Chapter 18

Joint Distributions, Marginals, Conditionals

Lecture 18 | Devore 9e: 5.1–5.2

18.1 Warm-up: linear transformations

Example 18.1 (Curving an exam)**.** Scores X_{old} are to be rescaled by $X_{\text{new}} = aX_{\text{old}} + b$ so that the top score of 100 stays 100 and the class average of 75 becomes 83 (a B). Solve

$$\begin{cases} 100a + b = 100 \\ 75a + b = 83 \end{cases} \implies 25a = 17, \quad a = 0.68, \quad b = 32.$$

Recall from Lectures 9–10 that this shifts the mean to $a\mu + b$ and *shrinks* the standard deviation to $a\sigma = 0.68\sigma$ — a curve of this kind compresses the class together as well as moving it up.

18.2 From one random variable to two

Everything so far concerned a single X. But the interesting questions are usually about relationships: height *and* weight, auto premium *and* home premium, almonds *and* cashews in the same can. So we now study pairs (X, Y) defined on the same sample space.

18.2.1 The discrete case

Definition 18.2 (Joint pmf)**.** For discrete X, Y on the same sample space,

$$p(x, y) := \mathbb{P}\{X = x \text{ and } Y = y\},$$

which satisfies $p \geq 0$ and $\sum_x \sum_y p(x, y) = 1$. For any set A of pairs,

$$\mathbb{P}\{(X, Y) \in A\} = \sum_{(x,y) \in A} p(x, y).$$

Example 18.3 (Insurance)**.** X is the auto deductible and Y the homeowner's deductible for a randomly chosen customer:

$p(x,y)$	$y=500$	$y=1000$	$y=5000$	$p_X(x)$
$x=100$	0.30	0.05	0	0.35
$x=500$	0.15	0.20	0.05	0.40
$x=1000$	0.10	0.10	0.05	0.25
$p_Y(y)$	0.55	0.35	0.10	1

Definition 18.4 (Marginal pmf)**.**

$$p_X(x) = \sum_y p(x,y), \qquad p_Y(y) = \sum_x p(x,y).$$

Marginals are the row and column sums — literally written in the margins of the table, which is where the name comes from. Note that the marginals do *not* determine the joint pmf: many tables have the same margins.

18.2.2 The continuous case

Definition 18.5 (Joint pdf)**.** $f(x,y)$ is a **joint pdf** if $f \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$; then

$$\mathbb{P}\{(X,Y)\in A\} = \iint_{(x,y)\in A} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Where a single-variable pdf was a curve and probability was area under it, a joint pdf is a *surface* and probability is the volume under it over a region.

Definition 18.6 (Marginal pdf)**.**

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x.$$

Example 18.7. $f(x,y) = \frac{6}{5}(x+y^2)$ for $0 \leq x \leq 1$, $0 \leq y \leq 1$. First check that this is a density: $\int_0^1\int_0^1 \frac{6}{5}(x+y^2)\,\mathrm{d}y\,\mathrm{d}x = \frac{6}{5}\big(\frac{1}{2}+\frac{1}{3}\big) = 1$. ✓ Then

$$\mathbb{P}\Big\{0 \leq X \leq \tfrac{1}{4},\ 0 \leq Y \leq \tfrac{1}{4}\Big\} = \int_0^{1/4}\int_0^{1/4} \tfrac{6}{5}(x+y^2)\,\mathrm{d}y\,\mathrm{d}x = \tfrac{6}{5}\left(\tfrac{1}{128}+\tfrac{1}{768}\right) = \frac{7}{640} \approx 0.0109.$$

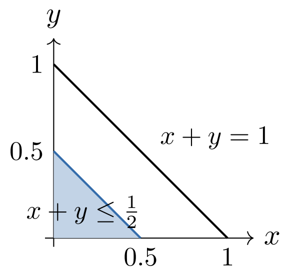
Example 18.8 (Can of nuts — the running example)**.** A 1-lb can of "deluxe mixed nuts" contains almonds, cashews and peanuts. Let X be the weight of almonds and Y the weight of cashews, so $1-X-Y$ is the weight of peanuts. Take

$$f(x,y) = 24xy \quad \text{for } x,y \geq 0,\ x+y \leq 1,$$

and 0 otherwise. Note the support is a *triangle*, not a rectangle — this is where most of the difficulty lives. It really is a density:

$$\int_0^1 \int_0^{1-x} 24xy \,\mathrm{d}y \,\mathrm{d}x = \int_0^1 12x(1-x)^2 \,\mathrm{d}x = 1. \quad \checkmark$$

What is $\mathbb{P}\{X + Y \leq 0.5\}$? The region is the triangle below the line $y = -x + \frac{1}{2}$:



$$\mathbb{P}\{X + Y \leq 0.5\} = \int_0^{1/2} \left(\int_0^{-x+1/2} 24xy \,\mathrm{d}y \right) \mathrm{d}x = \int_0^{1/2} 12x \left(\tfrac{1}{2} - x \right)^2 \mathrm{d}x = \frac{1}{16} = 0.0625.$$

Remark 18.9. The hard part of every joint-pdf problem is the *limits of integration*, not the integrand. Draw the region first. Always.

18.3 Independence

Definition 18.10 (Independent random variables)**.** X and Y are **independent** if

$$\text{(discrete)} \quad p(x, y) = p_X(x)\, p_Y(y) \qquad \text{or} \qquad \text{(continuous)} \quad f(x, y) = f_X(x)\, f_Y(y)$$

for all x, y. More generally $X_1, \ldots, X_n$ are independent if every subcollection has joint pmf/pdf equal to the product of its marginals.

Example 18.11. $X \sim \text{Exponential}(\lambda_1)$ and $Y \sim \text{Exponential}(\lambda_2)$ independent gives

$$f(x, y) = \lambda_1 e^{-\lambda_1 x} \cdot \lambda_2 e^{-\lambda_2 y} = \lambda_1 \lambda_2 e^{-(\lambda_1 x + \lambda_2 y)}, \qquad x, y \geq 0,$$

so, for instance, $\mathbb{P}\{X \geq 1500,\, Y \geq 1500\} = e^{-1500\lambda_1} e^{-1500\lambda_2}$.

Remark 18.12 (A fast test for dependence)**.** In the can-of-nuts example $f(x, y) = 24xy$ *looks* like a product — but the support $\{x + y \leq 1\}$ is not a rectangle, so knowing $X = 0.9$ forces $Y \leq 0.1$. The variables are dependent. Rule of thumb: if the support is not a product region, the variables cannot be independent no matter how the formula factors.

Example 18.13 (Maxima and minima)**.** If $X_1, \ldots, X_n$ are independent, then for $T_{\max} = \max(X_1, \ldots, X_n)$,

$$\mathbb{P}\{T_{\max} \leq t\} = \mathbb{P}\{X_1 \leq t, \ldots, X_n \leq t\} = \prod_{i=1}^{n} F_{X_i}(t),$$

and similarly $\mathbb{P}\{T_{\min} > t\} = \prod_i \bigl(1 - F_{X_i}(t)\bigr)$. This is the standard way to handle "the first component to fail" and "the last one standing".

18.4 Conditional distributions

Definition 18.14 (Conditional pdf/pmf)**.** For $f_X(x) > 0$,

$$f_{Y|X}(y \mid x) := \frac{f(x,y)}{f_X(x)}, \qquad \text{and symmetrically} \qquad f_{X|Y}(x \mid y) := \frac{f(x,y)}{f_Y(y)}.$$

The discrete version replaces f by p throughout.

Compare with $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ from Lecture 5: same shape, with the joint density playing the role of $\mathbb{P}(A \cap B)$ and the marginal that of $\mathbb{P}(B)$.

For each fixed x, $f_{Y|X}(\cdot \mid x)$ is a genuine pdf in y, so it can be integrated like any other:

$$\mathbb{P}\{Y \le 0.5 \mid X = 0.2\} = \int_{-\infty}^{0.5} f_{Y|X}(y \mid 0.2)\,\mathrm{d}y.$$

Remark 18.15. Conditioning on $\{X = 0.2\}$ is conditioning on an event of probability zero, which Lecture 5 did not allow. The definition above is how that is made legitimate: it is a limit of conditioning on $\{0.2 - \delta \le X \le 0.2 + \delta\}$.

Proposition 18.16. *X and Y are independent if and only if $f_{Y|X} = f_Y$ and $f_{X|Y} = f_X$; that is, conditioning changes nothing.*

Example 18.17 (Can of nuts, conditionals)**.** With $f(x,y) = 24xy$ on the triangle,

$$f_X(x) = \int_0^{1-x} 24xy\,\mathrm{d}y = 12x(1-x)^2, \qquad 0 \le x \le 1,$$

so for $0 < x < 1$,

$$f_{Y|X}(y \mid x) = \frac{24xy}{12x(1-x)^2} = \frac{2y}{(1-x)^2}, \qquad 0 \le y \le 1-x.$$

Given that the can holds x pounds of almonds, the cashew weight is Beta$(2,1)$-shaped on $[0, 1-x]$ — and the range depends on x, which is dependence made visible.