CSE 312

Foundations of Computing II

19: More Joint Distributions; Law of Total Expectation

Agenda

- Bit more on joint distributions
- Law of Total Expectation and LTP (continuous)

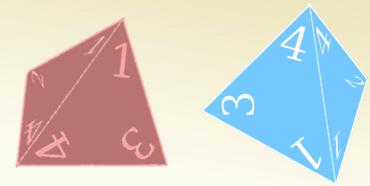
Discrete to Continuous

	Discrete	Continuous
PMF/PDF	$p_X(x) = P(X = x)$	$f_X(x) \neq P(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$
Normalization	$\sum_{x} p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
Expectation	$\mathbb{E}[g(X)] = \sum_{x} g(x)\, p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$

Jointly distributed random variables

X	Y
Grade on exam	Amount of sleep the night before
Performance of Microsoft stock	Performance of Amazon stock
Grade of person A on exam in 312	Grade of person B on exam in 332
Number of job interviews to get a job	State of the economy

X_1 blood pressure
X_2 temperature
X_3 blood glucose
X_4 kidney function

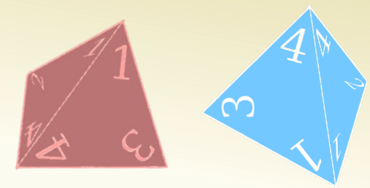


X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

- Valid joint probability mass function?

X/Y	Bad	Below Avg	Above Avg	Excellent
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

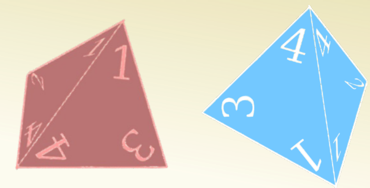


X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

X/Y	Bad	Below Avg	Above Avg	Excelle nt
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

- Probability a student gets a job in one interview and economy is excellent?

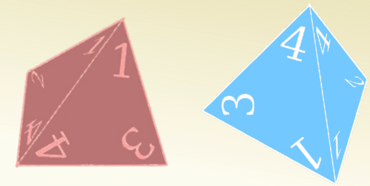


X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

X/Y	Bad	Below Avg	Above Avg	Excellent
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

- Probability a student gets a job in at most 3 interviews and economy below avg?

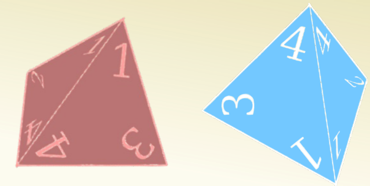


X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

X/Y	Bad	Below Avg	Above Avg	Excellent
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

- A student takes exactly 3 interviews to get a job?

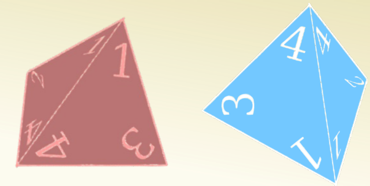


X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

X/Y	Bad	Below Avg	Above Avg	Excellent
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

- A student gets a job in one interview given that the economy is below average?



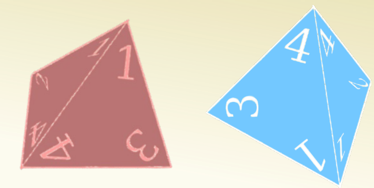
X is the number of of job interviews you need to do to get a job.

Y is the state of the economy. (Bad = 1, Below Avg = 2, Above Avg =3, Excellent = 4

X/Y	Bad	Below Avg	Above Avg	Excellent
1	0.01	0.05	0.08	0.12
2	0.03	0.05	0.06	0.12
3	0.03	0.12	0.04	0.05
4	0.08	0.12	0.03	0.01

- Are these random variables independent?

Example – Weirder Dice



Suppose I roll two fair 4-sided die independently. Let X be the value of the first die, and Y be the value of the second die. Let $U = \min(X, Y)$ and $W = \max(X, Y)$

$\Omega_U = \{1,2,3,4\}$ and $\Omega_W = \{1,2,3,4\}$

$$\Omega_{U,W} = \{(u, w) \in \Omega_U \times \Omega_W : u \leq w\} \neq \Omega_U \times \Omega_W$$

Are U and W independent?

U\W	1	2	3	4
1	1/16	2/16	2/16	2/16
2	0	1/16	2/16	2/16
3	0	0	1/16	2/16
4	0	0	0	1/16

A quick check for independence

The check: $\Omega_X \times \Omega_Y$ must equal $\Omega_{X,Y}$ for independence.

Suppose that there is some $(a, b) \in \Omega_X \times \Omega_Y$, but not in $\Omega_{X,Y}$.

Then: $p_{X,Y}(a, b) = 0$

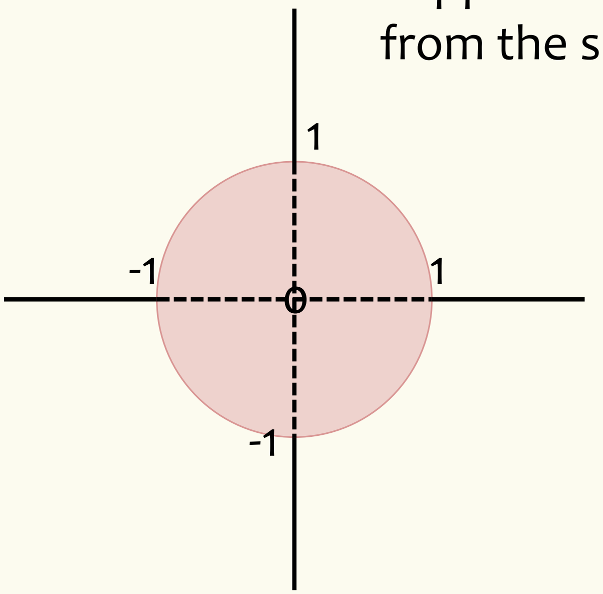
But: $p_X(a) > 0$ and $p_Y(b) > 0$

But beware, the converse is not true: $\Omega_X \times \Omega_Y = \Omega_{X,Y}$ does not imply independence!

Joint distributions

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leq x} \sum_{s \leq y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Example – Uniform distribution on a unit disk

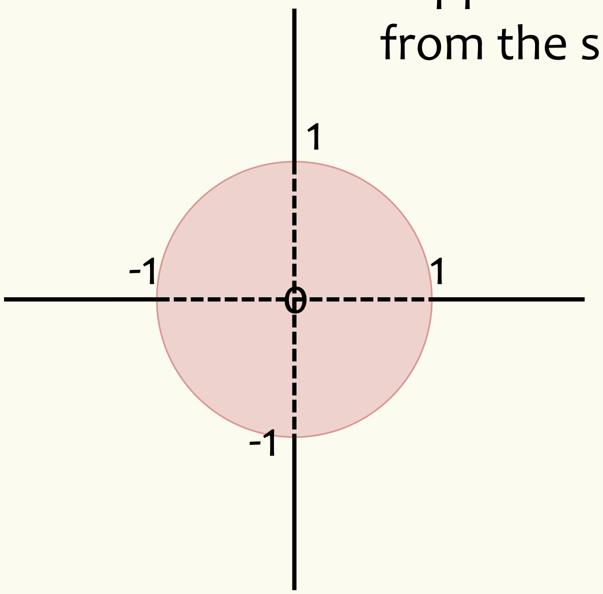


Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

What is the joint density?

$$f_{X,Y}(x,y) = \begin{cases} c & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example – Uniform distribution on a unit disk



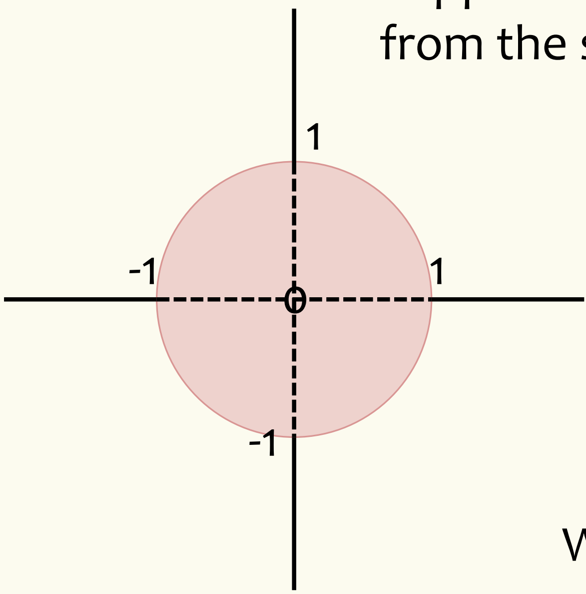
Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

Example – Uniform distribution on a unit disk



Suppose that a pair of random variabes (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + \mathrm{y}^2 \leq 1$

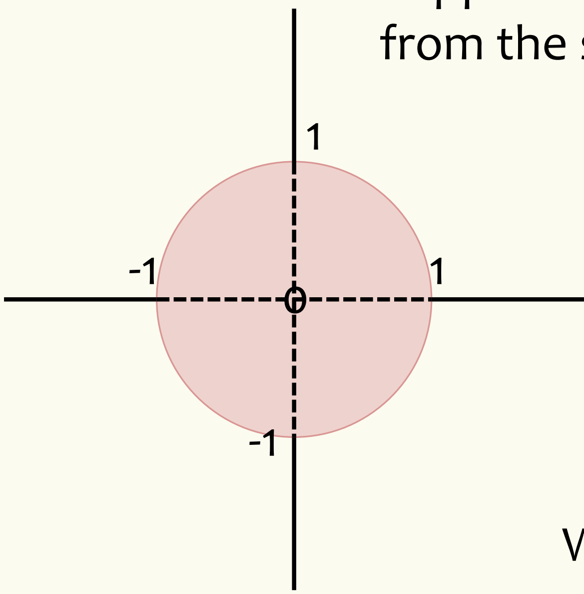
This is a disk of radius 1 which has area π

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $f_X(x)$?

Example – Uniform distribution on a unit disk

Suppose that a pair of random variabes (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$



This is a disk of radius 1 which has area π

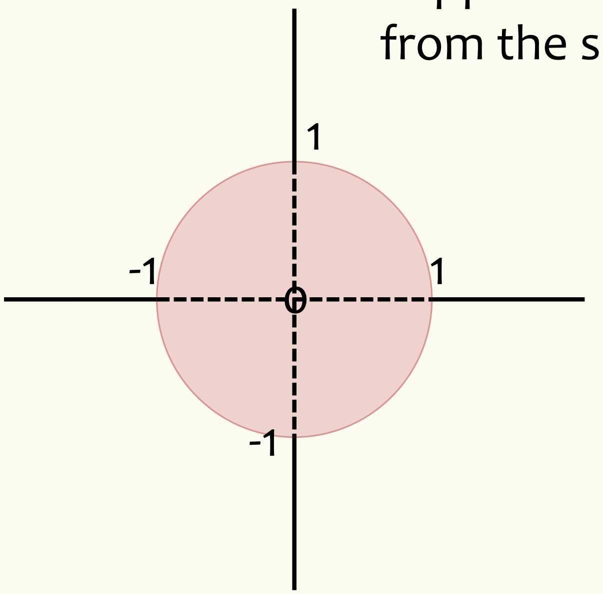
$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What is $f_X(x)$?

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} \, \mathrm{d}y$$

$$= 2\sqrt{1 - x^2}/\pi$$

Example – Uniform distribution on a unit disk



Suppose that a pair of random variables (X, Y) is chosen uniformly from the set of real points (x, y) such that $x^2 + y^2 \leq 1$

This is a disk of radius 1 which has area π

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent?

Agenda

- Bit more on joint distributions
- Law of Total Expectation and LTP (continuous)

Conditional Expectation and Law of Total Expectation

Suppose someone gave us $Y \sim \mathrm{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

Conditional Expectation

Definition. If X is a discrete random variable then the **conditional expectation** of X given event A is

$$\mathbb{E}[X \mid A] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid A)$$

Note:

- Linearity of expectation still applies here

$$\mathbb{E}[aX + bY + c \mid A] = a\,\mathbb{E}[X \mid A] + b\,\mathbb{E}[Y \mid A] + c$$

Law of Total Expectation

Law of Total Expectation (event version). Let X be a random variable and let events $A_1, \dots, A_n$ partition the sample space. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i] \cdot P(A_i)$$

Law of Total Expectation (random variable version). Let X be a random variable and Y be a discrete random variable. Then,

$$\mathbb{E}[X] = \sum_{y \in \Omega_Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y)$$

Proof of Law of Total Expectation

Follows from Law of Total Probability and manipulating sums

$$\mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot P(X = x)$$

$$= \sum_{x \in \Omega_X} x \cdot \sum_{i=1}^{n} P(X = x \mid A_i) \cdot P(A_i)$$ (by LTP)

$$= \sum_{i=1}^{n} P(A_i) \sum_{x \in \Omega_X} x \cdot P(X = x \mid A_i)$$ (change order of sums)

$$= \sum_{i=1}^{n} P(A_i) \cdot \mathbb{E}[X \mid A_i]$$ (def of cond. expect.)

Example – Flipping a Random Number of Coins

Suppose someone gave us $Y \sim \mathrm{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

By the Law of Total Expectation

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{E}[X \mid Y = i] \cdot P(Y = i) =$$

Example – Flipping a Random Number of Coins

Suppose someone gave us $Y \sim \mathrm{Poi}(5)$ fair coins and we wanted to compute the expected number of heads X from flipping those coins.

By the Law of Total Expectation

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} \mathbb{E}[X \mid Y = i] \cdot P(Y = i) = \sum_{i=0}^{\infty} \frac{i}{2} \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \sum_{i=0}^{\infty} i \cdot P(Y = i)$$

$$= \frac{1}{2} \cdot \mathbb{E}[Y] = \frac{1}{2} \cdot 5 = 2.5$$

Example -- Elevator rides

The number X of people who enter an elevator on the ground floor is a Poisson random variable with mean 10. If there are N floors above the ground floor, and if each person is equally likely to get off at any one of the N floors, independently of where others get off, compute the expected number of stops the elevator will make before discharging all the passengers.

Law of total probability

Definition. Let A be an event and Y a discrete random variable. Then

$$P[A] = \sum_{y \in \Omega_Y} P(A|Y = y) p_Y(y)$$

Definition. Let A be an event and Y a continuous random variable. Then

$$P[A] = \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y) \mathrm{d}y$$

Example use of law of total probability

Suppose that the time until server 1 crashes is $X \sim Exp\ (\lambda)$ and the time until server 2 crashes is independent, with $Y \sim Exp\ (\mu)$.

What is the probability that server 1 crashes before server 2?

Example use of law of total probability

$X \sim Exp\ (\lambda), Y \sim Exp\ (\mu)$.

What is the probability that $Y > X$?

$$P(Y > X) = \int_0^{\infty} \Pr(Y > X \mid X = x)\, f_X(x) dx$$

$$= \int_0^{\infty} \Pr(Y > x \mid X = x)\ \lambda e^{-\lambda x}\ dx$$

$$= \int_0^{\infty} \Pr(Y > x) \lambda e^{-\lambda x}\ dx$$

$$= \int_0^{\infty} e^{-\mu x}\, \lambda\, e^{-\lambda x}\ dx$$

$$= \frac{\lambda}{\lambda + \mu} \int_0^{\infty} (\lambda + \mu) \cdot e^{-\mu x}\, e^{-\lambda x}\, dx$$

$$= \frac{\lambda}{\lambda + \mu}$$

Alternative approach

$X \sim Exp\ (\lambda), Y \sim Exp\ (\mu)$.

What is the probability that $Y > X$?

$$P(Y > X) = \int_{x=0}^{\infty} \int_{y=x}^{\infty} f_{X,Y}(x, y) \mathrm{dy}\ \mathrm{d}x$$

$$= \int_{x=0}^{\infty} \int_{y=x}^{\infty} f_X(x) \cdot f_Y(y) \mathrm{dy}\ \mathrm{d}x$$

Reference Sheet (with continuous RVs)

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \leq x} \sum_{s \leq y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
Normalization	$\sum_{x} \sum_{y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Expectation	$E[g(X,Y)] = \sum_{x} \sum_{y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$
Conditional PMF/PDF	$p_{X \mid Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$	$f_{X \mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Conditional Expectation	$E[X \mid Y = y] = \sum_{x} x p_{X \mid Y}(x \mid y)$	$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x \mid y) dx$
Independence	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$