

## Section 9: Quiz #3 Review

---

### 1. Continuous Joint Distributions (courtesy of Oishi Banerjee)

- a. Let  $X, Y,$  and  $Z$  be independent Normal variables with means of  $\mu_X = 4, \mu_Y = 5,$  and  $\mu_Z = 6$  and variances  $\sigma_X^2 = 16, \sigma_Y^2 = 25,$  and  $\sigma_Z^2 = 36$ . If we assume  $A = X + Y$  and  $B = Y + Z$  are each sums of independent Normal variables, then what is the joint distribution of  $A$  and  $B$ ? Restated, what is their Bivariate Normal distribution?

$$(A, B) \sim N(\mu, \Sigma), \mu = \begin{bmatrix} \mu_X + \mu_Y \\ \mu_Y + \mu_Z \end{bmatrix}, \Sigma = \begin{bmatrix} \text{Var}(A) & \text{Cov}(A, B) \\ \text{Cov}(A, B) & \text{Var}(B) \end{bmatrix}$$

Now,  $\text{Var}(A) = \text{Var}(X+Y)$ , and because  $X$  and  $Y$  are independent,  $\text{Var}(A) = \text{Var}(X+Y) = \sigma_X^2 + \sigma_Y^2$ . Similarly,  $\text{Var}(B) = \text{Var}(Y+Z) = \sigma_Y^2 + \sigma_Z^2$ . Also,  $\text{Cov}(A, B) = \text{Cov}(X+Y, Y+Z)$ , but because  $X, Y,$  and  $Z$  are independent,  $\text{Cov}(A, B) = \text{Cov}(X+Y, Y+Z) = \text{Cov}(Y, Y) = \sigma_Y^2$ . Therefore,

$$\mu = \begin{bmatrix} \mu_X + \mu_Y \\ \mu_Y + \mu_Z \end{bmatrix} = \begin{bmatrix} 9 \\ 11 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_X^2 + \sigma_Y^2 & \sigma_Y^2 \\ \sigma_Y^2 & \sigma_Y^2 + \sigma_Z^2 \end{bmatrix} = \begin{bmatrix} 41 & 25 \\ 25 & 61 \end{bmatrix}$$

- b. Suppose hundreds of thousands (that is, a sufficiently large number) of student scores on a 150-question exam are distributed according to the following random variable:

$$R = \sum_{i=1}^{50} M_i + 0.5 \sum_{j=1}^{100} W_j \tag{1}$$

Each of the  $M_i$  are independent and identically distributed (IID) Beta random variables—yes, the questions are scored on a continuous scale from 0 to 1—and the  $W_j$  are separate IID Beta random variables, where all  $W_j$  are independent of all  $M_i$ . The Beta parameters are  $\alpha_M = 10, \beta_M = 2, \alpha_W = 8,$  and  $\beta_W = 4$ . If we sample 100 student scores  $R_1, \dots, R_n$  IID according to the distribution of  $R$  above, what is the distribution of the sample mean  $\bar{R}$ ?

$$E[M_i] = \frac{\alpha_M}{\alpha_M + \beta_M} = 0.83333$$

$$E[W_i] = \frac{\alpha_W}{\alpha_W + \beta_W} = 0.66667$$

$$\text{Var}(M_i) = \frac{\alpha_M \beta_M}{(\alpha_M + \beta_M)^2 (\alpha_M + \beta_M + 1)} = 0.01068$$

$$\text{Var}(W_i) = \frac{\alpha_W \beta_W}{(\alpha_W + \beta_W)^2 (\alpha_W + \beta_W + 1)} = 0.01709$$

We can compute  $R$ 's expectation using linearity of expectation. Because  $R$  is a sum of independent RVs, we can compute  $R$ 's variance by summing up the variance of the independent  $M_i$  and  $W_i$ 's as below:

$$\begin{aligned} E[R] &= 50 E[M_i] + 0.5 \cdot 100 E[W_i] = 75 \\ \text{Var}(R) &= 50 \text{Var}(M_i) + 0.25 \cdot 100 \text{Var}(W_i) = 0.961 \end{aligned}$$

As an aside,  $R$  can be approximated as  $R \sim N(75, 0.961)$ , since the sums of both question types  $M_i$  and  $W_i$  respectively approach Normal distributions according to the Central Limit Theorem, and the sum of independent Normal distributions is itself a Normal distribution.

The distribution of the sample mean  $\bar{R}$  is then given by:

$$\begin{aligned} \bar{R} &= \frac{1}{100} \sum_{i=1}^{100} R_i \sim N\left(75, \frac{1}{100} 0.961\right) \\ &\sim N(75, 0.0096) \end{aligned}$$

## 2. Bootstrapping and Null Hypotheses (courtesy of Oishi Banerjee)

While testing the efficacy of a new drug, Skylar Pharmaceuticals has collected 1000 data samples. Most of the samples came from patients who were treated with the drug, but the rest came from patients who received a placebo. Skylar observed that the sample mean blood pressure in the treated group was 80, while the sample mean blood pressure in the placebo group was 86. To demonstrate the difference is statistically significant, Skylar implemented the following to produce a p-value.

Data scientists at Skylar wrote the following bootstrapping code to arrive at a p-value to suggest the difference isn't statistically significant. Unfortunately (or fortunately, depending on your point of view), their code is not right.

```
import numpy as np

def resample(whole, num_samples):
    return np.random.choice(whole, num_samples, replace=True)

# list_treat is an ordinary 1-d numpy array
# it contains all the diastolic blood pressures of
# each patient who was treated

# list_placebo is an ordinary 1-d numpy array
# it contains the diastolic blood pressures of each
# patient who received a placebo
def pvalue(list_treat, list_placebo):
    # np.concatenate will make a 1000-element array #containing the
    # elements of both list_treat and list_placebo
    whole = np.concatenate([list_treat, list_placebo])
```

```

threshold = np.mean(list_treat) - np.mean(list_placebo)
counter, num_trials = 0, 100000
for trial in range(num_trials):
    sample_treat = resample(list_treat, 500)
    sample_placebo = resample(list_placebo, 500)
    mean_treat = np.mean(sample_treat)
    mean_placebo = np.mean(sample_placebo)
    new_diff = np.abs(mean_treat - mean_placebo)
    if new_diff == threshold: counter += 1
return counter/num_trials

```

Point out the algorithmic errors. Be clear what Skylar should do instead, explaining why each change you would make is necessary for correct bootstrapping.

- As written, the threshold will be -6, so new\_diff can never be smaller than threshold! Because we're really only concerned with magnitudes, Skylar should replace threshold with np.abs(threshold).
- To simulate the null hypothesis, we should sample from our new combined distribution. As a result both calls to resample should pass in whole, not list\_treat or list\_placebo.
- Though we're now sampling from our new combined distribution, we want to stay true to the design of the original experiment in every other way. Therefore we should make sure sample\_treat has as many elements as list\_treat and sample\_placebo has as many elements as list\_placebo. The 500s should be replaced with len(list\_treat) and len(list\_placebo) respectively.
- When bootstrapping, we count up how many times we see a result as dramatic or more dramatic than ours under the null hypothesis. As a result, we should check if new\_diff is greater than or equal to threshold.

### 3. Naïve Bayes (courtesy of David Varodayan and Lisa)

Suppose we observe two discrete input variables  $X_1$  and  $X_2$  and want to predict a single binary output variable  $Y$  (which can have values 0 or 1). We know that the functional forms for the input variables are  $X_1 \sim \text{Poi}(\lambda)$  and  $X_2 \sim \text{Ber}(p)$ , but we are not given the values of the parameters  $\lambda$  or  $p$ . We are, however, given a dataset of 9 training instances (shown at right.)

$X_1$	$X_2$	$Y$	$X_1$	$X_2$	$Y$
1	1	0	3	1	1
3	0	0	5	0	1
7	1	0	5	1	1
9	0	0	5	1	1
			7	1	1

- Use Maximum Likelihood Estimation to estimate the parameters  $\lambda$  and  $p$  in the case where  $Y = 0$  as well as the case  $Y = 1$ . You should have four parameter estimates:  $\lambda_0$  and  $p_0$  for when  $Y = 0$ , and  $\lambda_1$  and  $p_1$  for when  $Y = 1$ .

$$\lambda_0 = \frac{1}{4}(1 + 3 + 7 + 9) = \frac{20}{4} = 5 \qquad p_0 = \frac{1}{4}(1 + 0 + 1 + 0) = \frac{1}{2}$$

$$\lambda_1 = \frac{1}{5}(3 + 5 + 5 + 5 + 7) = \frac{25}{5} = 5 \qquad p_1 = \frac{1}{5}(1 + 0 + 1 + 1 + 1) = \frac{4}{5}$$

b. Use Maximum Likelihood Estimation to estimate the probability  $P(Y = 1)$ .

$$P(Y = 1) = 5/9.$$

c. You observe the following testing instance:  $(X_1, X_2) = (2, 0)$ . Using the Naïve Bayes assumption, predict the output  $Y$  for the testing instance. For this problem, showing how you computed your prediction is worth more points than the final answer.

We predict  $Y = 0$  if the following Naïve Bayes inequality holds:

$$P(Y = 1)P(X_1 = 2|Y = 1)P(X_2 = 0|Y = 1) \stackrel{?}{<} P(Y = 0)P(X_1 = 2|Y = 0)P(X_2 = 0|Y = 0)$$

$$\frac{5}{9} \left( \frac{\lambda_1^2}{2!} e^{-\lambda_1} \right) \left( 1 - \frac{4}{5} \right) \stackrel{?}{<} \frac{4}{9} \left( \frac{\lambda_0^2}{2!} e^{-\lambda_0} \right) \left( 1 - \frac{1}{2} \right)$$

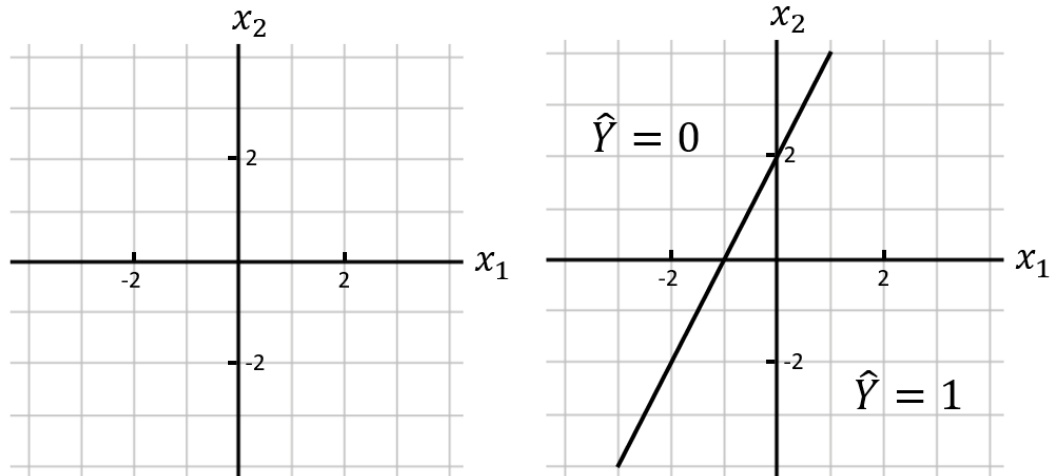
$$\frac{5}{9} \left( \frac{5^2}{2!} e^{-5} \right) \frac{1}{5} \stackrel{?}{<} \frac{4}{9} \left( \frac{5^2}{2!} e^{-5} \right) \frac{1}{2}$$

$$\frac{5}{9} \cdot \frac{1}{5} \stackrel{?}{<} \frac{4}{9} \cdot \frac{1}{2}$$

$$\frac{1}{9} < \frac{2}{9}$$

Since the last inequality is true, that means the first inequality was true, so we predict  $Y = 0$ .

#### 4. Logistic regression (courtesy of David Varodayan)



The two parts of this problem are unrelated.

- a. **Prediction.** Suppose you have trained a logistic regression classifier that accepts as input a data point  $(x_1, x_2)$  and predicts a class label  $\hat{Y}$ . The parameters of the model are  $(\theta_0, \theta_1, \theta_2) = (2, 2, -1)$ . On the axes, draw the decision boundary  $\theta^T \mathbf{x} = 0$  and clearly mark which side of the boundary predicts  $\hat{Y} = 0$  and which side predicts  $\hat{Y} = 1$ .

$\theta^T \mathbf{x}$  can be expanded as  $2 + 2x_1 - x_2 = 0$  because  $x_0 = 1$  by definition. The prediction is 1 when  $\theta^T \mathbf{x} > 0$ . For example, the origin  $(x_1, x_2) = (0, 0)$  yields  $\theta^T \mathbf{x} = 2$ , which gives us the prediction  $\hat{Y} = 1$ .

See the graph above, to the right of the original.

- b. **Training.** The logistic regression parameter update equation is

$$\theta_j^{\text{new}} = \theta_j^{\text{old}} + \eta \sum_{i=1}^n [y^{(i)} - \sigma(\theta^{\text{old}T} \mathbf{x}^{(i)})] x_j^{(i)}$$

Your training set consists of two data points  $(x_1^{(1)}, y^{(1)}) = (1, 1)$  and  $(x_1^{(2)}, y^{(2)}) = (-1, 0)$ . Given  $(\theta_0^{\text{old}}, \theta_1^{\text{old}}) = (0, 0)$  and  $\eta = 0.1$ , find  $(\theta_0^{\text{new}}, \theta_1^{\text{new}})$ .

First notice that  $(\theta_0^{\text{old}}, \theta_1^{\text{old}}) = (0, 0)$  implies that  $\sigma(\theta^{\text{old}T} \mathbf{x}^{(i)}) = \sigma(0) = 0.5$ . Therefore,

$$\begin{aligned} \theta_0^{\text{new}} &= 0 + 0.1 ([1 - 0.5] (1) + [0 - 0.5] (1)) && \text{since } x_0^{(i)} = 1 \text{ by definition} \\ &= 0 + 0.1(0.5 - 0.5) = 0 \\ \theta_1^{\text{new}} &= 0 + 0.1 ([1 - 0.5] (1) + [0 - 0.5] (-1)) \\ &= 0 + 0.1(0.5 + 0.5) = 0.1 \end{aligned}$$