CSE 312

Foundations of Computing II

Lecture 19: More Joint Distributions
Tail Bounds part I

Midterm

- Scores released at 2:30pm after class
  - Breathe & relax!

Average: 72.37 Standard Deviation: 17.85 (Median: 77)

Scores	90+	80s	70s	60s	50s	< 50
# of students	21	44	36	21	13	20

- Too much reading/harder than intended
- Solutions available on Canvas Pages
- Regrade requests via e-mail only (to me) for major issues.

*Average 76.43 for students answering ≥10 PollEv polls

Review Joint PMFs and Joint Range

Definition. Let X and Y be discrete random variables. The **Joint PMF** of X and Y is

$$p_{X,Y}(a,b) = P(X = a, Y = b)$$

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

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

Note that

$$\sum_{(s,t)\in\Omega_{X,Y}} p_{X,Y}(s,t) = 1$$

Review Marginal PMF

Definition. Let X and Y be discrete random variables and $p_{X,Y}(a,b)$ their joint PMF. The **marginal PMF** of X

$$p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a,b)$$

Similarly, $p_Y(b) = \sum_{a \in \Omega_X} p_{X,Y}(a,b)$

Review Continuous distributions on $\mathbb{R} \times \mathbb{R}$

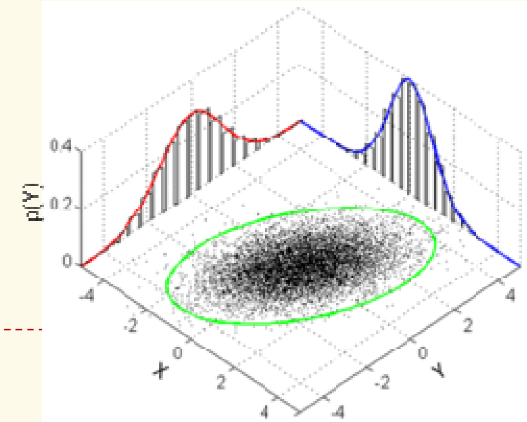
Definition. The **joint probability density function (PDF)** of continuous random variables X and Y is a function $f_{X,Y}$ defined on $\mathbb{R} \times \mathbb{R}$ such that

- $f_{X,Y}(x, y) \geq 0$ for all $x, y \in \mathbb{R}$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \mathrm{d}x \, \mathrm{d}y = 1$

for $A \subseteq \mathbb{R} \times \mathbb{R}$ the probability that $(X, Y) \in A$ is $\iint_A f_{X,Y}(x, y) \, \mathrm{d}x\mathrm{d}y$

The **(marginal) PDFs** f_X and f_Y are given by

- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}y$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}x$



Review Conditional Expectation

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

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

Notes:

- Can be phrased as a "random variable version"

$$\mathbb{E}[X \mid Y = y]$$

- Linearity of expectation still applies here

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

Review 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,

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

Agenda

- Joint Distributions
  - Another LTE example
  - Conditional expectation and LTE for continuous RVs
- Covariance
- Tail Bounds
  - Markov's Inequality

Example – Computer Failures (a familiar example)

Suppose your computer operates in a sequence of steps, and that at each step i your computer will fail with probability q (independently of other steps).

Let X be the number of steps it takes your computer to fail.

What is $\mathbb{E}[X]$?

What kind of RV is X?

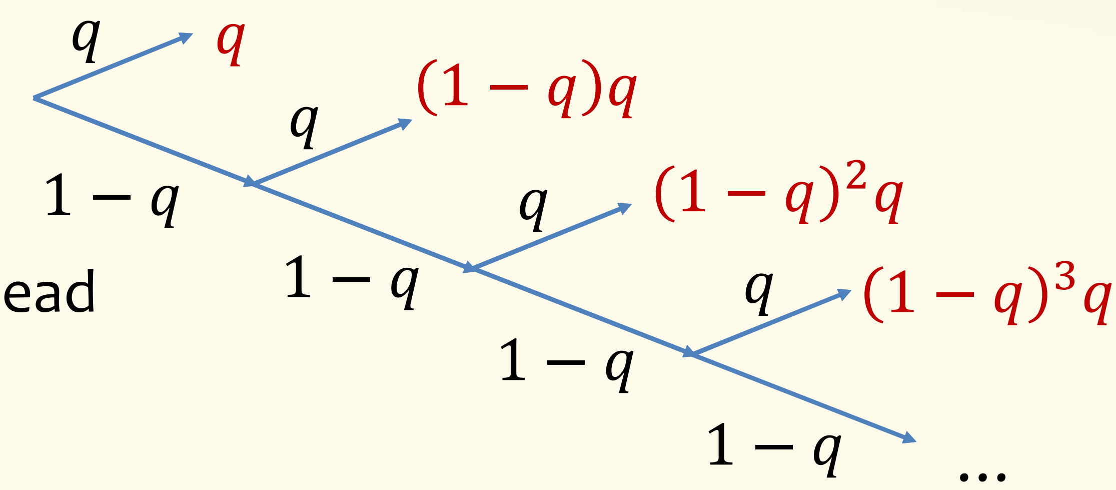
Recall – Flipping a biased coin until you see heads

- Biased coin:

  $P(H) = q > 0$

  $P(T) = 1 - q$

- X = # of coin flips until first head



$$P(X = i) = q\,(1-q)^{i-1}$$

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \cdot P(X = i) = \sum_{i=1}^{\infty} i \cdot q(1-q)^{i-1}$$

Converges, so $\mathbb{E}[X]$ is finite

Can calculate this directly …

Analysis – Flipping a biased coin until you see heads

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \cdot q(1-q)^{i-1} = q \sum_{i=1}^{\infty} i(1-q)^{i-1}$$

Converges, so $\mathbb{E}[X]$ is finite

So $\mathbb{E}[X] = q\,[1 + 2(1-q) + 3(1-q)^2 + \cdots \quad + i(1-q)^{i-1} + \cdots]$

Then $(1-q)\mathbb{E}[X] = q[\quad (1-q) + 2(1-q)^2 + \cdots + (i-1)(1-q)^{i-1} + \cdots]$

Subtracting gives

$$q\,\mathbb{E}[X] = q[\,1 + (1-q)\ + (1-q)^2 + \cdots \quad + (1-q)^{i-1} + \cdots]$$

$$q\,\mathbb{E}[X] = q\left[\frac{1}{1-(1-q)}\right] = 1$$

since for $0 < r < 1,\ \sum_{i=0}^{\infty} r^i = \frac{1}{1-r}$

Therefore $\mathbb{E}[X] = 1/q$

Same examples with the LTE

Suppose your computer operates in a sequence of steps, and that at each step i your computer will fail with probability q (independently of other steps).

Let X be the number of steps it takes your computer to fail.

What is $\mathbb{E}[X]$?

Let Y be the indicator random variable for the event of failure (heads) in step 1

Then by LTE, $\mathbb{E}[X] = \mathbb{E}[X \mid Y = 1] \cdot P(Y = 1) + \mathbb{E}[X \mid Y = 0] \cdot P(Y = 0)$

$= 1 \cdot q + \mathbb{E}[X \mid Y = 0] \cdot (1 - q)$

$= q + (1 + \mathbb{E}[X]) \cdot (1 - q)$ since if $Y = 0$ experiment starting at step 2 looks like original experiment

Solving we get $\mathbb{E}[X] = 1/q$

Conditional Expectation again…

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

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

Therefore for X and Y discrete random variables, the conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x \cdot P(X = x \mid Y = y) = \sum_{x \in \Omega_X} x \cdot p_{X|Y}(x|y)$$

where we **define** $p_{X|Y}(x|y) = P(X = x \mid Y = y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

Conditional Expectation – Discrete & Continuous

Discrete: Conditional PMF: $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$

Conditional Expectation: $\mathbb{E}[X \mid Y = y] = \sum_{x \in \Omega_X} x \cdot p_{X|Y}(x|y)$

Continuous: Conditional PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Conditional Expectation: $\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx$

Law of Total Expectation - continuous

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 and Y be continuous random variables. Then,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y) \; dy$$

Using LTE for Continuous RVs

PDF for $Exp(\lambda)$ is $\begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{o. w.} \end{cases}$

Expectation is $1/\lambda$

Suppose that we first choose $Y \sim Exp(1/2)$ and then choose $X \sim Exp(Y)$. What is $\mathbb{E}[X]$?

$$f_{X|Y}(x|y) = y\, e^{-x/y}$$

y is fixed here

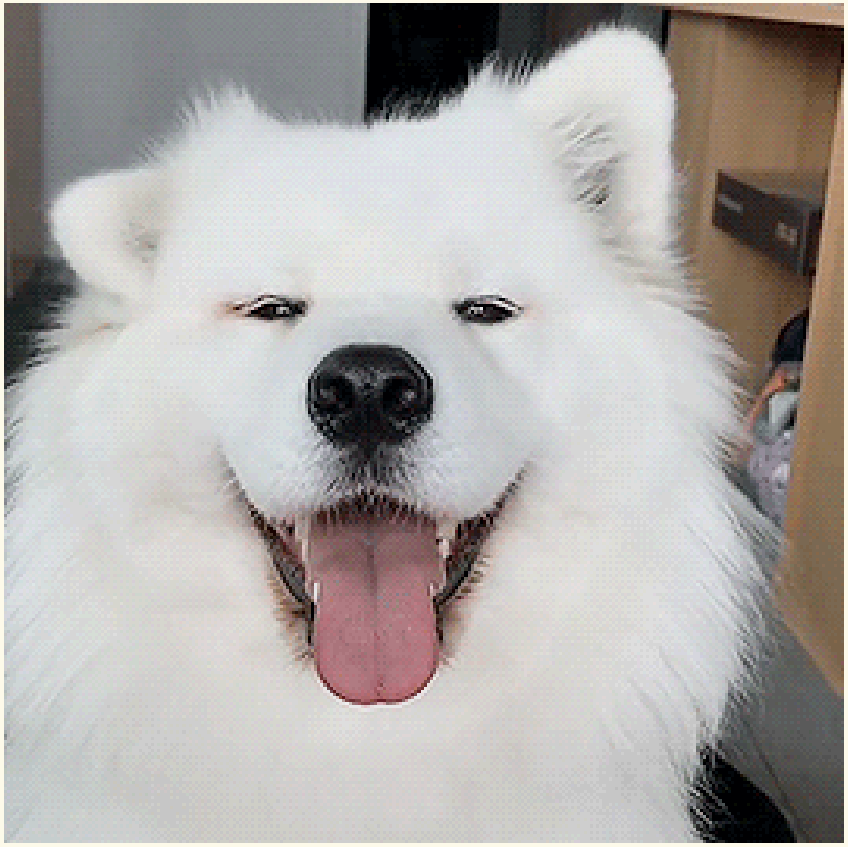
$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)\, dx = \int_{-\infty}^{\infty} x \cdot y\, e^{-x/y} dx = y$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y]\, f_Y(y)\, dy = \int_{-\infty}^{\infty} y \cdot 2\, e^{-y/2} dx = 2$$

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	$\forall x, y, p_{X,Y}(x,y) = p_X(x) p_Y(y)$	$\forall x, y, f_{X,Y}(x,y) = f_X(x) f_Y(y)$

Brain Break



Agenda

- Joint Distributions
  - Another LTE example
  - Conditional expectation and LTE for continuous RVs
- Covariance
- Tail Bounds
  - Markov's Inequality

Covariance: How correlated are X and Y?

Recall that if X and Y are independent, $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Definition: The **covariance** of random variables X and Y,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Unlike variance, covariance can be positive or negative. It has has value 0 if the random variables are independent.

$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Two Covariance examples:

Suppose $X \sim \mathrm{Bernoulli}(p)$

If random variable $Y = X$ then

$$\mathrm{Cov}(X, Y) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}(X) = p(1-p)$$

If random variable $Z = -X$ then

$$\begin{aligned} \mathrm{Cov}(X, Z) &= \mathbb{E}[XZ] - \mathbb{E}[X] \cdot \mathbb{E}[Z] \\ &= \mathbb{E}[-X^2] - \mathbb{E}[X] \cdot \mathbb{E}[-X] \\ &= -\mathbb{E}[X^2] + \mathbb{E}[X]^2 = -\mathrm{Var}(X) = -p(1-p) \end{aligned}$$

Agenda

- Joint Distributions
  – Another LTE example
  – Conditional expectation and LTE for continuous RVs
- Covariance
- Tail Bounds
  – Markov's Inequality

Tail Bounds (Idea)

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

$$P(X \geq a) \leq b$$
$$P(|X - \mathbb{E}[X]| \geq a) \leq b$$

Useful tool when

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

Markov's Inequality

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

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

(Alternative form) For any $k \geq 1$,

$$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$$

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 I

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

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

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

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

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

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

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

Follows by re-arranging terms …

Markov's Inequality – Proof II

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

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

$$\mathbb{E}[X] = \int_0^\infty x \cdot f_X(x)\, \mathrm{d}x$$

$$= \int_t^\infty x \cdot f_X(x)\, \mathrm{d}x + \int_0^t x \cdot f_X(x)\, \mathrm{d}x$$

$$\geq \int_t^\infty x \cdot f_X(x)\, \mathrm{d}x$$

$$\geq \int_t^\infty t \cdot f_X(x)\, \mathrm{d}x = t \cdot \int_t^\infty f_X(x)\, \mathrm{d}x = t \cdot P(X \geq t)$$

so $P(X \geq t) \leq \mathbb{E}[X]/t$ as before

Example – Geometric Random Variable

Let X be geometric RV with parameter p

$$P(X = i) = (1 - p)^{i-1}p \qquad\qquad \mathbb{E}[X] = \frac{1}{p}$$

"X is the number of times Alice needs to flip a biased coin until she sees heads, if heads occurs with probability p?

What is the probability that $X \geq 2\mathbb{E}[X] = 2/p$?

Markov's inequality: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Example

$$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$$

Suppose that the average number of ads you will see on a website is 25. Give an upper bound on the probability of seeing a website with 75 or more ads.

Poll: pollev.com/paulbeame028

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

Example

$P(X \geq k \cdot \mathbb{E}[X]) \leq \frac{1}{k}$

Suppose that the average number of ads you will see on a website is 25. Give an upper bound on the probability of seeing a website with 20 or more ads.

Poll: pollev.com/paulbeame028

a. $0 \leq p < 0.25$
b. $0.25 \leq p < 0.5$
c. $0.5 \leq p < 0.75$
d. $0.75 \leq p$
e. Unable to compute

Example – Geometric Random Variable

Let X be geometric RV with parameter p

$$P(X = i) = (1-p)^{i-1}p \qquad \mathbb{E}[X] = \frac{1}{p}$$

"X is … *e sees heads, if* … *n probability p?*

Next time we will see that we can get better tail bounds using variance

What is the probability that $X \geq 2\mathbb{E}[X] = 2/p$?

Markov's inequality: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$