قانون بن‌فورد و کاربردهای آن

افشین فلاح [۱] محمدرضا دانش‌زاد

چکیده:

در این مقاله، قانون بن‌فورد به عنوان یکی از قواعد جذاب نظریه احتمال مورد توجه قرار گرفته است. برخلاف تصور عمومی مبنی بر هم‌شانس بودن اعداد برای قرار گرفتن در رقم‌های مختلف، این قانون نشان می‌دهد که در مشاهدات حاصل از بسیاری پدیده‌های طبیعی، اعداد به صورتی خاص و غیر هم‌شانس در موقعیت‌های مختلف توزیع شده‌اند. قانون بن‌فورد توزیعی گسسته و نامتقارن برای ارقام معنی‌دار مشاهدات حاصل از شمارش یا برگرفته از رویدادهای طبیعی ارائه می‌دهد و دارای کاربردهای بسیار متنوعی در زمینه‌های مختلف است.

واژه‌های کلیدی: قانون بن‌فورد، دنباله اعداد، رقم آغازین، رقم معنی‌دار، پایایی.

۱ مقدمه

غالبا تصور می‌شود که رقم اول سمت چپ اعداد به عنوان یک متغیر تصادفی از توزیع یکنواخت گسسته پیروی می‌کند، یعنی اعداد ۱، ۲، ...، ۹ دارای احتمال یکسان $\frac{1}{9}$ برای قرار گرفتن در رقم آغازین مشاهدات مربوط به پدیده‌های مختلف هستند. امّا در بسیاری از مواقع رقم آغازین دارای توزیعی گسسته و چوله به راست است، که در آن بیشترین احتمال متعلق به رقم ۱ و کمترین احتمال متعلق به رقم ۹ است. بر اساس قانون بن‌فورد (۱۹۳۸) که به قانون رقم آغازین نیز معروف است، در مشاهداتی که در بسیاری از پدیده‌های طبیعی بدست می‌آیند، رقم آغازین بصورتی خاص توزیع شده است. این قانون به ظاهر عجیب برای بسیاری از داده‌ها برقرار است، که برای نمونه می‌توان به صورت حساب‌های مشترکین برق، قیمت سهام، تعداد جمعیت، آمار مرگ و میر، تعداد آرای نامزدهای انتخاباتی در حوزه‌ها و حتی صندوق‌های مختلف رای‌گیری، طول یا پهنای رودخانه‌ها، صورت حساب‌های مالی و اظهارنامه‌های مالیاتی، ثابت‌های موجود در علومی نظیر فیزیک و ریاضیات، سری‌های اعداد نظیر سری اعداد فیبوناچی، وزن مولکولی مواد و نیمه عمر مواد رادیواکتیوی، اعداد ارائه شده در روزنامه‌ها و مجلات، اشاره کرد. در نیم قرن گذشته مقالات متعددی در نشریات معتبر بین‌المللی در مورد قانون بن‌فورد چاپ شده است، که تعداد زیادی از این مقالات به کاربردهای گوناگون این قانون اختصاص دارد (دورتسکی و همکاران، ۲۰۰۴). این قانون در راستی‌آزمایی اطلاعات، کشف تقلب و اختلاس، فرارهای مالیاتی، شناسایی اشتباه یا تخلف در داده‌های حسابداری، بررسی درستی نتایج انتخابات و حتی کشف باگ‌های موجود در برنامه‌های کامپیوتری مورد استفاده قرار می‌گیرد. در کشورهای در حال توسعه که معمولاً اطلاعات از طریق نظرخواهی جمع‌آوری می‌شوند، همواره دغدغه درستی اطلاعاتی که در اختیار دولت‌مردان قرار می‌گیرد، وجود دارد. اهمیت این موضوع از آن جهت بیشتر می‌شود که نتایج مربوط به این اطلاعات پایه تصمیم‌گیری‌ها و سیاست‌گذاری‌های اجتماعی و اقتصادی دولت‌ها قرار می‌گیرند. هرچند برخی از آزمون‌هایی که بر پایه قانون بن‌فورد طراحی شده‌اند بسیار پیچیده‌اند، اما در بسیاری موارد کاربست این قانون به سادگی امکان پذیر است. کافی است رقم آغازین تعداد کمی از مشاهدات عدد ۱ و تعداد زیادی از مشاهدات عدد ۹، ۸ یا حتی ۵

[۱] گروه آمار، دانشگاه بین المللی امام خمینی

باشند، تا بتوان با اطمینان بالایی وجود انحراف یا غیرواقعی بودن این مشاهدات را تایید کرد. البته در برخی کاربردها به دلیل تاثیرگذاری عوامل مخدوش کننده، امکان استفاده از این قانون وجود ندارد. نیوکامب (۱۸۸۱)برای نخستین بار پی برد که ارقام آغازین اعداد دارای توزیع یکنواختی نیستند. هنگامی که وی برای انجام یک سری محاسبات مشغول استفاده از کتاب‌های لگاریتم بود، متوجه شد که صفحات شامل لگاریتم‌هایی که با رقم ۱ شروع می‌شوند، نسبت به دیگر صفحات بیشتر فرسوده شده‌اند. وی پس از بررسی‌های بیشتر به این موضوع پی‌برد که در بسیاری از پدیده‌های طبیعی اعداد کوچکتر شانس بیشتری برای ظاهر شدن در رقم آغازین مشاهدات دارند. وی به صورت تجربی قاعده‌ای پیدا کرد که به خوبی با مشاهداتش مطابقت داشت (هیل، ۱۹۹۸). بر اساس این قاعده، درصد اعدادی که با رقم d شروع می‌شوند، برابر

$$\log_{10}(1+\tfrac{1}{d})$$

است. اما نیوکامب هیچگاه موفق به ارائه تفسیر یا اثباتی برای این حدس نشد و در مواجهه با عدم توجه عمومی نسبت به موضوع، خیلی زود آن را به فراموشی سپرد (لیمس و همکاران، ۲۰۰). در سال ۱۹۳۸ فرانک بن‌فورد فیزیک‌دان شرکت جنرال الکتریک در مطالعاتی مستقل به الگوی مشابهی دست یافت، وی مجموعه‌ای بزرگ بالغ بر ۲۰,۲۲۹ مشاهده از پدیده‌های مختلف و اغلب نامربوط مانند اطلاعات مربوط به بازی بیس‌بال، پهنا و طول رودخانه‌ها، صورت‌حساب‌های برق، وزن مولکولی مواد، شماره خیابان محل سکونت و غیره را جمع‌آوری نمود و پیروی آنها از قانون بن‌فورد را مورد بررسی قرار داد. داده‌هایی که بن‌فورد مورد بررسی قرار داد با تقریب بسیار خوبی منطبق با قاعده‌ای بودند که نیم قرن پیش نیوکامب بدست آورده بود. البته بن‌فورد نیز نتوانست هیچگونه استدلالی برای درستی این قانون ارائه دهد (لاوو و همکاران، ۱۹۹۹). اولین قدم برای شرح این قانون در سال ۱۹۶۱ توسط پینکام ریاضی‌دان آمریکایی برداشته شد. اما وی تنها موفق به اثبات دلیل وجودی این قانون شد و نتوانست تحلیلی ریاضی برای آن ارائه کند. در نهایت هیل (۱۹۹۵b) این قانون را برای ارقام اول و دوم اعداد ثابت نمود (والسو و همکاران، ۱۹۹۹). در این مقاله قانون بن‌فورد به عنوان یکی از قواعد جذاب نظریه احتمال مورد توجه قرار گرفته است. در بخش دوم مبانی نظری قانون بن‌فورد بر اساس نظریه توزیع‌ها مورد بررسی قرار گرفته و برخی ویژگی‌های احتمالاتی آن تشریح شده است. استدلالهای ارائه شده برای قانون بن‌فورد و دامنه اعتبار این قانون بطور اجمالی در بخش سوم بررسی شده است. در بخش چهارم مثال‌های متنوعی از موارد کاربرد این قانون برای اهداف مختلف ارائه شده است.

۲ مبانی نظری قانون بن‌فورد

اگر متغیر تصادفی D_1 نشان دهنده‌ی اولین رقم یک عدد از سمت چپ باشد، در این صورت بر اساس قانون بن‌فورد D_1 دارای توزیعی گسسته و چوله به راست به صورت

$$P_{D_1}(d_1) \;=\; \log_{10}(1+\frac{1}{d_1}), \quad d_1 = 1,\ldots,9, \quad (1)$$

است. این قانون تنها محدود به اولین رقم معنی‌دار نبوده و می‌توان آن را به ارقام بالاتر تعمیم داد (هیل، ۱۹۹۵). فرض کنید $D_i\ ,\ i = 1,2,3,\ldots$ نشان دهنده $i-$امین رقم معنی‌دار از سمت چپ در اعداد در پایه ۱۰ باشند (به عنوان نمونه، $D_1(314) = 3$ ، $D_2(314) = 1$ ، $D_3(314) = 4$). در این صورت، به ازای عدد صحیح و مثبت k توزیع توام $(D_1,\ldots,D_k)$ به صورت

$$\begin{array}{l} P_{(D_1,D_2,\ldots,D_k)}(d_1,d_2,\ldots,d_k) = \\ \log_{10}\left[1+\left(\sum_{i=1}^{k} d_i\times 10^{k-i}\right)^{-1}\right], \quad j=2,3,\ldots,9; \quad (2) \\ \qquad d_j\in\{0,1,2,\ldots,9\};\ d_1\in\{1,2,\ldots,9\} \end{array}$$

خواهد بود.

به عنوان مثال، بر اساس این قانون احتمال مشاهده‌ی عددی مانند ۰/۳۱۴ برابر

$$\begin{aligned} P_{(D_1,D_2,D_3)}(3,1,4) &= \log_{10}\left[1+\frac{1}{314}\right] \\ &\simeq 0/014, \end{aligned}$$

است. به همین ترتیب، می‌توان احتمال اینکه دومین رقم برابر ۲ باشد را بدست آورد. چون اولین رقم یکی از اعداد ۱ ، ۲، ۳، ...، ۹ است، پس عددی که دومین رقم آن ۲ می‌باشد، با یکی از اعداد ۱۲ ، ۲۲، ۳۲، ...، ۹۲ آغاز می‌شود. از این رو، احتمال اینکه دومین رقم عدد ۲ باشد برابر با

$$\begin{aligned} P_{D_2}(2) &= \log_{10}(1+\frac{1}{12})+\ldots+ \\ &\quad \log_{10}(1+\frac{1}{92}) \simeq 0/109 \end{aligned}$$

خواهد بود. با توجه به این مثال، می‌توان رابطه‌ی

$$\begin{aligned} P_{D_2}(d_2) &= \sum_{k=1}^{9}\log_{10}\left(1+\frac{1}{10k+d_2}\right) \\ & d_2 = 0,1,\ldots,9, \qquad\qquad (3) \end{aligned}$$

را برای توزیع احتمال دومین رقم معنی‌دار ارائه نمود. در حالت کلی‌تر می‌توان رابطه‌ی

$$P(R \leq \frac{t}{۱۰}) = \log_{۱۰} t \qquad , t \in [۱, ۱۰), \tag{۴}$$

را برای این قانون ارائه کرد، که در آن منظور از R جزء اعشاری اعداد است (هیل، ۱۹۹۵). اگر x عددی در پایه ۱۰ باشد، آن‌گاه رابطه $x = R \times ۱۰^n$ برقرار خواهد بود، که در آن R عددی منحصر به فرد در فاصله $[\frac{۱}{۱۰}, ۱)$ است. به عنوان مثال، جزء اعشاری هر دو عدد ۳۱۴ و ۰/۰۳۱۴ برابر ۰/۳۱۴ است. یک نتیجه‌ی جالب از رابطه (۲) این است که ارقام معنی‌دار مستقل نیستند. یعنی احتمال مشاهده‌ی اعداد مختلف در رقم‌های مختلف به هم وابسته‌اند. به عنوان مثال، احتمال اینکه دومین رقم معنی‌دار ۲ باشد، به صورت غیر شرطی و با استفاده از رابطه (۳) تقریباً برابر ۰/۱۰۹ است. اما احتمال شرطی ۲ بودن دومین رقم به شرط ۱ بودن رقم اول، تقریباً برابر

$$\begin{aligned} P_{D_۲|D_۱} &= \frac{P_{(D_۱,D_۲)}(۱,۲)}{P_{D_۱}(۱)} \\ &= \frac{\log_{۱۰}(۱+(۱۲)^{-۱})}{\log_{۱۰}(۱+(۱)^{-۱})} \\ &\simeq ۰/۱۱۵ \end{aligned}$$

است. بر اساس رابطه (۲) این وابستگی با دور شدن موقعیت رقم از سمت چپ عدد کاهش می‌یابد. توزیع $k-$امین رقم معنی‌دار هنگامی که k به سمت بی‌نهایت میل می‌کند، بسیار نزدیک به یکنواخت گسسته روی مجموعه $\{۰, ۱, ۲, \ldots, ۹\}$ است. همان‌گونه که در جدول۱ مشاهده می‌کنید، هرچه موقعیت قرار گرفتن رقم در اعداد افزایش می‌یابد، توزیع آن‌ها به توزیع یکنواخت نزدیک‌تر می‌شود. برای استفاده از این قانون در پایه دیگری نظیر $b(>۱)$، روابط (۱) و (۴) را می‌توان به ترتیب به صورت

$$P_{D_۱}(d_۱) = \log_b(۱ + \frac{۱}{d_۱}),\ d_۱ = ۱, \ldots, b-۱$$
$$P\left(R(base(b)) \leq \frac{t}{b}\right) = \log_b t,\ t \in [۱, b),$$

بازنویسی نمود.

۳ استدلال‌های ارائه شده

پینکام (۱۹۹۶) با این استدلال که هر قانونی برای رقم اول می‌بایست در برابر مقیاس اندازه‌گیری پایا باشد، اولین قدم را در اثبات قانون بن‌فورد برداشت. فرض کنید واقعاً قاعده‌ای برای توزیع فراوانی اولین رقم مشاهدات وجود داشته باشد، در این صورت بدیهی است که این قاعده باید فراگیر بوده و به واحد اندازه‌گیری وابسته نباشد. پس اگر قیمت کالاها با واحدهایی نظیر دلار، دینار یا درهم و طول با واحدهایی نظیر یا پا اندازه‌گیری شود، این قانون باید برای اعدادی که از واحدهای مختلف بدست می‌آیند، نتیجه یکسان داشته باشد. وجود احتمال یکسان برای اعداد نشان دهنده نفی پایایی در برابر مقیاس‌های مختلف است. یک مثال مناسب در این زمینه تبدیلات داده‌های مالی از واحدی به واحد دیگر است. فرض کنید با ضرب اعداد با واحد A در ۲، این اعداد به واحد B تبدیل شوند. اگر رقم آغازین عددی در واحد A برابر ۱ باشد، آن عدد در واحد B با احتمالی یکسان با یکی از ارقام ۲ یا ۳ آغاز می‌شود. امّا اگر عدد در واحد A با یکی از اعداد ۵، ۶، ۷، ۸ یا ۹ آغاز شود، مطمئناً در واحد B با رقم ۱ آغاز خواهد شد. این مثال نشان می‌دهد که احتمال عدد ۱ برای قرار گرفتن در رقم آغازین بیشتر از بقیه است. از این رو توزیع یکنواخت نیست. برای بررسی قانون رقم آغازین برای اعدادی که به صورت نماد علمی ($۱ \leq x < ۱۰$ ،$x \times ۱۰^n$) بیان می‌شوند، رقم آغازین را d در نظر بگیرید. به سادگی می‌توان توزیعی برای d پیدا کرد که به واحد اندازه‌گیری وابسته نباشد. اگر توزیع مورد نظر به واحد اندازه‌گیری وابسته نبوده و رابطه $Y = \log_{۱۰} X$ برقرار باشد، در این صورت با ضرب عددی در X، در حقیقت مقداری به Y اضافه می‌شود. فرض کنید a عددی است که برای تبدیل یک مقیاس اندازه‌گیری به مقیاس دیگر به کار می‌رود، در این صورت رابطه

$$\begin{aligned} \log_{۱۰} aX &= \log_{۱۰} a + \log_{۱۰} X \\ &= \log_{۱۰} a + Y, \end{aligned}$$

برقرار است. امّا تنها توزیع احتمالی در فاصله $[۰, ۱)$ که پس از افزودن یک مقدار ثابت دلخواه به y بدون تغییر باقی می‌ماند، توزیع یکنواخت است. متغیر تصادفی Y دارای توزیع یکنواخت بین $۰ = \log_{۱۰} ۱$ و $\log_{۱۰} ۱۰ = ۱$ است. بنابراین احتمال اینکه رقم آغازین برابر ۱ باشد را می‌توان به صورت

$$\begin{aligned} P(D = ۱) &= P(۱ \leq X < ۲) \\ &= P(۰ \leq Y < \log_{۱۰} ۲) \\ &\simeq ۰/۳۰۱, \end{aligned}$$

نوشت. در حالت کلی احتمال اینکه رقم آغازین برابر d باشد، به صورت

$$\begin{aligned} P(D=d) &= P(d \leq X < d+1) \\ &= P(\log_{10} d \leq Y < \log_{10}(d+1)) \\ &= \log_{10}(1+\frac{1}{d}), \end{aligned}$$

است. این دقیقاً همان فرمولی است که ابتدا نیوکامب (۱۹۸۸) و بعدها بن‌فورد (۱۹۳۸) به صورت تجربی بدست آورده بودند. با این ایده که اعدادی که از قانون بن‌فورد پیروی می‌کنند، علاوه بر پایایی در برابر مقیاس، باید در برابر پایه اندازه‌گیری نیز پایا باشند و با بهره‌گیری از برخی مفاهیم آنالیز حقیقی، می‌توان اثباتی دقیق برای قانون بن‌فورد ارائه نمود. خواننده می‌تواند به هیل(۱۹۹۵a) و (۱۹۹۵b) مراجعه نماید.

۴ ویژگی‌های توزیع

در این بخش برخی ویژگی‌های توزیع احتمالی بن‌فورد مورد بررسی قرار می‌گیرد.

الف) امید ریاضی و واریانس توزیع

اگر D_k نشان دهنده $k-$امین رقم معنی‌دار باشد، در این صورت امید ریاضی و واریانس متغیر تصادفی D_k را می‌توان به صورت

$$E(D_k) = \sum_{d_k=1}^{9} d_k P_{D_k}(d_k)$$
$$V(D_k) = \sum_{d_k=1}^{9} d_k^2 P(D_k=d_k) - E^2(D_k),$$

نوشت. جدول۲ مقادیر امید ریاضی و واریانس توزیع بن‌فورد را برای ارقام مختلف نشان می‌دهد. میانگین و واریانس توزیع یکنواخت گسسته بر مجموعه‌ی $\{0,1,2,\ldots,\theta\}$ به ترتیب برابر $\frac{\theta}{2}$ و $\frac{\theta(\theta+2)}{12}$ است، از این رو به ازای $\theta=9$ میانگین برابر ۴/۵ و واریانس برابر ۸/۲۵ خواهد بود. ملاحظه می‌شود که در جدول ۲، میانگین D_k به سمت ۴/۵ و واریانس آن به سمت ۸/۲۵ میل می‌کند، که این دقیقاً برابر با میانگین و واریانس توزیع یکنواخت گسسته روی مجموعه‌ی $\{0,1,2,\ldots,9\}$ است. شکل‌های ۱ و ۲ به ترتیب بافت‌نگار توزیع بن‌فورد برای ارقام دوم تا پنجم را نمایش می‌دهند. ملاحظه می‌شود که برای ارقام بالاتر، توزیع بن‌فورد شبیه توزیع یکنواخت گسسته است.

ب) ضریب همبستگی بین ارقام

با توجه به اینکه ارقام معنی‌دار مستقل نیستند، اگر متغیرهای تصادفی D_i و D_j به ترتیب نشان دهنده $i-$امین و $j-$امین رقم معنی‌دار باشند، در این صورت ضریب همبستگی بین آن‌ها به صورت

$$\rho_{D_i,D_j} = \frac{Cov(D_i,D_j)}{\sqrt{Var(D_i)Var(D_j)}} \quad 0 < i < j$$

خواهد بود. در جدول ۳ مقادیر ضریب همبستگی خطی بین متغیرهای تصادفی D_i و D_j برای مقادیر مختلف i و j محاسبه شده است. همانطور که مشاهده می‌شود، هر چه که $|j-i|$ بیشتر می‌شود، میزان همبستگی کاهش می‌یابد.

ج) مجموع تغییرات فاصله‌ای D_k در مقایسه با توزیع یکنواخت

اگر D_k نشان دهنده‌ی kامین رقم معنی‌دار باشد، فاصله D_k با مقادیر مورد انتظار از توزیع یکنواخت گسسته روی مجموعه $\{0,1,2,\ldots,9\}$ به صورت

$$d(D_1,U) = \begin{cases} \frac{1}{2}\sum_{d_1=1}^{9} \left|P_{D_1}(d_1) - \frac{1}{9}\right| & ,k=1 \\ \frac{1}{2}\sum_{d_k=0}^{9} \left|P_{D_k}(d_k) - \frac{1}{10}\right| & ,k \neq 1, \end{cases}$$

تعریف می‌شود، که در آن $d_1 \in \{1,2,\ldots,9\}$ و $d_k \in \{0,1,\ldots,9\}$، $k \neq 1$ است. همانگونه که در جدول ۴ مشاهده می‌شود، مجموع تغییرات فاصله‌ای برای ارقام بالاتر به سمت صفر میل می‌کند، یعنی $\lim_{k\to\infty} d(D_k,U) = 0$. این بیانی دیگر از این واقعیت است که توزیع بن‌فورد برای رقم‌های بالاتر به توزیع یکنواخت همگرا می‌شود.

۵ شبیه‌سازی توزیع

در این بخش نحوه‌ی شبیه‌سازی اعدادی که رقم آغازین آن‌ها از توزیع بن‌فورد پیروی می‌کنند، شرح داده می‌شود. فرض کنید $Y \sim U(0,1)$ و تبدیل $X = 10^Y$ را در نظر بگیرید. اگر رقم آغازین متغیر تصادفی X دارای توزیع بن‌فورد باشد، در این صورت می‌بایست X برای یک مقدار صحیح n در فاصله

$$10^n \times d \leq X < 10^n \times (d+1), \qquad (5)$$

قرار گیرد. رابطه (۵) را می‌توان با لگاریتم‌گیری به صورت

$$n + \log_{10} d \leq Y < n + \log_{10}(d+1),$$

بازنویسی کرد. با توجه به اینکه طول فاصله فوق برابر

$$\log_{10}(d+1)-\log_{10} d=\log_{10}(1+\frac{1}{d}),$$

است و با در نظر گرفتن رابطه (۱)، می‌توان گفت که اگر Y به صورت یکنواخت روی بازه $(0,1)$ توزیع شده باشد، آن‌گاه رقم آغازین $X=10^{Y}$ دارای توزیع بن‌فورد است (فلر، ۱۹۷۱).

۶ دامنه اعتبار قانون بن‌فورد

اعتبار قانون بن‌فورد نیز مانند همه قوانین تابع شرایطی است. از این رو، این قانون همیشه و برای هر نوع مشاهده‌ای صادق نیست. بدیهی است که در پرتاب یک تاس سالم، اعداد تاس از شانس یکسانی برای ظاهر شدن برخوردارند و بنابراین فراوانی مشاهدات اعداد مختلف از توزیع یکنواخت پیروی می‌کند. بطور مشابه در بررسی سن سیاستمداران جهان، عدد ۱ دارای فراوانی زیادی در اولین رقم نیست. همچنین شماره‌های تلفن یک منطقه با عددی خاص و ثابت شروع می‌شوند. به طور کلی، قانون بن‌فورد در مورد مشاهداتی که زور یا مصلحت اندیشی بر آنها موثر باشد، برقرار نیست. به عبارت دیگر برای استفاده از قانون بن‌فورد به داده‌هایی نیاز است، که نه کاملاً تصادفی بوده و نه بر اساس اجبار و مصلحت اندیشی به دست آمده باشند، بلکه داده‌هایی که بر اساس شمارش عادلانه حاصل شده باشند.

۷ چند مثال کاربردی

در این بخش تلاش شده است نحوه‌ی کاربست قانون بن‌فورد با ارائه چند مثال متنوع در زمینه‌های مختلف، تبیین شود. در انتهای بخش، کاربردی عملی از این قانون با استفاده از نتایج انتخابات در ایالت فلوریدای آمریکا، شرح داده شده است. برای بررسی میزان همخوانی مشاهدات مربوط به هر مثال با قانون بن‌فورد از آزمون کای-دو و معیار مجموع تغییرات فاصله‌ای استفاده شده است. آزمون کای‌دو یکی از آزمون‌های بررسی نیکویی برازش یک نمونه‌ی تصادفی به توزیعی مفروض است، که به منظور بررسی میزان مشابهت توزیع برآورد شده و توزیع واقعی بکار می‌رود. برای این منظور از آماره‌ی آزمون

$$\chi^2_{n-1} \quad = \quad \sum_{i=1}^{n}\frac{(E_i-O_i)^2}{E_i}, \qquad (6)$$

استفاده می‌شود، که در آن n حجم نمونه، P_i مقدار احتمال در نقطه $i-$ام توزیع واقعی و $E_i=np_i$ بیانگر تعداد مشاهدات مورد انتظار در نقطه‌ی $i-$ام توزیع است. همچنین O_i نشان دهنده فراوانی مشاهده شده در نقطه $i-$ام است. مقادیر بزرگ آماره، از فرض عدم تطابق مشاهدات با بن‌فورد، به صورت

$$\chi^2_9=\sum_{d_k=0}^{9}\frac{(P_{D_k}(d_k)-\hat{P}_{D_k}(d_k))^2}{P_{D_k}(d_k)}\times n \qquad d_k\in\{0,1,2,\dots,9\},$$

است، که در آن $\hat{P}(D_k=d_k)$ احتمال آنکه $k-$امین رقم برابر d_k باشد را مشخص می‌کند و بر اساس نمونه موجود برآورد می‌شود. آزمون دیگری که برای بررسی تطابق اولین رقم اعداد با قانون بن‌فورد بکار می‌رود، مجموع تغییرات فاصله‌ای نمونه در مقایسه با توزیع بن‌فورد روی است، که به صورت

$$d=\frac{1}{2}\sum_{d_1=1}^{9}\left|P_{D_1}(d_1)-\log_{10}(1+\frac{1}{d_1})\right| \qquad d_1\in\{1,2,3,\dots,9\},$$

تعریف می‌شود. بیشترین میزان انحراف از توزیع بن‌فورد نیز می‌تواند معیار خوبی برای آزمون نیکویی برازش باشد. برای این منظور آماره d_{max} به صورت

$$d_{\max}=\max_{1\le d_1\le 9}\left|P_{D_1}(d_1)-\log_{10}(1+\frac{1}{d_1})\right| \qquad d_1\in\{1,2,3,\dots,9\},$$

تعریف می‌شود. واضح است که هر چه مقادیر دو آماره‌ی d و d_{max} کمتر باشد، داده‌ها انحراف کمتری از توزیع بن‌فورد خواهند داشت.

i) ثابت‌های فیزیک

در بسیاری از مقالاتی که پیرامون قانون بن‌فورد نوشته شده، از جدول ثابت‌های موجود در علم فیزیک به عنوان مثالی کاربردی از این قانون یاد شده است. بورک و کینکانون (۱۹۹۱) اولین کسانی بودند که به صورت جدی به بررسی قانون بن‌فورد برای مقادیر ثابت موجود در کتاب‌های فیزیک پرداختند. اما آن‌ها تنها ثابت‌هایی از کتاب‌های مقدماتی فیزیک را انتخاب کردند، بنابراین تعداد نمونه آن‌ها بسیار کوچک بود و از لحاظ آماری نتایج قابل اعتمادی در بر نداشت.

ii) داده‌های اقتصادی و اجتماعی

بسیاری از مشاهداتی که از پدیده‌های اقتصادی و اجتماعی مانند جمعیت، تولید ناخالص داخلی، میزان مصرف سوخت‌های فسیلی و تعداد

سالیانه قتل و خودکشی در کشورها حاصل می‌شوند،، از قانون بن‌فورد پیروی می‌کنند. البته میزان برازش هر دسته از این مشاهدات متفاوت است. یکی از نتایج جالب و قابل توجه قانون بن‌فورد در بررسی قیمت و شاخص روزانه سهام ظاهر می‌شود. قیمت سهام طی یک دوره زمانی بلند مدت دارای توزیع بن‌فورد است. می‌توان نشان داد که قیمت سهام شرکتها در یک روز مشخص، بهتر از قیمت آن‌ها در طول یک دوره زمانی طولانی‌تر از توزیع بن‌فورد پیروی می‌کند. دلیل این امر آن است که قیمت‌ها حول یک مقدار مشخص نوسان می‌کنند و دارای نوسانات شدید نیستند. اداره مالیات آمریکا این قانون را برای بررسی صحت ترازهای مالی بنگاه‌های بزرگ اقتصادی و اطلاعات تجاری و حسابداری آن‌ها مورد استفاده قرار می‌دهد (دورتسکی و همکاران، ۲۰۰۴). خواننده‌ی علاقه‌مند را برای جزییات بیشتر به لی (۱۹۹۶) ارجاع می‌دهیم.

iii) دنباله‌های اعداد

ژپینگ و همکاران (۲۰۰۴) میزان هم‌خوانی برخی دنباله‌های عددی از جمله چند دنباله مشهور ریاضی با قاعده بن‌فورد را مورد مطالعه قرار دادند.

– دنباله‌ی اعداد فیبوناچی

دنباله اعداد فیبوناچی با رابطه‌ای بازگشتی به صورت $a_{n+۲} = a_{n+۱} + a_n$ نمایش داده می‌شود. به سادگی می‌توان فرم کلی

$$a_n = \left(\frac{۱+\sqrt{۵}}{۲}\right)^{n-۱}\left(\frac{۵+۳\sqrt{۵}}{۱۰}\right) + \left(\frac{۱-\sqrt{۵}}{۲}\right)^{n-۱}\left(\frac{۵-۳\sqrt{۵}}{۱۰}\right), \quad (۷)$$

را برای حالتی که $a_۱ = ۱$ و $a_۲ = ۱$ می‌باشد، ارائه نمود. عبارت دوم در سمت راست رابطه (۷) همواره کوچکتر از ۱ است و برای nهای بزرگ به سمت صفر میل می‌کند. بنابراین برای بررسی رقم آغازین اعداد دنباله فیبوناچی تنها به جمله اول نیاز است. لگاریتم جمله اول را می‌توان به صورت

$$\ln a'_n = (n-۱)\ln(\frac{۱+\sqrt{۵}}{۲}) + \ln(\frac{۵+۳\sqrt{۵}}{۱۰}),$$

نوشت. هنگامی که تعداد نمونه بزرگ است، $\ln a'_n$ در فاصله $\left(\ln\frac{۵+۳\sqrt{۵}}{۱۰}, \ln\frac{۱+\sqrt{۵}}{۲}\right)$ دارای توزیع یکنواخت است. بنابراین $\{a'_n\}$ و $\{a_n\}$ در فضای لگاریتمی به صورت یکنواخت توزیع شده‌اند، یا به عبارتی از قانون بن‌فورد تبعیت می‌کنند.

– دنباله‌ی α^n در پایه b

دنباله‌ی α^n در پایه b را که در آن $n \in N$ و $\alpha \in R$ است، را در نظر بگیرید. چون توزیع بن‌فورد در برابر مقیاس اندازه‌گیری پایاست، شرط کافی برای اینکه رقم آغازین این دنباله d باشد، این است که رابطه $d \times b^k \leq \alpha^n < (d+۱) \times b^k$ برقرار باشد. به عبارتی باید برای هر k صحیح و مثبت رابطه

$$\log_b d + k \leq n\log_b \alpha < \log_b(d+۱) + k,$$

برقرار باشد. بنابراین برای هر k احتمال اینکه رقم آغازین عدد α^n رقم d باشد، به صورت

$$\frac{\left(\log_b(d+۱)+k\right) - \left(\log_b d + k\right)}{(k+۱)-k} = \log_b(\frac{d+۱}{d}),$$

بدست می‌آید، که دقیقاً همانند رابطه‌ی (۵) است. جدول ۵ نتایج حاصل از آزمون‌های سه‌گانه نیکویی برازش برای برخی از دنباله‌های عددی را نشان می‌دهد. مقادیر آماره‌های آزمون، مطابقت دنباله‌های یاد شده با توزیع بن‌فورد را مورد تایید قرار می‌دهند.

iv) اعداد ارائه شده در صفحات روزنامه‌ها و مجلات

در میان داده‌هایی که از قانون بن‌فورد پیروی می‌کنند، اعداد ارائه شده در صفحات روزنامه‌ها و مجلات موضوع جالبی است. این موضوع عنوان یکی از جدول‌های بیست‌گانه گردآوری شده توسط بن‌فورد (۱۹۳۸) است. در نگاه اول این گونه اعداد به نظر کاملاً تصادفی می‌آیند. اما این اعداد نیز کاملاً منطبق بر قانون بن‌فورد هستند. در واقع می‌توان نشان داد که اگر از توزیع‌های احتمالی مختلف به صورت تصادفی چند توزیع انتخاب و از هر توزیع نیز نمونه‌ای تصادفی به روش دلخواه انتخاب شود، آن‌گاه اعداد بدست آمده در این فرآیند در برابر پایه و مقیاس پایا است و در نتیجه ارقام آغازین آن‌ها از توزیع بن‌فورد پیروی می‌کند (هیل، ۱۹۹۸). اعداد ارائه شده در صفحات مختلف روزنامه‌ها و مجلات مخلوطی از توزیع‌های مختلف هستند و در نتیجه از توزیع بن‌فورد پیروی می‌کنند. در این میان اعداد مربوط به مجلات تخصصی نظیر مجلات حسابداری بیش از سایرین از توزیع بن‌فورد منحرف می‌شوند، زیرا به صورت کامل در شرایط فوق صدق نمی‌کنند.

v) راستی آزمایی مشاهدات مالی

راستی آزمایی مشاهده های مالی شرکت‌ها و موسسات، یکی دیگر از کاربردهای جذاب قاعده بن‌فورد است. این مشاهده ها معمولاً به دلایل مختلف از جمله فرار از پرداخت مالیات یا استفاده‌های سیاسی، مورد تحریف و دستکاری قرار می‌گیرند. در تمام این موارد تخطی از قانون

بن‌فورد هشداری جدی برای برای وقوع تخلف است. نیگرینی (۱۹۹۶) برنامه‌ای رایانه‌ای تهیه نمود که اظهارنامه‌های مالیاتی افراد یا شرکت‌ها را از طریق این قانون مورد ارزیابی قرار داده و تخلفات هفتگانه در تکمیل اظهارنامه‌های مالیاتی را که توسط سازمان بازرسی و تحقیقات مالی آمریکا مشخص شده‌اند، به خوبی تشخیص می‌دهد. براون (۱۹۹۸) از این برنامه برای بررسی عملکرد مالی بیل کلینتون رییس جمهور سابق ایالات متحده استفاده نمود.

vi) راستی آزمایی نتایج انتخابات

از قانون بن‌فورد برای راستی‌آزمایی نتایج انتخابات نیز استفاده می‌شود. مبین و همکاران (۲۰۰۶) و مبین و والتر (۲۰۰۹) راستی آزمایی نتایج انتخابات مختلف را مورد توجه قرار دادند. به عنوان مثال آنها تعداد آرای دو نامزد حزب جمهوری‌خواه و دموکرات (جرج بوش و جان کری) و همچنین دو نامزد مجلس سنای آمریکا (مل مارتینز و بتی کاستور) و آرای مخالف یا موافق برای ۸ سؤال پیرامون اصلاح قانون اساسی ایالتی که در سال ۲۰۰۴ انجام شده بود را مورد بررسی قرار دادند. این ۸سئوال و تعداد موافقین و مخالفین آن، در جدول ۶ آمده است. ملاحظه می‌شود که در این انتخابات همه اصلاحات پیشنهادی با اکثریت آراء مورد تأئید عمومی قرار گرفته‌اند، تنها در سوال چهارم پاسخ‌های مثبت و منفی به یکدیگر نزدیک هستند. در جدول ۷ آماره کای‌دو برای دو نوع آزمون مختلف محاسبه شده است. در آزمون اول میزان همخوانی رقم آغازین آراء بدست آمده از حوزه‌های مختلف رای‌گیری برای انتخابات ریاست جمهوری، انتخابات مجلس سنا و اصلاح قانون اساسی در مامی‌دیت، با قانون بن‌فورد بررسی شده است. در آزمون دوم یکنواخت بودن توزیع رقم آغازین مورد توجه قرار گرفته است. برای آزمون اول $q_{B_k i}$ فراوانی نسبی اینکه $k-$امین رقم معنی‌دار برابر عدد i باشد را نشان می‌دهد. d_{ki} بیانگر تعداد دفعاتی است که در $k-$امین رقم، عدد i مشاهده شده است. با فرض اینکه میان آراء بدست آمده از حوزه‌های مختلف رای‌گیری استقلال وجود دارد، دو آماره‌ی کای‌دو برای آزمون‌های اول و دوم به ترتیب به صورت

$$\chi^2_{B_1} = \sum_{i=1}^{9} \frac{(d_{1i} - d_1\, q_{B_1 i})^2}{d_1\, q_{B_1 i}}$$
$$\chi^2_{E_1} = \sum_{i=1}^{9} \frac{(d_{1i} - d_1/9)^2}{d_1/9},$$

هستند، که در آنها $d_1 = \sum_{i=1}^{9} d_{1i}$. مقادیر آماره‌های آزمون در جدول ۷ بطور قابل توجهی بزرگ هستند. بر این اساس، فرض‌های صفر پیروی اولین رقم معنی‌دار از توزیع بن‌فورد یا توزیع یکنواخت گسسته بر روی $\{1,2,\ldots,9\}$ رد می‌شوند. جدول ۸ آماره‌های آزمون کای‌دو پیرسن برای دومین رقم معنی‌دار میان حوزه‌های رای‌گیری را نشان می‌دهد. برای $k=2$ مقادیر $q_{B_2 i}$ ستون دوم جدول ۸ است. آماره‌ی آزمون برای دومین رقم معنی‌دار به صورت

$$\chi^2_{B_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2\, q_{B_2 i})^2}{d_2\, q_{B_2 i}}, \qquad (8)$$

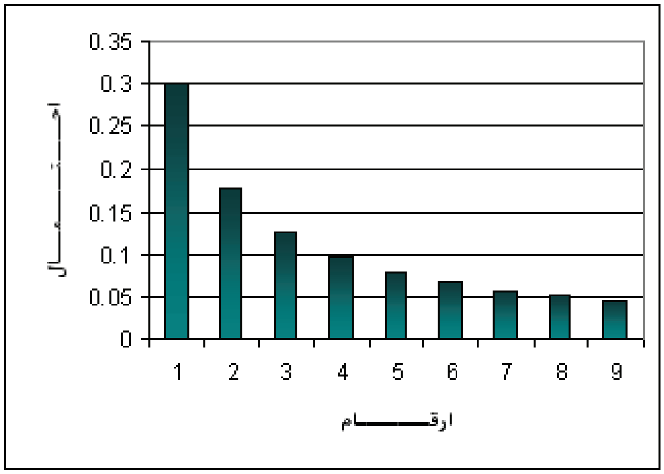
و برای آزمون یکنواختی اولین رقم معنی‌دار به صورت

$$\chi^2_{E_2} = \sum_{i=0}^{9} \frac{(d_{2i} - d_2/10)^2}{d_2/10}, \qquad (9)$$

است، که در روابط (۸) و (۹)، $d_2 = \sum_{i=0}^{9} d_{2i}$ می‌باشد. این آماره‌ها را می‌توان با توزیع کای‌دو با ۹ درجه آزادی مقایسه کرد. آماره‌های بدست آمده دلایل بسیار ناچیزی مبنی بر تخطی از قانون بن‌فورد در بر دارند. از بیست آماره‌ی موجود در جدول ۸، دو آماره بزرگتر از ۱۶/۹ هستند، اما هیچ آماره‌ای از ۲۵/۵ بزرگتر نیست و بزرگترین آماره در ستون اول برابر ۱۷/۹ است، در حالی که بزرگترین آماره در ستون دوم برابر ۲۵/۳ است. بنابراین می‌توان با سطح اطمینان قابل قبولی فرض وقوع تقلب در انتخابات سال ۲۰۰۴ در ایالت فلوریدای آمریکا را رد کرد (مبین و والتر، ۲۰۰۶).

۸ بحث و نتیجه‌گیری

توزیع بن‌فورد توزیعی گسسته و چوله به راست است که برای ارقام معنی‌دار اعداد حاصل از بسیاری از وقایع طبیعی بکار می‌رود. داده‌هایی که از توزیع بن‌فورد تبعیت می‌کنند در برابر مقیاس اندازه‌گیری و پایه‌ی اعداد پایا هستند. به عبارتی اگر اعدادی که بر اساس توزیع بن‌فورد توزیع شده‌اند، در عددی ثابت ضرب یا به پایه‌ی دیگری تبدیل شوند، در نتیجه‌ی آزمون تغییری حاصل نمی‌شود. کاربردهایی که در این مقاله برای قانون بن‌فورد ارائه شده، فقط بخش کوچکی از کاربردهای وسیع این قانون است.



شکل ۱: بافت‌نگار توزیع بن‌فورد برای اولین رقم.

جدول ۱: توزیع احتمالی ارقام معنی‌دار در موقعیت‌های مختلف.

	موقعیت قرار گرفتن عدد				
رقم	اول	دوم	سوم	چهارم	پنجم
۰	۰	۰/۱۱۹۶۸	۰/۱۰۱۸	۰/۱۰۰۲	۰/۱۰۰۰
۱	۰/۳۰۱۰۳	۰/۱۱۳۹	۰/۱۰۱۴	۰/۱۰۰۱	۰/۱۰۰۰
۲	۰/۱۷۶۰۹	۰/۱۰۸۸	۰/۱۰۱۰	۰/۱۰۰۱	۰/۱۰۰۰
۳	۰/۱۲۴۹۴	۰/۱۰۴۳	۰/۱۰۰۶	۰/۱۰۰۱	۰/۱۰۰۰
۴	۰/۰۹۶۹۱	۰/۱۰۰۳	۰/۱۰۰۲	۰/۱۰۰۰	۰/۱۰۰۰
۵	۰/۰۷۹۲۸	۰/۰۹۶۷	۰/۰۹۹۸	۰/۱۰۰۰	۰/۱۰۰۰
۶	۰/۰۶۶۹۵	۰/۰۹۳۴	۰/۰۹۹۴	۰/۰۹۹۹	۰/۱۰۰۰
۷	۰/۰۵۷۹۹	۰/۰۹۰۳	۰/۰۹۹۰	۰/۰۹۹۹	۰/۱۰۰۰
۸	۰/۰۵۱۱۵	۰/۰۸۷۶	۰/۰۹۸۶	۰/۰۹۹۹	۰/۱۰۰۰
۹	۰/۰۴۵۷۶	۰/۰۸۵۰	۰/۰۹۸۳	۰/۰۹۹۸	۰/۱۰۰۰

جدول ۲: امید ریاضی و واریانس توزیع بن‌فورد برای ارقام مختلف.

رقم	امید ریاضی	واریانس
اول	۳/۴۴۰۲	۶/۰۵۶۵
دوم	۴/۱۸۷۳	۸/۲۵۳۷
سوم	۴/۴۶۷۷	۸/۲۵۰۰
چهارم	۴/۴۹۶۷	۸/۲۵۰۰
پنجم	۴/۴۹۹۶	۸/۲۵۰۰
ششم	۴/۴۹۹۹	۸/۲۵۰۰
هفتم	۴/۴۹۹۹	۸/۲۵۰۰

جدول ۳: ضریب همبستگی بین ارقام مختلف در توزیع بن‌فورد.

	j			
i	۲	۳	۴	۵
۱	۰/۰۵۶۰۵۶۳	۰/۰۰۵۹۱۲۶	۰/۰۰۰۵۹۱۶	۰/۰۰۰۰۵۹۱
۲		۰/۰۰۲۰۵۶۶	۰/۰۰۰۲۰۵۹	۰/۰۰۰۰۲۰۵
۳			۰/۰۰۰۰۲۲۸	۰/۰۰۰۰۰۲۲
۴				۰/۰۰۰۰۰۰۲

جدول ۴: مجموع تغییرات فاصله‌ای برای ارقام مختلف.

رقم (k)	$d(D_k, U)$
اول	۰/۲۶۸۷۲۶۶۶
دوم	۰/۰۴۷۰۲۸۶۳
سوم	۰/۰۰۴۸۸۳۵۶
چهارم	۰/۰۰۰۴۸۸۵۸
پنجم	۰/۰۰۰۰۴۸۸۶
ششم	۰/۰۰۰۰۰۴۸۹
هفتم	۰/۰۰۰۰۰۰۴۹

جدول ۵: آزمون‌های نیکویی برازش برای دنباله‌های عددی.

$d_{\max}$	d	$\chi^۲$	اندازه نمونه	دنباله
$۱/۷ \times ۱۰^{-۴}$	$۳/۸ \times ۱۰^{-۴}$	۰/۰۱۲۵	۱۰,۳۱۷	اعداد فیبوناچی
۰/۱۵۲۲	۰/۲۲۷۱	۴۵/۰	۱۶۸	اعداد اول کوچکتر از ۱,۰۰۰
۰/۱۷۶۱	۰/۴۹۰۵	۳۲۴۷	۹۷۶۱	اعداد اول کوچکتر از ۱۰۰,۰۰۰
$۵/۹ \times ۱۰^{-۴}$	$۱/۲ \times ۱۰^{-۳}$	۰/۴۱۰	۳۰,۰۰۰	$۱/۰۰۷^n$ برای $n = ۱,\ldots,۳۰,۰۰۰$
$۱/۲ \times ۱۰^{-۴}$	$۲/۵ \times ۱۰^{-۴}$	۰/۰۳۲۹	۶۵,۰۲۸	$۱/۰۰۷^n$ برای $n = ۱,\ldots,۳۰,۰۰۰$
۰/۰۴۸۸۵	۰/۰۶۵۱	۶/۹۵	۱۰۰	فاکتوریل ۱ تا ۱۰۰
۰/۰۳۴۹۲	۰/۰۸۷۱	۸/۹۷	۱۳۰	فاکتوریل۱ تا ۱۳۰
۰/۰۳۶۳۵	۰/۰۸۳۴	۱۰/۱۰	۱۶۰	فاکتوریل۱ تا ۱۶۰
۰/۰۹۹۰۰	۰/۱۴۰۹	$۳/۱۶ \times ۱۰^۳$	۳۰,۰۰۰	$n^۲$ برای $n = ۱,\ldots,۳۰,۰۰۰$
۰/۰۳۶۴۰	۰/۰۳۹۴	$۲/۷۶ \times ۱۰^۲$	۳۰,۰۰۰	$n^۵$ برای $n = ۱,\ldots,۳۰,۰۰۰$
۰/۰۰۶۸۱	۰/۰۱۱۲	۲۰/۸	۳۰,۰۰۰	$n^{۲۰}$ برای $n = ۱,\ldots,۳۰,۰۰۰$
۰/۰۰۲۵۶	۰/۰۰۴۸	۳/۷	۳۰,۰۰۰	$n^{۵۰}$ برای $n = ۱,\ldots,۳۰,۰۰۰$

جدول ۶: اصلاح قانون اساسی ایالت فلوریدا، سال ۲۰۰۴ میلادی.

سوال	اصلاح مورد نظر	تعداد موافقین	تعداد مخالفین
۱	آگاه‌سازی والدین از سقط جنین در نوجوانان	۴,۶۳۹,۶۳۵	۲,۵۳۴,۹۱۰
۲	اصلاح قانون اساسی بر اساس درخواست و تقاضا برنامه‌ریزی شود	۴,۵۷۴,۳۶۱	۲,۱۰۹,۰۱۳
۳	اصلاح دستمزد پزشکان دارای تعهد خدمت	۴,۵۸۳,۱۶۴	۲,۶۲۲,۱۴۳
۴	آزمایش دستگاه‌های فروش خودکار در ادارات استان‌های مامی‌یت و بروارد	۳,۶۳۱,۲۶۱	۳,۵۱۲,۱۸۱
۵	اصلاح حداقل دستمزد در فلوریدا	۵,۱۹۸,۵۱۴	۲,۰۹۷,۱۵۱
۶	اصلاح حداکثر سرعت مجاز خطوط ریلی	۴,۵۱۹,۴۲۳	۲,۵۷۳,۲۸۰
۷	حق قانونی بیماران برای اطلاع از عوارض زیان‌بار جراحی	۵,۸۴۹,۱۲۵	۱,۳۵۸,۱۸۳
۸	حمایت دولت از مردم در برابر سهل‌انگاری در معالجات پزشکی	۵,۱۲۱,۸۴۱	۲,۰۸۳,۸۶۴

جدول ۷: مقدار آماره‌ی کای‌دو برای اولین رقم معنی‌دار در ۷۵۷ حوزه رای‌گیری.

موضوع			توزیع مورد مقایسه بن‌فورد	توزیع مورد مقایسه یکنواخت
انتخابات ریاست جمهوری	کاندید	بوش	۲۹/۳	۲۹۲/۵
		کری	۳۹/۹	۲۸۷/۰
انتخابات مجلس سنا	کاندید	مارتینز	۳۵/۶	۲۷۳/۸
		کاستور	۲۲/۰	۳۰۴/۷
موافقین اصلاح قانون ایالتی	سئوال	اول	۸۶/۲	۲۹۰/۵
		دوم	۹۵/۶	۳۶۲/۵
		سوم	۶۰/۵	۴۰۱/۳
		چهارم	۱۴۴/۸	۳۶۷/۰
		پنجم	۱۱۵/۴	۱۲۲/۲
		ششم	۹۸/۸	۳۹۵/۰
		هفتم	۱۳۰/۳	۱۱۲/۷
		هشتم	۱۲۳/۰	۲۱۰/۶
مخالفین اصلاح قانون ایالتی	سئوال	اول	۸۰/۵	۶۳۶/۵
		دوم	۶۰/۰	۷۲۲/۷
		سوم	۵۱/۵	۴۹۶/۵
		چهارم	۱۱۹/۶	۶۰۵/۶
		پنجم	۲۷/۶	۶۲۳/۴
		ششم	۸۴/۰	۵۳۲/۹
		هفتم	۴۹/۹	۵۸۲/۸
		هشتم	۱۰۲/۶	۸۳۱/۱

جدول ۸: مقدار آماره کای‌دو برای دومین رقم معنی‌دار در ۷۵۷ حوزه رای‌گیری.

موضوع			توزیع مورد مقایسه بن‌فورد	توزیع مورد مقایسه یکنواخت
انتخابات ریاست جمهوری	کاندید	بوش	۷/۹	۱۰/۸
		کری	۹/۵	۱۴/۴
انتخابات مجلس سنا	کاندید	مارتینز	۸/۹	۱۰/۸
		کاستور	۱۲/۰	۱۲/۸
موافقین اصلاح قانون ایالتی	سئوال	اول	۲/۵	۸/۰
		دوم	۱۶/۷	۲۳/۶
		سوم	۳/۳	۸/۵
		چهارم	۳/۳	۹/۰
		پنجم	۱۷/۹	۱۹/۶
		ششم	۴/۳	۱۰/۲
		هفتم	۱۷/۱	۱۶/۰
		هشتم	۱۲/۷	۲۵/۳
مخالفین اصلاح قانون ایالتی	سئوال	اول	۵/۵	۱۵/۵
		دوم	۷/۲	۱۶/۴
		سوم	۱۲/۹	۱۲/۷
		چهارم	۵/۷	۱۵/۴
		پنجم	۵/۸	۲۳/۳
		ششم	۹/۱	۱۱/۳
		هفتم	۸/۴	۱۶/۵
		هشتم	۶/۵	۱۰/۶

مراجع

[1] Benford, F. (1938). The Law of Anomalous Numbers, *Proceedings of the American Philosophical Society,* **78**, 551-572.

[2] Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate, A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B,* **57 (1)** , 289-300.

[3] Browne, M. W. (1998). *Following Benford's law, or Looking out for No. 1*, New York Times, August 4, 1998.

[4] Burke, J. and Kincanon, E. (1991). Benford's Law and Physical Constants: the Distribution of Initial Digits, *American Journal of Physics*, **59**, 952.

[5] Durtschi, C., Hillison, W. and Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data, *Journal of Forensic Accounting* 1524-5586 Vol. V, 17-34.

[6] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed.), John Wiley, **63**. New York.

[7] Hill, T. P. (1995a). Base-Invariance Implies Benford's Law, *Proceedings of the American Mathematical Society*, **123**, 887-895.

[8] Hill, T. P. (1995b). A Statistical Derivation of the Significant-Digit Law, *Statistical Science,* **10**, 354-363.

[9] Hill, T. P. (1998). The First Digit Phenomenon, *The American Scientist*, **10(4)**, 354-363.

[10] Leemis, L. M., Schmeier, B.W. and Evans, D.L. (2000). Survival Distributions Satisfying Benford's Law, *The American Statistician* **54**, 3, 1-6.

[11] Ley, E. (1996). On the Peculiar Distribution of the U. S. Stock Indices Digits. *The American Statistician,* **50**, 4, 311-313.

[12] Lowe, T., Murphy, S. and Hayward, J. (1999). The First Digit Law, URL: www.doc.ic.ac.uk/ldcl/JMC/group4.ps

[13] Mebane, Jr. Walter, R. (2006). Vote Counts and Benford's Law, URL: www.personal.umich.edu/wmebane/pm06.pdf

[14] Mebane, Jr. Walter, R. (2009). Note on the Presidential Election in Iran, URL: www-personal.umich.edu/wmebane/note18jun2009.pdf

[15] Newcomb, S. (1881). Note on the Frequency of Use of the Different Digits in Natural Numbers. *American Journal of Mathematics*, **4**, 39-40.

[16] Nigrini, M. (1996). A Taxpayer Compliance Application of Benford's Law, *Journal of the American Taxation Association*, **1**, 72-91.

[17] Pinkham, R. S. (1961). On the Distribution of First Significant Digits, *Annals of Mathematical Statistics*, **32**, 1223-1230.

[18] Walthoe, J. Hunt, R. and Pearson, M. (1999). Looking out for Number One, Plus Magazine, University of Cambridge, September 1999.

[19] Zhipeng, L. Cong, L. and Wang, H. (2004). Discussion on Benford's Law and its Application, URL: www.arxiv.org/abs/math.ST/0408057.