CSE 312

Foundations of Computing II

Lecture 16: Joint Continuous, Conditional Distributions, and Tail Bounds

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Aleks Jovcic

Slide Credit: Based on Stefano Tessaro's slides for 312 19au
incorporating ideas from Anna Karlin, Ryan O'Donnell, Alex Tsun, Rachel Lin, Hunter Schafer & myself ☺

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  – Law of Total Expectation
- Tail Bounds
  – Markov's Inequality
  – Chebyshev's Inequality

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X=x, Y=y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$
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	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$

Independence (continuous random variables)

Definition. Let X and Y be continuous random variables. The **joint pdf** of X and Y is

$$f_{X,Y}(a,b) \neq \Pr(X = a, Y = b)$$

Definition. The **joint range** of $p_{X,Y}$ is

$$\Omega(X,Y) = \{(c,d) : p_{X,Y}(c,d) > 0\} \subseteq \Omega(X) \times \Omega(Y)$$

Definition. X and Y are **independent** iff for all a, b

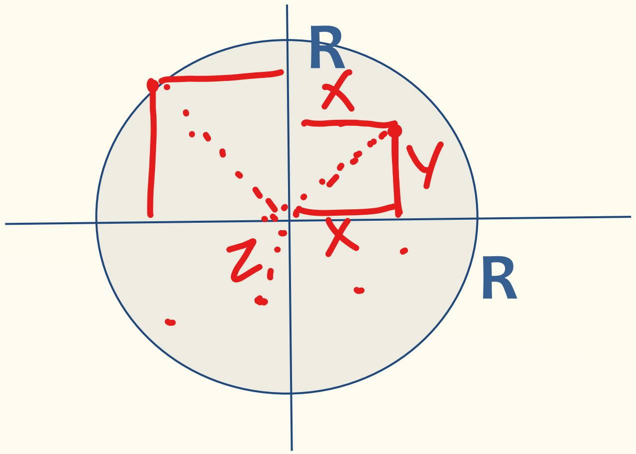
$$f_{X,Y}(a,b) = f_X(a) \cdot f_Y(b)$$

- Suppose that the surface of a disk is a circle with ~~area~~ radius R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - What is their joint density f(x,y)?

$$f_{X,Y} = c \quad x^2 + y^2 \le R^2 \qquad f_{X,Y}\left(\frac{R}{2}, \frac{R}{3}\right) = \frac{1}{\pi R^2}$$

$$\iint_{\pi R^2} c = 1 = c\pi R^2 = 1$$

$$c = \frac{1}{\pi R^2}$$



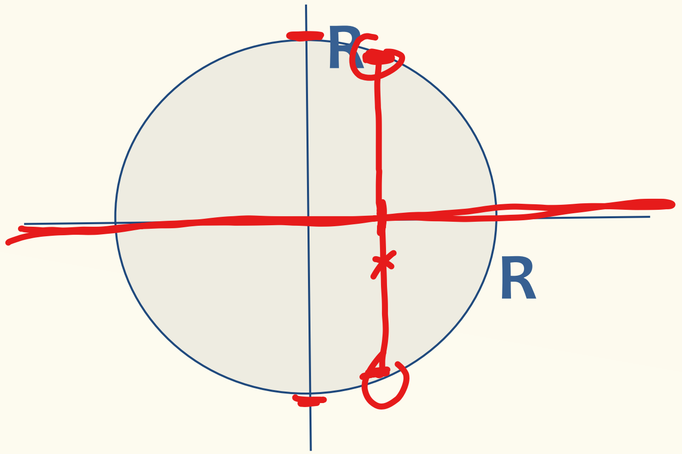
- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - What is the range of X & Y and the marginal density of X and of Y?

$x^2 + y^2 = R^2$

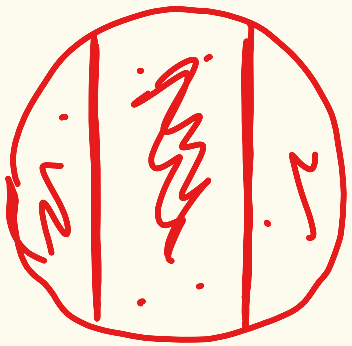
$\Omega_X : [-R, R]$

$\Omega_Y : [-R, R]$

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



$f_Y(y) = \dots$



Poll:

What is Ω_X?

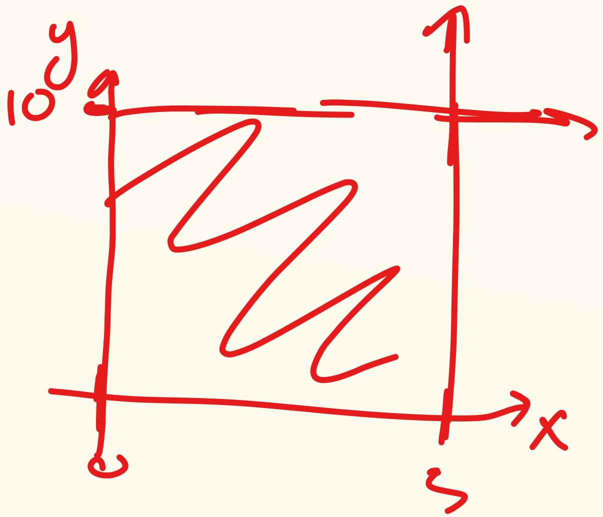
a. $[-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}]$

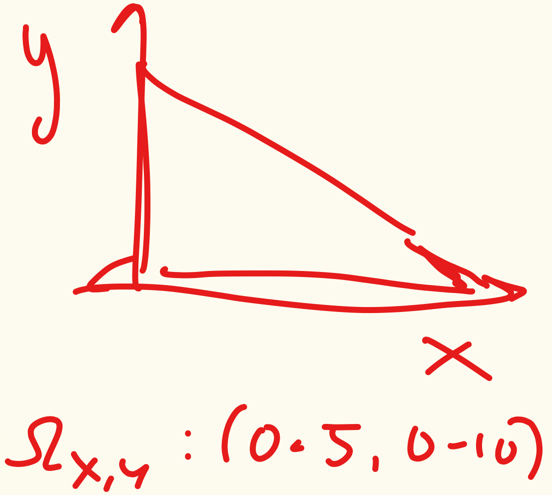
b. $[-R, R]$

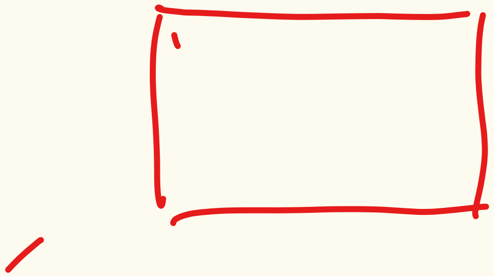
c. $[-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2}]$

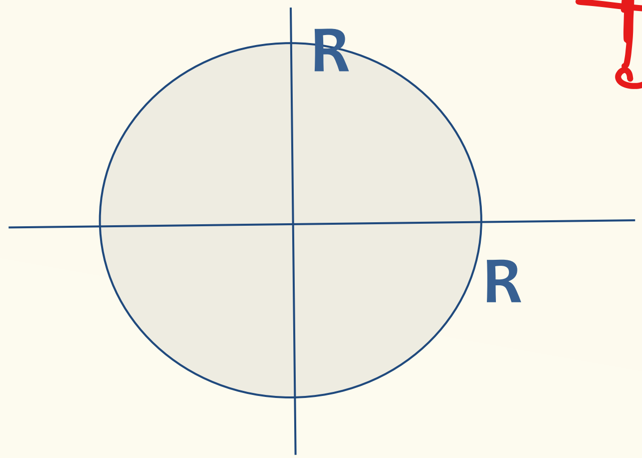
d. Not sure

- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - Are X and Y independent?









Poll:

Are X and Y independent?

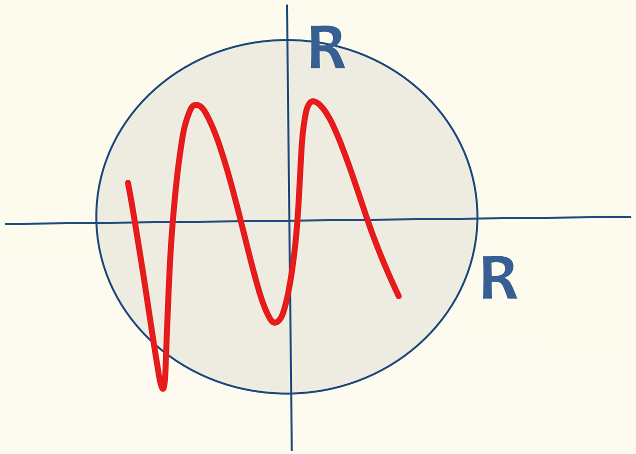
a. yes

b. no

- Suppose that the surface of a disk is a circle with area R centered at the origin and that there is a single point imperfection at a location with is uniformly distributed across the surface of the disk. Let X and Y be the x and y coordinates of the imperfection (random variables) and let Z be the distance of the imperfection from the origin.
  - What is E(Z)?

$g(x,y) = \sqrt{x^2+y^2}$

$$E[\sqrt{X^2+Y^2}] = \iint_{x^2+y^2 \le R^2} \sqrt{x^2+y^2}\,\frac{1}{\pi R^2}\,dx\,dy$$



$$E[g(X,Y)] = \iint g(x,y)\, f_{X,Y}(x,y)\,dx\,dy$$

All of this generalizes to more than 2 random variables

	Discrete	Continuous
Joint PMF/PDF	$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$
Joint range/support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s)\, ds dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx dy = 1$
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	$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  - Law of Total Expectation
- Tail Bounds
  - Markov's Inequality
  - Chebyshev's Inequality

Conditional Expectation

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

$$E[X \mid A] = \sum_{x \in \Omega(X)} x \Pr(X = x \mid A)$$

- Linearity of expectation still applies here

$$E[aX + bY + c \mid A] = aE[X \mid A] + bE[Y \mid A] + c$$

Conditional Expectation

Definition. Let X be a discrete random variable then the **conditional expectation** of X given event $Y = y$ is

$$E[X \mid Y = y] = \sum_{x \in \Omega(X)} x \Pr(X = x \mid Y = y)$$

- Linearity of expectation still applies here

$$E[aX + bY + c \mid Y = y] = aE[X \mid Y = y] + bE[Y \mid Y = y] + c$$

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  - Law of Total Expectation
- Tail Bounds
  - Markov's Inequality
  - Chebyshev's Inequality

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,

$$E[X] = \sum_{i=1}^{n} E[X|A_i]\Pr(A_i)$$

Proof of Law of Total Expectation

Follows from Law of Total Probability and manipulating sums

$$E[X] = \sum_{x \in \Omega(X)} x \Pr(X = x)$$

$$= \sum_{x \in \Omega(X)} x \sum_{i=1}^{n} \Pr(X = x \mid A_i) \Pr(A_i)$$ (by LTP)

$$= \sum_{i=1}^{n} \Pr(A_i) \sum_{x \in \Omega(X)} x \Pr(X = x \mid A_i)]$$ (change order of sums)

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

Law of Total Expectation

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

$$E[X] = \sum_{y \in \Omega(Y)} E[X|Y = y]\Pr(Y = y)$$

Example: Flipping Coins

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

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

$$= \sum_{y=0}^{\infty} yp \cdot e^{-5} \frac{5^y}{y!} = p \cdot e^{-5} \sum_{y=0}^{\infty} \frac{y \cdot 5^y}{y!}$$

$$E[XY] \overset{?}{=} E[X]E[Y]$$

Elevator rides

The number 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.

$Y_i = 1$ if stops at floor i

$$E[Y] = \sum_{i=1}^{N} E[Y_i]$$

$$P(Y_i = 1) = P(\text{at least one person gets off})$$
$$= 1 - P(\text{nobody gets off})$$
$$= 1 - P(\text{all choose another floor})$$
$$= 1 - \left(1 - \frac{1}{N}\right)^k$$

$$E[Y|X=k] = N\left(1 - \left(1 - \frac{1}{N}\right)^k\right)$$

$$E[Y] = \sum_{k=0}^{\infty} N\left(1 - \left(1 - \frac{1}{N}\right)^k\right) e^{-10} \frac{10^k}{k!}$$

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	$F_{X,Y}(x,y) = \sum_{t \le x} \sum_{s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,s) ds dt$
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	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  – Law of Total Expectation
- Tail Bounds
  – Markov's Inequality
  – Chebyshev's Inequality

Tail Bounds (Idea)

Bounding the probability a random variable is far from its mean. Usually statements of the form:

$$\Pr(X \geq a) \leq b$$
$$\Pr(|X - E[X]| \geq a) \leq b$$

Useful tool when

- An approximation that is easy to compute is sufficient
- The process is too complex to analyze exactly

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  - Law of Total Expectation
- Tail Bounds
  - Markov's Inequality
  - Chebyshev's Inequality

Markov's Inequality

Theorem. Let X be a random variable taking only non-negative values. Then, for any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

$$\mathbb{P}(X \geq 5) \leq \frac{\mathbb{E}[X]}{5}$$

$$\mathbb{P}\big(X \geq t \cdot \mathbb{E}(X)\big) \leq \frac{1}{t}.$$

Incredibly simplistic – only requires that the random variable is non-negative and only needs you to know <u>expectation</u>. You don't need to know **anything else** about the distribution of X.

Markov's Inequality – Proof

Theorem. Let X be a (discrete) random variable taking only non-negative values. Then, for any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

$$\mathbb{E}(X) = \sum_{x} x \cdot \mathbb{P}(X = x)$$

$$= \sum_{x \geq t} x \cdot \mathbb{P}(X = x) + \sum_{x < t} x \cdot \mathbb{P}(X = x)$$

Markov's Inequality – Proof

Theorem. Let X be a (discrete) random variable taking only non-negative values. Then, for any $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

$$\mathbb{E}(X) = \sum_{x} x \cdot \mathbb{P}(X = x)$$

$$= \sum_{x \geq t} x \cdot \mathbb{P}(X = x) + \sum_{x < t} x \cdot \mathbb{P}(X = x)$$

≥ 0 because $x \geq 0$ whenever $\mathbb{P}(X = x) \geq 0$ (takes only non-negative values)

$$\geq \sum_{x \geq t} x \cdot \mathbb{P}(X = x)$$

$$\mathbb{E}[X] \geq t \cdot \mathbb{P}(X \geq t)$$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

$$\geq \sum_{x \geq t} t \cdot \mathbb{P}(X = x) = t \cdot \mathbb{P}(X \geq t)$$

Follows by re-arranging terms …

Example – Binomial Random Variable

Markov's inequality
$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$.

Let X be Binomial RV with parameters. n, p

$$\mathbb{E}(X) = \frac{n}{2}$$

What is the probability that $X \geq \frac{3n}{4}$?

Markov's inequality: $\mathbb{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{3n} \cdot \frac{n}{2} = \frac{2}{3}$

Can we do better?

Agenda

- Joint Continuous Distributions
- Conditional Expectation
  - Law of Total Expectation
- Tail Bounds
  - Markov's Inequality
  - Chebyshev's Inequality

Using variance

- If we know more about the random variable, e.g. its variance, we can get a better bound!

Chebyshev's Inequality

Markov's inequality
$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$.

Theorem. Let X be a random variable. Then, for any $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof: Define $Z = X - \mathbb{E}(X)$

$$\mathbb{P}(|Z| \geq t) = \mathbb{P}(Z^2 \geq t^2) \leq \frac{\mathbb{E}(Z^2)}{t^2} = \frac{\mathrm{Var}(X)}{t^2}$$

Definition of Variance

$|Z| \geq t$ iff $Z^2 \geq t^2$

Markov's inequality ($Z^2 \geq 0$)

Example – Binomial Random Variable

Chebychev's Inequality

$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$

Let X be Binomial RV with parameters. $n, p = 0.5$

$$\mathbb{E}(X) = \frac{n}{2} \qquad Var(X) =$$

What is the probability that $X \geq \frac{3n}{4}$*?*

Chebychev's inequality: $\mathbb{P}\left(X \geq \frac{3n}{4}\right) \leq$

Markov's inequality: $\mathbb{P}\left(X \geq \frac{3n}{4}\right) \leq \frac{4}{3n} \cdot \frac{n}{2} = \frac{2}{3}$

Tail Bounds

Useful for approximations of complex systems. How good the approximation is depends on the actual distribution and the context you are using it in.

– Usually loose upper-bounds are okay when designing for worst-case

Generally, the more you know about your random variable the better tail bounds you can get.