CSE 312

Foundations of Computing II

Lecture 20: Tail Bounds

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

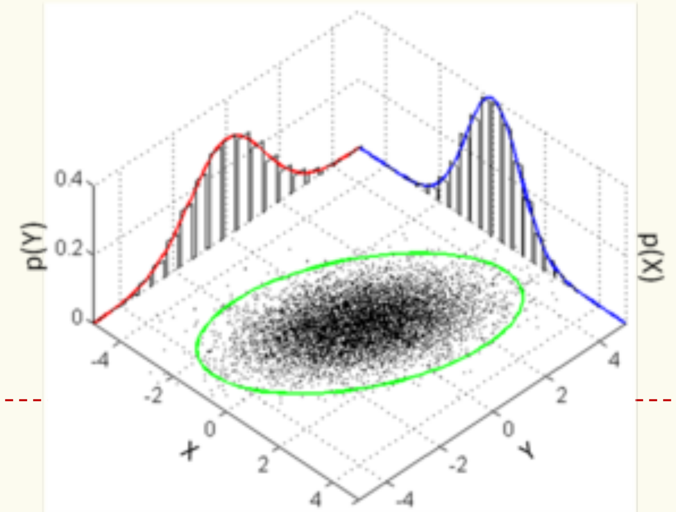
Definition. The **joint probability density function (PDF)** of continuous random variables X and Y is a function $f_{X,Y}$ defined on $\mathbb{R}\times\mathbb{R}$ such that

- $f_{X,Y}(x,y) \geq 0$ 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 Law of Total Expectation

Law of Total Expectation (event version). Let X be a random variable and let events $A_1, \ldots, A_n$ 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

- Covariance
- Markov's Inequality
- Chebyshev'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,

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

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

- Covariance
- Markov's Inequality
- Chebyshev'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 expectation. 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

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

"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: $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

Can we do better?

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/stefanotessaro617

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/stefanotessaro617

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

Brain Break

Agenda

- Covariance
- Markov's Inequality
- Chebyshev's Inequality

Chebyshev's Inequality

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

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

Proof: Define $Z = X - \mathbb{E}[X]$

Definition of Variance

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

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

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

Example – Geometric Random Variable

Let X be geometric RV *with* parameter p

$$P(X = i) = (1-p)^{i-1}p \qquad \mathbb{E}[X] = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

What is the probability that $X \geq 2\mathbb{E}(X) = 2/p$*?*

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

Chebyshev: $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = 1 - p$

Better if $p > 1/2$ ☺

Example

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

Suppose that the average number of ads you will see on a website is 25 and the standard deviation of the number of ads is 4. Give an upper bound on the probability of seeing a website with 30 or more ads.

Poll: pollev.com/stefanotessaro617

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

Chebyshev's Inequality – Repeated Experiments

"How many times does Alice need to flip a biased coin until she sees heads n times, if heads occurs with probability p?

X = # of flips until n times "heads"

X_i = # of flips between $(i-1)$-st and i-th "heads"

$$X = \sum_{i=1}^{n} X_i$$

Note: $X_1, \dots, X_n$ are independent and geometric with parameter p

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p}$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

Chebyshev's Inequality – Coin Flips

"How many times does Alice need to flip a biased coin <u>until she sees heads</u> n times, if heads occurs with probability p?

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p} \qquad \mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n(1-p)}{p^2}$$

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

<u>Markov:</u> $P(X \geq 2\mathbb{E}[X]) \leq \frac{1}{2}$

<u>Chebyshev:</u> $P(X \geq 2\mathbb{E}[X]) \leq P(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} = \frac{1-p}{n}$

Goes to zero as $n \to \infty$ ☺

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.

- Very often loose upper-bounds are okay when designing for the worst case

Generally (but not always) making more assumptions about your random variable leads to a more accurate upper-bound.