- [39] Ghahramani, Z. and Helller, K. A. (2006). *Bayesian sets*, *Advances in Neural Information Processing Systems*, **18**, 435-442, Cambridge, MA, USA, The MIT Press.
- [40] Ghosh, M., Hwang, J. T. and Tsui, K. W. (1984). Construction of improved estimators In multiparameter estimation for continuous exponential families, *Journal Of Multivariate Analysis*, **14**, 212-220.
- [41] Gombay, E. and Horvth, L. L. (1992). A goodness-of-fit test for exponential Families, *Statistics and Probability Letters*, **15**, 235-239.
- [42] Gunasekar, S., Ravikumar, P. and Ghosh, J. (2014). *Exponential family matrix completion under structural constraints*, Proceedings of the 31st International Conference on Machine Learning (ICML-14), 1917-1925.
- [43] Gupta, S. S. and Li, J. (2005). On empirical bayes procedures for selecting good populations in a positive exponential family, *Journal of Statistical Planning and Inference*, **129**, 3-18.
- [44] Hassairi, A. and Louati, M. (2009). Multivariate stable exponential families and Tweedie scale, *Journal of Statistical Planning and Inference*, **139**, 143-158.
- [45] Heller, K. A. and Ghahramani, Z. (2007). *A Nonparametric Bayesian Approach to Modeling Overlapping Clusters*. AISTATS.
- [46] Heller, K. A., Williamson, S. and Ghahramani, Z. (2008). Statistical models for partial membership, *Proceedings of the International Conference on Machine Learning*, 392-399.
- [47] Hougaard, P. (1985). Saddlepoint approximations for curved exponential families, *Statistics and Probability Letters*, **3**, 161-166.
- [48] Hwang, S. Y. and Basawa, I. V. (1994). Large sample inference for conditional exponential families with applications to nonlinear time series, *Journal of Statistical Planning and Inference*, **38**, 141-158, North-Holland.
- [49] Iverson, G. J. and Harp, S. A. (1987). A conditional likelihood ratio test for order restrictions in exponential families, *Mathematical Social Sciences*, **14**, 141-159, 1987.
- [50] Jafari Jozani, M., Nematollahi, N. and. Shafie, K. (2002). An admissible minimax estimator of a bounded scale-parameter in a subclass of the exponential family under scale-invariant squared-error loss, *Statistics and Probability Letters*, **60**, 437-444.

- [51] Johanson, S. (1979). *Introduction to the Theory of Regular Exponential Families*, Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark.
- [52] Joshi, V. M. (1976). On the attainment of the Cramer-Rao lower bound, *The Annals of Statistics*, **4**, 998-1002.
- [53] Kadane, J. B. and Olkin, I. (1990). Inequalities for predictive ratios and posterior variances in natural exponential families, *Journal of Multivariate Analysis*, **33**, 275-285.
- [54] Kallenberg, W. C. M. (1981). Bahadur deficiency of likelihood ratio tests in exponential families, *Journal of Multivariate Analysis*, **11**, 506-531.
- [55] Karakostas, K. X. (1985). On minimum variance estimators, *The American Statistician*, **39**, 303-305.
- [56] Kokonendji, C. C. and Khoudar, M. (2006). On Levy measures for infinitely divisible natural exponential families, *Statistics and Probability Letters*, **76**, 1364-1368.
- [57] Kuchler, U. and Sorensen, M. (1994). Exponential families of stochastic processes and Lévy processes, *Journal of Statistical Planning and Inference*, **39**, 211-237, North-Holland.
- [58] Kumon, M. (2009). On the conditions for the existence of ancillary statistics in a curved exponential family, *Statistical Methodology*, **6**, 320-335.
- [59] Lee, H., Raina, R., Teichman, A. and Ng., A. Y. (2009). *Exponential Family Sparse Coding with Applications to Self-taught Learning*, Proceeding of the 21st international joint conference on artificial intelligence, 1113-1119.
- [60] Letac, G. and Seshadri, V. (1988). Haight's distribution as a natural exponential family, *Statistics and Probability Letters*, **6**, 165-169.
- [61] Lee, J. Kaiser, M. S. and Cressie, N. (2001). Multiway dependence in exponential family conditional distributions, *Journal of Multivariate Analysis*, **79**, 171-190.
- [62] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition, Springer-Verlag, New York.
- [63] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd edition, Springer, New York.
- [64] Levy, M. S. (1985). A note on nonunique MLE's and sufficient statistics, *The American Statistician*, **39**, 66-75.
- [65] Li, J., Gupta, S. S. and Liese, F. (2005). Convergence rates of empirical Bayes estimation in exponential family, *Journal of Statistical Planning and Inference*, **131**, 101-115.

- [66] Liang, T. (2009). Empirical bayes estimation of in positive exponential families, *Journal of Statistical Planning and Inference*, **139**, 411-424.
- [67] Martin, C. and Shubov, V. (1993). Natural exponential families of probability distributions and exponential-polynomial approximation, *Applied Mathematics and Computation*, **59**, 275-297.
- [68] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Second edition, chapman and hall CRC monographs on statistics and applied probability, chapman and hall CRC.
- [69] Miao, W. and Hahn, M. (1996). Existence and strong consistency of maximum likelihood estimates for 1-dimensional exponential families, *Statistics and Probability Letters*, **28(1)**, 9-21.
- [70] Michel, R. (1979). Third order efficiency of conditional tests for exponential families, *Journal of Multivariate Analysis*, **9**, 401-409.
- [71] Mohamed, S., Helller, K. A. and Ghahramani, Z. (2009). Bayesian exponential family PCA, *In Advances in Neural Information Proceeding Systems*, **21**, 1089-1096.
- [72] Mohamed, S., Helller, K.A. and Ghahramani, Z. (2013). *A Simple and General Exponential Family Framework for Partial Membership and Factor; Handbook of Mixed-Membership Models and their Applications*, CRC Press.
- [73] Molenberghs, G. and Ryan, L. M. (1999). An exponential family model for clustered multivariate binary data, *Environmetrics*, **10**, 279-300.
- [74] Moore, D. S. (1971). Maximum likelihood and sufficient statistics, *The American Mathematical Monthly*, **78**, 50-52.
- [75] Morris, C. N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *The Annals of Statistics*, **11**, 515-529.
- [76] Nicolaou, A. (2002). Conditional score tests in the exponential family, *Statistics and Probability Letters*, **60**, 69-74.
- [77] Palmgren, J. and Ekholm, A. (1987). Exponential family non-linear models for categorical, *Applied Stochastic Models and Data Analysis*, **3**, 111-124.
- [78] Papathanasiou, V. (1990). Characterizations of multidimensional exponential families by cacoullos-type inequalities, *Journal of Multivariate Analysis*, **35**, 102-107.
- [79] Parsian, A. and Nematollahi, N. (1996). Estimation of scale parameter under entropy loss function. *Journal of Statistical Planning and Inference*, **52**, 77-91.

- [80] Pham-Gia, T. (1994). The hazard rate of the power-quadratic exponential family of distributions, *Statistics and Probability Letters*, **20**, 375-382.
- [81] Portnoy, S. (1977). Asymptotic efficiency of minimum variance unbiased estimators, *The Annals of Statistics*, **5**, 522-529.
- [82] Rahman, M. S. and Gupta, R. P. (1993). Family of transformed chi-square distributions. *Communications in Statistics - Theory and Methods*, **22(1)**, 135-146.
- [83] Rauh, J., Kahle, T. and Ay, N. (2011). Support sets in exponential families and oriented matroid theory, *International Journal of Approximate Reasoning*, **52**, 613-626.
- [84] Rufo, J., Martin, J. and Perez, C. J. (2009). Inference on exponential families with mixture of prior distributions, *Computational Statistics and Data Analysis*, **53**, 3271-3280.
- [85] Sampson, A. and Spencer, B. (1976). Sufficiency, minimal sufficiency and the lack thereof, *The American Statistician*, **30**, 34-35.
- [86] Sampson, A. R. (1976). Stepwise. BAN estimators for exponential families with multivariate normal applications, *Journal of Multivariate Analysis*, **6**, 167-175.
- [87] Seeger, M. (2005). *Expectation propagation for exponential families*, EPFL-REPORT-161464.
- [88] Shanmugam, R. (1989). Asymptotic homogeneity tests for mean exponential family distributions, *Journal of Statistical Planning and Inference*, **23**, 227-241, North-Holland.
- [89] Shao, J. (2003). *Mathematical Statistics*, Springer, New York.
- [90] Singh, R. S. (1995). Empirical bayes linear loss hypothesis testing in a nonregular exponential family, *Journal of Statistical Planning and Inference*, **43**, 107-120.
- [91] Stein, C. (1974). *Estimation of the Mean of a Multivariate Normal Distribution*. Stanford university technical report, **48**.
- [92] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution, *The Annals of Statistics*, **9**, 1135-1151.
- [93] Tahir, M. (1994). Sequential estimation of the mean of an exponential family in three stages, *Journal of Statistical Planning and Inference*, **38**, 1-12, North-Holland.
- [94] Terbeche, M. and Oluyede, B. O. (2003). Fixed and sequential designs for estimation in the exponential family with comparisons and applications to binomial proportions using beta priors, *Journal of Computational and Applied Mathematics*, **154(1)**, 175-182.

- [95] Tsai, M. T. M. (1993). Ui score tests for some restricted alternatives in exponential families. *Journal of Multivariate Analysis*, **45**(2), 305-323.
- [96] Vegas, E., del Castillo, J. and Ocana, J. (2000). Efficiency and exponential models in a variance-reduction technique for dichotomous response variables, *Journal of Statistical Planning and Inference*, **85**, 61-74.
- [97] Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families and variational inference, *Foundations and Trends in Machine Learning*, **1**(1-2), 1-305.
- [98] Welling, M., Rosen-Zvi, M. and Hinton, G. (2005). Exponential family harmoniums with an application to information retrieval, *Advances in Neural Information Processing Systems*, **17**, 1481-1488.
- [99] Wingate, D. and Singh, S. (2007). *Exponential Family Predictive Representations of State*, *Neural Information Processing Systems*, NIPS, (to appear).
- [100] Wisjman, R. A. (1973). On the attainment of the cramer-rao lower bound, *The Annals of Statistics*, **1**, 538-542.
- [101] Withers, C. S. and Nadarajah, S. (2008). Canonical regression models for exponential families, *Journal of the Korean Statistical Society*, **37**, 119-127.
- [102] Ycart, B. (1989). Markov processes and exponential families on a finite set, *Statistics and Probability Letters*, **8**, 371-376, North-Holland.
- [103] Ycart, B. (1992). Markov processes and exponential families, *Stochastic Processes and Their Applications*, **41**, 203-214, North-Holland.
- [104] Ycart, B. (1992). Integer valued markov processes and exponential families, *Statistics and Probability Letters*, **14**, 71-78, North-Holland, 1992.
- [105] Zacks, S. (1971). *The Theory of Statistical Inference*, Johns Wiley and Sons, New York.
- [106] Zografos, K. and Ferentinos, K. (1994). An information theoretic argument for the validity of the exponential model, *Metrika*, **41**, 109-119.