Probability 2 - Notes 9

Independence

Definition. Two jointly continuous random variables X and Y are said to be independent if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x,y.

It is easy to show that X and Y are independent iff any event for X and any event for Y are independent, i.e. for any measurable sets A and B $P\{(X \in A) \cap (Y \in B)\} = P(X \in A)P(Y \in B)$.

Note X and Y cannot be independent if their ranges are dependent. Independence of X and Y requires the support of the joint p.d.f. $f_{X,Y}$ to be just the Cartesian product of the support of f_X and the support of f_Y.

Theorem 1. *X and Y are independent iff $f_{X,Y}(x,y) = g(x)h(y)$ for all x,y for some functions g and h.*

Proof. If X and Y are independent then you need only take $g(x) = f_X(x)$ and $h(y) = f_Y(y)$.

If $f_{X,Y}(x,y) = g(x)h(y)$ then $f_X(x) = \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)H$, where $H = \int_{-\infty}^{\infty} h(y)dy$. Similarly $f_Y(y) = h(y)G$, where $G = \int_{-\infty}^{\infty} g(x)dx$. Since the marginal p.d.f. integrates to one you also have $HG = 1$. Therefore

$$f_X(x)f_Y(y) = g(x)Hh(y)G = g(x)h(y) = f_{X,Y}(x,y)$$

for all x,y. Hence X and Y are independent. $\square$

Note When $f_{X,Y}(x,y) = g(x)h(y)$ for all x,y you can easily write down the marginal p.d.f.'s. $f_X(x) = Cg(x)$ and $f_Y(y) = \frac{1}{C}h(y)$ for a suitable choice of C. You can find C by noting that the marginal p.d.f. integrates to one.

Examples

1. $f_{X,Y}(x,y) = 6x$ for $0 < x < y < 1$. X and Y are not independent since the ranges are dependent.

2. $f_{X,Y}(x,y) = \frac{3}{4} + xy$ for $0 < x < 1$ and $0 < y < 1$. In this case the ranges are not dependent but the joint p.d.f. cannot be written in the form $g(x)h(y)$ for any functions g and h. Hence X and Y are not independent.

3. $f_{X,Y}(x,y) = 2x$ for $0 < x < 1$ and $0 < y < 1$. X and Y are independent since the ranges are not dependent and $f_{X,Y}(x,y) = g(x)h(y)$ where we can choose $g(x) = Cx$ and $h(y) = \frac{2}{C}$. It is easy to see that if we set $C = 2$ then $f_X(x) = g(x) = 2x$ for $0 < x < 1$ and $f_Y(y) = h(y) = 1$ for $0 < y < 1$.

Expectation and measures over the joint distribution

In the sequel, the following important formula shall be used (no proof will be given). If a function $g(x,y$ is such that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|g(x,y)|f_{X,Y}(x,y)dxdy < \infty$ then

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dx\right]dy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dy\right]dx.$$

Results when X and Y are independent

Theorem 2. *If X and Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for any (suitably integrable) functions g and h.*

Proof.

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dx\right)dy = \int_{-\infty}^{\infty} h(y)f_Y(y)\left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)dy \\ &= E[g(X)]\int_{-\infty}^{\infty} h(y)f_Y(y)dy = E[g(X)]E[h(Y)] \quad \square \end{aligned}$$

Corollary. If X and Y are independent, $Z = X+Y$ then $M_Z(t) = M_X(t)M_Y(t)$. Indeed,

$$M_Z(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right] = E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)M_Y(t).$$

Example. If X and Y are independent with $X \sim Gamma(\theta,\alpha)$ and $Y \sim Gamma(\theta,\beta)$ and $U = X+Y$, then

$$M_U(t) = M_X(t)M_Y(t) = \left(1-\frac{t}{\theta}\right)^{-\alpha}\left(1-\frac{t}{\theta}\right)^{-\beta} = \left(1-\frac{t}{\theta}\right)^{-(\alpha+\beta)}$$

This is the m.g.f. of a $Gamma(\theta,\alpha+\beta)$. Hence from the uniqueness of the m.g.f., $U \sim Gamma(\theta,\alpha+\beta)$.

Joint Measures

The joint measure which is commonly used is the covariance $Cov(X,Y) = E[(X-E[X])(Y-E[Y])] \equiv E[XY]-E[X]E[Y]$. The dimensionless form (invariant to shift and positive scaling of X and/or Y) is the coefficient of correlation

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

When X and Y are independent $Cov(X,Y) = E[XY]-E[X]E[Y] = E[X]E[Y]-E[X]E[Y] = 0$. Hence independence implies covariance (and so correlation) zero. However it is not true that correlation zero implies independence.

Example. Let $X \sim U(-1,1)$ and $Y = X^2$. Then it is easily shown that $E[X] = 0$, $E[Y] = E[X^2] = \frac{1}{3}$ and $E[XY] = E[X^3] = 0$. Therefore $Cov(X,Y) = 0$. But clearly X and Y are not independent but have an exact relationship. The value of X completely determines the value of Y.

The correlation coefficient measures the degree of linear association. In the example there was no linear relation. X did not tend to increase as Y increased (positive correlation) nor did X tend to decrease as Y increased (negative correlation).

Example. $f_{X,Y}(x,y) = 2$ for $x > 0$, $y > 0$ and $x+y < 1$. Then $f_X(x) = 2(1-x)$ for $0 < x < 1$ and it is simple to show that $E[X] = \frac{1}{3}$, $E[X^2] = \frac{1}{6}$ and hence $Var(X) = \frac{1}{18}$. Also $f_Y(y) = 2(1-y)$ for $0 < y < 1$, so Y has the same marginal distribution as X. Then $E[Y] = \frac{1}{3}$ and $Var(Y) = \frac{1}{18}$.

$$E[XY] = \int_0^1 \left(\int_0^{1-y} 2xydx \right) dy = \int_0^1 y(1-y)^2 dy = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{1}{12}$$

Therefore $Cov(X,Y) = \frac{1}{12} - \frac{1}{9} = \frac{-1}{36}$. Hence $\rho(X,Y) = -\frac{1}{2}$.

Expectation, variance and covariance for linear functions of X and Y.

In Probability 1 you showed that $E[aX+bY+c] = aE[X]+bE[Y]+c$ and $Var(aX+bY+c) = a^2Var(X)+b^2Var(Y)+2abCov(X,Y)$. it is simple to obtain a similar result for the covariance of two linear functions of X and Y. Let $U = aX+bY+e$ and $V = cX+dY+f$. Then

$$\begin{aligned} Cov(U,V) &= E[((aX+bY+e)-(aE[X]+bE[Y]+e))((cX+dY+f)-(cE[X]+dE[Y]+f))] \\ &= E[(a(X-E[X])+b(Y-E[Y]))(c(X-E[X])+d(Y-E[Y]))] \\ &= E[ac(X-E[X])^2+bd(Y-E[Y])^2+(ad+bc)(X-E[X])(Y-E[Y])] \\ &= acVar(X)+bdVar(Y)+(ad+bc)Cov(X,Y) \end{aligned}$$

Theorem 3. *Provided $Var(X) > 0$ and $Var(Y) > 0$, $-1 \le \rho(X,Y) \le 1$.*

Proof. Set $\xi = \frac{X-E(X)}{\sqrt{Var(X)}}$ and $\eta = \frac{Y-E(Y)}{\sqrt{Var(Y)}}$. Note that then

(a) $E(\xi^2) = E(\eta^2) = 1$. Indeed $E(\xi^2) = E\left[\frac{(X-E(X))^2}{Var(X)}\right] = \frac{E[(X-E(X))^2]}{Var(X)} = 1$. The equality for η is proved similarly.

(b) $E(\xi\eta) = E\left[\frac{(X-E(X))(Y-E(Y))}{\sqrt{Var(X)Var(Y)}}\right] = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \rho(X,Y)$.

Consider now $E((\xi-\eta)^2) = E(\xi^2-2\xi\eta+\eta^2) = E(\xi^2)-2E(\xi\eta)+E(\eta^2) = 2-2\rho(X,Y)$.
Hence $2-2\rho(X,Y) \ge 0$ and $\rho(X,Y) \le 1$.
Similarly, $E((\xi+\eta)^2) = E(\xi^2+2\xi\eta+\eta^2) = 2+2\rho(X,Y) \ge 0$ and thus $\rho(X,Y) \ge -1$. □

Note. If $\rho(X,Y) = 1$, then $E((\xi-\eta)^2) = 2-2\rho(X,Y) = 0$ and thus $\xi = \eta$. In other words, $\frac{X-E(X)}{\sqrt{Var(X)}} = \frac{Y-E(Y)}{\sqrt{Var(Y)}}$. We can rewrite this as $Y = aX+b$, where $a = \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}$, $b = E(Y)-aE(X)$.
So there is an exact linear relation between X and Y (with positive a).
Similarly, if $\rho(X,Y) = -1$ then $\frac{X-E(X)}{\sqrt{Var(X)}} = -\frac{Y-E(Y)}{\sqrt{Var(Y)}}$ and $Y = aX+b$, where $a = -\frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}$, $b = E(Y)-aE(X)$. There is again an exact linear relation between X and Y (with negative a).

Note. In lectures when we considered X and Y having a trinomial distribution with parameters n, p and θ, we showed that $\rho(X,Y)$ was negative. The extreme case where $p+\theta = 1$ corresponded to $X+Y = n$, i.e. $Y = X-n$. In this case $\rho(X,Y) = -1$. This corresponded to an exact linear relation for Y in terms of X, where the coefficient of X was negative.