CMPSCI 240: Reasoning about Uncertainty

Lecture 11: More Tail Bounds and Multiple Random Variables

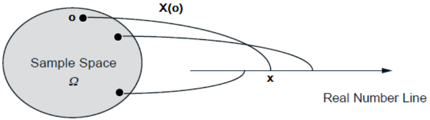
Andrew McGregor

University of Massachusetts

Last Compiled: March 2, 2017

Outline

Random Variables



- Formally, a random variable X is a mapping from Ω to $\mathbb{R}$.
- Given a random variable $X : \Omega \to \mathbb{R}$, we can define events $\{X = k\}$ for each value of $k \in \mathbb{R}$ and the probability of these events is the *probability mass function*.

$$P(X = k) = \sum_{o \in \Omega \text{ such that } X(o) = k} P(o)$$

- It's often sufficient to just consider the probability mass function rather than the probabilities of every event.

Expectation and Variance

- The expected value $E[X]$ is a probability-weighted average of the possible values of the random variable X:

$$E[X] = \sum_k k\, P(X = k)$$

- Variance measures how far we expect a random variable to be from its average.

$$\mathrm{var}(X) = \sum_k (k - E[X])^2 \cdot P(X = k)$$

- The **standard deviation** is the positive square root of the variance:

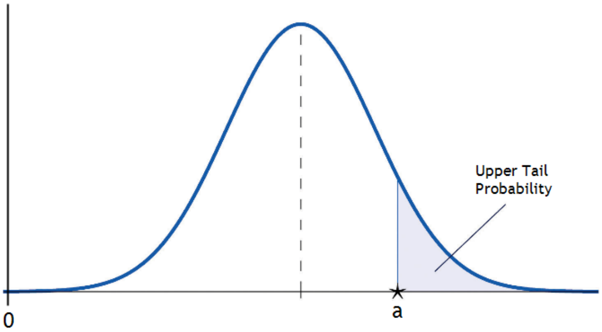
$$\mathrm{std}(X) = \sqrt{\mathrm{var}(X)}\ .$$

Outline

Tail Probabilities

- Given a random variable X, the **upper tail probability** is:

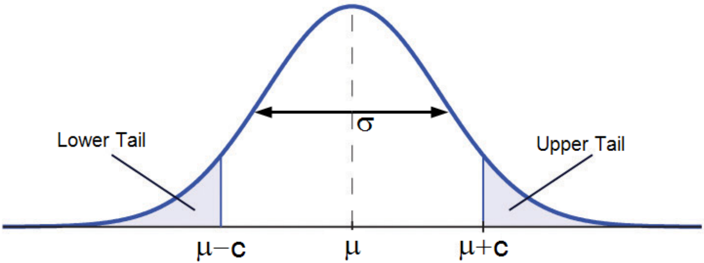
$$P(X \geq a) = \sum_{k \geq a} P(X = k)$$



Double Sided Risk

- Sometimes we want to estimate the lower tail and the upper tail. Write $\mu = E(X)$, then

$$P(|X - \mu| \geq c) \quad = \quad P(X \leq \mu - c) + P(X \geq \mu + c)$$



Markov's Inequality and Chebyshev's Inequality

- **Markov's Inequality:** For any non-negative random variable X and tail threshold $c > 0$:
$$P(X \geq c) \leq \frac{E[X]}{c}$$
- **Chebyshev's Inequality:** For any random variable X and tail threshold $c > 0$:
$$P(|X - \mu| \geq c) \leq \frac{var(X)}{c^2}$$
where $\mu = E(X)$.
- Intuition for Chebyshev: if random variable X has a small variance, then the probability that the value of X is far from μ is small.

Chebyshev's Inequality Proof

- First note that,

$$P(|X-\mu| \geq c) = P((X-\mu)^2 \geq c^2) = P(Y \geq c^2)$$

  where $Y = (X-\mu)^2$.
- Note that Y is a positive random variable and that

$$E[Y] = E[(X-\mu)^2] = var(X)\ .$$

- By Markov's Inequality, $P(Y \geq c^2) \leq \frac{E[Y]}{c^2}$ and hence

$$P(|X-\mu| \geq c) \leq var(X)/c^2$$

Outline

More Examples 1

- You turn up at 8pm to meet your friend but expect it'll be 5 minutes until he shows up. Which of the following is necessarily true?
  - A: The probability you'll wait ≥ 20 minutes is at most $1/4$.
  - B: The probability you'll wait < 20 minutes is at most $1/4$.
  - C: Neither of the above.
- Answer is A by appealing to the Markov Bound.

More Examples 2

- You're in chemistry class and need to measure the pH of a mysterious substance which is actually 7. When you take reading, the expected value is 7 but the standard deviation is 2. Which of the following statements is true
  - A: The probability your reading is between 4 and 10 is at most $4/9$.
  - B: The probability your reading is not between 4 and 10 is at most $4/9$.
  - C: Neither of the above.
- Answer is B by appealing to the Chebyshev Bound.

More Examples 3

- Suppose you are playing poker tomorrow and expect to earn \$10. Which of the following is necessarily true?
  - A: The probability you win $\geq$ \$30 is at most $1/3$.
  - B: The probability you win $<$ \$30 is at most $1/3$.
  - C: Neither of the above.
- Answer is C because earnings can be negative and we don't know the variance.

Outline

Multiple Random Variables

- Consider two random variables, X and Y mapping from Ω to $\mathbb{R}$.
- X defines events of the form $\{X = i\} = \{o \in \Omega \mid X(o) = i\}$ and Y defines events of the form $\{Y = j\} = \{o \in \Omega \mid Y(o) = j\}$
- For $i, j \in \mathbb{R}$, we can define the event

$$\{X = i, Y = j\} = \{X = i\} \cap \{Y = j\} = \{o \in \Omega \mid X(o) = i \text{ and } Y(o) = j\}$$

- The probabilities of these events give the joint PMF of X and Y:

$$P(X = i, Y = j) = P(\{X = i, Y = j\})$$

- Useful for describing **multiple properties** over the outcome space of a single experiment, e.g., pick a random student and let X be their height and Y be their test score.

Tabular Representation of Joint PMFs

P(X,Y)				
X\Y	$Y=1$	$Y=2$	$Y=3$	$Y=4$
$X=1$	0.1	0.1	0	0.2
$X=2$	0.05	0.05	0.1	0
$X=3$	0	0.1	0.2	0.1

- e.g., $P(X{=}2, Y{=}3) = 0.1$, $P(X{=}3, Y{=}1) = 0$, ...
- Given the joint PMF, how can we work out $P(X=i)$ and $P(Y=j)$?

$$P(X=i) = \sum_j P(X=i, Y=j)$$

$$P(Y=j) = \sum_i P(X=i, Y=j)$$

- If we start with the joint PMF of X and Y, we refer to $P(X)$ as the marginal PMF of X and $P(Y)$ as the marginal PMF of Y.

Marginal PMFs

P(X,Y)				
X\Y	1	2	3	4
1	0.1	0.1	0	0.2
2	0.05	0.05	0.1	0
3	0	0.1	0.2	0.1

X	P(X)
1	0.4
2	0.2
3	0.4

Marginal PMFs

P(X,Y)				
X\Y	1	2	3	4
1	0.1	0.1	0	0.2
2	0.05	0.05	0.1	0
3	0	0.1	0.2	0.1

Y	1	2	3	4
P(Y)	0.15	0.25	0.3	0.3

Computing Marginals from the Joint Distribution

- Suppose Y takes the values $j_1, j_2, \ldots, j_N$, then

$$\{Y = j_1\}, \{Y = j_2\}, \ldots, \{Y = j_N\}$$

  partition Ω.
- Hence, $\{X = i\}$ can be partitioned into

$$\{X = i\} \cap \{Y = j_1\}, \{X = i\} \cap \{Y = j_2\}, \ldots, \{X = i\} \cap \{Y = j_N\}$$

- Therefore,

$$\begin{aligned} P(X = i) &= P(\{X = i\}) \\ &= P(\{X = i\} \cap \{Y = j_1\}) + P(\{X = i\} \cap \{Y = j_2\}) + \ldots \\ &\qquad \ldots + P(\{X = i\} \cap \{Y = j_N\}) \\ &= \sum_j P(\{X = i\} \cap \{Y = j\}) = \sum_j P(X = i, Y = j) \end{aligned}$$

Another Example 1

P(X,Y)				
X\Y	1	2	3	4
1	0.1	0.1	0	0
2	0	0.05	0.1	0.05
3	0.1	0.2	0.2	0.1

What's the value of $P(X = 2, Y = 3)$?

A: 0

B: 0.1

C: 0.05

D: 0.2

E: 1

Answer is B.

Another Example 2

P(X,Y)				
X\Y	1	2	3	4
1	0.1	0.1	0	0
2	0	0.05	0.1	0.05
3	0.1	0.2	0.2	0.1

What's the value of $P(X = 3)$?

A: 0.1

B: 0.4

C: 0.05

D: 0.6

E: 1

Answer is D.

Outline

Functions of Two Random Variables

- Given two random variables X and Y and a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X, Y)$$

is a new random variable where

$$E(Z) = \sum_{x,y} f(x, y) P(X = x, Y = y) \text{ and } var(Z) = E(Z^2) - (E(Z))^2 .$$

- Linearity of Expectation: If $Z = X + Y$,

$$E(Z) = E(X + Y) = E(X) + E(Y)$$

- Linearity of Variance: If $Z = X + Y$,

$$var(Z) = var(X + Y) = var(X) + var(Y)$$

if X and Y are independent, i.e., for all i, j

$$P(X = i, Y = j) = P(X = i)P(Y = j) .$$

Functions of Multiple Random Variables

- Given random variables $X_1, X_2, \ldots, X_N$ and a function $f : \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} \to \mathbb{R}$,

$$Z = f(X_1, X_2, \ldots, X_N)$$

  is a new random variable.
- Linearity of Expectation: If $Z = \sum_{i=1}^{N} X_i$,

$$E(Z) = E(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} E(X_i)$$

- Linearity of Variance: If $Z = \sum_{i=1}^{N} X_i$,

$$var(Z) = var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} var(X_i)$$

  if $X_1, \ldots, X_N$ are pairwise independent, i.e., for all a, b

$$P(X_i = a, X_j = b) = P(X_i = a)P(X_j = b) \ .$$

Expectation and Variance of the Binomial Distribution

- Let X be the number of heads when tossing a coin n times that lands heads with probability p. Then X is a binomial random variable,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Let X_i be 1 if the ith coin toss is heads, then

$$X = X_1 + X_2 + X_3 + \ldots + X_n$$

- For each $E[X_i] = p$ and $var[X_i] = p(1-p)$
- Therefore,

$$E[X] = \sum_i E[X_i] = np$$

$$var[X] = \sum_i var[X_i] = np(1-p)$$