

اثبات تقلب در انتخابات ۲۲ خرداد ۸۸

با استفاده از یک تحلیل آماری

توضیح زیتون - نویسندگان در این مقاله علمی، با تحلیل‌های آماری بر آن است که نشان دهد مجموع شمار رای‌ها در انتخابات ریاست جمهوری ۱۳۸۸، «تصادفی» نیستند و لاجرم «واقعی» نیستند. لذا «نتایج اعلام شده از سوی وزارت کشور جمهوری اسلامی در انتخابات ۱۳۸۸ نمیتوانند نتیجه یک انتخابات سالم و عاری از دستکاری باشند.»

نویسندگان مبنای کار خود را چنین توضیح می‌دهد: «موضوع نوشته حاضر معطوف به وجود شمار زیادی صندوق‌های رای است که تعداد کل آرای آن‌ها ضریبی از ۱۰۰ بوده‌اند. آنچه که این تحلیل را از دیگر تحلیل‌ها که اشاره‌ای گذار به این یافته‌ها کرده‌اند، متمایز میکند این است که کوشش کرده‌ام به لحاظ آماری موضوع وجود چنین صندوق‌ها را ژرفتر ارزیابی کنم و دریابم که آیا نتایج اعلام شده رخدادی تصادفی ناشی از یک انتخابات سالم‌اند یا پیامد دستکاری‌ای هدفمند و تقلب. .. چنانکه ثابت شود که مجموع شمار رای‌ها در صندوق‌ها تصادفی نیستند - و لاجرم واقعی نیستند - ضرورتی برای ارزیابی چگونگی پخش ارقام غیر واقعی میان کاندیداها وجود ندارد.

با توجه به وجود جداول و علائم ریاضی و غیر قابل انتشار در ویراستار سایت زیتون، پس از انتشار مقدمه، متن کامل مقاله به صورت پی‌دی‌اف ارائه می‌شود.

زیتون - بهروز بیات

مقدمه: شاید ارائه این مطالعه آماری از انتخابات ریاست جمهوری خرداد ۱۳۸۸ هفت سال پس از آن وقایع فاجعه‌آمیز در ایران دیر هنگام بنظر آید. اما دو انگیزه مرا وادار به پرداختن به این موضوع کرده است:

یکم: هفت سال پس از ۲۲ خرداد ۱۳۸۸ هنوز سرکوب، زندان و حصر معترضین به تقلب انتخاباتی به عنوان «فتنه‌گر» ادامه دارد و

دوم: هر چند که جمهوری اسلامی تقریباً هرگز انتخابات آزاد و سالم که در خور سنجش با معیارهای رایج جهانی باشد به خود ندیده است اما تا پیش از خرداد ۱۳۸۸ تنها یکی از مولفه‌های انتخاباتی یعنی شمارش کم و بیش واقعی آرا را تا حدودی رعایت میکرد. انتخابات ۸۸ نقطه عطفی بود به سوی بازگشت از این حداقل انتخاباتی که پیامدش عملاً نفی وجودی انتخابات و تهی کردن آن از هرگونه معنائی میتواند باشد. ملاحظاتی مذکور مرا بر آن داشت که با انتشار تحلیلی ریاضی-آماري از نتیجه انتخابات خرداد ۱۳۸۸، به سویی‌ای از آن پردازم که تاکنون با ژرفای کافی به آن پرداخته نشده است.

پس از اعلام نتیجه انتخابات ۱۳۸۸ که برای بسیاری از شهروندان ایران گمان تقلب و دستکاری را برانگیخته بود متخصص‌های بسیاری از ایران و جهان به تحلیل این نتیجه‌ها دست‌یازیده‌اند که بخش در خور توجه‌ای از آن‌ها در ویکیپدیا منتشر شده است [اینجا](#).

موضوع نوشته حاضر معطوف به وجود شمار زیادی صندوق‌های رای است که تعداد کل آرای آن‌ها ضریبی از ۱۰۰ بوده‌اند. این واقعیت جلب توجه دیگر کارشناسان را نیز کرده بود و با نگاهی گذرا آن را به مثابه وجود تقلب ارزیابی کرده بودند. اما آنچه که این تحلیل را از دیگر تحلیل‌ها که اشاره‌ای گذار به این یافته‌ها کرده‌اند، متمایز میکند این است که کوشش کرده‌ام به لحاظ آماری موضوع وجود چنین صندوق‌ها را ژرفتر ارزیابی کنم و دریابم که آیا نتایج اعلام شده رخدادی تصادفی ناشی از یک انتخابات سالم‌اند یا پیامد دستکاری‌ای هدفمند و تقلب.

در تحلیل پیش‌رو به چگونگی توزیع آرا میان کاندیداها نپرداخته‌ام زیرا چنانکه ثابت شود که مجموع شمار رای‌ها در صندوق‌ها تصادفی نیستند - و لاجرم واقعی نیستند - ضرورتی برای ارزیابی چگونگی پخش ارقام غیر واقعی میان کاندیداها وجود ندارد.

نگارنده بلافاصله پس از انتخابات 1388 با همراهی آقای مسعود آذری تحلیلی به زبان انگلیسی از نتایج انتخابات منتشر کردیم که با در نظر گرفتن سویه های گوناگون موثر در انتخابات نشان میداد که تقلبی گسترده رخ داده است. اما در پی این بودم که در یابم آیا میتوان بدون در نظر گرفتن ملاحظات سیاسی، شیوه سازماندهی، نحوه و گستره اعمال نفوذ نهاد ها و دیگر عوامل اثرمند، صرفا با اکتفا به نتایج رسمی اعلام شده از سوی دولت جمهوری اسلامی، ادعای معترضین جنبش سبز را در باره تقلب در انتخابات راستی آزمایی کرد؟

پایه این تحلیل ارقامی است که وزارت کشور جمهوری اسلامی در وبسایت خود از استان های تهران، لرستان، یزد، آذربایجان شرقی، اصفهان، اردبیل و کرمانشاه منتشر کرده است. کوششی است مبتنی بر این اصل که با کمترین ملاحظات سیاسی در یابم که آیا این نتایج میتوانند ناشی از انتخاباتی سالم و دستکاری نشده باشند.

این نوشته را به دو بخش I و II تقسیم کرده ام: بخش I در برگزیده نتایج تحلیل به شیوه ای که برای عموم خواننده گان مفهوم باشد و بخش II را برای کسانی که علاقمند به جزئیات تئوریک آن باشند، تنظیم کرده ام.

شیوه پرداختن به موضوع

برای بررسی واقعی یا ساختگی بودن نتایج انتخابات 1388 مجموع آرای تک صندوق های اخذ رای را تحلیل کرده ام. اگر نتیجه این باشد که این مجموعه های آرا تصادفی نیستند و لاجرم دستکاری شده و ساختگی اند، پرداختن به جزئیات توزیع شان در میان نامزد ها - که معرف همگان است، معنی چندانی ندارد از این روی به توزیع آرا نپرداخته ام و تنها به کل آرای هر صندوق اکتفا کرده ام.

نگاهی به شمار کل آرای که در هر صندوق رای گیری ریخته شده است نشانه هایی از غیرعادی بودن آشکار میگرد. شاخص ترین نشانه ها این بود که تعداد صندوق هایی که تعداد کل آرای مأخوذه صریبی از 100 بود به طرز چشمگیری فراتر از انتظار بود.

این مشاهده مرا بران داشت که تکرار دو رقم دست راست یعنی یکان و دهگان تعداد کل آرای ریخته شده به همه تک صندوق های رای را برای استان های تهران، لرستان، اصفهان، یزد، آذربایجان، اردبیل و کرمانشاه تحلیل کنم و برای مقایسه با نتایج دو انتخابات ایالتی وین 2014 و 2015 و یکی در برلین که هر یک بزرگترین شهر آتریش یا آلمان نیز هستند، بسنجم.

I. **چکیده نتیجه:** پیش از پرداختن به جزئیات تئوریک رهیافتی که انتخاب کرده ام و برای خواننده گانی که ممکن است پرداختن به این جزئیات خارج از حوصله شان باشد چکیده یافته هایم را در بخش I در دو جدول 1 و 2 خلاصه کرده ام:

به دلیل آنچه در بخش II توضیح خواهم داد احتمال تکرار k بار دو رقم یکان و دهگان تعداد کل آرای هر صندوق رای را میتوان به میانجی توزیع بینومیال binomial distribution محاسبه کرد. برای تسهیل نمونه وار محتوای صندوق هائی مانند 200، 500، 900، 1500 ... را که یکان و دهگان شان صفر است E00 و 169، 869، 2069، 2369 ... را که یکان اشان 9 و دهگان اشان 6 است E69 مینامیم. احتمال تکرار k بار E00 یا E69 در هر دو رقم دیگر (m و n هر یک میتوانند ارقامی میان 0 تا 9 باشند) را میتوان از راه توزیع بینومیال محاسبه کرد.

در مورد انتخابات 1388 ایران آنچه که فوراً جلب نظر میکند فراوانی تکرار پاره ای از Emn ها به ویژه فراوانی چشمگیر تکرار E00 ها است. جدول 1 نتایج محاسبه این احتمالات را بازتاب داده است.

محل انتخابات	تعداد صندوق ها	تعداد E00 ها	% E00 از کل استان	% E00 مرکز استان	احتمال بروز E00	Emn	تعداد دیگر Emn	احتمال بروز Emn
وین 2014	1543	15	0,97	-	0,102	E96	23	0,0164
وین 2015	1518	20	1,32	-	0,044	E98	24	0,0091
برلین	2277	25	1,10	-	0,072	E86	39	0,0005

لرستان	1272	132	10,38	6,60	5,11E-87	E50	45	8,44E-13
یزد	692	25	3,61	7,20	5,13E-08	E50	16	0,0012
تهران	6085	216	3,55	4,50	7,55E-55	E50	82	0,0015
آذربایجان شرقی	2764	64	2,32	1,95	1,12E-09	E97	40	0,0055
اصفهان	2280	87	3,82	3,95	3,33E-25	E99	48	1,34E-06
اردبیل	1154	22	1,91	1,23	0,0019	E04	2	0,0006
کرمانشاه	۱۴۳۸	47	3,27	1,68	3,98E-12		4	0,0010

جدول 1: این جدول در برگزیده ستون های «محل انتخابات»، «تعداد صندوق های» رای در انتخابات محل مربوطه، «تعداد E00 ها» یعنی تعداد صندوق هایی که شمار کل رای های ریخته شده به آن ها ضریبی از ۱۰۰ باشد، «% E00 از کل استان» یعنی سهم در صد فراوانی صندوق های E00 از کل صندوق ها در استان مربوطه، «%E00 مرکز استان ها» یعنی سهم در صد صندوق های E00 از کل صندوقهای شهرهای مرکز استان، ستون «احتمال بروز E00» یعنی احتمال وقوع k بار تکرار E00 در استان ها، «Emn» دیگر شمار کل آرای اخذ شده در صندوق های رای با ارقام پایانی یکان و دهگان به ترتیب m و n است که فراوانی شان فراتر از حد انتظار بوده است، «تعداد Emn دیگر» شمار صندوق های که مجموع آرایشان به دو رقم m و n ختم میشوند و «احتمال بروز Emn» یعنی احتمال رخدادن این تعداد تکرار یک Emn معین - مثلا اعدادی که به 50 ختم میشوند.

از مشاهده جدول 1 میتوان به نتیجه گیری های زیر رسید:

- سه انتخابات در وین و برلین به مثابه نمونه انتخابات آزاد و سالم از یکسو و از دیگر سو هفت استان ایران مورد بررسی قرار گرفته اند. برای سه شهر نمونه درصد صندوق های E00 نسب به جمع کل تعداد همه صندوق ها میان 0.97% تا 1.32% قرار دارد که پیرامون اندازه مورد انتظار 1% در نوسان است در حالیکه در استان های ایران میان 1.91% برای اردبیل تا 3.55% برای تهران تا 10.38% برای لرستان در تغییر است که از 1% مورد انتظار بسیار فاصله دارد.
- برای همه موارد تکرار ها در جدول 1 احتمال وقوع محاسبه شده اند. در حالیکه احتمال این تعداد تکرار های رخ داده برای وین و برلین میان 0.04، 0.07 و 0.10 است برای استان های ایران میان 10⁻⁸⁷، 10⁻⁵⁵، 10⁻²⁵، 10⁻¹²، 10⁻⁹، 10⁻⁸ و 0.002، یعنی در سه مورد اول عملا احتمال وقوع صفر است و در بقیه موارد به استثنای اردبیل بسیار ناچیز. در مورد اردبیل نیز این احتمال وقوع E00، 20 تا 50 برابر کمتر از موارد وین و برلین است.
- این ارقام نشان میدهند که نتایج اعلام شده از سوی وزارت کشور جمهوری اسلامی نمیتواند نتیجه یک انتخابات سالم و عاری از دستکاری باشند. در همینجا میتوان دریافت که بالاترین اندازه تقلب در لرستان صورت گرفته است.
- پاره ای از **مدعیان** سالم بودن انتخابات بیان کرده اند که وجود بیش از انتظار صندوق های با شمار کل رای هائی که ضریبی از 100 (E00) هستند اینست که تعرفه های رای در دسته های صد تائی بسته بندی شده اند و مراکز اخذ رای دچار کمبود تعرفه شده باشند.
- a. به روایت سخنگوی وقت شورای نگهبان آقای کدخدائی برای تقریبا 40 میلیون شرکت کننده 60 میلیون تعرفه تهیه شده بود و به جز در دو منطقه اخذ رای آن هم با تاخیر 20 و 40 دقیقه ای هیچ شعبه رای گیری در ایران دچار کمبود تعرفه نبوده است. عین گفته ایشان به نقل از وبسایت **الف** چنین است:

«مجموع تعرفه هایی که برای این دوره از انتخابات چاپ شده بود، حدود ۶۰ میلیون برگه بود که شورای نگهبان در روزهای گذشته، برای اطمینان بیشتر، سه نفر از کارشناسان و بازرسان خود را به وزارت کشور فرستاد و آنها قراردادهای مربوط به چاپ، تعداد تعرفه هایی که در انبارها باقی مانده، تعداد تعرفه هایی که مصرف شده و تعداد تعرفه هایی که در صندوقها است را با آمار و ارقام و اسناد کافی جمعآوری کردند و گزارش آن را به اعضای شورای محترم نگهبان ارائه کردند و خوشبختانه هیچگونه مشکلی در این زمینه نبوده و این گزارش قابل ارائه است... روند رأیگیری، تنها در دو منطقه، آن هم به دلیل استقبال گسترده مردم، با تأخیر ۲۰ دقیقه و ۴۰ دقیقههای مواجه شد. غیر از این موارد هیچگونه گزارشی مبنی بر اینکه برگه تعرفه رای در حوزهای وجود نداشته باشد به ما واصل نشده است»

b. نگاهی به توزیع صندوق های ضریب 100 یا E00 نشان میدهد که در بیشتر استان ها در صد تعداد اینگونه صندوق ها در شهرهای مرکز بالاتر است تا در حاشیه های دور افتاده: استان تهران 3.55٪ در برابر شهر تهران 4.50٪، استان اصفهان 3.82٪ در برابر شهر اصفهان 3.95٪، استان یزد 3.61٪ در برابر شهر یزد 7.2٪. در موردی مانند استان لرستان که به شیوه ای مفرط یعنی تعداد E00 با 10.38٪ از انتظار بدور است، در مرکز ش خرم آباد هر چند که نسبت به حاشیه کمتر است اما با 6.6٪ یعنی 6.6 برابر 1٪ مورد انتظار در صدی بسیار بالا است. در استان آذربایجان شرقی 2.32٪ در مقابل تبریز با 1.95٪ صندوق های E00 اختلافی بزرگ میان مرکز و حاشیه وجود ندارد. تنها در استان کرمانشاه در صد E00 در شهر کرمانشاه با حاشیه اختلاف چشمگیر دارد (1.68٪ در برابر 3.28٪). در شهر مشهد، نیز سهم E00 از کل صندوق های شهر 4.3٪ است که احتمال رخدادنش با 3.2×10^{-13} بسیار ناچیز است- همین درصد برای آبادان 3.2٪ و برای اهواز 2.5٪ بوده است.

بنابراین اگر بیانات سخنگوی شورای نگهبان جدی تلقی شود نمیتوان وجود وافر صندوق هائی با ضریب 100 (E00) را با کمبود تعرفه توضیح داد؛ نخست اینکه تعرفه کافی در دسترس بوده است و دوم اینکه اگر کمبود جبران ناشدنی وجود میداشت قاعدتاً میبایستی در حاشیه باشد نه در مراکز استان ها - که با واقعیت های مندرج در جدول 1 مغایرت دارد .

وانگهی تعداد فراوان Emn، آنچنان که در جدول 1 میتوان مشاهده کرد، منحصر به E00 نیست بلکه E50 (ضریبی از 50)، E99، E45، E97 و E98 هم به حدی وافر پدیدار شده اند.

ضروری است خاطر نشان کنم که مثلاً افزایش E00 از 1٪ تعداد کل صندوق ها به 2٪، احتمال را دو برابر کم نمی کند زیرا که چنین رابطه ای خطی نیست. مثلاً برای اصفهان تغییر تعداد E00 از 1٪ به 2٪ احتمال وقوع آنرا تقریباً 14000 بار کمتر میکند.

آزمون chisquare :

در مورد انتخابات ایران کوشش در این است که نشان دهم که آیا با حذف صندوق های ضریب 100، E00، که به شدت از انتظار به دورند، رفتار دیگر داده ها در قیاس با انتخابات سالم وین و برلین چگونه است.

برای آزمودن اینکه آیا مجموعه ای از داده ها، مانند شمار آرای صندوق های اخذ رای، که باید خصلت تصادفی داشته باشند واقعاً به طور تصادفی بدست آمده اند یا مورد دستکاری قرار گرفته اند، روشی آماری وجود دارد که آنرا آزمون chi-square مینامند.

در جدول 2 نتیجه کاربرد این آزمون بازتاب یافته است. لازم به تذکر است که چه در انتخابات سالم مانند وین و برلین و چه در انتخاباتی که ظن دستکاری و عدد سازی در آن ها وجود دارد مانند دیگر انتخابات مندرج در این جدول 2، همه داده ها به درجات معینی تصادفی اند. یعنی حتی اگر اعداد ساختگی باشند باز هم روند ساختن شان تا حدودی تصادفی است. البته اعداد ساخته شده این قید را دارند که نتیجه تجمیع آن ها باید با نتیجه از پیش تعیین شده ای همپوش باشد. از اینرو قضاوت دقیق در پاره ای از موارد مشکل است اما میتوان گرایش به سوی واقعی یا دستکاری بودن را نشان داد.

این آزمون بدو شیوه صورت گرفته است:

1. **سنجش همشکلی Uniformity :** اگر انتخاباتی سالم باشد شمار آرا صندوق ها تصادفی اند. انتظار میرود که شمار تکرار ها k برای هر دو رقم دست راست معین Emn $1/100$ تعداد کل صندوق ها N باشد. این تعداد تکرار $N/100$ را اندازه منتظر مینامند expected value. از سوی دیگر ارقامی در دست اند که شمار واقعی این تکرار ها Emn را نشان میدهند. این آزمون اندازه های واقعی را با اندازه های منتظر می سنجد. در دو ستون جدول 2 «آزمون chisq» و «آزمون chisq بدون تکرار های نا محتمل» احتمال هائی درج شده اند که میتوانند نشان بدهند آیا اختلاف معنی داری میان اندازه واقعی و منتظر وجود دارد. در آمار متداول است که چنانچه این احتمال بزرگتر از 0.05 باشد اختلاف معنی دار نیست - یعنی تصادفی است - و اگر کمتر از آن باشد اختلاف معنی دار است - یعنی غیر تصادفی و میتواند نتیجه دستکاری

باشد. این آزمون نشان میدهد که اگر همه داده ها را ملحوظ کنیم انتخابات وین و برلین کاملاً خصلت تصادفی دارند در حالیکه هیچیک از ۷ انتخابات نامبرده ایران تصادفی نیستند. حال اگر موارد تکرار های بسیار نامتعارف مانند E00 را ملحوظ نکنیم باز هم نتایج لرستان، یزد، اصفهان و اردبیل غیر تصادفی اند یا به عبارت دیگر دستکاری شده اند.

2. در سنجش با توزیع بینومیال و آزمون همپوشی با آن binomial fit

در این روش اندازه های واقعی تکرار ها با اندازه منتظر از توزیع بینومیال سنجیده میشود. از آنجا که تعداد اندازه ها در این روش در قیاس با روش همشکلی کمتر است واکنش نتیجه محاسبه نسبت به تک داده هائی که خارج از انتظار هستند شدید تر است. (در مورد نخست 100 یا با حذف E00 ها 99 زوج از داده ها داریم در حالیکه در مورد دوم بسته به بزرگی استان شمار زوج های k واقعی و k منتظر میان 14 در یزد تا 40 در تهران تغییر میکند).
آنچنان که جدول دوم نشان میدهد اگر همه داده ها را محسوب کنیم بجز انتخابات وین 2014 هیچیک با توزیع بینومیال سازگار نیستند. اما اگر نسبی بسنجیم برای وین 2015 و برلین این احتمال در ابعاد 0,0001 تا 0,001 است در حالیکه برای ایران به استثنا اردبیل احتمال تصادفی بودن و همپوش بودن با توزیع بینومیال صفر است. حال چنانچه یکی دو زوج دور از انتظار و بی معنا را حذف کنیم مشاهده میکنیم که برای دو انتخابات وین و برلین احتمال تصادفی بودن به شدت افزایش مییابد اما هنوز برای لرستان، اصفهان و اردبیل و تا حدودی یزد احتمال تصادفی و همپوش بودن با توزیع بینومیال بسیار پائین است.

نتیجه این آزمون چنین است که نه تنها تعداد زیاد ضرایب 100 (E00) غیر عادی است بلکه در مجموع تکرار k صندوق های دیگر (Emn ها با دیگر دو رقم های دست راست) نیز در سنجش با انتخابات سالم در وین و برلین از حالت عادی بدور است.

	احتمال 1	احتمال 2	احتمال 3	احتمال 4
محل انتخابات	$N/100 = \text{عدد منتظر}$ همه عداد واقعی	$N/100 = \text{عدد منتظر}$ عداد واقعی بدون E00	سنجش با توزیع بینومیال با همه تکرار ها	سنجش با توزیع بینومیال منهای اندازه های بی معنا
وین 2014	0,091	0,091	0,12	0,153
وین 2015	0,599	0,60	0,0001	0,125
برلین	0,311	0,31	0,0012	0,434
لرستان	1,09E-222	1,35E-19	0	0
یزد	8,96E-08	0,0006	0	0,062
تهران	4,90E-56	0,38	0	0,196
آذربایجان شرقی	0,002	0,51	0	0,393
اصفهان	2,79E-24	0,01	0	0,004
اردبیل	0,027	0,03	0,02	0,020
کرمانشاه	0,00010	0,87	0	0,198

جدول 2 : این جدول در بر گیرنده محاسبه احتمال 1 تا احتمال 4 تصادفی یا تعددی بودن بر مبنای آزمون chi square برای فراوانی تکرار دو رقم دست راست صندوق های رای است. ستون نخست از چپ نام محل انتخابات است. ستون دوم « احتمال 1 » برای تعداد تکرار های معین از توزیع همشکلی Emn است در قیاس با اندازه منتظر expected value که یکصدم تعداد کل صندوق های مربوطه است (N/100). ستون سوم «احتمال 2» یعنی تکرار ستون دوم پس از حذف E00 ، ستون چهارم «احتمال 3» آزمون chisq در قیاس با عدد منتظر از توزیع بینومیال و در نظر گرفتن همه داده ها و ستون پنجم «احتمال 4» تکرار ستون چهارم است پس از حذف E00 ها و تعداد کمی از اندازه های واقعی برای همه مورد هائی که معنی دار نیستند و اثرشان بر نتیجه بسیار شدید است برای استان های ایران.

II. زمینه تئوریک کاربرد، توزیع بینومیال

با تقریبی نزدیک به یقین شمار همه آرای ماخوذه در صندوق ها در انتخابات نامبرده اعدادی سه و چهار رقمی هستند. احتمال اینکه دو رقم دست راست یعنی یکان و دهگان دو رقم معین مثلا ۰۰، ۵۹، ۰۵ و غیره... باشد یک صدم است. این ارقام را در نوشته حاضر با Emn خلاصه میکنم مثلا برای همه ضرایب 100 که به 00 ختم میشوند (یعنی $m=0$ و $n=0$) یا $E00$ یا $E59$ ($m=5, n=9$) یا $E05$ و ... فراوانی تکرار اعداد Emn در صندوق های رای از راه توزیع بینومیال $binomial\ distribution$ بیان میشود. این توزیع آماری برای مواردی به گار برده میشود که:

1. رخدادی مانند شمار آرای مأخوذه در یک صندوق به رخداد دیگر یعنی شمار آرا در دیگر صندوق بستگی نداشته باشد- مانند انداختن تاس در بازی تخت نرد که نتیجه نخستین پرتاب تاس اثری در پرتاب بعدی اش ندارد.
2. احتمال رخدادهای ممکن ثابت باشد مثلا یک ششم $1/6$ برای هر تعداد چشمهای تاس از یک آزمون به آزمون دیگر. در مورد دو رقم نامبرده مجموع آرای ماخوذه به صندوق رای یکصدم $1/100$ که برای همه صندوق ها یکسان و ثابت است
3. در هر آزمون بیش از دو امکان درست یا غلط (موفق یا ناموفق) موجود نباشد. مثلا تاس یا ۶ است (موفق) یا ۶ نیست (ناموفق) یا در مورد انتخابات مورد بحث، دو رقم $E00$ شمار آراء صندوق مثلا $E00$ است (موفق) یا $E00$ نیست (ناموفق)
4. شمار آزمون ها معین باشد که در این مورد به معنی تعداد معین صندوق های رای است.

همه این چهار شرط نامبرده برای Emn معین از آرای ماخوذه صادق اند و از اینروی کار برد توزیع بینومیال برای توضیح توزیع دو رقم دست راست مجموعه آرای صندوق ها مجاز است.

با توجه به اینکه شمار آراء صندوق های مورد نظر ۲ یا چهار رقمی اند وقوع Emn معین آنها در شرایطی عاری از دستکاری، تصادفی است و با احتمال وقوع $1/100$ (یکصدم) برای هر رخداد- مثلا احتمال اینکه $E00$ یا $E45$ یک یکبار رخ دهند $1/100$ (یکصدم) است.

فزون برین خروجی صندوق ها مستقل از یکدیگرند و رخدادن هر Emn دو امکان دارد مثبت یا منفی - یعنی مثلا تعداد کل آرای یک صندوق یا ضربی است از ۱۰۰ یا چنین ضربی نیست- و شمار دفعات آزمون هر بار معین است و این همان شمار صندوق های رای در هر محدوده جغرافیایی سیاسی است - مثلا در استان تهران.

آنچنان که در بالا اشاره شد اگر بخواهیم دریابیم که مثلا از هزار صندوق رای که هریک در بر گیرنده شماری از آراء حداقل سه رقمی باشند احتمال تکرار $E00$ چیست میتوان از توزیع بینومیال استفاده کرد.

اگر تعداد آزمون ها برای رخداد X را n ، احتمال یک رخداد را p ، احتمال k بار تکرار این رخداد X را $P(X=k)$ بنامیم فرمول توزیع بینومیال:

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

که در آن $\binom{n}{k}$ کمبیناسیون n آزمون است با k بار تکرار یک رخداد.

برای مثال اگر بخواهیم محاسبه کنیم که در تهران با $n=6085$ صندوق رای احتمال اینکه یک دو رقمی معین یکان و دهگان، مثلا $E63$ ، 50 بار تکرار شود باید در فرمول بالا برای $n=6085$ ، برای $k=50$ و برای $p=0.01$ (یکصدم) بگذاریم و احتمال وقوع این رخداد را که $P(E63=50)$ پیدا کنیم.

چنانچه احتمال $P(X=k)$ را برای وقوع تکرار Emn در شمار آراء میان ۰، ۱، ۲، ۳، ... و هر اندازه دیگر را محاسبه کنیم به نقطه هایی دست خواهیم یافت که همه باهم (مکان هندسی شان) منحنی توزیع را نشان میدهند.

هر توزیع بینومیال دارای یک ماکزیمم است که اندازه مورد انتظار (expectation value) نامیده میشود و برابر است با حاصل ضرب شمار آزمون ها و احتمال وقوع یک رخداد. مثلا در تهران با شمار آزمون 6085 و احتمال ۰/۰۱ بیشترین تعداد وقوع دو رقم در تکرار $k=60$ بار یا 61 بار است. تکرار های بیشتر یا کمتر از این تعدادشان کوچکتر از این هستند.

کاربرد توزیع بینومیال در ارزیابی نتایج انتخابات خرداد ۱۳۸۸

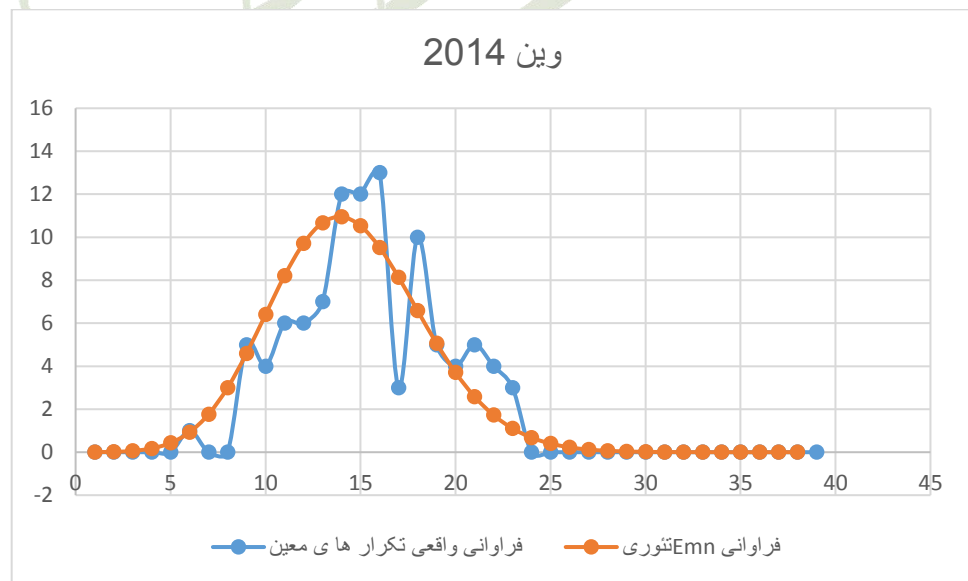
برای آزمایش این تئوری نخست کوشش کردم آنرا در انتخاباتی که شبهه ای در سلامت شان نباشد بکار بندم که همانا دو انتخابات شهر وین برای پارلمان اروپا 2014 و انتخابات ایالتی 2015 با حدود 1500 صندوق رای و انتخابات فدرال سال 2009 برلین با 2277 صندوق بوده اند.

این آزمون نشان میدهد که نتیجه عملی کسب شده در انتخابات با پیشبینی تئوریک توزیع بینومیال همخوانی دارد. نکته حائز اهمیت این است که در هیچیک از این سه انتخابات شمار تکرار Emn خارج از محدوده پیشبینی شده در تئوری وجود ندارد.

آزمودن فرض تئوریک در دو انتخابات وین و یک انتخابات از برلین به عنوان سه نمونه از انتخابات آزاد

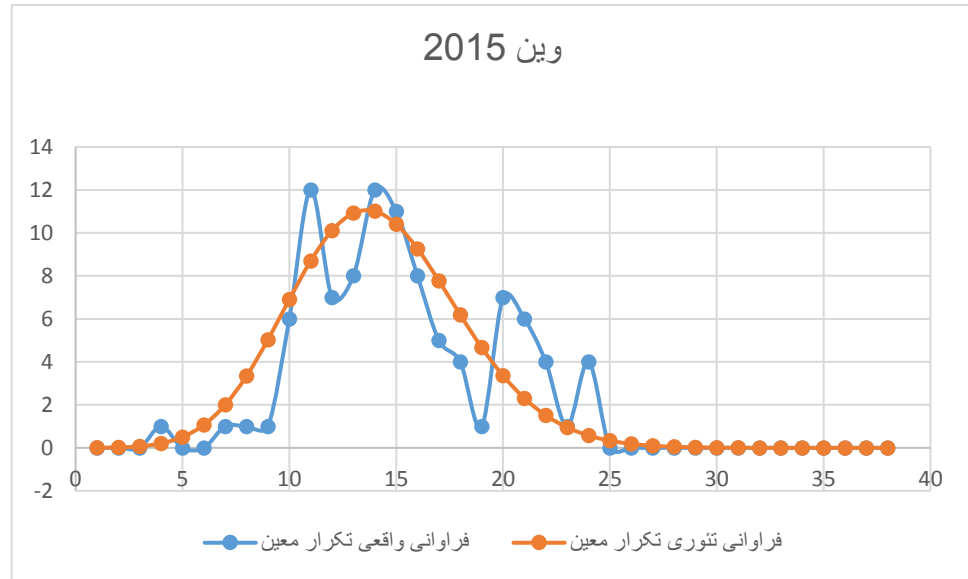
آنگونه که در سه نمودار 1، 2 و 3 زیرین برای انتخابات پارلمان اروپا در سال 2014، انتخابات مجلس ایالتی در وین و انتخابات فدرال 2009 برلین میتوان مشاهده کرد توزیع بینومیال برای دو رقم پایانی Emn از 1543 صندوق رای، خارج از محدوده ای که تئوری پیشبینی میکند یعنی تکرار Emn کمتر از 4-5 و بیشتر از 27-28 بار رخ نداده است. این امر را میتوان برای انتخابات سال 2015 وین نیز مشاهده کرد. شمار صندوق های رای در وین پیرامون ۱۵۰۰ بسیار نزدیک به کرمانشاه با بیش از 1400 صندوق و لرستان با نزدیک به 1300 صندوق. در مورد انتخابات 2009 برلین نیز با شمار 2277 بسیار نزدیک به اصفهان با 2280 صندوق است.

هر سه انتخابات نشان میدهند که فراوانی تکرار های دو رقم یکان و دهگان معین Emn در محدوده پیشبینی شده تئوری اتفاق افتاده اند.



نمودار 1 احتمال وقوع تکرار Emn های معین واقعی و تئوری در انتخابات 2014 ایالت وین را نشان میدهد. محور افقی اش تعداد تکرار ها و محور عمودی فراوانی این تکرارهای را نشان میدهد. میتوان مشاهده کرد که برای وین با حدود 1500 صندوق رای هیچ دو رقم آخری Emn شمار کل رای های

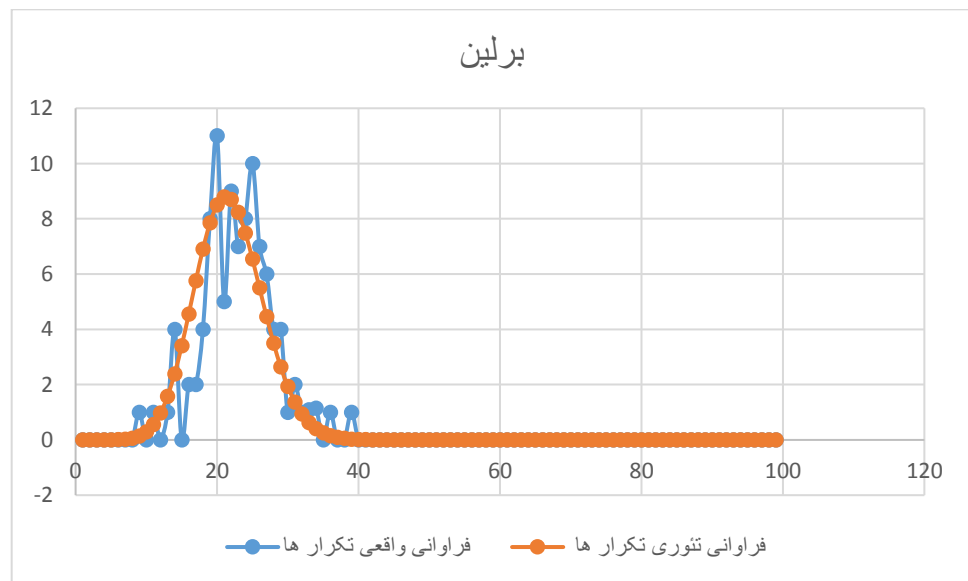
ماخوذه ای وجود ندارد که فراتر از پیشبینی تئوریک یعنی 5 بار تا 25 بار تکرار شده باشد. . فزون برین بیشترین دفعات تکرار ها با بیشترین احتمال 1٪ پیرامون 15 بار Emn رخ داده اند است- این را تعداد تکرار مورد انتظار مینامند expected value .



نمودار 2: احتمال وقوع تکرار Emn های معین واقعی (آبی) و تئوری (نارنجی) در انتخابات 2015 ایالت وین را نشان میدهد. محور افقی اش تعداد تکرار ها و محور عمودی فراوانی این تکرارهای را نشان میدهد. برای وین با حدود 1500 صندوق رای میتوان مشاهده کرد که هیچ دو رقم آخری Emn شمار کل رای های ماخوذه ای وجود ندارد که فراتر از پیشبینی تئوریک یعنی 4 بار تا 25 بار تکرار شده باشد. . فزون برین بیشترین دفعات تکرار ها با بیشترین احتمال 1٪ پیرامون 15 بار Emn رخ داده اند است- این را تعداد تکرار مورد انتظار مینامند expected value .

برلین انتخابات ایالتی 2009

برلین با نگاه به شمار 2277 صندوق های رای گیری کاملاً قابل سنجش با اصفهان (2280) است.



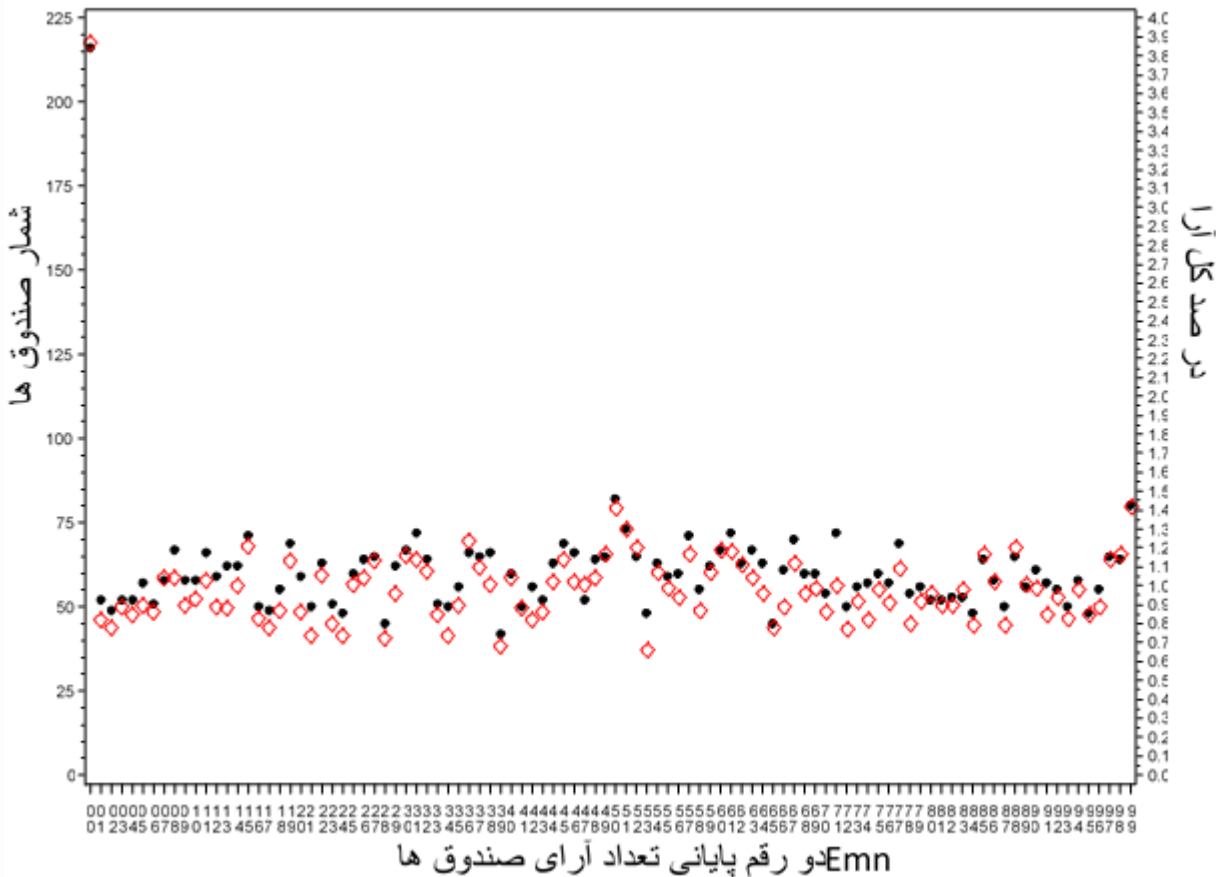
نمودار 3: نتایج انتخابات ایالتی برلین در سال 2009. محور افقی تعداد تکرار های معین k و محوری عمودی فراوانی این تکرار ها را نشان میدهد. در خور توجه است که فراوانی تکرار Emn از محدوده

قلمرو 10-40 پیشبینی شده به میانجی توزیع بینومیال فراتر نمیروند. نا محتمل ترین ها 9 و 39 با احتمال رخداد شان برابر با 0.0005 است. در سنجش با اصفهان میتوان مشاهده کرد که برای $k=46, 48, 85$ فراوانی بزرگتر از 0 ظاهر نشده است.

انتخابات تهران

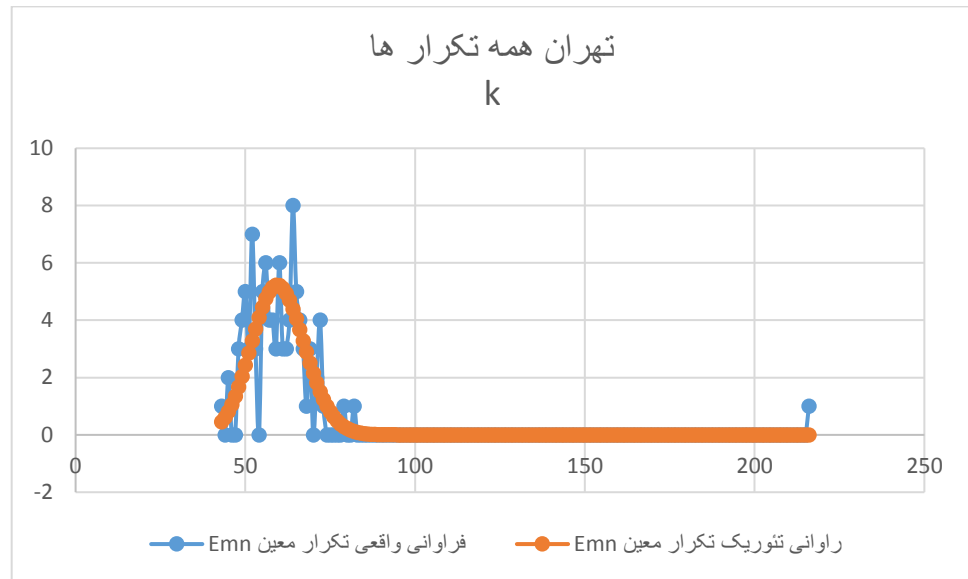
در انتخابات خرداد 1388 تهران دارای 6085 صندوق رای گیری بوده است. از این میان در 216 صندوق E00 ظاهر شده اند در حالیکه احتمال رخدادن چنین واقعه ای کمتر از 10^{-5} یعنی عملاً صفر است.

فزون بر این نگاهی به نمودار توزیع تئوریک و واقعی در تهران نشان میدهد که بقیه نقاط واقعی نیز با آنچه که تئوری پیشبینی میکند همخوانی ندارند. فزون برین برای E00 با 216 بار تکرار برای E50 با 82 بار تکرار و E99 با 79 بار تکرار نا محتمل اند.

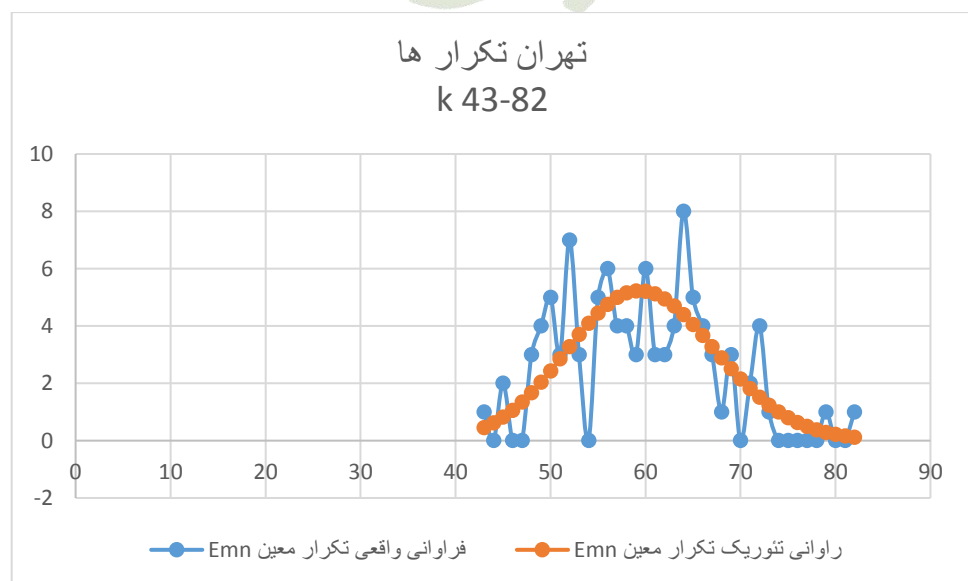


نمودار 4a: محور افقی صندوق ها با دو رقم پایانی مختلف از E00 تا ... تا E99 در انتخابات خرداد 1388 تهران را نشان میدهد. محور عمودی دست چپ تعداد تکرار ها k دو رقم پایانی و محور عمودی دست راست سهم در صد آرای هر صندوق از کل آرای ماخوذه در تهران. نقطه های سیاه تعداد صندوق ها با یک Emn معین را و نقطه های مربع شکل در صد هریک از صندوق ها از کل آرای ماخوذه تهران را نشان میدهد. آشکارا میتوان مشاهده کرد شمار صندوق ها منتهی به دو صفر E00 با 216 بار

، E50 با 82 بار تکرار و E99 با 79 بار تکرار و همچنین سهم در صدشان از مجموع آرای تهران برجسته اند. این نمودار را یکی از دوستان کارشناس آمار در اختیار م نهاده است



نمودار 4: توزیع Emn های مختلف برای انتخابات خردا 1388 در تهران. محور افقی تکرار ها k برای Emn ها و محور عمودی دفعاتی را که یک Emn معین برای یک تکرار معین ظاهر شده است (مثلا تکرار E00 برای 216 بار یکبار صورت گرفته است). از آنجا که احتمال تکرار 216 باره یک Emn عملا صفر است، میتوان نتیجه گرفت که چنین اتفاقی به طور تصادفی نمیتواند رخ دهد بلکه نتیجه دستکاری هدفمند است. توزیع تئوریک احتمال وقوع یک Emn معین نشان میدهد که در تهران تکرار های یک Emn معین در محدوده 40 تا 80 محتمل است



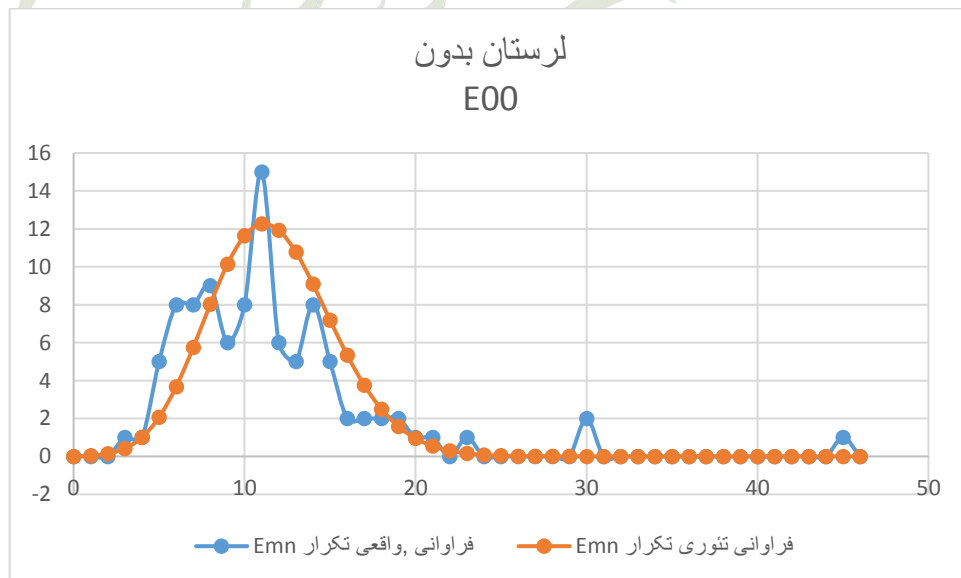
نمودار 4c: برای نشان دادن رفتار Emn در حیطه به لحاظ آماری محتمل (بدون E00) این نمودار تنها در بر گیرنده دامنه 40 تا 82 بار تکرار Emn های معین است. بروشنی میتوان مشاهده کرد که تعداد تکرار ها دارای نوساناتی شدید است. مثلا نقطه ماکزیموم تئوری در $k=60$ بکلی از نقاط ماکزیموم واقعی در تهران یعنی 52 و 65 بدورند.

انتخابات لرستان

استان لرستان دارای 1273 صندوق رای بوده است که از این میان 132 بار E00 یعنی $\frac{10.3}{100}$ درصد با دو صفر پایانی، 45 بار E50 و 30 بار E97 و 30 بار E99 که از این میان احتمال رخدادن شان به ترتیب 10^{-86} ، 10^{-13} و 10^{-5} است- دو احتمال اول عملا صفر و سومی برای وقوع، آن هم دوبار، بسیار نا محتمل. از اینرو آرای لرستان دلالت بر شدیدترین نوع دستکاری در نتایج آرا را در انتخابات خرداد 1388 به نمایش میگذارند.



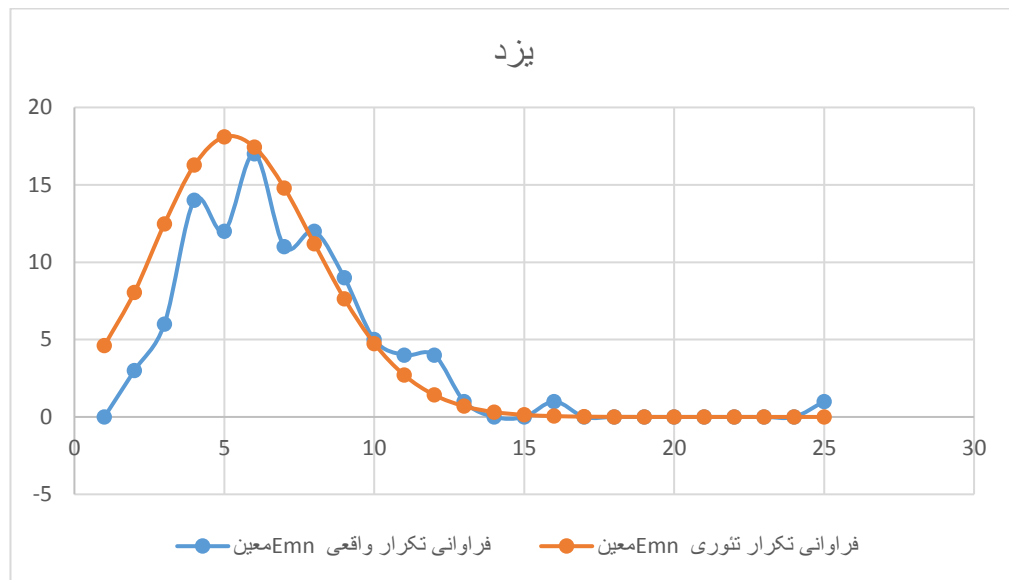
نمودار 5a: محور افقی تعداد تکرار های Emn های معین را برای صندوق های رای لرستان نشان میدهد. میتوان مشاهده کرد که تکرار 132 بار E00 و 45 بار E45 و 30 بار E97 و E98 با احتمالی به ترتیب برابر با 10^{-86} ، 10^{-13} و 10^{-5} بسیار دور از انتظار تا غیر ممکن اند.



نمودار 5b: محور افقی k یعنی تعداد تکرار های Emn های معین را برای صندوق های رای لرستان و محور عمودی فراوانی رخ دادن Emn های مختلف را بدون E00=216 نشان میدهد (E00=216 برای بهتر نشان دادن دیگر داده ها از گستره این نمودار خارج شده است). این نمودار پیشبینی تئوریک (نارنجی) را نشان میدهد که در انطباق با توزیع بینومیل فراوانی تکرار های محتمل میان 3 تا 23 دارای احتمالی عملا بزرگتر از صفر اند و خارج از این محدوده احتمالاتی عملا صفر.

استان یزد

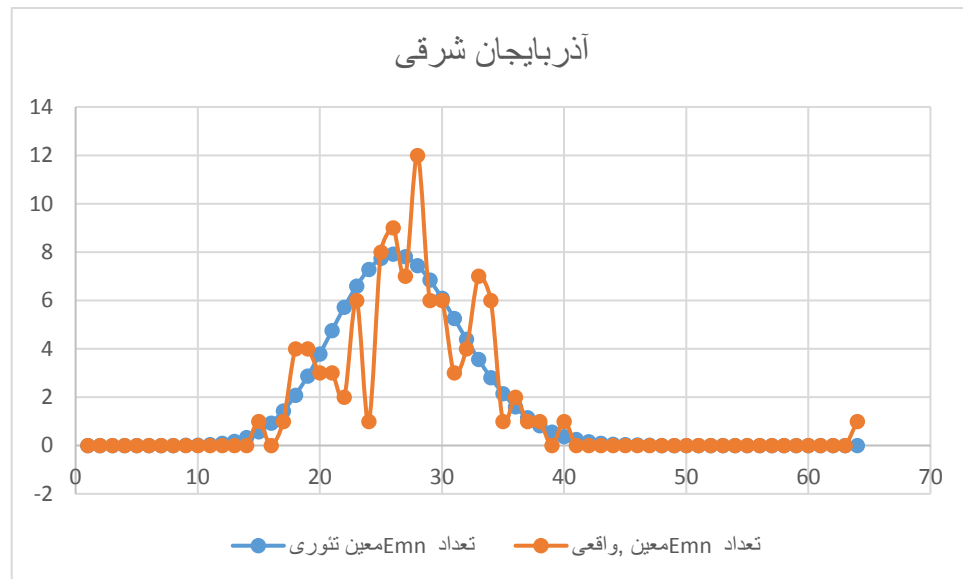
استان یزد دارای 692 صندوق اخذ رای بوده است که از این میان کل شمار ارای 25 صندوق از آن ها ضریبی از 100 یعنی E00 25 بار تکرار شده است که برابر است با 3.6٪. احتمال بروز چنین رخدادی $5,13E-08$ یا پنج صدمیلیونم است. 16 بار نیز E50 تکرار شده است با احتمالی برابر با 0,0012. جالب اینجا است که در بخش مرکزی شهر یزد با 207 صندوق 7.2٪ صندوق ها E00 هستند. اینکه در مراکز کلان شهرها مانند یزد سهم صندوق های E00 بیش از دهات و شهر های خاشیه ای و دور افتاده است ناقض این ادعا است که گویا وجود بیش از انتظار E00 ها پیامد کمبود برگ های تعرفه های رای بوده است.



نمودار 6: توزیع فراوانی واقعی (آبی) و تئوری (نارنجی) یک Emn معین در استان یزد. در این نمودار میتوان و جود صندوق هائی را مشاهده کرد مانند 25 بار و 16 بار تکرار یک Emn معین که کاملاً خارج از گستره پیشبینی شده تئوریک هستند با احتمالی به ترتیب $5,13E-08$ یا پنج صدمیلیونم و 0,0012 در قیاس با 0.01 احتمال مورد انتظار.

استان آذربایجان شرقی

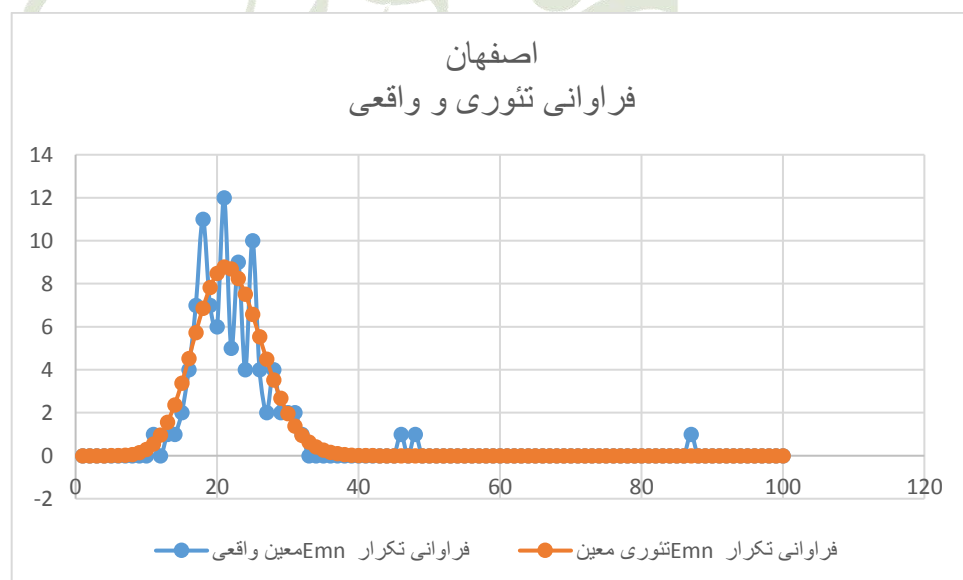
از ۲۷۶۵ صندوق رای 64 بار E00 وجود داشته اند (64 صندوق با ارقام پایانی 00 یعنی E00 که احتمال رخدادنش 1×10^{-9} - یک میلیاردم - است، وجود داشته اند). ارقام این استان نشان میدهند که در قیاس با دیگر استان های مورد مطالعه، دستکاری در آنجا نظر به ملاحظات قومی با احتیاط بیشتر صورت گرفته است.



نمودار 7: محور افقی تعداد تکرار های Emn های معین را برای صندوق های رای آذربایجان شرقی و محور عمودی فراوانی رخ دادن Emn های مختلف را نشان میدهد. میتوان مشاهده کرد که تکرار 64 باره E00 با احتمال یک میلیاردیم بسیار دور از انتظار تا غیر ممکن است همچنین میتوان مشاهده کرد که فارغ از E00 دیگر نقاط نیز تا چه اندازه پراکنده و بی سامانند اند.

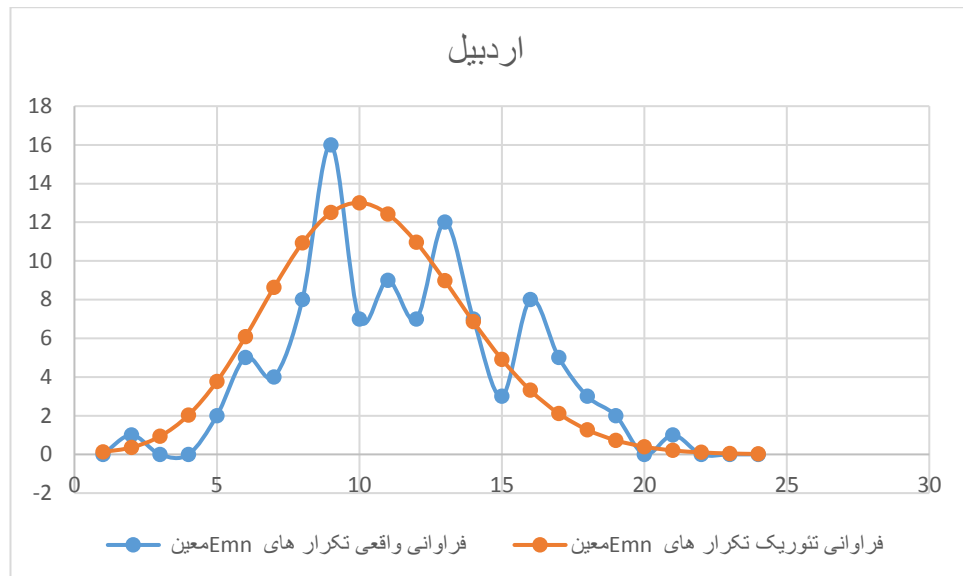
استان اصفهان:

استان اصفهان دارای 2280 صندوق اخذ رای بوده است که از میان آنها محتوای 87 صندوق ضریبی از 100 بوده است، یعنی 87 بار E00 تکرار شده است (3.82٪ کل صندوق ها). فزون برین 87 تکرار 48 بار E50 و 46 بار E99 نیز تکرار شده اند که احتمال وقوع شان به ترتیب 3×10^{-25} ، شش میلیونیم و 1.3 میلیونیم در برابر 0.01 احتمال مورد انتظار است. فزون برین از 532 صندوق مرکزی شهر اصفهان 3.95٪ ضریب 100 یعنی E00 بوده اند. در اینجا هم می بینیم که توضیح فزونی E00 ها از راه کمبود تعرفه های رای در حاشیه های دور افتاده دچار تناقض میشود.



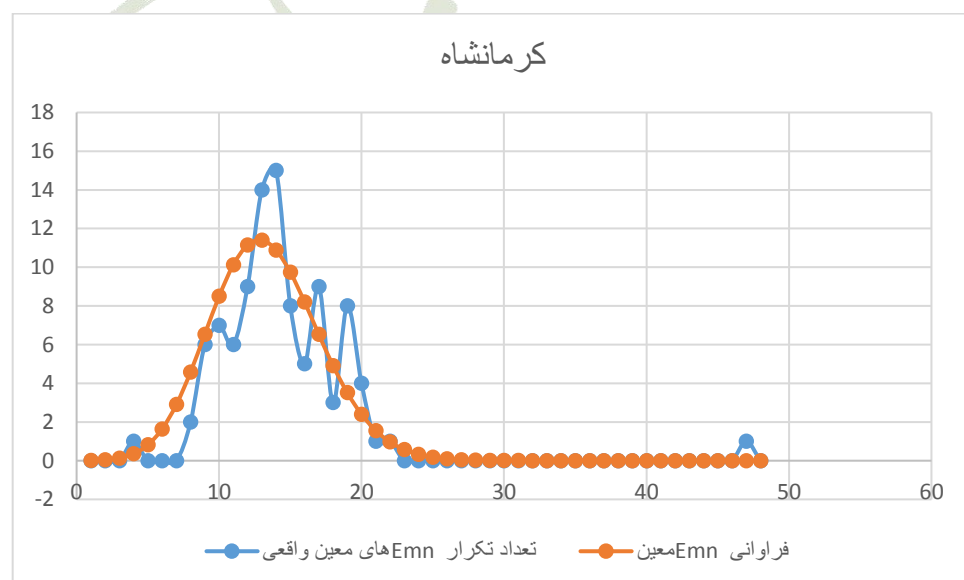
نمودار 8: توزیع فراوانی تکرار Emn های معین واقعی و تئوری برای استان اصفهان. تکرار های 87 ، 48 ، و 46 باره خارج از گستره پیشبینی شده از جانب تئوری هستند با احتمالاتی بسیار کم - به ترتیب 3×10^{25} ، 6×10^{-6} و 1.3 میلیونیم در برابر 0.01 احتمال مورد انتظار - در درون حیطه پیشبینی شده نیز تغییرات داده ها نسبت به تئوری بسیار شدید است (لطفاً این نمودار را با نمودار 3 مربوط به برلین مقایسه کنید.

استان اردبیل: این استان دارای 1156 صندوق رای گیری بوده است. از این میان تعداد تکرار ها تا حدودی در چهارچوب مورد انتظار بوده اند. همه تکرار ها در گستره تتوریک میان 1 بار تا بیست بار هستند با اندک تفاوتی برای E00 که با 22 بار تکرار احتمالی در حدود 0.001 دارد. اختلاف در در صد بروز E00 در شهر اردبیل و کل استان چندان بزرگ نیست (1.23٪ در برابر 1.91٪ به ترتیب)



استان کرمانشاه

استان کرمانشاه دارای 1435 صندوق رای بوده که از میان آن ها 47 صندوق E00 بوده اند. احتمال رخدادن چنین تکراری 3×10^{-12} و بدینگونه نزدیک به صفر است.



نمودار 9: محور افقی تعداد k تکرار های Emn های معین را برای صندوق های رای کرمانشاه و محور عمودی فرآوانی رخ دادن Emn های مختلف را نشان میدهد. میتوان مشاهده کرد که تکرار 47 باره E00 با احتمال 4×10^{-12} چهار تریلیاردیم بسیار دور از انتظار تا غیر ممکن است.

سنجش تصادفی بودن کل Emn ها

مقدمتا در خور توجه است که چه در انتخابات سالم مانند وین و برلین و چه در انتخاباتی که ظن دستکاری و عدد سازی در آن ها وجود داشته باشد، مانند انتخابات 1388 ایران (مندرج در جدول 2) ، همه داده ها به درجات معینی تصادفی اند. یعنی حتی اگر اعداد ساختگی هم باشند باز روند ساختن شان تا حدودی تصادفی است با این قید که نتیجه تجمیع آن ها باید با نتیجه از پیش تعیین شده ای همپوش باشد. از اینرو قضاوت دقیق در پاره ای از موارد مشکل است. از اینرو هدف از این آزمون تکمیل نکات مندرج در جدول 1 است که خود به اندازه کافی گویا هستند. میخواهیم بیازماییم که آیا میتوان گرایشی به سوی واقعی یا دستکاری بودن را شناسائی کرد.

برای آزمون این فرضیه که ارقام تکرار Emn نتیجه رخداد های تصادفی هستند یا نتیجه عوامل دیگر، در آمار روشی متداول است که «آزمون توان دوم شی» Chi-square Test نام دارد. در این آزمون داده های واقعی رخ داده با اندازه ای که انتظار می رود داشته باشند یعنی اندازه های منتظر expected value سنجیده میشوند و احتمال پذیرش یا رد فرضیه درستی و تصادفی بودن اعداد محاسبه میشود.

این آزمون دارای یک فرضیه صفر $H_0 = \text{null hypothesis}$ است دایر بر اینکه نتایج (ارقام) کسب شده (واقعی) اختلاف معنا داری با نتایج (ارقام) مورد انتظار (منتظر) ندارند. آنگاه یک احتمال تعریف میشود بدینگونه که اگر احتمال صحت این فرضیه بیشتر از این احتمال تعریف شده باشد در آن صورت فرضیه 0 معنا ندارد یا به عبارت دیگر نتایج کسب شده با آنچه انتظار میرفت اختلاف معنا داری ندارد و تصادفی اند (فارغ از دستکاری). برای انجام این محاسبات باید اندازه chi square محاسبه و احتمال مربوطه برای تصمیم گیری از جدول های ویژه این کار استخراج شوند. در مطالعات آماری متداول است که این احتمال مرجع 0.05 (5٪) است.

آزمون chi square را به دو شیوه انجام داده ام:

الف: آزمون همشکلی uniformity : شمار تکرار k همه دو رقم های سمت راست (Emn ها) در قیاس با عدد $n/100$ که در آن n تعداد کل صندوق های هر استان معین است.

ب: همپوشی با توزیع بینومیل binomial fit : فراوانی واقعی تکرار های Emn های معین در قیاس با اندازه منتظری که از توزیع بینومیل binomial distribution استخراج میشود.

این آزمون برای وین، برلین، لرستان، تهران، آذربایجان شرقی، یزد، اصفهان، اردبیل و کرمانشاه انجام گرفته نتایج اش در جدول 2 مندرج است.

نتیجه گیری:

نشانه های بارزی وجود دارد که تعداد آرای ریخته شده به صندوق ها اعداد تصادفی آنچنان که از یک انتخابات سالم انتظار می رود، نیستند. این واقعیت هر چند که در تعداد کل آرا که ضریبی از 100 هستند تجلی ای ویژه میابند اما منحصر به آن ها نیستند. رفتار دیگر شمار کل رای های صندوق ها که ضریبی از 100 نیستند هم خارج از انتظار است. به عبارت دیگر تقلب تنها در وجود ضریب های 100 خلاصه نمیشود بلکه در تقسیم عامدانه همه ارقام. ضریب 100 ها تنها سرخ دستکاری هدفمند را آشکار میکنند.

به احتمال قریب به یقین در نتایج انتخابات خرداد 1388 ریاست جمهوری ایران تقلب در حد عدد سازی صورت گرفته است. در لرستان بیشترین اندازه عدد سازی عریان صورت گرفته است. به نظر میاید که در مجموع، آرا صندوق ها به نحوی در میان کاندیدا ها تقسیم شده اند که به نتیجه از پیش اعلام شده برسند.

بهر روز بیات

خرداد 1395

