Tutorial/Worksheet on "M0a. Probability Background"

CS5691 PRML Jul–Nov 2025

August 9, 2025

1. [Probability Basics] Do all the tutorial/exercise problems in the slides on "M0a. Probability Background". The specific slides in this slide deck include:

   (a) Slides 10-11 on conditional probability and Bayes' theorem examples.

   Solution:
   Slide 10:
   $[Pr(Y=1|X=9), Pr(Y=2|X=9)]^T \propto [0\ 2]^T$. Therefore,
   $[Pr(Y=1|X=9), Pr(Y=2|X=9)]^T = [\frac{0}{2}\ \frac{2}{2}]^T = [0\ 1]^T$.
   We can calculate $Pr(Y|X=5)$ as above, or equivalently as:
   $Pr(Y=1|X=5) = \frac{Pr(Y=1,X=5)}{Pr(X=5)} = \frac{(5/N)}{(9/N)} = \frac{5}{9}$, and
   $Pr(Y=2|X=5) = 1 - Pr(Y=1|X=5) = \frac{4}{9}$.
   Slide 11:

   X

Y	R	G	B	
S	20	10	15	45
C	10	20	25	55
	30	30	40	

   (a) $P(X)$:

   $P(X=R) = \frac{30}{100} = 0.3$

   $P(X=G) = \frac{30}{100} = 0.3$

   $P(X=B) = \frac{40}{100} = 0.4$

   $P(Y)$:

   $P(Y=S) = \frac{45}{100} = 0.45$

   $P(Y=C) = \frac{55}{100} = 0.55$

   (b) $P(X/Y=C)$:

   $P(X=R/Y=C) = \frac{10}{55} = \frac{2}{11}$

   $P(X=G/Y=C) = \frac{20}{55} = \frac{4}{11}$

   $P(X=B/Y=C) = \frac{25}{55} = \frac{5}{11}$

   $P(Y/X=B)$

   $P(Y=S/X=B) = \frac{15}{40} = \frac{3}{8}$

   $P(Y=C/X=B) = \frac{25}{40} = \frac{5}{8}$

   (c) $P(Y=C/X=B) = \frac{P(X=B/Y=C)\ P(Y=C)}{P(X=B)}$

   $= \frac{\frac{25}{55} \cdot \frac{55}{100}}{\frac{40}{100}}$

   $= \frac{25}{40} = \frac{5}{8}$

   [Source: From [CC]Notes (by Prof. Chandra)]

   (b) Slide 19 on conditional distribution and conditional expectation questions.

   Solution: Discussed extensively in class. So explicit solutions not provided here.

   (c) Slides 32 and 35 on questions related to pmf (probability mass function) and pdf (probability density function) respectively.

Solution:
Slide 32:
(i) If $X \sim$ Bernoulli(θ), then $E[X] = \sum_{x=0,1} x\, Pr(X = x) = 0\, Pr(X = 0) + 1\, Pr(X = 1) = Pr(X = 1) = \theta$.
(ii) A Binomial-distributed rv can be expressed as a sum of n iid (independent and identically distributed) Bernoulli rvs as follows. For $i \in \{1, 2, \ldots, n\}$, let $X_i \sim$ Bernoulli(p). Then $X = \sum_{i=1}^{n} X_i$ follows a Binomial(n, p) distribution. So,
$E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$.
We have used linearity of expectation in the second step, and expectation of a Bernoulli rv from above in the third step.
(iii) Answer is $1/\theta$. You can compute mean of Geometric rv X either directly using the expectation formula, or using the conditional expectation property. For the latter, let $Y = 1$ if first toss is Heads, and 0 otherwise (ow). Then, use the Law of Total Expectation, i.e., compute $E[E[X|Y]]$ to obtain $E[X]$. Specifically,
$E[X] = E[E[X|Y]] = \theta\, E[X|Y = 1] + (1 - \theta)\, E[X|Y = 0] = \theta \cdot 1 + (1 - \theta) \cdot (1 + E[X])$
$\implies E[X] = 1 + (1 - \theta)E[X] \implies E[X] = 1/\theta$.

Slide 35: Mean and variance of uniform distribution can be derived from the formula directly.

(d) Slide 43 on covariance between two rvs (random variables) X, Y whose joint distribution is given.

Solution: In this example, $X = 3 - Y$, and so they are negatively related and the covariance is negative also then.

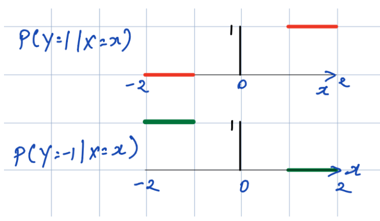
2. [Deriving Posteriors]

(a) Consider a continuous random variable X and a discrete random variable Y. Let

- $P_Y(Y = 1) = 0.5$ and $P_Y(Y = -1) = 0.5$, and
- $(X|Y = 1) \sim \text{Unif}(1, 2)$ and $(X|Y = -1) \sim \text{Unif}(-2, -1)$.

Draw the plots for $P(Y = 1|X = x)$ and $P(Y = -1|X = x)$ given the above assumptions.

Solution:



[Source: From [HG]Notes (by Harish Guruprasad)]

(b) Consider the following setting:

- $P_Y(Y = 1) = 0.7$ and $P_Y(Y = -1) = 0.3$
- $(X|Y = 1) \sim Unif(-1, 3)$ and $(X|Y = -1) \sim Unif(-2, 0)$

1. Derive $P(Y = 1|X = x)$ for different possible values of x.
2. Plot the derived $P(Y = 1|X = x)$.

Solution:

$$\eta(x) = P(Y=1 \mid X=x) = \frac{f_{X\mid Y}(x\mid 1)\cdot 0.7}{0.7\, f_{X\mid Y}(x\mid 1) + 0.3\, f_{X\mid Y}(x\mid -1)}$$

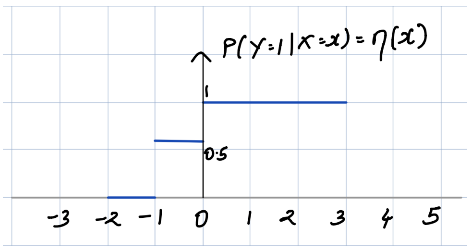
Case 1 : $x < -2$ $\Rightarrow \eta(x) = \frac{0}{0}$ = Doesn't matter

Case 2 : $x \in [-2,-1)$ $\Rightarrow \eta(x) = \frac{0}{0+\frac{1}{2}(0.3)} = 0$

Case 3 : $x \in [-1,0)$ $\Rightarrow \eta(x) = \frac{\frac{1}{4}\cdot(0.7)}{\frac{1}{4}(0.7)+\frac{1}{2}(0.3)} =$

Case 4 : $x \in [0,3)$ $\Rightarrow \eta(x) = \frac{\frac{1}{4}(0.7)}{\frac{1}{4}(0.7)} = 1$

Case 5 : $x \geq 3$ $\Rightarrow \eta(x) = \frac{0}{0}$ = Doesn't matter



[Source: From [HG]Notes (by Harish Guruprasad)]

3. [Bayes' Theorem on Events] It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

Solution:

Define events

$$A = \text{event that an email is detected as spam,}$$
$$B = \text{event that an email is spam,}$$
$$B^c = \text{event that an email is not spam.}$$

We know $P(B) = P(B^c) = .5, \quad P(A \mid B) = 0.99, \quad P(A \mid B^c) = 0.05.$
Hence by the Bayes's formula we have

$$\begin{aligned} P(B^c \mid A) &= \frac{P(A \mid B^c)P(B^c)}{P(A \mid B)P(B) + P(A|B^c)P(B^c)} \\ &= \frac{0.05 \times 0.5}{0.05 \times 0.5 + 0.99 \times 0.5} \\ &= \frac{5}{104}. \end{aligned}$$

[Source: https://www.studocu.com/row/document/united-international-university/data-structure-and-algorithm-ii-lab/ai-final-spring-2021/31515566]

4. [Random Variable Basics] A discrete random variable X is said to follow/have a Poisson distribution with parameter $\lambda > 0$ over the support $S = \{0, 1, 2, \dots\}$ if it has the following pmf:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

In this question, you are required to verify if the above pmf of X is indeed a valid pmf by verifying the following properties:

1. $P(X = x) \geq 0 \;\; \forall x \in S$, and
2. $\sum_{x \in S} P(X = x) = 1$.

In addition, derive $\mathbb{E}(X)$ and $Var(X)$.

[Hint: Use the Maclaurin's series of exponential function: $e^y = \sum_{n=0}^{\infty} \frac{y^n}{n!}$]

Solution:
Verifying that for probability mass function, $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$, follow the following properties:

$1) P(X = x) \geq 0 \;\; \forall x \in S$

This statement says that for every element x in the support S, all the probabilities must be positive.

Proof:
Given parameter $\lambda > 0$
$\implies \lambda^x > 0$ As any power of positive number is positive
As $x \in S$ and $S = \{0, 1, 2, \dots\}$ So, x ≥ 0
$\implies x! \geq 0$
As we know that any e is a constant with a positive value 2.71828.
$\implies e^{-\lambda} > 0$ As any power of positive number is positive

$\implies P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \geq 0$ As multiplication and division of 2 positive numbers is positive
Hence, $P(X = x) \geq 0 \ \ \forall x \in S$

2)$\sum_{x \in S} P(X = x) = 1$

This statement says that if we add up all the probabilities for all the possible values of x, in the support S, then that sum equals 1.

Proof:
Given $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$

$\implies \sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!}$

$\implies e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$

Since We know that, $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$

$\implies \sum_{x=0}^{\infty} P(X = x) = e^{-\lambda} e^{\lambda}$

$\implies \sum_{x=0}^{\infty} P(X = x) = 1$
Hence proved.

Calculating Expectation,

E(X) $= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}$

$= \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}$

$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{x!}$

$= \lambda e^{-\lambda} \sum_{x-1=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$

As, $\sum_{n=0}^{\infty} \frac{y^n}{n!} = e^y$

Thus, E(X) $= \lambda e^{-\lambda} e^{\lambda} = \lambda$

Calculating variance
Var(X) $= \sigma^2 = E(X^2) - (E(X))^2$

$E(X^2) = \sum_{x=1}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!}$

$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x^2 \frac{\lambda^{x-1}}{x!}$

$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!}$

$= \lambda e^{-\lambda} (\sum_{x=1}^{\infty} (x-1) \frac{\lambda^{x-1}}{(x-1)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!})$

$= \lambda e^{-\lambda} (\sum_{x=2}^{\infty} \frac{\lambda^{x-1}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!})$

$= \lambda e^{-\lambda} (\lambda \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!})$

Let $y = x - 1, z = x - 2$.

then, $\implies \lambda e^{-\lambda} (\lambda \sum_{z=0}^{\infty} \frac{\lambda^z}{z!} + \sum_{y=0}^{\infty} \frac{\lambda^y}{y!})$

Using the Maclaurin series of exponential function, we've:

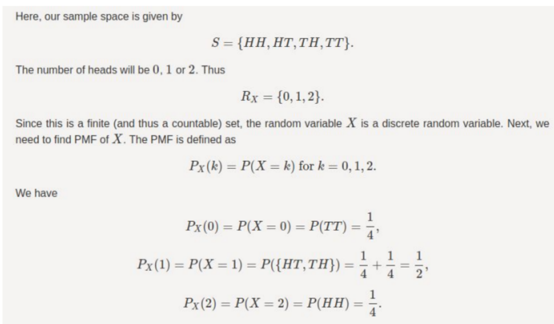
$E[X^2] = \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda})$

$= \lambda e^{-\lambda} e^{\lambda} (\lambda + 1) = \lambda^2 + \lambda$

Therefore, Var(X)= $E(X^2) - (E(X))^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$.

5. [RV Basics Contd.] A fair coin is tossed twice, and **X** is defined as the number of heads that are observed. Find the range of **X** (R_X) and the probability mass function (P_X).

Solution:



Here, our sample space is given by

$$S = \{HH, HT, TH, TT\}.$$

The number of heads will be 0, 1 or 2. Thus

$$R_X = \{0, 1, 2\}.$$

Since this is a finite (and thus a countable) set, the random variable X is a discrete random variable. Next, we need to find PMF of X. The PMF is defined as

$$P_X(k) = P(X = k) \text{ for } k = 0, 1, 2.$$

We have

$$P_X(0) = P(X = 0) = P(TT) = \frac{1}{4},$$

$$P_X(1) = P(X = 1) = P(\{HT, TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$P_X(2) = P(X = 2) = P(HH) = \frac{1}{4}.$$

Figure 1:

[Source: `https://www.probabilitycourse.com/chapter3/3_1_3_pmf.php`]

6. [Change of Variables vs. the Law of the Unconscious Statistican] Consider a continuous random variable X and a discrete random variable Y. Let

- $P_Y(Y = 1) = 0.5$ and $P_Y(Y = -1) = 0.5$, and
- $(X|Y = 1) \sim$ Unif$(1, 2)$ and $(X|Y = -1) \sim$ Unif$(-2, -1)$.

a. What is the marginal distribution of X? Specifically, plot the pdf of X denoted $f_X(x)$.

b. Write down the pdf $f_X(x)$ of X.

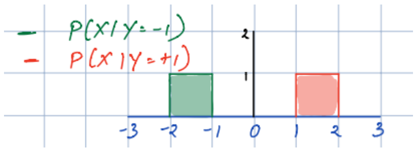
c. Let $Z = X^2$. What is the pdf of Z? Use it to compute $E[Z]$. (Hint: To obtain pdf of Z, you could simply derive the cdf of Z and differentiate it (or you could also use the change-of-variables formula).)

d. Now, use the pdf of X directly to compute $E[X^2]$ (using the law of the unconscious statistician). Does this give the same answer as the previous question? Which of these two methods do you prefer to compute the expectation of a function of a rv?

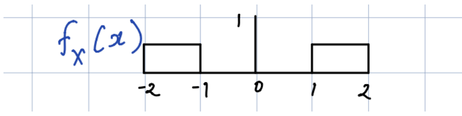
Solution:

a. Solution

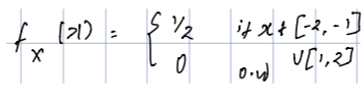
Class conditionals $f_{X|Y}(x)$:



Marginal $f_X(x)$:



b. See Figure b.



[SOURCE: Parts a,b from [HG]Notes]

c. Follow hint to obtain $f_Z(z)$; then use standard expectation formula, $E[Z] = \int_{z=-\infty}^{\infty} z f_Z(z) dz$.

The cdf of $Z = X^2$ is given by:

$$\begin{aligned}
F_Z(z) &= P(Z \leq z) \\
&= P(X^2 \leq z) \\
&= P(X \in [-\sqrt{z}, +\sqrt{z}]) \\
&= 2\ P(X \in [0, \sqrt{z}]) \\
&= \begin{cases} 2\ \frac{\sqrt{z}-1}{2} & \text{if } \sqrt{z} \in [1,2] \\ 2\ \frac{1}{2} & \text{if } \sqrt{z} > 2 \\ 0 & \text{o.w.} \end{cases} \\
&= \begin{cases} \sqrt{z} - 1 & \text{if } z \in [1,4] \\ 1 & \text{if } z > 4 \\ 0 & \text{o.w.} \end{cases}
\end{aligned}$$

Differentiating the above cdf gives:

$$f_Z(z) = F_Z'(z) \quad = \quad \begin{cases} \frac{1}{2\sqrt{z}} & \text{if } z \in [1,4] \\ 0 & \text{o.w.} \end{cases}$$

Therefore, $E[Z] = \int_{z=1}^{4} z \frac{1}{2\sqrt{z}} dz = [\frac{z^{3/2}}{3}]_{z=1}^{z=4} = (8-1)/3 \approx 2.33$.

Another way of deriving $f_Z(z)$ using change-of-variables formula: Let $Z = g(X) = X^2$. Assume $z \geq 0$ (otherwise, pdf is zero). Then, $dz = 2x\ dx \implies |dx/dz| = |1/(2x)| = 1/(2\sqrt{z})$. Since $g(.)$ is a two-to-one function, we use the following change-of-variables formula:

$$
\begin{aligned}
f_Z(z) &= f_X(x)\left|\frac{dx}{dz}\right| \text{ (at } x = -\sqrt{z}) + f_X(x)\left|\frac{dx}{dz}\right| \text{ (at } x = +\sqrt{z}) \\
&= f_X(\sqrt{z})\frac{1}{2\sqrt{z}} + f_X(-\sqrt{z})\frac{1}{2\sqrt{z}} \\
&= \begin{cases} 2\left(\frac{1}{2}\ \frac{1}{2\sqrt{z}}\right) & \text{if } \sqrt{z} \in [1,2] \\ 0 & \text{o.w.} \end{cases} \\
&= \begin{cases} \frac{1}{2\sqrt{z}} & \text{if } z \in [1,4] \\ 0 & \text{o.w.} \end{cases}
\end{aligned}
$$

d. By law of the unconscious statistician,

$$
\begin{aligned}
E[X^2] &= \int_{x=-\infty}^{\infty} f_X(x) x^2 dx \\
&= \int_{-2}^{-1} (x^2/2)dx + \int_{1}^{2} (x^2/2)dx \\
&= [x^3/6]_{-2}^{-1} + [x^3/6]_{1}^{2} \\
&= (-1 + 8 + 8 - 1)/6 \\
&= (8-1)/3 \approx 2.33
\end{aligned}
$$

The two answers match as expected. **Going forward, you can use the law of the unconscious statistician for all problems where expectation of a function of a rv is needed, as it greatly simplifies calculations as illustrated in this question.**