

~~*The sum of binomially distributed random variables*~~

Let the X_i ($i = 1, \dots, k$) be binomially distributed with the same parameter p (but with different n_i). Then the distribution of their sum is distributed as

$$X_1 + \dots + X_k \sim \text{Bin}(n_1 + \dots + n_k, p)$$

because the sum represents the number of successes in a sequence of $n_1 + \dots + n_k$ identical Bernoulli trials.

Multinomial distribution

Consider a sequence of n identical trials but now each trial has k ($k \geq 2$) different outcomes. Let the probabilities of the outcomes in a single experiment be $p_1, p_2, \dots, p_k$ ($\sum_{i=1}^k p_i = 1$)

Denote the number of occurrences of outcome i in the sequence by N_i . The problem is to calculate the probability $p(n_1, \dots, n_k) = P\{N_1 = n_1, \dots, N_k = n_k\}$ of the joint event $\{N_1 = n_1, \dots, N_k = n_k\}$.

Define the generating function of the joint distribution of several random variables $N_1, \dots, N_k$ by

$$G(z_1, \dots, z_k) = E[z_1^{N_1} \dots z_k^{N_k}] = \sum_{n_1=0}^{\infty} \dots \sum_{n_k=0}^{\infty} p(n_1, \dots, n_k) z_1^{n_1} \dots z_k^{n_k}$$

After one trial one of the N_i is 1 and the others are 0. Thus the generating function corresponding one trial is $(p_1 z_1 + \dots + p_k z_k)$.

The generating function of n independent trials is the product of the generating functions of a single trial, i.e. $(p_1 z_1 + \dots + p_k z_k)^n$.

From the coefficients of different powers of the z_i variables one identifies

$$p(n_1, \dots, n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

when $n_1 + \dots + n_k = n$,
0 otherwise

Perf Eval of Comp Systems

7.3. Geometric(p) distribution

Repeating independent Bernoulli(p) experiments until the first success.
 p is the probability of success

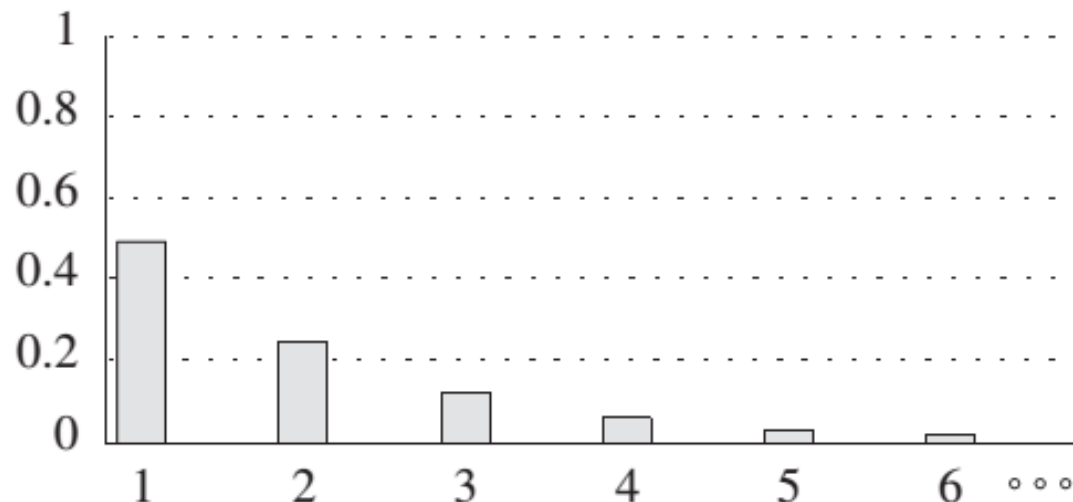
Definition: If $X \sim \text{Geometric}(p)$, then X represents the number of trials until we get a success. The p.m.f. of r.v. X is defined as follows:

$$p_X(i) = \mathbf{P}\{X = i\} = (1 - p)^{i-1} p, \quad \text{where } i = 1, 2, 3, \dots \quad (105)$$

Mean and variance take the following form:

$$E[X] = \frac{p}{1 - p} \quad V[X] = \frac{p}{(1 - p)^2} \quad (106)$$

Geometric(p), $p=0.5$



View 1- **Shifted Geometric distribution** $X \sim \text{Geom}(p)$

X represents the **number of failures** in a sequence of independent Bernoulli trials (with the probability of success p) needed before the first success occurs

$$p_i = P\{X = i\} = (1 - p)^i p \quad i = 0, 1, 2, \dots$$

$$G(z) = p \sum_{i=0}^{\infty} (1 - p)^i z^i = \frac{p}{1 - (1 - p)z}$$

$$E[X] = \frac{1 - p}{p}$$

$$V[X] = \frac{(1 - p)}{p^2}$$

View 2: **Geometric distribution** $X \sim \text{Geom}(p)$

X represents the number of **trials** in a sequence of independent Bernoulli trials (with the probability of success p) needed until the first success occurs

$$p_i = P\{X = i\} = (1 - p)^{i-1}p \quad i = 1, 2, \dots$$

Generating function

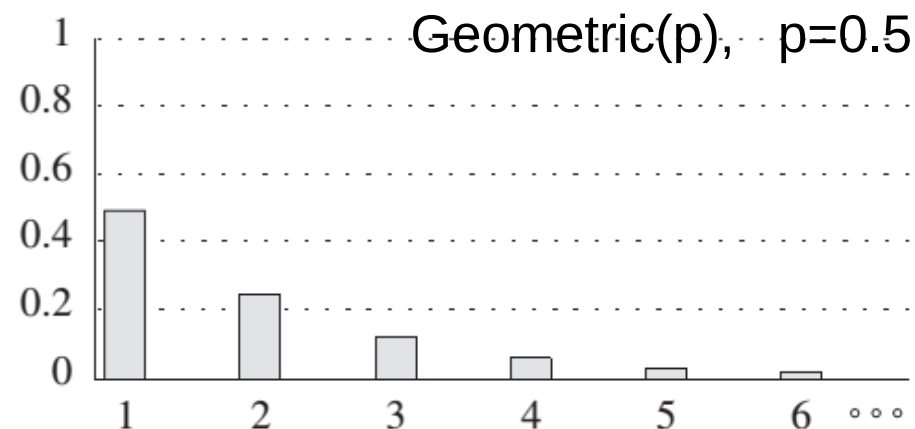
$$G(z) = p \sum_{i=1}^{\infty} (1 - p)^{i-1} z^i = \frac{pz}{1 - (1 - p)z}$$

This can be used to calculate the expectation and the variance:

$$E[X] = G'(1) = \left. \frac{p(1 - (1 - p)z) + p(1 - p)z}{(1 - (1 - p)z)^2} \right|_{z=1} = \frac{1}{p}$$

$$E[X^2] = G''(1) + G'(1) = \frac{1}{p} + \frac{2(1 - p)}{p^2}$$

$$V[X] = E[X^2] - E[X]^2 = \frac{(1 - p)}{p^2}$$



Geometric distribution (continued)

The probability that for the first success one needs more than n trials

$$P\{X > n\} = \sum_{i=n+1}^{\infty} p_i = (1-p)^n$$

Memoryless property of geometric distribution

$$\begin{aligned} P\{X > i+j \mid X > i\} &= \frac{P\{X > i+j \cap X > i\}}{P\{X > i\}} = \frac{P\{X > i+j\}}{P\{X > i\}} \\ &= \frac{(1-p)^{i+j}}{(1-p)^i} = P\{X > j\} \end{aligned}$$

If there have been i unsuccessful trials then the probability that for the first success one needs still more than j new trials is the same as the probability that in a completely new sequence of trials one needs more than j trials for the first success.

This is as it should be, since the past trials do not have any effect on the future trials, all of which are independent.

Negative binomial distribution $X \sim \text{NBin}(n, p)$

X is the number of trials needed in a sequence of Bernoulli trials needed for n successes.

If $X = i$, then among the first $(i - 1)$ trials there must have been $n - 1$ successes and the trial i must be a success. Thus,

$$\Pr\{X = i\} = \binom{i-1}{n-1} p^{n-1} (1-p)^{i-n} \cdot p = \binom{i-1}{n-1} p^n (1-p)^{i-n} \quad \begin{array}{l} \text{if } i \geq n \\ 0 \text{ otherwise} \end{array}$$

The number of trials for the first success $\sim \text{Geom}(p)$. Similarly, the number of trials needed from that point on for the next success etc. Thus,

$$X = X_1 + \cdots + X_n \quad \text{where } X_i \sim \text{Geom}(p) \quad (i.i.d.)$$

Now, the generating function of the distribution is

$$G(z) = \left(\frac{pz}{1 - (1-p)z} \right)^n \quad \begin{array}{l} \text{The point probabilities given above} \\ \text{can also be deduced from this g.f.} \end{array}$$

The expectation and the variance are n times those of the geometric distribution

$$E[X] = \frac{n}{p} \quad V[X] = n \frac{1-p}{p^2}$$