CHAPTER 7. JOINTLY DISTRIBUTED RANDOM VARIABLES

7.1 Introduction

We have so far focused on the description of one random variable at a time. Even when we had several random variables to work with, they were implicity taken to be independent of each other. In many situations of interest, a group of random variables may vary together, that is, they may exhibit dependence – whether associative or causal. The joint probabilistic behaviour need to be described in such cases. We will also formally look at the concept of independence among random variables.

7.2 Joint probability description

We start with the joint cumulative distribution function (JCDF) of two random variables *X* and *Y*. It is given by the probability:

$$F_{X,Y}(x,y) = P[X \le x, Y \le y] \tag{7.1}$$

It must be a monotone function taking values between 0 and 1. In the discrete case, it is given by the sum of the joint probabiliy mass function (JPMF):

$$F_{X,Y}(x,y) = P[X \le x, Y \le y] = \sum_{y_j \le y} \sum_{x_i \le x} p_{X,Y}(x_i, y_j) \tag{7.2}$$

while in the continuous case, it is given by the integration of the joint probability density function (JPDF), $f_{X,Y}(x,y)$,

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\,du\,dv \tag{7.3}$$

The JPMF is a non-negative function and sums to one. Its interpretation of the JPMF is as in the one variable case:

$$p_{X,Y}(x_i, y_j) = P[X = x_i, Y = y_j] \tag{7.4}$$

Likewise, the joint probability density function of two continuous random variables non-negative, contains a volume of unity under it, and is interpreted as:

$$f_{x,y}(x,y)\ \Delta x \Delta y = P[X \in (x, x+\Delta x) \cap Y \in (y, y+\Delta y)] \quad (7.5)$$

The probability content of a region *A* can be given by:

$$P[(x,y)\in A]=\begin{cases}\sum_{all\,y}\sum_{all\,x} I_A p_{X,Y}(x,y), & \text{discrete case}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_A f_{X,Y}(x,y)dx\,dy, & \text{continuous case}\end{cases} \tag{7.6}$$

where $\mathrm{I_A}=\begin{cases}1, & \text{if } (x,y)\in A\\ 0, & \text{otherwise}\end{cases}$

Since the treatments are similar, we focus on joint *continuous* random variables in this chapter, rather than the discrete case.

Recall, $f_X(x)=\dfrac{dF_{X(x)}}{dx}$. Similarly, the JPDF is the mixed partial derivative,

$$f_{X,Y}(x,y)=\frac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y) \tag{7.7}$$

When sevaral random variable are present, the word “marginal” is often used to denote the distribution (or density or mass function) of any one of them. The marginal distribution of X is thus the same as the distribution of X.

7.3 Characteristic fn of jointly distributed random variables

The characteristic function of jointly distributed RVs (or order n) is the function:

$$M_{X_1,X_2,\dots\dots,X_n}(\theta_1,\theta_2,\dots..,\theta_n)=E\left[\exp\left[i\theta_1X_1+i\theta_2X_2+\dots...\right]\right] \tag{7.8}$$

If $X_1,X_2,\dots...,X_n$ are independent, then the joint CF becomes the product of marginal CFs:

$$M_{X_1,X_2,\dots...,X_n}(\theta_1,\theta_2,\dots.,\theta_n)=M_{X_1}(\theta_1)M_{X_2}(\theta_2)\dots.M_{Xn}(\theta_n) \tag{7.9}$$

Consider the sum of 2 independent random variables:

$$Z=X_1+X_2,\ X_1\perp X_2 \tag{7.10}$$

The CF of Z is:

$$\begin{aligned}M_Z(\theta)&=M_{X_1+X_2}(\theta)=E\left[e^{i\theta Z}\right]=Ee^{i\theta(X_1+X_2)}=E\left[e^{i\theta X_1}e^{i\theta X_2}\right]\\ &=E\left[e^{i\theta X_1}\right]E\left[e^{i\theta X_2}\right]=M_{X_1}(\theta)M_{X_2}(\theta)\end{aligned} \tag{7.11}$$

Example, if X_1 and X_2 are each Normal with mean and variance μ_1, σ_1^2 and μ_2, σ_2^2 respectively, then it can be easily shown through the CF that Z too is Normal with mean and variance $\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2$.

In case of jointly distributed random variables, the covariance function can be obtained from:

$$E\left(X_1 X_2\right) = \frac{\partial^2}{\partial\theta_1 \partial\theta_2} M_{X1,X2}(0,0) \tag{7.12}$$

or equivalently in log space as:

$$\operatorname{cov}\left(X_{1,} X_2\right) = \frac{1}{i^2} \frac{\partial^2}{\partial\theta_1 \partial\theta_2} \ln M_{X1,X2}(0,0) \tag{7.13}$$

7.4 Joint moments

Define joint moments

7.4.1 MGF for joint RVs

The moment generating function of the jointly distributed random variables X and Y is given by:

$$g_{X,Y}(s_1, s_2) = E\left[e^{s_1 X + s_2 Y}\right] \tag{7.14}$$

so that the joint moments are recovered as:

$$E\left[X^m Y^n\right] = \frac{\partial^{m+n}}{\partial s_1{}^m \partial s_2{}^n} g_{X,Y}(s_1, s_2) \text{ at } s_1 = 0, s_2 = 0 \tag{7.15}$$

Show that the covariance between X and Y can be given by:

$$\operatorname{cov}(X,Y) = \frac{\partial^2}{\partial s_1 \partial s_2} \ln g_{X,Y}(s_1, s_2) \text{ at } s_1 = 0, s_2 = 0$$

Show that if X and Y are independent:

$$g_{X,Y}(s_1 s_2) = g_X(s_1) g_Y(s_2)$$

Show that if X and Y are independent and Z = X + Y,

$$g_Z(s) = g_X(s) g_Y(s)$$

.

7.5 Marginal probability description

The marginal PDF of X can be recovered from the JPDF using the theorem of total probability:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \qquad (7.16)$$

There is no difference between marginal pdf & individual or ordinary pdf. The marginal CDF of X is the joint CDF evaluated at the right end point of Y:

$$F_X(x) = F_{X,Y}(x,\infty) \qquad (7.17)$$

7.6 Conditional probability descriptions

7.6.1.1 Conditional pmf

The conditional PMF of X given Y has taken a particular value is:

$$p_{X|Y=y}(x,y) = P[X = x \mid Y = y] = \frac{p_{X,Y}(x,y)}{p_Y(y)} \qquad (7.18)$$

The JPMF can be written in terms of the conditional and marginal PMFs as:

$$p_{X,Y}(x,y) = p_{X|Y=y}(x,y)\,p_Y(y) = p_{Y|X=x}(y,x)\,p_X(x) \qquad (7.19)$$

7.6.1.2 Conditional pdf

The conditional PDF of X given a particular realization of Y is,

$$f_{X/Y=y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad (7.20)$$

The explanation can be given as:

$$\begin{aligned} f_{X|Y=y}(x,y)dx &\approx P[X \in (x,x+dx) \mid Y \in (y,y+dy)] \\ &= \frac{P[X \in (x,x+dx), y \in (y,y+dy)]}{P[Y \in (y,y+dy)]} \\ &= \frac{f_{X,Y}(x,y)dxdy}{f_Y(y)dy} \end{aligned}$$

The joint PDF can be given as the product of the conditonal and the corresponding marginal:

$$f_{X,Y}(x,y) = f_{X|Y=y}(x,y)\,f_Y(y) = f_{Y|X=x}(y,x)\,f_X(x) \qquad (7.21)$$

7.6.1.3 Conditional cdf

$$F_{X/Y=y}(x,y) = \begin{cases} \sum_{x_i \le x} p_{X/Y=y}(x_i,y) & \text{for discrete } X \\ \int_{-\infty}^{x} f_{X|Y=y}(u,y)du & \text{for continuous } X \end{cases}$$

In the continuous case, the conditional CDF becomes

$$F_{X/Y=y}(x,y)=\int_{-\infty}^{x}\frac{f_{X,Y}(u,y)}{f_Y(y)}du=\frac{1}{f_Y(y)}\int_{-\infty}^{x}f_{X,Y}(u,y)\,du$$

$$\begin{aligned}&\text{Note}: F_{XY}(x,y)\neq F_{X|Y=y}(x,y)F_Y(y)\\ &\because P(X\leq x, Y\leq y)=P(X\leq x\,|\,Y\leq y)P(Y\leq y)\\ &\text{and } P(X\leq x\,|\,Y\leq y)\neq F_{X|Y=y}(x,y)\end{aligned}$$

7.7 Independence

If *X* and *Y* are independent, the conditional distribution (or density or mass) of one is identical to its marginal. Equivalently, the joint distribution (or density or mass) is the product of the marginals. Each of these is both a necessary and sufficient condition for independence of X and Y.

$$X \text{ is independent of } Y \Leftrightarrow \begin{cases} F_{X|Y=y}(x,y)=F_X(x) \text{ for all } x,y \text{ (continuous or discrete)}\\ F_{X,Y}(x,y)=F_X(x)F_Y(y) \text{ for all } x,y \text{ (continuous or discrete)}\\ p_{X|Y=y}(x,y)=p_X(x) \text{ for all } x,y \text{ (discrete)}\\ f_{X,Y}(x,y)=f_X(x)f_Y(y) \text{ for all } x,y \text{ (continuous)}\end{cases}$$

(7.22)

If X and Y are independent, so are $g(X)$ and $h(Y)$

More from Papoulis p. 184.

7.7.1.1 Generalization to *n* dimensions

A vector $\{X_t, t\epsilon T\}$ of random variables is mutually independent iff for all subsets *J* of *T*, the joint CDF is the product of the marginal CDFs:

$$F_J(x_t, t\in J)=\prod_{t\in J}P[X_t\leq x_t] \tag{7.23}$$

7.8 Conditional moments

7.8.1 Conditional mean

The expecion of X given $Y=y$ is:

$$E[X \mid Y=y] = \mu_{X|Y=y} = \begin{cases} \sum_{all x} x p_{X|Y=y}(x,y) \text{ for discrete X} \\ \int_{-\infty}^{\infty} x f_{X|Y=y}(x,y)dx \text{ for continuous X} \end{cases} \tag{7.24}$$

This can be generalized to conditioning by any event A:

$$E[X \mid A] = \mu_{X|A} = \begin{cases} \sum_{all x} x p_{X|A}(x) \quad \text{for discrete X} \\ \int_{-\infty}^{\infty} x f_{X|A}(x)dx \quad \text{for continuous X} \end{cases} \tag{7.25}$$

Example: Find the conditional mean of the random variable *X* given that it is less than the number *a*:

$$E[X \mid X < a] = \frac{\int_{-\infty}^{a} x f_X(x)dx}{\int_{-\infty}^{a} f_X(x)dx} \tag{7.26}$$

The unconditional mean of *X* can be recovered from the conditional mean provided sufficient information is available:

$$\mu_X = \begin{cases} \sum_{\text{all } y} \mu_{X|Y=y} p_Y(y) \\ \int_{-\infty}^{\infty} \mu_{X|Y=y} f_Y(y)dy \end{cases} \tag{7.27}$$

If X and Y are independent, then $\mu_{X|Y} \equiv \mu_X$.

7.8.2 Conditional variance

The conditional variance of *X* given $Y = y$ is :

$\sigma^2{}_{X|Y=y} = E[(X - \mu_{X|Y=y})^2 \mid Y = y]$

When X is continuous:

$$= \int_{-\infty}^{\infty} (x - \mu_{X|Y=y})^2 f_{X|y=y}(x,y)dx$$

$$= \int_{-\infty}^{\infty} x^2 f_{X|Y=y}(x,y)dx - (\mu_{X|Y=y})^2$$

When X is discrete :

$$= \sum_{\text{all } x_i} \left(x_i - \mu_{X|Y=y} \right)^2 p_{X|y=y}(x_i, y)$$

$$= \sum_{\text{all } x_i} \left(x_i \right)^2 p_{X|y=y}(x_i, y) - \left(\mu_{X|Y=y} \right)^2$$

Unconditional variance in terms of conditional variance and conditional mean (Ross p 118)

$$\operatorname{var}(X) = E(\operatorname{var}(X \mid Y)) + \operatorname{var}(E(X \mid Y)) \qquad (7.28)$$

Proof:

$$\operatorname{var}(X) = \int (x - \mu_X)^2 f_X(x)\, dx$$

$$= \int_y \int_x (x - \mu_X)^2 f_{X,Y}(x, y)\, dx\, dy$$

$$= \int \int (x - \mu_X)^2 f_{X|Y=y}(x, y)\, dx\, f_y(y)\, dy$$

We can write $x - \mu_X = \left(x - \mu_{X|Y=y} \right) + \left(\mu_{X|Y=y} - \mu_X \right)$

Thus $\operatorname{var}(X)$ can be expressed as the sum of three integrals.

1st integral

$$= \int \int \left(x - \mu_{X|Y=y} \right)^2 f_{X|Y=y}(x, y)\, dx\;\; f_Y(y)\, dy$$

$$= \int \operatorname{var}\left(X \mid Y = y \right) f_Y(y)\, dy$$

$$= E\left(\operatorname{var}\left(X \mid Y \right) \right).$$

2nd integral

$$= 2 \int \int \left(x\mu_{X|Y=y} - x\mu_X - \mu_{X|Y=y}^2 + \mu_X \mu_{X|Y=y} \right) f_{X|Y}(x|y)\, dx\; f_Y(y)\, dy$$

$$= 2 \int \mu_{X|Y=y}\, \mu_{X|Y=y}\; f_Y(y)\, dy - 2 \int \mu_X \mu_{X|Y=y}\; f_Y(y)\, dy$$

$$-2 \int \mu_{X|Y=y}^2\; f_Y(y)\, dy + 2 \int \mu_X \mu_{X|Y=y}\; f_Y(y)\, dy$$

$= 0$ (first and third terms, and second and fourth terms, cancel each other)

3rd integral

$$= \int \int \left(\mu_{X|Y=y} - \mu_X \right)^2 f_{X|Y=y}(x, y)\, dx\;\; f_Y(y)\, dy$$

$= \int \left(\mu_{X|Y=y} - \mu_X \right)^2\; f_Y(y)\, dy$ (no x dependance in the squared term)

By definition of "variance" the 3rd integral is the variance of $\mu_{X|Y=y}$ which we denote as $\operatorname{var}\left(E\left(X|Y\right)\right)$.

Thus:

$$\operatorname{var}\left(X\right)=E\left(\operatorname{var}\left(X|Y\right)\right)+\operatorname{var}\left(E\left(X|Y\right)\right)$$

7.8.2.1 Example: sum of random variables

Find the mean and variance of the sum, *S*, of a random number (*N*) of iid RVs (X_i). *N* is independent of each X_i.

$$S=\sum_{i=1}^{N}X_i$$

Answer:

$$E(S)=\mu_N\mu_X$$
$$\operatorname{var}(S)=\sigma_X{}^2\mu_N+\mu_X{}^2\sigma_N{}^2$$

7.8.2.2 Example: joint CDF

$$F_{\underline{X}}(x_1,x_2)=1-e^{-x_1}-e^{-x_2}+e(-x_1-x_2-x_1x_2)$$

Find conditional density and conditional moments of X_1.

7.8.2.3 Example: proof loading

The theoretical strength of a beam is exponentially distributed with mean 10 kNm. A moment of 12 kNm is applied on the beam (at the critical cross section) and it is found to survive without damage. What can you now say about the mean and standard deviation of the beam's strength?

7.8.2.4 Example: flaws in a weld

Flaws of random size and number exist in a weld of length 1 m. Assume that the size of each flaw is Exponentially distributed with mean 2 mm, and the flaw sizes are independent of one another. The number of flaws in the weld is Poisson distributed with rate one every 15 cm.

(a) What is the probability that there are exactly 10 flaws in the weld?

(b) An ultrasonic test is used to detect flaws in the weld. The instrument has a resolution of 3 mm (i.e., it can detect weld flaws if they are 3mm or larger).

(i) What is the probability that no flaw will be detected?

(ii) If no flaw is detected, what is the probability that the weld is flawless?

7.9 Measure of dependence

The variance of a random variable can be expressed as $E[(X-\mu_X)(X-\mu_X)]$. In the same vein, the covariance between two random variables is defined as:

$$\text{Covariance} = \sigma_{XY}=E[(X-\mu_X)(Y-\mu_y)]$$

$$=E(XY) - E(X)E(Y)$$

$$= \int_{x-\infty}^{\infty} \int_{y-\infty}^{\infty} xy f_{X,Y}(x,y)dxdy - \mu_X \mu_Y$$

The correlation coefficient is:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

7.9.1 Properties of correlation coefficient

7.9.1.1 <u>Linearity and independence</u>

a) $|\rho_{XY}| \to 1 \Rightarrow$ high linear dependence between X and Y

b) $|\rho_{XY}| \to 0 \Rightarrow$ no linear dependence between X and Y

c) If X,Y are independent, $E(XY) = E(X)E(Y)$ and $\sigma_{XY} = 0$ i.e., $\rho_{XY} = 0$. The converse is not necessarily true.

Example:

Ex:

Let $Z \sim U(0,1)$. Define: $X = \sin 2\pi z,\ Y = \cos 2\pi z$. The means of X and Y are both zero:

$$E(X) = \int_0^1 \sin 2\pi z(1)dz = 0$$

$$E(Y) = \int_0^1 \cos 2\pi z(1)dz = 0$$

Further,

$$E(XY) = \int_0^1 \sin 2\pi z \cos 2\pi z(1)dz = \int_0^1 \frac{1}{2}(\sin 0 + \sin 4\pi z)dz = 0$$

Thus the correlation coefficient between X and Y $\rho = \dfrac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} = 0$

But, it is also true that, $X^2 + Y^2 = 1$ implying a definite dependence which the correlation coefficient is unable to capture.

7.9.1.2 <u>Linear transformation</u>

ρ is not affected by a linear transformation.

Proof:

Let $\rho_{UV} = \rho$

Define $X = aU + b, \quad Y = cV + d.$

$$E(X) = a\mu_U + b, E(Y) = c\mu_V + d, \sigma_X = a\sigma_U, \sigma_Y = c\sigma_V$$
$$E(XY) = E(acUV + adU + bcV + bd) = acE(UV) + ad\mu_U + bc\mu_V + bd$$
$$E(XY) - E(X)E(Y) = acE(UV) + ad\mu_U + bc\mu_V + bd - (a\mu_U + b)(c\mu_V + d)$$
$$= acE(UV) - ac\mu_U\mu_V$$
$$\therefore \text{cov}(X,Y) = ac\,\text{cov}(U,V)$$
$$\therefore \rho_{XY} = \frac{COV(XY)}{\sigma_X\sigma_Y} = \frac{ac\,\text{cov}(U,V)}{a\sigma_U b\sigma_V} = \frac{\text{cov}(U,V)}{\sigma_U\sigma_V} = \rho_{UV}$$

7.9.1.3 <u>Bounds on correlation coefficient</u>

The absolute value of the correlation coefficient is less than or equal to one.

$$|\rho_{XY}| \le 1 \tag{7.29}$$

Proof:

Define two normalized random variables, $X_0 = \dfrac{X - \mu_X}{\sigma_X}, Y_0 = \dfrac{Y - \mu_Y}{\sigma_Y}$. Their mean, and variance are: $\mu_{X_0} = \mu_{Y_0} = 0, \sigma_{X_0} = 1, \sigma_{Y_0} = 1$.

The correlation coefficient between X_0 and Y_0 is:

$$\rho_{X_0Y_0} = E\left[\frac{(X - \mu_X)(Y - \mu_Y)}{\sigma_X\sigma_Y}\right] = \frac{1}{\sigma_X\sigma_Y}\text{cov}(X,Y) = \frac{1}{\sigma_X\sigma_Y}\rho_{XY}\sigma_X\sigma_Y = \rho_{XY}$$

The variance of their sum,

$\text{var}(X_0 + Y_0) = \text{var}(X_0) + \text{var}(Y_0) + 2\sigma_{X_0}\sigma_{Y_0}\,\rho_{X_0Y_0} = 1 + 1 + 2\rho_{XY} = 2(1 + \rho_{XY})$ must be non-negative. Therefore:

$$1 + \rho_{XY} \ge 0 \tag{7.30}$$

Similarly, the variance of their difference,

$\text{var}(X_0 - Y_0) = 1 + 1 - 2\,\text{cov}(X_0, Y_0) = 2 - 2\rho_{XY} = 2(1 - \rho_{XY})$ must be non-negative as well, giving us:

$$1 - \rho_{XY} \ge 0 \Rightarrow \rho_{XY} \ge -1 \tag{7.31}$$

7.9.2 Limitations of the correlation coefficient

1) σ_1, σ_2 must be finite

2) ρ is not invariant under monotone transform

3) $\rho = 0 \not\Rightarrow$ independence

7.9.3 Desirable features of a correlation measure

Universal existence

invariance under monotone transformation

Zero $\Leftrightarrow$ independence

The following measures satisfy all three requirements

7.9.3.1 The sup correlation (or maximal correlation)

$\bar{\rho}(X_1, X_2) = \sup \rho(g_1(X_1), g_2(X_2))$

Where sup is taken over all Borel measureable functions $g_1(X_1), g_2(X_2)$ such that g_1, g_2 have finite positive variance, and ρ is the ordinary corr. coeff.

7.9.3.2 The monotone correlation

Same as $\bar{\rho}$ but sup is taken over all monotone functions $g_{1,}g_2$ only .

Relationship

$$|\rho| \le \rho^* \le \bar{\rho} \le 1$$

All 3 are equal for multivariate Normal rv's.

7.9.4 Examples: jointly distributed random variables

7.9.4.1 Example: Joint PMF

You and your friend go into a sports bar where a dart throwing competition is going on. You buy *m* darts at 1 rupee each. You throw these *m* darts at the board. Of these, *N* darts hit within the inner circle. Your friend picks up these *N* darts, and throws them at the board. *X* of them hit the inner circle. You and your friend earn 10 rupees for each of the *X* hits. Of course, *N* and *X* are random numbers. Assume that your throws are independent and each has a probability p_1 of hitting the inner circle. Your friend's throws are also independent, and each has a probability p_2 of hitting the inner circle.

a) What is the distribution of *N*?

b) What is the distribution of *X*?

c) How much do you expect to earn from this game?

d) Say, $p_1 > p_2$. Does it matter who goes first?

Answer:

Clearly, *N* is a binomial random variable with parameters *m* and p_1. Given $N = n$, *X* too is Binomial:

$$p_{X|N=n}(x;n) = P\left[X = x \mid N = n\right] = \binom{n}{x} p_2^x (1-p_2)^{n-x} \tag{7.32}$$

The unconditional PMF of *X* can be found by theorem of total probability:

$$p_X(x) = \sum_{\text{all}\, n} p_{X|N=n}(x;n)\, p_N(n) = \sum_{n=0}^{m} \tbinom{n}{x} p_2^x (1-p_2)^{n-x} \tbinom{m}{n} p_1^n (1-p_1)^{m-n} I(x \le n)$$

where the indicator function ensures that your friend can never have more successes than the darts you win. The PMF of X can be written as:

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

.

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

Substituting $v = n - x$ so that $n = x \Rightarrow v = 0,$ and $n = m \Rightarrow v = m - x$ allows us to rewrite the above summation as:

$$p_X(x) = \frac{m!}{x!} p_2^x \sum_{v=0}^{m-x} \frac{(1-p_2)^v p_1^{v+x} (1-p_1)^{m-v-x}}{v!(m-x-v)!}$$

Rerranging the terms,

$$p_X(x) = \frac{m!}{x!} p_2{}^x p_1{}^x \frac{1}{(m-x)!} \sum_{v=0}^{m-x} \frac{(m-x)!}{v!(m-x-v)!} (1-p_2)^v p_1^v (1-p_1)^{m-x-v}$$

and substituting $m' = m - x$, we obtain

$$p_X(x) = \frac{m!}{x!(m-x)!} (p_1 p_2)^x \sum_{v=0}^{m'} \binom{m'}{v} [p_1 - p_1 p_2]^v [1-p_1]^{m'-v}$$

$$= \binom{m}{x} (p_1 p_2)^x (1-p_1 p_2)^{m'} \sum_{v=0}^{m'} \binom{m'}{v} \frac{[p_1 - p_1 p_2]^v}{[1-p_1 p_2]^v} \frac{[1-p_1]^{m'-v}}{[1-p_1 p_2]^{m'-v}}$$

$$= \binom{m}{x} (p_1 p_2)^x (1-p_1 p_2)^{m-x} \sum_{v=0}^{m'} \binom{m'}{v} \left[\frac{p_1 - p_1 p_2}{1-p_1 p_2}\right]^v \left[\frac{1-p_1}{1-p_1 p_2}\right]^{m'-v}$$

Using the binomial identity, we replace the sum and obtain:

$$p_X(x) = \binom{m}{x} (p_1 p_2)^x (1-p_1 p_2)^{m-x} \left[\frac{p_1 - p_1 p_2}{1-p_1 p_2} + \frac{1-p_1}{1-p_1 p_2}\right]^{m'}$$

Since the two terms in the square brackets sum to one, the unconditional PMF of *X* simplifies to:

$$p_X(x) = \binom{m}{x} (p_1 p_2)^x (1-p_1 p_2)^{m-x}$$

Thus, *X* is Binomial with parameters $(m, p_1 p_2)$.

The expected earning is:

$$E[\text{earning}] = -m \times 1 + 10 \times E[X] = -m + 10 m p_1 p_2 = (10 p_1 p_2 - 1) m \tag{7.33}$$

Since the solution is symmetric in p_1 and p_2, it does not matter who goes first.

7.9.4.2 Example: marginal densities from joint PDF

Given the joint density function,

$$f_{X,Y}(x,y)=\begin{cases}(x+y)/3, & 0<x<1, 0<y<2\\ 0, & \text{elsewhere}\end{cases}$$

find the marginal density functions of X and Y.

Answer:

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy=\int_0^2 \frac{1}{3}(x+y)dy,\ \ 0<x<1$$

$$=\frac{1}{3}(xy+\frac{y^2}{2})_0^2=\frac{1}{3}(2x+\frac{2^2}{2}-0-0)=\frac{2}{3}(x+1),\ \ 0<x<1$$

Check $\int_{-\infty}^{\infty} f_X(x)dx=\int_0^1 \frac{2}{3}(x+1)dx=\frac{2}{3}(\frac{x^2}{2}+x)_0^1=\frac{2}{3}(\frac{1}{2}+1-0-0)=1$. OK.

$$f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx=\int_0^1 \frac{1}{3}(x+y)dx,\ 0<y<2$$

$$=\frac{1}{3}(x^2/2+xy)_0^1=\frac{1}{3}(y+\frac{1}{2}),\ \ 0<y<2$$

Check $\int_{-\infty}^{\infty} f_Y(y)dy=\int_0^2 \frac{1}{3}(y+\frac{1}{2})dy=\frac{1}{3}(y^2/2+y/2)_0^2=\frac{1}{6}[4+2-0-0]=1$. OK.

We can show that X and Y are not independent as $f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$ for $0<x<1, 0<y<2$. We can further show that $\sigma_{XY}=E[XY]-\mu_X\mu_Y=2/3-(5/9)(11/9)=-0.0123$.

7.9.4.3 Example: correlation coefficient from joint PDF

Given the joint density function of *X* and *Y*,

$$f_{XY}(x,y)=\begin{cases}8xy,\ 0\le x\le 1, 0\le y\le x\\ 0,\ \text{otherwise}\end{cases}$$

find the marginal densities and the correlation coefficient.

Answer:

$$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)dy=\int_0^x 8xy\,dy,\ 0\le x\le 1=8x\left[\frac{y^2}{2}\right]_0^x=4x^3,\ 0\le x\le 1$$

Cheek if $f_X\ge 0$ and $\int f_X dx=1$. Checked.

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)dx=\int_{\substack{0\\ y\le x}}^{1} 8xy\,dx$$

The variable upper limit on y can be taken care of by the Indicator function:

$$f_Y(y)=\int_0^1 8xy\ I(0\le y\le x)\ dx$$

The integral can now be split as:

$$f_Y(y)=\int_0^y 8xyI(0\le y\le x)dx+\int_y^1 8xyI(0\le y\le x)dx$$

$$=0+\int_y^1 8xy(1)\ dx=8y[x^2/2]_y^1=4y(1-y^2),\ 0\le y\le 1$$

Thus we find $f_{XY}(x,y)\neq f_X(x)f_Y(y)$ for some *x*,*y*.

So X and Y are not independent. Hence we find the correlation coefficient.

First, we determine the means of *X* and *Y*:

$$\mu_X=\int_0^1 x4x^3dx={}^4\!/_5,$$

$$\mu_Y=\int_0^1 y4y(1-y^2)=4(\frac{1}{3}-\frac{1}{5})={}^8\!/_{15}.$$

The variances are:

$$\sigma_X^2=\int_0^1 x^2 4x^3dx-(\frac{4}{5})^2=\frac{4}{6}x^6|_0^1-(\frac{4}{5})^2=2/75$$

$$\sigma_{Y^2}=\int_0^1 y^2 4y(1-y^2)dy-(\frac{8}{15})^2=\frac{4}{4}(y^4)_0^1-\frac{4}{6}|y^6|_0^1-(\frac{8}{15})^2=11/225$$

Finally, the expectation of the product,

$$E(XY)=\int_{y=0}^{1}\int_{x=y}^{1} xy8xy\,dx\,dy$$

$$=\int_0^1 8y^2\int_y^1 x^2dx\,dy=\int_0^1\frac{8}{3}y^2[x^3]_y^1dy=\int_0^1\frac{8}{3}y^2(1-y^3)dy=\frac{8}{3}[\frac{y^3}{3}-\frac{y^6}{6}]_0^1=4/9$$

yielding the covariance $\sigma_{XY}=\frac{4}{9}-\frac{4}{5}.\frac{8}{15}=4/225$

$$\Rightarrow \rho_{XY}=\frac{4/225}{\sqrt{2/75}\sqrt{11/225}}=0.4924$$

7.9.4.4 Example: effect of correlation coefficient under nonlinear monotonic transformation

Let us define two new random variables U and V by squaring X and Y respectively in Example 7.9.4.3. Since X and Y have non-zero probabilities only for positive values, this transformation is non-linear but monotonic. Does the correlation coefficient stay unchanged?

Answer:

We have, $U = X^2, V = Y^2$. Their means can be found from the first two moments of *X* and *Y*:

$$E(X^2) = \text{var}(X) + \mu_X{}^2 = \frac{2}{75} + \frac{16}{25} = \frac{2}{3}$$

$$E(Y^2) = \text{var}(Y) + \mu_Y{}^2 = \frac{11}{225} + \frac{64}{225} = \frac{1}{3}$$

The variance of *U* and *V* are:

$$\text{var}(U) = \text{var}(X^2) = E(X^4) - (E(X^2))^2 = \int_0^1 x^4 4x^3 dx - (2/3)^2 = 1/18$$

$$\text{var}(V) = E(Y^4) - (E(Y^2))^2 = \int_0^1 y^4 4y(1-y^2)dy - (1/3)^2 = 4[\frac{1}{6} - \frac{1}{8}] - 1/9 = 1/18$$

Finally, the mean of their product,

$$E(UV) = E(X^2Y^2) = \int_0^1\int_0^1 x^2y^2 8xy\,dx\,dy = \int_0^1 \frac{8}{4} y^3 [x^4]_y^1 dy = 2\int_0^1 y^3(1-y^4)dy = 2[\frac{y^4}{4} - \frac{y^8}{8}]_0^1 = \frac{1}{4}.$$

The correlation coefficient between U and V is then,

$$\rho_{UV} = \frac{E(UV) - E(U)E(U)}{\sigma_U \sigma_V} = \frac{1/4 - (2/3)\times(1/3)}{\sqrt{1/18}\sqrt{1/18}} = \frac{1}{2}$$

Thus, we find a slight change in the correlation coefficient between X and Y when they are squared. This is because ρ measures linear dependence between two random variables.

7.9.4.5 Example: conditional mean from joint PDF

Find the conditional mean of *X* and the conditional CDF of *X* given Y=$Y = \mu_Y$ in Example 7.9.4.3.

Answer:

We need to find the conditional density:

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

for which we need the marginal density of *Y* evaluated at its mean:

$f_Y(\mu_Y) = 4(\mu_Y)(1-\mu_Y{}^2) = 1.527$

which gives,

$$f_{X|Y=\mu_Y}(x) = \frac{f_{X,Y}(x,\mu_Y)}{f_Y(\mu_Y)} = \frac{8x\mu_Y}{1.527} = 2.795x, \qquad \frac{8}{15} < x < 1$$

Check that the conditional density integrates to one: $\int_{8/15}^{1} f_{X|Y=\mu Y}(x)dx = 1$.

The conditional mean of X is:

$$\mu_{X|Y=\mu_Y} = \int_{-\infty}^{\infty} x f_{X|Y=\mu_Y}(x)dx = \int_{8/15}^{1} x\ 2.795\ x\,dx = 2.795\,x^3\ |_{8/15}^{1} = 0.79$$

The conditional CDF is obtained by integrating the conditional density:

$$F_{X|Y=\mu_Y}(x) = \int_{-\infty}^{x} f_{X|Y=\mu_Y}(u)du$$

which needs to be described in three different regions:

$$F_{X|Y=\mu_Y}(x) = \begin{cases} 0, & x < 8/15 \\ \int_{8/15}^{x} 2.795u\,du = 1.3975(x^2 - .2844), & 8/15 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

7.9.4.6 Example: verifying independence from joint PDF

The joint density function of X and Y is:

$$f_{X,Y}(x,y) = \begin{cases} x(1+3y^2)/4, & 0 < x < 2,\ 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

a) Verify that the PDF integrates to one.

b) Find the probability that X is between 0 and 1 and Y is between 1/4 and 1/2.

c) Are X and Y independent?

Answer:

$$\int_{-\infty}^{\infty}\int f_{XY}(x,y)dydx = \int_0^1\int_0^2 \frac{x(1+3y^2)}{4}dxdy = \int_0^1 \left[\frac{x^2}{2}\frac{(1+3y^2)}{4}\right]_0^2 dy$$

$$= \int_0^1 \frac{1}{2}(1+3y^2)dy = \frac{1}{2}(y+y^3)_0^1 = \frac{1}{2}(1+1-0-0) = 1$$

Verified.

$P[0 < X < 1 \cap 1/4 < Y < 1/2]$

$$= \int_{1/4}^{1/2}\int_0^1 \frac{x(1+3y^2)}{4}\,dx\,dy = \int_{1/4}^{1/2} \left(\frac{1+3y^2}{4}.\frac{x^2}{2}/_0^1\right)dy = \int_{1/4}^{1/2}\frac{1}{8}.(1+3y^2)dy =$$

$$\frac{1}{8}[y+y^3]_{1/4}^{1/2} = \frac{1}{8}[\frac{1}{2}+\frac{1}{8}-\frac{1}{4}-\frac{1}{64}] = \frac{1}{8}\frac{32+8-16-1}{64} = \frac{23}{512}$$

In order to determine the independence between X and Y, we need their marginal density functions.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^1 \frac{x(1+3y^2)}{4}dy = x/2, \quad 0 < x < 2$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^2 \frac{x(1+3y^2)}{4}dx = \frac{1}{4}(1+3y^2)\frac{x^2}{2}\Big|_0^2 \qquad 0 < y < 1$$

$$= \frac{1}{2}(1+3y^2), \quad 0 < y < 1$$

We see that $f_{X,Y(x,y)} = f_X(x) f_Y(y)$ everywhere. Hence, in this example *X* and *Y* are independent.

7.10 Bayesian updating of distribution parameters

Let X be a continuous random variable with pdf f_X and cdf F_X. Let $\underline{\theta}$ be the parameters of the distribution. The type of this distribution may have been decided from analytical considerations or expert judgment.

Now, suppose the true value of $\underline{\theta}$ is unknown, and there is uncertainty about these parameters. It may then be useful to describe these parameters as random variables, i.e., as $\underline{\Theta}$, using our convention of upppercase letters for random variables. The earlier expressions, f_X and F_X, of the pdf and cdf of X, which had $\underline{\theta}$ as their parameters, now should be acknowledged as conditional on particular values of $\underline{\Theta} = \underline{\theta}$.

$$P[X \le x \mid \underline{\Theta} = \underline{\theta}] = F_{X|\underline{\Theta}}(x;\underline{\theta})$$
$$f_{X|\underline{\Theta}}(x;\underline{\theta}) = \frac{dF_{X|\underline{\Theta}}(x;\underline{\theta})}{dx} \tag{7.34}$$

Let the pdf of $\underline{\Theta}$ be $f_{\underline{\Theta}}$. It should be emphasized that $f_{\underline{\Theta}}$ depends on our current state of knowledge. We can do two things here. We can be happy about the state of affairs and get the unconditional distribution of X by the application of total probability. We can also try to improve our knowledge on $\underline{\Theta}$ by collecting more data. The second route takes us on a Bayesian updating of the distributions of X and $\underline{\Theta}$.

Suppose we make random samples on X, and end up with a set of observations $\underline{x}$. Can these observations help us improve our knowledge about $f_{\underline{\Theta}}$, and hence of f_X itself?

Based on the observations on X, the updated distribution of $\underline{\Theta}$ is:

$$f'_{\underline{\Theta}}(\underline{\theta}) = f_{\underline{\Theta}|\underline{X}=\underline{x}}(\underline{\theta};\underline{x}) = \frac{f_{\underline{X}=\underline{x}|\underline{\Theta}=\underline{\theta}}(\underline{x};\underline{\theta})\, f_{\underline{\Theta}}(\underline{\theta})}{f_{\underline{X}=\underline{x}}(\underline{x})} = \frac{f_{\underline{X}=\underline{x}|\underline{\Theta}=\underline{\theta}}(\underline{x};\underline{\theta})\, f_{\underline{\Theta}}(\underline{\theta})}{\int f_{\underline{X}=\underline{x}|\underline{\Theta}=\underline{\theta}}(\underline{x};\underline{\theta})\, f_{\underline{\Theta}}(\underline{\theta})d\underline{\theta}} \quad (7.35)$$

The superscript prime indicates the updated pdf of the parameters, $\underline{\Theta}$. For notational simplicity, let us rewrite Eq (7.35) as:

$$f'(\underline{\theta}) = f(\underline{\theta}\,|\,\underline{x}) = \frac{f(\underline{x}\,|\,\underline{\theta})\, f(\underline{\theta})}{f(\underline{x})} = \frac{f(\underline{x}\,|\,\underline{\theta})\, f(\underline{\theta})}{\int f(\underline{x}\,|\,\underline{\theta})\, f(\underline{\theta})d\underline{\theta}}$$

$$\text{i.e., } f'(\underline{\theta}) = \frac{\mathfrak{L}}{\mathfrak{C}} f(\underline{\theta}) \quad (7.36)$$

where all subscripts have been removed for improved readibility. It is clear that the updated density *f'* depends on the prior density *f* and the likelihood function, $\mathfrak{L}(\underline{x},\underline{\theta}) = f(\underline{x}\,|\,\underline{\theta})$. The normalizing constant is $\mathfrak{C}(\underline{x}) = f(\underline{x})$.

More generally, instead of observing $\underline{x}$ directly, we may be able to observe some function of $\underline{x}$ or some constraint on some function, i.e., $h(\underline{x}) < 0$. The conditioning in Eq (7.35) will then be $h(\underline{X}) < 0$.

Sequential updating is also possible. Suppose data are coming in batches, or periodically in time. We would like to keep current with the updating process. Let $\underline{x}^{(1)}$ be the first set of data, $\underline{x}^{(2)}$ be the second and so on. After the first data set arrives, the updated pdf of $\underline{\Theta}$, following Eq (7.36), is:

$$f'(\underline{\theta}) = \frac{f(\underline{x}^{(1)}\,|\,\underline{\theta})\, f(\underline{\theta})}{f(\underline{x}^{(1)})} = \frac{f(\underline{x}^{(1)}\,|\,\underline{\theta})\, f(\underline{\theta})}{\int f(\underline{x}^{(1)}\,|\,\underline{\theta})\, f(\underline{\theta})d\underline{\theta}} \quad (7.37)$$

After the second data set arrives, the updated pdf becomes:

$$f''(\underline{\theta}) = f(\underline{\theta}\,|\,\underline{x}^{(1)},\underline{x}^{(2)}) = \frac{f(\underline{x}^{(2)}\,|\,\underline{\theta},\underline{x}^{(1)})\, f'(\underline{\theta})}{f(\underline{x}^{(2)}\,|\,\underline{x}^{(1)})} = \frac{f(\underline{x}^{(2)}\,|\,\underline{\theta},\underline{x}^{(1)})\, f'(\underline{\theta})}{\int f(\underline{x}^{(2)}\,|\,\underline{\theta},\underline{x}^{(1)})\, f(\underline{\theta}\,|\,\underline{x}^{(1)})d\underline{\theta}} \quad (7.38)$$

The recursive nature of this relation is obvious. When estimating f'', f' becomes the prior density function. For a time series, there may be dependence in the sequential observations (i.e., between $x^{(1)}$ and $x^{(2)}$). If however, the two observation sets are independent, the updating relation simplifies to:

$$f''(\underline{\theta}) = \frac{f(\underline{x}^{(2)}\,|\,\underline{\theta})\, f'(\underline{\theta})}{f(\underline{x}^{(2)})} = \frac{f(\underline{x}^{(2)}\,|\,\underline{\theta})\, f'(\underline{\theta})}{\int f(\underline{x}^{(2)}\,|\,\underline{\theta})\, f(\underline{\theta}\,|\,\underline{x}^{(1)})d\underline{\theta}} \quad (7.39)$$

Likewise, the updated distribution of X can be given by:

$$
\begin{aligned}
f_X^{'}(x) = f_{X|\underline{X}=\underline{x}}(x) &= \frac{f_{X,\underline{X}}(x,\underline{x})}{f_{\underline{X}}(\underline{x})} \quad \text{(by definition)} \\
&= \int \frac{f_{X,\underline{X}|\Theta}(x,\underline{x})}{f_{\underline{X}}(\underline{x})} f_{\Theta}(\underline{\theta})\, d\underline{\theta} \\
&= \int\int \frac{f_{X,\underline{X}|\Theta}(x,\underline{x})}{f_{\underline{X}}(\underline{x})} f_{\Theta|\underline{X}=\underline{x}}(\underline{\theta})\, f_{\underline{X}}(\underline{x})\, d\underline{\theta}\; d\underline{x} \\
&= \int\int f_{X,\underline{X}|\Theta}(x,\underline{x}) f_{\Theta|\underline{X}=\underline{x}}(\underline{\theta})\, d\underline{\theta}\; d\underline{x} \\
&= \int f_{X|\Theta}(x) f_{\Theta|\underline{X}=\underline{x}}(\underline{\theta})\, d\underline{\theta}
\end{aligned}
\qquad (7.40)
$$

7.10.1 Example: Bayesian updating of an Exponential distribution

Let X be a random variable. Its distribution is exponential only if the parameter theta of the distribution can be known accurately. It so happens, that the parameter theta is uncertain, and it can be described as an exponential random variable with parameter λ. The parameter λ is known.

Therefore, $f(x\,|\,\theta) = \theta \exp(-x\theta), \quad f(\theta) = \lambda \exp(-\lambda\theta)$. Now suppose *n* random samples of X are obtained. Then the samples are iid with the same distribution as X. Their joint conditional density, which is also the likelihood function, is,

$$
\mathfrak{L}(\underline{x},\underline{\theta}) = f(\underline{x}\,|\,\theta) = \prod_{i=1}^{n} \theta \exp(-\theta x_i) = \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \qquad (7.41)
$$

The unconditional joint density of $\underline{X}$, which is also the normalizing constant, is

$$
f(\underline{x}) = \int_0^\infty \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \lambda \exp(-\lambda\theta) d\theta = \frac{\lambda}{(\lambda')^{n+1}} \Gamma(n+1) \qquad (7.42)
$$

where $\lambda' = \lambda + \sum x_i$. Thus, the updated density of theta is,

$$
f(\theta\,|\,\underline{x}) = \frac{\theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i\right) \lambda \exp(-\lambda\theta)}{\frac{\lambda}{(\lambda')^{n+1}} \Gamma(n+1)} = \frac{\lambda'(\lambda'\theta)^n \exp(-\lambda'\theta)}{\Gamma(n+1)} \qquad (7.43)
$$

which is clearly the gamma density function with parameters λ' and n+1. The updated density function of X can likewise be obtained as:

$$
f_X^{'}(x) = f_{X|\underline{x}}(x) = \int f(x\,|\,\theta,\underline{x}) f(\theta\,|\,\underline{x}) d\theta = \frac{n+1}{x+\lambda'} \left(\frac{\lambda'}{x+\lambda'}\right)^{n+1} \qquad (7.44)
$$