

MATH 180A: Introduction to Probability

Lecture A00 (Au)

www.math.ucsd.edu/~bau/w21.180a

Lecture B00 (Nemish)

www.math.ucsd.edu/~ynemish/teaching/180a

Today: Sums of random variables

Next: ASV 9.1-9.2

Video: Prof. Todd Kemp, Fall 2019

Week 9:

- Homework 7 (due Friday, March 5)
- CAPE at www.cape.ucsd.edu

We can now come back to questions about sums of random variables – in the context of their joint distribution.

8.1

Let X, Y be two (let's say discrete) random variables.

$$\begin{aligned}
 \underbrace{\mathbb{E}(X+Y)}_{g(X,Y)} &= \sum_{k,l} g(k,l) p_{X,Y}(k,l) \leftarrow p_{X,Y}(k,l) = P(X=k, Y=l) \\
 &= \sum_{k,l} (k+l) p_{X,Y}(k,l) = \sum_{k,l} k p_{X,Y}(k,l) + \sum_{k,l} l p_{X,Y}(k,l) \\
 &= \sum_k k \underbrace{\sum_l p_{X,Y}(k,l)}_{P_X(k)} + \sum_l l \underbrace{\sum_k p_{X,Y}(k,l)}_{P_Y(l)} \\
 &= \sum_k k p_X(k) + \sum_l l p_Y(l) = \mathbb{E}(X) + \mathbb{E}(Y).
 \end{aligned}$$

Theorem: For any random variables X_1, X_2, \dots, X_n ,

$$\mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n).$$

Eg. $S \sim \text{Bin}(n, p)$. $P(S=k) = \binom{n}{k} p^k (1-p)^{n-k}$ $0 \leq k \leq n$.

This means $S = X_1 + X_2 + \dots + X_n$ where $X_1, \dots, X_n \sim \text{Ber}(p)$

$$\begin{aligned}\therefore E(S) &= E(X_1) + E(X_2) + \dots + E(X_n) & E(X_j) &= P(X_j=1) = p \\ &= p + p + \dots + p & (\text{in of them}) \\ &= np.\end{aligned}$$

A binomial is a sum of Bernoullis (indicator r.v.'s).

Lots of problems can be solved when we can express desired events in terms of sums of indicators.

Eg. Suppose we put 200 balls randomly into 100 boxes. What

is the expected number of empty boxes?

$$X_i := \mathbb{1}_{\{\text{Box } i \text{ is empty}\}} \quad 1 \leq i \leq 100$$

$$\begin{aligned}X &:= \#\text{of empty boxes} = \sum_{i=1}^{100} X_i \\ \therefore E(X) &= \sum_{i=1}^{100} E(X_i) \quad \rightarrow 100(0.9)^{200}\end{aligned}$$

$$\mathbb{E}(X_i) = P(X_i=1) = P(\text{Box } i \text{ is empty}) = (0.9)^{200} \approx 13.4$$

Eg. Your favorite cereal (chocolate frosted sugar bombs) comes with a Pokémon figurine. There are ~~n=20~~ to collect. What is the expected number of boxes you need to buy to collect them all?

$X = \# \text{ of boxes you need to collect them all}$

$X_1 = " \quad " \quad \text{1st one} = 1$

$$Y \sim \text{Geom}(p)$$

$$\mathbb{E}(Y) = 1/p$$

$X_2 = \# \text{ of boxes after the } X_1^{\text{th}} \text{ needed to collect the 2nd. } \sim \text{Geom}\left(\frac{n-1}{n}\right)$

\vdots

$X_j = " \quad " \quad X_{j-1}^{\text{th}}$

$$X_j \sim \text{Geom}\left(\frac{n-j+1}{n}\right)$$

~ 0.577
E.M.
const.

$$X = X_1 + X_2 + \dots + X_n$$

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n) = 1 + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{1} \\ &= n \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right) \approx n \ln n + \gamma n \end{aligned}$$

$\underbrace{}_{H_n}$

\downarrow
 $+ O\left(\frac{1}{n}\right)$

$$(n=20) : \mathbb{E}(X) = 71.95$$

Sums & Variances

8.2 /

8.4

$$\begin{aligned}
 \text{Var}(X+Y) &= \mathbb{E}((X+Y)^2) - (\mathbb{E}(X+Y))^2 \\
 &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\
 &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2 \\
 &= \underbrace{\mathbb{E}(X^2) - \mathbb{E}(X)^2}_{\text{Var}(X)} + \underbrace{\mathbb{E}(Y^2) - \mathbb{E}(Y)^2}_{\text{Var}(Y)} + 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y))
 \end{aligned}$$

Def: $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$ $\stackrel{\uparrow}{=} \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
 (correlation)

Note: $\text{Cov}(X, X) = \text{Var}(X)$

Theorem: $\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$

$\hookrightarrow \sum_{i,j} \text{Cov}(X_i, X_j)$

$+ \sum_{i \neq j} \text{Cov}(X_i, X_j)$

Covariance & Independence

If X_1, X_2, \dots, X_n are independent, then for $i \neq j$

$$\text{Cov}(X_i, X_j) = \underbrace{\mathbb{E}(X_i X_j)}_{\mathbb{E}(X_i)\mathbb{E}(X_j)} - \mathbb{E}(X_i)\mathbb{E}(X_j) = 0.$$

Corollary: If X_1, X_2, \dots, X_n are independent

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

E.g. $S_n \sim \text{Bin}(n, p)$ $S_n = X_1 + X_2 + \dots + X_n$ $\leftarrow X_j \text{ i.i.d}$

$$\text{Var}(S_n) = n p(1-p) \quad ; \quad \text{Var}(S_n)$$

$$\begin{aligned}
 &= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) \quad \left| \quad \therefore \text{Var}(X_j) = \mathbb{E}(X_j^2) - \mathbb{E}(X_j)^2 \right. \\
 &= p(1-p) + p(1-p) + \dots + p(1-p) \quad \left| \quad = \mathbb{E}(X_j) - \mathbb{E}(X_j)^2 \right. \\
 &= np(1-p).
 \end{aligned}$$

Independent v.s. Uncorrelated

We've seen that independent rv's are uncorrelated.
The converse does not hold.

Eg. $X \sim \text{Unif}\{-1, 0, 1\}$ (ie. $P(X=+1)=P(X=0)=P(X=-1)=\frac{1}{3}$)

$$Y = X^2$$

$$\begin{aligned} E(X) &= \frac{1}{3}(-1) + \frac{1}{3}(0) + \frac{1}{3}(1) \\ &= 0. \end{aligned}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

$$= E(X^3) - E(X)E(X^2)$$

$$\begin{array}{c} \cancel{E(X)} \\ \cancel{0} \end{array}$$

$$X^3 = X.$$

$$= 0.$$

Eg. Coupon Collector (Revisited)

Let T_n be the number of cereal boxes it takes to collect n distinct toys.

$$T_n = 1 + W_1 + W_2 + \dots + W_{n-1}$$

$W_k \sim \text{Geom}\left(\frac{n-k}{n}\right)$ are all independent.

$$\text{Var}(T_n) \approx \frac{\pi^2}{6} n^2$$

Reversion to the Mean

Let X_1, X_2, \dots, X_n be i.i.d. random variables (i.e. sampling, but not just Bernoulli trials.)
Say $\mathbb{E}(X_j) = \mu$, $\text{Var}(X_j) = \sigma^2$.

The sample mean $\bar{X}_n = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$.

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} (\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)) = \mu \quad \text{indep.}$$

$$\begin{aligned}\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n} \sum_{j=1}^n X_j\right) = \frac{1}{n^2} \text{Var}\left(\sum_{j=1}^n X_j\right) \\ &= \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{1}{n^2} \cdot n \sigma^2 \\ &= \frac{\sigma^2}{n}\end{aligned}$$