MATH 180A: INTRO TO PROBABILITY
(FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: § 8.1 - 8.4

Next: § 9.1 - 9.2

Homework 7 Due Wednesday, Nov 27

Exam 1: graded & released. (Mean 63%, St.Dev. 20%)

↳ Topics to focus on reviewing:

- Chebyshev's Inequality
- Transforming densities
- MGF (esp. when it determines the distribution)

8.1

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

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

$\mathbb{E}(X+Y)$

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

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

Eg. $S \sim \text{Bin}(n,p)$.

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

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?

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

Sums & Variances

8.2/8.4

$\mathrm{Var}(X+Y)$

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

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

Theorem: $\mathrm{Var}(X_1+X_2+\cdots+X_n) = \mathrm{Var}(X_1)+\mathrm{Var}(X_2)+\cdots+\mathrm{Var}(X_n)$

$+$

Covariance & Independence

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

$$\mathrm{Cov}(X_i, X_j)$$

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

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) =$$

Eg. $S_n \sim \mathrm{Bin}(n, p)$

Independent v.s. Uncorrelated

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

E.g. $X \sim \text{Unif}\{-1,0,1\}$ (i.e. $\mathbb{P}(X=\pm 1)=\mathbb{P}(X=0)=\frac{1}{3}$)
$Y = X^2$.

E.g. Coupon Collector (Revisited)

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

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

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

Reversion to the Mean

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables (i.e. sampling, but not just Bernoulli trials.)

Say $\mathbb{E}(X_j) = \mu$, $Var(X_j) = \sigma^2$.

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

$\mathbb{E}(\bar{X}_n) =$

$Var(\bar{X}_n) =$

Moment Generating Function Revisited

8.3

Suppose X, Y are independent, and $M_X, M_Y < \infty$ on an interval containing 0. Then

$$M_{X+Y}(t)$$

E.g. $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ independent

$M_X(t) = e^{\lambda(e^t-1)}$ $\quad$ $M_Y(t) = e^{\mu(e^t-1)}$

Eg. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independent

Sample Variance

Collect some data $X_1, X_2, \ldots, X_n$ (random independent samples of the same distribution)

The true mean μ and variance σ^2 are unknown.

We've seen the best estimator for μ is $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$.

What about for σ^2?

$$\mathrm{Var}(X) = \mathbb{E}\left((X-\mu)^2\right)$$

$$\overline{\sigma}_n^2 \stackrel{?}{=} \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$$

replace w average?

replace w $\overline{X}_n$