Statistical Theory and Modeling (ST2601)
Joint distributions

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Overview

- **Joint, marginal and conditional distributions for discrete variables**
- **Double integrals**
- **Joint, marginal and conditional distributions for continuous variables**
- **Independent variables**
- **Covariance and Correlation**
- **Conditional expectation**

Joint distribution - discrete variables

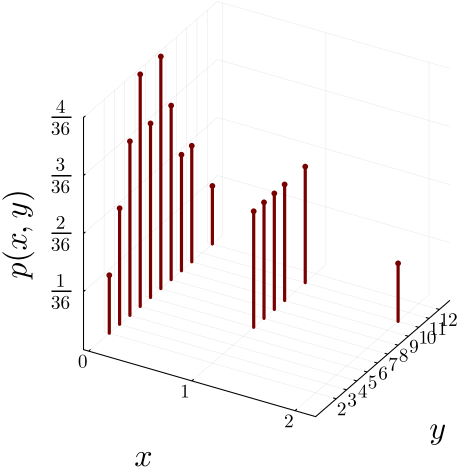
- Joint probability function for two discrete X and Y

$$p(x, y) = \Pr(X = x, Y = y)$$

- Example: Roll two dice.
  - X = the number of dice with 5
  - Y = sum of two dice

$X \backslash Y$	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{0}{36}$	$\frac{1}{36}$
1	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	$\frac{2}{36}$	0
2	0	0	0	0	0	0	0	0	$\frac{1}{36}$	0	0

Joint distribution - discrete variables



Marginal distribution - discrete variables

- **Marginal distribution** $p_X(x)$ for X: probability distribution for X *regardless of what happens to* Y.

$$p_X(x) = \sum_{\text{all } y} p(x, y)$$

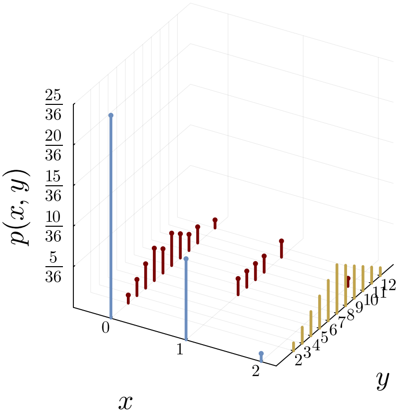
- **Marginal distribution** $p_Y(y)$ for Y

$$p_Y(y) = \sum_{\text{all } x} p(x, y)$$

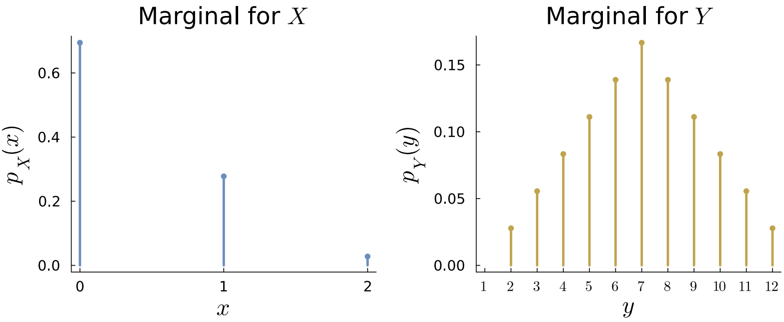
$X \backslash Y$	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	$p(x)$
0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	$\frac{1}{36}$	$\frac{25}{36}$
1	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	$\frac{2}{36}$	0	$\frac{10}{36}$
2	0	0	0	0	0	0	0	0	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$p(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	

Marginal distribution - discrete variables

$$p_X(x) = \sum_y p(x, y) = \begin{cases} \frac{25}{36} & \text{for } x = 0 \\ \frac{10}{36} & \text{for } x = 1 \\ \frac{1}{36} & \text{for } x = 2 \end{cases} \quad (1)$$

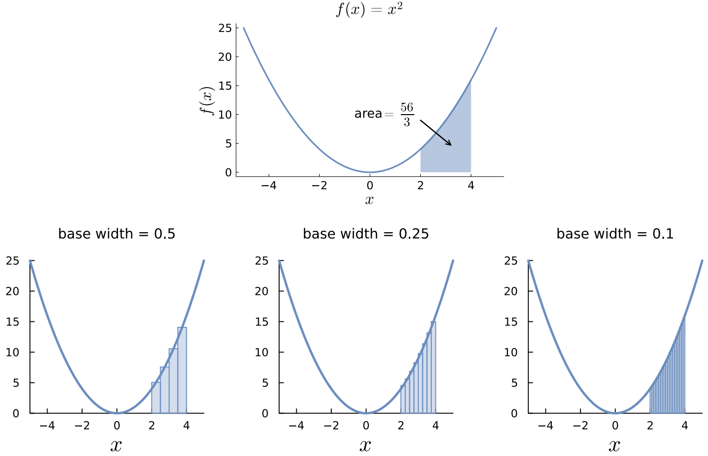


Marginal distribution - discrete variables



Single integral for function $f(x)$

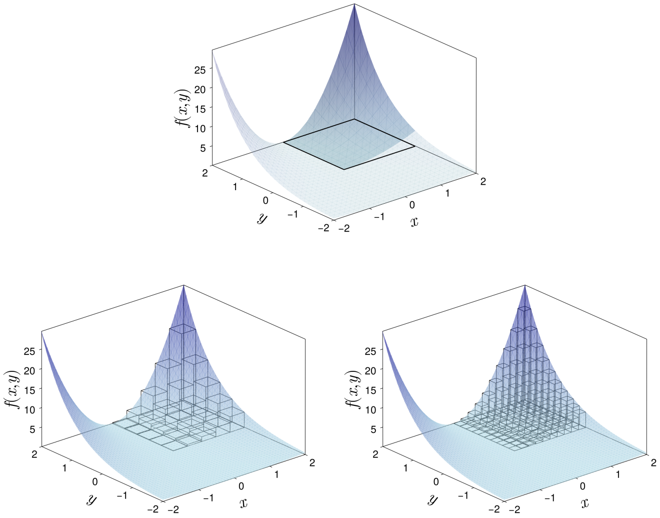
- **Integral** = **area** under curve $y = f(x)$



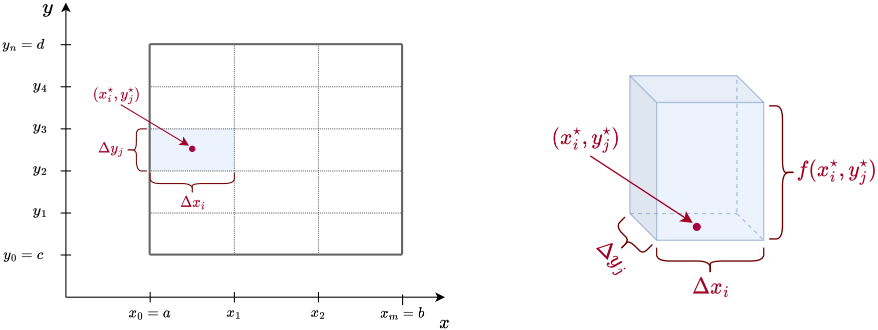
$$\sum_{i=1}^{n} f(x_i^\star)\Delta x_i \to \int_a^b f(x)\mathrm{d}x$$

Double integral for bivariate function $f(x, y)$

- **Double integral** $=$ **volume** under **surface** $z = f(x, y)$



Bivariate integrals



$$\sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i^\star, y_j^\star)\Delta x_i \Delta y_j \to \int_c^d \int_a^b f(x,y)\mathrm{d}x\mathrm{d}y$$

Double integrals in action

- Two-step approach:
  - first integrate with respect to x while treating y as a constant
  - then integrate with respect to y.
- Example: $f(x, y) = x^2 y$, integrate over $(x, y) \in (0, 1) \times (0, 1)$

$$\int_0^1 \int_0^1 x^2 y \mathrm{d}x \mathrm{d}y = \int_0^1 \left[\frac{1}{3} x^3 y\right]_0^1 \mathrm{d}y = \int_0^1 \left(\frac{1}{3} y\right) \mathrm{d}y = \left[\frac{1}{2 \cdot 3} y^2\right]_0^1 = \frac{1}{6}$$

Double integrals - non-rectangular integration region

- Integration region may not be rectangular.
- $f(x, y) = x^2 y$, integrate over triangular region:

$$(x, y) \in (0, 1) \times (0, 1) \quad \text{and } x \leq y$$

$$\int_0^1 \int_0^y x^2 y \mathrm{d}x \mathrm{d}y = \int_0^1 \left[\frac{1}{3} x^3 y\right]_0^y \mathrm{d}y = \int_0^1 \left(\frac{1}{3} y^4\right) \mathrm{d}y = \left[\frac{1}{5 \cdot 3} y^5\right]_0^1 = \frac{1}{15}$$

- General notation where R is some region in (x, y)-space

$$\iint_R f(x, y) \mathrm{d}x \mathrm{d}y$$

Joint cumulative distribution function

- **Joint cumulative distribution** for two random variables X and Y

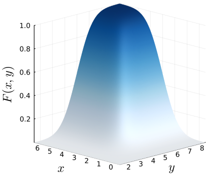
$$F(x, y) = \Pr(X \leq x, Y \leq y)$$

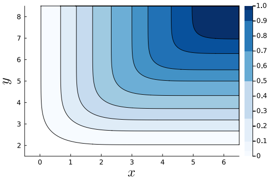
- Marginal distributions are special cases:

$$F(x, \infty) = \Pr(X \leq x, Y \leq \infty) = F_X(x)$$
$$F(\infty, y) = \Pr(X \leq \infty, Y \leq y) = F_Y(y)$$

- Other properties

$$F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0 \text{ and } F(\infty, \infty) = 1$$





Joint density function

- **Joint density function** for two random variables X and Y

$$f(x, y)$$

$$\Pr(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) \mathrm{d}x \mathrm{d}y$$

- Properties $f(x, y) \geq 0$ and

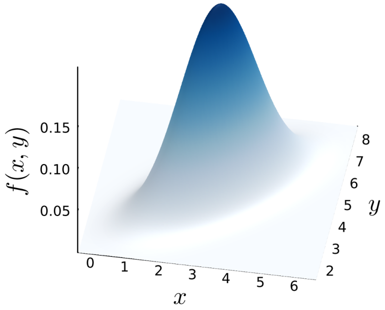
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \mathrm{d}x \mathrm{d}y = 1$$

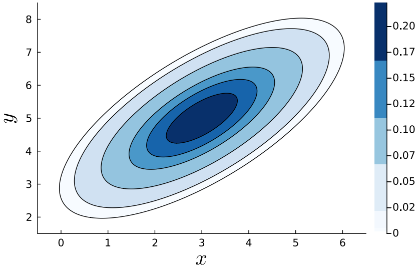
- Example: $f(x, y) = 6x^2 y$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$.
  Check:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 6x^2 y \, \mathrm{d}x \mathrm{d}y = \int_0^1 \left[6x^2 \frac{1}{2} y^2 \right]_0^1 \mathrm{d}x$$

$$= \int_0^1 3x^2 \, \mathrm{d}x = \left[x^3 \right]_0^1 = 1$$

Joint density function





Marginal distributions

- **Marginal density** for X

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

- **Marginal density** for Y

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

- Example: Marginal density for X

$$f_X(x) = \int 6x^2 y\,\mathrm{d}y = \left[6x^2 \frac{1}{2} y^2\right]_0^1 = 3x^2$$

- Example: Marginal density for Y

$$f_Y(y) = \int 6x^2 y\,\mathrm{d}x = \left[2x^3 y\right]_0^1 = 2y$$

Conditional distributions

- **Conditional probability events** for $\Pr(B) > 0$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- **Conditional distribution** of X given $Y = y$

$$p_{X|Y}(x|Y = y) = \frac{p(x, y)}{p_Y(y)}$$

- Continuous X and Y

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

- Example: $f(x, y) = 6x^2y$ and $f_Y(y) = 2y$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{6x^2y}{2y} = 3x^2$$

Marginal-conditional decomposition

- Conditional distribution

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

- **Marginal-conditional decomposition** of a joint density

$$\underbrace{f(x,y)}_{\text{joint}} = \underbrace{f_{X|Y}(x|y)}_{\text{conditional}} \cdot \underbrace{f_Y(y)}_{\text{marginal}}$$

- This is a great way to build models!
- Example: the joint density

$$f(x,y) = \frac{1}{x} e^{-\left(\frac{y}{x}+x\right)} \qquad 0 < x < \infty, 0 < y < \infty$$

$$X \sim \text{Expon}(1)$$
$$Y|(X = x) \sim \text{Expon}(x)$$

- The conditional $Y|(X = x)$ doesn't fit a scatter of y and x? Swap it out with something else, maybe a Gamma? 😍

Independent random variables

- **Independent events** if and only if

$$\Pr(A|B) = \Pr(A)$$

alternatively

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

- Knowing that B has occured has no affect on my beliefs about A.

- Two random variables are **independent** if and only if

$$p_{X|Y}(x|Y = y) = p_X(x)$$

alternatively

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

- Example: $f(x, y) = 6x^2y$, with $f_X(x) = 3x^2$ and $f_Y(y) = 2y$. X and Y are independent since

$$f_X(x)f_Y(y) = 3x^2 \cdot 2y = 6x^2y = f(x, y)$$

Multivariate distributions

- **Joint probability density** for $X_1, X_2, \dots, X_n$

$$f(x_1, x_2, \dots, x_n)$$

- **Marginal distribution** for X_1

$$f_{X_1}(x_1) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-1 \text{ integrals}} f(x_1, x_2, \dots, x_n) \underbrace{\mathrm{d}x_2 \cdots \mathrm{d}x_n}_{\text{all except } \mathrm{d}x_1}$$

- **Marginal distribution** for (X_1, X_2)

$$f_{X_1,X_2}(x_1, x_2) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-2 \text{ integrals}} f(x_1, x_2, \dots, x_n) \quad \underbrace{\mathrm{d}x_3 \cdots \mathrm{d}x_n}_{\text{all except } \mathrm{d}x_1 \text{ and } \mathrm{d}x_2}$$

- **Conditional distribution** for X_1

$$f(x_1 | X_2 = x_2, \dots, X_n = x_n) = \frac{f(x_1, x_2, \dots, x_n)}{f(x_2, \dots, x_n)}$$

Covariance and Correlation

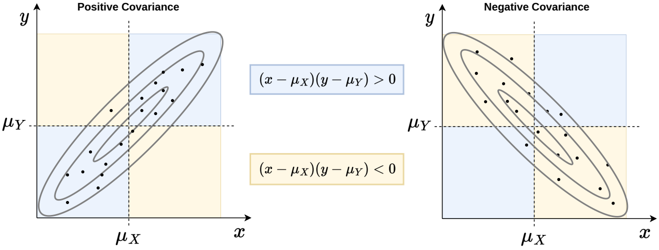
- **Covariance** between X and Y

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left((X - \mu_X)(Y - \mu_Y)\right)$$

where $\mu_X = \mathbb{E}(X)$ and $\mu_Y = \mathbb{E}(Y)$.

- **Correlation** between X and Y

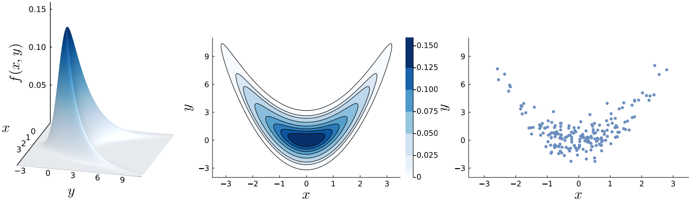
$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$



Covariance measures linear dependence

- **Covariance/Correlation** - measures **linear dependence**.
- Independent X and Y implies zero correlation $\rho_{XY} = 0$
- Zero correlation does **not** in general imply independence.
- Dependence can be non-linear.
- Example with $\rho_{XY} = 0$:

$$X \sim N(0,1) \qquad \text{and } Y|(X = x) \sim N(x^2, 1)$$



Conditional expectation

- **Conditional expectation**

$$\mathbb{E}(Y|X=x) = \begin{cases} \sum_y y \cdot p(y|x) & \text{if } x \text{ and } y \text{ discrete} \\ \int y \cdot f(y|x)\mathrm{d}y & \text{if } x \text{ and } y \text{ continuous} \end{cases}$$

- **Regression** and **classification** models the conditional expectation.
- Computing the expectation $\mathbb{E}(Y)$ directly is sometimes hard.
- But the conditional expectation $\mathbb{E}(Y|X=x)$ may be simpler.
- Two-step approach:
  1. Compute conditional expectation $\mathbb{E}(Y|X)$
  2. Undo the conditioning on X with $\mathbb{E}_X$
- **Law of iterated expectation**

$$\mathbb{E}(Y) = \mathbb{E}_X\left(\mathbb{E}_{Y|X}(Y|X)\right)$$

Law of iterated expectation in action

- Example:

$$X \sim \mathrm{Expon}(1)$$
$$Y|(X = x) \sim \mathrm{Expon}(x)$$

- Recall: if $X \sim \mathrm{Expon}(\beta)$ then $\mathbb{E}(X) = \beta$.
- Computing $\mathbb{E}(Y)$ directly requires marginal $f_Y(y)$. 🥺
- But the conditional expectation is easy:

$$\mathbb{E}_{Y|X}(Y|X) = X$$

- Finally, we undo the conditioning on X

$$\mathbb{E}(Y) = \mathbb{E}_X\left(\mathbb{E}_{Y|X}(Y|X)\right) = \mathbb{E}_X(X) = 1$$

- Ta-da! 🥳