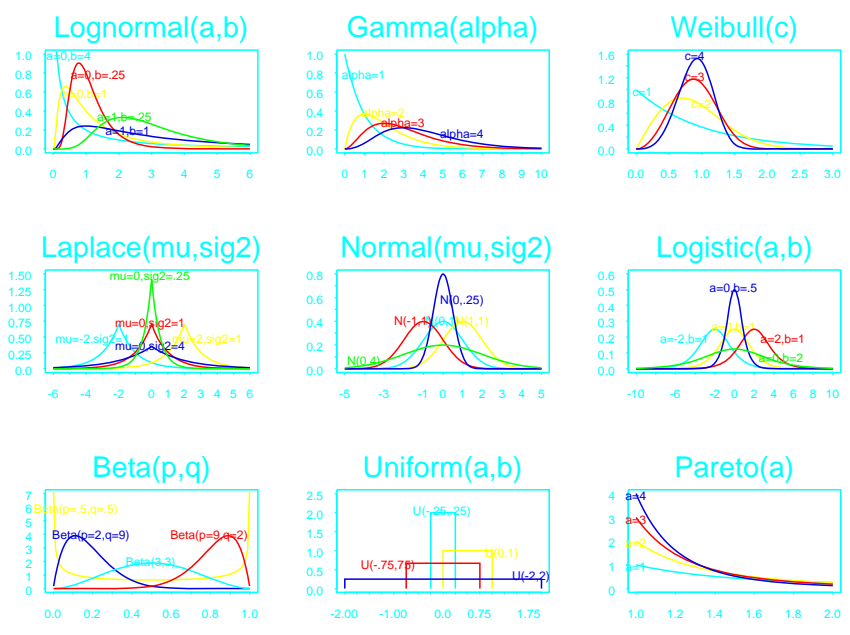




آز آمار

دستور کار آزمایشگاه مجازی روش‌های آماری با Stat Concept



عادل محمدپور
عضو هیات علمی دانشگاه صنعتی امیرکبیر

فهرست مندرجات

۹	مقدمه	۱
۱۳	نمونه	۲
۱۳	مقدمه	۱.۲
۱۳	پرسش‌ها	۲.۲
۱۹	جامعه	۳
۱۹	مقدمه	۱.۳
۲۰	پرسش‌ها	۲.۳
۲۵	آماره	۴
۲۵	مقدمه	۱.۴
۲۶	پرسش‌ها	۲.۴

لیست اشکال

۱۵	جامعه‌ی آماری با مقادیر ۰/۰۵، ۰/۱۵، ۰/۰۰۰، ۰/۹۵، ۰/۹۶	۱.۲
۱۶	بدون جایگذاری از ۵۰ گوی	۲.۲
۱۷	احتمال‌های واقعی و محاسبه شده از طریق فراوانی نسبی یک تاس	۳.۲
۱۸	نمونه «با» و «بدون» جایگذاری	۴.۲
۲۱	پنج عضو خانواده نرمال	۱.۳
۲۱	تابع چگالی و چگالی تجربی توزیع نرمال	۲.۳
۲۳	توزیع کی دو با درجات آزادی گوناگون	۳.۳
۲۸	توزیع نظری و تجربی آماره Z	۱.۴
۲۹	توزیع نمونه‌ایی \bar{X} و منحنی نرمال	۲.۴
۳۱	تقریب توزیع دو جمله‌ای با توزیع پواسن و نرمال	۳.۴

پیشگفتار

صاحب نظران در امر آموزش آمار معتقدند که شیوه‌های نوین در آموزش آمار به کندی جایگزین شیوه‌های کهنه می‌شوند. به عنوان مثال سالها طول کشید تا مفهوم جمله زیر در میان مدرسین آمار جا باز کرد:

آموزش آمار شبیه علوم آزمایشگاهی نظیر شیمی و فیزیک است،
نه شبیه ریاضیات سنتی

البته کندی تغییر در سیستم‌های آموزشی مختص رشته خاصی نیست ولی در رشته آمار گرفتار پایبندی بسیاری از مدرسین به تدریس مطالب به شیوه‌های سنتی و همچنین کمبود امکانات آزمایشگاهی می‌باشد. خوشبختانه در سالهای اخیر امکان ساخت آزمایشگاه‌های پیشرفته و درعین حال ارزان قیمت آمار فراهم شده است! توسعه فن آوری اطلاعات، شامل امکانات سخت‌افزاری و نرم‌افزاری رایانه‌ای، اینترنت و ...، امکان شبیه‌سازی آزمایشگاه‌های آمار را بر روی رایانه امکان‌پذیر نموده است، و هم اکنون دانشجویان از طریق اینترنت و یا CD می‌توانند به این آزمایشگاه‌های مجازی دسترسی داشته باشند.

کتاب حاضر دستور کار برای یک واحد عملی درس روش‌های آماری بر اساس یک آزمایشگاه مجازی به نام StatConcept (مفاهیم آماری) که بطور مختصر آنرا «آز» می‌نامیم، است. فصل‌های کتاب بر اساس رئوس مطالب درس روش‌های آماری تهیه شده که کاملاً با آز مطابقت دارد. بطور خلاصه ویژگی‌های آز را می‌توان بصورت زیر بیان نمود:

- کار کردن با آن بسیار ساده است و نیازی به آشنایی قبلی به یک نرم‌افزار و یا حتی سیستم عامل ندارد.

- برای درک بهتر اهداف آژ و استفاده بهینه از آن یک کتاب راهنما تهیه شده است.
- آژ کپی‌رایت ندارد و قابل توزیع می‌باشد.
- آژ به زبان S-PLUS که خود یک نرم‌افزار آماری است، نوشته شده است.
- متن برنامه‌های رایانه‌ای آژ در دسترس می‌باشد (CD ضمیمه کتاب) در نتیجه دانشجویان نکات بسیاری از برنامه‌های نوشته شده می‌آموزند و با کمی ابتکار می‌توانند با تغییر برنامه‌ها و یا نوشتن برنامه‌های جدید مفاهیم آماری دلخواه خود را در بوته آزمایش قرار دهند.
- نسخه‌ای از آژ تحت نرم‌افزار Stata به همراه کتاب راهنمای آن موجود است^۱.

البته این نرم‌افزار محدودیت‌های هم دارد. اگر از برخی اشکالهای محاسباتی که از نرم‌افزار S-PLUS به ارث برده بگذریم عدم توجه آن به داده‌های^۲ واقعی می‌باشد. در واقع همانگونه که از نام آن پیداست این نرم‌افزار برای تفهیم مفاهیم آماری طراحی شده نه کار کردن با داده‌های واقعی، و برای این منظور می‌توان از نرم‌افزارهای دیگری در این خصوص نظیر STEPS استفاده نمود.

امیدواریم که این کتاب بتواند دانشجویان را در درک مفاهیم روش‌های آماری بنحو مطلوبی یاری دهد. همچنین امید داریم که مدرسین محترم درس روش‌های آماری بیش از پیش بسوی روش‌های آموزشی نوین حرکت کنند و دانشجویان را در راه استفاده از این کتاب یاری دهند. از همه کسانی که این کتاب مطالعه می‌کنند خواهشمندیم که ما را از نظرات و پیشنهادات اصلاحی خویش بی‌بهره نسانند. نظرات و پیشنهادات سازنده خوانندگان محترم یقیناً ما را در ادامه حرکت بسوی تدوین یک ابزار آموزشی مفید یاری خواهد داد.

در پایان بر خود واجب می‌دانم از آقایان دکتر نواب‌پور (رییس پژوهشکده آمار) بخاطر حمایت بی‌دریغ‌شان، ایزدی و رنجی (سرپرست و عضو گروه آموزش آمار پژوهشکده آمار) بخاطر ویراستاری و کمک‌های فنی، امیری، جمالی، و کلانتری (دانشجویان کارشناس ارشد گروه آمار دانشکده ریاضی و علوم کامپیوتر دانشگاه

^۱ Newton, H.J. and Harvill, J.L. (1997) StatConcepts: A Visual Tour of Statistical Ideas. Duxbury Press: Pacific Grove, CA.

صنعتی امیرکبیر) بخاطر حروف چینی و ترجمه راهنمای Stat Concept تشکر کنم.

آدرس‌ها

• پژوهشکده آمار

<http://www.src.ac.ir>

E-mail: src@src.ac.ir

• مؤلف

E-mail: adel@aut.ac.ir

E-mail: mohammadpour_adel@yahoo.com

روزی که برای نخستین بار وارد گروه آمار شدم، نامی که در کنار دریکی از اتاق‌ها نوشته شده بود نظرم را جلب کرد. آزمایشگاه آمار نام آن اتاق بود. البته بعداً شایعاتی در مورد اتاق می‌شنیدم که گفته می‌شد در این اتاق چند سکه، تاس و ... وجود دارد. بعد از ورود رایانه به گروه قرار شده بود رایانه‌ها را در این اتاق بگذارند همه کنجکاو شده بودند که بالاخره در این اتاق چه وسایلی است. خیلی عجیب بود یک قفسه کتاب، که البته دوباره شایعه شده بود که تعداد کتابهای این قفسه ۵۲ است!

فصل ۱

مقدمه

بسیاری از تعاریف و تعابیر در آمار بر اساس عبارت زیر بیان می‌شوند:

«اگر این عمل را بارها و بارها انجام دهیم چنین خواهد شد»

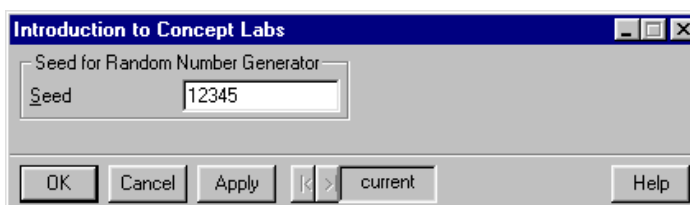
مانند تعبیر احتمال^۱ از طریق فراوانی نسبی^۲ و ولی ما برای انجام چنین کارهایی نیاز به مشاهدات^۳ زیادی از متغیرهای تصادفی^۴ گوناگون پیدا می‌کنیم، که مستلزم صرف زمان و هزینه زیادی می‌باشد. برای رفع این مشکل می‌توان به کمک رایانه مشاهدات تصادفی از توزیع‌های^۵ مختلف تولید کرد. به این کار اصطلاحاً شبیه‌سازی^۶ (شبیه‌سازی آماری^۷) و به مشاهدات تولید شده اعداد شبه‌تصادفی^۸ (به اختصار اعداد تصادفی) گفته می‌شود. دلیل به کار بردن واژه شبه‌تصادفی، تکرار اعداد تولید شده پس از یک دوره می‌باشد. یعنی یک عدد طبیعی N وجود دارد که مشاهده $x_{N+1} = x_1, x_{N+2} = x_2, \dots$ خواهد بود. هر چقدر N بزرگتر باشد تکرار اعداد تولید شده دیرتر اتفاق می‌افتد. طولانی شدن دوره تولید اعداد تصادفی بستگی به برنامه رایانه‌ای آن و عدد اولیه برنامه که دانه^۹ می‌نامیم دارد. معمولاً در نرم‌افزارها برنامه شبیه‌سازی قابل تغییر نیست ولی دانه مورد استفاده را می‌توان تغییر داد. در هر

Probability^۱
Relative Frequency^۲
Observations^۳
Random Variables^۴
Distributions^۵
Simulation^۶
Statistical Simulation^۷
Pseudo Random Number^۸
Seed^۹

برنامه رایانه‌ای دانه دارای شرایط متفاوتی است ولی معمولاً یک عدد طبیعی با ارقام زیاد و فرد (و یا حتی اول) می‌باشد اگر دانه توسط کاربر مشخص نشود خود نرم‌افزار این عدد را انتخاب می‌کند که معمولاً از تابع Time (زمان) رایانه استفاده می‌کند. در آراین عدد بصورت یک عدد ثابت در نظر گرفته شده است ولی کاربر می‌تواند آن را با استفاده از گزینه Introduction to Concept Labs^{۱۰} تغییر دهد.

با این مقدمه و عنایت به سرعت رایانه‌ها در انجام کارهای تکراری، دلیل اصلی گسترش آزمایشگاه‌های مجازی آمار تا حدودی روشنتر شده است. فصل‌های بعدی کتاب هر یک دستور کاری برای یک جلسه کار عملی در آزمایشگاه می‌باشند. برای کارایی بیشتر به نکات زیر توجه کنید:

- ترتیب فصل‌های کتاب مطابق اکثر کتاب‌های موجود تهیه شده ولی برخی از فصل‌ها جابجا پذیر می‌باشند.
- پرسش‌های هر فصل جابجا پذیر نیستند و غالباً به پرسش‌های قبلی خود در ارتباط می‌باشند، که پرسشی که مرتبط به پرسش قبل است دارای علامت ↑ می‌باشد.
- قبل از شروع هر جلسه آزمایشگاه، دانشجویان باید راهنمای آز مربوطه را از کتاب راهنمای آز مطالعه نمایند.
- معمولاً زمان کافی برای نوشتن گزارش کار در هر جلسه برای دانشجویان مهیا نمی‌باشد لذا توصیه می‌شود کلاس‌های آز یک هفته در میان برگزار شوند تا دانشجویان زمان کافی برای تهیه گزارش کار داشته باشند.
- از آنجا که پرسش‌های کتاب محدود است، طرح پرسش‌های دیگر توسط اساتید کمک زیادی به دانشجویان خواهد نمود و از طرفی ابتکاراتی که از ناحیه خود دانشجویان در انجام آزمایش‌های جدید زیر نظر اساتید صورت می‌گیرد کمک بیشتری به فهم مطالب توسط دانشجویان خواهد نمود.
- از آنجا که این کتاب یک درسی نیست استفاده از آن بدون معلم توصیه نمی‌شود.



شبیه‌سازی آماری چه وجه مشترکی با ماشین‌های تعلیم رانندگی دارد؟

سرویس‌های خوابگاه ۵ دقیقه ۵ دقیقه بود، ولی کم اتفاق می‌افتاد که این زمان توسط آنها حتی برای چند بار در روز رعایت می‌شد. هر بار که از سرویس جا می‌ماندم، بیشتر از ۱۰ دقیقه منتظر می‌شدم. در چنین مواقعی قضیه‌ای که در نظریه صف بندی خوانده بودیم یعنی: اگر زمان سرویس دهی مساوی باشد طول صف نصف می‌شود در ذهنم تداعی می‌شد (هر چند که چند بار به آنها تذکر دادیم.) و پیش خود می‌گفتم خوب که معاون دانشجویی و فرهنگی از بخش خودمان است وگرنه ۱۰۰ درصد به صورت تصادفی در می‌آمدند.

فصل ۲

نمونه

۱.۲ مقدمه

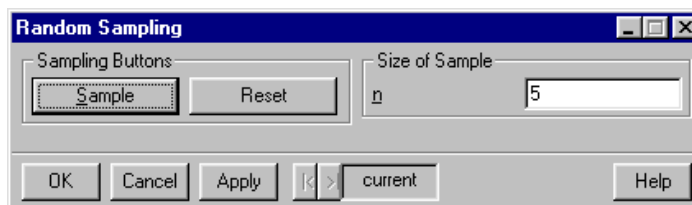
در درس روش‌های آماری مهمترین و قوی‌ترین فرض برای استفاده از روش‌ها، تصادفی بودن نمونه^۱ می‌باشد. متأسفانه هیچ روش توانمندی جهت تشخیص داده‌های مستقل^۲ و هم‌توزیع^۳ در اختیار نداریم؛ ولی برای رد کردن فرض استقلال^۴ و هم‌توزیعی داده‌ها چندین روش وجود دارد که در آینده با آنها آشنا خواهید شد.

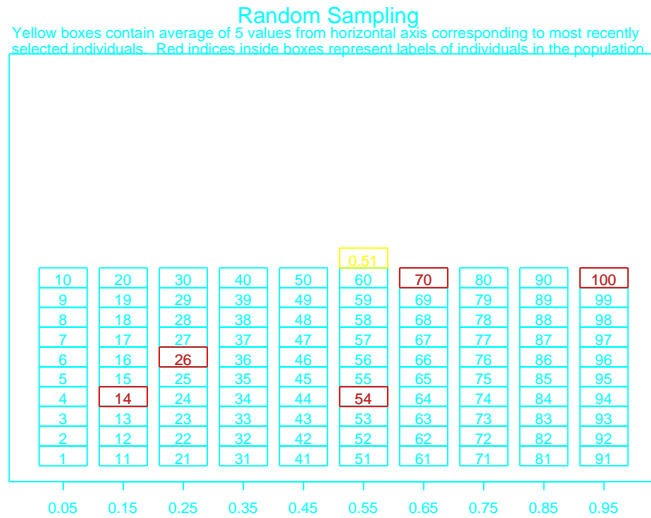
۲.۲ پرسش‌ها

Sample^۱
Independent^۲
Identical^۳
Independence^۴

- (۱) چگونه از یک جامعه^۵ متناهی^۶ یک نمونه تصادفی انتخاب می‌کنید؟ معمولاً به جای عبارت «نمونه‌ای که به روش نمونه‌گیری تصادفی ساده^۷ جمع آوری شده» از «نمونه تصادفی» استفاده می‌شود.
- (۲) چگونه از یک جامعه نامتناهی^۸ یک نمونه تصادفی انتخاب می‌کنید؟ یک مثال بیاورید.
- (۳) آیا روش پیشنهادی شما در دو مورد قبیل عملی می‌باشد؟ در صورت منفی بودن پاسخ، یک روش عملی بیان کنید.
- (۴) چگونه با استفاده از جدول اعداد تصادفی یک نمونه تصادفی از یک جامعه انتخاب می‌کنید؟
- (۵) آیا انتخاب یک نمونه تصادفی توسط رایانه امکان‌پذیر است؟
- (۶) آیا توسط ماشین حساب می‌توانید این کار را انجام دهید؟ چگونه؟
- (۷) چرا بر تصادفی بودن نمونه تاکید می‌کنیم؟
- (۸) در کلاس درس خود ۳ نفر را با سه روش متفاوت به‌طور تصادفی انتخاب کنید.
- (۹) اگر یک جامعه به صورت شکل ۱.۲ در اختیار داشته باشید چگونه یک نمونه تصادفی ۵ تایی از آن انتخاب می‌کنید؟
- (۱۰) ↑ با استفاده از گزینه Random Sampling^۹ (شکل ۱.۲) در آزار این کار را انجام دهید.

Population^۵
Finite^۶
Simple Random Sample^۷
Infinite^۸
۹





شکل ۱.۲: جامعه‌ی آماری با مقادیر ۰/۹۶، ۰/۹۵، ۰/۹۰، ۰/۸۵، ۰/۸۰، ۰/۷۵، ۰/۷۰، ۰/۶۵، ۰/۶۰، ۰/۵۵، ۰/۵۰، ۰/۴۵، ۰/۴۰، ۰/۳۵، ۰/۳۰، ۰/۲۵، ۰/۲۰، ۰/۱۵، ۰/۱۰، ۰/۰۵، ۰/۰۰.

۱۱) ↑ توزیع جامعه چیست؟ تابع جرم احتمال^{۱۰} (چگالی^{۱۱}) این جامعه را بنویسید؟

۱۲) ↑ آیا با استفاده از نمونه‌های تصادفی می‌توانید به سؤال قبل پاسخ دهید؟ (راهنمایی: از گزینه Random Sampling، حداقل ۲۵ نمونه یک‌تایی از جامعه انتخاب کرده و به هیستوگرام تولید شده نگاه کنید.)

۱۳) ↑ جدول فراوانی نسبی^{۱۲} نمونه انتخاب شده را رسم کنید، آیا می‌دانید که این جدول تقریب چیست.

۱۴) ↑ چه نتیجه‌ای از سوالات قبل گرفته‌اید؟

۱۵) آیا می‌توانید تعریفی برای نمونه خوب ارائه دهید؟

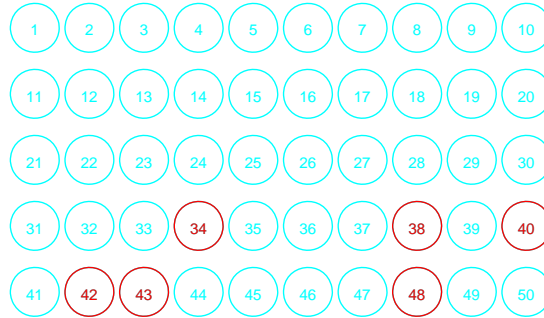
۱۶) با استفاده از نمونه تصادفی، احتمال را از طریق فراوانی نسبی تعبیر کنید. برای تعبیر احتمال از نمونه تصادفی با جایگذاری استفاده می‌کنید یا بدون

^{۱۰} Probability Mass Function

^{۱۱} Density

^{۱۲} Relative Frequency Table

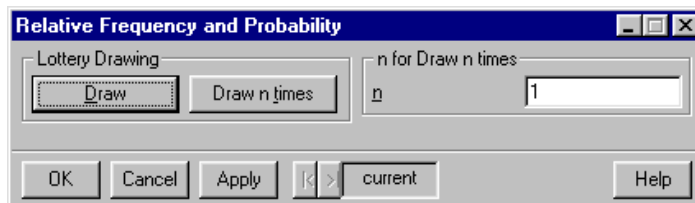
ndraws = 1 , npairs = 1 , relative frequency = 1

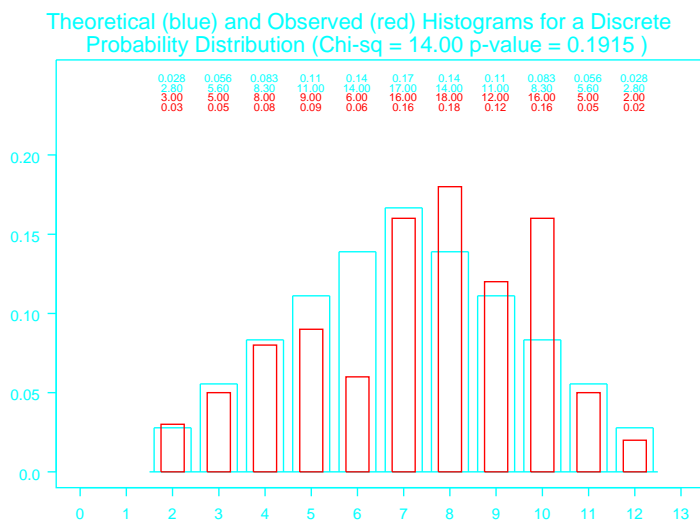


شکل ۲.۲: تعبیر احتمال از طریق فراوانی نسبی با استفاده از نمونه گیری بدون جایگذاری از ۵۰ گوی

جایگذاری؟

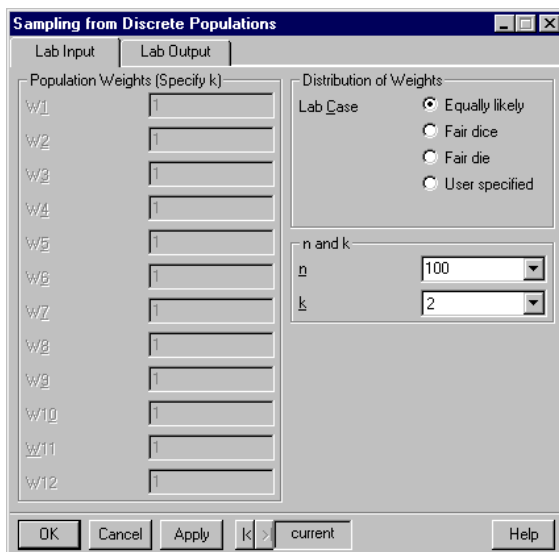
۱۷) ↑ با استفاده از گزینه Relative Frequency and Probability^{۱۲} (شکل ۲.۲) در آزدریک مسئله واقعی احتمال را از طریق فراوانی نسبی تعبیر کنید.

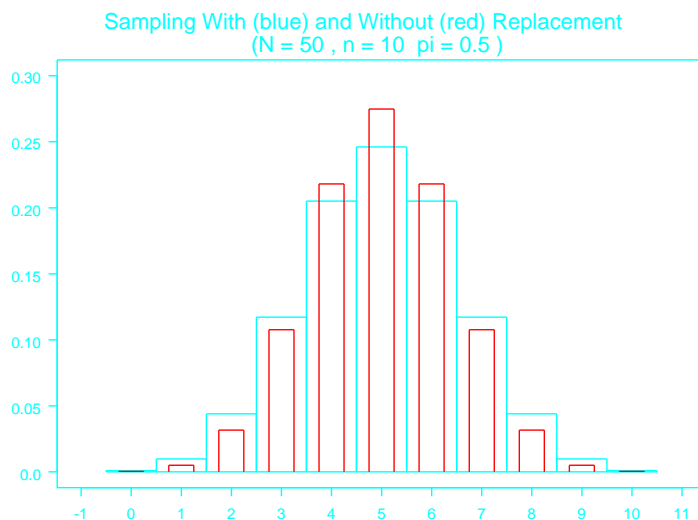




شکل ۳.۲: احتمال‌های واقعی و محاسبه شده از طریق فراوانی نسبی یک تاس

۱۸) با استفاده از گزینه Sampling from Discrete Populations^{۱۴} (شکل ۳.۲) در آزمون‌های خالهای تاس را از طریق فراوانی نسبی تعبیر کنید.





شکل ۴.۲: نمونه «با» و «بدون» جایگذاری

۱۹) با استفاده از گزینه

Sampling from 0-1 Population > Sampling With and Without Replacement

(شکل ۴.۲) در آزمون تفاوت بین نمونه‌گیری با جایگذاری^{۱۵} و بدون جایگذاری^{۱۶} را تحقیق کنید. (تفاوت بین توزیع دوجمله‌ای^{۱۷} و توزیع فوق هندسی^{۱۸})

پرسش من: تاسی را به هوا پرتاب می‌کنیم احتمال اینکه سر او بشکند را به چه روشی تعبیر می‌کنید؟
پاسخ او: از طریق فراوانی نسبی!
و گفتم در مورد خودت نیز به همین روش تعبیر را بکار می‌بری!

^{۱۵} With Replacement
^{۱۶} Without Replacement
^{۱۷} Binomial
^{۱۸} Hypergeometric

فصل ۳

جامعه

۱.۳ مقدمه

اگر گام نخست را در استفاده از روش‌های آماری، اطمینان از تصادفی بودن داده‌های جمع‌آوری شده بدانیم، گام بعدی تشخیص توزیع جامعه‌ای است که نمونه از آن انتخاب شده است. روش‌های بسیاری برای پی بردن به توزیع جامعه از طریق نمونه تصادفی وجود دارد، ولی نکته‌ای که در این روش‌ها نهفته است، و شاید کمتر به آن توجه می‌شود، این است که بیشتر آنها شرط لازم برای پی بردن به توزیع جامعه از طریق نمونه تصادفی را بررسی می‌کنند. بعنوان مثال یکی از روش‌هایی که ما برای پی بردن به توزیع جامعه استفاده می‌کنیم محاسبه و رسم تابع چگالی تجربی^۱ داده‌ها (که به آن هیستوگرام^۲ هم می‌گویند) می‌باشد نحوه صحیح استفاده از این روش به این صورت است که، اگر تابع چگالی تجربی داده‌ها را رسم کردیم و دیدیم که زنگی شکل نیست، فرض نرمال بودن جامعه را رد می‌کنیم ولی اگر زنگی شکل بود نمی‌توانیم بگوییم جامعه نرمال^۳ است، و اگر دلیل دیگری برای رد فرض نرمال بودن نیافتیم، می‌توانیم بگوییم که دلیلی برای رد فرض نرمال بودن جامعه نداریم. دلیل این

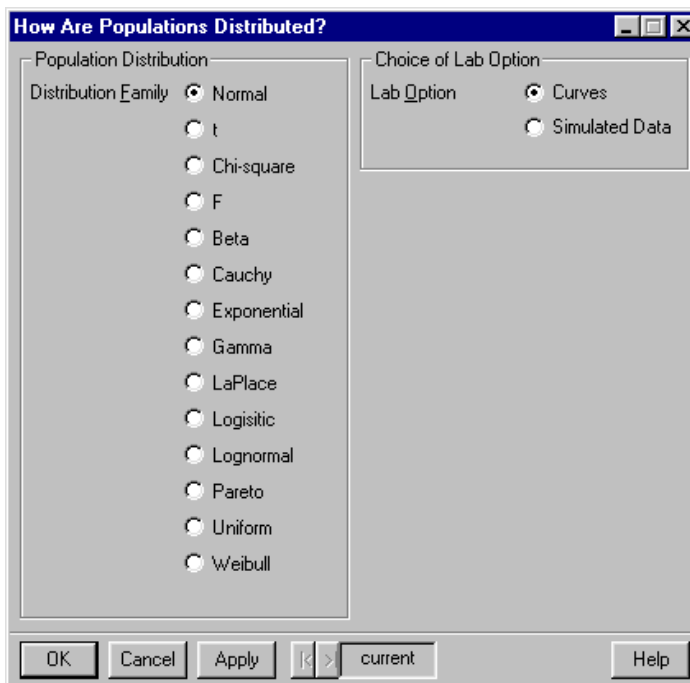
^۱ Empirical Density Function
^۲ Histogram
^۳ Normal

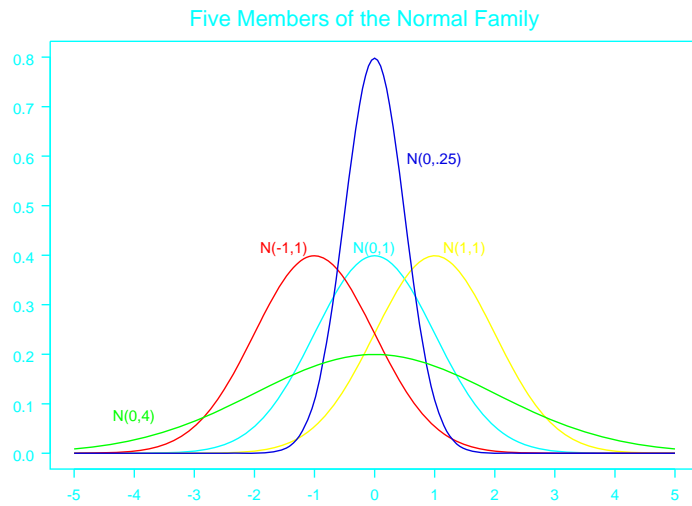
امر وجود توزیع‌های بی‌شماری است که شکل تابع چگالی آنها همگی آنها زنگی شکل است. در این فصل مروری بر توزیع‌های مشهور پیوسته و برخی از ویژگی‌های آنها داریم.

۲.۳ پرسش‌ها

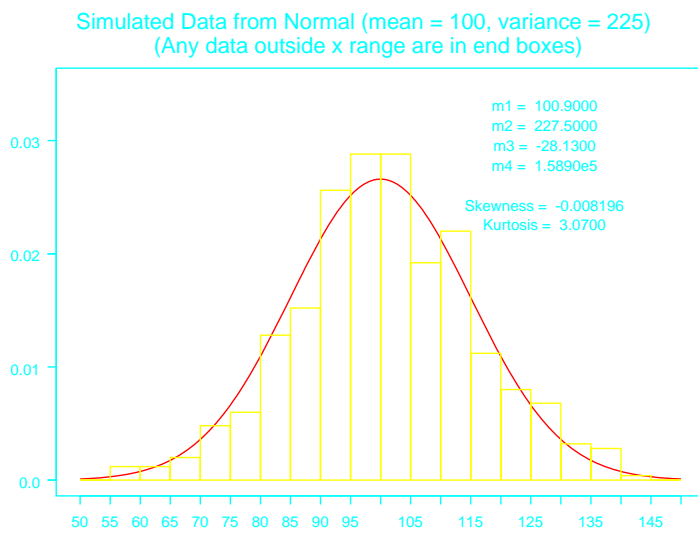
(۱) با استفاده از گزینه How are Population Distributed^۴ (شکل ۱.۳) در آژ تابع چگالی تعدادی از متغیرهای تصادفی را که قبلاً با آنها آشنا بوده‌اید رسم کرده و تفاوت آنها را با یکدیگر بیان کنید. چه معیارهایی مورد توجه شما قرار گرفته است؟

(۲) ↑ تابع چگالی تجربی هریک از متغیرهای تصادفی را همراه با تابع چگالی آنها، با انتخاب Simulated Data از قسمت Lab Option رسم کنید (شکل ۲.۳). توجه کنید که اگر تابع چگالی تجربی (هیستوگرام) را چند بار رسم کنیم، شبیه هم نخواهند شد. چرا؟





شکل ۱.۳: پنج عضو خانواده نرمال



شکل ۲.۳: تابع چگالی و چگالی تجربی توزیع نرمال

(۳) ↑ چه تعبیر و یا تفسیری برای تابع چگالی تجربی و تابع چگالی می‌توانید ارائه دهید.

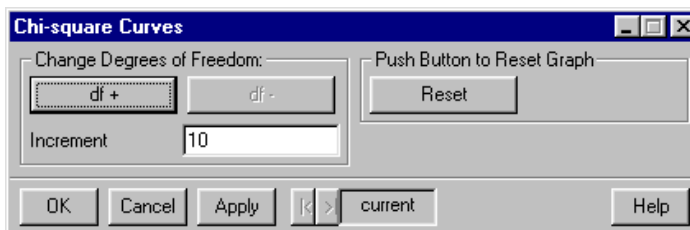
(۴) ↑ با استفاده از اعدادی که کنار تابع چگالی تجربی نرمال نوشته شده، معیارهای دیگری (غیر از شکل ظاهری) برای تشخیص نرمال بودن جامعه، از طریق محاسبه این معیارها از نمونه ارائه دهید. این معیارها عبارتند از گشتاورهای اول تا چهارم، چولگی و برجستگی.

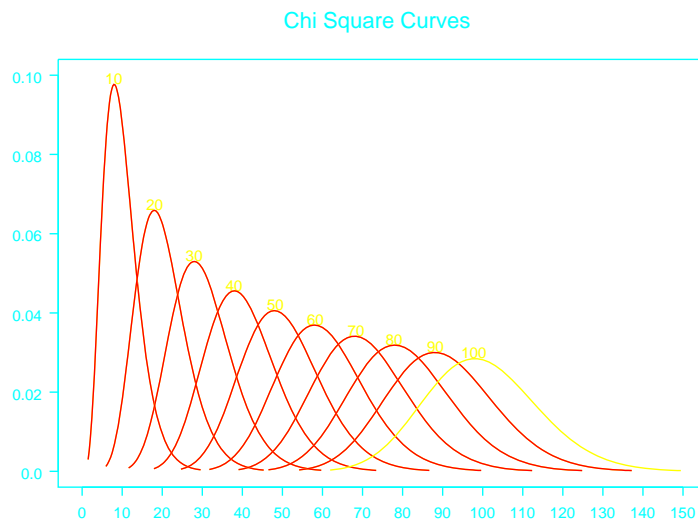
(۵) ↑ مقادیر این معیارها برای جامعه‌ی نرمال که نمودار تابع چگالی آن را رسم کرده‌اید چقدر است. (از معلم خود کمک بگیرید)

(۶) ↑ با علامت گذاشتن در Simulated Data از قسمت Lab Option به جای رسم منحنی متغیرهای تصادفی ذکر شده داده‌های مستقل و هم‌توزیع تولید و رسم کنید (نمونه تصادفی). اندازه نمونه‌ها را در این قسمت برابر ۵۰۰ قرار دهید. به نظر شما آیا داده‌های رسم شده با منحنی واقعی تفاوت زیادی دارند؟ چرا؟

(۷) با استفاده از گزینه Z , t , Chi-Square, F (شکل ۳.۳) در آزمون تاثیر تفسیر پارامترهای توزیع‌هایی را که می‌شناسید، از روی نمودار تابع چگالی بیان کنید.

سازمان ترافیک برای انجام طرحی که در دست داشت شروع به آمارگیری در پلیس راه‌های شهر کرده بود. نحوه نمونه‌گیری سیستماتیک بود. به این صورت که از هر پنج اتومبیل سواری که وارد شهر می‌شد، پنجمی به عنوان نمونه انتخاب و از فرد کنار دست راننده و پشت سری او سوال می‌شد که از کجا می‌آید و به کجا می‌رود. معمولاً همسر راننده کنار دست او نشسته بود، وقتی که آمارگر از همسر او سوال می‌کرد ... راننده پیاده می‌شد (البته در بعضی موارد) و با دعوا می‌گفت چرا از من سوال نمی‌کنی! تا آمارگر می‌آمد نمونه‌گیری سیستماتیک را برای او توضیح دهد دیگر کار از کار گذشته بود...!





شکل ۳.۳: توزیع کی دو با درجات آزادی گوناگون

فصل ۴

آماره

۱.۴ مقدمه

هر تابعی از نمونه تصادفی که به پارامتر مجهول بستگی نداشته باشد یک آماره^۱ نامیده می‌شود. اهمیت آماره‌ها از آنجا ناشی می‌شود که برخی از آنها اطلاعات موجود در نمونه تصادفی در خصوص یک یا چند پارامتر^۲ جامعه را تلخیص می‌کنند. از آنجا که آماره تابعی از نمونه تصادفی است، خود نیز یک متغیر تصادفی می‌باشد. توزیع آماره از اهمیت ویژه‌ای برخوردار است، زیرا برای استفاده بهتر از اطلاعات خلاصه شده در آماره نیاز به توزیع آماره داریم. در این فصل به بررسی توزیع آماره‌ها، توزیع نمونه‌ای^۳ آماره و دو قضیه که در مورد برخی آماره‌ها کاربرد بسیاری دارد یعنی قانون اعداد بزرگ^۴ و قضیه حد مرکزی^۵ می‌پردازیم.

^۱Statistic

^۲Parameter

^۳Sampling Distribution

^۴Weak Law of Large Number

^۵Central Limit Theorem

۲.۴ پرسش‌ها

• توزیع آماره

(۱) اگر X_1 و X_2 یک نمونه تصادفی به اندازه ۲ از توزیع برنولی^۶ با پارامتر p باشند توزیع آماره $T = X_1 + X_2$ چیست؟ اگر x_1 و x_2 مشاهدات X_1 و X_2 باشند آیا $t = x_1 + x_2$ حاوی اطلاعاتی در خصوص p می‌باشد؟

(۲) اگر X_1 و X_2 یک نمونه تصادفی به اندازه ۲ از توزیع نرمال با پارامترهای μ و σ^2 باشد توزیع آماره $\bar{X} = \frac{1}{2}(X_1 + X_2)$ چیست؟ این آماره حاوی اطلاعات در خصوص کدامیک از پارامترهای جامعه است؟

• توزیع نمونه‌ای آماره

(۳) با استفاده از گزینه Random Sampling از یک نمونه تصادفی به اندازه ۲ از جامعه بگیریید و به مقدار میانگین^۷ آن در شکل توجه کنید. اگر این عمل را چندین بار (مثلاً ۲۰ بار) انجام دهید چه نموداری مشاهده می‌کنید؟ به نمودار بدست آمده شکل توزیع نمونه‌ای آماره میانگین با استفاده از ۲ مشاهده (از جامعه‌ی ...) می‌گویید. تابع چگالی آماره میانگین با اندازه ۲ را با استفاده از جامعه‌ای که در اختیار دارید محاسبه و رسم کنید. چه رابطه‌ای بین نمودار رسم شده و نمودار قبلی وجود دارد؟ (گاهی به توزیع آماره، توزیع نمونه‌ای آماره و یا توزیع نمونه‌ای می‌گویند.)

(۴) ↑ اگر اندازه نمونه افزایش یابد مثلاً اندازه نمونه برابر ۵ باشد، چه تغییری در نمودار توزیع نمونه‌ای رخ می‌دهد؟

• قانون ضعیف اعداد بزرگ

(۵) در کلاس ۳ نفر را به تصادف انتخاب کرده و نمره یک درس را از آنها سؤال کنید. معدل این ۳ نمره را محاسبه کنید. حال از تمامی دانشجویان کلاس

Bernoli^۶
Mean^۷

نمره همان درس را سؤال کرده و معدل همه دانشجویان کلاس را محاسبه کنید. چه تفاوتی بین مفهوم این دو عدد بدست آمده وجود دارد. نام هر کدام را ذکر کنید.

(۶) با استفاده از گزینه Random Sampling در آزمون ۳۰ نمونه ۲۰ تایی انتخاب کرده و میانگین آنها را رسم کنید این شکل چه تفاوتی با نمونه‌های ۲ تایی و ۵ تایی در سؤال‌های ۳ و ۴ دارد.

(۷) ↑ قانون اعداد بزرگ: اگر اندازه نمونه تصادفی به سمت ∞ (اندازه جامعه) میل کند آنگاه آماره \bar{X} به سمت میل خواهد کرد. به عبارت دیگر \bar{X} به سمت یک متغیر تصادفی با واریانس میل خواهد نمود.

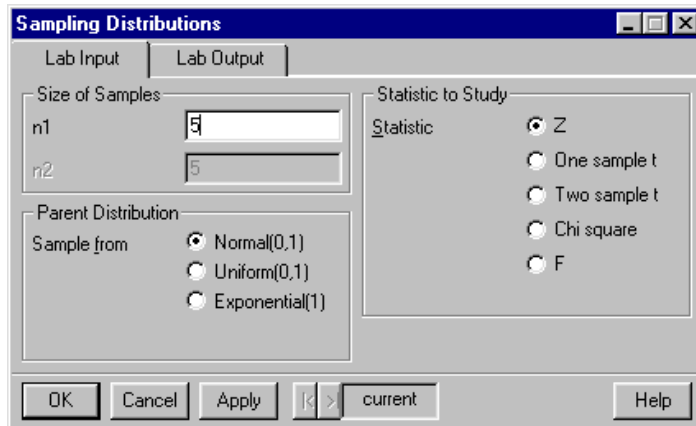
(۸) قانون اعداد بزرگ را می‌توان در مورد گشتاورها نیز بیان نمود. این قانون را برای واریانس بیان کنید.

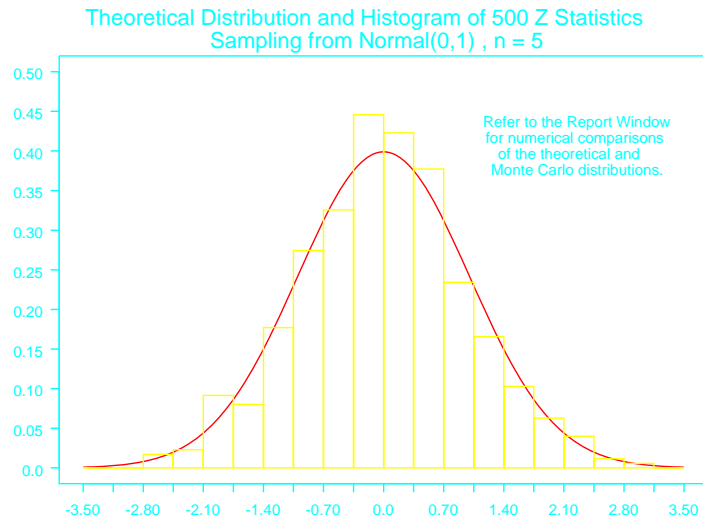
• قضیه حد مرکزی

(۹) با استفاده از گزینه Sampling Distributions^۱ (شکل ۱.۴) در آزمون می‌توانیم نمودار توزیع نمونه‌ای آماره

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

را برای توزیع‌های نرمال $N(0, 1)$ ، یکنواخت $U(0, 1)$ و نمایی $E(1)$ رسم کنیم. یعنی برای اندازه نمونه ثابت n مثلاً $n = 20$ ، $n = 500$ بار (یا بیشتر یا کمتر) از توزیع‌های ذکر شده نمونه‌های ۲۰ تایی بگیریم و هیستوگرام $z = \frac{\bar{x} - \mu}{\sigma}$ های



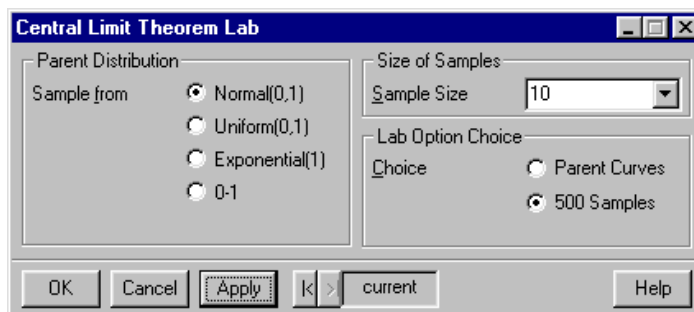


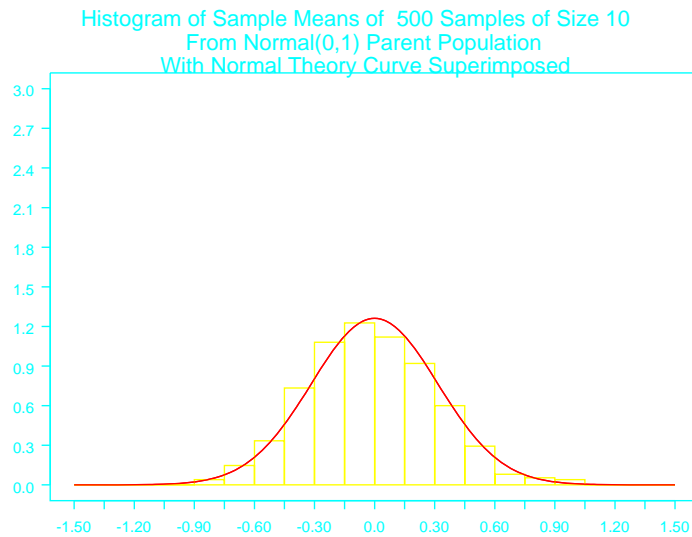
شکل ۱.۴: توزیع نظری و تجربی آماره Z

به دست آمده را رسم کنیم. (μ و σ به ترتیب میانگین و انحراف معیار جامعه می‌باشند)، چه نتیجه‌ای از انجام این کار می‌توانیم بگیریم؟ چه تفاوتی بین زمانی که از توزیع نرمال و غیر نرمال استفاده می‌کنیم وجود دارد؟ قضیه‌هایی را که در مورد این دو حالت خوانده‌اید بنویسید. آیا شما درستی این قضیه‌ها را تحقیق کرده‌اید؟

(۱۰) چه تفاوتی بین قضیه حد مرکزی و قانون اعداد بزرگ وجود دارد؟ کدامیک از شکل‌های رسم شده در این فصل را برای توجیه نظر خود استفاده می‌کنید.

(۱۱) تحقیق کنید! با استفاده از گزینه Central Limit Theorem^۹ (شکل ۲.۴)، چه چیزی را می‌توانید تحقیق کنید؟ قانون اعداد بزرگ یا قضیه حد مرکزی (با





شکل ۲.۴: توزیع نمونه‌ای \bar{X} و منحنی نرمال

توجه به محور افقی نمودارهای رسم شده به سؤال پاسخ دهید).

• تذکر

- (۱۲) آماره را تعریف کنید آیا قضیه حد مرکزی فقط برای آماره Z صحیح است؟
- (۱۳) پارامتر را تعریف کنید.
- (۱۴) صورت دیگری از قضیه حد مرکزی را بیان کنید.
- (۱۵) آیا جمله زیر صحیح است؟ اگر صحیح نیست صحیح آن را بنویسید.
«اگر n زیاد باشد \bar{X} دارای توزیع نرمال است.»
- (۱۶) آیا قضیه حد مرکزی تنها برای متغیرهای تصادفی پیوسته صحیح است؟
- (۱۷) به چه مقدار از n ، اندازه بزرگ نمونه می‌گوییم؟ آیا می‌توانیم بگوییم که اگر n مثلاً برابر ۳۰ باشد می‌توانیم از تقریب نرمال استفاده کنیم؟

۱۸) چه مسائلی بر خوب بودن تقریب نرمال (قضیه حد مرکزی) دخالت دارد؟ نام ببرید. (سوالات ۱۹ و ۲۰ به شما برای پاسخ دادن به این سؤال کمک می‌کنند)

۱۹) تقریب نرمال برای کدامیک از دو حالت زیر بهتر است. چرا؟ چه نتیجه‌ای می‌گیرید.

(a) اگر X_1, \dots, X_{100} یک نمونه تصادفی از توزیع پواسن $(1/0.01)$ باشند مقدار $P(\sum_{i=1}^{100} X_i > 0)$ را با استفاده از تقریب نرمال به دست آورید.

(b) Y یک مشاهده از توزیع پواسن (2) می‌باشد مطلوب است محاسبه $P(Y > 0)$ با استفاده از تقریب نرمال.

۲۰) تقریب زیر برای کدامیک از سه حالت زیر بهتر است. چرا؟ چه نتیجه‌ای می‌گیرید. برای محاسبه از گزینه Statistical Tables استفاده کنید.

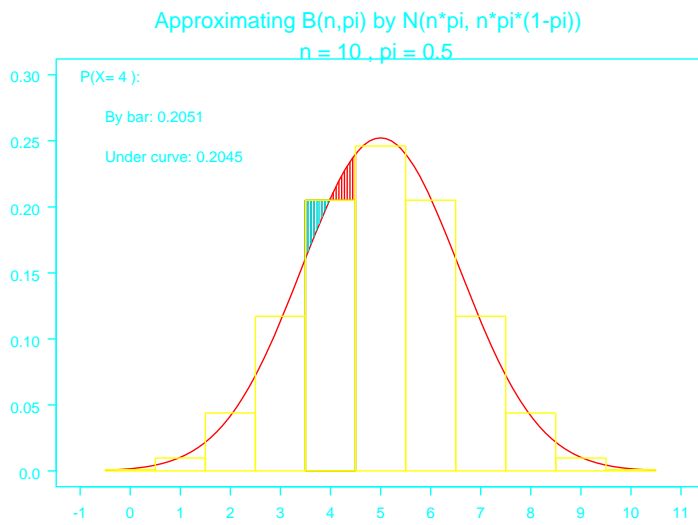
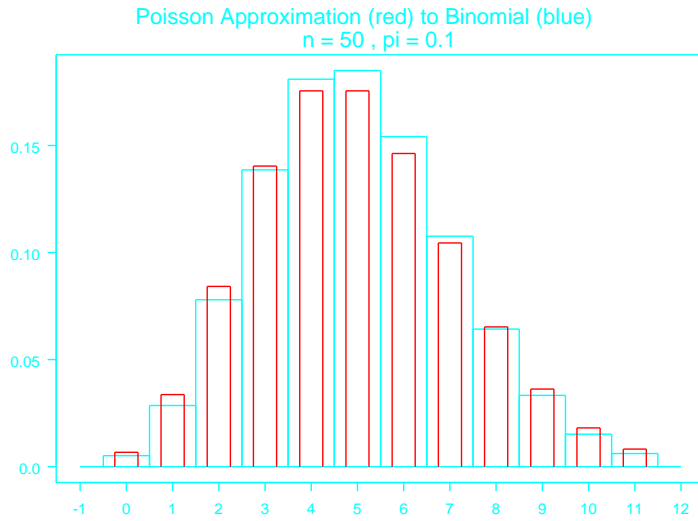
(a) X_1, \dots, X_{10} یک نمونه تصادفی از توزیع برنولی $B(10, 0.1)$ می‌باشند مطلوب است محاسبه $P(\sum_{i=1}^{10} X_i \geq 2)$ با استفاده از تقریب نرمال.

(b) X_1, \dots, X_{10} یک نمونه تصادفی از توزیع برنولی $B(10, 0.3)$ می‌باشند مطلوب است محاسبه $P(\sum_{i=1}^{10} X_i \geq 2)$ با استفاده از تقریب نرمال.

(c) X_1, \dots, X_{10} یک نمونه تصادفی از توزیع برنولی $B(10, 0.5)$ می‌باشند. مطلوب است محاسبه $P(\sum_{i=1}^{10} X_i \geq 2)$ با استفاده از تقریب نرمال.

۲۱) آیا می‌توانید بگویید چه وقت تقریب پواسن برای دو جمله‌ای بهتر از تقریب نرمال است؟ چرا؟ (از آزمون 0 -Sampling from 1 Population > Approximating Binomial Probabilities (شکل ۳.۴)، قسمت تقریب احتمال‌های دو جمله‌ای استفاده کنید.)

همگی در کتابخانه مشغول فکر کردن روی مسئله‌ای بودیم، ناگهان یکی از دانشجویان گفت: «یافتم». همه چشم‌ها به او خیره شد و با نگاهی خاموش به او فهماندند، پس چرا معطلی؟! سوال این بود: اگر $X \sim \chi_n^2$ باشد، \sqrt{X} دارای چه توزیعی می‌باشد؟ و پاسخ او: $\sqrt{X} \sim \chi_n$ و در پاسخ یکی از دانشجویان که گفت شاید منفی شود، گفت: « $|\chi_n|$ ، البته به خاطر شما!»



شکل ۳.۴: تقریب توزیع دو جمله‌ای با توزیع پواسن و نرمال