دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

تمارین نوبت اول درس یادگیری ماشین آماری

دانشجو:

سیّد احمد نقوی نوزاد

شماره دانشجوئی:

۹۴۱۳۱۰۶۰

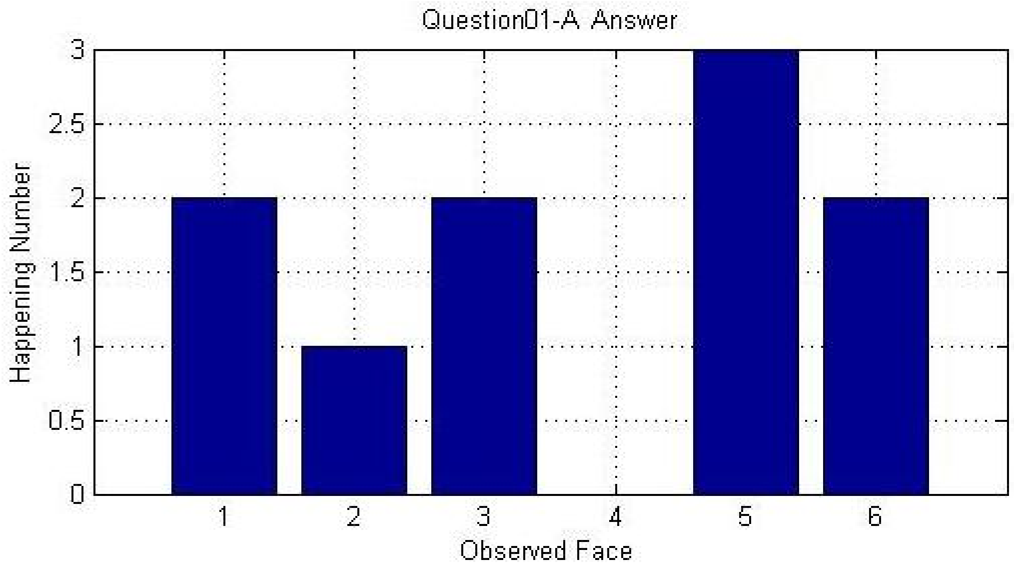
استاد:

دکتر نیک‌آبادی

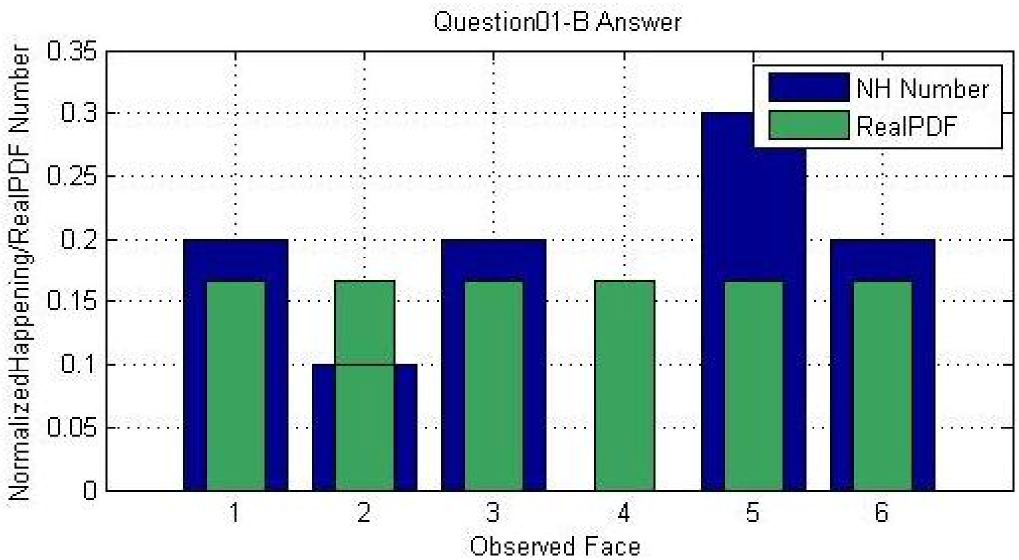
پاییز ۹۴

سوال اول:

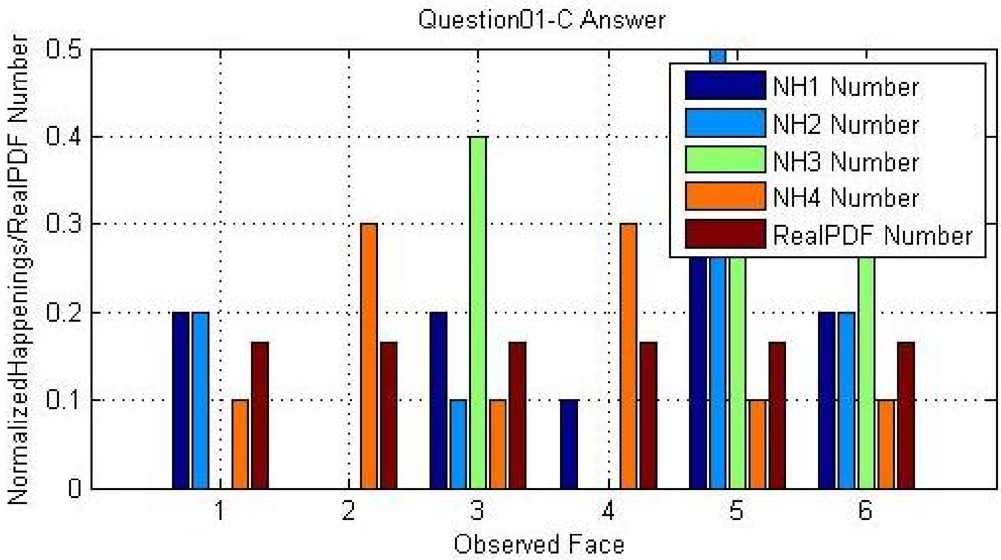
الف) یک تاس (با احتمال یکسان برای هر وجه) را به تعداد ده مرتبه پرتاب کرده و هیستوگرام یا همان نمودار ستونی را برای تعداد رو آمدن هر یک از وجوه مطابق شکل زیر رسم می‌کنیم:



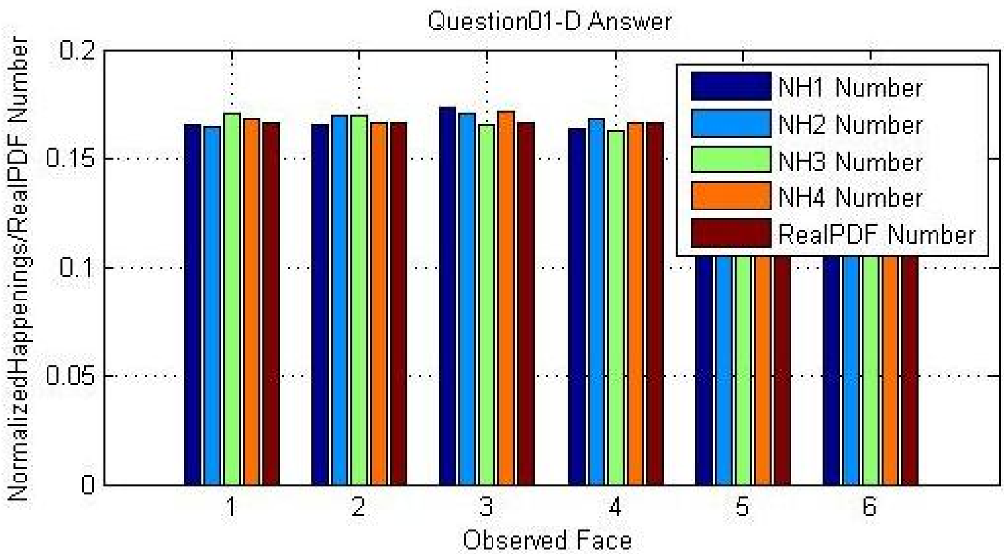
ب) در اینجا تعداد روآمدن به ازای هر وجه را بر تعداد کل پرتاب‌ها تقسیم می‌نمائیم که در نتیجه مقادیر چگالی احتمال مشاهده‌شده یا همان تجربی به دست می‌آید و نیز مقدار چگالی احتمال واقعی را نیز در کنار مقدار قبلی رسم می‌نمائیم:



ج) از آن‌جا که مقادیر چگالی احتمال واقعی با مقادیر چگالی احتمال تخمینی برابری نمی‌کنند، لذا این عمل را چند بار دیگر تکرار کرده و نتیجه را مشاهده می‌نمائیم که همانطور که از شکل پیداست در هر بار مقادیر متفاوتی حاصل می‌شوند که هیچ‌کدام با چگالی احتمال واقعی برابری نمی‌کنند:

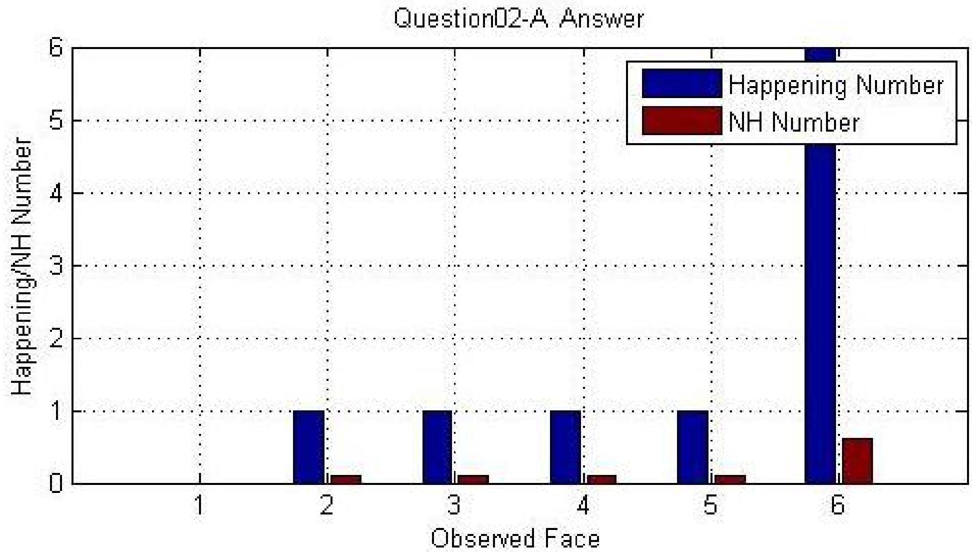


د) در اینجا تعداد دفعات آزمایش را تا ۱۰۰۰۰ بار افزایش داده و نیز این عمل را چندبار دیگر تکرار می‌کنیم که همانطور که از شکل پیداست نتایج تجربی با نتایج واقعی تقریبا برابری می‌نمایند:

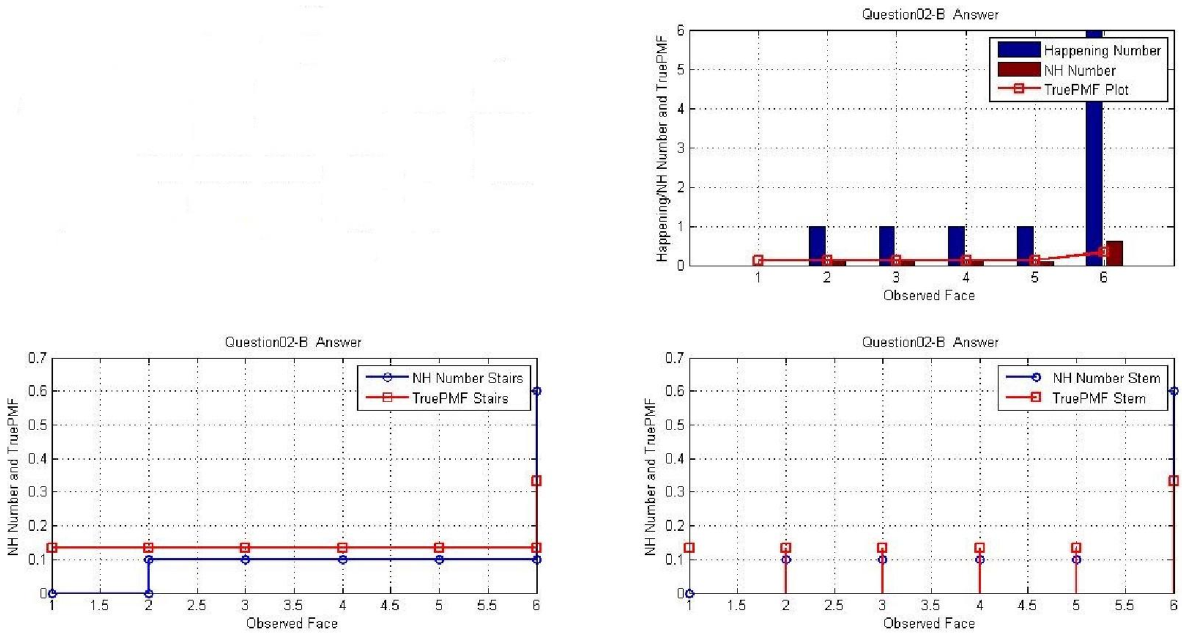


سوال دوم:

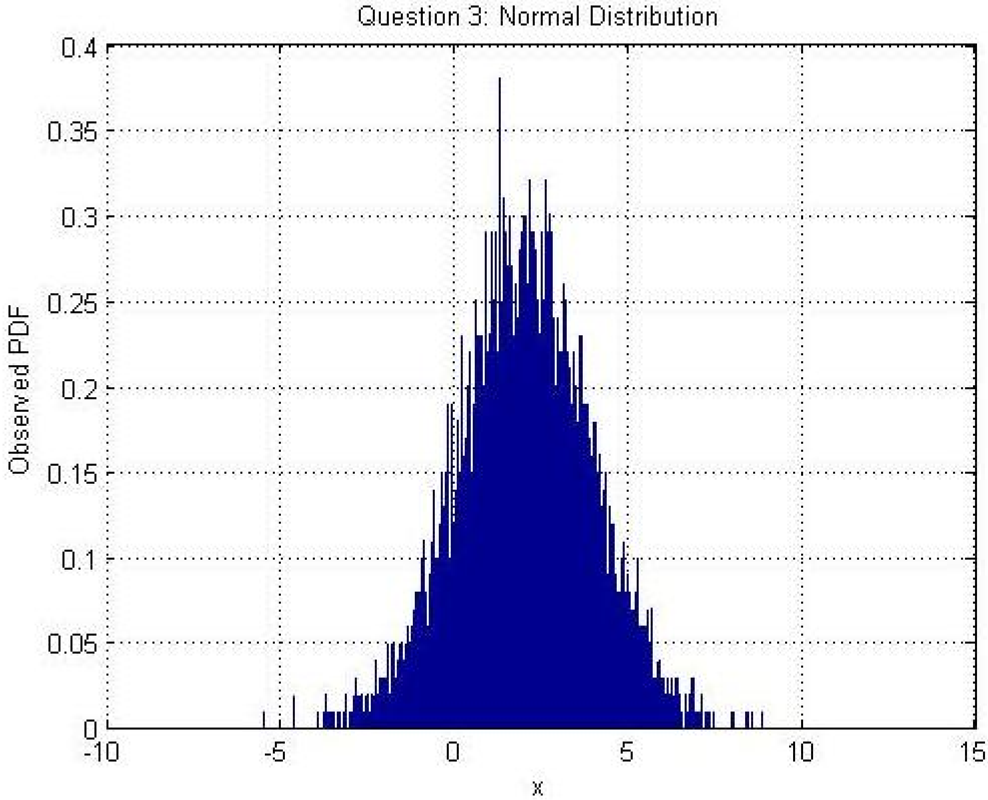
الف) در اینجا یک تاس وزن‌دار را پرتاپ می‌کنیم که احتمال ۶ آمدن برابر ۱/۳ و احتمال روآمدن بقیه‌ی وجوه برابر ۲/۱۵ می‌باشد. مانند سؤال اول تاس را تعداد ده مرتبه پرتاب کرده و نمودار هیستوگرام را برای نتایج اولیه و نیز نتایج نرمال شده رسم می‌نمائیم:



ب) در اینجا مقادیر تابع چگالی احتمال واقعی را برای تاس وزن‌دار به دست آورده و آن را در کنار مقادیر چگالی احتمال تجربی با استفاده از توابع plot و stairs و stem در نرم‌افزار MATLAB رسم می‌نمائیم:



سوال سوم:

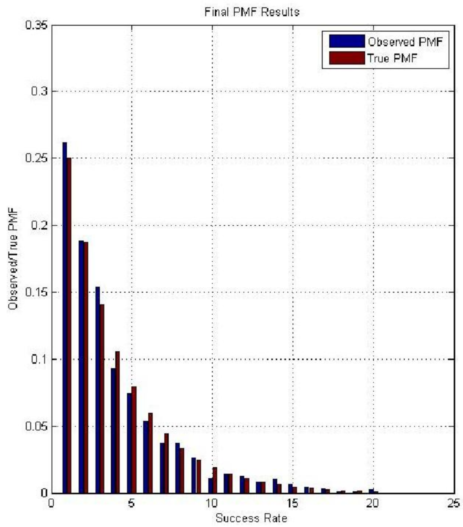


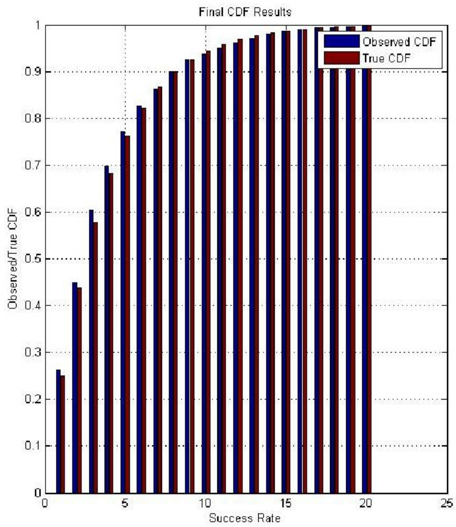
الف) در اینجا ابتدا از کاربر مقادیر میانگین و واریانس توزیع نرمال را از کاربر درخواست کرده (به عنوان مثال $\mu=۲$ و $\sigma=۳$) و سپس با استفاده از تابع randn مقادیر توزیع نرمال استاندارد را محاسبه کرده و با انجام محاسبات مربوطه آن‌ها را به مقادیر توزیع نرمال با پارامترهای قیدشده تبدیل می‌کنیم. حال مقادیر میانگین و واریانس را برای مقادیر توزیع نرمال محاسبه کرده و با مقادیر وارده مقایسه می‌کنیم که همانطور که مشاهده می‌شود با هم به طور تقریبی برابری می‌کنند. حال برای محاسبه مقادیر خواسته‌شده در قسمت الف، چگالی‌های تجربی به دست‌آمده را در بازه‌های مربوطه جمع زده و نتیجه را در خروجی چاپ می‌کنیم.

ب) در این قسمت نیز مقادیر احتمال به دست آمده را با یکدیگر جمع می‌زنیم و نتیجه‌ی صحیح که همان یک است را مشاهده می‌نمائیم.

سوال چهارم:

در اینجا آزمایش را به تعداد دفعات زیاد تکرار می‌کنیم و در هر دفعه نوبت موفقیت را درون یک بردار ذخیره می‌نمائیم. در نهایت برای مقادیر موفقیت ۱ تا ۲۰ میانگین تجربی را از روی بردار محاسبه کرده و با مقادیر PMF واقعی مقایسه می‌کنیم و همینطور با استفاده از تابع cumsum مقادیر CDF را محاسبه کرده و در نهایت با استفاده از آن احتمال X > 4 که برابر با مقدار ۰,۳۰۳۰۰ می‌باشد را به دست می‌آوریم:





سوال پنجم:

در اینجا با استفاده از توزیع دوجمله‌ای با پارامترهای n=100 و p=0.05 و تعداد آزمایش‌های بسیار زیاد، مقدار میانگین شمار موفقیت (در اینجا تعداد دفعات روآمدن سکه با احتمال p) را محاسبه می‌نمائیم که در اینجا برابر مقدار ۵,۰۱۰۰۰ یعنی همان مقدار مورد انتظار ما ۵=۱۰۰*۰,۰۵ می‌گردد. این نحوه تخمین‌زدن ارتباط تنگاتنگی با تخمین توزیع پواسون دارد که شرح آن در این مُقال نمی‌گنجد!

سوال ششم:

در اینجا از آنجا که ۶ توپ از ۹ توپ را با جایگذاری انتخاب می‌کنیم، با انجام تعداد دفعات آزمایش زیاد و میانگین گرفتن از تعداد دفعات رخ‌دادن هر کدام از توپ‌ها مشاهده می‌کنیم که محتمل‌ترین تعداد ممکن برای هر کدام از توپ‌ها برابر ۲ و با میزان احتمالی حدوداً برابر ۰,۳۳ می‌باشد.

سوال هفتم:

در اینجا آزمابش را به تعداد دفعات بسیار زیاد تکرار کرده و در هر بار آزمایش مقادیر روز تولدی از سال را با جایگذاری به هر کدام از ۲۳ دانش‌آموز نسبت می‌دهیم و همینطور بررسی می‌کنیم که آیا دو دانش‌آموزی وجود دارند که روز تولدشان اول ژانویه باشد یا نه و در صورت حصول موفقیت شمارنده‌ای را افزایش داده و در نهایت از تقسیم مقدار شمارنده بر تعداد دفعات کل آزمایش، احتمال رخداد مربوطه یعنی داشتن روز تولد یکسان را که حدودا برابر ۰,۰۰۱۸ می‌باشد را محاسبه می‌کنیم.

سوال هشتم:

در اینجا نیز مانند سؤال شش، از بین ۶ توپ ۲ توپ را بدون جایگذاری انتخاب می‌نمائیم و این کار را به تعداد دفعات بسیار زیاد انجام داده و در صورت حصول موفقیت شمارنده‌ای را افزایش می‌دهیم و در نهایت از تقسیم مقدار این شمارنده بر تعداد دفعات کل آزمایش، احتمال رخداد مربوطه که تقریبا برابر ۰,۴۴ می‌باشد را محاسبه می‌نمائیم.

سوال نهم:

در اینجا نیز برای محاسبه‌ی مقادیر میانگین نمونه و واریانس نمونه و با استفاده از تابع rand، یک بردار از نتایج حاصله را با توجه به مقادیر احتمالات تعریف‌شده به مقادیر −۱ و +۱ نگاشت کرده و سپس مقادیر میانگین تجربی و واریانس تجربی را با استفاده از فرمول‌های مربوطه محاسبه کرده و در نهایت با مقادیر میانگین و واریانس واقعی مقایسه می‌نمائیم. مشاهده می‌شود که شدیدا یه همدیگر نزدیک هستند.

سوال دهم:

در اینجا نیز همانند سؤال سوم ابتدا از کاربر مقادیر میانگین و واریانس توزیع نرمال را از کاربر درخواست کرده (به عنوان مثال μ=۲ و σ=۳) و سپس با استفاده از تابع randn مقادیر توزیع نرمال استاندارد را محاسبه کرده و سپس در یک حلقه به تعداد دفعات آزمایش زیاد و با انجام محاسبات مربوطه آن‌ها را به مقادیر توزیع نرمال با پارامترهای قیدشده تبدیل می‌کنیم و همین‌طور مقادیر میانگین نمونه و میانگین نرمال را در هر مرحله اجرای حلقه محاسبه کرده و نمودار مربوطه را مرحله به مرحله رسم می‌نمائیم. همانطور که مشاهده می‌شود مقدار میانگین نرمال شدیدا به سمت صفر همگرا بوده در حالی که مقدار میانگین نمونه تا حدی یه سمت صفر همگرا می‌باشد و شدت همگرائی میانگین نرمال را ندارد. با توجه به محدودیت شدیدا زمان‌بر بودن مسئله شاید با انجام تعداد دفعات آزمایش بسیار بیشتر، نتیجه‌ی بهتری حاصل گردد.