Densities and Expectation

Submit a PDF of your answers to Canvas.

1. *Unfair dice, convolution.*

a) Consider two independent dice $X \sim p(x)$ and $Y \sim q(y)$ where

$$p(x) = \begin{cases} p_1 & \text{if } x = 1 \\ \vdots & \\ p_6 & \text{if } x = 6 \\ 0 & \text{otherwise} \end{cases}$$

and

$$q(y) = \begin{cases} q_1 & \text{if } y = 1 \\ \vdots & \\ q_6 & \text{if } y = 6 \\ 0 & \text{otherwise.} \end{cases}$$

Find an expression for the pmf of $X + Y$. It may be helpful to specify the ways in which $X + Y = i$ for $i = 2, 3, \ldots, 12$.

b) Next consider the more general case. Let X and Y be integer valued independent random variables with pmfs given by p_X and p_Y, and define the random variable $Z = X + Y$. Show that

$$p_Z(z) = \sum_i p_X(x_i) p_Y(z - x_i).$$

SOLUTION: In general, to find $p(z)$, we have to sum the product of $p_X(x)p_Y(y)$ over all values of x, y such that $x + y = z$. One way to write this sum is:

$$p(z) = \sum_x \sum_y p_X(x) p_Y(y) \mathbb{I}\{x + y = z\}$$

where $\mathbb{I}$ is the indicator function. Note that both sums are over all integers. For a fixed x, the inner sum is only nonzero for one value of y, specifically, when $x + y = z$, or equivalently, when $y = z - x$, and the expression can be written with a single sum:

$$p(z) = \sum_x p_X(x) p_Y(z - x)$$

giving the result.

2. In homework 1, you wrote a few lines of code to find the minimum of n i.i.d. samples of a uniform random variable. Here, you will address the same problem analytically. More precisely, let

$$X_i \overset{i.i.d.}{\sim} U[0,1],$$

and define

$$Y = \min_{i=1,\dots,n} X_i.$$

a) Find an expression for the pdf of Y. Your answer should depend on n.

b) Find an expression for $E[Y]$.

c) Does this agree with the plot you made previously?

SOLUTION:

a) The crux is to note that events $\{\min_i X_i > a\}$ and $\{\cap_i X_i > a\}$ are equivalent. In words, if the minimum is greater than a, then all values $X_1, \dots$ are greater than a, and vice-versa. Hence, for $0 \leq a \leq 1$

$$\mathbb{P}(Y > a) = \mathbb{P}(X_1 > a, X_2 > a, \dots) = \prod \mathbb{P}(X_i > a) = (1-a)^n$$

Since we are interested in the pdf, we first find the cdf

$$F(y) = 1 - \mathbb{P}(Y > y) = 1 - (1-y)^n$$

and the pdf as

$$f(y) = \frac{dF}{dy} = n(1-y)^{n-1}.$$

b) The expected value can be found directly from the pdf:

$$E[Y] = \int_0^1 yn(1-y)^{n-1} = \left.\frac{-(1-y)^n(ny+1)}{n+1}\right|_0^1 = \frac{1}{n+1}$$

c) Yes.

3. *Expectation Basics.* Let $\mathbb{P}(X=1) = 1/2$, $\mathbb{P}(X=2) = 1/4$, and $\mathbb{P}(X=3) = 1/4$.

a) Compute $E[X]$.

b) Compute $E[g(X)]$ if $g(x) = x^2$.

c) The variance of an random variable is defined as $\mathrm{Var}(X) = E[(X - E[X])^2]$. Compute $\mathrm{Var}(X)$.

d) Write an expression for $E[-\log(g(X))]$ in terms of the function $g(X)$.

e) One particular function is $g(X) = \mathbb{P}(X)$, where $\mathbb{P}(X)$ is defined above. Write an expression for $E[-\log_2(\mathbb{P}(X))]$. This quantity is the *entropy* of X.

SOLUTION:

a) $E[X] = 1/2 \times 1 + 1/4 \times 2 + 1/4 \times 3 = 7/4$

b) $E[X^2] = 1/2 \times 1^2 + 1/4 \times 2^2 + 1/4 \times 3^2 = 15/4$

c) $\mathrm{var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2 = 11/16$

d) $E[-\log(g(X))] = -1/2\log(g(1)) - 1/4\log(g(2)) - 1/4\log(g(3))$

e) $E[-\log(\mathbb{P}(X))] = -1/2\log(1/2) - 1/4\log(1/4) - 1/4\log(1/4)$

4. *Bi-variate Random Variables.* The joint pmf $p(x, y)$ is defined as follows:

$$\mathbb{P}(X = 1, Y = 1) = 1/8$$
$$\mathbb{P}(X = 1, Y = 2) = 1/8$$
$$\mathbb{P}(X = 2, Y = 1) = 1/2$$
$$\mathbb{P}(X = 2, Y = 2) = 1/4$$

a) Are X and Y independent? Why or why not?

b) For any two random variables X and Y, the *covariance* is defined as $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$. Compute $\mathrm{Cov}(X, Y)$.

c) Define a new random variable $Z = X + Y$. Specify the pmf of $p(z)$.

d) Write a function (in code) that generates X and Y at random as specified by the joint PMF. Your function should take no input arguments, and should return $X, Y \in \{1, 2\}^2$ with probability specified above.

SOLUTION:

a) The marginals are $p(y = 1) = 5/8$, $p(y = 2) = 3/8$ and $p(x = 1) = 1/4$, $p(x = 2) = 3/4$, and we can see that $p(1, 1) = 1/8 \neq p(y = 1)p(x = 1)$. Also, it is clear that the distribution of y depends on the value of x (or vice-versa), which is enough to show they are not independent.

b) First, $E[X] = 7/4$ and $E[Y] = 11/8$. Then $\mathrm{Cov}(X, Y) = 1/8 \times (-3/6) + 1/8 \times (5/64) + 1/2 \times (-27/64) + 1/4 \times (81/64)$.

c) $p(z = 2) = 1/8$, $p(z = 3) = 1/8 + 1/2$, $p(z = 4) = 1/4$, otherwise zero. Note that since X, Y are not independent, it does not make sense to convolve the marginals.

d) One straight forward way is to generate a random variable $Z \sim U[0, 1)$ using `np.random.rand()`, and assign $(X, Y) = (1, 1)$ if $Z \leq 1/8$, $(X, Y) = (1, 2)$ if $1/8 < Z \leq 1/4$, $(X, Y) = (2, 1)$ if $1/4 < Z \leq 3/4$ and $(X, Y) = (2, 2)$ if $3/4 < Z \leq 1$.

5. *Total Probability.* Let X and Y be discrete random variables. Conditional expectation is defined as $E_X[X|Y] = \sum_x x\ \mathbb{P}(X = x|Y)$ for discrete random variables. Use this definition to show that

$$E[X] = E_Y[E_X[X|Y]].$$

SOLUTION: Using the definition of conditional expectation, we can write:

$$\begin{aligned}
E_Y[E_X[X|Y]] &= \sum_y p(y) \sum_x xp(x|y) \\
&= \sum_x x \sum_y p(y)p(x|y) \\
&= \sum_x xp(x) \\
&= E[X]
\end{aligned}$$

where the second to last inequality follows from the law of total probability.

6. Optional. *MAP classification and 1-d discriminant analysis.* Let $X \in \mathbb{R}$ represent a feature, and $Y = 0$ or $Y = 1$ the class label. The distribution of X depends on the label:

$$X|Y = 0 \sim \mathcal{N}(0, 1)$$

$$X|Y = 1 \sim \mathcal{N}(4, 1)$$

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian density: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Let the prior probability of the classes be $p(y = 0) = 3/4$ and $p(y = 1) = 1/4$.

a) Use a computer to create a plot with both pdfs $p(x|y = 0)$ and $p(x|y = 1)$ on the same axis.

b) Use Bayes and total probability to find an expression for the posterior $p(y|x)$.

c) Use a computer to evaluate $p(y = 0|x = 2)$ using your expression above. What is $p(y = 0|x = 2)$?

d) Recall that maximum a posteriori (MAP) classification rule predicts the label y as follows:

$$\widehat{y} = \arg\max_{y} p(y|x).$$

Use *maximum a posteriori* to design a classification rule that will predict if $Y = 0$ or $Y = 1$ given $X = x$.

e) What is the *true risk* of your MAP classifier? Use a computer to find a numerical answer.