Question 1 - Bayesian elements

A) به کمک تئوری بیز bionomial likelihood را به کمک Beta prior به روزرسانی می‌کنیم:

$$P(\theta|x) = \frac{P(x|\theta)\cdot P(\theta)}{\int_\theta P(x|\theta)\cdot P(\theta)} \propto P(x|\theta)\cdot P(\theta)$$

برای توزیع bionomial داریم ←
$$f(k,n,p) = Pr(k;n,p) = Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

برای توزیع Beta نیز داریم ←
$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

حال خواهیم داشت که
$$\begin{cases} x_i \sim Binomial(N_i, \theta) \\ \theta \sim Beta(\alpha,\beta) \end{cases}$$

$$P(\theta|x) \propto P(x|\theta)\,P(\theta) \propto \left[\binom{N}{x_i}\theta^t (1-\theta)^{N-t}\right]\left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\right]$$

$$\propto \theta^t (1-\theta)^{N-t}\,\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{\alpha+t-1}(1-\theta)^{\beta+N-t-1}$$

$$\boxed{t = \sum x_i \,,\; N = \sum N_i}$$

$$\begin{cases} \alpha' = \alpha+t-1 = \alpha + \sum x_i \\ \beta' = \beta + N - t - 1 = \beta + \sum N_i - \sum x_i \end{cases}$$

→ توزیع Posterior نیز از توزیع Beta پیروی می‌کند.

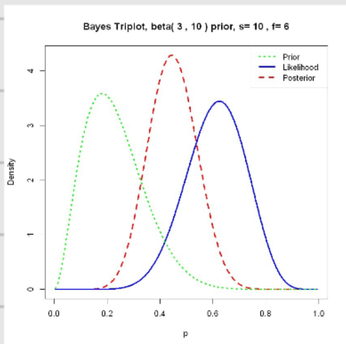
B) زمانی که توزیع پسین حاصل شده، از همان نوع توزیع قبلی باشد، توزیع پیشین و likelihood را مزدوج یا conjugate می‌نامیم. این بدین معنی است که اگر likelihood از نوع bionomial داشتیم می‌توانیم با قرار دادن یک توزیع Beta به جای prior انتظار یک Posterior از توزیع Beta داشته باشیم

C) توزیع Beta دو متغیر (α و β) دارد که مشخص می‌کنند داده‌ها چه شکلی به خود بگیرند. توزیع معادل uniform این است که به جای α و β مقدار یک قرار بدهیم ← Beta(1,1)

از قسمت قبل بدست آوردیم که در صورت داشتن Bionomial likelihood و Beta Prior توزیع Posterior به صورت زیر خواهد بود ←

$$P(\theta|x) = \theta^{\alpha+x}(1-\theta)^{\beta+n-x}$$

$$\alpha=\beta=1 \rightarrow P(\theta|x) = \theta^{1+x}(1-\theta)^{1+n-x}$$



D) پاسخ در Notebook

Question 2 - Dice Game

A) قبل از انجام هرگونه آزمایش و جمع آوری داده ما هیچ ایده ای نداریم که تاس خوب در کدام دست قرار دارد در نتیجه داریم:

$$\left\{\begin{array}{l} P(H_1) = 0.5 \\ P(H_2) = 0.5 \end{array}\right\} \longrightarrow \text{Prior probability} \longrightarrow$$ نشان دهنده احتمالات قبل از مشاهده هر داده

پس از اینکه دست راست را انتخاب کردیم و نتیجه بزرگتر از ۴ شد روابط به شکل زیر تفسیر می کند:

تاس خوب در دست راست

$$P(H_1 \mid T) = \frac{P(T \mid H_1)\, P(H_1)}{P(T)} = \frac{P(T \mid H_1)\, P(H_1)}{\sum p(T \mid H_i)\, P(H_i)} = \frac{0{,}75 \times 0{,}5}{0{,}5 \times 0{,}75 + 0{,}5 \times 0{,}5} = \frac{0{,}375}{0{,}375 + 0{,}25} = 0{,}6$$

(بزرگتر از 4) True

B) چهار حالت امکان دارد رخ دهد:

$$P(H_1 \mid T) = \frac{P(T \mid H_1)\, P(H_1)}{P(T)}$$

$$P(H_1 \mid F) = \frac{P(F \mid H_1)\, P(H_1)}{P(F)}$$

$$P(H_2 \mid T) = \frac{P(T \mid H_2)\, P(H_2)}{P(T)}$$

$$P(H_2 \mid F) = \frac{P(F \mid H_2)\, P(H_2)}{P(F)}$$

ادامه پاسخ داخل Notebook

C) به طور بسیار خلاصه می‌توان گفت در روش‌های Frequentist مقدار پارامتر یک کمیت ثابت است و احتمالی درمورد آن وجود ندارد و در مقابل در روش های Bayesian پارامتر مسأله یک متغیر تصادفی است و هیچ پاسخ کاملا دقیق برای مسأله تعریف نمی‌شود.

Question 3 - MAP Estimation

A) برای یافتن تخمین MAP

$$\arg\max_x f_{X|Y}(x,y) = \arg\max_x \frac{f_{Y|X}(y|x)\, f_X(x)}{\underbrace{f_Y(y)}_{\text{به } x \text{ وابسته نیست}}}$$

$$\arg\max_x f_{Y|X}(y|x)\, f_X(x)$$

احتمال MAP برای X به شرط مشاهده $Y=y$

برای مسائل مذکور داریم

$$\left.\begin{aligned} X &\sim N(0,\sigma_X^2) \\ W &\sim N(0,\sigma_W^2) \end{aligned}\right\} \rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\{-x^2/2\sigma_X^2\}$$

$$\rightarrow Y|X \sim N(0,\sigma_W^2) \longrightarrow f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma_W} \exp\{-(y-x)^2/2\sigma_W^2\}$$

مستقل از متغیرها و عدد ثابت

$$\longrightarrow \text{MAP Estimation}: \arg\max_x f_{Y|X}(y|x)\, f_X(x) = \overbrace{\frac{1}{2\pi\,\sigma_X\,\sigma_W}} \cdot \exp\left\{-\underbrace{\left[\frac{(y-x)^2}{2\sigma_W^2} + \frac{x^2}{2\sigma_X^2}\right]}\right\}$$

برای ماکسیمم شدن عبارت باید این قسمت کمینه شود.

$$\frac{\partial\left[\frac{(y-x)^2}{2\sigma_W^2} + \frac{x^2}{2\sigma_X^2}\right]}{\partial x} \overset{\text{set}}{=} 0 \longrightarrow \frac{-y-x}{\sigma_W^2} + \frac{x}{\sigma_X^2} = 0$$

$$\rightarrow -\sigma_X^2(y-x) + x\sigma_W^2 = 0 \longrightarrow -x\sigma_X^2 + y\sigma_X^2 + x\sigma_W^2 = 0 \longrightarrow \boxed{\hat{x}_{MAP} = \frac{y\,\sigma_X^2}{\sigma_X^2+\sigma_W^2}}$$

B) با توجه به مقدار تخمین MAP بدست آمده با افزایش σ_W^2 و میل به بی نهایت مقدار تخمین به صفر نزدیک می شود که منطقی نیز به نظر می رسد زیرا با افزایش واریانش Noise تشخیص سیگنال سخت تر شده و احتمال پایین می آید

Question 4 - Bayesian Hypothesis Testing

$$\begin{cases} P(H_0) = P_0 \\ P(H_1) = P_1 \\ \hline P_0 + P_1 = 1 \end{cases}$$

برای تصمیم گیری بین دو فرض H_0 و H_1 به کمک بیز، فرض می‌کنیم که مقادیر احتمال پیشین را از قبل می‌دانیم

با مشاهده متغیر Y توزیع متغیر Y تحت دو فرضیه به صورت روبرو است

$$\begin{cases} f_Y(y|H_0) \\ f_Y(y|H_1) \end{cases}$$

در نهایت به کمک قانون بیز احتمالات پسین بدست می‌آیند.

$$\begin{cases} P(H_0|Y=y) = \dfrac{f_Y(y|H_0)\,P(H_0)}{f_Y(y)} \\ P(H_1|Y=y) = \dfrac{f_Y(y|H_1)\,P(H_1)}{f_Y(y)} \end{cases}$$

پس از بدست آوردن احتمالات پسین فرضی را قبول می‌کنیم که احتمال بیشتری را دارد (MAP)

$$[Y = 2X + W]$$

در این مسئله با دو فرض روبرو هستیم ←

$$\begin{cases} H_0 : X = -1 \\ H_1 = X = 1 \end{cases}$$

تحت فرض H_0 داریم $Y = (-2) + W$ $\xleftarrow{W \sim N(0,\sigma^2)}$

$$Y|H_0 \sim N(-2, \sigma^2)$$

$$\rightarrow f_Y(y|H_0) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-(y+2)^2/2\sigma^2\right\}$$

تحت فرض H_1 نیز داریم $Y = 2 + W$ ←

$$Y|H_1 \sim N(2, \sigma^2)$$

$$\rightarrow f_Y(y|H_1) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-(y-2)^2/2\sigma^2\right\}$$

ما زمانی H_0 را رد کرده و H_1 را قبول می‌کنیم که ←

$$P(H_1)\cancel{\frac{1}{\sqrt{2\pi}\,\sigma}} \exp\left\{-(y-2)^2/2\sigma^2\right\} \geq P(H_0)\cancel{\frac{1}{\sqrt{2\pi}\,\sigma}} \exp\left\{-(y+2)/2\sigma^2\right\}$$

همچنین از فرض مسئله می‌دانیم که

$$\begin{cases} P(H_0) = (1-p) \\ P(H_1) = p \end{cases}$$

$$\rightarrow \exp\left\{\frac{\cancel{-y^2} + 4y \cancel{-4}}{2\sigma^2} - \frac{\cancel{-y^2} - 4y \cancel{-4}}{2\sigma^2}\right\} \geq \frac{P(H_0)}{P(H_1)} \rightarrow \exp\left\{\frac{8y}{2\sigma^2}\right\} \geq \frac{1-p}{p} \rightarrow \frac{8y}{2\sigma^2} \geq \ln\left(\frac{1-p}{p}\right)$$

$$\rightarrow \boxed{y \geq \frac{\ln\left(\frac{1-p}{p}\right)\sigma^2}{4}}$$

Question 5 - Metropolis Hastings Algorithm

در این مسأله قصد داریم تا با الگوریتم Metropolis Hastings آشنا شویم. این الگوریتم یک روش رایج برای تولید نمونه از یک توزیع پیچیده با استفاده از زنجیره MCMC (Markov chain Monte Carlo) می باشد.

در قانون بیز داریم که با داشتن یک احتمال پیشین، و likelihood قادر به استخراج احتمال پسین هستیم:

A)

استفاده برای استنباط پیرامون متغیر θ ← Posterior

$$\underbrace{\pi(\theta|x)}_{\text{Posterior}} = \frac{\overbrace{f(\theta|x)}^{\text{likelihood}}\ \overbrace{\pi(\theta)}^{\text{Prior}}}{\int f(\theta|x)\,\pi(\theta)\,d\theta}$$

در برخی حالات بدست آوردن احتمال پسین بصورت مستقیم، (به علت وجود انتگرال پیچیده در مخرج) سخت خواهد بود. در این حالات بهترین راهکار تخمین احتمال پسین به کمک روش های شبیه سازی monte carlo است.

الگوریتم Metropolis-Hastings یک روش تقریب توزیع با استفاده از MCMC است. تمام چیزی که این روش نیاز دارد عبارتی متناسب با چگالی (Density) توزیعی است که می خواهیم از آن نمونه گیری کنیم. این مقدار به طور معمول صورت کسر توزیع پیشین می باشد. برای توضیح الگوریتم فرض می کنیم توزیعی که می خواهیم از آن نمونه گیری کنیم دارای چگالی پیرو از تابع f است. همچنین نیاز به یک مقدار اولیه پیشنهادی برای توزیع داریم ($q(x|\theta)$). با توجه به توضیحات بالا سودوکد مربوطه به شکل روبرو است.

① انتخاب مقدار اولیه θ_0 (تخمین از روی داده یا تعریف توسط کاربر)

② برای $t = 1, 2, \dots$

a-2) نمونه گیری (θ^*) از توزیع پیشنهادی

b-2) محاسبه نسبت توزیع ها

$$r = \frac{f(\theta_t^*)\,q(\theta_{t-1}|\theta_t^*)}{f(\theta_{t-1})\,q(\theta_t^*|\theta_{t-1})}$$

③ مقداردهی مجدد (پرش) ←

$$\theta^t = \begin{cases} \theta^* & Pr = \min(r, 1) \\ \theta^{t-1} & \text{o.w.} \end{cases}$$

شهود پشت این الگوریتم این است که نسبت (r) زمانی بزرگتر از یک خواهد بود که توزیع مورد بررسی در θ_t^* بزرگتر از θ_{t-1} باشد. این بدین معنی است که مقدار θ_t که بدست آمده به سمت نقاط پرتراکم (high density) حرکت می کند و این یعنی که پس از چند تکرار توزیع بدست آمده نزدیک به توزیع Posterior مورد نظر خواهد شد.

$$\underbrace{\text{symmetric}}_{} \rightarrow q(x|y) = q(y|x)$$

در اغلب مواقع برای تخمین و انتخاب توزیع اولیه پیشنهادی از توزیع نرمال استفاده می شود و از آنجایی که این توزیع symmetric است کسر نسبت توزیع ها به شکل زیر تغییر خواهد کرد

$$r = \frac{f(\theta_t^*)}{f(\theta_{t-1})} \longrightarrow \text{Metropolis sampler}$$

B) ← داخل Notebook

Question 6 - Medical Test Paradox

1A) دلیل وقوع مسأله پارادوکس تفاوت مفهوم sensitivity یا accuracy و predictability

روش تست برای حل این مسائل احتمال ها را در نظر می گیرد اما نکته ای که وجود دارد این است که در این مسائل نمی توان برداشت دقیق کرد برای مثال در یک تست برای تشخیص وجود بیماری اگر تستی داشته باشیم که با دقت 95% بیماری را تشخیص می دهد و احتمال وجود بیماری در افراد 5% است نمی توانیم بگوییم فردی با آزمایش مثبت با 95% احتمال دچار بیماری است.

B) در ویدئو مذکور یک نمایش دیگر از قانون بیز را در حل دیگر مشاهده می کنیم که با مفهوم odds تعریف می شود

$$\underbrace{O(H|E)}_{\text{odd}} = O(H)\underbrace{\frac{P(E|H)}{P(E|\bar{H})}}_{\text{فاکتور بیز}}$$

برای درک بهتر یک آزمایش طراحی می کنیم. فرض می کنیم در یک نمونه گیری (1%) احتمال داشتن بیماری وجود دارد و تستی داریم که با دقت 90% قابلیت تشخیص دارند. مقادیر sensitivity و specificity برای این مسأله به شرح زیر است:

$$\begin{cases} \text{Sensitivity} = P(+|\text{بیماری}) = \frac{90}{100} \\ \text{Specificity} = P(-|\overline{\text{بیماری}}) = \frac{91}{100} \rightarrow P(+|\overline{\text{بیماری}}) = \frac{9}{100} \end{cases}$$

از مقدار sensitivity بدست آمده اینطور برداشت می شود که از هر 100 تست مثبت 90 تای آنها دچار بیماری هستند درصورتی که این با مقدار احتمال وجود بیماری که یک درصد هست در تناقض است و در نظر گرفته نشده است.

حال با روش جدید به بررسی می پردازیم. در روش odds از نسبت درصدها استفاده می شود یعنی برای مسأله ما که یک درصد احتمال بیماری وجود دارد odds به صورت [1 به 99] یا [1:99] است. در مرحله بعد فاکتور بیز را محاسبه می کنیم:

$\rightarrow$ Prior

موارد مثبت درست تشخیص داده شده $\leftarrow$

$$\text{Bayes factor} = \frac{P(+|\text{بیماری})}{P(+|\overline{\text{بیماری}})} = \frac{\%90}{\%9} = 10$$

موارد مثبت اشتباه تشخیص داده شده $\leftarrow$

درنهایت داریم $\leftarrow$

$$O(\text{بیماری}|+) = O(+|\text{بیماری}) \times \frac{P(+|\text{بیماری})}{P(+|\overline{\text{بیماری}})} = (1:99)\times 10$$

$$= 10:99 \rightarrow \frac{10}{109} = 0,09$$

پس احتمال اینکه فردی که آزمایش مثبت دارد واقعاً بیمار باشد تقریبا 9% است.

C) همانطور که در قسمت قبلی نیز دیدیم ←

$$\text{Byes factor} = \frac{P(+ \mid \text{بیمار})}{P(+ \mid \overline{\text{بیمار}})} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

این روش و استفاده از فاکتور بیز باعث کاهش حجم و هزینه و پیچیدگی در محاسبات می‌شود

Infection Probability

D) $\text{Prior} = \frac{8}{100} \rightarrow (8:92)$ $\quad \overset{\text{accuracy}}{\text{Sensitivity}} = 80\%$ $\quad P(+ \mid \overline{\text{infected}}) = \frac{10}{100} \rightarrow (10:90)$

$$\longrightarrow P(\text{infected} \mid +) = \text{Prior} \times \text{factor} = (8:92) \times 8 = (64:92) = \frac{64}{156} = 0{,}41$$