Continuous RVs, Bayes and Confidence Interval Practice Problems

Lab 3

Continuous RVs Practice Problem

P1. Let X and Y be two jointly continuous random variables with join PDF

$$f_{XY}(x,y) = \begin{cases} x + cy^2 & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

Find the constant c

P2. Suppose a straight stick is broken in three at two points chosen independently at random along its length. What is the chance that the three sticks so formed can be made into the sides of a triangle?

Continuous RVs Practice Problem

P3. X, Y be independent random variables that have normal distribution ~ (0,1)

U=min(X,Y), V=max(X,Y). Find E[U] and calculate the Cov(U,V)

Hints:

$$E[U] = E[\min(X,Y)] = E\left[\frac{1}{2}(X+Y-|X-Y|)\right]$$

$$\Pr_{X\sim N(0,\sigma^2)}[X=x] = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-x^2}{2\sigma^2}\right)$$

$$E_{Z\sim N(0,\sigma^2)}[Z] = \int_{-\infty}^{\infty} z \cdot \Pr[Z=z]dz$$

Bayes Practice Problem

Purpose of this exercise:

1. Practice Bayes with continuous random variables
2. Interpret equations intuitively
3. Intuition behind Bayes theorem

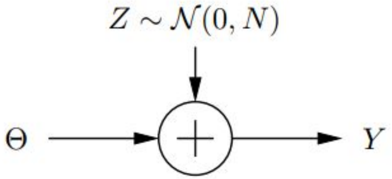
Review:

1. Bayes computes P(A | B) where A is a class, and B is an input
2. Prior - P(A)
3. Likelihood - P(B | A)
4. Posterior = Prior * Likelihood

Bayes Practice Problem

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1 - p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

Describe the situation, what is the diagram showing?

What is the difference between Y and y

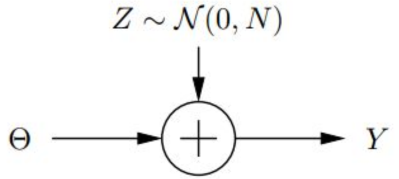
Give the equation and describe each of these terms in your own words

Prior: ?
Likelihood: ?
Posterior: ?

Bayes Practice Problem

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

Describe the situation, what is the diagram showing?

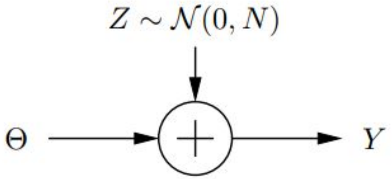
Signal (-1 or 1) sent from Theta to Y. Random interfering noise added to the original signal.

For instance, Y might receive 1, but does that mean Theta sent 1, or theta sent -1 and there was a noise value of 2?

Bayes Practice Problem

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

Give the equation and describe each of these terms in your own words

Prior:
P(Theta=1) = p

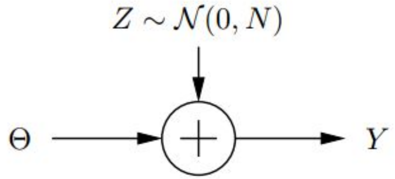
P(Theta = -1) = 1-p

The prior represents the underlying probability for the class' existence

Bayes Practice Problem

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

Give the equation and describe each of these terms in your own words

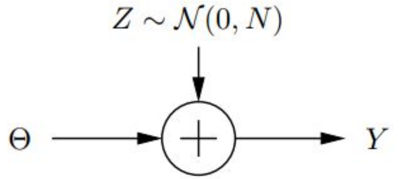
Likelihood:
P(Y | Theta)

For each class, we assume that if that class was the correct class, what are the odds of getting our observation, Y.

Bayes Practice Problem

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

Give the equation and describe each of these terms in your own words

Posterior:
P(Theta | Y)

Final probability: given our observation, Y, what is the most likely class it comes from

Solving the Posterior

Prior: P(Theta)

Likelihood: P(Y | Theta)

Posterior: P(Theta | Y)

$$p_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta)p_{\Theta}(\theta)}{\sum_{\theta'} p_{\Theta}(\theta')f_{Y|\Theta}(y|\theta')}$$

Need to find: Prior, Likelihood

Solving the Posterior - prior

$$p_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta)p_\Theta(\theta)}{\sum_{\theta'} p_\Theta(\theta')f_{Y|\Theta}(y|\theta')}$$

Prior: P(Theta)

Likelihood: P(Y | Theta)

Posterior: P(Theta | Y)

Prior:

We know the prior: P(Theta=1) = p, P(Theta= -1) = 1-p

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

Solving the Posterior - Likelihood

Prior: P(Theta)

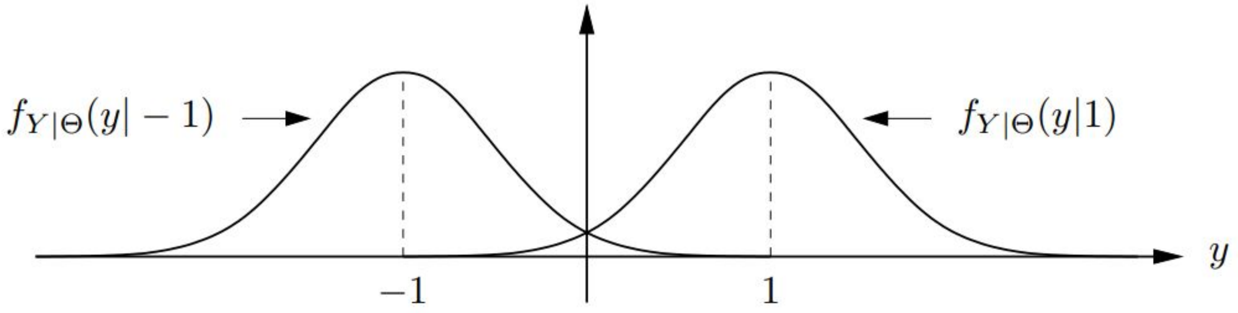
Likelihood: P(Y | Theta)

Posterior: P(Theta | Y)

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

$$Z \sim \mathcal{N}(0, N)$$

Likelihood: P(Y | Theta)



Therefore, $Y|\{\Theta = +1\} \sim \mathcal{N}(+1, N)$. Also, $Y|\{\Theta = -1\} \sim \mathcal{N}(-1, N)$

Solving the posterior - final

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

prior

$$p_{\Theta|Y}(1|y) = \frac{\frac{p}{\sqrt{2\pi N}} e^{-\frac{(y-1)^2}{2N}}}{\frac{p}{\sqrt{2\pi N}} e^{\frac{-(y-1)^2}{2N}} + \frac{(1-p)}{\sqrt{2\pi N}} e^{\frac{-(y+1)^2}{2N}}}$$

$$= \frac{pe^y}{pe^y + (1-p)e^{-y}} \quad \text{for} \quad -\infty < y < \infty$$

Application of central limit theorem

Conditions for CLT:

1. Distributions with finite variance
2. Independent and identically distributed random variables

Most probability distributions, such as Gaussian or uniform, fall under these two conditions.

Confidence Intervals

For a confidence interval from (a,b) with confidence x%:

There is a x% chance that the true value lies between (a,b)

CI according to CLT for **population mean and population SD**:

$$\Pr(-\sigma < X - \mu < \sigma) \approx 0.68$$
$$\Pr(-2\sigma < X - \mu < 2\sigma) \approx 0.95$$
$$\Pr(-3\sigma < X - \mu < 3\sigma) \approx 0.99$$

For computations on samples:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} \pm Z * \frac{\sigma}{\sqrt{n}}$$

Confidence	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291

Problem 1 - Taxes

Suppose that an accounting firm does a study to determine the time needed to complete one person's tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.

1. Computer the 95% confidence interval for the example above
2. Let us say that this study was conducted for a population of 10,000 people. If I wanted to find the 95% confidence interval for tax form time in a population of 100,000 people, should I sample less, the same, or more than 100 people in the new study?
3. If I wanted to increase the confidence, would the interval I report back be: Smaller, Equal to, or Larger w.r.t the initial interval

Confidence	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291

$$\bar{X} \pm Z * \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Problem 1 - Taxes Solutions

Suppose that an accounting firm does a study to determine the time needed to complete one person's tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.

1. **Computer the 95% confidence interval for the example above**
   a. (23.6)+2*(7.0/10) < 23.6 < (23.6)+2*(7.0/10)
   b. (22.228, 24.972)
2. **Does sample size w.r.t population size affect the confidence interval**
   a. No. Population sizes that are large enough do not affect affect confidence intervals. A sample of 100 of 10,000 is just as good as a sample of 100 in 100,000
3. **If I wanted to increase the confidence, would the interval I report back be: Smaller, Equal to, or Larger w.r.t the initial interval**
   a. Larger

Problem 2- Cookies

$$\bar{X} \pm Z * \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Suppose I am looking at grams of fat in a randomly selected set of 36 different cookie brands, and I report a confidence interval of (7.95 , 9.05) grams of fat with a known standard deviation of 2

1. **What is the confidence of this interval?**
   a. Mean = 8.5, sample std = 2/6 = ⅓
   b. Confidence = 90%
   c. 8.5 - 1.645(⅓) = 7.95, 8.5 + 1.645(⅓) = 9.05
2. **If I wanted to keep the confidence the same as the original problem, but decrease the interval, what would I do?**

   Sample more cookies brands

Confidence	Z
80%	1.282
85%	1.440
90%	1.645
95%	1.960
99%	2.576
99.5%	2.807
99.9%	3.291