DS-GA-1002 HW6 Yoon Tae Park (yp2201@nyu.edu)

1. (a) This model makes sense since we assume that the coin is more likely to land on heads. Also, we don't know the exact bias so it is reasonable to build a model as being uniform between 1/2 and 1. (coin always lands on heads)

Let's say that,

$\tilde{r}$: result of the coin flip

$\tilde{\theta}$: parameter of Bernoulli

v $p(\text{heads}) = P_{\tilde{r}}(1) = \int_{\theta=1/2}^{1} f_{\tilde{\theta}}(\theta) P_{\tilde{r}|\tilde{\theta}}(1|\theta)\, d\theta$

Since $\tilde{r}$ is uniform between $1/2$ and 1,

$$f_{\tilde{\theta}}(\theta) = \begin{cases} 2, & (\frac{1}{2} \leq \theta \leq 1) \\ 0 & \text{others} \end{cases}$$

$$= \int_{\theta=\frac{1}{2}}^{1} 2\theta\, d\theta = \left[\theta^2\right]_{\frac{1}{2}}^{1} = 1 - \frac{1}{4} = \frac{3}{4}$$

v $p(\text{tails}) = P_{\tilde{r}}(0) = 1 - p(\text{heads}) = 1 - \frac{3}{4} = \frac{1}{4}$

(b) i) Conditional pdf of coin flip conditioned on tails

$$f_{\tilde{\theta}|\tilde{r}}(\theta|0) = \frac{f_{\tilde{\theta}}(\theta) \cdot P_{\tilde{r}|\tilde{\theta}}(0|\theta)}{P_{\tilde{r}}(0)} = \frac{2 \times (1-\theta)}{\frac{1}{4}} = \begin{cases} 8(1-\theta), & (\frac{1}{2} \leq \theta \leq 1) \\ 0, & \text{otherwise} \end{cases}$$

ii) Conditional pdf of coin flip conditioned on heads

$$f_{\tilde{\theta}|\tilde{r}}(\theta|1) = \frac{f_{\tilde{\theta}}(\theta) \times P_{\tilde{r}|\tilde{\theta}}(1|\theta)}{P_{\tilde{r}}(1)} = \frac{2\theta}{3/4} = \begin{cases} \frac{8}{3}\theta, & (\frac{1}{2} \leq \theta \leq 1) \\ 0, & \text{otherwise} \end{cases}$$

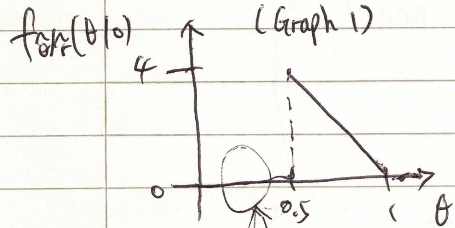
(b) (cont.) Both makes sense, as you see below graphs,

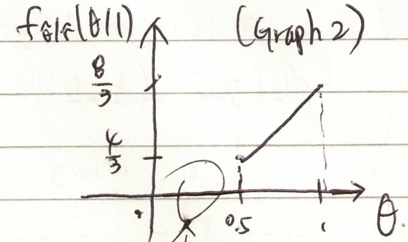
for graph 1 (conditional pdf conditioned on tails),

it is more skewed towards $1/2$. This happens because the coin flip is tails and model is being adjusted towards the coin flip being less biased

For graph 2 (conditional pdf conditioned on heads),

it is more skewed towards 1. This happens because the coin flip is heads and model is being adjusted towards the coin flip being more biased





(c) Yes, ~~Since~~ we'll always assign 0 probability density to any value of the bias below 0.5, we should reconsider our prior.

2.(a) $P_{\tilde{r}}(1) = 0.2$, $P_{\tilde{r}}(0) = 0.8$

$\begin{cases} P(\tilde{w}=1 \mid \tilde{r}=1) = 0.8 \\ P(\tilde{w}=0 \mid \tilde{r}=1) = 0.2 \end{cases}$ $\begin{cases} P(\tilde{w}=0 \mid \tilde{r}=0) = 0.75 \\ P(\tilde{w}=1 \mid \tilde{r}=0) = 0.25 \end{cases}$

$P(\tilde{w} \neq \tilde{r}) = P_{\tilde{w},\tilde{r}}(0,1) + P_{\tilde{w},\tilde{r}}(1,0)$

$= P_{\tilde{r}}(1) \times P_{\tilde{w}|\tilde{r}}(0,1) + P_{\tilde{r}}(0) \times P_{\tilde{w}|\tilde{r}}(1|0)$

$= 0.2 \times 0.2 + 0.8 \times 0.25$

$= 0.24$

(b) It is more reasonable to assume that $\tilde{h}$ and $\tilde{w}$ are conditionally independent given $\tilde{r}$.

Since $\tilde{h}$ and $\tilde{w}$ are both linked through the rain, if $\tilde{w}=1$ than $\tilde{h}$ will be high since it will probably rain and if $\tilde{w}=0$ than $\tilde{h}$ will be low since it will probably not rain. So, it is not reasonable to assume that $\tilde{h}$ and $\tilde{w}$ are independent.

But, if we condition on raining, we can say that $\tilde{h}$ and $\tilde{w}$ are conditinally independent given $\tilde{r}$, since $\tilde{h}$ is not used to produce $\tilde{w}$.

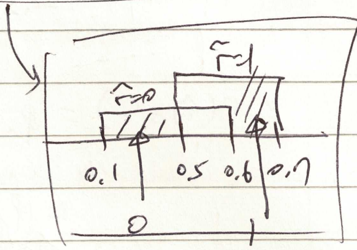
(c) $\begin{cases} \tilde{r}=1 , \ 0.5 \leq \tilde{h} \leq 0.7 \\ \tilde{r}=0 , \ 0.1 \leq \tilde{h} \leq 0.6 \end{cases}$ $\tilde{h}, \tilde{w} \xrightarrow{\text{cond'l. independent}} \tilde{r}$

$$P_{\tilde{r}|\tilde{w},\tilde{h}}(r|w,h) = \frac{f_{\tilde{h}|\tilde{w},\tilde{r}}(h|w,r) \cdot P_{\tilde{w}|\tilde{r}}(w|r) \cdot P_{\tilde{r}}(r)}{P_{\tilde{w}}(w) \times f_{\tilde{h}|\tilde{w}}(h|w)}$$

$$= \frac{f_{\tilde{h}|\tilde{r}}(h|r) \times P_{\tilde{w}|\tilde{r}}(w|r) \times P_{\tilde{r}}(r)}{P_{\tilde{w}}(w) \times f_{\tilde{h}|\tilde{w}}(h|w)}$$

c) (cont.) since $\begin{cases} \tilde{r}=1 \rightarrow 0.5 \leq \tilde{h} \leq 0.7 \rightarrow f_{\tilde{h}|\tilde{r}}(h|1) = 5 \\ \tilde{r}=0 \rightarrow 0.1 \leq \tilde{h} \leq 0.6. \rightarrow f_{\tilde{h}|\tilde{r}}(h|0) = 2 \end{cases}$

✓ $P_{\tilde{r}|\tilde{w},\tilde{h}}(1|w,h)$ is divided by variable ranges



$$P_{\tilde{r}|\tilde{w},\tilde{h}}(1|1,h) = \frac{f_{\tilde{h}|\tilde{r}}(h|1) \times P_{\tilde{w}|\tilde{r}}(1|1) \times P_{\tilde{r}}(1)}{f_{\tilde{h}|\tilde{r}}(h|0) \times P_{\tilde{w}|\tilde{r}}(1|0) \times P_{\tilde{r}}(0) + f_{\tilde{h}|\tilde{r}}(h|1) \times P_{\tilde{w}|\tilde{r}}(1|1) \times P_{\tilde{r}}(1)}$$

$$= \frac{5 \times 0.8 \times 0.2}{2 \times 0.25 \times 0.8 + 5 \times 0.8 \times 0.2} = 0.667$$

$$P_{\tilde{r}|\tilde{w},\tilde{h}}(1|0,h) = \frac{f_{\tilde{h}|\tilde{r}}(h|1) \times P_{\tilde{w}|\tilde{r}}(0|1) \times P_{\tilde{r}}(1)}{f_{\tilde{h}|\tilde{r}}(h|0) \times P_{\tilde{w}|\tilde{r}}(0|0) \times P_{\tilde{r}}(0) + f_{\tilde{h}|\tilde{r}}(h|0) \times P_{\tilde{w}|\tilde{r}}(0|1) \times P_{\tilde{r}}(1)}$$

$$= \frac{5 \times 0.2 \times 0.2}{2 \times 0.75 \times 0.8 + 5 \times 0.2 \times 0.2} = 0.143$$

$$\therefore P_{\tilde{r}|\tilde{w},\tilde{h}}(1|w,h) = \begin{cases} 0 & (0.1 \leq h \leq 0.5) \\ 0.667 & (0.5 \leq h \leq 0.6,\ w=1) \\ 0.143 & (0.5 \leq h \leq 0.6,\ w=0) \\ 1 & (0.6 \leq h \leq 0.7) \end{cases}$$

→ We predict no rain if $0.1 \leq h \leq 0.5$ or $(0.5 \leq h \leq 6$ and $w=0)$

Otherwise, we predict rain

2-(d) There would be two cases that would be an error.

$$P(\text{Error}) = P(\tilde{r}=0, w=1, 0.5 \leq \tilde{h} \leq 0.6) + P(\tilde{r}=1, w=0, 0.5 \leq \tilde{h} \leq 0.6)$$

(If $0.1 \leq h \leq 0.5$ or $0.6 \leq h \leq 0.7$, we are always correct)

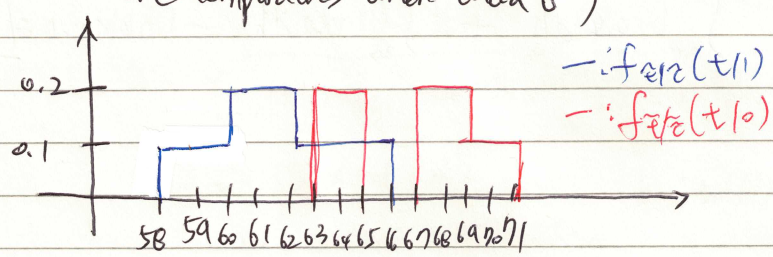
$$= \int_{h=0.5}^{0.6} \underbrace{f_{\tilde{h}|\tilde{r}}(h|0)}_{2} \underbrace{P_{\tilde{w}|\tilde{r}}(1|0)}_{0.25} \cdot \underbrace{P_{\tilde{r}}(0)}_{0.8}\, dh + \int_{h=0.5}^{0.6} \underbrace{f_{\tilde{h}|\tilde{r}}(h|1)}_{5} \cdot \underbrace{P_{\tilde{w}|\tilde{r}}(0|1)}_{0.2} \cdot \underbrace{P_{\tilde{r}}(1)}_{0.2}\, dh$$

$$= \int_{0.5}^{0.6} 0.4\, dh + \int_{0.5}^{0.6} 0.2\, dh = 0.6 \times 0.1 = \boxed{0.06}$$

3.(a) KDE :

$$\begin{cases} f_{\tilde{t}|\tilde{c}}(t|0) = \frac{1}{n_0}\sum_{n=1}^{n_0}\frac{1}{2}\Pi\left(\frac{t-d_{i0}}{2}\right) \\ f_{\tilde{t}|\tilde{c}}(t|1) = \frac{1}{n_1}\sum_{n=1}^{n_1}\frac{1}{2}\Pi\left(\frac{t-d_{i1}}{2}\right) \end{cases}$$

(Π : rectangular kernel with unit depth $d_{10}, \ldots d_{n_0 0}$ the temperatures when chad x, and $d_{11}, \ldots d_{n_1 1}$ the temperatures when chad 0)



(b) $\left(\begin{array}{l} f_{\tilde{t}|\tilde{c}}(68|0) = 0.2 \\ f_{\tilde{t}|\tilde{c}}(68|1) = 0 \end{array}\right)$, therefore ML estimate is that Chad is not at the office

(c) For empirical pmf, we just count.

$\therefore \left(P_{\tilde{c}}(0) = \frac{5}{15} = \frac{1}{3},\quad P_{\tilde{c}}(1) = \frac{10}{15} = \frac{2}{3} \right)$

(d) Applying Bayes' rule,

$$P_{\tilde{c}|\tilde{t}}(0|64) = \frac{P_{\tilde{c}}(0)\cdot f_{\tilde{t}|\tilde{c}}(64|0)}{P_{\tilde{c}}(0)\cdot f_{\tilde{t}|\tilde{c}}(64|0) + P_{\tilde{c}}(1)\cdot f_{\tilde{t}|\tilde{c}}(64|1)}$$

$$= \frac{\frac{1}{3}\times 0.2}{\frac{1}{3}\times 0.2 + \frac{2}{3}\times 0.1} = \boxed{\frac{1}{2}}$$

$$P_{\tilde{c}|\tilde{t}}(1|64) = 1 - P_{\tilde{c}|\tilde{t}}(0|64) = \boxed{\frac{1}{2}}$$

$\therefore$ There is a 50% chance that Chad is there

(c) Both $f_{\tilde{x}|\tilde{z}}(57|0)$ and $f_{\tilde{x}|\tilde{z}}(57|1)$ are zero, so

ML estimate and Bayesian posterior are inconclusive.

So, we can alleviate the problem by using parametric distribution of Gaussian. Then, $f_{\tilde{x}|\tilde{z}}(57|0)$ and $f_{\tilde{x}|\tilde{z}}(57|1)$ would not be at least zero as long as the distribution has nonzero values for all the real line.

This would make us conduct ML estimate or MAP.

(+) for non-parametric solution, using a kernel with a larger width would help.

4. (a) $P_{\tilde{h}|\tilde{s},\tilde{c}}(h|s,c) = \dfrac{P_{\tilde{h}}(h) \times P_{\tilde{s},\tilde{c}|\tilde{h}}(s,c|h)}{P_{\tilde{s},\tilde{c}}(s,c)} = \dfrac{P_{\tilde{h}}(h) \times P_{\tilde{s}|\tilde{h}}(s|h) \cdot P_{\tilde{c}|\tilde{h}}(c|h)}{P_{\tilde{s},\tilde{c}}(s,c)}$

Denominator doesn't depend on h, so MAP rule can be applied as below,

$$\hat{h}(s,c) = \begin{cases} 0 & \text{if } P_{\tilde{h}}(1) \times P_{\tilde{s}|\tilde{h}}(s|1) \times P_{\tilde{c}|\tilde{h}}(c|1) < P_{\tilde{h}}(0) \times P_{\tilde{s}|\tilde{h}}(s|0) \times P_{\tilde{c}|\tilde{h}}(c|0), \\ 1 & \text{otherwise} \end{cases}$$

(b)

QUESTION (b)

Non-parametric model Compute empirical pmf, derive the conditional pmf, and estimate the MAP decision by the mode of posterior distribution $p_{\tilde{h}|\tilde{s},\tilde{c}}$. (The MAP estimates should be $\tilde{h} = 0$ or 1)

In [2]:

```
# Estimate the pmf of H, i.e. P(H=0) and P(H=1)
P_H0 = count_x_eq_val(data['heart_disease'], 0.0)
P_H1 = count_x_eq_val(data['heart_disease'], 1.0)

# Estimate the conditional pmf of S given H, i.e. P(S|H=0) and P(S|H=1)
P_S_H0 = np.zeros(2)
P_S_H1 = np.zeros(2)

h0 = ind_x_eq_val(data['heart_disease'], 0.0)
h1 = ind_x_eq_val(data['heart_disease'], 1.0)

for ind_S in range(2):
    P_S_H0[ind_S] = len(set(h0) & set(ind_x_eq_val(data['sex'], ind_S))) / len(h0)
    P_S_H1[ind_S] = len(set(h1) & set(ind_x_eq_val(data['sex'], ind_S))) / len(h1)

# Estimate the conditional pmf of C given H, i.e. P(C|H=0) and P(C|H=1)
P_C_H0 = np.zeros(4)
P_C_H1 = np.zeros(4)
for ind_C in range(4):
    P_C_H0[ind_C] = len(set(h0) & set(ind_x_eq_val(data['chest_pain'], ind_C))) / len(h0)
    P_C_H1[ind_C] = len(set(h1) & set(ind_x_eq_val(data['chest_pain'], ind_C))) / len(h1)

# Calculate the MAP estimate
MAP_estimate_S_C = []
for i in range(len(data['heart_disease_test'])):
    h1_based = P_H1 * P_S_H1[int(data['sex_test'][i])] * P_C_H1[int(data['chest_pain_test'][i])]
    h0_based = P_H0 * P_S_H0[int(data['sex_test'][i])] * P_C_H0[int(data['chest_pain_test'][i])]

    if h1_based < h0_based:
        MAP_estimate_S_C.append(0)
    else:
        MAP_estimate_S_C.append(1)

# Calculate the error rate - i.e. the proportion of all predictions that were incorrect
error_rate_S_C = 0

for i in range(len(data['heart_disease_test'])):
    if MAP_estimate_S_C[i] != data['heart_disease_test'][i]:
        error_rate_S_C += 1

error_rate_S_C /= len(data['heart_disease_test'])

print("Probability of error " + str(error_rate_S_C))
```

```
Probability of error 0.18
```

(c) Same logic as (a).

$$P_{\tilde{h}|\tilde{s},\tilde{c},\tilde{x}}(h|s,c,x) = \frac{P_{\tilde{h}}(h)\cdot P_{\tilde{s},\tilde{c}|\tilde{h}}(s,c|h)\cdot f_{\tilde{x}|\tilde{h},\tilde{s},\tilde{c}}(x|h,s,c)}{f_{\tilde{x}|\tilde{s},\tilde{c}}(x|s,c)\cdot P_{\tilde{s},\tilde{c}}(s,c)}$$

$$= \frac{P_{\tilde{h}}(h)\cdot P_{\tilde{s}|\tilde{h}}(s|h)\cdot P_{\tilde{c}|\tilde{h}}(c|h)\cdot f_{\tilde{x}|\tilde{h}}(x|h)}{f_{\tilde{x}|\tilde{s},\tilde{c}}(x|s,c)\cdot P_{\tilde{s},\tilde{c}}(s,c)}$$

Denominator doesn't depend on h, so MAP rule can be applied as below,

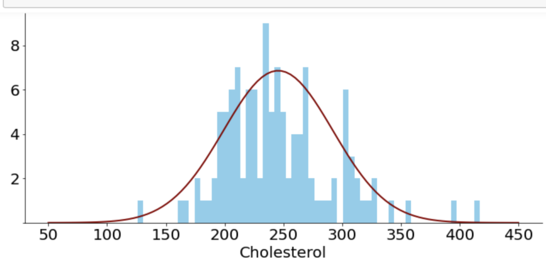
$$\hat{h}(s,c) = \begin{cases} 0 & \text{if } P_{\tilde{h}}(1)\cdot P_{\tilde{s}|\tilde{h}}(s|1)\cdot P_{\tilde{c}|\tilde{h}}(c|1)\cdot f_{\tilde{x}|\tilde{h}}(x|1) \\ & < P_{\tilde{h}}(0)\cdot P_{\tilde{s}|\tilde{h}}(s|0)\cdot P_{\tilde{c}|\tilde{h}}(c|0)\cdot f_{\tilde{x}|\tilde{h}}(x|0) \\ 1 & \text{otherwise.} \end{cases}$$

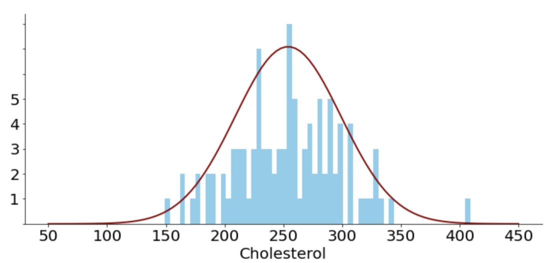
QUESTION (d)

Maximum likelihood estimates Find the parameters of two normal distributions ($\tilde{x}|\tilde{h}=1$ and $\tilde{x}|\tilde{h}=0$) that maximize the likelihood functions.

In [13]:

```
## Estimate MLE of X given H
mean_X_H = np.zeros(2)
std_X_H = np.zeros(2)
H = [[0 for j in range(1)] for i in range(2)]
H[0] = ind_x_eq_val(data['heart_disease'], 0.0)
H[1] = ind_x_eq_val(data['heart_disease'], 1.0)

mean_X_H[0]= np.mean(data['cholesterol'][H[0]])
std_X_H[0] = np.std(data['cholesterol'][H[0]])
mean_X_H[1]= np.mean(data['cholesterol'][H[1]])
std_X_H[1] = np.std(data['cholesterol'][H[1]])

n_plot = 100
for i in range(2):
    plt.figure(figsize=(12, 5))
    ax = plt.subplot(111)
    ax.spines["top"].set_visible(False)
    ax.spines["right"].set_visible(False)
    ax.get_xaxis().tick_bottom()
    ax.get_yaxis().tick_left()
    yticks = ax.yaxis.get_major_ticks()
    yticks[0].label1.set_visible(False)
    plt.xticks(fontsize=20)
    plt.yticks(fontsize=20)
    plt.xlabel("Cholesterol", fontsize=20)

    plt.hist(data['cholesterol'][H[i]],
             60, edgecolor = "none", color="skyblue")

    plt.plot(np.linspace(50, 450, n_plot), 800*gaussian(np.linspace(50, 450, n_plot),
                      mean_X_H[i], std_X_H[i]), color="darkred", lw=2)
```





QUESTION(e)

MAP decision compute posterior $p_{\tilde{h}|\tilde{s},\tilde{c},\tilde{x}}$ and derive MAP

In [15]:
```
# Calculate the MAP estimate
MAP_estimate_S_C_X = []
for i in range(len(data['heart_disease_test'])):
    h1_based = P_H1 * P_S_H1[int(data['sex_test'][i])] * P_C_H1[int(data['chest_pain_test'][i])] \
               * gaussian(data['cholesterol_test'][i], mean_X_H[1], std_X_H[1])
    h0_based = P_H0 * P_S_H0[int(data['sex_test'][i])] * P_C_H0[int(data['chest_pain_test'][i])] \
               * gaussian(data['cholesterol_test'][i], mean_X_H[0], std_X_H[0])

    if h1_based < h0_based:
        MAP_estimate_S_C_X.append(0)
    else:
        MAP_estimate_S_C_X.append(1)

# Calculate the error rate
error_rate_S_C_X = 0

for i in range(len(data['heart_disease_test'])):
    if MAP_estimate_S_C_X[i] != data['heart_disease_test'][i]:
        error_rate_S_C_X += 1

error_rate_S_C_X /= len(data['heart_disease_test'])

print("Probability of error using cholesterol " + str(error_rate_S_C_X))
```

```
Probability of error using cholesterol 0.14
```

(f) If we have a lot of data to estimate the joint distribution, then this approach would work well. However, we would need more than 218 patients data. We have already lots of parameters, (ex) joint pmf of $\tilde{c}, \tilde{s}, \tilde{h}$ has $2 \times 2 \times 4 - 1 = 15$ parameters) and there would be about 15 patients per parameter

It should be more than that to differentiate the parameters.

(We need more data?!)

So, with limited data, decreasing number of parameters by using some assumptions as above questions would work better.