Joint densities

Def. Random variables X and Y have *joint density function* $f(x,y)$ if

$$\mathbb{P}\big((X,Y)\in A\big) = \iint_A f(x,y)\,dx\,dy \qquad (*)$$

for every "reasonable" set $A \subseteq \mathbb{R}^2$.

Properties: (i) $f(x,y) \geq 0 \quad \forall x,y$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$ (take $A=\mathbb{R}^2$ in $(*)$).

Note that in particular,

$$\begin{aligned}\mathbb{P}(a \leq X \leq b) &= \mathbb{P}(a \leq X \leq b,\ -\infty < Y < \infty) \\ &= \mathbb{P}\big((X,Y)\in [a,b]\times\mathbb{R}\big) \\ &= \int_a^b \Big(\int_{-\infty}^{\infty} f(x,y)\,dy\Big)dx\end{aligned}$$

so that X has density

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

Similarly, Y has density

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.}$$

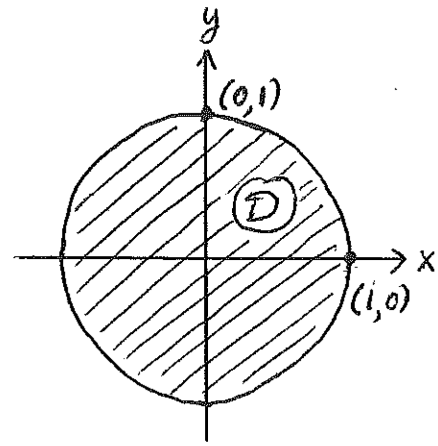
We call $f_X(x)$, $f_Y(y)$ the *marginal densities* of X and Y.

Def. Suppose X and Y have joint density $f(x,y)$. We say X and Y are <u>independent</u> if

$$f(x,y) = f_X(x)\, f_Y(y) \qquad \text{for all } x \text{ and } y.$$

Ex. Let $D = \{(x,y) : x^2+y^2 \le 1\}$, and

$$f(x,y) = \begin{cases} c, & \text{if } (x,y) \in D \\ 0, & \text{otherwise} \end{cases}$$



(<u>Uniform distribution</u> on D). Since

$$\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = \iint_D c\,dx\,dy = c \iint_D 1\,dx\,dy = c \cdot \text{area}(D) = c\pi,$$

we have $c = \frac{1}{\pi}$. The range of X is $[-1,1]$, and for $-1<x<1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,dy = \frac{2}{\pi}\sqrt{1-x^2}$$

Similarly, $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$ for $-1<y<1$. Thus

$$f_X(x) f_Y(y) = \frac{4}{\pi^2}\sqrt{(1-x^2)(1-y^2)} \neq f(x,y)$$

so X and Y are dependent.

Note that if $A \subseteq D$, then

$$P((X,Y)\in A) = \frac{\text{area}(A)}{\text{area}(D)}$$

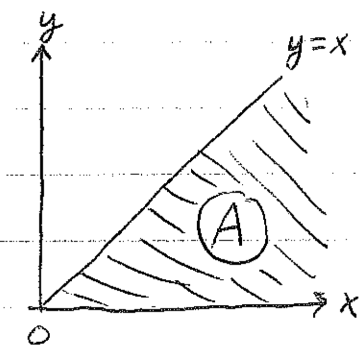
e.g. $P(X \ge 0, Y \ge 0) = \dfrac{\text{area of quarter disk}}{\text{area}(D)} = \dfrac{1}{4}.$

Ex. Let X, Y be indep., $X \sim Exp(\mu)$, $Y \sim Exp(\lambda)$. Find $P(X>Y)$.

Solution: $f_X(x) = \mu e^{-\mu x}$, $f_Y(y) = \lambda e^{-\lambda y}$ and by independence,

$$f(x,y) = \lambda\mu e^{-\mu x} e^{-\lambda y}, \qquad x>0,\ y>0.$$

Let $A = \{(x,y):\ x>y>0\}$. Then



$$P(X>Y) = P((X,Y)\in A) = \iint_A f(x,y)\,dx\,dy$$

$$= \int_0^\infty \left(\int_y^\infty \lambda\mu e^{-\mu x} e^{-\lambda y}\,dx \right) dy$$

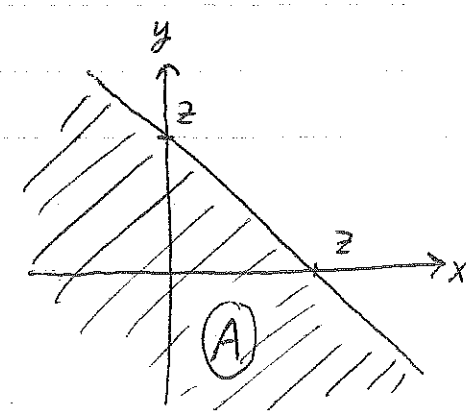
$$= \int_0^\infty \lambda e^{-\lambda y} \left(\int_y^\infty \mu e^{-\mu x}\,dx \right) dy = \int_0^\infty \lambda e^{-\lambda y} e^{-\mu y}\,dy$$

$$= \lambda \int_0^\infty e^{-(\lambda+\mu)y}\,dy = \frac{\lambda}{\lambda+\mu} \int_0^\infty (\lambda+\mu) e^{-(\lambda+\mu)y}\,dy$$

$$= \frac{\lambda}{\lambda+\mu}\cdot 1 = \frac{\lambda}{\lambda+\mu}.$$

Distribution of a sum.

Let X, Y be indep. r.v.'s with densities f_X and f_Y, and let $Z := X+Y$. Then Z has c.d.f.



$$F_Z(z) = P(X+Y \le z)$$

$$= P((X,Y)\in A)$$

$$= \iint_A f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_X(x) f_Y(y)\,dy \right) dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\ F_Y(z-x)\,dx$$

and differentiating with respect to z gives

$$\boxed{f_Z(z) = F_Z'(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx}$$

This is the <u>convolution formula</u> for densities.

<u>Ex.</u> Let X, Y be indep., $X \sim \text{Normal}(\mu, \sigma^2)$, $Y \sim \text{Normal}(\nu, \tau^2)$. Then $X+Y \sim \text{Normal}(\mu+\nu, \sigma^2+\tau^2)$. (What about $X-Y$?)

<u>Pf.</u> Calculus exercise! Remember this important result.

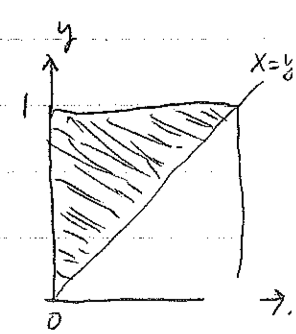
<u>Expectation</u>

If (X,Y) has joint density $f(x,y)$, and if $g(x,y)$ is any function, then

$$\boxed{E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f(x,y)\,dx\,dy}$$

<u>Ex.</u> Let

$$f(x,y) = \begin{cases} 10\,x y^2, & \text{if } 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$



Find:

a) the marginal densities of X and Y

b) $E(X)$ and $E(Y)$

c) $E(XY)$.

Solution: a) Range (X) = Range $(Y) = (0,1)$. For $0<x<1$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^1 10xy^2\,dy = 10x \cdot \tfrac{1}{3}y^3\Big|_{y=x}^{y=1} = \tfrac{10}{3}x(1-x^3)$$

and for $0<y<1$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^y 10xy^2\,dx = 5x^2y^2\Big|_{x=0}^{x=y} = 5y^4.$$

(Note that X and Y are dependent: $f(x,y) \neq f_X(x) f_Y(y)$.)

b) $$E(X) = \int x f_X(x)\,dx = \int_0^1 \tfrac{10}{3}x^2(1-x^3)\,dx = \tfrac{10}{3}\int_0^1 (x^2-x^5)\,dx$$

$$= \tfrac{10}{3}\left(\tfrac{1}{3}-\tfrac{1}{6}\right) = \tfrac{10}{18} = \tfrac{5}{9}$$

$$E(Y) = \int y f_Y(y)\,dy = \int_0^1 5y^5\,dy = \tfrac{5}{6}$$

c) $$E(XY) = \iint_{\mathbb{R}^2} xy\, f(x,y)\,dx\,dy = \int_0^1\left(\int_0^y 10x^2y^3\,dx\right)dy$$

$$= \int_0^1 \tfrac{10}{3}x^3\Big|_{x=0}^{x=y} y^3\,dy = \tfrac{10}{3}\int_0^1 y^6\,dy = \tfrac{10}{21}.$$

Covariance and correlation.

Def. Let X, Y be r.v.'s with $E(X) = \mu_X$, $E(Y) = \mu_Y$. The covariance of X and Y is defined by

$$(**) \qquad \mathrm{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right],$$

and the correlation of X and Y is defined by

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{SD(X)\,SD(Y)}$$

(provided the standard deviations exist).

Rewriting (**) gives a shortcut formula:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y \\ &= E(XY) - E(X)E(Y).\end{aligned}$$

Ex. In the previous example, $E(XY) = \frac{10}{21}$, $E(X) = \frac{5}{9}$ and $E(Y) = \frac{5}{6}$, so

$$\mathrm{Cov}(X,Y) = \frac{10}{21} - \frac{5}{9}\cdot\frac{5}{6} = \frac{10}{21} - \frac{25}{54} = \frac{5}{378}.$$

Def. X and Y are

a) positively correlated if $\mathrm{Cov}(X,Y) > 0$

b) negatively correlated if $\mathrm{Cov}(X,Y) < 0$

c) uncorrelated if $\mathrm{Cov}(X,Y) = 0$ (i.e. if $E(XY) = E(X)E(Y)$).

Note: If X and Y are independent, they are uncorrelated. But uncorrelated r.v.'s need not be independent!

Ex. Let $f(x,y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2+y^2 \le 1 \\ 0, & \text{otherwise.}\end{cases}$

Recall X and Y are dependent. But (check!) $E(XY) = 0 = E(X)E(Y)$, so X and Y are uncorrelated.

Ex. (Discrete case)

A box has 3 tickets labeled 1, 2 and 3. Two tickets are drawn from the box without replacement. Let $X = 1^{st}$ number drawn, $Y = 2^{nd}$ number.

Clearly $E(X) = E(Y) = 2$. Next,

$$E(XY) = \sum_{i=1}^{3} \sum_{j=1}^{3} ij\ P(X=i, Y=j) = \sum_{i=1}^{3} \sum_{\substack{j=1 \\ j\neq i}}^{3} ij \cdot \frac{1}{3} \cdot \frac{1}{2}$$

$$= \frac{1}{6} \sum_{i=1}^{3} i \Big(\sum_{\substack{j=1 \\ j\neq i}}^{3} j \Big) = \frac{1}{6} \left[1\cdot(2+3) + 2\cdot(1+3) + 3\cdot(1+2) \right]$$

$$= \frac{1}{6}(5+8+9) = \frac{22}{6} = \frac{11}{3}.$$

Thus, $Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{11}{3} - 4 = -\frac{1}{3}$.

(Why could negative correlation have been expected?) □

<u>Theorem</u>: $-1 \le Corr(X,Y) \le 1$ for all pairs of r.v.'s (X,Y).

<u>Pf</u>. See book, p.

<u>Covariance properties</u>:

(a) $Cov(X,X) = Var(X)$

(b) $Cov(X,Y) = Cov(Y,X)$ (symmetry)

(c) $Cov(aX+b, cY+d) = ac\ Cov(X,Y)$

(d) $Cov\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j\ Cov(X_i, Y_j)$

Combining (a), (b) and (d) we get a general formula for the variance of a sum of two (not necessarily independent!) r.v.'s:

$$\boxed{Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)}$$

Pf.
$$\begin{aligned} Var(X+Y) &= Cov(X+Y, X+Y) && \text{by (a)} \\ &= Cov(X,X) + Cov(X,Y) + Cov(Y,X) + Cov(Y,Y) && \text{by (d)} \\ &= Var(X) + 2\,Cov(X,Y) + Var(Y) && \text{by (b) \& (a)} \end{aligned}$$

Note: If X and Y are indep., then $Cov(X,Y) = 0$ and we get, again,

$$Var(X+Y) = Var(X) + Var(Y).$$

Suggested exercises:

§5.1: 1, 7
§5.2: 1, 5
§5.3: 2, 4
§5.4: 8, 9
§6.4: 4, 15^{a-c}