**Journal of Education and Educational Research**
ISSN: 2957-9465 | Vol. 13, No. 3, 2025

# Understanding Mixed Probability Expressions in Engineering Courses

**Xiaolei Jiang**

School of Control and Computer Engineering, North China Electric Power University, Baoding Hebei, 071003, China

**Abstract:** The gap between basic and specialized probability knowledge for engineering students is addressed by extending core concepts. In the basic course, students typically encounter probability expressions involving either random events or continuous random variables, but not both simultaneously. However, specialized courses in fields such as communication systems and machine learning require handling both types together. The meanings and properties of mixed joint and conditional probability (density) are given, and the product rule is extended to cover these cases. Introducing these topics can help students to transition smoothly and prepares them for more complex applications.

**Keywords:** Mixed Joint Probability; Mixed Conditional Probability; Product Rule.

## 1. Introduction

In probability theory courses, students majoring in engineering encounter probability expressions involving either continuous random variables or random events, but not both simultaneously [1]. However, in specialized courses such as communication systems, pattern recognition, and machine learning, the probability expressions often include both continuous random variables and random events [2,3,4,5]. This leads to a gap between the knowledge acquired in basic courses and what is required in subsequent specialized courses. This article aims to bridge this gap by building a connecting link between them.

First, let us clarify the meaning of the symbol $p(\alpha)$. We make the following conventions:

(1) If $\alpha$ is a random event $A$, then $p(A)$ gives the probability of event $A$.

(2) If $\alpha$ is a discrete random variable $X$, then $p(x) = p(X = x)$ gives the probability of the event $X = x$.

(3) If $\alpha$ is a continuous random variable $X$, then $p(x)$ gives the probability density of the random variable $X$ at $x$.

In other words, we allow the notation $p$ to be overloaded. Depending on the argument $\cdot$, $p(\cdot)$ can represent probability, probability mass, or probability density. In the following discussion, we primarily focus on random events and continuous random variables. However, since discrete random variables can be considered as random events, the results obtained are applicable to random events, discrete random variables, and continuous random variables.

For random events $A$, $B$, and continuous random variables $X$, $Y$, we already know the meanings of $p(A,B)$, $p(x,y)$, $p(A|B)$, $p(x|y)$, where either random events or continuous random variables are present, but not both. In this article, we explore the meanings of mixed probability expressions, which involve both random events and continuous random variables, such as $p(x,B)$, $p(x|B)$, $p(A|y)$, $p(x,y,C,D)$, $p(A|y,z,D)$, $p(x,D|y,C)$. We will see that the product rule $p(AB) = p(B)\,p(A|B)$ and the marginalization formula $\int p(x,y)\,\mathrm{d}x = p(y)$ hold in their more general forms.

Unless otherwise specified, uppercase letters $A$, $B$, $C$, $D$ represent random events, uppercase letters $X$, $Y$, $Z$ represent random variables, and $x$, $y$, $z$ represent the values taken by random variables. Greek letters $\alpha$, $\beta$, $\gamma$ represent random events or random variables, or multiple random events or random variables separated by commas.

The rest of this article is organized as follows. Based on the meaning of probability density of continuous random variables, the meaning and properties of mixed joint probability (density) are presented in Section 2. Building on Section 2, Section 3 presents the meaning and properties of mixed conditional probability (density). In Section 4, the probability product rule is extended to cases involving multiple random events and continuous random variables separated by commas. We draw our conclusion in Section 5.

## 2. Mixed Joint Probability (Density)

If $\boldsymbol{\alpha}$ is a collection of random events and continuous random variables separated by commas, then $\boldsymbol{p(\alpha)}$ can be called a mixed joint probability expression. For continuous random variable $\boldsymbol{Y}$, we have

$$\boldsymbol{p(y) = \lim_{\Delta y \to 0} \frac{p(y < Y \le y + \Delta y)}{\Delta y}}.$$

By following the same pattern, the meaning of $\boldsymbol{p(A,y)}$ is given by

$$\boldsymbol{p(A,y) = \lim_{\Delta y \to 0} \frac{p(A, y < Y \le y + \Delta y)}{\Delta y}}, \qquad \textbf{(1)}$$

From the above it follows that

$$\boldsymbol{\int_{\mathcal{B}} p(A,y)\,dy = p(A, Y \in \mathcal{B})}.$$

Hence $\boldsymbol{p(A,\cdot)}$ is an unnormalized probability density. Equation (1) also leads to the following result.

$$\boldsymbol{p(\emptyset, y) = 0}$$

$$\boldsymbol{p(A_1 \cup A_2, y) = p(A_1, y) + p(A_2, y), \quad A_1 \cap A_2 = \emptyset}$$

$$\boldsymbol{p(\Omega, y) = 1}$$

So $\boldsymbol{p(\cdot, y)}$ is an unnormalized probability. Similar results hold for multiple random events and continuous random variables. For example,

$$\boldsymbol{p(x,y,C,D)} = \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{\boldsymbol{p(x < X \le x+\Delta x, y < Y \le y+\Delta y, C, D)}}{\boldsymbol{\Delta x\,\Delta y}}$$

$$\iint_{\mathcal{S}} \boldsymbol{p(x,y,C,D)\,dx\,dy = p((X,Y) \in \mathcal{S}, C, D)}$$

$$\int_{\mathcal{A}} \boldsymbol{p(x,y,C,D)\,dx = p(X \in \mathcal{A}, y, C, D)}$$

$$\int_{\mathcal{B}} \boldsymbol{p(x,y,C,D)\,dy = p(x, Y \in \mathcal{B}, C, D)}$$

$$\boldsymbol{p(x,y,\emptyset,D) = 0}$$

$$\boldsymbol{p(x,y,C_1 \cup C_2,D) = p(x,y,C_1,D) + p(x,y,C_2,D),}$$
$$\boldsymbol{C_1 \cap C_2 = \emptyset}$$

$$\boldsymbol{p(x,y,\Omega,D) = p(x,y,D)}$$

$$\boldsymbol{p(x,y,C,\emptyset) = 0}$$

$$\boldsymbol{p(x,y,C,D_1 \cup D_2) = p(x,y,C,D_1) + p(x,y,C,D_2),}$$
$$\boldsymbol{D_1 \cap D_2 = \emptyset}$$

$$\boldsymbol{p(x,y,C,\Omega) = p(x,y,C)}$$

In other words, $\boldsymbol{p(\cdot,\cdot,C,D)}$ is an unnormalized joint probability density, $\boldsymbol{p(x,y,\cdot,D)}$ and $\boldsymbol{p(x,y,C,\cdot)}$ are unnormalized probabilities. It can be seen that the marginalization formula $\int \boldsymbol{p(x,y)\,dx = p(y)}$ is a special case of the above discussion.

# 3. Mixed Conditional Probability (Density)

If $\alpha$ and $\beta$ are a collection of random events and continuous random variables separated by commas, then $p(\alpha|\beta)$ can be called a mixed conditional probability expression. The expression $p(\alpha|\beta)$ appears frequently in fields such as electronic communication and machine learning. If height and weight are represented by continuous random variables $h$ and $w$, and $M$ represents males, then $p(h|M)$ is the likelihood of the height being $h$ given that the person is male, $p(M|h)$ is the likelihood of being male given a height of $h$, $p(w|h,M)$ is the likelihood of weight $w$ given a height of $h$ and that the person is male, and $p(M|h,w)$ is the likelihood of being male given a height of $h$ and a weight of $w$. From everyday life experiences, these quantities are meaningful and should be given precise definitions.

First, let us consider the meaning of $p(\cdot\,|\beta)$, where $\cdot$ represents a random event.

(1) Let $B$ be a random event. We all know that when $p(B) \neq 0$, $p(A|B) = \frac{P(AB)}{p(B)}$. Because

$$\boldsymbol{p(\emptyset|B) = 0}$$

$$\boldsymbol{p(A_1 \cup A_2|B) = p(A_1|B) + p(A_2|B), \quad A_1 \cap A_2 = \emptyset}$$

$$\boldsymbol{p(\Omega|B) = 1}$$

$p(\cdot\,|B)$ satisfies the axioms of probability.

(2) Let $Y$ be a continuous random variable. The meaning of $p(A|y)$ is given by

$$p(A|Y=y) = \lim_{\Delta y\to 0} p(A|y < Y \le y+\Delta y)$$
$$= \lim_{\Delta y\to 0} \frac{p(A, y < Y \le y+\Delta y)}{p(y < Y \le y+\Delta y)}$$

This definition leads to

$$\boldsymbol{p(\emptyset|y) = 0}$$

$$\boldsymbol{p(A_1 \cup A_2|y) = p(A_1|y) + p(A_2|y), \quad A_1 \cap A_2 = \emptyset}$$

$$\boldsymbol{p(\Omega|y) = 1}$$

hence $p(\cdot\,|y)$ satisfies the axioms of probability.

(3) From the above, it is easy to understand that when $\beta$ is a collection of multiple random events or continuous random variables separated by commas, $p(\cdot\,|\beta)$ is also a probability. For example,

$$p(A|y,D) = \lim_{\Delta y\to 0} p(A|y < Y \le y+\Delta y, D)$$

$$p(A|y,z) = \lim_{(\Delta y,\Delta z)\to(0,0)} p(A|y < Y \le y+\Delta y, z < Z \le z+\Delta z)$$

From these, we can obtain new probabilities $p(\cdot\,|y,D)$ and $p(\cdot\,|y,z)$, respectively.

In summary, when $\beta$ contains only random events, $p(A|\beta)$ is the conditional probability of random events, and when $\beta$ contains continuous random variables, the meaning of $p(A|\beta)$ can also be given by taking limits. Since $p(\cdot\,|\beta)$ defined in this way satisfies the axioms of probability, it is a new probability derived from $p(\cdot)$.

Now, the meaning of $p(\alpha|\beta)$ can be clearly stated. Since $p(\cdot\,|\beta)$ is a probability, the meaning of $p(\alpha|\beta)$ can be obtained from the meaning of $p(\alpha)$ by simply replacing one probability $p(\cdot)$ with another probability $p(\cdot\,|\beta)$. For example,

$$p(y,D|x,C) = \lim_{\Delta y\to 0} \frac{p(y < Y \le y+\Delta y, D|x,C)}{\Delta y}$$

$$= \lim_{\Delta y\to 0} \frac{1}{\Delta y} \lim_{\Delta x\to 0} p(y < Y \le y+\Delta y, D|x < X \le x+\Delta x, C)$$

$$= \lim_{\Delta y\to 0} \frac{1}{\Delta y} \lim_{\Delta x\to 0} \frac{p(y < Y \le y+\Delta y, D, x < X \le x+\Delta x, C)}{p(x < X \le x+\Delta x, C)}$$

The difference between $p(\alpha|\beta)$ and $p(\alpha)$ lies solely in the fundamental probability, and the properties of $p(\alpha)$ outlined in Section 2 are also applicable to $p(\alpha|\beta)$. For example,

$$\iint_{\mathcal{S}} \boldsymbol{p(x,y,C,D|z,B)\,dx\,dy = p((X,Y) \in \mathcal{S}, C, D|z,B)}$$

$$\int_{\mathcal{A}} \boldsymbol{p(x,y,C,D|z,B)\,dx = p(X \in \mathcal{A}, y, C, D|z,B)}$$

$$\int_{\mathcal{B}} \boldsymbol{p(x,y,C,D|z,B)\,dy = p(x, Y \in \mathcal{B}, C, D|z,B)}$$

$$\boldsymbol{p(x,y,\emptyset,D|z,B) = 0}$$

$$\boldsymbol{p(x,y,C_1 \cup C_2,D|z,B)}$$
$$\boldsymbol{= p(x,y,C_1,D|z,B)}$$
$$\boldsymbol{+ p(x,y,C_2,D|z,B), \quad C_1 \cap C_2 = \emptyset}$$

$$\boldsymbol{p(x,y,\Omega,D|z,B) = p(x,y,D|z,B)}$$

$$\boldsymbol{p(x,y,C,\emptyset|z,B) = 0}$$

$$\boldsymbol{p(x,y,C,D_1 \cup D_2|z,B)} \\ \boldsymbol{= p(x,y,C,D_1|z,B)} \\ \boldsymbol{+ p(x,y,C,D_2|z,B), \quad D_1 \cap D_2 = \emptyset}$$

$$\boldsymbol{p(x,y,C,\Omega|z,B) = p(x,y,C|z,B)}$$

In other words, $p(\cdot,\cdot,C,D|z,B)$ is an unnormalized joint probability density, $p(x,y,\cdot,D|z,B)$ and $p(x,y,C,\cdot\,|z,B)$ are unnormalized probabilities.

# 4. Mixed Product Rule

We consider the product factorization of $p(\alpha,\beta)$, such as $p(A,y)$, $p(x,y,C)$, $p(x,y,C,D)$. Here, we first consider the product factorization of $p(\alpha,\beta)$ when $\alpha$ and $\beta$ are a single random event or continuous random variable.

(1) If $A$ and $B$ are both random events, then

$$\boldsymbol{p(A,B) = p(B)p(A|B) = p(A)p(B|A).}$$

(2) If $Y$ is a continuous random variable, then

$$p(y)p(A|y) \\ = \lim_{\Delta y\to 0}\frac{p(y<Y\le y+\Delta y)}{\Delta y}\lim_{\Delta y\to 0}\frac{p(A,y<Y\le y+\Delta y)}{p(y<Y\le y+\Delta y)} \\ = \lim_{\Delta y\to 0}\frac{p(A,y<Y\le y+\Delta y)}{\Delta y} = p(A,y)$$

Similarly,

$$p(A)p(y|A) \\ = p(A)\lim_{\Delta y\to 0}\frac{p(y<Y\le y+\Delta y|A)}{\Delta y} \\ = \lim_{\Delta y\to 0}\frac{p(A,y<Y\le y+\Delta y)}{\Delta y} = p(A,y)$$

Therefore,

$$p(y)p(A|y) = p(A)p(y|A) = p(A,y).$$

(3) If $X$ and $Y$ are both continuous random variables, then

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y).$$

In summary, whether $\alpha$ and $\beta$ are random events or random variables, the product rule holds:

$$p(\alpha,\beta) = p(\alpha)p(\beta|\alpha) = p(\beta)p(\alpha|\beta). \qquad (2)$$

We now further consider the product factorization of $p(\alpha,\beta)$, where $\alpha$ and $\beta$ are multiple random events and continuous random variables separated by commas. Taking $p(x,y,C,D)$ for an example, one of multiple possible product factorizations is

$$p(x,C)p(y,D|x,C) \\ = \lim_{\Delta x\to 0}\frac{p(x<X\le x+\Delta x,C)}{\Delta x}\lim_{\Delta y\to 0}\frac{1}{\Delta y}\lim_{\Delta x\to 0}\frac{p(y<Y\le y+\Delta y,D,x<X\le x+\Delta x,C)}{p(x<X\le x+\Delta x,C)} \\ = \lim_{\Delta y\to 0}\frac{1}{\Delta y}\lim_{\Delta x\to 0}\frac{p(y<Y\le y+\Delta y,D,x<X\le x+\Delta x,C)}{\Delta x} \\ = \lim_{(\Delta x,\Delta y)\to(0,0)}\frac{p(x<X\le x+\Delta x,y<Y\le y+\Delta y,C,D)}{\Delta x\,\Delta y} = p(x,y,C,D)$$

Note that here we assume the iterated limit equals the double limit, and that the order of iterated limits can be interchanged. Hence, it is easy to understand that the product rule (2) still holds when $\alpha$ and $\beta$ represent multiple random events and random variables separated by commas. For example,

$$p(x,B,z,D) = p(x)p(B,z,D|x) = p(x,B)p(z,D|x,B) \\ = p(x,B,z)p(D|x,B,z) = \cdots.$$

From the product rule it may be easier to see that $p(\cdot,\cdot,C,D)$ is an unnormalized joint probability density and $p(x,y,\cdot,D)$ is an unnormalized probability, since $p(x,y,C,D) = p(C,D)p(x,y|C,D) = p(x,y,D)\,p(C|x,y,D)$. Additionally, we can determine that the normalization factors are $p(C,D)$ and $p(x,y,D)$, respectively.

# 5. Conclusion

The probability expressions encountered by students in basic courses, which involve either random events or continuous random variables, are extended to include both random events and continuous random variables simultaneously. The meanings and properties of mixed joint probability (density) and mixed conditional probability (density) are presented. The product rule is also extended, applying not only to multiple random events or continuous random variables, but also to combinations of both random events and continuous random variables. These topics can be included in specialized courses, and after teaching them, the content of the specialized course can then be covered straight away. This approach can help bridge the gap between basic and specialized courses.


# Acknowledgments

The author thanks his colleges and students for their kindness and support.


# References


[1] Athanasios Papoulis and S. Unnikrishna Pillai (2002). Probability, random variables, and stochastic processes (fourth edition). McGraw-Hill.

[2] John A. Gubner (2006). Probability and random processes for electrical and computer engineers. Cambridge University Press.

[3] John M. Wozencraft and Irwin Mark Jacobs (1965). Principles of communication engineering. John Wiley & Sons.

[4] Christopher M. Bishop (2006). Pattern recognition and machine learning. Springer.

[5] Carl Edward Rasmussen and Christopher K. I. Williams (2006). Gaussian processes for machine learning. MIT Press.