ICS 435 HW3

In Woo Park

March 8, 2022

1 Problem Set: Probability

1. **Are X and Y independent?**

   By proof of contradiction they are not independent because if we assume they are independent then,

   $$P(X|Y) = P(X)$$
   $$P(Y|X) = P(Y)$$
   $$P(X,Y) = P(X)P(Y)$$

   these properties must be true. However, take the case where we choose an arbitrary integer such as the value 2,

   $$P(X = 2|Y = 2) = \frac{1}{3}$$

   This suggests that the value of P(X=2) = $\frac{1}{3}$ but that is not the case as P(X=2) = $\frac{3}{8}$.

2. **Find the joint probability P(X=1,Y=3)**.

   The joint probability is:

   $$P(X = 1, Y = 3) = \frac{1}{24}$$

   based on the provided figure.

3. **Find the marginal probability $\mathbf{P(X = 1)}$**.

   $$\begin{aligned}\frac{2}{24} + \frac{1}{24} + \frac{1}{24} &= \frac{4}{24}\\ &= \frac{1}{6}\end{aligned}$$

4. **Find the conditional probability $\mathbf{P(Y = 3|X = 1)}$**.

   $$\frac{1}{24} \div \frac{4}{24} = \frac{1}{4}$$

5. **Find the marginal probability distribution P(X). Remember that a probability distribution is a function**.

   $$P(X) = \sum_{y_j \in Y} P(X = x, Y = y_j)$$

6. **Find the conditional probability distribution $\mathbf{P(Y \mid X = 1)}$**.

   $$\begin{aligned}P(Y|X = 1) &= \frac{P(X = 1, Y = y)}{P(X = 1)}\\ &= \frac{P(y, 1)}{P(1, 1) + P(2, 1) + P(3, 1)}\end{aligned}$$

1.1 References

Probability Course
faculty.math.illinois.edu

2 Problem Set: Applied Probability

1. **Write the marginal distribution $\mathbf{P}(X_d)$ as a function.**

$$P(X_d = x_d) = \theta^{x_c}(1-\theta)^{1-x_c}$$

2. **Write the joint distribution P(Xc, Xd) as a table.**

	$X_d = 1$	$X_d = 2$	$X_d = 3$	$X_d = 4$	$X_d = 5$	$X_d = 6$
$X_c = 0$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$	$\frac{1-\theta}{6}$
$X_c = 1$	$\frac{\theta}{6}$	$\frac{\theta}{6}$	$\frac{\theta}{6}$	$\frac{\theta}{6}$	$\frac{\theta}{6}$	$\frac{\theta}{6}$

3. **Write the distribution $P(X_{product})$ as a piecewise-defined function.**

$$P(X_{product}|X_d = 6) = \begin{cases} \frac{1-\theta}{6} & \text{if } X_c = 0, \\ \frac{\theta}{6} & \text{if } X_c = 1, \\ 0 & \text{otherwise} \end{cases}$$

4. **Write the conditional distribution $P(X_{product}|X_d = 6)$ as a piecewise-defined function.**

$$P(X_{product}|X_d = 6) = \begin{cases} \frac{1-\theta}{6} & \text{if } X_c = 0, \\ \frac{\theta}{6} & \text{if } X_c = 1, \\ 0 & \text{otherwise} \end{cases}$$

2.1 References

Statistics How To
I snooped around the discord for clues.
I got help from Michael Rogers who took ICS 435 2021.

3 Problem Set: Bayes rule

1. **Suppose you receive a positive test result, what is the probability of having the variant?**

$$P(Z = 1|T = 1) = \frac{P(T = 1|Z = 1)P(Z = 1)}{P(T = 1|Z = 1)P(Z = 1) + P(T = 1|Z = 0)P(Z = 0)}$$

$$= \frac{0.999 \cdot 0.0005}{(0.999 \cdot 0.0005) + (0.01 \cdot 0.9995)}$$

$$= 0.04759\ldots$$

$$\approx 4.76\%$$

2. **Suppose you repeat the test, and the second test is also positive. Now what is the probability of having the disease? Assume that the tests are conditionally independent**

$$P(T1, T2|Z) = P(T1|Z)P(T2|Z)$$

$$= \frac{0.999^2 \cdot 0.0005}{(0.999^2 \cdot 0.0005) + (0.01^2 \cdot 0.9995)}$$

$$= 0.83312\ldots$$

$$\approx 83.33\%$$

3.1 References

Towards Data Science
Probablistic World
stat.cmu.edu
I used Symbolab to check my math.

4 Problem Set: Maximum Likelihood Estimate

1. **The coin is flipped once and comes up Heads. What is the maximum likelihood estimate (MLE) of θ**

We know that:

$$P(X_c = x_c) = \theta^{x_c}(1-\theta)^{1-x_c}$$

So if we used the Bernoulli distribution:

$$L_x(p) = P(x|p) = p^x(1-p)^{1-x}$$

We can write the likelihood of a particular sequence as the product of each flip:

$$L_x(p) = P(x|p) = \prod_{x \in X} p^x(1-p)^{1-x}$$

If we generalized this for n coin flips with h heads:

$$L(p) = p^h \cdot (1-p)^{n-h}$$

To simplify, we take the log of both sides and then take the derivative:

$$l(p) = h \cdot log(p) + (n-h) \cdot log(1-p)$$
$$l'(p) = \frac{h}{p} - \frac{n-h}{1-p}$$

We set the derivative to 0 to find the maximum of the function:

$$p = \frac{h}{n}$$

Here we see that the MLE for a coin is the number of heads divided by the number of flips:

$$MLE = \frac{1}{1} = 1$$

2. **The coin is flipped again and comes up Tails; so the the data is now Heads, Tails. What is the MLE of θ? Prove your answer.**

   We know that the MLE for a coin is the number of heads divided by the number of flips:

$$MLE = \frac{1}{2} = 0.5$$

3. **The coin is flipped N times, and comes up heads NH times and tails NT times. What is the MLE of θ? For those enrolled in ICS635: prove your answer.**

   We know that the MLE for a coin is the number of heads divided by the number of flips:

   You can actually prove this with induction using our scenario of $N_H = 1$ and $N_T = 1$ as base cases and that any further randomized iteration will follow suit:

$$MLE = \frac{N_H}{N_H + N_T} = \frac{N_H}{N}$$

4.1 References

Towards Date Science

5 Problem Set: PDF and MLE

1. **Suppose X is a continuous random variable drawn from the probability density function $f(X = x) = -\frac{x}{2} + 1$ defined in the range $x \in [0, 2]$. What is the probability that X is less than 1, i.e. $P(x < 1)$?**

   Let X be a continuous r.v. Then a probability distribution or probability density function (pdf) of X is a function f(x) such that for any two numbers a and b with a $\leq$ b, we have:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

   Where:

$$f_X(x) = \begin{cases} -\frac{x}{2} + 1 & \text{if } 0 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

   Therefore:

$$\begin{aligned} \int_0^2 (-\frac{x}{2} + 1)dx &= -\int_0^2 \frac{x}{2}dx + \int_0^2 1 \cdot dx \\ &= -\frac{1}{2}\int_0^2 xdx + 2 \\ &= -\frac{1}{2} \cdot \frac{x^2}{2}\Big|_0^2 + 2 \\ &= -\frac{1}{2} \cdot 2 + 2 \\ &= 1 \end{aligned}$$

Where:

$$
\begin{aligned}
P(X < 1) &= \int_{-\infty}^{1} (-\frac{x}{2} + 1)dx \\
&= -\frac{1}{2}\int_0^1 xdx + \int_0^1 1 \cdot dx \\
&= -\frac{1}{2} \cdot \frac{x^2}{2}\Big|_0^1 + 1 \\
&= -\frac{1}{4} + 1 \\
&= \frac{3}{4}
\end{aligned}
$$

2. **Say we observe three data points, $\{1.2, 3.6, 1.2\}$, and we model these as i.i.d. samples from a uniform distribution $U(0, a)$ parameterized by positive real number $a \in \mathbb{R}^+$. What is the maximum likelihood estimate of a?**

   The product of individual marginals gives us the joint density. Given the conditions provided:

$$
f(X|a) = \begin{cases} \frac{1}{a} & \text{for } 0 \le X_i \le a \\ 0 & \text{otherwise.} \end{cases}
$$

   Where:

$$
\begin{aligned}
\mathcal{L}(a|X_1, ...X_n) &= \prod_{i=1}^{n} \frac{1}{a^i} \\
&= \frac{1}{a^n} \\
&\Rightarrow \hat{a} \\
&= X_n \\
&= \max\{X_1, X_2, ...X_n\}
\end{aligned}
$$

   Therefore from the three data points 1.2, 3.6, 1.2 the maximum likelihood estimate of a is 3.6.

5.1 References

colorado.edu
I used Symbolab to check my math.
I snooped around the discord for clues.
I got help from Michael Rogers who took ICS 435 2021.

6 Problem Set: Expectation

1. **Prove that expectation is linear, i.e., that $E[aX + b] = aE[X] + b$ where a and b are constants.**

$$
\begin{aligned}
E[aX+b] &= \int p(x)(ax+b)dx \\
&= \int axp(x)dx + \int bp(x)dx \\
&= a\int xp(x)dx + b\int p(x)dx \\
&= aE[X] + b\int p(x)dx \\
&= aE[X] + b(1) \\
&= aE[X] + b
\end{aligned}
$$

2. **Prove that $var(cX) = c^2 var(X)$ where c is a constant.**

$$
\begin{aligned}
var(cX) &= E[(cX)^2] - [E(cX)]^2 \\
&= c^2 E[X^2] - c^2 E[X]^2 \\
&= c^2(E[X^2] - E[X]^2) \\
&= c^2 var(X)
\end{aligned}
$$

3. **Prove that $var(X) = E[X^2] - (E[X])^2$.**

$$
\begin{aligned}
var(X) &= \int p(x)(x-\mu_x)^2 dx \\
&= \int p(x)(x^2 - 2x\mu_x + \mu_x^2)dx \\
&= \int p(x)(x^2)dx - \int p(x)2x\mu_x dx + \int p(x)\mu_x^2 dx \\
&= E[X^2] - 2\mu_x \int p(x)xdx + \mu_x^2 \int p(x)dx \\
&= E[X^2] - 2\mu_x \cdot \mu_x + \mu_x^2(1) \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}
$$

6.1 References

brilliant.org
I used Symbolab to check my math.
I snooped around the discord for clues.
I got help from Michael Rogers who took ICS 435 2021.

7 Problem Set: MLE of Gaussian

1. **First, show that the data likelihood p(x1, x2, ..., xN) is the product of the individual likelihoods.**

$$
p_{\mu,\sigma}(\mathcal{D}) = \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-(X_n-\mu)^2/2\sigma^2}
$$

2. **Take the logarithm of the data likelihood. Explain why finding the x that maximizes the log-likelihood also maximizes the original likelihood.**

$$\begin{aligned}\ln p_{\mu,\sigma}(\mathcal{D}) &= \ln\left(\prod_{n=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}e^{-(X_n-\mu)^2/2\sigma^2}\right)\\ &= \sum_{n=1}^{N}\ln\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-(X_n-\mu)^2/2\sigma^2}\right)\end{aligned}$$

3. **Rearrange the terms so that the terms that contain x are separated from the terms that do not.**

$$\begin{aligned}\sum_{n=1}^{N}\ln\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-(X_n-\mu)^2/2\sigma^2}\right) &= \sum_{n=1}^{N}\left[\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{(X_n-\mu)^2}{2\sigma^2}\right]\\ &= \sum_{n=1}^{N}\left[-\ln(\sigma\sqrt{2\pi})-\frac{(X_n-\mu)^2}{2\sigma^2}\right]\\ &= -N\ln(\sigma\sqrt{2\pi})-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(X_n-\mu)^2\end{aligned}$$

4. **Take the derivative of the log-likelihood with respect to μ to find the where the derivative is 0. Confirm that this is a maximum by computing the sign of the double derivative.**

$$\begin{aligned}\frac{\partial}{\partial\mu}\ln p_{\mu,\sigma}(\mathcal{D}) &= 0+\frac{\partial}{\partial\mu}\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(X_n-\mu)^2\right)\\ &= \sum_{n=1}^{N}\frac{\partial}{\partial\mu}\left(-\frac{1}{2\sigma^2}(X_n-\mu)^2\right)\\ &= \sum_{n=1}^{N}\frac{(X_n-\mu)}{\sigma^2}=0\\ &\Rightarrow\sum_{n=1}^{N}X_n=N\mu\\ &\Rightarrow\hat{\mu}_{MLE}\\ &= \frac{1}{N}\sum_{n=1}^{N}X_n\\ &= \overline{X}\end{aligned}$$

5. **Suppose someone tells you the true value of σ. Will this change your MLE of μ? Explain why this makes sense in terms of the shape of the Gaussian function.**

   If someone tells you the true value of σ, the MLE of μ will not change as the MLE of μ does not depend on σ. This makes sense in terms of the shape of the Gaussian function because it determines the highest point.

6. **Suppose someone tells you the true value of µ. Will this change your MLE of σ? Explain why this makes sense in terms of the shape of the Gaussian function.**

   If someone tells you the true value of μ, the MLE of σ will change as the MLE of σ depends on μ. The value of $\overline{X}$ is used to find the MLE of σ. This makes sense in terms of the shape of the Gaussian function because it estimates the width and the variance would change based on the location of the mean.

7.1 References

jrmeyer.github.io
Towards Data Science
statlect.com
Symbolab and WolframAlpha were of no help.
I snooped around the discord for clues.
I got help from Michael Rogers who took ICS 435 2021.

8 Problem Set: Bayesian Based Coins

1. **We begin by assuming a uniform prior distribution $\mathbf{P}(\theta) \approx \mathbf{U}(0,\ 1)$. Write the uniform prior in terms of a Beta distribution. This is useful for subsequent computations. To do this, determine the Beta parameters α and β that result in the uniform distribution being equivalent to the Beta distribution**

$$P(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1 - \theta)^{\beta-1}$$

where Γ is the gamma function that ensures the function integrates to one for any value of α, $\beta > \mathbf{0}$.

If α and $\beta > 1$:

$$\begin{aligned} P(\theta) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{0}(1 - \theta)^{0} \\ &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \\ &= \frac{1}{1} \\ &= 1 \end{aligned}$$

2. **Next, we flip the coin 10 times and observe 8 heads followed by 2 tails. Use Bayes' Rule to write an expression for the posterior in terms of the likelihood, prior, and marginal P(data).**

$$P(Bias = 0.8|Heads) = \frac{P(Heads|Bias = 0.8) \cdot P(Bias = 0.8)}{P(Heads)}$$

3. **Simplify the numerator to obtain an expression in the form of a Beta pdf. Observe that the marginal in the denominator does not depend on θ, so we can treat it as a constant factor. Since the posterior you have calculated must integrate to one over the range [0,1], multiplying all the constant factors together must constitute a normalizing factor for the pdf. Conclude that the posterior is a Beta distribution. What are Beta parameters α', β' of this posterior?**

$$P(Bias = 0.8|Heads = 1, \theta) = \binom{Bias}{Heads}\theta^{0.8}(1-\theta)^{0.2}$$

Given a Binomial $(n, k|\theta$ likelihood, and a Beta(a, b) prior on θ, the posterior will be:

$$Beta(a + k, b + n - k)$$

Or in our case:

$$Beta(a + (Bias||0.8), b + (Heads||1) - (Bias||0.8))$$

Because the normalizing factor equals 1, you don't have to do the integral.

$$10 - 1 = 9$$
$$10 - (8 - 1) = 3$$

Where α' is 9 and β' is 3.

$$\frac{(\Gamma 12)}{(\Gamma 9) \cdot (\Gamma 3)} = 495$$

8.1 References

Beta Distribution Wikipedia
stat.cmu.edu
Stack Exchange
Towards Data Science
Probabilistic World