50008 - Probability and Statistics - Lecture 4

Oliver Killane

01/02/22

0.1 Joint Distributions

CDF

Lecture Recording

Lecture recording is available here

Suppose we have random variables X and Y such that:

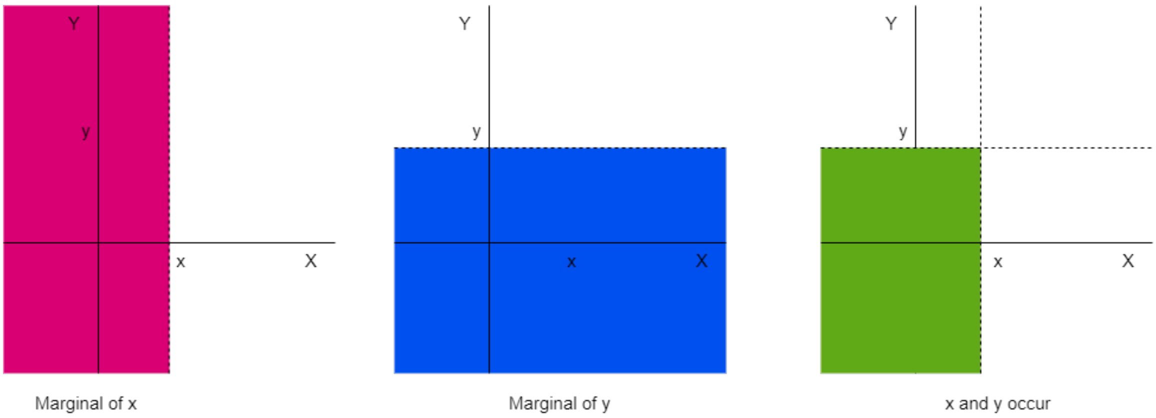
$$X : S_X \to \mathbb{R} \quad \text{and} \quad Y : S_Y \to \mathbb{R}$$

We can define Z operating on sample space S such that:

$$S = S_1 \times S_2 \quad S = \{(s_X, s_Y) | s_X \in S_X \land s_Y \in S_Y\} \quad Z = (X, Y) : S \to \mathbb{R}^2$$

Hence we have a mapping from joint random variable $Z(s)$ onto $(X(s), Y(s))$.

We can consider this using a graph of the sample space:



Hence the induced probability function for Z will be:

$$F(x, y) = P_Z(X \leq x, Y \leq y) = P_Z((-\infty, x], (-\infty, y]) = P(S_{XY})$$

Hence we can use the marginals of the joint distribution to get the distribution of the two random variables:

$$F_X(x) = F(x, \infty) \quad \text{and} \quad F_Y(y) = F(\infty, y)$$

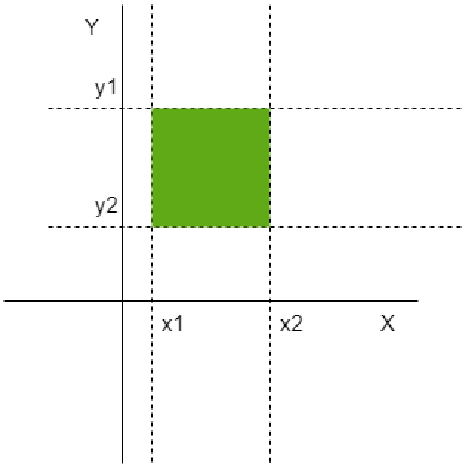
To be a valid **joint cumulative distribution function**:

- $\forall x, y \in \mathbb{R}.\ 0 \leq F(x, y) \leq 1$
- **Monotonicity**

$$\forall x_1, x_2, y_1, y_2 \in \mathbb{R}.\ [x_1 < x_2 \Rightarrow F(x_1, y_1) \leq F(x_2, y_1) \land y_1 < y_2 \Rightarrow F(x_1, y_1) \leq F(x_1, y_2)]$$

- $\forall x, y \in \mathbb{R}.\ F(x - \infty) = F(-\infty, y) = 0$
- $F(\infty, \infty) = 1$

For the probability of intervals we can use the graph mapping concept again:



$$P_Z(x_1 < X \leq x_2, Y \leq y) = F(x_2, y) - F(x_1, y)$$

Hence we can get the interval:

$$P_Z(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

PMF

Definition: Joint Probability Mass Function

$$p(x, y) = P_Z(X = x, Y = y) \text{ where } x, y \in \mathbb{R}$$

We can get the original **pmfs** of the two variables as:

$$p_X(x) = \sum_y p(x, y) \quad \text{and} \quad p_Y(y) = \sum_x p(x, y)$$

To be a valid **pmf**:

- $\forall x, y \in \mathbb{R}.\ 0 \leq p(x, y) \leq 1$
- $\sum_y \sum_x p(x, y) = 1$

PDF

Lecture Recording

Lecture recording is available here

Fundamental Theorem of Caculus

The fundamental law that integration and differentiation and the inverse of each other (except for constant added in integration c, which does not affect definite integrals).

Definition: Joint Probability Density Function

When the variables being *joined* are continuous we have $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, in this case:

$$F(x, y) = \int_{a=-\infty}^{y} \int_{b=-\infty}^{x} f(b, a)\ db\ da$$

The sum of the probability density function from $(x, y) \to (-\infty, -\infty)$

Hence by the fundamental theorem of calculus:

$$f(x, y) = \frac{\sigma^2}{\sigma x \sigma y} F(x, y)$$

We can differentiate to go get the PMF from the PDF.

To be valid:

- $\forall x, y \in \mathbb{R}. f(x, y) \geq 0$
- $\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f(x, y)\ dx\ dy$

Definition: Marginal Density Functions

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} F(x, \infty) \\ &= \frac{d}{dx} \int_{y=-\infty}^{\infty} \int_{s=-\infty}^{x} f(s, y)\ ds\ dy \end{aligned}$$

And likewise for y:

$$f_Y(y) = \frac{d}{dy} \int_{x=-\infty}^{\infty} \int_{s=-\infty}^{y} f(x, s)\ ds\ dx$$

Hence by applying the fundamental theorem of calculus:

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

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

Example: Marginal pdf

Given continuous variables $(X, Y) \in \mathbb{R}^2$:

$$f(x, y) = \begin{cases} 1 & |x| + |y| < \frac{1}{\sqrt{2}} \\ 0 & otherwise \end{cases}$$

To determine the marginal **pdf**s for X and Y:

First notice that: $|x| + |y| < 1\frac{1}{\sqrt{2}} \Leftrightarrow |y| < \frac{1}{\sqrt{2}} - |x|$.

Hence given an x we can see that for the first case of the probability density function to match, y must be between:

$$\frac{1}{-\sqrt{2}} + |x| < y < \sqrt{2} - |x|$$

$$\begin{aligned} f_X(x) &= \int_{y=-\infty}^{\infty} f(x, y)\ dy \\ &= \int_{y=-\sqrt{2}+|x|}^{\sqrt{2}-|x|} 1\ dy \\ &= [y]_{-\sqrt{2}+|x|}^{\sqrt{2}-|x|} \\ &= \left(\sqrt{2} - |x|\right) - \left(-\sqrt{2} + |x|\right) \\ &= 2\sqrt{2} - 2|x| \end{aligned}$$

Similarly for y:

$$f_Y(y) = 2\sqrt{2} - 2|y|$$

Joint Conditional Random Variables

Lecture Recording

Lecture recording is available here

Given random variables X and Y:

$$\text{variables independent } \Leftrightarrow F(x, y) = F_X(x)F_Y(y)$$

(For both continuous and discrete)

More specifically:

For Discrete Variables	$p(x, y) = p_X(x)p_Y(y)$	(probability mass function)
For Continuous Variables	$f(x, y) = f_X(x)f_Y(y)$	(Probability density function)

Example: Diamond at origin

Consider **pdf**:

$$f(x,y) = \begin{cases} 1 & |x| + |y| < \frac{1}{\sqrt{2}} \\ 0 & otherwise \end{cases}$$

By the previous example:

$$f_X(x) = 2\sqrt{2} - 2|x|$$
$$f_Y(y) = 2\sqrt{2} - 2|y|$$

Hence as $f(x,y) \neq f_X(x)f_Y(y)$ and hence X and Y are not independent.

Example: Independent variables

Given two continuous random variables X and Y:

$$f(x,y) = \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} \text{ given } x, y > 0$$

We can get the marginal **pdf** by integrating over all of y:

$$\begin{aligned}
f(x) &= \int_{y=-\infty}^{\infty} f(x,y)dy \\
&= \int_{y=0}^{\infty} f(x,y)dy \\
&= \lim_{t\to\infty} \int_{y=0}^{t} \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} dy \\
&= \lim_{t\to\infty} \int_{y=0}^{t} \lambda_1\lambda_2 e^{-\lambda_1 x} \times e^{-\lambda_2 y} dy \\
&= \lim_{t\to\infty} \left[-\lambda_1 e^{-\lambda_1 x - \lambda_2 y}\right]_{y=0}^{y=t} \\
&= \lim_{t\to\infty} \left(-\lambda_1 e^{-\lambda_1 x - \lambda_2 t}\right) - \left(-\lambda_1 e^{-\lambda_1 x - \lambda_2 0}\right) \\
&= \lim_{t\to\infty} \left(-\lambda_1 e^{-\lambda_1 x - \lambda_2 t}\right) - \left(-\lambda_1 e^{-\lambda_1 x - \lambda_2 0}\right) \\
&= 0 - \left(-\lambda_1 e^{-\lambda_1 x}\right) \\
&= \lambda_1 e^{-\lambda_1 x}
\end{aligned}$$

We can do the same for $f_Y(y)$ to get $\lambda_2 e^{-\lambda_2 y}$.

Hence the events are independent as:

$$\lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} = \lambda_1 e^{-\lambda_1 x} \times \lambda_2 e^{-\lambda_2 y}$$

Conditional PMF

For discrete random variables we can define the joint **pmf** as:

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} \text{ where } \forall y . p_Y(y) > 0$$

Definition: Baye's Theorem

Baye's theorem states that on some partition of the sample space S, $P_1, \dots P_k$:

$$P(X) = \sum_{i=1}^{k} P(X|E_i)P(E_i)$$

Given each partition the probability of some X occurring sums to the total probability of X occurring.

Using the conditional joint **pmf** we can also express this theorem (over a single partition) as:

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

Definition: Conditional PMF Marginal Joint Probabilities

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

(Go through every y, summing the probability of x occurring with that y, multiplied by the probability of that y)

Conditional PDF

For continuous random variables we can define the joint **pdf** as:

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

$$X \text{ and } Y \text{ independent } \Leftrightarrow \forall x, y \in \mathbb{R}.\ f_{X|Y}(x,y) = f_X(x)$$

And we can now have **bayes theorem** as:

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

Definition: Conditional PDF Marginal Joint Probabilities

$$f_X(x) = \int_{y=-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) \; dy$$

and with the cumulative distribution:

$$F_X(x) = \int_{y=-\infty}^{\infty} F_{X|Y}(x|y) f_Y(y) \; dy$$

Example: Independent exponential random variables

Given $X \sim Exp(\lambda)$ and $Y \sim Exp(\mu)$ what is $P(X < Y)$.

$$
\begin{aligned}
P(X < Y) &= \int_{x<y} f(x,y)\ dx\ dy \\
&= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{y} f(x,y)\ dx\ dy \text{ (go over all } y\text{s, for each take the } x\text{s that are less)} \\
&= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{y} f_X(x) f_Y(y)\ dx\ dy \text{ (} X \text{ and } Y \text{ are independent)} \\
&= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{y} f_X(x) f_Y(y)\ dx\ dy \text{ (} X \text{ and } Y \text{ are independent)} \\
&= \int_{y=-\infty}^{\infty} F_X(y) \times (\mu e^{-\mu y})\ dx\ dy \text{ (Integrate } f_X \text{ to get } F_X \text{ and then get all below } y\text{)} \\
&= \int_{y=-\infty}^{\infty} (1 - e^{-\lambda y}) \times (\mu e^{-\mu y})\ dx\ dy \text{ (Substitute definitions)} \\
&= \int_{y=0}^{\infty} (1 - e^{-\lambda y}) \times (\mu e^{-\mu y})\ dx\ dy \text{ (exponential cut at 0)} \\
&= \lim_{t\to\infty} \int_{y=0}^{t} (1 - e^{-\lambda y}) \times (\mu e^{-\mu y})\ dx\ dy \\
&= \lim_{t\to\infty} \int_{y=0}^{t} (\mu e^{-\mu y}) - e^{-\lambda y} \times (\mu e^{-\mu y})\ dx\ dy \\
&= \lim_{t\to\infty} \int_{y=0}^{t} (\mu e^{-\mu y}) - \mu e^{(-\lambda-\mu)y}\ dx\ dy \\
&= \lim_{t\to\infty} \left[-e^{-\mu y} + \frac{-\mu}{-\lambda-\mu} e^{(-\lambda-\mu)y} \right]_{y=0}^{y=t} \\
&= \lim_{t\to\infty} \left[-e^{-\mu y} + \frac{\mu}{\lambda+\mu} e^{(-\lambda-\mu)y} \right]_{y=0}^{y=t} \\
&= \lim_{t\to\infty} \left(-e^{-\mu t} + \frac{\mu}{\lambda+\mu} e^{(-\lambda-\mu)t} \right) - \left(-e^{\mu 0} + \frac{\mu}{\lambda+\mu} e^{(-\lambda-\mu)0} \right) \\
&= (0 - 0) - \left(-1 + \frac{\mu}{\lambda+\mu} \right) \\
&= 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}
\end{aligned}
$$

Expectation and Variance for Joint Random Variables

Lecture Recording

Lecture recording is available here

Definition: Joint Expectation

Where g is a **bivariat function** on the random variables X and Y:

For **discrete variables**:

$$E(g(X,Y)) = \sum_{y} \sum_{x} g(x,y)p(x,y)$$

For **continuous variables**:

$$E(g(X,Y)) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x,y)f(x,y)\ dx\ dy$$

Hence we have the following:

- **For all** $g(X,Y) = g_1(X) + g_2(Y) \Rightarrow E(g_1(X) + g_2(Y)) = E_X(g_1(X)) + E_Y(g_2(Y))$
- **If X and Y are independent** $E(g_1(X) \times g_2(Y)) = E_X(g_1(X))) \times E_Y(g_2(Y))$
  Hence where $g(X,Y) = X \times Y$ we have $E(XY) = E_X(X) \times E_Y(Y)$

Definition: Covariance

Covariance measures how two random variables change with respect to one another.

For a single random variable we consider expected value of the difference between the mean and the value, squared.

$$\text{Expectation of } g(X) = (X - \mu_X)^2 = \sigma_X^2$$

For a bivariate we consider the expectation:

$$\text{Expectation of } g(X,Y) = (X - \mu_X)(Y - \mu_Y)$$

We can then defined the covariance as:

$$\begin{aligned}\sigma_{XY} = Cov(X,Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - E_X[X] \times E_Y[Y] \\ &= E[XY] - \mu_X \mu_Y\end{aligned}$$

When X and Y are independent so:

$$\sigma_{XY} = Cov(X,Y) = E[XY] - E_X[X] \times E_Y[Y] = E[XY] - E[XY] = 0$$

Definition: Correlation

Much like covariance, however is invariant to the scale of X and Y.

$$\rho_{XY} = Cor(X,Y) = \frac{\sigma_{XY}}{\sigma_X \times \sigma_Y}$$

If the variables are independent then $\rho_{XY} = \sigma_{XY} = 0$.

Multivariate Normal Distribution

Definition: Multivariate Normal Distribution

Given a random vector $X = (X_1, \ldots, X_n)$ with means $\mu = (\mu_1, \ldots, \mu_n)$ has joint **pdf**:

$$f_X = \frac{1}{\sqrt{(2\pi)^n det \sum}} exp(-1/2(x-\mu)^T \sum^{-1}(x-\mu))$$

Where $\sum$ is the covariance matrix:

$$\sum_{(i,j)} = Cov(X_i, X_j) \quad \text{where } 1 \leq i, j \leq n$$

The covariance matrix must be **positive-definite** for a **pdf** to exist Note that the random variables do not need to be independent.

Positive Definite real Matrices

$$M \text{ is positive-definite} \Leftrightarrow \forall x \in \mathbb{R}^{\ltimes} \backslash \{0\}.\ x^T M x > 0$$

Conditional Expectation

Definition: Conditional Expectation

In general $E(XY) \neq E_X(X)E_Y(Y)$
For discrete random variables the **conditional expectation** of Y given that $X = x$ is:

$$E_{Y|X}(Y|x) = \sum_y y p_{y|X}(y|x)$$

For continuous random variables:

$$E_{Y|X}(Y|x) = \int_{y=-\infty}^{\infty} y f_{Y|X}(y|x)\ dy$$

In both cases the conditional expectation is a function of x and not Y. We are getting the weighted sum over all Ys, for a single value (x) of X.

Definition: Expectation of a Conditional Expectation

We can define random variable W such that:

$$W = E_{Y|X}(Y|X)$$

W is effectively a function of the random variable $X : S \to \mathbb{R}$ by $W(s) = E_{Y|X}(Y|x)$ where $X(s) = x$.

Using this we can determine that:

$$E_Y(Y) = E_X(E_{Y|X}(Y|X))$$

(Expectation of Y is the same as the expectation function of X, of the expected value of Y given X)

This holds for both discrete and continuous.

$$\int_y \int_x y f_{Y|X}(y|x) f_X(x)\ dx\ dy = \int_y \int_x y f(x,y)\ dx\ dy = \int_y y f_Y(y)\ dy$$

Definition: Tower Rule

The expectation of a conditional expectation rule extends to chains of expectations:

$$\begin{aligned}
E(Y) &= E_{X_1}(E_Y(Y|X_1)) \\
&= E_{X_2}(E_{X_1}(E_Y(Y|X_1, X_2)|X_2)) \\
&= \dots \\
&= E_{X_n}(E_{X_{n-1}}(\dots E_{X_1}(E_Y(Y|X_1, \dots, X_n)|X_2, \dots, X_n) \dots |X_n))
\end{aligned}$$

This is a generalisation of the **partition rule** for conditional expectations.

Markov Chains

Lecture Recording

Lecture recording is available here

Definition: Discrete Time Markov Chain (DTMC)

- A series of random variable modelling the state at a time step: $X_0, X_1, X_2, \dots$
- The state space J (all states), where $J = sipp(X_i)$ (contains all states that we can be in at any step)
- We can take a sequence (sample path) through the states $(X_0, X_1, X_2, \dots)$
- We denote the state taken at step n as state J_n

We use an initial probability vector π to determine the start state:

$$\pi_0 = [\dots \text{ probability of starting in state i } \dots]$$

We determine the probability of each next state through the transition probability matrix r:

$$r_{ij} = P(X_{n+1} = j | X_n = i)$$

For a markov chain the probability of being in any next state is **only** dependent on the current state (memoryless, history of previous states does not matter).

$$P(X_{i+1} = J_{n+1} | X_i = J_i) = P(X_{i+1} = J_{n+1} | X_i = J_i) = P(X_{i+1} = J_{n+1} | X_0 = J_0, \dots, X_i = J_i)$$

To get the probability we can use power of the matrix:

$$P(X_n = j | X_0 = i) = (R^n)_{ij}$$

If we have the initial probability vector we can calculate:

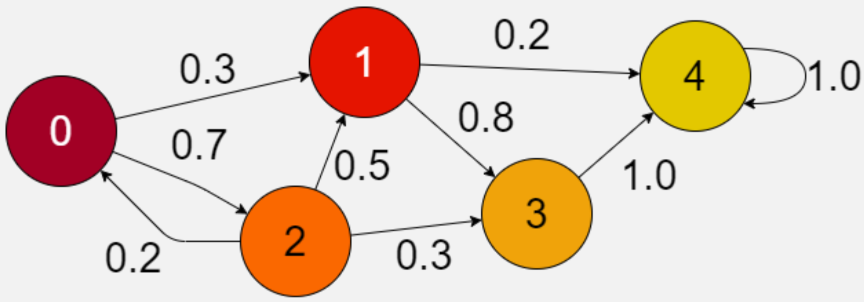
$$\begin{aligned} P(X_n = j) &= \sum_{i \in J} P(X_0 = i) \times P(X_n = j | X_0 = i) \\ &= \sum_{i \in j} \pi_{0_i} (R^n)_{ij} \\ &= (\pi_0 R^n)_{ij} \end{aligned}$$

We can obtain the long term probabilities by using the ∞th step:

$$\lim_{t \to +\infty} \pi_0 R^n = \pi_\infty$$

Note that since $\pi_\infty R = \pi_\infty$ we have eigenvector π_∞ and eigenvalue 1.

Example: Probabilistic Finite State Machine



$$J = \{\ 0 \quad 1 \quad 2 \quad 3 \quad 4\ \}$$ State Space

$$\pi_0 = [\ 1.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0\ ]$$ Initial Probability Vector

100% chance we start at 0

To

$$\text{From}\quad \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0.0 & 0.3 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.8 & 0.2 \\ 0.2 & 0.5 & 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \end{array}$$ Transition Probability Matrix

$\boxed{P(X[i+1] = 1 \mid X[i] = 2) = 0.5}$

Can get permentantly stuck at state 4

Example: Modelling Climate