



LECTURE NOTE

ON

PROBABILITY AND STATISTICS 2

BY

ASSIST. PRF. DR. MUSTAFA I. NAIF

DEPARTMENT OF MATHEMATICS COLLEGE OF EDUCATION FOR PURE SCIENCE UNIVERISTY OF ANBAR

Dutline :- LECTURE 6#

Discrete distributions
 6- Geometric distribution

Definition

Expected value Variance

Moment generating function

Characteristic function

Distribution function

Solved exercises

Exercises

• If X represents the total number of successes in n independent Bernoulli trials, then the random variable X~BIN(n,p), where p is the probability of success of a single Bernoulli trial and the probability mass function of X is given by:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, ..., n.$$

Now, Let X denote the trial number on which the first success occurs. Hence the probability that the first success occurs on x *th* trial is given by:

$$f(x) = P(X = x) = (1 - p)^{x-1} p.$$



Definition: A random variable X has a geometric distribution if its probability mas function is given by : $f(x) = (1-p)^{x-1}p \qquad x = 1, 2, 3, ..., \infty,$ Check that

where p denotes the probability of success in a single Bernoulli trial. If X has a geometric distribution we denote it as $X \sim GEO(p)$.

Example: The probability that a machine produces a defective item is 0.02. Each item is checked as it is produced. Assuming that these are independent trials, what is the probability that at least 100 items must be checked to find one that is defective?



Answer: Let X denote the trial number on which the first defective item is observed. We want to find :

$$P(X \ge 100) = \sum_{x=100}^{\infty} f(x)$$

= $\sum_{x=100}^{\infty} (1-p)^{x-1} p$
= $(1-p)^{99} \sum_{y=0}^{\infty} (1-p)^{y} p$
= $(1-p)^{99}$
= $(0.98)^{99} = 0.1353.$

Hence the probability that at least 100 items must be checked to find one that is defective is 0.1353.

Theorem: If X is a geometric random variable with parameter p, then the mean, variance and moment generating functions are respectively given by: $\mu_X = \frac{1}{p}$, $\sigma_X^2 = \frac{1-p}{p^2}$, $M_X(t) = \frac{p e^t}{1-(1-p) e^t}$, if t < -ln(1-p).

Proof: First, we compute the moment generating function of X and then we generate all the mean and variance of X from it.

$$M(t) = \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p$$

$$= p \sum_{y=0}^{\infty} e^{t(y+1)} (1-p)^{y}, \quad \text{where } y = x-1$$

$$= p e^{t} \sum_{y=0}^{\infty} (e^{t} (1-p))^{y}$$

$$= \frac{p e^{t}}{1-(1-p) e^{t}}, \quad \text{if } t < -ln(1-p).$$
Pifferentiating M(t) with respect to t, we
$$M'(t) = \frac{(1-(1-p) e^{t}) p e^{t} + p e^{t} (1-p) e^{t}}{[1-(1-p) e^{t}]^{2}}$$

$$= \frac{p e^{t}}{[1-(1-p) e^{t}]^{2}}.$$
Hence
$$\mu_{X} = E(X) = M'(0) = \frac{1}{p}.$$

D

obtain

Similarly, the second derivative of M(t) can be obtained from the first derivative as:

$$M''(t) = \frac{\left[1 - (1 - p)e^{t}\right]^{2} p e^{t} + p e^{t} 2 \left[1 - (1 - p)e^{t}\right] (1 - p)e^{t}}{\left[1 - (1 - p)e^{t}\right]^{4}}$$
 Hence, $M''(0) = \frac{p^{3} + 2 p^{2}(1 - p)}{p^{4}} = \frac{2 - p}{p^{2}}$.
Therefore, the variance of X is:
$$\sigma_{X}^{2} = M''(0) - (M'(0))^{2}$$
$$= \frac{2 - p}{p^{2}} - \frac{1}{p^{2}}$$
$$= \frac{1 - p}{p^{2}}.$$

Theorem. The cumulative distribution function of a geometric random variable *X* is: $F(X) = P(X \le x) = 1 - (1 - p)^{x}$

Proof: $P(X \le k) = 1 - P(X > k)$ But $P(X > k) = P(X \ge k+1) = \sum_{x=k+1}^{\infty} (1-p)^{x-1}p$ $= p\left[(1-p)^k \left[1+(1-p) + (1-p)^2 + \cdots\right]\right]$ $= p\left[(1-p)^k \left[\frac{1}{1-(1-p)}\right]\right] = (1-p)^k \longrightarrow P(X \le k) = 1 - (1-p)^k$

Theorem: The characteristic function of a geometric random variable *X* is:

$$\varphi(t) = \frac{pe^{it}}{1 - (1 - p)e^{it}}$$

Proof: Similar to the proof of m.g.f.

Example: If the probability of engine malfunction during any one-hour period is p = .02 and Y denotes the number of one-hour intervals until the first malfunction, find the mean and standard deviation of Y.

Solution : Y has a geometric distribution with p = .02. Then: E(Y) = 1/p = 1/(.02) = 50,

V(Y) = .98/.0004 = 2450, and the standard deviation of Y is $\sigma = \sqrt{2450} = 49.497$.

Exercises

1- Suppose that *Y* is a random variable with a geometric distribution. Show that

a
$$\sum_{y} p(y) = \sum_{y=1}^{\infty} q^{y-1} p = 1.$$

b $\frac{p(y)}{p(y-1)} = q$, for y = 2, 3, ... This ratio is less than 1, implying that the geometric probabilities are monotonically decreasing as a function of y. If Y has a geometric distribution, what value of Y is the most likely (has the highest probability)?

2- Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Find the probability that the first applicant with advanced training in programming is found on the fifth interview.

3- Suppose that X has the geometric distribution with parameter p. Show that for every nonnegative integer k,

$$\Pr(X \ge k) = (1-p)^k.$$

SEE YOU IN THE NEXT LECTURE