STATISTICAL METHODS FOR ENGINEERING
Discrete random variables

Roberto Fontana
Department of Mathematical Sciences
Politecnico di Torino
roberto.fontana@polito.it

Data Spaces, October 2018 – January 2019

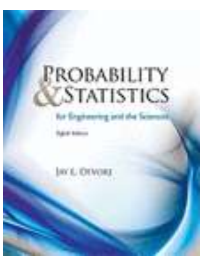
These slides are mainly based on the supporting material of Devore J., Probability and Statistics for Engineering and the Sciences (8th edition)

I also thank Professors Gianfranco Genta, Franco Pellerey, Suela Ruffa and Marina Santacroce for having provided me with useful material for the preparation of these notes

WARNING: The slides only contain brief reminders of topics; PLEASE REFER TO THE BOOK FOR STUDYING!

Textbook

- Devore J., Probability and Statistics for Engineering and the Sciences (8th edition)



Bernoulli Distribution

An experiment consists of one trial. It can result in one of 2 outcomes: Success or Failure (or a characteristic being Present or Absent).

Probability of Success is p $(0<p<1)$

Y = 1 if Success (Characteristic Present), 0 if not

$$p(y)=\begin{cases} p & y=1 \\ 1-p & y=0 \end{cases}$$

$$E(Y)=\sum_{y=0}^{1} y p(y) = 0(1-p)+1p=p$$

$$E\left(Y^2\right)=0^2(1-p)+1^2 p=p$$

$$\Rightarrow V(Y)=E\left(Y^2\right)-[E(Y)]^2=p-p^2=p(1-p)$$

$$\Rightarrow \sigma=\sqrt{p(1-p)}$$

Binomial Experiment

- Experiment consists of a series of n identical trials
- Each trial can end in one of 2 outcomes: Success or Failure
- Trials are independent (outcome of one has no bearing on outcomes of others)
- Probability of Success, p, is constant for all trials
- Random Variable Y, the number of Successes in the n trials, is said to follow Binomial Distribution with parameters n and p
- Y can take on the values $y=0,1,\ldots,n$
- Notation: $Y \sim Bin(n,p)$

Binomial Random Variable

- A sequence of n independent Bernoulli trials (Y_i) are performed, where the probability of success on each trial is p
- X is the number of successes: $X = Y_1+Y_2+\ldots+Y_n$

Then for $i = 0, 1, \ldots, n$,

$$p(i) = P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

where

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

X is a binomial random variable with parameters n and p.

Binomial Distribution

Consider outcomes of an experiment with 3 Trials:

$SSS \Rightarrow x=3 \quad P(SSS) = P(X=3) = p(3) = p^3$

$SSF, SFS, FSS \Rightarrow x=2 \quad P(SSF \cup SFS \cup FSS) = P(X=2) = p(2) = 3p^2(1-p)$

$SFF, FSF, FFS \Rightarrow x=1 \quad P(SFF \cup FSF \cup FFS) = P(X=1) = p(1) = 3p(1-p)^2$

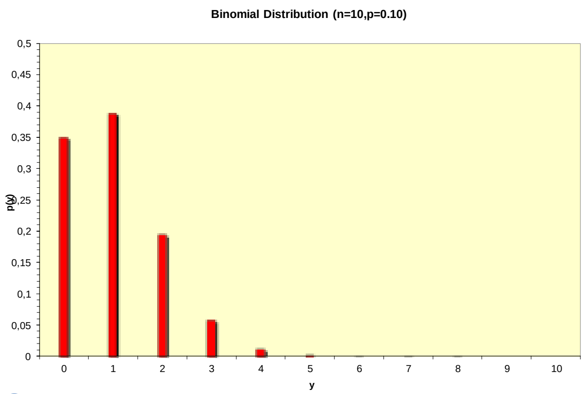
$FFF \Rightarrow x=0 \quad P(FFF) = P(X=0) = p(0) = (1-p)^3$

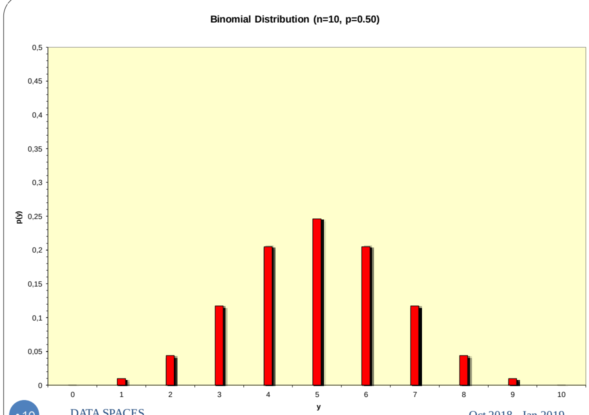
In General:

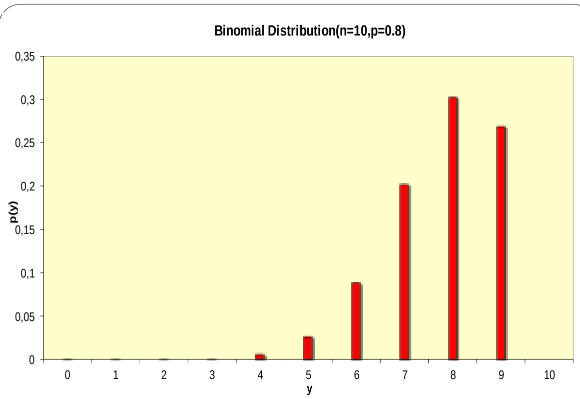
1) # of ways of arranging x S^s (and $(n-x)$ F^s) in a sequence of n positions $\equiv \binom{n}{x} = \frac{n!}{x!(n-x)!}$

2) Probability of each arrangement of x S^s (and $(n-x)$ F^s) $\equiv p^x(1-p)^{n-x}$

3) $\Rightarrow P(X=x) = p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,\ldots,n$







Binomial Distribution – Expected Value and Variance

$$\begin{aligned} E[X] &= E[Y_1 + Y_2 + \ldots + Y_n] \\ &= E[Y_1] + E[Y_2] + \ldots + E[Y_n] = np. \end{aligned}$$

- $V(X) = E\left(X^2\right) - \left[E(X)\right]^2 = n^2p^2 + np(1-p) - (np)^2 = np(1-p)$
- $\sigma = \sqrt{np(1-p)}$

Example

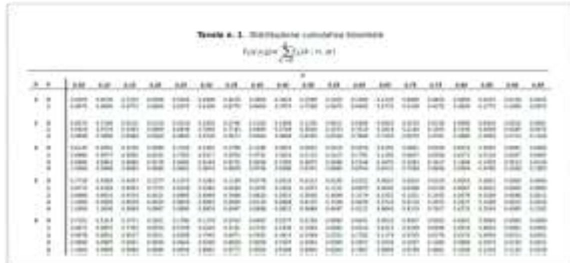
Suppose that 20% of all copies of a particular textbook fail a certain binding strength test. Let X denote the number among 15 randomly selected copies that fail the test.

Then X has a binomial distribution with n=15 and p=0.2

1. The probability that at most 8 fail the test is…
2. The probability that exactly 8 fail is…
3. The probability that at least 8 fail is…
4. The probability that between 4 and 7, inclusive, fail is…

→ cdf table

Table of the cdf of binomial rv's



→available at the course page on the TTPU portal

n=15

		p=0.05	p=0.10	p=0.15	p=0.20	p=0.25	p=0.30
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Ex. 49 p. 120

A company that produces fine crystal knows from experience that 10% of its goblets have cosmetic flaws and must be classified as "seconds."

a. Among six randomly selected goblets, how likely is it that only one is a second?

b. Among six randomly selected goblets, what is the probability that at least two are seconds?

c. If goblets are examined one by one, what is the probability that at most five must be selected to find four that are not seconds?

Ex. 55 p. 121

Twenty percent of all telephones of a certain type are submitted for service while under warranty. Of these, 60% can be repaired, whereas the other 40% must be replaced with new units. If a company purchases ten of these telephones, what is the probability that exactly two will end up being replaced under warranty?

Ex. 59, p. 121

An ordinance requiring that a smoke detector be installed in all previously constructed houses has been in effect in a particular city for 1 year. The fire department is concerned that many houses remain without detectors. Let p=the true proportion of such houses having detectors, and suppose that a random sample of 25 homes is inspected. If the sample strongly indicates that fewer than 80% of all houses have a detector, the fire department will campaign for a mandatory inspection program. Because of the costliness of the program, the department prefers not to call for such inspections unless sample evidence strongly argues for their necessity. Let X denote the number of homes with detectors among the 25 sampled. Consider rejecting the claim that $p \geq 0,8$ if $X \leq 15$.

Ex. 59, p. 121

a. What is the probability that the claim is rejected when the actual value of p is 0.8?

b. What is the probability of not rejecting the claim when $p = 0.7$. When $p=0.6$?

c. How do the "error probabilities" of parts (a) and (b) change if the value 15 in the decision rule is replaced by 14?

Poisson Distribution

Distribution often used to model the number of occurrences of some characteristic in time or space:

- Arrivals of customers in a queue
- Numbers of flaws in a roll of fabric
- Number of typos per page of text.

Distribution obtained as follows:

- Break down the "*area*" into many small "pieces" (n pieces)
- Each "piece" can have only 0 or 1 occurrences (p=P(1))
- Let $\lambda = np \equiv$ Average number of occurrences over "area"
- $Y \equiv$ # occurrences in "area" is sum of 0^s & 1^s over "pieces"
- $Y \sim \text{Bin}(n,p)$ with $p = \lambda/n$
- Take limit of Binomial Distribution as $n \to \infty$ with $p = \lambda/n$

Poisson Distribution

$$p(y)=\frac{n!}{y!(n-y)!}p^{y}(1-p)^{n-y}=\frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}$$

Taking limit as $n\to\infty$:

$$\lim_{n\to\infty}p(y)=\lim_{n\to\infty}\frac{n!}{y!(n-y)!}\left(\frac{\lambda}{n}\right)^{y}\left(1-\frac{\lambda}{n}\right)^{n-y}=\frac{\lambda^{y}}{y!}\lim_{n\to\infty}\frac{n(n-1)...(n-y+1)(n-y)!}{n^{y}(n-y)!}\left(1-\frac{\lambda}{n}\right)^{n}\left(\frac{n-\lambda}{n}\right)^{-y}=$$

$$=\frac{\lambda^{y}}{y!}\lim_{n\to\infty}\frac{n(n-1)...(n-y+1)}{(n-\lambda)^{y}}\left(1-\frac{\lambda}{n}\right)^{n}=\frac{\lambda^{y}}{y!}\lim_{n\to\infty}\left(\frac{n}{n-\lambda}\right)\left(\frac{n-1}{n-\lambda}\right)...\left(\frac{n-y+1}{n-\lambda}\right)\left(1-\frac{\lambda}{n}\right)^{n}$$

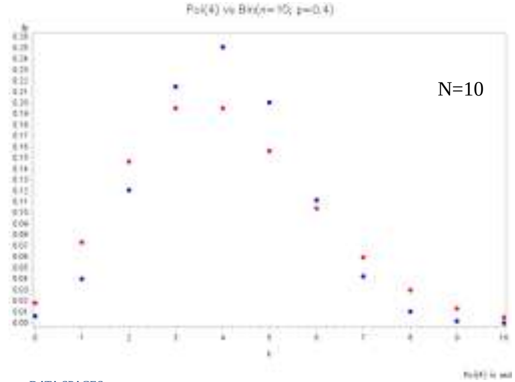
Note: $\lim_{n\to\infty}\left(\frac{n}{n-\lambda}\right)=...=\lim_{n\to\infty}\left(\frac{n-y+1}{n-\lambda}\right)=1$ for all fixed y

$$\Rightarrow\lim_{n\to\infty}p(y)=\frac{\lambda^{y}}{y!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}$$

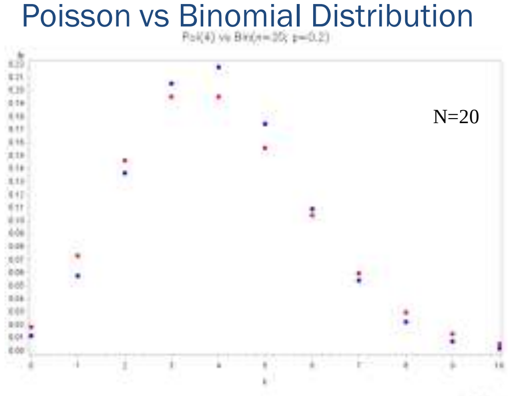
From Calculus, we get: $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^{n}=e^{a}$

$$\Rightarrow\lim_{n\to\infty}p(y)=\frac{\lambda^{y}}{y!}e^{-\lambda}=\frac{e^{-\lambda}\lambda^{y}}{y!}\quad y=0,1,2,...$$

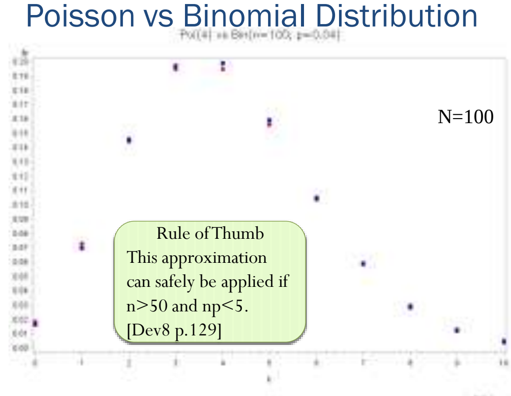
Poisson vs Binomial Distribution



Poisson vs Binomial Distribution



Poisson vs Binomial Distribution



Rule of Thumb
This approximation can safely be applied if n>50 and np<5.
[Dev8 p.129]

Poisson Distribution - Expectations

$$f(y)=\frac{e^{-\lambda}\lambda^{y}}{y!}\qquad y=0,1,2,\ldots$$

$$E(Y)=\sum_{y=0}^{\infty}y\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=1}^{\infty}y\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=1}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{(y-1)!}=\lambda e^{-\lambda}\sum_{y=1}^{\infty}\frac{\lambda^{y-1}}{(y-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$$E\big(Y(Y-1)\big)=\sum_{y=0}^{\infty}y(y-1)\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=2}^{\infty}y(y-1)\left[\frac{e^{-\lambda}\lambda^{y}}{y!}\right]=\sum_{y=2}^{\infty}\frac{e^{-\lambda}\lambda^{y}}{(y-2)!}=$$

$$=\lambda^{2}e^{-\lambda}\sum_{y=2}^{\infty}\frac{\lambda^{y-2}}{(y-2)!}=\lambda^{2}e^{-\lambda}e^{\lambda}=\lambda^{2}$$

$$\Rightarrow E\left(Y^{2}\right)=E\big(Y(Y-1)\big)+E(Y)=\lambda^{2}+\lambda$$

$$\Rightarrow V(Y)=E\left(Y^{2}\right)-[E(Y)]^{2}=\lambda^{2}+\lambda-[\lambda]^{2}=\lambda$$

$$\Rightarrow \sigma=\sqrt{\lambda}$$

Hypergeometric Distribution

Finite population generalization of Binomial Distribution

Population:

- N Elements
- k Successes (elements with characteristic if interest, M in place of k in the textbook)

Sample:

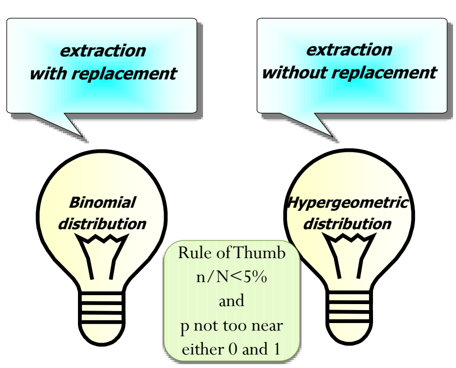
- n Elements
- X = # of Successes in sample ($x = 0,1,\ldots,\min(n,k)$)

$$p(x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}\qquad x=0,1,\ldots,\min(n,k)$$

It should be also $x \geq \max(0, n-N+k)$

$$E(X)=n\left(\frac{k}{N}\right)$$

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



Ex. 85 p. 135

Suppose small aircraft arrive at a certain airport according to a Poisson process with rate α=8 per hour, so that the number of arrivals during a time period of t hours is a Poisson rv with parameter μ=8 t.

a. What is the probability that exactly 6 small aircraft arrive during a 1-hour period? At least 6? At least 10?

b. What are the expected value and standard deviation of the number of small aircraft that arrive during a 90-min period?

c. What is the probability that at least 20 small aircraft arrive during a 2.5-hour period? That at most 10 arrive during this period?

Ex. 97 p. 133

Of all customers purchasing automatic garage-door openers, 75% purchase a chain-driven model. Let X = the number among the next 15 purchasers who select the chain-driven model.

a. What is the pmf of *X*?

b. Compute P(X>10)

c. Compute P(6≤X ≤ 10)

d. Compute μ and σ^2.

e. If the store currently has in stock 10 chain-driven models and 8 shaft-driven models, what is the probability that the requests of these 15 customers can all be met from existing stock?

The Negative Binomial Distribution

The Negative Binomial Distribution

- The negative binomial rv and distribution are based on an experiment satisfying the following conditions:
- **1.** The experiment consists of a sequence of independent trials.

  2. Each trial can result in either a success (*S*) or a failure (*F*).

  3. The probability of success is constant from trial to trial, so for *i* = 1, 2, 3, ...

The Negative Binomial Distribution

- **4.** The experiment continues (trials are performed) until a total of *r* successes have been observed, where *r* is a specified positive integer.

- The random variable of interest is *X* = the number of failures that precede the *r*th success; *X* is called a **negative binomial random variable** because, in contrast to the binomial rv, the number of successes is fixed and the number of trials is random.

The Negative Binomial Distribution

- Possible values of X are 0, 1, 2, ... Let $nb(x; r, p)$ denote the pmf of X. Consider $nb(7; 3, p) = P(X = 7)$, the probability that exactly 7 F's occur before the 3rd S.
- In order for this to happen, the 10th trial must be an S and there must be exactly 2 S's among the first 9 trials. Thus

$$nb(7; 3, p) = \left\{\binom{9}{2} \cdot p^2(1-p)^7\right\} \cdot p = \binom{9}{2} \cdot p^3(1-p)^7$$

- Generalizing this line of reasoning gives the following formula for the negative binomial pmf.

The Negative Binomial Distribution

- **Proposition**
- The pmf of the negative binomial rv X with parameters r = numberof S's and $p = P(S)$ is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r(1-p)^x \quad x = 0, 1, 2, \ldots$$

Example 38

- A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let $p = P$(a randomly selected couple agrees to participate).
- If $p = .2$, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with S = {agrees to participate}, what is the probability that 10 F's occur before the fifth S?
- Substituting $r = 5$, $p = .2$, and $x = 10$ into $nb(x; r, p)$ gives

$$nb(10; 5, .2) = \binom{14}{4}(.2)^5(.8)^{10} = .034$$

Example 38 cont'd

- The probability that at most 10 F's are observed (at most 15 couples are asked) is

$$P(X \le 10) = \sum_{x=0}^{10} nb(x; 5, .2)$$

$$= (.2)^5 \sum_{x=0}^{10} \binom{x+4}{4} (.8)^x$$

$$= .164$$

The Negative Binomial Distribution

- In some sources, the negative binomial rv is taken to be the number of trials $X + r$ rather than the number of failures.
- In the special case $r = 1$, the pmf is
-

$$nb(x; 1, p) = (1 - p)^x p \quad x = 0, 1, 2, \ldots \qquad \textbf{(3.17)}$$

- In earlier Example, we derived the pmf for the number of trials necessary to obtain the first S, and the pmf there is similar to Expression (3.17).

The Negative Binomial Distribution

- Both X = number of F's and Y = number of trials ($= 1 + X$) are referred to in the literature as **geometric random variables,** and the pmf in Expression (3.17) is called the **geometric distribution.**
- The expected number of trials until the first S was shown earlier to be $1/p$, so that the expected number of F's until the first S is $(1/p) - 1 = (1 - p)/p$.
- Intuitively, we would expect to see $r \cdot (1 - p)/p$ F's before the rth S, and this is indeed $E(X)$. There is also a simple formula for $V(X)$.

The Negative Binomial Distribution

- **Proposition**
- If X is a negative binomial rv with pmf $nb(x; r, p)$, then

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

- Finally, by expanding the binomial coefficient in front of $p^r(1 - p)^x$ and doing some cancellation, it can be seen that $nb(x; r, p)$ is well defined even when r is not an integer. This *generalized negative binomial distribution* has been found to fit observed data quite well in a wide variety of applications.