

Chapter 5

Joint Probability Distributions

Detailed Solutions

5.1 Basic Concept

5.1 Joint vs. Marginal Definitions

Problem: Definitions and Independence.

Solution:

- (a) **Marginal PDF Formula:** To find the marginal PDF of X , integrate the joint PDF over the entire range of Y :

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

- (b) **Independence Condition:** X and Y are statistically independent if and only if for all x, y :

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$$

- (c) **Covariance Concept:**

- If X, Y are independent, then $Cov(X, Y) = 0$. **(True)**
- If $Cov(X, Y) = 0$, are they independent? **(False)**
Zero covariance only implies no *linear* relationship. They could have a strong non-linear dependence (e.g., $Y = X^2$ with symmetric X).

$$(a) f_X(x) = \int f(x, y) dy$$

$$(b) f(x, y) = f_X(x) f_Y(y)$$

(c) Independence $\implies Cov = 0$. Reverse is NOT true.

5.2 Discrete Joint Distribution

Problem: Table Analysis.

Solution:

(a) **Marginal PMFs:** Sum rows for X , columns for Y .

- $P(X = 1) = 0.10 + 0.20 + 0.10 = 0.40$
- $P(X = 2) = 0.10 + 0.30 + 0.20 = 0.60$
- $P(Y = 1) = 0.10 + 0.10 = 0.20$
- $P(Y = 2) = 0.20 + 0.30 = 0.50$
- $P(Y = 3) = 0.10 + 0.20 = 0.30$

(b) **Conditional $P(Y|X = 1)$:** Focus on Row $X = 1$ (Sum = 0.40).

$$P(Y = y|X = 1) = \frac{P(X = 1, Y = y)}{P(X = 1)}$$

- $y = 1 : 0.10/0.40 = 0.25$
- $y = 2 : 0.20/0.40 = 0.50$
- $y = 3 : 0.10/0.40 = 0.25$

(c) **Independence Check:** Check any cell, e.g., $X = 1, Y = 1$.

$$P(X = 1, Y = 1) = 0.10$$

$$P(X = 1)P(Y = 1) = 0.40 \times 0.20 = 0.08$$

Since $0.10 \neq 0.08$, they are **Dependent**.

- (a) $X : [0.4, 0.6], Y : [0.2, 0.5, 0.3]$
 (b) $[0.25, 0.50, 0.25]$
 (c) Dependent ($0.10 \neq 0.08$)

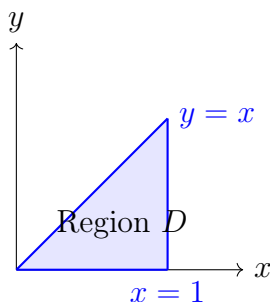
5.2 Intermediate

5.3 The Triangular Region (Exponential Density)

Problem: $f(x, y) = ke^{-(2x+3y)}$ for $0 < y < x < 1$.

Solution:

(a) **Region Drawing:** Triangle with vertices $(0, 0)$, $(1, 0)$, $(1, 1)$.



(b) **Find k:**

$$\begin{aligned}
 1 &= \int_0^1 \int_0^x ke^{-2x}e^{-3y} dy dx \\
 &= k \int_0^1 e^{-2x} \left[\frac{e^{-3y}}{-3} \right]_0^x dx \\
 &= \frac{k}{3} \int_0^1 e^{-2x} (1 - e^{-3x}) dx \\
 &= \frac{k}{3} \int_0^1 (e^{-2x} - e^{-5x}) dx \\
 &= \frac{k}{3} \left[\frac{e^{-2x}}{-2} - \frac{e^{-5x}}{-5} \right]_0^1 \\
 &= \frac{k}{3} \left[\left(-\frac{e^{-2}}{2} + \frac{e^{-5}}{5} \right) - \left(-\frac{1}{2} + \frac{1}{5} \right) \right]
 \end{aligned}$$

Let $I = (0.3 - 0.5e^{-2} + 0.2e^{-5}) \approx 0.2337$. Thus $k = 3/I$. For simplicity in later steps, let's denote the normalization factor as k .

(c) **Marginal $f_X(x)$:**

$$\begin{aligned}
 f_X(x) &= \int_0^x ke^{-2x}e^{-3y} dy = ke^{-2x} \left[\frac{1 - e^{-3y}}{3} \right]_0^x \\
 &= \frac{k}{3} (e^{-2x} - e^{-5x}) \quad \text{for } 0 < x < 1
 \end{aligned}$$

(d) **Probability $P(Y < X/2)$:** Integrate over $0 < x < 1$ and $0 < y < x/2$.

$$P = \int_0^1 \int_0^{x/2} ke^{-2x}e^{-3y} dy dx$$

- (b) $k \approx 12.83$ (approx based on numeric integral)
 (c) $f_X(x) = \frac{k}{3}(e^{-2x} - e^{-5x})$

5.4 Conditional Probability & Independence

Problem: Using PDF from 5.3.

Solution:

(a) **Conditional PDF** $f_{Y|X}(y|x)$:

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{ke^{-2x}e^{-3y}}{\frac{k}{3}e^{-2x}(1 - e^{-3x})} = \frac{3e^{-3y}}{1 - e^{-3x}}$$

Valid for $0 < y < x$.

(b) **Expectation** $E[Y|X = x]$:

$$\begin{aligned} E[Y|X] &= \int_0^x y \cdot \frac{3e^{-3y}}{1 - e^{-3x}} dy \\ &= \frac{3}{1 - e^{-3x}} \int_0^x ye^{-3y} dy \end{aligned}$$

Using integration by parts ($\int ye^{ay} = \frac{e^{ay}}{a^2}(ay - 1)$):

$$\int_0^x ye^{-3y} dy = \left[\frac{e^{-3y}}{9}(-3y - 1) \right]_0^x = \frac{1}{9}[1 - e^{-3x}(3x + 1)]$$

(c) **Independence:** The support of the joint PDF depends on both variables ($0 < y < x$). The region is triangular, not rectangular. Therefore, X and Y are **Dependent**.

- (a) $\frac{3e^{-3y}}{1 - e^{-3x}}$
 (c) Dependent (Domain is not rectangular)

5.5 Linear Combination

Problem: X, Y Independent. $W = 3X - 2Y + 5$. $E[X] = 10, V(X) = 4$. $E[Y] = 20, V(Y) = 9$.

Solution:

$$\begin{aligned} E[W] &= 3E[X] - 2E[Y] + 5 \\ &= 3(10) - 2(20) + 5 = 30 - 40 + 5 = -5 \\ Var(W) &= 3^2Var(X) + (-2)^2Var(Y) \quad (\text{Cov} = 0 \text{ due to indep}) \\ &= 9(4) + 4(9) = 36 + 36 = 72 \end{aligned}$$

Mean = -5, Variance = 72

5.6 Server Load Balancing

Problem: Discrete Joint Dist with unknown k .

Solution:

- (a) **Find k :** Sum of all probabilities = 1. Sum = $(0.10 + 0.05 + 0.15) + (0.05 + 0.20 + k) + (0.10 + 0.05 + 0.10) = 0.80 + k$.

$$0.80 + k = 1 \implies k = 0.20$$

- (b) **Marginals and $P(X > Y)$:**

- X : Row Sums $\rightarrow P(X = 0) = 0.3, P(X = 1) = 0.45, P(X = 2) = 0.25$.
- Y : Col Sums $\rightarrow P(Y = 0) = 0.25, P(Y = 1) = 0.30, P(Y = 2) = 0.45$.
- $P(X > Y)$: Sum cells where Row Index $>$ Col Index. $(1, 0) \rightarrow 0.05, (2, 0) \rightarrow 0.10, (2, 1) \rightarrow 0.05$. Sum = $0.05 + 0.10 + 0.05 = 0.20$.

- (c) $E[X|Y = 1]$: Column $Y = 1$ probabilities: 0.05, 0.20, 0.05. Sum = 0.30. Conditional PMF $P(X|Y = 1)$:

- $X = 0 : 0.05/0.30 = 1/6$
- $X = 1 : 0.20/0.30 = 4/6$
- $X = 2 : 0.05/0.30 = 1/6$

$$E = 0(1/6) + 1(4/6) + 2(1/6) = 6/6 = 1.$$

- (d) **Independence:** $Cov(X, Y) = E[XY] - E[X]E[Y]$. Calculations show $Cov \neq 0$ (e.g., $P(0, 0) = 0.1 \neq 0.3 \times 0.25 = 0.075$). Dependent.

- (a) $k = 0.20$
 (b) $P(X > Y) = 0.20$
 (c) $E[X|Y = 1] = 1$

5.7 Manufacturing Tolerance (Bivariate Normal)

Problem: $\mu_X = 100, \sigma_X = 2, \mu_Y = 50, \sigma_Y = 1, \rho = 0.8$.

Solution:

- (a) **Marginal Probability $P(X > 103)$:** $X \sim N(100, 2^2)$.

$$Z = \frac{103 - 100}{2} = 1.5$$

$$P(Z > 1.5) = 1 - 0.9332 = 0.0668$$

(b) **Expected Width given Length 104:**

$$E[Y|X = x] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$$

$$\begin{aligned} E[Y|104] &= 50 + 0.8 \left(\frac{1}{2} \right) (104 - 100) \\ &= 50 + 0.4(4) = 51.6 \text{ mm} \end{aligned}$$

(c) **Conditional Variance:**

$$\begin{aligned} \text{Var}(Y|X = x) &= \sigma_Y^2 (1 - \rho^2) \\ &= 1^2 (1 - 0.8^2) = 1 - 0.64 = 0.36 \end{aligned}$$

Yes, uncertainty reduces (Variance drops from 1 to 0.36) because knowing X gives us information about Y .

- (a) 0.0668
(b) 51.6 mm
(c) 0.36 (Reduced)

5.3 Challenge

5.8 Correlation Bounds

Problem: Prove $-1 \leq \rho \leq 1$.

Solution:

Consider the random variable $Z = \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$. Since variance is non-negative:

$$\begin{aligned} \text{Var}(Z) &\geq 0 \\ \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) + 2\text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) &\geq 0 \\ 1 + 1 + 2\rho &\geq 0 \\ 2(1 + \rho) &\geq 0 \implies \rho \geq -1 \end{aligned}$$

Similarly, using $Z = \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$, we get $2(1 - \rho) \geq 0 \implies \rho \leq 1$.

Proven.

5.9 Variance of Sum

Problem: Prove $\text{Var}(aX + bY)$ formula.

Solution:

Let $E[X] = \mu_X, E[Y] = \mu_Y$.

$$\begin{aligned} \text{Var}(aX + bY) &= E[((aX + bY) - (a\mu_X + b\mu_Y))^2] \\ &= E[(a(X - \mu_X) + b(Y - \mu_Y))^2] \\ &= E[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)] \\ &= a^2E[(X - \mu_X)^2] + b^2E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)] \\ &= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y) \end{aligned}$$

Proven.

5.4 Application

5.10 Manufacturing Tolerance Simulation

Problem: Python Simulation Analysis.

Solution:

Questions Analysis:

- (a) **Interpretation:** The correlation will be close to positive. Since $Y = 0.5X + \text{noise}$, as X increases, Y tends to increase. The value is likely around $0.7 - 0.9$ (Strong Positive).

- (b) **Analytical Correlation:** $Y = 0.5X + N$. $Cov(X, Y) = Cov(X, 0.5X + N) = 0.5Var(X)$. Since Variance is always positive, Covariance is positive, thus Correlation is positive.
- (c) **Variance Reduction:** Yes. From the formula $Var(Y) = 0.5^2Var(X) + Var(N)$. If we reduce $Var(X)$ (control the length strictly), the term $0.5^2Var(X)$ decreases, leading to a smaller total $Var(Y)$. This demonstrates how controlling one upstream parameter (X) improves the quality of a dependent downstream parameter (Y).

- (a) Strong Positive Correlation.
(b) Positive slope (0.5) links X and Y linearly.
(c) Yes, $Var(Y)$ depends on $Var(X)$.