Probability & Statistics for EECS: Homework #7 Solutions

Professor Ziyu Shao

Problem 1

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y\|X=x) = \frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x) = \frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
X **continuous**	$P(Y=y\|X=x) = \frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x) = \frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

- X discrete, Y continuous:
  According to the continuous Bayes' rule, we have

$$P(Y \in (y-\varepsilon, y+\varepsilon)|X=x) = \frac{P(X=x|Y \in (y-\varepsilon, y+\varepsilon)P(Y \in (y-\varepsilon, y+\varepsilon))}{P(X=x)}.$$

  By letting $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \to 0} P(Y \in (y-\varepsilon, y+\varepsilon)|X=x) = \lim_{\varepsilon \to 0} f_Y(y|X=x) \cdot 2\varepsilon,$$

  and

$$\lim_{\varepsilon \to 0} \frac{P(X=x|Y \in (y-\varepsilon, y+\varepsilon)P(Y \in (y-\varepsilon, y+\varepsilon))}{P(X=x)} = \lim_{\varepsilon \to 0} \frac{P(X=x|Y=y)f_Y(y) \cdot 2\varepsilon}{P(X=x)}.$$

  Therefore, we can finish the proof by canceling the term 2ε in the following equation:

$$\lim_{\varepsilon \to 0} f_Y(y|X=x) \cdot 2\varepsilon = \lim_{\varepsilon \to 0} \frac{P(X=x|Y=y)f_Y(y) \cdot 2\varepsilon}{P(X=x)}$$
$$\Rightarrow f_Y(y|X=x) = \frac{P(X=x|Y=y)f_Y(y)}{P(X=x)}.$$

- X continuous, Y discrete:

$$\begin{aligned} P(Y=y|X=x) &= \lim_{\varepsilon \to 0} P(Y=y|X \in (x-\varepsilon, x+\varepsilon)) \\ &= \lim_{\varepsilon \to 0} \frac{P(X \in (x-\varepsilon, x+\varepsilon)|Y=y)P(Y=y)}{P(X \in (x-\varepsilon, x+\varepsilon))} \\ &= \lim_{\varepsilon \to 0} \frac{2\varepsilon \cdot f_X(x|Y=y)P(Y=y)}{2\varepsilon \cdot f_X(x)} \\ &= \frac{f_X(x|Y=y)P(Y=y)}{f_X(x)} \end{aligned}$$

	Y **discrete**	Y **continuous**
X **discrete**	$P(X=x) = \sum_y P(X=x \mid Y=y)P(Y=y)$	$P(X=x) = \int_{-\infty}^{\infty} P(X=x \mid Y=y) f_Y(y)dy$
X **continuous**	$f_X(x) = \sum_y f_X(x \mid Y=y)P(Y=y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X\mid Y}(x \mid y) f_Y(y)dy$

- X discrete, Y continuous:

$$P(X=x \mid Y \in (y-\varepsilon, y+\varepsilon)) = \frac{P(Y \in (y-\varepsilon, y+\varepsilon) \mid X=x)P(X=x)}{P(Y \in (y-\varepsilon, y+\varepsilon))}.$$

By letting $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \to 0} P(X=x \mid Y \in (y-\varepsilon, y+\varepsilon)) = P(X=x \mid Y=y),$$

and

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \frac{P(Y \in (y-\varepsilon, y+\varepsilon) \mid X=x)P(X=x)}{P(Y \in (y-\varepsilon, y+\varepsilon))} \\
=&\lim_{\varepsilon \to 0} \frac{f_Y(y \mid X=x) \cdot 2\varepsilon \cdot P(X=x)}{f_Y(y) \cdot 2\varepsilon} \\
=&\frac{f_Y(y \mid X=x)P(X=x)}{f_Y(y)}.
\end{aligned}$$

By combining the two equations, we can get

$$\begin{aligned}
P(X=x \mid Y=y) &= \frac{f_Y(y \mid X=x)P(X=x)}{f_Y(y)} \\
\Rightarrow P(X=x \mid Y=y) f_Y(y) &= f_Y(y \mid X=x)P(X=x).
\end{aligned}$$

By integrating on both sides of the equation with respect y, we can get

$$\begin{aligned}
\int_{-\infty}^{\infty} P(X=x \mid Y=y) f_Y(y)dy &= \int_{-\infty}^{\infty} f_Y(y \mid X=x)P(X=x)dy \\
&= P(X=x) \int_{-\infty}^{\infty} f_Y(y \mid X=x)dy \\
&= P(X=x).
\end{aligned}$$

- X continuous, Y discrete:

$$P(X \in (x-\varepsilon, x+\varepsilon)) = \sum_y P(X \in (x-\varepsilon, x+\varepsilon) \mid Y=y)P(Y=y).$$

By letting $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \to 0} P(X \in (x-\varepsilon, x+\varepsilon)) = \lim_{\varepsilon \to 0} f_X(x) \cdot 2\varepsilon,$$

and

$$\lim_{\varepsilon \to 0} \sum_y P(X \in (x-\varepsilon, x+\varepsilon) \mid Y=y)P(Y=y) = \lim_{\varepsilon \to 0} \sum_y f_X(x \mid Y=y) \cdot 2\varepsilon \cdot P(Y=y).$$

By combining the two equations, we can get

$$\begin{aligned}
\lim_{\varepsilon \to 0} f_X(x) \cdot 2\varepsilon &= \lim_{\varepsilon \to 0} \sum_y f_X(x \mid Y=y) \cdot 2\varepsilon \cdot P(Y=y) \\
f_X(x) &= \sum_y f_X(x \mid Y=y)P(Y=y).
\end{aligned}$$

Problem 2

A chicken lays a Pois(λ) number N of eggs. Each egg hatches a chick with probability p, independently. Let X be the number which hatch, and Y be the number which do NOT hatch.

(a) Find the joint PMF of N, X, Y. Are they independent?

(b) Find the joint PMF of N, X. Are they independent?

(c) Find the joint PMF of X, Y. Are they independent?

(d) Find the correlation between N (the number of eggs) and X (the number of eggs which hatch). Simplify; your final answer should work out to a simple function of p (the λ should cancel out).

Solution

Using the chicken-egg story, we can obtain that X is distributed Pois($p\lambda$) and Y similarly Pois($q\lambda$) with $q = 1 - p$ and that these random variables are independent!

(a) For non-negative integer i, j, n, if $i + j \neq n, P(X = i, Y = j, N = n) = 0$.
If $i + j = n$, then

$$P(X = i, Y = j, N = n) = P(X = i, Y = j \mid N = n)P(N = n) = \binom{n}{i} p^i q^{n-i} \cdot \frac{\lambda^n}{n!} e^{-\lambda}.$$

X, Y, and N are not independent because we have that for $i, j, n > 0$ such that $i + j \neq n$

$$P(X = i, Y = j, N = n) = 0.$$

But obviously we have that

$$P(X = i)P(Y = j)P(N = n) > 0.$$

(b) For $n \geq i \geq 0$,

$$P(X = i, N = n) = P(X = i \mid N = n)P(N = n) = \binom{n}{i} p^i q^{n-i} \cdot \frac{\lambda^n}{n!} e^{-\lambda}, n \geq i \geq 0.$$

Otherwise, $P(X = i, N = n) = 0$.

X and N are not independent since from the story, $X \sim \text{Pois}(p\lambda)$, then we have

$$P(X = i)P(N = n) = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \cdot \frac{\lambda^n}{n!} e^{-\lambda}$$

which is obviously not equal to the joint PMF. (We could also see this by observing that for $i > n$ we have that $P(X = i, N = n) = 0$.)

(c) As we know from the chicken-egg story, we have that X and Y are independent, so the joint distribution is

$$P(X = i, Y = j) = P(X = i)P(Y = j) = \frac{(\lambda p)^i}{i!} e^{-\lambda p} \frac{(\lambda q)^j}{j!} e^{-\lambda q}, i, j \geq 0.$$

(d) By the property of covariance,

$$\text{Cov}(N, X) = \text{Cov}(X + Y, X) = \text{Cov}(X, X) + \text{Cov}(X, Y) = \text{Var}(X) = \lambda p$$

Since $N \sim \text{Pois}(\lambda), \text{Var}(N) = \lambda$, we have

$$\text{Corr}(N, X) = \frac{\text{Cov}(N, X)}{\sqrt{\text{Var}(N)\text{Var}(X)}} = \frac{\lambda p}{\sqrt{\lambda \cdot \lambda p}} = \sqrt{p}.$$

Problem 3

Let X and Y be i.i.d. Expo(λ), and $T = X + Y$.

(a) Find the conditional CDF of T given $X = x$. Be sure to specify where it is zero.

(b) Find the conditional PDF $f_{T|X}(t \mid x)$, and verify that it is a valid PDF.

(c) Find the conditional PDF $f_{X|T}(x \mid t)$, and verify that it is a valid PDF.

Hint: This can be done using Bayes' rule without having to know the marginal PDF of T, by recognizing what the conditional PDF is up to a normalizing constant-then the normalizing constant must be whatever is needed to make the conditional PDF valid.

(d) In Example 8.2.4, we will show that the marginal PDF of T is $f_T(t) = \lambda^2 t e^{-\lambda t}$, for $t > 0$. Give a short alternative proof of this fact, based on the previous parts and Bayes' rule.

Solution

(a)

$$F_{T|X}(t|x) = P(T \leq t|X = x) = P(X + Y \leq t|X = x) = P(Y \leq t - x) = \left(1 - e^{-\lambda(t-x)}\right) \cdot \chi_{t \geq x}.$$

P.S., view $\chi_{\{\cdot\}}$ as the indicator function $\mathbb{I}\{\cdot\}$.

(b) Take derivative from $F_{T|X}$ respective to t.

$$f_{T|X}(t|x) = \frac{\partial}{\partial t} F_{T|X}(t|x) = \frac{\partial}{\partial t}\left[\left(1 - e^{-\lambda(t-x)}\right) \cdot \chi_{t \geq x}\right] = \lambda e^{-\lambda(t-x)} \cdot \chi_{t \geq x}.$$

- Non-negativity:

$$f_{T|X}(t|x) = \lambda e^{-\lambda(t-x)} \cdot \chi_{t \geq x} = \begin{cases} 0 \geq 0, & t < x \\ \lambda e^{-\lambda(t-x)} \geq 0, & t \geq x \end{cases}$$

- Integrates to 1:

$$\int_{-\infty}^{\infty} f_{T|X}(t|x)dt = \int_{x}^{\infty} \lambda e^{-\lambda(t-x)}dt = -\, e^{-\lambda(t-x)}\Big|_{t=x}^{t=\infty} = 1$$

Therefore, $f_{T|X}(t|x)$ is valid PDF.

(c)

$$\begin{aligned} f_{X|T}(x|t) =& \frac{f_{T|X}(t|x) f_X(x)}{f_T(t)} \\ =& \frac{1}{f_T(t)} \lambda e^{-\lambda(t-x)} \cdot \lambda e^{-\lambda x} \cdot \chi_{t \geq x} \\ =& \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t \geq x} \end{aligned}$$

- Non-negativity:

$$f_{X|T}(x|t) = \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t \geq x} = \begin{cases} 0 \geq 0, & t < x \\ \dfrac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \geq 0, & t \geq x \end{cases}$$

- Integrates to 1: Note that $f_{X|T}(x|t) = \frac{1}{f_T(t)} \lambda^2 e^{-\lambda t} \cdot \chi_{t \geq x}$ is constant with respect to x. In particular, $f_{X|T}(x|t)$ is a non-zero constant respect to x over support $(0, t)$ and zero otherwise. By definition of Uniform distribution, we have $X|T = t \sim \text{Unif}(0, t)$, hence a valid PDF $f_{X|T}(x|t)$ over support $(0, t)$.

(d) Recall in part (c) that $X|T = t \sim \text{Unif}(0, t)$, hence $f_{X|T}(x|t) = \frac{1}{t} \cdot \chi_{t \geq x}$. Therefore, we have

$$f_T(t) = \frac{f_{T|X}(t|x) f_X(x)}{f_{X|T}(x|t)} = \frac{\lambda e^{-\lambda(t-x)} \cdot \chi_{t \geq x} \cdot \lambda e^{-\lambda x}}{\frac{1}{t} \cdot \chi_{t \geq x}} = \lambda^2 t e^{-\lambda t}, t > 0.$$

Another "solution" to (c) and (d)

Using Bayes' rule we have that

$$f_{X|T}(x \mid t) = \frac{f(x,t)}{f_T(t)} = \frac{f_{T|X}(t|x) f_X(x)}{f_T(t)} = \alpha f_{T|X}(t \mid x) f_X(x) = \alpha \lambda e^{-\lambda(t-x)} \lambda e^{-\lambda x} \cdot \chi_{t \geq x} = \alpha \lambda^2 e^{-\lambda t} \cdot \chi_{t \geq x}$$

for some $\alpha > 0$. Observe that $f_{X|T}(x \mid t)$ is a constant function respective to x. In order to be a valid PDF, $f_{X|T}(x \mid t)$ has to satisfy following

$$1 = \int_{\mathbb{R}} f_{X|T}(x \mid t) dx = \int_0^t \alpha \lambda^2 e^{-\lambda t} dx = t \alpha \lambda^2 e^{-\lambda t}$$

So, for every $t > 0$ there has to be

$$\alpha = \frac{1}{t \lambda^2 e^{-\lambda t}}$$

and in this case it is a valid PDF.

Observe that in part (c) we have that in fact $f_T(t) = \frac{1}{\alpha}$. So, we can easily obtain that

$$f_T(t) = \lambda^2 t e^{-\lambda t}.$$

Problem 4

Let U_1, U_2, U_3 be i.i.d. Unif$(0,1)$, and let $L = \min(U_1, U_2, U_3)$, $M = \max(U_1, U_2, U_3)$.

(a) Find the marginal CDF and marginal PDF of M, and the joint CDF and joint PDF of L, M.
Hint: For the latter, start by considering $P(L \geq l, M \leq m)$.

(b) Find the conditional PDF of M given L.

Solution

(a) The event $M \leq m$ is the same as the event that all 3 of the U_j are at most m, so the CDF of M is $F_M(m) = m^3$ and the PDF is $f_M(m) = 3m^2$, for $0 \leq m \leq 1$. The event $L \geq l$, $M \leq m$ is the same as the event that all 3 of the U_j are between l and m (inclusive), so

$$P(L \geq l, M \leq m) = (m-l)^3$$

for $m \geq l$ with $m, l \in [0,1]$. By the axioms of probability, we have

$$P(M \leq m) = P(L \leq l, M \leq m) + P(L > l, M \leq m)$$

So the joint CDF is

$$P(L \leq l, M \leq m) = m^3 - (m-l)^3,$$

for $m \geq l$ with $m, l \in [0,1]$. The joint PDF is obtained by differentiating this with respect to l and then with respect to m (or vice versa):

$$f(l,m) = 6(m-l),$$

for $m \geq l$ with $m, l \in [0,1]$. As a check, note that getting the marginal PDF of M by finding $\int_0^m f(l,m)dl$ does recover the PDF of M (the limits of integration are from 0 to m since the min can't be more than the max).

(b) The marginal PDF of L is $f_L(l) = 3(1-l)^2$ for $0 \leq l \leq 1$ since $P(L > l) = P(U_1 > l, U_2 > l, U_3 > l) = (1-l)^3$ (alternatively, use the PDF of M together with the symmetry that $1 - U_j$ has the same distribution as U_j, or integrate out m in the joint PDF of L, M). So the conditional PDF of M given L is

$$f_{m|L}(m|l) = \frac{f(l,m)}{f_L(l)} = \frac{2(m-1)}{(1-l)^2},$$

for all $m, l \in [0,1]$ with $m \geq l$.

Problem 5

This problem explores a visual interpretation of covariance. Data are collected for $n \geq 2$ individuals, where for each individual two variables are measured (e.g., height and weight). Assume independence across individuals (e.g., person l's variables gives no information about the other people), but not within individuals (e.g., a person's height and weight may be correlated).

Let $\left(x_1, y_1\right), \ldots, \left(x_n, y_n\right)$ be the n data points. The data are considered here as fixed, known numbers-they are the observed values after performing an experiment. Imagine plotting all the points (x_i, y_i) in the plane, and drawing the rectangle determined by each pair of points. For example, the points $(1,3)$ and $(4,6)$ determine the rectangle with vertices $(1,3),(1,6),(4,6),(4,3)$.

The signed area contributed by (x_i, y_i) and (x_j, y_j) is the area of the rectangle they determine if the slope of the line between them is positive, and is the negative of the area of the rectangle they determine if the slope of the line between them is negative. (Define the signed area to be 0 if $x_i = x_j$ or $y_i = y_j$, since then the rectangle is degenerate.) So the signed area is positive if a higher x value goes with a higher y value for the pair of points, and negative otherwise. Assume that the x_i are all distinct and the y_i are all distinct.

(a) The sample covariance of the data is defined to be

$$r = \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

are the sample means. (There are differing conventions about whether to divide by $n-1$ or n in the definition of sample covariance, but that need not concern us for this problem.)

Let (X, Y) be one of the (x_i, y_i) pairs, chosen uniformly at random. Determine precisely how $\operatorname{Cov}(X, Y)$ is related to the sample covariance.

(b) Let (X, Y) be as in (a), and $(\tilde{X}, \tilde{Y})$ be an independent draw from the same distribution. That is, (X, Y) and $(\bar{X}, \tilde{Y})$ are randomly chosen from the n points, independently (so it is possible for the same point to be chosen twice).

Express the total signed area of the rectangles as a constant times $E((X-\bar{X})(Y-\tilde{Y}))$. Then show that the sample covariance of the data is a constant times the total signed area of the rectangles.

Hint: Consider $E((X-\tilde{X})(Y-\tilde{Y}))$ in two ways: as the average signed area of the random rectangle formed by (X, Y) and $(\bar{X}, \bar{Y})$, and using properties of expectation to relate it to $\operatorname{Cov}(X, Y)$. For the former, consider the n^2 possibilities for which point (X, Y) is and which point $(\tilde{X}, \bar{Y})$; note that n such choices result in degenerate rectangles.

(c) Based on the interpretation from (b), give intuitive explanations of why for any r.v.s W_1, W_2, W_3 and constants a_1, a_2, covariance has the following properties:

(i) $\operatorname{Cov}\left(W_1, W_2\right) = \operatorname{Cov}\left(W_2, W_1\right)$;

(ii) $\operatorname{Cov}\left(a_1 W_1, a_2 W_2\right) = a_1 a_2 \operatorname{Cov}\left(W_1, W_2\right)$;

(iii) $\operatorname{Cov}\left(W_1 + a_1, W_2 + a_2\right) = \operatorname{Cov}\left(W_1, W_2\right)$;

(iv) $\operatorname{Cov}\left(W_1, W_2 + W_3\right) = \operatorname{Cov}\left(W_1, W_2\right) + \operatorname{Cov}\left(W_1, W_3\right)$.

Solution

(a) Since (X, Y) is chosen uniformly at random, we have

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}; \quad \mathrm{E}(Y) = \sum_{i=1}^{n} y_i P(Y = y_i) = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}.$$

By definition, we know

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))] \\
&= \sum_{i=1}^{n} (x_i - \mathrm{E}(X))(y_i - \mathrm{E}(Y)) P(X = x_i, Y = y_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = r.
\end{aligned}$$

Thus we prove that $\mathrm{Cov}(X, Y)$ equals the sample covariance r.

(b)
- Denote the total signed area of the rectangles as S, then

  $$S = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j).$$

  Since (X, Y) and $(\tilde{X}, \tilde{Y})$ are independent, we have

  $$\begin{aligned}
  \mathrm{E}((X - \tilde{X})(Y - \tilde{Y})) &= \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j) P(X = x_i, Y = y_i) P(\tilde{X} = x_j, \tilde{Y} = y_j) \\
  &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)(y_i - y_j) = \frac{S}{n^2}.
  \end{aligned}$$

  Thusly we have $S = n^2 \mathrm{E}((X - \tilde{X})(Y - \tilde{Y}))$.
- By the properties of expectation and considering that (X, Y) and $(\tilde{X}, \tilde{Y})$ are identically and independently sampled, we have

  $$\begin{aligned}
  \mathrm{E}((X - \tilde{X})(Y - \tilde{Y})) &= \mathrm{E}(XY) - \mathrm{E}(\tilde{X}Y) - \mathrm{E}(X\tilde{Y}) + \mathrm{E}(\tilde{X}\tilde{Y}) \\
  &= \mathrm{E}(XY) - \mathrm{E}(\tilde{X})\mathrm{E}(Y) - \mathrm{E}(X)\mathrm{E}(\tilde{Y}) + \mathrm{E}(\tilde{X}\tilde{Y}) \\
  &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(\tilde{X})\mathrm{E}(\tilde{Y}) + \mathrm{E}(\tilde{X}\tilde{Y}) \\
  &= 2[\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)] \\
  &= 2\,\mathrm{Cov}(X, Y) = 2r.
  \end{aligned}$$

  Thusly we have $r = \frac{S}{2n^2}$.

(c)
- The claim (i) is true because it doesn't matter what is the base and what is the height of the rectangle, we can switch them.
- The claim (ii) is true because rescaling the one coordinate by the factor c yields that the total area of the rectangle rescales for c.
- The claim (iii) is true since the area of the rectangle is invariant on linear translation.
- The claim (iv) is true because the distributive property of the area: it doesn't matter if we calculate two areas with the same base and then sum them or first we add heights and then calculate the total area.

Problem 6

We use the notation $X \perp\!\!\!\perp Y \mid Z$ to represent the statement: random variables X and Y are conditionally independent given random variable Z. Now given any four continuous random variables X, Y, Z, W, show the following properties of conditional independence:

1. Symmetry:
$$X \perp\!\!\!\perp Y \mid Z \iff Y \perp\!\!\!\perp X \mid Z.$$
2. Decomposition:
$$X \perp\!\!\!\perp (Y, W) \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z.$$
3. Weak Union:
$$X \perp\!\!\!\perp (Y, W) \mid Z \Rightarrow X \perp\!\!\!\perp (Y, W) \mid (Z, W).$$
4. Contraction:
$$X \perp\!\!\!\perp Y \mid Z \ \&\ X \perp\!\!\!\perp W \mid (Y, Z) \iff X \perp\!\!\!\perp (Y, W) \mid Z.$$
5. Intersection: For any positive joint PDF of X, Y, Z, W,
$$X \perp\!\!\!\perp Y \mid (Z, W) \ \&\ X \perp\!\!\!\perp Z \mid (Y, W) \iff X \perp\!\!\!\perp (Y, Z) \mid W.$$

In fact, these properties are found by Judea Pearl, who won 2011 Turing Award for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning. As Judea Pearl commented: "Exploiting conditional independence to generate fast probabilistic computations is one of the main contributions CS has made to probability theory."

Solution

1. According to the problem, proving one side is enough. By definition of conditional densities, given $X \perp\!\!\!\perp Y \mid Z$, we have
$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z &\Leftrightarrow f_{XY|Z}(x, y, z) = f_{X|Z}(x, z) f_{Y|Z}(y, z). \\ &\Leftrightarrow f_{XYZ}(x, y, z) f_Z(z) = f_{XZ}(x, z) f_{YZ}(y, z). \\ &\Leftrightarrow \exists f, a, b : f(x, y, z) = a(x, z) b(y, z). \end{aligned}$$
This immediately proves the symmetry:
$$\begin{aligned} & X \perp\!\!\!\perp Y \mid Z \\ \Leftrightarrow\ & f(x, y, z) = a(x, z) b(y, z). \\ \Leftrightarrow\ & \exists a^*, b^* : f(x, y, z) = a^*(x, z) b^*(y, z). \\ \Leftrightarrow\ & Y \perp\!\!\!\perp X \mid Z \end{aligned}$$
2. Given $X \perp\!\!\!\perp (Y, W) \mid Z$, we have $f(x, y, z, w) = a(x, z) b(y, z, w)$. By definition, we have
$$\begin{aligned} f_{XYZ}(x, y, z) = & \int_w f_{XYZW}(x, y, z, w) = \int_w f(x, y, z, w) \\ = & \ a(x, z) \int_w b(y, z, w) = a^*(x, z) b^*(y, z), \end{aligned}$$
which shows $X \perp\!\!\!\perp Y \mid Z$.

3. Given $X \perp\!\!\!\perp (Y, W) \mid Z$, we have $f(x, y, z, w) = a(x, z)b(y, z, w)$. Therefore, we have

$$\begin{aligned} f(x,y,z,w) &= a(x,z)b(y,z,w) \\ &= a^*(x,z,w)b^*(y,z,w), \end{aligned}$$

where the last equality holds by defining $a^*(x, z, w) \propto a(x, z), \forall w$. Therefore, this shows $X \perp\!\!\!\perp (Y, W) \mid (Z, W)$.

4.
   - $\Rightarrow$ Given $X \perp\!\!\!\perp Y \mid Z$, we have $f_{XY|Z}(x, y, z) = f_{X|Z}(x, z)f_{Y|Z}(y, z)$; $X \perp\!\!\!\perp W \mid (Y, Z)$ means $f_{XW|YZ}(x, y, z, w) = f_{X|YZ}(x, y, z)f_{W|YZ}(y, z, w)$. Therefore, by definition, we have

$$\begin{aligned} f_{XYW|Z}(x,y,z,w) &= f_{XW|YZ}(x,y,z,w)f_{Y|Z}(y,z) \\ &= f_{X|YZ}(x,y,z)f_{W|YZ}(y,z,w)f_{Y|Z}(y,z) \\ &\overset{X \perp\!\!\!\perp Y|Z}{=} f_{X|Z}(x,z)f_{YW|Z}(y,z,w), \end{aligned}$$

   which shows $X \perp\!\!\!\perp (Y, W) \mid Z$.
   - $\Leftarrow$ Given $X \perp\!\!\!\perp (Y, W) \mid Z$, we have

$$\begin{aligned} &X \perp\!\!\!\perp (Y,W) \mid Z \overset{\text{Decomposition}}{\longrightarrow} X \perp\!\!\!\perp Y \mid Z \\ &X \perp\!\!\!\perp (Y,W) \mid Z \overset{\text{WeakUnion}}{\longrightarrow} X \perp\!\!\!\perp (Y,W) \mid (Y,Z) \overset{\text{Decomposition}}{\longrightarrow} X \perp\!\!\!\perp W \mid (Y,Z) \end{aligned}$$

5.
   - $\Rightarrow$ Given $X \perp\!\!\!\perp Y \mid (Z, W)$, we have $f(x, y, z, w) = a(x, z, w)b(y, z, w)$. Similarly, $X \perp\!\!\!\perp Z \mid (Y, W)$ means $f(x, y, z, w) = g(x, y, w)h(y, z, w)$. If $f(x, y, z, w) > 0$ for all (x, y, z, w), it follows that

$$g(x,y,w) = \frac{a(x,z,w)b(y,z,w)}{h(y,z,w)}.$$

   Since the left-hand side does not depend on z, So for fixed $z = z_0$, we have

$$g(x,y,w) = \tilde{a}(x,w)\tilde{b}(y,w).$$

   Insert this into the second expression for f to get

$$f(x,y,z,w) = \tilde{a}(x,w)\tilde{b}(y,w)h(y,z,w) = a^*(x,w)b^*(y,z,w),$$

   which shows $X \perp\!\!\!\perp (Y, Z) \mid W$.
   - $\Leftarrow$ Given $X \perp\!\!\!\perp (Y, Z) \mid W$, we have

$$\begin{aligned} &X \perp\!\!\!\perp (Y,Z) \mid W \overset{\text{WeakUnion}}{\longrightarrow} X \perp\!\!\!\perp (Y,Z) \mid (Z,W) \overset{\text{Contraction}}{\longrightarrow} X \perp\!\!\!\perp Y \mid (Z,W) \\ &X \perp\!\!\!\perp (Y,Z) \mid W \overset{\text{WeakUnion}}{\longrightarrow} X \perp\!\!\!\perp (Y,Z) \mid (Y,W) \overset{\text{Contraction}}{\longrightarrow} X \perp\!\!\!\perp Z \mid (Y,W) \end{aligned}$$