Some standard discrete distribution

Bernoulli & Binomial

Lecture-9

Example on MGF

EX: On any randomly chosen ball, the probability that a player hits a six is 0.6. Assume that player’s performance on previous balls have no effect on the next ball.

Let X = the number of balls required for the player to hit a six. Find the mgf of X and hence its mean and variance.

Bernoulli Distribution

- Named after Swiss mathematician Jacob Bernoulli
- The discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$
- The probability distribution of any single experiment that asks a yes–no question; the question results in a boolean-valued outcome, a single bit whose value is success/yes/true/one with probability p and failure/no/false/zero with probability q.

Bernoulli Distribution

If X is a random variable with this distribution, then

$$P(X = 1) = p = 1 - P(X = 0) = 1 - q$$

The probability mass function f of this distribution, over possible outcomes x, is

$$f(x;p) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

This can also be expressed as

$$f(x;p) = p^x(1-p)^{1-x}; x \in \{0,1\}$$

Bernoulli Distribution

- Parameter $0 \leq p \leq 1; q = 1 - p$
- Support $k \in \{0,1\}$
- $f(x.p) = \begin{cases} 1 - p; x = 0 \\ p; x = 1 \\ 0; \text{elsewhere} \end{cases}$
- $F(x) = \begin{cases} 0; x < 0 \\ 1 - p; 0 \leq x < 1 \\ 1; x \geq 1 \end{cases}$

Bernoulli Distribution

- $E(X) = p$
- $E(X^2) = p$
- $Var(X) = pq$
- $m_X(t) = q + pe^t$

Binomial experiment

(i) The experiment consists of a sequence of n smaller experiments called *trials,* where n is fixed in advance of the experiment.

(ii) Each trial can result in one of the same two possible outcomes which we generally write as success (S) and failure (F).

(iii) The trials are independent, so that the outcome on any particular trial does not influence outcome on any other trial.

(iv) The probability of success $P(S)$ =p is constant from trial to trial.

Binomial random variable

The **binomial rv *X*** associated with a binomial experiment consisting of n trials is defined as the number of ***X*** Successes among n trials

A discrete random variable

$$X \sim \text{Bin(n, p)}$$

has *binomial* distribution with parameters n and p, n is a positive integer and $0<p<1$, its density function is

$$p(x) = P(X = x) = b(x; n, p)$$

$$\text{p(x)} = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{x=0,1,2,..........n}$$

$$= 0 \qquad \text{otherwise}$$

Because the pmf of a binomial rv X depends on the two parameters n and p, we denote the pmf by $b(x; n, p)$.

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Theorem: Let X be a binomial random variable with parameters n and p. Then show that

mgf of X is: $(q + pe^t)^n, q = 1 - p$.

Also show that $E(X) = np$ and Var $(X)=npq$. Compute these using mgf as well.

Ex 1. Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours ?

Ex 2. The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

Ex3. The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that

(a) at least 10 survive (b) from 3 to 8 survive

(c) exactly 5 survive?

Solution:

Ex 4. A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. Find the prob. that there will be at least one defective item among these 20?

Solution.

(b) Suppose the retailer receives 10 shipments in a month and inspector randomly tests 20 devices per shipment. Find the prob. that there will be exactly 3 shipments each containing at least one defective device among 20 that are selected and tested from the shipment?

Solution:

Ex 5. In a bombing attack, there is a 50% chance that any bomb will strike the target. At least two direct hits are required to destroy the target. How many minimum number of bombs must be dropped so that the prob. of hitting the target at least twice is more than 0.99?

Solution:

Q: In one out of 6 cases, material for bulletproof vests fails to meet puncture standards. If 405 specimen are tested, what does Chebyshev's theorem tell us about the probability of getting at most 30 or more than 105 cases that do not meet puncture standards.

Hypergeometric Distributions
Lecture-10

The Hypergeometric Distribution

Assumptions leading to the hypergeometric distribution are as follows:

1. The population or set to be sampled consists of N individuals, objects or elements.
2. Each individual can be characterized as a success (S) or failure (F) and there are M successes.

3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

The Hypergeometric Distribution

The rv of interest is X = number of S's in the sample.

The probability distribution of X depends on parameters n, M and N. So, $P(X = x) = h(x; n, M, N)$.

Ex1:

During a particular period, a university's information technology office received 20 service orders for problems with printers, of which 8 were laser printers and 12 were inkjet models. A sample of 5 of these service orders is to be selected for inclusion in a customer satisfaction survey.

Suppose that 5 are selected in a completely random fashion, so that any particular subset of size 5 has the same chance of being selected as does any other subset.

Then, what is the probability that exactly

x (x = 0, 1, 2, 3, 4 or 5) of the selected service orders were for inkjet printers?

Ex1:cont-

Here, the population size $N = 20$, sample size $n = 5$

Number of S's (inkjet = S), $M = 12$

Number of F's, $N - M = 8$.

Consider for one value $x = 2$. Because all outcomes are equally likely,

$$P(X = 2) = h(2; 5, 12, 20)$$

$$= \frac{\text{number of outcomes having } X = 2}{\text{number of possible outcomes}}$$

Number of possible outcomes in experiment is number of ways of selecting 5 from 20 objects without regard to order: $\binom{20}{5}$

Ex1:cont-

Number of outcomes having $X = 2$, there are $\binom{12}{2}$ ways of selecting 2 of inkjet orders

and for each such way there are $\binom{8}{3}$ ways of selecting 3 laser orders to fill out the sample.

Hence, from product rule $\binom{12}{2}\binom{8}{3}$ as the number of outcomes with $X = 2$, so

$$h(2;\, 5,\, 12,\, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}} = \frac{77}{323} = .238$$

Observation:

> If sample size n < number of successes in the population (M), then largest possible X value is n.

> However, if $M < n$ (e.g., a sample size of 25 and only 15 successes in the population), then X can be at most M.

> Whenever no. of population failures ($N - M$) > n, smallest possible X value is 0 (all sampled individuals might be failures).

> However, if $N - M < n$, smallest possible X value is $n - (N - M)$.

Thus, possible values of X satisfy the restriction:

$$\max(0, n - (N - M)) \leq x \leq \min(n, M).$$

For instance:

Minimum value of x could be n+M-N instead of 0. To understand this, let N = 30, M = 20 and n = 15.

Minimum value of x is n+M-N = 15+20-30 = 5. There are only N-M = 10 objects without S's (with F's) in the 30 items.

So, a sample of 15 items certainly contains at least 5 objects with S's. So, the rv X takes the values 5, 6, ..., 15.

Notice that, maximum value of x is min(n, M) = min(20, 15) = 15.

Similarly, if we choose n = 25, rv X takes the values 15, 16, 17, 18, 19 and 20. In case, we choose n = 8, rv X takes the values 0, 1, 2,..., 8.

The Hypergeometric Distribution

pmf:

If X is the number of S's in a completely random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the **hypergeometric distribution,** is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

for x, an integer, satisfying

$$\max(0, n - N + M) \le x \le \min(n, M).$$

The Hypergeometric Distribution

The mean and variance of the hypergeometric rv X having pmf $h(x; n, M, N)$ are

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

The ratio M/N is the proportion of S's in the population. If we replace M/N by p in $E(X)$ and $V(X)$, we get

$$E(X) = np$$

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1-p)$$

Ex 2:

During a course of an hour, 1000 bottles of juices are filled by a particular machine. Each hour a sample of 20 bottles is randomly selected and number of ounces of juices per bottle is checked.

Let X : number of bottles selected that are under filled. Suppose during a particular hour, 100 under filled bottles are produced.

Find the Prob. that at least 3 under filled bottles will be among those sampled.

Solution: (using Hypergeometric)

Advantage:

Sometimes population size is large but not known and proportion of favorable population is given. Then we can use Binomial distribution for both sampling with or without replacement where prob. p is the proportion of favorable population.

Ex:

Suppose the population size N is not actually known, so the value x is observed and we wish to estimate N.

It is reasonable to equate the observed sample proportion of S's, x/n, with the population proportion, M/N, giving the estimate

$$\hat{N} = \frac{M \cdot n}{x}$$

If $M = 100$, $n = 40$, and $x = 16$, then $\hat{N} = 250$.

Geometric and Poisson Probability Distribution
Lecture-11

Geometric random variable:

If we perform a series of identical & indep^n trials and X = no. of trials reqd. to get the first success then X is called a g.r.v.

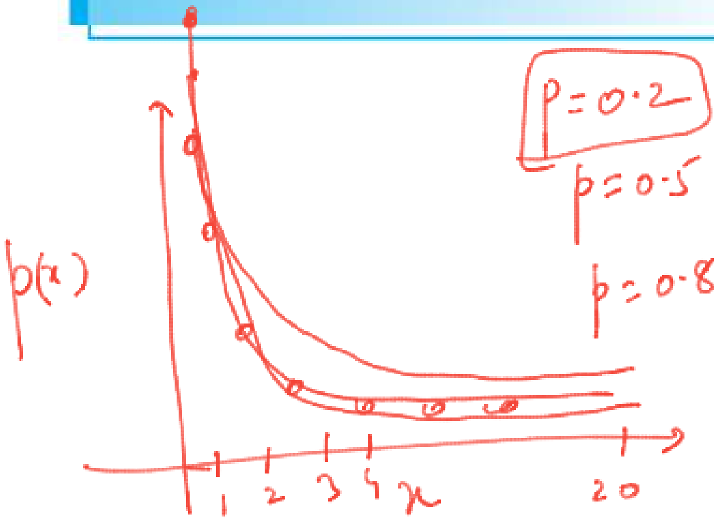
$S = \{S, FS, FFS, \ldots\ldots\}$ * Area of games, statistical quality control.

p = prob. of success on any trial (same)

$$p(x) = \begin{cases} (1-p)^{x-1} p = q^{x-1} p, & x = 1, 2, 3, \ldots \\ 0, & \text{else} \end{cases} \quad \boxed{f(x;p)} \quad 0<p<1$$

→ verify this is a density of a r.v. $\Rightarrow \sum_{x=1}^{\infty} q^{x-1} p = 1$

Geometric random variable:



Results:- let X be a g.r.v. with parameter p. Then

(a) mgf of $X = m_X(t) = \dfrac{pe^t}{1-qe^t}$, $t < -\ln q$

(b) $E(X) = \dfrac{1}{p}$, $\operatorname{var}(X) = q/p^2$

Geometric random variable:

Proof:- $m_X(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} q^{x-1} p$

$= \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x = \frac{p}{q}\left[qe^t + (qe^t)^2 + \cdots\right]$

G-series with $a = qe^t$, c.r. $= qe^t$

when $|qe^t| < 1$

$\Rightarrow qe^t < 1$

$\Rightarrow t < \ln\frac{1}{q} = -\ln q$.

$= \frac{p}{q}\left[\frac{qe^t}{1-qe^t}\right]$

$= \frac{pe^t}{1-qe^t}$ provided $t < -\ln q$

$E(X) = \frac{d}{dt} m_X(t)\Big|_{t=0} \Rightarrow E(X) = \frac{pe^t}{1-qe^t} + \frac{pe^t(-1)(-qe^t)}{(1-qe^t)^2}\Big|_{t=0}$

$= \frac{p}{1-q} + \frac{pq}{(1-q)^2} = 1 + \frac{q}{p} = \frac{1}{p}$ Ans

$E(X^2) = \frac{d^2}{dt^2} m_X(t)\Big|_{t=0} = (1+q)/p^2$; $Var(X) = q/p^2$.

Geometric random variable:

$E(X) = ?$ By defn.

$$E(x) = \sum_{x=1}^{\infty} x q^{x-1} p = \frac{p}{q} \underbrace{\sum_{x=1}^{\infty} x q^{x}}_{[q + 2q^2 + 3q^3 + \cdots]}$$

$$= p \underbrace{[1 + 2q + 3q^2 + \cdots]}$$

$$= p \cdot \frac{1}{p^2} = \frac{1}{p}$$

Ans

$$S = 1 + 2q + 3q^2 + 4q^3 + \cdots$$

$$Sq = \quad q + 2q^2 + 3q^3 + \cdots$$

$$- \quad - \quad - \quad -$$

$$(1-q) S = 1 + q + q^2 + \cdots$$

$$= \frac{1}{1-q} \Rightarrow S = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

Poisson Probability Distribution

Binomial and hypergeometric distributions were derived by starting with an experiment consisting of trials or draws and applying the laws of probability to various outcomes of the experiment. .

There is no simple experiment on which the Poisson distribution is based, though shortly we will describe how it can be obtained by certain limiting operations.

Poisson Probability Distribution

Definition

A discrete random variable X is said to have a **Poisson distribution** with parameter μ ($\mu > 0$) if the pmf of X is

CDF: $F(x) = P(X \le x)$

$$\boxed{p(x;\mu) = \frac{e^{-\mu} \cdot \mu^x}{x!}} \qquad x = 0, 1, 2, 3, \ldots$$

$$= \sum_{y=0}^{x} \frac{e^{-\mu}\mu^y}{y!}$$

$y = 0, 1, \ldots$

Appendix Table A.2 exhibits the cdf $F(x; \mu)$ for $\mu = .1, .2, \ldots, 1, 2, \ldots, 10, 15$, and 20.

For example, if $\mu = 2$, then $P(X \le 3) = F(3; 2) = .857$,

whereas $P(X = 3) = F(3; 2) - F(2; 2) = .180$.

EX. Let X, the number of flaws on the surface of a randomly selected boiler of a certain type, have a Poisson distribution with parameter $\lambda = 5$. Use Appendix Table A.2 to compute the following probabilities:

a. $P(X \le 8)$ b. $P(X = 8)$ c. $P(9 \le X)$
d. $P(5 \le X \le 8)$ e. $P(5 < X < 8)$

$\mu = \lambda = 5$

a) $F(8)$

b) $F(8) - F(7)$

c) $1 - P(X \le 8)$
$1 - F(8)$

d) $F(8) - F(4)$

e) $F(7) - F(5)$

Ex:

Let X denote the number of creatures of a particular type captured in a trap during a given time period.

Suppose that X has a Poisson distribution with μ = 4.5, so on average traps will contain 4.5 creatures.

The probability that a trap contains exactly five creatures is

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = .1708$$

$$\frac{e^{-\mu}\mu^{x}}{x!}$$

The probability that a trap has at most five creatures is

$$P(X \le 5) = \sum_{x=0}^{5} \frac{e^{-4.5}(4.5)^x}{x!} = .7029$$

$$= e^{-4.5}\left[1 + 4.5 + \frac{(4.5)^2}{2!} + \cdots + \frac{(4.5)^5}{5!}\right]$$

Poisson Distribution as a Limit

Poisson Distribution as a Limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

Proposition

Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n, p) \to p(x; \mu)$.

$np \to \mu > 0$

According to this proposition, *in any binomial experiment in which n is large and p is small*, $b(x; n, p) \approx p(x; \mu)$, *where* $\mu = np$. As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$.

$p < \frac{5}{50}$ $.1$

Ex:

If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is .005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors? At most three pages with errors?

$np = 2$ $\quad n \to \infty$ $\quad p \to 0$ $\quad np = \mu$

Solution: Let *S* denotes a page containing at least one error

F an error-free page

X = *number* of pages containing at least one error is a binomial rv with $n = 400$ and $p = .005$, so $np = 2$.

Ex:

cont'd

$P(x=1)$

$e^{-2}2^{1}/1!$

So,

$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = .270671$

x

The binomial value is $b(1; 400, .005) = .270669$, so the approximation is very good.

$= {}^{400}C_1 (.005)^1 (.995)^{399}$

Similarly,

$\mu = 2$, $F(3) = ?$

$$P(X \leq 3) \approx \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} e^{-2} \frac{2^x}{x!}$$

$= .135335 + .270671 + .270671 + .180447$

$= .8571$

and this again is quite close to the binomial value

$P(X \leq 3) = .8576$

$.0005$ = error

Poisson Distribution as a Limit

Table shows the Poisson distribution for $\mu = 3$ along with three binomial distributions with $np = 3$

n is Large

np=3

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391	0.047553	0.049041	0.049787
1	0.141304	0.147070	0.148609	0.149361
2	0.227656	0.225153	0.224414	0.224042
3	0.236088	0.227474	0.225170	0.224042
4	0.177066	0.170606	0.168877	0.168031
5	0.102305	0.101308	0.100985	0.100819
6	0.047363	0.049610	0.050153	0.050409
7	0.018043	0.020604	0.021277	0.021604
8	0.005764	0.007408	0.007871	0.008102
9	0.001565	0.002342	0.002580	0.002701
10	0.000365	0.000659	0.000758	0.000810

Comparing the Poisson and Three Binomial Distributions

Mean, Variance, MGF

Proposition: If X has a Poisson distribution with parameter μ, then

(a) $E(X) = V(X) = \mu$.

(b) $m_X(t) = e^{\mu(e^t-1)}$

$$\sum_{x=0}^{\infty} \frac{e^{-\mu}\mu^x}{x!} = 1$$

Proof: $m_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\mu}\mu^x}{x!}$

$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{-\mu}\left[1 + \frac{\mu e^t}{1} + \cdots\right]$$

$$= e^{-\mu} e^{\mu e^t} = e^{\mu(e^t-1)}$$

$$E(X) = \frac{d}{dt} m_X(t)\Big|_{t=0} = \underline{e^{\mu(e^t-1)} \mu e^t}\Big|_{t=0} = \mu$$

$$E(X^2) = \mu^2 + \mu, \quad var(X) = \mu^2 + \mu - \mu^2 = \mu$$

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\mu}\mu^{x}}{x!} \qquad x-1=y$$

$$= \sum_{x=1}^{\infty} \frac{e^{-\mu}\mu^{x}}{(x-1)!} = e^{-\mu}\left[\sum_{y=0}^{\infty} \frac{\mu^{\cancel{y}+1}}{y!}\right]$$

$$= e^{-\mu}\mu \sum_{y=0}^{\infty} \frac{\mu^{y}}{y!} = \cancel{e^{-\mu}}\,\mu\,\cancel{e^{\mu}}$$

$$= \mu \quad \underline{\text{Ans}}$$

$$\boxed{\begin{array}{l} X \qquad Y \\ m_X(t) = m_Y(t) \Rightarrow X \ \& \ Y \text{ are same.} \end{array}}$$

$$X \to e^{\mu(e^{t}-1)} \leftarrow e^{3(e^{t}-1)} \to X(\mu=3)$$

Mean and Variance

Since $b(x; n, p) \to p(x; \mu)$ as $n \to \infty$, $p \to 0$, $np \to \mu$,

mean and variance of a binomial variable should approach those of a Poisson variable.

These limits are $np \to \mu$ and $np(1 - p) \to \mu$.

$b(x; n, p)$; np = Mean $\to \mu$

npq = Var $\to \mu(1-p) = \mu$

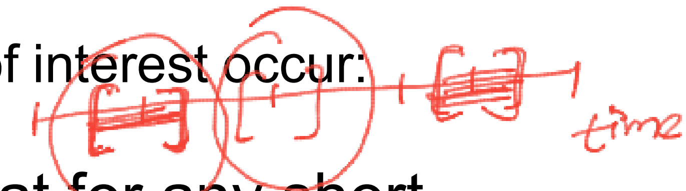
Poisson Process

A very important application of the Poisson distribution arises in connection with the occurrence of events of some type over time.

Events of interest might be visits to a particular website, email messages sent to a particular address, accidents in an industrial facility, or cosmic ray showers observed by astronomers at a particular observatory.

Poisson Process

Assumptions: about the way in which the events of interest occur:



1. There exists a parameter $\alpha > 0$ such that for any short time interval of length Δt, the probability that exactly one event occurs is $\alpha \cdot \Delta t + o(\Delta t)$*

$\frac{o(\Delta t)}{\Delta t} \to 0$ as $\Delta t \to 0$

> for a short interval of time, probability of a single event occurring is approximately proportional to the length of the time interval, where α is the constant of proportionality.

$o(\Delta t)$ decays to zero faster than Δt.

2. The probability of more than one event occurring during Δt is $o(\Delta t)$ [which, along with Assumption 1, implies that the probability of no events during Δt is $1 - \alpha \cdot \Delta t - o(\Delta t)$]

3. The number of events occurring during the time interval Δt is independent of the number that occur prior to this time interval.

The Poisson Process

Proposition

Let $P_k(t)$ denote the probability that k events will be observed during any particular time interval of length t.

$$e^{-\alpha t}(\alpha t)^K / K!$$

$$\Delta t = t$$

$$P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k / k!$$

So that the number of events during a time interval of length t is a Poisson rv with parameter $\mu = \alpha t$.

$$\sum_{K=0}^{\infty} \frac{e^{-\alpha t}(\alpha t)^K}{K!} = 1$$

The expected number of events during any such time interval is then αt, so the expected number during a unit interval of time is α.

The occurrence of events over time as described is called a *Poisson process*; the parameter α specifies the *rate* for the process.

Ex:

Suppose pulses arrive at a counter at an average rate of six per minute, so that $\alpha = 6$.

To find the probability that in a .5-min interval at least one pulse is received, note that the number of pulses in such an interval has a Poisson distribution with parameter $\alpha t = 6(.5) = 3$ (.5 min is used because α is expressed as a rate per minute).

Then with X = the number of pulses received in the 30-sec interval,

$\mu = 3$

$$P(1 \le X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = .950$$

Ans