Lecture 6

Binomial Distribution

X — # of heads out of n coin flips with $P(H) = p$

$$X_i = \begin{cases} 1 & \text{if coin } i \text{ is head} \\ 0 & \text{oth} \end{cases}$$

$X = \sum_{i=1}^{n} X_i$ ✓

$E[X] = \sum_{i=1}^{n} E[X_i] = np$

$E[X_i] = p$

$E[X^2] = E\left[\sum_{i=1}^{n} X_i \sum_{i'=1}^{n} X_{i'} \right]$

$= \sum_{i=1}^{n} \sum_{i'=1}^{n} E[X_i X_{i'}] = \sum_{i=1}^{n} p + \sum_{i=1}^{n} \sum_{i'=1, i' \neq i}^{n} p^2 = np + n(n-1)p^2$

$Var(X) = np - np^2 = np(1-p)$

$E[X_i X_{i'}] = P[X_i X_{i'} = 1] = P[X_i = 1 \wedge X_{i'} = 1]$

this includes $i = i'$

$$= \begin{cases} \text{case } i = i' & P[X_i = 1] = p \\ i \neq i' & P[X_i = 1] \, P[X_{i'} = 1] = p^2 \quad \text{indep} \end{cases}$$

Q: is $X_i X_{i'}$ a 0-1 R.V?

Property of 0-1 R.V:

products (powers) of 0-1 R.V are 0-1 R.V

Result is 1 iff all variables are 1

Sampling

D is a set of reals

$f(D) = \sum_{x \in D} x$

Take D' a sample

$\hat{f}(D) = \hat{f}(D')$

want $\hat{f}$ to reasonably approximate $f(D)$

$X = \hat{f}(D') = \frac{|D'|}{|D|} \sum_{x \in D'} x$

α

Case 1: D' is a sample with replacement.

Idea: Express X in terms of simpler RV

X_i - R.V. that gives the outcome (x) in position i of D'

$X = \alpha \sum_{i=1}^{|D'|} X_i$

$\sum_{x \in D} x$

Q: What is known about X_i's

X_i's are independent

X_i is uniform random over D

$E(X_i) = \sum_{x \in D} \underbrace{P(X_i = x)}_{\frac{1}{|D|}} x = \frac{\sum_{x \in D} x}{|D|}$

X_i's are i.i.d.

$E(X) = \alpha \sum_{i=1}^{|D'|} E[X_i]$

$= \alpha \cdot \frac{|D'|}{|D|} \sum_{x \in D} x$

$= f(D)$

$\text{Var}(X) = \alpha^2 \sum_{i=1}^{|D'|} \text{Var}(X_i)$

$= \alpha^2 |D'|$

$E[X_i^2] = \sum_{x \in D} P(X_i = x) x^2$

$= \frac{\sum_{x \in D} x^2}{|D|}$

$\text{Var}(X_i) = \frac{\sum_{x \in D} x^2}{|D|} - \left(\frac{\sum x}{|D|}\right)^2$

Case 2: D' is a sample without replacement

Same idea

X_i – r.v. that gives outcome (val) in position i of D'

$X = \sum_{i=1}^{|D'|} X_i$

what is the distrib of X_i?

X_1 is uniform distributed

Q: X_2

A: X_2 is uniform if X_1 is not known

$E(X_i) = \frac{\sum_{x \in D} x}{|D|}$ $E[X] = f(D)$

What about interaction?

Alternative modeling:

Important info: what is in D' (irrespective of position)

For every $y \in D$

Let $X_y = \begin{cases} 1 & \text{if } y \in D' \\ 0 & \text{oth.} \end{cases}$

$X = \sum_{y \in D'} y = \sum_{y \in D} X_y \cdot y$

$E(X) = \sum_{y \in D} E(X_y) \cdot y = \sum_{y \in D} y = f(D)$

$E(X_y) = P[y \in D'] = \frac{|D'|}{|D|}$

$E[\sum_{x \in D} \ldots] \neq \sum_{x \in D'} E(\ldots)$

$Var(X)$