مروری بر کاربردهای خانوادهٔ نمایی در استنباط آماری

مهدی شمس [۱]

تاریخ دریافت: ۱۳۹۶/۸/۲۴
تاریخ پذیرش: ۱۳۹۷/۱۰/۲۱

چکیده:
در این مقاله پس از معرفی خانوادهٔ نمایی و تاریخچه‌ای از کارهای انجام‌شده توسط محققان در شاخه‌های مختلف آمار، به شرح کاربردهایی از این خانواده در استنباط آماری به‌ویژه در مسئلهٔ برآورد، آزمون‌های فرضیهٔ آماری و مفاهیم نظریهٔ اطلاع آماری پرداخته می‌شود.

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

۱ مقدمه

خانوادهٔ نمایی یک ردهٔ 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، نظریهٔ 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 خانوادهٔ نمایی و شرایط خانوادهٔ 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 نااریب با کمترین واریانس، نامساوی کرامر-رائو و کاکولوس، بهترین برآوردگر مجاناً نرمال، ماکسیمم کردن آنتروپی یک خانوادهٔ نمایی، محاسبهٔ فاصلهٔ کولبک-لایبلر بین دو خانوادهٔ نمایی و آزمون‌های فرضیهٔ آماری است. در مراجع [۱۰، ۱۵] می‌توان 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 خانوادهٔ نمایی شرطی و کاربرد در سری‌های زمانی غیرخطی [۴۸]، توزیع‌های پوآسون آمیخته [۳۱]، کاربرد در معادلات دیفرانسیل تصادفی [۹]، بررسی نتایج هندسی در خانوادهٔ 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 مؤلفهٔ اصلی [۱۸، ۷۱]، تحلیل مدل‌های لجستیک [۷]، روش کاهش واریانس برای متغیرهای پاسخ دوحالتی [۹۶]، داده‌های رسته‌ای [۲۹]، داده‌های دودویی چندمتغیره خوشه‌ای [۷۳]، خوشه‌بندی فازی

[۱] هیأت علمی گروه آمار، دانشگاه کاشان، ایران

[۲] Generahzed Linear Model

[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 نظریهٔ ماتروید جهت‌دار [۸۳]، بسته بودن تحت آماره‌های ترتیبی [۹]، معیاری برای تشخیص متعلق بودن به خانوادهٔ نمایی یک توزیع و مقدار فاصله یک توزیع دلخواه از نمایی بودن [۲۷]، ماتریس خانوادهٔ نمایی تحت محدودیت‌های ساختاری [۳۷]، نرخ خطر برای خانواده‌های نمایی با توان درجهٔ دو [۸۰]، تعمیم معادلهٔ اشتاین [۱۴]، کارایی آزمون‌های نسبت درست‌نمایی [۵۴]، آزمون نسبت درست‌نمایی برای آزمون‌های فرضیه با فرضیهٔ 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]، دنبالهٔ مجموعهٔ اطمینان محدب [۲۳]، وجود، سازگاری و مجانباً نرمال بودن MLE [۳] [۴، ۷، ۶۹]، خواص مجانبی MLE شرطی [۳۳]، تقریب تابع چگالی MLE در خانوادهٔ نمایی خمیده با استفاده از بسط نقطهٔ 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 خانوادهٔ نمایی خمیده [۵۸]، بهترین برآوردگر مجانباً نرمال گام به گام [۸۶]، تقریب توسط چندجمله‌ای‌های دریکله [۶۷]، معیاری برای نیرومندی برآوردگرهای نااریب خانوادهٔ نمایی با تابع واریانس درجهٔ 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 از روش مونته‌کارلو برای برآورد فاصلهٔ کولبک-لایبلر [۴۸]، توزیع‌های پیشینی غیرآگاهنده در خانوادهٔ نمایی واریانس درجهٔ 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 [۱۳، ۳۸]، 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 خانواده توابع چگالی یا جرم احتمال مهم در آمار خانوادهٔ نمایی هستند که یک ردهٔ کلّی از اکثر توابع چگالی یا جرم احتمال معروف در آمار را شامل می‌شوند و به خانواده کوپمن-پیتمن-داریموس (به واسطهٔ تحقیقات مستقل این دانشمندان در این زمینه) نیز معروف هستند. یک خانواده توابع چگالی یا توابع جرم احتمال:

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

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

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

که توابع c، h، t_i و w_i توابع با مقادیر حقیقی هستند. بازپارامتریدن [۷] رابطهٔ (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:

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

[۳] Maximum Likelihood Estimator

[۴] Uniformly Minimum Variance Unbiased Estimator

[۵] Minimum Risk Equivariant

[۶] Bayes Information Criterion

[۷] Reparametrization

که در آن $s(x)=\log(g(x))$، $d(\boldsymbol{\theta})=\log(c(\boldsymbol{\theta}))$ و A بستگی به $\boldsymbol{\theta}$ ندارد. بنا بر این خانوادهٔ $U(0,\theta)$ متعلق به خانوادهٔ نمایی نیستند، زیرا تکیه‌گاه آن به θ بستگی دارد.

بازپارامتریدنی که کاربردهای نظری دارد استفاده از پارامتر طبیعی $\boldsymbol{\eta}$ است. فضای پارامتر طبیعی Ω یک مجموعه است که انتگرال تابع هسته متناهی باشد، یعنی:

$$\Omega=\Big\{\boldsymbol{\eta}=(\eta_1,\ldots,\eta_k):\frac{1}{c^*(\boldsymbol{\eta})} \equiv\int_{-\infty}^{\infty}h(x)\exp\left[\sum_{i=1}^{q}\eta_i t_i(x)\right]dx<\infty\Big\},$$

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

$$f(x|\boldsymbol{\eta}h(x)c^*(\boldsymbol{\eta})\exp\left[\sum_{i=1}^{q}\eta_i t_i(x)\right] \qquad (۲)$$

که $h(x)$ و $t_i(x)$ همان توابع در رابطهٔ (۱) هستند و $\boldsymbol{\eta}\in\Omega$. مشابه قبل بازپارامتریدن در این‌جا نیز یکتا نیست.

اکنون به تعریف خانوادهٔ نمایی منظم (یا خانوادهٔ نمایی پر رتبه) پرداخته می‌شود. اگر $d_i(y)$ به‌صورت یکی از عبارت‌های t_i، $w_i(\theta)$ یا η_i تعریف شود، یک قید خطی (وابستهٔ خطی) توسط $d_q(y),\ldots,d_1(y)$ به‌صورت $\sum_{i=1}^{q}a_i d_i(y)=c$ تعریف می‌شود که a_iها و c ثابت‌اند به‌طوری که همهٔ a_iها صفر نیستند. اگر تساوی بالا فقط وقتی که $a_1=\ldots=a_k=0$ برقرار باشد، آن‌گاه $d_i(y)$ یک قید خطی نیستند (مستقل خطی هستند). در حالتی که یک وابستگی خطی بین t_iها برقرار باشد، برخی از اعضای فضای پارامتر اضافی هستند و در این حالت خانوادهٔ نمایی را بیش‌کامل 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 هیچ‌یک قید خطی نباشند.

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

اگر شرط ۱ و ۲ برقرار باشد، خانوادهٔ نمایی منظم است (REF[۸]). برای خانوادهٔ نمایی یک پارامتری، مستطیل یک‌بعدی، فاصله است و فقط تابعی از یک متغیر که قید خطی داشته باشد، تابع ثابت است.

مثال ۱.۱. اگر $X\sim N(\mu,\sigma^2)$، $\boldsymbol{\theta}=(\mu,\sigma)\in\mathbb{R}\times\mathbb{R}^+$، از این‌که بازپارامتریدن معمولی به‌صورت رابطهٔ (۱) یعنی

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

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

به‌طور مشابه توزیع‌های گاما، بتا، پوآسون، دوجمله‌ای، پاسکال و هندسی خانوادهٔ نمایی هستند ولی توزیع‌هایی نظیر کوشی و یکنواخت متعلق به خانوادهٔ نمایی نیستند زیرا تابع چگالی آنها را نمی‌توان به‌صورت رابطهٔ (۱) نوشت. اگر t_i یا η_i در قید خطی صدق کنند، آن‌گاه می‌توان تعداد جملات توان در رابطهٔ (۱) را کاهش داد. به‌عنوان مثال، فرض کنید $(X_1\ldots,X_j)\sim M_j(n,p_1,\ldots,p_j)$ که در آن $\dim\Theta=j$. در این صورت $\sum_{i=1}^{j}X_i=n$ و $\sum_{i=1}^{j}p_i=1$. اگر $h(\boldsymbol{x})=n!/\prod_{i=1}^{j}x_i!$، آن‌گاه

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

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

[۸] Regular Exponential Family

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

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

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

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

در حد امکان بزرگ باشد و قرار دهید:

$\Theta_{\boldsymbol{\eta}} = \{\boldsymbol{\theta} \in \Theta :$ تعریف شده باشند $w_q(\boldsymbol{\theta}), \ldots, w_1(\boldsymbol{\theta})\}$.

در این‌صورت فرض کنید فضای پارامتر طبیعی

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

باشد. به‌عبارت دیگر برای سادگی تعریف کنید $\boldsymbol{\eta}_i = w_i(\boldsymbol{\theta})$. به‌ازای هر توزیع عمومی، $\boldsymbol{\eta}$ یک تابع یک به یک از $\boldsymbol{\theta}$ است.

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

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

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

خانوادهٔ نمایی یک‌پارامتری کانونی

$$f(x|\boldsymbol{\eta}) = h(x)\exp\left(\sum_{i=1}^{q}\eta_i t_i(x) - \Lambda(\boldsymbol{\eta})\right) \qquad (۳)$$

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

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

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

$$E\left\{\left[\frac{h'(X)}{h(X)} + \sum_{i=1}^{q}\eta_i t'_i(X)\right]g(X)\right\} = -E\left(g'(X)\right) \qquad (۴)$$

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

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

در حالت توزیع نرمال $N(\mu, \sigma^2)$ رابطهٔ بالا به‌صورت زیر

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

آشنای بدل می‌شود که در حالت اخیر با قرار دادن $g(x) = 1$ و $g(x) = x$ به‌ترتیب گشتاورهای $E(X) = \mu$ و $E(X^2) = \sigma^2 + \mu^2$ به دست می‌آیند.

در پایان یک مثال نقض از [۱۴] ارائه می‌شود که در خانوادهٔ نمایی یک پارامتری کانونی رابطهٔ (۳)، به‌صورت زیر

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

معادلهٔ اشتاین برقرار نیست. برای $n = 2, 3, \ldots$ تعریف می‌کنیم:

$$h(x) = \begin{cases} \frac{1}{2}\sin\left(\frac{\pi}{4}x\right) & 0 \leq x \leq 2 \\ \left(\frac{1}{n} + c_n(x)\right) + \left(\frac{1}{n} - c_n(x)\right) & n \leq x \leq n + \frac{2}{n^2} \\ \quad \times\cos\left(\pi n^2(x-n) - \pi\right) & \\ \frac{2}{(n+1)^2} & n + \frac{2}{n^2} \leq x < n+1 \end{cases}$$

که در آن

$$c_n(x) = \begin{cases} \frac{1}{n^2} & n \leq x < n + \frac{1}{n^2} \\ \frac{1}{(n+1)^2} & n + \frac{1}{n^2} \leq x < n + \frac{2}{n^2}. \end{cases}$$

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

۲ خانواده‌های نمایی و بسندگی

در تلخیص داده‌ها توسط یک آماره ممکن است اطلاعات نهفته در نمونه از بین برود و از این رو آماره‌ای که برای استنباط به کار برده می‌شود نباید باعث تلخیص نمونهٔ اوّلیه به قسمی شود که دیگر برای بررسی مسئله کفایت نکند. علت مطالعهٔ بسندگی در استنباط آماری این است که خلاصه کردن و فشرده کردن داده‌ها توسط آماره‌ای مورد علاقه هست که بدون از دست دادن اطلاعات، شامل تمام اطلاعات نمونه پیرامون پارامتر باشد. آمارهٔ بسندهٔ مینیمال، حداکثر تلخیص را روی نمونه انجام می‌دهد، ولی ممکن است شامل اطلاعات اضافی (اطلاعات فرعی یا کمکی) باشد که برای تخمین پارامتر مجهول مفید نیست و در اصطلاح گوییم آمارهٔ بسنده هنوز به‌طور کامل فشرده نشده است و به‌عبارت دیگر ممکن است یک تابع از آمارهٔ بسنده که کمکی است وجود داشته باشد. وجود آمارهٔ بسنده کامل تضمین می‌کند که تابعی از آمارهٔ بسندهٔ کامل که کمکی باشد وجود ندارد و این بدین معنی است که این آماره شامل اطلاعات کمکی و غیر لازم در بارهٔ پارامتر مجهول نیست. به‌عبارت عامیانه‌تر کامل بودن یک آمارهٔ بسنده، منجر به بیشترین فشردگی در داده‌ها می‌شود و یا این که آمارهٔ بسندهٔ کامل در ایجاد بیشترین فشردگی در داده‌ها موفق است.

فرض کنید $X_n, \ldots, X_1$ یک نمونهٔ تصادفی از خانواده‌های نمایی با بازپارامتریدن معمولی داده شده در رابطهٔ (۲) باشند در این صورت تابع چگالی احتمال یا جرم احتمال توأم آن‌که در ذیل آمده است یک خانواه نمایی k-پارامتری است:

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

در این صورت

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

الف) یک آمارهٔ بسنده برای $\boldsymbol{\eta}$ است.

ب) یک آمارهٔ بسندهٔ مینیمال برای $\boldsymbol{\eta}$ است، هرگاه $\eta_k, \ldots, \eta_1$ قید خطی نباشند.

ج) یک آمارهٔ کامل برای $\boldsymbol{\eta}$ است، هرگاه Ω شامل یک مستطیل k-بعدی باشد.

قسمت (الف) با استفاده از قضیهٔ تجزیهٔ فیشر-نیمن قابل بررسی است [۶۲]، ص ۳۵). برای اثبات قسمت (ب)، با توجه به این‌که رابطهٔ زیر

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

به $\boldsymbol{\eta}$ بستگی ندارد اگر و تنها اگر به‌ازای هر η_i و یک ثابت d،

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

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

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

مثال ۱.۲. اگر $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \gamma_0^2 \mu^2)$ که $\gamma_0^2 > 0$ معلوم و $\mu > 0$ است، آن‌گاه $f(x|\mu)$ یک خانوادهٔ نمایی یک پارامتری با $\Theta = (0, \infty)$ (که شامل یک مستطیل یک‌بعدی است) هست و $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ یک آمارهٔ بسندهٔ مینیمال است و از آن‌جایی که به‌ازای هر μ عبارت زیر برقرار است:

$$E_\mu\left[\frac{n+\gamma_0^2}{1+\gamma_0^2} \sum_{i=1}^{n} X_i^2 - \left[\sum_{i=1}^{n} X_i\right]^2\right] = 0,$$

آمارهٔ بسندهٔ مینیمال بالا کامل نیست. این مثال نشان می‌دهد که مستطیل باید در عوض Θ، شامل Ω باشد. شایان ذکر است که

یک آمارهٔ بسندهٔ مینیمال k-پارامتری برای یک پارامتر j-بعدی که $j < k$، کامل نیست. در این مثال $j = ۱ < ۲ = k$ ([۲۲]، ص ۳۱).

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

مثال ۲.۲. اگر $X \sim N(۰, \sigma^۲)$، این خانواده یک *REF* با آمارهٔ بسندهٔ مینیمال کامل $X^۲$ است. آمارهٔ X بسنده است ولی تابعی از $X^۲$ نیست و بنا بر این X بسندهٔ مینیمال نیست و طبق قضیهٔ بهادر [۵] کامل هم نیست.

مثال ۳.۲. اگر $X_۱, \ldots, X_n \overset{i.i.d.}{\sim} E(\beta)$ و $Y_۱, \ldots, Y_m \overset{i.i.d.}{\sim} E(\frac{\beta}{۲})$ دو نمونه از هم مستقل باشند، آن‌گاه تابع چگالی توأم $f(\boldsymbol{x}, \boldsymbol{y}|\beta)$ متعلق به *REF* یک‌پارامتری با آمارهٔ بسندهٔ کامل $\sum_{i=۱}^{n} X_i + ۲\sum_{i=۱}^{m} Y_i$ است.

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

لم ۴.۲. اگر $X_n, \ldots, X_۱$ یک نمونهٔ تصادفی از (۱) باشد، تابع احتمال آمارهٔ بسندهٔ زیر

$$\begin{aligned} T(\boldsymbol{X}) &= (T_۱(\boldsymbol{X}), \ldots, T_q(\boldsymbol{X})) \\ &= \left(\sum_{j=۱}^{n} t_۱(X_j), \ldots, \sum_{j=۱}^{n} t_q(X_j)\right) \end{aligned}$$

به‌صورت زیر خواهد بود:

$$f_{\boldsymbol{T}}(\boldsymbol{t}|(\boldsymbol{\theta}) = h_۰(\boldsymbol{t})c_۰(\boldsymbol{\theta}) \exp\left[\sum_{i=۱}^{q} w_i(\boldsymbol{\theta})t_i\right],$$

که متعلق به خانوادهٔ نمایی است و همچنین $\sum_{i=۱}^{q} t_i(X_i)$ نیز متعلق به یک خانوادهٔ نمایی است ([۳۰]، ص ۱۲۷).

نشان داده می‌شود اگر $P = \{f(x|\theta) : \theta \in \Theta\}$ یک خانواده از توابع چگالی احتمال باشد که به‌ازای هر $\theta \in \Theta$، $f(x|\theta)$ یک تابع پیوسته در x باشد و برای $X_۱, X_۲ \overset{i.i.d.}{\sim} f(x|\theta)$، $X_۱ + X_۲$ برای θ بسنده باشد، آن‌گاه P متعلق به یک خانوادهٔ 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 خانوادهٔ نمایی $P = \{P_\theta : \theta \in \Theta\}$ باشد، آن‌گاه توزیع X که روی مجموعهٔ بورل A بریده شده است (یعنی توزیع شرطی X به شرط $X \in A$) نیز متعلق به خانوادهٔ نمایی با تابع چگالی احتمال $f(x|\theta)I_A(x)/P_\theta(A)$ است ([۸۹]، ص ۱۴۴). همچنین به‌راحتی نشان داده می‌شود که اگر $T(X)$ یک آمارهٔ بسندهٔ کامل برای خانواده P باشد، آن‌گاه این آماره برای خانواده توزیع بریده شده روی مجموعهٔ بورل A نیز بسندهٔ کامل است ([۸۹]، ص ۱۴۸).

۳ خانواده‌های نمایی و MLE

روش ماکسیمم درست‌نمایی برای اوّلین بار توسط ریاضیدانانی چون لاگرانژ، برنولی، اویلر، لاپلاس و گاوس مورد استفاده قرار گرفت ولی سازماندهی اصلی و محبوبیت این روش به واسطهٔ تحقیقات فیشر است و واژه «ماکسیمم درست‌نمایی» اوّلین بار توسط این دانشمند مشهور مورد استفاده قرار گرفت. کسلا و برگر([۱۲]، ص۳۱۷) یک نتیجه مفید برای یافتن برآورد ماکسیمم درست‌نمایی (MLE) در حالت یک‌پارامتری 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 آمارهٔ بسندهٔ مینیمال است [۶۴، ۷۴]. این حقیقت در خانواده‌های نمایی دارای اهمیت است، زیرا در آنها لگاریتم تابع درست‌نمایی صورت ساده‌ای دارد و به‌راحتی می‌توان آمارهٔ بسندهٔ مینیمال را پیدا کرد. به‌عنوان نمونه در مثال ۳.۲، MLE برای β عبارت است از $\hat{\beta} = T/(n+m)$.

در حالت کلّی با اختیار کردن یک نمونهٔ n تایی از خانوادهٔ نمایی یک‌پارامتری

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

لگاریتم تابع درست‌نمایی به‌صورت زیر است:

$$l(\eta) = \sum_{i=1}^{n} \ln\left(h(x_i)\right) + n\ln\left(c^*(\eta)\right) + \eta \sum_{i=1}^{n} t(x_i)$$
$$= \sum_{i=1}^{n} \ln\left(h(x_i)\right) - n\Lambda(\eta) + \eta \sum_{i=1}^{n} t(x_i)$$

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

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

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

$$\hat{\eta}\left(T(\boldsymbol{x})\right) = \dot{\Lambda}^{-1}\left(T(\boldsymbol{x})\right)$$
$$= \arg\max_{\eta\in\Omega}\left(\eta T(\boldsymbol{x}) - \Lambda(\eta)\right).$$

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

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

از طرف دیگر با به‌ترتیب 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:

$$\dot{\Lambda}(\eta) e^{\Lambda(\eta)} = \int_\Omega h(x)t(x)e^{\eta t(x)}dx = e^{\Lambda(\eta)} E_\eta\left(t(X)\right),$$
$$\left(\ddot{\Lambda}(\eta) + \dot{\Lambda}^2(\eta)\right) e^{\Lambda(\eta)} = \int_\Omega h(x)t^2(x)e^{\eta t(x)}dx = e^{\Lambda(\eta)} E_\eta\left(t^2(X)\right).$$

بنا بر این:

$$E_\eta\left(t(X)\right) = \dot{\Lambda}(\eta),$$
$$\mathrm{Var}_\eta\left(t(X)\right) = \ddot{\Lambda}(\eta).$$

با توجه به قضیهٔ حدی مرکزی می‌توان نتیجه گرفت که $T(\boldsymbol{x})$ به‌طور تقریبی دارای توزیع $N(\dot{\Lambda}(\eta), \frac{1}{n}\ddot{\Lambda}(\eta))$ است. همچنین به‌راحتی می‌توان ثابت کرد که $\Lambda(\eta)$ بی‌نهایت بار مشتق‌پذیر و در حقیقت یک تابع محدب است. در تحلیل محدب لگاریتم تابع درست‌نمایی ماکسیمم شده یعنی

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

را تبدیل فنچل-لژاندر Λ گویند. همچنین در تحلیل محدب داریم $\frac{d}{dt}\Lambda^*(T) = \arg\max_{\eta\in\Omega}\left(\eta T - \Lambda(\eta)\right)$. 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 خانوادهٔ نمایی یک‌پارامتری این اتفاق نمی‌افتد. مشاهده می‌شود که Λ^* یک MLE را محاسبه می‌کند، یعنی $\eta(\mu) = \dot{\Lambda}^*(\mu)$. همچنین داریم $\Lambda^*(\mu) = \eta(\mu)\mu - \Lambda(\eta(\mu))$ و در پی آن $\dot{\Lambda}\left(\dot{\Lambda}^*(\mu)\right) = \dot{\Lambda}\left(\eta(\mu)\right) = E_{\eta(\mu)}\left(t(X)\right) = \mu$ که 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

$$\frac{d}{d\mu}\tilde{l}(\mu) = n\frac{\frac{d}{d\eta}\left(\eta T(\boldsymbol{X}) - \Lambda(\eta)\right)}{\frac{d\mu}{d\eta}}$$
$$= n\frac{T(\boldsymbol{X}) - \mu}{\mathrm{Var}_{\eta(\mu)}(t(X))}$$
$$= n\ddot{\Lambda}^*(\mu)\left[T(\boldsymbol{X}) - \mu\right]$$

که نشان می‌دهد $\hat{\mu} = T(\boldsymbol{X})$. در [۶۹] برای یک خانوادهٔ نمایی به‌صورت $f(x|\eta) = \exp(\eta x - \Lambda(\eta))$، شرط وجود، سازگاری و مجانباً نرمال بودن MLE مورد تحلیل قرار می‌گیرد.

در پایان با گرفتن لگاریتم از دو طرف رابطهٔ 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 خانوادهٔ نمایی یک پارامتری به‌ترتیب به‌صورت

زیر محاسبه می‌شوند:

$$\gamma(\eta) = \frac{\kappa_۳}{\kappa_۲^{۳/۲}} = \frac{\Lambda^{(۳)}(\eta)}{\ddot{\Lambda}(\eta)^{۳/۲}},$$
$$\delta(\eta) = \frac{\kappa_۴}{\kappa_۲^{۲}} = \frac{\Lambda^{(۴)}(\eta)}{\ddot{\Lambda}(\eta)^{۲}}.$$

۴ خانواده‌های نمایی و UMVUE CRLB

در انتخاب یک برآوردگر نقطه‌ای به‌دنبال برآوردگری هستیم که به‌ازای 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 نااریب در نظر گرفته شود، برآوردگر بهینه با این ساختار را برآوردگر نااریب با کمترین واریانس (UMVUE) نامند. در ابتدا به یافتن یک UMVUE در یک کلاس کلّی از خانوادهٔ نمایی گسسته پرداخته می‌شود.

مثال ۱.۴. فرض کنید $X_n, \ldots, X_۱$ یک نمونهٔ تصادفی از خانوادهٔ نمایی یک‌پارامتری با تابع احتمال زیر باشد.

$$P(X_i = x) = \frac{\gamma(x)}{c(\theta)}\theta^x, \quad x = ۰, ۱, \ldots$$

با استفاده از لم ۴.۲، توزیع آمارهٔ بسنده $T(\boldsymbol{X}) = \sum_{i=۱}^{n} X_i$ نیز یک خانوادهٔ نمایی یک‌پارامتری با تابع احتمال زیر است:

$$P(T = t) = \frac{\gamma_n(t)}{[c(\theta)]^n}\theta^t, \quad t = ۰, ۱, \ldots$$

از این‌که رابطهٔ زیر برقرار است:

$$\sum_{t=۰}^{\infty} h(t)\gamma_n(t)\theta^t = [c(\theta)]^{n-p}\theta^r,$$

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

$$h(T) = \begin{cases} ۰ & T < r \\ \frac{\gamma_n - p(T-r)}{\gamma_n(T)} & T \geq r \end{cases}$$

و از این رو با اختیار کردن $p = ۱$، *UMVUE* برای

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

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

برخی از مرجع‌های آماری پیشنهاد می‌کنند که UMVUE توسط تعیین این‌که یک برآوردگر نااریب 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 آمارهٔ بسندهٔ کامل T استخراج شوند.

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

در حالتی‌که داده‌ها از یک REF یک پارامتری یا REF، q-پارامتری با $q > ۱$ بیایند، عموماً برای توابع غیرخطی از T، نامساوی کرامر-رائو اکید است. اگر T کامل باشد، $g(T)$، UMVUE امید خودش است و تعیین این‌که T یک آمارهٔ بسندهٔ کامل از یک REF یک‌پارامتری است نسبت به محاسبهٔ $\mathrm{Var}_\theta W$ و CRLB ساده‌تر است. اگر خانوادهٔ نمایی نباشد، CRLB ممکن است یک کران پایین برای واریانس برآوردگر نااریب 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 کلّی تابع امتیاز برای یک تابع وارون‌پذیر η یعنی $\xi = h(\eta)$ برابر است با:

$$\frac{d}{d\xi} l\left(h^{-۱}(\xi)\right) = \frac{\frac{d}{d\eta} l(\eta)|_{\eta = h^{-۱}(\xi)}}{\frac{dh}{d\eta}|_{\eta = h^{-۱}(\xi)}}$$

که دارای خاصیت اصلی زیر است:

$$\begin{aligned} E_\eta\left[\frac{d}{d\zeta} l(\zeta)\,|_{\zeta=\eta}\right] &= n\left[E_\eta\left(T(\boldsymbol{X})\right) - \dot{\Lambda}(\eta)\right] \\ &= n\left[\mu(\eta) - \mu(\eta)\right] \\ &= ۰. \end{aligned}$$

[۹] Cramer Rao Lower Bound

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

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

در عبارت بالا

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

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

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

$\mathrm{Var}_\eta\left(t(X)\right)$ تحت شرایط نظم برابر است با اطلاع فیشر برای پارامتر طبیعی η و یا برابر است با وارون اطلاع فیشر برای پارامتر $E_\eta\left(t(X)\right)$ ([۸۹]، ص ۱۷۱). همچنین بر اساس یک نمونهٔ تصادفی n تایی از توزیع بالا، CRLB برای برآوردگر ناریب $\hat{\xi}$ از $\xi=h(\eta)$ برابر است با:

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

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

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

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

با این‌که در حالت کلّی MLE یعنی $\hat{\xi}=h(\hat{\eta})$ ناریب نیست، با استفاده از تقریب دلتا تساوی تقریبی زیر

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

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

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

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

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

که در آن A به پارامتر $\boldsymbol{\theta}\in\Theta\subseteq\mathbb{R}^q$ بستگی ندارد و یک مجموعهٔ باز است، $h(x),c(\boldsymbol{\theta})>0$ و $Q(\boldsymbol{\theta})$ اولین مشتقات جزئی پیوسته دارد حاصل می‌شود و مقدار مینیمم برابر با ماتریس $M^t(\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1}M(\boldsymbol{\theta})$ است که در آن $\boldsymbol{\Sigma}$ ماتریس کواریانس $T(X)=(T_1(X),\ldots,T_k(X))$ و $M(\boldsymbol{\theta})=\left[\frac{\partial E_{\boldsymbol{\theta}}(T_i(X))}{\partial\theta_j}\right]_{k\times q}$ ماتریس مشتقات جزئی است [۱۰۶].

همچنین ثابت می‌شود دنبالهٔ ماتریس‌های اطلاع فیشر $\{I_n(\boldsymbol{\theta})\}_{n\in\mathbb{N}}$ صعودی است از این جهت که به‌ازای هر $n_1,n_2\in\mathbb{N}$ که $n_1\leq n_2$، $I_{n_2}(\boldsymbol{\theta})-I_{n_1}(\boldsymbol{\theta})\geq 0$، یعنی ماتریس $I_{n_2}(\boldsymbol{\theta})-I_{n_1}(\boldsymbol{\theta})$ نیمه‌معین مثبت است [۱۰۶]. در حالت چندمتغیره نیز مشابه حالت یک‌متغیره نامساوی کرامر-رائو به تساوی تبدیل می‌شود اگر و تنها اگر:

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

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

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

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

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

پیدا کرد. برای سادگی حالت $q=۲$ بررسی می‌شود. اگر $X_n,\ldots,X_۱$ یک نمونهٔ تصادفی از عبارت زیر باشد:

$$f(x|\boldsymbol{\theta})=h(x)\,c(\boldsymbol{\theta})\exp\left\{\sum_{j=۱}^{۲}Q_j(\boldsymbol{\theta})\,T_j(x)\right\}I_A(x)$$

باشد، برآوردگر $\boldsymbol{T}^t=(T_۱,T_۲)$ از دیدگاه نامساوی کرامر-رائو یک UMVUE برای $\boldsymbol{\theta}^t=(\theta_۱,\theta_۲)$ است هرگاه:

$$\begin{aligned}&ET_۱=\theta_۱,\\&ET_۲=\theta_۲,\\&VarT_۱=I^{-۱}_{۱۱}(\boldsymbol{\theta})=\tfrac{I_{۲۲}(\boldsymbol{\theta})}{\det(I(\boldsymbol{\theta}))},\\&VarT_۲=I^{-۱}_{۲۲}(\boldsymbol{\theta})=\tfrac{I_{۱۱}(\boldsymbol{\theta})}{\det(I(\boldsymbol{\theta}))}\end{aligned}$$

که در آن $I_{ii}(\boldsymbol{\theta})=E\Big(\frac{\partial}{\partial\theta_i}f(x|\boldsymbol{\theta})\Big)^۲$ و $I^{-۱}_{ii}(\boldsymbol{\theta})$، i-امین درایهٔ قطری ماتریس $I^{-۱}(\boldsymbol{\theta})$ است.

با توجه به نتیجه الف داریم:

$$\begin{aligned}\tfrac{\partial}{\partial\theta_۱}\log\textstyle\prod_{i=۱}^{n}f(x_i|\boldsymbol{\theta})=&nQ'_{۱\theta_۱}\left[\tfrac{۱}{n}\sum_{i=۱}^{n}T_۱(x_i)-E_{\boldsymbol{\theta}}\left(T_۱(X)\right)\right]\\&+nQ'_{۲\theta_۱}\left[\tfrac{۱}{n}\sum_{i=۱}^{n}T_۲(x_i)-E_{\boldsymbol{\theta}}\left(T_۲(X)\right)\right],\\\tfrac{\partial}{\partial\theta_۲}\log\textstyle\prod_{i=۱}^{n}f(x_i|\boldsymbol{\theta})=&nQ'_{۱\theta_۲}\left[\tfrac{۱}{n}\sum_{i=۱}^{n}T_۱(x_i)-E_{\boldsymbol{\theta}}\left(T_۱(X)\right)\right]\\&+nQ'_{۲\theta_۲}\left[\tfrac{۱}{n}\sum_{i=۱}^{n}T_۲(x_i)-E_{\boldsymbol{\theta}}\left(T_۲(X)\right)\right],\end{aligned}\qquad(۵)$$

که در آن

$$\left(E_{\boldsymbol{\theta}}\left(T_۱\right),E_{\boldsymbol{\theta}}\left(T_۲\right)\right)^t=[Q(\boldsymbol{\theta})]^{-۱}c_*(\boldsymbol{\theta}),$$

با $c^t_*(\boldsymbol{\theta})=(c'_۱,c'_۲)$، $Q(\boldsymbol{\theta})=\left[Q'_{i\theta_j}\right]_{۲\times۲}=\left[\frac{\partial}{\partial\theta_j}Q_i(\boldsymbol{\theta})\right]_{۲\times۲}$ که $c'_i=\frac{\partial}{\partial\theta_i}\log c(\boldsymbol{\theta})$ و $i,j=۱,۲$. بنا بر این برآوردگر $T^t=(T^*_۱,T^*_۲)$ با $T^*_j=\frac{۱}{n}\sum_{i=۱}^{n}T_j(x_i)$، UMVUE برای $\left(E_{\boldsymbol{\theta}}\left(T_۱\right),E_{\boldsymbol{\theta}}\left(T_۲\right)\right)$ است. UMVUE (برآوردگر ناریب با واریانس به‌طور یکنواخت مینیمم)های دیگر از توابع پارامتری دیگر تبدیلات خطی از $\left(E_{\boldsymbol{\theta}}\left(T_۱\right),E_{\boldsymbol{\theta}}\left(T_۲\right)\right)$ هستند. به‌عبارت دیگر از دیدگاه نامساوی کرامر-رائو فقط توابع پارامتری برآوردپذیر به‌صورت $\left(k_۱+k_۲E_{\boldsymbol{\theta}}\left(T_۱\right),k_۳+k_۴E_{\boldsymbol{\theta}}\left(\boldsymbol{T}_۲\right)\right)$ هستند که UMVUE نظیر آن $(k_۱+k_۲T^*_۱,k_۳+k_۴T^*_۲)$ است و k_iها ثابت‌هایی هستند که به $\boldsymbol{\theta}$ و x_iها بستگی ندارند. از این رو برای خانواده‌های غیرنمایی نمی‌توان UMVUE ها را از روش نامساوی کرامر-رائو پیدا کرد.

مثال ۲.۴. در توزیع سه‌جمله‌ای با تابع جرم احتمال زیر

$$\begin{aligned}&P\big(X_۱=x_۱,X_۲=x_۲,X_۳=x_۳\,|\theta_۱,\theta_۲,\theta_۳\big)\\&\quad=\tfrac{n!}{x_۱!x_۲!x_۳!}\theta_۱^{x_۱}\theta_۲^{x_۲}(۱-\theta_۱-\theta_۲)^{x_۳},\\&\quad x_۱+x_۲+x_۳=n,\quad \theta_۱+\theta_۲+\theta_۳=۱\end{aligned}$$

داریم، $Q_۱(\boldsymbol{\theta})=\frac{\theta_۱}{۱-\theta_۱-\theta_۲}$، $c(\boldsymbol{\theta})=(۱-\theta_۱-\theta_۲)^n$، $Q_۲(\boldsymbol{\theta})=\frac{\theta_۲}{۱-\theta_۱-\theta_۲}$، $T_۱(X)=X_۱$، $T_۲(X)=X_۲$ و $h(x)=\frac{n!}{x_۱!x_۲!x_۳!}I_A(x)$. بنا بر این با توجه به رابطهٔ (۵) داریم:

$$\begin{aligned}&ET_۱=n\theta_۱,\\&ET_۲=n\theta_۲,\\&\tfrac{\partial}{\partial\theta_۱}\log P=\tfrac{n(۱-\theta_۲)}{\theta_۱(۱-\theta_۱-\theta_۲)}\left(\tfrac{X_۱}{n}-\theta_۱\right)+\tfrac{n}{۱-\theta_۱-\theta_۲}\left(\tfrac{X_۲}{n}-\theta_۲\right),\\&\tfrac{\partial}{\partial\theta_۲}\log P=\tfrac{n}{۱-\theta_۱-\theta_۲}\left(\tfrac{X_۱}{n}-\theta_۱\right)+\tfrac{n(۱-\theta_۱)}{\theta_۲(۱-\theta_۱-\theta_۲)}\left(\tfrac{X_۲}{n}-\theta_۲\right),\\&I^{-۱}(\boldsymbol{\theta})=\tfrac{۱}{n}\begin{bmatrix}\theta_۱(۱-\theta_۱)&-\theta_۱\theta_۲\\-\theta_۱\theta_۲&\theta_۲(۱-\theta_۲)\end{bmatrix}.\end{aligned}$$

بنا بر این برآوردگر $T^t=(T^*_۱,T^*_۲)=\left(\frac{X_۱}{n},\frac{X_۲}{n}\right)$، $UMVUE$ برای $\boldsymbol{\theta}^t=(\theta_۱,\theta_۲)$ است و

$$\begin{aligned}&VarT^*_۱=I^{-۱}_{۱۱}(\boldsymbol{\theta})=\tfrac{\theta_۱(۱-\theta_۱)}{n},\\&VarT^*_۲=I^{-۱}_{۲۲}(\boldsymbol{\theta})=\tfrac{\theta_۲(۱-\theta_۲)}{n}.\end{aligned}$$

به غیر از نامساوی کرامر-رائو، نامساوهای دیگری نیز هستند که کران پایینی برای واریانس برآوردگر ارائه می‌کنند. از جمله معروف‌ترین آنها نامساوی کاکولوس است که پاپاتاناسیو [۷۸] آن را برای خانوادهٔ نمایی تعمیم داد. فرض کنید $(X_q,\ldots,X_۱)$ یک خانوادهٔ نمایی q-بعدی پیوسته با تابع چگالی زیر باشد:

$$f(\boldsymbol{x}|\boldsymbol{\theta})=c(\boldsymbol{\theta})\exp(\boldsymbol{\theta}.\boldsymbol{x}-s(\boldsymbol{x}))I_A(\boldsymbol{x})$$

که در آن $\boldsymbol{\theta}=(\theta_۱,\ldots,\theta_q)$ یک بردار پارامتر طبیعی و «.» نشان‌دهندهٔ ضرب اسکالر و مجموعهٔ A یک اجتماع متناهی از مجموع‌های همبند باز در $\mathbb{R}^q$ و $s(\boldsymbol{x})$ دارای مشتقات جزئی پیوسته باشد. تحت شرایط نظم مثل متناهی بودن قدر مطلق امید ریاضی ([۱۴]) نشان داده می‌شود:

$$E\left[(\nabla s(\boldsymbol{X})-\boldsymbol{\theta})\,g(\boldsymbol{X})\right]=E\left[\nabla g(\boldsymbol{X})\right]$$

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

$$\mathrm{Var}\left(g(\boldsymbol{X})\right) \geq E\left[\nabla g(\boldsymbol{X})\right] J^{-1}.E\left[\nabla g(\boldsymbol{X})\right]$$

که در آن J^{-1} وارون ماتریس نامنفرد $J=(s_{ij})$ با درایه‌های $s_{ij}=E\left(s_{ij}(\boldsymbol{X})\right)$ است که در آن $s_{ij}(\boldsymbol{X})=\frac{\partial^2 s(\boldsymbol{X})}{\partial x_i \partial x_j}$ هست. شایان ذکر است که پاپاتاناسیو [۷۸] نشان داد این مطلب هم صحیح است و با برقراری نامساوی بالا و داشتن شرایط نظم، خانوادهٔ نمایی به‌صورت داده شده است. به‌عنوان نمونه در توزیع نرمال چندمتغیره نامساوی به‌صورت زیر بدل می‌شود:

$$\mathrm{Var}\left(g(\boldsymbol{X})\right) \geq E\left[\nabla g(\boldsymbol{X})\right] \Sigma.E\left[\nabla g(\boldsymbol{X})\right].$$

برای اطلاعات بیشتر درمورد نامساوی‌های مشابه در حالت خانوادهٔ نمایی گسسته می‌توانند به [۷۸] مراجعه کنند.

در انتهای این بخش با انجام محاسبات می‌توان اطلاع متقابل (فاصلهٔ کولبک-لایبلر) بین دو خانوادهٔ نمایی 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 به‌صورت 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

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

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

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

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

در حالتی که یک نمونهٔ تصادفی n تایی از خانوادهٔ نمایی یک پارامتری در اختیار باشد، مجموع فاصله به‌صورت زیر خواهد بود:

$$D^{(n)}(\eta_1;\eta_2)=2nD(\eta_1;\eta_2)$$

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

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

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

$$\left|D(\eta_1;\eta_2)-I_{\eta_1}(\eta_1-\eta_2)^2\right|\leq\tfrac{1}{6}C\left(\eta_1;|\eta_1-\eta_2|\right)(\eta_1-\eta_2)^3$$

که در آن

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

یک ثابت لیپ‌شیتز موضعی یا هنگ پیوستگی برای I در η است.

اگر متغیر تصادفی X متعلق به خانوادهٔ نمایی 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

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

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

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

از حل معادلهٔ زیر

$$\frac{d\psi(f,\theta_0,\dots,\theta_k,\lambda)}{df(x|\eta)} = 1 + \log f(x|\eta) - \sum_{i=1}^{k}\theta_i t_i(x) + \theta_0 - \lambda(x) = 0$$

مقدار ماکسیمم آنتروپی برابر است با:

$$f(x|\eta) = \exp\left[\sum_{i=1}^{k}\eta_i t_i(x) - 1 - \theta_0 - \lambda(x)\right].$$

با توجه به برقراری شرط $f(x|\eta) > 0$ با اختیار کردن $\lambda(x) = 0$ و همچنین انتخاب $\theta_0 = -1 + \Lambda(\eta)$ 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 خانوادهٔ 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 رابطهٔ 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 این‌که $D(g;f) > 0$ 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 اندازهٔ شمارشی μ روی مجموعهٔ $\{-1,1\}$ به‌طوری که $\mu(\{-1\}) = \mu(\{1\}) = 1$، ماکسیمم آنتروپی برای توزیع یکنواخت گسسته $DU\{-1,1\}$ رخ می‌دهد.

ب) برای اندازهٔ لبگ μ روی $\mathbb{R}$ یعنی $\mu([a,b]) = b-a$ که $a \le b$، 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 نمونهٔ تصادفی مثل $X_1,\dots,X_n$ از خانوادهٔ نمایی چندپارامتری در نظر بگیرید که تابع چگالی آن به‌صورت زیر باشد:

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

فرض کنید به‌ازای هر $i = 1,\dots,q$، $E(t_i(\boldsymbol{X})) = c_i\theta + d_i$ 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 $v > 0$، $\sum_{j=1}^{q}E\left|c_j\theta + d_j - t_j(\boldsymbol{X})\right|^{2+v} < \infty$.

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:

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

آن‌گاه در [۸۶] ثابت می‌شود برآوردگر گام به گام $\hat{\theta}_n$ برای θ که به‌صورت بازگشتی به‌ازای هر $n \in \mathbb{N}$ به‌صورت زیر تعریف می‌شود:

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

به‌طوری‌که $\hat{\theta}_0$ یک مقدار اولّیه است. بهترین برآوردگر مجانباً نرمال برای θ است، یعنی

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

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

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

ذکر این نکته جالب است که برای خانوادهٔ نمایی یک‌پارامتری ($k=1$)، با قرار دادن $\hat{\theta}_0 = 0$، 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 $\boldsymbol{X} = (X_1, \ldots, X_r)^t$ (بردار $r \times 1$) دارای توزیع $N(\boldsymbol{0}, \Sigma)$ با تابع چگالی زیر باشد:

$$c(\rho) \exp\left[\sum_{i=1}^{3} w_i(\rho) t_i(\boldsymbol{x})\right],$$

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

$$w_1(\rho) = -(2 - 2\rho^2)^{-1},$$
$$w_2(\rho) = -\rho^2(2 - 2\rho^2)^{-1},$$
$$w_3(\rho) = -\rho(1 - \rho^2)^{-1},$$
$$t_1(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{x},$$
$$t_2(\boldsymbol{x}) = \sum_{i=2}^{r-1} x_i^2,$$
$$t_3(\boldsymbol{x}) = \sum_{i=1}^{r-1} x_i x_{i+1}.$$

به‌راحتی مشاهده می‌شود که

$$E(t_1(\boldsymbol{X})) = r,$$
$$E(t_2(\boldsymbol{X})) = r - 2,$$
$$E(t_3(\boldsymbol{X})) = (r-1)\rho$$

و با قرار دادن

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

$$\bar{w}_3(x) = \begin{cases} -(1+x^2)(1-x^2)^{-2} & |x| < 1 \\ -(1+x^2) & |x| \geq 1 \end{cases}$$

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

برای مشاهدهٔ 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 $X_1, \ldots, X_n$ یک نمونهٔ تصادفی از خانواده‌های نمایی با تابع چگالی زیر باشد:

$$f(x|\boldsymbol{\theta}) = h(x) \exp(\boldsymbol{\theta}.\boldsymbol{t}(x) - \Lambda(\boldsymbol{\theta}))$$

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

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

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

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

خواهد شد که در آن $T_n(\boldsymbol{x}) = \sum_{i=1}^{n} t(x_i)$. 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 به‌دست‌آمده است. با در نظر گرفتن پارامتر $\eta = h(\boldsymbol{\theta})$ که در آن h یک تبدیل یک به یک به‌طور پیوسته دو بار مشتق‌پذیر است (و از این به بعد آن را به‌طور خلاصه بازپارامتریدن هموار نامیم)، توزیع‌های پیشینی مزدوج زیر پیشنهاد می‌شوند:

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

$$\pi_{\boldsymbol{\alpha},\beta}(\eta) \propto \exp\left\{\boldsymbol{\theta}(\eta).\boldsymbol{\alpha} + \beta\Lambda\left(\boldsymbol{\theta}(\eta)\right)\right\} \qquad (۷)$$

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

در [۲۵] خانوادهٔ جدیدی از پیشین‌های مزدوج برای خانوادهٔ نمایی که نسبت به بازپارامتریدن ناوردا هستند معرفی می‌شود. به‌ازای هر بازپارامتریدن هموار $\eta(\theta)$، پیشینی مزدوج جفریز (JCP[۱۰]) متناظر به‌صورت زیر تعریف می‌شود:

$$\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\eta}) \propto \exp\left\{\boldsymbol{\theta}(\boldsymbol{\eta}).\boldsymbol{\alpha} + \beta\Lambda\left(\boldsymbol{\theta}(\boldsymbol{\eta})\right)\right\}\sqrt{|I_{\boldsymbol{\eta}}(\boldsymbol{\eta})|},$$

که در آن $I_{\boldsymbol{\eta}}(\boldsymbol{\eta})$ ماتریس اطلاع فیشر برای $\boldsymbol{\eta}$ است. در حالت خاص $\boldsymbol{\eta} = \theta$ داریم:

$$\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\theta}) \propto \exp\left\{\boldsymbol{\theta}.\boldsymbol{\alpha} + \beta\Lambda\left(\boldsymbol{\theta}\right)\right\}\sqrt{|I_{\boldsymbol{\theta}}(\boldsymbol{\theta})|}.$$

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

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

نتیجه گرفته می‌شود:

$$\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\eta}) = \left|\frac{d\boldsymbol{\theta}(\boldsymbol{\eta})}{d\boldsymbol{\eta}}\right| \pi^J_{\boldsymbol{\alpha},\beta}\left(\boldsymbol{\theta}(\boldsymbol{\eta})\right)$$

که نشان‌دهندهٔ یکسان بودن توزیع‌های $\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\eta})$و $\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\theta})$ یا به‌عبارت دیگر خاصیت ناوردایی JCP ها تحت بازپارامتریدن هموار است. این حقیقت نشان می‌دهد که شرایط روی $\boldsymbol{\alpha}$ و β که منجر به توزیع‌های پیشینی سره برای $\pi^J_{\boldsymbol{\alpha},\beta}(\boldsymbol{\eta})$ می‌شود، نیاز به انتخاب بازپارامتریدن هموار $\boldsymbol{\eta}$ ندارد. برای خانوادهٔ نمایی طبیعی درجهٔ دو یک‌متغیره بیان‌کنندهٔ این است که JCP های متناظر با پارامتر میانگین $m(\theta) = (X_i\,|\theta)$ معادل با پیشین‌های مزدوج استاندارد رابطهٔ (۶) هستند [۲۵]. بنا بر این در یک خانوادهٔ نمایی طبیعی درجهٔ دو بین پیشین‌های مزدوج استاندارد و JCP ها برای پارامتر میانگین، تمایزی نیست.

مثال ۱.۶. توزیع گوسین معکوس $IG(\mu, ۱)$ با تابع چگالی

$$f(x\,|\mu) = \sqrt{\frac{۱}{۲\pi x^۳}}\exp\left\{-\frac{(x-\mu)^۲}{۲\mu^۲ x}\right\},\ x > ۰$$

نسبت به اندازهٔ لبگ روی $\mathbb{R}^+$ را در نظر بگیرید. ماتریس اطلاع فیشر برای μ عبارت است از $I_\mu(\mu) = \mu^{-۳}$ و *JCP* به‌صورت زیر خواهد بود:

$$\pi^J_{\alpha,\beta}(\mu) \propto \mu^{-۳}\exp\left\{\frac{-\alpha}{۲\mu^۲} + \frac{\beta}{\mu}\right\} I_{(۰,\infty)}(\mu).$$

برای $\alpha = ۰$، $\pi^J_{\alpha,\beta}$ دارای توزیع $IG(\frac{۱}{۲}, -\beta)$ خواهد بود که در حالت $\beta < ۰$ یک توزیع پیشینی سره است [۲۵].

در حالت کلّی توزیع پسینی متناظر

$$\pi^J_{\alpha,\beta}(\mu|\boldsymbol{x}) = \pi^J_{\alpha+S,\beta+n}(\mu)$$

[۱۰] Jeffreys Conjugate Prior

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

برای یک خانوادهٔ نمایی 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 $\vartheta=E\,(t(X))$ 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

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

که در آن $\theta_\circ\in(\theta_-,\theta_+)$ [۸۹]، ص ۲۶۵). همچنین در حالتی که $\theta_-=-\infty$ و $\theta_+=+\infty$ در برآورد $\vartheta=E\,(t(X))$ و تحت تابع زیان $L(\vartheta,a)=\frac{1}{\mathrm{Var}(T)}(a-\vartheta)^2$، با قرار دادن $\lambda=\circ$ در قسمت بالا به‌راحتی نتیجه گرفته می‌شود که $t(X)$ یک برآوردگر مجاز و در پی آن از این‌که 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 $X_n,\ldots,X_1$ نمونهٔ 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 نمایی با پارامتر مقیاس کران‌دار و تحت تابع زیان ناوردای مقیاسی برای توانی از پارامتر برآوردگر مجاز و مینیماکس پیدا می‌شود.

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

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

تعریف می‌شود. در برخی مواقع این مدل به‌صورت زیر

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

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

برآوردگرهای خطی مجاز برای $\theta=\tau^r$ تحت زیان آنتروپی در خانوادهٔ نمایی کلّی (۸) در رابطهٔ [۷۹] به‌دست‌آمده است. در این‌جا به‌عنوان نمونه در خانوادهٔ نمایی در رابطهٔ (۸) برآوردگر مجاز و مینیماکس برای $\theta=\tau^r$ که $\theta\in[a,+\infty)$ $(a>\circ)$ محاسبه می‌گردد.

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

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

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

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

و علاوه بر آن مخاطرهٔ بیزی این برآوردگر متناهی است و

$$\lim_{m\to\infty}r_m\,(\delta_m)=\tfrac{1}{\nu+1}.$$

همچنین ثابت می‌شود که:

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

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

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

$T(\boldsymbol{x}) = \frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i}$ و $c(\boldsymbol{x},n) = \left(\prod_{i=1}^{n} 2\pi x_i^3\right)^{-1/2}$ تابع چگالی توأم نمونهٔ تصادفی برای $i = 1,\ldots,n$ به‌صورت ذیل خواهد بود:

$$f(\boldsymbol{x}\,|\lambda) = \left(\prod_{i=1}^{n} 2\pi x_i^3\right)^{-1/2} \lambda^{-n/2} e^{-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}},\ x_i > 0.$$

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

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

در ادامه نمونهٔ تصادفی $X_1,\ldots,X_n$ از یک خانوادهٔ نمایی با تابع چگالی توأم نمونه به‌صورت زیر را در نظر بگیرید:

$$f(\boldsymbol{x}\,|\eta) = e^{a(\boldsymbol{x})b(\eta)+c(\eta)+h(\boldsymbol{x})}. \qquad (۹)$$

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

$$2c'(\eta)b(\eta) = kb'(\eta) \qquad (۱۰)$$

که در آن k مثبت و به η بستگی ندارد. در حالت خاص که k یک عدد طبیعی باشد، $-2a(\boldsymbol{X})b(\eta)$ دارای توزیع χ^2_k خواهد بود. خانواده‌های نمایی (۹) که در شرط (۱۰) برای $k \in \mathbb{N}$ صدق کنند را توزیع‌های تبدیل‌یافته به خی‌دو 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 معادلهٔ دیفرانسیل (۱۰)، داریم، $c(\eta) = \frac{k}{2}\ln|b(\eta)| + k_1$ و در پی آن

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

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

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

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

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

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

بنا بر این اگر شرط (۱۰) برقرار باشد، خانوادهٔ یک‌پارامتری (۹) به‌شکل خانوادهٔ نمایی با پارامتر مقیاس (۸) با $\nu = k/2$، $T(\boldsymbol{x}) = a(\boldsymbol{x})$ 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]:

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

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

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

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

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

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

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

[11] Uniformly Most Powerful Test

۷ خانواده‌های نمایی و آزمون‌های فرضیهٔ بهینه

آزمون‌های فرضیهٔ آماری یکی از شاخه‌های مورد توجه در استنباط آماری به حساب می‌آید و آزمون‌های بهینه مثل پرتوان‌ترین آزمون یکنواخت (UMPT[۱۱])، پرتوان‌ترین آزمون یکنواخت ناااریب (UMPUT[۱۲]) و آزمون نسبت درست‌نمایی (LRT[۱۳]) 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 $X_n,\ldots,X_1$ یک نمونهٔ تصادفی با تابع چگالی احتمال یا تابع جرم احتمال از یک خانوادهٔ 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 آمارهٔ بسندهٔ کامل $T(\boldsymbol{X})=\sum_{i=1}^{n}t(X_i)$ دارای خاصیت MLR[۱۴] است. در این صورت تابع آزمون UMPT برای آزمون فرضیهٔ آماری $H_\circ:\ \theta\le\theta_\circ$ در مقابل $H_1:\theta>\theta_\circ$ برابر است با

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

و $\alpha=P_{\theta_\circ}(T>k)+\gamma P_{\theta_\circ}(T=k)$.

فرض کنید بازپارامتریدن به‌صورت زیر باشد:

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

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

در حالتی‌که با آزمون‌های دوطرفه $H_\circ:\ \theta_1\le\theta\le\theta_2$ در مقابل $H_1:\theta<\theta_1\ or\ \theta>\theta_2$ روبرو هستیم قضیه زیر می‌تواند راه‌گشا باشد:

قضیهٔ 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

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

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

$$\varphi' E_\theta\varphi'(X)-E_\theta\varphi(X)\begin{cases}\le\circ & \theta_1<\theta<\theta_2\\ \ge\circ & \theta<\theta_1\,,\ \theta>\theta_2.\end{cases}$$

همچنین φ' یک *UMPU* در اندازهٔ 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 $H_\circ:\theta=\theta_\circ$ در مقابل $H_1:\theta\ne\theta_\circ$ روبرو هستیم قضیه زیر می‌تواند راه‌گشا باشد:

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

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

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

$$\tfrac{d}{d\theta}E_\theta\varphi'(X)\,|_{\theta=\theta_\circ}=\tfrac{d}{d\theta}E_\theta\varphi(X)\,|_{\theta=\theta_\circ}$$

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

$$E_\theta\varphi'(X)\ge E_\theta\varphi(X).$$

[۱۲] Uniformly Most Powerful Unbiased Test

[۱۳] Likelihood Ratio Test

[۱۴] Monotone Likelihood Ratio

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

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

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

از این‌که (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 خانوادهٔ نمایی یک‌پارامتری است. بنا بر این UMPU در اندازهٔ α در برای آزمون فرضیهٔ آماری $H_0 : \theta \le \theta_0$ 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 به‌ازای هر u، $E_{\theta_0}[\varphi(T,U)\,|U=u] = \alpha$. برای آزمون‌های دوطرفه $H_0 : \theta_1 \le \theta \le \theta_2$ $(H_0 : \theta = \theta_0)$ در مقابل $H_1 : \theta < \theta_1 \; or \; \theta > \theta_2$ $(H_1 : \theta \ne \theta_0)$، UMPU در اندازهٔ α برابر است با:

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

که در آن $k_i(u)$ و $\gamma_i(u)$ به گونه‌ای محاسبه می‌شوند که به‌ازای 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 مشاهدهٔ مثال در این زمینه به مرجع‌های لیمان و رومانو ([۶۳]، صفحه‌های ۱۲۴-۱۱۹) و شائو ([۸۹]، صفحه‌های ۴۰۶-۴۱۷) مراجعه شود.

با توجه به ارتباط فاصلهٔ اطمینان و آزمون‌های فرضیهٔ آماری می‌توان با استفاده از توابع آزمون بالا فواصل اطمینان مناسب برای پارامترهای جامعه پیدا کرد. به‌عنوان نمونه خانوادهٔ نمایی یک‌پارامتری زیر در نظر بگیرید:

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

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

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

برابر است با:

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

که در آن $k(\theta_0) = k$. به‌راحتی ثابت می‌شود $k(\theta)$ اکیداً صعودی است. بنا بر این $C(\boldsymbol{X}) = \left[\underline{\theta}(\boldsymbol{X}), \infty\right)$ یک فاصلهٔ اطمینان یک‌طرفه برای θ در سطح معنی‌داری $1-\alpha$ است که در آن $\underline{\theta}(X) = \inf\{\theta : T(x) \le k(\theta)\}$. به‌طور مشابه برای آزمون فرضیهٔ آماری $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta < \theta_0$، فاصلهٔ اطمینان یک‌طرفه برای θ در سطح معنی‌داری $1-\alpha$ به‌صورت $C(\boldsymbol{X}) = \left(-\infty, \bar{\theta}(\boldsymbol{X})\right]$ است که در آن $\bar{\theta}(\boldsymbol{X}) = \sup\{\theta : T(\boldsymbol{x}) \ge k(\theta)\}$. به‌طور مشابه متناظر با آزمون‌های دوطرفه نیز می‌توان فاصلهٔ 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 خانوادهٔ نمایی مطالب ارزنده‌ای را پیدا کرد. در [۲۳] ثابت می‌شود که اگر $X_1, \ldots, X_n$ یک نمونهٔ تصادفی با تابع چگالی احتمال یا تابع جرم احتمال از یک خانوادهٔ نمایی

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = c(\boldsymbol{\theta})\exp(\boldsymbol{\theta}.\boldsymbol{x})$$

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

باشد، به‌ازای هر $\varepsilon > ۰$ و $n \in \mathbb{N}$:

$$G_n = \left\{ \boldsymbol{\lambda} \in \Theta : \frac{\int_\Theta \prod_{i=۱}^n f_{\boldsymbol{\theta}}(X_i) dF(\boldsymbol{\theta})}{\prod_{i=۱}^n f_{\boldsymbol{\lambda}}(X_i)} < \varepsilon \right\}$$

یک دنباله از فاصله‌های اطمینان محدب در سطح $۱ - \varepsilon^{-۱}$ است و به‌ازای هر $\boldsymbol{\theta}_۰ \in \Theta$، $\lim_{n\to\infty} \rho(G_n, \boldsymbol{\theta}_۰) = ۰$ که در آن $\rho(G, x) = \sup\{|y - x| : y \in G\}$.

در [۲۸] ثابت می‌شود آزمون نسبت درست‌نمایی (LRT) برای فرضیهٔ ساده در مقابل مرکب برای پارامترهای یک خانوادهٔ نمایی چندپارامتری مجاز است. برای این منظور فرض کنید X_iها ($i = ۱, \ldots, q$) متغیرهای تصادفی مستقل از توزیع زیر باشند:

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

تابع چگالی توأم $(X_۱, \ldots, X_q)$ به‌صورت قضیهٔ بالا که در آن $A(\boldsymbol{x}) = \prod_{i=۱}^q A(x_i)$، $B(\boldsymbol{\theta}) = \prod_{i=۱}^q B(\theta_i)$ باشند، اگر $\theta_{i۰}$ها نقاط درونی Θ_iها باشند برای $i = ۱, \ldots, q$، می‌خواهیم LRT برای آزمون فرضیهٔ آماری $H_۰ : \theta = \theta_{۱۰}, \ldots, \theta_{q۰}$ در مقابل $H_۱ : \theta \neq \theta_{۱۰}, \ldots, \theta_{q۰}$ پیدا کنیم. بدون این‌که به کلّیت مسئله لطمه‌ای وارد شود با تبدیل مناسب می‌توان فرض کرد به‌ازای هر $i = ۱, \ldots, q$، $\theta_{i۰} = ۰$. بنا بر این LRT برای آزمون فرضیهٔ آماری به‌صورت زیر خواهد بود:

$$\varphi(\boldsymbol{x}) = \begin{cases} ۱ & \lambda(\boldsymbol{x}) > k \\ ۰ & o.w. \end{cases}$$

و

$$\lambda(\boldsymbol{x}) = \frac{L(\boldsymbol{x}, \hat{\theta}_۱, \ldots, \hat{\theta}_q)}{L(\boldsymbol{x}, ۰)}$$

که در آن $\hat{\theta}_i$ها MLE برای θ_iها هستند که در آن $i = ۱, \ldots, q$.

با توجه به این‌که

$$\ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i=۱}^q \ln B(\theta_i) + \sum_{i=۱}^q \ln A(x_i) + \sum_{i=۱}^q \theta_i x_i$$

داریم:

$$\frac{\partial}{\partial \theta_i} \ln L(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{B'(\theta_i)}{B(\theta_i)} + x_i = ۰ \Leftrightarrow \theta_i = \theta_i^* = H(x_i)$$

که در آن $H(.)$ وارون تابع

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

است. به‌راحتی می‌توان نشان داد $h'(\boldsymbol{\theta}) = Var_{\boldsymbol{\theta}}(X) > ۰$ و در پی آن h یک تابع اکیداً صعودی و لذا H یکتاست. برای اثبات این‌که برای $i = ۱, \ldots, q$، $\hat{\theta}_i = H(x_i)$ تابع $L(x, \boldsymbol{\theta})$ را ماکسیمم می‌کند از این‌که $\frac{\partial^۲}{\partial \theta_i^۲} L(\boldsymbol{x}, \boldsymbol{\theta}) = -Var_{\boldsymbol{\theta}}(X) < ۰$ و $\frac{\partial^۲}{\partial \theta_i \partial \theta_j} L(\boldsymbol{x}, \boldsymbol{\theta}) = ۰$ برای $i \neq j$ نتیجه می‌گیریم که ماتریس مشتقات جزئی مرتبهٔ دوم یک ماتریس معین نامنفی باشد و از این رو نقطهٔ زیر

$$(\hat{\theta}_۱, \ldots, \hat{\theta}_q) = (H(x_۱), \ldots, H(x_q))$$

یک MLE برای $(\theta_۱, \ldots, \theta_q)$ خواهد بود. بنا بر این:

$$\lambda(\boldsymbol{x}) = \frac{L(\boldsymbol{x}, \hat{\theta}_۱, \ldots, \hat{\theta}_q)}{L(x, ۰)} = \prod_{i=۱}^q \left(\frac{B\left(H(x_i)\right)}{B\left(۰\right)} \right) \exp\left(\sum_{i=۱}^q x_i H(x_i) \right)$$

و در پی آن LRT به‌صورت زیر خواهد بود:

$$\varphi(\boldsymbol{x}) = \begin{cases} ۱ & \prod_{i=۱}^q B\left(H(x_i)\right) \exp\left(\sum_{i=۱}^q x_i H(x_i)\right) > k \\ ۰ & o.w. \end{cases} \quad (۱۱)$$

قضیه‌های زیر می‌تواند برای اثبات مجاز بودن آزمون‌ها مورد استفاده قرار گیرند.

قضیهٔ ۳.۷. اگر $f(x\,|\theta) = h(x)c(\theta)\exp[\theta x]$ و تابع زیان برای $\theta_۱ < \theta_۲$ در نامساوری زیر صدق کند:

$$L(\theta, a_۱) - L(\theta, a_۰) \begin{cases} \geq ۰ & \theta_۱ < \theta < \theta_۲ \\ \leq ۰ & \theta < \theta_۱ \; or \; \theta > \theta_۲ \end{cases}$$

کلاس آزمون‌های دوطرفه

$$\varphi'(x) = \begin{cases} ۱ & x < x_۱ or x > x_۲ \\ \gamma_i & x = x_i, i = ۱, ۲ \\ ۰ & x_۱ < x < x_۲ \end{cases}$$

اساساً کامل است. اگر علاوه بر شرایط بالا $\theta'_۱ < \theta'_۲ < \theta'_۳$ وجود داشته باشند که

$$L(\theta'_i, a_۱) - L(\theta'_i, a_۰) \begin{cases} < ۰ & i = ۱, ۳ \\ > ۰ & i = ۲ \end{cases}$$

هر آزمون دوطرفه

$$\varphi'(x) = \begin{cases} ۱ & x < x_۱ \, or \, x > x_۲ \\ \gamma_i & x = x_i, \; i = ۱, ۲ \\ ۰ & x_۱ < x < x_۲ \end{cases}$$

مجاز است([۳۰]، ص ۲۱۲).

قضیهٔ ۴.۷. فرض کنید $\boldsymbol{X}$ دارای توزیع

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

باشد که در آن $\boldsymbol{\theta}\in\Theta\subseteq\mathbb{R}^q$ شرط لازم و کافی برای این‌که تابع آزمون φ برای آزمون فرضیهٔ آماری $H_0:\theta_1=\ldots=\theta_p$ در مقابل نفی آن‌که $p\le q$ مجاز باشد آن است که

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

که در آن $S_p=(x_1,\ldots,x_q)$ و $t_P=(x_{p+1},\ldots,x_q)$ و به‌ازای هر t_P، $A(t_P)$ یک مجموعهٔ محدب بسته در $\mathbb{R}^p$ است [۲۶].

در [۲۸] به‌کمک قضیهٔ ۴.۷ نشان داده می‌شود که برای آزمون فرضیهٔ آماری $H_0:\theta=\theta_{10},\ldots,\theta_{q0}$ در مقابل $H_1:\theta\neq\theta_{10},\ldots,\theta_{q0}$، LRT که با استفاده از رابطهٔ (۱۱) محاسبه شد یک آزمون مجاز نیز هست.

در پایان در مورد آزمون نیکویی برازش در خانوادهٔ نمایی به‌طور خلاصه مطالبی بیان می‌شود [۳۶]. اگر $X_1,\ldots,X_n$ نمونهٔ تصادفی با تابع توزیع F_0 باشد، می‌خواهیم آزمون فرضیهٔ آماری مرکب $H_0: F_0\in F=\{F_{\boldsymbol{\theta}}(x):\boldsymbol{\theta}\in\Theta\subseteq\mathbb{R}^q\}$ را انجام دهیم. فرض کنید مقدار واقعی پارامتر $\boldsymbol{\theta}_0$ باشد، یعنی تحت فرضیهٔ H_0، تابع توزیع واقعی $F_0(x)=F_{\boldsymbol{\theta}_0}(x)$ است. روش‌های متعددی برای بررسی این آزمون به کار گرفته شده است. در یکی از این روش‌ها می‌توان آمارهٔ آزمون را بر حسب تابع توزیع تجربی و توابع چندک نوشت و در پی آن با استفاده از خواص مجانبی، توزیع‌های حدی را پیدا کرد. اغلب پارامتر مجهول تحت فرضیهٔ H_0، توسط دنباله‌ای از برآوردگرها مثل $\hat{\boldsymbol{\theta}}_n$ برآورد می‌شود. در این حالت به جای $F_{\boldsymbol{\theta}}(x)$ از تابع توزیع برآوردشده یعنی $F_{\hat{\boldsymbol{\theta}}}(x)$ استفاده می‌شود. از دیگر روش‌های پیشنهادی تصادفی‌کردن و بازنمونه‌گیری است. در مقاله [۳۶] روش جدیدی برای حالتی که $F_{\boldsymbol{\theta}}(x)$ به‌ازای هر $\boldsymbol{\theta}\in\Theta$ در x پیوسته است پیشنهاد شده که در آن از متغیرهای تصادفی مستقل $U_1=F_{\boldsymbol{\theta}_0}(X_1),\ldots,U_n=F_{\boldsymbol{\theta}_0}(X_n)$ که دارای توزیع $U(0,1)$ هستند، استفاده می‌شود. با جایگزینی $\hat{\boldsymbol{\theta}}_n$ به جای پارامتر مجهول $\boldsymbol{\theta}_0$، متغیرهای تصادفی $\hat{U}_1=F_{\hat{\boldsymbol{\theta}}_n}(X_1),\ldots,\hat{U}_n=F_{\hat{\boldsymbol{\theta}}_n}(X_n)$ و در پی آن آماره‌های ترتیبی $U_1,\ldots U_n$ و $\hat{U}_1,\ldots,\hat{U}_n$ که به‌ترتیب برابر $U_{(1)},\ldots,U_{(n)}$ و $\hat{U}_{(1)},\ldots,\hat{U}_{(n)}$ هستند را در نظر می‌گیریم. به‌راحتی ثابت می‌شود اگر $k(n),\ n-k(n)\to\infty$، آن‌گاه

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

به‌طور مجانبی دارای توزیع نرمال استاندارد است.

می‌توان توزیع حدی

$$Z_n=\sqrt{n^3}\max_{1\le k<n}\left|\frac{U_{(k(n))}-\frac{k(n)}{n}}{\sqrt{k(n)\,(n-k(n))}}\right|$$

و

$$\hat{Z}_n=\sqrt{n^3}\max_{1\le k<n}\left|\frac{\hat{U}_{(k(n))}-\frac{k(n)}{n}}{\sqrt{k(n)\,(n-k(n))}}\right|$$

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

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

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

که در آن D به پارامتر مجهول $\boldsymbol{\theta}$ بستگی ندارد.

۲. مشتقات جزئی $c(\boldsymbol{\theta})$ و $w_i(\boldsymbol{\theta})$ ها نسبت به θ_j که در آن $j=1,\ldots,q$ تا مرتبهٔ سوم وجود داشته باشند و در x و یک همسایگی حول $\boldsymbol{\theta}_0$ به‌طور یکنواخت کران‌دار باشند.

۳. $\max_{1\le i<q}\int_{-\infty}^{+\infty}(t_i(x))^4 f(x|\boldsymbol{\theta}_0)dx<\infty$.

۴. $\sqrt{n}\left\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right\|=O_P(1)$ که در آن $\|.\|$ نرم ماکسیمم در $\mathbb{R}^q$ است.

در [۳۶] نشان داده می‌شود Z_n و $\hat{Z}_n$ داری توزیع مجانبی یکسانی هستند و همچنین تحت فرضیهٔ H_0 و برقراری شرایط نظم بالا به‌ازای هر x داریم:

$$\lim_{n\to\infty}P\left\{a\,(\log(n))\,\hat{Z}_n\le x+b\,(\log(n))\right\}=\exp\left(-2e^{-x}\right)$$

که در آن

$$a\ (x) = \sqrt{۲\log(x)},$$

$$b\ (x) = ۲\log(x) + \tfrac{۱}{۲}\log\left(\log(x)\right) - \tfrac{۱}{۲}\log(\pi).$$

با توجه به این حقیقت فرضیهٔ $H_{\circ}$ را برای مقادیر بزرگ $a\ (\log(n))\,\hat{Z}_n - b\ (\log(n))$ رد می‌کنیم. اکنون فرض کنید تحت فرضیهٔ مقابل، x و $\delta > ۰$ وجود داشته باشد به گونه‌ای که

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

که رابطهٔ اخیر برای n بزرگ و $k(n) = [nF_{\circ}(x_{\circ})] + ۱$ منجر به عبارت زیر می‌شود:

$$P\left\{\left|\hat{U}_{(k(n))} - U_{(k(n))}\right| > \tfrac{۱}{۲}\delta\right\} > ۱ - \delta$$

بنا بر این به‌ازای هر $\varepsilon \in (۰,۱)$، $\delta = \delta(\varepsilon) > ۰$ وجود دارد

به‌طوری که

$$\liminf_{n\to\infty} P\left\{\tfrac{۱}{\sqrt{n}}\hat{Z}_n \geq \delta\right\} \geq \varepsilon$$

و این حقیقت سازگاری مجانبی آزمون داده شده را اثبات می‌کند.

۸ نتیجه‌گیری

در این مقاله مروری، تاریخچه‌ای از کارهای انجام شده راجع به خانوادهٔ نمایی که توسط محققان در شاخه‌های مختلف آمار انجام‌شده ارائه شد و مهم‌ترین کاربردهایی از این خانواده که در استنباط آماری از آنها استفاده می‌شود به تفضیل شرح داده شد. علاقه‌مندان می‌توانند از این مقاله به‌عنوان شروعی برای تحقیقات مرتبط با خانوادهٔ نمایی بهره گیرند و همچنین کاربردهای این ردهٔ مهم در شاخه‌های 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 Tourneret, 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 Horvhth, 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 Multivariats 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.