STOR 435.001 Lecture 9

Random Variables - III: Geometric random variables. Sums of Random variables

Jan Hannig

UNC Chapel Hill

Plan for this lecture

- We have seen some examples of “important” families of discrete random variables (Binomial, Poisson).
- Plan: see one more example.
- Then come to a super useful fact about expectations.
- Starting from **next Lecture**: continuous random variables; *Calculus strikes back!*

Definition

A random variable X is called geometric if it takes values $1, 2, \ldots$ with probabilities

$$P(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, \ldots$$

for some $p \in (0, 1)$.

Note: $\sum_{n=1}^{\infty}(1-p)^{n-1}p = 1$. What is the interpretation of X?

☼

Random variables

X is geometric with parameter p: What are EX and $Var(X)$?

1. MGF $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=1}^{\infty} e^{tk}\mathbb{P}(X = k) =$

2. So $\mathbb{E}(X) = \frac{d}{dt} M_X(t)|_{t=0} =$

3. So $\mathbb{E}(X^2) = \frac{d^2}{dt^2} M_X(t)|_{t=0} =$

4. So Var$(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 =$

☼

(Just as we did for Binomial/Poisson, can also get above formulae by direction calculation instead if using MGF.)

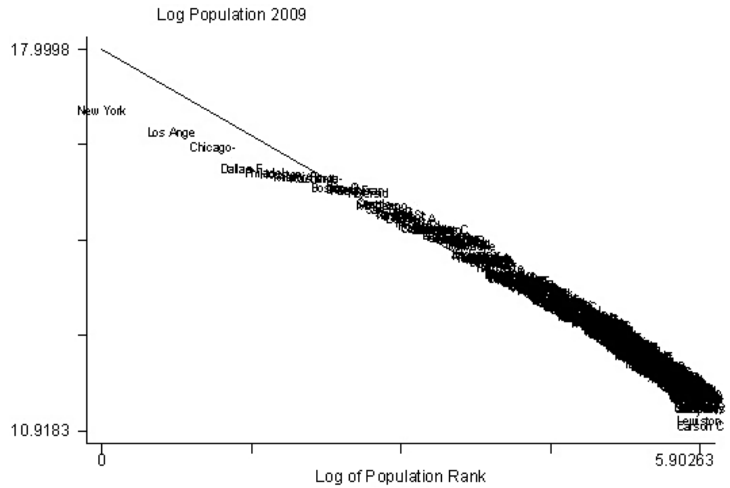
Other discrete distributions not covered in this course but important for applications:

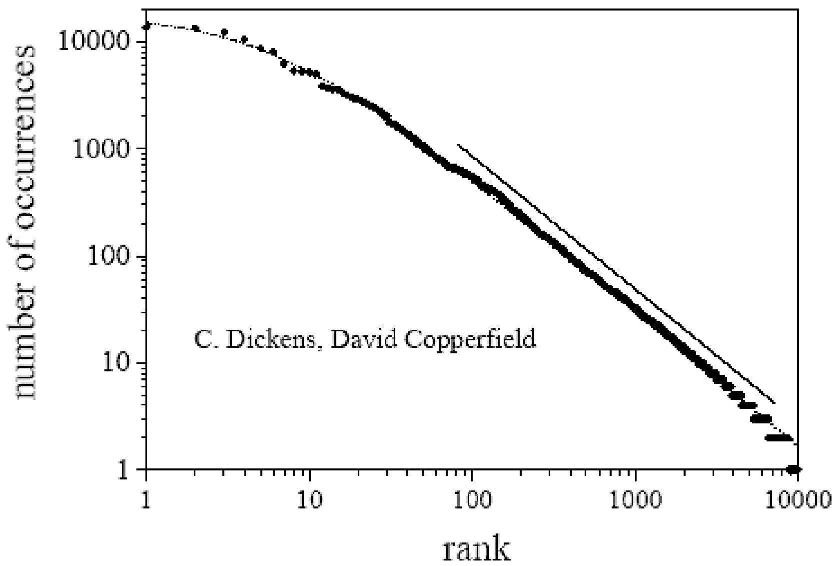
- Negative Binomial
- Hypergeometric random variable
- Zipf's distribution[1]
- Benford's law[2]
- ...

[1] Extras on Sakai: “ZIPF'S LAW”, “Zipfs law unzipped”, “There is More than a Power Law in Zipf”, “Power laws Pareto distributions and Zipfs law”

[2] Extras on Sakai: “The first digit problem”, “The First Digit Phenomenon”, “The Significant-Digit Phenomenon”, “Breaking the Benford Law”

Zipf's law:


Log Population 2009
17.9998
New York
Los Ange
Chicago-
10.9183
0
5.90263
Log of Population Rank


number of occurrences
10000
1000
100
10
1
C. Dickens, David Copperfield
1
10
100
1000
10000
rank

Super Duper Useful fact

For random variables $X_1, \ldots, X_n$,

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} EX_i.$$

Example: Jan has 10 students in his class. The probability that person i will do well in the class is given by p_i (for $1 \leq i \leq 10$). Find the expected total number of students who do well in his class.

☼

Example

1. You Throw a fair dice 3 times. Let X be the sum of the values of the first 2 throws and Y be the sum of the values on the last 2 throws. What is $\mathbb{E}(X+Y)$?
2. You throw a fair dice n times. Let X denote the total of all your throws. What is $\mathbb{E}(X)$?

Coupon collector problem

A brand of cereal called "Super sugary cereal" had decided to package each of its boxes with one of 100 different X-men character cards. Each box of cereal, one of the 100 different cards are randomly chosen and packaged in the box. You keep buying boxes till you get all 100 cards.

1. Let X be the number of boxes you need to buy to collect all 100 cards. What is $\mathbb{E}(X)$?
2. Suppose each box costs $ 2.50. Let Z be the amount you spend to get all 100 cards. What is $\mathbb{E}(Z)$?

☼

Sums of random variables: Coupon collector problem

Concluding Remarks about Coupon collector

(a) While it seems like a “made up” problem, turns out to be super important in a wide array of applications.

(b) See “Extras” for two papers: one on applications on CS and another for applications in genetics.

(c) **Punchline:** The key math tool that helped us solve this problem was: “$\mathbb{E}$(sum of a bunch of random variables) $=$ sum of the individual expectations”

(d) You will see more examples in your HW why this is such an important technique.