Multivariate distributions

Cumulative distribution functions

The joint cdf of X and Y is: $F_{X,Y}(x,y) = F(x,y) = P(X \le x, Y \le y)$, for $x, y \in \mathbb{R}$

1. rarely used in the discrete case but is computed as

$$F(x,y) = P(X \le x, Y \le y) = \sum_{x_1 \le x, y_1 \le y} f(x_1, y_1)$$

2. may be useful in the continuous case and is computed as

$$(*) \quad F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)\,ds\,dt$$

where also in the continuous case

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{d \frac{dF(x,y)}{dy}}{dx} = f(x,y).$$

Properties: A function $F(x,y)$ is a cdf for some r.v. (X,Y) if and only if

1. ✓ $\lim\limits_{x \to -\infty} F(x,y) = \lim\limits_{y \to -\infty} F(x,y) = 0$ $\quad \lim\limits_{x \to -\infty} F(x,y) = \lim\limits_{x \to -\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,ds\,dt = 0$
2. ✓ $\lim\limits_{x,y \to \infty} F(x,y) = 1$ $\quad F(x,y) = \lim\limits_{(x,y) \to +\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,ds\,dt = 1$
3. Right continuous in each argument: $\lim\limits_{h \downarrow 0} F(x+h,y) = \lim\limits_{h \downarrow 0} F(x,y+h) = F(x,y)$
4. "nondecreasing" (the probability assigned to any rectangle is ≥ 0)

$$\left\{ \begin{aligned} & P(x < X \le x+\Delta_1, y < Y \le y+\Delta_2) \\ = \; & F(x+\Delta_1, y+\Delta_2) - F(x+\Delta_1, y) - F(x, y+\Delta_2) + F(x,y) \ge 0 \end{aligned} \right.$$

for all $x, y \in \mathbb{R}$ and for any $\Delta_1, \Delta_2 > 0$

$$P(x < X \le x+\Delta_1, y < Y \le y+\Delta_2) \overset{\text{def}}{=} \int_{x}^{x+\Delta_1} \underbrace{\int_{y}^{y+\Delta_2} f(s,t)\,ds\,dt}$$

$$= \int_{x}^{x+\Delta_1} \left[\int_{-\infty}^{y+\Delta_2} f(s,t)\,ds - \int_{-\infty}^{y} f(s,t)\,ds \right] dt = \cdots \quad ✓$$

Multivariate distributions

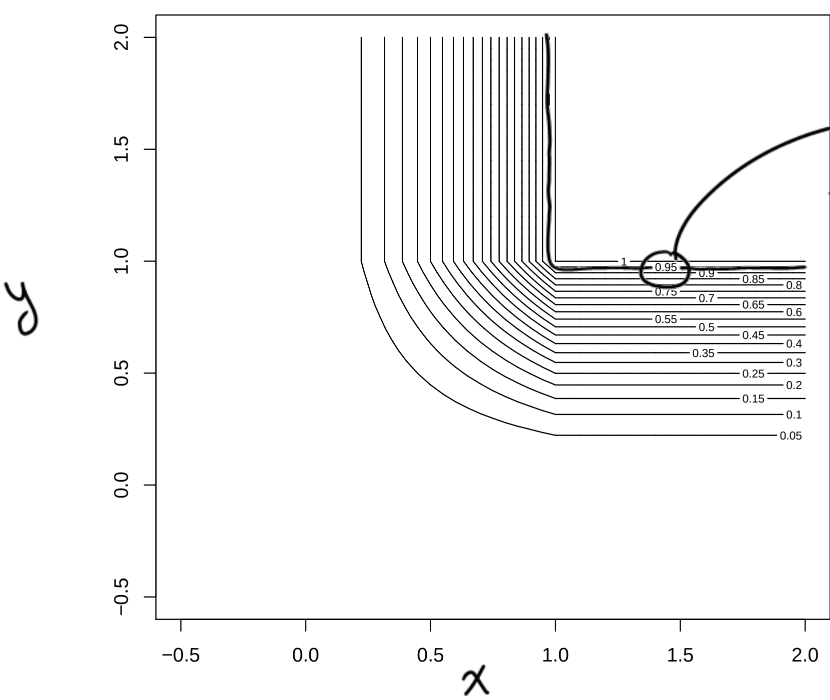
Cumulative distribution functions (cont'd)

e.g., $F(x,y) = \begin{cases} 0 & \text{if } x+y < -1 \\ 1 & \text{if } x+y \geq -1 \end{cases}$ is *not* a valid cdf, though

Properties 1-3 hold, and $F(x,y)$ is nondecreasing in its arguments x or y (but "probabilities of rectangles" are not always non-negative).

Cdf example (continuous case): $f(x,y) = 4xy$, for $0 < x, y < 1$

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \underbrace{f(s,t)dsdt}_{4st\,dsdt} = \begin{cases} 1 & x \geq 1, y \geq 1 \\ y^2 & x \geq 1, 0 < y < 1 \\ x^2 & 0 < x < 1, y \geq 1 \\ x^2y^2 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$



$\{(x,y): F(x,y) = .95\}$

Contour plot of $F(x,y)$

$\dfrac{\partial^2 F(x,y)}{\partial x\, \partial y} =$

Multivariate distributions

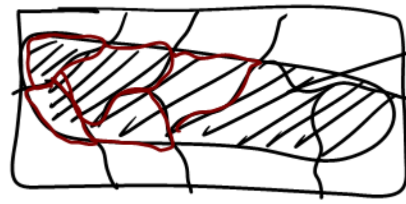
Marginal distributions

- Marginal distribution of X (or Y) describes the distribution of a single random variable that is implied by the joint distribution

  $(X,Y) \overset{?}{\Longrightarrow} X, Y$

  In other words, the joint distribution of (X, Y) must determine individual distributions for X and Y

- For discrete and continuous distributions, we often get the marginal distribution from the joint pmf or pdf $f(x, y)$



$\mathbb{P}(\underbrace{X=x}_{A})$

$= \mathbb{P}(\{X=x, Y=y_1\} \cup \{X=x, Y=y_2\} \cup \{X=x, Y=y_3\} \cup \vdots \cup \{X=x, Y=y_n\})$

Marginal Distributions in Discrete Case:

pmf $\longrightarrow$
$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f(x, y), \quad x \in \mathbb{R}$$

fix/given

similarly,

$\longrightarrow$
$$f_Y(y) = P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x f(x, y), \quad y \in \mathbb{R}$$

fix/given

$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \cdots + \mathbb{P}(A \cap B_n) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(X=x, Y=y_i)$

Discrete Example (artificial): (X, Y) with joint pmf given in tabular form as

		x		
		1	2	3
y	3	1/12	1/12	1/6
	2	1/12	1/6	1/12
	1	1/6	1/12	1/12

$$f_X(x) = \mathbb{P}(X=x) = \begin{cases} 1/12 + 1/12 + 1/6 & x = 1 \\ 1/12 + 1/6 + 1/12 & x = 2 \\ 1/6 + 1/12 + 1/12 & x = 3 \end{cases}$$

Multivariate distributions

Discrete marginal distributions: another example

Recall previous discrete example:

- Suppose there are n independent trials, where each trial has three possible outcomes: outcome a with prob. p_1, outcome b with prob. p_2 and outcome c with prob. $1-p_1-p_2$
- Let $X = \#$ of trials having outcome a & $Y = \#$ of trials having outcome b
- Then the joint pmf of (X,Y) is

$$\boxed{f(x,y) = \frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y (1-p_1-p_2)^{n-x-y}}$$

for $0 \le x, y, x+y \le n$ (assuming $0 < p_1, p_2, p_1+p_2 < 1$)

- Find the marginal pmf of X. Fix a x value in $\{0, 1, \ldots, n\}$.

$$\mathbb{P}(X=x) = f_X(x) = \sum_{y=0}^{n-x} \mathbb{P}(X=x, Y=y)$$

fix

$$= \sum_{y=0}^{n-x} \frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

$$= \frac{n!}{x!(n-x)!}p_1^x (1-p_1)^{n-x} \sum_{y=0}^{n-x} \frac{(n-x)!}{y!(n-x-y)!} \frac{p_2^y (1-p_1-p_2)^{n-x-y}}{(1-p_1)^{n-x}}$$

$$= \frac{n!}{x!(n-x)!}p_1^x (1-p_1)^{n-x} \underbrace{\sum_{y=0}^{n-x} \underbrace{\frac{(n-x)!}{y!(n-x-y)!}}_{\binom{n-x}{y}} \left(\underbrace{\frac{p_2}{1-p_1}}_{\tilde{p}}\right)^y \left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-x-y}}_{1}$$

$$\tilde{p} := \frac{p_2}{1-p_1}$$

$$1-\tilde{p} = 1 - \left[\frac{p_2}{1-p_1}\right] = \frac{1-p_1-p_2}{1-p_1}$$

$$\Rightarrow \mathbb{P}(X=x) = \binom{n}{x} p_1^x (1-p_1)^{n-x}$$

$$\Rightarrow \boxed{X \sim \text{Bin}(n, p_1)}$$

$$= \underbrace{\sum_{y=0}^{n-x} \binom{n-x}{y} \tilde{p}^y (1-\tilde{p})^{n-x-y}}_{*} = 1$$

$\longrightarrow$ ⊛ is the pmf of
$\text{Bin}\left(n-x, \frac{p_2}{1-p_1}\right)$

Multivariate distributions

Marginal distributions: $\boxed{\text{continuous case}}$

In the continuous (X,Y) case, note that, for $x \in \mathbb{R}$,

$$P(X \le x) = P\big((X,Y) \in (-\infty, x] \times (-\infty, \infty)\big) = \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{\infty} f(s,t)\,ds}_{g(t)}\,dt$$

$$\int_{-\infty}^{x} g(t)\,dt \equiv h(x)$$

Marginal Distributions in Continuous Case:

$$\longrightarrow \quad f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad \text{for given } x \in \mathbb{R}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \quad \text{for given } y \in \mathbb{R}$$

We often speak of "integrating out" x or y to obtain a marginal pdf

Continuous example: joint pdf $f(x,y) = 1/x$, for $0 < y < x < 1$

$$f_X(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^x \frac{1}{x}\,dy = \frac{1}{x}\,y\Big]_0^x = 1 \quad \forall \quad x \in (0,1)$$

$$f_Y(y) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x,y)\,dx = \int_y^1 \frac{1}{x}\,dx = \log x\Big]_y^1 = \log 1 - \log y = -\log y \quad \forall\, y \in (0,1)$$

$$f_Y(y) = \begin{cases} -\log y & y \in (0,1) \\ 0 & \text{o.w.} \end{cases} \qquad f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{o.w.} \end{cases}$$

Multivariate distributions

From bivariate to multivariate

Now consider the n-dimensional case: $\mathbb{R}^n$-valued r.v. $(X_1, \ldots, X_n)$

Cdf: $F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ for $x_1, \ldots, x_n \in \mathbb{R}$

1. $\displaystyle\lim_{\text{all } x_i \to \infty} F(x_1, \ldots, x_n) = 1$
2. $\displaystyle\lim_{\text{any } x_i \to -\infty} F(x_1, \ldots, x_n) = 0$
3. right continuous in each argument
4. "monotonicity" condition from bivariate case generalized to n-dim rectangle

Pdf or pmf: $f(x_1, \ldots, x_n) \geq 0$ with integral (or sum) equal to 1

$$P((X_1, \ldots, X_n) \in A) = \int \cdots \int_{\{(x_1, \ldots, x_n) \in A\}} f(x_1, \ldots, x_n) dx_1 \ldots, dx_n, \quad A \subset \mathbb{R}^n$$

$$P((X_1, \ldots, X_n) \in A) = \sum_{(x_1, \ldots, x_n) \in A} f(x_1, \ldots, x_n)$$

Marginal pdf or pmf for *any* subset of $(X_1, \ldots, X_n)$ → n – R.V.

e.g., if you want the marginal distribution of $(X_1, \ldots, X_k)$ from $(X_1, \ldots, X_n)$, fix values of $x_1, \ldots, x_k \in \mathbb{R}$, and sum or integrate out the remaining variables

$$f(x_1, \ldots, x_k) = \int \cdots \int f(x_1, \ldots, x_n) dx_{k+1} \ldots, dx_n$$

$$f(x_1, \ldots, x_k) = \sum_{(x_{k+1}, \ldots, x_n)} f(x_1, \ldots, x_n)$$

We'll still mostly focus on the bivariate case, though the same ideas apply to more than 2 variables (though more computationally involved than we presently need)

Multivariate distributions

Expectations of several random variables

Recall: $n=1$, $\mathbb{E}[g(X)] = \int g(x) f_X(x)dx$

Provided $\int |g(x) f_X(x)| < \infty$

Definition: extension of the univariate definition

If $(X_1, \ldots, X_n)$ are jointly discrete with pmf f and $g : \mathbb{R}^n \to \mathbb{R}$, we define

$$\mathrm{E}g(X_1, \ldots, X_n) = \sum_{(x_1,\ldots,x_n)} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n)$$

$$\sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

provided that

$$\sum_{(x_1,\ldots,x_n)} |g(x_1, \ldots, x_n)| f(x_1, \ldots, x_n) < \infty$$

If $(X_1, \ldots, X_n)$ are jointly continuous with pmf f and $g : \mathbb{R}^n \to \mathbb{R}$, we define

$$\mathrm{E}g(X_1, \ldots, X_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

provided that

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(x_1, \ldots, x_n)| f(x_1, \ldots, x_n) dx_1 \ldots dx_n < \infty$$

In the bivariate case, these become

$n=2$

$$\mathrm{E}g(X, Y) = \begin{cases} \sum_{(x,y)} g(x, y) f(x, y) & \text{discrete case} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy & \text{continuous case} \end{cases}$$

Multivariate distributions

Expectations of several random variables (cont'd)

Theoretical properties as before

e.g., if $a_1, \ldots, a_k, b \in \mathbb{R}$ and each $g_i : \mathbb{R}^n \to \mathbb{R}$

$$\mathrm{E}\left(\sum_{i=1}^{k} a_i g_i(X_1, \ldots, X_n) + b\right) = b + \sum_{i=1}^{k} a_i \mathrm{E} g_i(X_1, \ldots, X_n)$$

Marginal moments:

$$\mathrm{E}X_i = \mu_{X_i} = \begin{cases} \displaystyle\sum_{(x_1,\ldots,x_n)} x_i f(x_1, \ldots, x_n) & \text{discrete case} \\ \displaystyle\int \cdots \int x_i \underbrace{f(x_1, \ldots, x_n)} \underbrace{dx_1 \ldots, dx_n} & \text{continuous case} \end{cases}$$

$$\mathbb{E}X = \iint x f(x,y)\, dx\, dy \qquad \mathbb{E}Y = \iint y f(x,y)\, dx\, dy$$

Discrete Example: (X, Y) with joint pmf given in tabular form as

		x = 1	2	3
	3	1/12	1/12	1/6
y	2	1/12	1/6	1/12
	1	1/6	1/12	1/12