به نام خدا

دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس مبانی یادگیری آماری

استاد نیک‌آبادی

تمرین دوم

علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

توجه: برای حل سوالات این تمرین از پاسخ‌نامه موجود در کورسز و گیت‌هاب https://github.com/telmo-correa/all-of-statistics الهام گرفته شده است.

قسمت تئوری

سوال ۱

با توجه به آنکه تنها لازم است تا دو قسمت مربوط به همگرایی را ثابت کنیم، می‌توان از دو رابطه مربوط به امیدریاضی و واریانس در تئوری استفاده کرد:

$$MSE = bias^2\left(\widehat{F}_n(x)\right) + \mathbb{V}\left(\widehat{F}_n(x)\right) = \mathbb{E}\left(\widehat{F}_n(x) - \mathbb{E}\left(\widehat{F}_n(x)\right)\right)^2 + \mathbb{V}\left(\widehat{F}_n(x)\right)$$
$$= \mathbb{E}\left(\widehat{F}_n(x) - F(x)\right)^2 + \frac{F(x)\big(1-F(x)\big)}{n} = \frac{F(x)\big(1-F(x)\big)}{n}$$

مقدار $F(x)$ قطعا محدود است، مقدار $\big(1-F(x)\big)$ هم محدود است ولی مخرج از n است و وقتی به بی‌نهایت میل کند حاصل کل کسر یعنی MSE به صفر میل می‌کند.

$$\mathbb{V}\left(\widehat{F}_n(x)\right) = \frac{F(x)\big(1-F(x)\big)}{n}$$
$$\mathbb{E}\left(\widehat{F}_n(x)\right) = F(x)$$

به طور مشابه داریم:

$$\mathbb{V}\left(\widehat{F}_n(x)\right) \to 0 \Rightarrow se\left(\widehat{F}_n(x)\right) \to 0$$

و همچنین:

$$MSE \to 0 \Rightarrow bias \to 0$$

طبق قضیه 6.10 اگر با $n \to \infty$ داشته باشیم $bias \to 0$ و $se \to 0$ می‌توان نتیجه گرفت که $\widehat{F}_n(x)$ سازگار (consistence) است و در نتیجه:

$$\widehat{F}_n(x) \xrightarrow{P} F(x)$$

سوال ۲

مطابق با CLT می‌دانیم به صورت تقریبی رابطه زیر برقرار است:

$$\widehat{F_n}(x) \rightsquigarrow N\left(\mathbb{E}\left(\widehat{F_n}(x)\right), \frac{\mathbb{V}\left(\widehat{F_n}(x)\right)}{n}\right)$$

مطابق تئوری 7.3 که در سوال ۱ بود خواهیم داشت:

$$\widehat{F_n}(x) \rightsquigarrow N\left(F(x), \frac{F(x)\left(1 - F(x)\right)}{n^2}\right)$$

طبیعتا خواهیم داشت:

$$n \to \infty: \widehat{F_n}(x) \rightsquigarrow N(F(x), 0) \Rightarrow \widehat{F_n}(x) \rightsquigarrow F(x)$$

سوال ۳

$$f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

تخمین با استفاده از مومان‌ها:

$$\alpha_1\left(\hat{\theta}\right) = \widehat{\alpha_1}$$

$$\alpha_j(\theta) = \mathbb{E}_\theta\left(X^j\right)$$

$$\widehat{\alpha_j} = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

$$\alpha_1\left(\hat{\theta}\right) = \mathbb{E}_\theta(X) = \hat{\lambda}$$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

تخمین با استفاده از بیشینه درست‌نمایی:

$$\mathcal{L}(\theta) = f(X^n|\theta) = \prod_{i=1}^{n} f(X_i|\theta) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!} e^{-\lambda} \propto \prod_{i=1}^{n} \lambda^{X_i} e^{-\lambda}$$

$$\ell(\theta) = \log\big(\mathcal{L}(\theta)\big) \propto \sum_{i=1}^{n} \log\big(\lambda^{X_i} e^{-\lambda}\big) = \sum_{i=1}^{n} (X_i \log(\lambda) - \lambda)$$

$$= \left(\log(\lambda) * \sum_{i=1}^{n} X_i\right) - n\lambda$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0 \rightarrow \frac{1}{\lambda} \sum_{i=1}^{n} X_i = n \rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

امتیاز فیشر:

$$\log f(X; \lambda) = x \log(\lambda) - \log(x!) - \lambda$$

$$s(X; \lambda) = \frac{\partial \log f(X; \lambda)}{\partial \lambda} = \frac{x}{\lambda} - 1$$

$$\frac{\partial^2 \log f(X_i; \lambda)}{\partial \lambda^2} = -\frac{x}{\lambda^2}$$

$$I(\lambda) = -\mathbb{E}\left(\frac{\partial^2 \log f(X_i; \lambda)}{\partial \lambda^2}\right) = \mathbb{E}\left(\frac{x}{\lambda^2}\right) = \frac{1}{\lambda^2} \mathbb{E}(x) = \frac{1}{\lambda}$$

سوال ۴

برای تست این مورد از آزمون Wald استفاده می‌کنم. فرض این آزمون با تعداد بالای نمونه تحقق پیدا می‌کند که به نظر می‌رسد ۱۰۰ نمونه کافی باشد. تخمین اولیه مطابق با MLE برای مقدار $\hat{p}$ برابر خواهد بود با 0.6. مطابق آزمون داریم:

$$\frac{\hat{p}-p_0}{\widehat{se}} \rightsquigarrow N(0,1)$$

مقدار $\widehat{se}$ هم برابر است با:

$$\widehat{se} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.6*0.4}{100}} \approx 0.049$$

پس نهایتا داریم:

$$W \approx \frac{0.6-0.5}{0.049} = 2.041$$

آزمون Wald در این سوال که یک طرفه (one-sided) هست با مقدار α زمانی رد می‌شود که داشته باشیم:

$$W > z_\alpha$$

الف) برای $\alpha = 0.05$ داریم:

$$z_\alpha = \Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) \approx 1.65$$

طبیعتا فرض H_0 رد می‌شود.

ب) برای $\alpha = 0.01$ داریم:

$$z_\alpha = \Phi^{-1}(1-\alpha) = \Phi^{-1}(0.99) \approx 2.33$$

فرض H_0 رد نمی‌شود.

ج) باید کمترین α را پیدا کرد که با آن H_0 رد شود پس:

$$W = z_\alpha \rightarrow 2.041 = z_{alpha} \rightarrow \Phi(2.041) = 1-\alpha \rightarrow \alpha \approx 1-0.979 \rightarrow \alpha \approx 0.021$$

پس p-value در حدود 0.021 است.

سوال ۵

الف)

$$\mathbb{E}(\bar{X}_n^*|X_1,X_2,\ldots,X_n) = \mathbb{E}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum\mathbb{E}(X_i) = \frac{1}{n}\sum\mathbb{E}(X) = \mathbb{E}(x)$$

ب)

$$\mathbb{V}(\bar{X}_n^*|X_1,X_2,\ldots,X_n) = \mathbb{V}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum\mathbb{V}(X_i) = \frac{1}{n^2}\sum\mathbb{V}(X) = \frac{\mathbb{V}(X_i)}{n}$$

ج)

$$\mathbb{E}(\bar{X}_n^*) = \mathbb{E}\left(\frac{1}{n}\sum X_i^*\right) = \frac{1}{n}\sum\mathbb{E}(X_i^*) = \frac{1}{n}\sum\mathbb{E}(X) = \mathbb{E}(X)$$

د)

$$\mathbb{V}(\bar{X}_n^*) = \mathbb{E}\left({\bar{X}_n^*}^2\right) - \mathbb{E}(\bar{X}_n^*)^2 = \mathbb{E}\left(\left(\frac{1}{n}\sum X_i^*\right)^2\right) - \mathbb{E}(X)^2$$
$$= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^*X_j^*\right) - \mathbb{E}(X)^2$$

حال باید به قسمت $\mathbb{E}\left(X_i^*X_j^*\right)$ توجه کرد. اگر دو متغیر یکسان باشند، می‌توان این عبارت را به صورت $\mathbb{E}\left({X_i^*}^2\right)$ و سپس به $\mathbb{E}(X^2)$ تبدیل کرد. و اگر یکسان نباشند، این عبارت به $\mathbb{E}(X_i^*)\mathbb{E}\left(X_j^*\right)$ و سپس به $\mathbb{E}(X)^2$ تبدیل کرد. اگر $i = j$ باشد قطعا یک داده عینا تکرار شده است. ولی حتی اگر $i \neq j$ باشد باز محتمل است که یک داده عینا تکرار شده باشد، چراکه نمونه‌برداری با جایگزینی را در پیش گرفته‌ایم. با توجه به آنکه داده‌های اولیه برابر نیستند این احتمال برابر با $\frac{1}{n}$ خواهد بود. پس اگر i را ثابت فرض کنیم و jها را بررسی کنیم، یک جمله از n به علاوه $\frac{1}{n}$ از n-1 جمله دارای مقدار یکسانی است پس داریم:

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbb{E}\left(X_i^*X_j^*\right) = \sum_{i=1}^{n}\left(\mathbb{E}(X^2) + \frac{n-1}{n}\mathbb{E}(X^2) + \frac{(n-1)^2}{n}\mathbb{E}(X)^2\right)$$
$$= (2n-1)\mathbb{E}(X^2) + (n-1)^2\mathbb{E}(X)^2$$

نهایتا داریم:

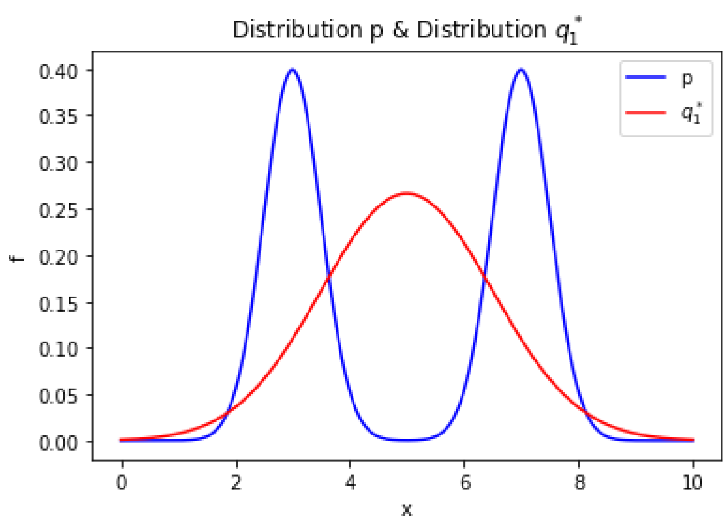
$$\mathbb{V}(\bar{X}_n^*) = \frac{1}{n^2}\left((2n-1)\mathbb{E}(X^2) + (n-1)^2\mathbb{E}(X)^2\right) - \mathbb{E}(X)^2$$
$$= \frac{2n-1}{n^2}\mathbb{E}(X^2) + \frac{-2n+1}{n^2}\mathbb{E}(X)^2 = \frac{2n-1}{n^2}(\mathbb{E}(X^2) - E(X)^2)$$
$$= \frac{2n-1}{n^2}\mathbb{V}(X)$$

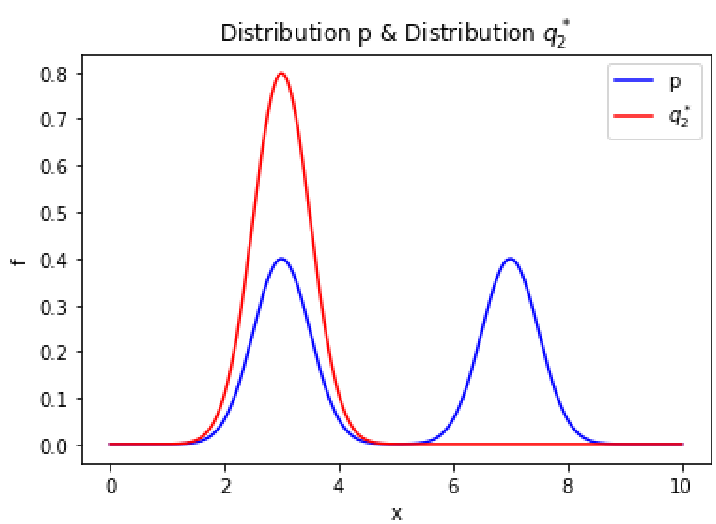
قسمت پیاده‌سازی

سوال ۱

برای محاسبه واگرایی KL از تکنیک نمونه‌برداری استفاده کردم؛ به این شکل که احتمال ده هزار نقطه با فاصله یکسان از هم از ۰ تا ۱۰ را در نظر گرفتم و فرمول KL را در همان نقاط محاسبه کردم. مقادیر $D_{KL}(p,q_1)$، $D_{KL}(p,q_2)$ و $D_{KL}(p,q_3)$ به ترتیب برابر است با: ۱۵۳۰۵، ۸۴۹ و ۱۵۳۰۵. مقادیر $D_{KL}(q_1,p)$، $D_{KL}(q_2,p)$ و $D_{KL}(q_3,p)$ به ترتیب برابر است با: ۶۹۲، ۱۹۷۸ و ۶۹۲. پس بر این اساس $q_1^* = q_2$ و $q_2^* = q_1$ یا $q_2^* = q_3$.

در دو تصویر بعد توزیع p و q_1^* و توزیع p و q_2^* آورده شده است.





به نظر می‌رسد که وقتی $D_{KL}(p,q)$ را در نظر گرفته‌ایم، توزیعی کمترین فاصله را داشته است که میانگین تمام اجزا بوده است ولی وقتی $D_{KL}(q,p)$ را در نظر گرفته‌ایم توزیعی پیشنهاد شده است که با یک جز به خوبی تطبیق داشته است. چنین چیزی با فرمول هم سازگار است. مطابق فرمول ضریب پشت لگاریتم f(x) است. پس اگر q را به عنوان ورودی اول بدهیم در جاهایی که مقدار ندارد، مقدار q هم مهم نخواهد بود چراکه ضریب صفر می‌شود و صرفا تطبیق در جاهایی که q مقدار بالایی دارد اهمیت دارد. وقتی هم p را اول می‌دهیم در تمام نقاطی که احتمال بالا است باید تابع q مقدار نزدیک به آن را داشته باشد که در این صورت مجبور خواهد بود به تمام اجزا نزدیک باشد. با این تفاصیل در حوزه تصویر اگر p شامل نمونه‌هایی متنوع باشد. اگر $D_{KL}(p,q)$ را به کار ببریم، توزیعی داریم که میانگین تمام توزیع‌های اجزاست ولی اگر $D_{KL}(q,p)$ را استفاده کنیم یکی از اجزا که با شکل توزیع q مطابق‌تر است به عنوان توزیع نهایی پیشنهاد خواهد شد.

سوال ۲

این سوال را با کمک روش bootstrap حل کردم. یعنی ابتدا تابع خواسته‌شده (میانگین و میانه) را بدست آوردم و سپس با روش bootstrap انحراف معیار تخمین را بدست آوردم. نهایتا با روش Normal Interval یک بازه اطمینان ارائه خواهم داد. توجه کنید که این روش ممکن است کمی خطا داشته باشد ولی از آن چشم‌پوشی می‌کنیم. در این حالت بازه اطمینان برابر است با $T_n \pm z_{\frac{\alpha}{2}} * \hat{s}_{boot}$

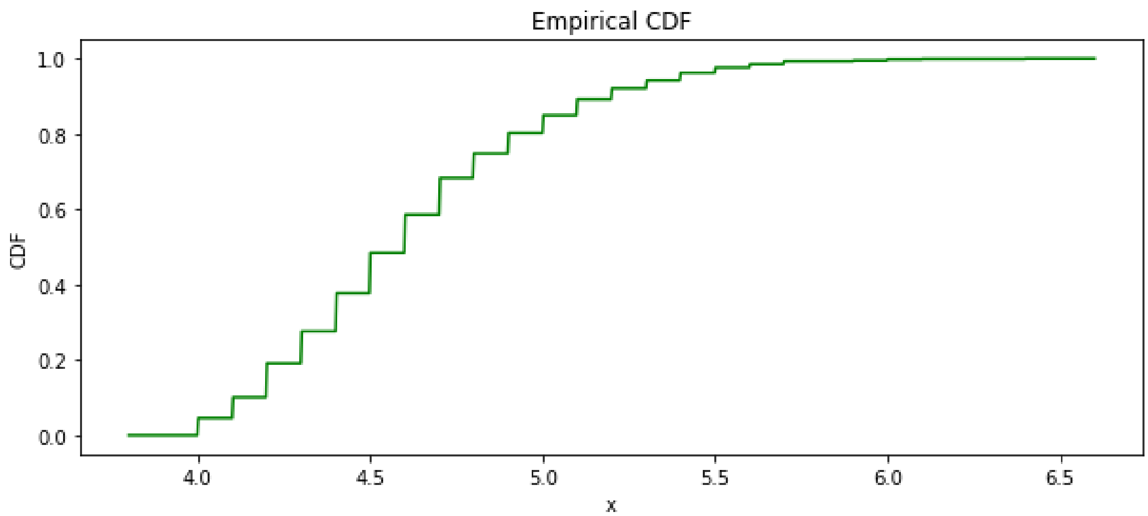
الف) میانگین ۳.۴۸ و انحراف معیار تخمین ۰.۰۶۸ بدست آمد.

ب) $3.48 \pm 1.65 * 0.068 = 3.48 \pm 0.112$

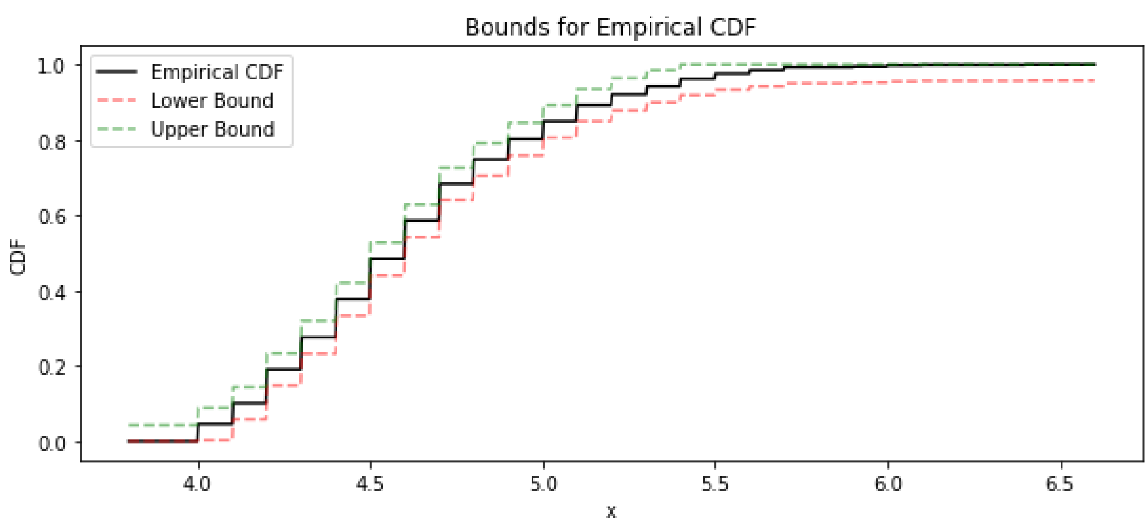
ج) میانه ۴.۰ و انحراف معیار تخمین ۰.۰۷۷ بدست آمد.

سوال ۳

الف) برای تخمین CDF از CDF تجربی استفاده کرده‌ام. نمودار آن مطابق تصویر زیر بدست آمده است.



ب)



ج) برای این قسمت تخمین‌گر ما $F(4.9) - F(4.3)$ است. پس این تابع را باید بر روی کل نمونه‌ها و نمونه‌های بوت‌استرپ اعمال کنیم. مطابق فرمول‌های بخش ۸.۳ پیش می‌رویم.

تخمین با روش نرمال: روال شبیه سوال ۲ است. تخمین: 0.526، انحراف معیار: 0.015 و متناسب با آن‌ها بازه (0.495,0.556) را داریم.

تخمین با روش Percentile Interval: تعداد هزار نمونه بوت‌استرپی آماده کردیم و تابع تخمین‌گر را بر آن اعمال. سپس نمونه ۲۵ ام (0.025*1000) از اول و نمونه ۲۵ ام از آخر را به عنوان حد بالا و پایین بازه پیشنهادی بدست آوردیم که شد: (0.494,0.557)

تخمین با روش Pivotal Interval: در این حالت مانند Percentile Interval هزار نمونه بوت‌استرپی آماده می‌شود و تابع تخمین‌گر بر آن اعمال می‌گردد. یک بار هم تابع تخمین‌گر بر کل نمونه‌ها اعمال می‌شود تا $\hat{\theta}_n$ بدست آید. پس از آن می‌توان از فرول این روش استفاده کرد که بازه‌ی (0.495,0.556) حاصل می‌شود.

بازه خروجی هر سه روش بسیار به هم شبیه است که نشان می‌دهد در عمل می‌توان هر یک از سه روش را استفاده کرد.

سوال ۴

الف) مطابق فرمول مثال ۸.۶ می‌توان تخمین correlation را بدست آورد. این میزان برابر با ۰.۵۴۶ شد.

ب) مطابق با توضیحات و کد سوالات قبل داریم: ۰.۲۲۳

ج) مطابق توضیحات و کد سوالات قبل داریم:

بازه روش نرمال: (0.108,0.983)

بازه روش Percentile Interval: (0.180, 0.943)

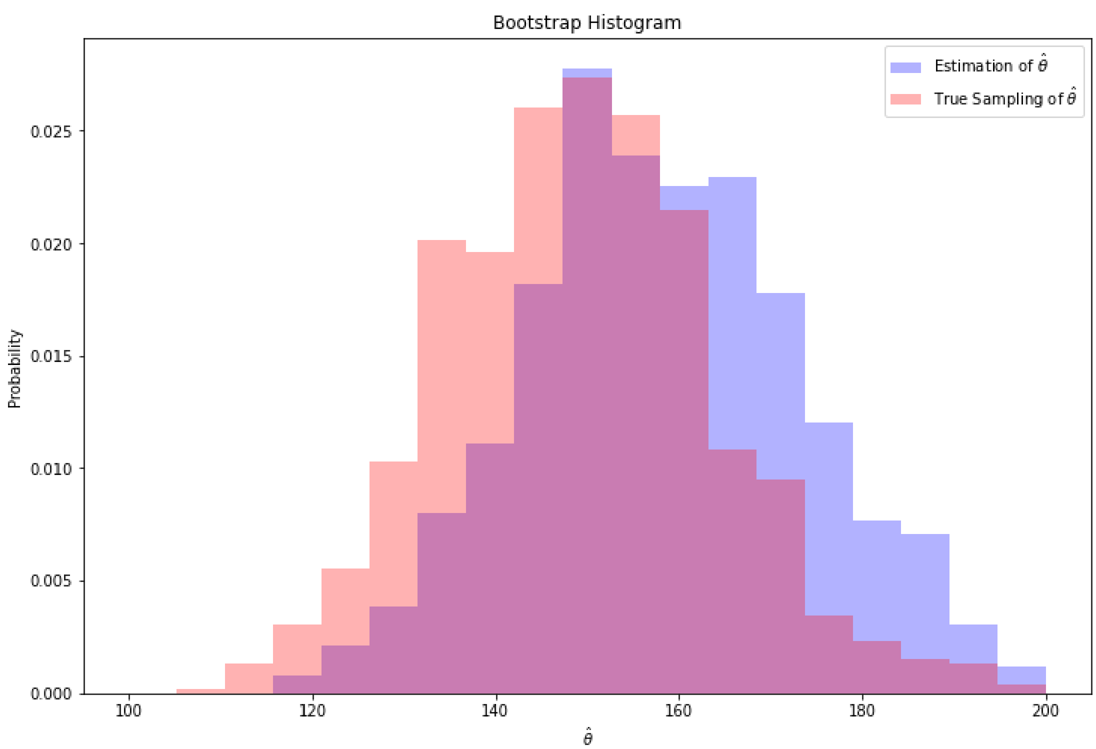
بازه روش Pivot Interval: (0.161, 0.888)

همچنان می‌بینیم که بازه‌های پیشنهادی به هم نزدیک است. اما اختلاف در این سوال بیشتر از سوال پیشین است. به نظر می‌رسد تعداد کم داده (۱۵) در این سوال نسبت به تعداد بسیار زیاد داده (۱۰۰۰) در سوال قبل علت این تفاوت باشد.

سوال ۵

ب) برای این سوال هم از همان کد بوت استرپ سوالات پیشین و روش تخمین نرمال برای محاسبه بازه اطمینان استفاده کرده‌ام. تخمین تتا و انحراف معیار تخمین به ترتیب برابر شد با ۱۵۷.۵۶ و ۱۶.۶۶. نهایتا بازه‌ی (126.23, 188.88) به عنوان بازه اطمینان بدست آمد.

ج) در تصویر زیر هر دو نمودار هیستوگرام خواسته‌شده در کنار هم آورده شده است. برای ترسیم تخمین بوت‌استرپی $\hat{\theta}$ تعداد هزار نمونه از مجموعه‌داده ساخته‌شده ایجاد کردم و تابع مدنظر را بر روی آن اعمال کردم. برای بدست آوردن توزیع واقعی $\hat{\theta}$ به تعداد هزار بار، هر بار یک مجموعه صد تایی مشابه مجموعه‌داده ساخته شده ایجاد کردم و تابع $\hat{\theta}$ را بر روی آت اعمال کردم.



با توجه به نمودار می‌توان دید که تخمین بوت‌استرپ عالی نبوده و خطای آن قابل چشم‌پوشی نیست ولیکن با توجه به حجم محدود داده خروجی آن قابل قبول است و در خیلی از کاربردها احتمالا مناسب باشد.

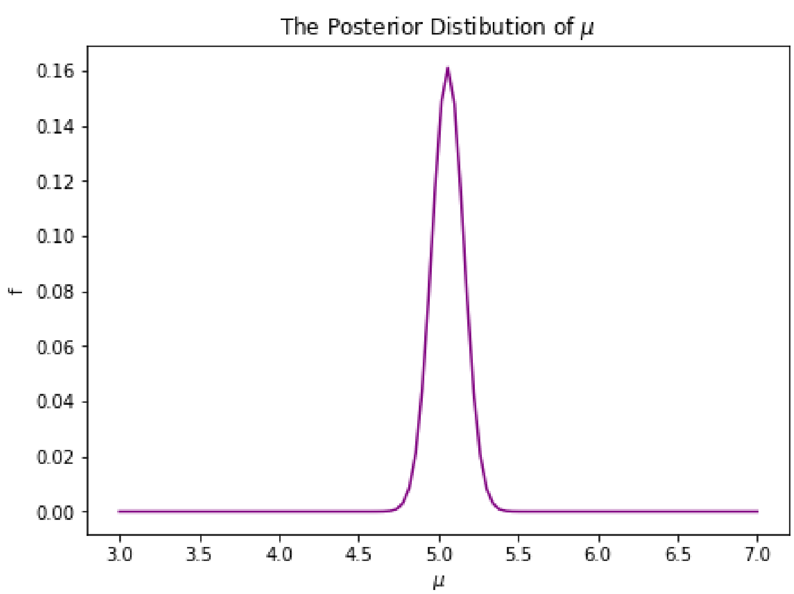
سوال ۶

ب) مطابق رابطه زیر برای توزیع posterior خواهیم داشت:

$$f(\mu|X^n) \propto f(\mu) * f(X^n|\mu) = f(X^n|\mu) = \prod f(X_i|\mu)$$

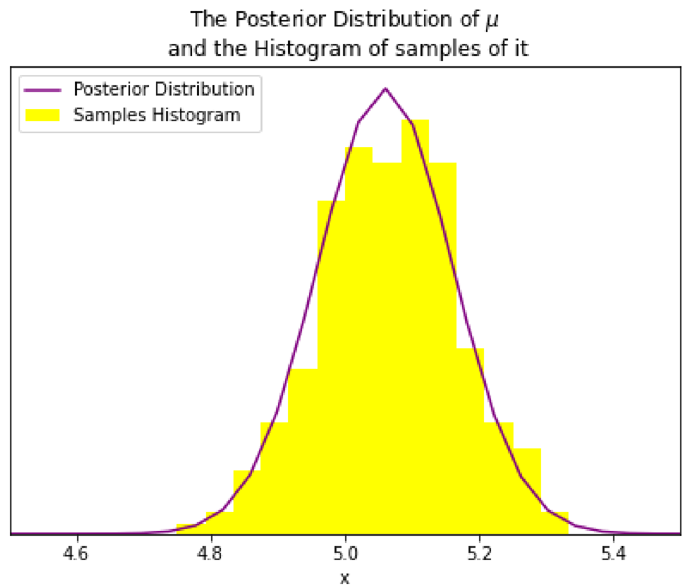
توجه کنید که مقدار واریانس توزیع نرمال ثابت و شناخته شده است و بر اساس مقدار میانگین توزیع نرمال احتمالات مختلفی برای توزیع posterior بدست می‌آید. برای ترسیم این توزیع بایستی رنجی از مقدار μ مختلف را بررسی کرد و مقدار احتمال آن را بدست آورد. همچنین باید اعداد را بر یک عدد ثابت تقسیم کرد. این عدد باید به گونه‌ای پیدا شود که انتگرال چگالی احتمال نهایی برابر با یک شود. طبیعتا مجموع احتمالات پیدا شده همان عدد ثابت خواهد بود. توزیع بدست آمده با استفاده از شبیه‌سازی متناسب با نمودار زیر خواهد بود.

مطابق انتظار می‌بینیم احتمال آنکه $\mu = 5$ باشد بسیار زیاد است و اعداد بسیار نزدیک به آن نیز همچنان محتمل هستند. چون تعداد نمونه در نظر گرفته شده نسبتا زیاد بوده است تنها با فاصله یکی دو دهم از ۵ احتمال رخداد از صفر فاصله معنی‌دار دارد و بعد از احتمال تقریبا با صفر برابر شده است.



ج) در تصویر زیر هیستوگرام پانصد نمونه به همراه توزیع احتمالاتی قسمت ب آورده شده است. به نظر می‌رسد هیستوگرام تطبیق قابل قبولی با توزیع مطابق انتظار دارد.

شایان ذکر است که برای ترسیم دو نمودار در قالب یک نمودار مجبور به اسکیل کردن مقیاس y شدیم.



د) ابتدا به صورت تئوری مقدار آن را بدست می‌آوریم و سپس از طریق پیاده‌سازی. فرض کنید داشته باشیم: $Y = e^X$

$$F_Y(y) = F_Y(y) = \mathbb{P}(Y \leq y) \leq \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log y)$$
$$= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{\log y - \mu}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{\log y - \mu}{\sigma}\right) = \Phi\left(\frac{\log y - \mu}{\sigma}\right)$$

$$P$$

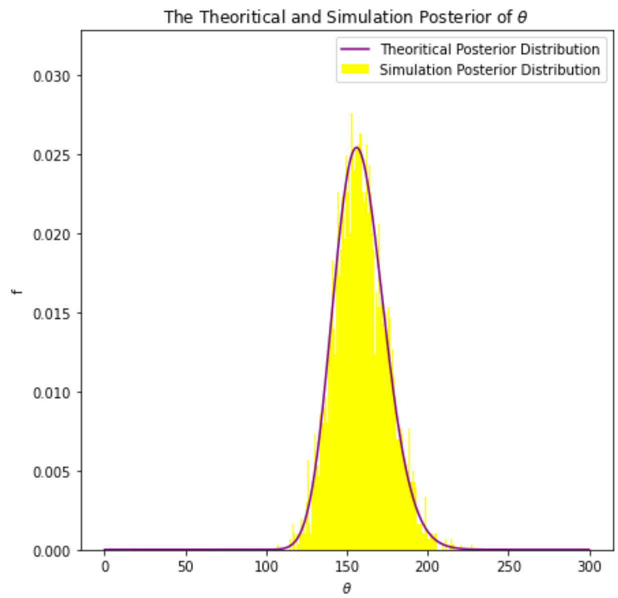
قبل از هر چیز باید توجه داشت که θ برابر است با e^μ و نه e^X. در این حالت میانگین میانگین همان μ است ولی انحراف معیار $\frac{\sigma}{\sqrt{n}}$ یعنی 0.1 است. پس:

$$F_Y(y) = \Phi\big(10(\log(y) - \hat{\mu})\big)$$

از آنجایی که به دنبال PDF و نه CDF هستیم باید از عبارت فوق مشتق گرفت. در نتیجه داریم:

$$f_Y(y) = \frac{10}{y} * \phi\big(10(\log(y) - \hat{\mu})\big)$$

حال باید دید آیا در عمل هم توزیعی مشابه به چیزی که بدست آوردیم را می‌بینیم یا خیر. برای بدست آوردن نمونه‌های عملی باید تعدادی داده μ تولید کرد و سپس e را به توان آن‌ها رساند. توزیع میانگین نمونه‌های یک نرمال همچنان یک توزیع نرمال است ولی انحراف معیار آن مطابق چیزی که در قسمت تئوری گفته‌شد متفاوت است. در تصویر زیر نمودار $f_Y(y) = \frac{10}{y} * \phi\big(10(\log(y) - \hat{\mu})\big)$ به عنوان جواب تئوری و هیستوگرام‌ها که تخمین با استفاده از پیاده‌سازی هستند آورده شده است. همانطور که مشاهده می‌شود تخمین عملی و تئوری تطبیق بسیار خوبی باهم دارند.



۵) برای بدست آوردن بازه اطمینان برای θ می‌توان مثلا روش عملی را ملاک قرار داد و $\frac{\alpha}{2}$ تا داده اول و آخر را کنار گذاشت و از مابقی داده‌ها بازه را ارائه داد. مطابق با این روش بازه (126.5, 195.4) برای اطمینان ۹۷٪ و بازه (131.5, 189.1) برای اطمینان ۹۳٪ بدست آمد. این اعداد هم منطقی هستند؛ وقتی بازه اطمینان را بخواهیم با دقت بیشتری بگوییم ناچارا باید طول بازه را بیشتر کرد و از طرفی این اعداد حول $e^5 = 148.41$ یعنی مقدار واقعی θ هستند.

سوال ۷

یک آزمون Wald ترتیب می‌دهیم به گونه‌ای که $H_0: \theta = \theta_0 = \frac{1}{2}$ و $H_1: \theta \neq \frac{1}{2}$

بهترین تخمین برابر است با $\hat{\theta} = \frac{997}{997+922}$ و همچنین داریم: $\widehat{se} = \sqrt{\frac{\hat{\theta}*(1-\hat{\theta})}{2}}$

مطابق آزمون Wald خواهیم داشت: $W = \frac{\hat{\theta}-\theta_0}{\widehat{se}}$ و فرض پایه رد می‌شود اگر و فقط اگر $|W| > z_{\frac{\alpha}{2}}$

مطابق محاسبات داریم، مقدار p-value برابر با حدود ۹٪ بدست آمد که نشان می‌دهد مطابق جدول کتاب شواهد ضعیفی علیه فرض پایه و به نفع ادعای مطرح شده در ابتدای سوال وجود دارد. همچنین بازه اطمینان ۹۵٪ برابر با (0.497,0.541) بدست آمد که در این بازه فرض صفر هم وجود دارد. طبیعتا مطابق با $\alpha = 5\%$ شرط پایه رد نمی‌شود.

سوال ۸

مشابه با سوال ۷ آزمون Wald رو ترتیب دادم. تنها تفاوت مهم محاسبه $\hat{\lambda}$ و $\widehat{se}$ هست. می‌دانیم امید ریاضی تابع پواسون برابر با λ است. پس در اینجا هم برای تخمین مقدار λ از میانگین نمونه‌های ساخته‌شده استفاده می‌کنیم. برای تخمین انحراف معیار هم می‌توان از فرمول زیر استفاده کرد:

$$\widehat{se} = \sqrt{\frac{\hat{\lambda}}{n}}$$

به تعداد صد هزار بار نمونه تولید کردم و آزمون Wald را انجام دادم و نزدیک به ۵.۰٪ دفعات فرض پایه رد شد. در این حالت فرض پایه واقعا درست بود و رد شدن آن خطا محسوب می‌شود؛ خطای نوع ۱. همان‌طور که انتظار داشتیم در حدود ۵٪ موارد مرتکب این خطا شدیم که مطابق با همان $\alpha = 0.05$ است.